TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

Impacts of Climatic Oscillations on Precipitation in an extended Mediterranean area

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Geodäsie und Geoinformation

eingereicht von

Laura Crocetti
Matrikelnummer 01226985

ausgeführt am Department für Geodäsie und Geoinformation
der Fakultät für Mathematik und Geoinformation der Technischen Universität Wien

Betreuung
Betreuer: Univ.Prof. Dr.rer.nat. Wouter Arnoud Dorigo MSc

Wien, 04.04.2019

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

Abstract

Oceanic-atmospheric oscillation patterns, described by so-called climate modes, have a strong impact on the variability in the terrestrial water cycle. However, the relation between climatic oscillations and hydrology is not yet fully understood due to uncertainties in the observations and the co-varying behavior of multiple oscillations. A better knowledge about these connections is needed to provide better predictions about climate and hydrology.

In this study, the impact of 17 major climate modes on monthly precipitation anomalies in an extended Mediterranean area between 28.5°N - 56.5°N and 10°W - 46°E is analyzed. The climate modes are expressed through their corresponding Climate Oscillation Index (COI), used to describe the state of the atmospheric-oceanic circulations. A supervised learning approach, called least absolute shrinkage and selection operator (LASSO) regression is used to quantify the influence of these teleconnection patterns (e.g., North Atlantic Oscillation, East Atlantic West Russia Pattern) on precipitation anomalies. Precipitation is an important component of the hydrological cycle and one of the most dominant climatic drivers for water availability besides potential evaporation.

The LASSO regression is a data-driven method that uses automatic feature selection and regularization, which in this study is used, to identify oceanic-atmospheric controls on precipitation anomalies and to disentangle the impact of individual climate modes. The methodology considers cross-correlations in the features, i.e. Climate Oscillation Indices. Time lags ranging between zero and five months are introduced in every feature to account for potential lagged response of precipitation anomalies to ocean-atmospheric oscillations. The LASSO model is fitted for each grid point in two ways. Once, by only using the time series of the grid point and additionally by adding the information of the eight neighboring grid points. Besides using all months of the year to build the model, the analysis is also performed for each season separately. Both of these steps increase the coefficient of determination R^2 derived from the LASSO regression and therefore improve the predictive performance of the LASSO model. For validation of the regression models two cross-validations and a significance test using the Benjamini-Hochberg procedure are applied.

The results gained by the LASSO regression show that in specific hot spot regions up to 70% of the precipitation anomalies can be explained by the modes of climate variability. Adding the information of the neighborhood into the model increases the explained variance R^2 significantly. Analyzing the influence of each COI shows that the signal of the East Atlantic Pattern (EA), East Atlantic West Russia Pattern (EAWR), Northern Annular Mode (NAM), and North Atlantic Oscillation (NAO) have a significant impact in the western parts of the investigated area during wintertime. These results help to improve the general understanding of how the individual climate modes affect different parts of the extended Mediterranean area.

Kurzfassung

Ozeanisch-atmosphärische Oszillationen, welche durch sogenannte Klimamodi beschrieben werden, haben einen starken Einfluss auf die Variabilität des terrestrischen Wasserkreislaufs. Der Zusammenhang zwischen diesen Oszillationen und der Hydrologie ist jedoch aufgrund von Unsicherheiten in den Beobachtungen und des veränderlichen Verhalten von mehreren Oszillationen noch nicht vollständig verstanden. Bessere Kenntnisse sind erforderlich, um bessere Vorhersagen über Klima und Hydrologie liefern zu können.

In dieser Studie wird der Einfluss von 17 wesentlichen Klimaoszillationen auf monatliche Niederschlagsanomalien in einem erweiterten Mittelmeerraum zwischen 28.5°N - 56.5°N und 10°W - 46°O analysiert. Die Klimaoszillationen werden durch ihren entsprechenden Klimaoszillationsindex, der zur Beschreibung des Zustands der atmosphärisch-ozeanischen Zirkulationen verwendet wird, ausgedrückt. Ein überwachtes maschinelles Lernverfahren, namens least absolute shrinkage and selection operator (LASSO) Regression, wird verwendet, um den Einfluss von diesen Klimaoszillationen (wie zum Beispiel: North Atlantic Oscillation, East Atlantic West Russia Pattern) auf Niederschlagsanomalien zu bestimmen. Niederschlag ist eine wichtige Komponente des Wasserkreislaufs und neben der Verdunstung einer der wichtigsten klimatischen Treiber für die Wasserverfügbarkeit.

Die LASSO Regression ist eine datengetriebene Methode, welche eine automatische Merkmalsauswahl und Regularisierung verwendet. Dies wird in dieser Studie genutzt, um die ozeanisch-atmosphärischen Oszillationen auf Niederschlagsanomalien zu identifizieren, und den Einfluss individueller Klimaoszillationen zu entwirren. Die Methodik berücksichtigt Kreuzkorrelationen zwischen Klimaoszillationsindizes. Für jedes Feature werden Zeitverzögerungen zwischen null und fünf Monaten eingeführt, um die eventuell verzögerte Reaktion der Niederschlagsanomalien auf die ozeanisch-atmosphärischen Oszillationen zu berücksichtigen. Das LASSO Modell ist auf zwei Arten für jeden Gitterpunkt trainiert: einmal nur durch Verwendung der Zeitserie des entsprechenden Gitterpunkts, und zusätzlich durch Hinzufügen der Informationen der acht benachbarten Gitterpunkte. Die Analyse wird sowohl für alle Monate eines Jahres gleichzeitig als auch für jede Jahreszeit separat durchgeführt. Diese beiden Schritte erhöhen den aus der LASSO Regression abgeleiteten Bestimmungskoeffizienten R^2 und verbessern somit die Vorhersageleistung des LASSO Modells. Zur Validierung der LASSO Modelle werden zwei Kreuzvalidierungen und ein Signifikanztest mit Benjamini-Hochberg Prozedur eingesetzt.

Die Ergebnisse der LASSO Regression zeigen, dass in speziellen Hotspot Regionen bis zu 70% der Niederschlagsanomalien durch Klimaoszillationen erklärt werden können. Durch das Hinzufügen der Informationen der Nachbarschaft zum Modell erhöht sich die erklärte Varianz R^2 signifikant. Wird der Einfluss von jedem Klimaoszillationsindex analysiert, zeigt sich, dass East Atlantic Pattern (EA), East Atlantic West Russia Pattern (EAWR), Northern Annular Mode (NAM), und North Atlantic Oscillation (NAO) während der Winterzeit signifikante Auswirkungen in Teilen des westlichen Untersuchungsgebiets haben. Diese Ergebnisse tragen dazu bei, das allgemeine Verständnis dafür zu verbessern, wie sich die einzelnen Klimaoszillationen auf verschiedene Teile des erweiterten Mittelmeerraums auswirken.

Acknowledgements

First of all, I would like to sincerely thank my supervisor, Prof. Wouter Dorigo, without whom this thesis would not have been possible. His excellent ideas, the inspiring discussions, and his continuous support led to the final result. I am very grateful to have had the chance to work in his research group, the Climate and Environmental Remote Sensing (CLIMERS) Research Group at the Department of Geodesy and Geoinformation at TU Wien. A big thank you to my colleagues, who were always ready to help me with any difficulties.

Moreover, I want to express my gratitude to all my friends, who have always been there for me. I have always enjoyed spending time with them, laughing together and talking hours about everything and anything. I am also very happy to have found good friends at university who accompanied me over the years. However, the biggest thanks go to my family, including my boyfriend Matthias. My grandmother, parents, and sisters have always had an open ear and because they know me best, they had just the right words in every situation. Without the great support of my family, I would not be at the point where I am now. And last but not least, I cannot thank Matthias enough for his endless support, for the many hours discussing all kind of ideas, concerns, and dreams, for making everyday life better, and for just being such an amazing person.

Thank you for making it all possible!

Contents

List of Figures

List of Tables

List of abbreviations

AMM Atlantic Meridional Mode

AMMSST AMM sea-surface temperature

AMO Atlantic Multidecadal Oscillation

AMOC Atlantic meridional overturning circulation

COI Climate Oscillation Index

CRU Climatic Research Unit

DMI Dipole Mode Index

EA East Atlantic Pattern

EAWR East Atlantic West Russia Pattern

ENSO El Niño-Southern Oscillation

EPNP East Pacific North Pacific Pattern

ERSST Extended Reconstructed Sea Surface Temperature

ESRL Earth System Research Laboratory

IOD Indian Ocean Dipole

JAMSTEC Japan Agency for Marine-Earth Science and Technology

KNMI Royal Meteorological Institute of the Netherlands

LASSO least absolute shrinkage and selection operator

NAM Northern Annular Mode

NAO North Atlantic Oscillation

NCEP National Centers for Environmental Prediction

NOAA National Oceanic and Atmospheric Administration

PDO Pacific Decadal Oscillation

PEA Polar Eurasia Pattern

PNA Pacific-North American Pattern

SAM Southern Annular Mode

SCAND Scandinavian Pattern

SOI Southern Oscillation Index

SST Sea Surface Temperature

TNA Tropical Northern Atlantic Dipole

TSA Tropical Southern Atlantic Dipole

WP West Pacific Pattern

1 Introduction

1.1 Climate of the Mediterranean area

The climate of the Mediterranean basin, which stretches over 3800 km west-east and 1000 km north-south, between 30°N and 45°N [Blondel and Aronson, 1999], is of high interest to environmental scientists and has been the focus of many research studies. The attraction for this region can be explained by the important social-economical-environmental impacts of climate variability and change, and the forecasted global warming and dryness, that are likely to greatly affect the environment and human activities [Lionello, 2012; Field et al., 2014].

According to the most frequently used Köppen-Geiger climate classification [Köppen, 1900; Kottek et al., 2006], the Mediterranean climate is defined by a warm temperate climate with hot and dry summers. The Mediterranean region is characterized by the Mediterranean Sea, an almost completely closed sea, serving as a large reservoir of heat and moisture for the surrounding land areas. The geographical location of the Mediterranean Sea is what makes it particularly unique. It is located in a transitional band between humid and dry areas. This transition zone between subtropical and mid-latitude systems leads to large climate variability and a strong seasonal variability of precipitation in many areas [Lionello et al., 2012]. The Mediterranean Sea is an almost isolated oceanic system, thus many processes that are fundamental to the global ocean circulation are also taking place within the Mediterranean basin [Robinson et al., 2001]. Strong air-sea interactions occurring due to mesoscale processes and wind regimes considerably affect the thermohaline circulation of the Mediterranean Sea [Lionello et al., 2012]. Even small modifications in the general circulation, e.g. shifts in the location of mid-latitude storm tracks or sub-tropical high pressure cells, can have a big impact on the climate and natural environment of the Mediterranean [Giorgi and Lionello, 2008]. This makes the region very vulnerable to climatic changes. Giorgi [2006] refers to the Mediterranean as one of the most responsive regions to global change, and as a climate change hot spot with the trend to a hotter and drier climate.

A dominant role for Earth's climate plays the Atlantic meridional overturning circulation (AMOC) [Cheng et al., 2013; Pohlmann et al., 2006; Sutton, 2005]. The AMOC is one of Earth's major ocean circulation systems in the North Atlantic, redistributing heat on our planet and thereby affecting climate all over the world [Caesar et al., 2018]. Caesar et al. [2018] show that in recent years the AMOC has weakened and that it is very likely that the decline since the 1950s is largely anthropogenic due to rising CO_2 levels. A weakening of the AMOC, which may already have an impact on weather in

Europe, is accompanied by a northward shift of the Gulf Stream [Caesar et al., 2018]. The Gulf Stream is highly relevant to the European climate. Its variability manifests in changes in European temperature, precipitation, and storminess [Palter, 2015].

The assessments of projected climate change show with high confidence that temperatures will increase over the Mediterranean region and that it is likely that Mediterranean summer mean precipitation decreases, while annual mean precipitation in Northern and Central Europe will very likely increase [Stocker et al., 2013]. Hence, it is expected that the hydrological cycle in the Mediterranean basin, which is sensitive to changes in temperature and precipitation, will be affected by global warming. The expected future changes in the Mediterranean water budget will have important consequences in the terrestrial and marine environment in the Mediterranean region [Sanchez-Gomez et al., 2009]. Major concerns are the increasing frequency of extreme climate events, like heat waves or floods and a significant reduction of water availability [Barros et al., 2014].

Therefore, it is important to improve the knowledge of the current Mediterranean hydrological cycle and its interactions and processes to assess the potential implications of climate change on its future climate.

1.2 Climate variability in the Mediterranean area

A large portion of the natural variability of climate is controlled by the coupled state of ocean and atmosphere, the so-called oceanic-atmospheric oscillations. The climate system largely depends on large-scale phenomena that occur more or less regularly, which are called modes of climate variability [de Viron et al., 2013]. These climate modes characterized by a repeating time-space pattern are important drivers of the hydrology and can even influence regions that are far apart. Therefore, they are known as teleconnections [Wallace and Gutzler, 1981; Wang and Schimel, 2003]. One of the most relevant teleconnections affecting Mediterranean hydrology is the link between normalized sea-level pressure at Lisbon, Portugal and Reykjavik, Iceland which defines the well-known North Atlantic Oscillation (NAO). However, not only the NAO influences the Mediterranean climate but also other large-scale phenomena like the East Atlantic Pattern (EA), the East Atlantic West Russia Pattern (EAWR) or the Scandinavian Pattern (SCAND).

Mikhailova and Yurovsky [2016] have analyzed the mechanism of the EA and its impact on the atmospheric circulation patterns, as well as the surface air temperature and precipitation in the European region in winter. They found that $15 - 25\%$ of the precipitation interannual variability is explained by the EA. Chronis et al. [2011] point

out the importance of the summer NAO over the eastern Mediterranean. They detected that higher NAO summer values are related to the relative cooling of the surface, and lower troposphere, respectively, which leads to an enhanced meridional circulation and cloudiness over Greece, the Anatolian Plateau, and the surrounding waters. The study of Bladé et al. [2012] has focused on the impacts of the summer NAO on precipitation in Europe. By computing linear correlations and regressions between the summer NAO and precipitation they found that the influence of summer NAO is characterized by a strong north-south dipole in precipitation between north-west Europe and the Mediterranean. These findings are consistent with Mariotti and Arkin [2007]. Bueh and Nakamura [2007] describe the impact of the SCAND on the climate, whereas Krichak and Alpert [2005] demonstrate the existence of a significant relationship between precipitation and the EAWR.

A lot of studies concerning the variability of individual climate modes and their impact on the Mediterranean climate have been conducted. It is shown that teleconnection patterns exert a significant impact on climatic drivers [Krichak et al., 2014; Martens et al., 2018]. However, several climate modes show a co-varying behaviour as will be later shown in Figure 3. Individual climate modes are highly correlated. An extreme example is the correlation between AMM and TNA with a pearson correlation coefficient of 0.8. While looking only at correlations of individual modes with climate variables, the coupled interactions between individual climate modes are not addressed. However, these interactions are important and their consideration improves the accuracy of the results compared to correlations. Therefore, further research is necessary to further clarify the relation between precipitation anomalies and climate modes.

1.3 Objective of this study

The goal of this study is to reveal the dominant modes of climate variability controlling precipitation anomalies in different parts of an extended Mediterranean area. Thus, the general understanding of the connection between ocean-atmospheric oscillations and precipitation anomalies can be improved. This is essential to gain new insights into the interactions between ocean, atmosphere, land and ultimately the biosphere to accurately predict the impact of climate change on hydrology and ecosystem response in the Mediterranean region.

For this purpose, a supervised machine learning technique called least absolute shrinkage and selection operator (LASSO) regression is used. It is a multivariate method, that analyses the joint behavior of multiple variables concurrently. This is important be-

cause many climate modes are highly correlated, up to a factor of 0.8. That is, if the analysis would be performed separately for each climate mode, as it is the case for a correlation, the cross-correlations between the single climate modes would not be taken into account. However, when using a LASSO regression, the co-varying behavior of the individual climate modes is considered and the impact of many climate modes on precipitation anomalies is determined simultaneously.

2 Data

2.1 Climate modes

Climate modes describe the state of the atmospheric-oceanic circulations and follow distinct repeating spatio-temporal patterns. The strength of each individual climate mode is generally monitored by a so-called Climate Oscillation Index (COI), a simple diagnostic quantity that is used to characterize an aspect of a geophysical system such as a circulation pattern.

In this study, 17 major ocean-atmospheric oscillations, expressed through their associated COI, listed in Table 1, are used as predictive features in a LASSO regression framework to analyze the effects of climate variability on the Mediterranean hydrology.

The following subchapters describe the 17 climate modes used and their associated COI in more detail. A complete list of the used time series and the associated references including the data origin can be found in Table 3 and 4 in the appendix.

Climate mode	Climate Oscillation Index (COI)
Atlantic Meridional Mode (AMM)	AMM sea-surface temperature (AMMSST) index
Atlantic Multidecadal Oscillation (AMO)	AMO index
Indian Ocean Dipole (IOD)	Dipole Mode Index (DMI)
East Atlantic Pattern (EA)	EA index
East Atlantic West Russia Pattern (EAWR)	EAWR index
East Pacific North Pacific Pattern (EPNP)	EPNP index
Northern Annular Mode (NAM)	NAM index
North Atlantic Oscillation (NAO)	NAO index
Pacific Decadal Oscillation (PDO)	PDO index
Polar Eurasia Pattern (PEA)	PEA index
Pacific-North American Pattern (PNA)	PNA index
Southern Annular Mode (SAM)	SAM index
Scandinavian Pattern (SCAND)	SCAND index
El Niño-Southern Oscillation (ENSO)	Southern Oscillation Index (SOI)
Tropical Northern Atlantic Dipole (TNA)	TNA index
Tropical Southern Atlantic Dipole (TSA)	TSA index
West Pacific Pattern (WP)	WP index

Table 1: Climate modes used in this study with their corresponding COI.

2.1.1 Atlantic Meridional Mode (AMM)

The Atlantic Meridional Mode (AMM) is the leading mode of coupled ocean-atmosphere variability in the tropical Atlantic Ocean. It is a dipole mode that has a typical oscillation period of about a decade [Chang et al., 1997]. A Maximum Covariance Analysis between the Sea Surface Temperature (SST) and the zonal and meridional components of 10-m wind defines the spatial structure [Chiang and Vimont, 2004]. The AMM is coupled with an anomalous displacement of the intertropical convergence zone toward the warmer hemisphere.

A positive phase of the AMM is associated with anomalously warm SSTs in the tropical North Atlantic and anomalously cool SSTs in the tropical South Atlantic. These changing SSTs lead to higher than normal surface air pressure over the cold SSTs and lower than normal surface air pressure over the warm SSTs. That, in turn, leads to strengthened south-easterly trade winds in the South Atlantic and weaker north-easterly trade winds in the North Atlantic. Moura and Shukla [1981] observed that these effects produce severe drought conditions over north-east Brazil, whereas Vimont and Kossin [2007] found out that seasonal hurricane activity in the Atlantic is strongly related to the AMM.

The AMMSST index used in this study is obtained from the Earth System Research Laboratory (ESRL) of the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce. The data is originally from the University of Wisconsin-Madison using SST from the National Centers for Environmental Prediction (NCEP).

2.1.2 Atlantic Multidecadal Oscillation (AMO)

The Atlantic Multidecadal Oscillation (AMO) is characterized by North Atlantic SSTs containing a 65-80 year cycle [Enfield et al., 2001].

The AMO has alternating warm and cold phases, which are causing interannual climate variability, influencing rainfall in certain areas. When looking at the United States, warm phases lead to less precipitation than normal in the eastern Mississippi basin, whereas in Florida more rainfall occurs. A cold phase in turn leads to less precipitation in most of the southern parts of the United States [Enfield et al., 2001].

There are two major issues when calculating the AMO index. First, the tropical region is not included in the selected area because it is influenced by the ENSO. Second, global warming is affecting the average of SSTs in a non-linear way. Therefore, van Oldenborgh et al. [2009] defined the AMO index as the averaged SST in the North Atlantic, at latitudes north of 25°N, minus the regression of the SST on global mean temperature.

The AMO index used in this study is obtained from the Royal Meteorological Institute of the Netherlands (KNMI) based on HadSST 3.1.1.0 data of the Met Office Hadley Centre for Climate Science and Services.

2.1.3 Indian Ocean Dipole (IOD)

The Indian Ocean Dipole (IOD) is a climate mode which shows patterns of internal variability with anomalously low SSTs off Sumatra and high SSTs in the western Indian Ocean, with accompanying wind and precipitation anomalies [Saji et al., 1999].

The IOD mode is responsible for 12% of the SST variability in the Indian Ocean and brings heavy rainfall to Eastern Africa and causes severe droughts in Indonesia [Saji et al., 1999]. Black et al. [2003] show that during an extreme IOD mode event a weakening of the westerly flows over the northern Indian Ocean is happening. The consequence is strong East African rainfall because these winds normally transport moisture away from the African continent out over the Indian Ocean.

The associated COI is called Dipole Mode Index (DMI). It is defined as the difference in SST anomaly between the tropical western Indian Ocean (50°E - 70°E, 10°S - 10°N) and the tropical south-eastern Indian Ocean (90°E - 110°E, 10°S - Equator) [Saji et al., 1999].

The DMI used in this study is obtained from the ESRL of the NOAA of the U.S. Department of Commerce. The data is originally from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) using SST based on HadISST data of the Met Office Hadley Centre for Climate Science and Services.

2.1.4 East Atlantic Pattern (EA)

The East Atlantic Pattern (EA) is the second leading mode over the North Atlantic besides NAO [Barnston and Livezey, 1987]. Due to structural similarities with the NAO, the EA is often referred to as southward shifted NAO pattern.

During a positive phase of the EA pattern, Europe experiences above average surface temperatures over the whole year, whereas below average temperatures occur during January to May over the southern United States and during July to October in the north-central United States. Furthermore, above average precipitation is expected over northern Europe and Scandinavia and below average precipitation across southern Europe [Wallace and Gutzler, 1981].

The EA index is originally defined by Wallace and Gutzler [1981], based on normalized 500 hPa height anomalies at three specific centres: one located southwest of the Canary Islands, another west of Great Britain, and a third near the Black Sea. The positive phase is characterized by anomalously high 500 hPa height fields over the North Atlantic and low heights over the subtropical Atlantic and Eastern Europe. The negative pattern index shows a reversed behaviour [Wallace and Gutzler, 1981].

The EA index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using the 1981-2010 base period monthly means and standard deviations. The data is based on height anomalies from the NCEP.

2.1.5 East Atlantic West Russia Pattern (EAWR)

The East Atlantic West Russia Pattern (EAWR) is one of three major teleconnections influencing Eurasia the whole year. Barnston and Livezey [1987] defined this pattern as Eurasian pattern Type 2. It is an east-west pattern with three centres, one near England or Denmark, another north or north-east the Caspian Sea which is oppositely signed, and the third in north-east China having the same sign as the British centre [Barnston and Livezey, 1987].

During a positive phase, positive height anomalies are located over Europe and Northern China, and negative height anomalies are located over the central North Atlantic and north of the Caspian Sea. A positive phase indicates higher surface temperatures over Eastern Asia, and lower surface temperatures over large portions of Western Russia and North-Eastern Africa. Furthermore, below average precipitation across Central Europe and above average precipitation in Eastern China is occurring.

The EAWR index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using

the 1981-2010 base period monthly means and standard deviations. The data is based on height anomalies from the NCEP.

2.1.6 East Pacific North Pacific Pattern (EPNP)

The East Pacific North Pacific Pattern (EPNP) is a pattern appearing mostly in spring, summer, and fall with three main anomaly centres: one located in the central North Pacific, another near Korea and the last near Eastern Alaska [Barnston and Livezey, 1987]. Barnston and Livezey [1987] show that the EPNP is reflecting the strength of the westerlies in the 50° - 70°N zone in Eastern Asia and across the Pacific.

The positive phase of this pattern is characterized by positive height anomalies located over Alaska and Western Canada, and negative anomalies over the central North Pacific and eastern North America. A persistent positive phase of the EPNP is associated with an anomalous circulation with higher cyclone activity over the middle latitudes of the North Pacific and with lower cyclone activity over the western and central United States. Furthermore, it leads to an intensification of the Pacific jet stream from Eastern Asia to the eastern North Pacific. A strong negative phase of this pattern features circulation anomalies of opposite sign in these areas.

During a positive phase, the surface temperatures are above average over the eastern North Pacific, and below average over the central North Pacific and eastern North America. Furthermore, higher precipitation is arising in the region north of Hawaii, whereas lower precipitation is occurring over south-western Canada.

The EPNP index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using the 1981-2010 base period monthly means and standard deviations. The data is based on height anomalies from the NCEP.

2.1.7 Northern Annular Mode (NAM)

The Northern Annular Mode (NAM), also known as Arctic Oscillation, is a leading mode in the Northern Hemisphere during wintertime [Thompson and Wallace, 1998]. It is characterized by winds circulating counterclockwise around the Arctic at around 55°N latitude. The NAM is related to the NAO which shows a similar structure.

The negative phase is defined by high pressure over the polar region and low pressure at mid latitudes. The positive phase shows a reversed behaviour. Due to the weaker pressure system during a negative phase, the cold arctic air masses move south, causing more storms in the mid latitudes. In contrast, in a positive phase the cold arctic air does not move across polar regions. Consequently, the northern Hemisphere, except regions as Greenland or Newfoundland, experiences temperatures warmer than usual. Additionally, the positive phase, characterized by higher pressure at mid latitudes, brings wetter conditions to northern countries, like Scotland and Scandinavia, whereas the Mediterranean or the western United States experience drier conditions.

In the past centuries the NAM alternated between its positive and negative phases. However, since the 1970s, a trend to a lower than normal arctic air pressure can be seen, meaning that the NAM tends to stay in the positive phase.

The NAM index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The data is based on height anomalies from the NCEP.

2.1.8 North Atlantic Oscillation (NAO)

The North Atlantic Oscillation (NAO) is the dominant climate mode in the North Atlantic. It contains a strong center over Greenland, and another of opposite sign over the Atlantic, Europe or the United States [Barnston and Livezey, 1987].

The positive phase of the NAO is associated with weaker than usual heights and pressure across the high latitudes of North Atlantic and stronger than usual heights and pressure over the central North Atlantic, the eastern United States and Western Europe. These height and pressure patterns lead to warmer and wetter conditions over northern Europe and most of north-eastern North America, whereas below normal temperatures and precipitation are observed over Greenland and southern Europe. The negative phase of the NAO shows a reversed behaviour. It brings cold air to northern Europe and the north-eastern North America, and moist air to southern Europe.

The NAO index is defined as the difference of normalized surface sea-level pressure at Lisbon, Portugal and Reykjavik, Iceland. The NAO index used in this study is obtained from the KNMI. The data is originally from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using the 1981-2010

base period monthly means and standard deviations. The data is based on sea-level pressure anomalies from the NCEP.

2.1.9 Pacific Decadal Oscillation (PDO)

The Pacific Decadal Oscillation (PDO) is often described as a ENSO-like pattern of Pacific climate variability [Zhang et al., 1997]. The spatial pattern are similar, however, the main difference between these two modes of climate variability is their timescale. The ENSO is an interannual phenomena, while the PDO is decadal in scale. Therefore, observations for a long time are needed to understand the PDO and analyze its phases.

During a positive phase of the PDO the SSTs are anomalously warm along the eastern Pacific Coast and sea-level pressures are below normal over the North Pacific. A reversed behaviour is associated in a negative phase. Extremes in the PDO are associated with widespread variations in the Pacific Basin and the North American climate [Mantua et al., 1997].

The PDO index used in this study is obtained from the National Centers for Environmental Information of the NOAA of the U.S. Department of Commerce. The PDO index is based on NOAA's extended reconstruction of SSTs (Extended Reconstructed Sea Surface Temperature (ERSST) Version 4).

2.1.10 Polar Eurasia Pattern (PEA)

The Polar Eurasia Pattern (PEA) is a teleconnection pattern, that is apparent over the whole year [Gao et al., 2016].

The positive phase of the PEA is associated with positive height anomalies over Northern China and Mongolia, and negative height anomalies over the polar region [Barnston and Livezey, 1987]. These height anomalies lead to above normal temperatures in Siberia and below normal temperatures in Eastern China. In addition, more than usual precipitation is expected in the polar region north of Scandinavia during a positive phase.

The PEA index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using the 1981-2010 base period monthly means and standard deviations. The data is based on height anomalies from the NCEP.

2.1.11 Pacific-North American Pattern (PNA)

The Pacific-North American Pattern (PNA) is a dominant teleconnection over the North Pacific-North American region [Chen and van den Dool, 2003]. It is characterized by alternating pressure patterns in the central Pacific Ocean and over Western Canada and the south-eastern United States [Wallace and Gutzler, 1981].

The positive phase of the PNA is associated with above average geopotential heights over the western United States and below average geopotential heights over the eastern United States. These height field pattern result in above normal temperatures in Western Canada and the western United States, whereas the south-central and south-eastern United States experience below normal temperatures. Moreover, above average precipitation is identified in the Gulf of Alaska extending into the Pacific north-western United States, and below average precipitation over the upper Midwestern United States.

The PNA index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using the 1981-2010 base period monthly means and standard deviations. The data is based on height anomalies from the NCEP.

2.1.12 Southern Annular Mode (SAM)

The Southern Annular Mode (SAM), also known as Antarctic Oscillation is the leading mode of variability in the southern hemisphere [Marshall, 2003].

The SAM influences the westerly wind belt that circles Antarctica. The changes in the position and strength of the westerlies can have a major impact on rainfall and temperature variability.

In a positive phase of the SAM stronger than average westerlies occur over the mid-high latitudes and weaker westerlies in the mid-latitudes. This leads to higher pressures and drier conditions over Southern Australia. During a negative phase, the westerly winds expand towards the equator, resulting in low pressure systems over Southern Australia. Consequently, increased storms and rain are expected over this area.

The SAM index is based on the zonal pressure differences between 40°S and 60°S. The observations are based on 12 stations. The SAM index used in this study is obtained from Marshall [2003].

2.1.13 Scandinavian Pattern (SCAND)

The Scandinavian Pattern (SCAND) was originally defined by Barnston and Livezey [1987] as Eurasian Type 1 pattern. Its primary centre of action is located around the Scandinavian Peninsula and two other centres with opposite sign are found over the North-Eastern Atlantic and over Central Siberia to the south-west of Lake Baikal.

The positive phase of the SCAND is characterized by positive height anomalies over the Scandinavian Peninsula and Western Russia. These patterns lead to lower than usual temperatures across Central Russia and Western Europe. Furthermore, the positive phase is associated with higher than usual precipitation across central and southern Europe, and lower than usual precipitation across Scandinavia [Bueh and Nakamura, 2007]. The negative phase of the SCAND is associated with a reversed behaviour in these regions.

The SCAND index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using the 1981-2010 base period monthly means and standard deviations. The data is based on height anomalies from the NCEP.

2.1.14 El Niño-Southern Oscillation (ENSO)

The El Niño-Southern Oscillation (ENSO) is a well-known teleconnection pattern and one of the most prominent global drivers of interannual variability [Trenberth and Caron, 2000]. The ENSO describes periodic fluctuations in SST and air pressure across the equatorial Pacific Ocean. Due to the changing water temperatures in the eastern tropical Pacific, there are far-reaching consequences to weather patterns.

The negative phase of the SOI is associated with below normal air pressure at Tahiti and above normal air pressure at Darwin. Long periods of negative SOI values lead to abnormally warm ocean waters across the eastern tropical Pacific, referred to as El Niño. During El Niño enhanced cloudiness and rainfall occur along the coasts of Ecuador and northern Peru because of the abnormally warm waters. In contrast, at the same time rainfall is reduced over Indonesia, Malaysia and Northern Australia, leading to drier than normal conditions [Ropelewski and Halpert, 1987].

In contrast, prolonged positive periods of the SOI come with abnormally cold ocean waters across the eastern tropical Pacific, referred to as La Niña. During La Niña

the situation is reversed: rainfall is enhanced over Indonesia, Malaysia and Northern Australia, whereas less rainfall is occurring along the eastern tropical Pacific [Ropelewski and Halpert, 1989].

The corresponding index is called Southern Oscillation Index (SOI) [Chen, 1982]. It measures the differences in air pressure anomaly between Tahiti and Darwin, Australia. The SOI used in this study is obtained from the Australian Government Bureau of Meteorology. The data is based on means and standard deviations calculated over the period 1933 to 1992 inclusive.

2.1.15 Tropical Northern Atlantic Dipole (TNA)

The Tropical Northern Atlantic Dipole (TNA) is defined by Enfield et al. [1999] as the average of SSTs over the region 5°N - 25°N, 55°W - 15°W. It is associated with variations in the so called Hadley circulation [Wang, 2002]. The TNA shows a connection to other teleconnection patterns, namely ENSO and NAO [Andreoli, 2004; Rajagopalan et al., 1998].

The TNA index used in this study is obtained from the ESRL of the NOAA of the U.S. Department of Commerce. The indices are standardized by the 1971-2000 climatology. The data is based on HadISST and NOAA OI 1x1 datasets.

2.1.16 Tropical Southern Atlantic Dipole (TSA)

The Tropical Southern Atlantic Dipole (TSA) is defined by Enfield et al. [1999] as the average of SSTs over the region 20°S - 0, 30°W - 10°E. As the TNA, also the TSA is associated with variations in the so called Hadley circulation [Wang, 2002].

The TSA index used in this study is obtained from the ESRL of the NOAA of the U.S. Department of Commerce. The indices are standardized by the 1971-2000 climatology. The data is based on HadISST and NOAA Optimal Interpolation (OI) 1x1 datasets.

2.1.17 West Pacific Pattern (WP)

The West Pacific Pattern (WP) is a teleconnection pattern with two centers of action [Wallace and Gutzler, 1981]. One center is located east of Kamchatka and another broad center of opposite sign in the east-west Pacific [Barnston and Livezey, 1987].

The positive phase of the WP leads to higher than average temperatures over the lower latitudes of the western North Pacific in winter and spring, and to lower than average temperatures over eastern Siberia over the whole year. While the high latitudes of the North Pacific experience above normal precipitation in all seasons, below normal precipitation is expected across the central North Pacific, in particular, in winter and spring.

The WP index used in this study is obtained from the Climate Prediction Center of the NOAA of the U.S. Department of Commerce. The indices are normalized using the 1981-2010 base period monthly means and standard deviations. The data is based on height anomalies from the NCEP.

2.2 Precipitation data

One key component of the hydrological cycle is precipitation [Bonan, 2015]. Many studies require precipitation estimates, but the non-homogenous spatial and temporal precipitation patterns, as well as the limitations of the observing systems, make it difficult to provide high-quality datasets covering long time series [Stephens and Kummerow, 2007; Tang et al., 2009; Sun et al., 2018; Sene, 2013].

In this study, precipitation provided by the Climatic Research Unit (CRU) TS v.3.23 [Harris et al., 2013] is used. In the LASSO framework, precipitation anomalies are used as the target variable and the climate modes are the features. Thus, the impact of the individual climate modes on precipitation anomalies can be quantified.

2.3 Data preparation

2.3.1 Climate Oscillation Index (COI) data

The 17 climate modes discussed from section 2.1.1 to 2.1.17 include all important oscillations covering the globe. The COIs have a monthly temporal resolution. The time series are pre-processed based on the multi-year climatology to eliminate the effect of seasonal variations. Therefore, the raw time series of the COIs are standardized by month. This means that the mean of all January values is calculated and subtracted from these values. Afterwards, the January values are standardized to give them a standard deviation of one. This process is repeated for every month. As an example Figure 1 shows the time series of the NAM index before and after the pre-processing. Finally, the full time series is normalized by dividing it by the $l2$ norm for better numerical behavior during the

regression algorithm and easier interpretation during the LASSO regression, see section 3.1.

There are no data gaps for the climate modes, except the East Pacific North Pacific Pattern (EPNP). For the entire time span of the EPNP the December values are missing. These data gaps are filled with the arithmetic mean of the respective November and January value.

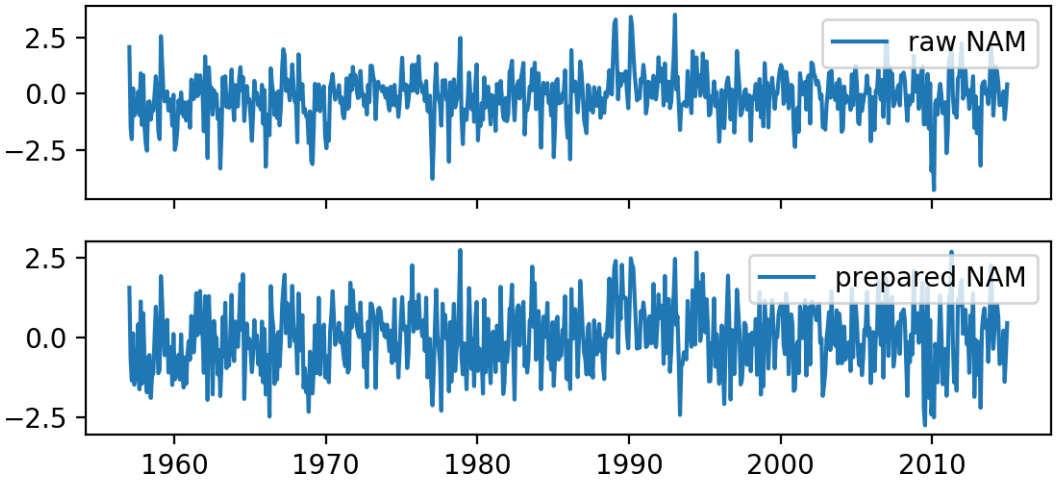

Figure 1: Illustration of the time series of the NAM index before and after pre-processing.

2.3.2 Precipitation data

CRU TS is a gridded high-resolution climate dataset from monthly observations at meteorological stations across the world's land areas. It is provided on a regular latitude-longitude grid with a common spatial resolution of 0.5°. This dataset covers all land areas but Antarctica over the period 1901 to 2014.

The LASSO regression is trained by the anomalies of the precipitation data instead of the raw time series to remove seasonal signals. First, the linear trend is subtracted, then the anomalies are calculated based on the multi-year climatology similar to the climate modes, see section 2.1, but without standardizing the monthly data. As an example Figure 2 shows the raw time series of precipitation and the derived precipitation anomalies at the location 38.75°N and 7.25°W. This specific location is further investigated in section 4.3. In contrast to COI data, occurring data gaps are not filled and thus not included in the analysis leaving the original time series unchanged.

Anomalies are used because the time series are dominated by a yearly signal due to seasonal effects as can be seen in Figure 2. It is known that the winters in the Mediterranean area are wetter than the summers [Kottek et al., 2006; Kelley et al., 2012]. However, the seasonal signals are not of interest and should therefore be removed during the analysis. This is achieved by using anomalies.

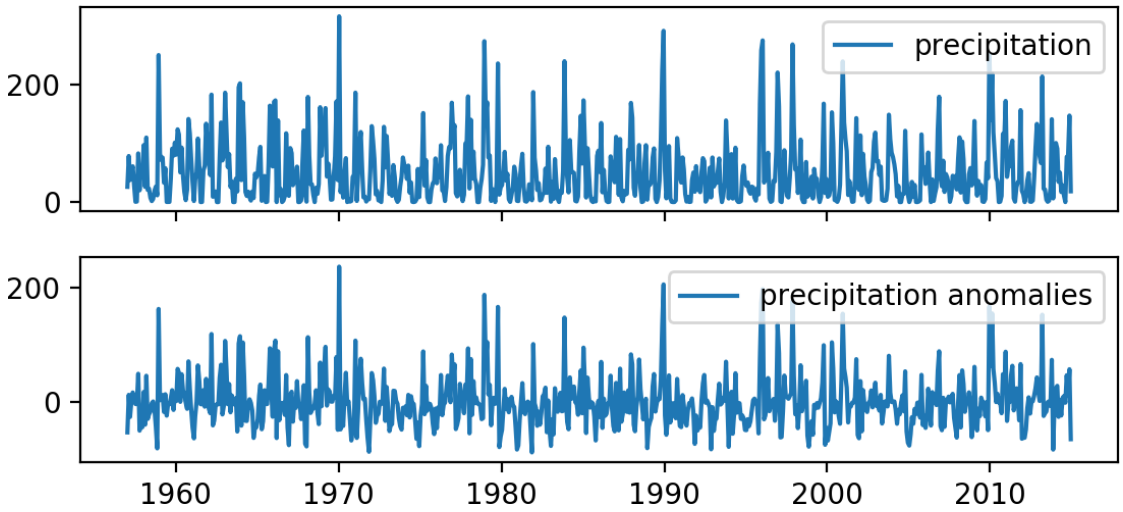

Figure 2: Example time series of precipitation and precipitation anomalies at 38.75°N and 7.25°W

3 Methods

3.1 LASSO regression

For this study, a supervised machine learning algorithm based on least absolute shrinkage and selection operator (LASSO) models [Tibshirani, 1996] is used to disentangle the impact of COIs on precipitation anomalies. In our framework, 17 COIs (see Table 1) are used as features, and anomalies of CRU TS precipitation (see section 2.2) are used as the target variable. The LASSO regression is chosen because it is a relatively simple but more advanced method than correlation. However, compared to other machine-learning algorithms like random forest, the LASSO regression is easier to understand and evaluate. Furthermore, providing a simple model helps interpreting the result. Another advantage is that the LASSO supervised learning approach takes cross-correlations between the single predictors into account, which leads to an improvement in the quality of the model compared to correlations.

The cross-correlations need to be considered because the features, i.e. the COIs, are showing a co-varying behavior, as shown in Figure 3. Some COIs are correlated up to a factor of 0.8. If seasonal models are computed the cross-correlations increase even more.

The LASSO regression is a data-driven method that uses automatic feature selection and regularization to quantify the impact of each feature on the target variable. The regularization helps to prevent overfitting and helps to provide a simple model by setting the impact of weak features to zero. This means that only a subset of features is left for the final predictive model, which improves its prediction accuracy and simplifies the interpretation of the result.

The LASSO regression finds the estimated regression coefficients $\hat{\beta}$ by minimizing a penalized residual sum of squares [Tibshirani, 1996].

$$\hat{\beta} = argmin \left\{ \sum_{i=0}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \alpha \sum_{j=1}^{p} |\beta_j| \right\} \tag{3.1}$$

In equation (3.1), $\hat{\beta}$ is the p-dimensional vector with the estimated regression coefficients, n the number of training samples in the dataset, y_i the value of the target variable in sample i, p the number of features, x_{ij} the value of feature j in sample i, and α a hyper-parameter controlling the amount of shrinkage [Martens et al., 2018].

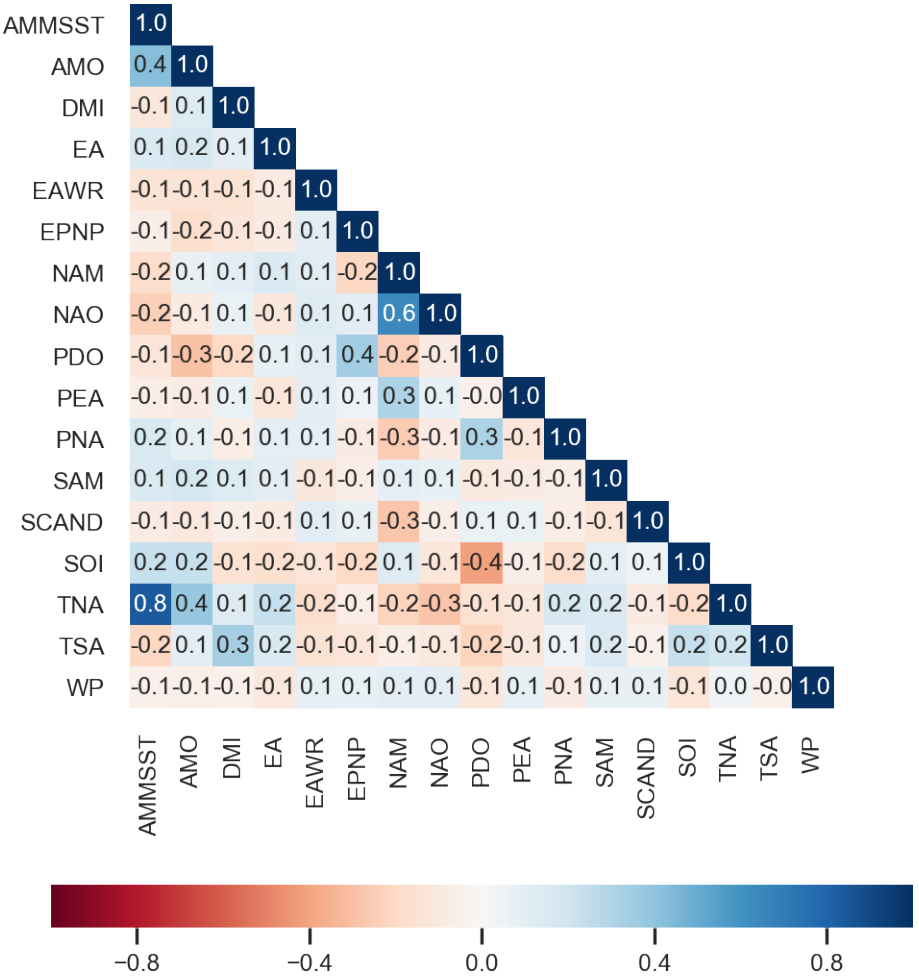

Figure 3: Cross-correlation of COI anomalies

Concentrating the individual feature vectors x_i lead to the feature matrix X (3.2) shown in Figure 5, 6 and 8.

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \tag{3.2}$$

With the help of the resulting regression coefficients $\hat{\beta}$ of the LASSO regression the target variable can be predicted,

$$y_{pred} = \sum_{feature=1}^{17} \left\{ \sum_{lag=0}^{5} \left(x_{feature_{lag}} \cdot \hat{\beta}_{feature_{lag}} \right) + bias \right\} \tag{3.3}$$

where y_{pred} is the value of the predicted target variable, x is the feature vector for a given time lag, and $\hat{\beta}$ the vector with the corresponding estimated regression coefficients.

For the evaluation of the predictive model the performance measure R^2, the coefficient of determination, is used, which is the squared correlation between our target variable y and the predicted target variable y_{pred}.

$$R^2 = corr(y, y_{pred})^2 \tag{3.4}$$

3.2 Cross-validation

To determine the regularization parameter α and evaluate the coefficient of determination R^2, two five-fold cross-validations are applied for each grid point. First, one five-fold cross-validation is used for the determination of the regularization parameter α, after this parameter is fixed, the second five-fold cross-validation is applied for the calculation of the coefficient of determination R^2.

During the cross-validation, the data is split randomly into five equally sized folds. Four of these folds are used to train the model and calculate predictions for the remaining test fold. This step is repeated five times until predictions for every fold are calculated. The five predictions are put together and the coefficient of determination R^2 is calculated based on the predicted and original values.

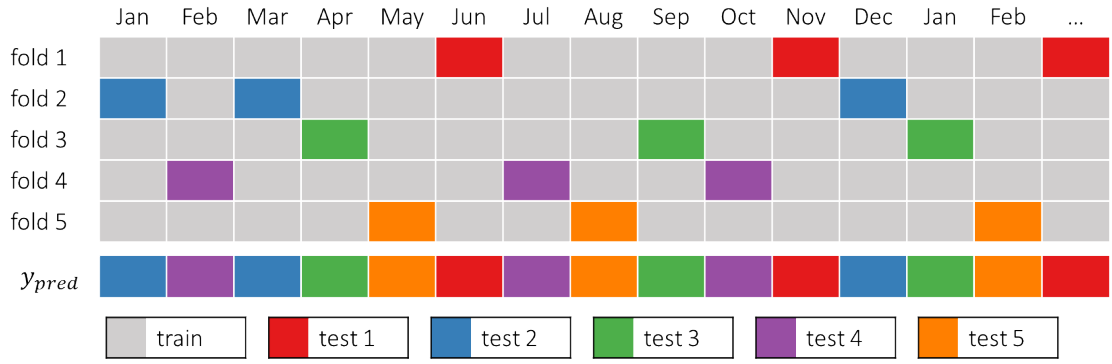

Figure 4: Concept of five-fold cross-validation for R^2. Every time stamp is used once for testing.

Figure 4 illustrates the scheme of the five-fold cross-validation for R^2. The data is shuffled and spread among five equally sized folds. Four of these folds are used for training, illustrated as grey time stamps, whereas the fifth fold, including the coloured

time stamps, remains for testing. A LASSO model is fitted based on the training data and used to calculate predictions for the testing time points. This procedure is repeated five times, until every time stamp is used once as testing data and, therefore, predictions are available for the entire time span. If, for example, folds 1 to 4 are the training folds, the orange time stamps in fold 5 would be the ones predicted. If the folds 2 to 5 are used to train the model, the red time stamps in fold 1 would be the calculated predictions, and so on. In the end, every time stamp is predicted once, based on one of the five models. This information is combined to produce the predicted time series y_{pred} as shown in the last row of Figure 4 which is used to calculate the explained variance R^2, see equation (3.4).

3.3 Significance test

To ensure that the results are robust and trustful, a significance test at a significance level of 95% is applied. First, a nonparametric permutation test is used to obtain the null distribution of the test statistic (R^2 values). LASSO models are fitted using the same data as target variable but shuffled in time. Therefore, the nonparametric null distribution of the R^2 values is gained, which is compared with the distribution of the original R^2 values. As a second step the false discovery rate at a level of 5% is controlled by adjusting the p-values according to the Benjamini-Hochberg procedure [Benjamini and Hochberg, 1995]. The procedure is carried out by arranging the p-values in ascending order, then assigning ranks to the p-values, so that the smallest p-value starts with rank one, the second smallest with rank two etc., and finally calculating the Benjamini-Hochberg critical value for each individual p-value. The Benjamini-Hochberg critical value is calculated with the formula,

$$\frac{i}{m} \cdot Q \tag{3.5}$$

where i is the rank of the individual p-values, m is the total number of tests, and Q is the false discovery rate, in our case 5%. The largest p-value that is smaller than the critical value (3.5), is significant and defines the threshold. All the p-values smaller than that are significant too, the remaining are rejected. As a result, the significant R^2 values can be declared.

3.4 Experimental set up

In this study, an extended Mediterranean area spanning from 28.5°N to 56.5°N and from 10°W to 46°E is investigated. A LASSO regression model (see section 3.1) is fitted individually for every grid point using 17 major COIs as features (see section 2.1) and precipitation anomalies as target variable (see section 2.2). The overlapping time period for the climate modes and the precipitation dataset is from 1957 to 2014. This time span is used for the analysis.

In order to consider possible delayed responses of the COIs on the Mediterranean hydrology, six time lags ranging between zero and five months are introduced for every predictor, resulting in a total of 102 predictive features, i.e. 17 COIs times six time lags.

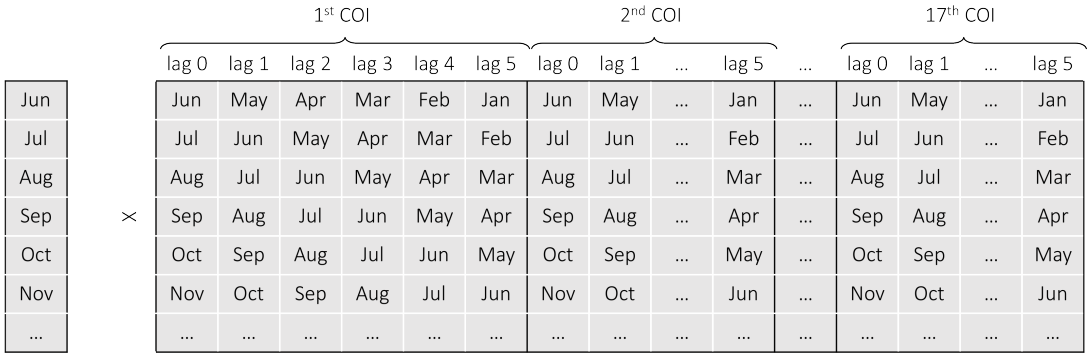

Figure 5: Basic framework of the LASSO regression

Figure 5 shows the concept of the LASSO regression when using all months of a year. The left block is the target variable vector y (3.1), in our case precipitation anomalies from CRU, and the right block represents the feature matrix X (3.2), containing 17 features, i.e. 17 COIs listed in Table 1, each with six time lags. The data is sampled with a monthly temporal resolution.

In general, if more than 50% of the target data are missing values or if less than 10 samples are available, the LASSO regression is not performed for that particular grid point due to a lack of reliability.

To optimize and validate the predictive LASSO regression models, two five-fold cross-validations are applied (see section 3.2), one for the regularization parameter α, see equation (3.1), and the second for the coefficient of determination R^2, see equation (3.4), a metric indicating the performance measure of the model. Furthermore, a significance

test using the Benjamini-Hochberg procedure (see section 3.3) [Benjamini and Hochberg, 1995] is applied to verify the results.

3.4.1 Seasonal models

Additionally, seasonal models are calculated as it was done in similar studies [Martens et al., 2018; Mariotti et al., 2002; Bladé et al., 2012]. By using seasonal models a higher coefficient of determination R^2 is achieved, meaning that the derived models from the COIs fit better to the precipitation anomalies. The seasonal models are trained using only data of three months:

- winter model: December, January, February
- spring model: March, April, May
- summer model: June, July, August
- autumn model: September, October, November

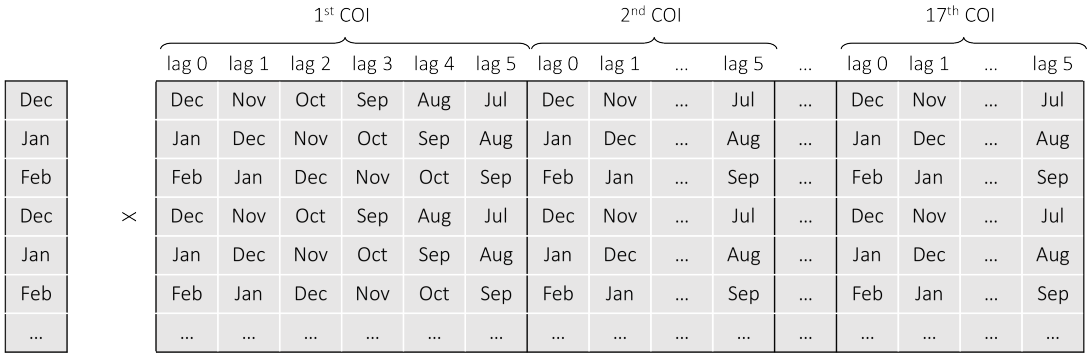

Figure 6: Framework of the LASSO regression for the winter model. Due to the six time lags also months before December occur in the framework for the features.

Figure 6 shows the concept of the LASSO regression for the winter model. The winter model covers the time span between December and February of each year for the target variable. As already described, time lags are introduced for the feature vectors x (3.1). Therefore, only the feature vector with $lag = 0$ covers the same period. However, when time lags are applied, the months shift and the data is used up to July.

3.4.2 Neighborhood information

To increase the predictive performance of the model and make the results more robust, the 3×3 neighborhood is included. The framework is extended by the data of the eight

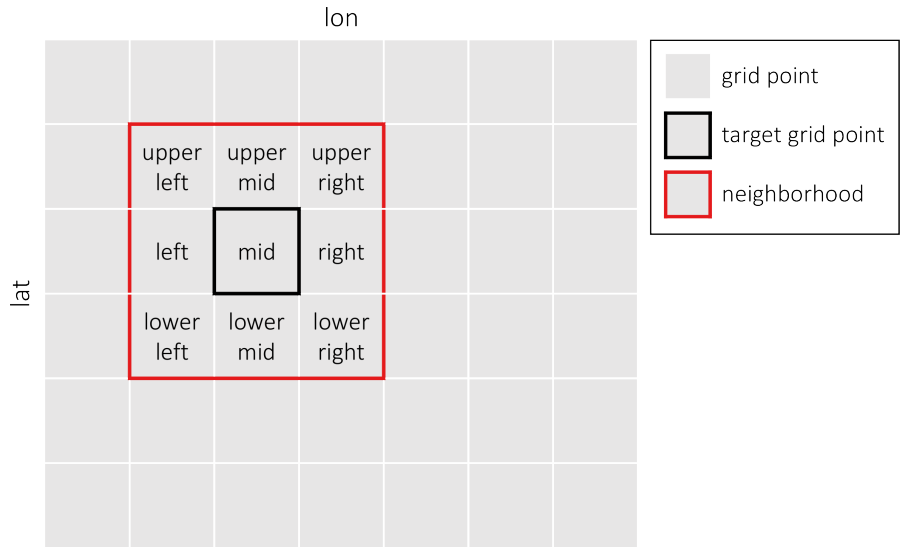

Figure 7: Illustration of the 3×3 neighborhood

neighboring grid points. Since the variables are provided with a monthly resolution and the time series covers 58 years, the number of samples is relatively short for using supervised machine learning algorithms. This is especially true for the seasonal models when keeping the additional 20% data reduction due to five-fold cross-validation in mind. It is expected that by including the neighborhood information and therefore extending the time series by a factor of nine, more information can be extracted from the time series and fit into the model. If one individual grid point does not provide enough information to create a good model, at least the neighboring grid points can help improving the prediction.

Figure 7 shows the concept of the 3×3 neighborhood. The grey grid illustrates the spatial expansion of the examined area with its respective latitude and longitude. An individual LASSO regression model is computed for each target grid point. The black square marks the target grid point for which the LASSO regression model is currently computed, while the red square marks the corresponding 3×3 neighborhood. The time series of the target grid point is extended by the eight neighboring grid points as shown in Figure 8. Due to the neighborhood, the time span typically increases by a factor of nine. Since the CRU data set only contains values over land areas, grid points near the coast typically have fewer neighbors with data and, therefore, a fewer amount of samples can be used. At the corners of the investigated area, only three neighbors are available, while at the edges only five neighbors can be used.

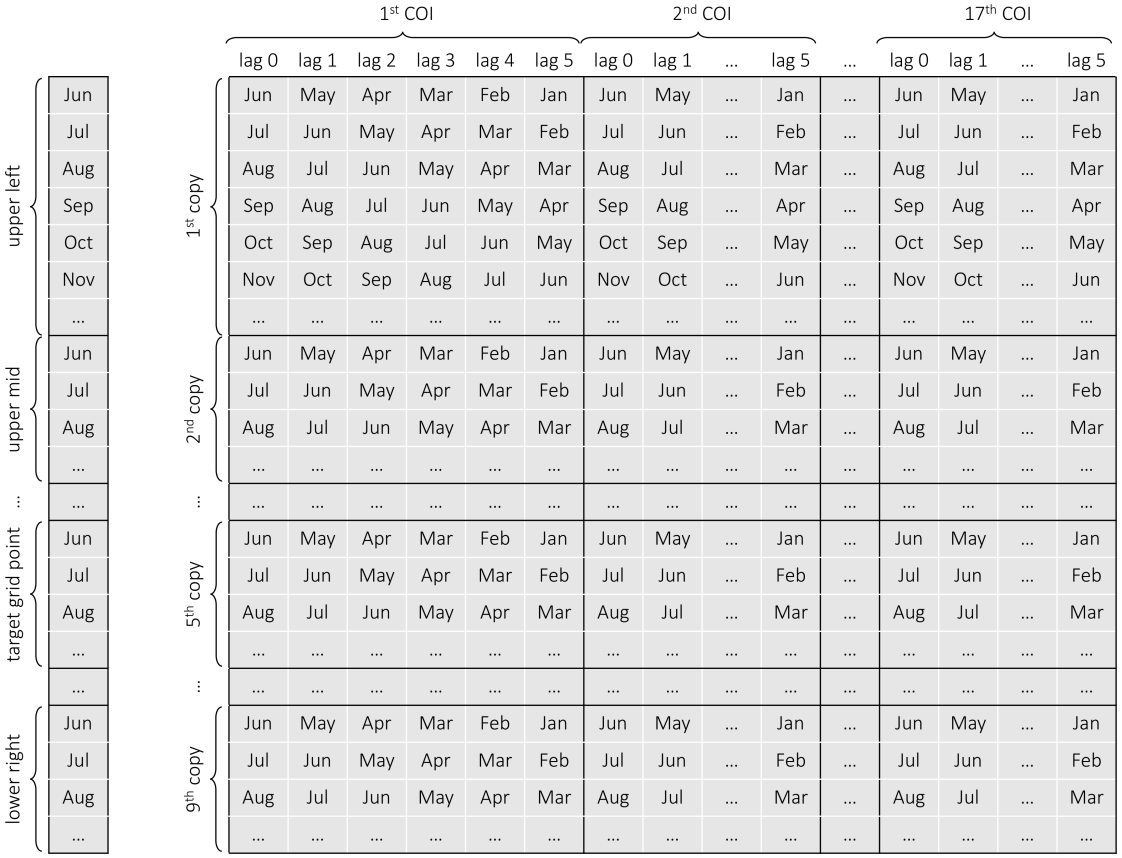

Figure 8: Framework of the LASSO regression for the full time series including the 3×3 neighborhood. The terms upper left, upper mid and lower right next to the target variable y indicate the position of the specific grid point in the 3×3 neighborhood as shown in Figure 7

Figure 8 illustrates how the time series for the target variable y (3.1) is extended by up to eight neighbours. In detail, this means that the time series of all nine grid points are concatenated for the target variable. Consequently, the time series for one grid point is artificially prolonged by its eight neighbors, resulting in nine times more samples for the target variable. The 102 predictive COI features x (3.1), which have no spatial affiliation, remain the same but are copied nine times, resulting in the same number of samples as for the target variable.

It is assumed that the areas lying close together have similar climatic conditions influenced by precipitation in a similar way. Therefore, it is reasonable to include the 3×3 neighborhood to get more appropriate results. From a mathematical point of view, the results get smoothed because the noise is reduced.

4 Results and Discussion

4.1 Analysis based on coefficient of determination

4.1.1 Full-year model

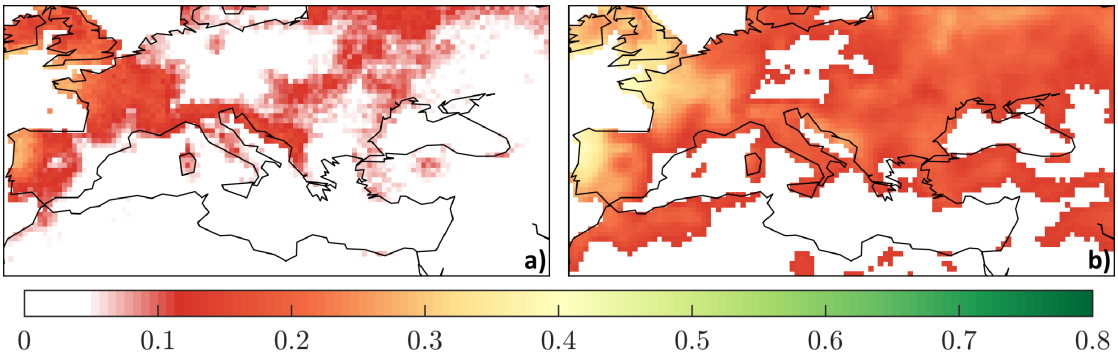

Figure 9: Explained variance R^2 of the LASSO regression with monthly precipitation anomalies as the target and 17 climate modes with six time lags as features. Areas that are classified as non-significant are marked white. a) the model is trained for each pixel individually, b) additional 3×3 neighborhood is taken into account

Figure 9 shows the explained variance R^2 of the precipitation anomalies obtained by the LASSO regression using all months of a year from 1957 to 2014. In the left plot, the LASSO regression is trained for each pixel individually, while in the right plot additionally the 3×3 neighborhood is taken into account. By comparing the two maps in Figure 9 it can be seen that the explained variance R^2 gets significantly higher when the 3×3 neighborhood is included. The reason for the higher coefficient of determination R^2 when using the 3×3 neighborhood is, that the impact of the noise in the data is reduced due to the nine times longer time series. Thus, the time series gets more robust. If the information provided by one pixel is not sufficient for deriving a good model, at least the information of neighboring pixels can help to improve the model. Therefore, it gets easier for the LASSO regression to predict the precipitation anomalies and consequently, the model gets significantly better.

The white areas in the maps in Figure 9 have been declared as non-significant by the permutation test using the Benjamini-Hochberg procedure at a significance level of 95%. Both results in Figure 9 agree that precipitation anomalies in the northern and eastern region of the extended Mediterranean area, as well as precipitation anomalies in

Northern Africa are not significantly driven by any climate mode.

The hot spot regions, where climate modes have a high influence on precipitation anomalies, occur in particular over the Iberian Peninsula, Ireland, the United Kingdom, France, big parts of Italy, and the eastern Adriatic coast. In these regions, up to 40% of the variability can be explained. These results provide evidence that the monthly precipitation anomalies are significantly driven by the main modes of climate variability.

4.1.2 Seasonal models

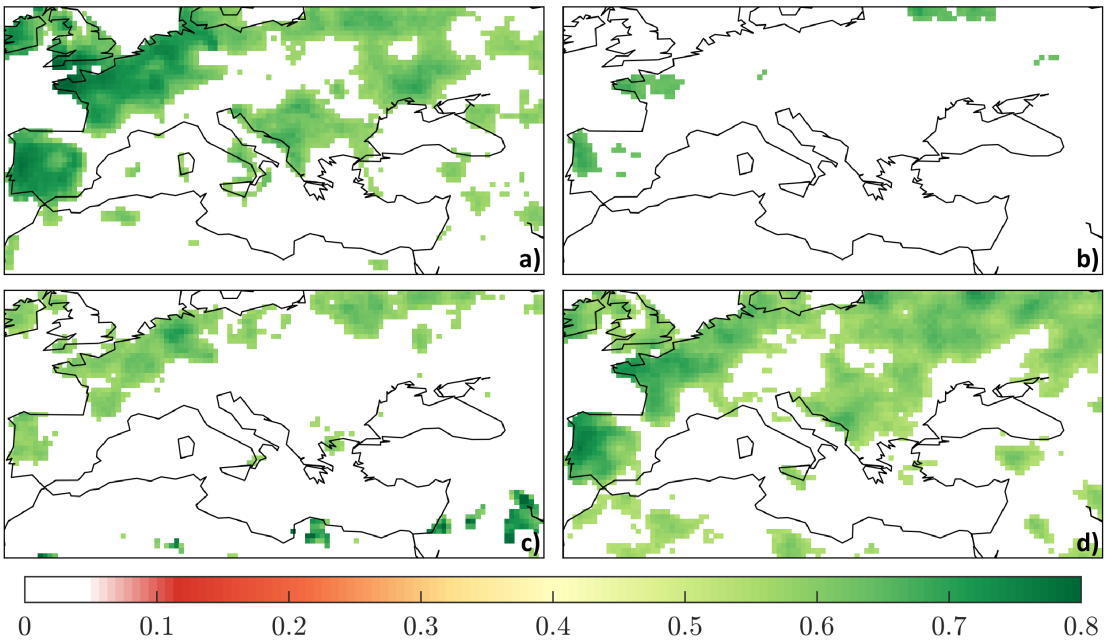

Figure 10: Explained variance R^2 for seasonal models including the 3×3 neighborhood. a) winter model, b) spring model, c) summer model, d) autumn model

The seasonal models show a considerably higher coefficient of determination R^2 than the model that uses precipitation anomalies of all months of a year shown in Figure 9. This means that the seasonal signals can be better explained by the COI anomalies. When computing separate models, as shown in Figure 10, variabilities up to 70% can be estimated. Especially, precipitation anomalies in the winter model show a high sensitivity to the climate modes. In particular, the Iberian Peninsula, Ireland, the United Kingdom, France, Belgium, the Netherlands, as well as some parts of Italy, and the eastern coast of the Adriatic Sea are influenced by the modes of climate variability. The autumn model looks similar to the winter model but less strong, highlighting the same

regions responding to the climate modes. In summer only in a few regions precipitation anomalies are driven by the COIs. The western Iberian Peninsula, Ireland, northern France, Belgium, and the Netherlands are showing a high explained variance R^2. In contrast, the spring model is showing hardly any explained variance R^2, meaning that there is nearly no influence of teleconnection patterns on precipitation anomalies during that season.

The patterns seen in Northern Africa, and the Middle East, especially in the summer model, have to be interpreted with caution. In these regions, no precipitation occurs. Therefore, the precipitation signal is constant zero, which, obviously, makes it easy for the LASSO regression to create a good model and thus leading to a high explained variance R^2.

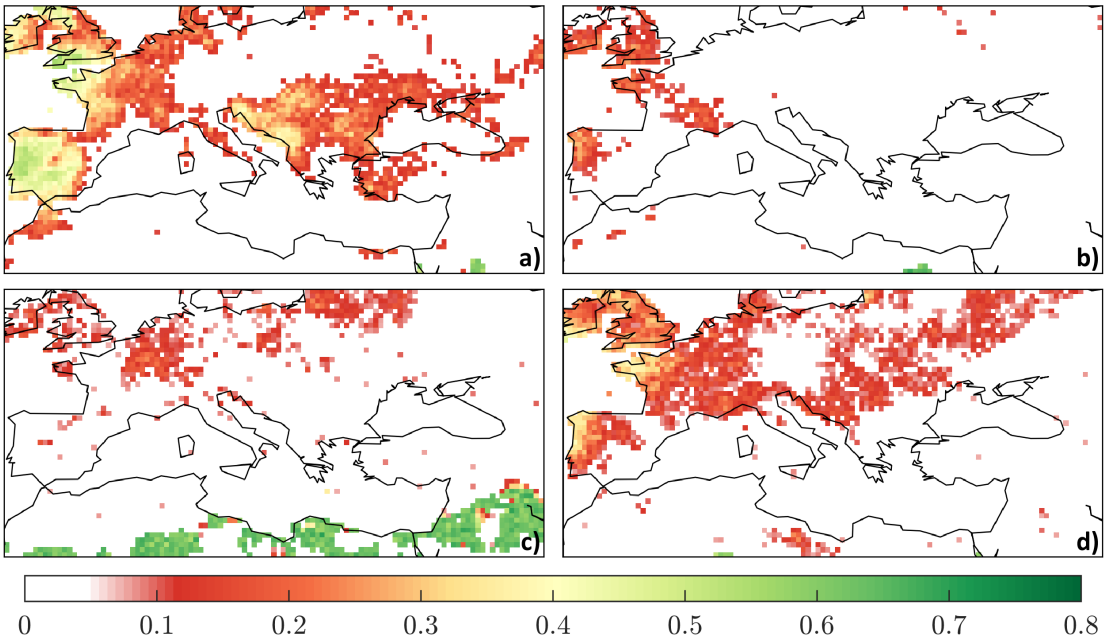

Figure 11: Explained variance R^2 for seasonal models without using the 3×3 neighborhood. a) winter model, b) spring model, c) summer model, d) autumn model

Figure 11 shows the explained variances R^2 of the precipitation anomalies when the data is separated for each season but without using the 3×3 neighborhood. The spatial patterns detected look very similar to the ones identified by the LASSO regression that includes the information of the 3×3 neighborhood shown in Figure 10. However, the explained variances R^2 are significantly higher, when the 3×3 neighborhood is taken into

account. This confirms our findings in Figure 9, where the full-year model is computed.

4.2 Impact of individual COIs

Figure 9, 10 and 11 show that teleconnection patterns influence precipitation anomalies in the extended Mediterranean area.

The impact of each individual COI is shown by calculating the explained variance for each individual COI R^2_{COI}. This value is calculated by taking the six regression coefficients $\hat{\beta}$ of one individual COI from the LASSO regression and based on these coefficients the predicted target variable $y_{pred_{COI}}$ and finally the explained variance R^2_{COI} is calculated. Therefore, equation (3.4) changes to (4.2), and equation (3.3) changes to (4.1). This is done for each COI individually.

$$y_{pred_{COI}} = \sum_{lag=0}^{5} x_{COI_{lag}} \cdot \hat{\beta}_{COI_{lag}} \tag{4.1}$$

$$R^2_{COI} = corr(y, y_{pred_{COI}})^2 \tag{4.2}$$

4.2.1 Full-year model

Looking at the explained variances R^2_{COI} of the full-year model, the same five COIs are detected having a significant impact as for the winter or autumn model, namely EA, EAWR, NAM, NAO, and SCAND. Interestingly, the remaining 12 COIs do not show any influence at all, the explained variances R^2_{COI} are zero. This can be explained by the automatic feature selection of the LASSO regression. Since all four seasons are put together into one model, it is significantly harder for the LASSO regression to disentangle the impact of individual features. The explained variances R^2_{COI} detected for each COI for the full-year model are significantly lower than the ones when separating the time period into the four seasons.

Each climate mode drives precipitation anomalies in regions of the extended Mediterranean area. The patterns detected by the model that uses all months of a year correspond to the patterns detected by one or more of the seasonal models, shown in Figure 13, 15, 16, and 17.

The EA drives precipitation anomalies in several regions of the extended Mediterranean area. Some patterns seen in the model that uses all months of a year correspond to the patterns found in the winter, spring and autumn model. The impact west of the Black Sea can be seen in the autumn model, whereas the patterns in northern France

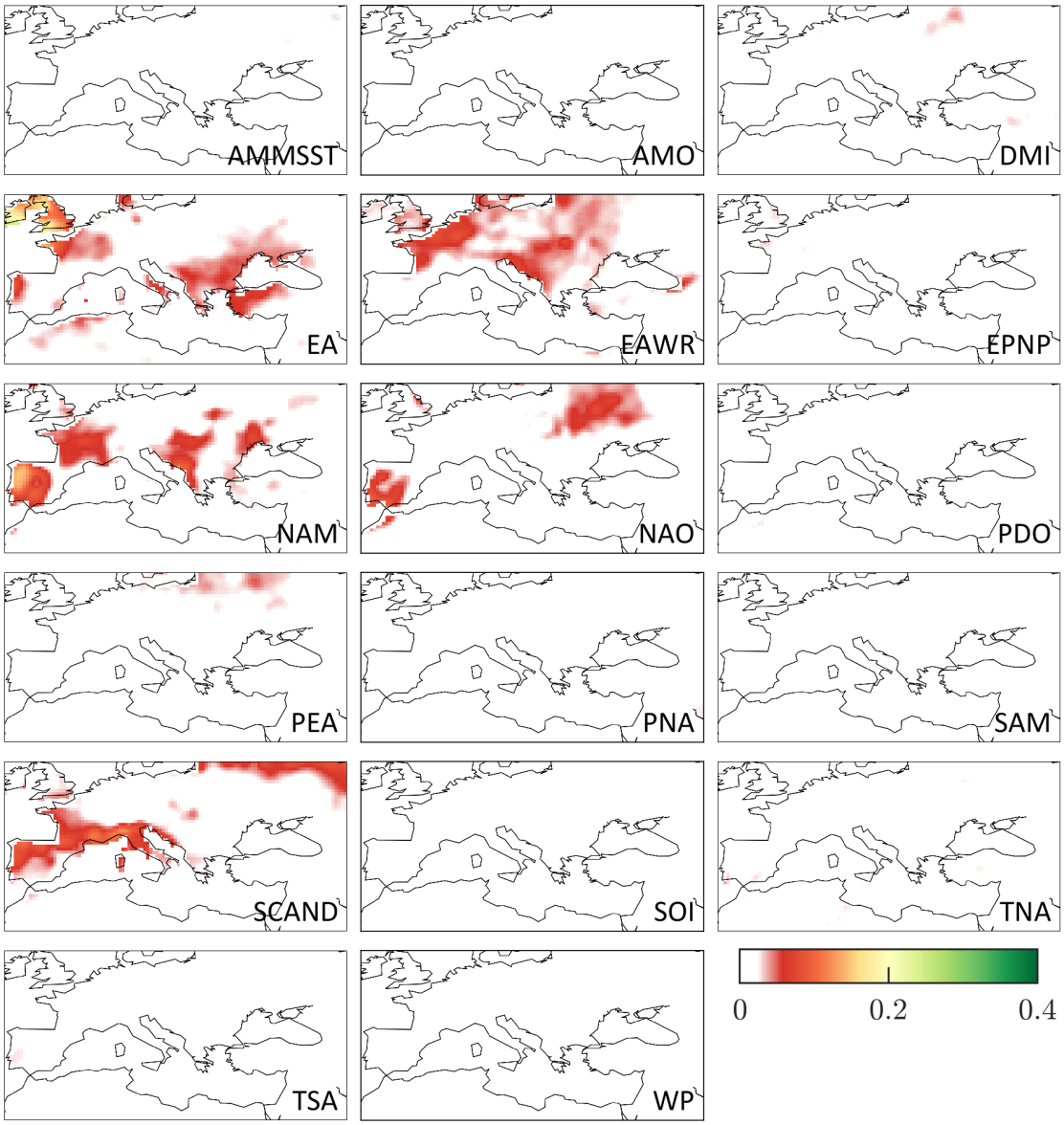

Figure 12: Impact of individual climate modes on precipitation anomalies R^2_{COI} for the model using all months of the year with additional neighborhood information.

correspond to the patterns seen in the winter and spring model. The patterns identified in Ireland and the United Kingdom are also seen in the winter, spring and autumn model.

Especially, precipitation anomalies in the northern part of France, Belgium, the Netherlands, as well as in the eastern coast of the Adriatic Sea are driven by the EAWR. The regions influenced by the EAWR correspond to the ones in the winter and autumn model.

The regions where precipitation anomalies are driven by the NAM are the same as in the winter or autumn model. In particular, the Iberian Peninsula and France, but also the eastern coast of the Adriatic Sea show a high response.

The NAO has an impact in the Iberian Peninsula, as well as in the north-eastern part of the extended Mediterranean area. A high explained variance R^2_{COI} at the Iberian Peninsula is also seen in the winter model, whereas the impact in the north-eastern part of the extended Mediterranean area is detected in the autumn model.

Precipitation anomalies in the northern Iberian Peninsula and France are driven by the SCAND. Precipitation anomalies in these regions show the same response as in the autumn model.

4.2.2 Winter model

Figure 13 shows the explained variance for each COI R^2_{COI} (4.2) for the winter model. The explained variance of each individual COI R^2_{COI} (Figure 13) is considerably lower than the explained variance R^2 of the complete model (Figure 10a). This can be explained by the now missing interaction of the individual COIs. As already mentioned in section 1.3 the LASSO regression is a multivariate method, having the unique characteristic of considering all features simultaneously. Due to the interaction of individual features, the model improves, resulting in a high coefficient of determination R^2. However, the partial R^2_{COI} values of each COI shown in Figure 13 are lower. The LASSO model is created based on the information of all features, but only the regression coefficients $\hat{\beta}$ of a single feature are used to calculate the coefficient of determination R^2_{COI}. Moreover, since the model is based on all features and cross-correlation is taken into account the partial R^2_{COI} values get even smaller compared to correlation results which do not take cross-correlation into account.

Nevertheless, the regions in which a particular COI has a significant impact can be identified and the received results are reasonable and most likely more accurate than correlation results.

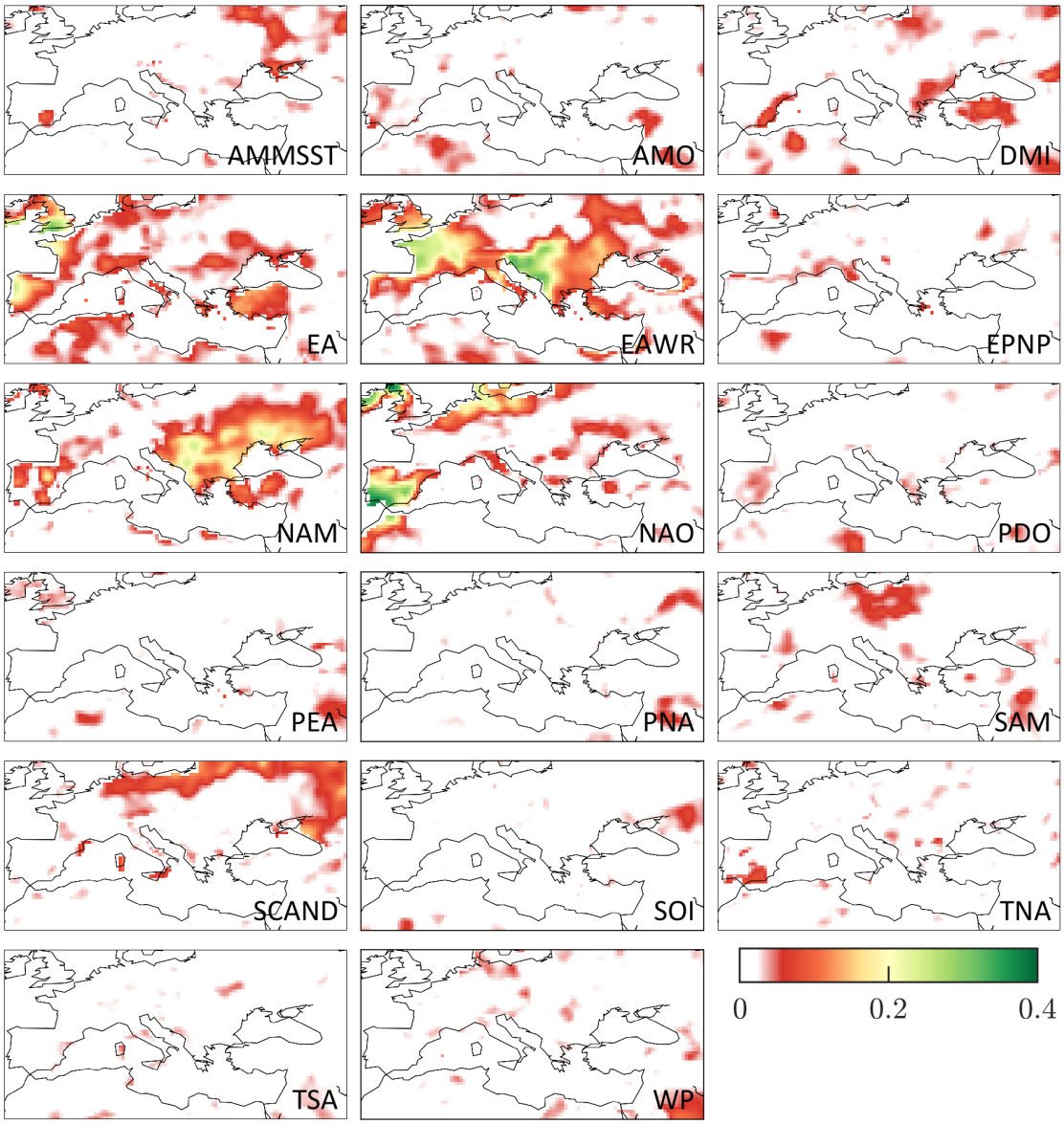

Figure 13: Impact of individual climate modes on precipitation anomalies R^2_{COI} for the winter model with additional neighborhood information. The respective abbreviations of each climate mode are written in the bottom right corner.

Figure 13 reveals five COIs that have a significant impact on precipitation anomalies in winter, i.e. the East Atlantic Pattern (EA), the East Atlantic West Russia Pattern (EAWR), the Northern Annular Mode (NAM), the North Atlantic Oscillation (NAO), and the Scandinavian Pattern (SCAND).

Based on the LASSO model the EA shows the highest impact in the southern United Kingdom, at the western coast of France and at the western Iberian Peninsula. Up to 30% of the precipitation anomalies can be explained. A less strong influence can be seen south and west of the Black Sea. These results correspond very well to the studies conducted by Mikhailova and Yurovsky [2016] and Casanueva et al. [2014]. In both studies correlations between the EA and winter precipitation from E-OBS gridded dataset [Haylock et al., 2008] are computed over Europe. The identified areas of highest influence are very similar to our results, moreover, they correspond quite well to the areas used by the definition of the EA index, which is based on normalized 500 hPa height anomalies at three specific centres, one located southwest of the Canary Islands, another west of Great Britain, and a third near the Black Sea, see section 2.1.4.

Furthermore, the LASSO result shows that the EAWR explains a lot of variance in France and the eastern coast of the Adriatic Sea. These findings agree well with the results shown by Ionita [2014] and Casanueva et al. [2014]. Both studies investigated the relationship between winter precipitation and the EAWR, showing high negative correlations up to 0.5, i.e. a R^2 of 0.25, over central Europe and the Balkans. Moreover, the patterns correspond well to the areas used by the definition of the EAWR, which is defined as an east-west pattern with three centers, one near England and Denmark, another northeast the Caspian Sea, and the third in northeast China, see section 2.1.5.

In addition, with the help of the LASSO regression, it is found that NAM drives especially precipitation anomalies in the eastern investigated area and to a less extent parts of the Iberian Peninsula. The study of Kryzhov and Gorelits [2015], as well as the study of Givati and Rosenfeld [2013] are confirmed by our findings by detecting a significant dependence of precipitation anomalies on the NAM. Both studies show strong negative correlations up to 0.7 between the NAM and wintertime precipitation in the same areas.

Furthermore, the results of the LASSO regression show that the NAO has the highest explained variance R^2_{COI} of all climate modes and is, therefore, the dominant climate mode, especially in the southern Iberian Peninsula, Ireland, and Scotland. These

results are reasonable when looking at its definition in section 2.1.8. Several studies describe the relationship between the NAO and precipitation. Mariotti and Arkin [2007] analyzed the behavior between the NAO and winter precipitation using different precipitation datasets. All results show strong positive correlations in the northern Mediterranean, whereas over the southern Mediterranean strong negative correlations are detected. These patterns correspond well to the patterns that we reveal.

Finally, the SCAND is detected by the LASSO regression as a driver of precipitation anomalies of the northern investigated area, however its influence is weaker than that of EA, EAWR, NAM or NAO. The correlation map in the study of Comas-Bru and McDermott [2013] between winter precipitation from the CRU and the SCAND is similar to the pattern seen in our study.

All other climate modes show no strong significant influence individually but are useful for possible interactions with other modes as shown in Figure 14. Figure 14a illustrates the coefficient of determination R^2 when using only the five most significant COIs listed above while Figure 14b illustrates the same by using all 17 COIs.

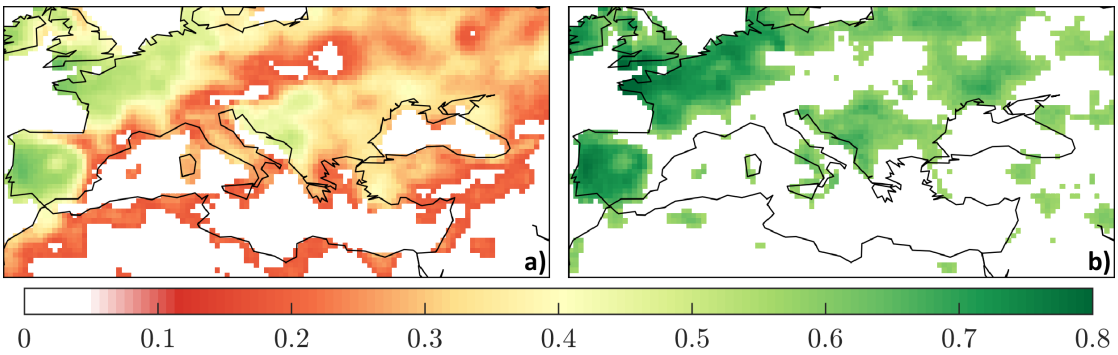

Figure 14: Explained variance R^2 for the winter model including the 3×3 neighborhood. a) LASSO model is trained by using only EA, EAWR, NAM, NAO, and SCAND, b) LASSO model is trained by using all 17 COIs

4.2.3 Spring model

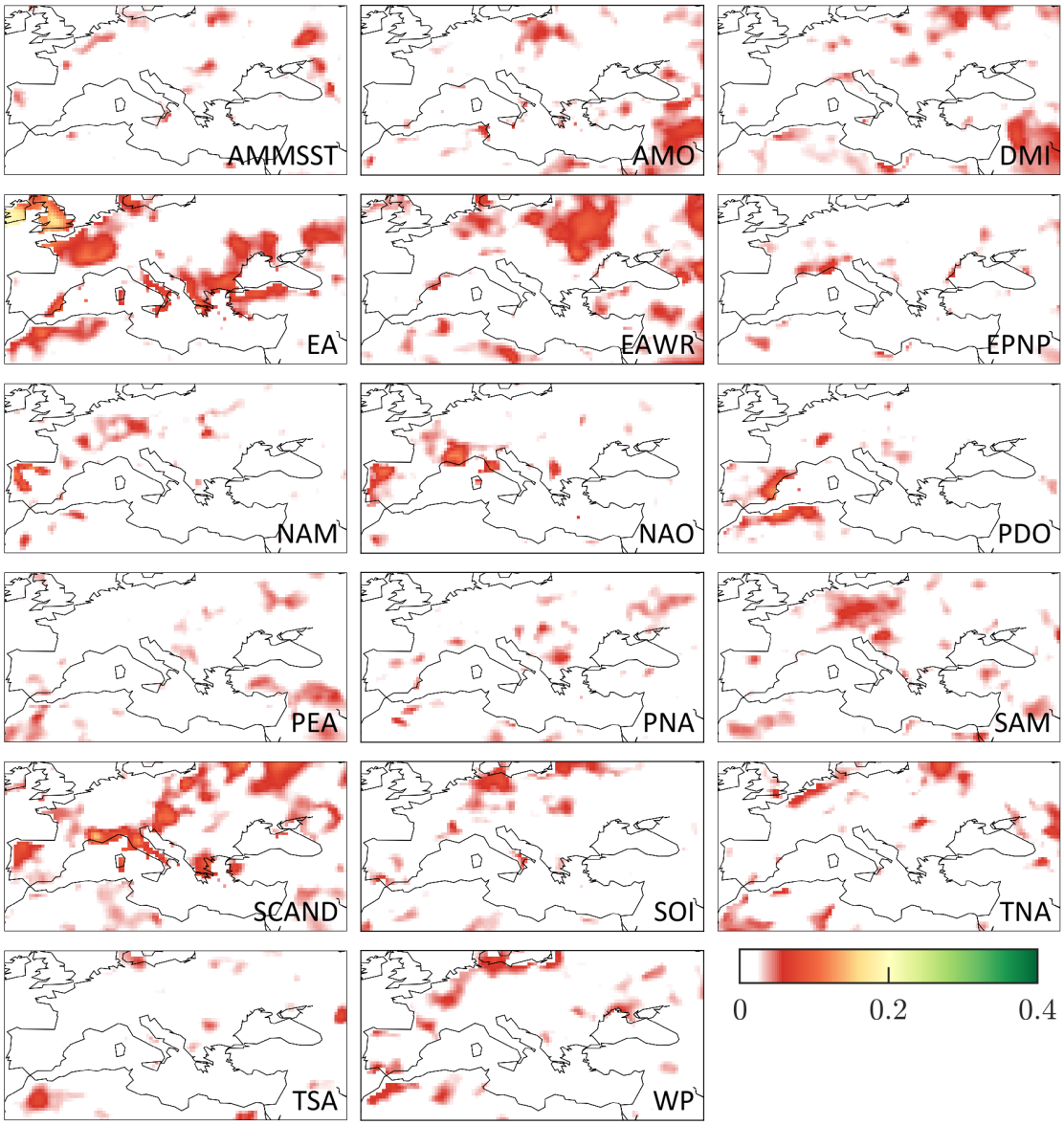

Figure 15: Impact of individual climate modes on precipitation anomalies R^2_{COI} for the spring model with additional neighborhood information.

Figure 15 shows the influence of each individual COI on precipitation anomalies for the spring model, using data of March, April and May. During this time of the year the impact of climate modes on precipitation anomalies is rather low. This manifests in a low

explained variance R^2_{COI}. Only the EA index shows a raised coefficient of determination R^2_{COI} over Ireland, the United Kingdom, France, parts of Italy and the areas west of the Black Sea. The precipitation anomalies in the eastern part of the investigated area seem to be affected to a low extent by the EAWR, whereas precipitation anomalies over Italy and the southern part of France are driven by the SCAND. All the other COIs do not show any strong patterns and, therefore, have no influence during spring.

4.2.4 Summer model

In the summer model which is trained using data of June, July and August three climate modes are detected to be the drivers of precipitation anomalies, i.e. EA, NAO and SCAND. This can be seen in Figure 16. However, in general, the impact of the individual COIs is rather low compared to the winter model shown in Figure 13. The EA is showing an influence over Ireland, the northern part of the United Kingdom and Italy. The NAO is driving the precipitation anomalies in the northern part of the extended Mediterranean area, including Ireland, the southern part of the United Kingdom, Belgium, the Netherlands and the very northern Germany. In contrast, the SCAND drives the precipitation anomalies over Spain, northern Italy, Poland and Czech Republic. The explained variance R^2_{COI} of the other COIs is close to zero, thus, no impact of these COIs is detected during summer.

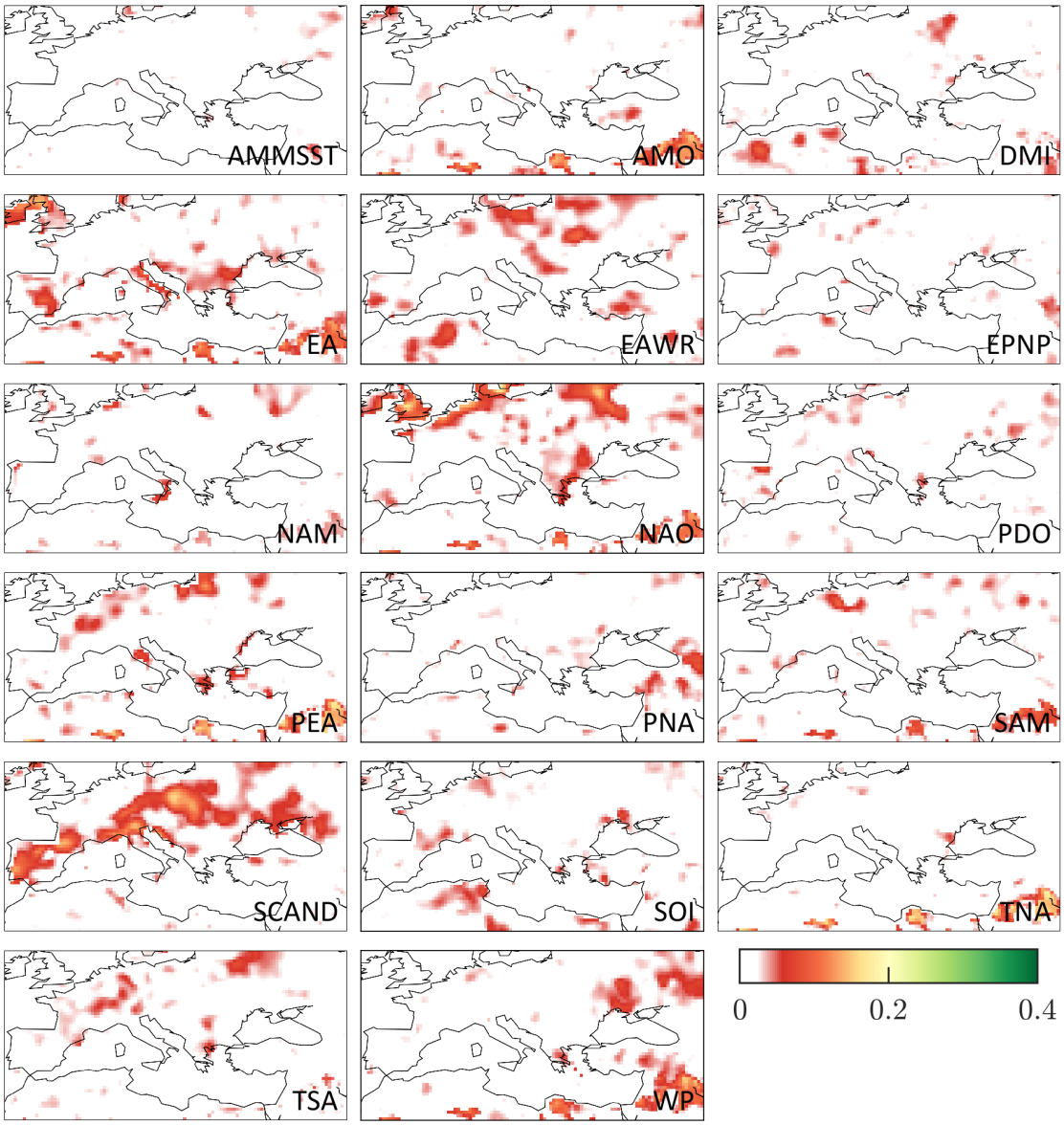

Figure 16: Impact of individual climate modes on precipitation anomalies R^2_{COI} for the summer model with additional neighborhood information.

4.2.5 Autumn model

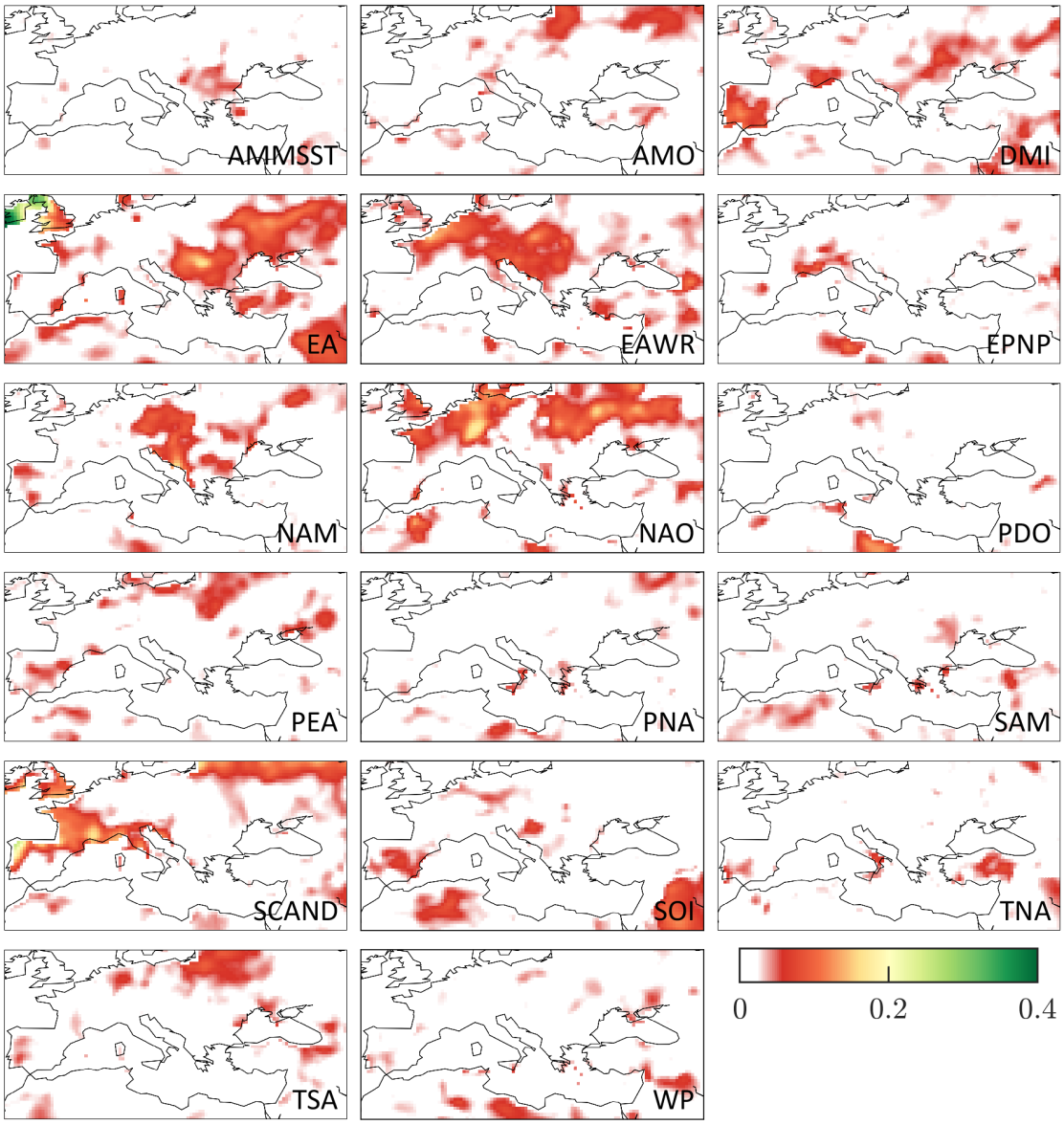

Figure 17: Impact of individual climate modes on precipitation anomalies R^2_{COI} for the autumn model with additional neighborhood information.

The autumn model which is trained using data of September, October and November has the second highest average coefficient of determination R^2 after the winter model, which can be seen in Figure 10. Figure 17 displays the impact of the individual climate

modes during autumn. The same COIs are driving precipitation anomalies as in the winter model shown in Figure 13. However, the patterns are slightly different and the explained variance for each COI R^2_{COI} is lower. The strongest influence occurs in Ireland and Scotland and is caused by the EA. Additionally, precipitation anomalies in regions west of the Black Sea are driven by this climate mode. The EAWR, in contrast, has its influence at the eastern coast of the Adriatic Sea and in the northern parts of France, Belgium and the Netherlands. The NAM is driving the eastern coast of the Adriatic Sea too, but also the eastern part of Germany and the western part of Poland. The NAO shows an impact, especially, in the northern part of the investigated area, whereas SCAND influences precipitation anomalies over the northern Iberian Peninsula, France and the southern United Kingdom. The other COIs do not have any impact on the investigated area during autumn.

4.3 Analysis over single location in the Iberian Peninsula

A LASSO model is fitted for each grid point individually. To better understand how the LASSO model works and what the coefficient of determination R^2 shows, the results of one randomly chosen grid point are discussed in more detail. The displayed grid point is located at a latitude of 38.75°N and 7.25°W. Figure 18 shows its location on a map. It lies in the south-western region of the Iberian Peninsula.

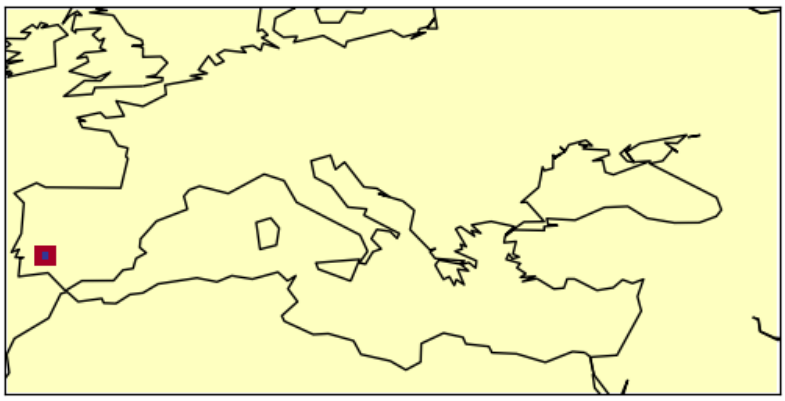

Figure 18: Location of the LASSO regression model that is discussed in more detail.

4.3.1 Without neighborhood information

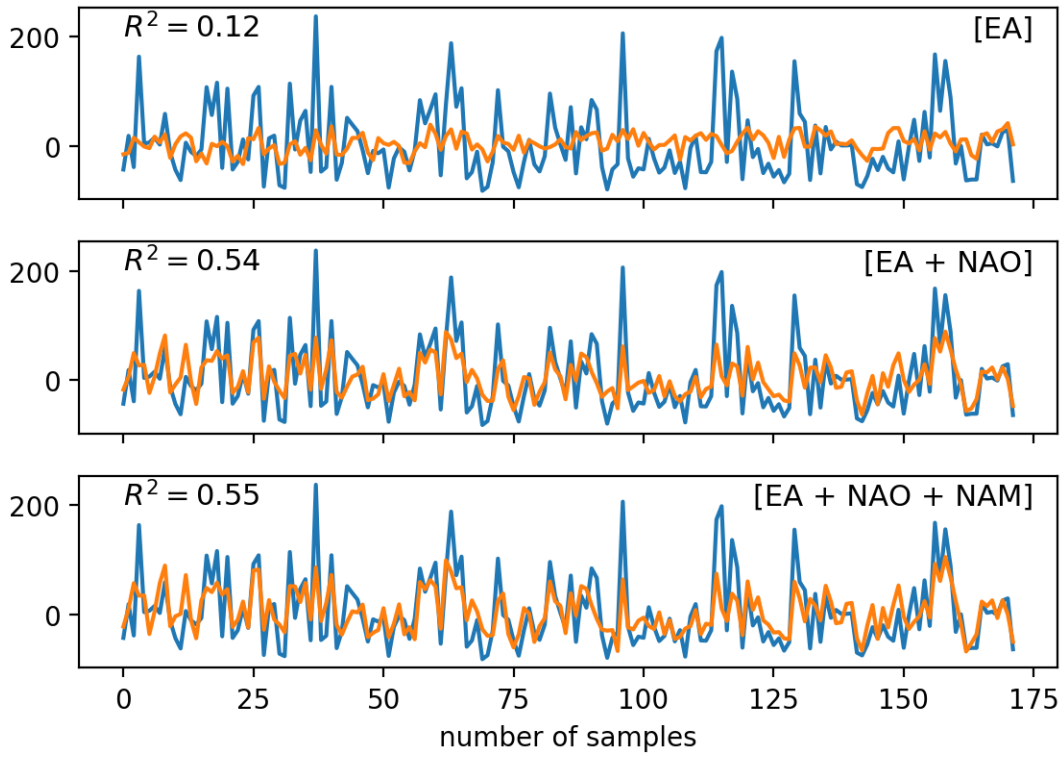

Figure 19: Step-wise addition of COIs building the LASSO winter model for precipitation anomalies at 38.75°N and 7.25°W. Blue: target variable y, orange: predicted target variable y_{pred}. Only three COIs are selected by the LASSO regression model.

Figure 19 shows the results of a step-wise addition of COIs to the LASSO regression model. It is calculated with the help of equation (3.1). The LASSO model is trained by using all 17 COIs. In case of the no neighborhood model, for this specific grid point only three features, namely EA, NAO, and NAM are used by the automatic feature selection of the LASSO regression to build the final model. The result of the LASSO regression are the regression coefficients $\hat{\beta}$ of the features x. The prediction y_{pred} can be build based on the calculated regression coefficients $\hat{\beta}$ as shown in equation (3.3).

The first plot in Figure 19 shows the result, if only the regression coefficients $\hat{\beta}$ from the full LASSO regression model of one feature, namely EA, would be used to predict the precipitation anomalies. The corresponding coefficient of determination would result

in $R^2 = 0.12$. By adding the regression coefficients $\hat{\beta}$ of the second feature, i.e. NAO, the coefficient of determination would increase considerably to $R^2 = 0.54$. Adding the third feature, i.e. NAM, shows the final LASSO result with a R^2 of 0.55. The patterns of the predictions y_{pred} of the final LASSO model follow the temporal patterns of the target variable y well but the amplitude is lower.

In contrast, Figure 20 shows the impact of each individual COI without interactions with other COIs. It is clearly visible that taking only the regression coefficients $\hat{\beta}$ from the full LASSO regression model of one COI at a time leads to a significantly worse prediction and, therefore, low explained variance R^2. While the R^2 of the final LASSO regression is 0.55, the highest possible obtained R^2 by only using the regression coefficients $\hat{\beta}$ from the full LASSO regression model of one feature at a time is 0.40. The same would occur by using correlations instead of a LASSO regression because a correlation also only takes one feature into account.

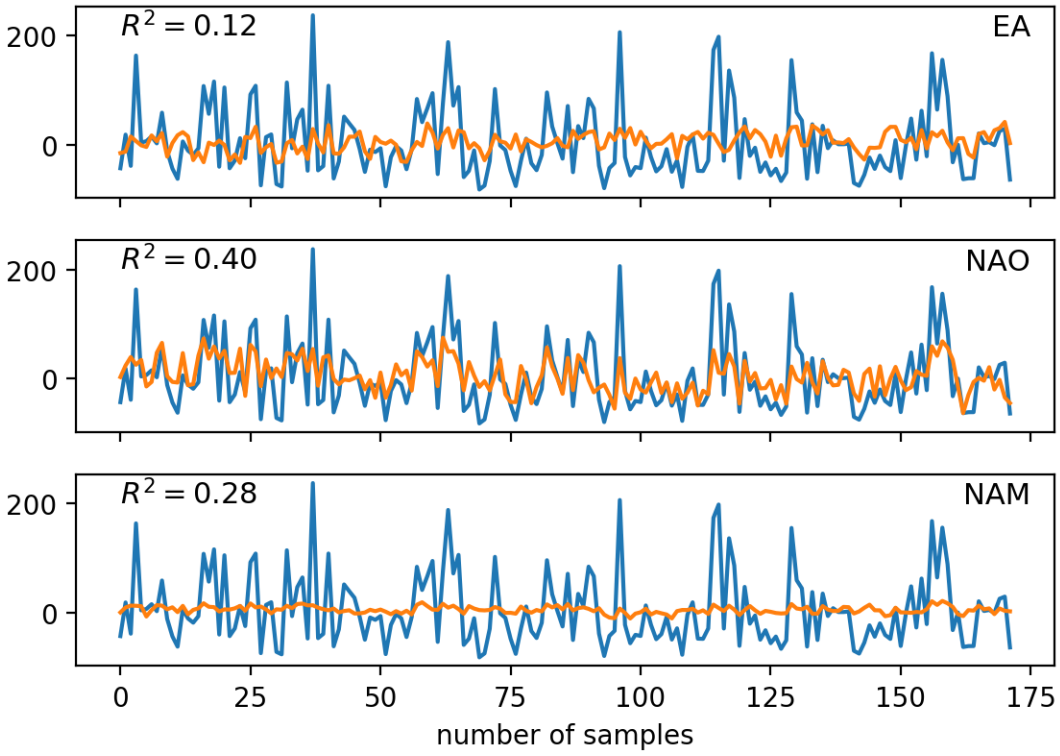

Figure 20: Individual COIs multiplied with its coefficients $\hat{\beta}$ from the LASSO winter model at 38.75°N and 7.25°W. Blue: target variable y, orange: predicted target variable y_{pred}

The possibility to analyze all features simultaneously and, therefore, also account for their cross-correlations and interactions, is one of the biggest benefits of the LASSO regression.

4.3.2 With neighborhood information

Figure 21 shows the same step-wise addition of individual features as Figure 19 but includes the 3×3 neighborhood. By adding the neighborhood information the LASSO regression uses a lot more features to build its prediction model. Therefore, three to four features are added simultaneously in Figure 21 to reduce the number of necessary plots. The final LASSO regression model is the last plot in Figure 21. It is clearly visible, that the final result including the neighborhood corresponds better to the target variable y as it is the case in Figure 19. This time the temporal patterns and amplitudes are represented better. This can also be verified by the higher explained variance R^2 of 0.77. However, the contribution per COI is lower as shown in Table 2.

COI	R^2	COI	R^2	COI	R^2	COI	R^2
AMMSST	0.00	EPNP	0.00	PEA	0.00	SOI	0.00
AMO	0.03	NAO	0.31	PNA	0.00	TNA	0.00
DMI	0.01	NAM	0.00	SAM	0.00	TSA	0.01
EA	0.13	PDO	0.02	SCAND	0.00	WP	0.00
EAWR	0.00						

Table 2: Explained variances by each individual COI R^2_{COI} for the grid point at 38.75°N and 7.25°W including the neighborhood.

Table 2 shows the explained variance by each COI R^2_{COI} for the single grid point at 38.75°N and 7.25°W including the neighborhood. It can be seen that the explained variance for one individual COI R^2_{COI} is rather low, with $R^2_{COI} = 0.31$ being the highest for the NAO. However, the coefficient of determination R^2 of the final LASSO model is significantly higher with a value of 0.77, as it can be seen in Figure 21. This can be explained by the interaction of the individual COIs. Even if one COI alone has a low explained variance R^2_{COI}, it still contributes to the accuracy of the final LASSO regression model. Thus, the explained variance R^2 of the final LASSO model is higher than the one for each individual COI.

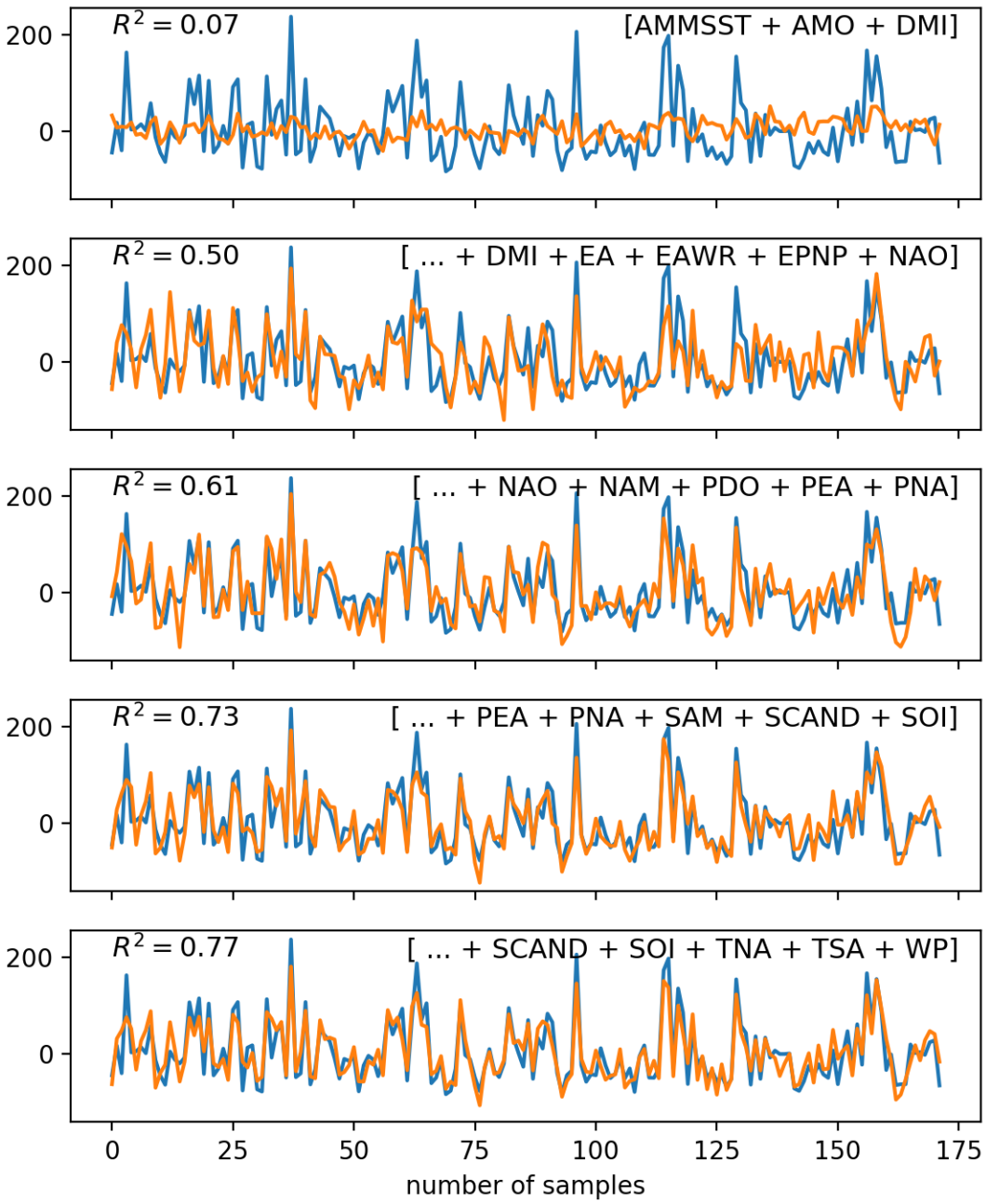

Figure 21: Step-wise addition of COIs building the LASSO model for precipitation anomalies at 38.75°N and 7.25°W including the neighborhood. Blue: target variable y, orange: target variable y_{pred}.

Although this is only the result of one grid point, the same can be seen at nearly every location. As shown in Figure 9 and when comparing Figure 10 with 11 the explained variance R^2 is almost always higher when adding the information of the neighborhood.

4.4 Effect of cross-validation on coefficient of determination

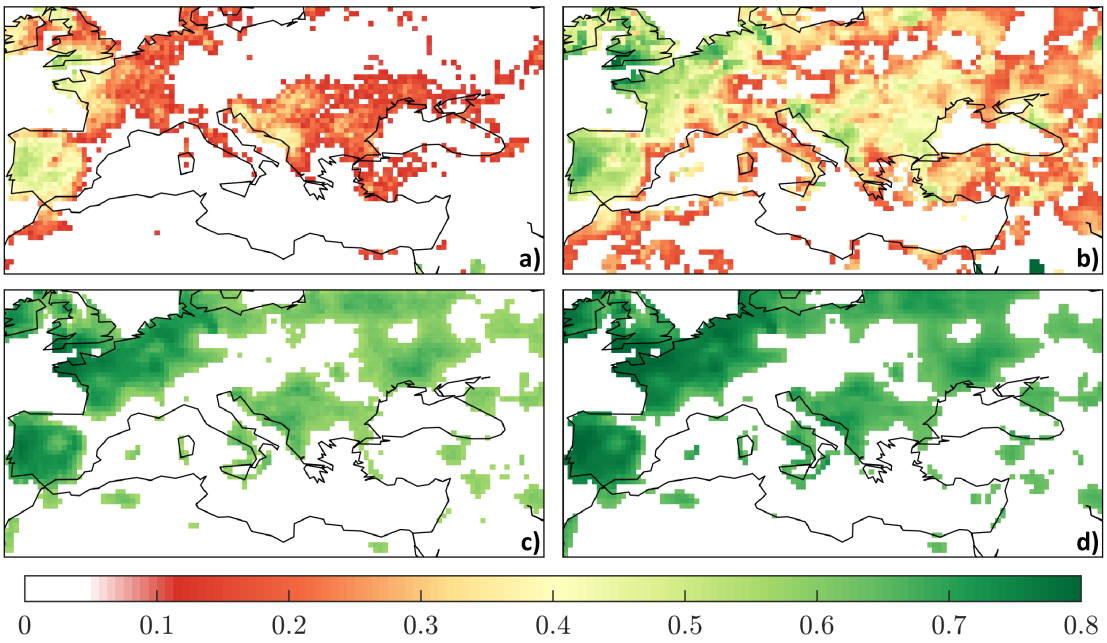

Figure 22: Comparison of the effect of five-fold cross-validation on R^2 for the winter model. a) R^2 with cross-validation without neighborhood, b) R^2 without cross-validation without neighborhood, c) R^2 with cross-validation with neighborhood, d) R^2 without cross-validation with neighborhood

Figure 22 compares the results gained with and without cross-validation for the winter model. The first row shows the explained variance R^2 without neighborhood, the second row with neighborhood. The first column is with cross-validation and the second is without cross-validation. The results without five-fold cross-validation have a significantly higher explained variance R^2 than the ones with cross-validation. Although the explained variance R^2 gets higher, the significance test marks roughly the same areas as significant. Only areas where the explained variance R^2 gets too low, like the northern parts of Europe without using the neighborhood information, get additionally marked as non-significant.

Cross-validation is a model validation technique that estimates the model's predictive performance. To validate the LASSO regression model, a five-fold cross-validation is applied for the coefficient of determination R^2 (section 3.2). This is done to avoid effects of data-overfitting. Without cross-validation the model would be based on 100% of the data and it is more likely that overfitting would occur. By using a cross-validation, the explained variance R^2 is based on predictions and, thus, it is unlikely that strong overfitting occurs.

Furthermore, the results are more reasonable using a cross-validation. If the model is fitted to 100% of the data, as it is the case without cross-validation, the model parameters are optimized fitting this data as good as possible. However, it is not clear how accurately the predictive model will perform on an independent data set. Therefore, a cross-validation needs to be applied to show a more realistic model.

4.5 Comparison of LASSO regression with correlation

A correlation coefficient is a very common metric used to determine the relationship between two variables. There are several correlation coefficients that are assuming different data distributions. In this study, the so-called pearson correlation is used, that measures the linear relationship between two variables. It assumes two normally distributed datasets. The pearson correlation coefficient ranges between -1 and 1. High positive coefficients indicate that the two tested variables show the same behaviour, whereas high negative coefficients indicate an opposite behavior of the two tested variables.

Correlations between climate modes and precipitation have been computed in several studies [Steirou et al., 2017; Comas-Bru and McDermott, 2013; Bladé et al., 2012; Mariotti and Arkin, 2007].

As the winter model shows the strongest impact on precipitation anomalies, correlations are computed between climate modes and winter precipitation anomalies, using data of December, January, and February.

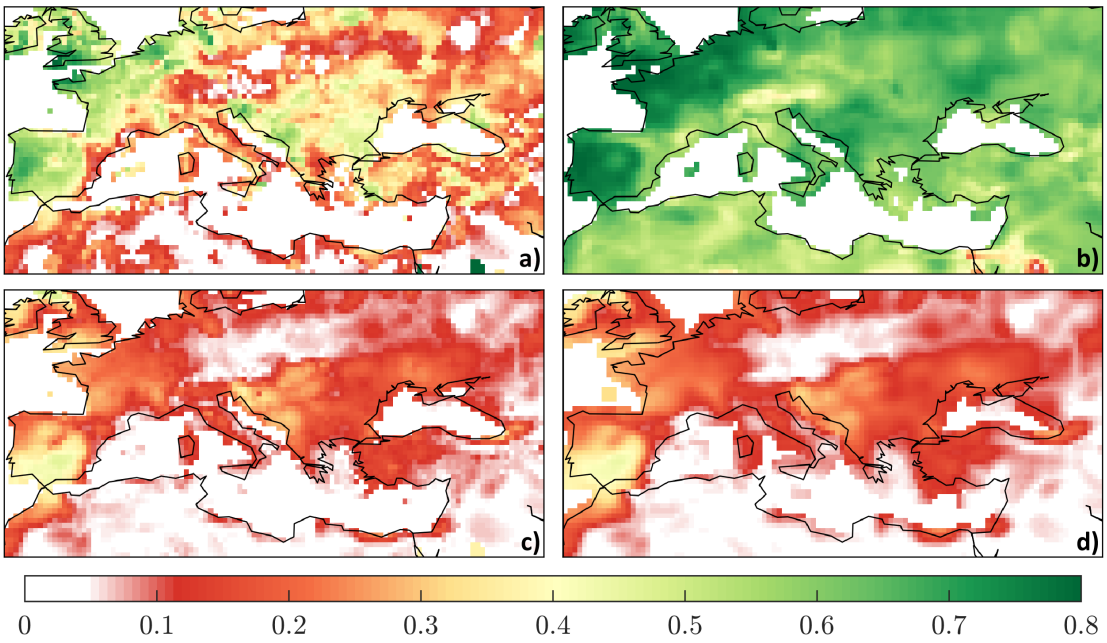

Figure 23: Comparison of the explained variance R^2 of LASSO regression and maximum squared correlation coefficients $corr^2$ for the winter model. No cross-validation and no significance test is applied. a) R^2 without neighborhood, b) R^2 with neighborhood c) $corr^2$ without neighborhood, d) $corr^2$ with neighborhood

Figure 23 compares the LASSO regression without cross-validation and without significance test and correlation results of the winter model with and without using the 3×3 neighbourhood. As already discussed for Figure 9 the explained variance R^2 gained by the LASSO regression gets significantly higher when the 3×3 neighbourhood is used.

The correlation plots in Figure 23 show the maximum squared correlation over all COIs and lags

$$corr^2 = \max_{COI} \left\{ \max_{lag} \left\{ corr(x_{COI_{lag}}, y)^2 \right\} \right\} \tag{4.3}$$

where x are the COI anomalies with a specific time lag and y are the precipitation anomalies.

In contrast to the LASSO regression, the correlation coefficients show no improvement when including the 3×3 neighbourhood. The result remains the same, apart from the smoothing effect resulting from the 3×3 neighbourhood.

To compare the LASSO results with the correlation coefficients in the best possible manner, all 17 correlation coefficients of the individual climate modes are combined. This

means in detail that in each case the maximum correlation coefficient of the 17 COIs is taken per grid point. Hence, comparable maps are created. Since no significance test and no five-fold cross-validation is applied for the correlation results, the LASSO results are also shown without significance test and cross-validation. Nevertheless, it is obvious that the LASSO regression provides better results than the correlation, especially, if the 3×3 neighborhood is included.

4.5.1 Without neighborhood information

In order to compare the LASSO regression results with correlations in more detail, Figure 24 and 26 were created.

The correlations are computed for each COI for six time lags.

$$corr^2_{COI} = \max_{lag} \left\{ corr(x_{COI_{lag}}, y)^2 \right\} \tag{4.4}$$

The square of the highest correlation coefficient per grid point $corr^2_{COI}$ is displayed to obtain comparable results similar to the explained variances R^2_{COI} of the LASSO regression shown in Figure 25 and 13.

The impact of each COI on precipitation anomalies R^2_{COI} for the winter model without using the 3×3 neighbourhood shown in Figure 25 is almost the same as the correlation results shown in Figure 24.

The climate modes EA, EAWR, NAM, NAO, and SCAND show high correlation coefficients (Figure 24). These COIs are the same as the ones identified by the LASSO regression (Figure 25).

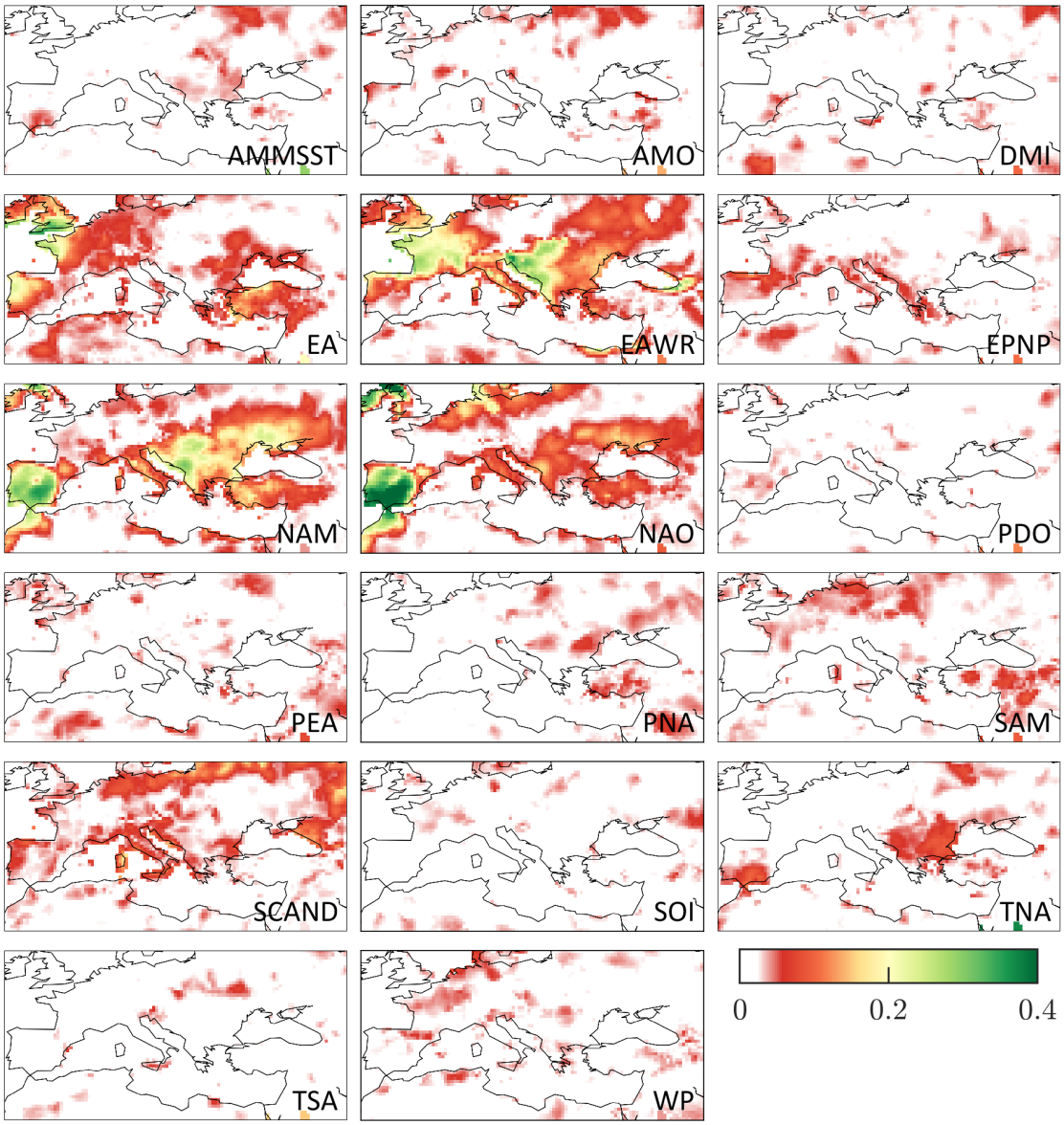

Figure 24: Squared correlations between climate modes and winter precipitation anomalies $corr^2_{COI}$ for winter model without using the neighborhood.

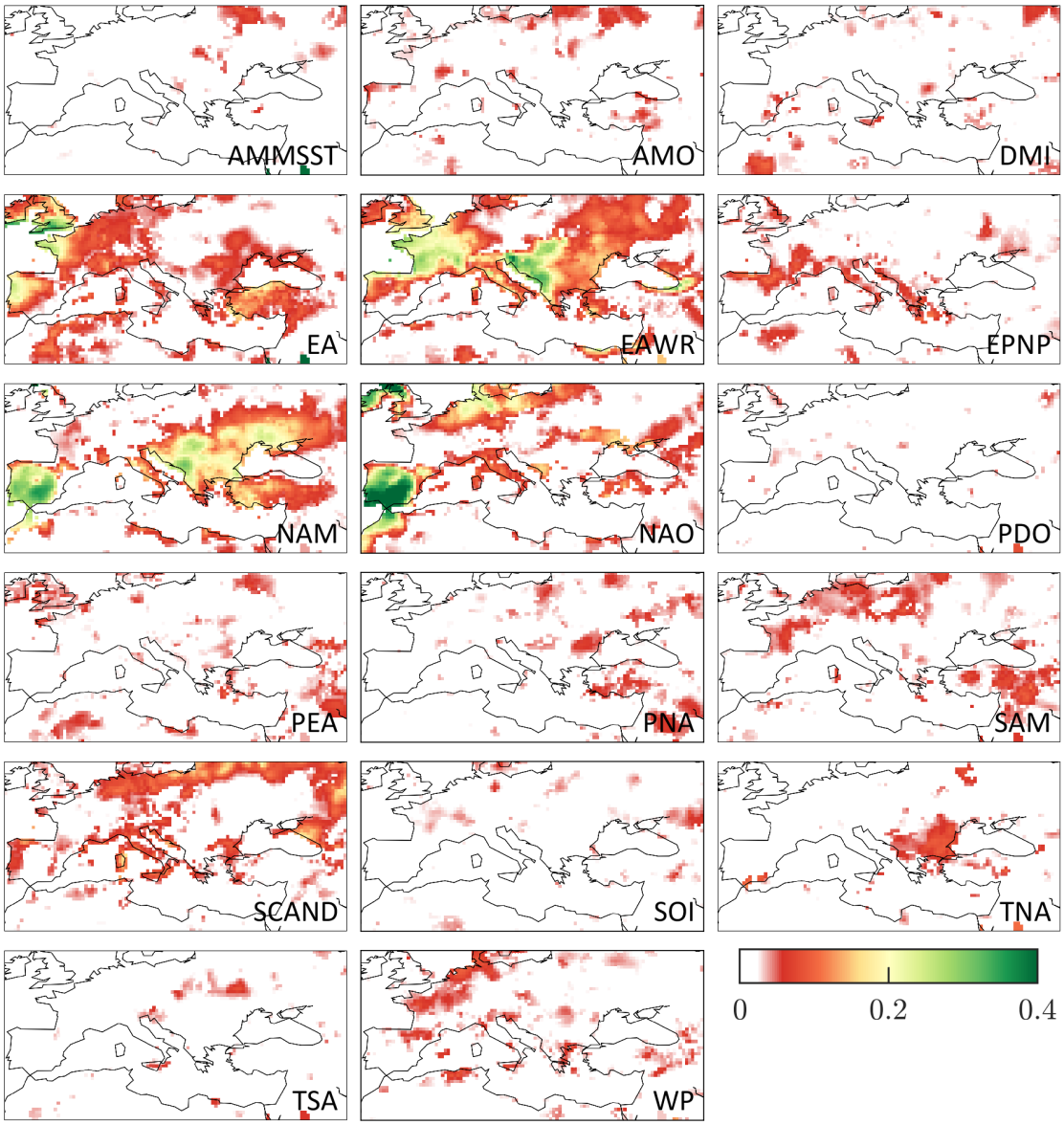

Figure 25: Impact of individual climate modes on precipitation anomalies R^2_{COI} for the winter LASSO model without using the neighbourhood.

4.5.2 With neighborhood information

Comparing Figure 24 and 26, which show the squared correlations with and without including the neighborhood, no differences can be seen. Consequently, adding the neighborhood information does not make a difference for correlation.

In contrast to the correlation, the final model of the LASSO regression benefits greatly by including the information of the 3×3 neighborhood, as shown in Figure 23. However, the explained variance for each COI R^2_{COI} does not increase as shown when comparing Figure 25 and Figure 13.

Since correlation indicates the strength of association between variables, extending the time series does only help by reducing the noise of the computation. Unlike the LASSO regression, where the model is derived and optimized using the information of all features simultaneously, the correlation coefficient still relies only on the two tested variables and no additional information can be added.

However, the squared correlation coefficients including the 3×3 neighbourhood, shown in Figure 26 are stronger than the explained variances R^2_{COI} of the LASSO regression shown in Figure 13. This can be explained by the interaction of the individual COIs which is not considered for the correlation and thus leads to higher correlation coefficients. Additionally, the results gained by the LASSO regression are computed using a five-fold cross-validation for the coefficient of determination R^2, meaning that only 80% of the data is used at once, while the correlation is using 100% of the available data without any cross-validation. Moreover, a significance test is applied in case of the LASSO regression. A more detailed explanation concerning the consideration of the cross-correlations and the significance test can be found in section 3.

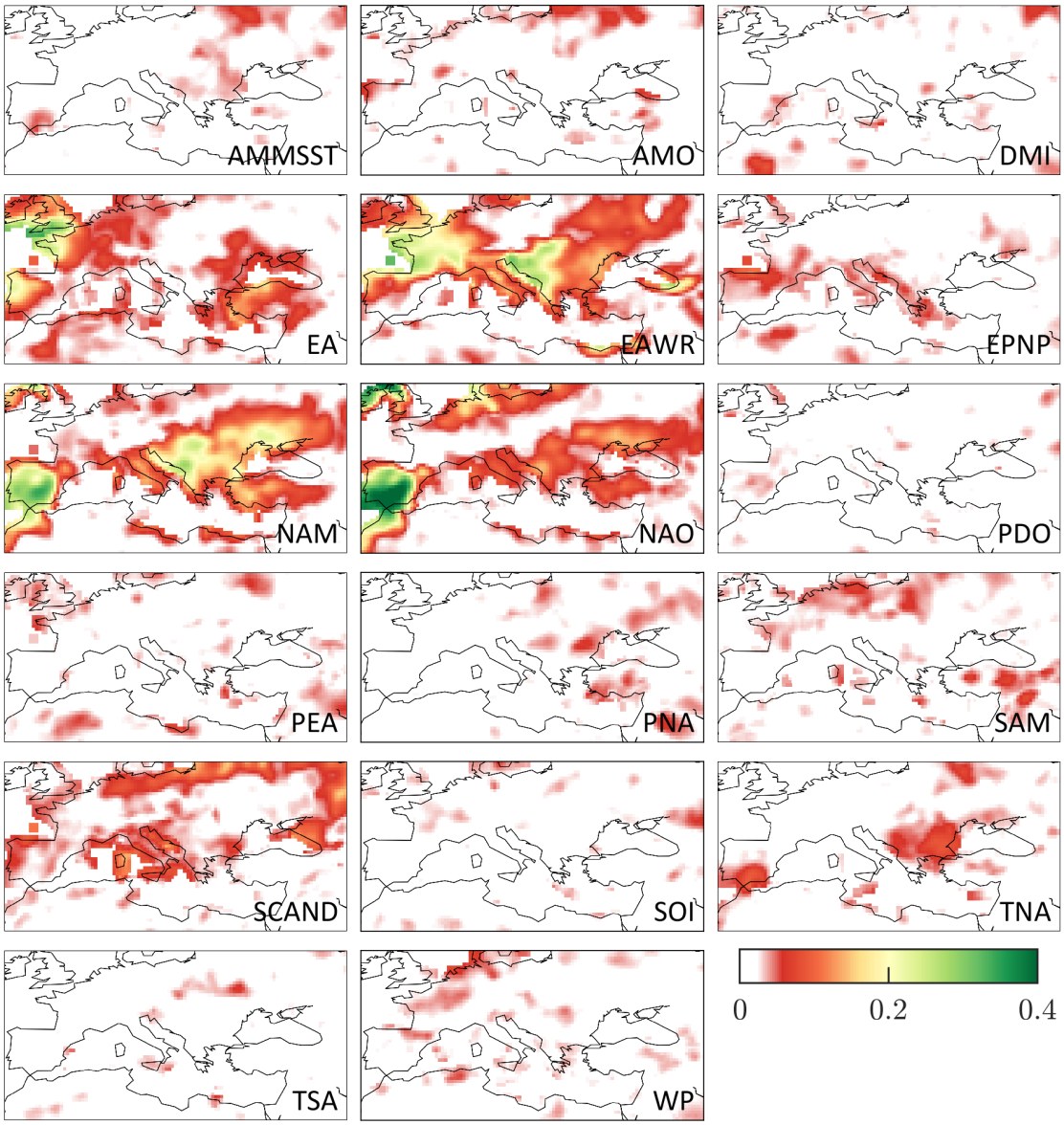

Figure 26: Squared correlations between climate modes and winter precipitation anomalies $corr^2_{COI}$ for winter model including the neighborhood.

4.5.3 Detailed analysis at the Iberian Peninsula

As already mentioned in section 3.1 one improvement of the LASSO regression model over correlation is, that the influence of multiple COIs is calculated simultaneously and cross-correlations between COIs are taken into account. Therefore, the result is expected to be more accurate and the real impact of individual climate modes can be shown better. This is explained by looking at the study of Comas-Bru and McDermott [2013] where correlations between precipitation of CRU TS v3.1 and the climate modes NAO, EA, and SCAND are computed. Their results show that the NAO and the EA have a high correlation over the Iberian Peninsula, while the SCAND seems to have a low influence in this area.

These results are different from the patterns found by the LASSO regression (Figure 13). The LASSO result shows that precipitation anomalies only in the southern part of the Iberian Peninsula are dominated by the NAO, whereas precipitation anomalies only in the western part are driven by the EA, and the SCAND is not driving precipitation anomalies at all in this area. Precipitation anomalies in the northern part of the Iberian Peninsula is mainly driven by the EAWR and the NAM.

The reason for the differences between the correlation results and the LASSO regression result is the cross-correlation of individual COIs. For example, the correlation between NAO and NAM is 0.75 during the winter months. By calculating correlations between NAO and precipitation anomalies, and NAM and precipitation anomalies, both results would show a high correlation coefficient in the northern part of the Iberian Peninsula. In contrast, the LASSO regression uses feature selection to only select the feature with the highest predictive power, in case of those cross-correlated features. Therefore, in our results, only the NAM shows an impact in the northern part of the Iberian Peninsula and the NAO does not. This confirms that the LASSO regression is able to disentangle individual climate modes in a more sophisticated manner.

5 Conclusion

In this study, the impact of 17 climate modes on precipitation anomalies over an extended Mediterranean area is analyzed. The results show that the variability in precipitation anomalies are strongly affected by ocean-atmospheric oscillations.

A supervised learning approach, called LASSO regression (section 3.1) is used to disentangle the impact of individual climate modes. The analysis is performed in two ways, once by analyzing the time series of each grid point individually and, furthermore, by adding the information of the 3×3 neighborhood, see section 3.4.2. If the LASSO framework is extended by the 3×3 neighborhood, the model improves significantly, resulting in a high explained variance R^2, as it can be seen in section 4.1. Additionally, seasonal models are calculated (section 3.4.1) which lead to a significantly higher explained variance R^2 compared to the full-year model that uses all months of a year (section 4.1). Especially, the winter model shows a high coefficient of determination R^2. In some regions, up to 70% of the precipitation anomalies can be explained by the combination of specific climate modes. For the winter model, the LASSO regression detected that EA, EAWR, NAM, NAO, and SCAND have the greatest impact, meaning that these climate modes are mostly responsible for interannual precipitation anomalies, as shown in section 4.2.

The LASSO regression is a method that can definitely help revealing the dominant modes of climate variability by disentangling their complex impacts and interactions. New insights have been gained by using the LASSO regression over correlation. In general, the results from the LASSO regression correspond quite well with correlation results between climate modes and winter precipitation anomalies computed in previous studies. However, certain differences can be seen, described in section 4.5, which can be explained by the fact that the LASSO regression is considering cross-correlations between individual climate modes and uses the information of all climate modes simultaneously to create the predictive model. If several climate modes are correlated, the automated feature selection carried out by the LASSO regression selects the ones with the highest predictive power, whereas the others are set to zero, meaning that they show no influence in that area. The regularization of the LASSO regression and the fact that predictions based on cross-validation are calculated helps preventing overfitting and thus the impacts of each COI in specific regions can be detected more accurately than it is done by correlation in previous studies.

Furthermore, data quality is highly important to achieve meaningful results. This is confirmed by the findings in the analysis when including the neighborhood. As already mentioned, the coefficient of determination R^2 improves significantly by adding the neighborhood information due to the noise reduction in the data. If the data is too noisy, less information can be extracted and it gets difficult to build a good regression model. Moreover, long time series are absolutely necessary to build a good and sophisticated model. Especially for machine learning algorithms the length of the data plays a crucial role. The algorithms need enough training data to perform well.

Based on the results found in this study, further research can be carried out to improve the understanding of the relation between climatic oscillations and hydrology.

6 Outlook

The LASSO regression worked quite well, showing reasonable results, with a higher explained variance R^2 than correlations. Therefore, further studies should be conducted using the LASSO regression. Although a lot of investigations have been performed regarding the methodology, possible improvements or new ideas could be incorporated, making the LASSO regression perform even better.

The regularization parameter α is set by the algorithm automatically. A better understanding on how the model is affected by this parameter could be helpful to improve the model. Additionally, the estimation of the regularization parameter α and the coefficient of determination R^2 could be determined together in a single nested five-fold cross-validation.

The limits of the method regarding the length of the data or the number of features still need to be investigated further. Further analysis could be conducted to know whether the method fails at some point because of numeric reasons or if the data itself is the pivotal factor to achieve good results. This could be determined by using different data sets with different data quality.

As the noise in the data affects the results, the pre-processing of the data is important to achieve good results. Therefore, further investigations on how the data should be prepared could be carried out.

Additionally, other regression methods like elastic net regression or ridge regression could be used which might create a better predictive model and help interpreting the results. Results from other regression models could be used to validate the results found by using the LASSO regression.

Another idea would be to separate the investigated area into groups with similar hydro-climatic properties and train and use the regression models based on these groups instead of calculating a new regression model for each grid point individually. Since it is found out that adding the neighborhood information helps the LASSO regression, this would be the next logical step. This idea would also help in case of short time series with only few samples since the information of a whole region is used simultaneously for the training of the regression model.

Furthermore, the study region could be extended to the global scale, making it possible to explore the impact of large-scale climate modes on precipitation anomalies over the whole globe. Thus, hot spot regions would be highlighted, showing which climate drivers are dominant in which region of the world.

This study focused on precipitation anomalies as response variable. However, different hydrological variables, as well as vegetation variables, could be used as the target variable. These experiments are required to gain new insights in the interactions between ocean, atmosphere, land, and ultimately the biosphere. Improving this knowledge is needed to assess the potential implications of climate change on our future climate and to provide predictions about the evolution of climate and hydrology.

Appendix

Climate mode, Climate Oscillation Index (COI)	Reference
Atlantic Meridional Mode (AMM), AMM sea-surface temperature (AMMSST)	[Chiang and Vimont, 2004; Kossin and Vimont, 2007; Vimont and Kossin, 2007]
Atlantic Multidecadal Oscillation (AMO), AMO index	[van Oldenborgh et al., 2009; Enfield et al., 2001]
Indian Ocean Dipole (IOD), Dipole Mode Index (DMI)	[Saji et al., 1999; Black et al., 2003]
East Atlantic Pattern (EA), EA index	[Barnston and Livezey, 1987; Wallace and Gutzler, 1981]
East Atlantic West Russia Pattern (EAWR), EAWR index	[Barnston and Livezey, 1987]
East Pacific North Pacific Pattern (EPNP), EPNP index	[Barnston and Livezey, 1987; Bell and Janowiak, 1995]
Northern Annular Mode (NAM), NAM index	[Higgins et al., 2002, 2000; Larson et al., 2005; Zhou et al., 2001]
North Atlantic Oscillation (NAO), NAO index	[Walker and Bliss, 1932; van Loon and Rogers, 1978; Barnston and Livezey, 1987; Hurrell, 1995]
Pacific Decadal Oscillation (PDO), PDO index	[Zhang et al., 1997; Mantua et al., 1997]
Polar Eurasia Pattern (PEA), PEA index	[Barnston and Livezey, 1987; Gao et al., 2016]
Pacific-North American Pattern (PNA), PNA index	[Barnston and Livezey, 1987; Wallace and Gutzler, 1981; Chen and van den Dool, 2003; van den Dool et al., 2000]
Southern Annular Mode (SAM), SAM index	[Marshall, 2003]
Scandinavian Pattern (SCAND), SCAND index	[Barnston and Livezey, 1987]
El Niño-Southern Oscillation (ENSO), Southern Oscillation Index (SOI)	[Walker and Bliss, 1932; Chen, 1982; Ropelewski and Jones, 1987; Trenberth and Caron, 2000]
Tropical Northern Atlantic Dipole (TNA), TNA index	[Barnston and Livezey, 1987; Enfield et al., 1999]
Tropical Southern Atlantic Dipole (TSA), TSA index	[Enfield et al., 1999]
West Pacific Pattern (WP), WP index	[Barnston and Livezey, 1987; Wallace and Gutzler, 1981]

Table 3: The climate modes used in this study are listed with their corresponding COI. Additionally, the references are provided for each climate mode.

COI	Data origin
AMMSST index	http://www.esrl.noaa.gov/psd/data/timeseries/monthly/AMM
AMO index	http://climexp.knmi.nl/amo.cgi
DMI	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/DMI/
EA index	http://www.cpc.ncep.noaa.gov/data/teledoc/ea.shtml
EAWR index	http://www.cpc.ncep.noaa.gov/data/teledoc/eawruss.shtml
EPNP index	http://www.cpc.ncep.noaa.gov/data/teledoc/ep.shtml
NAM index	http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml
NAO index	http://climexp.knmi.nl/getindices.cgi?WMO=NCEPData/cpc_nao
PDO index	https://www.ncdc.noaa.gov/teleconnections/pdo/
PEA index	http://www.cpc.ncep.noaa.gov/data/teledoc/poleur.shtml
PNA index	http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/pna.shtml
SAM index	http://www.nerc-bas.ac.uk/icd/gjma/sam.html
SCAND index	http://www.cpc.ncep.noaa.gov/data/teledoc/scand.shtml
SOI	http://www.bom.gov.au/climate/current/soi2.shtml
TNA index	http://www.esrl.noaa.gov/psd/data/climateindices/list/
TSA index	http://www.esrl.noaa.gov/psd/data/climateindices/list/
WP index	http://www.cpc.ncep.noaa.gov/data/teledoc/wp.shtml

Table 4: Data origin for each COI is listed.

References

Andreoli, R. V. (2004). Multi-scale variability of the sea surface temperature in the tropical atlantic. *Journal of Geophysical Research*, 109(C5).

Barnston, A. G. and Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, 115(6):1083–1126.

Barros, V. R., Field, C. B., Dokke, D. J., Mastrandrea, M. D., Mach, K. J., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C., Girma, B., Kissel, E. S., Levy, A. N., MacCracken, S., Mastrandrea, P. R., and White, L. L. (2014). *Climate change 2014: impacts, adaptation, and vulnerability - Part B: regional aspects - Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Bell, G. D. and Janowiak, J. E. (1995). Atmospheric circulation associated with the midwest floods of 1993. *Bulletin of the American Meteorological Society*, 76(5):681–696.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Black, E., Slingo, J., and Sperber, K. R. (2003). An observational study of the relationship between excessively strong short rains in coastal east africa and indian ocean sst. *Monthly Weather Review*, 131(1):74–94.

Bladé, I., Liebmann, B., Fortuny, D., and van Oldenborgh, G. J. (2012). Observed and simulated impacts of the summer nao in europe: implications for projected drying in the mediterranean region. *Climate Dynamics*, 39(3):709–727.

Blondel, J. and Aronson, J. (1999). *Biology and Wildlife of the Mediterranean Region*. Oxford University Press.

Bonan, G. (2015). *Global Cycles*, page 40–58. Cambridge University Press, 3 edition.

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

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V. (2018). Observed fingerprint of a weakening atlantic ocean overturning circulation. *Nature*, 556(7700):191–196.

Casanueva, A., Rodríguez-Puebla, C., Frías, M. D., and González-Reviriego, N. (2014). Variability of extreme precipitation over europe and its relationships with teleconnection patterns. *Hydrology and Earth System Sciences*, 18(2):709–725.

Chang, P., Ji, L., and Li, H. (1997). A decadal climate variation in the tropical atlantic ocean from thermodynamic air-sea interactions. *Nature*, 385(6616):516–518.

Chen, W. Y. (1982). Assessment of southern oscillation sea-level pressure indices. *Monthly Weather Review*, 110(7):800–807.

Chen, W. Y. and van den Dool, H. (2003). Sensitivity of teleconnection patterns to the sign of their primary action center. *Monthly Weather Review*, 131(11):2885–2899.

Cheng, W., Chiang, J. C. H., and Zhang, D. (2013). Atlantic meridional overturning circulation (AMOC) in CMIP5 models: RCP and historical simulations. *Journal of Climate*, 26(18):7187–7197.

Chiang, J. C. H. and Vimont, D. J. (2004). Analogous pacific and atlantic meridional modes of tropical atmosphere–ocean variability. *Journal of Climate*, 17(21):4143–4158.

Chronis, T., Raitsos, D. E., Kassis, D., and Sarantopoulos, A. (2011). The summer north atlantic oscillation influence on the eastern mediterranean. *Journal of Climate*, 24(21):5584–5596.

Comas-Bru, L. and McDermott, F. (2013). Impacts of the EA and SCA patterns on the european twentieth century NAO-winter climate relationship. *Quarterly Journal of the Royal Meteorological Society*, 140(679):354–363.

de Viron, O., Dickey, J. O., and Ghil, M. (2013). Global modes of climate variability. *Geophysical Research Letters*, 40(9):1832–1837.

Enfield, D. B., Mestas-Nuñez, A. M., Mayer, D. A., and Cid-Serrano, L. (1999). How ubiquitous is the dipole relationship in tropical atlantic sea surface temperatures? *Journal of Geophysical Research: Oceans*, 104(C4):7841–7848.

Enfield, D. B., Mestas-Nuñez, A. M., and Trimble, P. J. (2001). The atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental u.s. *Geophysical Research Letters*, 28(10):2077–2080.

Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J., Mastrandrea, M. D., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C., Girma, B., Kissel, E. S., Levy, A. N., MacCracken, S., Mastrandrea, P. R., White, L. L., Field, C. B., Barros, V. R., Mach, K. J., Mastrandrea, M. D., Aalst, M. v., Adger, W. N., Arent, D. J., Barnett, J., Betts, R., Bilir, T. E., Birkmann, J., Carmin, J., Chadee, D. D., Challinor, A. J., Chatterjee, M., Cramer, W., Davidson, D. J., Estrada, Y. O., Gattuso, J.-P., Hijioka, Y., Hoegh-Guldberg, O., Huang, H.-Q., Insarov, G. E., Jones, R. N., Kovats, R. S., Lankao, P. R., Larsen, J. N., Losada, I. J., Marengo, J. A., McLean, R. F., Mearns, L. O., Mechler, R., Morton, J. F., Niang, I., Oki, T., Olwoch, J. M., Opondo, M., Poloczanska, E. S., Pörtner, H.-O., Redsteer, M. H., Reisinger, A., Revi, A., Schmidt, D. N., Shaw, M. R., Solecki, W., Stone, D. A., Stone, J. M. R., Strzepek, K. M., Suarez, A. G., Tschakert, P., Valentini, R., Vicuña, S., Villamizar, A., Vincent, K. E., Warren, R., White, L. L., Wilbanks, T. J., Wong, P. P., and Yohe, G. W. (2014). Technical summary. In *Climate Change 2014, Impacts, Adaptiation, and Vulnerability, Part A: Global and Sectoral Aspects*, pages 35–94. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Gao, T., yi Yu, J., and Paek, H. (2016). Impacts of four northern-hemisphere teleconnection patterns on atmospheric circulations over eurasia and the pacific. *Theoretical and Applied Climatology*, 129(3-4):815–831.

Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letters*, 33(8).

Giorgi, F. and Lionello, P. (2008). Climate change projections for the mediterranean region. *Global and Planetary Change*, 63(2-3):90–104.

Givati, A. and Rosenfeld, D. (2013). The arctic oscillation, climate change and the effects on precipitation in israel. *Atmospheric Research*, 132-133:114–124.

Harris, I., Jones, P., Osborn, T., and Lister, D. (2013). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *International Journal of Climatology*, 34(3):623–642.

Haylock, M. R., Hofstra, N., Tank, A. M. G. K., Klok, E. J., Jones, P. D., and New, M. (2008). A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113(D20).

Higgins, R. W., Leetmaa, A., and Kousky, V. E. (2002). Relationships between climate variability and winter temperature extremes in the united states. *Journal of Climate*, 15(13):1555–1572.

Higgins, W., Leetmaa, A., Xue, Y., and Barnston, A. (2000). Dominant factors influencing the seasonal predictability of u.s. precipitation and surface air temperature. *Journal of Climate - J CLIMATE*, 13:3994–4017.

Hurrell, J. W. (1995). Decadal trends in the north atlantic oscillation: Regional temperatures and precipitation. *Science*, 269(5224):676–679.

Ionita, M. (2014). The impact of the east atlantic/western russia pattern on the hydroclimatology of europe from mid-winter to late spring. *Climate*, 2(4):296–309.

Kelley, C., Ting, M., Seager, R., and Kushnir, Y. (2012). Mediterranean precipitation climatology, seasonal cycle, and trend as simulated by CMIP5. *Geophysical Research Letters*, 39(21):n/a–n/a.

Kossin, J. P. and Vimont, D. J. (2007). A more general framework for understanding atlantic hurricane variability and trends. *Bulletin of the American Meteorological Society*, 88(11):1767–1782.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the köppen-geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263.

Krichak, S. O. and Alpert, P. (2005). Decadal trends in the east atlantic-west russia pattern and mediterranean precipitation. *International Journal of Climatology*, 25(2):183–192.

Krichak, S. O., Breitgand, J. S., Gualdi, S., and Feldstein, S. B. (2014). Teleconnection–extreme precipitation relationships over the mediterranean region. *Theoretical and Applied Climatology*, 117(3):679–692.

Kryzhov, V. N. and Gorelits, O. V. (2015). The arctic oscillation and its impact on temperature and precipitation in northern eurasia in the 20th century. *Russian Meteorology and Hydrology*, 40(11):711–721.

Köppen, W. (1900). Versuch einer klassifikation der klimate, vorzugsweise nach ihren beziehungen zur pflanzenwelt. *Geographische Zeitschrift*, 6(11):593–611.

Larson, J., Zhou, Y., and Higgins, R. W. (2005). Characteristics of landfalling tropical cyclones in the united states and mexico: Climatology and interannual variability. *Journal of Climate*, 18(8):1247–1262.

Lionello, P. (2012). Preface. In Lionello, P., editor, *The Climate of the Mediterranean Region*, pages xiii – xix. Elsevier, Oxford.

Lionello, P., Abrantes, F., Congedi, L., Dulac, F., Gacic, M., Gomis, D., Goodess, C., Hoff, H., Kutiel, H., Luterbacher, J., Planton, S., Reale, M., Schröder, K., Struglia, M. V., Toreti, A., Tsimplis, M., Ulbrich, U., and Xoplaki, E. (2012). Introduction: Mediterranean climate—background information. In Lionello, P., editor, *The Climate of the Mediterranean Region*, pages xxxv – xc. Elsevier, Oxford.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., and Francis, R. C. (1997). A pacific interdecadal climate oscillation with impacts on salmon production*. *Bulletin of the American Meteorological Society*, 78(6):1069–1080.

Mariotti, A. and Arkin, P. (2007). The north atlantic oscillation and oceanic precipitation variability. *Climate Dynamics*, 28(1):35–51.

Mariotti, A., Struglia, M. V., Zeng, N., and Lau, K.-M. (2002). The hydrological cycle in the mediterranean region and implications for the water budget of the mediterranean sea. *Journal of Climate*, 15(13):1674–1690.

Marshall, G. J. (2003). Trends in the southern annular mode from observations and reanalyses. *Journal of Climate*, 16(24):4134–4143.

Martens, B., Waegeman, W., Dorigo, W. A., Verhoest, N. E. C., and Miralles, D. G. (2018). Terrestrial evaporation response to modes of climate variability. *npj Climate and Atmospheric Science*, 1(1).

Mikhailova, N. and Yurovsky, A. (2016). The east atlantic oscillation: Mechanism and impact on the european climate in winter. *Physical Oceanography*, (4):25–33.

Moura, A. D. and Shukla, J. (1981). On the dynamics of droughts in northeast brazil: Observations, theory and numerical experiments with a general circulation model. *Journal of the Atmospheric Sciences*, 38(12):2653–2675.

Palter, J. B. (2015). The role of the gulf stream in european climate. *Annual Review of Marine Science*, 7(1):113–137. PMID: 25560606.

Pohlmann, H., Sienz, F., and Latif, M. (2006). Influence of the multidecadal atlantic meridional overturning circulation variability on european climate. *Journal of Climate*, 19(23):6062–6067.

Rajagopalan, B., Kushnir, Y., and Tourre, Y. M. (1998). Observed decadal midlatitude and tropical atlantic climate variability. *Geophysical Research Letters*, 25(21):3967–3970.

Robinson, A., Leslie, W., Theocharis, A., and Lascaratos, A. (2001). Mediterranean sea circulation. In *Encyclopedia of Ocean Sciences*, pages 1689–1705. Elsevier.

Ropelewski, C. F. and Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the el niño/southern oscillation. *Monthly Weather Review*, 115(8):1606–1626.

Ropelewski, C. F. and Halpert, M. S. (1989). Precipitation patterns associated with the high index phase of the southern oscillation. *Journal of Climate*, 2(3):268–284.

Ropelewski, C. F. and Jones, P. D. (1987). An extension of the tahiti–darwin southern oscillation index. *Monthly Weather Review*, 115(9):2161–2165.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T. (1999). A dipole mode in the tropical indian ocean. *Nature*, 401:360.

Sanchez-Gomez, E., Somot, S., and Mariotti, A. (2009). Future changes in the mediterranean water budget projected by an ensemble of regional climate models. *Geophysical Research Letters*, 36(21).

Sene, K. (2013). *Flash Floods*. Springer Netherlands.

Steirou, E., Gerlitz, L., Apel, H., and Merz, B. (2017). Links between large-scale circulation patterns and streamflow in central europe: A review. *Journal of Hydrology*, 549:484 – 500.

Stephens, G. L. and Kummerow, C. D. (2007). The remote sensing of clouds and precipitation from space: A review. *Journal of the Atmospheric Sciences*, 64(11):3742–3765.

Stocker, T., Qin, D., Plattner, G.-K., Alexander, L., Allen, S., Bindoff, N., Bréon, F.-M., Church, J., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory, J., Hartmann, D., Jansen, E., Kirtman, B., Knutti, R., Krishna Kumar, K., Lemke, P., Marotzke, J., Masson-Delmotte, V., Meehl, G., Mokhov, I., Piao, S., Ramaswamy, V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L., Vaughan, D., and Xie, S.-P. (2013). *Technical Summary*, book section TS, page 33–115. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1):79–107.

Sutton, R. T. (2005). Atlantic ocean forcing of north american and european summer climate. *Science*, 309(5731):115–118.

Tang, Q., Gao, H., Lu, H., and Lettenmaier, D. P. (2009). Remote sensing: hydrology. *Progress in Physical Geography: Earth and Environment*, 33(4):490–509.

Thompson, D. W. J. and Wallace, J. M. (1998). The arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, 25(9):1297–1300.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Trenberth, K. E. and Caron, J. M. (2000). The southern oscillation revisited: Sea level pressures, surface temperatures, and precipitation. *Journal of Climate*, 13(24):4358–4365.

van den Dool, H. M., Saha, S., and Johansson, A. (2000). Empirical orthogonal teleconnections. *Journal of Climate*, 13(8):1421–1435.

van Loon, H. and Rogers, J. C. (1978). The seesaw in winter temperatures between greenland and northern europe. part i: General description. *Monthly Weather Review*, 106(3):296–310.

van Oldenborgh, G. J., te Raa, L. A., Dijkstra, H. A., and Philip, S. Y. (2009). Frequency- or amplitude-dependent effects of the atlantic meridional overturning on the tropical pacific ocean. *Ocean Science*, 5(3):293–301.

Vimont, D. J. and Kossin, J. P. (2007). The atlantic meridional mode and hurricane activity. *Geophysical Research Letters*, 34(7).

Walker, G. and Bliss, E. (1932). World weather v. *Memoirs of the Royal Meteorological Society*, 4(36):53–58.

Wallace, J. M. and Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, 109(4):784–812.

Wang, C. (2002). Atlantic climate variability and its associated atmospheric circulation cells. *Journal of Climate*, 15(13):1516–1536.

Wang, G. and Schimel, D. (2003). Climate Change, Climate Modes,and Climate Impacts. *Annual Review of Environment and Resources*, 28(1):1–28.

Zhang, Y., Wallace, J. M., and Battisti, D. S. (1997). Enso-like interdecadal variability: 1900–93. *Journal of Climate*, 10(5):1004–1020.

Zhou, S., Miller, A. J., Wang, J., and Angell, J. K. (2001). Trends of NAO and AO and their associations with stratospheric processes. *Geophysical Research Letters*, 28(21):4107–4110.