Indecis
Sectorial Climate Services

Integrated approach for the **d**evelopment across **E**urope of user oriented **c**limate **i**ndicators for GFCS high-priority **s**ectors: Agriculture, disaster risk reduction, energy, health, water and tourism

Work Package 5

Deliverable 5.3

Report on the relation between INDECIS-QCHDS and INDECIS-ISD and atmospheric patterns

Richard P. Allan[1], Albert Ossó[1,2], Philip Craig[1,3], Len Shaffrey[1,3], Ed Hawkins[1,3], Pedro M. Sousa[4], Ricardo M. Trigo[4], Ana Casanueva[5], Joaquín Bedia[5], Sixto Herrera[5], Mª Dolores Frías[5], Jesús Fernández[5], Antonio S. Cofiño[5], José Manuel Gutiérrez[5], Llorenç Lledó[6], Jaume Ramon Gamon[6], Roberto Coscarelli[7], Tommaso Caloiero[8], Giulio Nils-Caroletti[7], Manuel Del Jesus[9], Salvador Navas[9], Diego Urrea[9]

[1] Department of Meteorology, University of Reading, Reading – United Kingdom

[2] Center for Climate and Global Change University of Graz, Austria

[3] National Centre for Atmospheric Science, University of Reading, Reading – United Kingdom

[4] Instituto Dom Luiz, Faculdade de Ciencias, Universidade de Lisboa, Lisbon, Portugal

[5] University of Cantabria, Santander, Spain

[6] Barcelona Supercomputing Center (BSC-CNS)

[7] Consiglio Nazionale delle Ricerche - Istituto di Ricerca per la Protezione Idrogeologica, C.N.R.-I.R.P.I.

[8] Istituto per i Sistemi Agricoli e Forestali del Mediterraneo (CNR-ISAFOM)

[9] Environmental Hydraulics Institute. Universidad de Cantabria

ERA4CS European Research Area for Climate Services

JPI Climate

This report arises from the Project INDECIS which is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR), with co-funding by the European Union's Horizon 2020 research and innovation programme

TABLE OF CONTENTS

1. Overview

The goal of WP5 is to use the data developed in WPs 2 and 3 and indices developed in WP4 to quantify their variability and change over time, and their link to atmospheric circulation patterns. The deliverable 5.2 presented analysis of the INDECIS-ISD climatology, temporal evolution, and assessed their Time of Emergence. In this report we explore the relationships between the indices and patterns of atmospheric circulation. This covers analysis of the relationships of INDECIS-ISD with atmospheric teleconnection indices (Section 2) with an additional focus also on growing season characteristics and teleconnection patterns (Section 6) and extension to prediction of sectoral indices from atmospheric circulation metrics (Section 7). Links to sectoral impacts are further covered through an analysis of how Euro-Atlantic teleconnections affect wind power indices (Section 3). In addition, a process-based analysis further outlines the impact of more complex circulation metrics associated with blocking high pressure characteristics and their impact on sectoral indices (Section 4). These Europe-wide investigations are complimented with a focused case study on how atmospheric circulation teleconnections influence precipitation across Calabria, a region of southern Italy (Section 5).

2. Relationships of INDECIS-ISD with atmospheric teleconnection indices

Ana Casanueva , Joaquín Bedia, Sixto Herrera, Mª Dolores Frías, Jesús Fernández, Antonio S. Cofiño, José Manuel Gutiérrez (University of Cantabria)

2.1 Summary

Climate impact indices, such as INDECIS-ISD, are effective tools to assess local and regional sector-specific climate impacts such as droughts, heat waves or wildfires among others. Their temporal and spatial variability are therefore impact-relevant and can be related to climate oscillations, such as changes in atmospheric patterns. The relationships between impact indices and atmospheric patterns can therefore shed light on the mechanisms controlling the variability of extreme events. Those links can be analyzed either in temperature and precipitation, but also in indices related to precipitation extremes, river flows, snow cover or cereal production (see e.g. Rodríguez-Puebla *et al.* 2007; Carril *et al.* 2008; Vicente-Serrano *et al.* 2009; Brands *et al.* 2013, Casanueva *et al.* 2014).

INDECIS D51 compiled a list of atmospheric patterns which influence European climate variability. INDECIS D52 gave an overview on temporal evolution of the INDECIS-QHDS and INDECIS-ISD, including the time-emergence of climate-change signals and relation with atmospheric patterns. The UC contribution to D5.3 is focused on the analysis of the relationship between the main Euro-Atlantic circulation patterns and the variability of a selected subset of INDECIS-ISD.

In addition, UC is currently working on the development of an R package for the calculation of several INDECIS-QHDS indices and atmospheric patterns, as well as towards its integration within the climate4R open-source framework for climate data access and analysis (Iturbide *et al.* 2019), in order to demonstrate the reproducibility and possible integration of INDECIS scientific activities in the framework of user-oriented climate services.

2.2 Data and Methods

Data

E-OBS daily observational dataset (v17) on a 0.5x0.5º regular grid is used to obtain a subset of impact indices (INDECIS-ISD) for the period 1950-2017. It was retrieved from the data services provided by University of Cantabria Meteorology Group (UCMG), built on a THREDDS Data Server (TDS) by means of the user authorization web application User Data Getaway - Thredds Access Portal (UDG-TAP). Data retrieval, further postprocessing and indices calculation were performed with the climate4R bundle for climate data analysis and processing (Iturbide *et al.* 2019). The indices (Table 1) span simple indicators expressing, in most cases, threshold exceedances and rely on the three primary ECVs, namely precipitation (Pr), minimum temperature (Tmin) and maximum temperature (Tmax). They are computed on a seasonal basis (see details in Table 2.1). Grid boxes with less than 50% of data in the full period are left blank. Climatological means and variability (expressed by the standard deviation normalized by the mean, as percentage) are shown in order to give an overview of the selected indices.

Index Code	Index Name	Description	Units	Seasonal agg. fun	ECV
RR1	Wet day frequency	Nº of days with Pr ≥ 1 mm	Days/season	sum	Pr
SDII	Simple precipitation intensity index	Mean wet-day precipitation	mm/day	mean	Pr
TR	Tropical nights	Nº of days with Tmin>20°C	Days/season	sum	Tmin
SU	Summer days	Nº of days with Tmax>25°C	Days/season	sum	Tmax
FD	Frost days	Nº days with Tmin< 0°C	Days/season	sum	Tmin
ID	Ice days	Nº of days with Tmax<0°C	Days/season	sum	Tmax

Table 2.1. Description of considered INDECIS-ISD.

Data for the teleconnection indices is available from the US Climate Prediction Centre (CPC, https://www.cpc.ncep.noaa.gov/data/teledoc/teleindcalc.shtml). CPC indices are derived from the first 10 Rotated Principal Components (rPCA) of monthly mean standardized 500-mb height (hgt500) anomalies. They are associated with the main modes of atmospheric variability in the northern hemisphere: NAO, EA, WP, EP/NP, PNA, EA/WR, SCA, TNH, POL, PT (Table 2.2). We additionally consider two teleconnection indices related to sea surface temperature (SST) anomalies in the Pacific Ocean, namely NINO3.4 and ONI, available from the National Center for Atmospheric Research (NCAR, https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni).

The monthly values from 1950 to 2017 are aggregated into mean seasonal values for the sake of comparison with the climate impact indices.

Index Code	Index Name	Variable	Source
NAO	North Atlantic Oscillation	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
EA	East Atlantic Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
WP	West Pacific Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
EP/NP	East Pacific/ North Pacific Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
PNA	Pacific/ North American Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
EA/WR	East Atlantic/West Russia Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
SCA	Scandinavian Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
TNH	Tropical/ Northern Hemisphere Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
POL	Polar/ Eurasia Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
PT	Pacific Transition Pattern	hgt500 (or z500)	https://www.cpc.ncep.noaa.gov/data/tel
NINO3.4	El Niño/Southern Oscillation	SST	https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni
ONI	Oceanic Niño Index	SST	https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni

Table 2.2 Description of the teleconnection indices used.

Correlation analysis

The relationship between impact and teleconnection indices is assessed through regression models, considering in all cases detrended time series, in order to avoid misleading results due to temporal trends. A linear regression is fitted between each teleconnection index (as predictor) and each impact index (predictand). The regression coefficient shows the sign and strength of the relation between predictor and predictand, i.e. how much the mean of the dependent variable changes given a one-unit shift in the independent variable. Statistically significant coefficients (at a significance level of 0.05) are shown.

2.3 Results

Relationships with teleconnection indices

Regression coefficients for all 12 teleconnection indices (see Table 2) are analyzed for the six climate impact indices (see Table 1). Here we show the most relevant results for each index. The four seasons are presented for the precipitation-derived indices, whereas results for the most relevant seasons for the temperature indices are included.

RR1

The main modes of variability associated with RR1 are NAO, EA, SCA and NINO3.4 in different parts of the continent, in agreement with previous literature (e.g. Casanueva *et al.* 2014). A dipole structure is evident when relating winter NAO and RR1 (Fig.2.1): positive phases of NAO favour higher frequency of wet days in northern Europe and the British Isles, whereas the opposite behaviour is found in southern Europe. In summer, negative correlation is found between positive phases of NAO and RR1 in the Atlantic watershed. Some significant negative correlations also arise in the intermediate seasons in Central and eastern Europe.

The most important relationships between RR1 and EA are found in western Europe in winter, in the Atlantic watershed in spring and in the British Isles and southern Norway in summer (Fig.8). Some negative correlations are found in the eastern Mediterranean area. SCA largely drives the frequency of wet days in the Scandinavian Peninsula and northeastern Europe along the four seasons (negative correlation, Fig.2.3). RR1 in southern and southwestern Europe are positively related to SCA positive phases in summer and the intermediate seasons, respectively. Although the effect of Pacific SST anomalies over Europe is limited, el Niño 3.4 phases are related to wetter conditions in the Iberian Peninsula in autumn, whereas la Niña events are associated with wetter springs (Fig. 2.4).

SDII

Regression coefficients for SDII are between ±1 mm/day for all teleconnection patterns. Statistically significant values are obtained in winter, in southwestern Iberian Peninsula (negative correlation) and north and east of Europe with NAO (Fig.2.5), and in eastern Europe and Norway with EA/WR (Fig.2.6).

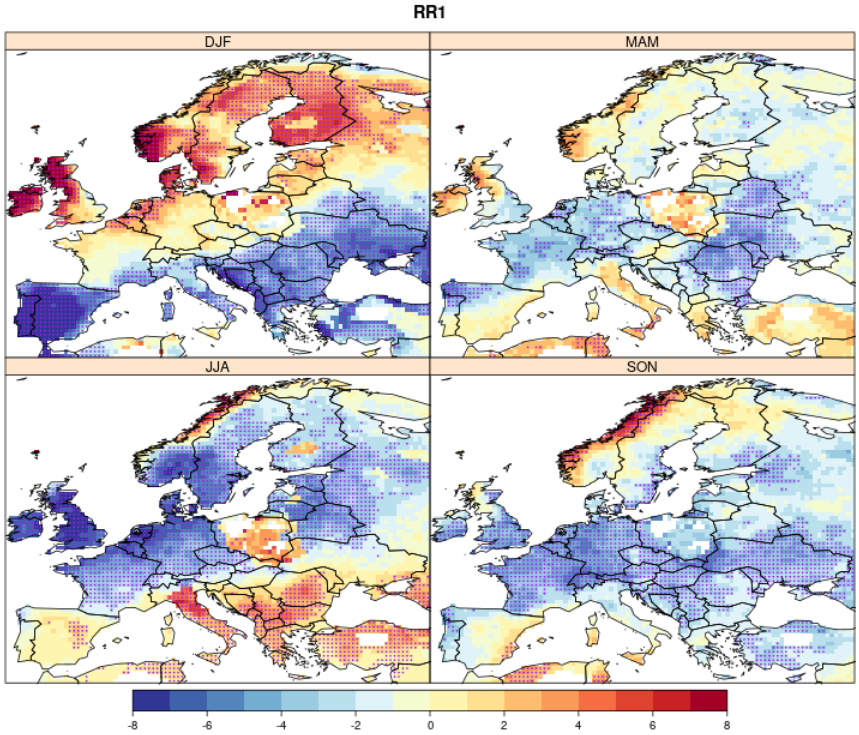

Fig. 2.1 Regression coefficients for RR1 ~ NAO (days/season). Statistically significant values (p-value < 0.05) are depicted in purple.

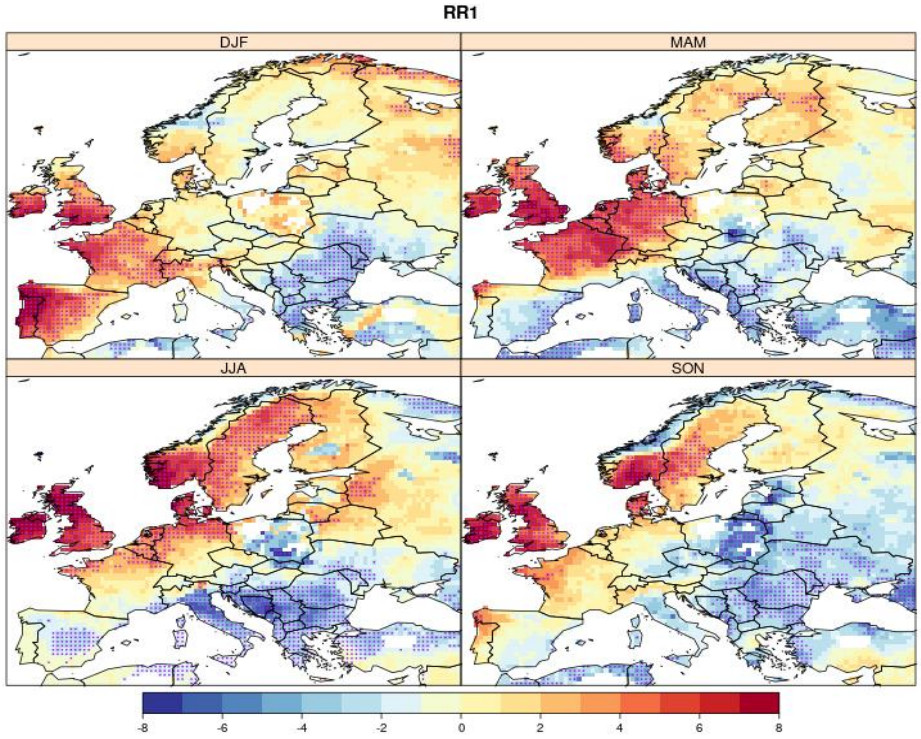

Fig. 2.2 As Fig.2.1 for RR1 ~ EA.

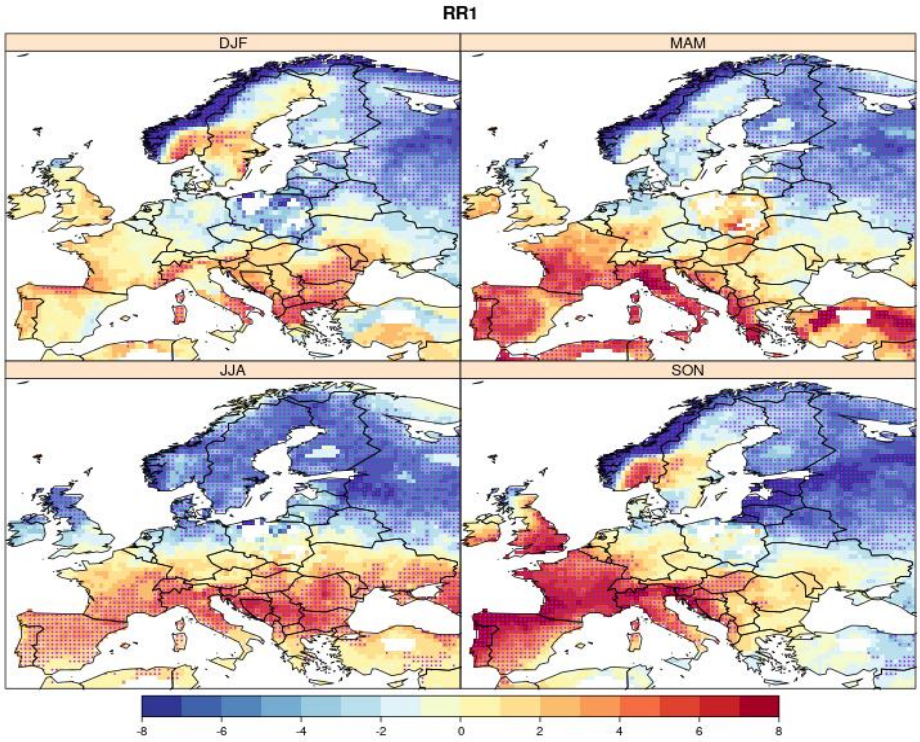

Fig. 2.3 As Fig.2.1 for RR1 ~ SCA.

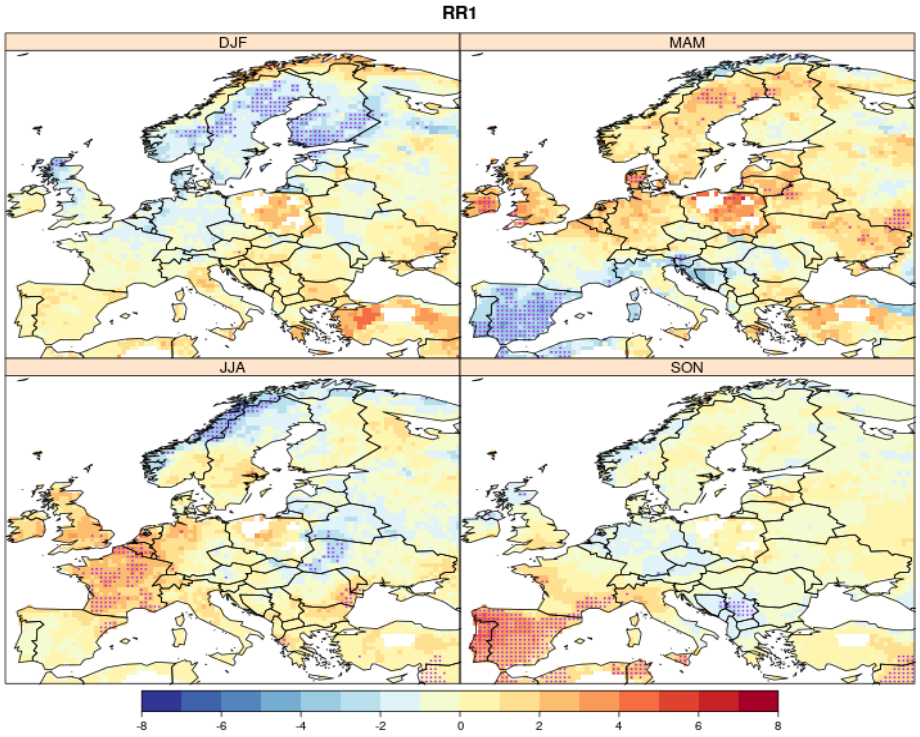

Fig. 2.4 As Fig.2.1 for RR1 ~ NINO3.4.

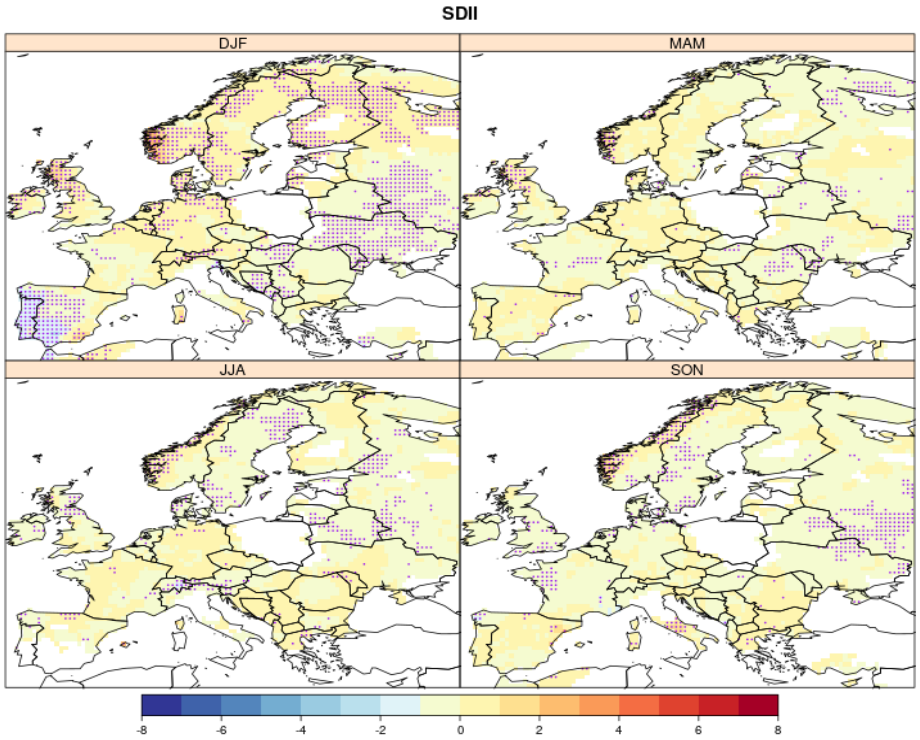

Fig.2.5 As Fig.2.1 for SDII ~ NAO (units of the coefficients are mm/day in this case).

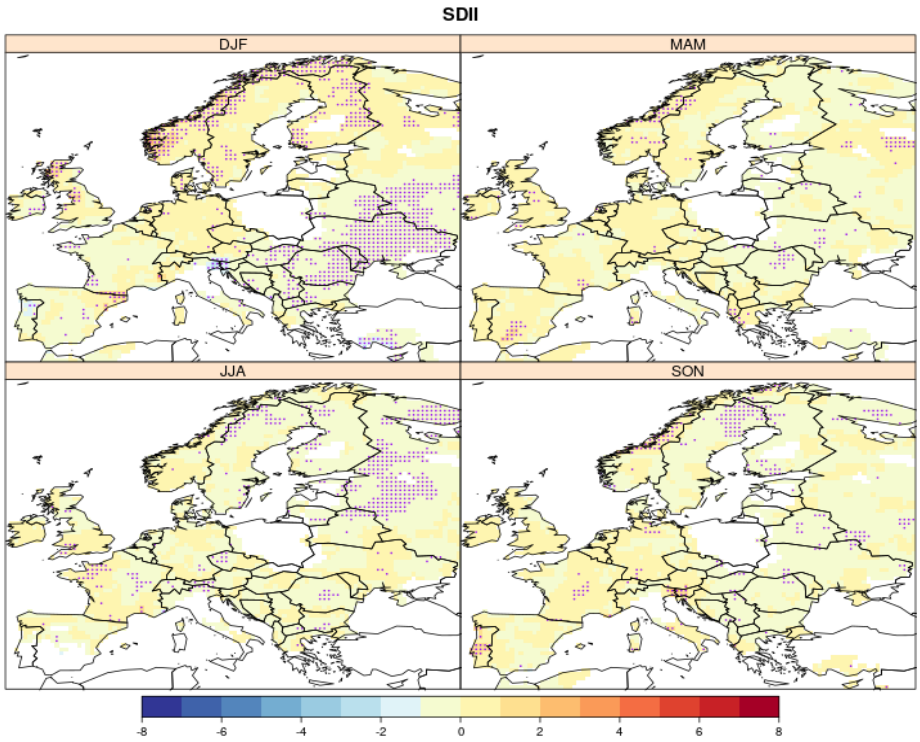

Fig.2.6 As Fig.2.1 for SDII ~ EA/WR (units of the coefficients are mm/day in this case).

TR

Unlike precipitation indices, results for temperature indices are more localized in space and time (i.e. season). The number of tropical nights is negatively (positively) related to NAO and SCA (EA) in several regions in Spain, Italy and Romania (Figs. 2.7-2.9).

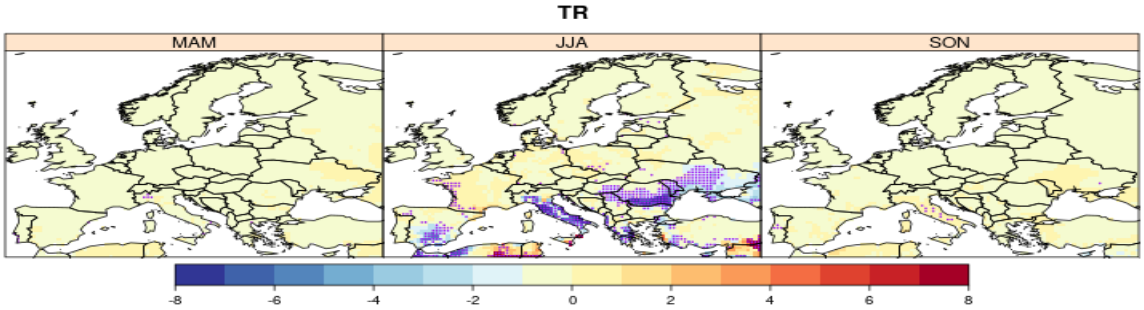

Fig.2.7 As Fig. 2.1 for TR ~ NAO.

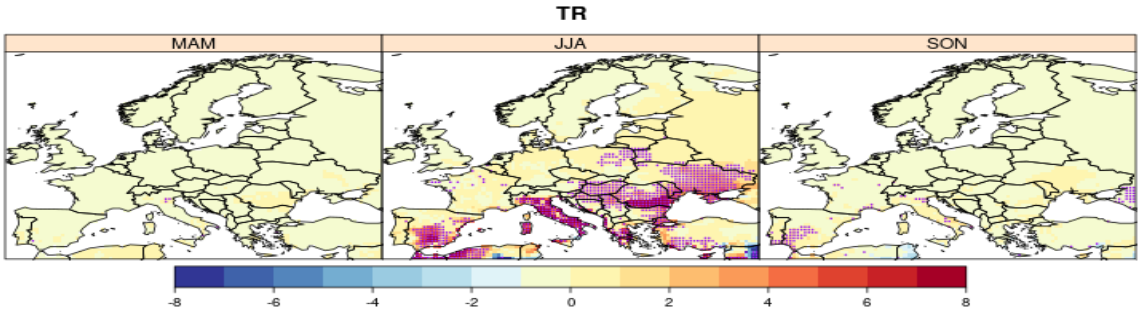

Fig.2.8 As Fig.2.1 for TR ~ EA.

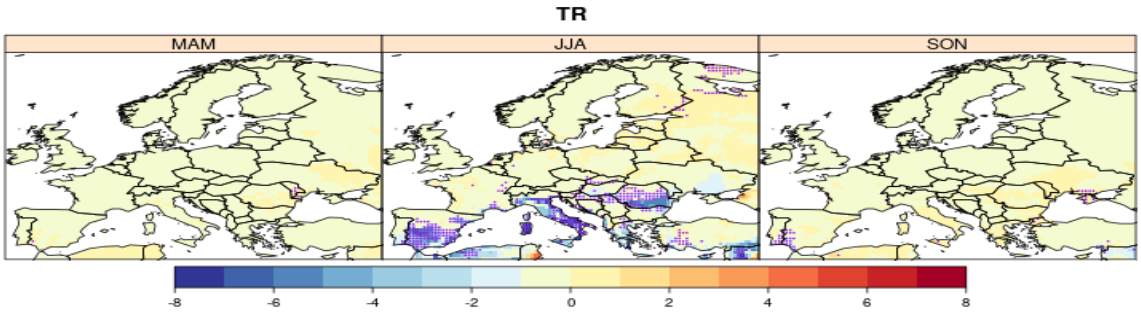

Fig.2.9 As Fig.2.1 for TR ~ SCA.

SU

Dipole structures are also evident for the summer relationships (Figs.2.10-2.13). NAO and EA in summer present high correlation (of opposite sign) with the number of summer days in the eastern Mediterranean, whereas the opposite is found in the Atlantic watershed. East-west and south-north dipoles are found for EA/WR and SCA. Therefore, more summer days in southwestern Europe occur under negative NAO, EA/WR and SCA and positive EA. In the intermediate seasons SCA and EA/WR have an influence on the number of summer days in the Iberian Peninsula.

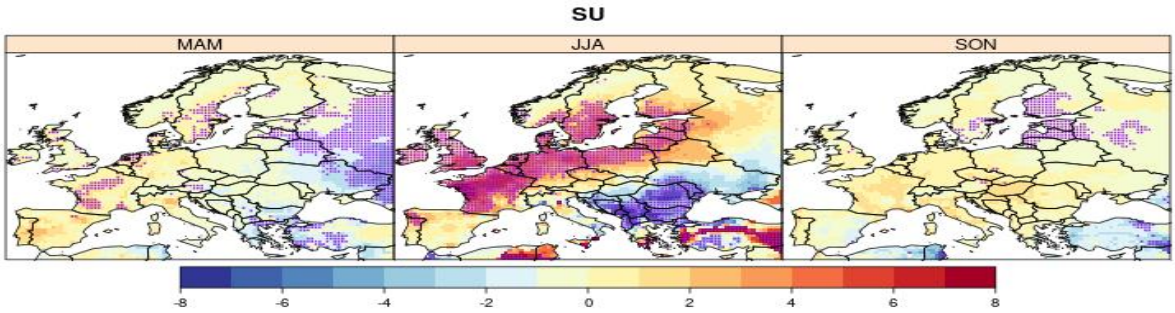

Fig.2.10 As Fig.2.1 for SU ~ NAO.

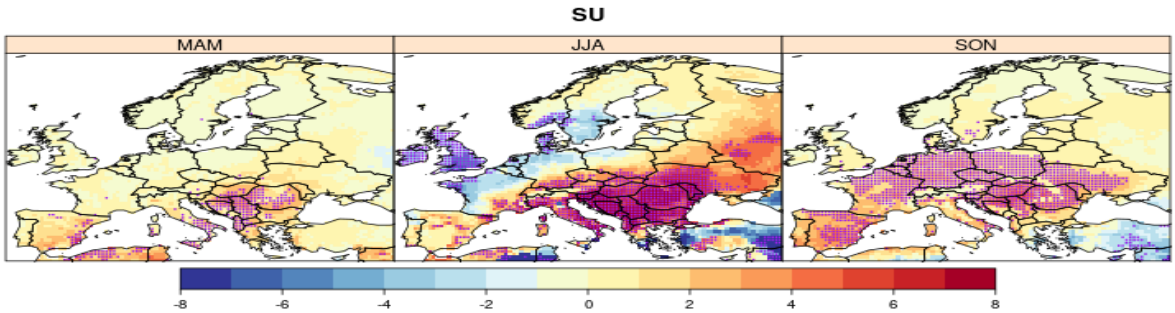

Fig. 2.11 As Fig.2.1 for SU ~ EA.

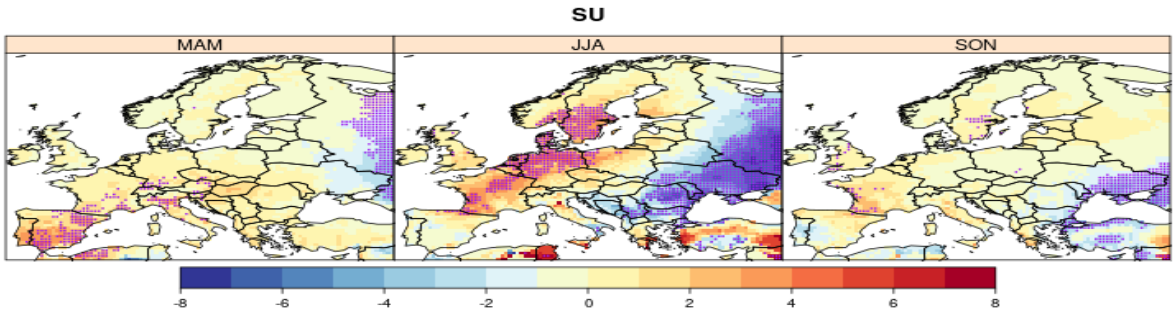

Fig. 2.12 As Fig.2.1 for SU ~ EA/WR.

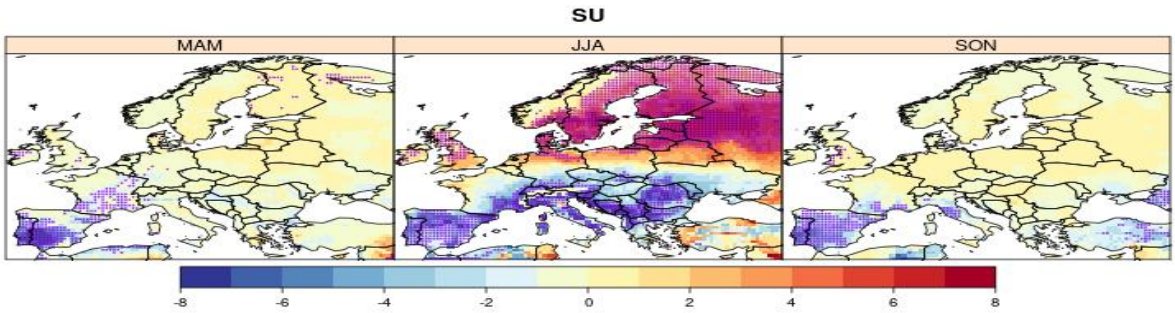

Fig. 2.13 As Fig. 2.1 for SU ~ SCA.

FD

Negative winter NAO is associated with larger number of frost days in large parts in Central and northern Europe (Fig.2.14), whereas EA has a large influence in the Central and southern parts of the continent (Fig.2.15). EA/WR and SCA present larger effects around the Baltic Sea in winter (Figs. 2.16, 2.17). These four teleconnection indices also present some association to FD in the intermediate seasons.

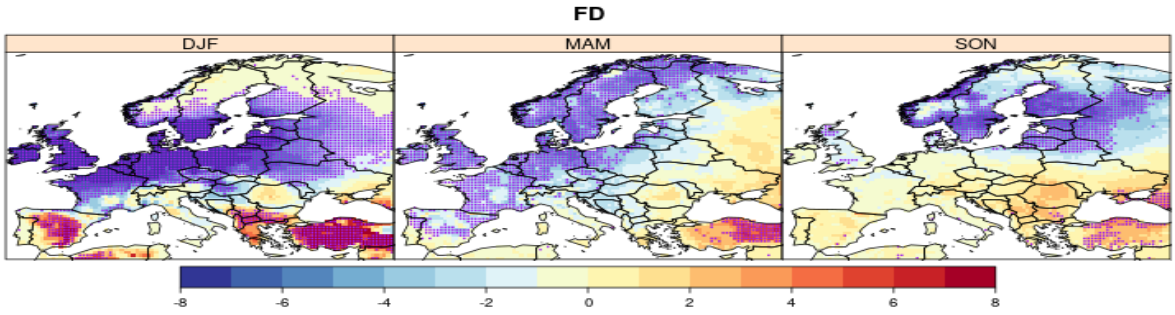

Fig.2.14 As Fig. 2.1 for FD ~ NAO.

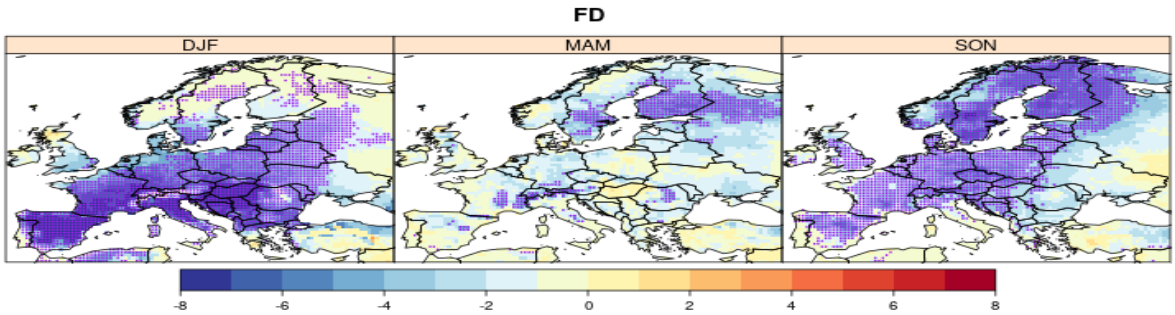

Fig. 2.15 As Fig.2.1 for FD ~ EA.

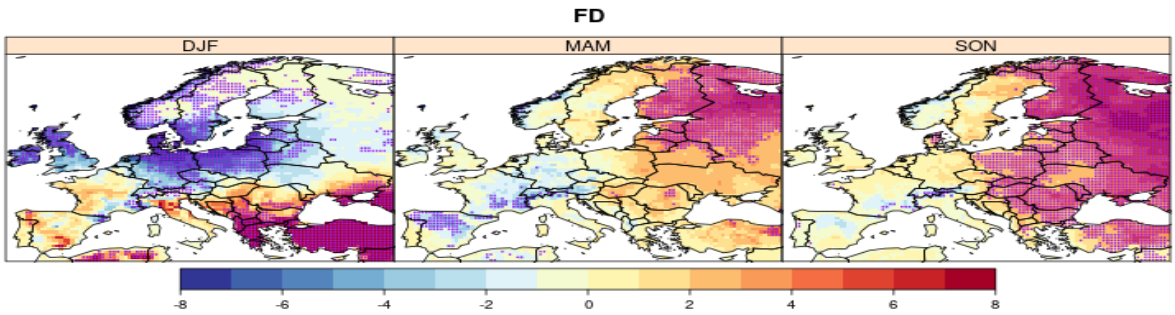

Fig. 2.16 As Fig.2.1 for FD ~ EA/WR.

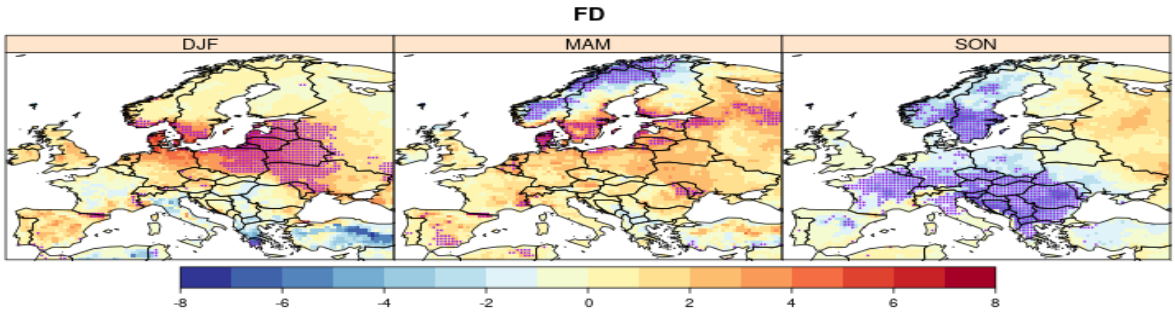

Fig.2.17 As Fig.2.1 for FD ~ SCA.

ID

The frequency of ice days in winter is modulated by negative phase of the NAO (Fig.2.18) in large parts of the continent. Likewise, negative EA and EA/WR and positive SCA are associated with more ID in the southeast, north and east of the continent, respectively (Figs. 2.19-2.21). ID in the intermediate seasons in most influenced by negative NAO in Scandinavia and positive EA/WR in eastern Europe.

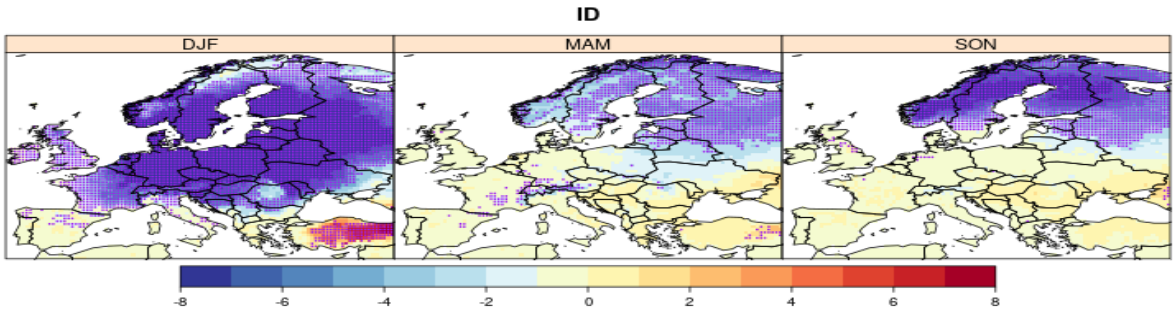

Fig. 2.18 As Fig.2.1 for ID ~ NAO.

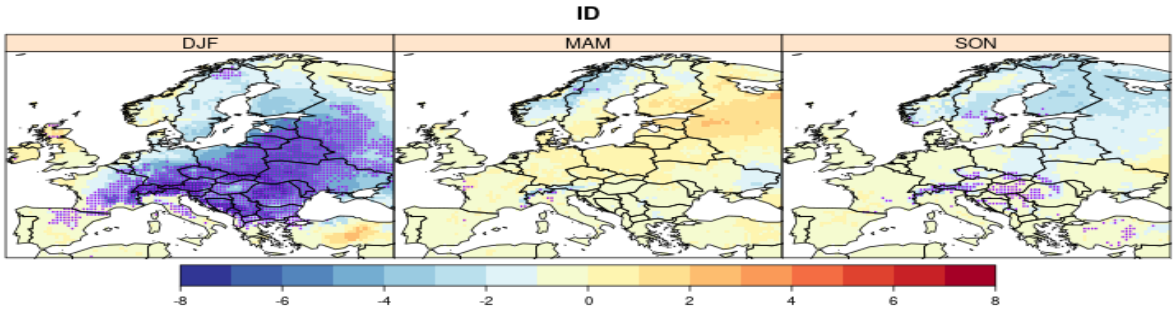

Fig. 2.19 As Fig.2.1 for ID ~ EA.

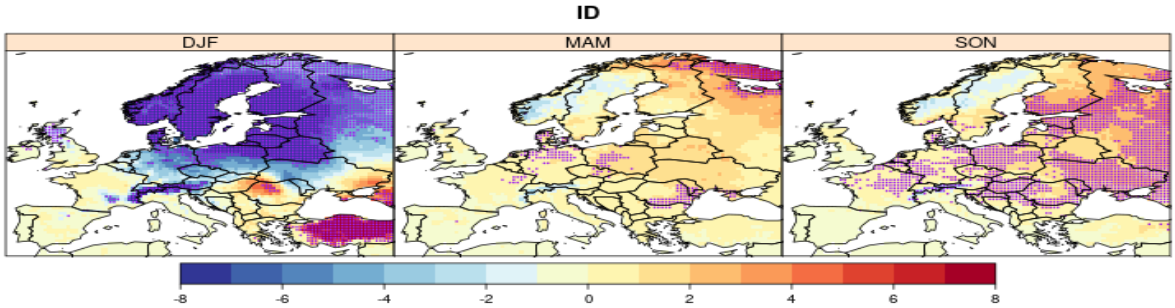

Fig.2.20 As Fig.2.1 for ID ~ EA/WR.

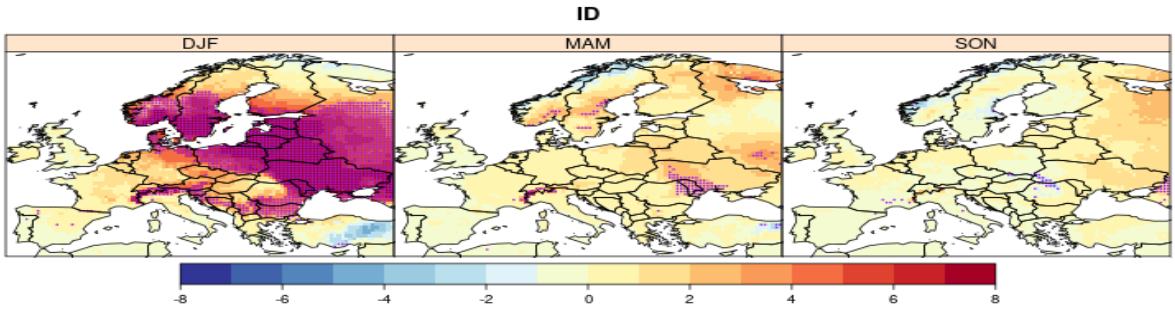

Fig.2.21 As Fig.2.1 for ID ~ SCA.

Also statistically significant trends are found for the teleconnection indices for the period 1950-2017 (Figs. 2.22-2.26). For the sake of brevity, we focus on the teleconnection indices which show the largest relations with the selected INDECIS-ISD. Significant trends (at 95% confidence level) are found for the NAO in winter, spring and autumn (Fig.2.22), EA in the four seasons (Fig.35), EA/WR in summer and autumn (Fig. 2.23), SCA in spring and summer (Fig.2.24) and NINO3.4 in winter, spring and summer (Fig.2.25).

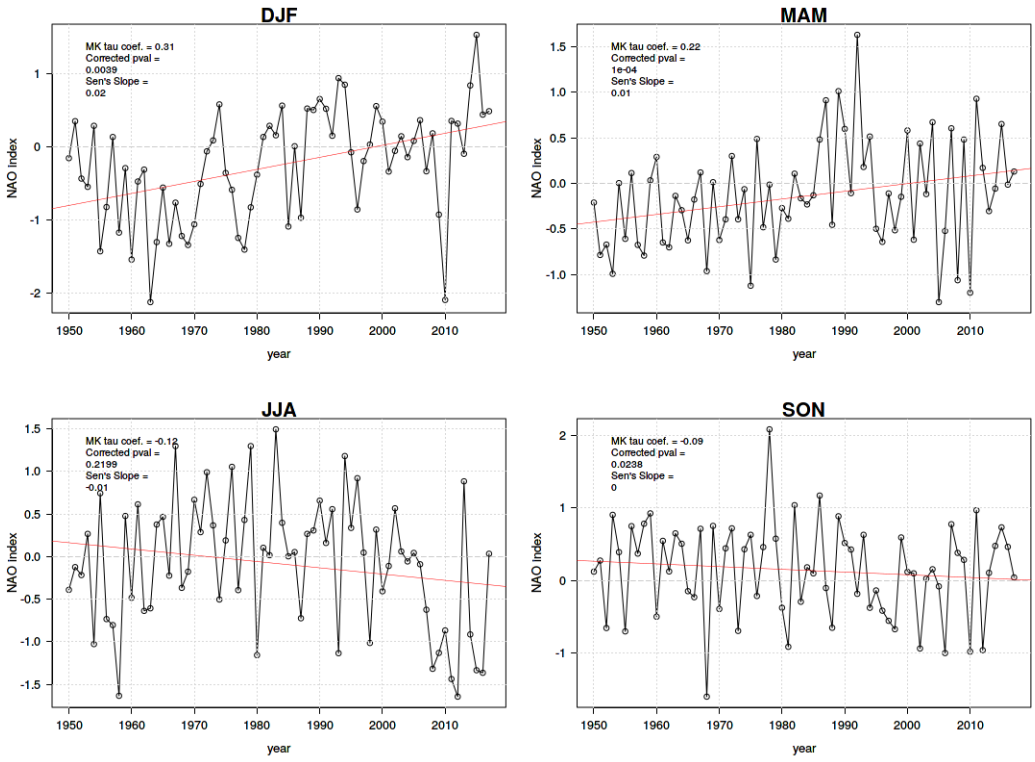

Fig. 2.22 Seasonal time series (1950-2017) for the NAO, including the Kendall's Tau coefficient and its corrected p-value and the Sen's slope.

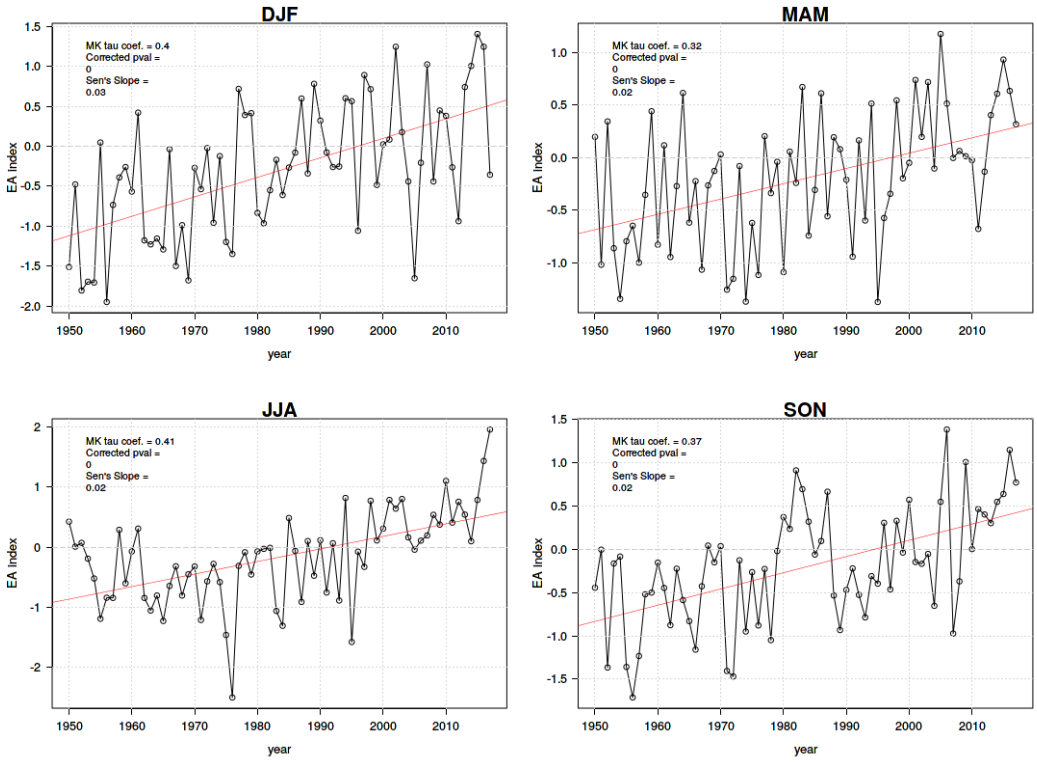

Fig. 2.23 As Fig. 2.22, for EA.

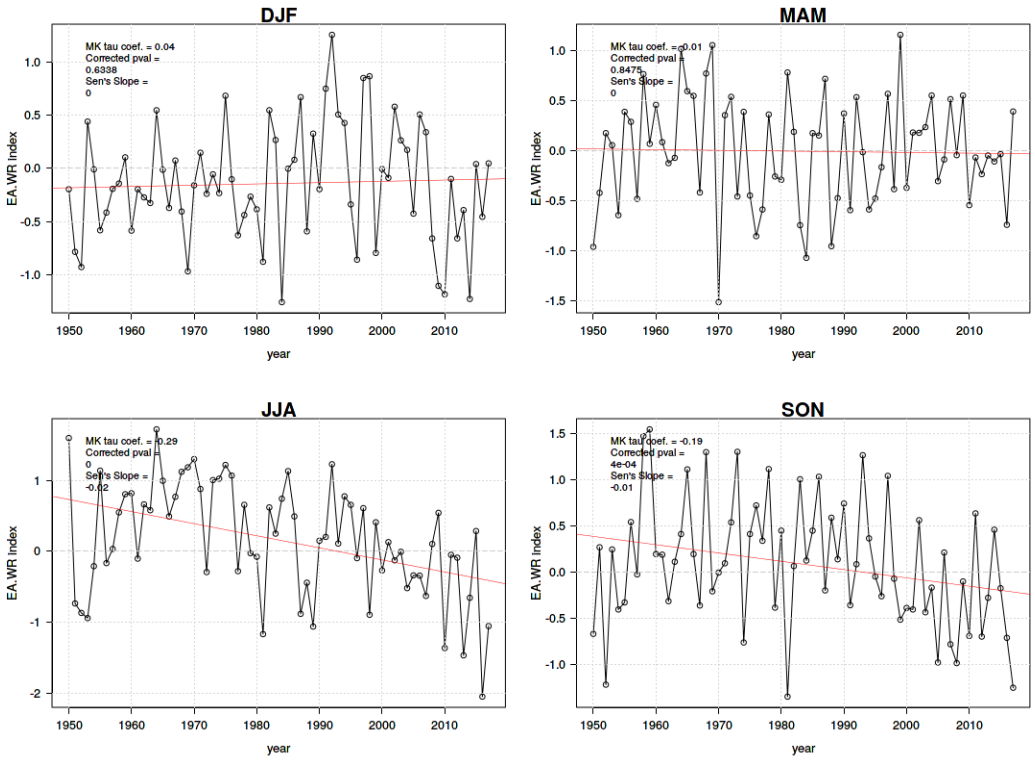

Fig.2.24 As Fig.2.22, for EA/WR.

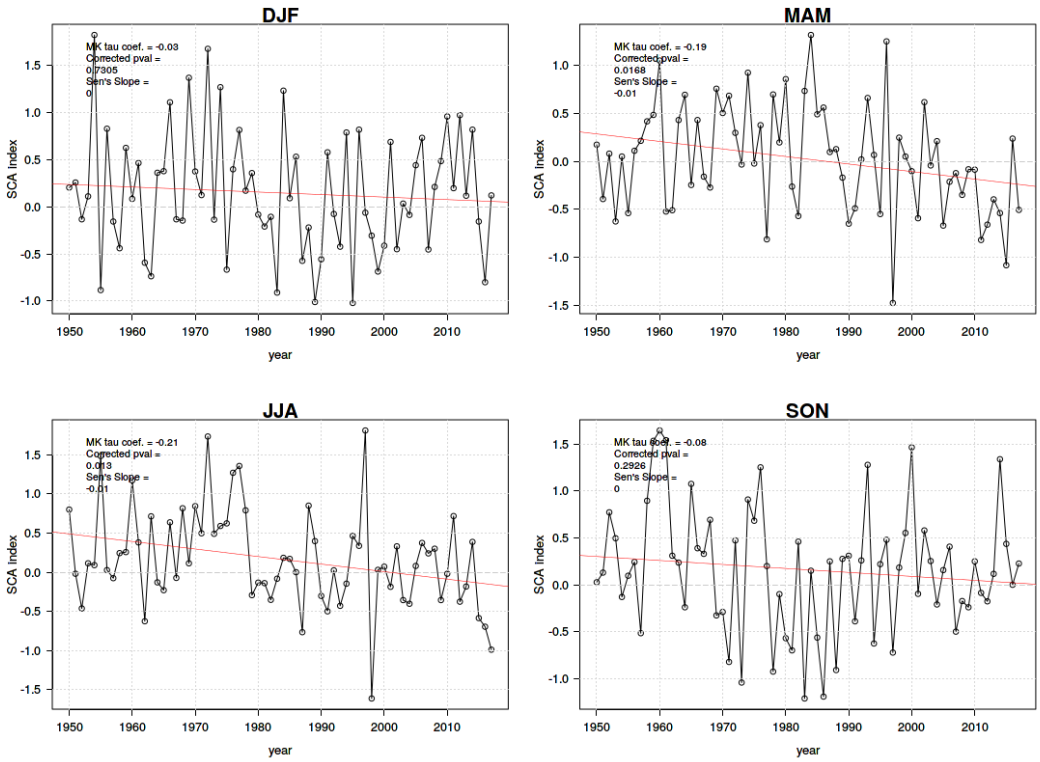

Fig.2.25 As Fig. 2.22, for SCA.

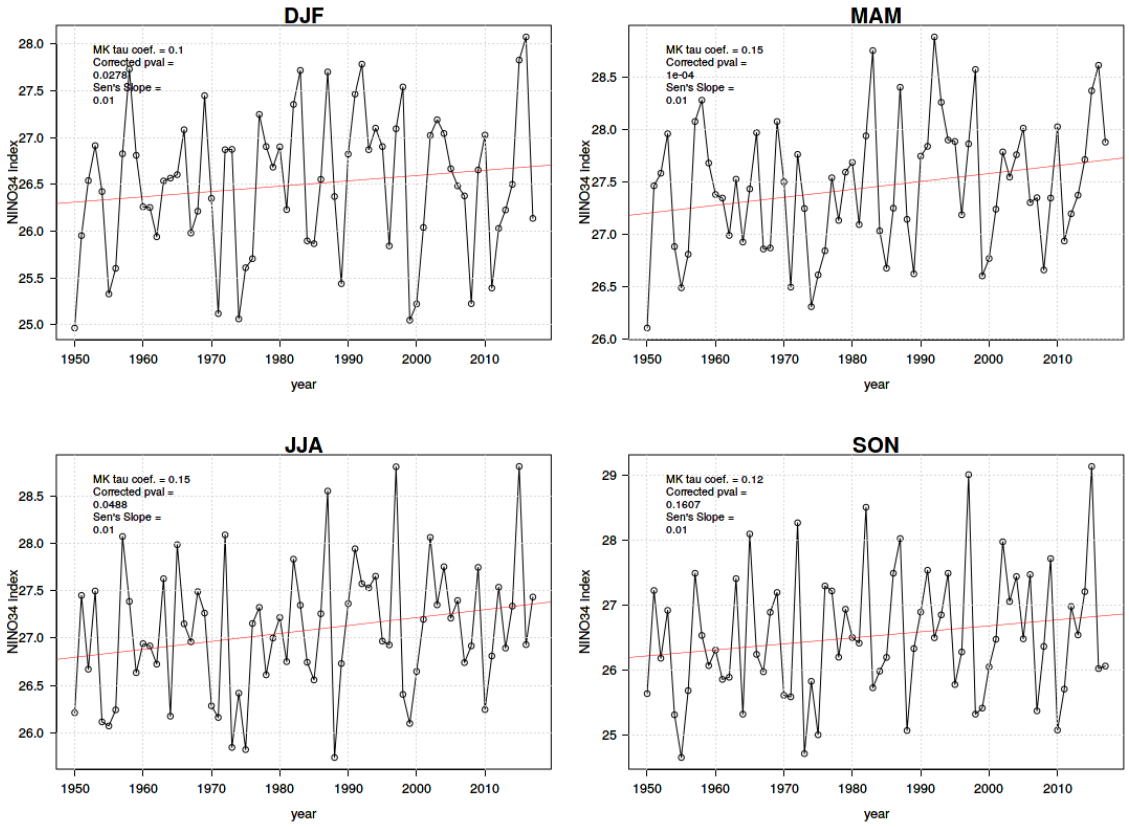

Fig.2.26 As Fig. 2.22, for NINO3.4.

2.4 Conclusions

- The teleconnection indices which present the largest influence over climate indices in Europe are NAO, EA, SCA and EA/WR, with a seasonal- and region-dependent effect.
- Oceanic indices such as el Niño3.4 show little impact on the European climate, except for wet-day frequency in the Iberian Peninsula (in agreement with Rocha 1999, Casanueva *et al.* 2014), which has been found to be a potential source of conditional predictability in the context of seasonal forecasts (Frias *et al.* 2010).
- The largest observed trends concerning teleconnection indices correspond to NAO (especially winter NAO) and EA, which are closely related to precipitation-derived indices and cold extremes in most continental Europe, and more locally associated with temperature-derived (hot extremes) indices in the Mediterranean basin.
- In this work the relationships between climate indices and concurrent teleconnections are shown. Lagged teleconnections might additionally shed light on the mechanisms controlling the variability of extreme events and will be further analyzed in this WP.

2.5 Software development

A large effort within INDECIS WP5 has been devoted to the development of an R package that allows the calculation of teleconnection patterns, thus setting the ground for new windows of opportunity within INDECIS works. In particular, the representation of atmospheric patterns by climate models (e.g. in a seasonal forecasting context) and their relations with climate impact indices will be further evaluated in future INDECIS deliverables (D63).

climate4R.indices is an open-source R package included in climate4R framework for climate data analysis and processing (Iturbide *et al.* 2019). It allows the calculation of climate impact indices (such as number of frost days, hot days, etc.). An updated version of the package, developed in the frame of INDECIS, offers also the possibility of calculating atmospheric indices. In particular, the CPC indices and ENSO indices described in Table 2 are available. Further developments will include other atmospheric patterns such as weather typing techniques. The code can be found in a GitHub repository (see https://github.com/SantanderMetGroup/climate4R.indices/tree/devel for the latest version). The code can be found in the public climate4R.indices package repository (https://github.com/SantanderMetGroup/climate4R.indices/tree/devel for the latest version). Worked-out examples of several INDECIS-QCHDS and INDECIS-ISD calculations are also available in the package wiki (https://github.com/SantanderMetGroup/climate4R.indices/wiki).

3. Impact of Euro-Atlantic teleconnections in wind power indices

Llorenç Lledó, Jaume Ramon Gamon

3.1 Objective

The aim of this section is to evaluate the existing correlations between (1) the near-surface wind speeds and wind-energy indices, i.e. Capacity Factor (CF) and Wind Power Density (WPD), and (2) the most relevant teleconnection indices in the Euro-Atlantic region.

3.2 Methods

Four atmospheric patterns and its associated teleconnection indices are derived from the four main modes of variability in the Euro-Atlantic region. These four Euro-Atlantic Teleconnections (EATC) are referred to as North Atlantic Oscillation (NAO), East Atlantic (EA), East Atlantic/Western Russia (EA/WR) and Scandinavian pattern (SCA). The four EATC patterns and indices have been computed from ERA5 500 hPa geopotential height fields employing a Rotated EOF (REOF) analysis over the Euro-Atlantic domain [90ºW-60ºE; 20ºN-80ºN]. Further details on the computation of these variability modes and the obtained patterns for each season can be found in Lledó et al. (2020) or S2S4E (2019).

The EATC indices have been correlated against station data measured at 17 tall towers distributed over a region in northern Europe [11ºW-37ºE; 47ºN-70ºN], which concentrates an important part of the installed wind power capacity in Europe. The set of tall towers is included within the Tall Tower Dataset (Ramon et al., 2020), and they measure wind speeds at heights ranging from 10 to 200 meters above the ground. In this study, only wind speeds measured at the nearest 100-metre level (which is also the typical height where modern turbine hubs are located) in each tall tower are used. All the time series span at least 6 years within the 1980-2017 period.

In addition to wind speed, we select wind-energy indices that are of interest to the wind power sector. Capacity Factor (CF) and Wind Power Density (WPD) have been computed at the 17 tall tower locations. Capacity factors have been computed for three different turbines suitable for high, medium or low wind speed conditions (CF1, CF2 and CF3). To this end, three power curves from three different turbine classes according to the IEC-61400-1 standard have been employed. CFs are derived from the nearest 100-metre tall tower winds, which are converted into power production according to the methodology described in D6.2. In the computation of the WPD, the same hub-height winds are used, and the air density is assumed to have the constant value of 1.22 kg/m^3. We note that temperature, pressure and relative humidity data (variables on which air density depends) are not always available in the Tall Tower Dataset, and no quality-control has been performed in the cases where they are. Therefore, a fixed value has been used for this parameter. Both CF and WPD indices are computed using hourly wind speed data and then averaged into seasonal values. Finally, seasonal anomalies have been computed with respect to the mean value obtained from each tower's available period.

Additionally, the experiment has been repeated using the 100m wind speeds from ERA5 for the 1980-2017 period in northern Europe, and used as a background to complete the large-scale picture of correlations where towers are not available. In this case, the reference period in which anomalies are computed is 1980-2017.

3.3 Results

Results are presented here only for the boreal winter (December-January-February, DJF). Results for the other season are similar so they are not shown here to avoid repetition. Figures 3.1, 3.2 and 3.3 show the correlations between the four EATC and the hub-height winds, CFs and WPD measured at the set of tall towers in northern Europe. In the case of CF, results for CF1 are representative for the CF2 and CF3 so that the last two are not included for discussion in the report to avoid repetition.

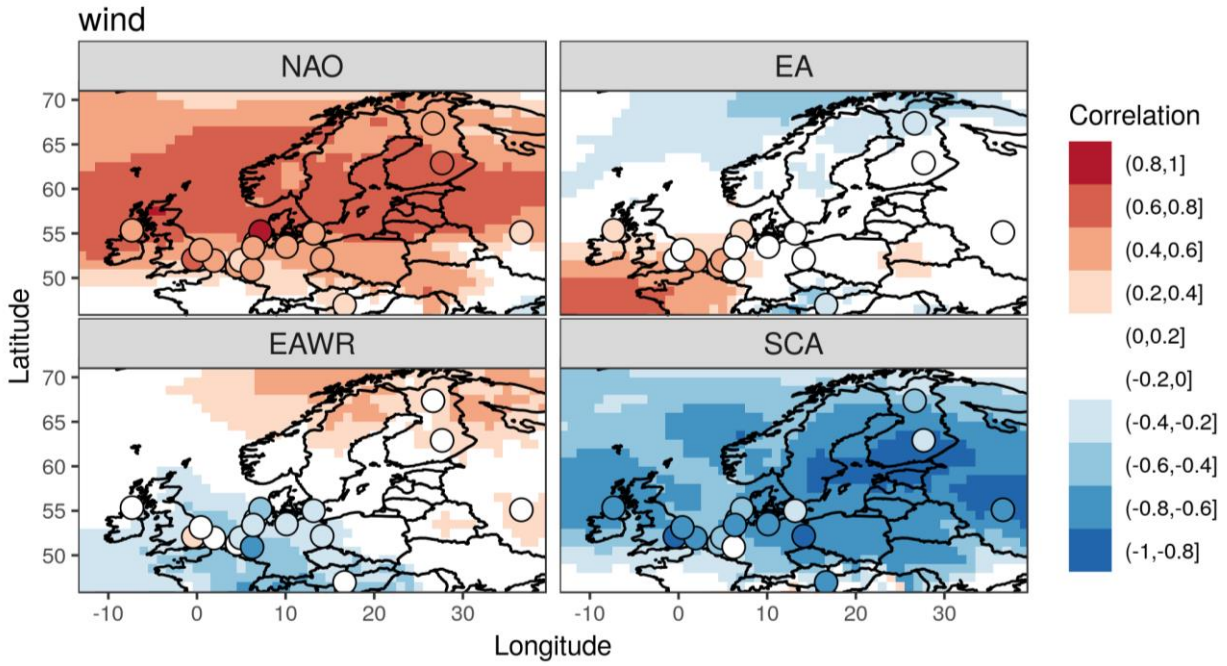

Figure 3.1.- Correlation coefficient between the teleconnection indices of the four main EATC indices and 1) the hub-height winds from the tower winds (points) and 2) the ERA5 reanalysis 100-metre winds (background).

Positive correlations between the NAO index and the three parameters are observed over the British Isles, North Sea, Scandinavia, northern Germany and the Baltic Sea countries. The correlation coefficients reach values of 0.8. Correlation coefficients at the tall towers do agree with the ERA5 correlations, being in some cases above 0.8. On the other hand, negative correlations in the region are noted for SCA, with values below -0.8. Similar values to those observed for the reanalysis are noted in most of the 17 tall tower locations.

Complementary correlation patterns are displayed for EA and EA/WR. Whereas EA shows negative correlations in the northernmost area and positive values in the southwest corner, the EA/WR depicts the positive correlations in northern Europe and the negative over the North Sea region. We also note a stripe from the northern part of the North Sea through mainland eastern Europe where there is no correlation between EA or EA/WR and the three parameters. Most of the towers in this band also display close-to-zero correlation values, thus agreeing with the ERA5 data.

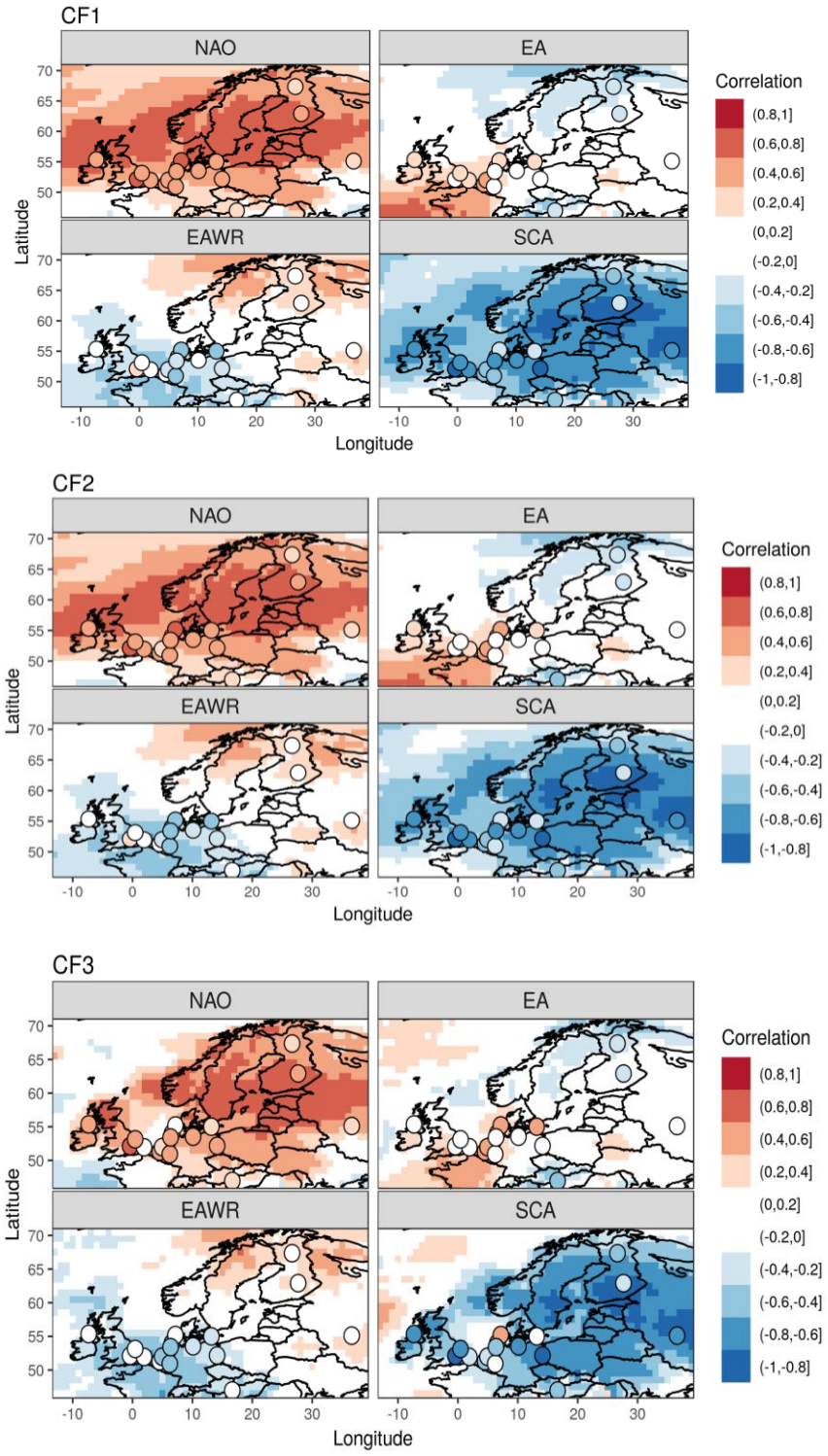

Figure 3.2.- Correlation coefficient between the teleconnection indices of the four main EATC indices and 1) the CF1 (top left), CF2 (top right) and CF3 (bottom) computed from the tower winds (points) and 2) derived from the ERA5 reanalysis 100-metre winds (background).

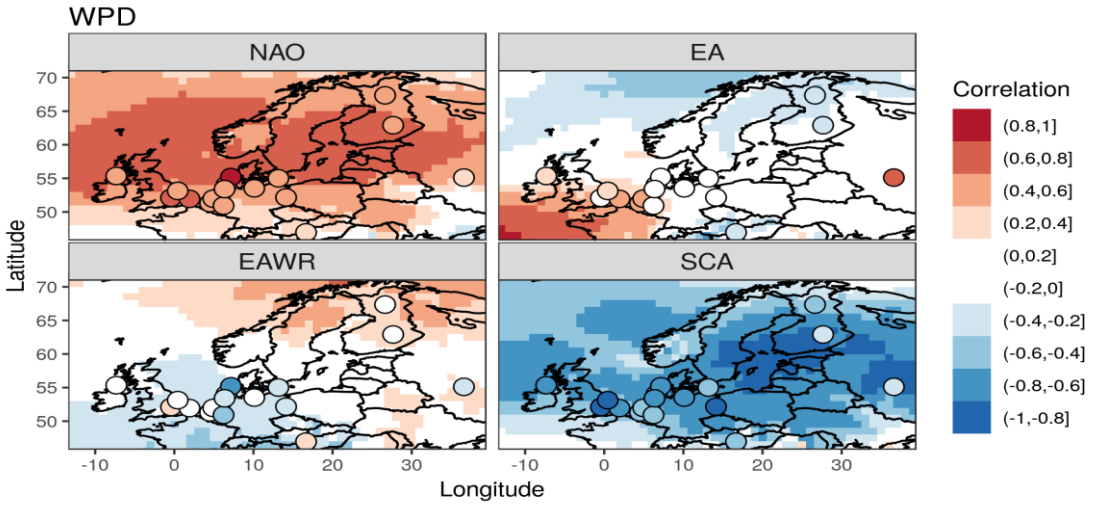

Figure 3.3- Correlation coefficient between the teleconnection indices of the four main EATC indices and 1) the WPD computed from the tower winds (points) and 2) derived from the ERA5 reanalysis 100-metre winds (background).

Overall, correlations at the tall tower locations do agree with the ones depicted by the ERA5. However, there are some specific cases where discrepancies between the two datasets are noted. This is the case, for instance, of the correlations with the three parameters and the EA/WR in the North Sea. Even though the ERA5-derived parameters correlate negatively with the EA/WR, some locations show zero or even positive correlations with the mentioned EATC. This is explained by the short length of the tall tower series, which is in some cases below 10 years.

3.4 Conclusions

There exists correlation between the four EATC indices and the hub height wind speeds, CF and WPD in northern Europe, and all three parameters show similar patterns and values of correlation. Tall tower series do confirm the correlation patterns depicted by the ERA5-derived parameters of interest, although tower series are relatively short (sometimes spanning from 6 to 10 years) and local wind phenomena are not taken into account in the reanalysis. Positive phases of NAO and EA/WR increase winds, CF and WPD in northern Europe, and so does the wind power production. Conversely, SCA and EA reduce the wind power production in the previously mentioned regions, so we can expect less production.

In this regard, accurate seasonal forecasts of the EATC indices will allow anticipating periods of higher-than-normal or below-than-normal wind energy production, and thus, scheduling activities accordingly (see D6.3).

4. Influence of blocking high pressure characteristics on INDECIS variables

Pedro M. Sousa, Ricardo M. Trigo

4.1 Summary

Blocking high pressure systems are large-scale atmospheric circulation patterns with meteorological impacts that vary across regions and seasons, depending on the blocking location, spatial characteristics and temporal length. While blocking episodes are an important component of intra-seasonal and inter-annual variability at mid-latitudes, particularly in Europe, previous studies have mostly focused on characterizing European blocking impacts on either winter or summer seasons. A characterization of Euro-Atlantic blocking occurrence within different longitudinal sectors (Atlantic, European and Russian) was performed, followed by a comprehensive analysis of seasonal impacts on temperature and precipitation regimes. We distinguished high-latitude blocking from sub-tropical ridges, which do not require a wave-breaking occurrence as blockings do, although they are frequent precursors of wave-breaking, which may lead to blocking.

This distinction clarifies that most extreme heat episodes in southern Europe and Mediterranean areas should not be attributed to blockings, but rather to ridges. In central and northern areas of the continent, both regimes are responsible for warm conditions in summer, due to enhanced radiative heating and increased subsidence. During winter, blocking and ridges lead to opposite temperature responses. Blocking reinforces cold northerly advection in its eastern flank, thus promoting European cold winter spells, especially those located in the eastern Atlantic and western Europe, while mild Atlantic flows associated to ridge patterns result in warmer conditions. Regarding the impacts on precipitation, blocking and ridges are associated with a marked north-south dipole for the three considered longitudinal sectors of occurrence. While blocking patterns force a split of the storm-track (Trigo et al., 2006), ridges are associated with a stronger zonal flow at higher latitudes. Thus, negative (positive) precipitation anomalies during blocks occur at higher (lower) latitudes. Enhanced atmospheric instability and cyclonic activity south of blocking centres relate very well with increased rainfall in southern Europe, including Iberia, where torrential regimes are more relevant in the precipitation totals. This dipole is reversed during ridges, which lead to dry conditions in southern Europe. The seasonal analysis further reveals that winters characterized by high frequencies of blocking (ridge) occurrence present above (below) average snow covered soils.

4.2 Data and Methods

Data

The E-OBS daily dataset (Haylock et al. 2008) was used to obtain a surface responses and impacts from blocking/ridge patterns occurring in distinct Euro-Atlantic sectors. This European land-only daily high-resolution gridded dataset is accessible throughout the European Climate Assessment and Dataset (ECA&D), and is available on a grid with a horizontal resolution of 0.25°×0.25°, based on the interpolation of daily observations from meteorological stations. Despite

the overall good quality of this dataset we must acknowledge that this dataset has caveats in areas where the spatial distribution of stations is sparser, such as biased and over smoothed precipitation fields in some cases, as well as possible effects in extremes.

The NCEP/NCAR reanalysis dataset (Kalnay et al., 1996), on a horizontal resolution of 2.5°×2.5° and relying on daily fields of the 500hPa geopotential height (Z500) was used to compute: i) blocking events (as in Barriopedro et al., 2006); ii) ridge events (as in Sousa et al., 2017a,b).

Methods

The methodology for the catalogue of high-latitude blocking days (hereafter only referred as blocking) was developed by Barriopedro et al. (2006). The algorithm is an adapted version of the Tibaldi and Molteni (1990) index based on the reversal of the meridional Z500 gradient around the typical latitudes of the extra-tropical jet stream. It further imposes spatial (minimum longitudinal extension of 12.5º) and temporal (minimum duration of 5 days) criteria to account for the characteristic spatio-temporal scales of blocking. The algorithm also enables the characterization of useful daily parameters, such as the location of the blocking centre, the intensity or the spatial extension. Three non-overlapping blocking sectors covering the Europe and the eastern Atlantic are defined. More specifically, daily blocking occurrences are assigned to one of the following spatial sectors according to the location of their blocking centres (Figure 1): ATL (30ºW–0º), EUR (0º–30ºE) and RUS (30ºE–60ºE).

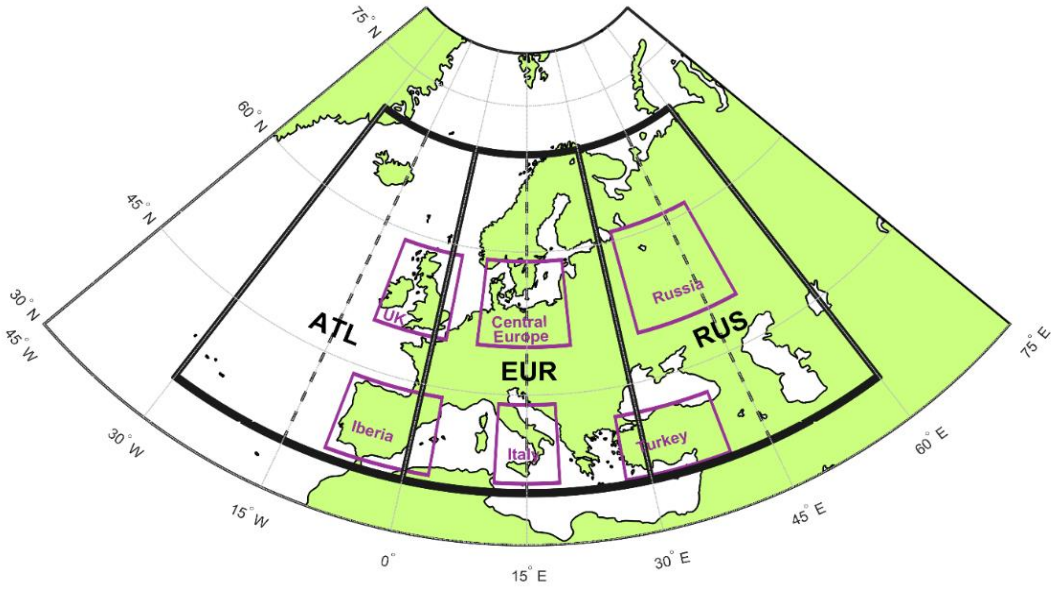

Figure 4.1 - *Geographical representation of the considered sectors for blocking location (thick black frames): Atlantic (ATL) – from 30ºW to 0º; European (EUR) – from 0º to 30ºE; Russian (RUS) – from 30ºE to 60ºE. Each of these sectors was also sub-divided into two smaller 15º longitude-wide sub-sectors (west and east, dashed black lines). Magenta boxes identify areas for regional-scale assessments.*

The catalogue of sub-tropical ridge days (hereafter only referred to as ridges), we followed a similar methodology as in Santos et al. (2009), which is based on daily anomalies of the Z500 field. In order to compare it with the blocking catalogue, we also classified ridge occurrence into the same three sectors (ATL, EUR and RUS). Furthermore, each sector is split into two halves: south (30ºN–50ºN) and north (50ºN–70ºN). These latitudinal bands are used for winter and, to

accommodate the annual cycle, they are shifted 5º northward for summer. This partition enables classifying ridges as strong Z500 positive departures in sub-tropical and mid-latitudes that do not extend significantly northwards, thus avoiding overlapping days between blocking and ridge patterns. For each grid point, we computed Z500 departures for each specific day, and a 30-day running threshold based on the 80th percentile of the daily Z500 series. We then obtained, on a daily basis and for each longitudinal sector, the percentage of area above that threshold in its northern and southern halves. To classify a ridge day in one of the three considered longitudinal sectors the following criteria are employed: 1) at least 75% of the area in the southern half is above the threshold; 2) less than 50% of the area in the northern half is above the same threshold. These percentages and thresholds were tested and calibrated in order to avoid overlaps between blocking and ridge dates, and furthermore to obtain climatological ridge frequencies comparable to previous studies (e.g. Santos et al. 2009).

4.3 Results

Blocking and ridge climatology

The classification of blocking/ridge events according to their location provides a simple and objective way of grouping blocking and ridge days (and consequent impacts) in each sector and season. The winter and summer composites of Z500 anomalies for blocking and ridge days of each sector are shown in Figure 2. Overall, blocking and ridges display a clear difference in the latitude of their maximum Z500 anomalies. The composites for blocking days show an omega-like structure, which is distinguishable from the non-wave-breaking pattern that is evident in the composites corresponding to days of ridge. Absolute anomalies are larger in winter, and for blocking regimes (Figures 4.2a-c). During winter, around one third of the days comprises blocking occurrence in at least one sector of the Euro-Atlantic region, while summer frequencies are smaller, particularly over the ATL sector (Figure 2g). It is worth noting that some events contribute to the composites of more than one sector during their lifecycle (Sousa et al., 2017a). Ridge frequencies are more equally distributed throughout the three sectors and seasons, with closer values to those of blocking during summer. The largest Z500 anomalies under ridging patterns are found for ATL and EUR ridges in winter (Figures 2d-e). The positive anomalies of Z500 during blocking and ridge regimes are often accompanied by negative anomalies, but much less pronounced. The most relevant negative Z500 anomalies occur southwards (northwards) of the blocking (ridges) centres, mainly for ATL structures in winter (Figures 2a and 2d).

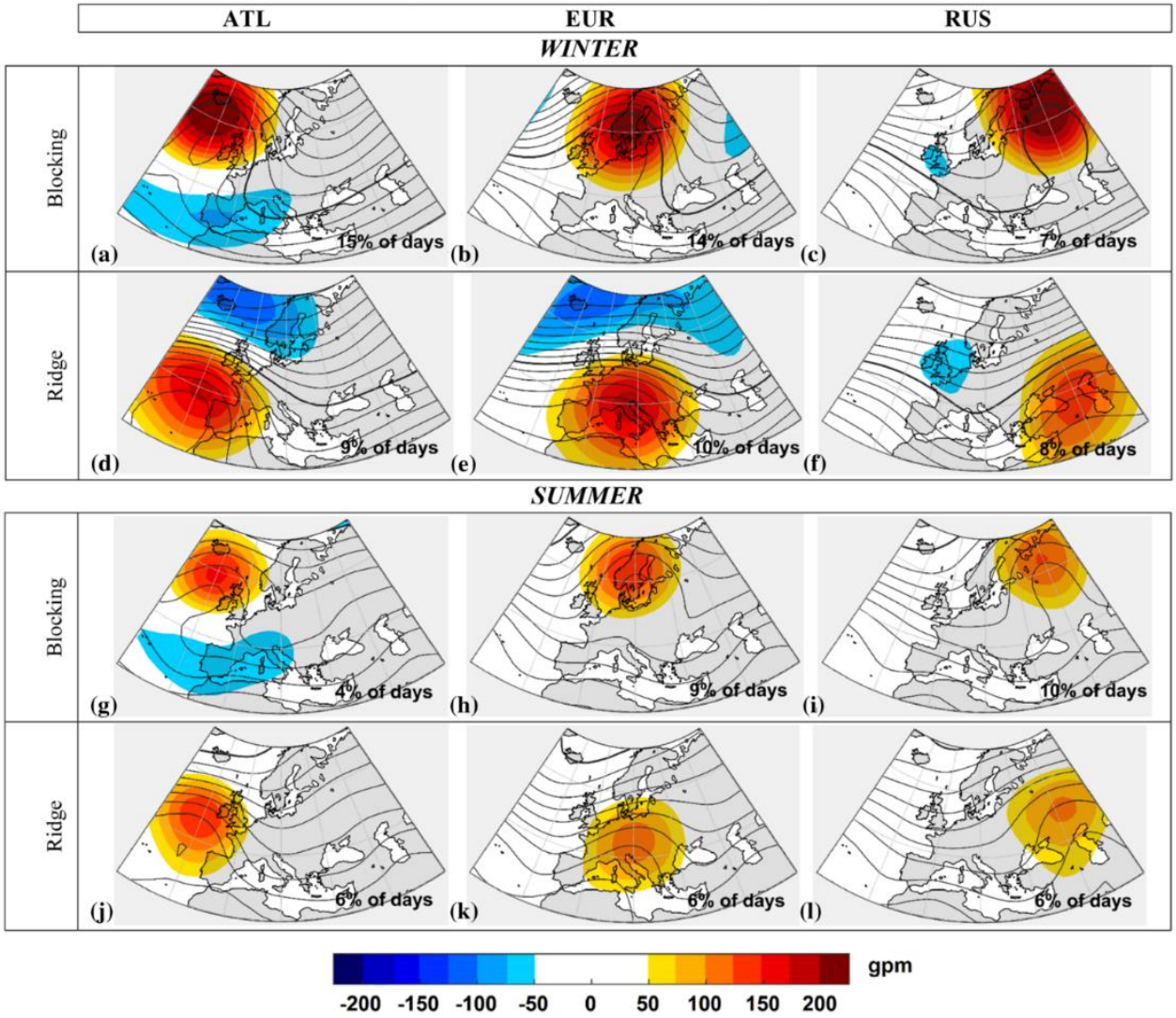

Figure 4.2 - *Composites of the daily anomalies (shaded areas) and absolute values (contours) of 500 hPa geopotential height for blocking centres and ridges in each sector, during winter (upper panels, a-c and d-f, respectively) and summer (lower panels, g-i and j-l, respectively). All values are in gpm and the thick line represents the 5500 isohypse (the thinner contours are separated by 50 gpm*

Impacts in near-surface temperature

Composites for the maximum and minimum T2m anomalies (TX and TN) were computed for blocking/ridge patterns, as shown in Figures 4.3-4.4.

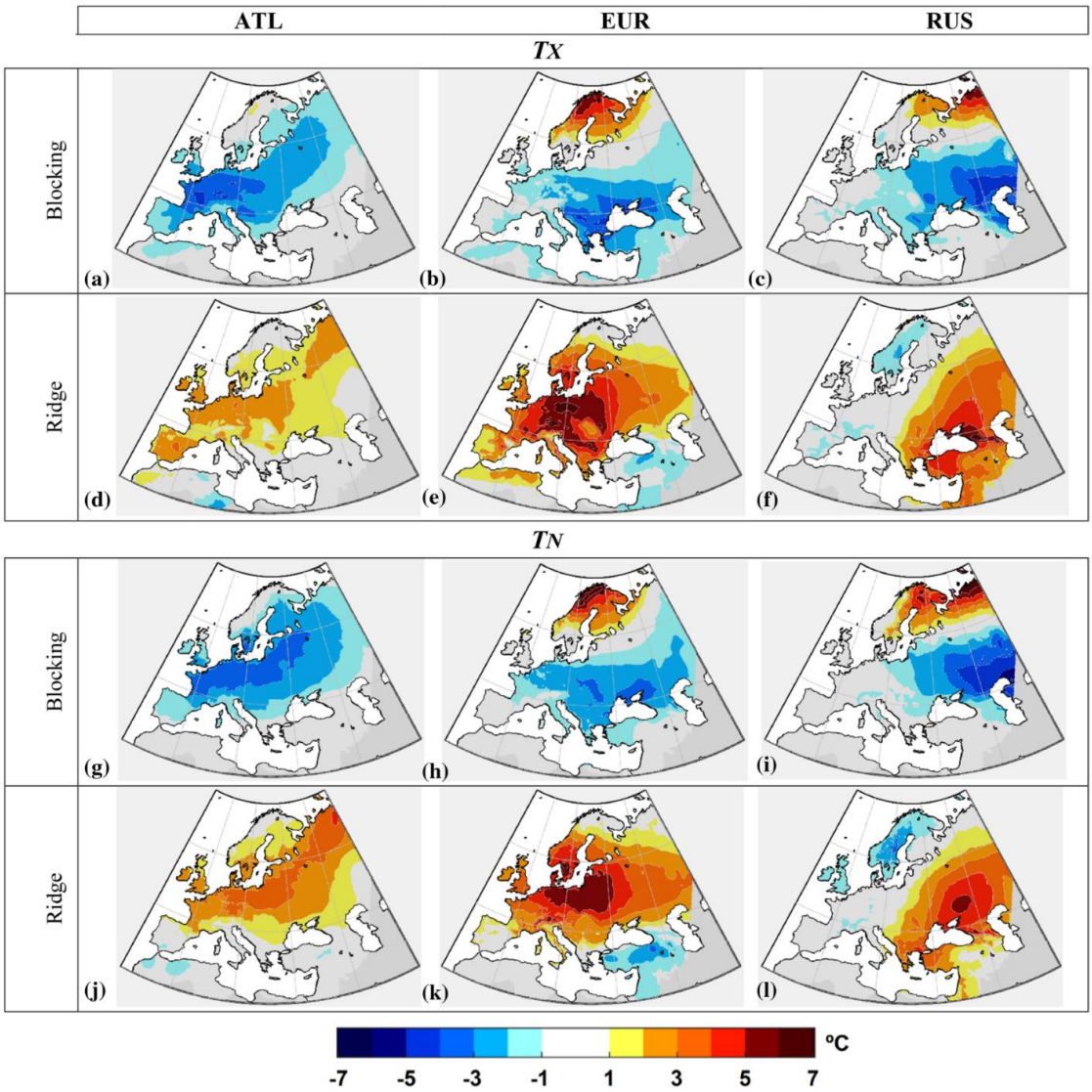

Figure 4.3 - *Composites for blocking and ridge days occurring in each sector (ATL, EUR and RUS) of winter 2 meters above ground maximum (upper panels – TX; a-c and d-f, respectively) and minimum temperature (lower panels – TN; g-i and j-l, respectively) anomalies (in ºC). Only statistically significant anomalies at 5% significance level are depicted.*

The responses in TN and TX are generally coherent for all sectors and regimes during winter (Figure 3) and reveal highly contrasting patterns between blocking and ridges (Figure 3a-c versus 3d-f). While during blocking most of Europe experiences well below average temperatures, ridge days are characterized by extensive above average temperatures. Negative anomalies exceeding -3ºC tend to occur southward and eastward of the blocking centres, with ATL blocking (Figure 3a) revealing the largest widespread signal over the continent. During blocking episodes, strong positive temperature anomalies are found in land areas under the highest Z500 anomalies (Figure 2), i.e. northern half of Scandinavia for EUR blocks (Figure 3b) and northern Russia and eastern Scandinavia for RUS blocks (Figure 3c). Conversely, winter ridges in both ATL and EUR sectors (Figures 3d-e) are responsible for anomalously warm conditions in almost all regions of Europe. These anomalies are particularly striking for EUR ridges (Figure 3d), when Central Europe

experiences positive TX anomalies reaching up to 7ºC. There are some areas on the ridge's eastern and western flanks that experience slightly below average temperatures, mainly in Mediterranean regions. This is particularly noteworthy in Turkey during EUR ridges (Figure 3k), though these negative anomalies are smaller in magnitude and spatial extension than their positive counterparts.

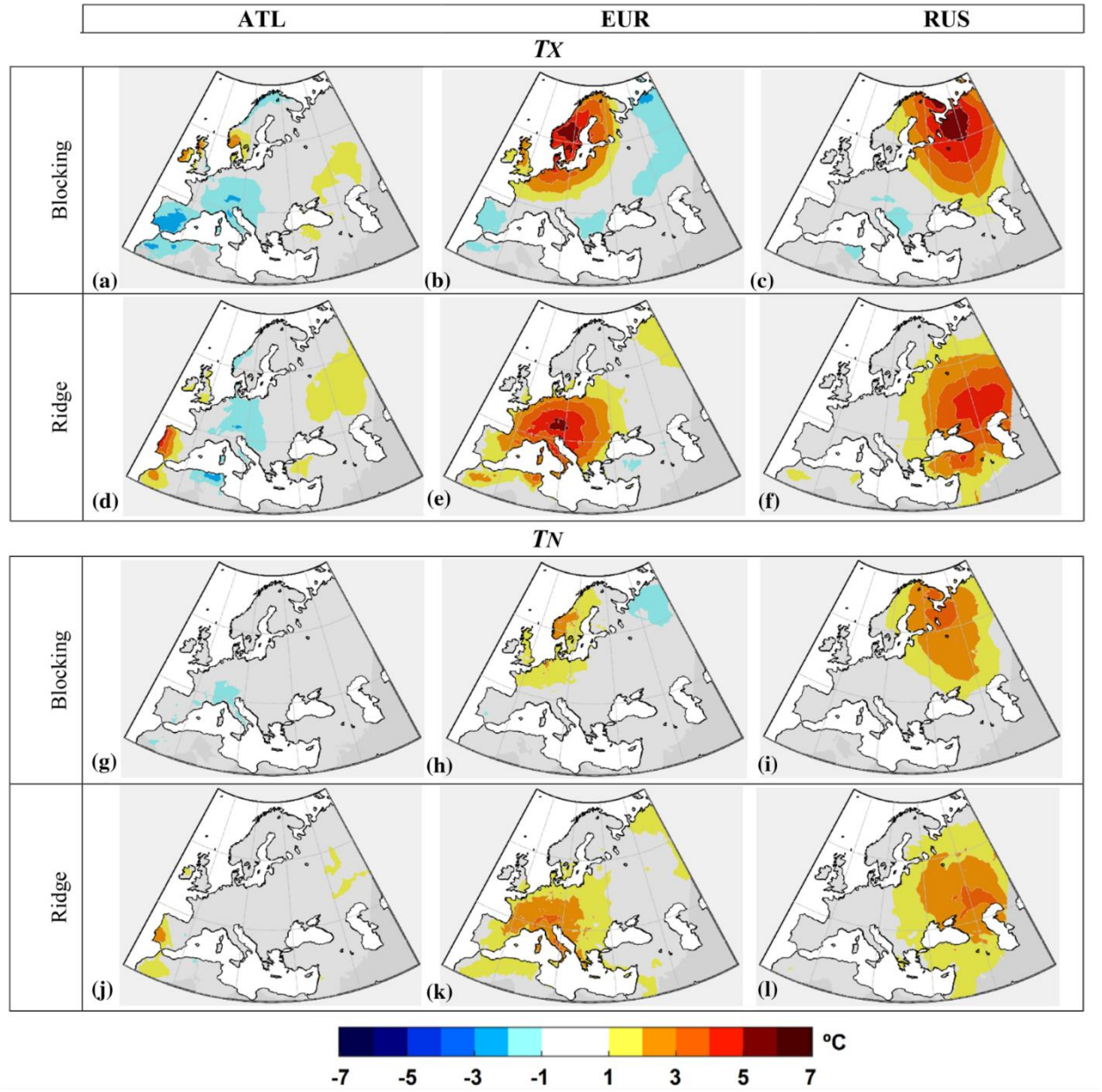

Figure 4.4 - *Same as in Figure 4.3, but for summer.*

Summer temperature anomalies (Figure 4.4) are more spatially confined than in winter and the opposite temperature response to blocking and ridge patterns is no longer observed (Fig. 4.3). In the case of blocking systems, positive anomalies are again cantered under the maximum Z500 anomaly area, but now affecting larger areas. In particular, during EUR (RUS) blocking, extensive areas of Central Europe and Scandinavia (Eastern Europe and Russia) experience anomalously

warm conditions, with TX anomalies >5ºC, as seen in Figure 4.4b (4.4c). Temperature anomalies for ATL blocking (Figure 4a) are much less pronounced. As in winter, negative anomalies are found in the southern and eastern flanks of blocking systems, but they are small in magnitude, and mostly restricted to TX. Still, southern areas of Europe (e.g., Iberia, Balkans) display negative temperature anomalies associated to blocking in both winter and summer.

Summer ridges are associated with above normal temperatures over a more confined area than in winter. In particular, they do not have significant effects in temperature over northernmost areas of Europe during this season. The lack of opposite signed responses to blocking and ridges during summer is evident in some regions, such as Central Europe or Russia, which experience above average surface temperatures under both regimes, albeit at different latitudes. On the contrary, in southernmost areas, particularly the Iberian Peninsula for ATL regimes (Figures 4.4d and 4.4a), positive temperature anomalies during ridge days tend to be replaced by negative anomalies during blocking days. Furthermore, summer temperature anomalies over southern Europe critically depend on the specific location of ridge structures. For detailed regional changes in temperature distributions and other related indices can be found in Sousa et al. (2017b).

Impacts in precipitation

The precipitation anomalies due to the presence of blocking patterns (Figure 4.5) clearly show that these systems lead to well below average precipitation in a wide region under direct influence of the anticyclonic circulation. Very significant decreases are recorded in northern Europe near the blocking centres, which correspond to an almost complete cease of precipitating days. On the other hand, the increase in precipitation in southern areas of Europe, as well as in the Atlantic strip of Scandinavia matches well with the negative Z500 anomalies, which arise northward and southward of blocking systems (as shown in Figure 4.2). These positive precipitation anomalies shift eastwards with the considered blocking sector, thus following the relative west-east migration of the maximum Z500 positive anomaly (blocking centre). Throughout the year, the widest positive anomalies tend to be found for ATL blocks, during which a large portion of Iberia, Italy, and the Balkans region experience wetter than usual conditions. In particular, daily precipitation rates in southeastern Iberia almost double. When blocked conditions prevail, precipitation losses in the UK and central Europe are quite striking throughout the year. Summer precipitation responses often embrace lower spatial extensions; however, in some regions such as the northernmost countries and Russia–where precipitation in form of snow is often recorded in colder months–the summer deficits represent the largest changes.

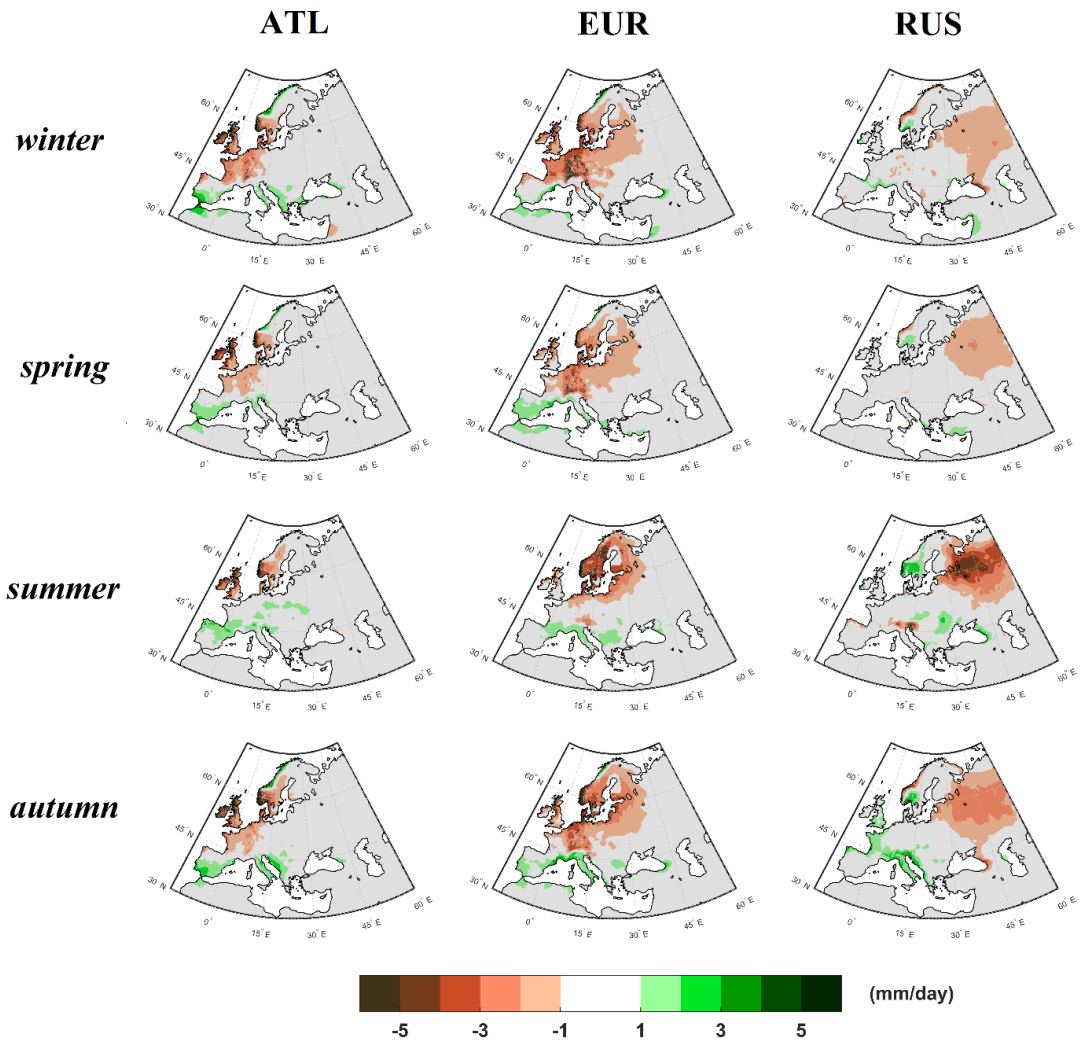

Figure 4.5 - *Seasonal composites for the differences in absolute anomalies of daily precipitation (mm/day) between blocking and strong zonal flow patterns in the ATL (left column), EUR (central column) and RUS (right column) sectors. Only statistically significant anomalies at the 5 % level are shown.*

Similarly to the temperature anomalies (particularly during winter) presented previously, the precipitation changes during ridge days (Figure 4.6) are essentially opposite to those found for blocking days (Figure 4.5). Consequently, ridge patterns are associated to well below average precipitation in most of southern areas, where the maximum geopotential height anomalies of the ridge structure are located (see Figure 4.2). The exception can be found during RUS ridges, which are located sufficiently to the east to still allow the penetration of Atlantic frontal systems towards the Mediterranean area. It should be noted that ridge patterns tend to enhance zonal flow northwards, thus explaining the positive rainfall anomalies at higher latitudes, contrarily to the very striking drying pattern found for southern Europe. A more detailed characterization of changes in precipitation distributions at the regional scale can be found in Sousa et al. (2017a)

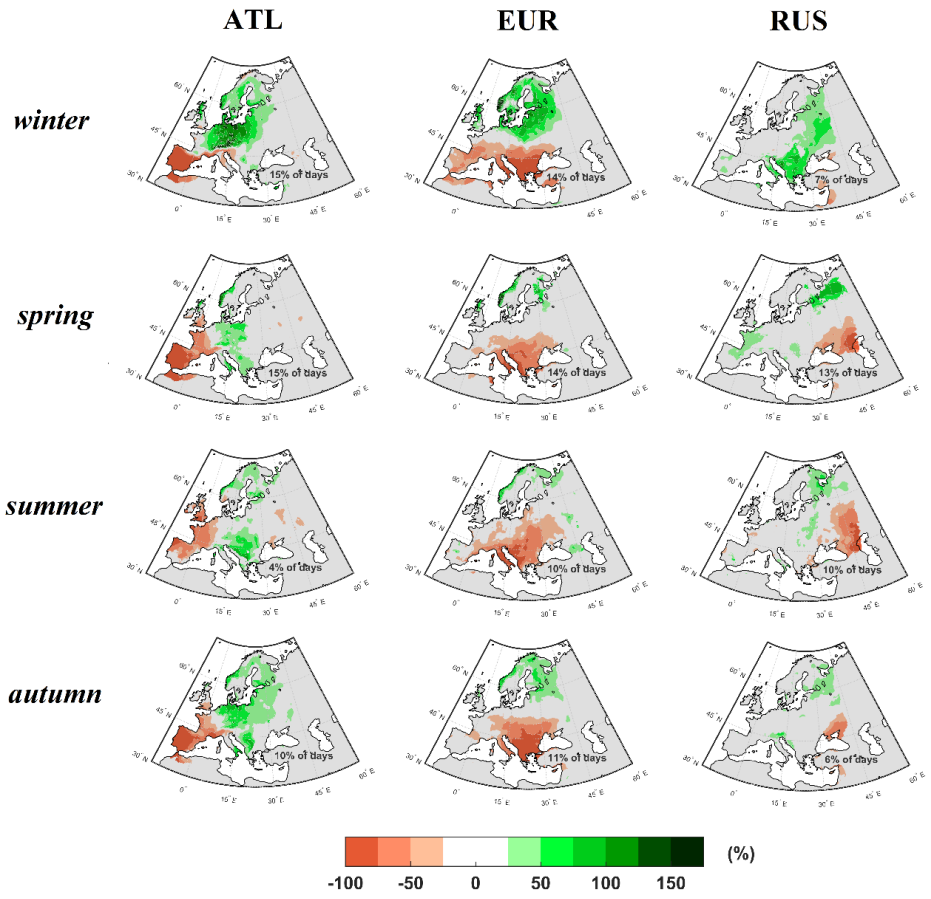

Figure 4.6 - *Seasonal composites of daily precipitation anomalies (%) in Europe for ridges in the ATL (left column), EUR (central column) and RUS (right column) sector. Only statistically significant anomalies at the 5 % level are shown (two-sample Kolmogorov-Smirnov test).*

4.4 Data availability

Blocking and Ridge catalogues

The catalogues computed for blocking and ridge days, computed using the NCEP reanalysis, are available in their full content (with blocking/ridge metrics, such as extension, intensity, etc.) for different reanalysis datasets (NCEP, ERA-Interim, 20CR, ERA20C), and also in summarized simple ".txt" files, containing simply a list of days where each one of these regimes affects each one of the regional sectors presented in Figure 1. In fact, these summarized files provide central locations for blockings and ridges in finer 15º longitudinal sectors: i) ATL_W (30ºW-15ºW); ii) ATL_E (15ºW-0ºW); iii) EUR_W (0ºE-15ºE); iv) EUR_E (15ºE-30ºW); vi) RUS_W (30ºE-45ºE); vii) RUS_E (45ºE-60ºE).

In Table 4.1, an example of the summarized catalogue for ridges affecting the European sectors (EUR_W and EUR_E) is presented. All summarized catalogues for each of the patterns, and for each sub-sector, will present the same format (yyyy / mm / dd).

All these complete/summarized catalogues can be provided by us for the datasets referred above.

Table 4.1 – List of days where a ridge was present over the European sectors (0ºE-30ºE) during the 2010 summer, according to the NCEP reanalysis.

EUR_W ridges			**EUR_E ridges**		
yyyy	mm	day	yyyy	mm	day
2010	6	6	2010	6	11
2010	7	4	2010	6	12
2010	7	7	2010	6	16
2010	7	8	2010	7	12
2010	7	9	2010	7	15
2010	7	10	2010	7	18
2010	7	11	2010	7	23
2010	7	16	2010	8	1
2010	7	18	2010	8	2
2010	7	19	2010	8	15
2010	8	20	2010	8	16
2010	8	21	2010	8	21
2010	8	22	2010	8	22

4.5 Conclusions

- The disentanglement of high-latitude blocking and low-latitude subtropical ridges is crucial to comprehend a large part of surface impacts in variables such as temperature and precipitation (and other derived), as these patterns reveal a north-south dipole.
- Blocks tend to be the main driver for extreme cold winter events throughout most of Europe, while ridges relate to well above average temperature conditions and mild winters;
- During summer, both patterns related to warm events, however, in Mediterranean areas, ridges are the pattern responsible for extreme heat events;
- Rainfall in Southern Europe is strongly affected by these patterns, as strong ridges lead to extremely dry conditions, with almost complete absence of rainfall for prolonged periods;
- On the other hand, while blocks tend to cause drier conditions in Central/Northern Europe, they are associated with increased instability and increased frequency of extreme precipitation events in Southern Europe;
- Further analysis of more indices may and should be performed regarding the impacts of these patterns, at several time-scales and regional scales for further sensibility tests;
- Links with other atmospheric circulation features (e.g. Atmospheric Rivers) should be crucial for a deeper understanding of the role that atmospheric blocking plays over the European climate, and how it will relate with future changes. These will be explored in the short-term.

4. A local analysis of teleconnections: a case study for Calabria (Southern Italy).

Roberto Coscarelli, Tommaso Caloiero, Giulio Nils-Caroletti

4.1 **Overview**

Teleconnection indices provide a numerical evaluation of the state of major circulation patterns (Glantz, 1994). They allow an easy way to quantify and measure teleconnections and their influence on local weather and climate through the use of statistical approaches, for instance like various correlation methods. Among the most used approaches is the comparison of seasonal teleconnection patterns with seasonal variables, e.g. seasonal cumulated precipitation. A teleconnection index is thus compared to the precipitation at a grid point or at a weather station.
However, one of the main limitations of this method is its low statistical significance. If we consider, for instance a 30-years period (the usual time frame for climate analysis), we have only 30 values over which correlation is calculated.
In the following report, we present an alternative approach that aims to improve statistical significance through the use of data from grid points or gauge stations all belonging to the same climatic zone. Although this approach has been developed for a scenario of precipitation over Calabria (southern Italy), this is a general method that can be reproduced for any variable for any area of Europe, as long as local climates/micro-climates regional boundaries are available.

4.2 **Methodology**

This method builds on the premise that large scale influences over a small region characterized by the same climatic pattern will be, at first approximation, very similar. If we take into account precipitation, for instance, various studies have shown that in many cases, especially at mid-latitudes, while orography and local wind patterns can influence the distribution of precipitation, these will work to modify a general large scale moisture transport which is essentially the same for a certain region (e.g., Smith and Barstad, 2004). This is especially true for orographic precipitation, but local features are also prominent in determining the formation of local convective clouds.

Given this premise, for each time step (i.e., a season or a month) we can consider the measure taken at every station or grid point belonging to the same local scale region with the same climate as an independent measure of the same quantity.
For instance, if we have five stations over a region, we will have for each time step five measures instead of one, changing the number of measures used in correlation studies over a 30-years time frame from a mere 30 to 150, and greatly increasing the statistical significance of our results.

For our case study, we used 79 rain gauges in Calabria (southern Italy), subdivided into 5 distinct homogeneous areas, characterized by different climatic conditions (Brunetti et al. 2012). The subregions were obtained through the Principal Component Analysis, including

Varimax Rotation of the Principal Components, applied to the correlation matrix of daily records, thus obtaining five Rain Zones (RZs). Specifically, the first five Empirical Orthogonal Functions on the western, Tyrrhenian side, which explained 90% of the total variance, led to the identification of the North-Western Zone (T1) and the South Zone (T2). On the eastern, Ionian side, the North Zone (I1), the Central Zone (I2) and the South Zone (I3) were identified (Figure 1).

We compared the monthly and seasonal cumulated precipitation from 1981 to 2010 with 7 teleconnection indices indicated as the most relevant for European climate (Deliverable 5.1). The teleconnection indices used were NAOI (North Atlantic Oscillation index (see e.g., Barnston and Livezey, 1987; Lamb and Peppler, 1987), MOI (Mediterranean Oscillation index; Conte et al., 1989; Palutikof et al., 1996), ONI (Oceanic Niño index, a measure of El Nino-Southern Oscillation, ENSO; Ropelewski and Halpert, 1996), WeMOI (Western Mediterranean Oscillation index; Martin-Vide and Lopez-Bustins, 2006), EA (East Atlantic pattern index; Barnston and Livezey, 1987), EA/WR (East Atlantic/West Russian pattern index; introduced as Eurasia-2 in Barnston and Livezey, 1987) and SCAND (Scandinavian pattern index; introduced as Eurasia-1 in Barnston and Livezey, 1987).

We compared monthly teleconnection indices, from the current month and from all the previous 11 months, with monthly cumulated precipitation. We also compared seasonal teleconnection indices for the four previous seasons with seasonal cumulated precipitation. It was also possible to compare seasonal teleconnections with monthly precipitation.

The correlation technique used was Pearson correlation. The significance threshold was set to 95%. Only correlation values with significance over the threshold were taken into account for the results.

5.3 Results

Seasonal teleconnection indices vs seasonal cumulated precipitation

The most used correlation studies compare seasonal indices with seasonal cumulated precipitation. In Table 5.1, the highest correlation values for each season-season combination are shown. Results show that all teleconnections except ENSO are especially relevant to the five climate regions. A few indices were particularly relevant: EA for the Ionio 1 region; MOI for the Ionio 2 and Ionio 3 Rainfall Zones; WeMOI, EA and MOI on the Tirreno 1 RZ; NAO, WeMOI and EA for the Tirreno 2 RZ.

Three-months teleconnection indices vs monthly precipitation

Correlation results for three-months averaged indices against monthly precipitation are shown in Table 5.2. The year was broken down into three-month periods that didn't equate with stations, but were instead taken as the previous three months (current month included) and then the following three-month periods (from 3 to 5 months back, 6 to 8 months back and 9 to 11 months back).

The NAO and MOI emerged as the most relevant teleconnection indices for the Ionian

rainfall zones, both in magnitude and in frequency. Contributions from ENSO and EA were not negligible either in the Ionio 1 RZ; however, it must be pointed out that they were relevant for the May-September months which have the least precipitation, and the correlation never topped 0.3 in magnitude. EA/WR was important for Ionio 2 and WeMOI for Ionio 3.

For the west coast RZs, MOI was relevant to precipitation from January to April; EA/WR and WeMOI were also important teleconnections for Tirreno 1, while relevant correlations with precipitation in Tirreno 2 was also found with NAO (October, December) and ENSO (November).

Monthly teleconnection indices vs monthly cumulated precipitation

At last, we compared monthly precipitation with teleconnection indices from the current month and back for the other 11 previous months (examining thus a full year, including current month; Table 5.3).
As the number of correlation products is huge, we examined the number of instances in which each teleconnection index has the strongest correlation to monthly precipitation, and in how many of these instances the correlation is strong (i.e., equal or above 0.4). This is shown in Table 4. From this examination, the Mediterranean Oscillation emerges as the most important teleconnection for the various Calabrian climate regions. The North Atlantic Oscillation is often represented, although its magnitude rarely rises above 0.4.
We also performed this examination taking into account an extended period from 1951 to 2010, to determine whether there as a change in teleconnection patterns during the last sixty years. What emerged (Table 5.5) is that teleconnections seem to have become more correlated precipitation in Calabria in general, although the proportion of relevant correlations with precipitation among teleconnections remains roughly the same. Thus, we searched for an explanation to this finding by calculating also the 1951-1980 results for the correlation and comparing the relevant teleconnections between the two time periods (Table 5.6). The changes seen between the 30-years time frames mean that the 1951-2010 correlations seem weaker because results from the two time periods cancel out over the 60-years correlation analysis.
A possible explanation for the changes is that the change in local precipitation response to global teleconnections have changed either as a result of natural variation or because of climate change.

Precipitation anomalies as a regression function of teleconnections

Considering the relevant number of significant correlations, it might be possible to devise a predictive function through multivariate regression to estimate average regional seasonal precipitation anomaly. The variables used would only be the teleconnection anomalies which correlate with a magnitude over 0.2.
For instance, in the case of Tirreno 2 RZ, these correlations and the relevant teleconnections are shown in Table 7.

The regression function would thus become:

$$SON\ precipitation\ anomaly = EA/WR\ (SON-1) * k1 + WeMOI(SON-2) * k2 + NAO(SON-3) * k3 + MOI(SON-3) * k4 + residual$$

where k1, k2, k3 and k4 are the regression coefficient for the variables taken into account.

For the SON months (autumn), 'season -1' is JJA (summer), 'season -2' is MAM (spring) and 'season -3' is winter (DJF). Thus, the regression function becomes:

$$SON\ precipitation\ anomaly = EA/WR(summer) * k1 + WeMOI(spring) * k2 + NAO(winter) * k3 + MOI(winter) * k4 + residual$$

5.4 Conclusions

Many studies have been published about the possible influence of teleconnections on European and especially Mediterranean precipitation (e.g., Fraedrich, 1994; Kutiel et al., 1996; Maheras et al., 1999; Mariotti et al., 2002; Dunkeloh and Jacobeit, 2003; Xoplaki et al., 2004; Shaman and Tziperman, 2011; Shaman, 2014). In this work, correlation analysis at different time frames of influencing teleconnections with different accumulated precipitation times show that the Mediterranean Oscillation is the most significant and important teleconnection with regard to precipitation in Calabria for the 1981-2010 time period.

Results have also shown that from 1951-1980 to 1981-2010 there is a clear change in the correlation between the various teleconnection indices and the regional precipitation values, for instance in the impact that the Mediterranean Oscillation has on the East coast precipitation (Ionio RZs 1-3). Further studies should be required to suggest a possible explanation of the phenomenon.

The method has been used for a network of sparse rain gauges, however similar studies can easily be performed on regular gridded data networks like ECA&D. Although the method has been used for precipitation in Calabria, it can be easily applied to any meaningful climatic variable which is relevant at the local level, e.g., temperature, and at any region in Europe, provided that the assumption of similar climate patterns (i.e., belonging to the same microclimate or local climate region) holds true.

It is also possible to compare results from this method with other tests: for instance, one could compare correlation from individual stations and/or grid points with the results from regional correlation. On one hand, this can test how well the assumption of similar climate holds for a region; on the other, it can find outlier stations or grid points that can point out for some special feature of the area taken into account. Another possibility is to perform a different regional correlation by averaging precipitation over a climatic area and compare the results with those from the regional correlation introduced in this study

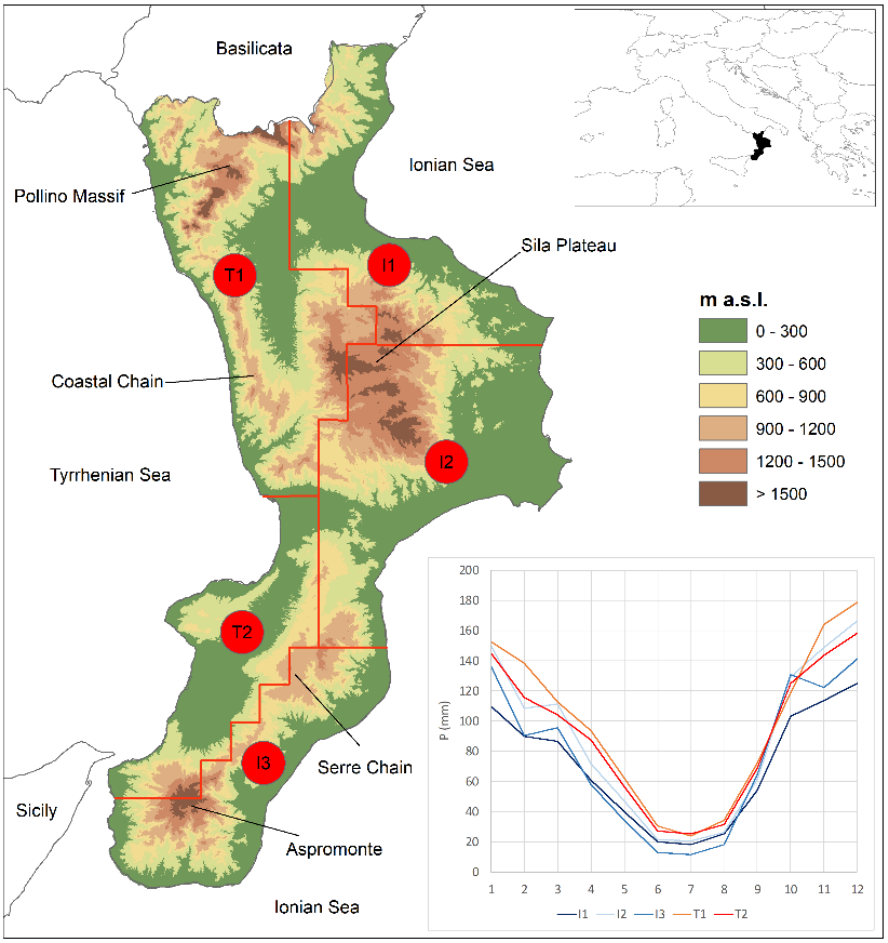

Figure 5.1. The five Rainfall Zones over a DEM of Calabria.

Ionio 1		RAIN	SEASON	
TELECON	DJF	MAM	JJA	SON
MONTHS				
Season -1	0	-0,27	-0,24	0
Season -2	0	-0,21	0	0
Season -3	0	0,31	0	-0,23
Season -4	0,26	0	0,47	0

Ionio 2				
TELECON	DJF	MAM	JJA	SON
MONTHS				
Season -1	0,22	-0,23	0	0
Season -2	-0,24	-0,21	0	-0,26
Season -3	-0,29	-0,32	0	-0,36
Season -4	0,24	0	0,24	-0,22

Ionio 3				
TELECON	DJF	MAM	JJA	SON
MONTHS				
Season -1	0	-0,22	0	0
Season -2	-0,22	0	0	-0,26
Season -3	-0,24	-0,23	0	-0,32
Season -4	0,26	0	0	-0,25

Tirreno 1				
TELECON	DJF	MAM	JJA	SON
MONTHS				
Season -1	0	-0,26	0	0
Season -2	-0,26	-0,24	0	-0,33
Season -3	0	-0,24	0	-0,44
Season -4	0	0	0,4	0

Tirreno 2				
TELECON	DJF	MAM	JJA	SON
MONTHS				
Season -1	0	0	0,21	-0,29
Season -2	-0,23	0	-0,26	-0,35
Season -3	0	0,29	-0,24	-0,29
Season -4	0	0	0	0

NAO
MOI
ONI/ENSO
WeMOI
EA
EA/WR
SCAND

Table 5.1. Seasonally averaged teleconnections with the highest, significant correlation with seasonally accumulated precipitation. Calabria rain zones, 1981-2010.

PRECIPITATION MONTHS

Ionio 1

TELECON MONTHS	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current to month-2	-0,47	0	-0,25	-0,37	-0,24	0	0,23	0,27	-0,3	-0,22	0	0,28
Month -3 to -5	-0,24	0	0	0	0,22	-0,22	-0,3	0,3	-0,25	-0,3	0	0
Month -6 to -8	0	0	0,36	0	-0,26	0,21	-0,22	0,24	-0,24	0	0	0,35
Month -9 to -11	0	0	-0,38	0	0,22	0,26	0,26	0,37	-0,26	-0,28	0,33	0,27

Ionio 2

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current to month-2	-0,47	0	-0,32	0	0,32	0,22	0,2	0	-0,22	-0,22	0	0,21
Month -3 to -5	-0,24	0	0	0,23	0,22	0	0	-0,21	-0,3	0	0	-0,27
Month -6 to -8	-0,2	0,24	-0,41	0,23	-0,25	0,23	0	0	0,29	0,33	0	-0,3
Month -9 to -11	0	-0,27	-0,45	0	0,21	0	-0,2	0	-0,29	-0,23	0,25	0,22

Ionio 3

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current to month-2	-0,46	0	-0,35	0	0,25	0	0,2	0	-0,25	-0,23	0	0
Month -3 to -5	-0,28	0	0	0	0,29	-0,23	0	0	-0,26	0	0,25	0
Month -6 to -8	-0,21	0	-0,22	0	0	0	0	0	0,26	0,32	0	0
Month -9 to -11	-0,24	-0,24	-0,46	0	0,21	-0,21	-0,23	0,2	-0,25	-0,32	0,22	0,23

Tirreno 1

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current to month-2	-0,45	-0,25	-0,36	-0,2	0,27	0	0	0,21	0,28	-0,32	0	-0,23
Month -3 to -5	0	0	0	-0,26	0,24	-0,2	0	0	-0,35	-0,3	0	0
Month -6 to -8	0	0,22	-0,34	-0,24	-0,22	-0,29	0	0	-0,31	-0,22	0,22	-0,34
Month -9 to -11	0	-0,33	-0,3	-0,23	0,21	0,23	0,23	0,23	0,3	0	-0,25	0,28

NAO
MOI
ONI/ENSO
WeMOI
EA
EA/WR
SCAND

	Tirreno 2											
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current to month-2	-0,36	-0,26	0	0	0,27	0	0,23	0	0,26	-0,35	0,22	0
Month -3 to -5	-0,22	0	0	-0,24	0,25	-0,35	0	-0,2	-0,29	-0,23	0,22	0
Month -6 to -8	-0,22	0,21	0,28	0	0	0,25	0	0	0,25	-0,21	0,22	0,26
Month -9 to -11	0	-0,31	-0,39	-0,29	0	-0,27	-0,21	0,33	0	0	0,26	0,25

Table 5.2. Three-months averaged teleconnections with the highest, significant correlation value with monthly precipitation. Calabria rain zones, 1981-2010.

Ionio 1

PRECIPITATION
MONTHS

TELECONNECTION MONTHS	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,49	-0,4	0,43	0,25	-0,31	0	-0,24	-0,43	0	-0,25	-0,25	0
Month -1	-0,27	0	-0,35	0,25	0,31	0,3	0,34	0	0	0	0	0
Month -2	0,25	0,42	-0,25	0	0,24	-0,22	0	-0,27	-0,3	0	0	0
Month -3	0,25	0	0	-0,31	0,23	0,34	-0,28	-0,29	0,27	0	0	0,23
Month -4	0	0	0,2	0,23	-0,25	-0,31	0,22	-0,23	-0,2	0	0	0
Month -5	-0,31	0	0	0	0	0	0	0,25	-0,32	0	-0,24	0
Month -6	0	0	0,36	0,26	0,23	0	0,23	0	-0,29	0	-0,32	0
Month -7	0	-0,23	0	0,21	0,35	-0,22	-0,29	0,27	-0,28	0,24	0	0,22
Month -8	0,27	-0,23	-0,36	-0,27	-0,24	0	-0,3	0,21	-0,28	-0,35	0,29	0,21
Month -9	0	0	0,34	0	0	0,22	0,24	0,31	-0,29	-0,27	0,21	-0,27
Month -10	0	0	0,27	0	0,23	-0,28	-0,22	-0,24	-0,25	0,25	0	0,24
Month -11	0,25	0,31	0	-0,21	0,28	0,26	0,41	-0,24	-0,24	-0,31	-0,24	0,3

Ionio 2

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,49	-0,47	0,47	0,24	0,33	-0,22	-0,22	-0,36	0	0	-0,33	-0,26
Month -1	-0,28	0	-0,35	-0,25	0,35	0,25	0,37	0	0,3	0	-0,23	0
Month -2	-0,24	0,34	-0,33	0,21	0,3	0,28	0	0	-0,34	0	0	0
Month -3	0	0	0	-0,21	0,28	0	-0,32	0	0	0	-0,25	0,27
Month -4	0,21	0	0	0	0,32	-0,21	0,3	-0,26	0,28	0	0	0,24
Month -5	-0,32	0,21	-0,21	0,24	0,29	0,21	-0,28	-0,24	0	0,25	0,22	0,24
Month -6	-0,24	0	-0,41	0	0,21	-0,29	0	0,28	-0,23	-0,26	-0,33	0
Month -7	0	-0,21	-0,36	0,26	0,31	0	0,24	0	-0,2	-0,29	0	0
Month -8	-0,25	-0,24	-0,43	0,24	-0,25	0	0	0	0	-0,4	0	0,29
Month -9	0	0,25	0,28	0	0	0,23	0	-0,23	-0,33	-0,3	0	0
Month -10	-0,33	0	-0,27	0,25	0,21	-0,23	0	0,26	-0,3	-0,32	0	0,27

NAO
ONI/ENSO
WeMOI
EA
EA/WR
SCAND

Month -11	0	0	0	0	0,28	0	0,34	0	-0,21	0	-0,2	0,24

Ionio 3

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,46	-0,39	0,47	0	0,29	0	0	0,29	0	0	0,32	-0,23
Month -1	-0,32	-0,22	-0,31	-0,21	0,32	0,22	0,33	0,29	0	0	0	0
Month -2	-0,28	-0,22	-0,37	0	0,28	0	0	0,22	0,31	0,3	0,25	-0,26
Month -3	0	0	0	0	0,24	0	0,22	0,25	0,22	0	-0,26	-0,24
Month -4	-0,2	0	0	0	0,25	0	0	-0,21	0,22	0	-0,21	0,24
Month -5	-0,31	0	-0,2	0	0,31	0	0	0,28	-0,3	0,21	-0,27	0
Month -6	-0,26	0	-0,24	0	0,28	0	0	0,26	0	-0,29	-0,29	0
Month -7	0	-0,24	-0,27	0	-0,24	-0,23	0	0	0	-0,31	0	0
Month -8	-0,28	0	-0,43	0	0	-0,27	0	-0,21	0	-0,4	0,21	0
Month -9	0	-0,23	0	0	0,2	0	0	0	-0,31	-0,28	-0,26	0
Month -10	-0,35	0	0	-0,23	0,21	-0,23	0	0	-0,27	-0,45	0	0
Month -11	-0,33	0,25	0	0	0,24	-0,21	-0,28	-0,21	0	-0,2	0	0

Tirreno 1

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,45	-0,6	-0,37	0,35	-0,44	0	-0,32	-0,32	0,38	0,28	-0,3	-0,35
Month -1	0,29	-0,26	-0,32	0	0,23	0,24	0,37	0,24	-0,27	0,24	-0,26	0
Month -2	0,22	-0,21	-0,36	0,27	0,26	0,24	-0,22	-0,3	0,41	0	-0,25	0,24
Month -3	-0,27	0	0	0	0,3	0,24	-0,35	-0,26	-0,35	0,25	-0,22	0,23
Month -4	0	0	0	-0,2	0,31	-0,31	0,3	0	-0,25	0,4	0	0
Month -5	-0,21	-0,26	-0,29	-0,26	0,31	0	-0,25	-0,22	-0,31	0,28	-0,22	0
Month -6	-0,22	-0,25	-0,34	-0,21	0,28	0	0	0,26	-0,29	-0,3	-0,42	0
Month -7	0	0,29	-0,24	-0,24	0,35	-0,22	0	0,25	-0,27	-0,35	0	0,26
Month -8	0,33	-0,24	-0,29	0	0,24	-0,28	0	0	0,31	0,22	-0,25	-0,22
Month -9	0,23	-0,29	-0,33	0	0,23	-0,36	0	0	-0,28	-0,38	0	-0,3
Month -10	0	0	0,24	0	0,23	-0,32	0	0	-0,26	-0,31	-0,26	0,3
Month -11	-0,28	0	0	-0,22	0,24	0,24	0,33	0	0	-0,31	-0,2	-0,23

Tirreno 2

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,37	-0,55	0,41	0,22	0,3	0	-0,21	-0,32	0,28	0	-0,3	-0,31
Month -1	0,22	-0,27	-0,32	-0,28	0,25	0,24	0,29	0	-0,24	0	0	0
Month -2	0,23	-0,27	0	0,38	0,25	0,24	0	-0,24	-0,28	0	0	-0,2
Month -3	0	0	0,27	0	0,27	-0,2	0,22	0,24	-0,36	-0,22	0	0,22
Month -4	0	0,23	0	-0,32	0,27	-0,29	0,23	-0,22	0	0	0	0
Month -5	-0,27	0,22	0	-0,26	0,31	0	0	0,21	0,33	0,22	-0,2	0
Month -6	0	-0,3	0,28	-0,22	0,25	-0,23	0,22	0,27	0	-0,29	-0,29	0
Month -7	0	-0,27	-0,22	0,21	0,35	-0,35	-0,24	0	0	-0,33	0	-0,21
Month -8	-0,31	-0,21	-0,38	0,32	0	-0,29	0	0	0	-0,23	0,23	0
Month -9	0	-0,3	0,25	-0,3	0,2	0,28	0	-0,22	0	-0,33	0,25	-0,22
Month -10	-0,33	0,22	0	0,27	0	-0,33	0	0,21	0	-0,41	0,25	0,22
Month -11	0	0,26	0	0	0	-0,31	0,36	0	0	-0,23	0,24	0

Table 5.3. Monthly teleconnections with the highest, significant correlation value with monthly precipitation. Calabria rain zones, 1981-2010.

IONIO 1	# strongest correlation	# magnitude above 0.4
NAO	12	0
MOI	16	2
ONI/ENSO	9	0
WeMOI	10	1
EA	16	2
EA/WR	20	0
SCAND	11	1
IONIO 2	# strongest correlation	# magnitude above 0.4
NAO	17	0
MOI	19	3
ONI/ENSO	4	0
WeMOI	16	1
EA	12	0
EA/WR	16	0
SCAND	6	2
IONIO 3	# strongest correlation	# magnitude above 0.4
NAO	13	0
MOI	17	3
ONI/ENSO	5	0
WeMOI	12	1
EA	10	0
EA/WR	13	0
SCAND	8	1
TIRRENO 1	# strongest correlation	# magnitude above 0.4
NAO	17	0
MOI	12	1
ONI/ENSO	17	0
WeMOI	19	1
EA	11	3
EA/WR	17	1
SCAND	12	0
TIRRENO 2	# strongest correlation	# magnitude above 0.4
NAO	17	0
MOI	16	2
ONI/ENSO	10	0
WeMOI	15	0
EA	12	0
EA/WR	13	0
SCAND	11	1

Table 5.4. Number of months (#) when a teleconnection (left) has the strongest correlation with monthly precipitation (center) and the number of months when this correlation is equal or above 0.4. Calabria rain zones, 1981-2010.

TELECONNECTION MONTHS	PRECIPITATION MONTHS											
Ionio 1	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,37	-0,27	0,27	0,23	-0,27	0	0	-0,32	-0,23	0,24	-0,3	0
Month -1	0	0	0	0,25	0	0,24	0	0	0	-0,2	0	0
Month -2	0	0	0	0,2	0,26	0	0	-0,21	0	0	0	0
Month -3	0	0	0	-0,24	0	0,25	0,2	-0,21	0,23	0	-0,21	0
Month -4	-0,24	0	0	0	0	0	0	0	0,24	0,2	0,21	0
Month -5	0	0	0	0	0	0	0	0	0	0,25	0	0
Month -6	0	0	0	0	0	0	0	0	0	0	-0,22	0
Month -7	0	0	0	0,2	0,2	0	0	0	0	0	0	0,24
Month -8	0,28	0,22	0	0	0	0	-0,21	0	0	-0,3	0	0
Month -9	0	0	-0,23	0	0	0	0	0	-0,25	0	0,23	0
Month -10	0	0	0,2	0	0	0	0	0,28	0	0	0	0
Month -11	0	0	0	-0,21	0,24	0	0,28	0	0	0	0	0,24

Ionio 2	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,39	0,31	0,38	-0,29	0,26	0,2	-0,23	-0,29	-0,24	0,3	-0,36	-0,24
Month -1	0	0	-0,22	0	0,25	0	0	0	0,27	0	0	0
Month -2	0	0	-0,21	0	0,25	0	0	0,21	0	0	0,27	0
Month -3	0	0	0	0	0	0	0	0	0	0	-0,21	-0,22
Month -4	-0,25	0	0	0	-0,21	0	0	0	0	0	-0,21	-0,25
Month -5	-0,24	0	0	0	0	0	0	-0,2	0	0,23	0,22	0
Month -6	0	0	0	0	0	0	0	0	0	0	-0,21	0
Month -7	0	0	-0,26	0,26	0,24	0	0	0	0	0	0	0
Month -8	0,21	0,25	-0,2	0	0	0	0	0	0	-0,24	0	0,22
Month -9	0	0,24	-0,2	0	0	0,22	0	0	-0,25	0	0	0
Month -10	-0,22	0	0	0	0	0	0	0,27	0	-0,2	0	0
Month -11	0	0	0	0	0,23	0	0	-0,23	0	0,26	0	0

NAO
MOI
ONI/ENSO
WeMOI
EA
EA/WR
SCAND

Ionio 3	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,33	0,27	0,36	0	0,24	0	0	0,24	0	0,37	-0,35	0
Month -1	0	0	-0,21	0,22	0,26	0	0	0,22	0	0,24	0	0
Month -2	0	0	-0,24	0	0,22	0	0	0,23	0	-0,27	0	0
Month -3	0	0	0	0	0	0	0	0	0	0,21	0	0
Month -4	0	0	0	0	0	-0,2	0	0	0	0,22	-0,23	0
Month -5	0	0	0	0	0	0	0	0	0	0,25	0	0
Month -6	0	0	0	0	0	0	0	0	0	-0,25	0	0
Month -7	0	0	0	0,21	0	0	0	0	0	-0,21	0	0
Month -8	0	0	0	0	0	0	0	0	0	-0,26	0	0
Month -9	0	0	0	0,2	0	0	0	0	-0,23	-0,21	0	0
Month -10	0	0	0	0	0	0	0	0	0	-0,28	-0,22	0
Month -11	0	0	0	0	0,22	0	0	0	0	-0,24	0	0,21

Tirreno 1	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	-0,35	-0,46	-0,37	0,21	-0,36	0	0	-0,28	-0,23	0,31	-0,26	-0,43
Month -1	0,22	-0,23	0	0	0	0	0	0,25	0,25	0	0,23	0
Month -2	0	0	0	0	0	-0,21	0	0	0	0	0	0
Month -3	0	0	0	0	0	0	0	0	-0,22	0,23	0	0
Month -4	0	0	0	0	0	-0,2	0	0	-0,22	0,26	0	0
Month -5	0	0	-0,2	0	0	0	0	0,22	0	0	0	0
Month -6	0	0	0	0	0	0	0	-0,22	0	0	-0,21	0
Month -7	0	0	0	0	0,21	0	0	0	-0,21	-0,21	0	0
Month -8	0,24	0	0,26	0	0	-0,2	0	0	0	0	0	0,23
Month -9	0	0	0	0,2	0	-0,24	0	0	0	-0,26	0	0
Month -10	0	0	0	0	0	0	0	0,3	0	0	0	0
Month -11	-0,22	0	0,23	0	0	0	0,23	0,25	0	0,28	0	0

Tirreno 2	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Current month	**-0,27**	**-0,36**	**-0,37**	**-0,2**	**0,23**	**0**	**0**	**-0,24**	**0**	**0,21**	**-0,33**	**-0,32**
Month -1	**0**	**0**	**0**	**0**	**0,2**	**0**	**0**	**0,25**	**0**	**0**	**0**	**0**
Month -2	**0**	**0**	**0**	**0**	**0,21**	**0**	**0,21**	**0,24**	**0**	**0**	**0**	**0**
Month -3	**0**	**0**	**0,21**	**0**	**0**	**0**	**0**	**0**	**-0,23**	**0**	**0**	**0**
Month -4	**-0,21**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Month -5	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Month -6	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**-0,23**	**0**
Month -7	**0**	**-0,22**	**0**	**0,22**	**0,22**	**-0,21**	**0**	**0**	**-0,2**	**0**	**0**	**0**
Month -8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**-0,2**	**0**	**0**	**0**	**0**
Month -9	**0**	**0**	**0**	**-0,2**	**0**	**0,27**	**0**	**0**	**0**	**0**	**0**	**0**
Month -10	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0,32**	**0**	**-0,22**	**0**	**0**
Month -11	**0**	**-0,21**	**0**	**0**	**0**	**0**	**0,22**	**0,22**	**0**	**-0,24**	**0**	**0**

Table 5.5. Monthly teleconnections with the highest, significant correlation value with monthly precipitation. Calabria rain zones, 1951-2010.

IONIO 1	**1951-1980**	**1981-2010**
NAO	9 (0)	12 (0)
MOI	10 (1)	16 (2)
ONI/ENSO	22 (0)	9 (0)
WeMOI	17 (0)	10 (1)
EA	14 (0)	16 (2)
EA/WR	14 (1)	20 (0)
SCAND	19 (2)	11 (1)
IONIO 2		
NAO	9 (0)	17 (0)
MOI	8 (0)	19 (3)
ONI/ENSO	6 (0)	4 (0)
WeMOI	20 (0)	16 (1)
EA	14 (1)	12 (0)
EA/WR	19 (2)	16 (0)
SCAND	19 (1)	6 (2)
IONIO 3		
NAO	13 (2)	13 (0)
MOI	8 (0)	17 (3)
ONI/ENSO	12 (0)	5 (0)
WeMOI	9 (0)	12 (1)
EA	11 (0)	10 (0)
EA/WR	17 (3)	13 (0)
SCAND	18 (0)	8 (1)
TIRRENO 1		
NAO	13 (0)	17 (0)
MOI	15 (1)	12 (1)
ONI/ENSO	7 (0)	17 (0)
WeMOI	15 (0)	19 (1)
EA	16 (0)	11 (3)
EA/WR	14 (1)	17 (1)
SCAND	20 (0)	12 (0)
TIRRENO 2		
NAO	12 (1)	17 (0)
MOI	9 (2)	16 (2)
ONI/ENSO	16 (0)	10 (0)
WeMOI	17 (0)	15 (0)
EA	16 (1)	12 (0)
EA/WR	17 (1)	13 (0)
SCAND	21 (1)	11 (1)

Table 5.6. Number of months (#) when a teleconnection (left) has the strongest

correlation with monthly precipitation (and, in brackets, the number of months when this correlation is equal or above 0.4). Calabria rain zones; 1951-1980 (center), 1981-2010 (right).

Tirreno 2

INFLUENCE	on DJF	on MAM	on JJA	on SON
-1	0	0	0,21	-0,29
-2	-0,23	0	-0,26	-0,35
	0,2			
-3	0	0,29	-0,24	-0,29
				-0,28
-4	0	0	0	0

NAO
MOI
ONI/ENSO
WeMOI
EA
EA/WR
SCAND

Table 5.7. Seasonal teleconnection indices which correlate significantly with seasonal accumulated precipitation, with correlation values of 0.4 or above. Tirreno 2 region, 1981-2010.

6. Linking variability and trends in thermal growing season onset to the NAO and EA patterns

Philip Craig, Richard P. Allan

6.1 Introduction

Agriculture is strongly affected by weather and climate with up to 80% of variability in crop production linked to weather conditions (Hoogenboom, 2000). Climate change is also likely to affect European agriculture with the increasing likelihood of droughts in southern Europe during summer and longer growing seasons in northern Europe due to warmer winters (Bindi & Olesen, 2011).

INDECIS has various temperature- and precipitation-based variables related to the agriculture sector (Dominguez-Castro *et al*, 2020). For example, growing season onset (ogs6, ogs10), growing season length (gsl), growing season rainfall (gsr) and mean growing season temperature (t_ao, t_ms) are all included in the dataset.

Here, we focus on the growing season onset as this is the first gridded dataset of growing season onsets for all of Europe. Previous studies have generally focused on station data over specific regions within Europe and their relationships with teleconnections, e.g. Finland (Irannezhad & Klove, 2015) and Poland (Tomcyzk *et al*, 2019). Cornes et al. (2018) used the ECA&D homogenized observations (Klok and Klein Tank, 2009) which had sparse coverage in some regions and has since been improved upon as part of INDECIS (Coll et al., 2019).

INDECIS provides two definitions of thermal growing season onset (Klein Tank *et al.*, 2009): one using the first six day span with each daily mean temperature above 5°C (ogs6) and another using the first ten day span with each daily mean temperature above 10°C (ogs10). We have used ogs10 because synoptic scale temperature events which influence growing season onset typically last around 10 days (Cornes *et al*., 2018). This avoids any unusually early growing season onsets followed by late frosts to which the early growing season will become increasingly susceptible to as the onset gets earlier with climate change (Liu et al., 2018; Ceglar *et al.*, 2019).

By using Principal Component Analysis, Cornes *et al.* (2018) found that 34.7% of interannual variability in ogs6 can be explained by the NAO and EA teleconnections, and also that they explain 57% of interannual variability when a time series decomposition method is used to define growing season onset. Irranezhad & Klove (2015) found that the EAWR pattern is the most significantly correlated with variations in the Finnish growing season onset. Tomcyzk found some influence from the NAO in North-West Poland and the EA in the south of Poland.

From INDECIS, ogs10 is earliest in southern and western Europe, increasing towards the north and east with mountainous regions also showing late onset (Figure 6.1a). Some of these regions are not relevant to agriculture anyway. The most variable regions are Ireland, South England, North France through Belgium & Holland into North Germany with very low variability in Portugal and South Spain where the growing season is likely to last all year (Figure 6.1b).

There is a general increasing trend over Europe which is strongest around the English Channel (Figure 6.1c). Most of Eastern Europe, which has late onset and low variability, has weak but statistically significant increasing trend.

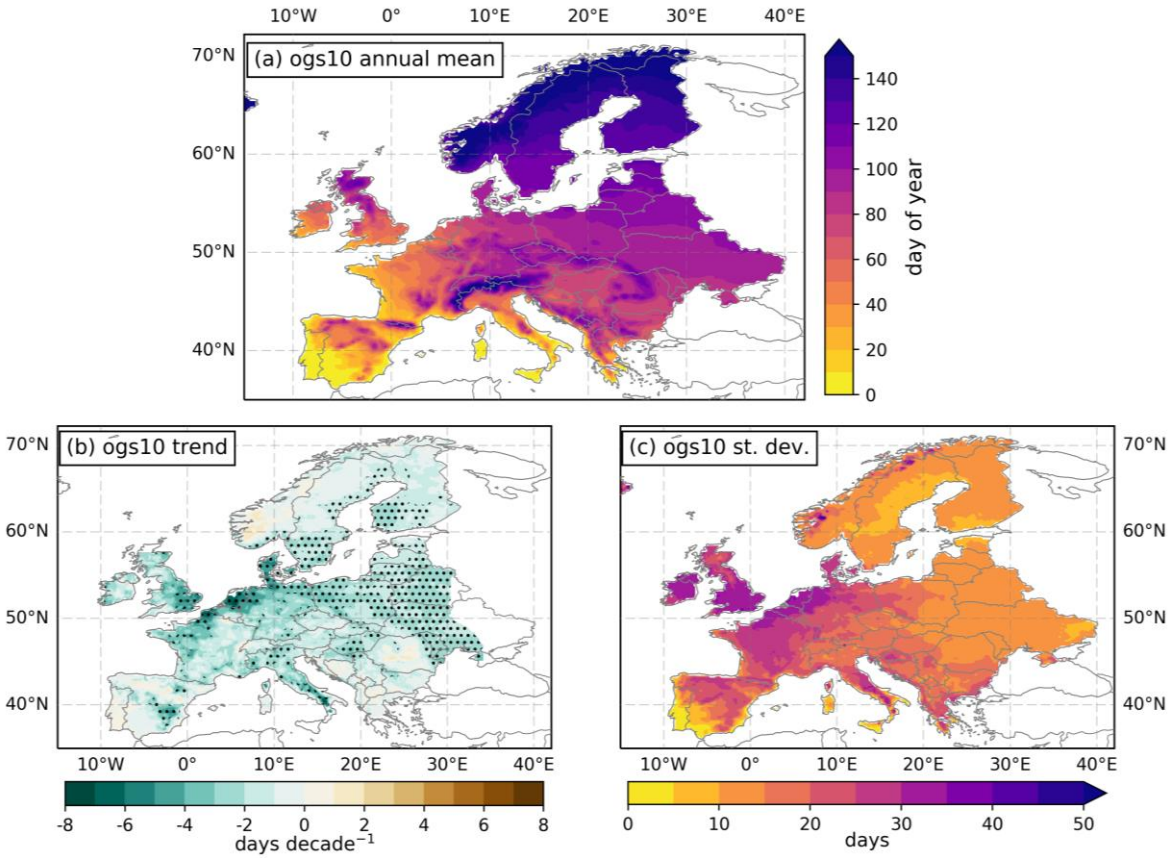

Figure 6.1: (a) the 1950-2016 annual mean ogs10, (b) the trend in ogs10 over 1950-2016 where stippling indicates statistical significance at the 95% level using a Wald Test, and (c) the standard deviation of ogs10 across 1950-2016.

6.2 Methods

We have investigated the relationship between growing season onset and four teleconnection patterns relevant to European weather and climate (NAO, SCA, EA, EAWR) by removing each teleconnection signal from the ogs10 time series at each grid point with the method described by Bhend & von Storch (2008) and Iles & Hegerl (2017). The signal is removed by regressing the detrended ogs10 time series on the detrended teleconnection index with least squares linear regression. The resulting slope at each grid point is the teleconnection signal in ogs10. The teleconnection component of ogs10 interannual variability is calculated by multiplying each year's ogs10 by the signal and the residual variability is simply the difference between the

ogs10 time series and the teleconnection component of interannual variability. The residual variability represents the variability when the circulation changes associated with each teleconnection pattern are removed.

We have focused on the NAO and EA signals for ogs10 (Figure 6.2) since these two teleconnections have the strongest relationships with the interannual variability of ogs10 and have more statistical significance than EAWR and SCA (not shown). We also focus on the three months at the beginning of the year (JFM; January-February-March) as the annual mean ogs10 for the main agricultural regions of Europe is generally within the first 90 days of the year (Figure 6.1(a)) and the teleconnection signals after March are generally weak with little statistical significance.

6.3 Interannual variability

The JFM NAO and EA both generally have a negative signal in ogs10 across Europe (Figure 6.2) i.e. the positive NAO and EA phases result in an earlier growing season onset. The EA signal is strong across much of continental Europe with statistically significant negative signal (earlier onset) over France, Germany, Italy and South-East Europe (Figure 6.2d). The NAO signal is strongest around the English Channel with a region of statistical significance in South England and North France (Figure 6.2a). On continental Europe, the NAO signal is much weaker than the EA signal although there is a region of statistically significant NAO signal over Hungary. The area along the southern coast of the English Channel occupied by statistically significant NAO signal has a much weaker EA signal which is not statistically significant.

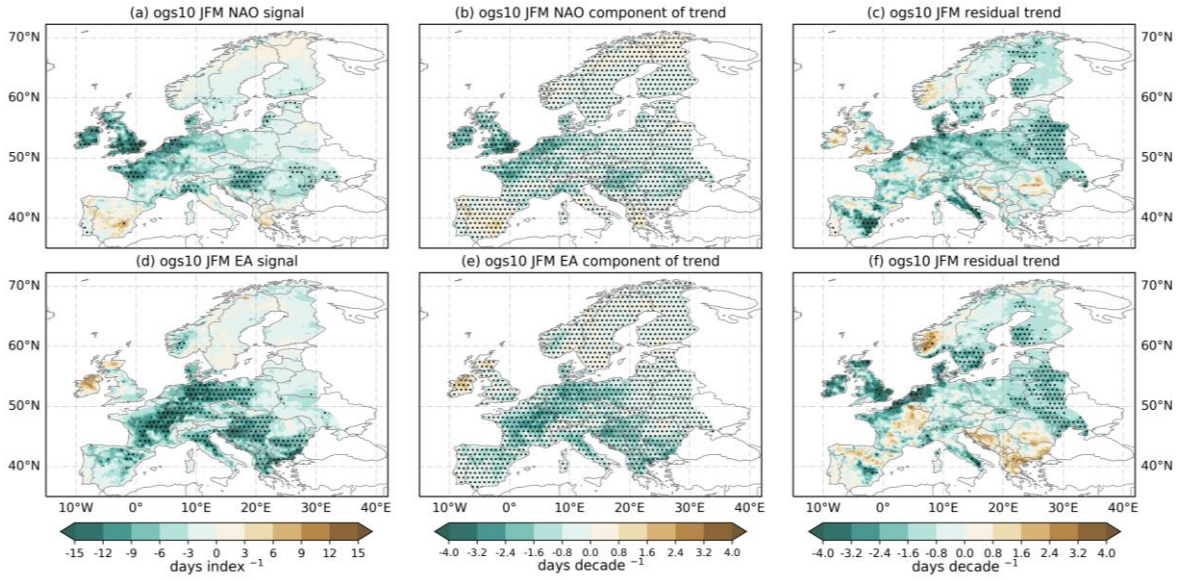

Figure 6.2: relationships between the JFM (January-February-March) NAO and EA teleconnection indices and the ogs10. The top row corresponds to the NAO and the bottom row corresponds to the EA: (a) and (d) show the teleconnection signals, (b) and (e) show the JFM teleconnection patterns contributions to the overall trend, (c) and (f) show the residual trend when the teleconnection contribution is removed. Green contours indicate a reduction in ogs10 (earlier onset) and brown contours indicate an increase in ogs10 (later onset). Stippling indicates statistical significance at the 95% level using a Wald Test.

The positive phases of both NAO and EA in JFM are associated with earlier ogs10 and the negative phases are associated with later ogs10 (Figure 6.3). The influence of the NAO on ogs10 decreases rapidly with increasing distance from the Atlantic coast (Scheifinger et al., 2002). The regions most influenced by NAO+ are confined to Western Europe (Ireland, England, France, Belgium, Holland) with onset about 20 days earlier for years with NAO+ greater than one standard deviation above the mean (Figure 6.3a). A narrow corridor along the axis of the English Channel from Cornwall to Denmark is most strongly influenced by NAO- (Figure 6.3b). A broader region stretching across continental Europe is strongly affected by EA+ with earlier ogs10, but EA- is associated with later ogs10 in France and Spain with a much weaker influence into Central Europe (Figure 6.3c,d).

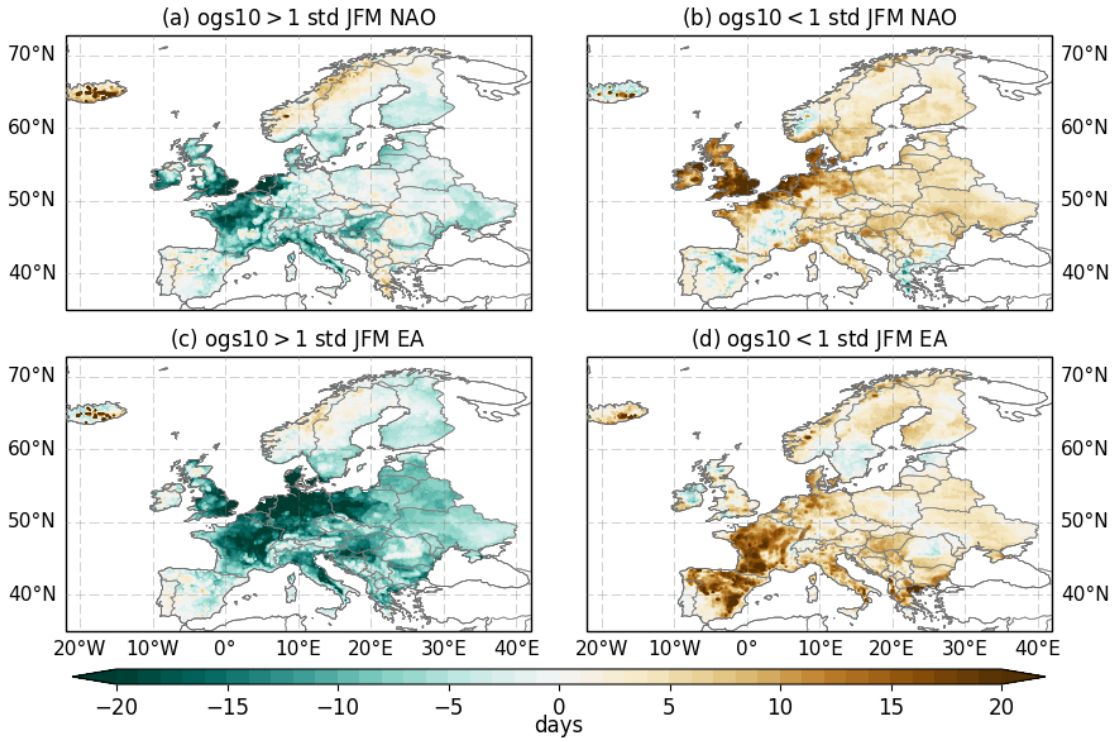

Figure 6.3: composites of anomalies from ogs10 annual mean for (a) all years where JFM NAO index is greater than one standard deviation above the 1950-2016 mean JFM NAO index; (b) all years where JFM NAO index is less than one standard deviation below the 1950-2016 mean JFM NAO index; (c) same as panel (a) but for the EA index; and (d) same as panel (b) for the EA index. Green contours indicate a reduction in ogs10 (earlier onset) and brown contours indicate an increase in ogs10 (later onset).

To investigate the role of the NAO and EA in ogs10 interannual variability, six regions with statistically significant signals for one or both teleconnections were selected (Figure 6.4(a)): Great Britain and Ireland (GBI), France, Italy, Poland and North Germany (PNG), Hungary and the Balkans. Hungary has been kept separate from the Balkans as it has strong and statistically significant signals in both teleconnections (Figure 6.2). The JFM NAO and EA signals were averaged over these regions and compared to the area-averaged anomalies from the ogs10 climatology (Figure 6.4b-f). Various years with particularly late or early growing season onsets can therefore be linked to either the NAO or EA component of inerannual variability.

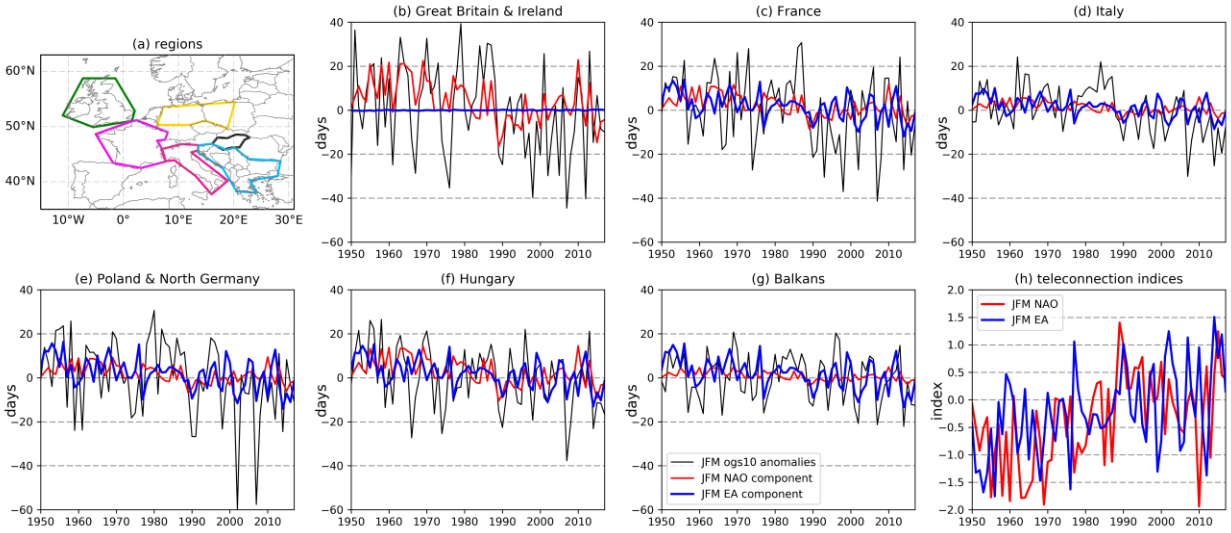

Figure 6.4: (a) regions over which the JFM NAO and EA ogs10 signals from Figure 6.2 are averaged; (b)-(g) yearly ogs10 anomalies (black lines) with the yearly NAO (red lines) and EA (blue lines) components; and (h) the NAO and EA indices averaged across JFM. In panels (b)-(g) negative values represent an earlier growing season onset and positive values represent a later growing season onset.

Interannual variability in area-averaged GBI ogs10 can be partially linked to the JFM NAO but not the EA. The NAO component closely follows the anomalies around 1960 and 2010 with the positive (late onset) peaks sometimes entirely explained, or exceeded, by the NAO component (Figure 6.4b). In contrast, the EA component is almost completely flat reflecting the weak EA signal for this region of varying sign (Figure 6.2d). Furthermore, the NAO component has an explained variance of 23% compared to 1% for the EA component (Table 6.1).

Across France, the NAO and EA components each have a more balanced contribution with 13% and 17% explained variance respectively (Table 6.1), although they each influence different parts of France (Figure 6.2a,d). The interannual variability in the 1950s looks to be mostly explained by the EA component but the largest ogs10 anomalies, particularly for early onset, do not appear to be explained by either teleconnection (Figure 6.4c).

For Italy, the EA component often follows the interannual variability of the ogs10 anomalies (Figure 6.4(d)) and it has explained variance of 28% compared to 16% from the NAO component which does not appear to match much of the ogs10 variability. This is to be expected since the region of strong, statistically significant NAO signal only is small area of northwest Italy, whereas the EA signal is strong over a larger area (Figure 6.2a,d). However, the EA component does not explain the post-1990 early onset (negative) peaks and neither component gives any representation of the consistent late onset for a few years around 1980.

The interannual variability for the PNG region is more closely linked to the EA component than the NAO component. Some of the variability in the EA component corresponds to the variability in the ogs10 anomalies but the major peaks in the anomalies are not explained by either teleconnection component (Figure 6.4e). Specifically, two early onset peaks above 20 days

either side of 1960, three late onset spells around 1980 and two huge early onset peaks above 50 days after 2000.

Hungary has the best overall statistical relationships with the two teleconnection indices and a total explained variance of 49% (Table 6.1), reflecting the strong signals (Figure 6.2a,d). The interannual variability of the EA component corresponds well with the ogs10 anomalies and matches some of the lower peaks exactly (Figure 6.4f). The NAO component has a slightly lower explained variance but it does correspond with some of the variability in ogs10 anomalies where the EA component does not.

The EA component for the Balkans has the highest explained variance for any region (30%, Table 6.1) and the interannual variability of the EA component corresponds well with the ogs10 anomalies throughout the INDECIS time period. This is particularly notable after 1990 (Figure 6.4g). The strong statistical relationship between the EA component and the ogs10 anomalies for the Balkans reflects the strong JFM signal (Figure 6.2d) across much of the region, particularly Northern Croatia/Serbia and the Greek Aegean coast. Romania appears to be an outlier in the Balkans as it has a weak signal for both teleconnections and has areas of statistical significance on two sides (Figure 6.2a,d).

Table 6.1: correlation coefficients (R) between ogs10 interannual variability and its NAO and EA components averaged over each region highlighted in Figure 6.3(a). Numbers in italics indicate R which are not statistically significant at the 95% level.

	Great Britain & Ireland		France		Italy		Poland & North Germany		Hungary		Balkans	
	NAO	EA	NAO	EA	NAO	EA	NAO	EA	NAO	EA	NAO	EA
R	0.48	*-0.12*	0.36	0.41	0.4	0.53	0.33	0.47	0.48	0.51	0.21	0.55
R^2	0.23	*0.01*	0.13	0.17	0.16	0.28	0.11	0.22	0.23	0.26	0.05	0.3

6.4 Trends

The trend in the JFM NAO and EA components of ogs10 interannual variability are visibly similar to their signals and statistically significant across Europe (Figure 6.2b,e) since the JFM means for each teleconnection index are increasing (Figure 6.3h). The residual trends for both teleconnections are weak in regions where there is strong signal in the interannual variability but statistically significant in regions with a weak signal, specifically Ukraine, Belarus, South Sweden and South Finland (Figure 6.3c,f). Part of the region around the English Channel with strong NAO signal also has the strongest residual trend when the JFM EA is removed (Figure 6.3f). The contributions from the trends in the NAO and EA components are more balanced for PNG

The strongest overall area-averaged trends are across the GBI (-2.8 days decade^{-1}) and PNG (-3.0 days decade^{-1}) regions (Table 6.2).The GBI trend towards earlier onset is dominated by the trend in the NAO component of ogs10 (-2.3 days decade^{-1}) which is consistent with the trend in the JFM NAO index since 1950 (Figure 6.4h). Only the Balkans has a residual trend which is statistically significant. The changes in ogs10 for the Balkans not associated with the NAO or EA cause a later onset – this is particularly notable when the EA is removed (Figure 6.2f). The result is that the overall ogs10 trend for the Balkans is a weak trend towards earlier onset (-0.8 days decade^{-1}) which is not statistically significant.

Table 6.2: area-averaged trends for all six regions shown in Figure 6.4(a). The actual trend is the ogs10 trend from Figure 6.1(a), the NAO and EA columns represent the trends in the teleconnection component of ogs10 (Figure 6.2 (b,e), and the residual (Figure 6.2(c,f)) is the remaining trend when changes associated with the NAO and EA are removed. Negative trends indicate an earlier growing season onset and positive trends indicate a later growing season onset. Units are days decade^{-1} (numbers in italics are not statistically significant at 95% level).

	actual	NAO	EA	residual
Great Britain & Ireland	-2.8	-2.3	0.0	*-0.5*
France	-2.3	-1.2	-1.5	*0.4*
Italy	-2.4	-0.6	-1.1	*-0.8*
Poland & North Germany	-3.0	-1.0	-1.7	*-0.3*
Hungary	-2.2	-1.5	-1.6	*0.9*
Balkans	*-0.8*	-0.5	-1.6	1.3

6.4 Summary

Six regions have been identified where interannual variability of growing season onset is sensitive to one or both of the JFM NAO and EA. The strongest NAO signal is found around the English Channel and the EA signal is strong across much of continental Europe. The positive phases of each teleconnection are associated with earlier growing season onsets with the influence of the NAO mostly confined to Western Europe.

The strongest statistical relationship between the NAO and interannual variability of ogs10 is found across GBI where the EA has a weak influence. The EA pattern has much strongest statistical relationships with ogs10 variability on the continent, particularly for Italy, Hungary and the Balkans. Five of the six regions have trends towards an earlier growing season onset greater than 2 days decade^{-1} with only the Balkans showing a weak and non-statistically significant trend. The trend in the JFM NAO component dominates the GBI trend and the EA trends are more important for the five continental regions.

7. Prediction of sectoral indices from atmospheric indices through the implementation of statistical models

Manuel Del Jesus, Salvador Navas, Diego Urrea

7.1 INTRODUCTION

Currently, there are multiple methodologies and tools used to make predictions of data series for different time and spatial horizons, which are based on statistical models. These models are mainly based stablishing statistical links among variables, which have the essential function of predicting and accurately representing the phenomena that could happen in future scenarios.

Data series predictions have great relevance and applicability, because using them, events related to agriculture, climate, tourism and disaster risk, among others, can be predicted. An example is the high global demand for future climate data series for the development of studies that measure the impact of climate change.

It is here where, through this study, we want to contribute and give an answer to this high demand, carrying out the construction, validation and the execution of four statistical prediction models, *Random Forest (RF), K-nearest Neighbours Regression (KNN), Linear Regression (LR) and Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX)*, which are considered relevant about the wide range of models that exist.

For the construction of these models, the goal is to use global atmospheric indices as predictors, such as ENSO (El Niño-Southern Oscillation) and NAO (North Atlantic Oscillation), and as predictand sectoral indices related to agriculture, climate or tourism, among others. Another characteristic of this type of model is that they are based on the historical data from both the sectoral data series and the atmospheric data series for a given climatic period.

This study has been developed covering three fundamental approaches: the processing of information, the implementation and validation of the statistical model and the analysis of the results. This report contains, in addition to this introductory chapter, a chapter in which the general and particular objectives of the study are presented (chapter 2), a chapter where the description of the study area is presented (chapter 3), a chapter where reference is made to the description of the initial data and the sources of information used (chapter 4), a chapter that describes the implemented methodology to carry out the analysis and obtain the results of the study (chapter 5) and a chapter where the results obtained when applying the models are presented and analyzed (chapter 6). Finally, Chapters 7 and 8 present the general conclusions of the study and the bibliographic references used for its development.

7.2 OBJECTIVES AND CONTENTS OF THE STUDY

For the preparation of this study, a general objective has been determined, which in turn is complemented by specific objectives to specifically encompass and define the scope of the study. The objectives described are set out below.

General objective

Construction of statistical predictive models using series of the main global atmospheric indices as predictors, to predict other types of indices called sectoral, that is, starting from indices such as ENSO and NAO, the goal is to evaluate their ability to predict related sectoral indices with agricultural, climatic or tourist aspects, among others, from the construction of regression models and time series. Likewise, once the models are built, the goal is to evaluate the validity of the results obtained through the use of different statistical correlation and error metrics.

Particular objectives

- Collection and analysis of information related to atmospheric and sectoral indices available in the study area.
- Implementation and construction of the statistical prediction models *Random Forest (RF), K-nearest Neighbours Regression (KNN), Linear Regression (LR) and Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX)*, for time lags between 1 and 3 months.
- Evaluation of the predictive capacity of the models by means of the validation and evaluation of the determination coefficient (R^2), the Pearson correlation coefficient (R) and the mean squared error (MSE), for the different modelled scenarios.
- Based on the predictive evaluation of the models using the above coefficients, the goal is to evaluate which of them provides better results.

7.3 SOURCES OF INFORMATION AND INITIAL DATA

The sources of information refer to the entities responsible for providing the data for atmospheric indices and sectoral indices. In the following lines, the information used in the study is presented.

Information sources related to atmospheric indices

To carry out the development of the study, two entities dedicated to the collection, treatment and generation of atmospheric indices have been consulted. *The National Weather Service Climate Prediction Center* (*NWS*) was consulted in the first instance and *The National Center for Atmospheric Research* (*NCAR*), both entities of the United States, in the second instance.

The *NWS* compiles data related to historical and current atmospheric and ocean conditions. Additionally, it records information on the behaviour and weather patterns such as the decadal oscillations of the North Atlantic and the Pacific, also containing data on ozone concentrations and temperature, at different heights of the atmosphere (Climate Prediction Center, 2020). Within the teleconnection patterns of the northern hemisphere are those described in Table 7.3, which record data related to different global climate behaviours. The information used was downloaded from the NWS website and corresponds to the monthly record of teleconnection patterns collected from January 1950 to December 2017.

NORTHERN HEMISPHERE TELECONNECTION PATTERNS	
Acronym	Teleconnection Index Name
NAO	North Atlantic Oscillation
EA	East Atlantic Pattern
WP	West Pacific Pattern
EP/NP	East Pacific/North Pacific Pattern
PNA	Pacific/North American Pattern
EA/WR	East Atlantic/West Russia Pattern
SCA	Scandinavia Pattern
TNH	Tropical/Northern Hemisphere Pattern
POL	Polar/Eurasia Pattern
PT	Pacific Transition Pattern

Table 7.3 Atmospheric indices consulted at The National Weather Service Climate Prediction Center (NWS)

On the other hand, the NCAR databases of atmospheric indices have been consulted, which are used to monitor climatic anomalies that occur in the tropical areas of the Pacific and are calculated with a base period of 30 years (NCAR, 2020). Within these indices, the Oceanic Niño Index (ONI) and the Niño Index 3.4 can be found, which are used to define the occurrence of natural events such as The Niño and Niña. For the reasons stated, and like the previous

indices, they have been selected as the fundamental basis of this study. The registration period used in the study corresponds to the monthly data recorded, which dates from January 1950 to December 2017. The data was downloaded from the website[1] and is presented in Table 7.4.

NINO STT INDICES	
NIÑO 3.4	El Niño 3.4
ONI	Oceanic Niño Index

Table 7.4. Atmospheric indices consulted at The National Center for Atmospheric Research *(NCAR)*.

Also, these indices can be obtained through the tool climate4R.indices [2] developed by the meteorology group at the University of Cantabria (Iturbide et al., 2019).

Information sources related to sector indices

The sectoral indices have been developed under the INDECIS project (JPI Climate, 2017) and include 129 climate indicators geared towards various sectors such as agriculture, disaster risk reduction, energy, health, water and tourism (Clime Indices of INDECIS and their links to sectorial data, 2020). The description of each of the indices, as well as the database, is available on the website[3]and they have a wide temporal and spatial resolution, covering the vast majority of the European continent.

Within the indices, multiple subgroups stand out, mainly related to precipitation, temperature, wind, solar radiation, snow, continental aridity, drought, fires and tourism. In total, all indicators that are presented by subgroups in Table 7.5 have been used. Likewise, the information consulted has a monthly registration period and has been compiled from January 1950, until December of the year 2017.

[1] https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni
[2] https://github.com/SantanderMetGroup/climate4R.indices
[3] http://www.indecis.eu/

SECTORAL INDEX GROUPS	SECTORIAL INDEX
Precipitation	*cdd, cwd, d50mm, d95p, dd, dr1mm, dr3mm, ep, gsr, mfi, ngsr, pci, r10mm, r20mm, r95tot, r99tot, rti, rx1day, rx5d, sdii.*
Temperature	*cfd, csd, csdi, d32, dd17, dtr, etr, fd, gd4, gsl, gtg, gtn, gtx, hd17, id, ntg, txn, ogs10, ogs6, ptg, stn10, stn15, stx32, su, ta_o, tm_s, tn10p, tn90p, tr, tx10p, tx90p, vcd, vdtr, vwd, wki, ws, wsdi, xtg, txx, zcd.*
Wind	*dfx21, fg, fg6bft, fgcalm, fxx.*
cloud_radiation	*aci, cc, fod, snd, ssd, ssp.*
Snow	*asd, fpsc, fsc, fsd, hsd, lpsc, ms, msd, scd, sd0_10, sd10_20, ss.*
Bioclimatic	*at, bio10, bio11, bio13, bio14, bio15, bio16, bio17, bio18, bio19, bio20, bio4, bio5, bio6, bio7, bio8, bio9, hi, mi, wci.*
Aridity continentality	*bi, cmd, eai, eto, jci, koi, mai, moi, pici, uai*
Drought	*spei1, spei12, spei3, spei6, spi1, spi12, spi3, spi6*
Fire	*fwi, kbdi, ffdi, mni, fffi*
Tourism	*utci*

Table 7.5. Groups and subgroups derived from sector indices (Clime Indices of INDECIS and their links to sectorial data, 2020)

Within the scope of the project, each of the indices was calculated in each of the cells of the mesh, with a resolution of 0.25º, which covers the entire European continent. These databases have a temporal resolution of 1 month and in some indices of 1 year. Figure 7.1 shows an example of the database used.

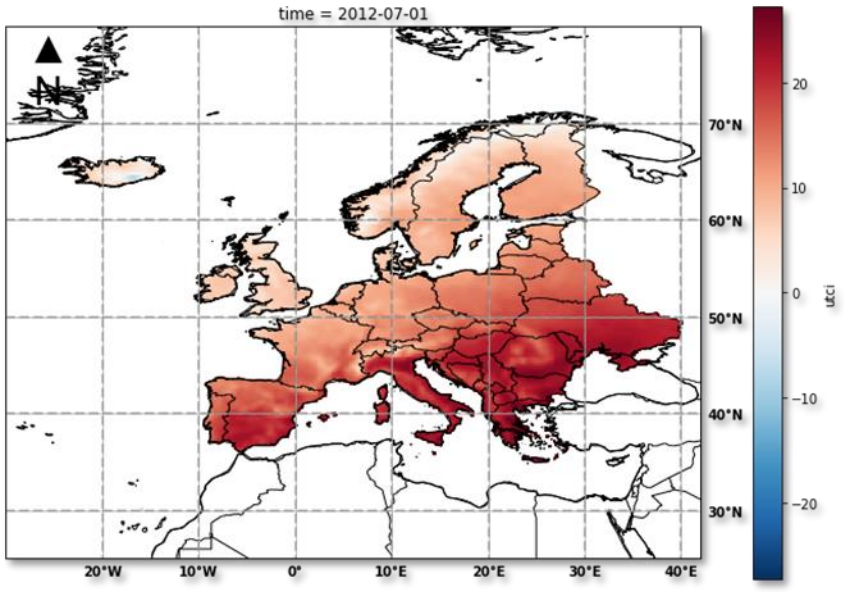

Figure 7.1. Example of data mesh for a sector index and a given month

7.4 DESCRIPTION OF THE STUDY AREA

The European continent has been selected as the study area for two fundamental reasons: firstly, due to the climatic diversity that can be found in the different regions that comprise it, which cover a wide range of climatic types; and secondly, due to the wide spatial and temporal resolution available for each of the atmospheric and sector indices, covering a registration period of approximately 67 years between 1950 and 2017. Figure 7.1 shows the spatial coverage of the study domain.

In order to have a greater coverage of the analysis and taking into account the available computational capacity and the modelling time required by each statistical model, a total of 378 homogeneously spatially distributed cities in Europe have been selected to carry out the Random Forest analysis. (RF), K-nearest Neighbours Regression (KNN) and Linear Regression (LR), to have representative results in modelling. The selected European cities are represented in red points in Figure 7.2. At these points, the corresponding sector index series will be extracted with the nearest cell of the data mesh elaborated on the project.

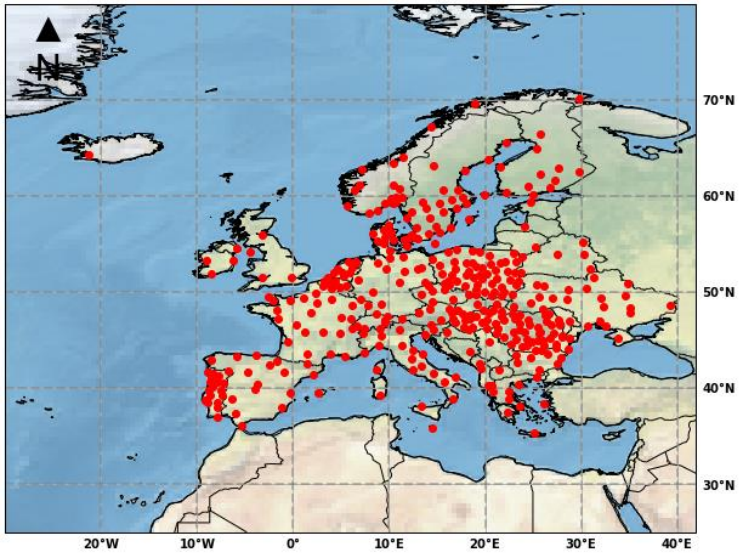

Figure 7.2 Cities of Europe selected to carry out the execution of the Random Forest, KNearest Neighbours Regression and Linear regression model.

Similarly, and taking into account the computational capacity demanded by the SARIMAX model, which is quite high compared to the other three models, a total of three cities in Europe have been selected, distributed in three strategic areas, to cover a more representative area of the continent. The selected European cities -Madrid (Spain), Vienna (Austria) and Vadso (Finland)- are represented by the red dots in Figure 7.3.

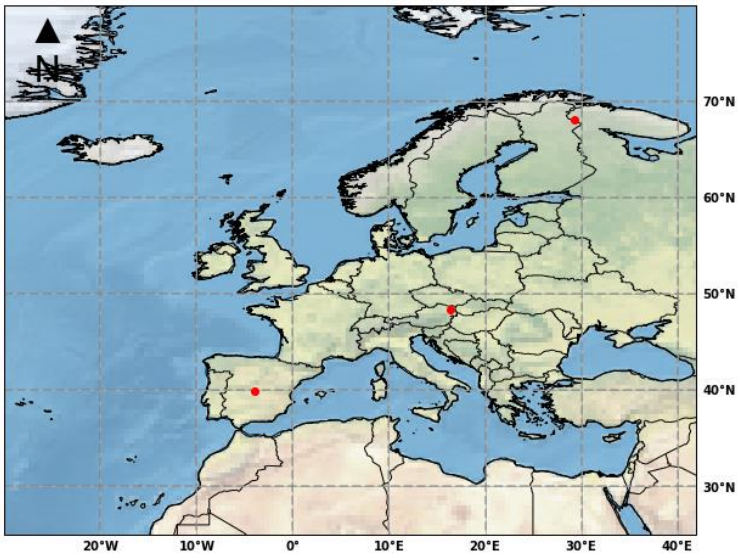

Figure 7.3. Cities of Europe selected to carry out the execution of the SARIMAX regression model.

7.5 METHODOLOGY

This section describes the methodological process used to evaluate the predictive capacity of atmospheric indices on sectoral indices, based on the application of regression models and time series: Random Forest (RF), K-nearest Neighbours Regression (KNN), Linear Regression (LR) and Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX).

Additionally, a theoretical description of the different implemented models is made, as well as an explanation of the valuation and qualification methods of these, among which are the Pearson's correlation coefficient, the determination coefficient R^2 and the mean square error.

Description of the methodological process

The description of the methodological process refers to the presentation of a diagram that synthetically explains the process used to obtain the results, which is presented in Figure 7.4.

The process of generating prediction models begins with the selection of the predictor variables and those that are to be predicted. In this case, as previously stated, the predictor variables will be the atmospheric indices and the variables to predict the sectoral indices elaborated on the INDECIS project (Clime Indices of INDECIS and their links to sectorial data, 2020). For each of the selected points (Figure 7.2-7.3) the time series of the closest cell is extracted from the corresponding index database. The atmospheric indices the index does not vary spatially, therefore, it will be the same time series for each of the points. At each of the points, the prediction model is constructed.

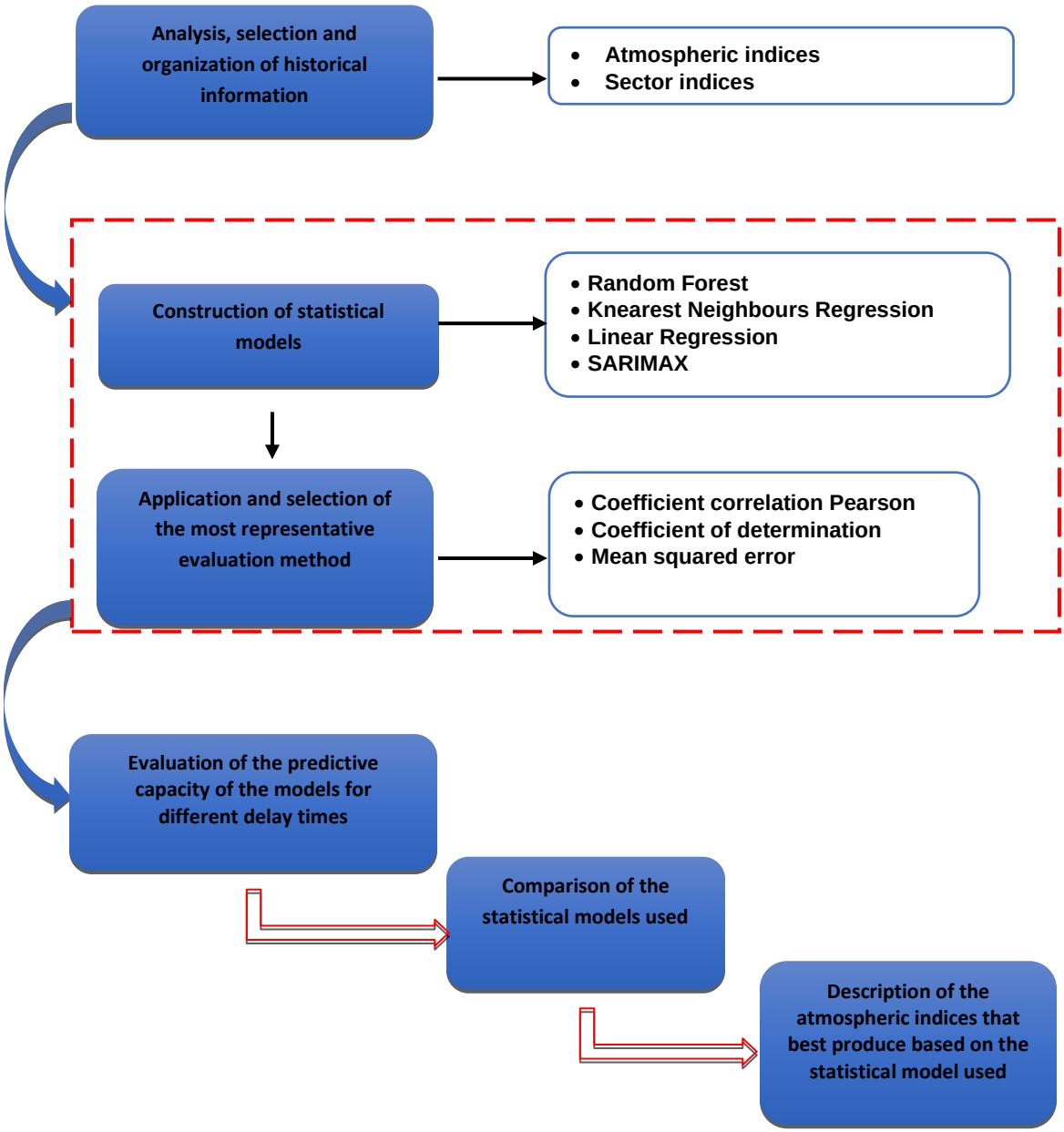

Figure 7.4. Methodological description of the study.

To build the model, not all the data available from the series is used, but a certain amount of data (80%) is used for training the model, and the rest of the data (20%) are used for validation. From the results obtained in the validation, the different coefficients that provide the predictive capacity of the model in each one of the points will be calculated and the final result will be the average of the values obtained in the set of points. Once the metrics are evaluated, the different models used in the study must be compared to see which one has the best predictive capacity.

Predictive statistical models

Random Forest Model

The Random forest model is a method derived from the *bagging process*[4], which starts from a long variety of interrelated trees and then averages them. This method gives very good results because it is a combination of predictor trees in such a way that each tree depends on the values of a random vector tested independently and with the same distribution for each of these (Breiman, 2001).

The algorithm with which this method is built follows the same methodology as *bagging*, however, the difference is that the *Random Forest* makes a random selection of m predictors before evaluating each division. What this generates is that the model is simpler to train and adjust. As a consequence of this, this methodology is widely used (Amit & Geman, 1997).

Trees are an ideal solution for *bagging,* because it has the uniqueness of registering complex data interaction structures. Bagging also averages many noisy, but roughly unbiased, models and is looking to reduce variation. Since trees carry a lot of noise, they benefit from averaging. Each of the trees is built according to the following algorithm.

- N is defined as the number of test cases, while M is defined as the number of variables in the classifier.
- m is defined as the number of input variables to be used to find the decision at a given node; it should be noted that m must be much less than M.
- You must choose a training set for this tree and use the rest of the test cases, this way you can estimate the error.
- For each node in the tree, you must randomly choose m variables on which to base the decision. Finally, the best partition of the training set must be calculated from the m variables.

For prediction a new case is pushed down the tree. Then it is assigned the label of the terminal node where it ends. This process is iterated by all the trees in the assembly, and the label that gets the most incidents is reported as the prediction.

Among the advantages of the Random Forest, it stands out that it is very useful for modelling very large databases, it provides an estimate of which of the variables are most important in the classification process, and it also offers an experimental methodology to detect the interaction that the variables under study can probably have.

Other advantages it brings is that it computes the prototypes that give information about the relationship between the variables and the classification, as well as the proximity between the

[4] The bagging technique was created by Leo Breiman and consists of creating different models using random samples with replacement and then assembling or combining the results. The main objective of bagging algorithms corresponds to the reduction of variance (Breiman, 1996).

pairs of cases that can be used in the groups, identifying outliers or giving interesting perspectives on the variables.

K-nearest Neighbours Model

The k-nearest Neighbours model is a non-parametric classification method that corresponds to a supervised classification, based on a learning process which is based on a set of training and prototypes. The concept of the algorithm is simple, because it classifies each data according to the distance of a new element with the existing ones, ordering these distances from smallest to largest in order to select the group to which it belongs, this group will be, for Therefore, the most frequent with the shortest distances (Silverman & Jones, 1989).

The method has several applications, such as pattern recognition, classification and regression, the last one being the one we will use in this study. This model is classified as a type of lazy learning, where the function is approached only locally and all the computation is deferred to the classification.

The training spaces correspond to the vectors found in a multidimensional space, which is described in terms of p attributes considering q classes for classification. The attribute values of the i-th example are represented by the p-dimensional vector.

$$x_i = (x_{1i}, x_2, \ldots, x_{pi}) \in X$$

Space should be determined by in localized areas containing the training examples. A point in the same space can then be classified into the class that contains the highest frequency of data and is closest to the designated point. Generally, to perform the statistical procedure, the Euclidean distance is used, then its equation is presented.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{p} (x_{ri} - x_{rj})^2}$$

To train the algorithm, you must store the vectors and class labels from the training examples. During the classification period, the evaluation of the point starts from the representation of a vector in a given space. Subsequently, the Euclidean distance between the stored vectors and the new vector must be calculated, and the closest vectors are selected. The new point is classified with the class that has the highest frequency in the selected vectors.

This method assumes that the closest neighbours give the best classification, however, the disadvantage of this is that there are many irrelevant attributes that give an incorrect classification. Despite the aforementioned, it is possible to correct the bias, by assigning a weight to the distances of each attribute, thus increasing the relevance of the most significant attributes. There are also other possibilities, such as adjusting weights with known training vectors. The most convenient way, before assigning the degree of importance, is to identify and eliminate the attributes that are considered irrelevant.

Linear Regression Model

A linear regression or fit model is intended to model a dependent random variable named as Y_t, from a set of independent predictor variables named as X_1, X_2, X_3, …, X_n all of a random nature. The basic structure of this model can be expressed by the following equation:

$$Y_t = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \cdots + X_n\beta_n + \varepsilon \quad (2)$$

Where:

Y_t: explained or dependent variable.

X: explanatory or independent variables.

β: regression coefficients that measure the influence that explanatory variables have on those explained.

n: number of independent parameters to consider in the regression.

ε: random variable that describes the disturbance or residual of the model. This variable collects the factors of reality that are not observable and therefore are associated with chance.

The application of linear regression models refers to the study of the relationships that exist between the explanatory variables with the explained variables. This relationship defines the general structure of the data and through it, future observations can be predicted and the relevance of the effect of the predictor variables on the response variable can be assessed (Rencher, 2005). In order to create a linear regression model, the following assumptions must be met:

- That the relationship between the variables is linear.
- That the errors ε1, ε2, ε3, …, εn are random and independent of each other, that is, that any error εi does not influence the value of the following error ε(i+1).
- That the errors ε1, ε2, ε3, …, εn have the same variance (Homoscedasticity).
- That the errors ε1, ε2, ε3, …, εn are normally distributed.
- That the errors ε1, ε2, ε3, …, εn have a mean of 0.
- That the correlation between the errors ε1, ε2, ε3, …, εn and the adjusted values is 0.

One way to check that the above conditions are met is to plot the residuals εi against the fitted values. When the linear model is valid, and the described conditions are fulfilled, the graph will not indicate an important pattern, the vertical dispersion of the points will not have significant variation with respect to the horizontal dispersion, except for the edges. When these conditions are satisfied there is no reason to doubt the assumptions of the linear model.

In the case of the specific study, the linear regression statistical model will be constructed based on a series of predictors that correspond to atmospheric indices (variables X), to predict other types of indices called sectoral (variables Y). The results of the model will be analyzed in a later chapter.

Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors Model (SARIMAX)

The Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (*SARIMAX*) model is derived from the integrated moving average autoregressive model commonly known as *ARIMA*. This model is based on the search for regressions and variations of the variables under study, in order to find future prediction patterns. This type of model explains future estimates through the observed data.

The *ARIMA* model is generally expressed as ARIMA (p, d, q) where p, d, and q are characterized by being positive integer terms and are an indicator of the order of the different components of the model and represent the autoregressive component, the integrated component, and the moving average. The *ARIMA* model is described as follows:

$$Y_t = -(\Delta^d Y_t - Y_t) + \emptyset_0 + \sum_{i=1}^{p} \emptyset_i \Delta^d Y_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Where:

Y_t: predicted value.

d: corresponds to the d differences that are necessary to convert the original series to stationary.

$\emptyset_i$: parameters belonging to the "autoregressive" part of the model.

θ: parameters belonging to the "moving averages" part of the model.

ϵ: associated error.

It should be noted that:

$$\Delta Y_t = Y_t - Y_{t-1}$$

In case of working with stationary series and with an external predictor, also called an exogenous variable, the *ARIMA* model can be transformed into the already described *SARIMAX* model. It is important that the exogenous variable has a registration period equal to that of the endogenous variable.

Evaluation of the predictive capacity of the models

To evaluate the predictive capacity of the models, it is necessary to use an objective scoring methodology, which starts from the comparison between the observed results and the data obtained from the different prediction models in the validation period. The metrics used as the endpoint correspond to the Pearson correlation coefficient, the determination coefficient R^2 and the mean squared error.

Pearson correlation coefficient

Pearson correlation coefficient measures the linear dependence between two random variables which must be quantified and have to be normally distributed. An important characteristic of this method is the independence it has with respect to the units of measurement of the analyzed variables since they can have different units. Similarly, the Pearson correlation coefficient defines the degree of relationship of two variables as long as they are both continuous.

When obtaining the Pearson's correlation coefficient, it is important to calculate the P-value, which measures the correlation significance. If the p-value is significant, it can be accepted that there is a correlation and that its magnitude corresponds to the magnitude indicated by the coefficient. In the event that the correlation is not significant, it can be interpreted as 0, that is, there is no linear correlation between both variables. The calculation of this coefficient starts from the following equation:

$$\rho X,Y = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X-\mu x)(Y-\mu y)]}{\sigma_X \sigma_Y}$$

Where:

σxy: covariance of the study variables.

σY: standard deviation of the variable Y.

σX: standard deviation of the variable X.

The range of the correlation index varies in the range [-1,1], where the sign indicates the meaning of the relationship. To understand better, the result is described below:

- If r = 1, there is a perfect positive correlation. This means that there is a total dependency between the two variables which is called a direct relationship. Because the sign is positive, it means that when one of the variables under study increases, the other does too.
- If 0 < r < 1, there is a positive correlation.
- If r = 0, there is no correlation between the variables under study. However, this does not necessarily imply that the variables are independent, because non-linear relationships may exist between the two variables.
- If -1 < r < 0, there is a negative correlation.
- If r = -1, there is a perfect negative correlation. The index indicates a total dependency between the two variables which is called the inverse relationship: when one of them increases. Because the sign is negative, it means that when one of the variables under study increases, the other decreases in a constant proportion.

Determination coefficient R^2

This coefficient determines the quality of the model to replicate the results and the proportion of the variation of the results that can be explained by the model. In the coefficient of determination R^2, it expresses the relationship between the error obtained when comparing the values measured in the field (y), against the simulated values (x) and the difference between the values taken in the field are their average (yo- y). Utilizing its formulation, an attempt is made to verify that the modelled data to explain better the data taken in the field than the simple average of these. Next, the corresponding formulation for obtaining it is presented.

$$R^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2}$$

Where:

σxy: covariance of the study variables.

σY: variance of the variable Y.

σX: variance of the variable X.

When the difference between the data measured in the field and the average is very significant and the error between the data measured in the field and the simulated ones is small, the quotient will tend to zero, that is, the coefficient of determination will tend to be one.

- *The mean squared error (MSE)*

The mean square error (*MSE*) is defined as an estimator in charge of measuring the average of the squared errors, that is, using this statistical parameter, the predicted values can be

compared with the observed or known values. Its calculation is derived from the following formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_p)^2$$

Where:

Y_i: observed variable.

Y_o: predicted variable.

n: data number.

The smaller the mean square error, the less difference there will be between the predicted values with respect to the observed values, that is, when a smaller mean square error is obtained, the better the data reproduced by the regression method will be. In the case where the MSE is equal to zero, it means that the actual or observed value is the same as the calculated value.

7.6 ANALYSIS OF RESULTS

This chapter describes and analyzes the results obtained by applying the methodologies, regression models and time series: Random Forest (RF), K-nearest Neighbours Regression (KNN), Linear Regression (LR) and Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors. In addition, some recommendations are made regarding the different groups of sector indices on which it could be implemented in an operational forecasting system.

Results of the Random Forest model

The *Random Forest* statistical model was built using atmospheric indices as predictors and sector indices as predictands. The model was reproduced for three different time lags of the series of predictands over the series of predictors, corresponding to 1, 2 and 3 months. The cities of Europe selected to carry out the regression correspond to 378, which are represented in Figure .

According to the description in sub-section 7.5 and to evaluate the predictive capacity of the *Random Forest* model, the determination coefficients (R^2), the Pearson correlation coefficients (R) and the mean square error were calculated (*R*). Taking into account the values obtained and to explain the results more precisely, the determination coefficient (R^2) was taken as representative.

Subsequently, the R^2 values were represented by graphs with colour scales, to observe which of the sectoral indices responded better to the predictive capabilities of the model, having atmospheric indices as predictors. The results obtained are presented in Figure 7.5 for the

model built with one month as the selected time offset, in Figure 7.6 with two months and in Figure 7.7 with three months, respectively.

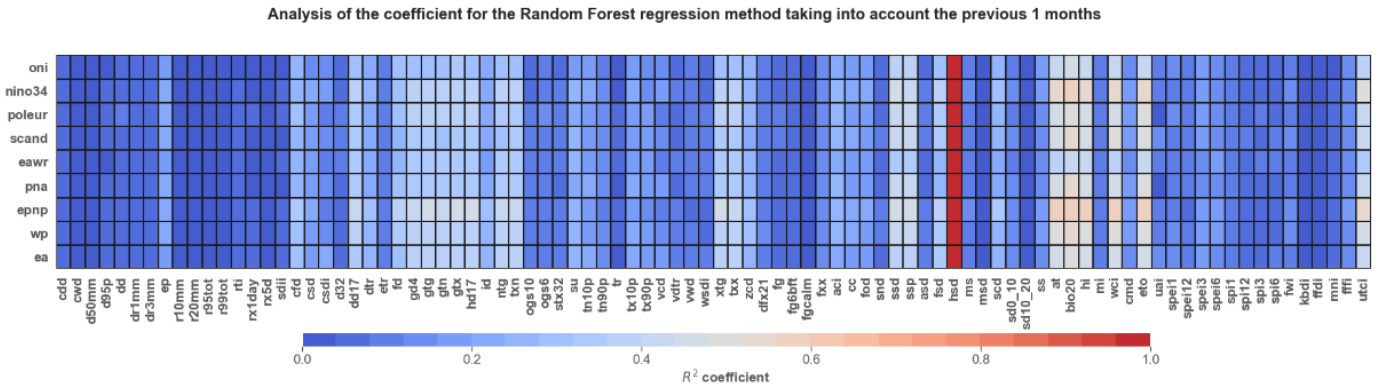

Figure 7.5. Analysis of the coefficient for the *Random Forest regression method* taking into account the previous 1 months.

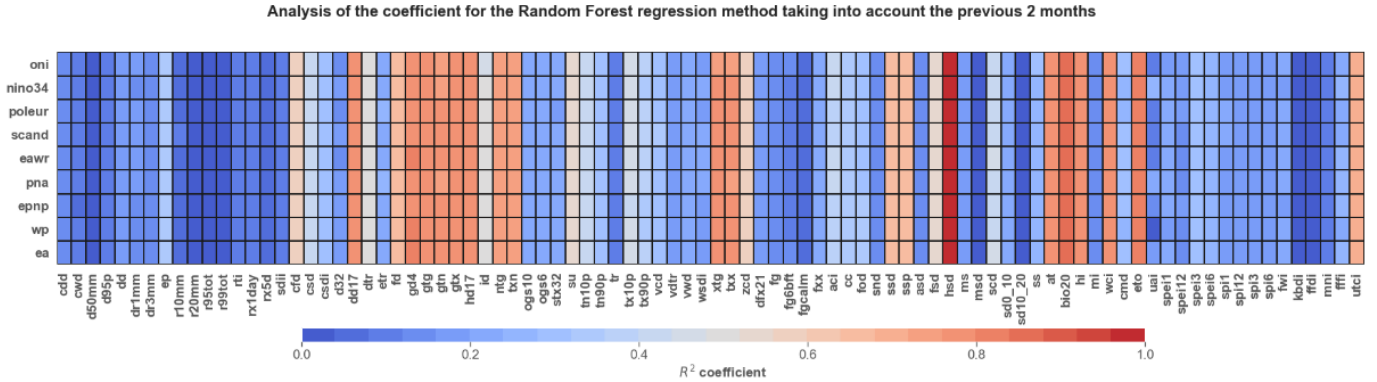

Figure 7.6 Analysis of the coefficient for the *Random Forest regression method* taking into account the previous 2 months.

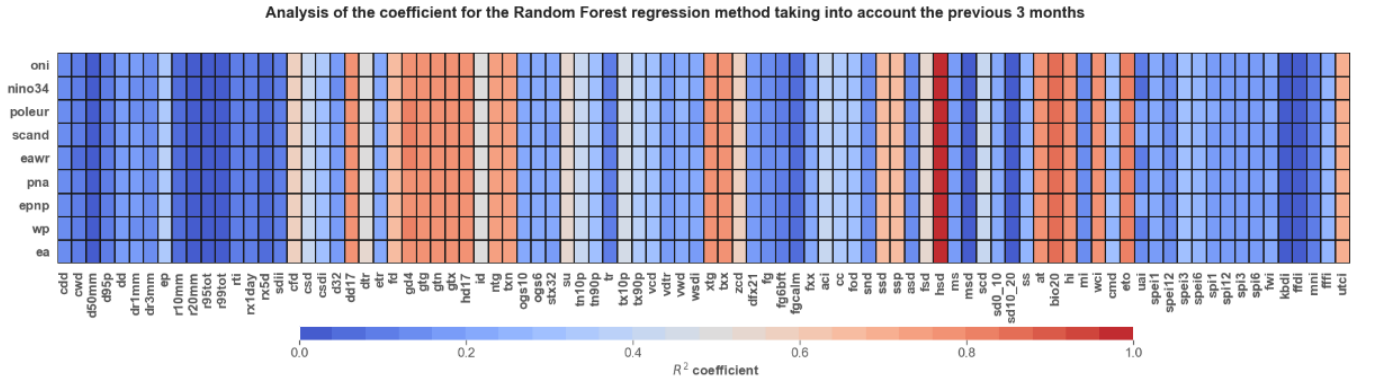

Figure 7.7. Analysis of the coefficient for the *Random Forest* regression method taking into account the previous 3 months.

Prior to the analysis, it is important to mention that a warmer colour, with tendencies to red colour, indicates a coefficient of determination closer to 1, while a colder colour, with tendencies to blue, indicates a coefficient of determination closer to 0.

In Figure 7.5-7.7, it can be seen that the models that showed warmer colours with tendencies to red colour correspond to those that were calculated with 2 and 3 months of time lag. This phenomenon is accentuated a little more in the indexes of the groups corresponding to temperature, tourism, bioclimatic and snow that were presented in Table 7.3. This indicates that the simulations with greater time lag obtained determination coefficients with values closer to 1, which can be translated as a better predictive capacity of the model.

Analyzing the predictive behaviour of atmospheric indices, homogeneity is observed in the results, which indicates that there is no significant variation in their predictive capacity over a given sector index. An example of this can be seen in Figure 7.6 where the predictive capacity of the atmospheric indices in relation to the *utci* tourism index, corresponding to the last box of the graph, does not present colour variations, indicating that the determination coefficient has a very small variation for all predictions calculated from the evaluated atmospheric phenomena. Due to the homogeneity of the results obtained in each of the sectorial indices with respect to the atmospheric ones, and bearing in mind that the best coefficients of determination were obtained from the model constructed with a lag time of 2 and 3 months, improving a little more the R^2 value for the 2-month model, a box diagram was prepared from it, the results are presented in Figure .

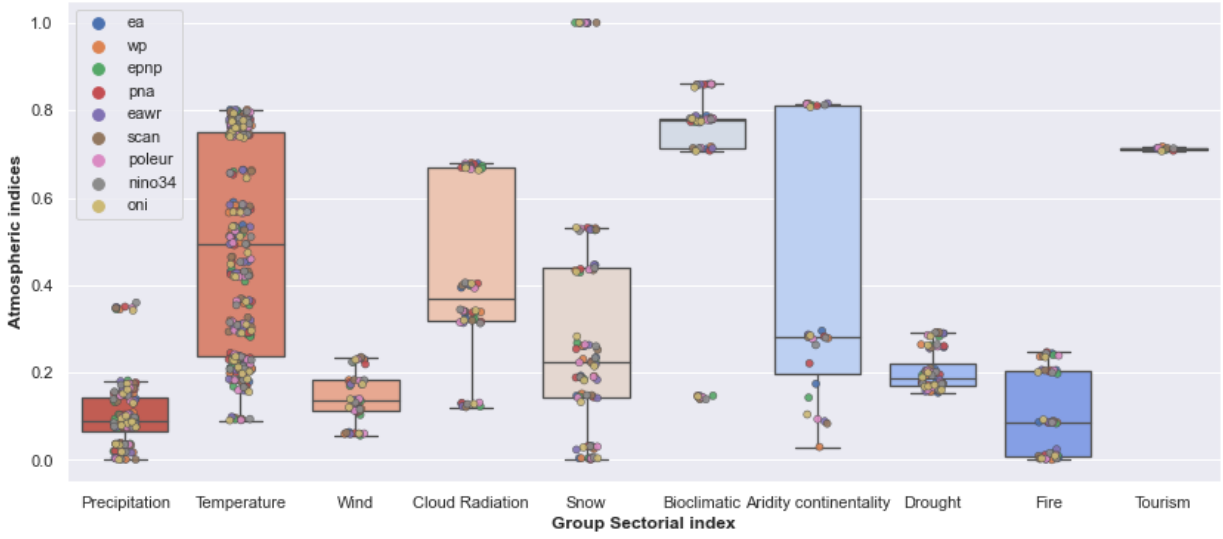

Figure 7.8. Boxplot for Analysis of the coefficient for the *Random Forest* method.

In Figure 7.8 different coloured points can be visualized, which represent the results of the coefficient of determination for each of the groups. The determination coefficients calculated for the sectoral indices of the Precipitation group are also contained, which vary between 0.0 and 0.19, presenting a homogeneous distribution around the median that corresponds to 0.1. Some

extreme values of the order of 0.38 can be observed. Likewise, you can see groups that have a behaviour similar to that of Precipitation, such as Wind, Drought and Fire, where more than 95% of their R^2 values are below 0.3.

In the case of the indices corresponding to the temperature group, the results of R^2 can be seen with a very wide range of variation that ranges between 0.1 and 0.81, presenting a homogeneous distribution around the median that corresponds to 0.50. Other groups with a similar behaviour correspond to Cloud Radiation, Snow and Aridity Continentality which have very wide ranges of variation and a median below 0.4. Some extreme values of the Snow group stand out, which reach values equal to 1.

Finally, there are the Bioclimatic and Tourism groups, which have a median above 0.7, indicating that at least 50% of the data had coefficients of determination above this value. In the case of tourism, there is no dispersion of the data, since there is only one index for each atmospheric group. Some extreme values of the Bioclimatic group stand out, which reach minimum values equal to 0.17.

Suggestions for implementation of an operational forecasting system for the Random Forest model

Taking into account the methodological description described in Section 7.5 and based on the results obtained for the *Random Forest model*, a value of the determination coefficient has been defined above which the implementation of an operational prediction system can be recommended, which corresponds to the values of R^2 that are above 0.8. Taking this criterion into account, Table 7.6 was prepared, which presents the sector indices by group that meet the described condition and therefore have a good assessment to recommend the implementation of said system.

SECTORAL INDEX GROUPS	ATMOSPHERIC INDEX	SECTORAL INDEX	R^2
Snow	ea	hsd	1.00
Snow	wp	hsd	1.00
Snow	epnp	hsd	1.00
Snow	pna	hsd	1.00
Snow	eawr	hsd	1.00
Snow	scan	hsd	1.00

Snow	poleur	hsd	1.00
Snow	nino34	hsd	1.00
Snow	oni	hsd	1.00
Bioclimatic	epnp	bio20	0.86
Bioclimatic	ea	bio20	0.86
Bioclimatic	poleur	bio20	0.86
Bioclimatic	eawr	bio20	0.86
Bioclimatic	scan	bio20	0.86
Bioclimatic	pna	bio20	0.86
Bioclimatic	wp	bio20	0.86
Bioclimatic	nino34	bio20	0.86
Bioclimatic	oni	bio20	0.85
Aridity continentality	eawr	eto	0.82
Aridity continentality	scan	eto	0.82
Aridity continentality	wp	eto	0.81
Aridity continentality	ea	eto	0.81
Aridity continentality	poleur	eto	0.81
Aridity continentality	nino34	eto	0.81
Aridity continentality	epnp	eto	0.81
Aridity continentality	pna	eto	0.81
Aridity continentality	oni	eto	0.81
Temperature	wp	gd4	0.80
Temperature	eawr	gd4	0.80
Temperature	ea	gd4	0.80

Table 7.6. Recommended sector indices for the implementation of a system for the Random Forest regression method taking into account the previous 2 months.

Results of the K-nearest Neighbours regression model

The *K-nearest Neighbours regression* statistical model, was built using atmospheric indices as predictors and sector indices as predictands. The model was reproduced for three different time lags of the series of predictands over the series of predictors, corresponding to 1, 2 and 3 months. The cities of Europe selected to carry out the regression correspond to 378, which are represented in Figure .

2

According to the description in Section 7.5 and to evaluate the predictive capacity of the *K-nearest Neighbours regression* model, the determination coefficients (R^2), the Pearson correlation coefficients (R) and the mean square error were calculated (*R*). Taking into account the values obtained and to explain the results more precisely, the determination coefficient (R^2) was taken as representative. Subsequently, the R^2 values were represented by graphs with colour scales, to observe which of the sectoral indices responded better to the predictive capabilities of the model, having atmospheric indices as predictors. The results obtained are presented in Figure 7.9 for the model built with one month as the selected time offset, in Figure 7.10 with two months and in Figure 7.11 con with three months, respectively.

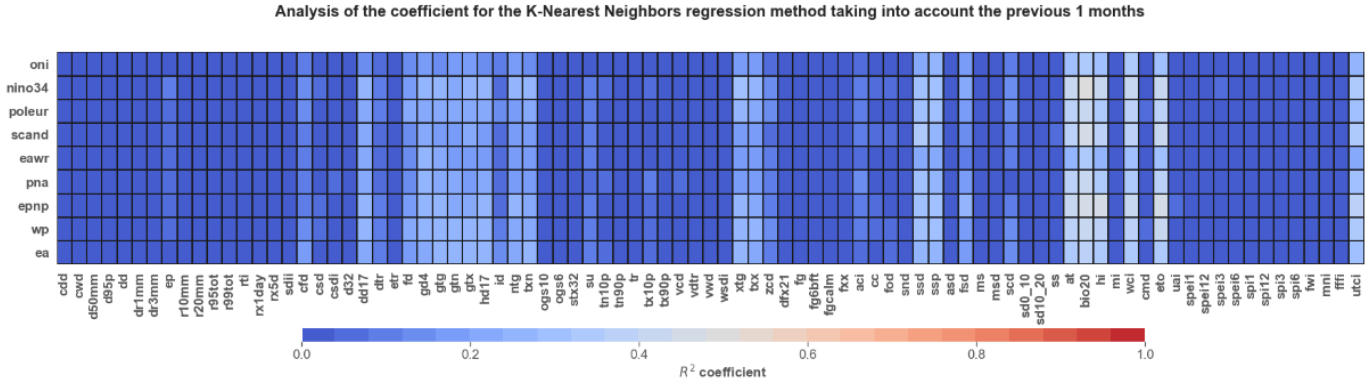

Figure 7.9. Analysis of the coefficient for the *Knearest Neighbours regression* method taking into account the previous 1 months.

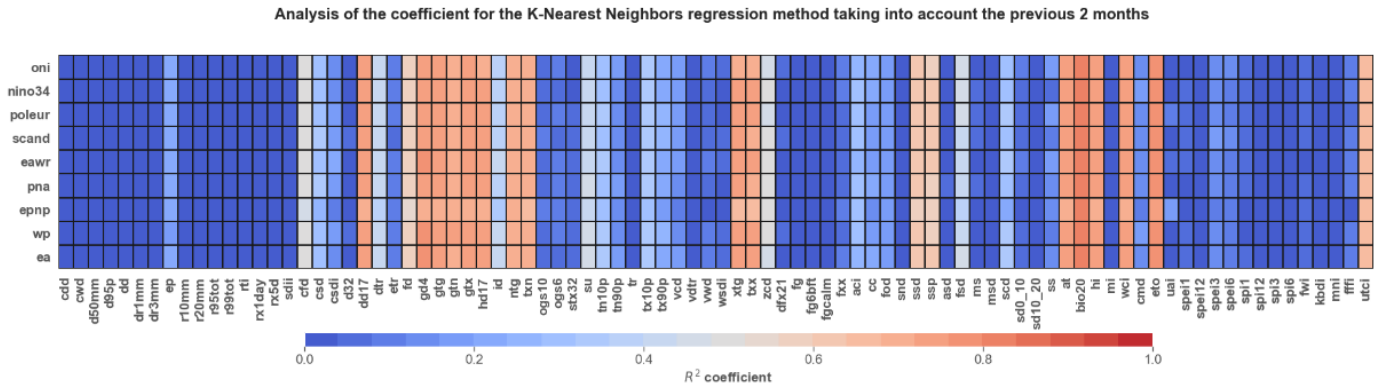

Figure 7.10. Analysis of the coefficient for the *Knearest Neighbours regression* method taking into account the previous 2 months.

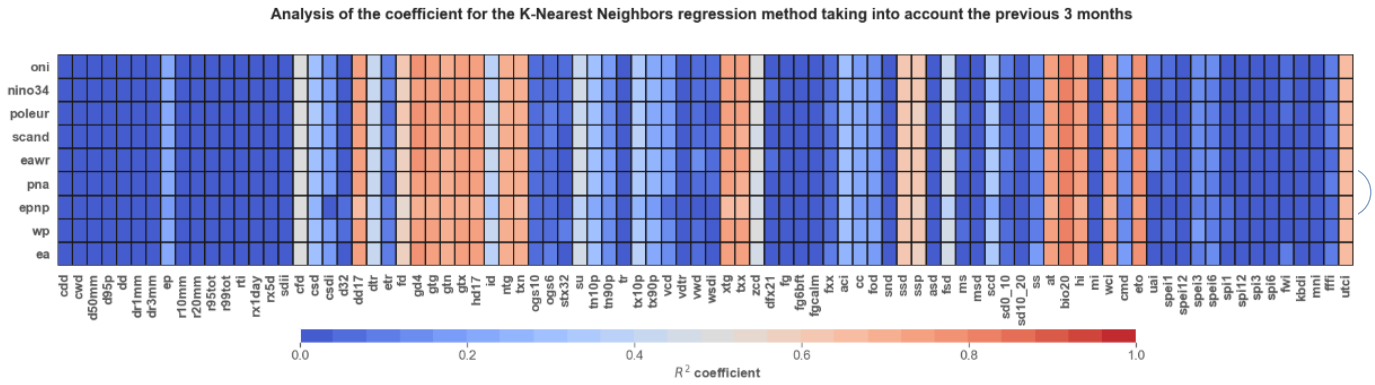

Figure 7.11. Analysis of the coefficient for the *Knearest Neighbours regression* method taking into account the previous 3 months.

Prior to the analysis, it is important to mention that a warmer colour, with tendencies to red colour, indicates a coefficient of determination closer to 1, while a colder colour, with tendencies to blue, indicates a coefficient of determination closer to 0. In Figure 7.12, it can be seen that the models that showed warmer colours with tendencies to red colour correspond to those that were calculated with 2 and 3 months of time lag. This phenomenon is accentuated a little more in the indexes of the groups corresponding to temperature, tourism, bioclimatic and snow that were presented in Table 7.5. This indicates that the simulations with greater time lag obtained determination coefficients with values closer to 1, which can be translated as a better predictive capacity of the model.

Analyzing the predictive behaviour of atmospheric indices, homogeneity is observed in the results, which indicates that there is no significant variation in their predictive capacity over a given sector index. An example of this can be seen in Figure 7.12 where the predictive capacity of the atmospheric indices in relation to the *utci* tourism index, corresponding to the last box of the graph, does not present colour variations, indicating that the determination coefficient has a very small variation for all predictions calculated from the evaluated atmospheric phenomena.

Due to the homogeneity of the results obtained in each of the sectorial indices with respect to the atmospheric ones, and bearing in mind that the best coefficients of determination were obtained from the model constructed with a lag time of 2 and 3 months, improving a little more the R^2 value for the 2-month model, a box diagram was prepared from it, the results are presented in Figure 7.12.

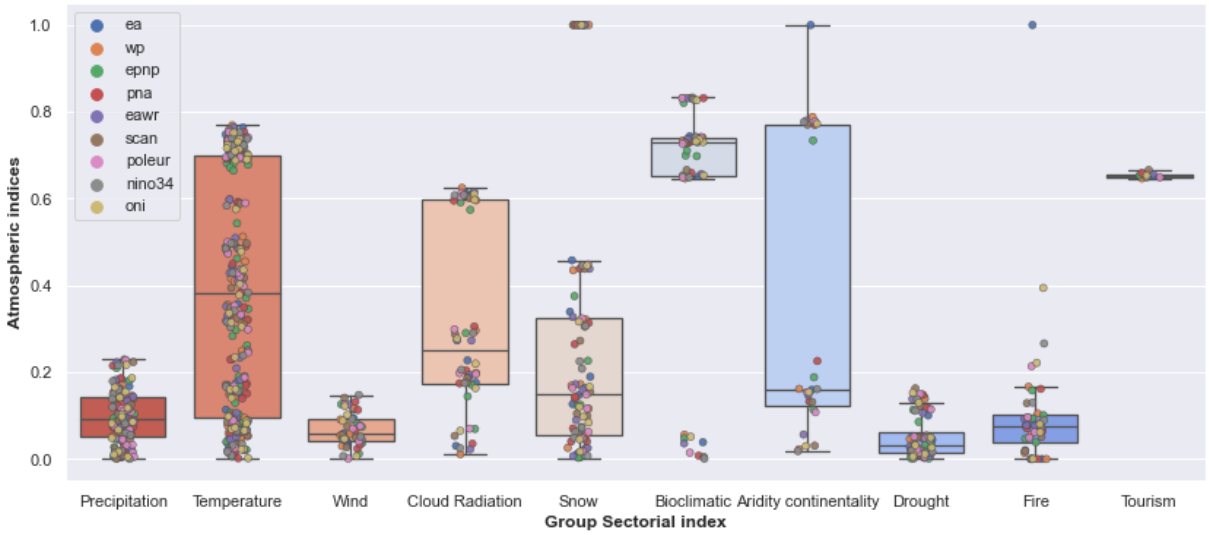

Figure 7.12. Boxplot for Analysis of the coefficient for the K-nearest Neighbours regression method.

In Figure 7.12 different coloured points can be visualized, which represent the results of the coefficient of determination for each of the groups. The determination coefficients calculated for the sectoral indices of the Precipitation group are also contained, which vary between 0.0 and 0.22, presenting a homogeneous distribution around the median that corresponds to 0.1. You can also see groups that have a behaviour similar to that of precipitation, such as Wind, Drought and Fire, where more than 95% of their R^2 values are below 0.2, except for the fire group, which reaches values extremes of up to 0.4 and 1.0, for the values predicted from the atmospheric index oni and ea.

In the case of the indices corresponding to the temperature group, the results of R^2 can be seen with a very wide range of variation that ranges between 0.0 and 0.78, presenting a homogeneous distribution around the median that corresponds to 0.39. Other groups with a similar behaviour correspond to Cloud Radiation, Snow and Aridity Continentality which have very wide ranges of variation and a median below 0.4. However, some extreme values of the Snow and Aridity Continentality group stand out, which reach values equal to 1. Finally, there are the Bioclimatic and Tourism groups, which have a median above 0.6, indicating that at least 50% of the data had coefficients of determination above this value. In the case of tourism, there is no dispersion of the data, since there is only one index for each atmospheric group. Some extreme values of the Bioclimatic group stand out, which reach minimum values equal to 0.

Suggestions for implementation of an operational forecasting system for the K-nearest Neighbours regression model

Taking into account the methodological description described in Section 7.5 and based on the results obtained for the *K-nearest Neighbours regression model*, a value of the determination coefficient has been defined above which the implementation of an operational prediction system can be recommended, which corresponds to the values of R^2 that are above 0.8. Taking this criterion into account, Table 7.7 was prepared, which presents the sector indices by group

that meet the described condition and therefore have a good assessment to recommend the implementation of said system.

SECTORAL INDEX GROUPS	ATMOSPHERIC INDEX	SECTORAL INDEX	R^2
Snow	ea	hsd	1.00
Snow	wp	hsd	1.00
Snow	epnp	hsd	1.00
Snow	pna	hsd	1.00
Snow	eawr	hsd	1.00
Snow	scan	hsd	1.00
Snow	poleur	hsd	1.00
Snow	nino34	hsd	1.00
Snow	oni	hsd	1.00
Aridity continentality	ea	uai	1.00
Fire	ea	kbdi	1.00
Bioclimatic	scan	bio20	0.83
Bioclimatic	wp	bio20	0.83
Bioclimatic	eawr	bio20	0.83
Bioclimatic	pna	bio20	0.83
Bioclimatic	ea	bio20	0.83
Bioclimatic	poleur	bio20	0.83
Bioclimatic	nino34	bio20	0.83
Bioclimatic	oni	bio20	0.83
Bioclimatic	epnp	bio20	0.82

Table 7.7. Recommended sector indices for the implementation of a system for the K-nearest Neighbours regression method taking into account the previous 2 months.

Results of the Linear Regression Model

The *Linear Regression* statistical model, was built using atmospheric indices as predictors and sector indices as predictands. The model was reproduced for three different time lags of the series of predictands over the series of predictors, corresponding to 1, 2 and 3 months. The cities of Europe selected to carry out the regression correspond to 378, which are represented in Figure .

2

According to the description in sub-section 7.5. and to evaluate the predictive capacity of the *Linear Regression* model, the determination coefficients (R^2), the Pearson correlation coefficients (R) and the mean square error were calculated (*R*). Taking into account the values obtained and to explain the results more precisely, the determination coefficient (R^2) was taken as representative. Subsequently, the R^2 values were represented by graphs with colour scales, to observe which of the sectoral indices responded better to the predictive capabilities of the model, having atmospheric indices as predictors. The results obtained are presented in Figure for the model built with one month as the selected time offset, in Figure 7.14 with two months and in Figure with three months.

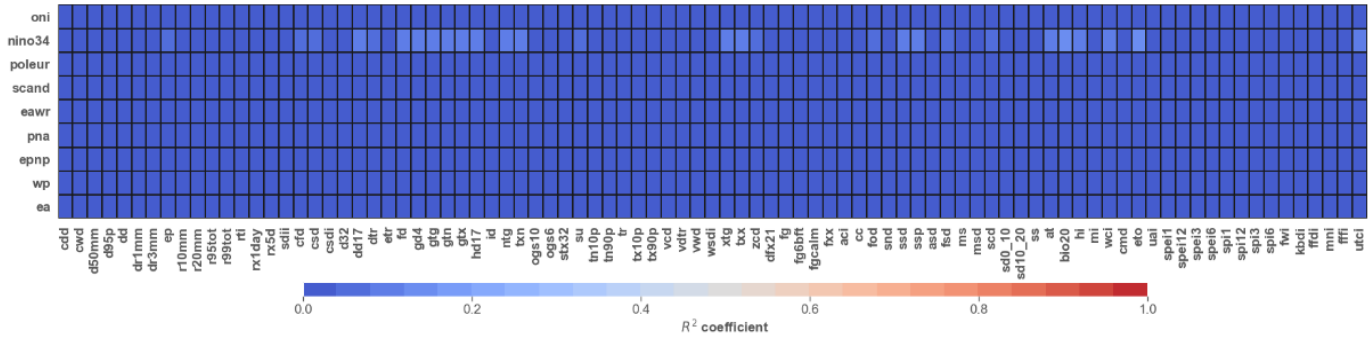

Figure 7.13. Analysis of the coefficient for the *Linear Regression* method taking into account the previous 1 months.

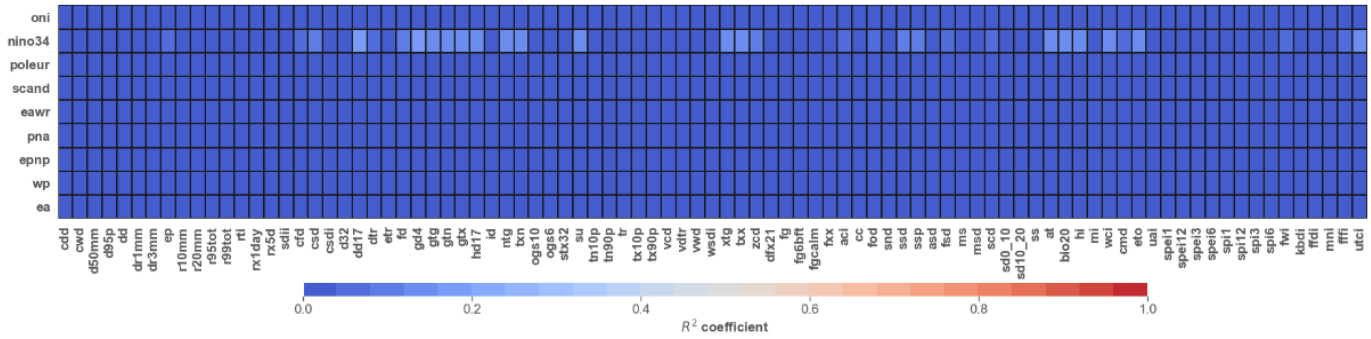

Figure 7.14. Analysis of the coefficient for the *Linear Regression* method taking into account the previous 2 months.

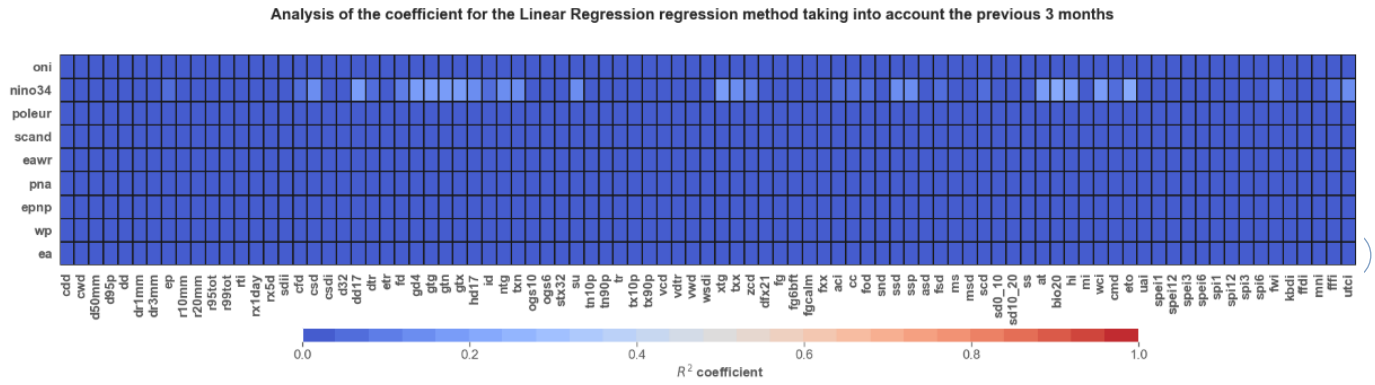

Figure 7.15. Analysis of the coefficient for the *Linear Regression* method taking into account the previous 3 months.

Prior to the analysis, it is important to mention that a warmer colour, with tendencies to red colour, indicates a coefficient of determination closer to 1, while a colder colour, with tendencies to blue, indicates a coefficient of determination closer to 0.

In this model, R^2 values close to 1 are not observed, since the figures constructed for each lag time present a colour scale with a blue gradient, indicating that the coefficients of determination vary in a range of 0.0 to 0.4, respectively. The values obtained for the Niño atmospheric index are highlighted, which reaches values of up to 0.4.

Results of the Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors Model (SARIMAX)

The *SARIMAX* statistical model, was built using atmospheric indices as predictors and sector indices as predictands. The model was reproduced for three different time lags of the series of predictands over the series of predictors, corresponding to 1, 2 and 3 months. The cities of Europe selected to carry out the regression correspond unlike the other methods applied only to 3, due to the computational capacity demanded by the model. The described cities are represented in Figure.

According to the description in section 7.5 and to evaluate the predictive capacity of the *SARIMAX* model, the determination coefficients (R^2), the Pearson correlation coefficients (R) and the mean square error were calculated (*R*). Taking into account the values obtained and to explain the results more precisely, the determination coefficient (R^2) was taken as representative. Subsequently, the R^2 values were represented by graphs with colour scales, to observe which of the sectoral indices responded better to the predictive capabilities of the model, having atmospheric indices as predictors.

The results obtained are presented in Figure for the model built with one month as the selected time offset, in Figure 7.17 with two months and in Figure 7.18 con with three months, respectively.

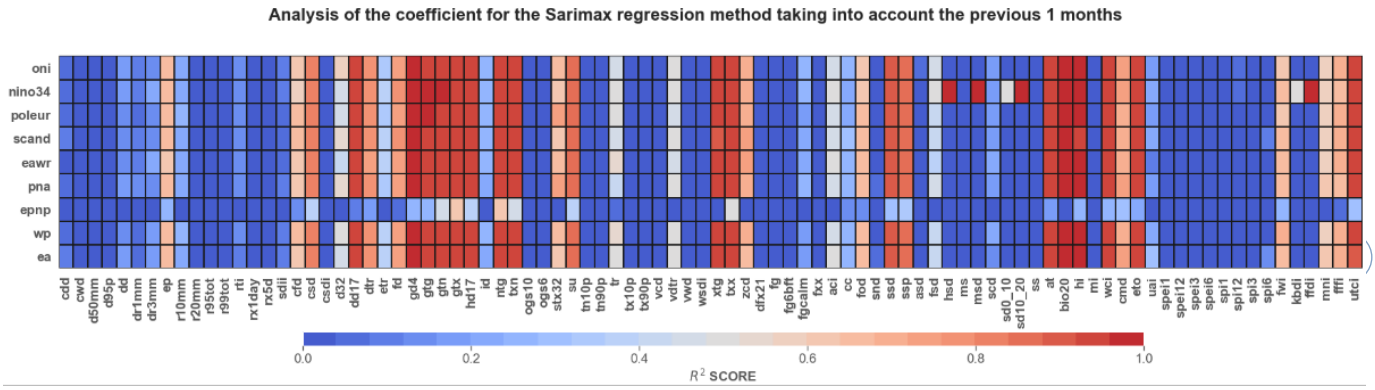

Figure 7.16. Analysis of the coefficient for the *SARIMAX* method taking into account the previous 1 months.

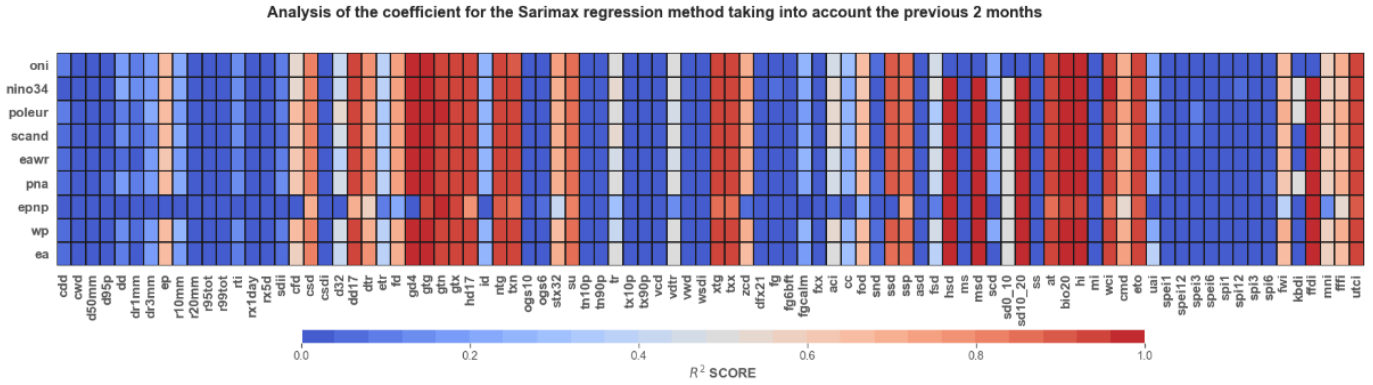

Figure 7.17. Analysis of the coefficient for the *SARIMAX* method taking into account the previous 2 months.

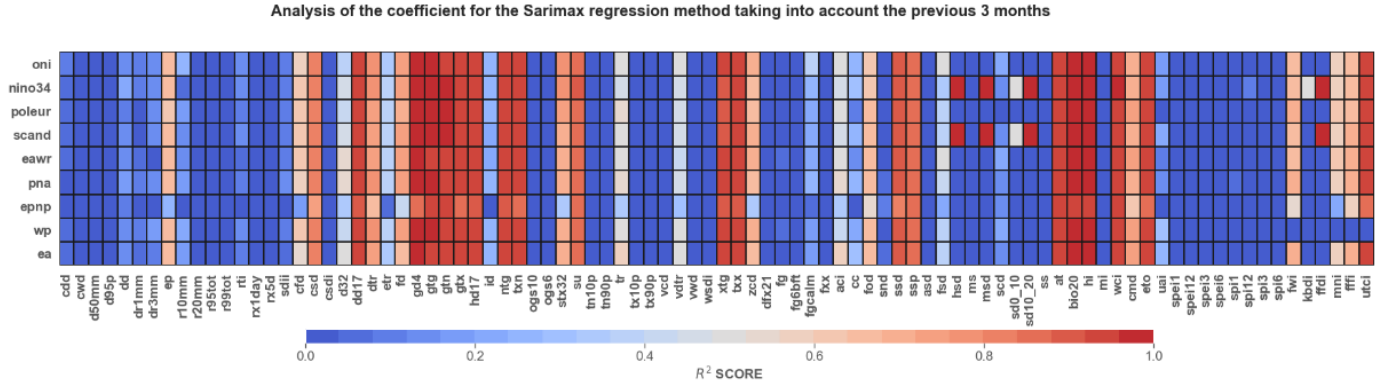

Figure 7.18. Analysis of the coefficient for the *SARIMAX* method taking into account the previous 3 months.

In Figure , you can see that the colours vary depending on the lag time used. The largest number of squares in a warm tone with more red tendencies, are presented in the model made after a delay time of 2 months. This phenomenon is accentuated a little more in the indexes of the groups corresponding to temperature, tourism, bioclimatic and snow that were presented in

Table 7.5. This indicates that the simulations with greater time lag obtained determination coefficients with values closer to 1, which can be translated as a better predictive capacity of the model.

Regarding the predictive behaviour of atmospheric indices over sectoral indices, there are important differences in the results, especially those predicted by the *epnp* atmospheric index. An example of this can be seen in Figure 7.19, column 8 (precipitation group index *ep*), where the colour scale obtained shows variations just in the atmospheric index *epnp*, indicating that the coefficient of determination changes, with respect to the other indices calculated.

Taking into account that the best determination coefficients were obtained from the model built with a lag time of 2 months, a boxplot was prepared in order to observe the behaviour of the indices by group, the results are presented in the Figure .

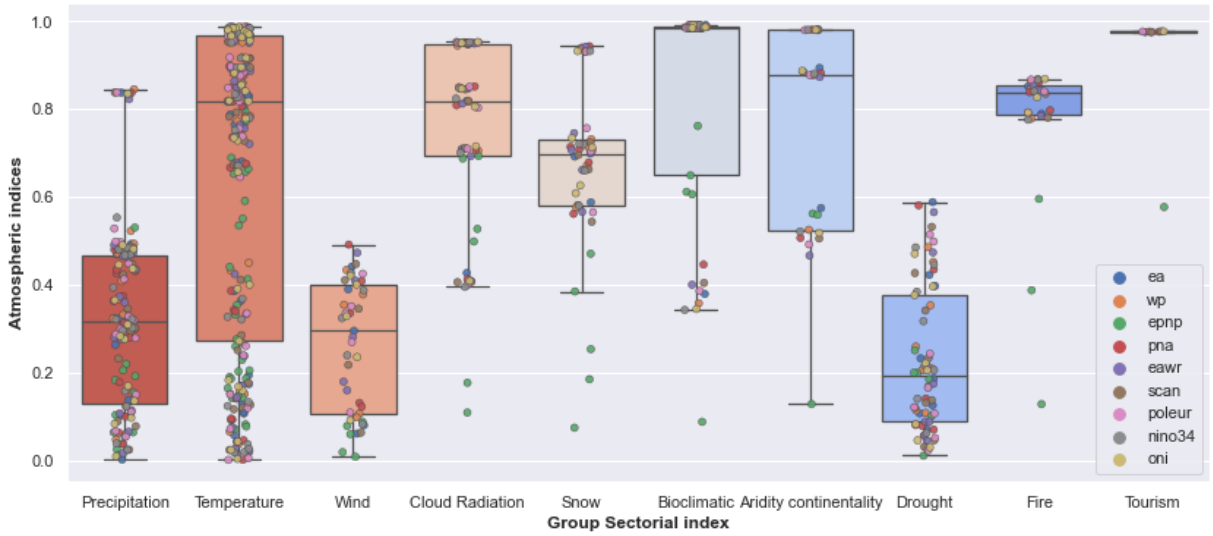

Figure 7.19. Boxplot for Analysis of the coefficient for *SARIMAX* regression method.

In Figure different coloured points can be visualized, which represent the results of the coefficient of determination for each of the groups. The determination coefficients calculated for the sectoral indices of the Precipitation group are also contained, which vary between 0.0 and 0.82, presenting a homogeneous distribution around the median that corresponds to 0.34. You can also see groups that have a behaviour similar to that of precipitation, such as Wind and Drought, where more than 95% of their R^2 values are below 0.34.

For the indices corresponding to the temperature group, 50% of the R^2 results oscillate in the range of 0.81 and 1.00 with a grouped distribution, while the other 50% oscillate in the range of 0.81 to 0.0 with a distribution more dispersed. The groups corresponding to Cloud Radiation, Snow, Aridity and Fire have ranges of variation with less amplitude than the group of temperature and all have a median below 0.85.

The Bioclimatic and Tourism groups stand out, which have a median above 0.9, which indicates that at least 50% of the data had coefficients of determination above this value. In the case of

tourism, since there is only one index for each atmospheric group, there is no dispersion in the results. Finally, and as in Figure 7.19, an atypical behaviour is observed in the determination coefficients calculated from the atmospheric index *epnp*, in relation to the others.

7.6.1 Suggestions for implementation of an operational forecasting system for the K-nearest Neighbours regression model

Taking into account the methodological description described in sub-section 7.5 and based on the results obtained for the *SARIMAX*, a value of the determination coefficient has been defined above which the implementation of an operational prediction system can be recommended, which corresponds to the values of R^2 that are above 0.8. Taking this criterion into account, Table 7.8 was prepared, which presents the sector indices by group that meet the described condition and therefore have a good assessment to recommend the implementation of said system.

SECTORAL INDEX GROUPS	ATMOSPHERIC INDEX	SECTORAL INDEX	R^2
Precipitation	oni, scan, pna, nino34, ea, poleur, wp, eawr.	ep	0.845-0.823
Temperature	oni, scan, pna, nino34, ea, poleur, wp, eawr.	gd4, gnt gtg, gtx hd17, dd17 ntg, hi, wci, at, txx, xtg, txn, su, csd, dtr, fd, stx32, zcd, d32, cfd.	0.987-0.804
Cloud radiation	oni, scan, pna, nino34, ea, poleur, wp, eawr.	ssd,ssp, fod, aci	0.953-0.803
Snow	oni, scan, pna, nino34, ea, poleur, wp, eawr.	scd	0.943-0.931
Bioclimatic	oni, scan, pna, nino34, ea, poleur, wp, eawr.	Bio20, hi, wci, at,	0.992-0.984
Aridity continentality	oni, scan, pna, nino34, ea, poleur, wp, eawr.	eto, cmd	0.981-0.873
Fire	oni, scan, pna, nino34, ea, poleur, wp, eawr.	Fffi, fwi	0.869–0.827
Tourism	oni, scan, pna, nino34, ea, poleur, wp, eawr.	utci	0.977-0.975

Table 7.8. Recommended sector indices for the implementation of a system for the SARIMAX regression method taking into account the previous 2 months.

7.7 CONCLUSIONS

Statistical models were constructed for three time lags, corresponding to 1, 2 and 3 months. To evaluate the predictive capacity of the models, the coefficient of determination (R^2), the Pearson's correlation coefficient (R) and the mean square error were calculated. In accordance with the obtained values and in order to explain the results more precisely, the determination coefficient (R^2) was taken as representative. With the coefficients obtained and for each one of the models, a graph with a colour scale was constructed to determine which of the models gave better results. Once these graphs were analyzed, the lag time with the best predictive capacity was selected for each model, and a boxplot was constructed from it to evaluate the results by groups. Through the procedure described, the following conclusions were reached:

- The *Random Forest* model built from 1 month of time lag, was the one that obtained lower coefficients of determination, compared to the 2 and 3 month models, therefore it is concluded that it has a lower predictive capacity. Subsequently, the elaborated model is located with a time lag of 3 months, which gave R^2 values slightly lower than those calculated with the model constructed with 2 months. Finally, the model with the best predictive capacity corresponds to the one developed with a time lag of 2 months.

- The *Random Forest* model elaborated with 2 months of time lag, obtained determination coefficients below 0.3 for the indices contained within the Precipitation, Wind, Drought and Fire groups, which translates as low predictive capacity. For the groups Temperature, Cloud Radiation, Snow and Aridity Continentality, R^2 values were found with a lot of variability, which in relation to the groups described at the beginning obtained average values of 0.5, with which it can be concluded that for these groups the model had an intermediate predictive capacity. Finally, there are the Tourism and Bioclimatic groups where more than 90% of the data obtained coefficients of determination above 0.7, which indicates that the model explains the results well and therefore its predictive capacity is good.

- The *K-nearest Neighbours* model built from 1 month of time lag, was the one that obtained lower coefficients of determination, compared to the 2 and 3 month models, therefore it is concluded that it has a lower predictive capacity. Subsequently, the elaborated model is located with a time lag of 3 months, which gave R^2 values slightly lower than those calculated with the model constructed with 2 months. Finally, the model with the best predictive capacity corresponds to the one developed with a time lag of 2 months.

- The *K-nearest Neighbours* model elaborated with 2 months of time lag, obtained determination coefficients below 0.25 for the indices contained within the Precipitation, Wind, Drought and Fire groups, which translates as low predictive capacity. For the groups Temperature, Cloud Radiation, Snow and Aridity Continentality, R^2 values were found with a lot of variability, which in relation to the groups described at the beginning obtained average values of 0.4, with which it can be concluded that for these groups the model had a low predictive capacity. Finally, there are the Tourism and Bioclimatic groups where more than 90% of the data obtained coefficients of determination above

0.6, which indicates that the model explains the results well and therefore its predictive capacity is good.

- The *Linear Regression* model did not show satisfactory results, since the values obtained in the determination coefficients for the three models constructed from the different lag times presented values close to 0, highlighting those obtained from the Niño34 atmospheric index, the which reached values of up to 0.4. For this reason, it is concluded that the predictive capacity of the model is low, almost null for this model.

- The *SARIMAX* model built from 1-month time lag presented a lower number of determination coefficients with values close to 1, compared to the model developed for 2-month time lags. The model built from 3 months of lag, presented a behaviour similar to that of 1 month. In conclusion, the model that best explains the results and that therefore yielded a greater number of determination coefficients with values close to 1, corresponded to the model developed for a delay time of 2 months.

- The *SARIMAX* model elaborated with 2 months of time lag, obtained determination coefficients below 0.35 for the indices contained within the Precipitation, Wind and Drought, which translates as low predictive capacity. For the groups Temperature, Cloud Radiation, Snow and Aridity Continentality, R^2 values were found with a lot of variability, which in relation to the groups described at the beginning obtained average values of 0.85, with which it can be concluded that for these groups the model had a good predictive capacity. Finally, there are the Tourism and Bioclimatic groups where more than 90% of the data obtained coefficients of determination above 0.9, which indicates that the model explains the results very well and therefore its predictive capacity is very good.

- There is a pattern that is repeated in all the models built and is related to the small difference that can be observed between the determination coefficients (R^2) calculated from the relationship of any sector index, with respect to each atmospheric index, which in most cases they have a variation percentage of less than 1%. An example of this pattern can be seen in the R^2 results corresponding to the utci tourism index obtained for a lag time of 2 months (last column of Figure , Figure and Figure), where the colour gradient it remains almost constant, which indicates a minimal variation in the determination coefficients.

- Making a joint evaluation of the models that gave the best results, built with a lag time of 2 months, it can be concluded that *SARIMAX* is located first, which had the best predictive capacity, presenting very good determination coefficients for the groups evaluated. Secondly, the *Random Forest* model, which has a much lower predictive quality in relation to *SARIMAX*, is located for each of the groups evaluated. In third place is the *K-nearest Neighbours* model, which presented slightly lower R^2 results, but without significant differences compared to the *Random Forest* model. Finally, there is the *Linear Regression* model, which in general did not present satisfactory results for the analysis.

8. CONCLUSIONS

Some key results associated with D5.3 are listed below:

1) The teleconnection indices which present the largest influence over climate indices in Europe are North Atlantic Oscillation (NAO), East Atlantic pattern (EA), Scandinavian pattern (SCA) and the East Atlantic/West Russia pattern (EA/WR), with a seasonal- and region-dependent effect. The largest observed trends concerning teleconnection indices correspond to NAO (especially winter NAO) and EA, which are closely related to precipitation-derived indices and cold extremes in most continental Europe, and more locally associated with temperature-derived (hot extremes) indices in the Mediterranean basin. Details are discussed in Section 2.

2) Six regions have been identified where interannual variability of growing season onset is sensitive to one or both of the winter NAO and EA. The strongest NAO signal is found around the English Channel and the EA signal is strong across much of continental Europe (particularly for Italy, Hungary and the Balkans). The positive phases of each teleconnection are associated with earlier growing season onsets with the influence of the NAO mostly confined to Western Europe. Five of six regions considered have trends towards an earlier growing season onset greater than 2 days/decade with only the Balkans showing a weak and non-statistically significant trend. The trend in the JFM NAO component dominates the GBI trend and the EA trends are more important for the five continental regions. Further details are outlined in Section 6.

3) Atmospheric Blocks tend to be the main driver for extremely cold winter events throughout most of Europe, while ridges relate to well above average temperature conditions and mild winters. During summer, both patterns related to warm events but in Mediterranean areas, ridges are the pattern responsible for extreme heat events; Rainfall in Southern Europe is strongly affected by these patterns, as strong ridges lead to extremely dry conditions, with almost complete absence of rainfall for prolonged periods. While blocks tend to cause drier conditions in Central/Northern Europe, they are associated with increased instability and increased frequency of extreme precipitation events in Southern Europe. More details are in Section 4.

4) Correlation between four primary teleconnection indices (NAO, EA/WR, SCA and EA) and the hub height wind speeds, Capacity Factor (CF) and Wind Power Density (WPD) in northern Europe. Tall tower series confirm the correlation patterns depicted by the ERA5-derived parameters of interest. Positive phases of NAO and EA/WR increase winds, CF and WPD in northern Europe, and so does the wind power production. Conversely, SCA and EA reduce the wind power production in the previously mentioned regions, so we can expect less production. Accurate seasonal forecasts of the atmospheirc indices will allow anticipating periods of higher-than-normal or below-than-normal wind energy production, and thus, scheduling activities accordingly. For more details, see Section 3 and D6.3.

5) Based on a case study focusing on southern Italy, it is found that links between impacts and teleconnection patterns are not stationary over time. From 1951-1980 to 1981-2010 there is a clear change in the correlation between the various teleconnection indices and the regional precipitation values (e.g. the impact of the Mediterranean Oscillation on the East coast precipitation). Details are outlined in Section 5.

6) Statistical and physically-based models have been developed (Sections 2-6) with a particular advance in the prediction of sectoral indices from atmospheric indices through the implementation of lagged statistical models. The *SARIMAX* model is found to produce the best predictive capacity, with the *Random Forest* and *K-nearest Neighbours* models also performing skilfully while the Linear model is found to be deficient. Recommendations are made to build an operational prediction system for indices that performed well as described in detail in Section 7.

References

Amit, Y., & Geman, D. (1997). Shape Quantization and Recognition with Randomized Trees. Neural Computation, 9(7), 1545–1588. https://doi.org/10.1162/neco.1997.9.7.1545

Barnston, A.G.; Livezey, R.E.: Classifications, Seasonality, and Persistence of Low-Frequency Atmospheric Circulation Patterns, Monthly Weather Review, 115, 1083-1126, 1987

Barriopedro D, García-Herrera R, Lupo AR, Hernández E (2006) A climatology of northern hemisphere blocking. J Clim 19:1042–1063. doi:10.1175/JCLI3678.1

Bedia, J., Herrera, S., Gutiérrez, J.M., Zavala, G., Urbieta, I.R., Moreno, J.M., 2012. Sensitivity of fire weather index to different reanalysis products in the Iberian Peninsula. Nat. Hazards Earth Syst. Sci. 12, 699–708. https://doi.org/10.5194/nhess-12-699-2012

Bhend, J. and von Storch, S. (2008) Consistency of observed winter precipitation trends in northern Europe with regional climate change projections. Climate Dynamics, 31, 17-28

Bindi,M. and Olesen,J.E. (2011) The responses of agriculture in Europe to climate change. Regional Environmental Change, 11, 151-158

Brands, S., Herrera, S., and Gutierrez, J.: Is Eurasian snow cover in October a reliable statistical predictor for the wintertime climate on the Iberian Peninsula?, Int. J. Climatol.,doi:10.1002/joc.3788, 2013.

Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. https://doi.org/10.1007/bf00058655

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brunetti M.; Caloiero T.; Coscarelli R.; Gullà G.; Nanni T.; Simolo C.: Applicazione di test di omogeneità e tecniche di ricostruzione dei dati mancanti alle serie pluviometriche giornaliere calabresi per l'individuazione di aree omogenee e la stima dei trend, L'Acqua, 2, 35-44, 2012 (in Italian)

Carril, A. F., Gualdi, S., Cherchi, A., and Navarra, A.: Heatwaves in Europe: areas of homogeneous variability and links with theregional to large-scale atmospheric and SSTs anomalies, Clim.Dynam., 30, 77–98, 2008.

Casanueva, A., Rodríguez-Puebla, C., Frías, M. D., and González-Reviriego, N.: Variability of extreme precipitation over Europe and its relationships with teleconnection patterns, Hydrol. Earth Syst. Sci., 18, 709-725, https://doi.org/10.5194/hess-18-709-2014, 2014.

Ceglar,A., Zampieri,M., Toreti,A. and Dentener,F. (2019) Observed Northward Migration of Agro-Climate Zones in Europe Will Further Accelerate Under Climate Change. Earth's Future, 7, 1088-1101

Climate Prediction Center. (n.d.). Oceanic & Atmospheric Climate Data.

Clime Indices of INDECIS and their links to sectorial data. (2020). This report arises from the Project INDECIS which is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR), with co- funding by the European Union's Horizon 2020 research and. 72.

Coll, J.R., van der Schrier, G., Aguilar, E., Rasol, D., Coscarelli, R. and Bishop, A. (2019) Data rescue of daily climate station-based observations across Europe. Earth System Science Data, 1-30.

Conte, M.; Giuffrida, A.; Tedesco, S.: The Mediterranean Oscillation. Impact on precipitation and hydrology in Italy Climate Water, Publications of the Academy of Finland, Helsinki, 1989

Cornes, R.C., van der Schrier, G. and Squintu, A.A. (2018) A reappraisal of the thermal growing season length across Europe. International Journal of Climatology, 39, 1787-1795

Domínguez-Castro,F., Reig,F., Vicente-Serrano,S.M., Aguilar,E., Peña-Angulo,D., Noguera,I., Revuelto,J., van der Schrier,G. and El Kenawy,A.M. (2020) A multidecadal assessment of climate indices over Europe. Scientific Data, 7, 125

Dunkeloh, A.; Jacobeit, J.: Circulation dynamics of Mediterranean precipitation variability 1948-98, International Journal of Climatology, 23, 1843-1866, 2003

Fraedrich, K.: An ENSO impact on Europe? A review, Tellus, 46A, 541-552, 1994

Frias, M. D., Herrera, S., Cofino, A. S., and Gutierrez, J. M.: Assessing the Skill of Precipitation and Temperature Seasonal Forecasts in Spain: Windows of Opportunity Related to ENSO Events, J.Climate, 23, 209–220, 2010.

Glantz, M. (ed): Usable Science: Food Security, Early Warning, and El Niño, Proceedings of the Workshop on ENSO/FEWS, Budapest, Hungary, October 1993, UNEP, Nairobi; and NCAR, Boulder, Colorado, 1994

Hamed, K.H., Rao, A.R., 1998. A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology 204, 182–196.

Haylock MR, Hofstra N, Klein Tank AMJ, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded dataset of surface temperature and precipitation. J Geophys Res 113:D20119. doi:10.1029/2008JD10201

Hoogenboom, G. (2000) Contribution of agrometeorology to the simulation of crop production and its applications. Agricultural and Forest Meteorology, 103, 137-157

Iles, C. and Hegerl, G. (2017) Role of the North Atlantic Oscillation in decadal temperature trends. Environmental Research Letters, 12, 114010

Irannezhad,M. and Klove,B. (2015) Do atmospheric teleconnection patterns explain variations and trends in thermal growing season parameters in Finland? International Journal of Climatology, 35, 4619-4630

Iturbide, M., J. Bedia, S. Herrera, J. Baño-Medina, J. Fernández, M.D. Frías, R. Manzanas, D. San-Martín, E. Cimadevilla, A.S. Cofiño, J.M. Gutiérrez, The R-based climate4R open framework for reproducible climate data access and post-processing, Environmental Modelling & Software, 111, 42-54, https://doi.org/10.1016/j.envsoft.2018.09.009, 2019.

JPI Climate. (2017). Indecis Project.

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77(3): 437–471. doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Klok,E.J. and Klein Tank,A.M.G. (2009) Updated and extended European dataset of daily climate observations. International Journal of Climatology, 29, 1182-1191

Klein Tank,A.M.G., Zwiers,F.W. and Zhang,X. 2009. Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation, climate data and monitoring WCDMP-No 72, WMO-TD No 1500, p 5

Kutiel, H.; Maheras, P.; Guika, S.: Circulation indices over the Mediterranean and Europe and their relationship with rainfall conditions across the Mediterranean, Theoretical and Applied Climatology, 54, 125-138, 1996;

Lamb, P. and Peppler, R.: The North Atlantic Oscillation: concept and an application, Bull. Amer. Meteor. Soc., 68, 1218-1225, 1987

Liu,Q., Piao,S., Janssens,I.A., Fu,Y., Peng,S., Lian,X., Ciais,P., Myneni,R.B., Peñuelas,J. and Wang,T. (2018) Extension of the growing season increases vegetation exposure to frost. Nature Communications, 9, 426

Lledó, Ll., I. Cionni, V. Torralba, P.A. Bretonnière, M. Samsó, 2020: Seasonal prediction of Euro-Atlantic teleconnections from multiple systems. Environmental Research Letters, https://doi.org/10.1088/1748-9326/ab87d2

Maheras, P.; Xoplaki, E.; Kutiel, H.: Wet and dry monthly anomalies across the Mediterranean basin and their relationship with circulation, 1860-1990, Theoretical and Applied Climatology, 64, 189-199, 1999

Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 245–259.

Mariotti, A.; Zeng, N.; Lau, K.-M.: Euro-Mediterranean rainfall and ENSO - a seasonally varying relationship, Geophysical Research Letters, 29(12), doi:10.1029/2001GL014248, 2002.

NCAR, T. N. C. for A. R. (2020). Nino SST Indices | NCAR - Climate Data Guide.

Palutikof, J. P.; Conte, M.; Casimiro Mendes, J.; Goodess, C. M.; Espirito Santo, F.: Climate and climate change, in Mediterranean desertification and land use, John Wiley and Sons, London, 1996

Ramon, J., Ll. Lledó, N. Pérez-Zanón, A. Soret, and F. J. Doblas-Reyes, 2020: The Tall Tower Dataset. A unique initiative to boost wind energy research. Earth Syst. Sci. Data, 12, 429–439, https://doi.org/10.5194/essd-12-429-2020.

Rencher, A. C. (2005). A Review Of "Methods of Multivariate Analysis, Second Edition." In IIE Transactions (Vol. 37, Issue 11). https://doi.org/10.1080/07408170500232784

Rocha, A.: Low-frequency variability of seasonal rainfall over theIberian Peninsula and ENSO, Int. J. Climatol., 19, 889–901,1999.

Rodríguez-Puebla, C., Ayuso, S., Frias, M., and Garcia-Casado, L.:Effects of climate variation on winter cereal production in Spain,Clim. Res., 34, 223–232, 2007.

Ropelewski, C. F.; Halpert, M. S.: Quantifying Southern Oscillation-precipitation relationships, J. Climate, 9, 1043-1059, 1996.

Sen P. K.: Estimates of regression coefficient based on Kendalls Tau, J. Am. Stat. Assoc., 63, 1379–1389, doi:10.2307/2285891,1968.

Sheng, Y., Wang, C., 2004. The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. Water Resour. Manag. 201–218.

Silverman, B. W., & Jones, M. C. (1989). E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). International Statistical Review / Revue Internationale de Statistique, 57(3), 233. https://doi.org/10.2307/1403796

S2S4E, 2019: Impact and assessment of weather regimes on the energy sector. 1st version of deliverable D6.2 of the HORIZON 2020 project S2S4E. EC Grant agreement no.: 776787.

Santos JA, Pinto JG, Ulbrich U (2009) On the development of strong ridge episodes over the eastern North Atlantic. Geophysical Research Letters 36: L17804. doi:10.1029/2009GL039086

Shaman, J.; Tziperman, E.: An Atmospheric Teleconnection Linking ENSO and Southwestern European Precipitation, Journal of Climate, 24, 124-139, 2011;

Shaman, J.: The Seasonal Effects of ENSO on European Precipitation: Observational Analysis, Journal of Climate, 27, 6423-6438, 2014

Scheifinger,H., Menzel,A., Koch,E., Peter,C. and Ahas,R. (2002) Atmospheric mechanisms governing the spatial and temporal variability of phenological phases in central Europe. International Journal of Climatology, 22, 1739-1755

Smith and Barstad, A Linear Theory of Orographic Precipitation, Jorunal of the Atmospheric Sciences, 61, 1377-1391, 2004

Sousa, P.M., Trigo, R.M., Aizpurua, P., Nieto, R., Gimeno, L., García-Herrera, R., 2011. Trends and extremes of drought indices throughout the 20th century in the Mediterranean. Natural Hazards and Earth System Sciences 33–51.

Sousa PM, Trigo RM, Barriopedro D, Soares PMM, Ramos AM, Liberato MLR (2017a) Responses of European precipitation distributions and regimes to different blocking locations. Clim. Dyn. doi: 10.1007/s00382-016-3132-5

Sousa PM, Trigo RM, Barriopedro D, Soares PMM, Santos JA (2017b) European temperature responses to blocking and ridge regional patterns. Climate Dynamics. doi: 10.1007/s00382-017-3620-2

Tibaldi S and Molteni F (1990) On the operational predictability of blocking. Tellus 42A: 343–365, doi:10.1034/j.1600-0870.1990.t01-2-00003.x

Tomcyzk,A.M., Szyga-Pluta,K. and Bednorz,E. (2019) The effect of macro-scale circulation types on the length of the growing season in Poland. Meteorology and Atmospheric Physics, 131, 1315-1325

Trigo, R. M.; D. Pozo-Vazquez; T. J. Osborn; Y. Castro-Diez; S. Gamiz-Fortis; M. J. Esteban-Parra: North Atlantic Oscillation influence on precipitation, river flow and water resources in the Iberian Peninsula, Int. J. Climatol., 24, 925–944, 2004

Trigo RM, Pereira JMC, Pereira MG, Mota B, Calado MT, DaCamara CC, Santo FE (2006) Atmospheric conditions associated with the exceptional fire season of 2003 in Portugal. International Journal of Climatology 26 (13): 1741-1757. doi: 10.1002/joc.1333

Vicente-Serrano, S., Beguería, S., López-Moreno, J., El Kenawy,A., and Angulo, M.: Daily atmospheric circulation events and extreme precipitation risk in Northeast Spain: the role of the North Atlantic Oscillation, Western Mediterranean Oscillation, and Mediterranean Oscillation, J. Geophys. Res.-Atmos., 114,D08106, doi:10.1029/2008JD011492, 2009.

Xoplaki, E.; Gonzalez-Rouco, J.; Luterbacher, J.; Wanner, H.: Wet season mediterranean precipitation variability: inuence of large-scale dynamics, Climate Dynamics, 23, 63-78, 2004.