S2S4E

Climate Services for Clean Energy

Research and Innovation action

H2020-SC5-2017

Impact and assessment of weather regimes on the energy sector

Deliverable D4.2

Version N°1

Authors: David Brayshaw (UREAD), Hannah Bloomfield (UREAD), Irene Cionni (ENEA), Nicola Cortesi (BSC), Llorenç Lledó (BSC), Andrea Manrique-Suñén (BSC), Nathalie Schaller (CICERO), Verónica Torralba (BSC), Wei Yang (SMHI)

This project received funding from the Horizon 2020 programme under the grant agreement n°776787.

Disclaimer

Document Information

Grant Agreement	776787
Project Title	Subseasonal to Seasonal climate forecasting for Energy
Project Acronym	S2S4E
Project Start Date	01/12/2017
Related work package	WP 4. S2S Climate predictions
Related task(s)	Task 4.3: Predictability of energy-relevant climate variability Task 4.4: Conditional predictability and weather regimes
Lead Organisation	UREAD
Submission date	30/11/2019
Dissemination Level	PU

History

Date	Submitted by	Reviewed by	Version (Notes)
26/11/19	David Brayshaw	Cross-reviewed by author team	Version 1

Table of content

List of figures

List of tables

Summary

Climate information on subseasonal (up to 6 weeks ahead) to seasonal (up to 7 months ahead) timescales is needed for decision-making in a number of sectors, yet skill – and consequently user uptake – of forecasts in Europe has been limited. The present deliverable seeks to explore the potential for "pattern based" forecasting using a variety of different techniques applied to a range of leading numerical weather prediction (NWP) systems at subseasonal and seasonal time-horizons. The central concept of pattern-based forecasting is to use the NWP system to predict the large-scale atmospheric conditions, while utilizing statistics based on observational records to link the circulation to surface climate impacts (and hence energy impacts). The analysis presented here seeks to provide two key insights into the capabilities of subseasonal and seasonal prediction systems: assessing their ability to (a) represent and predict key patterns of large-scale atmospheric circulation in the Euro-Atlantic sector and (b) to faithfully capture the observed surface impacts of the large-scale circulation.

It is demonstrated that subseasonal and seasonal NWP systems offer modest skill in predicting the large-scale circulation (particularly in week 1-2 and month 1, decaying thereafter) though this depends strongly on the circulation typing scheme, season, and detailed methodology. It is also demonstrated that the link between the large-scale circulation and surface impacts appears to be imperfect and degrades with lead time. Both findings provide supporting evidence for the potential value in pattern-based forecasting for energy applications (as opposed to "grid point" surface meteorological data from the NWP models). Such techniques are developed and evaluated subsequently in Deliverable D4.3)

Keywords

Climate prediction; Forecasting; Subseasonal prediction; Seasonal prediction; Energy; Electricity; Renewables; Power; Wind; Hydro; Solar; Demand; Weather patterns; Weather regimes; Teleconnection patterns

Glossary

EATC	Euro Atlantic Teleconnections. A set of patterns and corresponding indices describing the large-scale atmospheric circulation over the Euro-Atlantic region, usually applied at monthly/seasonal timescales.
ECV	Essential Climate Variable. Examples include surface temperature, precipitation, near-surface wind speeds.
Forecast month *n*	Validity time for a given seasonal prediction. Month 1 corresponds to the first full calendar month after the last member of the ensemble forecast was launched. Full discussion is provided in Deliverable D4.1 but, for illustration, month 1 corresponds to December for both of the following examples: E.g. 1, a forecast containing a set of ensemble members launched in the window Nov 1st – Nov 30th. E.g. 2, a forecast containing a set of ensemble members, all of which are launched on Nov 1st.
Forecast week *n*	Validity time for a given subseasonal prediction. Week 1 corresponds to an aggregation of days 5-11, etc. Full discussion is provided in Deliverable D4.1.
HWR	Hydrological Weather Regime. A set of tailored patterns describing the atmospheric circulation linked to local precipitation.
NWP	Numerical Weather Prediction (usually referring to the process of producing forecasts using gridded NWP models). In the present context, NWP models include atmospheric models which are coupled to representations of additional Earth system

	components evolving on timescales relevant to subseasonal and seasonal forecasting (e.g., sea ice, oceans).
Reanalysis	A 3-D gridded reconstruction of the atmosphere spanning a few or several decades combining model simulations and observations.
S2S	Subseasonal to seasonal
Seasonal forecast	A meteorological forecast targeting a lead-time of a few to several months. In the present context, this usually refers to a forecast generated by an atmospheric NWP model which has been coupled to appropriate models of Earth system components evolving on relevant timescales (e.g., sea ice, oceans).
Subseasonal forecast	A meteorological forecast targeting a lead-time of a few to several weeks. In the present context, this usually refers to a forecast generated by an atmospheric NWP model which has been coupled to appropriate models of Earth system components evolving on relevant timescales (e.g., sea ice, oceans).
TCT	Targeted Circulation Type (formerly "Impact Pattern" in D3.2). A set of tailored patterns describing the atmospheric circulation linked to nationally-aggregated energy indicators (e.g., demand, wind power) over Europe.
WR	Weather Regime. A set of patterns describing the atmospheric circulation over the Euro-Atlantic region, usually applied at daily/weekly timescales and typically following the general method of Cassou (2008).

1 Introduction

Climate information on subseasonal (up to 6 weeks ahead) to seasonal (up to 7 months ahead) timescales is needed for decision-making in a number of sectors. Compared to short to medium-range (up to 10 days ahead), subseasonal to seasonal (S2S) time-scales hold potential value for a wide range of users who are affected by variability in climate, water and energy and who would benefit from understanding and better managing climate-related risks (Bruno Soares et al., 2017; Stoft, 2002; Green, 2005). Wind, solar, hydro and energy demand are examples of renewable energy applications in which S2S information can affect decision making.

In Europe, there has been relatively little uptake and use of S2S forecasts by users for decision making, compared to other parts of the world, such as the USA and Australia, possibly associated with the relatively limited inherent predictability and limited quality of forecasts (Bennett et al., 2017; Mendoza et al., 2017; Arnal et al., 2018). However, recent advances in our understanding and forecasting of climate have begun to result in somewhat skillful predictions, which can consequently lead to improvements in awareness, preparedness and decision-making from a user perspective (Bruno Soares and Dessai, 2016).

A first assessment of the forecast skill achievable in the European Centre for Medium Range Forecasting (ECMWF) subseasonal and seasonal forecast systems applied to energy was provided in Deliverable D4.1. That deliverable focused on evaluating skill in directly forecasting surface meteorological variables (e.g., wind, temperature, precipitation, insolation; often referred to as Essential Climate Variables or ECVs) and their subsequent conversion into energy-relevant quantities (wind power, demand, hydrology, solar power). In general, skill was shown to exist over some regions of Europe - but at rather modest levels - for multi-week (subseasonal) and multi-month (seasonal) lead times. In this document, such forecasts of surface climate or energy impacts are referred to as "grid point" forecasts.

The present deliverable seeks to explore a different angle concerning the predictive capabilities of numerical weather prediction (NWP) systems at subseasonal and seasonal time-horizons. Rather than predicting gridded surface variables and converting these into estimates of energy impacts, the ability of the models to predict the large-scale circulation of the atmosphere in the Euro-Atlantic sector is instead interrogated. The rationale for doing so is that (a) atmospheric circulation patterns are known to have a strong influence on surface weather conditions and consequent impacts on the energy sector (see, e.g., Deliverable D3.2); and, (b) it is widely believed that NWP forecast skill at subseasonal and seasonal timescales lies in slowly-evolving aspects of the large-scale atmospheric circulation rather than the specific details of local (i.e., grid-point) atmospheric conditions. The ability to skillfully forecast the large-scale circulation therefore appears to open opportunities for hybrid forecasts, whereby the NWP system is used to forecast the large-scale circulation while statistical techniques are then used to link the circulation state to surface climate and energy impacts. A classic (and simple) example of this can be found in, e.g., seeking to assess or predict the state of the North Atlantic Oscillation pattern, and linking that state forecast to an

energy impact using observed historic data (e.g., Brayshaw et al.2011; Ely et al., 2013; Thornton et al.2019). In this document, such hybrid methods are referred to as "pattern-based" forecasts.

The aim of the present document is therefore to provide an assessment of the ability of subseasonal and seasonal forecast systems to simulate large-scale drivers of Euro-Atlantic climate, along with their impact on European surface climate and consequent energy system impacts. Conceptually, this may be viewed as breaking the assessment of NWP forecast skill into two parts: the ability to correctly represent and predict particular states of large-scale circulation, and the ability to correctly associate circulation states to surface climate/energy impacts. It therefore builds upon previous deliverables in terms of technique and analysis (particularly Deliverables D3.1, D3.2 and D4.1). It should also be read alongside Deliverable D4.3 which builds upon this deliverable to compare the overall performance of "pattern based" forecast schemes (for surface meteorological variables and their energy impacts) versus other forecasting approaches (including, amongst others, grid-point based methods).

In the discussion that follows, several different and complementary approaches to pattern-based forecasting are tested and reported. A common language and methodology has been followed as much as possible throughout this document but, in order to facilitate rapid scientific exploration of this research topic, the nature of exploratory research, the individual chapters should be viewed as a set of parallel investigations, each adopting its own specific methodological innovations.

The ability of seasonal and subseasonal forecasts to predict circulation patterns using traditional meteorological classification techniques (Euro-Atlantic Teleconnections, EATCs; Weather Regimes, WRs) are first investigated in Chapters 3 and 4 respectively. This is followed by an analysis of forecasts using more tailored pattern classification schemes specifically targeting energy applications (Targeted Circulation Types, TCTs; Hydrological Weather Regimes, HWR) in Chapters 5 and 6 respectively. Chapter 7 explores the links between a specific large-scale remote climate driver, Stratospheric Sudden Warmings (SSWs), known to drive European surface climate and energy impacts. Chapter 8 returns to seasonal forecasts of EATCs, exploring the robustness of their predictability and structure against a long baseline of 20th Century climate. As noted above, each chapter operates within a common analysis framework (which has evolved from Deliverables D3.2 and D4.1) which is outlined briefly in Chapter 2 (e.g., forecast and observational datasets, pattern identification methods), though important methodological distinctions are reported in detail within the Method section of each individual Chapter. A concluding discussion is provided in Chapter 9.

2 Datasets and Methods

2.1 Introduction

Many of the datasets and tools used in this document follow closely the methods developed in earlier deliverables (particularly D3.1, D3.2 and D4.1) and in the "partner" deliverable D4.3. As these datasets and tools are central to the science that follows, the following sections seek to provide a high-level overview of the datasets and methods involved. More comprehensive discussion of each dataset/tool can be found in previous documentation (references provided).

Consistent with the research objectives of this deliverable, however, it is noted that different research activities (generally corresponding to the individual chapters) have introduced a range of experimental innovations to the basic techniques in order to advance understanding and/or improve predictive skill. As such, the detailed implementation of each dataset or tool for a particular research task is provided separately within each of Chapters 3 to 8.

2.2 Meteorological datasets and forecasting systems

2.2.1 Reanalysis data

Reanalysis products have appeared as an efficient alternative to in-situ observations to investigate the past atmospheric conditions, both for monitoring and research purposes (Gregow et al., 2016; Compo et al., 2011; Dee et al., 2011). These global datasets are the result of combining a state-of-the-art numerical model with the assimilation of past observations from several sources to recreate the state of the atmosphere in a gridded three-dimensional mesh (Fujiwara et al., 2017). Many different reanalysis products exist and a full description and intercomparison of their properties is provided in D3.1 (see also Ramon et al., 2019). Based on this evaluation, in this report, four reanalyses products are selected for use. The first two, ERA-Interim and ERA5, are modern-era reanalyses constrained by a full suite of observational data, but the third, ERA-20C is a reconstruction of a longer period using a consistent (but restricted) set of observations. Each dataset is described briefly below.

ERA-Interim (Dee et al., 2011) has become a very used dataset for the energy sector (Gregow et al. 2016; Bett and Thornton 2016). The dataset covers the 1979-2017 period (and has latterly been continued to 2019), and the temporal resolution is either 3 h (forecast) or 6 h (analysis), depending on the variable (see Dee et al., 2011, for details). The spatial resolution of the data set is 0.75 ° (approximately 80 km) on 60 vertical levels from the surface up to 0.1 hPa. ERA-Interim's data assimilation includes near-surface air temperature, pressure and relative humidity, upper-air temperature, wind, specific humidity and rain-affected SSM/I radiances.

ERA5 is the new climate reanalysis dataset from ECMWF, intended to improve on the earlier ERA-Interim dataset. It features a spatial resolution of ~31 km and 137 vertical levels, high time frequency output for surface fields (typically hourly), uses a newer version of the ECMWF

IFS numerical model (Cycle 41r2), and assimilates a full range of available observation data (enhanced from ERA-Interim). The dataset is intended to cover 1950 to near real time though, in general, more limited periods are used in the present analysis (e.g., only the period 1980-2018 was available when the analysis was conducted and more restricted periods are used when comparing against, e.g., subseasonal hindcast datasets) . ERA5 includes a 10-member ensemble (seeking to represent observational uncertainty), though here the deterministic high-resolution version is used.

ERA-20C is the 20th century reanalysis from ECMWF, covering the period 1900 to 2010 and assimilating surface observations only (Poli et al. 2013, 2015). It was produced using the IFS model (Cycle 38R1), with horizontal resolution of about 125km 91 vertical model levels and 3h output. The main target of ERA-20C was to demonstrate the feasibility of reanalyzing the whole 20th century with new assimilation methods when you have varying observing systems, rather than provide a "best-product" dataset, such as ERA-Interim or ERA5.

2.2.2 Sudden Stratospheric Warmings database

Sudden stratospheric warmings are defined as times when there is a large and rapid temperature increase in the winter polar stratosphere, associated with a reversal of the climatological westerly winds (Butler et al., 2017). These extreme events can have substantial impacts on winter surface climate, including increased frequency of cold air outbreaks over Europe, hence the investigation into their predictability conducted here. A database of these events has been compiled by Butler et al. (2017, available at: https://www.esrl.noaa.gov/csd/groups/csd8/sswcompendium/majorevents.html), defined using an index based on the ERA-interim re-analysis (equivalent dates for the ERA5 re-analysis are not yet available). This results in 23 dates in the period (1980-2018) and12 dates in the common hindcast period for the ECMWF and NCEP models (1999-2010).

2.2.3 Seasonal forecast systems

Several European national meteorological centres and institutions produce operational seasonal predictions. Seven different seasonal prediction systems have been employed in this report, from the European Center for Medium-Range Weather Forecasts (ECMWF), Deutscher Wetterdienst (DWD), Meteo France (MF), UK Met Office (UKMO) and Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC). Many of those predictions can be obtained from the Climate Data Store (CDS) of the Copernicus Climate Change Service (C3S) initiative, which provides a unified access point, and a common hindcast period and spatial resolution. Other predictions have been obtained from ECMWF's Meteorological Archival and Retrieval System (MARS).. Some details of each of the prediction systems employed here, as the number of ensemble members, the hindcast period or the spatial grid are detailed in Table 1. Notice that not all the datasets have been employed for all the different analyses available in this report.

Center	*Prediction system*	*Data source*	*Analyzed period*	*Ensemble members*	*Horizontal grid*
CMCC	SPS3	CDS	1993-2018	40	Regular 360x180
DWD	System2	CDS	1993-2018	30	Regular 360x180
UKMO	GLOSEA5 System13	CDS	1993-2018	28	Regular 360x180
MF	System6	CDS	1993-2018	25	Regular 360x180
ECMWF	SEAS5	CDS	1993-2018	25	Regular 360x180
ECMWF	SEAS5	MARS	1981-2018	51	Regular Gaussian F160 (640x320)
ECMWF	SEAS5	MARS	1981-2016	25 (re-forecasts)	Regular Gaussian F160 (640x320)
ECMWF	System4	MARS	1982-2016	15	Regular Gaussian F128 (512x256)

Table 1: Technical details of the seasonal prediction systems

In addition to the operational forecast systems, to investigate the long-term variability of atmospheric teleconnections and their predictability, we use the ERA-20C reanalysis (see above) and the seasonal hindcasts from the ASF-20C dataset (Weisheimer et al., 2017). Both ERA-20C and ASF-20C use an atmosphere-only model, and the simulations of the ASF-20C dataset were designed to match the set up of the ERA-20C reanalysis. ASF-20C consists of 51 ensemble members, similar to the operational seasonal forecasts produced by ECMWF. There are 4 start dates, November/February/May/August, and the forecast period is 4 months. Both ERA-20C and ASF-20C span the period 1900-2010.

2.2.4 Subseasonal forecast systems

In the subseasonal time range, two systems from the S2S database (Vitart et al. 2017) have been employed: ECMWF monthly forecast system (MFS or extended range) and NCEP CFSv2. ECMWF-MFS (Vitart 2004) runs coupled ocean-atmosphere integrations up to 46 days issued every Monday and Thursday. Operational configuration consists of 51 ensemble members while the hindcasts consist of 11 members. It has been described in Deliverable 4.1 (Section 3.1.2) and its skill for surface variables over Europe was assessed in Deliverable 4.1 (Section 5).

National Centers for Environmental Prediction's (NCEP) Climate Forecast System (Saha et al. 2014) is a coupled system to both an ocean model (GFDL MOM4) and an ice model. The forecast length for subseasonal predictions is 45 days and the system is run every 6 hours.

The real-time forecast runs three perturbed members and one control run initialised four times a day (00 UTC, 06 UTZ, 12 UTC and 18 UTC). The hindcast period is fixed and spans 12 years (1999-2010). The hindcast is also initialised daily, four times a day, but only one simulation at the time, producing a lagged ensemble of 4 members daily.

2.2.5 Forecast horizons

In terms of the temporal scale analysis, forecast skill is evaluated using the following lead-time convention. For seasonal forecasts, month 1 corresponds to one month after initialization (so for an ensemble of forecasts launched at any point in November, forecast month 1 is December). For subseasonal forecasts, week 1 is defined as the week starting at day 5 (i.e., forecast week 1 is the period day 5 to 11, week 2 is days 12-18 etc). A complication occurs in the case of lagged ensemble forecasts, which may include ensemble members launched earlier (in which case day 5 is defined relative to the most recently launched ensemble member).

2.3 Pattern identification methods

2.3.1 Euro-Atlantic Teleconnection Patterns

European climate variability is often analyzed through the role of atmospheric teleconnections. The rationale behind this is to find a set of fixed atmospheric circulation patterns and corresponding temporally varying indices that can be used to describe monthly or seasonal circulation anomalies and their surface impacts in a simplified way. A common method to define teleconnections is through Rotated Empirical Orthogonal Function (REOF) analysis (Barnston and Livesy, 1987). This dimensionality reduction technique allows approximating geopotential height anomaly fields as a linear combination of only a few variability modes or spatial patterns:

$$Anom(t,x,y) = \sum_{i=1}^{nmodes} TCI_i(t) \cdot TCP_i(x,y) + Residuals$$

Teleconnection patterns (TCP) and indices (TCI) --i.e. the weights in the linear combination-- are chosen so that their combination minimizes the residual term. Over the Euro-Atlantic region, the first four modes are typically employed and are commonly referred to as the North Atlantic Oscillation (NAO), East Atlantic (EA), East-Atlantic/Western Russia (EAWR) and Scandinavian pattern (SCA) teleconnections respectively:

$$Anom = NAO_i * NAO_p + EA_i * EA_p + EAWR_i * EAWR_p + SCA_i * SCA_p + Residuals$$

The observed patterns and indices for these four Euro-Atlantic Teleconnections (EATC) have been obtained from ERA-Interim reanalysis. Each EATC pattern is associated with a set of surface temperature, wind speed, solar radiation and precipitation impacts, which can, in turn, be associated with particular impacts on the European energy system (see Deliverable D3.2

and also Brayshaw et al., 2011, Cradden et al., 2017, Zubiate et al., 2017). Forecasts of each EATC index are obtained from seasonal prediction systems by projecting predicted anomalies onto the observed patterns (i.e. by making a scalar product between the anomaly fields and the patterns).

2.3.2 Weather Regimes

Weather regimes (WRs) are large scale recurrent and persistent circulation patterns (Vautard 1990; Michelangeli et al., 1995, Cassou et al. 2004). They are defined by algorithms that classify daily atmospheric circulation fields (geopotential height, sea level pressure or wind speed and direction). The methodology employed is the K-means clustering as described in Deliverable 3.2 (Section 3.2.1 therein).

In this Deliverable, the ability of the prediction systems to predict the WRs has been explored. Different WRs classifications have been tested, the different classifications are a result of different variables or different windows over which the clustering is applied (e.g., a separate set of WRs may be defined for each individual week or month in the year, or a single set of WRs may be defined over an entire season or year).

2.3.3 Targeted Circulation Types

Targeted Circulation Types (TCTs; formerly referred to as "Impact Patterns" or IP in D3.2) are constructed analogously to the weather regimes above (i.e., using k-means clustering) but rather than applying the clustering to gridded meteorological data, the input is instead the set of 29 nationally-aggregated daily timeseries of power system balance indicators (e.g., national demand, or demand-net-wind; see D3.2 and Bloomfield et al., in review for a full discussion of the method). In this case, a single clustering is applied to the extended winter period (November-March) from 1980-2018.

2.3.4 Hydrological weather regimes

Hydrological weather regimes (HWR) are large-scale circulation patterns, classified on the concept of fuzzy sets (Zadeh, 1965) which uses imprecise statements to describe a certain system (in this case the climate system). Daily mean sea level pressures or geopotential heights are normally used as predictor to classify daily atmospheric field. By using local observations (e.g., daily precipitation, daily temperature) each fuzzy rule is able to describe a type of "average" variability of local climate in terms of the frequency and magnitude of observed events (normal events and extreme events) via an iteration optimization process. The details can be found in Bárdossy et al.(2002).

Here, an ensemble of daily HWR is generated over whole hindcast period and all members with ECMWF SEAS5 pressure field (see D3.2 section 3.2.2). The HWRs are investigated to

understand whether they provide additional information to improve forecast skill in the hydrological seasonal forecast.

3 Euro-Atlantic teleconnections in seasonal prediction systems

3.1 Predictive ability of forecasting Euro-Atlantic Teleconnection indices

3.1.1 Introduction

The North Atlantic Oscillation (NAO), as well as the other three main Euro-Atlantic Teleconnections (EATCs), namely the East Atlantic (EA), East Atlantic/Western Russia (EAWR) and Scandinavian pattern (SCA), have a strong impact on surface variables that affect the European energy sector (see Deliverable 3.2). Several authors have recently shown that the state of the NAO during winter (DJF) can be skillfully predicted months ahead employing dynamical seasonal prediction systems (Scaife et al. 2014, Dunstone et al. 2016, Johnson et al. 2019). The use of multisystem prediction ensembles can achieve even better results in terms of ensemble mean correlation (Athanasiadis et al. 2017, Baker et al. 2018). Those authors show that the results are sensitive to the NAO index definition, and the highest NAO skill is obtained when using box-based or zonally-averaged indices. However, these NAO indices do not fully characterise surface variability in Europe (e.g. Zubiate et al. 2017), and the impact of other EATCs is relevant in several regions, as was shown in Deliverable 3.2. In this section, we analyse the skill of several seasonal prediction systems in predicting winter NAO, EA, EAWR and SCA teleconnection indices one month ahead, and we extend the analysis to the four seasons.

3.1.2 Computation of EATC patterns and indices

Although the NAO index can be computed in many different ways, (essentially point-based, box-based or EOF-based methods), the other EATCs are not so straightforward to compute. Here, an approach that allows computing NAO, EA, EAWR and SCA indices from both the observations and the forecasts has been developed. The methodology mimics well-known Climate Prediction Center patterns and indices as much as possible (see Deliverable 3.2 for a detailed comparison).

The four observed EATC patterns and indices have been computed from ERA-Interim 500 hPa geopotential height fields employing a Rotated EOF (REOF) analysis over the Euro-Atlantic domain [90ºW-60ºE; 20ºN-80ºN]. First, seasonal (DJF/MAM/JJA/SON) anomalies for the 1981-2018 period have been computed with respect to the same period mean. Then an EOF analysis is performed, and the first four variability modes are retained. After that, a Varimax rotation is applied. The four obtained REOF modes have been reordered and their sign has been adjusted when needed so that the EATC patterns resemble as much as possible the positive phases of the NAO, EA, EAWR and SCAN patterns (in this order) as computed by CPC. Figure 1 shows the patterns obtained for DJF.

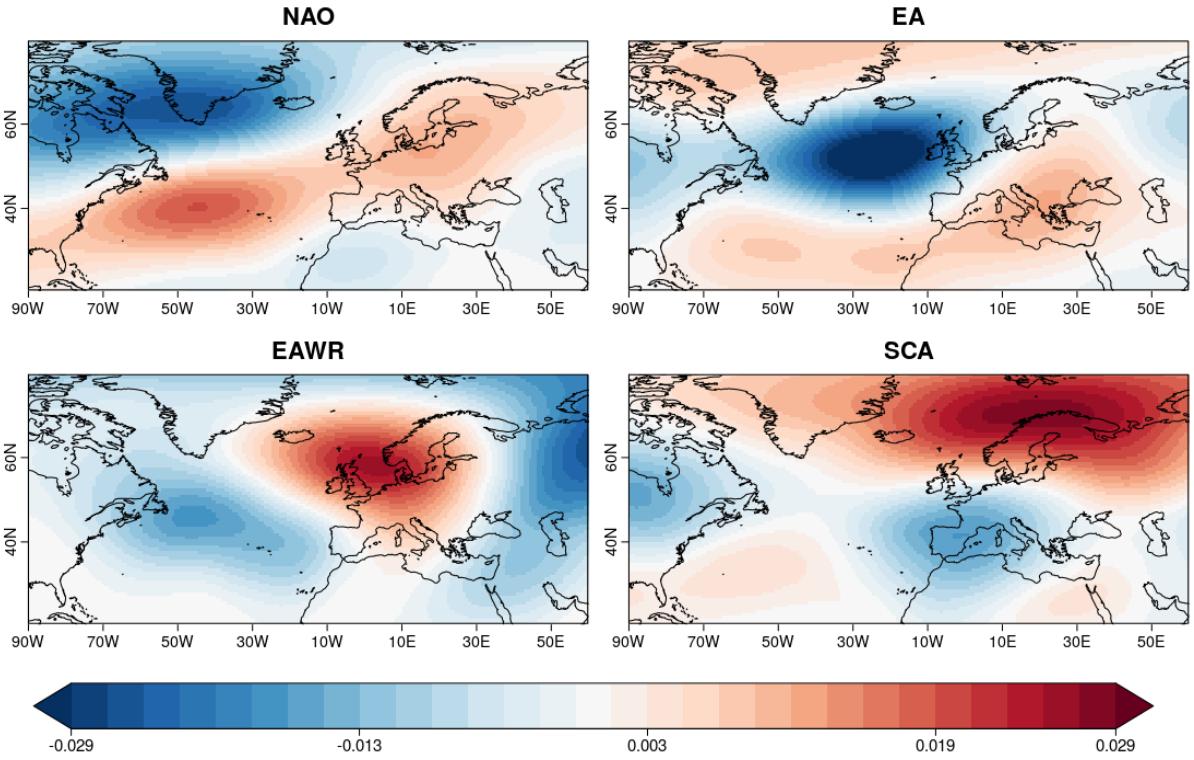

Figure 1. The four observed Euro-Atlantic Teleconnection patterns obtained from REOF analysis of ERA-Interim DJF seasonal mean 500 hPa geopotential height anomalies for the 1981-2018 period.

To obtain (retrospective) forecasts of the EATC indices, the 500 hPa geopotential height fields from several seasonal prediction systems in the 1993-2018 period have been used. The seasonal mean (DJF/MAM/JJA/SON) anomalies of each system (with respect to 1993-2018 climatology) have been projected onto the observed patterns, individually for each ensemble member. The forecasts need to be regrided to match the ERA-Interim grid before the projection. A conservative interpolation method has been used. The projected indices have not been normalised, so that the explained variances of each decomposition can be computed by just doing the scalar product of each projected EATC index by itself. The total variance of each prediction system and also of the reanalysis have also been computed and normalised per number of grid points, years and ensemble members, so that results are comparable amongst different datasets.

In some cases, the results have been compared to a simpler analysis where only NAO forecasts are obtained via an unrotated EOF procedure. This is especially useful in terms of analysing the residual variances that are not explained by the teleconnections. The observed NAO indices computed with EOF and REOF methods have a correlation of 0.94 indicating very good correspondence.

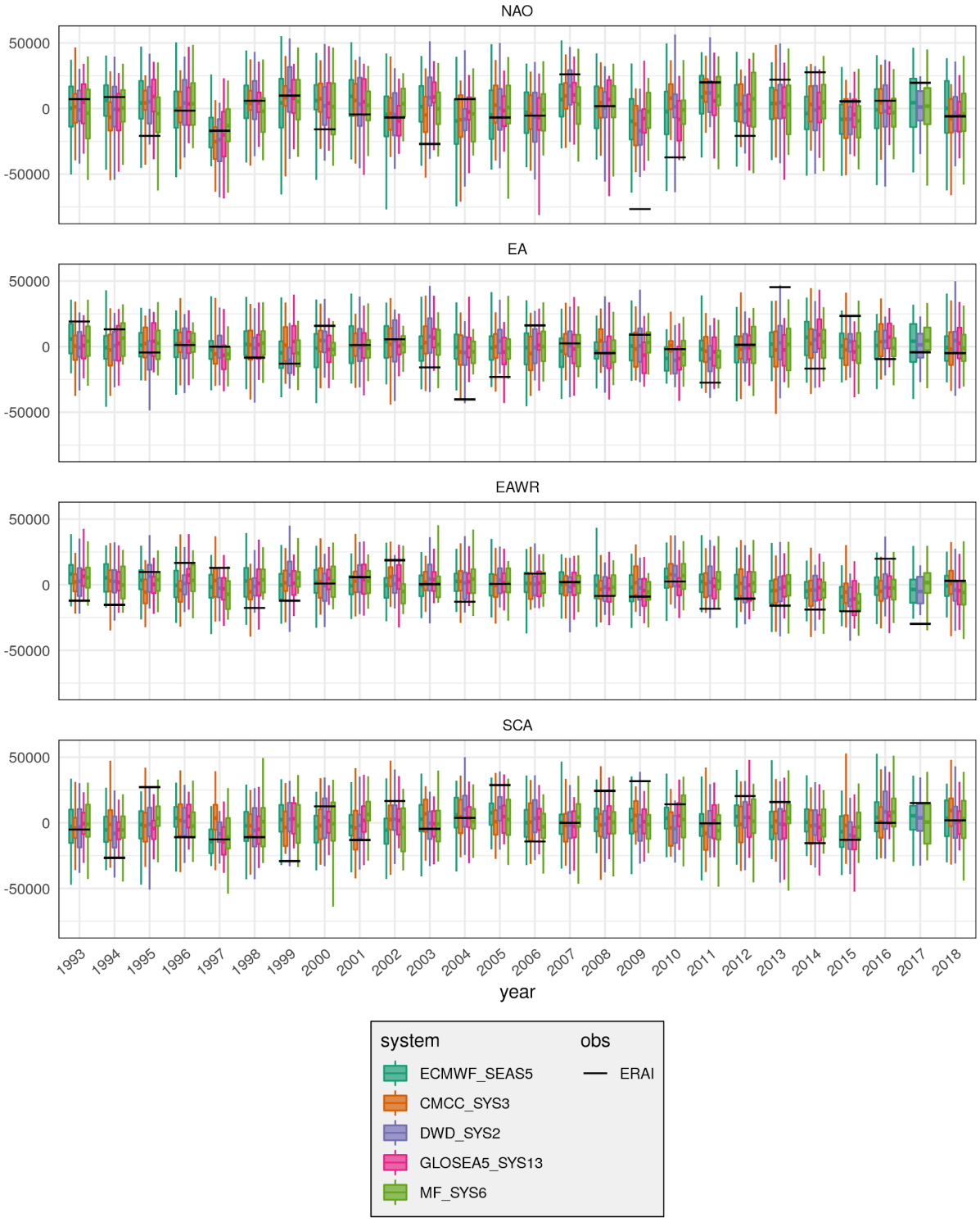

Figure 2. EATC forecasts issued in November and valid for DJF (i.e. 1 month of lead time) for each of the systems, for the 1993-2018 period. Each boxplot shows the maximum and minimum forecast value (whiskers), the first and third quartiles (box) and the ensemble (color line). The black line denotes the observed value according to ERA-Interim. The year labels correspond to the forecast start date (i.e. November). CMCC and GLOSEA5 forecasts for 2017 were not available.

Figure 2 shows the four EATC retrospective forecasts issued in November and valid for DJF (i.e. 1 month of lead time) for each of the systems for the 1993-2018 period. For instance, in the first panel we can see how models anticipated a negative NAO for 2009/2010 winter, although the actual observed value was well below the range of any prediction. Less well-known events such as a strong EA in 2013/2014 or a negative EAWR in 2017/2018 can also be seen. As the forecasted teleconnection indices have not been normalised, a wider range can be seen for NAO than for the other three EATCs.

3.1.3 Variance explained by each EATC in the prediction systems

In order to understand how well the observed teleconnection patterns describe the variability in each of the prediction systems, the explained variance of each EATC pattern as a percentage of the total variance in the system is analysed for the 1993-2018 period. The explained variances have also been computed for ERA-Interim for the same period. Table 2 presents these results for DJF with one month of lead time (i.e., issued in November). The NAO represents a 32% of the variance in the observations (first row), while in the predictions (i.e., after projecting the predicted anomalies) this pattern represents a bit less (ranging from 23 to 29%). EA and SCA explain a 19% in the observations, and the systems also have less variability of these types. For the EAWR the situation is different, and predictions have more EAWR-shaped variability than observations. These differences show that variability has a different structure in the prediction systems, and probably its own internal variability modes do not match the observed ones, i.e. there are biases in location and shape of the modeled variability modes. The same situation is also seen when using an (unrotated) EOF analysis for computing NAO.

Although in the EOF analysis the NAO explains more variance than in the REOF analysis (both in observations and in forecasts), the residual variance that the four REOF patterns do not explain is much lower than the residual variance of the single EOF approach. This means that the REOF analysis with four teleconnections describes much more variability in the observations and predictions than the NAO-only EOF analysis and allows a more comprehensive view. Regarding absolute values of variance (last column), the forecasts show close values to the observations, some slightly above and some others slightly below, i.e. the total amount of variability in the systems seems to be quite correct.

	Percentage of variance (%)							*Total sqrt variance*
	REOF analysis					*EOF analysis*		
	NAO	*EA*	*EAWR*	*SCA*	*res.*	*NAO*	*res.*	
ERA-Interim	32	19	9	19	21	39	61	304
ECMWF_SEAS5	29	14	11	14	32	30	70	300
CMCC_SYS3	23	15	11	18	32	29	71	292
DWD_SYS2	27	14	11	16	33	31	69	317
GLOSEA5_SYS12	26	15	11	14	34	28	72	292
MF_SYS6	26	12	10	18	34	33	67	310

Table 2. Total variance, explained variance percentages and residual variance percentage for each prediction system, both for the REOF and the EOF analysis.

3.1.4 Skill assessment of EATC forecasts in winter

The verification metric that is typically employed for NAO predictions, the ensemble mean correlation, is used here for the four EATCs. The process of averaging EATC predictions from individual members produces the effect of cancelling the noise present in each single realisation and allows extracting any forcing signal that is common to the different members. Table 3 presents the results again for DJF and one month of lead time. We see that all the EATCs can be predicted by some of the systems with statistically significant correlations. EA has the weakest (or even negative) correlations overall, although MF SYS6 produces a statistically significant correlation of 0.41. On the other hand, the SCA forecast quality is as good as that of the NAO. The best system in predicting each EATC is displayed in bold in the table. There is not a single system that has the best performance for all the EATCs, although DWD System2 seems to be the more consistent across the different EATCs. The simpler EOF analysis did not produce systematically better results than the more complex REOF in terms of NAO correlations.

Multisystem combinations have also been produced by gathering all the available ensemble members together (pooling), which gives more weight to the systems with more members, or by averaging the ensemble mean of each of the systems (equal weighting) (see Athanasiadis

et al. 2017). The pooling and equal weighting methods deliver similar levels of performance, but the equal weight seems to be slightly superior. For the SCA, the multisystem performance is better than any of its contributors.

	REOF analysis				*EOF analysis*
	NAO	*EA*	*EAWR*	*SCA*	*NAO*
ECMWF_SEAS5	0.35	0.18	0.13	0.36	0.35
CMCC_SYS3	0.36	-0.08	0.32	0.21	0.30
DWD_SYS2	**0.47**	0.32	0.33	0.39	0.44
GLOSEA5_SYS12	0.38	-0.18	**0.42**	0.35	0.39
MF_SYS6	0.29	**0.41**	0.07	0.40	0.28
Pooling	0.41	0.18	0.30	0.46	0.42
Equal weight	0.42	0.19	0.31	**0.47**	0.44

Table 3. EATC index ensemble mean correlations for each of the analysed systems, valid for DJF and a lead time of one month. Grey values are not statistically significant at a 95% of confidence level. Bold values signal the best result for each EATC.

3.1.5 Sensitivity of ensemble mean correlation to hindcast period and number of members

Previous research has found that ensemble mean correlation is sensitive to hindcast length and ensemble size (Kumar 2009, Baker et al. 2018, Manzanas et al. 2019). The ECMWF SEAS5 hindcast from MARS has 51 ensemble members and a longer hindcast period than the other systems (1981-2018). By subsetting the number of years and the number of members included in the ensemble mean correlation computation, we can evaluate the sensitivity of previous results to the ensemble size and the hindcast length. For each Euro-Atlantic Teleconnection Index, a collection of 10000 random subset of 25 members out of the 51 available are used to draw a distribution of results. In the same way, a collection of 10000 subsets of 25 years out of the 38 available is used to produce one distribution for each EATC.

Figure 3 shows the results for DJF and issued in November (i.e., a lead time of 1 month) in violin plots. The variability when subsampling the period is a bit larger than the variability when subsampling the members for all EATCs. Indeed, the results in terms of sensitivity (i.e. the distribution spread) are very consistent across different EATCs. The violins on the right panel (period subsetting) are shifted to slightly higher correlations, because they employ 51 ensemble members and there is more noise cancellation. It is interesting to note that depending on the selection of members and years, ensemble mean correlations can either be statistical significant or not. From this plot we deduce that ensemble mean correlation is not very robust to evaluate differences among systems, and all the results presented before have to be taken with a grain of salt. In conclusion, large ensembles and large hindcasts are highly needed for evaluating EATC forecasts.

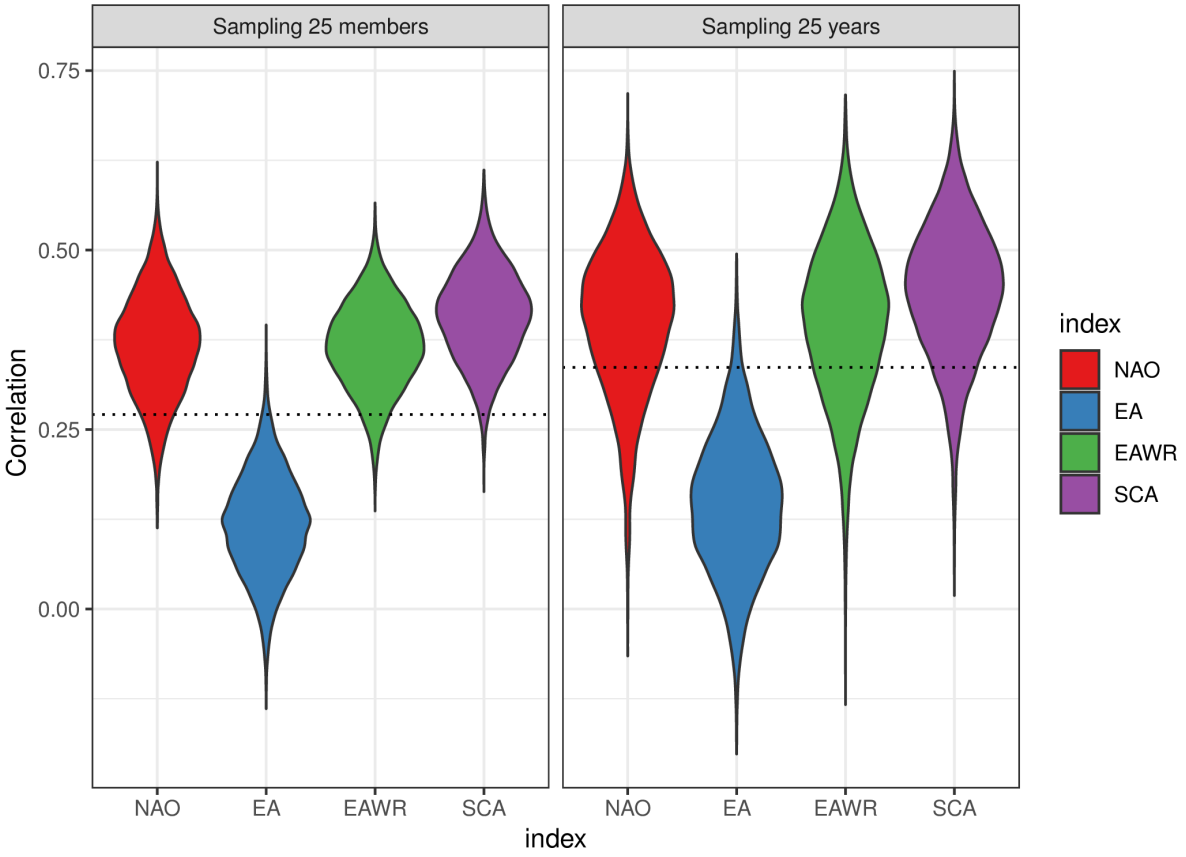

Figure 3. Distribution of the ensemble mean correlation values for SEAS5 when randomly subsampling its hindcast to 25 members (left) and 25 years (right), for each of the four EATCs in DJF and a lead time of one month. The dotted lines indicate the threshold of statistical significance at a confidence level of 95% for 38 years (left) and 25 years (right).

3.1.6 Skill assessment for Spring, Summer and Autumn

Forecasts of the four EATC indices have also been computed for Spring (MAM), Summer (JJA) and Fall (SON) with a lead time of one month (i.e. issued in February, May and August

respectively). Table 4 shows the total explained and residual variance percentage at lead one. In all the seasons, Era-Interim patterns (first row) describe more variability than any of the system projections. The EA, EAWR and SCA indices describe good amounts of variability in spring summer and autumn, and the residuals after using four REOFs are much lower than just employing one EOF, as was the case for winter. The systems have a total variance slightly higher than observation in MAM and SON (except for SEAS5) and lower than observed in JJA (except for CMCC).

The ensemble mean correlation in MAM, JJA and SON (Table 5) is generally lower than in winter (see Table 3). In this case, the NAO obtained with the unrotated EOF methodology has higher values than the REOF-based NAO. The patterns that show higher correlations are NAO in spring, EA/WR in summer and SCA (except for SEAS5) in autumn. Autumn is the season with the lowest correlations.

		Percentage of variance (%)							*Total sqrt variance*
		REOF analysis					*EOF analysis*		
		NAO	*EA*	*EAWR*	*SCA*	*res.*	*NAO*	*res.*	
ERA-Interim	MAM	25	12	18	7	37	38	64	219
	JJA	26	14	20	12	28	30	70	178
	SON	19	12	17	16	36	21	79	208
ECMWF_SEAS5 (C3S)	MAM	20	9	9	10	49	23	77	224
	JJA	13	11	10	10	56	11	89	150
	SON	17	10	17	11	45	18	82	207
CMCC	MAM	19	11	11	11	48	22	78	231
	JJA	10	16	36	8	30	46	54	328
	SON	15	12	15	8	50	15	85	224
DWD_SYS2	MAM	17	9	12	9	53	21	79	240
	JJA	13	11	10	10	57	12	88	151
	SON	15	10	18	10	46	17	83	230
GLOSEA5_SYS12	MAM	20	9	15	9	47	26	74	237
	JJA	15	12	10	9	55	12	88	153
	SON	15	11	16	10	48	16	84	209
MF_SYS6	MAM	20	12	12	9	48	23	77	241
	JJA	13	11	9	9	57	11	89	150
	SON	16	9	18	10	47	17	83	220

Table 4. Total variance, explained variance percentages and residual variance percentage for each prediction system (MAM, JJA, SON)

		REOF analysis				*EOF analysis*
		NAO	*EA*	*EAWR*	*SCA*	*NAO*
ECMWF_SEAS5 (C3S)	MAM	0.33	-0.13	0.24	0.15	0.33
	JJA	0.28	0.40	0.69	0.34	0.83
	SON	0.16	0.25	-0.19	-0.01	0.07
CMCC_SYS3	MAM	0.35	0.42	0.26	0.20	0.37
	JJA	0.32	-0.25	0.15	-0.18	0.37
	SON	-0.05	-0.04	0.01	0.23	-0.03
DWD_SYS2	MAM	0.46	-0.05	0.25	0.23	0.37
	JJA	0.57	0.12	0.70	0.32	0.86
	SON	0.13	0.04	-0.28	0.23xx	-0.03
GLOSEA5_SYS12	MAM	0.32	0.34	0.52	0.22	0.43
	JJA	0.44	-0.22	0.58	0.15	0.84
	SON	-0.18	0.01	0.28	0.26	-0.21
MF_SYS6	MAM	0.34	0.08	0.34	0.34	0.34
	JJA	0.27	0.29	0.79	0.17	0.81
	SON	0.20	0.25	0.20	0.28	0.24
Pooling	MAM	0.41	0.21	0.35	0.33	0.41
	JJA	0.61	0.11	0.66	0.17	0.85
	SON	0.10	0.10	-0.03	0.24	0.09
Equal weight	MAM	0.41	0.19	0.35	0.32	0.41
	JJA	0.60	0.16	0.69	0.21	0.88
	SON	0.12	0.12	-0.02	0.24	0.10

Table 5. EATC index ensemble mean correlations for each of the analysed systems and for MAM, JJA and SON with a lead time of one month. Grey values are not statistically significant at a 95% of confidence level.

3.1.7 Skill assessment at longer lead times

The four observed EATC patterns and indices have been computed for DJF at a lead time of 2 (i.e. predictions initialized in 1st October). Table 6 shows the total explained and residual variance percentage at lead 2 for DJF. For all the systems the total variance explained is lower than at lead 1 (Table 6) with some differences among the teleconnections: for NAO, the

explained variance is generally lower for lead 2 than for lead 1, but for EA and particularly for EA/WR the explained variance is higher and comparable with the observed variance. Correlations for NAO and SCA at lead 2 (Table 7) are lower than at lead 1.

	Percentage of variance (%)							*Total sqrt variance*
	REOF analysis					*EOF analysis*		
	NAO	*EA*	*EAWR*	*SCA*	*res.*	*NAO*	*res.*	
ERA-Interim	32	18	9	19	21	39	61	304
ECMWF_SEAS5 (C3S)	23	15	13	10	39	25	75	285
CMCC_SYS3	26	16	14	8	36	31	69	285
DWD_SYS2	28	14	13	8	38	30	70	313
GLOSEA5_SYS12	24	17	13	9	36	28	72	288
MF_SYS6	24	15	12	10	39	28	72	306

Table 6. Total variance, explained variance percentages and residual variance percentage for each prediction system, both for the REOF and the EOF analysis for DJF at lead 2.

	REOF analysis				*EOF analysis*
	NAO	*EA*	*EAWR*	*SCA*	*NAO*
ECMWF_SEAS5 (C3S)	0.17	0.32	0.27	0.26	0.16
CMCC_SYS3	0.11	-0.10	0.18	0.33	-0.01
DWD_SYS2	0.36	-0.11	0.24	0.26	0.40
GLOSEA5_SYS12	0.13	0.10	0.53	0.48	0.10
MF_SYS6	0.30	0.08	0.20	0.29	0.28
Pooling	0.28	0.04	0.39	0.34	0.24
Equal weight	0.29	0.10	0.38	0.34	0.26

Table 7. EATC index ensemble mean correlations for each of the analysed systems for DJF at a lead time of 2 months. Grey values are not statistically significant at a 95% of confidence level.

3.1.8 Conclusions

A method that allows computing NAO, EA, EAWR and SCA index forecasts has been developed. Atmospheric variability and its energy impacts in the Euro-Atlantic region are better characterized by the four EATC indices than just by NAO alone, therefore forecasts of those four indices might be useful to understand surface impacts. We have shown that EA, EAWR and SCA winter forecasts issued one month ahead can be as skillful as NAO forecasts. Generally speaking, skill for the spring and summer seasons is slightly lower than for winter, while the predictions for autumn have a negligible skill. At longer lead times, correlation values also decrease.

We have also shown that multisystem ensembles can be better than the best individual system sometimes. It is important to note that skill estimates of ensemble mean correlation are very sensitive to ensemble size and period.

3.2 Relationship between EATC patterns and surface variables in the seasonal systems

To understand how well the seasonal prediction systems reproduce the observed EATC impacts on surface wind, radiation and temperature, correlations between forecasted EATCs and its respective surface variable predictions have been produced (separately for each ensemble member and then averaged together). The same correlations have been performed with observations (i.e. ERA-Interim) as a benchmark for comparison.

3.2.1 Wind

Figure 4 shows the surface wind correlations with each EATC for DJF for ECMWF SEAS5 and DWD System2. Although in general terms the correlation patterns are quite similar in shape and position, overall the correlations are stronger for the observations than for the systems, with differences of up to 0.4 points in some patterns and regions. The most notable differences can be seen for the EAWR teleconnection. Those differences show that systems are not able to exactly reproduce the teleconnection impacts that are expected from the observations. It is worth noting how both analysed systems show similar biases in terms of wind correlation with the forecasted EATCs.

Regression coefficients have also been computed and compared between forecasts and observations (see Figure 5). The EATC index forecasts have been scaled before regressing wind speed onto them. These results are similar to the correlations above but highlight the regions where the impact of EATCs on the absolute wind anomalies is bigger. In general, regression coefficients for the prediction systems are smaller than observed, indicating that EATC anomalies produce weaker effects on wind speed anomalies in the forecasts than in the observations.

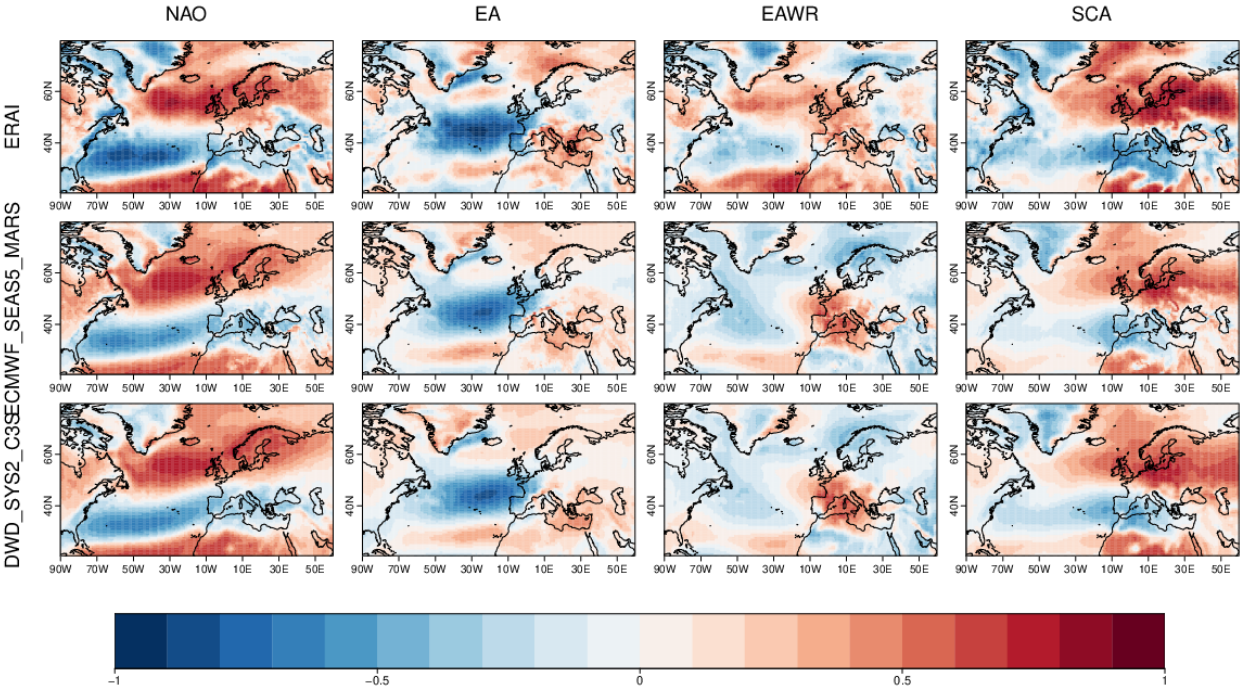

Figure 4. Correlation between EATCs and sfcWind for ERA-Interim (first row) and two seasonal prediction systems (SEAS5 and DWD_Sys2, second and third rows).

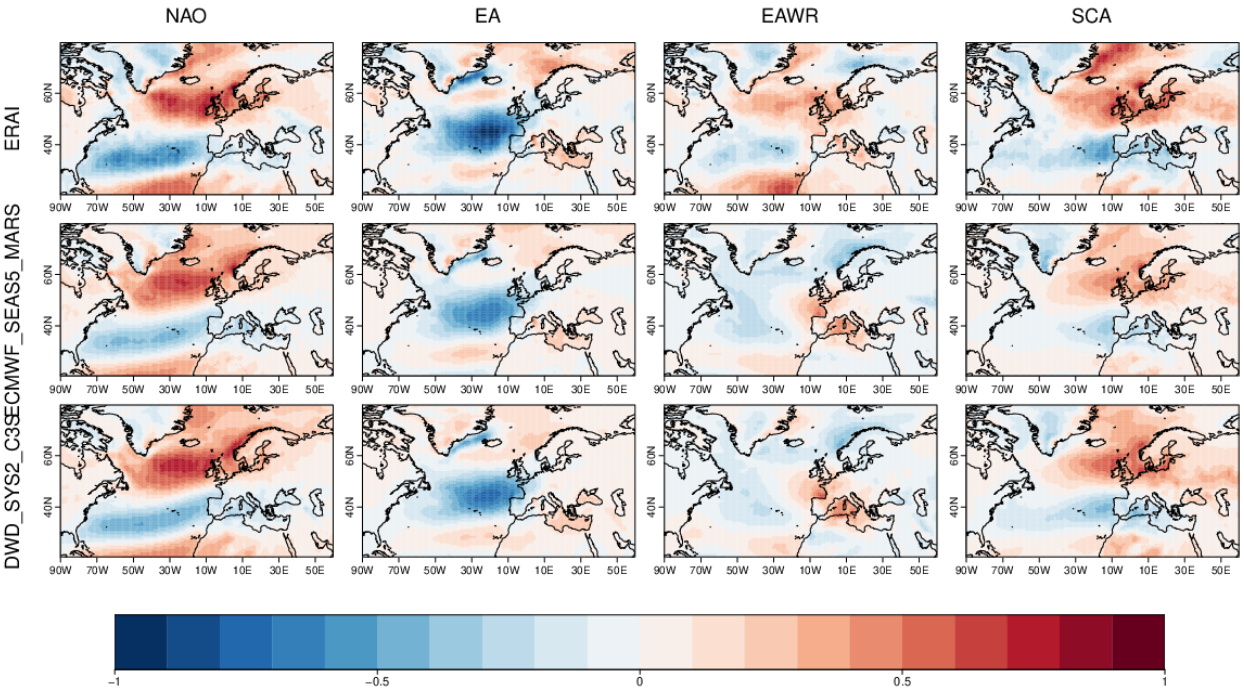

Figure 5. Regression patterns [sfcWind~A*EATC+B] for ERA-Interim (first row) and two seasonal prediction systems (SEAS5 and DWD_Sys2, second and third rows).

3.2.2 Solar radiation

Figure 6 shows correlations between the EATC indices and surface solar radiation anomalies for ERA-Interim, SEAS5 (from C3S), DWD, CMCC, GLOSEA5 and MF_SYS6 during DJF. The patches of correlation are similar for observations and prediction systems. NAO is generally correlated with an increase of SSRD over Southern Europe and SCA is correlated with a decrease of SSRS over continental Europe. The prediction systems reproduce this behaviour but fail to reproduce the intensity.

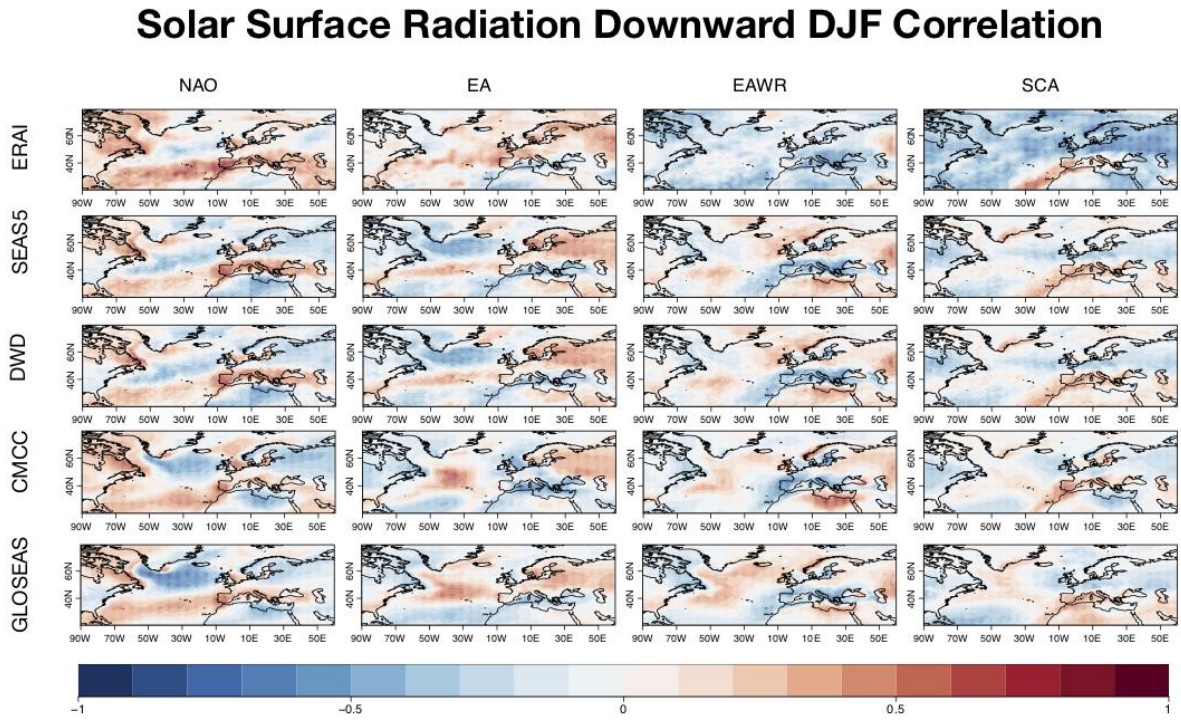

Figure 6. Correlation between EATCs and surface solar radiation downward for ERA-Interim (1st row) and four seasonal prediction systems: SEAS5 (2nd row), DWD_Sys2 (3rd row), CMCC (4th row) and GLOSEAS (5th row).

3.2.3 Surface temperature

The position and the shape of the correlation patterns are quite similar between observations and seasonal prediction systems for 2 m temperature (Figure 7), although the observations have always larger values. NAO positive correlation and EA negative correlation are well reproduced by the models over Europe. EAWR negative correlation is generally comparable with observed correlation. SCA correlation over Europe is not as strong as in ERA-interim.

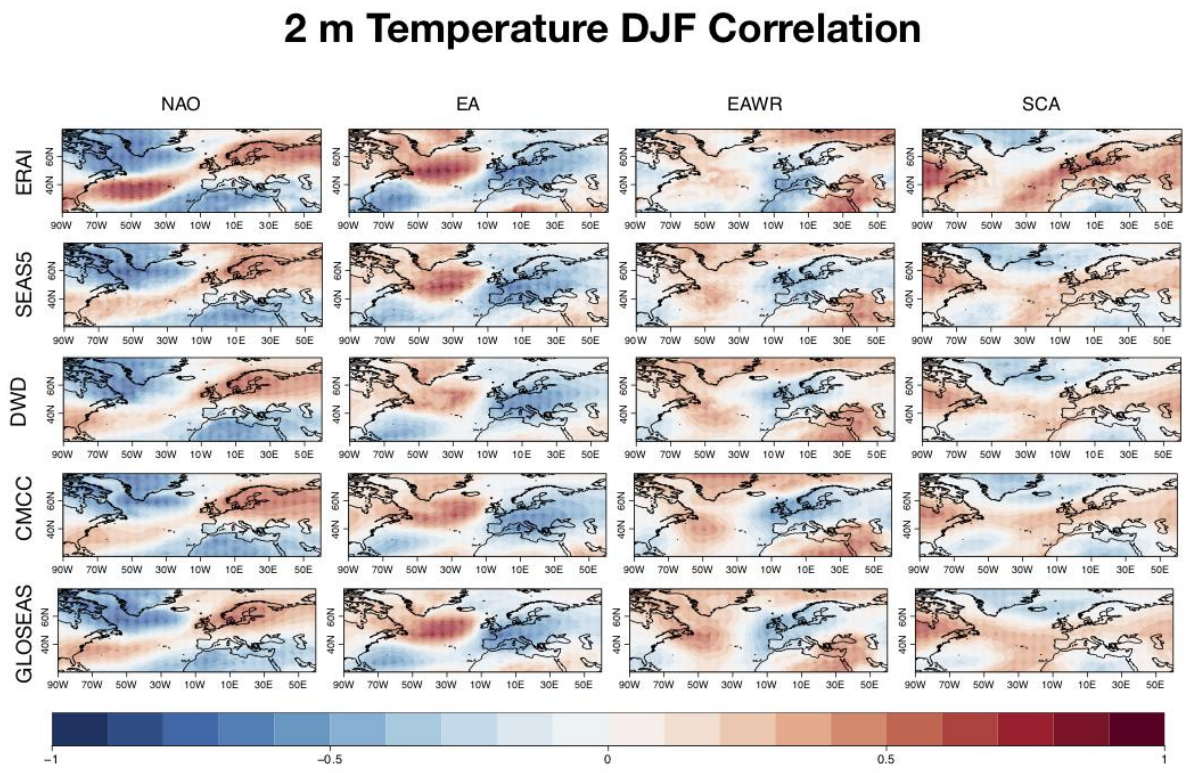

Figure 7. Correlation between EATCs and 2m temperature for ERA-Interim (1st row) and four seasonal prediction systems: SEAS5 (2nd row) DWD_Sys2 (3rd row), CMCC (4th row) and GLOSEAS (5th row).

3.2.4 Conclusions

For all the analysed variables the relationship between EATCs and surface variables is weaker in prediction systems than in observations. There are also biases in the exact position of the correlation patterns. These deficiencies in reproducing impacts of each EATC in energy-related variables opens the door to produce reconstructions of surface variables combining forecasts of those four EATC indices -which have proven skillful in the previous section- and the observed impacts.

4 Weather regimes predictability in seasonal and subseasonal systems

4.1 Weather regimes predictability in seasonal forecast systems

4.1.1 Introduction

In this section, the ability of ECMWF System 4 to reproduce the observed monthly Euro-Atlantic weather regimes (WRs) defined in ERA-Interim is explored. System 4 has been the operational seasonal system at ECMWF from 2011 until 2017 and is a state-of-the-art coupled ocean-atmosphere model. Two different approaches have been used to identify the WRs in the prediction system, replicating the k-means clustering method (KM) and an assignation method using the minimum root mean square distance (RMSD). The analysis examines the correspondence between the spatial patterns and the monthly frequency of occurrence of the WRs obtained for the predictions for all months and for lead times from 0 to 6 months with the regimes obtained for ERA-Interim. For more details of this work see Torralba (2019).

4.1.2 Methodology

This analysis is based on daily mean anomalies of sea level pressure (MSLP) in the period 1982-2016 (35 years) for the Euro-Atlantic region (27°N–81°N, 85.5°W–45°E). WRs are defined on a monthly basis by the K-means clustering algorithm as detailed in Deliverable 3.2 (Section 3.2.1). The seasonal predictions of the ECMWF System 4 have been employed and ERA-Interim has been used as a reference dataset. The sensitivity of the WRs classification to the choice of the reanalysis has been explored in Cortesi et al. (2019) and it is demonstrated that different reanalyses produce equivalent WR classifications. As this analysis has been performed for each individual month of the year, 15 members have been used to obtain consistent results among the months.

To obtain the WRs based on the seasonal forecasts from ECMWF System 4 (Figure 8, right side) two different approaches have been adopted. The first method is the same KM cluster analysis that has been used with the ERA-Interim reanalysis, which is useful to investigate how the seasonal forecast systems simulate the weather regimes. As an alternative approach, a WR classification based on the minimum root-mean-square distance (RMSD) between the predicted MSLP daily anomalies and the defined clusters from the ERA-Interim reanalysis has been obtained. Although the fair verification of the WRs computed from seasonal forecasts could require that these regimes would be obtained independently from the observational reference, the KM approach presents some limitations when applied to seasonal forecasts. To deal with these limitations, the RMSD methodology is proposed.

RMSD is based on the sum of squares differences between the MSLP daily anomaly maps from ECMWF System 4 (corresponding to each day, member and forecast time) and a set of clusters employed as a reference. In this work, the reference is the four clusters obtained from

the ERA-Interim reanalysis for each separate month. Then, the particular day is assigned to the cluster for which the RMSD is minimized (Neal et al. 2016). This method guarantees that the predicted WRs have very similar spatial structures to the observed regimes, which is essential for their verification.

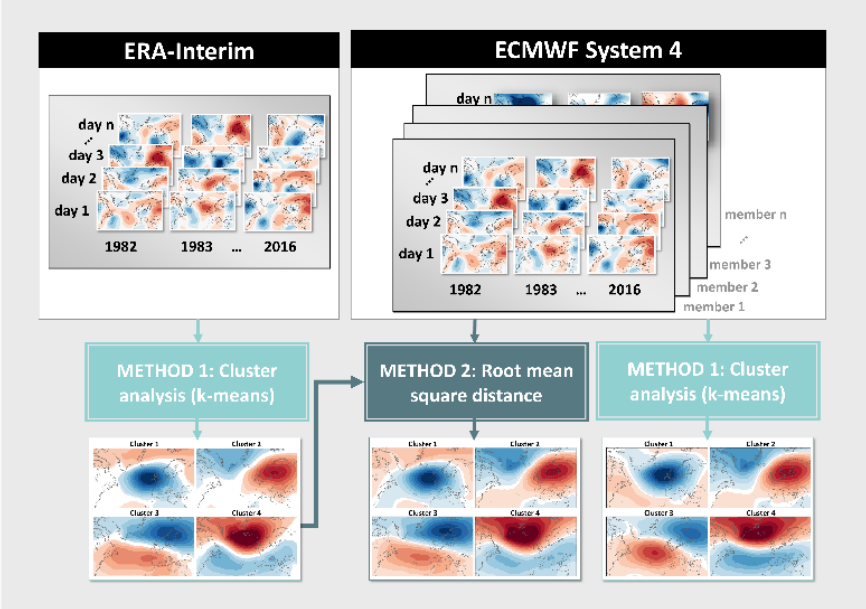

Figure 8. Workflow used to perform weather regime classifications for each month of the year from daily sea level pressure anomalies. The k-means (KM) method has been applied over the ERA-Interim daily anomalies to obtain a reference classification. Then two different sets of WRs have been obtained from the ECMWF System 4 seasonal forecasts.

4.1.3 Results

The ability of the KM and RMSD methods to produce similar spatial patterns to those observed has been summarised in terms of spatial correlations between the WRs based on ECMWF System 4 seasonal forecasts and the corresponding WRs from ERA-Interim for all the available months and lead times (Figure 9).

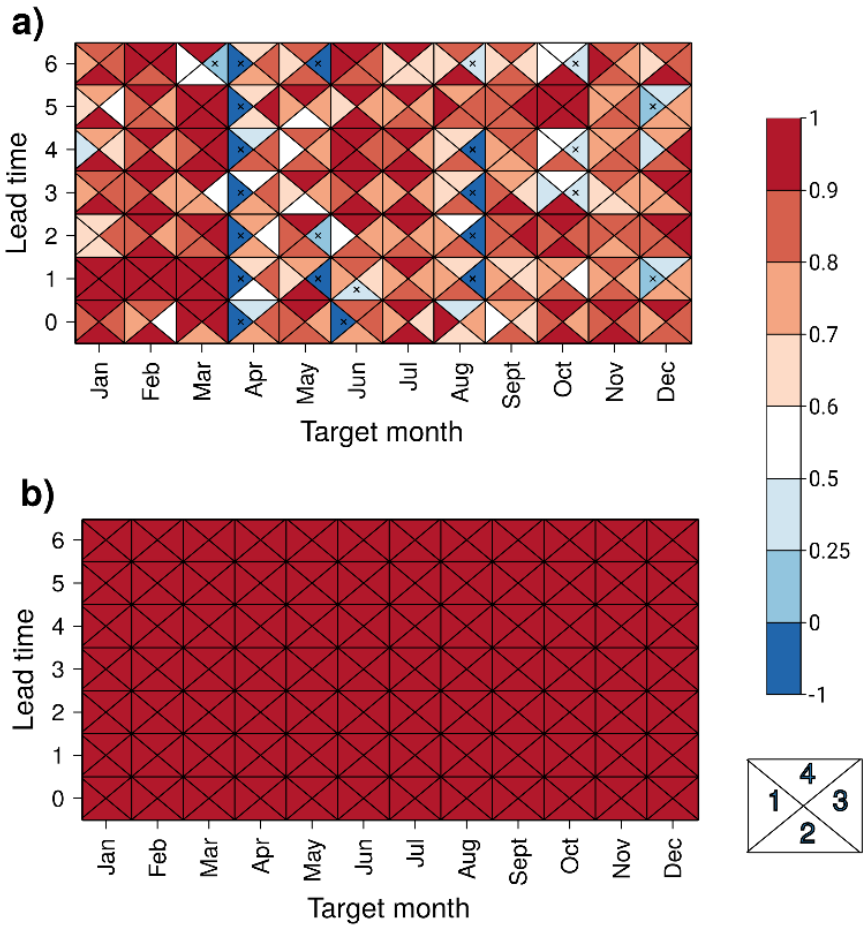

Figure 9. Spatial correlation between the ECMWF System 4 seasonal predictions and ERA-Interim reanalysis monthly weather regimes patterns obtained by the a) KM and b) RMSD methods. The classifications have been performed with daily sea level pressure anomalies in the 1982-2016 period. Each triangle represents the correlations for each of the four clusters as indicated in the bottom right legend. The spatial correlations are shown as a function of the target month (x-axis) and lead time (y-axis). Crosses denote non-significant correlation values (two-tailed t-test at a 95% confidence level).

The spatial patterns of the WRs computed from the seasonal forecasts with the KM method (Figure 9a) have some differences with the ERA-Interim WRs, as indicated by the negative and non-significant spatial correlations in certain months and leads. For example, Cluster 1 displays negative and non-significant correlations with the ERA-Interim Cluster 1 for most of the lead times in April. This is caused by a complete mismatch in the spatial pattern. Cluster 1 found in ERA-I cannot be found in any of the April clusters from ECMWF System 4 (For details see Torralba (2019) Appendix B, Figure B27 therein). The difficulties in matching the WRs from the seasonal predictions with those from the observational reference illustrate the difficulties that arise in the validation of predicted WRs as a consequence of the systematic errors affecting these forecasts. To bypass this problem the RMSD methodology is used, by which each day is assigned to one of the reference clusters defined in in ERA-Interim. The

RMSD method produces a WR classification with spatial patterns very consistent with the ERA-Interim patterns for all months and lead times, with spatial correlations above 0.9 (Figure 9b). This occurs because the RMSD approach seeks to group simulated daily MSLP anomalies similar to those from the ERA-Interim reanalysis. Consequently, it is expected that WR maps obtained for the predictions show a high resemblance to those in the observational reference.

To assess the efficiency of the WRs derived from the seasonal predictions to reproduce the influence of the atmospheric circulation on the wind speed field discussed in D3.2-3.2.1, spatial correlations have been used. These correlations allow to quantify the correspondence between the composite maps of wind speed obtained for the WR classification in ECMWF System 4 and the equivalent maps in the ERA-Interim reanalysis (Figure 10).

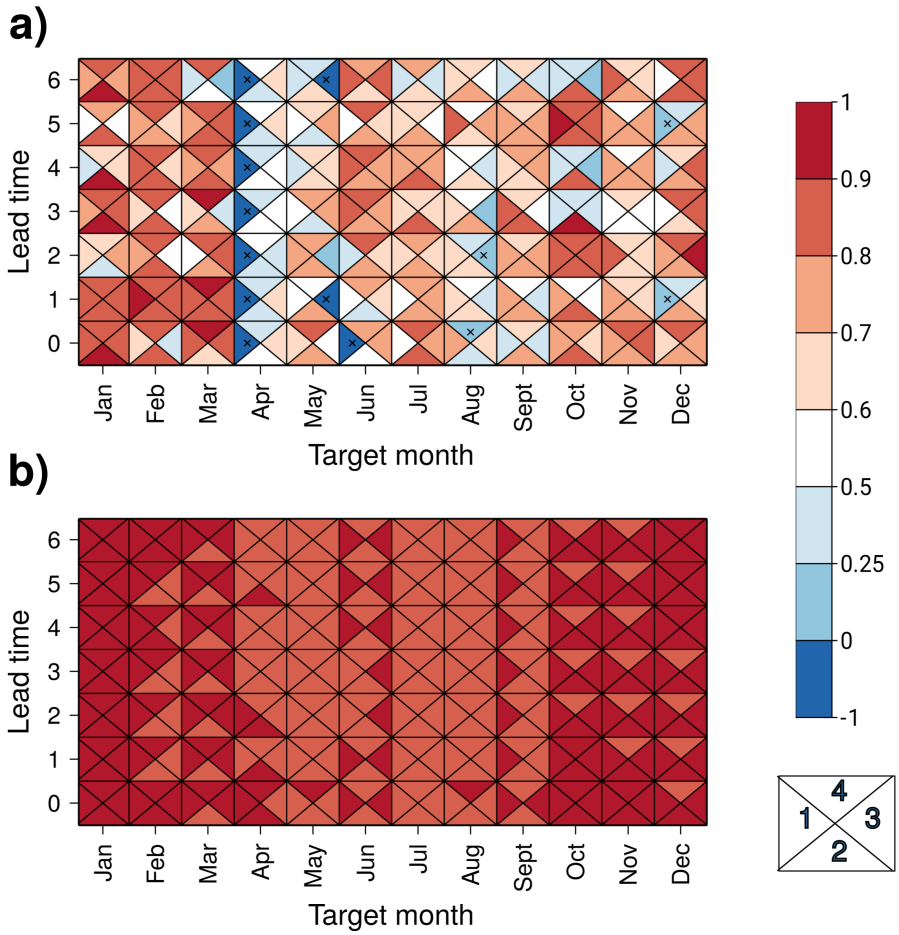

Figure 10. Spatial correlation between the ECMWF System 4 seasonal predictions and ERA-Interim reanalysis composites of the 10-m wind speed with monthly WRs obtained by the KM (a) and RMSD (b) methods. The classifications have been performed with daily sea level pressure anomalies in the 1982-2016 period. Each triangle represents the correlations for a specific cluster as indicated in the bottom right legend. The spatial correlations are shown as a function of the target month (x-axis) and lead time (y-axis). Crosses denote non-significant correlation values (two-tailed t-test at a 95% confidence level).

The results show that the impact of those WRs obtained for the seasonal forecasts from the RMSD method (Figure 10b) show higher resemblance to the wind speed composites obtained for ERA-Interim (as their spatial correlations are above 0.8 for all months and lead times) than those obtained by the KM method (Figure 10b). The small differences in windspeed spatial correlations that still appear when employing the RMSD method are caused by systematic errors of the seasonal forecasts and/or local effects.

The analysis of the WR spatial patterns and their influence on the wind speed daily anomalies, has been complemented with the assessment of the seasonal forecast system ability to predict the year-to-year variations in the frequency of occurrence of a specific regime. To investigate the correspondence between the monthly frequency of occurrence of the clusters obtained from the seasonal predictions and those from the observational reference, Pearson correlation has been calculated.

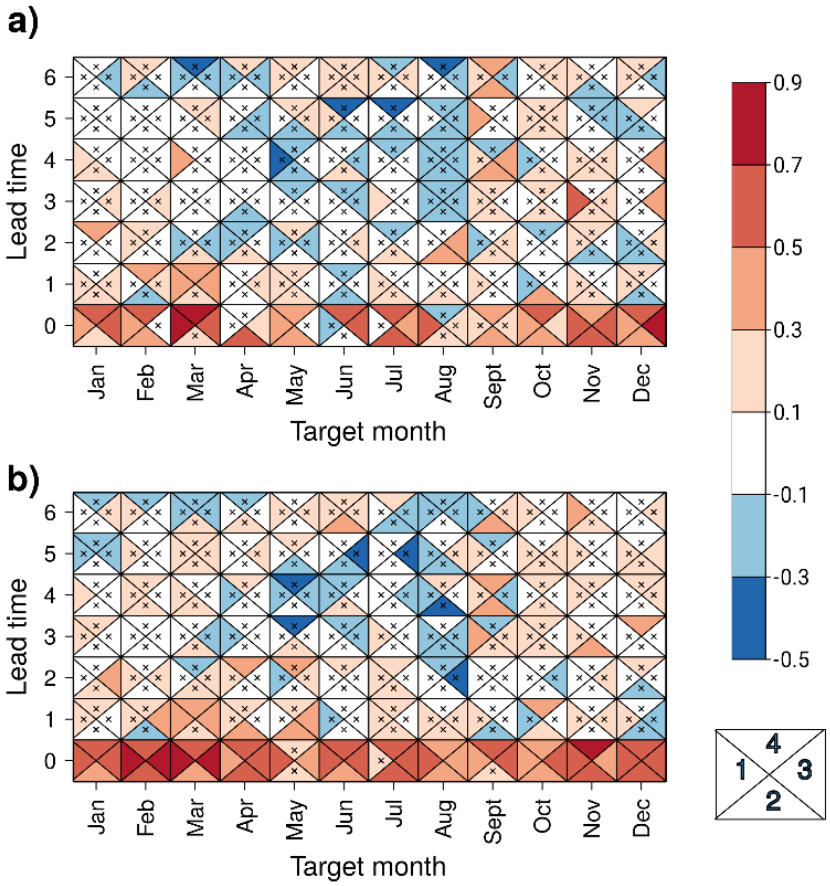

Figure 11. Pearson correlation between the ECMWF System 4 seasonal predictions and ERA-Interim reanalysis monthly frequencies of occurrence corresponding to the WRs obtained by the (a) KM method and (b) RMSD methods. The classifications have been performed with daily sea level pressure anomalies in the 1982-2016 period. Each triangle represents the correlations for a specific cluster as indicated in the bottom right legend. The spatial correlations are shown as a function of the target month (x-axis) and the lead time (y-axis). Crosses denote non-significant correlation values (two-tailed t-test at a 95% confidence level).

Figure 11 shows that RMSD method provides slightly better correlations than the KM method for all target months and clusters. Nevertheless, positive and significant correlations are mostly restricted to lead 0 for all the target months. Beyond lead 0, the ability of the seasonal forecast system to reproduce the monthly frequencies of the WRs drops, and only positive correlations are found for the frequencies of some clusters for some leads in what looks like a random distribution. However, one exception can be found in March, when the WRs obtained by the two different methods show significant positive correlations for Clusters 1, 3 and 4 for the predictions initialised in February (1 lead month).

4.1.4 Conclusions

Of the two methods employed to derive the WRs from the seasonal prediction system, the KM may produce spatial patterns that do not resemble those in the reference patterns and is therefore not suitable verification purposes. However, it is a useful diagnostic tool to identify systematic errors in seasonal predictions. The RMSD method is most suitable to perform an effective verification and therefore to generate a product based on seasonal predictions of WRs in an operational context, which is essential to distribute climate information tailored to specific users' needs.

The spatial correlations between predicted and observed MSLP and wind speed composite fields associated to each WR with the RMSD method are considerably high (Figure 9 and Figure 10) indicating a good (but not perfect) match between the impact of large scale circulation patterns and the surface variables in model and observations.. The deviations between the forecasts and observations can be attributed to local effects or systematic errors in the forecasts. However, we find that the main limitation for the use of WRs as predictors for surface variables lies in the poor skill in predicting their frequency of occurrence for lead times. ECMWF System 4 accurately predicts the monthly frequencies of occurrence of the WRs only for lead time 0 which reflects the difficulties to predict the seasonal climate in the North Atlantic region.

As future work, it would be interesting to repeat this analysis in the current ECMWF seasonal forecast system SEAS5. Nevertheless, System 4 is a state-of-the-art global atmosphere-ocean coupled model, so the present results offer valid conclusions.

4.2 Weather regime predictability in subseasonal forecast systems

4.2.1 Introduction

In this section, the predictability of WRs in subseasonal timescales is analysed. The observational reference chosen is ERA-Interim and the selected prediction system is the ECMWF monthly prediction system 2018 (version CY43r3/CY45R1) with its associated

hindcast for the period 1998-2017. In the case of subseasonal predictions, the forecast window is of one week, with the following forecast times: week 1: days 5-11, week 2: days 12-18, week 3: days 19-25 and week 4: days 26-32. To avoid most of the discontinuities that a monthly or seasonal WR classification often introduces when moving from one month to the following, a weekly classification has been designed, producing 52 different classifications, one for each week of the year. An annual classification is another alternative choice. However, the weekly approach favours the emergence of less-frequent regimes which might be relevant in some periods of the year.

4.2.2 Methodology

The k-means cluster algorithm has been applied to daily mean sea level pressure anomalies from ERA-Interim reanalysis over the same region employed in Section 4.1 (27°N–81°N, 85.5°W–45°E) for a running window of 5 weeks, centered in the week of interest. Therefore, the period is similar to the one in monthly WRs but allowing for a smoother transition. Following the same approach that has been applied to obtain the monthly WR, which was already described in D3.2-3.2.1, a LOESS polynomial regression with a degree of smoothing α = 0.15 (Mahlstein et al. 2015) has been applied on the daily climatologies to remove the influence of the annual cycle and to minimize the short-term variability of the climatological estimates. The KM cluster algorithm was applied with 50 random starts and a maximum of 200 iterations. The number of clusters was set to 4 as in Michelangeli (1995). The resulting classification consists of 52 sets of 4 patterns as displayed in Figure 12. Week 1 refers to the first week of January, week 2 to the second one of the same month, and so on until week 52, the last one of December. The names of the months are included to indicate the position of the weeks during the year. The WRs of each week were ordered to try to match the patterns throughout the year. We can see how most of the time there is a continuous evolution of the MSLP, with weekly patterns evolving with slight changes. However, some unavoidable discontinuities appear at some times of the year. For example, the patterns of WR1 and WR2 in week 16 do not have correspondence with those patterns obtained for week 17 (Figure 12).

Assignation of daily MSLP anomalies from ECMWF monthly forecast system to the ERA-Interim WRs was done by root mean square distance (RMSD) method. As it has been discussed in Section 4.1, this method allows a direct matching between the observed WR patterns from ERA-Interim with the classification obtained from the subseasonal forecasts. The daily MSLP anomaly fields for each member and forecast time are associated to one of the 4 observed WRs, the one that minimises the RMSD.

4.2.3 Results

As a consequence of the approach chosen for WR assignation, the RMSD method, the simulated WR patterns are very close to the observed ones, with spatial correlations between them always higher than 0.9 (not shown).

In terms of predicted WRs frequency of occurrence, we look at two indices: the Pearson correlation between the frequency of occurrence of predicted and observed WRs, and the bias in the probability of transition of one WR into the other three and itself. The frequency of occurrence of the WRs is calculated as the number of days a certain WR occurs in a moving window of 5 consecutive weeks (to increase statistical robustness), for every week and lead time. Figure 13 shows the Pearson correlation between predicted and observed WR frequency of occurrence. It can be seen that for forecast days 5-11 (week 1) correlations in frequency are always significant and above 0.6 for the four WRs. Additionally, many weeks during winter show correlations above 0.8 for all WRs, especially in the case of WR 1 (NAO-), which appears to be the most skillfully predicted. From the second forecast week, skill drops considerably, except for winter weeks. Notably, there are some periods of the year that stand out with significant skill. Particularly, correlations of WR 1 in February persist with $r > 0.4$ even until the fourth forecast week (forecast days 26-32). In January and December WR1 shows correlations between 0.2 and 0.4 in weeks 3 and 4. For the other WR, positive correlations beyond week 1 are also found for the period February–March. Overall, forecasts of WR 1 (NAO-) are most skillful, which is consistent with results by Matsueda and Palmer (2018) for the medium-range.

The transition probability is the probability of a WR to evolve to another one (including itself) on a daily basis. For a given WR, start date and forecast week, its transition probability was measured by counting the number of times it transitions to another WR and normalising it by the total number of days. Figure 14 shows the bias (difference) between predicted and observed transition probability. Red boxes indicate that the forecasts overestimate that particular transition and blue boxes indicate that the forecasts underestimate that particular transition. The most interesting feature of Figure 14 is that for each transition bias it is possible to identify the compensating transition. Any overestimation of transition probabilities corresponds to an underestimation of one or more transitions probabilities. For example, transitions from WR 2 to WR 4 are greatly overestimated (red boxes) for start dates 25-50; on the contrary, transitions from WR 2 to itself (no WR change) are highly underestimated (blue boxes) during the same start dates. Thus, ECMWF-MFS favours more an evolution from WR 2 to WR 4 than from WR 2 to itself, compared to observed transitions. A similar "compensation" mechanism is clearly visible for start dates 41-47 of WR 1, when the system favours transitions between WR 1 and WR 4 and in favour of WR 1 with itself.For WR 4, there is an underestimation of the transitions between WR 4 and itself during start date 34-52 are almost evenly compensated by an overestimation of the transitions to the other three WRs. Globally, highest biases are found for the transition of a WR to itself, and usually they represent an underestimation of the transition probabilities. Thus, WRs simulated by ECMWF-MFS are usually less prone to transition to themselves than the observed WRs, favouring instead transitions to other WRs.

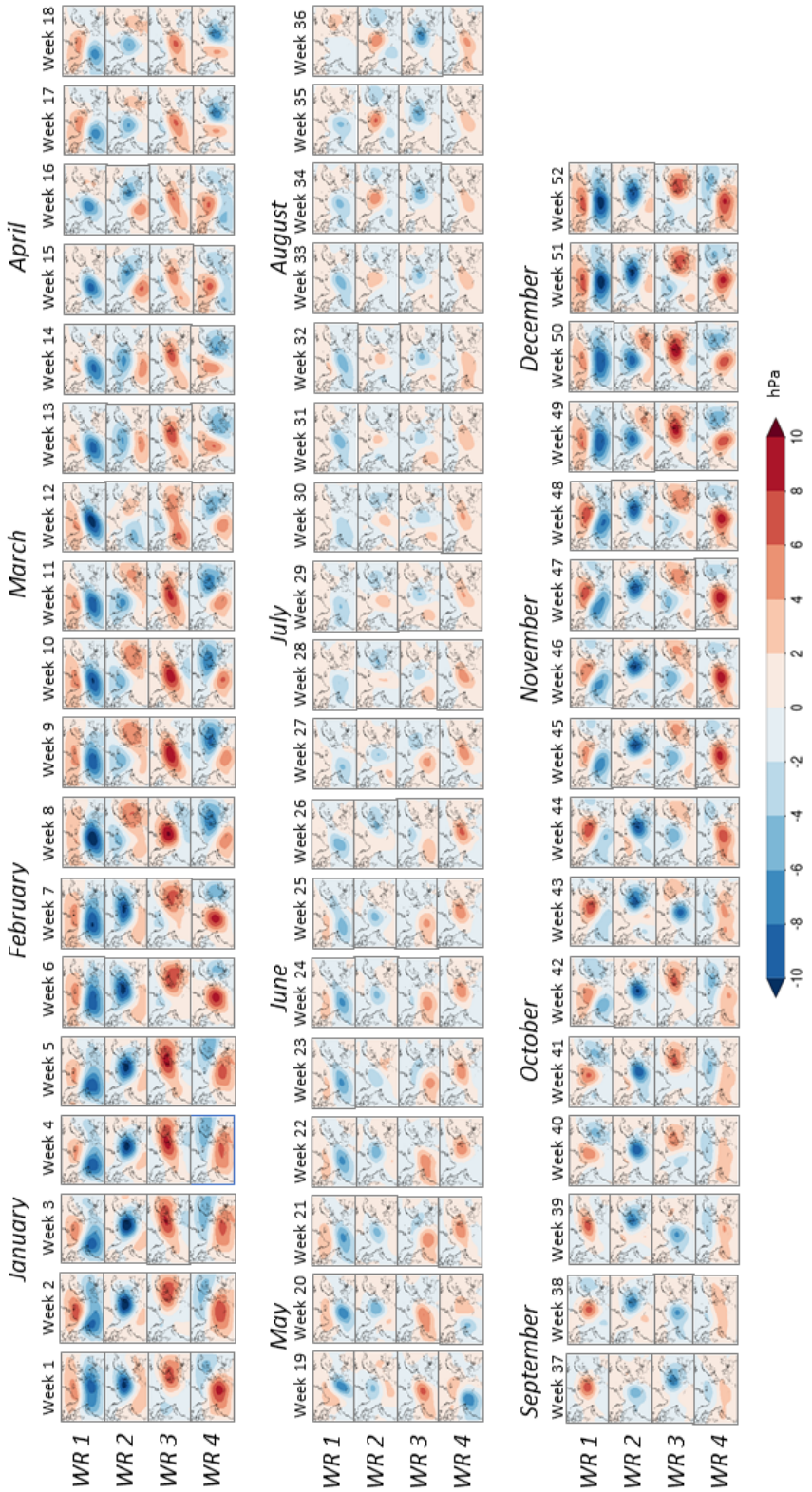

Figure 12. WR patterns of the weekly classification derived from ERA-Interim MSLP daily anomalies (1981-2017).

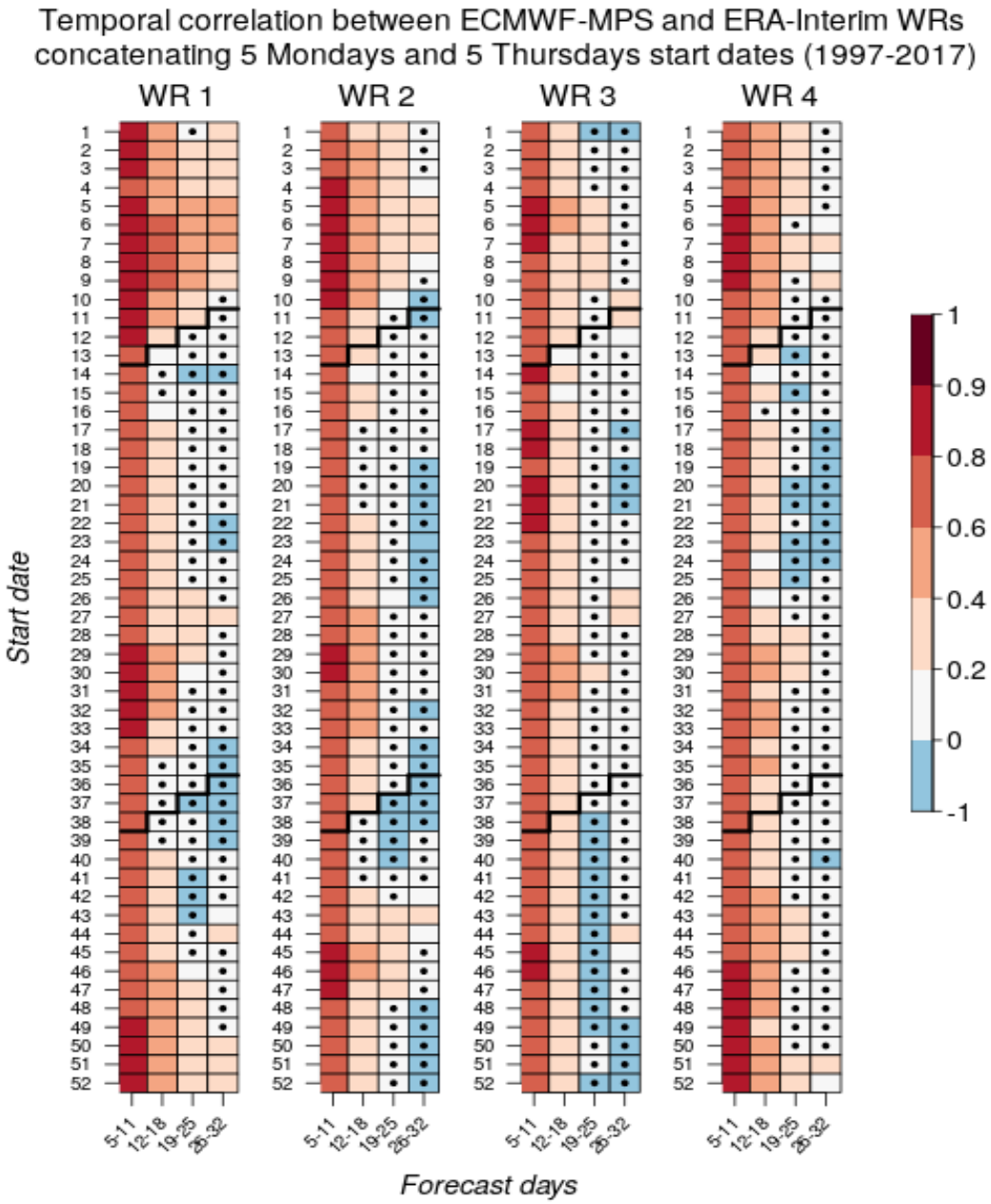

Figure 13. Pearson correlations between predicted (ECMWF-MPS) and observed (ERA-Interim) weekly frequency of occurrence of WRs based on 5-weeks mobile k-means clustering from MSLP anomalies of ECMWF subseasonal forecasts and of ERA-Interim (1998-2017). Correlations refer to the forecasts concatenating ten consecutive Mondays and Thursdays start dates. The x axis indicates both the WR and the forecast time while the y axis corresponds to the week of the year of the start date, from 1 to 52. Black dots indicate not-significant correlation (for a paired t-test with α = 0.01). Black bold lines separate weeks of the winter months (ONDJFM) and summer ones (AMJJAG).

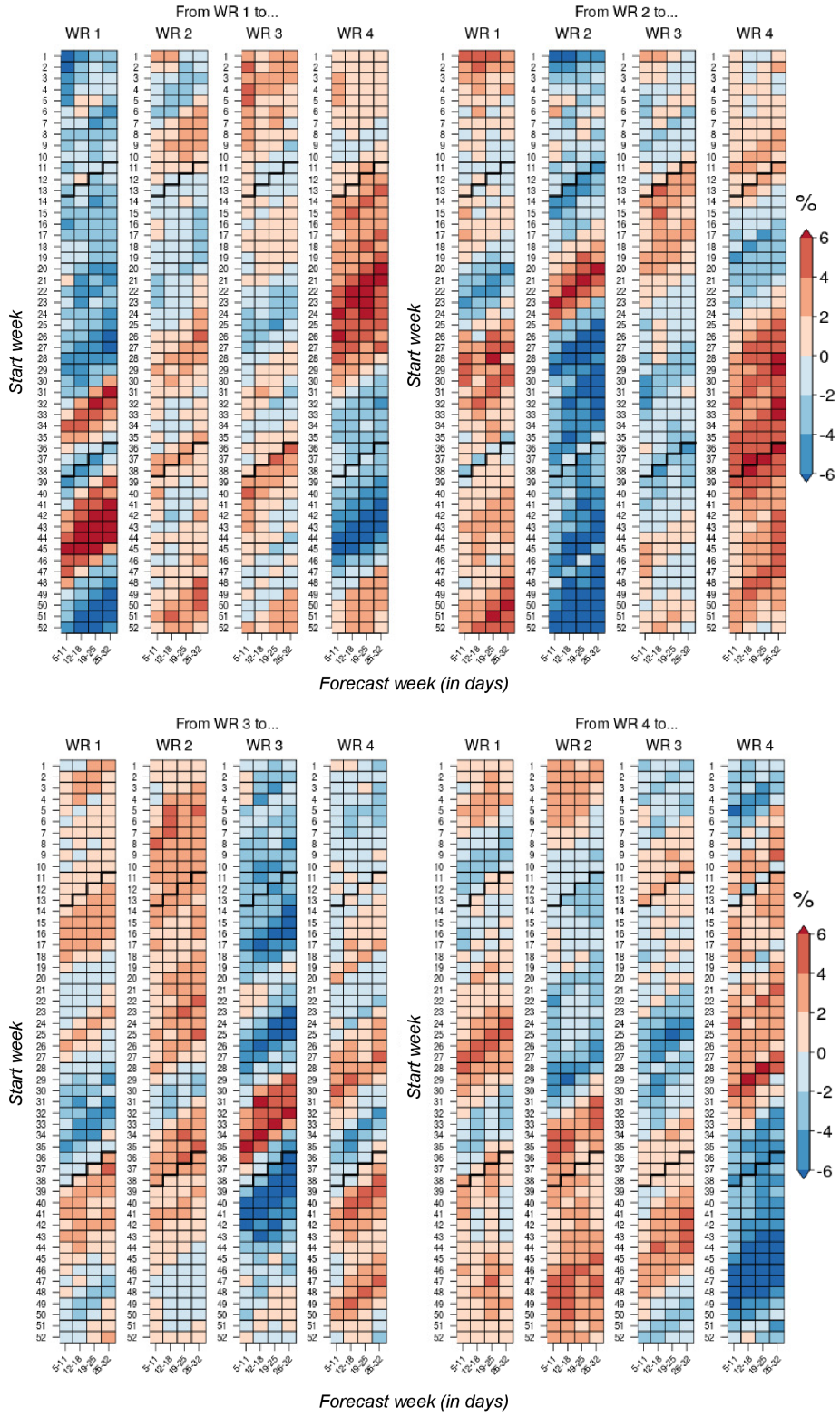

Figure 14. Transition probability bias. The x axis indicates both the pair of transitioning WRs and the forecast time, while the y axis corresponds to the week of the year of the start date, from 1 to 52. Black lines separate winter weeks (ONDJFM) from summer ones (AMJJAG). Source: ERA-Interim and ECMWF subseasonal forecasts 1998-2017.

4.2.4 Conclusions

In this section the ability of ECMWF-MFS to predict different atmospheric WRs in the Euro-Atlantic region (1998-2017) has been analysed. A novel weekly classification based on KM clustering of daily MSLP over a running window of 5 weeks has been employed. The objective of such classification is to allow the emergence of less-frequent regimes which might be relevant in some periods of the year. The running window aims to minimise discontinuities in the WR spatial patterns, allowing an operational all year round implementation of WR forecast product. There is good evidence that there is some skill in predicting WR occurrence in weeks 1-2 all year round, but there is marked seasonal difference, the highest skill is found for winter. The Pearson correlation of the predicted weekly frequencies of occurrence has values of $r > 0.6$ for week one. Skill decreases with lead time but, for some WR and seasons, there is modest level of skill even in weeks 3 and 4. In particular, forecasts of WR 1 (NAO-) have the highest skill. ECMWF-MFS tends to have more skill in predicting WR frequencies in winter especially early (Dec) and late winter (Feb). A window of opportunity was detected in February, as all its weeks show the highest correlations: $r > 0.8$ for the first forecast week (5-11 days) and $r > 0.6$ for the second one (12-18 days). However, during the other weeks of the year, especially transition seasons, correlations drops below $r = 0.6$ from the first to the second forecast week. The analysis of the transition probability bias shows some that ECMWF-MFS tends to overestimate transitions from one WR to another while underestimating transitions from one WR to itself.

5 Weather regime and TCT representation in subseasonal forecasts – European "national-scale" energy balance indicators

5.1 Introduction

The analysis in this section shows how the representation of the weather regimes (WRs) and Targeted Circulation Types (TCTs) compares between ERA5 and the subseasonal hindcasts from ECMWF and NCEP. After recapping the methods used to calculate country-aggregate energy variables, weather regimes and TCTs (Section 5.2) we assess the ability of the forecasts to replicate the WR and TCT frequencies seen in ERA5 at lead weeks 1-4 (Section 5.3). Following this the connections between the forecast WRs and TCTs and the surface energy indicators of interest are shown (Section 5.4) with a focus on demand-net-wind (DNW). To conclude, the predictability of the two types of patterns is assessed in Section 5.5.

In this chapter the analysis presented is limited to the November-March period (rather than each week or month individually as presented in Chapter 4) and to the demand-net-wind national energy balance indicators (DNW). Further analysis has been completed for other time periods and for 2m temperature, 10m wind speeds, demand, and demand-net-solar but these results are omitted for brevity.

5.2 Methods

A brief overview of the country aggregate demand and wind power capacity factor models is given below. Full details of the country aggregate models can be found in Bloomfield et al. (submitted) and S2S4E Deliverable 3.2 Annex A.

5.2.1 Meteorology-to-power conversion: ERA5 reanalysis

Electricity demand is calculated with a country-level multiple-linear regression model containing parameters to capture both meteorological and human behaviour. Each country has a unique regression model, which is trained on two years of measured demand data (2016-2017) from the ENTSOe transparency platform (ENTOSe, 2019), and is then applied retrospectively to the full ERA5 reanalysis period (1980-2018). Two versions of the model output are created, the "full" demand (using all of the available regression parameters) and the "weather-dependent" demand (which includes only the weather-dependent terms, heating-degree-days and cooling-degree-days – i.e., removes the impacts of the day-of-week behavioural patterns and long term socio-economic trends). In this chapter the weather-dependent model is used to highlight the meteorologically driven power system variability.

Wind power capacity factor is calculated based on the methodology of Lledó, et al. (2017) and Lledó et al. (2019) which calculates gridded capacity factor using three different power curves corresponding to three turbine classes, and is then aggregated to country level (similar to Cannon et al. 2015 and Bloomfield et al. 2016). To calculate country aggregate capacity factor, firstly the most appropriate wind turbine for each grid box is calculated based on the 1980-2018 mean ERA5 100m wind speed. Previous work in Deliverable 4.1 highlighted that ERA5 produces anomalously low 100m wind speeds over large regions of Europe; therefore prior to use within the wind power model the ERA5 100m wind speeds are bias corrected to the global wind atlas (Global Wind Atlas, 2019). The country aggregate capacity factor is calculated by passing bias corrected 100m wind speeds through each curve and aggregating based on the locations in installed turbines taken from thewindpower.net (2019).

5.2.2 Meteorology-to-power conversion: hindcasts

Hindcasts of energy demand and wind power are calculated from hindcasts of 2m temperature and 10m wind speed. Hindcasts from both ECMWF and NCEP are used from the S2S database (Vitart et al., 2017). The ECMWF hindcasts cover the years 1996 to 2015, and the NCEP hindcasts cover the years 1999 to 2010. Only hindcasts initialised in the months November to March are used. Since the wind power model requires 100m wind speed, the 10m wind speed is raised to 100m using a 1/7 power law (100m winds are not available from the S2S database).

Before the energy variables are calculated, the hindcasts of 2m temperature and 100m wind speed are bias corrected using ERA5 as the reference. The ERA5 100m wind speed has itself been bias corrected first (see Section 5.2.1). A separate bias correction is applied for each hindcast start day of the year and each lead time. A leave-one-out scheme is employed in which each hindcast is omitted from the set of hindcasts used to calculate its correction parameters. There is both a mean correction and an optional variance inflation. The mean correction ensures that the hindcast mean agrees with the ERA5 mean. The variance inflation employs the scheme described by Doblas-Reyes et al. (2005), which ensures not only that the hindcast variance agrees with the ERA5 variance, but also that the correlation between the hindcast and ERA5 is preserved.

The bias corrected meteorological variables are converted to energy variables using the models described in Section 5.2.1 with the slight difference that the wind power capacity factor is converted to wind power at the end. The regression coefficients and turbine choices applied to the hindcasts are those derived from ERA5.

The ECMWF hindcasts are issued twice a week and have 11 ensemble members, whilst the NCEP hindcasts are issued daily and have 4 ensemble members. In this chapter, in order for a fair comparison between the ECMWF and NCEP hindcasts, a lagged NCEP ensemble is constructed at the energy variable level by combining the NCEP hindcasts initialised on ECMWF start days with the NCEP hindcasts initialised on the previous two days, giving a 12-

member ensemble. This method is used in order to mirror that used for the S2S4E demonstrator.

5.2.3 Weather regimes (WRs)

A modification to the method used in S2S4E Deliverable 3.2 is used to calculate the WRs in this section. Following the method of Cassou, 2008 the k-means clustering algorithm is applied to area-weighted, gridded, November-March daily-mean 500hPa geopotential height anomalies from 1980-2018 over the domain 90W-30E, 20N-80N to obtain four circulation types. The clustering is performed in empirical-orthogonal-function (EOF) phase space to speed up the computation (14 modes retained, corresponding to 90% of variance). The patterns calculated from ERA5 match those found in other studies (Cassou, 2008; Van Der Wiel et al., 2019) and are qualitatively similar to the regimes from D3.2 used in Chapter 4 (compare Figure 15 to Figure 12). This method has been chosen to have a consistent set of regimes from November-March, for easier computation and skill assessment of the forecasts.

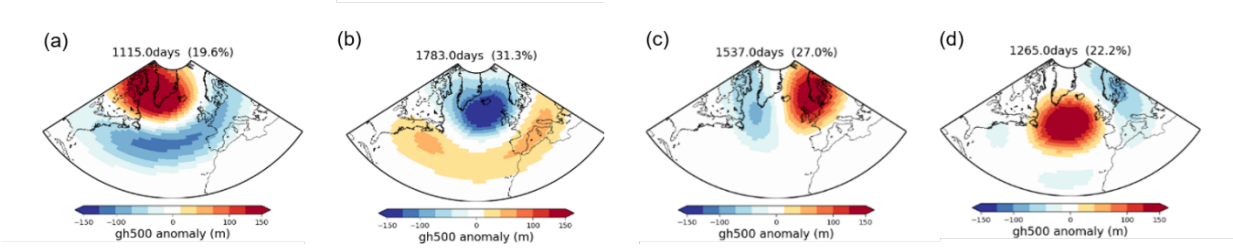

Figure 15. The weather regimes used in this study using the method of Cassou (2008). Panels: (a) the negative phase of the NAO (NAO-), (b) the positive phase of the NAO (NAO+), (c) Scandinavian blocking, and (d) Atlantic Ridge.

5.2.4 Targeted circulation types (TCTs)

The principle of the TCT method is described in Section 2.3.3. The difference between the methodology used to assign the TCT's between D3.2 and this deliverable is that in order to take advantage of the subseasonal models potential for increased predictability of large scale features (Robertson and Vitart, 2018) the TCTs from the hindcasts will be assigned based on the 500hpa geopotential height (Z500) pattern composites from ERA5 (i.e., the TCT Z500 patterns are first identified in the reanalysis and then these patterns are used as the basis to assign a given Z500 forecast into a particular TCT: this is done by determining which cluster centroid Z500 minimises the Euclidean distance to the Z500 circulation on a particular day). An example set of Z500 composites for the DNW TCTs are shown in Figure 16. The Z500 signals are much weaker for the TCT's than the signals seen for the weather regimes, with the main signal of the TCT patterns located over the European continent (compare Figure 15 and Figure 16).

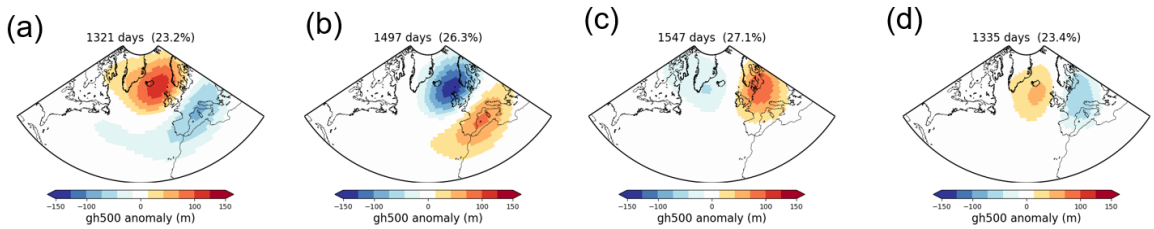

Figure 16. 500hPa geopotential height anomaly composites for the November-March DNW TCTs. Panels: (a) Blocked, (b) Zonal, (c) European High, and (d) European Trough.

As an initial test of the classification system, we first seek to recreate the TCT classification using the Z500 field from the ERA5 reanalysis. That is, the DNW data derived from ERA5 surface variables is first used to generate a set of TCTs and create the corresponding Z500 composites; ERA5 is then re-classified using Z500 applied to these composites. If the Z500 assignment process is able to "perfectly" capture the TCT assignment, then the re-assignment of day should match the original assignment.

Figure 17 shows the differences between the classifications of the DNW TCTs in ERA5 depending on whether the normalised DNW anomaly composites are used to assign the TCT, or whether the Z500 anomaly composites are used. The re-assignment of the TCTs achieves a decent performance but the overall success ratio (the percentage of correct assignments out of all assignments) is modest at 51%. This is important to note as it suggests that the Z500 circulation associated with a TCT provides only a modest constraint for TCT assignment, which might be improved through the addition of further large-scale circulation properties (e.g., temperature on 850hPa or mean sea level pressure). For this reason, the "perfect forecast" for which the hindcasts are compared to in the following subsections is the ERA5 data re-assigned to TCTs from the Z500 anomalies (rather than the original TCT assignment) in order to not unfairly penalise the hindcasts. Future work will look to find methods to improve this assignment success rate and is intended to form the basis of a publication.

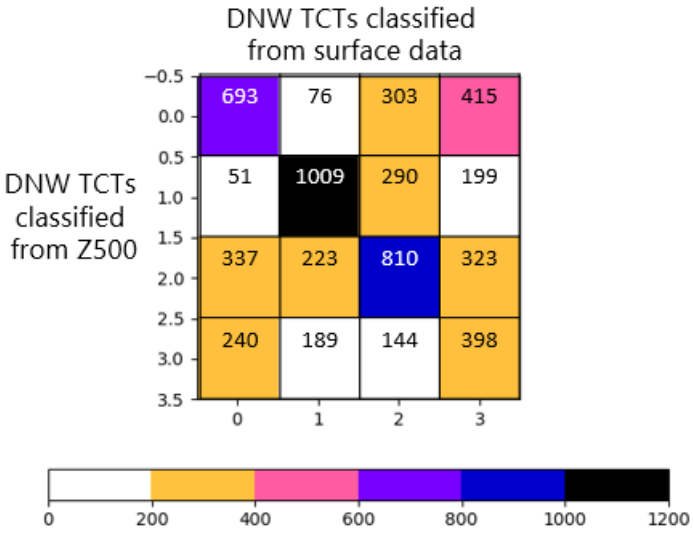

Figure 17. Hit rate table of the DNW TCTs classified from the composites of 500hPa geopotential height anomalies (Z500; columns) vs. DNW TCT's classified from the composites of normalised surface DNW anomalies (rows).

In summary, TCT's in the hindcasts are assigned as follows:

1. Calculate the DNW TCT patterns, as described in section 2.3.3. This results in four country-level DNW anomaly patterns.
2. Calculate the Z500 anomalies present during each of the DNW TCT's in ERA5. These are used to assign each hindcast day to a cluster.
3. Assign each hindcast days Z500 data, based on the ERA5 Z500 anomalies present during each of the DNW TCT's.

5.3 Impact regime frequencies in subseasonal forecasts

Figure 18a shows the frequency of occurrence of the WRs seen in ERA5, the hindcast ensemble members from ECMWF and NCEP for lead weeks 1 – 6 (note only 5 weeks are shown for NCEP due to the shorter hindcast length). The forecasts used are all available forecasts from the ECMWF hindcast in the period, and a corresponding twice-weekly lagged ensemble is constructed over a 3-day window from NCEP (see Section 2.2.4). In the ECMWF hindcast we see generally a good agreement in the regime classification with an under-representation of the NAO+ and Atlantic ridge regimes, and over-representation of NAO- regimes compared to ERA5, which is worsening with increasing lead time. These differences in frequency are similar but accentuated in the NCEP hindcast.

Figure 18b shows the corresponding frequency of occurrence but for the DNW TCTs. The zonal regime is represented well by both models, however both models tend to underestimate the occurrence of the Blocked regime and over-estimate the occurrence of the

European trough. The European high (which looks similar to Scandinavian blocking is well represented in the ECMWF hindcast.

Figure 18 (panels c and d) shows the frequency of occurrence of the regimes if the ensemble-mean is taken of the hindcast before the classification (i.e., in contrast to the upper panels which classify each ensemble member's Z500 to a weather pattern, the lower panels classify the ensemble mean Z500). A day is classified into a WR/TCT by finding the minimum of the Euclidean distances between the day and each pattern. Taking the ensemble-mean results in a much weaker Z500 anomaly to be classified than is seen in individual members, which leads to a preference for the patterns which have smaller anomalies (e.g., Figure 18d shows a preference for European High and European Trough which have weaker Z500 anomalies in Figure 16 panels c and d). This has consequences for how TCT-based forecast are generated (in particular the use of ensemble-averaging to increase the signal to noise ratio) and is discussed further in Section 5.5.

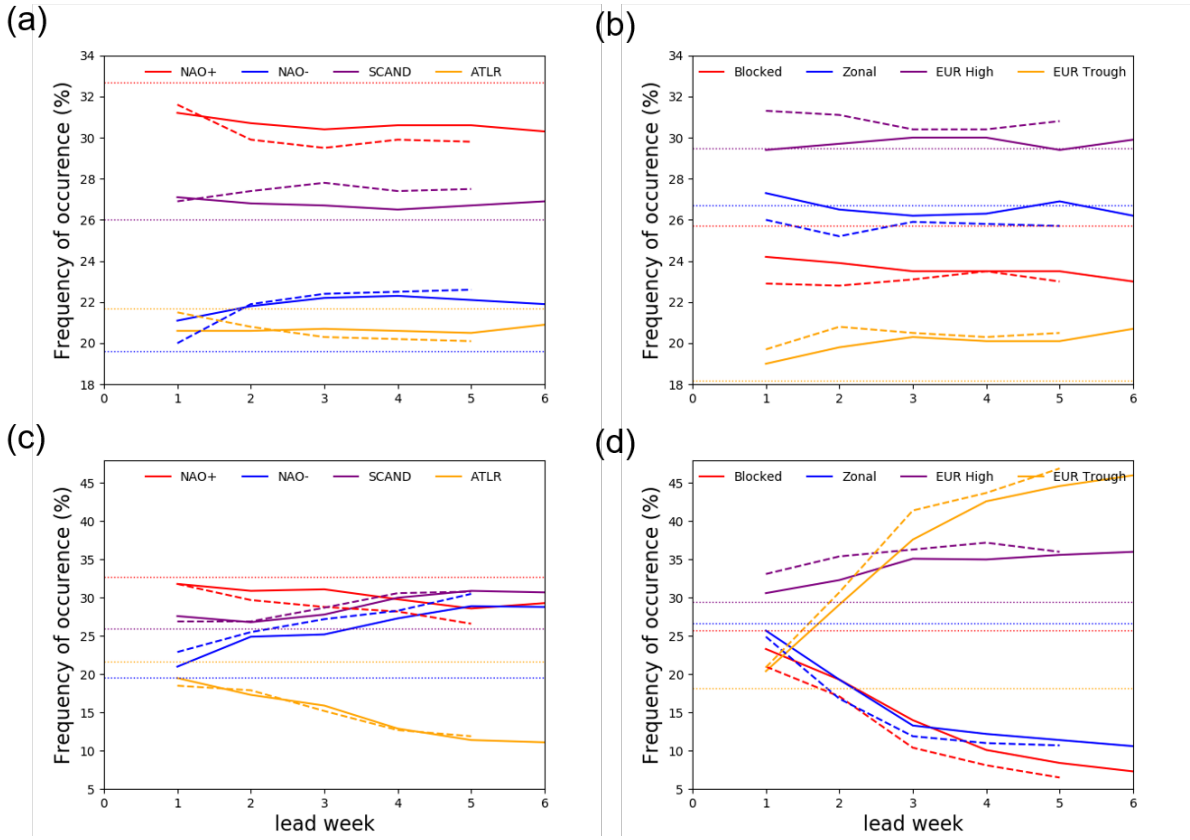

Figure 18. Frequency of Occurrence of Weather regimes (a and c) and Demand-net-wind TCTs (b and d) in ERA5 (dotted line) and with increasing forecast lead time for the ECMWF (solid lines) and NCEP hindcasts (dashed lines). Subplots (a-b) show results for all ensemble members and subplots (c-d) show the results when the ensemble-mean is taken before classification.

5.4 Impact regime connection to national energy indicators in subseasonal forecasts

In this section we examine the extent to which the ECMWF and NCEP hindcasts are able to reproduce the surface impacts from ERA5 which are associated with each of the WRs and DNW TCTs. Composites are made of the DNW for days within each regime (as classified by the Z500 field produced by the NWP forecast). The relationship is investigated for lead weeks 1-4 to see if there is any degradation with lead time.

Figure 19 shows the normalised DNW anomalies present during each of the weather regimes for ERA5 and the first four weeks of the ECMWF hindcast. We note that during the weather regimes the response of DNW is relatively small (i.e. max of 1 standard deviation anomaly from the mean), but it is captured well for all regimes during week 1. This is consistent with the forecast model is capturing the 2m temperature and 10m wind speeds responses at the surface (not shown). As the lead time increases the spatial pattern of the DNW response is well represented, however the magnitude of the patterns reduces and is too small compared to the observations. The NAO- regime is best represented at long lead times in the hindcasts. Figure 20 shows the equivalent results but for the NCEP hindcast, which are very similar to those in Figure 19 suggesting the results are robust to the choice of model.

Figure 21 shows the DNW anomalies during the DNW TCT patterns. Comparing subplots (a-d) with (e-h) shows the impact of the Z500-reclassification process on the resultant surface impacts. Perhaps surprisingly, in most cases, stronger surface anomalies seen in the Z500-reclassification, so for fair comparison the hindcast should be compared to subplots (e-h). In lead week 1 the forecasts are representing the DNW anomalies seen in the reanalysis well, however, even in lead week 1 the anomalies of the zonal pattern are too small (compare subplots j and f). With increasing lead time the spatial structure of the patterns is consistent but the magnitude is significantly reduced compared to reanalysis. Similar results are seen for the NCEP hindcast (not shown).

Figure 19 to Figure 21 show that the ECMWF and NCEP hindcasts capture the near-surface response associated with each the WRs and TCTs reasonably well, though at longer lead times the link between assigned circulation patterns and the surface impact degrades. This suggests that if the large-scale WRs and TCTs were assigned for the correct days (i.e. there is some predictability of the patterns), then "downscaling" to the surface impacts using the observed TCT-surface impact relationship should offer a forecast skill improvement over the use of raw surface variables forecasted by the same system. The ability of the NWP forecasts to predict WR and TCT occurrence is investigated in the next section.

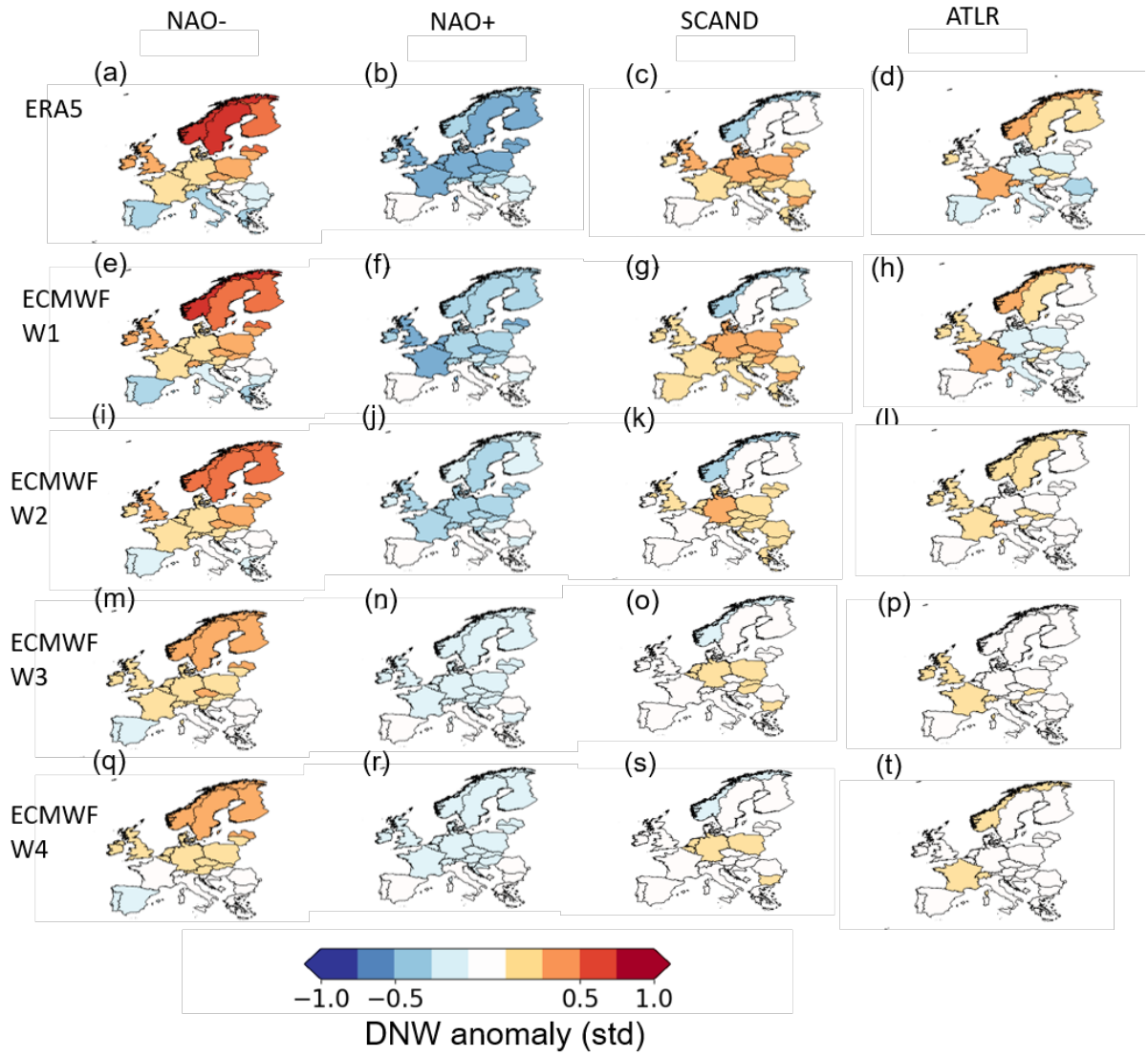

Figure 19. Normalised DNW anomalies during each of the weather regimes for ERA5 (a-d) and ECMWF hindcast week1 (e-h), week 2 (i-l), week 3 (m-p), and week 4 (q-t)

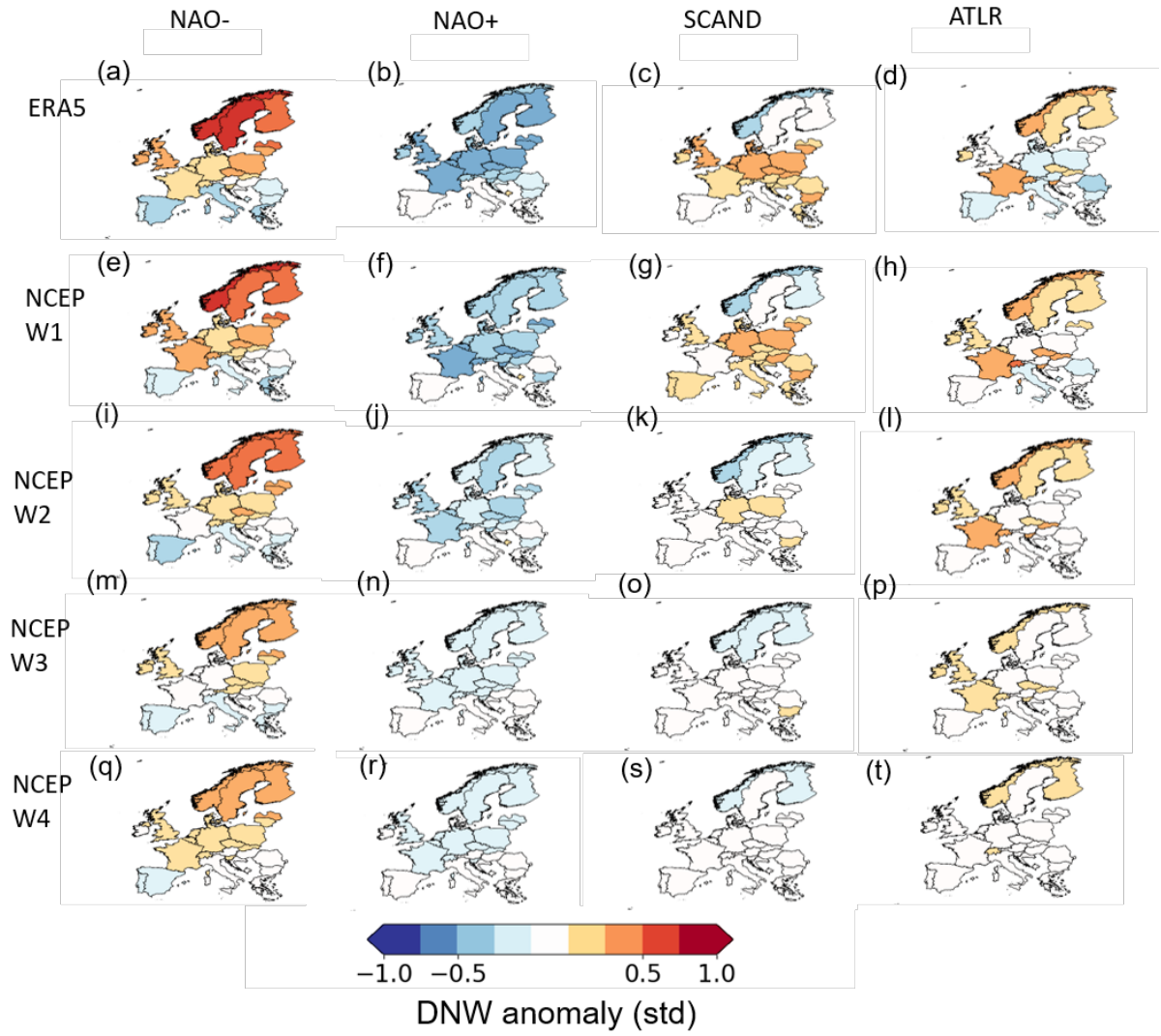

Figure 20. Normalised DNW anomalies during each of the weather regimes for ERA5 (a-d) and NCEP hindcast week1 (e-h), week 2 (i-l), week 3 (m-p), and week 4 (q-t).

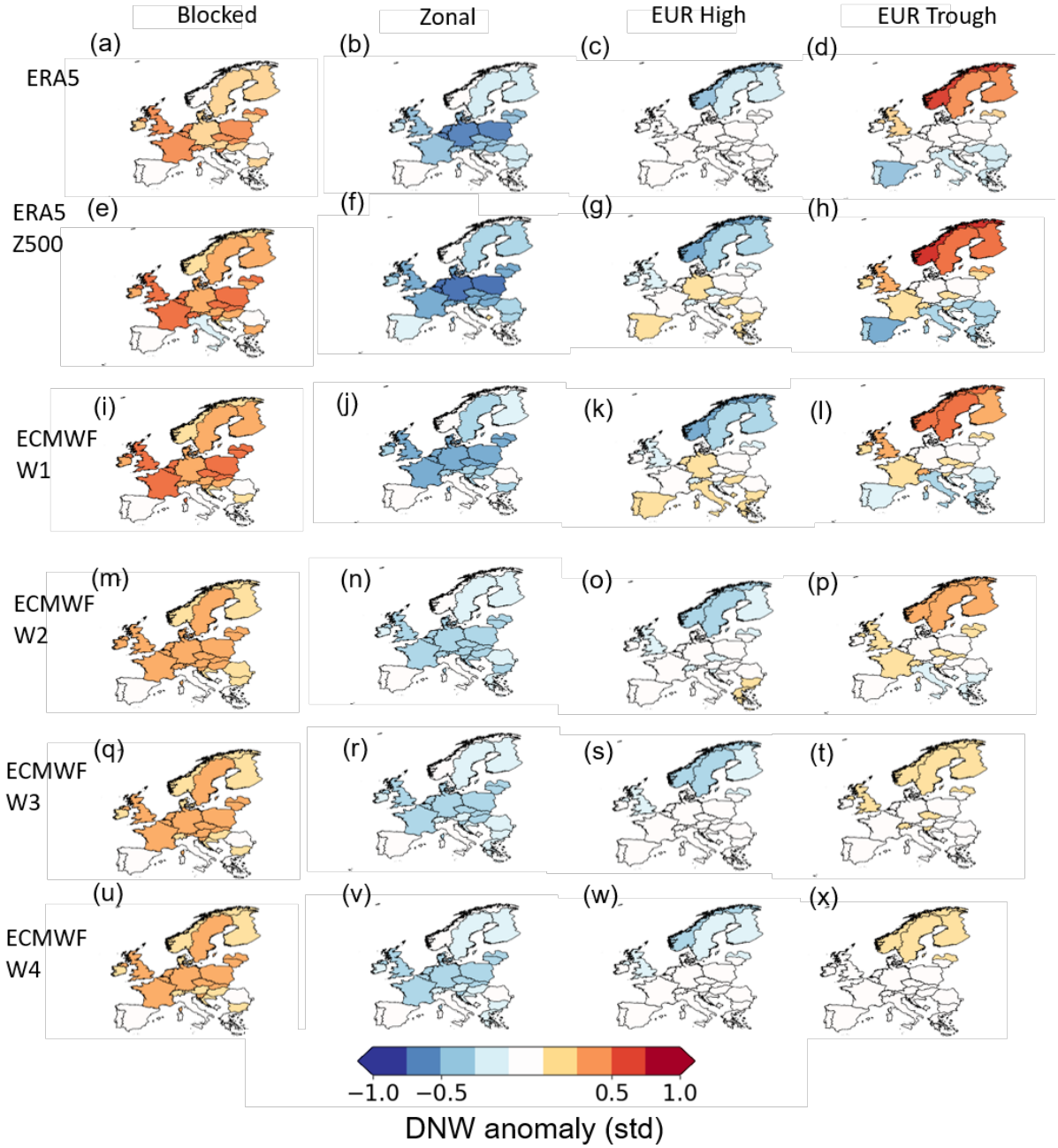

Figure 21. Normalised DNW anomalies during each of the DNW TCTs for ERA5 TCTs assigned using surface DNW data (a-d), ERA5 patterns assigned using 500hPa geopotential height data (e-h), and ECMWF hindcast week 1 (i=l), week 2 (m-p), week 3 (q-t), and week 4 (u-x).

5.5 Predictability of WR and TCTs in subseasonal forecasts

In order to assess the predictability of the weather regimes and TCTs in the hindcasts the success ratio of the daily WR assignment (Figure 22a) and DNW TCT (Figure 22b) for the ECMWF and NCEP hindcasts are shown. The success ratio is the percentage of correct assignments out of all assignments. In lead week 1 success ratios of ~60% are seen for the

WRs when averaged over both hindcasts with a reduction to ~40% in week 2 and ~30% in week 3 (Figure 22a). Similar results are seen for the DNW TCTs at all lead times, suggesting that both types of circulation pattern have similar levels of predictability from the NWP forecast output. For both TCT and WR methods, and all lead weeks, the ECMWF hindcast has a higher success ratio than the NCEP hindcast. However, the reduction in success ratio with increasing lead time is very similar in both of the hindcasts.

A common process for enhancing the signal-to-noise ratio in ensemble NWP is to take the ensemble average. Two methods of TCT and WR assignment are therefore shown: classifying each ensemble member, and taking the ensemble mean before assignment. Taking the ensemble mean before classifying the regime leads to an increased success ratio in weeks 1-4 for both hindcasts, though at longer lead times it has little benefit and leads to an overall frequency-of-occurrence bias that prefers the weaker patterns (EuTr and EuHi; Figure 18d).

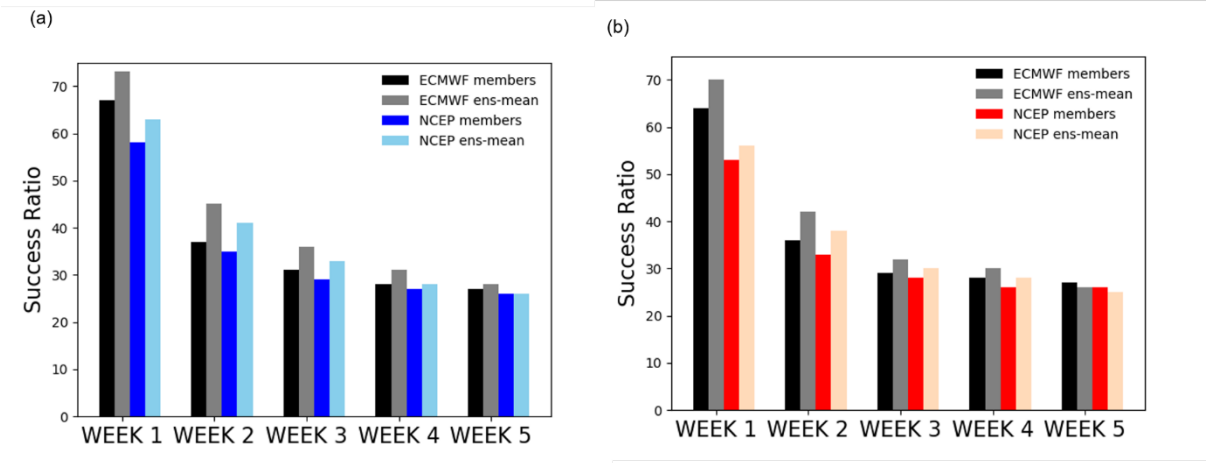

Figure 22. The success ratio between assignment of the Weather regimes (a) and DNW TCTs (b) for ECMWF (greys) and NCEP (blues) hindcasts. Dark bars show success ratios from assignments using all available ensemble members, with lighter bars showing assignments using the ensemble mean. A 25% occurrence in each regime would be expected by random chance.

5.6 Conclusions

In summary, we have found that the spatial structure of the surface DNW signals associated with the WRs and DNW TCTs are well represented in lead weeks 1-4 of the ECMWF and NCEP hindcasts. As expected, this confirms that the forecast models broadly preserve the observed fundamental relationships between surface impacts and large-scale circulation. However, the magnitude of the surface impact signal (associated with a given TCT circulation pattern) degrades with increasing lead time. This therefore suggests that, if it is believed that the current generation of NWP subseasonal forecasts offer more predictive skill in the large-scale circulation (compared to local surface variables), then using the NWP models to predict the large-scale circulation (TCT or WR) and then downscaling using the "observed" (i.e., ERA5) surface responses (to a TCT or WR) may offer opportunities for enhanced forecast skill.

The WRs and DNW TCTs have comparable frequency of occurrence to that seen in ERA5 if each ensemble member is used to classify the pattern. However, if the ensemble-mean Z500 field is used to classify the regime then some preferences are seen for patterns with weaker magnitudes at longer lead times.

In forecast week 1 success ratios of ~60% are seen for assigning regimes in the ECMWF and NCEP hindcasts (with ~5% more skill in the ECMWF hindcast than for NCEP). This reduces to ~40% in week 2. These results suggest that the WRs and DNW TCTs could both be used to forecast European country-level DNW, as the models can assign the correct pattern in week 1 (and to some extent week 2).

This work has highlighted that improved methods for WR/TCT assignment are perhaps required in order to increase the success ratio in weeks 2-4 and to enhance the extent to which assignment based on the large-scale circulation constrains the corresponding surface impact. Possible future work for this includes assignment based on both 500hPa geopotential height and a field lower in the atmosphere such as 850hPa temperature of mean-sea level pressure.

6 Hydrological weather regimes representation in seasonal forecast models

6.1 Introduction

The hydrological weather regimes (HWR) are weather regimes classified based on large-scale climate information and optimized with local hydrological information. They are designed to include specific information at local in addition to general circulation movement at large scale. In this study the HWRs are produced for a river basin of interest, the Ume River in the Northern Sweden. The observed HWRs have been produced based on reanalysis data, ERA-Interim (ERAI). They have shown usefulness in explaining rainfall variability in amount and frequency (see D3.2, Section 3.2.2 therein). Also, the HWRs have presented their skills in selecting the most relevant analogue years employed in common Ensemble streamflow prediction (ESP) approach using historical observations (see D3.2, Section 6.3.2 therein).

In this section we aim to assess the skill of ECMWF seasonal forecast model, ECMWF SEAS5, in reproducing the observed HWRs. Analyses are conducted in terms of occurrence frequency and persistence to see how skillful the ECMWF SEAS5 seasonal forecast system is in simulating the observed HWR. The set of re-forecasts is used here. It starts on the 1st of every month for the years since 1981 and have 25 ensemble members (see Table 1). The consequent influence of the HWRs on hydrological runoff forecast will be reported in D4.3.

6.2 Data and Methodology

The classification of daily HWR follows the methodology described in D3.2 (section 3.2.2 therein). Twelve daily HWR are classified using ERAI daily MSLP anomalies covering the Euro-Atlantic domain (27°-75°N, 100.5°W-73.5°E) for a period from 1981 to 2016. Each HWR is described with a fuzzy rule k, represented by a vector V(k) = ($v(1)^k$, $v(2)^k$, . . . , $v(i)^k$, i=1, n), where n stands for the number of locations (grid points) and $v(i)^k$ are the indices of membership functions corresponding to multi-grid locations.

ECMWF SEAS5 Daily-mean SLP (MSLP) is computed as an average of 12-hourly raw data (00, 12UTC on the same day and 00UTC on the next day). The climatology has been computed using all ensemble members for the same day of the year. The daily anomalies of MSLP covering the same domain and the same time period are consequently calculated for individual member and used as predictor to generate corresponding HWR time series.

The same fuzzy rules describing the observed HWRs are applied to each day in an ensemble consisting of different start date and different lead time. Based on the membership functions membership grades for a given time t and a given location i are computed and combined to obtain the degree of fulfilment (DOF) for each fuzzy rule. The rule with the highest DOF is selected as a HWR for a specific day. The composite maps of the HWRs based on the ECMWF SEAS5 are therefore expected to have similar spatial structures to those based on the reanalysis data, ERA Interim.

As monthly frequency occurrence of individual HWR is used as a criterion to identify the analogue years used in hydrological seasonal forecast, we therefore evaluate the skills of ECMWF SEAS5 by investigating its reproduction of monthly frequencies occurrence of individual HWRs in terms of inter-annual variability and bias in monthly average. A two-tailed t-test is used to assess the level of significance of correlation between statistics derived from seasonal forecast, ECMWF SEAS5, and statistics derived from reference data, ERAI. Persistence is also an important property when evaluating the weather regimes and its influence onto the local hydrology.

The assessments are conducted for each start time (i.e., individual month of the year) and each lead time (i.e., 0-6 months) for each HWR using all members available in hindcast mode (i.e., 25 members). We focus on the dominant HWRs (i.e., 02, 03, 06 and 12) identified in D3.2, (see Figure 23 in this document for a recap) which are the HWRs that are likely to cause extreme rainfall conditions.

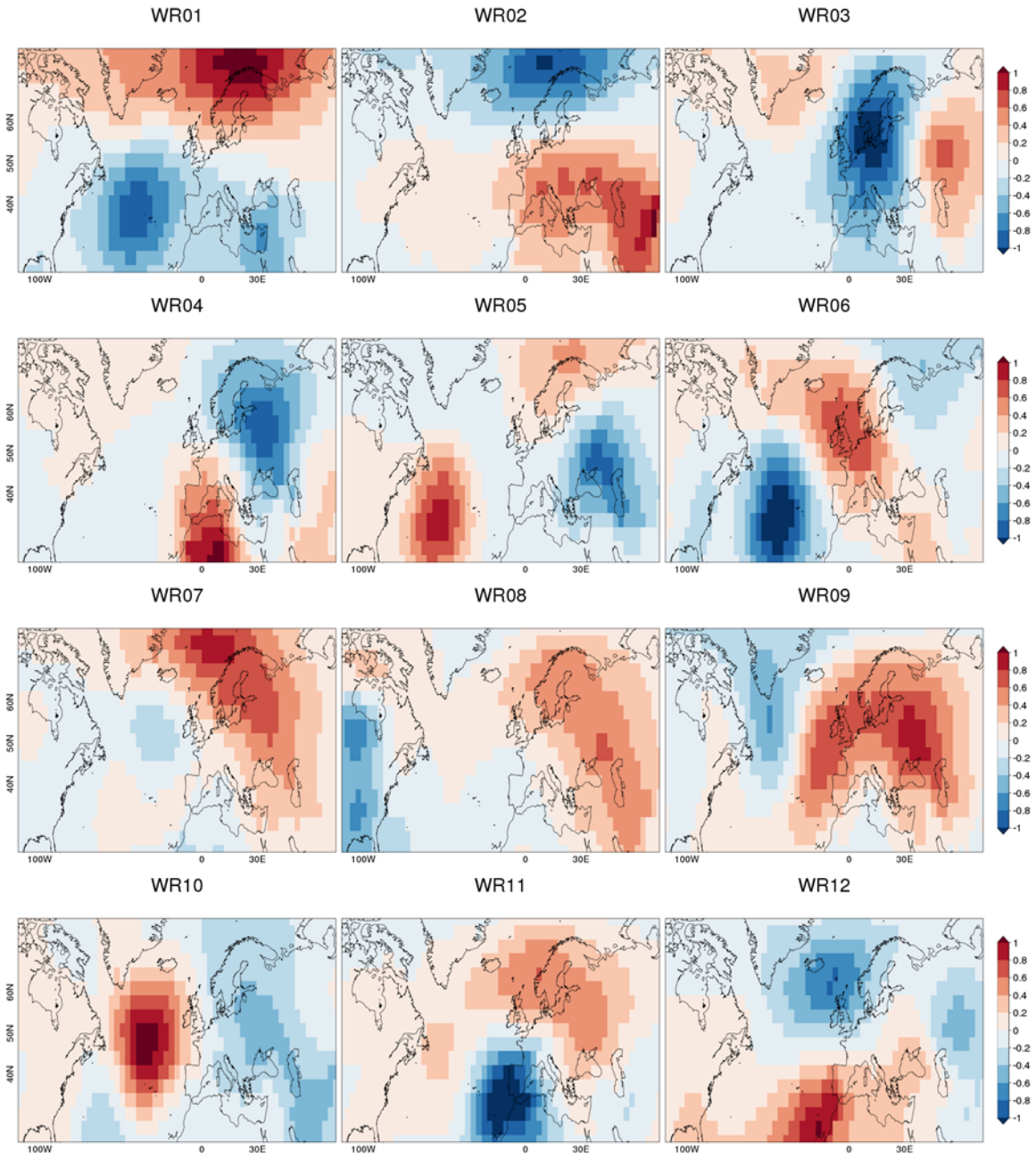

Figure 23. Anomaly maps of hydrological WRs derived from fuzzy classification using the MSLP during 1981-2016. Source: ERA-interim (see Deliverable D3.2 for details).

6.3 Results

To evaluate the capability of ECMWF SEAS 5 to reproduce the observed inter-annual monthly frequencies of HWRs, Pearson correlation is used as a measure to investigate the monthly frequency time series based on seasonal forecast data and reference data. It was calculated over all ensemble members.

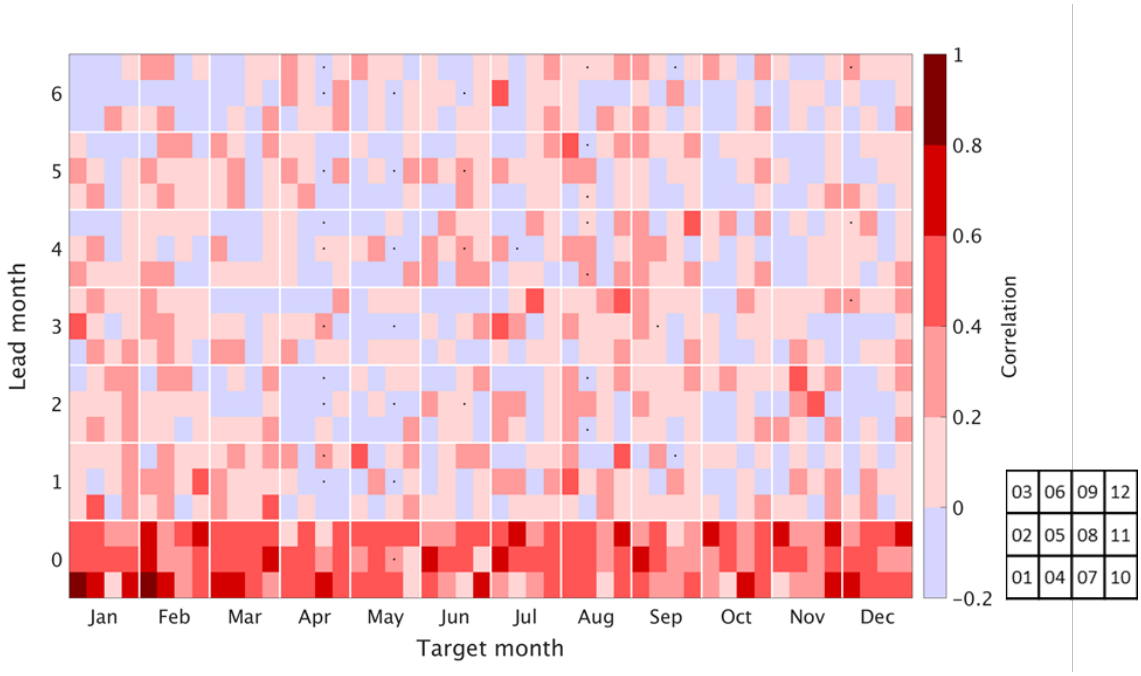

Figure 24. Pearson correlation between the ECMWF SEAS5 seasonal predictions and ERA-Interim reanalysis monthly frequency occurrence corresponding to the HWRs. The correlations in each block are shown as a function of the target month (x-axis) and the lead time (y-axis). Each sub-square represents the correlations for a specific HWR as indicated in the bottom right legend. Points denote non-significant correlation values (two-tailed t-test at a 95% confidence level).

Figure 24 shows the influence of lead time on different target months concerning different HWR. Positive and significant correlations appear in lead month 0 for all target months and for majority of HWRs. However, the capability of seasonal forecast system in reproducing the monthly frequencies drops abruptly with increase of lead time ahead of the target month. For majority of HWRs the correlation coefficients are reduced from 0.4~0.8 to -0.2~0.4. Non-significant correlations often appear with low and negative correlation coefficents. These occur predominantly in months Apr – Aug. In general, the representation of individual HWR in ECMWF SEAS5 appears random. No clear pattern is found..

The bias in monthly average frequency is evaluated for each HWR by comparing the difference between average monthly frequency occurrence based on ECMWF SEAS5 and ERAI. The differences pending on different HWRs are presented in Figure 25. The bias is in a range of -10% ~+10%. The bias pattern for lead month 0 shows clear difference from other lead months except summer season. For lead month 0 HWR02 is in general underpredicted by the seasonal forecast model in all target month except July and November. HWR03 is often overpredicted in winter and autumn except October, while is underpredicted in spring and summer except March. HWR06 is similar to HWR02, which is underpredicted in all target month except July-Sept. HWR12 is overpredicted in spring, while is underpredicted in all rest

target month. For lead month from 1 to 6 the bias patterns look similar for each target month. No clear change is found with increase of lead time.

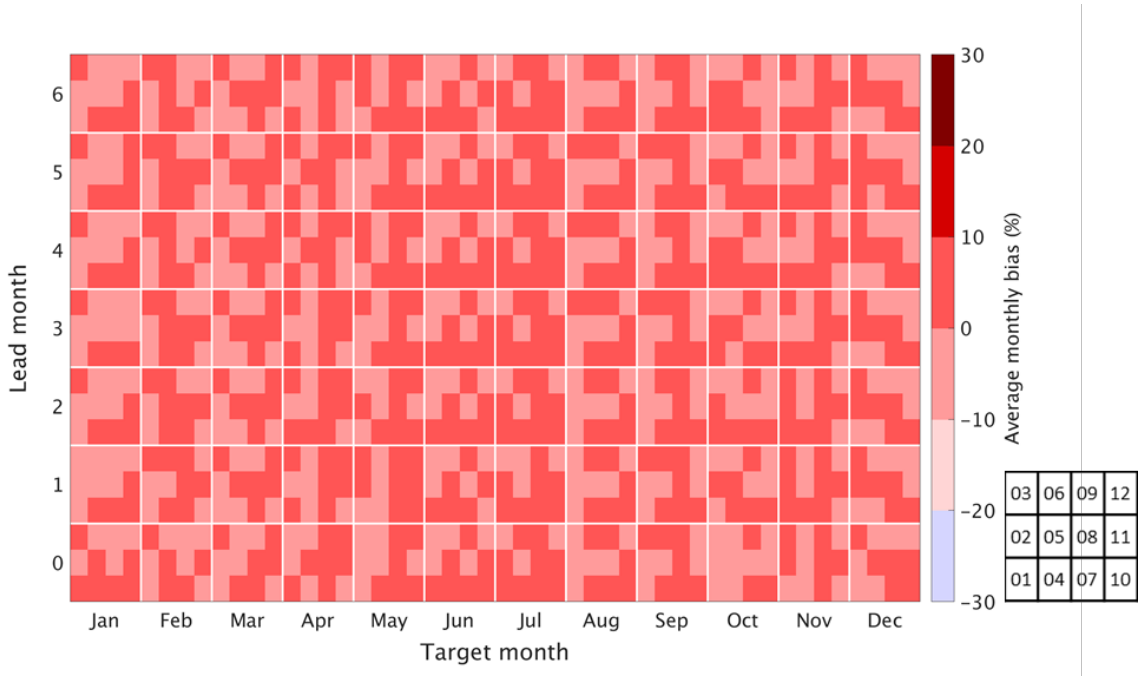

Figure 25. Bias in average monthly frequency between the ECMWF SEAS5 seasonal predictions and ERA-Interim reanalysis monthly frequencies occurrence corresponding to the HWRs. The correlations in each block are shown as a function of the target month (x-axis) and the lead time (y-axis). Each sub-square represents the correlations for a specific HWR as indicated in the bottom right legend.

The bias in average monthly persistence bias for individual HWR is calculated as the difference between the number of days based on seasonal forecast and those based on the reference data (in days). The influence from different lead time is shown in Figure 26. The bias for every target month is quite limited. The corresponding values are all located in a range of -0.4 to 0.8 day/month when considering different lead time and different HWR. No clear change is found with increase of lead time or with the change of season.

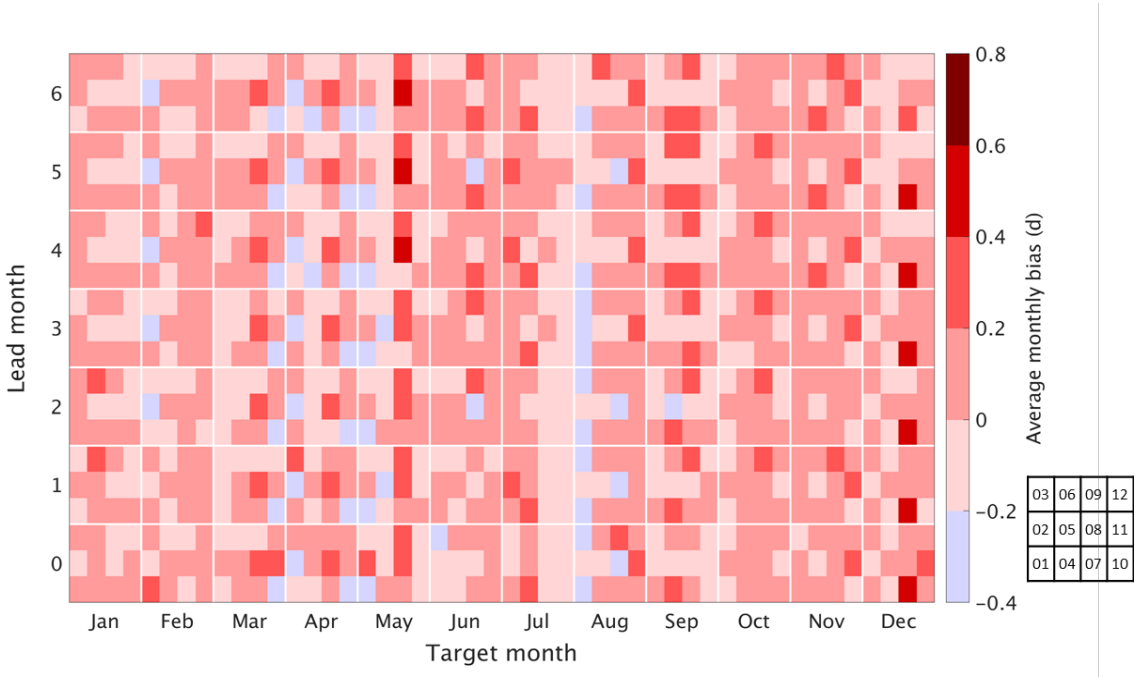

Figure 26. Bias in average monthly persistence between the ECMWF SEAS5 seasonal predictions and ERA-Interim reanalysis monthly frequencies occurrence corresponding to the HWRs. Each square represents the correlations for a HWR. The correlations are shown as a function of the target month (x-axis) and the lead time (y-axis).

6.4 Conclusions

The HWRs based on seasonal forecast, ECMWF SEAS5, are reproduced using optimized fuzzy rules based on ECMWF reanalysis product, ERAI. The spatial structure of HWRs based on seasonal forecast is highly correlated to those based on reference data (not shown here). The composite maps of individual HWR classified based on reference data can be find in D3.2 (section 3.2.2 therein).

Inter-annual variability and average bias of monthly frequency occurrence and monthly persistence of individual HWR are investigated.

The ECMWF SEAS5 shows capability in predicting the inter-annual monthly frequency occurrence for all target months when lead time is 0. With increase of lead time its predictability decreases quickly. The average monthly frequency is well reproduced by seasonal forecast model with bias of ±10%. The bias pattern for lead time 0 slightly differs from that for the other lead time. No dependence on the lead time 1-6 is found. Also, the seasonal forecast model shows skill in reproducing the residence time of HWRs with bias of -0.4 to 0.8 day/month.

7 Remote large-scale climate drivers

7.1 Introduction

A number of important sources of subseasonal predictability have been previously highlighted in the literature including: The Madden-Julian Oscillation, soil moisture memory, snow cover, ocean sea-surface temperature conditions and the interactions between the stratosphere and the troposphere (Robertson and Vitart 2018). In this section we focus on the stratosphere-troposphere interactions through analysing the impact of sudden stratospheric warmings (SSWs) on the European energy system and their ability to improve forecast skill by comparing the essential climate variable and energy indicator responses following an SSW for ERA5 (Section 7.2). Following this, the response of the weather regimes and DNW TCTs following an SSW are investigated for ERA5 (Section 7.3). Section 7.4 shows these results again for the ECMWF and NCEP hindcasts to see if the observed responses are well represented by the models over the common period (1999-2010).

7.2 Impact on European national energy balance indicators

The surface response of an SSW is usually felt 5-60 days after the event as the signal slowly propagates towards the surface (Butler et al., 2017). Maps of the near-surface, country-aggregate response 30 days (days 5-35) and 60 days (days 5-65) after an SSW for ERA5 are shown in Figure 27. Following an SSW there are below average near-surface temperatures and wind speeds in northern Europe, extending into parts of central Europe (Figure 27). These are accompanied by anomalously high demand (due to the cold temperatures) and anomalously high DNW (due to the cold temperatures and low wind speeds). In South East Europe above average 2m temperatures are seen with average 10m wind speeds, resulting in below average demand and DNW. In South West Europe average temperatures are seen with above average wind speeds, resulting in near-normal demand and DNW.

Analysis of the individual events shows that some SSW events are associated with significant warming in all countries and some with significant cooling (suggesting that tropospheric variability dominates over the SSW signal in many cases). There are, however, a number of events which resemble the composite in spatial structure. The 30-day mean and 60-day mean responses are similar for all of the fields analysed, showing the response is not particularly sensitive to the time frame used for the analysis. Previous studies have shown a link between SSW's and ENSO state (Ineson and Scaife, 2009, Butler and Polvani 2011, Richter et al., 2011). We find that SSW's during a La Niña event lead to slightly colder conditions over Europe than those during an El Niño (not shown). However, there are only 7 events in each category due to the small number of events in the observational period, with the events in the 1980's tending to result in much colder conditions over Europe than those from the late 1990's.

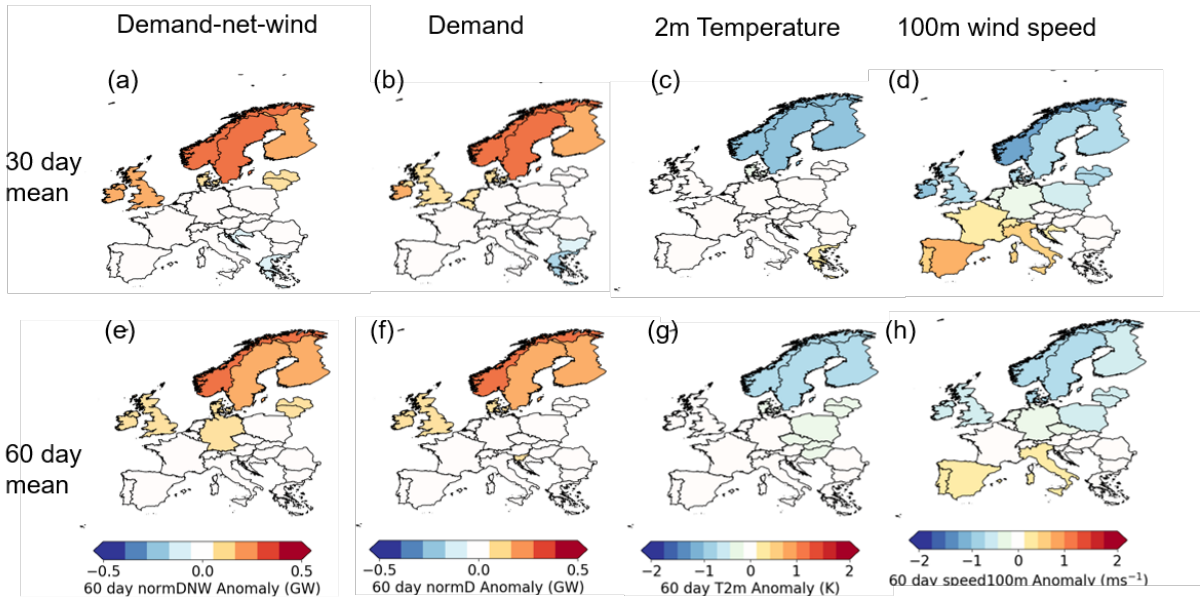

Figure 27. The 30day (a-d) and 60 day (e-h) mean response of ECV's and energy indicators to sudden stratospheric warmings compared to the November-March climatology.

7.3 WR and TCT frequencies

Figure 28a shows the dominant WR (i.e. the WR which occurs most frequently) after an SSW event for events within the full ERA5 period (1980-2013). Compared to the climatological regime occurrence (blue bars in Figure Figure 28a) the NAO- regime occurs much more frequently, with a reduced occurrence of the Atlantic Ridge as seen in Charlton-Perez et al. (2018). The NAO+ and Scandinavian blocking regimes occur in similar frequency to climatology.

Figure 28b shows the same results but for the DNW TCT patterns following an SSW. An increase in European Trough is seen in weeks 2-5 following an SSW, resulting in increased DNW over northern Europe (see Figure 21d) and a reduction in the European High is seen, which corresponds to reduced DNW in northern Europe (see Figure 21d)., the region where the response to SSW's is strongest in Figure 27. The results for the Zonal and Blocked TCTs following an SSW are inconclusive with varying increases/decreases in pattern frequency depending on the lead time, suggesting this link is not robust.

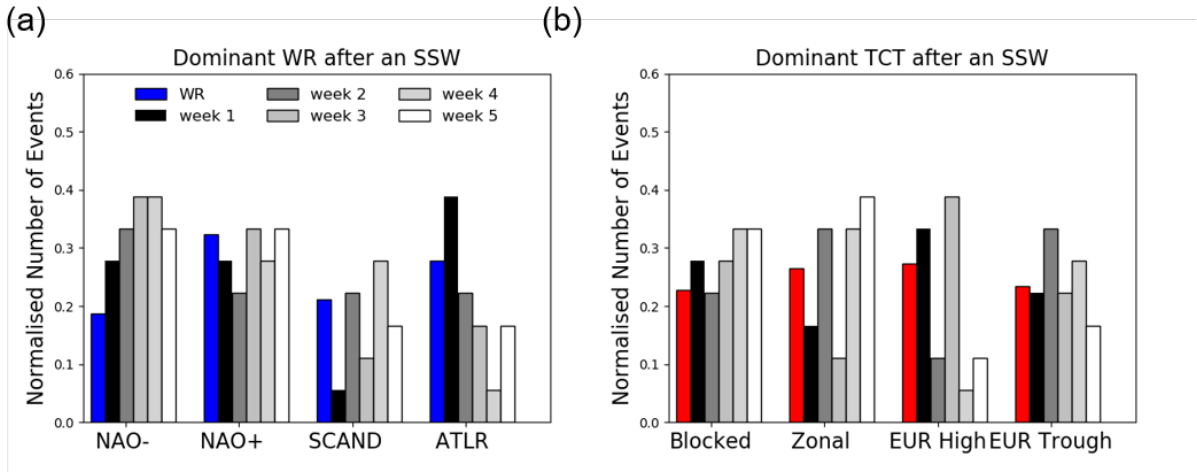

Figure 28. The frequency of occurrence of WRs (a) and DNW TCTs (b) in the weeks following an SSW in ERA5 (1980-2013). Dominant weekly regime is taken as the modal regime. Coloured bars are the November-March climatology for each WR/TCT.

7.4 Representation of SSW connections in subseasonal forecast systems

Figure 29 shows the response of demand and DNW following an SSW for ERA5 and the ECMWF and NCEP hindcasts. In order to calculate this for the hindcasts, the first hindcast launched on either the initial day of an SSW (or the next available forecast) is taken, and days 5-32 from the hindcast are used for analysis. For a fair comparison the ERA5 responses are restricted to those from within the common hindcast period (1999-2010) resulting in 12 events. The spatial structure of the composites in Figure 28 (a, d) look very similar to those in Figure 27 with all the SSW's from 1980-2013 included, albeit with a weaker magnitude of the response and a few countries, such as Spain, have a different sign of response. The ECMWF hindcast captures the spatial structure of the demand response following an SSW well in central and northern Europe though the magnitude is however a bit weak, perhaps associated with taking the average over all ensemble members' responses. The NCEP hindcast has lower demands in Southern Europe following an SSW than ERA5. Both hindcast models capture the response of DNW following an SSW well in central and northern Europe, particularly the increased DNW in northern Europe. However, they do not capture the reduced DNW in Spain. These results suggest the potential for increased predictability after an SSW in northern Europe, due to the hindcasts capturing the response seen in the reanalysis well.

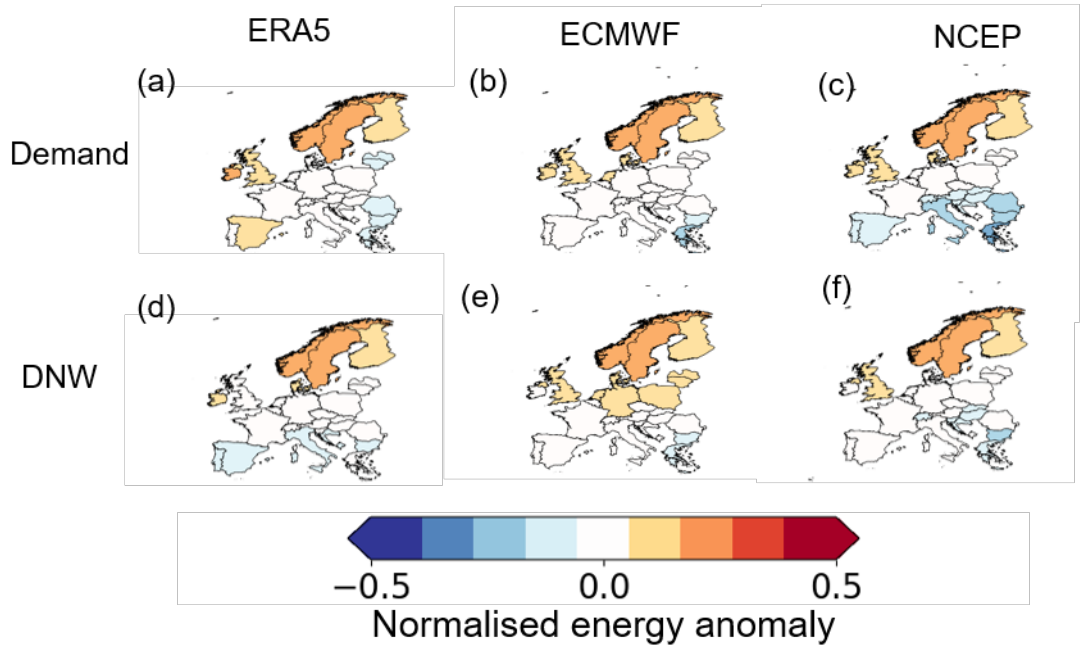

Figure 29. The 28 day mean response of demand (a-c) and DNW (d-f) following an SSW compared to the reanalysis' or models' own November-March climatology. The first column shows the ERA5 response during the common hindcast period (1999-2010), with the ECMWF and NCEP hindcasts in the middle and right columns respectively.

Figure 30 compares the dominant WRs and DNW TCTs found after a SSW between ERA5 and the two hindcast models over the common hindcast period (1999-2010). The first point to note it that the results seen in Figure 30b for the ERA5 DNW TCTs restricted period look very similar to Figure 28b, which showed all of the available SSWs, with increases in European Trough and reductions in European High. However, for the weather regimes we no longer see an increase in NAO- events but instead see an increase in NAO+. It is therefore worth emphasising that the sample sizes are very small and the results presented are likely to be very sensitive to individual events in the observational record.

In the ECMWF and NCEP hindcasts a general increase in NAO- and NAO+ is seen following SSWs with a reduction in the Atlantic Ridge WR (Figure 30, first column). For the DNW TCTs an increase in European Trough is seen in both hindcasts, and a reduction in the Zonal and Blocked patterns (with the Blocked TCT response contrasting to what is seen in observations). The DNW TCT responses after an SSW in the hindcast models are stronger than seen in the reanalysis, with some differences between the ERA5 response following an SSW and the hindcast (particularly for the Blocked TCT pattern). However, there may the potential for increased predictability following these events.

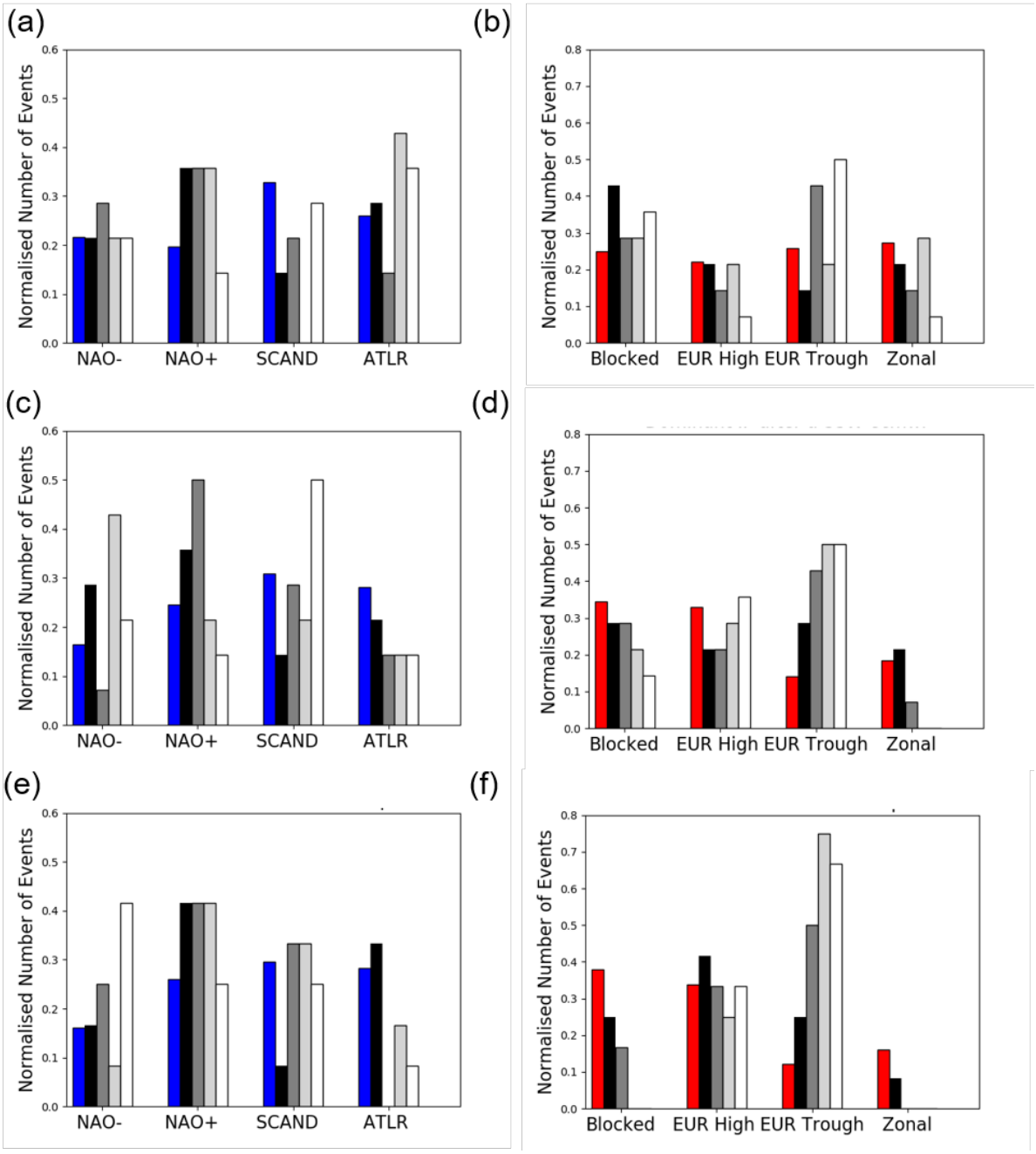

Figure 30. The frequency of occurrence of WRs (left) and demand-net-wind TCTs (right) in the weeks following an SSW. Dominant weekly regime is taken as the modal regime. Coloured bars are the November-March climatology for each TCT. (a-b) ERA5 over the common hindcast period (c-d) ECMWF hindcast (e-f) NCEP hindcast.

7.5 Conclusions

In this section we have shown that following an SSW, northern Europe experiences anomalously cold temperatures and low wind speeds resulting in increased demand and DNW compared to climatology. The spatial structure of these results are well reproduced by the ECMWF and NCEP hindcasts, albeit with a reduced magnitude.

In the observations an increase in the NAO- weather regime and the European Trough TCT (which are typically likely to be associated with increased DNW across Europe) are found following an SSW. A reduction in the Atlantic ridge (little relation to European energy variables) and the European High TCT (which results in high temperatures and low DNW over Scandinavia, Figure 28) are seen following an SSW suggesting the potential for increased predictability over Northern Europe following an SSW. These results are, however, not seen in the restricted ERA5 period period (1999-2010) suggesting that in this period (Figure 30) there is not the potential for increased predictability. It is, however, emphasised that this analysis is based on a very small sample size and is therefore subject to considerable uncertainty.

8 Winter time EATC patterns and their decadal variability

8.1 Introduction

In this section, we focus on the long-term variability of the Euro-Atlantic teleconnections (EATCs) for the December-January-February (DJF) period, the skill of seasonal forecast models in predicting them, and how they relate to extreme temperature indices. A further question is whether the seasonal forecast model is able to capture this relationship between large-scale teleconnection patterns and extreme temperatures, and whether this relationship changes over time. The motivation for this work comes from studies such as Shi et al. (2015) and Weisheimer et al. (2017) which showed that even older seasonal forecast models had skill in predicting the NAO in recent decades, but not in the middle of the 20th century. While several current operational seasonal forecast systems report skillful prediction of the NAO (e.g. Scaife et al., 2014), the question is whether they would also have skill during periods where the NAO is more negative, and the intraseasonal Z500 variance is higher, as was the case in the middle of the 20th century (Weisheimer et al., 2017), and might happen again in the future.

8.2 Decadal variability and skill

Prior to discussing the evolution of the forecast skill in predicting EATCs in the ASF-20C reforecasts, we first review the findings of D3.2 (Section 3.1.2 therein) where the time evolution of the 4 main EATCs in the ERA-20C reanalysis (Poli et al. 2013, 2015) and the 10 member ensemble of the CERA-20C reanalysis (Laloyaux et al. 2018) for the 1900-2010 period was presented. A key result from D3.2 was that while the evolution of the EATC was very similar in the common period for both reanalysis datasets, there were discrepancies in the 4 spatial patterns revealed by the ERA-20C/CERA-20C datasets compared to the ERA-Interim dataset.

Subsequent investigations have shown that EATC patterns produced by the rotated EOF method (See Chapter 3) is rather sensitive to the years considered (rather than the underlying dataset itself). For the 1981-2010 period, the EATC patterns derived from both ERA-Interim and ERA-20C were very similar, but both were different to the EATCs patterns derived from ERA-Interim using the period 1981-2016. For this reason, to study the forecast skill in the long seasonal hindcast dataset ASF-20C (Weisheimer et al., 2017), the forecast is projected onto a pre-defined set of EATC patterns (derived from ERA-Interim using the 1981-2016 period, similar to the patterns shown in Figure 1) rather than re-calculating the EATCs for the longer base-line period of ASF-20C. For this reason, to study the forecast skill in the long seasonal hindcast dataset ASF-20C (Weisheimer et al., 2017), the forecast for the whole 1900-2010 period is projected onto a pre-defined set of EATC patterns (derived from ERA-Interim using the 1981-2016 period, similar to the patterns shown in Figure 1) rather than re-calculating the EATCs for the longer base-line period of ASF-20C.

Figure 31 shows the time evolution of the four main patterns (all in their positive phase): North Atlantic Oscillation (NAO), East Atlantic (EA), East Atlantic/Western Russia (EA/WR) and Scandinavia (SCA) for DJF for 1901-2010 in ERA-20C and the ensemble mean forecast launched in November of ASF-20C. DJF 1901 means December 1900 and January and February 1901.

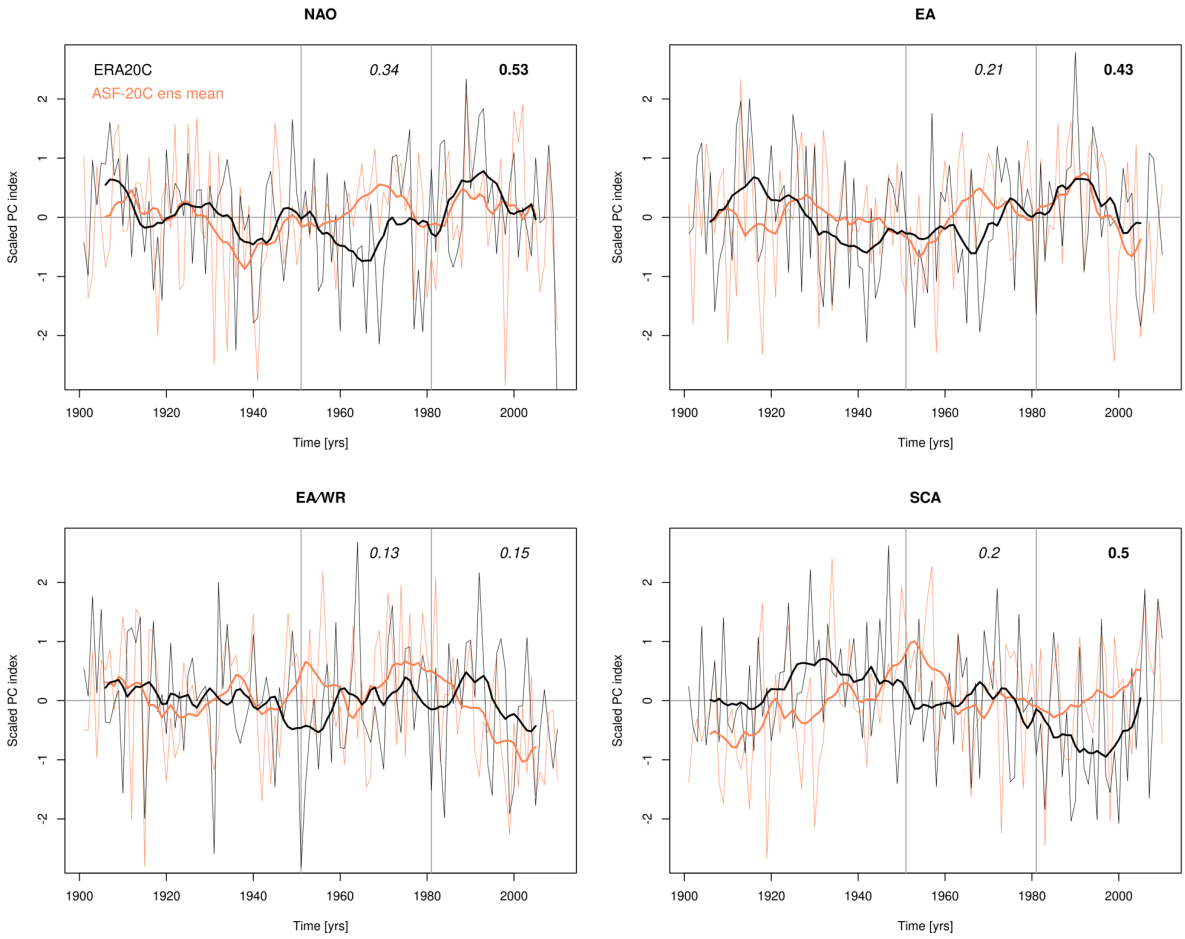

Figure 31. Time series of the 4 main EATC patterns in ERA-20C (black) and the ensemble mean of ASF-20C (coral). The thin lines represented the scaled and centered time series and the thick lines represent the scaled and centered time series with a 10yrs smoothing filter. The vertical lines indicate the two periods of interest, 1951-1980 and 1981-2010. The numbers on the top are the Pearson's correlation coefficients between ERA-20C and the ensemble mean of ASF-20C for each period, numbers in italic are non-significant on the 5% level, while those in bold are.

There is decadal variability in all 4 EATCs, and the ASF-20C (the ensemble average as well as the individual ensemble members, not shown) appears able to reproduce this variability. We focus the rest of this analysis on two periods that show different phases of the NAO: 1951-1980 with more negative NAO and 1981-2010 with more positive NAO. These two periods also coincide with mostly negative and positive EA phases respectively, and neutral and

negative SCA respectively. These two periods however do not coincide well with different phases of the EA/WR, but this pattern is the one explaining the least of the variance (12.3% compared to 34.4% for NAO, 18.3% for EA and 14.7% for SCA[1]), and appears to be an outlier in the rest of the analysis as well. The modelled time series for the NAO, EA and SCA further correlate significantly with ERA-20C for the 1981-2010 period, but not for the 1951-1980 period. As noted in Weisheimer et al. (2017), the significant correlations cannot be explained simply by the availability of better observation data since there are clear periods with significant correlations in the 1920s and 1930s, as shown in Figure 32.

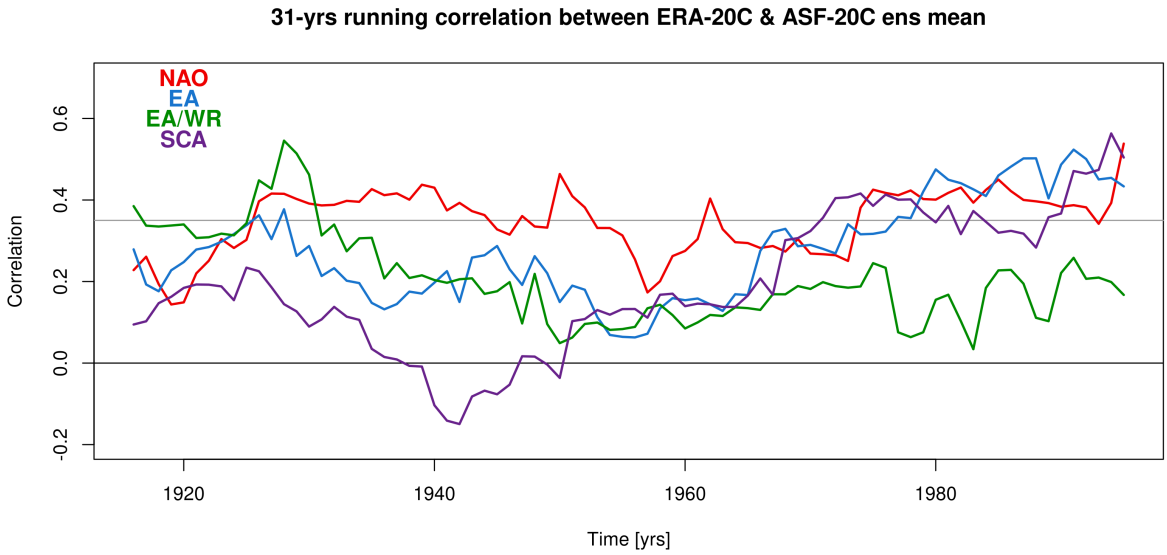

Figure 32. Time series of the 31-yrs running mean Pearson's correlation coefficients for each of the four main EATC patterns between ERA-20C and the ensemble mean of ASF-20C. The light grey line shows the t-test 95% significance level of the correlations.

Figure 32 shows the 31-yrs running mean correlation between the ASF-20C ensemble mean and ERA-20C to illustrate that the significant and non-significant correlations in the two chosen periods are not occurring by chance. The EA/WR is the only EATC that appears to have significant correlations in the 1930s, but not in the recent period. However, as mentioned above, this EATC seems to be an outlier and it could be that the EA/WR pattern in ERA-Interim, onto which ERA-20C and ASF-20C anomalies were projected, does not match well with the variability in ERA-20C and ASF-20C.

This analysis therefore suggests that the strength of the atmospheric circulation's projection onto the modern-day EATCs varies substantially over the 20th Century. Moreover, while the ASF-20C reforecasts appear able to represent the decadal-scale variability in the projection of the atmospheric circulation onto modern-day EATC patterns (e.g., the NAO defined as a REOF on 1981-2016), it is less clear that the patterns themselves are consistent over longer

[1] Note that these calculations of explained variance refer to a different period than those reported in Chapter 3.

timescales. The skill in seasonal forecasts of EATC indices defined in this way seems to vary somewhat on decadal timescales, with some periods (and patterns) demonstrating more skill than others. This may, in part, be indicative of periods in time which are inherently more (or less) predictable, but the extent to which this is also linked to the evolving nature of the EATCs themselves is unclear. Overall, this analysis suggests that caution needs to be taken in evaluating the long-term performance of REOF-based pattern predictions in seasonal forecasts: quantitative patterns and forecast skill evaluations derived over a relatively modest "modern" period (circa 20-30 years) may or may not be representative over longer periods.

8.3 Relationship between temperature extremes and EATCs in the two chosen periods

The 4 EATCs have known climate impacts in Europe (e.g., van der Weil et al., 2019). Thus, in addition to assessing the ability of forecasts to predict the EATC state, a key concern is the extent to which the occurrence of a particular EATC state is correctly linked to surface weather (e.g., is the NAO- pattern in the ASF-20C forecast linked to the correct set of surface weather impacts over Europe).

We are here using an adapted version of the daily Heat Wave Magnitude Index (HWMId) (Russo et al. 2015) adapted to capture cold extremes, i.e. the daily Cold Wave Magnitude Index (CWMId) (Brunner et al. 2018). However, due to the relatively narrow validity time window of the forecasts in the ASF-20C, i.e. months 1-3, the 30-day window to calculate the daily 90th percentile could not reasonably be applied. Instead we use a 10-day window for DJF, with the first and last 5 days of the season having consequently less robust statistics but this does not affect the main conclusions. At each grid point, each day is classified either as a cold extreme (obtains a value of 1), or as a "normal" day (obtains a value of 0). Then for each DJF period, we obtain the sum of cold extremes at each grid point. The Spearman's rank correlation coefficient between this timeseries and the patterns indices is then calculated. Results for the CWMId correlate very well with the sum of DJF heating degrees (not shown, but correlation coefficients significant at the 1% level for almost every grid point in Europe) and therefore relates to energy demand (see, e.g., Deliverable D3.2).

Figure 33 and Figure 34 show the correlation between each of the EATC's index DJF average value and the sum of extreme cold days in ERA-20C for the 1951-1980 and 1981-2010 period respectively.

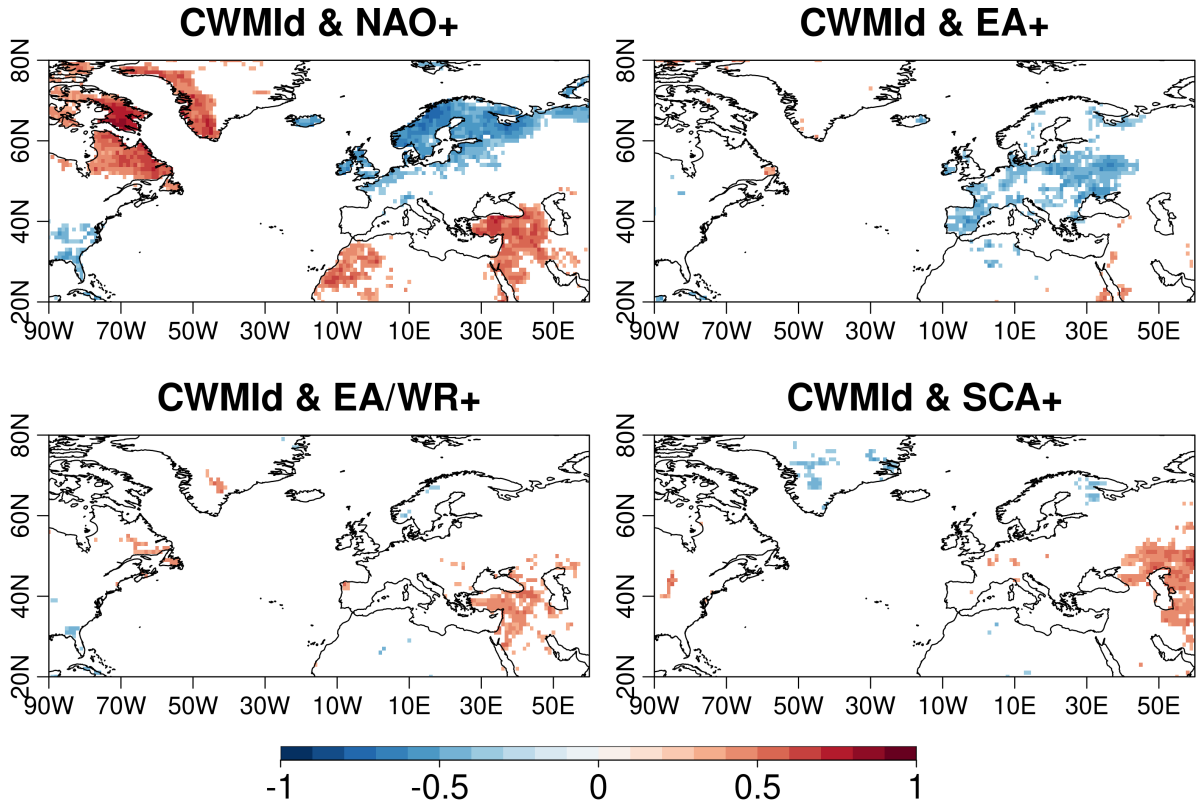

Figure 33. Maps of the Spearman's rank correlation coefficient between the teleconnection index of each of the four main patterns in ERA-20C and the sum of cold wave days during DJF for the period 1951-1980. Only correlation coefficients significant at the 5% level are shown.

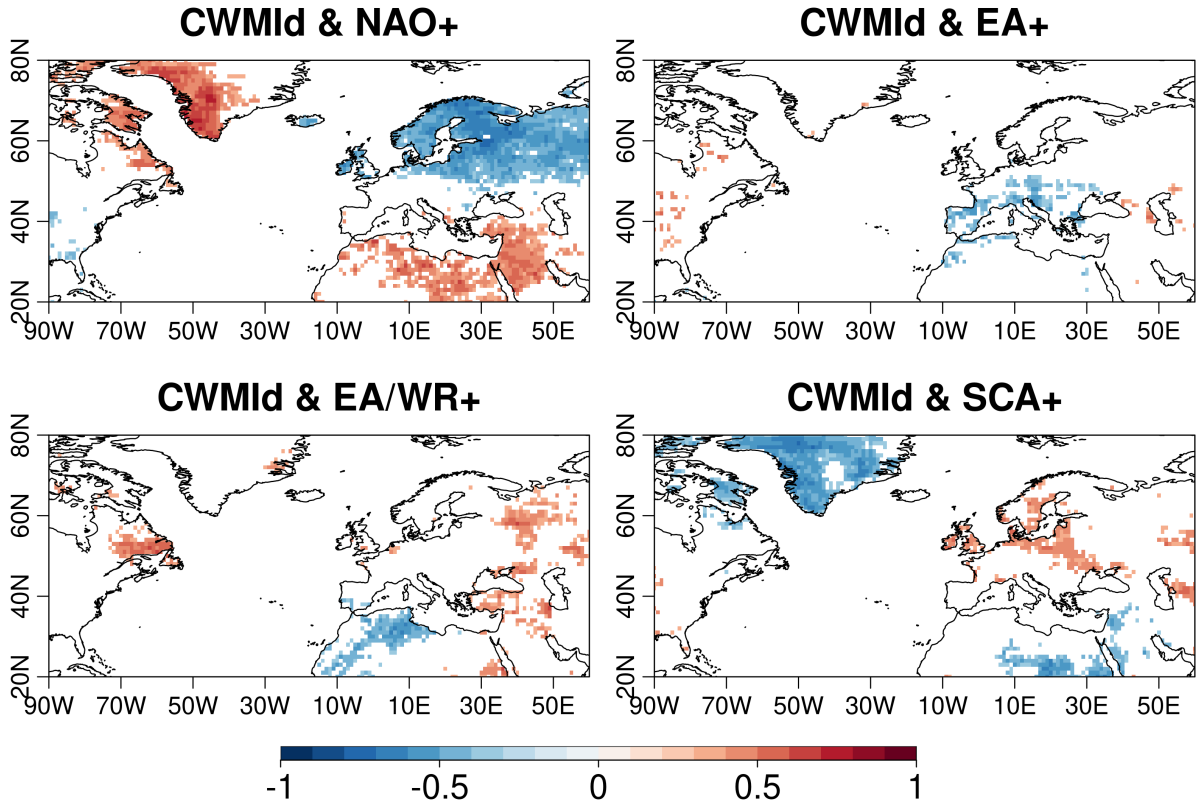

Figure 34. As for Figure 33 but for the 1981-2010 period in ERA-20C.

The positive phase of NAO is anti-correlated with cold extremes in Northern Europe, and correlated with more cold extremes south of the Mediterranean in both periods. A larger extent of western Siberia seems to show significant negative correlations in the earlier period, however it is not clear whether this has a real physical meaning. The positive phase of the EA is also associated with less cold extremes for large parts of Europe in both periods, in line with the patterns derived in D3.2. In the 1951-1980 period, the positive phase of SCA is associated with more cold extremes in Northern and eastern Europe, but there is almost no significant correlation for the later period, which could be due to the fact that the SCA was mostly negative during the 1981-2010. The EA/WR pattern appears not the correlate with cold extremes in ERA-20C.

These correlation maps were also calculated for each of the 51 ensemble members of the ASF-20C, and in Figure 35 and Figure 36, we present the correlation between the ensemble average EATCs' index time series and the average sum of cold extreme days.

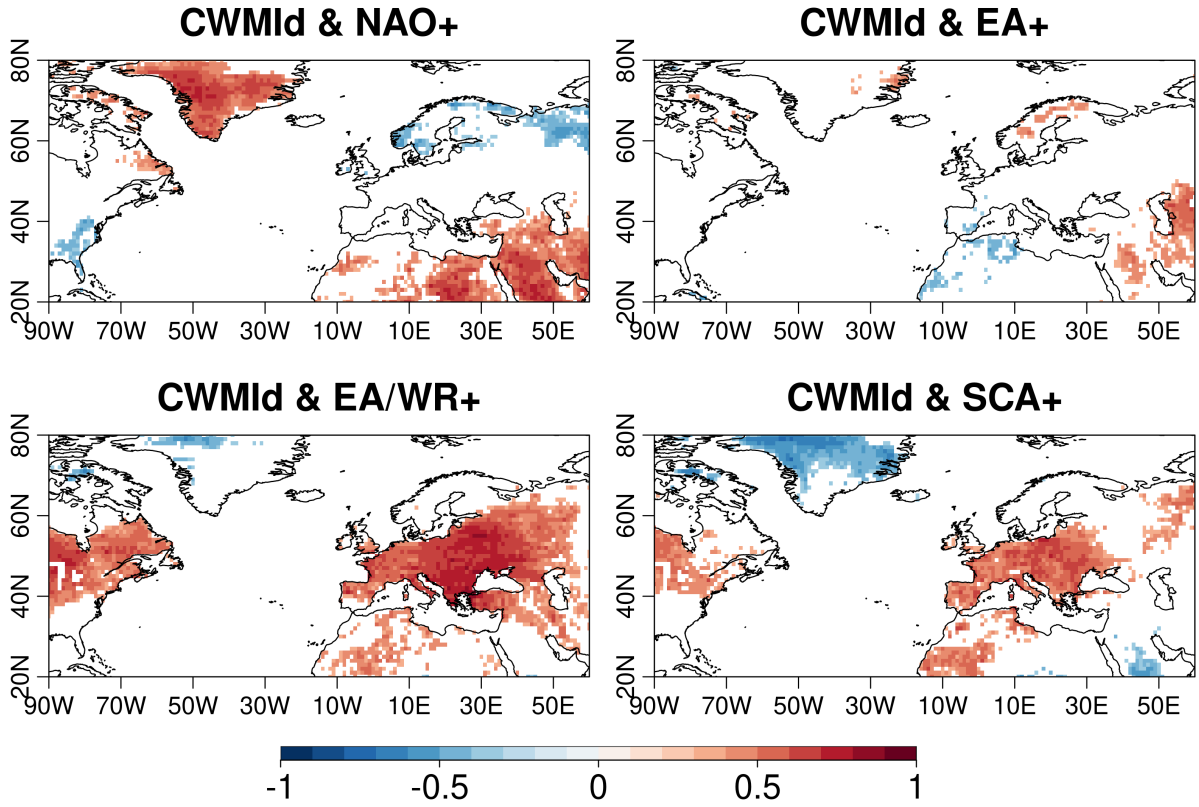

Figure 35. Maps of the Spearman's rank correlation coefficient between the teleconnection index of each of the four main patterns in the ASF-20C ensemble average and the ensemble average of the sum of cold wave days during DJF for the period 1951-1980. Only correlation coefficients significant at the 5% level are shown.

From looking at the correlation maps in each ensemble member, it seems that some members show similar relationships between extreme cold days and the EATCs' indices, while others do not. It is therefore not clear how well the model is able to reproduce the relationship. The ensemble average though seems to partly capture the negative and positive correlations in Northern and Southern Europe between the NAO and cold extremes and shows a stronger positive relationship between cold extremes and positive phases of the SCA in central Europe compared to ERA-20C. There is however little correlation between cold extremes and the EA in the model average, but a suspiciously strong correlation between the EA/WR and cold extremes for large parts of Europe that will require further investigation.

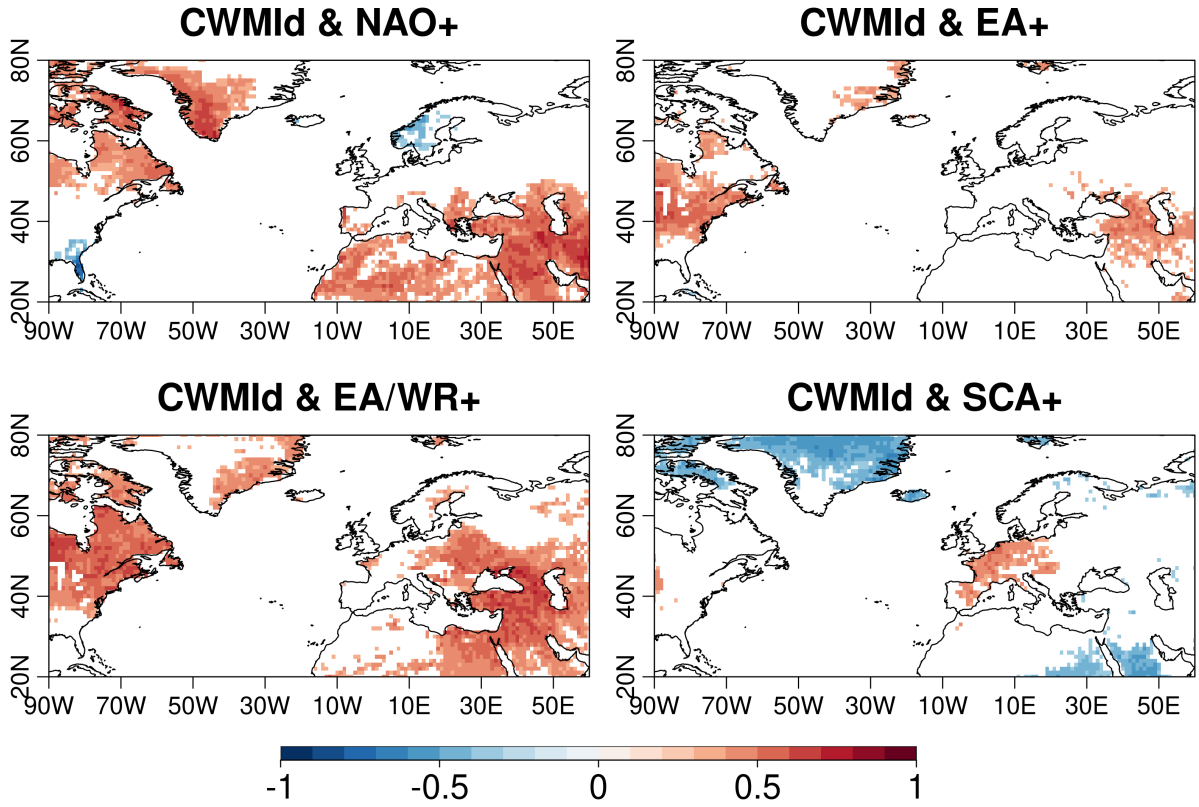

Figure 36. As Figure 35 but for the 1981-2010 period.

8.4 Conclusions

All 4 EATCs show decadal variability for DJF in ERA-20C and also the ASF-20C ensemble. The NAO and EA are often in similar phases throughout the 20th century. The skill of the ASF-20C to predict the EATC is estimated by computing a 31-yr running window correlation during the modelled period, 1900-2010. ASF-20C shows significant correlations for the NAO, EA and SCA in recent period, and partly towards the beginning of the century, while generally, correlations are non-significant in the middle of the century, similar to what Weisheimer et al. (2017) described for the NAO. It therefore seems that since the 1980s, the EuroAtlantic region has been in a relatively predictable state, but it might be that in the future, it will again be in a less predictable state, which is important to keep in mind. In addition, more work is needed to understand how the EATCs patterns themselves evolve over time.

We then focus the analysis on the relationship between DJF temperature extremes and the EATCs on two periods showing different EATC phases, i.e. 1951-1980 and 1981-2010. In ERA-20C, we find regions with significant correlations between EATC and cold (and warm, not shown) extremes. The extreme cold temperature index used here is highly correlated with heating degree days, hence energy demand, and is therefore relevant to the energy sector.

ASF-20C is however only partly able to represent the relationships between cold extremes and the 4 EATCs' indices but more work is required to understand the reason. From both

ERA-20C and ASF-20C however, there does not appear to be strong differences in this relationship in the two periods considered, which is similar to what was found in Schaller et al. (2018) between summer warm extremes and atmospheric blocking. Further work will focus on quantifying this relationship on composites of years with strongly positive or negative EATC phases rather than periods to better understand the physical mechanisms at play.

9 Conclusions

The analysis presented above seeks to provide two key insights into the capabilities of subseasonal and seasonal NWP systems, assessing their ability to (a) represent and predict key patterns of large-scale atmospheric circulation in the Euro-Atlantic sector and (b) to faithfully capture the observed surface impacts of the large-scale circulation. As discussed in Chapter 1, the rationale for doing this lies in the potential for hybrid "pattern-based" forecasts to offer skill improvements over direct "grid-point" forecasts of surface meteorological conditions (and consequent energy impacts): in effect, a typical hybrid pattern-based forecast seeks to use an NWP system to predict the large-scale circulation which are linked to surface climate and energy impacts using statistical methods. Several such hybrid forecast schemes are subsequently developed and their performance evaluated in D4.3.

Several different approaches to circulation typing are tested, spanning a range of forecast systems and time-scales. As the details of each methodological scheme are somewhat unique to each study, readers of this document are referred to the relevant sections in each individual chapter for detailed discussion of each method and its principle conclusions. A brief recap of the main findings from each chapter is, however, given below before a concluding with a synthesis.

EATCs and seasonal NWP prediction systems (Chapter 3). The method to express the atmospheric circulation forecast in terms of four EATC pattern indices (NAO, EA, EAWR and SCA) developed in D3.2 is shown to characterize Euro-Atlantic atmospheric variability better than using just a single index (NAO). NWP forecasts of all four patterns in winter (DJF-average) are shown to have skill at a lead time of 1 month, with indications of lead-1 skill also found in summer and spring, though not autumn. Importantly for the hybrid pattern-based approach, the NWP models are shown to have deficiencies in their simulated relationship between the EATCs and surface climate: while the surface impact is qualitatively similar to observations, it is often found to be weaker and spatially shifted in the seasonal NWP models.

WRs in subseasonal and seasonal NWP prediction systems (Chapter 4). The method to express the atmospheric circulation in terms of four time-evolving WRs developed in D3.2 is applied to ECMWF subseasonal and seasonal forecasts (the method is conceptually similar to Cassou (2008) but is applied on a rolling five-week window, such that WR patterns vary each week across the year). Applying the clustering scheme to the ECMWF seasonal forecast system revealed very different WR patterns compared to observations: it is therefore concluded that, in the present context, the WR patterns (into which NWP forecast days are assigned) should themselves be defined on observational datasets (i.e., reanalysis), rather than defined from the model data itself. The resulting analysis suggests that the ECMWF subseasonal and seasonal NWP models have limited skill in predicting the frequency of the time-evolving WRs (forecast week 1 and forecast month 1 respectively). Although for some WRs and times of the year some skill is found in ECMWF-MFS into weeks 3 and 4, especially in winter weeks. Analysis of the ECMWF subseasonal NWP model also suggests a transition

probability bias: the likelihood of a transition from one WR to another is overestimated while persistence in the same WR is underestimated.

Winter TCTs and subseasonal NWP (Chapter 5). The TCT method developed in D3.2 is applied to the ECMWF and NCEP subseasonal forecasts in extended winter, along with a "simple" WR method (a single set of 4 regimes defined over the entire extended winter). The frequency of occurrence of and the spatial structure of the associated surface conditions is found to be well represented in both systems though the use of ensemble-averaging (of the circulation prior to classification) leads to a bias towards "weaker" circulation patterns, and the magnitude of the surface impact decays with lead time (even in individual ensemble members). Both NWP models are found to have some skill in predicting weekly frequencies of occurrence in forecast weeks 1 and 2, reducing thereafter (similar for both WR and TCT). A key limitation for TCTs is found to be the weak constraint on the surface impact conditions provided by Z500 (geopotential height in the mid-troposphere): methods to enhance the strength of this constraint (e.g., through use of additional large-scale circulation fields) are highlighted as a promising area for further consideration.

HWR and seasonal NWP (Chapter 6). The HWR classification scheme presented in D3.2 is applied to the ECMWF seasonal forecast system. The spatial structure, frequency of occurrence, and monthly persistence of the HWRs in the NWP forecast is found to correspond well with reanalysis, suggesting that the forecast model provides a good representation of the underlying large-scale circulation drivers. In terms of predictions, however, the NWP forecast skill is found to be limited: while there is clear skill in predicting the frequency of occurrence of HWR in month 0, the results are less clear for longer lead times.

Winter SSWs and subseasonal NWP (Chapter 7). Winter-time SSWs, defined by a catalogue of past recorded events, are shown to influence European climate with cold temperatures and low wind speeds resulting in increased demand and DNW compared to normal winter conditions (and associated with an increased likelihood of the NAO- WR and EurTr TCT). The spatial structure of the surface impact of SSWs and their impact on WR/TCT frequencies is well represented in both the ECMWF and NCEP subseasonal forecasts, though the surface impact has reduced magnitude. The analysis, however, relies on a small sample of events and is consequently subject to some uncertainty.

Winter EATCs in 20th century retrospective seasonal forecast reconstruction (Chapter 8). The EATC classification scheme (as in Chapter 3) is applied to the longer ERA-20C reanalysis and ASF-20C reforecast ensemble. All 4 EATCs are shown to contain substantial decadal variability for DJF in ERA-20C and also the ASF-20C ensemble (both in terms of the projection index onto a fixed pattern and in the patterns themselves if re-calculated over different periods). The skill of ASF-20C to predict the EATC is also shown to exhibit decadal variability with significant correlation skill in the recent period and early 20C, but not in mid-20C: this suggests that the predictability (and therefore NWP forecast performance with any given model) may change over time. The ASF-20C forecast, however, seems to be able to capture the relationship between EATCs and cold extremes.

Throughout the analysis presented above is possible to identify a number of important general lessons relevant to both ongoing research and subsequent operationalization. Most importantly, the overall analysis supports the hypothesis that pattern-based forecasting methods may potentially offer additional skill over grid-point methods. This relies on two arguments. Firstly, the ability of seasonal and subseasonal NWP models to represent the link between the large-scale circulation and surface impacts appears to be imperfect and degrades with lead time (Chapters 3, 4, 5 and 7) and, secondly, the large-scale circulation itself appears to be somewhat predictable at least at modest lead times (Chapters 3, 4, and 5). This central theme will be returned to in Deliverable D4.3. It is, however, important to emphasize that the skill advantages are likely to be quantitative and incremental rather than a pronounced (i.e., qualitative) step-change. The ability of the NWP forecast systems to predict weather types remains strongest at short lead times (week 1-2 or month 0-1 in subseasonal and seasonal respectively), decaying thereafter.

Several different circulation typing schemes have been tested, and the performance of the forecast systems in predicting the types is shown to vary considerably across the year. For example, subseasonal forecasts are shown to have little skill in predicting WR occurrence beyond week 1 in the transition seasons (Section 4.2), whereas for winter (and to a lesser extent summer) WR can be predicted with modest levels of skill in week 2 and perhaps even week 3 (Section 4.2 and Chapter 5). Moreover, the analysis suggests that the detailed structure of the circulation patterns may be highly sensitive to the model selection (Section 4.1), period of definition (seasonal variations Chapters 3, 5 and 6; decadal variability Chapter 8), tailoring to target specific variables or properties (TCT Chapter 5 and HWR Chapter 6), and methodological choices (e.g., the use of ensemble averaging, Section 5.5). It also suggests that increased modelling complexity does not always lead to enhanced forecast skill (e.g., the evolving WR vs the fixed WR). Overall, this strongly confirms the principle that the process of design of the circulation patterns for operational use should be closely integrated with the process forecast skill assessment (e.g., through repeated iterations in design).

Despite a widespread focus on seasonal forecasting of the NAO in the NWP forecasting community, an interesting observation from the present analysis is that it appears to suggest that other atmospheric circulation patterns may have comparable levels of forecast skill (e.g., the four EATCs in month 1 seasonal forecasts, Chapter 3). Similarly, at subseasonal timescales, the highly tailored TCTs appear to have comparable skill to traditional WR (Chapter 5). This may suggest an area where value can be obtained from the climate service "co-design" principle: if the atmospheric circulation patterns to which a user is most sensitive can be identified, it may be possible to seek to forecast those patterns with comparable skill levels to more classically defined meteorological patterns.

Finally, the results of Chapter 3 highlight the potential benefits of combining several NWP forecast systems into a single multi-model ensemble. It is anticipated that this will be further explored in Deliverable D4.4.

Bibliography

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., … Pappenberger, F. (2018). Skilful seasonal forecasts of streamflow over Europe? Hydrology and Earth System Sciences, 22, 2057–2072. https://doi.org/10.5194/hess-22-2057-2018

Athanasiadis, P. J., Bellucci, A., Scaife, A. A., Hermanson, L., Materia, S., Sanna, A., … Gualdi, S. (2017). A multisystem view of wintertime NAO seasonal predictions. Journal of Climate, 30(4), 1461–1475. https://doi.org/10.1175/JCLI-D-16-0153.1

Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., & Scaife, A. A. (2018). An Intercomparison of Skill and Overconfidence/Underconfidence of the Wintertime North Atlantic Oscillation in Multimodel Seasonal Forecasts. Geophysical Research Letters, 45(15), 7808–7817. https://doi.org/10.1029/2018GL078838

Bárdossy, A., Stehlík, J., & Caspary, H. (2002). Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules. *Climate Research, 23:* 11-22.

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

Bennett, J. C., Wang, Q. J., Robertson, D. E., Schepen, A., Li, M., & Michael, K. (2017). Assessment of an ensemble seasonal streamflow forecasting system for Australia. Hydrology and Earth System Sciences, 21(12), 6007–6030. https://doi.org/10.5194/hess-21-6007-2017

Bett, P. E., & Thornton, H. E. (2016). The climatological relationships between wind and solar energy supply in Britain. *Renewable Energy*, *87*, 96-110.

Bloomfield, H. C., Brayshaw, D. J., Shaffrey, L. C., Coker, P. J., & Thornton, H. E. (2016). Quantifying the increasing sensitivity of power systems to climate variability. *Environmental Research Letters*, *11*(12).

Bloomfield, HC, Brayshaw, DJ, Charlton-Perez, A. (in press) Characterising the meteorological drivers of the European electricity system using Targeted Circulation Types. Meteorological Applications

Brayshaw, D. J., Troccoli, A., Fordham, R., & Methven, J. (2011). The impact of large scale atmospheric circulation patterns on wind power generation and its potential predictability: A case study over the UK. *Renewable Energy*, *36*(8), 2087-2096.

Brunner, L., Schaller, N., Anstey, J., Sillmann, J., & Steiner, A. K. (2018). Dependence of Present and Future European Temperature Extremes on the Location of Atmospheric Blocking. *Geophys. Res. Letters, 45 (12),* 6311-6320.

Bruno Soares, M., Alexander, M., & Dessai, S. (2017). Sectoral use of climate information in Europe: A synoptic overview. Climate Services, 1–16. https://doi.org/10.1016/j.cliser.2017.06.001

Bruno Soares, M., & Dessai, S. (2016). Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. Climatic Change, 137(1–2), 89–103. https://doi.org/10.1007/s10584-016-1671-8

Butler, A. H., & Polvani, L. M. (2011). El Niño, La Niña, and stratospheric sudden warmings: A re-evaluation in light of the observational record. *Geophysical Research Letters*, *38*(13).

Butler, A.H., Sjoberg, J.P., Seidel, D.J. and Rosenlof, K.H., 2017. A sudden stratospheric warming compendium. Earth System Science Data, 9(1).

Cannon, D. J., Brayshaw, D. J., Methven, J., Coker, P. J., & Lenaghan, D. (2015). Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain. *Renewable Energy*, *75*, 767-778.

Cassou, C. (2008). Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. Nature, 455, 523.

Charlton-Perez, A. J., Ferranti, L., & Lee, R. W. (2018). The influence of the stratospheric state on North Atlantic weather regimes. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1140-1151.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.(2011). The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, https://doi.org/10.1002/qj.776.

Cortesi, N., Torralba, V., González-Reviriego, N., Soret, A., & Doblas-Reyes, F. J. (2019). Characterization of European wind speed variability using weather regimes. *Climate Dynamics*, 1–16. https://doi.org/10.1007/s00382-019-04839-5

Cradden, L. C., McDermott, F., Zubiate, L., Sweeney, C., & O'Malley, M. (2017). A 34-year simulation of wind generation potential for Ireland and the impact of large-scale atmospheric pressure patterns. Renewable Energy, 106, 165–176. https://doi.org/10.1016/j.renene.2016.12.079

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the

data assimilation system, Quarterly Journal of the Royal Meteorological, 137, 553– 597, https://doi.org/10.1002/qj.828

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., … Knight, J. (2016). Skilful predictions of the winter North Atlantic Oscillation one year ahead. Nature Geoscience, 9(11), 809–814. https://doi.org/10.1038/ngeo2824

Ely, C. R., Brayshaw, D. J., Methven, J., Cox, J., & Pearce, O. (2013). Implications of the North Atlantic Oscillation for a UK–Norway renewable power system. *Energy Policy*, *62*, 1420-1427.

ENTSO (2019). European network of transmission system operators for electricity: data platform [cited 17th Jan 2017]. Available from http://entsoe.eu/data/Pages/default.aspx.

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Harvey, V. L., Hegglin, M. I., Homeyer, C. R., Knox, J. A., Kruger, K., Lambert, A., Long, C. S., Martineau, P., Molod, A., Monge-Sanz, B. M., San- tee, M. L., Tegtmeier, S., Chabrillat, S., Tan, D. G. H., Jack- son, D. R., Polavarapu, S., Compo, G. P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Paw- son, S., Simmons, A., Wargan, K., Whitaker, J. S., and Zou, C.-Z. (2017). Introduction to the SPARC reanalysis intercomparison project (S-RIP) and overview of the reanalysis systems, Atmospheric Chemistry and Physics, 17, 1417–1452, https://doi.org/10.5194/acp- 17-1417-2017.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M. , Gu, W. , Kim, G. , Koster, R. ,Lucchesi, R., Merkova, D., Nielsen, J.E. , Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B. (2017).The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate, 30, 5419–5454, doi:10.1175/JCLI-D-16-0758.1

Global Wind Atlas (2019). The Global Wind Atlas [cited 1st March 2019]. Available from https://globalwindatlas.info.

Green, R. (2005). Electricity and Markets. Oxford Review of Economic Policy, 21, 67–87.

Gregow, H., Jylhä, K., Mäkelä, H. M., Aalto, J., Manninen, T., Karlsson, P., Kaiser-Weiss, A.K., Kaspar, F., Poli, P., Tan, D. G. H., Obregon, A., and Su, Z. (2016). Worldwide survey of awareness and needs concerning reanalyses and respondents views on climate services, Bulletin of the American Meteorological Society, 97(8), 1461–1474, doi:10.1175/BAMS-D-14-00271.1.

Ineson, S., & Scaife, A. A. (2009). The role of the stratosphere in the European climate response to El Niño. *Nature Geoscience*, *2*(1), 32.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., … Monge-Sanz, B. M. (2019). SEAS5: The new ECMWF seasonal forecast system. Geoscientific Model Development, 12(3), 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Kumar, A. (2009). Finite Samples and Uncertainty Estimates for Skill Measures for Seasonal Prediction. Monthly Weather Review, 137(8), 2622–2631. https://doi.org/10.1175/2009MWR2814.1

Lledó, Ll. (2017). CLIM4ENERGY technical note no.1: computing capacity factor. BSC-ESS Technical Memorandum 2017-001, 9 pp.

Lledó, L., V. Torralba, A. Soret, Ramon. J., and F.J. Doblas-Reyes (2019). Seasonal forecasts of wind power generation. Renewable Energy, 143, 91-100. https://doi.org/10.1016/j.renene.2019.04.135

Mahlstein, I., Spirig, C., Liniger, M. A., & Appenzeller, C. (2015). Estimating daily climatologies for climate indices derived from climate model data and observations. *Journal of Geophysical Research: Atmospheres*, *120*(7), 2808–2818. https://doi.org/10.1002/2014JD022327

Manzanas, R., Gutiérrez, J. M., Bhend, J., Hemri, S., Doblas-Reyes, F. J., Torralba, V., … Brookshaw, A. (2019). Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. Climate Dynamics, 53(3–4), 1287–1305. https://doi.org/10.1007/s00382-019-04640-4

Matsueda, M., & Palmer, T. N. (2018). Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1012–1027. https://doi.org/10.1002/qj.3265

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., … Arnold, J. R. (2017). An intercomparison of approaches for improving operational seasonal streamflow forecasts. Hydrology and Earth System Sciences, 21(7), 3915–3935. https://doi.org/10.5194/hess-21-3915-2017

Michelangeli, P.-A., Vautard, R., & Legras, B. (1995). Weather Regimes: Recurrence and Quasi Stationarity. *Journal of the Atmospheric Sciences*. https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2

Neal, R., Fereday, D., Crocker, R., & Comer, R. E. (2016). A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, *23*(3), 389–400. https://doi.org/10.1002/met.1563

Poli, P., Hersbach, H., Tan, D., Dee, D., Thepaut, J.-N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P., Dragani, R., Treemolet, Y., Holm, E., Bonavita, M., Isaksen, L., & Fisher, M. (2013). The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20thcentury assimilating surface observations only (ERA-20C). *ECMWF ERA Report Series 14: Reading, UK*.

Poli, P., Hersbach, H., Berrisford, P., Dee, D., Simmons, A. & Laloyaux, P. (2015). ERA-20C deterministic. *ECMWF ERA Report Series 20: Reading, UK*.

Ramon, J., Lledó, L., Torralba, V., Soret, A., & Doblas-Reyes, F. J. (2019). What global reanalysis best represents near-surface winds? Quarterly Journal of the Royal Meteorological Society. https://doi.org/10.1002/qj.3616

Richter, J. H., Matthes, K., Calvo, N., & Gray, L. J. (2011). Influence of the quasi-biennial oscillation and El Niño–Southern Oscillation on the frequency of sudden stratospheric warmings. *Journal of Geophysical Research: Atmospheres*, *116*(D20).

Robertson, A. and Vitart, F. eds., 2018. Subseasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting. Elsevier.

Russo, S., Sillmann, J., & Fischer, E. M. (2015). Top ten European heatwaves since 1950 and their occurrence in the coming decades. Environmental Research Letters 10(12), 124003.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., … Becker, E. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., ... & Williams, A. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*(7), 2514-2519.

Schaller, N., Sillmann, J., Anstey, J., Fischer, E. M., Grams, C. M., & Russo, S. (2018). Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles, *ERL, 13 (5).*

Shi W., Schaller N., MacLeod D., Palmer T. N. & Weisheimer A. (2015). Impact of hindcast length on estimates of seasonal climate predictability. *Geophys. Res. Lett. 42*, 1554–1559.

Stoft, S. (2002). Power System Economics. IEEE Press Wiley, Piscataway, NJ.

Thornton, H. E., Scaife, A., Hoskins, B. J., Brayshaw, D. J., Smith, D., Dunstone, N., Stringer, N. & Bett, P. E. (2019). Skilful seasonal prediction of winter gas demand. *Environmental Research Letters*, 14, 024009.

Torralba, V. (2019). *Seasonal climate prediction for the wind energy sector: methods and tools for the development of a climate service*. Universidad Complutense de Madrid. https://eprints.ucm.es/56841/

Van der Wiel, K., H.C. Bloomfield, R.W. Lee, L.P. Stoop, R. Blackport, J.A. Screen, and F.M. Selten (2019). The influence of weather regimes on European renewable energy production and demand. Environmental Research Letters, submitted.

Vautard, R. (1990). Multiple Weather Regimes over the North Atlantic: Analysis of Precursors and Successors. *Monthly Weather Review*, *118*(10), 2056–2081. https://doi.org/10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;2

Vitart, F. (2004). Monthly Forecasting at ECMWF. *Mon. Wea. Rev.*, *132*(12), 2761–2779. https://doi.org/10.1175/mwr2826.1

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., … Zhang, L. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., & Palmer, T. N. (2017). Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *QJ. RMetSoc, 143 (703)*, 917-923.

Windpower.net (2019). Thewindpower.net database. [Cited: 1st March 2017]. Available for purchase from: https://www.thewindpower.net.

Zadeh, L. (1965). Fuzzy sets. *Information and control*, 8:338–353.

Zubiate, L., McDermott, F., Sweeney, C., & O'Malley, M. (2017). Spatial variability in winter NAO–wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns. Quarterly Journal of the Royal Meteorological Society, 143(702), 552–562. https://doi.org/10.1002/qj.2943