S2S4E

Climate Services for Clean Energy

Research and Innovation action

H2020-SC5-2017

Assessment of conditional predictability and use of teleconnections to construct energy-relevant variables

Deliverable D4.3

Version N°1

Authors: David Brayshaw (UREAD), Hannah Bloomfield (UREAD), Irene Cionni (ENEA), Kean Foster (previously with SMHI), David Livings (UREAD), Llorenç Lledó (BSC), Andrea Manrique-Suñén (BSC), Verónica Torralba (BSC), Nicola Cortesi (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

1 Introduction 12

2 Data and Methods 14

2.1 Introduction 14

2.2 Meteorological datasets and forecasting systems 14

2.2.1 Reanalysis data 14
2.2.2 Sudden Stratospheric Warmings 15
2.2.3 Seasonal forecast systems 15
2.2.4 Subseasonal forecast systems 16
2.2.5 Forecast horizons 17

2.3 Pattern identification methods 17

2.3.1 East Atlantic Teleconnection Patterns 17
2.3.2 Weather Regimes 18
2.3.3 Targeted Circulation Types 18
2.3.4 Hydrological Weather Regimes 18

3 Employing seasonal forecasts of Euro-Atlantic Teleconnection indices to reconstruct surface variable forecasts 20

3.1 Introduction 20

3.2 Producing hybrid predictions 20

3.3 Performance under perfect knowledge of EATCs 22

3.4 Skill assessment of hybrid forecasts compared to dynamical predictions 23

3.5 Conclusions 26

4 Weather regimes for wind forecasting 27

4.1 Introduction 27

4.2 Method 27

4.3 Results 28

4.4 Conclusions 30

5 Weather Regime and Targeted Circulation Types for European national-scale energy balance indicator forecasts at subseasonal timescales during winter 32

5.1 Introduction 32

5.2 Forecast methodology for national energy indicators 32

5.2.1 Meteorology-to-power conversion models: ERA5 32
5.2.2 Meteorology-to-power conversion:hindcasts 33

5.3 Weather regimes and targeted circulation types (TCTs) 33

5.4 Pattern forecasting methodologies 34

5.5 A perfect forecast model of National DNW 36

5.6 Subseasonal forecasts of DNW 38

5.6.1 The impact of variance inflation on forecast skill 38

5.6.2 WR and TCT pattern forecasts in subseasonal models 39

5.7 Conclusions 43

6 Hydrological catchment-scale forecasting using hydrological weather regimes 44

6.1 Introduction 44

6.2 Experimental Setup 44

6.3 Case study: Ume River 47

6.4 Conclusions 49

7 Use of remote climate indices to condition European national-scale energy balance indicator forecasts 50

7.1 Introduction 50

7.2 Method 50

7.3 Results 51

7.4 Conclusions 56

8 Conclusions 57

Bibliography 60

List of figures

Figure 1. Schematic representation of the process used to build hybrid dynamical-statistical forecasts of surface variables from the EATC forecasts. 22

Figure 2. Goodness of fit of a multilinear model to observed anomalies of 500 hPa geopotential height, 2m temperature, surface wind and solar radiation in DJF..... 23

Figure 3. Ranked Probability Skill Score of hybrid forecasts of surface variables compared to the reference climatology. EATC forecasts in DJF from DWD System2 (first column) and ECMWF SEAS5 (second column) have been employed to reconstruct forecasts of surface temperature (first row), surface wind (second row) and surface solar radiation (third row). Forecasts issued November 1st. 24

Figure 4. Ranked Probability Skill Score of hybrid forecasts of surface variables compared to the dynamical prediction forecasts over Europe. EATC forecasts in DJF from DWD System2 (first column) and ECMWF SEAS5 (second column) have been employed to reconstruct forecasts of surface temperature (first row), surface wind (second row) and surface solar radiation (third row). Raw forecasts of the three surface variables have been employed as benchmark. Black dots indicate grid points where although the hybrid predictions are better than the dynamical predictions, its quality is still worse than the climatology reference..... 25

Figure 5. Reconstructed 10-m wind speed for the months of January, April, July and October. Four error metrics are shown from top to bottom, respectively: the temporal correlation between the time series of the reconstructed and observed monthly mean wind speed anomalies, the ratio between the standard deviations of the reconstructed and observed time series of monthly mean speed anomalies, the correlation between the reconstructed and observed monthly 95th percentile of wind speed anomalies, and the mean wind speed bias. These results are based on the WRs and 10-m wind speed from the JRA-55 reanalysis..... 30

Figure 6. Schematic description of the forecast process (as presented in Chapter 5)..... 36

Figure 7. ERA5 DNW perfect forecast tests for the European total (a and c) and six case study countries (b and d). (a-b) Mean Absolute Error (MAE) (c-d) Anomaly Correlation Coefficient (ACC). Error bars show significance of results based on 2000 bootstrapped samples of each year of data. 37

Figure 8. Difference in Mean Absolute Error (MAE) between forecasts of European total DNW with and without variance inflation included in the bias correction (see Section 5.2 for details of the method). Positive numbers imply that variance inflation has resulted in a larger MAE (and therefore has degraded forecast quality)..... 39

Figure 9. Anomaly correlation between normalised DNW forecasts and ERA5. ECMWF (left) and NCEP (right). (a-b) European total (c-d) France (e-f) Germany (g-h) United Kingdom.

Note that, by construction, a climatological forecast has an ACC of zero for all lead times. 41

Figure 10. Anomaly correlation between normalised DNW forecasts using ERA5 and ECMWF hindcasts (left) and NCEP hindcasts (right) for lead weeks 1-4. (a-b) Sweden (c-d) Romania (e-f) Spain. Note that, by construction, a climatological forecast has an ACC of zero for all lead times. 42

Figure 11. Schematic presentation of the HBV-96 model for a single basin (Lindström et al., 1997). 46

Figure 12. Differences in the verification metrics for ECMWF demand hindcasts using mean bias adjustment. Top left: Brier score for upper decile. Bottom left: Brier score for lower decile. Top right: RPS for terciles. Bottom right: CRPS. 51

Figure 13. Statistically significant differences in the verification metrics for ECMWF demand hindcasts using mean bias adjustment. Layout as in Figure 12 52

Figure 14. Statistically significant differences in the verification metrics for NCEP demand hindcasts without variance inflation. Layout as in Figure 12. 53

Figure 15. Statistically significant differences in the verification metrics for ECMWF wind power hindcasts without variance inflation. Layout as in Figure 12. 54

Figure 16. Statistically significant differences in the verification metrics for the NCEP wind power hindcasts without variance inflation. Layout as in Figure 12. 55

List of tables

Table 1: Technical details of the seasonal prediction systems 16

Table 2. The validation metrics used to evaluate performance. The threshold for skill is 50% for FY^+ and 0 for all the other metrics. 45

Table 3. Bootstrapped (n=5000) FY^+ and MAESS, aggregate of all reservoirs, for the standard ESP approach (without use of HWR information) with climatology being used as the reference. The threshold for FY+ skill is 50% and MAESS skill is 0. Red shading indicates results below the skill threshold and blue indicates results above the skill threshold. 47

Table 4. Bootstrapped (n=5000) FY^+, aggregate of all reservoirs, for the analogue approach with the 'standard' ESP approach being used as the reference. Rows: month of hindcast initialization. Columns: indicates the forecast months for which the HWR were defined (e.g., column "2" indicates HWRs were defined across forecast months 0 and 1; column "5" indicates HWRs across forecast months 0, 1, 2, 3 and 4; see main text) . The threshold for skill is 50%. Red shading indicates results below the skill threshold and blue indicates results above the skill threshold. 48

Table 5. Bootstrapped (n=5000) MAESS, aggregate of all reservoirs, for the analogue approach with the reference being the 'standard' ESP approach. Rows and columns: as Table 4. The threshold for skill is 0. Red shading indicates results below the skill threshold, white indicate results around the skill threshold, and blue indicates results above the skill threshold (the intensity of the shading indicates relatively how far away the results are from 0, light = near, dark = far). 48

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 systems at subseasonal and seasonal time-horizons. The central concept of pattern-based forecasting is to use the NWP forecast 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 aim of this document is therefore two-fold. Firstly, it seeks to extend the pattern-forecasting techniques developed in Deliverable D4.2 to produce quantitative forecasts of surface meteorological and energy-system impacts, comparing their performance against alternative forecasting methodologies (e.g., grid-point NWP and simpler benchmark forecasts such as climatology and persistence). Secondly, it investigates the closely-related subject of conditional predictability: here interpreted to mean changes in the predictive skill produced by forecast models during different prevailing atmospheric conditions (here, Stratospheric Sudden Warmings). It therefore builds upon previous deliverables in terms of technique and analysis (particularly Deliverable D4.2 but also D3.1, D3.2 and D4.1).

A series of different pattern-based forecast methods are tested across multiple different NWP forecast systems. It is demonstrated that pattern-based methods can be utilized with NWP systems on both seasonal and subseasonal timescales to derive modest skill improvements in energy forecasts over equivalent forecasts based on direct grid-point predictions of surface climate variables. The precise nature of this skill improvement is, however, highly dependent on the methodology used and the geographical region considered. A key conclusion therefore is that the process of the design of the pattern-based forecast must be closely integrated with the process of forecast skill assessment (e.g., through repeated iterations in design). Furthermore, skill must be assessed continuously and on a case-by-case basis: skill in a particular performance metric or region should not be taken to imply skill in all regions or metrics.

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 the 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.

Deliverable D4.2 began 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 was instead interrogated. Modest levels of skill were found to exist in predicting the large-scale circulation patterns (particularly for weeks 1-2 and months 0-1, decreasing with lead-time), though this depended on the detailed methodology employed. In general, however, the NWP models were found to have subtle deficiencies in the link between the circulation patterns and their consequent surface impacts: typically the surface impact associated with a given circulation type was too week or spatially shifted (on average). This therefore appears to open an opportunity for hybrid forecasting, 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). As in D4.2, in this document such hybrid methods are referred to as "pattern-based" forecasts.

The aim of this document is therefore two-fold. Firstly, it seeks to extend the pattern-forecasting techniques developed in Deliverable D4.2 to produce quantitative forecasts of surface meteorological and energy-system impacts, comparing their performance against alternative forecasting methodologies (e.g., grid-point NWP and simpler benchmark forecasts such as climatology and persistence). Secondly, it investigates the closely-related subject of conditional predictability: here interpreted to mean changes in the predictive skill produced by forecast models during different prevailing atmospheric conditions (here, Stratospheric Sudden Warmings). It therefore builds upon previous deliverables in terms of technique and analysis (particularly Deliverable D4.2 but also D3.1, D3.2 and D4.1).

As in D4.2, several different and complementary approaches to pattern-based forecasting are tested and reported in the following discussion. 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 forecast models to predict surface weather conditions using a combination of Euro-Atlantic Teleconnection (EATC) patterns is examined first (Chapter 3). This is followed by an analysis of the extent to which Weather Regimes (WR) and Targeted Circulation Types (TCTs) can be used to predict surface climate and energy impacts (Chapters 4 and 5; note that the WR methodology differs substantially between the two chapters). Hydrological Weather Regimes (HWR) are investigated in Chapter 6, followed by the impact of SSW on conditional predictability (Chapter 7). A concluding discussion and synthesis is provided in Chapter 8.

As noted above, each chapter operates within a common analysis framework (which has evolved from Deliverables D3.2 and D4.1, similarly to D4.2) 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.

2 Data 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, D4.2. 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 7.

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. Based on this evaluation, in this report, three reanalyses products are selected for use: ERA-Interim, ERA5 and JRA55. All three, are modern-era reanalyses constrained by a full suite of observational data. 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 to represent observational uncertainty), though here the deterministic high-resolution version is used.

JRA-55 (Kobayashi et al. 2015) is a reanalysis produced by the Japan Meteorological Agency (JMA) operational data assimilation system, which is based on the operational system as of December 2009 with a 4D-VAR scheme with a six-hourly update cycle. This reanalysis starts in 1958 and provides data with six-hourly temporal resolution, a T319 spectral truncation (55 km) and 60 hybrid vertical levels. This reanalysis is distributed in real-time making it suitable for the development of products in an operational context. …

2.2.2 Sudden Stratospheric Warmings

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 and a southward shift of the storm track, 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), 14 dates in the ECMWF hindcast period (1996-2015), and 12 dates in the NCEP hindcast period (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, we use the ERA-20C reanalysis (Poli et al. 2013, 2015) as reference dataset, 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 further 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). It is a coupled system to both the 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 (in some cases, a larger ensemble is created by including a wider lagging period, e.g., up to 3 days previous).

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 East 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, HWRs are investigated to understand whether they provide additional information to improve forecast skill in the hydrological seasonal forecast. The hypothesis is that the HWRs may aid to select analogue years out of the historical ensembles used in the Ensemble Streamflow Prediction (ESP; Day, 1985) approach commonly employed in the hydropower sector in Sweden. The details can be found in Olsson et al., 2016 and its application to S2S4E can be found in Deliverable D3.2 (Section 3.2.2 therein).

An ensemble of daily HWR is generated over whole hindcast period and all members with ECMWF SEAS5 pressure field (see Deliverable D3.2 section 3.2.2 therein). The influence of HWR on hydrological runoff seasonal forecast has been conducted and the results are reported in Section 6. There, a rainfall-runoff model, HBV model (Bergström, 1976; Lindström et al., 1997), has been set up and calibrated for study area, the Ume river in northern Sweden.

3 Employing seasonal forecasts of Euro-Atlantic Teleconnection indices to reconstruct surface variable forecasts

3.1 Introduction

In D4.2 we have shown that EATC indices can be skillfully predicted one month in advance. Additionally, in D3.2 we found that observed EATC indices are strongly correlated with surface variables. However, we have also seen in D4.2 that the observed relationship between EATC patterns and surface impacts is not accurately reproduced by seasonal prediction systems. This opens the door to employ hybrid dynamical-statistical methods. The idea consists of combining the dynamical seasonal predictions of EATC indices with the observed relationship between EATCs and surface variables. This method can be thought of as a perfect prognosis method (Wilks 2011, section 7.5.2 therein) where the observed relationship between EATCs and impact variables has been modeled through a multilinear regression.

Multiple studies have related surface anomalies with teleconnections indices (NAO: Hurrell 1995, Hurrell and Deser 2009; EA: Moore and Renfrew 2012; EA/WR: Kutiel, H., Y. Benaroch, 2002; SCAND: Bueh and Nakamura, 2007). Several of those studies attempt to reconstruct surface variables through the use of EATC indices. For instance, Castro-Diez et al. 2002, limiting their study to the use of NAO, tried to reconstruct temperature anomalies and, more recently Riaz et al. 2017, using NAO and its constituent centers of action tried to reconstruct the climate in Germany. Rust et al. (2015) systematically studied the effect of a set of teleconnection patterns on European temperature and seek for a quantitative description of their individual contribution to temperature anomalies. In addition, northern hemisphere teleconnections patterns have been linked to climate anomalies in Northern America (Yu et al. 2019) and East Asia (Park et al.2011). However, none of those studies explores the possibility to combine those empirical models with seasonal predictions of the teleconnection indices.

In this chapter, we try to improve the utility of seasonal forecasts by incorporating skillful information of the Euro Atlantic Teleconnection patterns. The methodology is presented in Section 3.2, the performance of the statistical multi linear regression model is presented in Section 3.3. In Section 3.4, we present the skill assessment of the forecast obtained with this methodology (hybrid forecasts) in comparison with the raw seasonal predictions (dynamical forecasts) for three variables that impact the energy system (2 meters temperature, surface wind and surface solar radiation downward) at lead1 for the DJF season.

3.2 Producing hybrid predictions

A perfect prognosis method consists of two separate steps. Firstly, a statistical relationship between a circulation index and a surface variable is established from observations. Then a dynamical forecast of the circulation index is combined with the statistical relationship to derive a forecast for the surface variable. In this case four circulation indices, namely the four EATC indices described in Deliverables D3.2 and D4.2 have been employed.

A schematic representation of the methodology is shown in Figure 1. A statistical relationship between the four observed EATC indices and a surface variable (either surface wind, surface solar radiation or 2 meters temperature) has been obtained from ERA-Interim gridded data. The surface anomalies at each grid point have been expressed as a linear combination of the EATC indices. For each time (t) and location (x,y), the observed anomalies have been approximated as:

$$Anom(t,x,y) \approx a(x,y) + b(x,y) * NAO_i(t) + c(x,y) * EA_i(t) + d(x,y) * EAWR_i(t) + e(x,y) * SCA_i(t)$$

where NAO_i, EA_i, $EAWR_i$, and SCA_i are the observed EATC indices calculated following the methodology described in Section 2.3.1 above and in Deliverable D4.2. Least squares estimation allows to obtain the multilinear regression coefficients *a*, *b*, *c*, *d*, and *e* that minimize the approximation error for each grid point. Determination coefficients of each multilinear model fit have been used as an indication of goodness of fit.

Ensemble forecasts of EATC indices have been obtained from DWD System2 and ECMWF SEAS5 following the methodology in D4.2. The four predicted EATC indices for each ensemble member have been employed to reconstruct an ensemble of forecasts for the surface variable by using the multilinear model.

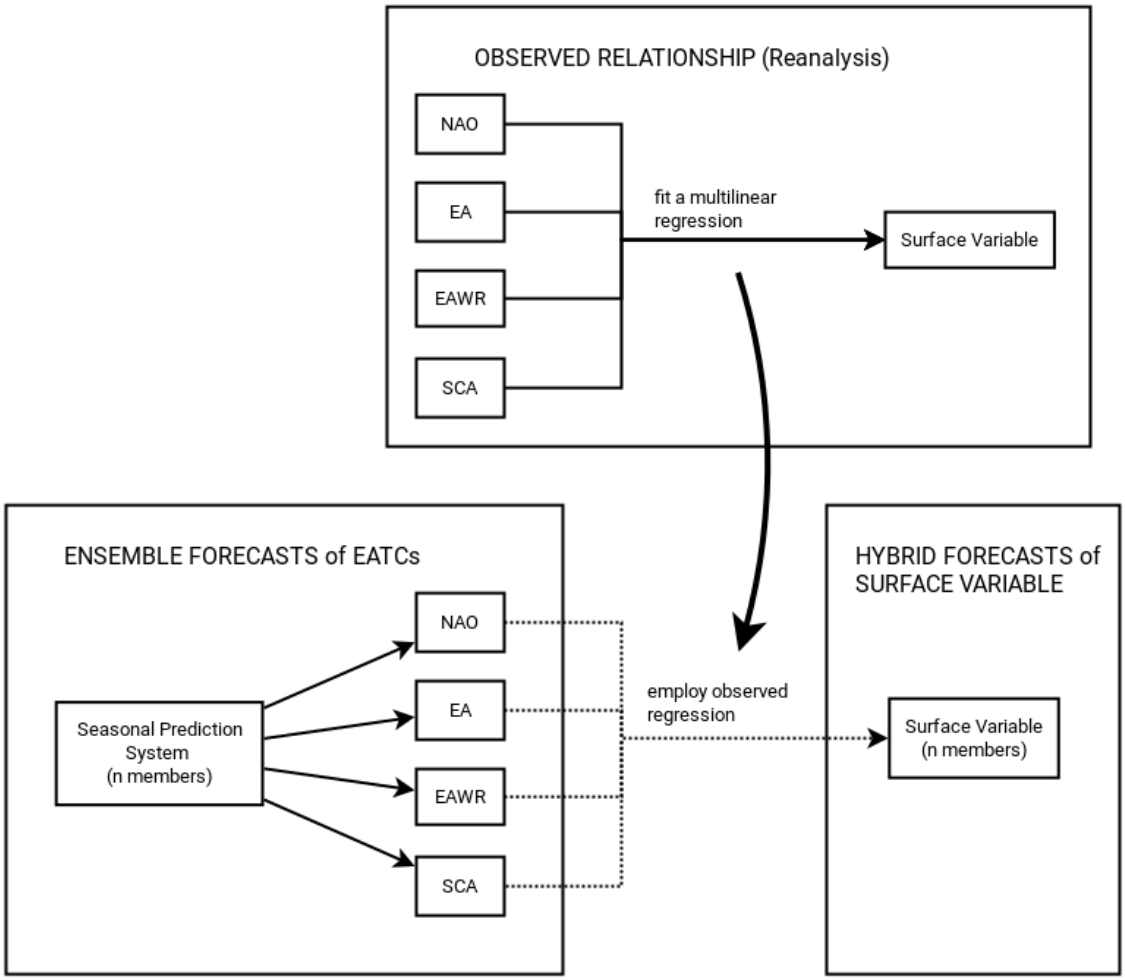

Figure 1. Schematic representation of the process used to build hybrid dynamical-statistical forecasts of surface variables from the EATC forecasts.

The quality of the hybrid predictions of surface variables has been assessed by computing Ranked Probability Skill Scores of tercile forecasts. Both the reference climatology and the dynamical predictions of the surface variables have been used as benchmark to understand the performance of this method. In both cases, the evaluation is made by comparing the hybrid forecasts with observed anomalies in ERA-Interim.

3.3 Performance under perfect knowledge of EATCs

A separate multilinear model has been fitted in each grid point for the three surface variables, and also for 500 hPa geopotential height (Z500) anomalies. The determination coefficient (i.e. the squared Pearson correlation coefficient) of the multilinear model fit is often used as an indicator of the goodness of the fit. It indicates the percentage of variability in the observed predictand (surface anomaly) that is explained by changes in the observed predictors (the EATC indices). This can also be thought of as the maximum quality that one can expect from

the hybrid forecasts if the state of the four EATCs is known. Figure 2 shows the determination coefficients for each variable and grid point for DJF. The determination coefficient is very high for Z500 (especially over the Atlantic), as expected, because the EATC indices were specially designed to represent as much Z500 variability as possible in the domain. For surface variables, the coefficients are still high in many European regions. For instance, 2m temperature can be very well determined in France and Germany from the four EATC indices, and surface winds can be well determined in the British Isles or north of Germany. Determination coefficients of solar radiation are weaker than determination coefficients of the previous variables, however they are useful in the Iberian Peninsula and central Europe. Overall, the figure shows that a good knowledge of the four EATCs could translate in skillful forecasts of the surface variables.

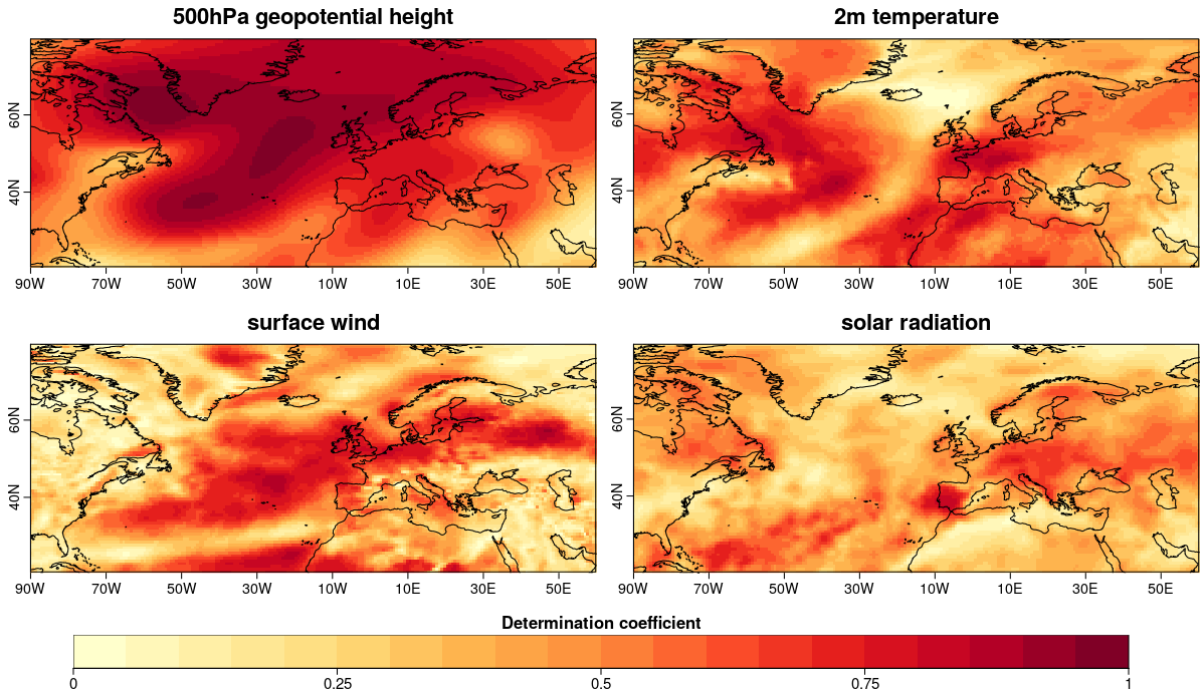

Figure 2. Goodness of fit of a multilinear model to observed anomalies of 500 hPa geopotential height, 2m temperature, surface wind and solar radiation in DJF.

3.4 Skill assessment of hybrid forecasts compared to dynamical predictions

Hybrid dynamical-statistical predictions have been produced from DWD System2 and ECMWF SEAS5 predictions of the four EATC indices for DJF and initialized in November. The quality of those forecasts has been compared to the usual climatology in terms of Ranked Probability Skill Score (RPSS) for tercile categories. Figure 3 shows the results for the two systems and the three surface variables analyzed. For the three variables, the DWD system produces better forecasts than the SEAS5. This is directly linked with the quality of the EATC forecasts employed, because the observed relationship (the multilinear model) is the same in both cases. In D4.2 we saw that DWD SYS2 was the system with the best predictions when

looking at all the four EATCs. The hybrid forecasts of temperature have a positive RPSS almost all over Europe. For surface wind, positive skill is found mostly in the north of the continent, while for solar radiation the better results are seen in the Mediterranean and Balkans regions. It is interesting to note that regions were the statistical relationship between EATCs and surface variables is not good (i.e. low determination coefficients in Figure 2), have negative RPSS (Figure 3). This can be seen, for instance, in a narrow band crossing the Atlantic in the temperature panels for both systems.

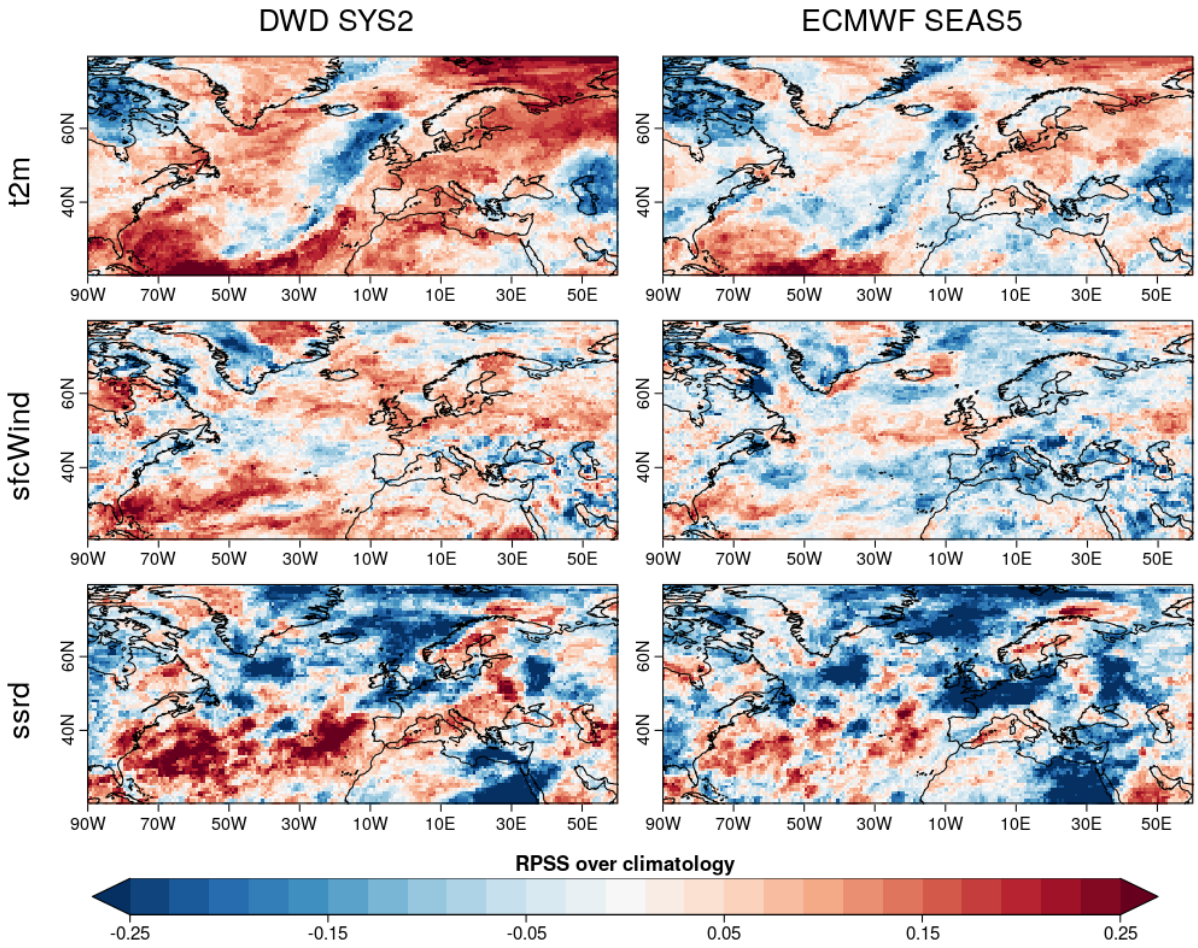

Figure 3. Ranked Probability Skill Score of hybrid forecasts of surface variables compared to the reference climatology. EATC forecasts in DJF from DWD System2 (first column) and ECMWF SEAS5 (second column) have been employed to reconstruct forecasts of surface temperature (first row), surface wind (second row) and surface solar radiation (third row). Forecasts issued November 1st.

In order to understand if the hybrid predictions are better than the dynamical predictions of surface variables directly output by the systems, the RPSS has also been computed employing the dynamical predictions as the benchmark forecast to beat. Figure 4 shows the results over Europe for the two systems and the three variables. In this case, a positive value indicates that hybrid predictions are better than dynamical predictions. For temperature, the hybrid forecasts from both prediction systems produce better results than the dynamical forecasts in the Mediterranean area and Central Europe. For surface wind the improvements

are more modest, but can still be seen for DWD in parts of Germany, Poland, Spain, Italy and the Balkans. Regarding solar radiation, a strong improvement is found over the Mediterranean basin, eastern Europe and parts of Scandinavia.

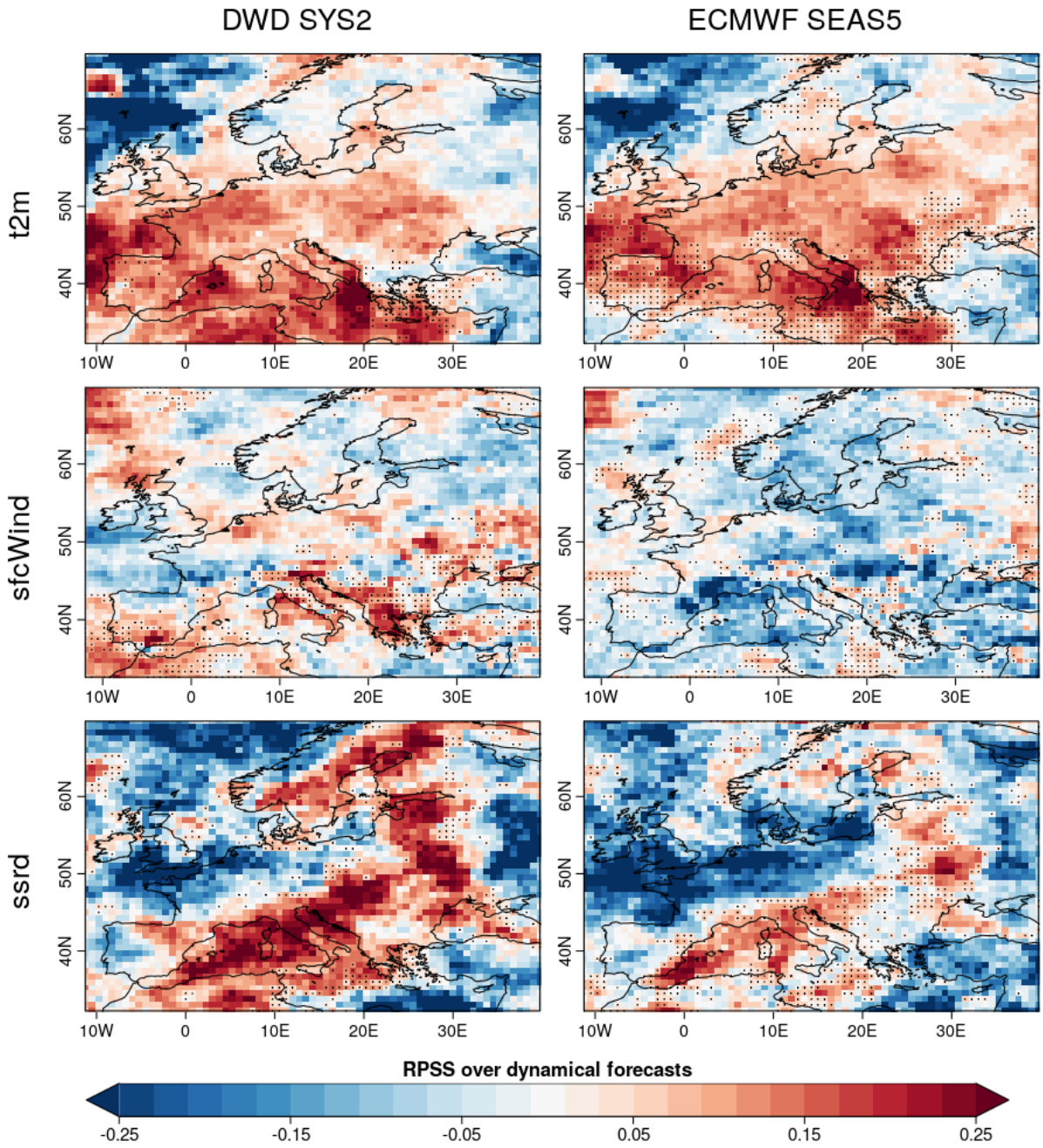

Figure 4. Ranked Probability Skill Score of hybrid forecasts of surface variables compared to the dynamical prediction forecasts over Europe. EATC forecasts in DJF from DWD System2 (first column) and ECMWF SEAS5 (second column) have been employed to reconstruct forecasts of surface temperature (first row), surface wind (second row) and surface solar radiation (third row). Raw forecasts of the three surface variables have been employed as benchmark. Black dots indicate grid points where although the hybrid predictions are better than the dynamical predictions, its quality is still worse than the climatology reference.

3.5 Conclusions

We attempted to improve the seasonal forecast of energy relevant essential climate variables by using statistical information of the Euro Atlantic Teleconnections and its relation to the surface variables. Our method is based on perfect prognosis method that takes advantage of a strong multilinear relationship between observed EATCs indices and observed impact variables.

The multilinear model fit has shown that the relationship is very strong for the geopotential at 500 hPa all over the domain (90° W - 60°E and 50°S-50°N) in DJF, which is strictly related with the way the EATCs indices were computed (Section 2.3.1). More interesting are the results for 2 meter temperature, surface wind and solar radiation, where the multilinear model fit works well in many European regions (France and Germany for temperature, British Isles and North of Germany for wind, Central Europe and Iberian Peninsula for surface solar radiation). With these promising results for the fitting model, we reconstructed the surface anomalies for two seasonal prediction systems and benchmarked these hybrid forecasts with the direct variable forecasts from the systems and also with the climatology. In both analyzed systems, temperature predictions are improved by the hybrid methodology over most of Europe. Wind predictions show patches of modest improvements while solar radiation has improvements in southern Europe. DWD SYS2, the system with the best winter EATC predictions overall (D4.2), shows the best results for the three analyzed surface variables.

4 Weather regimes for wind forecasting

4.1 Introduction

To understand the amount of wind speed variability that can be explained in terms of changes in the frequency of occurrence weather regimes, monthly wind speed has been reconstructed from the monthly impact maps of each WR (as described in D3.2) and the observed frequency of occurrence of each WR (as described in D4.2). This study can be defined as a 'perfect model approach' as it is based on reanalysis data and aims to provide information on the role of the WRs as a source of predictability of wind speed. Several aspects of the reconstructed fields are analysed in comparison with the wind speed directly obtained from the reanalysis: Pearson correlation of the mean wind speed, the ratio between standard deviations, Pearson correlation of the 95th percentile and mean wind speed bias. The JRA-55 reanalysis has been employed as reference (Kobayashi et al. 2015). Since this reanalysis is produced in near-real time it is useful for a posteriori analysis of WRs impact on particular anomalous wind events, which is sometimes required by energy companies to understand what flow configuration is behind particular events. An intercomparison of WR patterns as produced from different reanalyses can be found in Cortesi et al.(2019) as well as more details on the work presented in this section. A similar reconstruction using ERA-Interim reanalysis can be found in Torralba (2019).

4.2 Method

The reconstruction has been performed in a leave-one-out cross-validation framework, for which the year to be reconstructed has not been considered to estimate the observed impact of WR on 10-m wind speed. Firstly, the observed impact $I_{r,m,y}\,(lat,lon)$ of a given regime r in month m and year y on wind speed anomalies was measured by averaging the normalised wind speed anomalies $w_{r,m}\,(d,\ lat,\ lon)$ for all days d associated to regime r and month m during 1981-2016, excluding year y:

$$I_{r,m,y}\,(lat,lon) = \frac{1}{N}\sum_{d=1}^{N} w_{r,m}\,(d,lat,lon) \qquad (1)$$

with N the number of days belonging to regime r and month m during 1981, … y_{-1}, y_{+1}, …, 2016.

Subsequently, the WRs were employed to reconstruct the mean monthly wind speed anomaly $wRecon_{m,y}\,(lat,lon)$ as the linear combination of the WRs monthly impact $I_{r,m,y}$ and the WRs frequencies:

$$wRecon_{m,y}\,(lat,lon) = \frac{1}{N_{m,y}}\sum_{r=1}^{4} I_{r,m,y}(lat,lon)\cdot N_{rmy} \qquad (2)$$

being N_{rmy} the number of days belonging to the weather regime r, month m and year y and N_{my} the total number of days in month m and year y.

To assess the wind speed reconstruction, anomalies of reconstructed monthly mean wind speed (*wRecon)* were compared to the JRA-55 wind speed over the period 1981-2016 with four different metrics. The first one is the Pearson correlation between the reconstructed and observed time series. It ranges from -1 to 1, values closer to 1 indicate the reconstructed value is similar to the observed one. The second metric is the standard deviation ratio between the reconstructed and observed series and it ranges from 0 to 1, with values close to 1 indicating a high similarity between the variability of the reconstructed and observed wind speeds. The third one is the Pearson's correlation between the reconstructed and observed time series of the 95th percentile of 6-hourly wind speed and it ranges from -1 to 1. The last metric is the difference between the mean of reconstructed and original wind speed time series, therefore a difference equal to zero shows a perfect agreement between the means of the reconstructed and original wind speed anomalies.

The ability to reconstruct the highest wind speeds is studied by analysing the 95th percentile of the distribution, calculated in cross-validation with a similar approach to that used to reconstruct the time series of the mean wind speed. However, in order to improve the robustness of the measure of the 95th percentile thresholds, they were not reconstructed from the daily wind speed anomalies but from the 6-hourly wind speed anomalies instead. Notice that in this case, as the k-means clustering was still performed with daily MSLP anomalies, all the four 6-hourly wind speed values within the same day were associated to the same daily WR.

4.3 Results

Figure 5 shows the four aspects analysed for the selected months: January, April, July and October. The top row of Figure 5 shows correlations between observed mean wind speed and the one reconstructed by WRs. The highest correlations are found in January and in north-western Europe, particularly in the British Isles, Spain, Portugal, Denmark and in the northern part of France and Germany. Lowest correlations are measured in eastern Europe, North Africa and Middle East. In the North Sea, a key region for European wind power generation, correlations are often positive and in January are always above 0.5. The second row of Figure 5 presents the ratio between the reconstructed and observed wind speed standard deviation. It is always lower than 0.6, and it has a spatial distribution very similar to that of the mean wind speed correlations previously described. The ability to reconstruct not only the mean values but also the tails of the distribution is shown by correlation of the observed and reconstructed 95th percentiles, as shown in third row. Correlation values present a similar spatial distribution to those of the mean wind speed (first row), even though positive correlations are ~0.1 lower, while negative ones are ~0.1 higher (i.e.: less negative than correlations of the first row). Finally, the mean wind speed bias is shown in the fourth row of Figure 5. It is generally lower than 0.15% of mean wind speed. Moreover, the bias is never significant (for a paired t-test with a 90% confidence level). Figures for the other months can be found in Cortesi et al. (2019).

The more interesting feature of Figure 5 from the point of view of wind energy generation, is that it identifies countries in in which WRs play a dominant role on wind speed variability, like the North Sea region, but also other areas like the Iberian Peninsula, Iceland, northern Scandinavia and to a lesser extent the Gulf of Lion (southern France), the Black Sea and the Aegean Sea (between Greece and Turkey). In such regions, during many months of the year a large part of wind variability can be attributed to the change of the monthly frequencies of occurrence of the WRs, a key factor that could be exploited for the development of tailored products for the wind energy sector. Some of these regions, particularly northern Scandinavia, southern Spain and the Aegean Sea, are also characterised by high average yearly wind speeds (7-10 m/s) and are far enough from the North Sea to present positive wind anomalies when they are negative over central Europe. Thus, they might play an important role in future electricity generation, by reducing the high intermittency of total produced European wind power.

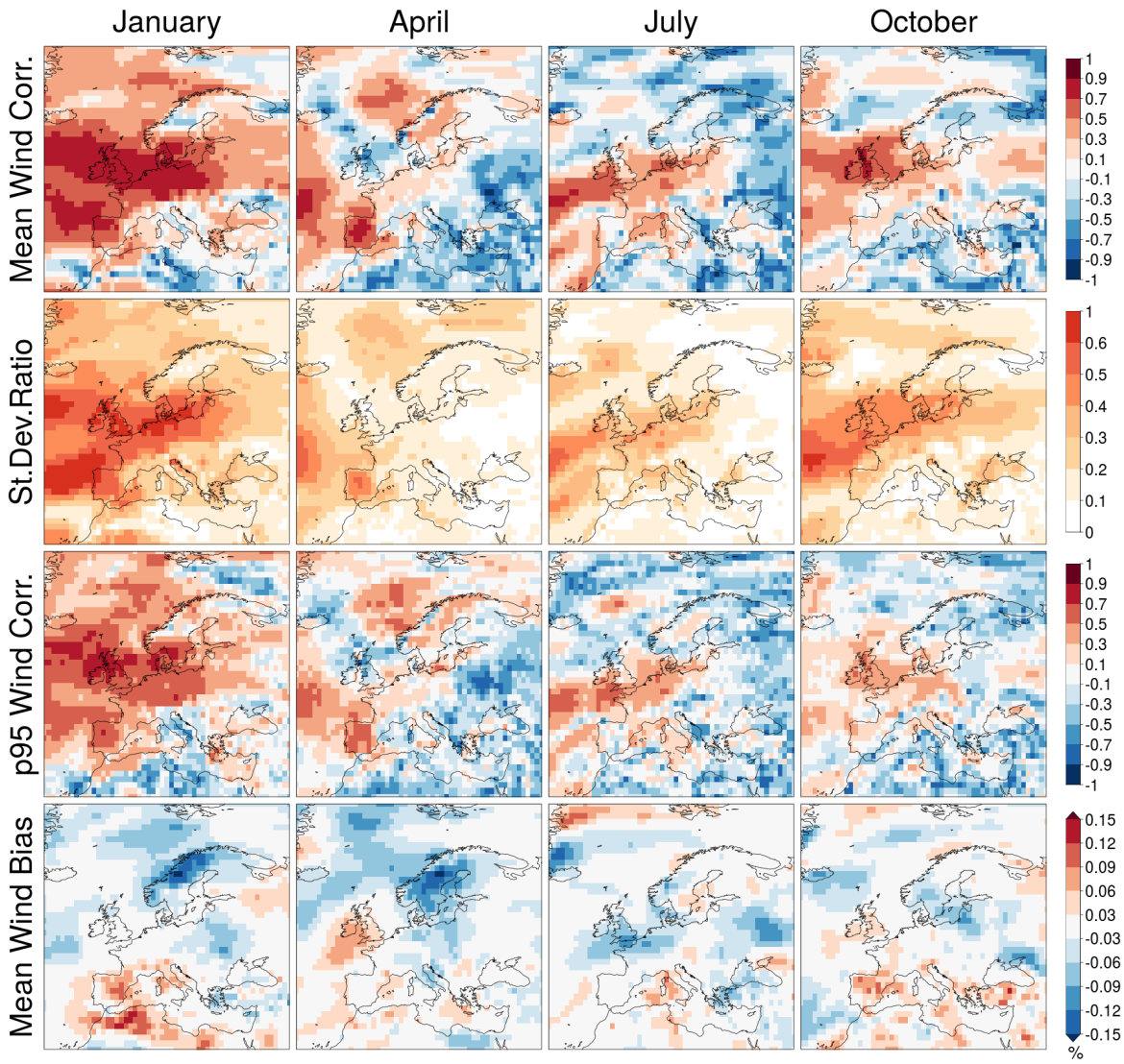

Figure 5. Reconstructed 10-m wind speed for the months of January, April, July and October. Four error metrics are shown from top to bottom, respectively: the temporal correlation between the time series of the reconstructed and observed monthly mean wind speed anomalies, the ratio between the standard deviations of the reconstructed and observed time series of monthly mean speed anomalies, the correlation between the reconstructed and observed monthly 95th percentile of wind speed anomalies, and the mean wind speed bias. These results are based on the WRs and 10-m wind speed from the JRA-55 reanalysis.

4.4 Conclusions

This analysis investigates the potential of WRs to reconstruct wind speed in Europe. It complements the impact of each single WR on wind speed and it is critical to identify areas and months where WRs can be considered sources of predictability of wind speed.

The area around the North Sea, the UK, northern and western Europe appear to have wind speeds most impacted by the large scale circulation with highest correlations for both mean values and the 95th percentiles. This result is important since these are areas of large wind generation. The time of the year when better reconstructions are achieved is November to March.

There are regions as northern Scandinavia, southern Spain and the Aegean Sea with high average yearly wind speed and where the overall influence of WRs on windspeed is moderate to high. Moreover, due to their distance from the North Sea, wind generation in these regions is not negatively affected by the passage of anticyclones over central Europe, where most of the European wind power generation in concentrated. Thus, these regions may play an important role in future electricity generation, by reducing the high intermittency of total produced European wind power, as already demonstrated by Grams et al. (2017).

This section has shown that there are European areas where there is potential to use the large scale patterns as predictors for surface wind speeds in long range predictions. However, the limitations still lie in the poor predictability of the WRs beyond 1 month.

5 Weather Regime and Targeted Circulation Types for European national-scale energy balance indicator forecasts at subseasonal timescales during winter

5.1 Introduction

In this Chapter we investigate the potential for both WRs and TCTs to be used to produce subseasonal forecasts of national scale energy balance indicators for lead weeks 1-4. The chapter begins by outlining the methods used to get national forecasts of electricity demand and wind power generation as well as the WRs and TCTs (Section 5.2) from the hindcast simulations. Following this a set of methods are discussed which are used to assess the predictability of the WR and TCT methods (Section 5.4). The potential of the methods to be used in a perfect forecast model setting (similar to that discussed in Chapter 4) are shown in Section 5.5. The ability of the ECMWF and NCEP hindcasts to predict demand-net-wind (DNW) in lead weeks 1 to weeks 4 is assessed in Section 5.6. Concluding remarks and avenues for future research are presented in Section 5.7.

5.2 Forecast methodology for national energy indicators

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 D3.2 Annex A.

5.2.1 Meteorology-to-power conversion models: ERA5

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 D4.1

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.

5.3 Weather regimes and targeted circulation types (TCTs)

Weather regimes: Weather regimes are calculated closely following the methods outlined in Deliverable D3.2 (using 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 2019). This method used here matches that in D4.2 (i.e., is an update to the method originally described in Deliverable D3.2) and provides a consistent single set of regimes relevant to the whole period November-March, rather than changing patterns each month. This enables easier computation and provides a larger amount of data for skill assessment of the hindcasts which only have an 11 year common period (1999-2010).

Targeted Circulation Types - definition: TCTs are constructed analogously to the weather regimes (i.e., using k-mean clustering) but with input time-series corresponding to daily maps of an impact variable rather than a gridded meteorological data set. In the present application this impact corresponds to national power system indicators (e.g., national demand) for each of 29 European countries over the extended winter November-March from 1980-2018. Each power system variable (i.e., demand or residual load) is normalised prior to the calculation of the TCTs such that a value of +1 in the resulting time series therefore indicates a +1 standard deviation departure above the norm for that day of the year for that country (see Deliverable 3.2 for further details).

Targeted Circulation Types – allocation in NWP forecasts: In order to allocate the TCT patterns in the hindcasts the 500hPa geopotential height composite from each TCT is calculated and the TCTs are then each day is assigned to the Z500 pattern with minimum Euclidean distance (a similar approach is used for WR assignment in operational forecasting). In the analysis below, the ensemble mean Z500 is used (rather than classifying individual ensemble members). This was shown in Deliverable 4.2 (chapter 4) to provide the highest success ratio for both WR and TCT assignment in both the ECMWF and NCEP hindcasts, though performance degrades at longer lead times (beyond 1-2 weeks).

5.4 Pattern forecasting methodologies

To assess the performance of TCT-based forecasts, a number of forecast methods are defined. In each case, the target predictand is the European national weekly-mean normalised demand-net-wind (DNW), though this process has also been completed for normalised demand forecasts (not shown as the results are similar to those for DNW). The methods are shown schematically in Figure 6 and a description of each is given below:

1. **Grid point forecast:** A direct forecast of national-aggregate DNW using gridded surface weather variables from the NWP forecast, (e.g., the NWP forecast surface data for week 3 is converted into a DNW estimate for week 3 following the methods described in Section **Error! Reference source not found.**). Grid point forecasts are

included in this chapter as a benchmark to indicate how well pattern-forecast based methods perform against raw grid-point methods (the general performance of grid-point forecasts has already been assessed in Deliverable 4.1).

2. **Climatology:** The weekly-mean climatology of DNW is calculated from the national level normalised DNW data from ERA5. This seasonally-varying climatological mean can be interpreted as the simplest "no new information" forecast.
3. **Persistence:** This is calculated by taking the weekly average of the seven days of "observed" DNW prior to the forecast intitialisation date (i.e., it effectively corresponds to persisting last week's weather/DNW forward to the relevant forecast validity time).
4. **Persistence 52:** This is calculated by using the weekly averages from the previous year of ERA5 for DNW and using this as the weekly-mean forecast for lead weeks 1-4.
5. **Dominant WR:** This is the WR that occurs most frequently during the forecast week (i.e. the forecast's modal regime from the seven day period). To calculate this forecast, each day's ensemble-mean Z500 anomaly (from the NWP model's lead dependent seasonally-varying climatology) is first assigned into one of the four canonical WR patterns (themselves defined from ERA5). Once a dominant WR is determined, the weekly mean DNW forecast is given by the corresponding the ERA5 DNW associated with that WR.
6. **Mixed Weather regime:** This forecast is similar to the "dominant WR" but corresponds to weighting the ERA5 surface impact responses of the WRs by the number of times they are forecast to occur in the target week. This is calculated from the ensemble-mean hindcast of Z500, with weightings giving to each of the seven days in the forecast week.
7. **Dominant TCT:** This is calculated in the same way as dominant WR but using the ERA5 TCT Z500 patterns and corresponding DNW responses.
8. **Mixed impact pattern:** This is calculated in the same way as mixed WR but using the ERA5 TCT Z500 patterns and corresponding DNW responses.

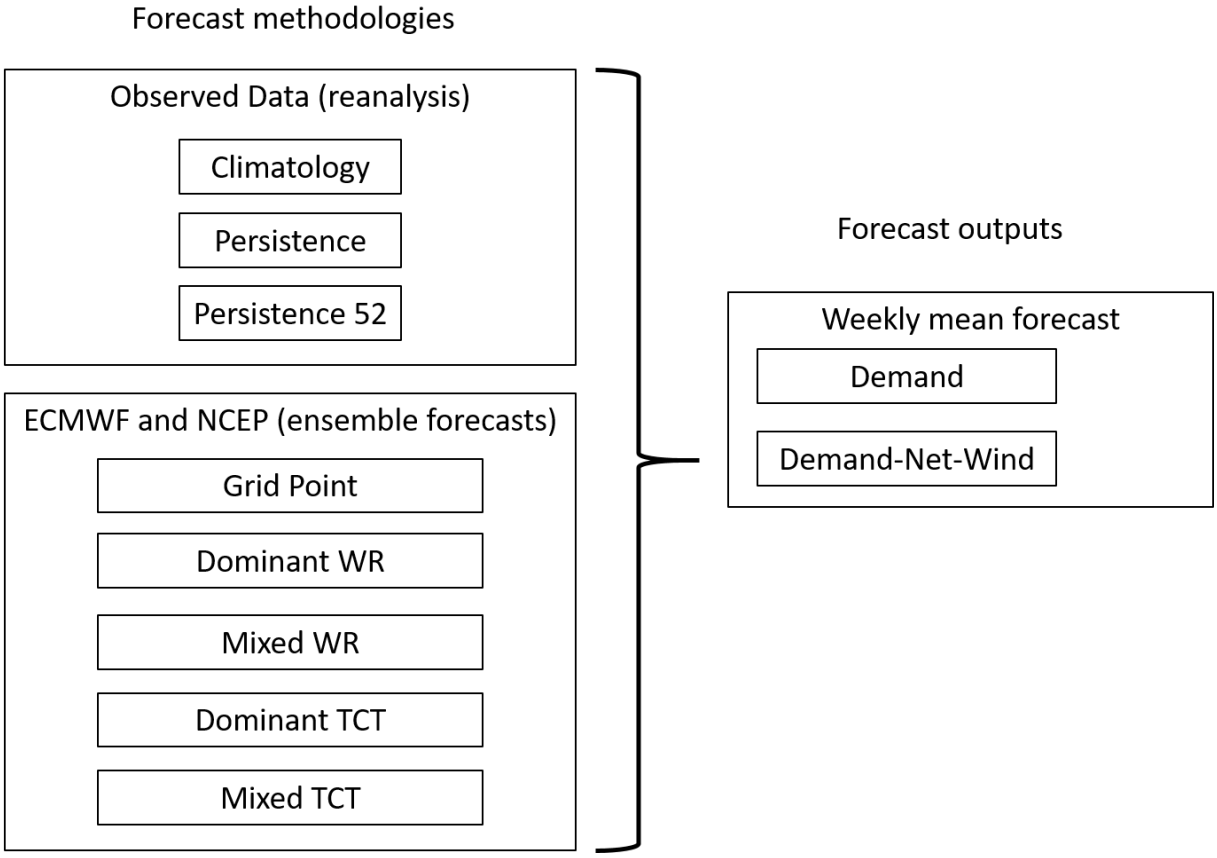

Figure 6. Schematic description of the forecast process (as presented in Chapter 5).

In order to assess the skill of these methods of forecasting the Mean Absolute Error (MAE) and Anomaly Correlation Coefficient (ACC) are used. These two metrics are chosen as the ACC gives a sense of if the methods are able to forecast the correct sign of the DNW anomaly, whereas MAE provides more information on the magnitude of the error associated with the forecast.

The analysis focuses on the skill of the European total DNW forecast, as well as a set of six case study countries which have been chosen due to (a) their geographic diversity, and (b) their contrasting magnitudes of DNW anomalies in both the weather regimes and TCTs (i.e., some case study countries exhibit a strong response to WRs/TCTs whereas others are rather weaker; see Deliverable 4.2 Chapter 5 for more details).

5.5 A perfect forecast model of National DNW

Section 5.5 first assesses the ability of the methods in a perfect forecast model using the ERA5 data. (i.e. if we have a perfect forecast of the WRs/TCTs, how well are we able to capture the DNW response?). Following this the ECMWF and NCEP hindcasts are assessed in Section 5.6.

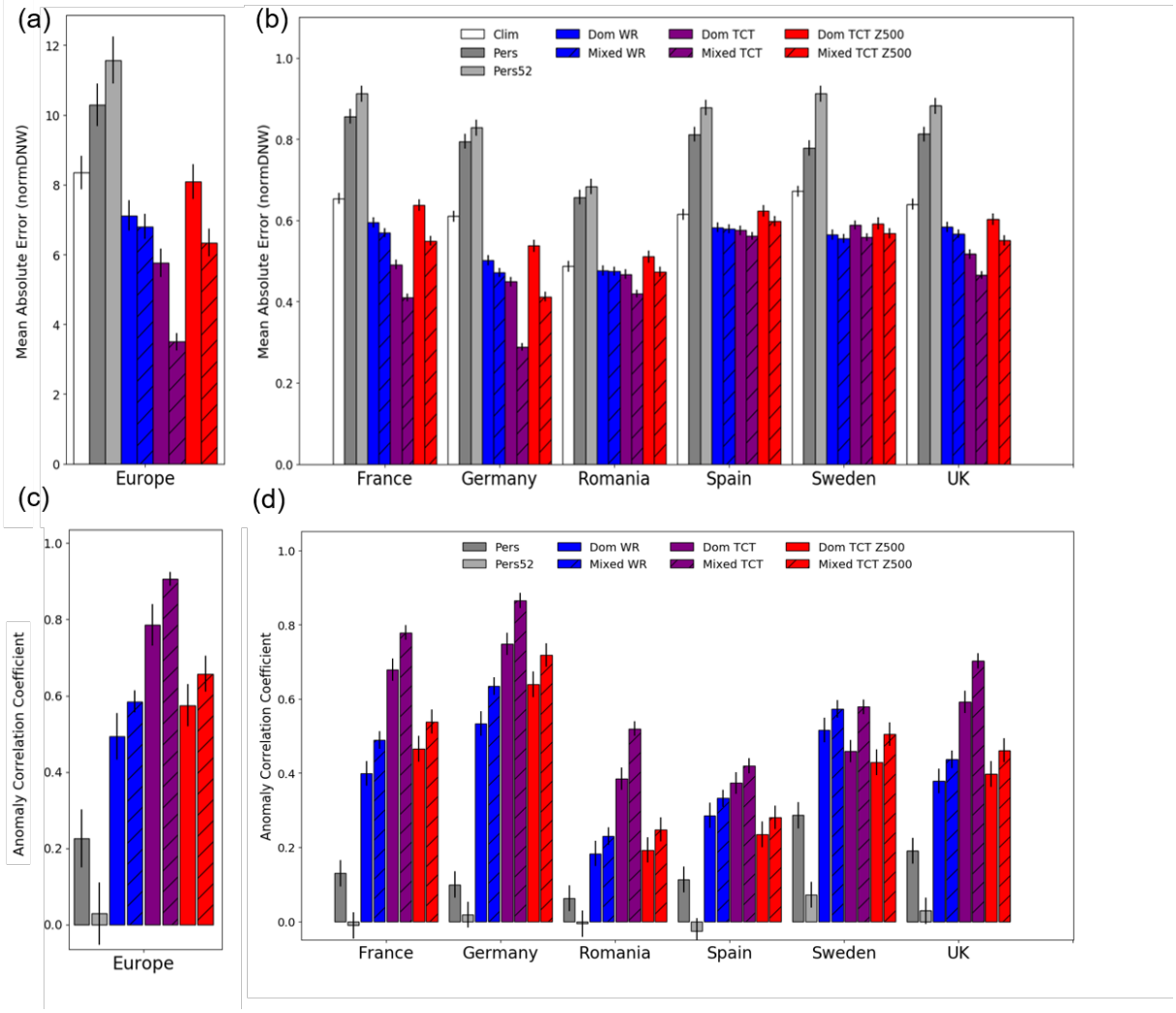

Figure 7. ERA5 DNW perfect forecast tests for the European total (a and c) and six case study countries (b and d). (a-b) Mean Absolute Error (MAE) (c-d) Anomaly Correlation Coefficient (ACC). Error bars show significance of results based on 2000 bootstrapped samples of each year of data.

Figure 7 shows the ability of the methods outlined in Section 5.4 to forecast national normalised DNW if the WR/TCT were known. These tests therefore demonstrate the maximum potential of the WR/TCT based forecasting methods. Focusing first on the European total, we see that both the WR and TCT based methods are able to outperform both a climatological and persistence forecast. The climatology forecast is significantly more skillful than persistence (for MAE, panel a) so this is taken as our benchmark against which the WR/TCT methods are tested. The Mixed WR and mixed TCT method perform better than the Dominant weather regime and Dominant TCT method in terms of both MAE and ACC. In general, however, both the WR/TCT methods dramatically outperform climatology (and persistence), suggesting a potential use for subseasonal forecasting – provided NWP forecasts are themselves able to assign the correct WR or TCT.

Similar results are seen for the six individual case study countries with marked improvements over climatology when using TCTs in Germany, France and the United Kingdom (UK) and some improvements in the other countries. It is noticeable when all 29 countries are examined that the largest potential improvement for the WR/TCT methods appears to be in Central Europe. This could be due to this being geographically where the bulk of the 29 countries are located (due to a number of small countries being located here), compared to the larger, more sparsely located Southern and Northern European countries. The WR and TCT methods are not equally skillful in all countries. For example, in Spain and Sweden the WRs have higher ACC and lower MAE than the TCT patterns assigned on Z500 (see Figure 5.1 subplots b and d). Whereas in most other countries the TCT method offer greater potential improvements compared to a climatological forecast than the WRs.

A key limitation with TCTs, however, lies in the ability to allocate forecast days to particular patterns. Here, a method resembling the WR method (Cassou 2008) is adopted based on the use of Z500. In particular, each TCT (defined by applying clustering algorithms to maps of national-aggregate DNW) is associated with a composite Z500 circulation pattern: it is this Z500 circulation pattern (rather than DNW data directly) which is used to select which TCT is said to be occurring on any given forecast day. As Z500 does not uniquely constrain the surface meteorological conditions (and hence the DNW), this leads to some days being "misclassified" in terms of the TCT patterns assigned (see Deliverable D4.2). The result of this is that the skill levels achieved using the Z500 TCT assignment (shown by the red bars in Figure 7) are comparable to (and not much better than) the skill levels achievable from the original WR patterns in most countries. It is, however, noted that the choice of using Z500 for TCT pattern assignment is arbitrary. As such, other meteorological fields could be considered or included in order to achieve "better" classification system. In the limiting case of a perfect classification scheme (whereby all forecast days are correctly assigned to TCTs), the performance is significantly enhanced, as shown by the purple bars in Figure 7.

In summary, in a perfect forecast simulation the WR and TCT methods provide more skill than both climatology and persistence forecasts for European and national, normalised DNW. The mixed TCT method is generally the best performing of all of the method but the results are country dependent and, in the case of TCTs, highly dependent on accurately assigning forecast days to the correct TCT types.

5.6 Subseasonal forecasts of DNW

5.6.1 The impact of variance inflation on forecast skill

Two methods of calibration have been implemented on the essential climate variables from the hindcasts used to create national demand and wind power generation. The first corresponds to the standard approach of a lead time dependent mean correction, whereas the second combines the lead time dependent mean correction with variance inflation (see Section 5.2 for more details). Figure 8 shows the impact of including variance inflation on the

MAE for a whole European forecast of normalised DNW for weeks 1-4. In all cases the more sophisticated bias correction scheme including variance inflation leads to increased MAE (i.e., less skill). The grid point forecasts are more sensitive to the calibration scheme (especially in lead weeks 1 and 2) compared to the pattern-based methods, though all forecasts methods become more strongly affected by the inclusion of variance inflation at longer lead times.

As the inclusion of variance inflation leads to increased forecast errors, for the remainder of this chapter the "standard" calibration is used (i.e., forecasts are subjected to a lead-time dependent mean bias correction but not variance inflation).

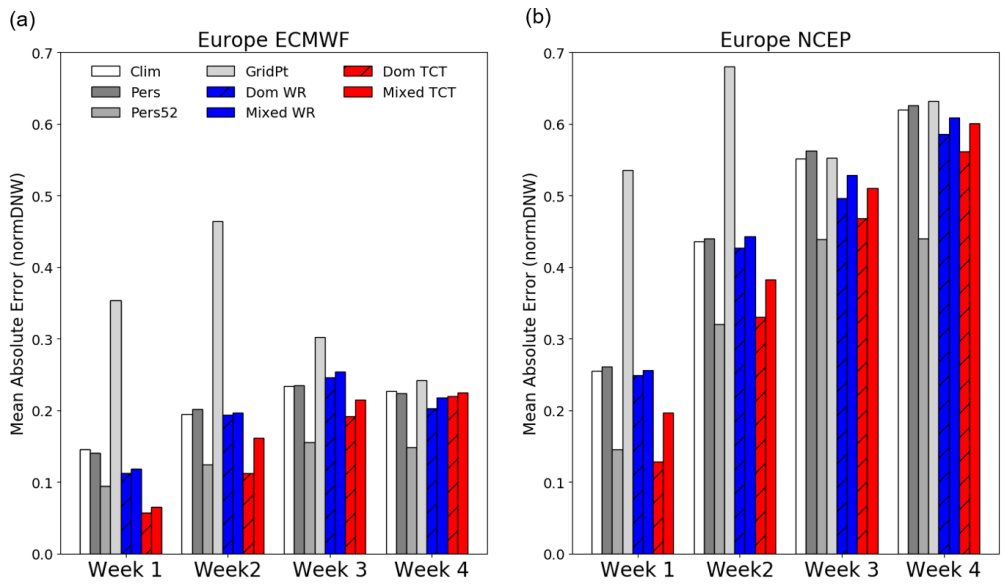

Figure 8. Difference in Mean Absolute Error (MAE) between forecasts of European total DNW with and without variance inflation included in the bias correction (see Section 5.2 for details of the method). Positive numbers imply that variance inflation has resulted in a larger MAE (and therefore has degraded forecast quality).

5.6.2 WR and TCT pattern forecasts in subseasonal models

Figure 9 shows the anomaly correlation coefficient for the various forecasting methods described in Section 5.4 for lead weeks 1-4, for Europe, and the three Central European case study countries. All results are in generally good agreement between the ECMWF and NCEP forecast systems, however there tends to be moderately more skill in the ECMWF system.

Focusing first on the European total skill (Figure 9 a and b), we see that for all lead times the grid point forecast provide skill relative to a pure climatological forecast (i.e., ACC values are always positive) and, for most lead times, the grid-point forecast is the most skillful forecast method. However, the skill of the grid-point method decays rapidly with lead time and the pattern-based methods (both WR and TCT) have a comparable (or better) performance to the grid-point forecast at longer lead times. In the NCEP system, greater skill than the grid point forecast is found for the WR and TCT methods in week 3 and 4 (Figure 9b) whereas for the

ECMWF system the pattern methods have very limited success (only the mixed WR method in week 4 outperforms the grid-point forecast; Figure 9a). Additionally, all the forecast methods outperform persistence forecasts at all lead times and the "mixed-pattern" methods outperform the "dominant-pattern" method in general (as seen in the perfect forecast simulations, Figure 7). At the gross pan-European scale, the WR and TCT methods offer only modest benefits over direct conversion of grid-point surface weather variables into energy-prediction at longer lead-times.

Greater benefit of the pattern-based approaches is, however, found if individual countries are assessed, particularly those in a zonal band across the middle of Europe (see Figure 9c-h). In France, Germany and the UK, the pattern-based methods (WR and TCT) already have comparable skill to the grid-point forecast in week 2, and either match or outperform the grid-point forecast at longer leads. The performance of the WR and TCT methods is similar but, as noted above (Section 5.5), the skill of the TCT forecast could be further improved with a "better" classification system creating a stronger large-scale constraint on the climate than provided by Z500 alone (e.g., in the "perfect pattern forecast" assessment the ACC was found to be ~10-50% higher for the "perfect classification" scheme compared to assignment using Z500; Section 5.5 and Figure 7).

For completeness Figure 10 shows the ACC for the other three case study countries, located in Northern, Eastern and Southern Europe. There is generally less skill in these countries than seen for the Central European countries, perhaps suggestive that the construction of the TCTs (which equally weight the nationally-aggregated DNW timeseries) emphasizes regions with clusters of small countries. There are areas/lead-times where the WR/TCT methods have more skill than the grid point forecast (e.g., Sweden lead week 3 and Spain lead week 4), indicating that there may be opportunities to exploit pattern based forecasts, but it is clearly important to evaluate this on a case-by-case basis and (for TCTs) to ensure that the initial design of the weather patterns is appropriate.

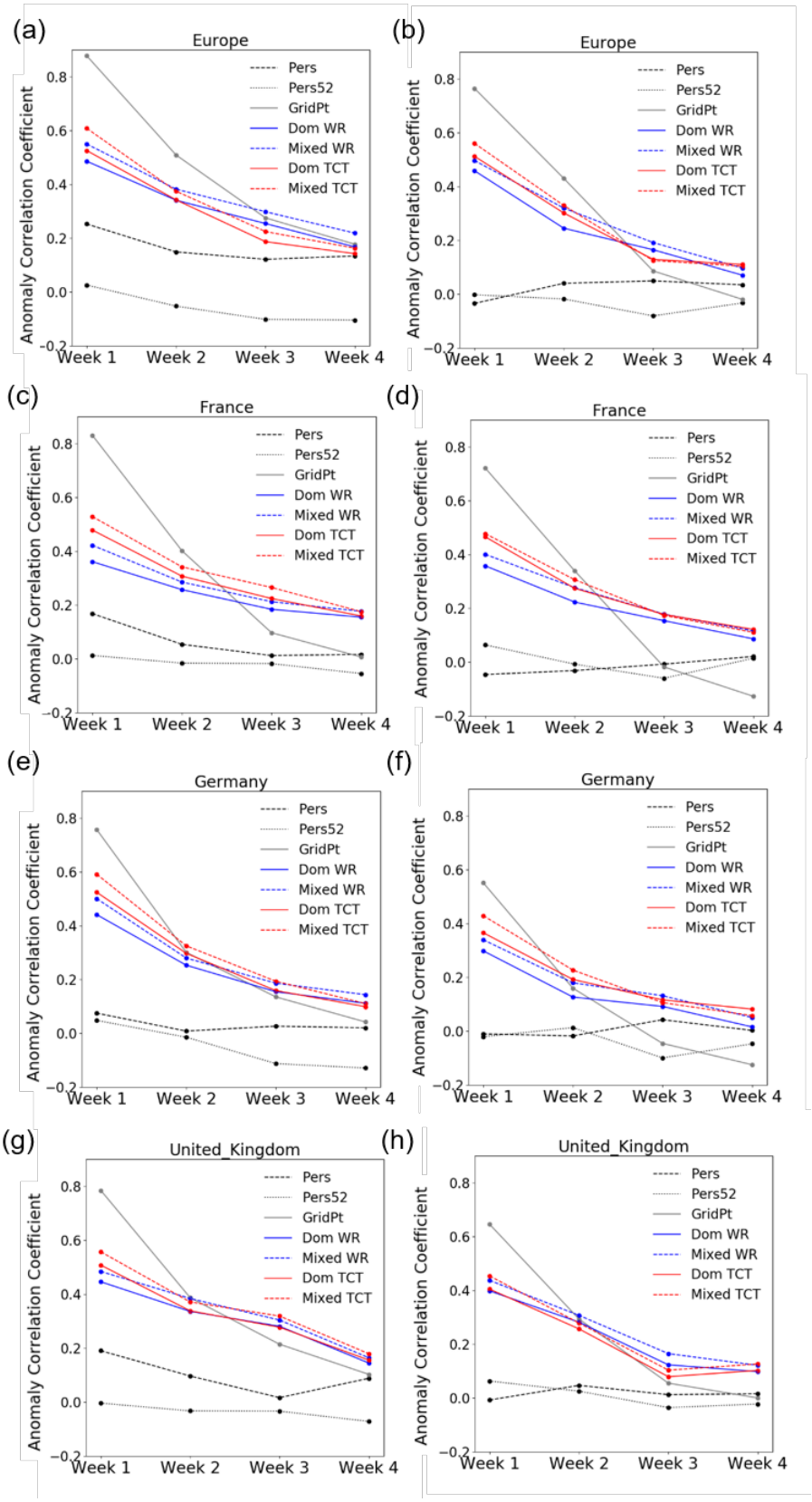

Figure 9. Anomaly correlation between normalised DNW forecasts and ERA5. ECMWF (left) and NCEP (right). (a-b) European total (c-d) France (e-f) Germany (g-h) United Kingdom. Note that, by construction, a climatological forecast has an ACC of zero for all lead times.

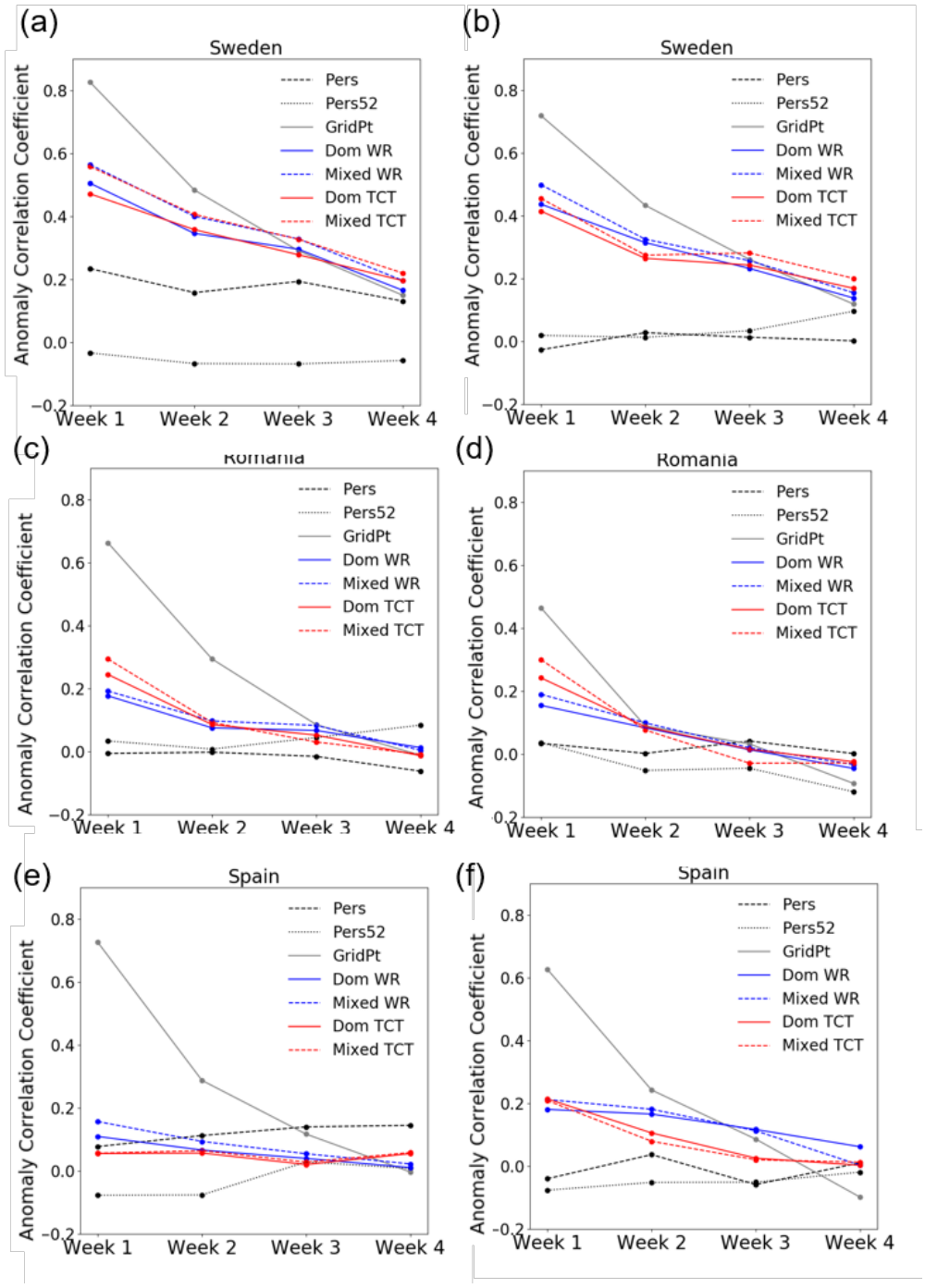

Figure 10. Anomaly correlation between normalised DNW forecasts using ERA5 and ECMWF hindcasts (left) and NCEP hindcasts (right) for lead weeks 1-4. (a-b) Sweden (c-d) Romania (e-f) Spain. Note that, by construction, a climatological forecast has an ACC of zero for all lead times.

5.7 Conclusions

The preceding sections have assessed the performance of pattern-based methods (WR and TCT) to produce skillful forecasts of winter time European-total and national-total DNW (and demand; not shown), providing a benchmark of their performance against alternative methods (grid-point, climatology and persistence). At short lead times (week 1 to 2) and for larger domains (whole-Europe) the most skillful forecast is typically provided by grid-point methods, i.e., using surface meteorological variables and converting them to DNW estimates. However, at longer lead times, the pattern-based methods can begin to outperform grid-point methods. In all cases, grid-point and pattern-based methods are typically better (or no worse) than climatological or persistence forecasts. Several points relevant to developing a practical application of the pattern-based methods are, however, worthy of note.

Firstly, regarding NWP forecast bias correction for surface weather variables. The method of bias correction used to calibrate the essential climate variables that are used as inputs for the demand and wind power models results in significant differences in forecast skill. Interestingly the more sophisticated bias correction method (including variance inflation) appears to result in a reduction in skill at all lead times. The reason for this reduction is not known at the present time but suggests that application of complex bias-correction and calibration schemes should be undertaken with great care.

Secondly, the assessment in this chapter focusses primarily on the simplest skill metric (anomaly correlation co-efficients): essentially a forecast of whether DNW is above or below normal. More complex metrics (e.g., MAE, or RMSE) are associated with generally lower skill scores (for all forecast methods). The existence of ACC skill in forecast weeks 3 and 4 does not therefore imply skill across these different metrics.

Thirdly, the skill by all the forecast methods is far from uniform across the European domain and between NWP forecast systems, suggesting careful assessment is needed on a case by case basis. It can also be observed that large-area spatial averaging does not always lead to enhanced forecast skill for pattern-based methods (e.g., where the pattern corresponds to signals of opposing across the domain). An interesting area for investigation is therefore the ability of grid-point vs pattern-based methods in forecasting the "structure" of energy anomalies across Europe (i.e., the spatial imprint of the weather) as opposed to viewing each country's DNW time series as independent.

Finally, while the performance of the WR and TCT approaches (where TCTs are assigned using a single field representing the atmospheric circulation, Z500) was found to be broadly similar, there are opportunities to improve TCT-based forecast skill through improved pattern-assignment methods (i.e., the method through which an NWP-forecast day is assigned into a particular TCT pattern). The mid-tropospheric circulation (represented by Z500) appears to be a fairly weak constraint on the surface weather and additional atmospheric variables (e.g., temperature in the lower troposphere) may therefore provide opportunities for more refined classification schemes.

6 Hydrological catchment-scale forecasting using hydrological weather regimes

6.1 Introduction

The aim of our work is to test hydrological seasonal forecasts that make use of hydrological weather regimes (HWR) to improve forecast skill. The hypothesis is that HWR information can be used to select analogue years from an ensemble of historical precipitation (P) and temperature (T) data with which to force a hydrological model to give an improved seasonal forecast of reservoir inflows (e.g. Foster et al. 2018; Olsson et al., 2016). The predominant seasonal forecasting method employed by the hydropower sector in Sweden is the Ensemble Streamflow Prediction (ESP) approach (Day, 1985). ESP uses an ensemble of historical observations of P and T, over the forecast period, to force a hydrological model resulting in a seasonal forecast which has a climatological-like evolution from the initial conditions. There have been attempts to improve on this approach, however it has been shown to be difficult to improve on (e.g. Foster et al. 2018; Olsson et al., 2016). This work compares cross validated hindcasts of inflows to 12 hydropower reservoirs in the Ume River system to determine if using HWRs, calculated from ECMWF seasonal forecasts, can improve the skill of seasonal forecasts of inflows to hydropower reservoirs.

6.2 Experimental Setup

To assess forecast skill, two sets of cross-validated hindcasts of seasonal inflows (Q) were made for 12 different hydropower reservoirs in the Ume River system in Northern Sweden. The first set was made using the well-established Ensemble Streamflow Prediction approach and the second set was made using a weighted ESP approach.

For the ESP hindcasts, a well-calibrated setup of the rainfall-runoff model HBV (Hydrologiska Byråns Vattenbalansavdelning; Bergström, 1976; Lindström et al., 1997) is initialized by running it up to the forecast issue date using observed P and T. This is done to ensure that the HBV model states reflect the current hydrological conditions in the basin with respect to e.g. streamflow, snow pack and soil moisture at the time of the forecast initialisation. This properly initialised model is then forced with an ensemble of historical T and P which covers the period from the forecast issue date and extending seven months forward. The same procedure was followed for the analogue or weighted ESP hindcasts. However instead of using the historical ensemble of P and T, a weighted historical ensemble was used to force the model.

The weighted historical ensemble was made by identifying analogues from the historical data and replicating their timeseries until the number of members in the ensemble matched that of the historical ensemble and then the two ensembles were pooled together. This gives an ensemble that still has the same spread as the historical ensemble but is weighted towards

the identified analogues. The analogues are selected by comparing the two leading HWRs in terms of occurrence frequency over the forecast period with the leading HWRs over the same period from the years in the historical period. We defined an analogue year to be one where the two leading HWRs over the forecast period are the same as those predicted in the seasonal forecast. In the case where no analogues are found, the weighted analogue approach uses the same forcing data as the 'standard' ESP i.e. the full historical ensemble.

The HWRs were classified using a fuzzy-rule-based classification to determine the circulation patterns common to the region during in the one to seven months following the forecast initialisation date (i.e., from month 0 up to month 6 of the NWP forecast) . An analogue year is defined as any historical year for which the two dominant HWRs in the same one to seven months period are the same as those found for the year in question (see Deliverable 4.2 for more information).

The evaluation first evaluates the naive skill of the standard ESP approach, this is to establish that this approach is skillful and a worthy benchmark for the new approaches. Afterwards, the seasonal forecasts were evaluated with skill scores, using the ESP approach as the benchmark, to investigate the skill of the new approaches. Two validation metrics were used in this work (Table 2). The mean absolute error skill score (MAESS) is a skill score that quantifies the relative forecast error against a reference forecast. The frequency of years (FY^+) is a metric that quantifies how often the forecast outperforms a reference forecast. The validation metrics are bootstrapped 5000 times to ensure robust results.

Name	Equation	Description
Mean absolute error skill score (**MAESS**)	$$\boldsymbol{MAESS = 1 - \frac{MAE_f}{MAE_r}}$$ where *f* is the modelled forecast and *r* reference or benchmark forecast.	Measure of the model's general performance; it quantifies the relative forecast error against a reference forecast.
Frequency of Years (**FY^+**)	$$\boldsymbol{FY^+ = \frac{100}{n}\sum_{y=1}^{n} H^y},$$ where *y* is the timestep and *n* is the total number of timesteps. *H* is the Heaviside function defined by $$H^y = \begin{cases} 0, & AE_r^y < AE_f^y \\ 1, & AE_r^y > AE_f^y \end{cases},$$ where *AE* is the absolute error, *y* is the timestep, *f* is the modelled forecast, and *r* reference or benchmark forecast.	Measure of the model's general performance; it quantifies how often the forecast outperforms a reference forecast.

Table 2. The validation metrics used to evaluate performance. The threshold for skill is 50% for FY^+ and 0 for all the other metrics.

The hydrological model used in this experiment is the Hydrologiska Byråns Vattenbalansavdelning hydrological model (HBV). It is a semi-distributed conceptual rainfall-runoff model which includes numerical descriptions of hydrological processes at the basin scale. The general water balance in the HBV-96 model can be expressed as:

$$P - E - Q = \frac{d}{dt}(SP + SM + UZ + LZ + LV)$$

where,

P = precipitation E = evapotranspiration

Q = runoff SP = snow pack

SM = soil moisture UZ = upper groundwater zone

LZ =lower groundwater zone LV = lake volume

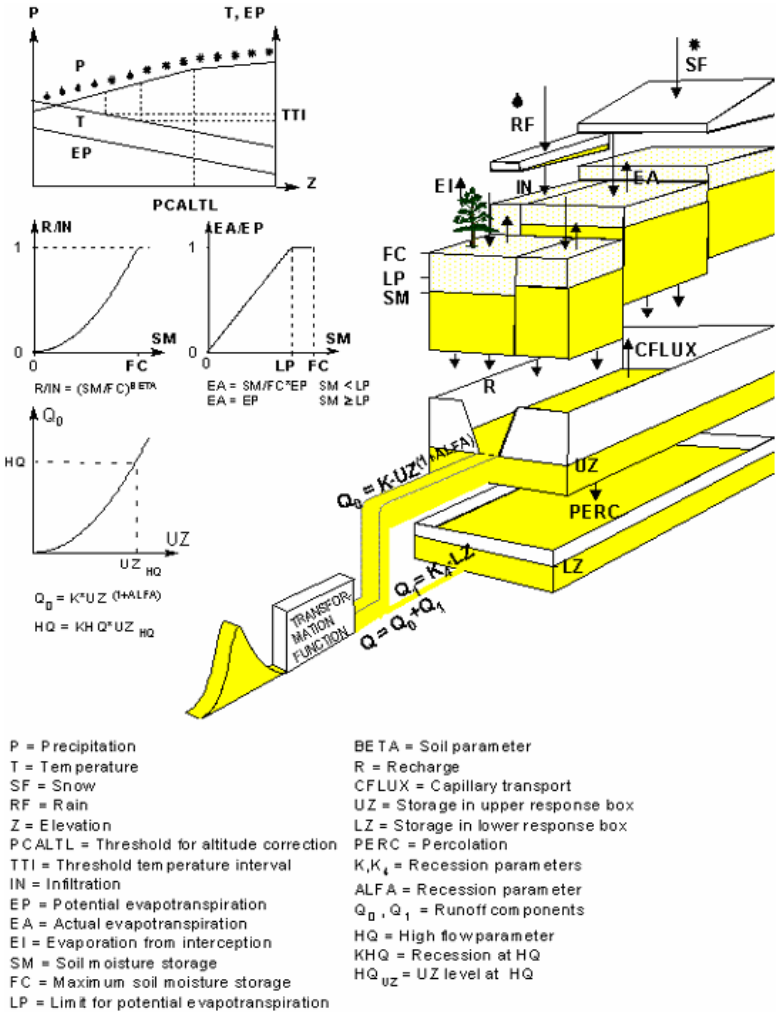

Figure 11. Schematic presentation of the HBV-96 model for a single basin (Lindström et al., 1997).

The model is normally forced with daily observations of P, T and monthly estimates of potential evapotranspiration. The model consists of subroutines for meteorological interpolation, snow accumulation and melt, evapotranspiration estimation, soil moisture accounting procedure, routines for runoff generation and finally, a simple routing procedure between sub-basins and in lakes. Basins with considerable elevation ranges can be subdivided into elevation zones which, if needed, can be further divided into different

vegetation zones (e.g., forested and non-forested areas). These subdivisions are made for the snow and soil moisture routines only. The model structure of HBV-96, with the most important characteristics, is presented schematically in Figure 11. For a more comprehensive model description readers are referred to Lindström et al. (1997).

6.3 Case study: Ume River

As a benchmark, the skill of the "standard" ESP method is calculated (Table 3) using climatology as the reference. In general, that the ESP outperforms climatology at predicting seasonal inflows to the reservoirs, both with respect to error and frequency. These results, while modest, have been shown to be very difficult to improve on.

	FY+	
Jan	50	0
Feb	51	9
Mar	56	18
Apr	56	27
Maj	61	36
Jun	61	45
Jul	56	55
Aug	52	64
Sep	61	73
Oct	56	82
Nov	51	91
Dec	61	100

	MAESS	
Jan	-0.054	-1.0
Feb	0.021	-0.8
Mar	0.099	-0.6
Apr	0.096	-0.5
Maj	0.110	-0.3
Jun	0.170	-0.1
Jul	0.089	0.1
Aug	0.037	0.3
Sep	0.074	0.5
Oct	0.064	0.6
Nov	0.004	0.8
Dec	0.123	1.0

Table 3. Bootstrapped (n=5000) FY^+ and MAESS, aggregate of all reservoirs, for the standard ESP approach (without use of HWR information) with climatology being used as the reference. The threshold for FY+ skill is 50% and MAESS skill is 0. Red shading indicates results below the skill threshold and blue indicates results above the skill threshold.

To assess the skill of the new weighted analogue approach, the validation metrics are calculated using the 'standard' ESP as the reference. The FY^+ results (Table 4) show the frequency with which the weighted analogue approach outperforms the ESP approach. The rows are the months when the hindcasts were initialized. The columns indicate the forecast months following the initialization date for which the HWR were defined (so, month 0 indicates the HWR were defined on the month 0 forecast data alone, whereas month 5 indicates the HWR were defined in based on the data from forecasts months 0, 1, 2, 3 and 4).

The MAESS results (Table 5) show how the weighted analogue approach performed against the ESP with respect to hindcast error.

	HWR classification period (months)						
	0	1	2	3	4	5	6
Jan	53	45	54	56	59	53	38
Feb	38	38	38	38	42	50	50
Mar	50	47	43	44	47	43	41
Apr	53	50	50	46	50	41	55
Maj	56	53	47	44	42	47	47
Jun	63	56	53	50	47	40	41
Jul	47	46	47	44	36	42	50
Aug	50	53	57	56	44	50	50
Sep	56	39	43	44	39	44	41
Oct	56	56	50	53	50	59	56
Nov	47	46	63	67	72	65	63
Dec	56	56	50	50	50	50	50

0 9 18 27 36 45 55 64 73 82 91 100

Table 4. Bootstrapped (n=5000) FY$^+$, aggregate of all reservoirs, for the analogue appro with the 'standard' ESP approach being used as the reference. Rows: month of hindcast initialization. Columns: indicates the forecast months for which the HWR were defined (e.g., column "2" indicates HWRs were defined across forecast months 0 and 1; column "5" indicates HWRs across forecast months 0, 1, 2, 3 and 4; see main text) . The threshold for skill is 50%. Red shading indicates results below the skill threshold and blue indicates results above the skill threshold.

	HWR classification period (months)						
	0	1	2	3	4	5	6
Jan	0.010	-0.012	0.010	0.008	0.015	0.039	-0.026
Feb	-0.036	-0.026	-0.006	-0.017	-0.005	-0.009	-0.016
Mar	-0.006	0.004	-0.009	-0.006	-0.013	-0.017	-0.018
Apr	0.013	-0.009	0.003	0.002	0.004	-0.006	0.012
Maj	0.003	0.015	-0.015	-0.003	-0.027	-0.005	-0.011
Jun	0.036	0.044	0.013	0.002	-0.007	-0.026	-0.023
Jul	-0.005	-0.023	0.000	-0.009	-0.038	-0.081	-0.016
Aug	0.001	0.044	0.017	0.010	-0.062	0.004	0.005
Sep	0.015	-0.028	-0.018	-0.008	-0.068	-0.010	0.001
Oct	0.006	0.006	-0.008	0.008	0.013	0.005	0.009
Nov	0.002	-0.017	0.016	0.015	0.046	0.033	0.035
Dec	0.006	0.024	0.000	0.007	0.006	0.004	0.004

-1.0 -0.8 -0.6 -0.5 -0.3 -0.1 0.1 0.3 0.5 0.6 0.8 1.0

Table 5. Bootstrapped (n=5000) MAESS, aggregate of all reservoirs, for the analogue approach with the reference being the 'standard' ESP approach. Rows and columns: as Table 4. The threshold for skill is 0. Red shading indicates results below the skill threshold, white indicate results around the skill threshold, and blue indicates results above the skill threshold (the intensity of the shading indicates relatively how far away the results are from 0, light = near, dark = far).

These evaluation results of the evaluation of the weighted analogue approach against the ESP show mixed outcomes with the most consistent improvements in skill being found for the approach that uses HWRs classified in the one month following the forecast initialisation date. This is consistent with the findings reported in Deliverable 4.2 where the correlation between the ECMWF SEAS5 seasonal predictions and ERA-Interim reanalysis monthly frequency occurrence corresponding to the HWRs was highest for the first month after the initialization date (i.e., forecast month 0). This suggests that the ECMWF SEAS5 seasonal predictions over northern Sweden have limited value past the first month.

6.4 Conclusions

The use of HWRs, derived from ECMWF SEAS5 seasonal predictions, can improve the forecast skill of seasonal inflows to reservoirs in the Ume River system. The improved skill is most consistently achieved when analogues are selected using HWRs for the month following the forecast initialization date (i.e., forecast month 0). However, these results are modest. As the analogues are selected based on the weather regimes in the first month following the forecast initialization date only, the analogue selection is based on a very limited selection of the hydroclimatic period during which the inflows are generated. Future studies should evaluate whether analogues selected using HWRs classified over a period both before and after the forecast initialization date.

7 Use of remote climate indices to condition European national-scale energy balance indicator forecasts

7.1 Introduction

As was noted in Section 2.2.2, sudden stratospheric warmings can be followed by an increased frequency of cold air outbreaks over northern Europe and a southward shift of the storm track. These in turn will affect European energy demand and wind power. Since the tropospheric effects of an SSW event can persist for up to 60 days (Baldwin and Dunkerton, 2001), there is potential for enhanced predictability of energy variables in the period following an SSW. This chapter investigates whether such an enhancement can be seen in the subseasonal hindcasts from ECMWF (elsewhere referred to as MFS) and NCEP in the S2S database.

7.2 Method

Subseasonal hindcasts of energy demand and wind power were calculated from hindcasts of 2m temperature and 10m wind speed as described in Section 5.2 following the mean bias correction process. Hindcasts from both ECMWF (i.e., MFS) and NCEP were used. The ECMWF hindcasts cover the years 1996 to 2015 (encompassing 14 SSWs), and the NCEP hindcasts cover the years 1999 to 2010 (12 SSWs) Since SSWs are a winter phenomenon, only hindcasts initialised in the months November to March were considered. Unlike in Chapter 5, all hindcasts initialised during these months were used, and no lagged ensembles were created. A 28 day mean of the energy variables was taken over the full four weeks (days 5-32) of the hindcast, and it was this mean that was verified.

The hindcasts were verified against energy demand and wind power calculated from ERA5. Four verification metrics were computed: the Brier score for the lower decile, the Brier score for the upper decile, the ranked probability score (RPS) for terciles, and the continuous ranked probability score (CRPS). The scores for hindcasts initialised at or soon after an SSW event were compared to the scores for hindcasts initialised at other times. The dates of SSWs were taken from the catalogue described in Section 2.2.2. For a hindcast to be associated with an SSW, the SSW had to satisfy three conditions: (i) it had to occur at or before the start date of the hindcast, (ii) it had to occur after the start date of the preceding hindcast, and (iii) it had to occur at most five days before the start date of the hindcast.

Note that the verification metrics chosen were plain scores, not skill scores. The reason for this is that such scores are sample means of values calculated for each individual hindcast. This enables a two-sample t-test to be applied to gauge the significance of the difference of the scores. A difference is considered statistically significant if it fails the t-test at the 5% level. The results of this test should be treated with caution because the samples could deviate from the null hypothesis by being non-Gaussian or serially correlated as well as by having

different means, but nevertheless provide some suggestion of where more (or less) confidence can be placed in the observed signals.

7.3 Results

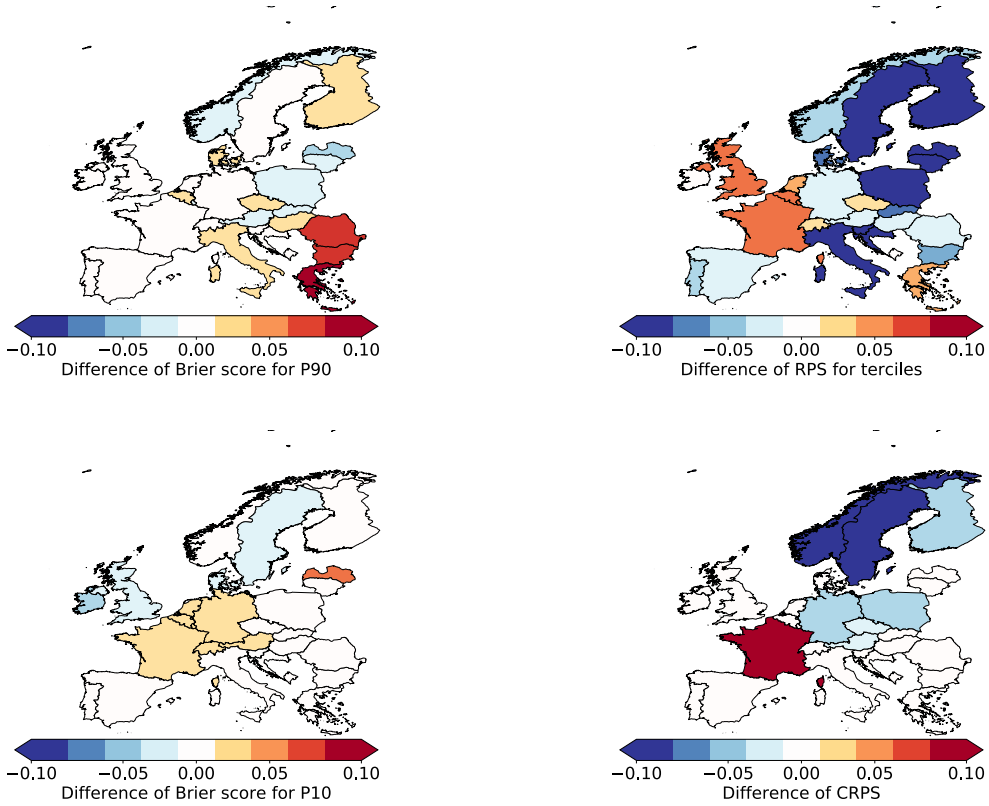

Figure 12. Differences in the verification metrics for ECMWF demand hindcasts using mean bias adjustment. Top left: Brier score for upper decile. Bottom left: Brier score for lower decile. Top right: RPS for terciles. Bottom right: CRPS.

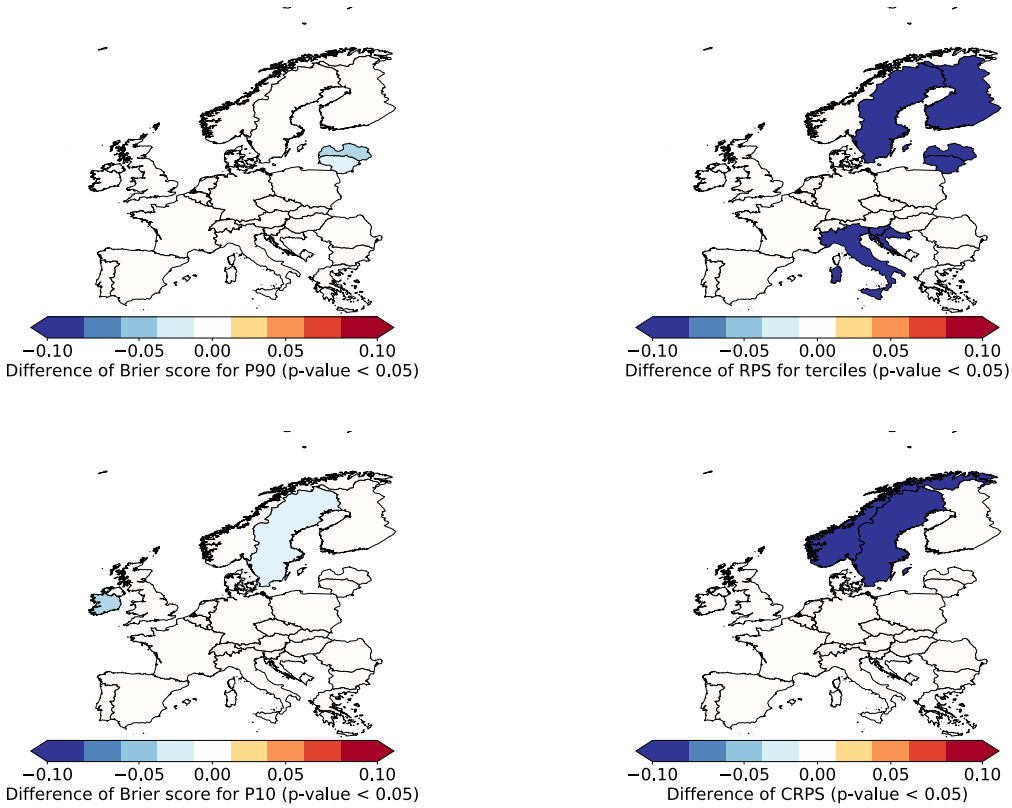

Figure 13. Statistically significant differences in the verification metrics for ECMWF demand hindcasts using mean bias adjustment. Layout as in Figure 12.

Figure 12 shows the differences in the four metrics for the ECMWF demand hindcasts using mean bias adjustment (i.e., without the optional variance inflation described in Section 5.2)without variance inflation. The metric in the no SSW case is subtracted from the metric in the SSW case. Since these are negatively oriented metrics (smaller is better), a negative difference means that the hindcasts are better in the SSW case. It can be seen that the differences can be positive or negative. However, when we concentrate on the statistically significant differences, all the differences are negative, indicating an improvement in hindcasts in the SSW case (Figure 13). There is little consistency between the different metrics as to which countries improve in the SSW case. Sweden improves in three of the four metrics, but otherwise the improvements are scattered between Ireland in the west and Finland in the east, and Italy in the south and Norway in the north.

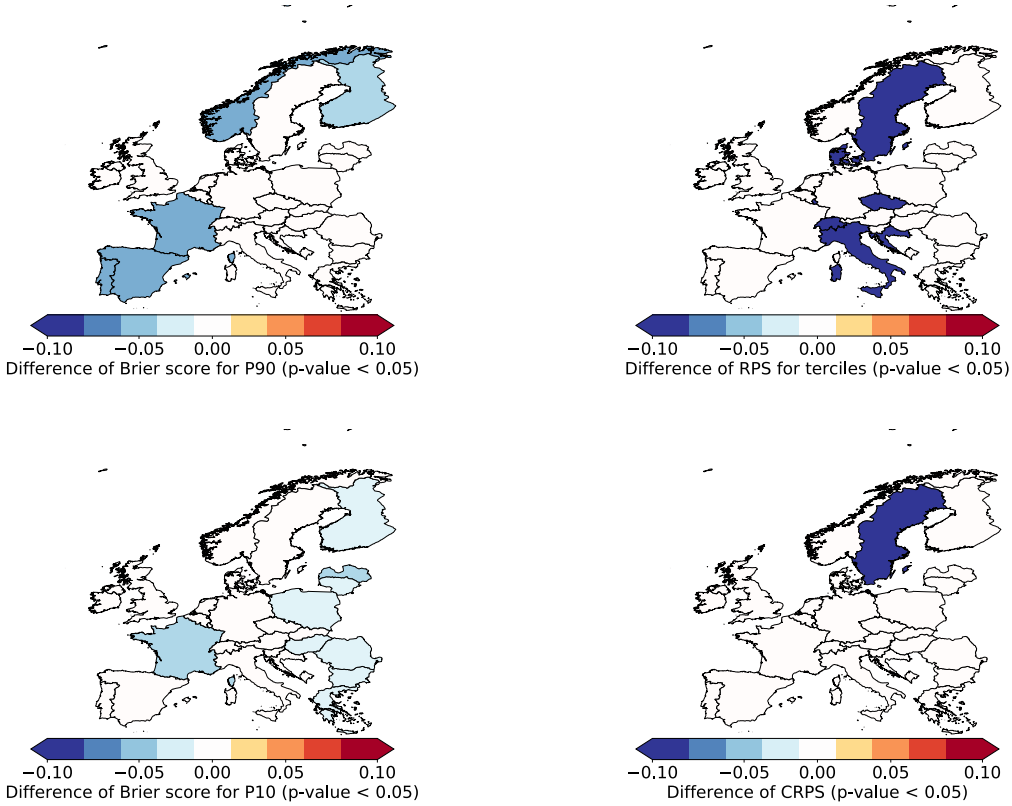

Figure 14. Statistically significant differences in the verification metrics for NCEP demand hindcasts without variance inflation. Layout as in Figure 12.

Figure 14 shows the statistically significant differences in the four metrics for the NCEP demand hindcasts without variance inflation. Again they are all negative, and again there is little spatial consistency between the four metrics. If anything, the improvements are more scattered than in the ECMWF hindcasts, with no country showing an improvement in more than two metrics, and the south western countries of Portugal, Spain, and France now added to those showing improvements.

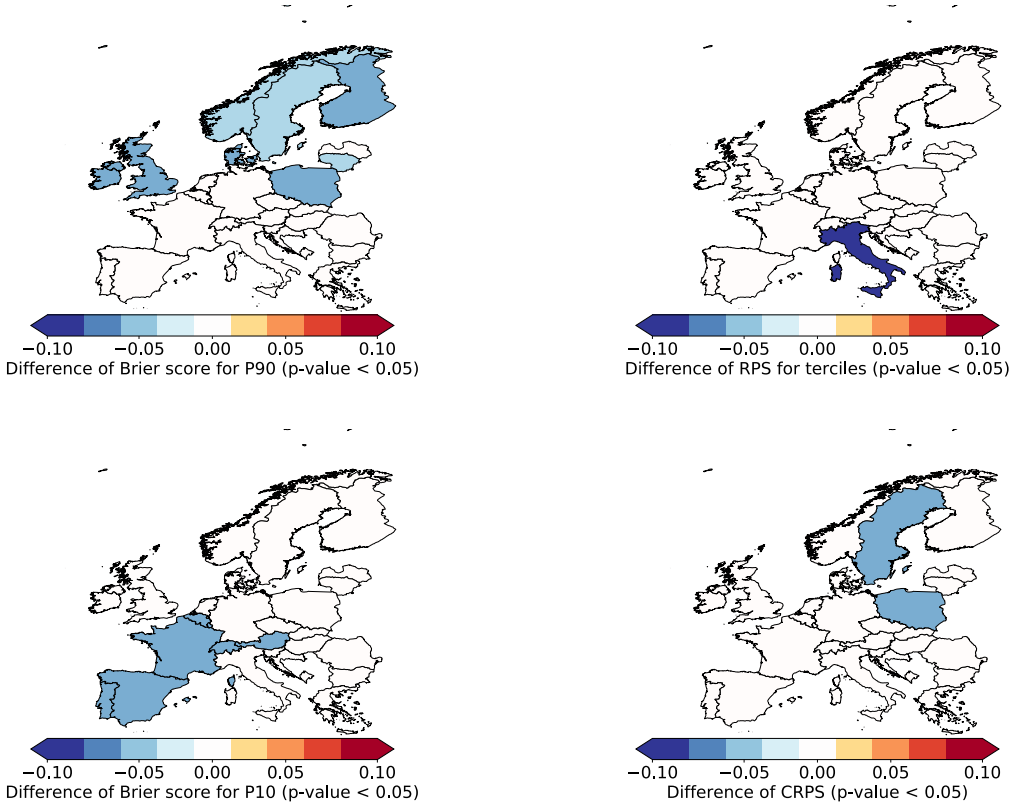

Figure 15. Statistically significant differences in the verification metrics for ECMWF wind power hindcasts without variance inflation. Layout as in Figure 12.

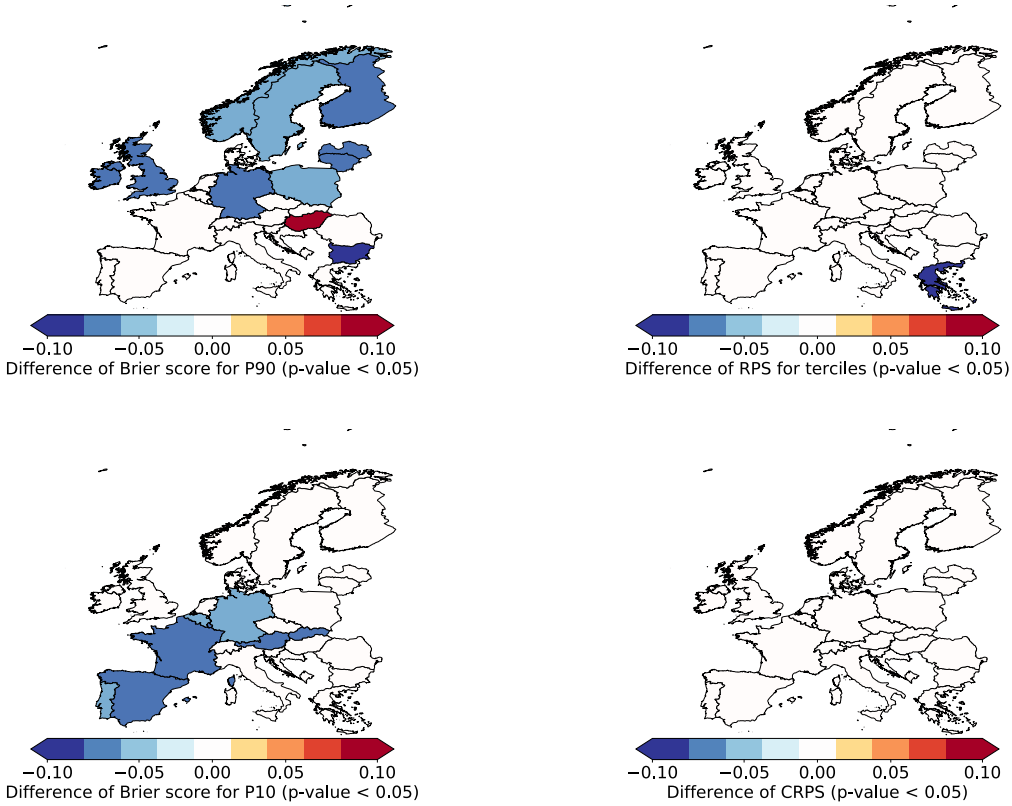

Figure 16. Statistically significant differences in the verification metrics for the NCEP wind power hindcasts without variance inflation. Layout as in Figure 12.

Figure 15 and Figure 16 show the statistically significant differences in the four metrics for the ECMWF and NCEP wind power hindcasts without variance inflation. In all but one case the differences are negative. There is some spatial coherence in the Brier scores, with the Brier score for the upper decile showing an improvement in the SSW case over the northern countries of Ireland, the UK, Norway, Sweden, and Finland, and the Brier score for the lower decile showing an improvement in the SSW case over the south western countries of Portugal, Spain, France.

The results from the hindcasts with variance inflation are qualitatively the same as those from the hindcasts without, although the details differ. All statistically significant differences are negative, indicating an improvement in the hindcasts in the SSW case. There is not much sign of a coherent spatial pattern in the improvements in the demand hindcasts, but the wind power hindcasts show an improvement in the Brier score for the upper decile over northern countries and an improvement in the Brier score for the lower decile over south western countries.

7.4 Conclusions

The hindcasts in the SSW case may be better or worse than in the no SSW case, but almost all the statistically significant differences in the verification metrics indicate an improved hindcast in the SSW case.

There is no coherent spatial pattern in the improvements in the demand hindcasts. This may be due to the sensitivity of demand to the details of each country's power system.

There is some spatial coherence in the improvements in the wind power hindcasts: in the SSW case the Brier score for the upper decile is better over northern countries and the Brier score for the lower decile is better over south western countries. Note that we expect lower winds over northern Europe and higher winds over south western Europe following an SSW, so these improvements indicate that the models are successfully capturing the reduced probability of the opposite extreme.

8 Conclusions

The analysis presented above investigates two areas where additional forecast skill may potentially be available in subseasonal and seasonal NWP systems. Firstly, the use of hybrid pattern-based forecasts (based on large-scale atmospheric circulation as opposed to use of grid-point surface climate variables or simpler benchmark forecasts such as climatology and persistence). Secondly, the potential for conditional predictability, here interpreted as changes in the predictive skill produced by forecast models during different prevailing atmospheric conditions (in this case, Stratospheric Sudden Warmings). This work builds upon previous deliverables in terms of technique and analysis (particularly Deliverable D4.2 but also D3.1, D3.2 and D4.1).

Several different approaches are tested, spanning a range of forecast systems and forecast time-scales. As the details of each scheme are 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 detailed findings. A brief recap of the main findings from each chapter is, however, given below before a concluding synthesis.

EATCs and seasonal NWP (Chapter 3). A hybrid pattern-based seasonal forecast is constructed by first projecting the atmospheric circulation onto the four EATC patterns (NAO, EA, EAWR, SCA; see Deliverables D3.2 and D4.2), then exploiting observed statistical relationships between the EATCs and surface essential climate variables (wind, temperature, insolation). In perfect model experiments (where the EATC state is fully known), the multilinear model fit works well in many European regions (France and Germany for temperature, British Isles and North of Germany for wind, Central Europe and Iberian Peninsula for surface solar radiation). Correspondingly, the hybrid forecast (using predicted circulation from two different NWP seasonal forecast systems) is shown to offer improvements over grid-point forecasts and climatological forecasts over many areas of Europe (though the details depend on the variable and forecast system studied).

Seasonally-evolving WR and monthly wind speeds (Chapter 4). The potential utility of seasonally evolving WR patterns (a separate set of 4 WR is defined for each calendar month) in forecasting monthly-mean European surface wind speeds is investigated in a perfect model experiment (i.e., where the WR classification is known). The area around the North Sea, the UK, northern and western Europe appear to be most impacted by WR state, with this impact more pronounced in wind extremes and during winter. Other regions where WRs offer significant explanatory power are found in the Iberian Peninsula, Iceland, northern Scandinavia and the Agean Sea and to a lesser extent the Gulf of Lion (southern France). Nevertheless, a key limitation – as identified in Deliverable D4.2 - lies in the limited ability of NWP models to successfully predict these seasonally evolving WRs at lead times greater than week 1 or month 1 (subseasonal and seasonal respectively).

Winter TCTs and subseasonal NWP (Chapter 5). A hybrid pattern-based seasonal forecast is constructed by first assigning the atmospheric circulation into one of four TCT patterns (see Deliverables D3.2 and D4.2), then exploiting observed statistical relationships between the

TCTs and national-balance indicators (in this case, DNW). This is benchmarked against the performance of climatology, persistence, grid-point and "fixed" WR forecasts (i.e. a single set of four WR is defined for the whole extended winter season). For perfect-model experiments (where the TCT or WR assignment is known) TCTs are shown to offer considerable benefits over WRs. In experiments where NWP subseasonal forecasts are used to assign WRs and TCTs, the most skillful forecast is typically provided by grid-point methods at short lead times (week 1-2). However, at longer lead times the pattern-based methods begin to slightly outperform grid-point methods (the precise details depend on NWP forecast system, verification metric, methodology and geographical region). In all cases, grid-point and pattern-based methods are typically better (or no worse) than climatological or persistence forecasts. The WR and TCT pattern-based methods are found to have comparable performance though it is noted that a key limitation affecting TCTs is the weak constraint on the surface weather provided by geopotential height on 500 hPa (Z500): it is suggested that use of additional atmospheric variables (e.g., temperature in the lower troposphere) may provide opportunities for more refined TCT assignment and hence increase the available skill from a hybrid pattern-forecast based on TCT-style approaches.

HWR and seasonal forecasts (Chapter 6). A hybrid pattern-based forecasting method is applied to predict streamflow in a test river catchment in Sweden. The HWR-based method is used to provide "weights" to members of a streamflow ensemble forecast. This method is compared to existing streamflow ensemble forecast techniques. These existing techniques have historically proven to be a hard benchmark to outperform but, through the use of HWRs, the ECMWF seasonal forecast is shown to provide modest improvements in forecast skill in the first month after the forecast initialisation (i.e., lead time 0).

SSW-conditional predictability in subseasonal forecasts (Chapter 7). The impact of SSW events on the forecast skill is investigated in the ECMWF and NCEP subseasonal systems. The effect is shown to be subtle and, during SSWs, forecast performance in any given country may be slightly improved or degraded. However, in almost all cases where the difference in performance was shown to be statistically significant, the verification metrics indicate improved performance following an SSW event. The spatial pattern of improvement is found to be more coherent for wind power - particularly extremes - than demand. A key limitation, however, is the small sample size of SSW events for which archived hindcasts are available.

In general, the results suggest that pattern-based forecast schemes can indeed offer modest potential benefits over direct grid-point based methods, though the details of this are highly dependent on the precise methods and verification statistics applied. Compared to a purely grid-point approach the EATC method predicting the 3 month average circulation appears to offer additional skill in month 1 (3 month average seasonal forecasts, Chapter 3). The HWR method offers some skill improvements in month 0 (monthly streamflows from seasonal forecasts, Chapter 6). TCTs and to a lesser extent the simpler "fixed" WR scheme appear to offer the potential for additional skill in each of weeks 2-4 (1-week average subseasonal forecasts, Chapter 5). SSWs appear to be associated with forecast windows of slightly more

skill in some regions of Europe (monthly timescales, Chapter 7). Each technique therefore offers some opportunity for skill improvements if carried through to an operational setting.

As emphasized in Deliverable D4.2, however, it important to recall that the skill advantages provided by pattern-based techniques are generally quantitative and incremental in nature rather than representing a qualitative step-change. Moreover, as demonstrated both here and in Deliverable D4.2, there are many subtleties in the detailed design of these pattern forecasting techniques that can strongly influence the resultant forecast skill. This therefore confirms again that the process of the design of the circulation patterns must be closely integrated with the process of forecast skill assessment (e.g., through repeated iterations in design) if these developments are to be carried through to an operational setting. Furthermore, skill must be assessed on a case-by-case basis: skill in a particular performance metric or region should not be taken to imply skill in all regions or metrics.

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/doi.org/10.5194/hess-22-2057-2018

Baldwin, M.P. and Dunkerton, T.J. (2001). Stratospheric harbingers of anomalous weather regimes. Science, 294, 581-584.

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

Bergström, S. (1976). Development and application of a conceptual runoff model for Scandinavian catchments (No. 52). Department of Water Resources Engineering, Lund Institute of Technology, University of Lund.

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.

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

Bueh, C., Nakamura H., (2007). Scandinavian pattern and its climatic impact. Quart. J. Roy. Meteor. Soc. 133, 2117–2131.

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., Terray, L., Hurrell, J. W., & Deser, C. (2004). North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing. *Journal of Climate*, *17*(5), 1055–1068. https://doi.org/10.1175/1520-0442(2004)017<1055:NAWCRS>2.0.CO;2

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

Castro-Diez, Y., Pozon-Vasquez, D., Rodrigo, F. and Esteban-Parra, M. (2002). NAO and winter temperature variability in southern Europe. Geophys. Res. Lett. 29, 1-1–1-4. doi:10.1029/2001GL014042.

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

Day, G. N. (1985). Extended Streamflow Forecasting Using NWSRFS, *J. Water Resour. Plan. Manag.*, 111, 642–654.

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

Doblas-Reyes, F.J., Renate, H., and Palmer, T.N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination. Tellus, 57A, 234-252.

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.

Foster, K., Uvo, C.B. and Olsson, J. (2018). The development and evaluation of a hydrological seasonal forecast system prototype for predicting spring flood volumes in Swedish rivers. *Hydrology and Earth System Sciences, 22(5),* pp.2953-2970.

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.

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

Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I., & Wernli, H. (2017). Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*. https://doi.org/10.1038/nclimate3338

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.

Hurrell J.W. (1995). Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. Science 269, 676–679.

Hurrell J.W. and Deser, C. (2009). North Atlantic climate variability: The role of the North Atlantic Oscillation. J. Mar. Syst. 78, 28–41.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., … Takahashi, K. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kutiel, H., Maheras, P. Türkes, M., Paz, S. (2002). North Sea– Caspian Pattern (NCP) – an upper level atmospheric teleconnection affecting the eastern Mediterranean - implications on the regional climate. Theor. Appl. Climatol. 72, 173–192.

Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of hydrology, 201(1-4)*, pp.272-288.

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.

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

Moore G.W.K. and Renfrew I.A. (2012). Cold European winters: interplay between the NAO and the East Atlantic mode. Atmos. Sci. Lett. 13, 1–8.

Olsson, J., Uvo, C.B., Foster, K. and Yang, W. (2016). Initial assessment of a multi-method approach to spring-flood forecasting in Sweden. *Hydrology and Earth System Sciences, 20(2)*, pp.659-667.

Park, T.W., Ho, C.H., Yang, S. (2011). Relationship between the Arctic Oscillation and cold surges over East Asia. J. Clim. 2011, 24, 68–83.

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*.

Riaz S. M. F., Iqbal M. J. and Hameed S. (2017). Impact of the North Atlantic Oscillation on winter climate of Germany, Tellus A: Dynamic Meteorology and Oceanography, 69:1, 1406263, DOI: 10.1080/16000870.2017.1406263.

Rust, H. R., A. Richling, P. Bissolli and U. Uwe. (2015). Linking teleconnection patterns to European temperature - A multiple linear regression model. Meteorologische Zeitschrift. 24. 411-423. 10.1127/metz/2015/0642.

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

Thornton, H. E., Scaife, A., Hoskins, B., Brayshaw, D., Smith, D., Dunstone, N., … & Bett, P. E. (2019). Skilful seasonal prediction of winter gas demand. *Environmental Research Letters*, *14*(2).

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

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.

Wilks, D.S. Statistical Methods in the Atmospheric Sciences. 3rd ed, Elsevier/Academic Press, 2011.

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

Yu, B., Lin, H. and Soulard, N. (2019). A Comparison of North American Surface Temperature and Temperature Extreme Anomalies in Association with Various Atmospheric Teleconnection Patterns. Atmosphere. 10. 172. 10.3390/atmos10040172.

Zadeh, L.A., (1965). Fuzzy sets. *Information and control, 8(3)*, pp.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