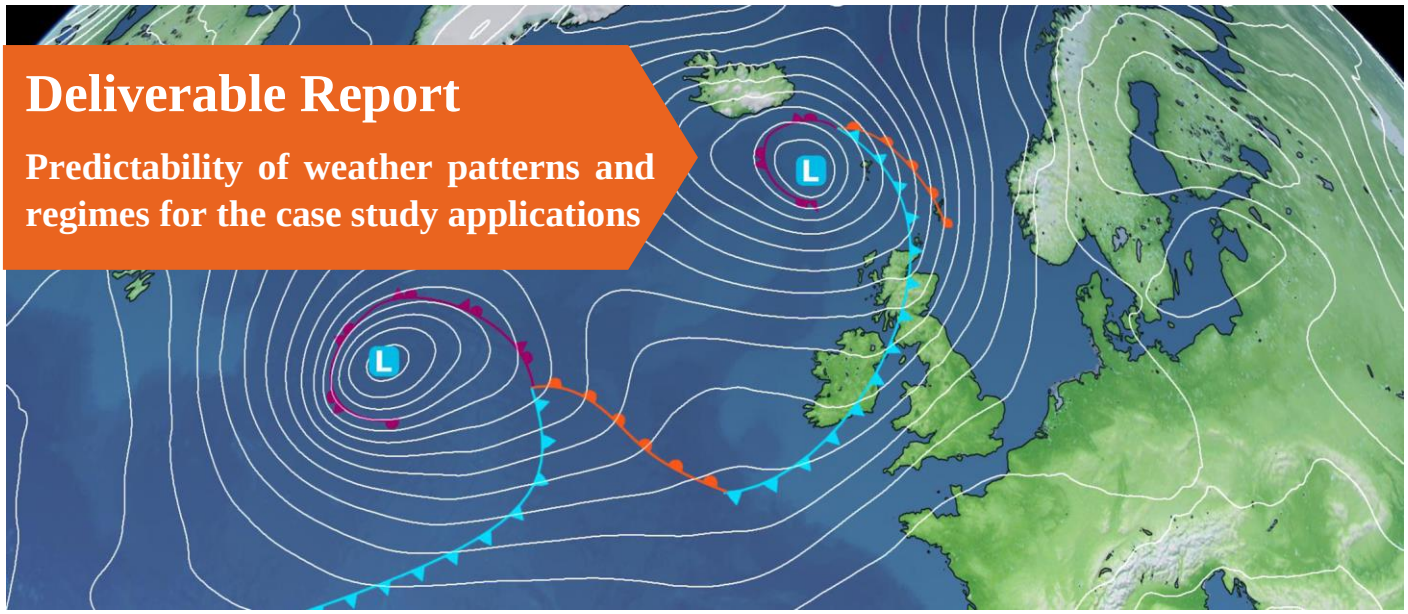


The Added Value of Seasonal Climate Forecasts for Integrated Risk Management Decisions

(SECLI-FIRM)

EU H2020 Project (ref. n. 776868)

D2.3: Report on the predictability of weather patterns and regimes of relevance for the case study applications

[Dissemination level: Public]

Version Table

Name/Party	Description	Date
Philip Bett (MO)	First Draft	09/03/2021
Philip Bett (MO)	Second Draft	22/03/2021
Philip Bett (MO)	Final Draft	29/03/2021

Internal Review Table

Name/Party	Description	Date
Emily Wallace (MO)	MO leader review	24/03/21
José Vidal (UL)	WP2 Leader review	26/03/21
Alberto Troccoli (UEA)	Project Leader's final review	30/3/2021

Contributors (Consortium Party/Person)

UEA	
ENEL	
ENEA	
Met Office	Philip Bett, Hazel Thornton, Tim Mitchell, Emily Wallace
UL – AWS Truepower	Victor Estella-Perez
KNMI	
WEMC	
Eurac	
Alperia	
Météo France	Christian Viel, Paola Marson, Lucas Grigis, Jean-Michel Soubeyroux

This document has been produced within the scope of the SECLI-FIRM project. The utilisation and release of this document is subject to the conditions of the grant agreement no. 776868 within the H2020 Framework Programme and to the conditions of the SECLI-FIRM Consortium Agreement.

Image on front cover produced by the Met Office

Table of Contents

1 Executive Summary

This report describes a series of investigations undertaken in the SECLI-FIRM project on the use of weather types or regimes in seasonal forecasts. Each section describes a different approach to the construction and application of weather types, based on different data sets and algorithms, and in the context of different Case Studies.

Section 2 describes work using a system of weather types currently used operationally at the Met Office for sub-seasonal forecasting. A methodology is developed to utilize these weather types on seasonal timescales, to forecast quantities relevant to UK winter electricity supply and demand, in support of SECLI-FIRM Case Study 8. It was found that only the weather types representing the North Atlantic Oscillation (NAO) pattern could be forecast skilfully, and furthermore, this only resulted in significant skill for forecasting wind speed and not temperature. There was therefore no benefit in terms of skill to using weather types over a simple NAO index in this context, and in particular they would not be useful for seasonal forecasts of electricity demand based on temperature.

Section 3 was also led by the Met Office, and tests a different weather regime-like method – canonical correlation analysis on a principal component analysis, CCA on PCA – to produce forecasts of significant wave height (SWH) in the North Sea, at long lead times, supporting Case Study 7. It was found that there is scope for forecasting mean SWH in May in the southern North Sea, from 1st March, using wind speeds as the predictor variable, which could be used to extend the summer operational windows for asset maintenance.

Section 4 investigates the use of a statistical bias correction and downscaling method, ADAMONT, in conjunction with weather regimes used operationally by Météo-France. The ADAMONT method on its own is shown to considerably improve the skill of seasonal forecasts of temperature, wind and precipitation. However, conditioning the ADAMONT downscaling on weather regimes does not yield any improvement in skill, due to the lack of skill in forecasting the weather regime frequencies themselves. Methods are discussed for improving the weather regime forecast skill.

Section 5, led by UL, explores how well weather types based on *k*-means clustering are able to represent the different wind regimes observed in Spain and Italy, in support of Case Study 4. Although the clusters are able to distinguish different distributions of daily PMSL, and the mean wind speed patterns in each cluster are distinct, the distributions of daily wind speeds in each cluster are very broad, and the clusters do not separate different regimes. Furthermore, as the distributions are very skewed, the mean wind speeds in each cluster are not representative of the most frequent wind speeds in those weather regimes. This means that this method on its own would not be helpful in forecasting within-season wind variability in Spain and Italy, regardless of the skill of seasonal forecast systems.

Overall, this report demonstrates an array of techniques for using North Atlantic/European atmospheric circulation patterns to make probabilistic forecasts of quantities of interest to users at long lead times. Under perfect forecast conditions, weather regimes could benefit forecasts in some cases, for example by allowing better quantification of within-season variability. However, this is not straightforward or universal. Furthermore, the limited skill in seasonal forecasts of weather regimes means that any possible benefits typically fail to be realised in practice. Nevertheless, the door remains open to alternative regime-based approaches in specific cases, tailored to specific user needs.

2 Application of Met Office weather types to UK winter energy requirements

2.1 Overview

This section investigates whether using patterns of atmospheric circulation in the North Atlantic/European sector can be used to improve seasonal forecasts of use to the UK electricity network. These “weather types” could provide several advantages over using meteorological variables directly from seasonal forecasting models: For example, they provide a natural way of providing coherent multivariate forecasts, and would capture within-season variability rather than just the mean state.

We developed a weather type-based forecast methodology, based on existing well-tested weather types and the Met Office GloSea5 seasonal forecasting system. The method was designed to integrate straightforwardly into the existing electricity demand forecasting process of National Grid ESO, the electricity grid system operator in Great Britain. Following the requirements of the corresponding SECLI-FIRM Case Study 8, we focus on forecasting for winter (December to February) from a November start date.

The method is based on forecasting the number of days of each weather type in the winter. This is used to make composite forecasts of winter means and winter daily standard deviations of relevant quantities such as temperature and wind speed. We then create forecast distributions of daily means of these quantities, based on transformations of the climatological distributions of daily winter values, but with the mean and standard deviation prescribed by the seasonal forecasts. These forecast distributions could then be used by National Grid ESO in place of the climatologies currently used.

Many variations on the method were tested, including varying the region, construction of the hindcast ensemble, and number of weather types. It was found that GloSea5 can only forecast NAO-like weather types, and with limited skill. This translates into only being able to provide skilful forecasts of seasonal mean wind speed. Further tests showed that these forecasts offer no benefit over using a seasonal-mean NAO index to forecast wind speed. We conclude that seasonal forecasts of weather type frequency cannot be used at present to improve electricity demand predictions in Britain.

Our tests imply that for seasonal forecasting timescales we should rule out the use of daily weather types such as these. However, these weather types do have proven skilful on sub-seasonal timescales. Furthermore, other possibilities remain open for seasonal forecasts based on large-scale circulation patterns, through defining weather regimes themselves on longer timescales, removing the synoptic-scale variability; or developing other metrics from large-scale climate features that capture the within-season variability we wish to forecast. Finally, it should be noted that there may be other benefits to using daily weather types (e.g.

in terms of communication or processing) that mean that they are an appropriate tool for forecasting, even though they do not offer a skill increase compared to traditional methods.

2.2 Introduction

Weather forecasters traditionally used maps of the synoptic-scale atmospheric circulation, characterised by fields such as the air pressure at mean sea level (PMSL), or the height of constant-pressure surfaces, to diagnose meteorological fields of interest, such as temperature, rainfall and wind speed. It is clear that certain circulation patterns – certain geographical configurations of low- and high-pressure synoptic systems – occur more frequently than others. Furthermore, particular weather conditions can be linked to particular circulation patterns.

While modern weather forecasting uses physically-based numerical models to predict all quantities of interest, it has long been recognised that focusing on metrics of atmospheric circulation can still give an advantage. For example, at long lead times – beyond the weather forecasting horizon of around 10 days – circulation types might be more skilfully predicted than the detailed rainfall or temperature fields themselves. The forecast of the likelihood of a particular circulation type could then be translated into forecasts of temperature, rainfall etc. based on the relationships of those quantities with that circulation pattern as seen in historical observations. This hybrid dynamical—statistical forecasting procedure allows the prediction of many different fields simultaneously in a physically consistent way. Furthermore, as the circulation patterns are typically characterised at relatively high temporal frequency (e.g. daily, rather than monthly or seasonal), they offer a possible way of forecasting high-frequency extremes in a probabilistic manner at long lead times. Hence, being able to predict these "weather types" is seen as a potential route to improved forecast skill on sub-seasonal-to-seasonal timescales.

The first step in using weather types for forecasting is to define the types themselves. There are many different approaches, but most are based on variants of either empirical orthogonal functions, or *k*-means clustering, or both (e.g. Huth et al., 2008, Neal et al., 2016, Matsueda & Palmer, 2018, and references therein). For the work in this section, we have chosen to use the existing weather types defined at the Met Office and currently used operationally in its medium-range forecasts (i.e. on lead times up to about 50 days), in particular in the *Decider*[1] forecast product. These are based on *k*-means clustering of daily PMSL fields, described in more detail below. However, these weather types have had relatively little testing in practical applications on the longer timescales required for seasonal forecasting.

Seasonal forecasting, particularly in extratropical regions such as Europe, is subject to much lower levels of skill, and higher levels of noise, than is typical for shorter range weather forecasts (e.g. Bett, Thornton, Troccoli, 2018). Dynamical seasonal forecast systems are

[1] https://www.metoffice.gov.uk/services/business-industry/energy/decider

typically based on coupled climate models, simulating the physics of the atmosphere, ocean, land and sea ice together, following an initialisation procedure. Forecasts are produced as ensembles, allowing probabilistic assessments to be made. This is essential given the comparatively low levels of skill involved. Although seasonal forecasts can be used in many different ways, we will be focusing on forecasts of the climate of 3-month periods (i.e. statistics of the weather within a season), at lead times of approximately 1 month.

In this section, we will be focusing on SECLI-FIRM's Case Study 8: winter electricity demand in Britain, for National Grid ESO (hereafter National Grid, or NG). While our forecast methodology has been developed with National Grid in mind specifically, many of the lessons learned from this report are applicable much more generally. The *Decider* weather types are also being used on sub-seasonal timescales for SECLI-FIRM Case Studies 7 and 9 (North Sea wind and wave conditions, and UK water demand, respectively), and these are described in detail in Deliverables D3.8 and D3.10, respectively. Other approaches to the application of weather types, considering the needs of different case studies, are covered in other sections of this report.

We outline in Section 2.3 the data sets we use, and briefly describe the methodology behind the weather types. We then go on to describe the forecast methodology we have developed, using weather types to forecast quantities of interest on seasonal timescales. In Section 2.4, we assess the performance of this new method, and discuss reasons behind its shortcomings. Section 2.5 describes alternative approaches based on what we have learned, and Section 2.6 describes our conclusions and proposes next steps in the context of the SECLI-FIRM project.

2.3 Data and Methodology

2.3.1 Data sets

We use hindcast data from the Met Office GloSea5 system (MacLachlan et al., 2015, Williams et al., 2015), produced as part of the operational forecast runs in 2017 and 2018[2]. We pool hindcast ensemble members initialised from three weeks around the start of November, from both runs, to give a 42-member ensemble. The hindcast of this size covers 23 winters, from 1993/94 to 2015/16.

To assess the skill of the forecast system, we use ERA-Interim reanalysis data (Dee et al., 2011) as a proxy for observations, covering 38 winters (1980/81 to 2017/18). National Grid use time series of site-based observations, selected and weighted to provide a population-weighted average for Great Britain. We use the same locations from ERA-Interim with the

[2] These correspond to "system" 12 and 13 in the Copernicus Climate Data Store. See https://confluence.ecmwf.int/display/CKB/Summary+of+available+data for details.

same weights to simulate the observational data used by NG for the initial assessments presented here.

National Grid use temperature, wind speed and irradiance in their long-term forecasts of electricity demand.[3] In addition to these variables, we also consider precipitation rate, to test it for other SECLI-FIRM case studies.

2.3.2 Weather Type methodology

The weather types we use are fully described in Neal et al. (2016), based on the method of Fereday et al. (2008). They use a variation of *k*-means clustering that uses simulated annealing, with the number of clusters set at 30 (other methodological choices are described in those papers).

Using the EMSLP data set (Ansell et al., 2006) over the period 1850—2003, the algorithm clusters together similar patterns of daily mean PMSL fields in a broad North Atlantic and European region. The clustering is based on data throughout the year rather than a particular season, and each daily PMSL field is assigned to a single cluster. The mean PMSL pattern of all the days assigned to each cluster (its centroid) defines that weather type. The 30 resulting weather types are shown in Figure 1 below.

These 30 types are intended for use in medium-range weather forecasts (up to 15 days' lead time). However, the distinction between the 30 types is too subtle to be skilfully forecast at longer lead times. Therefore, a reduced set of 8 types is derived from the 30, by repeatedly merging the most highly correlated pairs (Neal et al., 2016). Figure 2 shows the resulting 8 patterns. Although this reduced set of weather types has been successfully used on sub-seasonal timescales (up to about 50 days' lead time), there has been limited testing on seasonal timescales.

Given a set of weather types such as these, a forecast daily-mean PMSL anomaly field can be assigned to a weather type using a particular metric. The standard approach is to use a "distance" metric: the area-weighted sum of the squares of the differences between the forecast field and the weather type at each grid cell. This is used for the operational WT assignment in the observational data. An alternative method would be to use the spatial correlation of the PMSL anomaly fields. This focuses on the similarity of the spatial pattern, attaching less significance to the absolute magnitudes of the anomalies in the patterns. This could be useful if the data being assigned to the WTs has significantly different variability to the WT centroids themselves, which can be the case in the context of seasonal forecasting.

[3] See, e.g. https://www.emrdeliverybody.com/Lists/Latest News/Attachments/189/SC4L12 ACS Methodology.pdf

This assignment can be done separately for each member of an ensemble forecast: The fraction of ensemble members in a given weather type on a given day can be interpreted as the forecast probability of that day having that weather type. Knowing the historical distributions of other quantities such as temperatures under each weather type then allows us to use weather types to make probabilistic forecasts of those quantities.

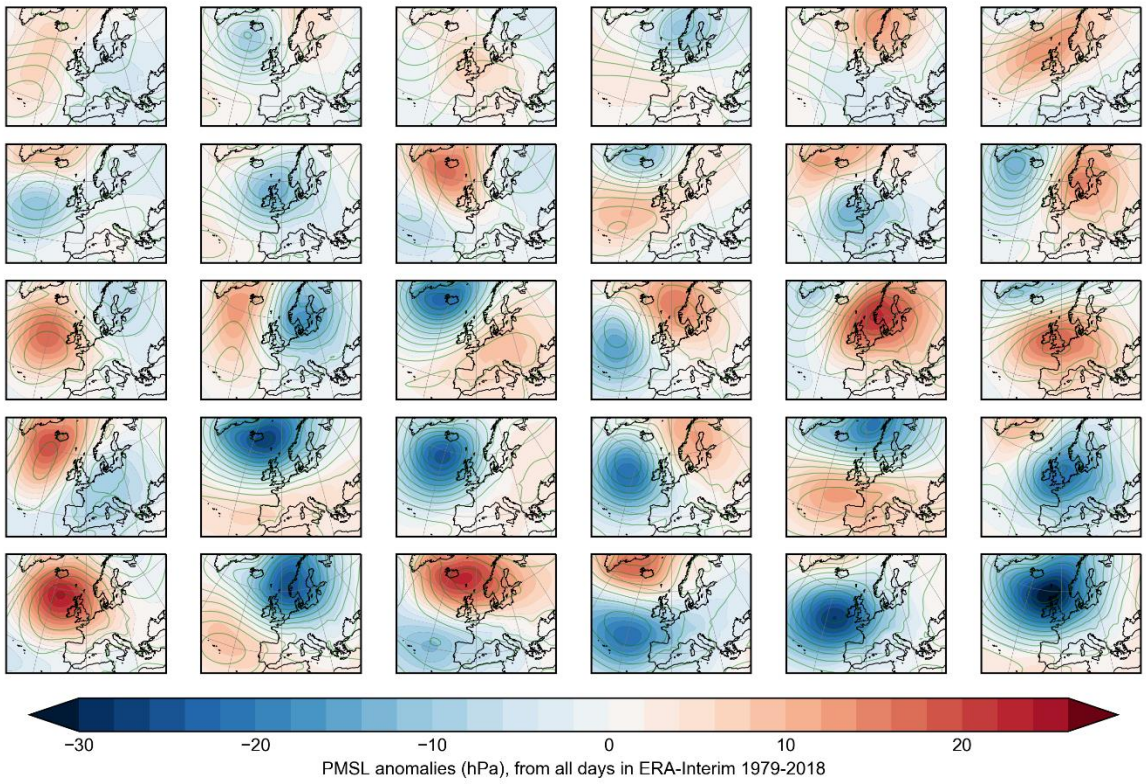


Figure 1. The 30 weather types of the Neal et al. (2016) method (similar to their Figure 1). Colours show PMSL anomalies from ERA-Interim, and the green contours show the actual PMSL levels. The WTs are ordered according to their frequency in the Neal et al. analysis.

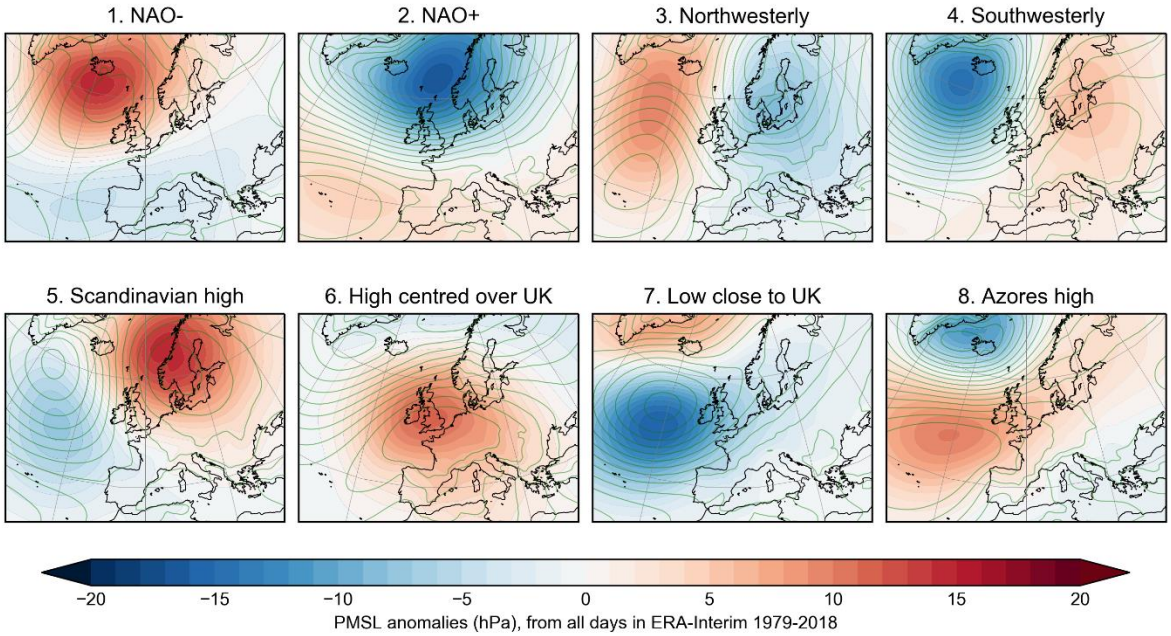


Figure 2: The 8 reduced weather types. Colours indicate PMSL anomalies, and the green contours show the actual PMSL patterns. The WT names follow those of Neal et al. (2016), and they are ordered by their year-round occurrence frequency in that paper.

2.3.3 *Proposed forecasting methodology*

There are many ways that the daily weather types described above could be used in a seasonal forecasting context. We have developed the following method with the specific needs of National Grid in mind: We aim to produce forecast information in a format that matches, as closely as possible, current practices at National Grid, without reducing the available forecast skill. This tailoring to meet the user needs is critically important: it means we can make a fair comparison when assessing whether the new method represents an improvement in skill, and it makes it easier for National Grid to implement and use the forecasts if they do prove to be beneficial.

National Grid base their demand forecasts on GB-averaged temperature, wind speed and irradiance. These data are used as climatological distributions, which can be Monte Carlo sampled to provide probabilistic estimates of possible outcomes. Our aim here is to use weather types (WTs) to forecast how those observed climatologies should be modified to provide a forecast for a coming winter (December–January–February, DJF), for example by applying mean shifts or scaling factors to them. Thus, rather than using the climatological distribution as an indicator of future conditions, they would use the distribution derived from a seasonal climate prediction model; hopefully, this would be better than just using climatology.

Our use of weather types is intended to allow us to:

- Make forecasts coherently for several meteorological variables at once.
- Forecast changes in the distribution of daily data within a season, rather than just the seasonal mean. (For example, capturing seasons with periods of both blocked and stormy conditions, rather than relying on a season-wide tendency towards one or the other.)

For this to work however, we require two things:

- The seasonal forecasting system needs to be able to skilfully forecast the distribution of weather types within a given season.
- The observed meteorological variables of interest need to exhibit clear, distinct responses to the different weather types. (Note that this is unrelated to forecast model skill.)

As described above, each winter day in the historical period is assigned to one of the 30 weather types (WTs) according to their PMSL anomaly pattern. The established mapping between the 30 base WTs and the 8 reduced WTs is then used to assign each day to one of the 8 types, which we use going forward.

For each meteorological variable, we consider the distributions of their daily-mean GB-mean values under each weather type separately. We can then characterise those distributions under each WT in terms of their means and standard deviations. We characterise the overall climatological distribution, covering all WTs, in the same way. Figure 3 shows the climatological distributions of temperature and wind speed in each WT as an example.

From Figure 3, we can see certain expected behaviours under different weather types. Weather types 1 and 2, which look like “NAO-” and “NAO+” patterns respectively, show the expected shifts to cool+calm and warm+windy conditions respectively. Other types show more or less variation from climatology.

When making a forecast of a given winter, each day in that winter is assigned to a weather type according to its forecast PMSL anomaly pattern. Due to the well-known signal-to-noise problem of seasonal forecasts, in North Atlantic/European winters in particular (e.g. Scaife & Smith, 2018), we expect that the forecast anomalies in the daily ensemble mean PMSL fields will be smaller than observed anomalies, and that the relevant forecast signal is in the pattern rather than the magnitudes. We therefore use the pattern correlation to make the WT assignments, rather than the distance metric (see earlier). In practice however, we found that this choice makes negligible difference to the forecast skill.

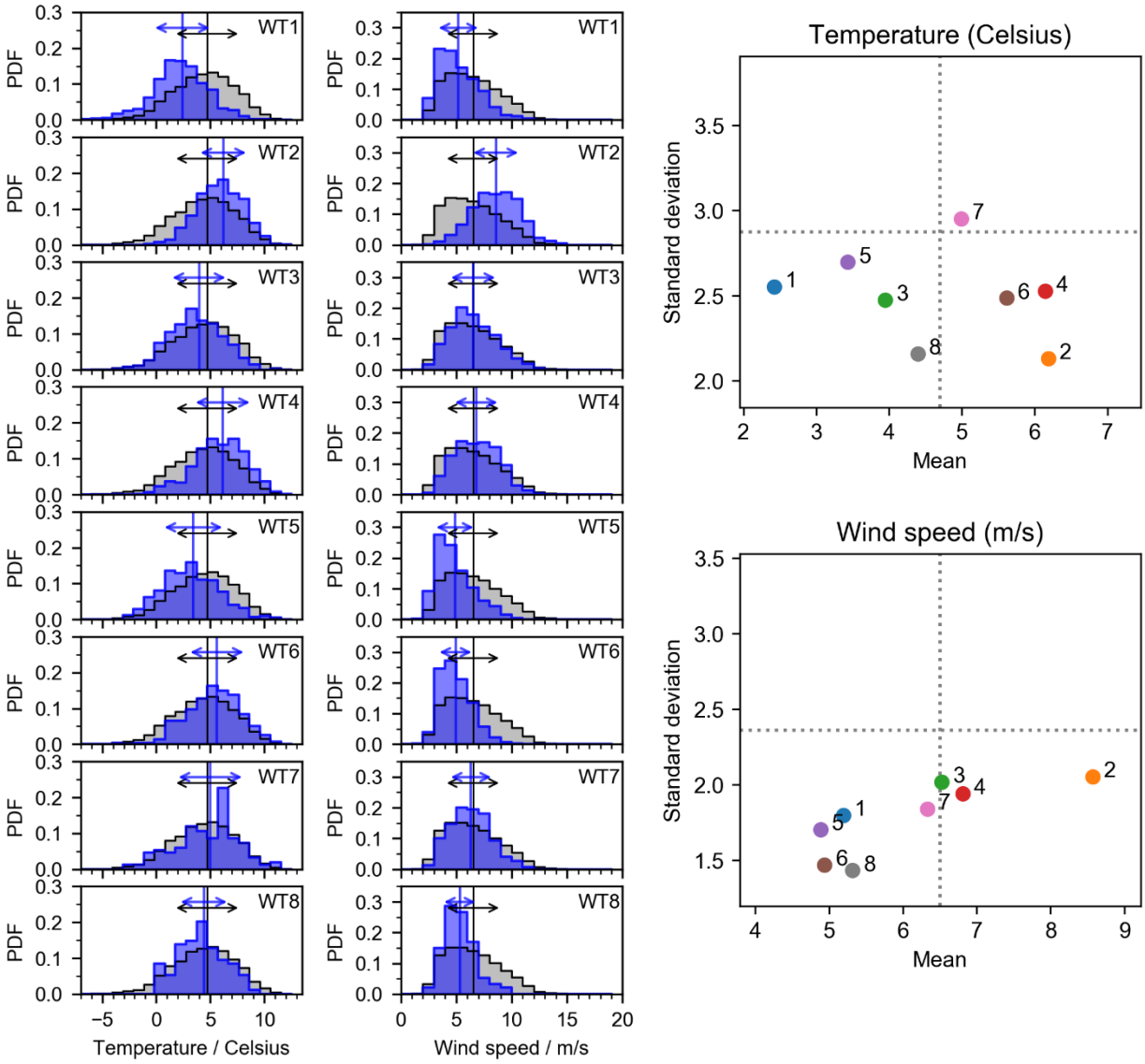


Figure 3. Variation of daily-mean UK-mean temperatures and wind speeds under different weather types. The histograms show the distributions of temperatures (left) and wind speeds (right) under each weather type (blue), with the climatological distribution for all weather types shown in grey (the same in each panel). Vertical lines show the means of both distributions in each panel, and the horizontal arrows show the ± 1 standard deviation about the means. The scatter plots on the right show these means and standard deviations under each weather type as coloured points, with the all-types climatological values indicated by dashed lines.

Counting the number of days in each weather type in the season gives us our forecast weather type distribution. For example, we might forecast the winter to be composed of 40 days in WT1,

35 days in WT2, 5 days in WT3, etc. Note that the time-ordering of the WTs within the season is not considered significant, as this will not be forecastable on these time scales: we only consider the WT frequencies.

We then create forecasts of the seasonal mean and variance of the distribution of daily-mean temperature, wind speed, irradiance and precipitation. This is done by compositing the seasonal means and variances of the WT-based climatologies for each variable, weighted according to how many days of each weather type were forecast. For example, considering just temperatures, if we forecast that $f_{\mathrm{WT1}} = {}^{n_{\mathrm{WT1}}}/_{N_{\mathrm{days}}} = 45\%$ of the winter will be WT1, then the mean temperature under WT1, $\mu_{\mathrm{WT1}}(T)$, will contribute to the composite seasonal mean $\mu_{\mathrm{fc}}(T)$, and variance $\sigma_{\mathrm{fc}}(T)$, with a weight of 45%.

In particular, if we write the observed climatological mean and variance of daily temperature under WT i as $\mu_{\mathrm{WT}i}$ and$\sigma^2_{\mathrm{WT}i}$, then the forecast seasonal mean temperature is just the mean over all WTs weighted by the forecast WT frequencies $f_{\mathrm{WT}i}$:

$$\mu_{fc} = \frac{1}{N_{\mathrm{WTs}}} \sum_{i=1}^{N_{\mathrm{WTs}}} f_{\mathrm{WT}i} \cdot \mu_{\mathrm{WT}i}$$

The forecast variance of days in the season is the mean of the variances from the WT assigned to each day, plus the variance of the WT-based means for each day; this can also be written in terms of weighted means:

$$\sigma^2_{\mathrm{fc}} = \frac{1}{N_{\mathrm{WTs}}} \sum_{i=1}^{N_{\mathrm{WTs}}} f_{\mathrm{WT}i} \cdot \sigma^2_{\mathrm{WT}i} \quad + \frac{1}{N_{\mathrm{WTs}}} \sum_{i=1}^{N_{\mathrm{WTs}}} f_{\mathrm{WT}i} \cdot \mu^2_{\mathrm{WT}i} - \left(\frac{1}{N_{\mathrm{WTs}}} \sum_{i=1}^{N_{\mathrm{WTs}}} f_{\mathrm{WT}i} \cdot \mu_{\mathrm{WT}i} \right)^2$$

These composite forecasts of seasonal mean and variance are the primary outputs of this method, which is summarised in Figure 4. Note that the only contribution from the seasonal climate forecast model is from the PMSL fields in the WT assignment step.

Given the forecast seasonal mean and variance for each variable, we can then apply this to each variable's overall observed climatological distribution: we can shift and scale the climatological distributions so that they match the forecast in terms of mean and variance. In an "operational" context, it is imagined that this would be done by National Grid; the resulting forecast distribution could then be used in the rest of their electricity demand forecast model, without any further modification to their method, and only the mean and variance information would need to be provided, rather than large data sets. For our purposes in this Task, we can examine the forecast distributions to see how skilful this method is at forecasting standard statistics, such as the seasonal mean of each variable.

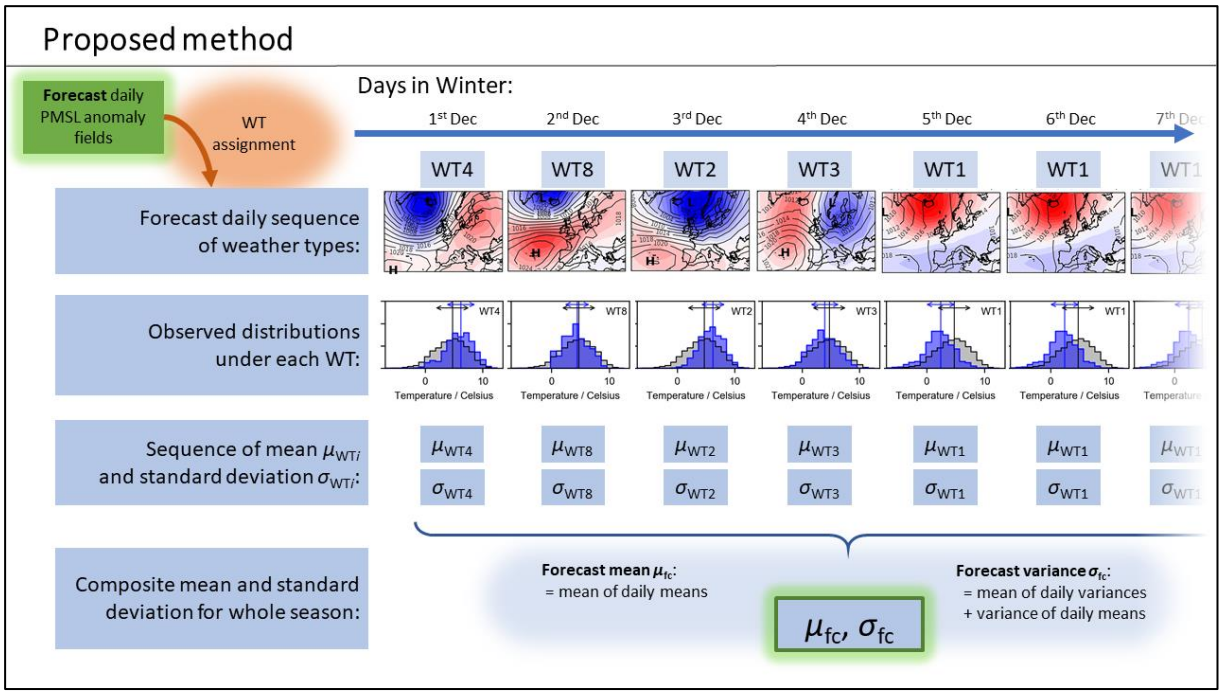


Figure 4. Illustration of the forecast method. The forecast daily PMSL anomaly fields are the only components from the seasonal climate prediction model; all other components use observed relationships.

2.4 Assessment of WT-based forecast skill

2.4.1 Assessment of Method

We have tested the above method using ERA-Interim data alone, i.e. using the "observed" weather type occurrence in each season to "forecast" the distributions of the variables each year (based on the climatologies of the remaining years). We examine the skill of these "perfect forecasts", in terms of the correlation of either the time series of forecast seasonal means, or the time series of forecast seasonal standard deviations of daily data, with their actual observed values.

Statistical significance of these correlations can be assessed using a Fisher z test. For the 38 years of data here, a correlation will be significantly (i.e., measurably) different to zero at the 5% level if it exceeds ±0.32. The time series, and their correlations, are shown in Figure 5 below.

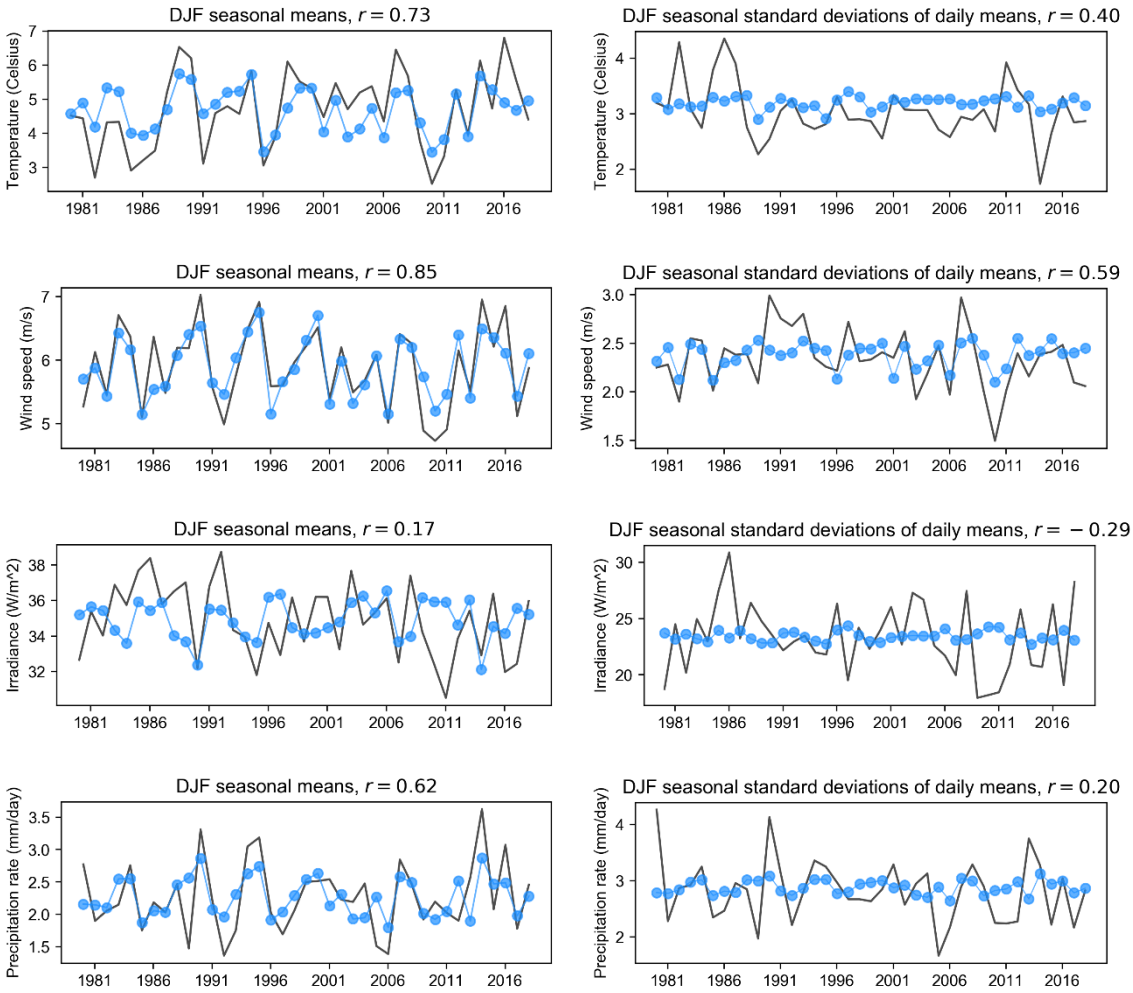


Figure 5. Time series of seasonal characteristics of the composite distributions of daily mean values within each winter. The actual value in ERA-Interim each winter is shown in black. The "forecast" for each year (blue) is calculated from the WT distribution for that winter, and the climatological distribution of each variable under each WT excluding the forecast year.

This demonstrates that the method can work extremely well for forecasting seasonal mean wind speed: the correlation between the actual and "forecast" seasonal mean wind speeds is $r = 0.85$. It is equally important to note that even in this case it is not perfect: even the best seasonal forecast system could not improve on this skill.

Our method also works well for temperature and precipitation rate (both correlations > 0.6). It will not work for forecasting irradiance however ($r = 0.17$). Our investigations show that the distribution of daily mean irradiance each winter does not vary strongly, in terms of its mean and variance, from year to year; and what variability it does have is not well captured by the

variation in weather type occurrence. However, it might be possible to refine this method for irradiance: The minimum of the annual cycle for irradiance occurs around the solstice in December, and by February there is a clear trend towards brighter days. Since this is a strong signal, and identical every year, it could be swamping the interannual variability. It might be that removing the annual cycle, or simply focusing on a solstice-centric season (e.g. November—December—January) could improve the skill.

The right-hand column in Figure 5 shows that our method is unlikely to capture the within-season variability very well, for most variables apart from wind speed (r = 0.59; other variables show much lower correlations).

These observation-based results represent the maximum possible skill, i.e. for a perfect seasonal forecasting system, and provide context when interpreting the GloSea5-based results.

2.4.2 Skill of GloSea5 in forecasting the 8 weather types

Having established the observed link between weather types and meteorological variables, we now test how well the frequency of weather types within a winter can be predicted by GloSea5. This skill is shown in Figure 6, as the fraction of days each winter in each weather type, for each ensemble member.

A convenient summary skill score is the correlation of the ensemble mean with the observed frequencies. However, because these frequencies are constrained to be between 0 and 1, just taking a simple ensemble mean could lead to biases for frequencies that are at the high or low end, as they are more constrained. Instead, we use a logit transformation, a common way of transforming functions from the range $[0,1]$ to $[-\infty,+\infty]$, applied to the frequencies f in the observations and in the ensemble members: $f' = \log\left(\frac{f}{1-f}\right)$.

The ensemble mean is then calculated as the mean of the transformed frequencies f', and the correlations are taken between the transformed observed frequencies and the ensemble mean of the transformed frequencies. The ensemble mean is then back-transformed into a proper frequency for plotting. (In practice, we find that this makes only small differences to the correlations compared to using the “raw” frequency values.)

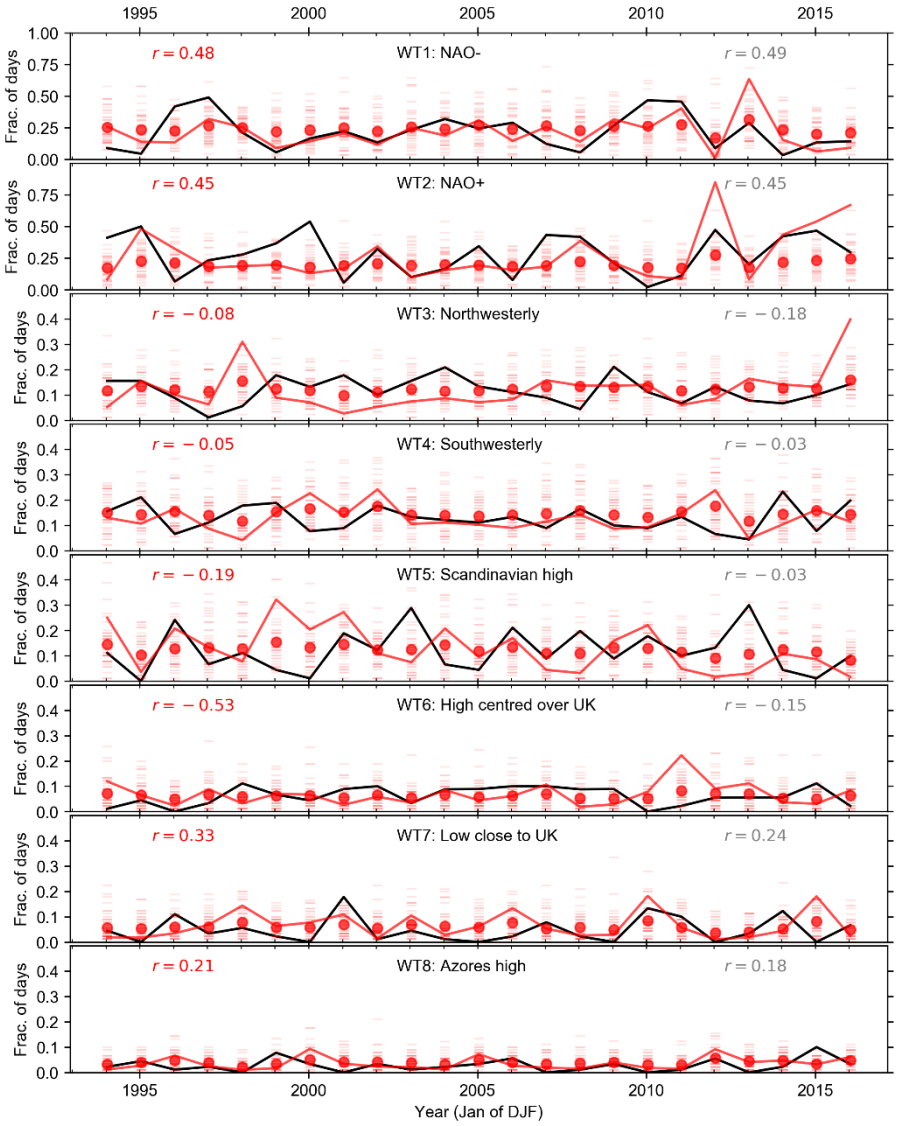


Figure 6. Frequency of the 8 reduced weather types in winters, from ERA-Interim (black), and forecast from GloSea5 (red). Note that WT1 and WT2 are shown on a different axis scale to the other WTs. The frequencies in individual ensemble members are marked as horizontal lines each winter. The ensemble mean is shown in red; note this was calculated under a logit transformation (see text). The correlation skill *r* of the ensemble mean is given in red in the top-left of each panel; again, this was calculated using logit-transformed observed and ensemble mean frequencies. For comparison, the correlation of the raw frequency data is shown in grey in the top-right of each panel. Using a Fisher z-test on the 23 winters here, we can say that correlations are significantly different to zero at the 5% level if they exceed ±0.41.

These skill scores should immediately temper our expectations of using weather types with seasonal forecasts. The only significant skill is for the NAO-like weather types, WT1 and WT2. WT6 (*High centred over UK*) in fact has significantly negative skill.

While these results are disappointing, they only reflect one component of our method. Many of the weather types have very low frequencies per winter, and trying to forecast small numbers of occurrences will always be difficult. In practice though, we will be trying to forecast the *combinations* of weather types in a given winter, not the absolute number, and the skill scores shown in Figure 6 do not reflect this.

In assembling the hindcast data to assess the skill, there is a choice of which initialisation dates to use. The GloSea5 hindcasts are initialised on 4 dates each month (1st, 9th, 17th, 25th), which contrasts with the forecast production of 2 members every day. In an operational context, the forecast is built from three weeks of forecasts (42 members), and a corresponding hindcast dataset is assembled from the three nearest hindcast start dates to each of those 21 forecast start dates (weighted appropriately, see MacLachlan et al., 2015). To assess skill, it is usual to use choose three hindcast start dates around a hypothetical forecast date. Thus, the natural choice for forecasts of DJF with a 1-month lead time would be to choose hindcast start dates of 25th Oct, 1st Nov and 9th Nov. However, we find that we achieve slightly higher apparent skill when using start dates about a week later, i.e. the 1st, 9th and 17th Nov. In practice, the results are not significantly different (the correlation skill for WTs 1 and 2 for the earlier dates are 0.45 and 0.38 respectively), and in any case neither choice corresponds exactly to the lags we would have in real-time forecasts. We used the slightly later start dates in Figure 6, and continue to use them throughout this study.

2.4.3 Skill of the full forecasting method using GloSea5

The results from applying the forecast methodology to the GloSea5 seasonal hindcast data are shown in Figure 7.

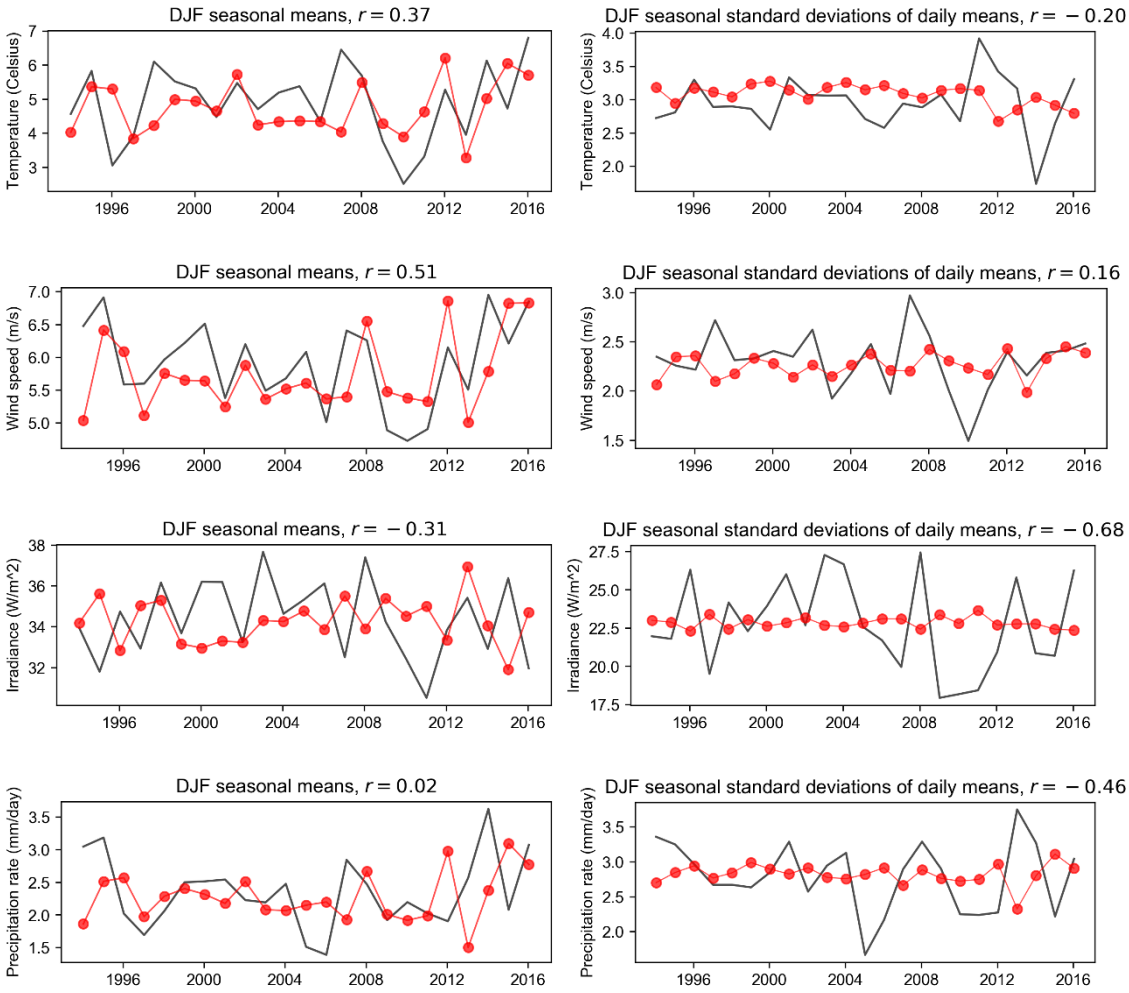


Figure 7. Time series of seasonal characteristics of daily mean values each winter, as in Figure 5. The observation-based value from ERA-Interim each winter is again shown in black. In this case however, we have used the ensemble-mean PMSL fields from the GloSea5 hindcasts to provide the weather type distribution in each winter, mirroring the procedure in a real-time seasonal forecast. The resulting "forecasts" for each variable are shown in red. The threshold for statistical significance for a correlation of 23 years at the 5% level is $r = \pm 0.412$.

The results show good skill for the seasonal mean wind speed, but more disappointing results for the other variables. Seasonal mean temperature in particular, the primary meteorological driver of electricity demand in Britain, has correlation with the observations of $r = 0.37$, which is formally statistically indistinguishable from zero at the 5% level.

The seasonal mean irradiance is not skilful, as expected from the ERA-Interim results. The seasonal mean precipitation is also not skilful in this case, although we have found that this

can be improved by varying the region used. As anticipated from Figure 5, the within-season variability also cannot be predicted using our method.

The seasonal mean precipitation is also poor. However, we found this to be strongly affected by the use of the site-based GB region chosen to match National Grid procedures. For example, we found an increase in precipitation skill if we used a UK region defined by international borders, or a “British Isles” region defined by a simple latitude—longitude rectangle including sea points (following Clark et al. 2017 and Thornton et al. 2019). As precipitation is not needed for National Grid forecasts, a different region should be chosen for other applications. Changing the region had little impact on the other variables.

Note that the results shown in Figure 7 do not give a probabilistic assessment: We are using the daily ensemble mean PMSL fields for the WT assignment. This is because taking the ensemble mean will maximise any available forecast signal, whereas individual ensemble members would be extremely noisy. This also provides a simpler comparison with the ERA-Interim results in section 2.4.1. However, any real forecast system would have to be probabilistic in nature. The deterministic correlation skill we present here can therefore be seen as necessary but not sufficient criteria for having a skilful forecast system: If there is insufficient skill in the ensemble mean, then probabilistic methods will also fail.

Indeed, much of our initial testing was based on the ensemble members themselves, examining the distribution of weather type combinations for each winter across the ensemble. However, we found that using daily ensemble means resulted in a marked improvement in skill. This was somewhat surprising, as for seasonal forecasts we would not expect the synoptic weather to be synchronised on day-to-day timescales across the ensemble: events like low-pressure systems passing over the UK, or blocking highs breaking down, could easily be out by a day, if they exist consistently at all across the ensemble. We assume that enough similar-looking events being synchronised across enough of the ensemble to improve the skill, while the other cases where events are less correlated in time don’t act to reduce the skill further.

As well as testing different regions and use of the ensemble, we also tested the seasonal compositing process, by penalising days where the PMSL pattern does not correspond closely to any of the weather types. When calculating the composite seasonal mean and variance of meteorological quantities, we applied weights to the contributions from each day according to the spatial correlation with their assigned WT pattern: If a day showed an exact match with the pattern, then it contributed with a weight of 1, but anything less was downweighted. In practice, the observed daily WT assignments, calculated at the Met Office operationally, do not use weights. Therefore the variation in the resulting temperature etc. distributions due to possible poor matches to the WTs is already included in the procedure. The PMSL patterns in GloSea5 could span a broader range, including patterns not seen in observations, so it is not unreasonable to imagine this weighting could improve things. However, in practice we found

that using weights made negligible difference to the skill, and we did not use weights after this test.

Finally, we also tested whether using fewer weather types could improve the forecast skill. The choice of reducing the 30 base types down to 8 (see section 2.3.2) is not fixed, and it represents a balance between capturing the diversity of possible weather situations, and being able to skilfully distinguish between them in forecasts at long lead times. For example, if a few WTs differ only in the detailed locations of a low or high pressure system, then there might be a benefit in combining those WTs: we wouldn't believe that a seasonal forecast would genuinely be able to distinguish between them, and higher frequencies of fewer types is likely to be more skilful than very low frequencies across several types. On the other hand, if we merge WTs that have quite different temperature distributions for example, then we will no longer be able to distinguish between those temperature regimes, and our skill will fall. So, while using 30 types on seasonal forecast timescales would clearly be inappropriate, it might also be that 8 types is still too detailed; but we should also be wary of reducing them too far. Many studies typically use 4 weather regimes on seasonal forecast timescales (e.g. Cassou, 2008; Dawson, Palmer, and Corti, 2012; Matsueda & Palmer, 2018; van der Wiel et al. 2019): These usually describe NAO+ and NAO-, Scandinavian Blocking and Atlantic Ridge patterns, and are closely related to the first three EOFs of the North Atlantic circulation (the NAO, East Atlantic and Scandinavian patterns, e.g. Zubiate et al., 2017).

We have tested reducing the number of weather types we use by continuing the merging process used to derive the standard 8 types, providing sets of 7, 6, 5, etc, types down to a final set of 2. The set of 4 types and the set of 2 types this yielded are shown in Figure 8. The four types roughly correspond to the first three EOF patterns: the first two types correspond to NAO+ and NAO-, the third type presents an east–west gradient like Scandinavian Blocking, and the fourth pattern presents a monopole in the Atlantic (although ours is a low anomaly, rather than the Atlantic Ridge pattern with a high pressure anomaly). The set of two types are simply NAO-like patterns. Note that the WT assignment was still performed using the base 30 types, but the seasonal compositing, and temperature etc. responses, are provided by the reduced types.

Prior to testing this with the GloSea5 ensemble, we tested it using observations (i.e. "perfect" WT assignment, as in section 2.4.1): The wind speed skill was reduced when using fewer than below 8 WTs, and was worst for 2 WTs, although temperature skill was largely unaffected by the number of WTs. In terms of actual seasonal forecast skill using GloSea5, we found that using fewer weather types hindered the forecasts more than it helped: There was a marked reduction in skill in seasonal mean temperature, due to reducing the freedom of temperature to vary in different situations. Using fewer WTs didn't increase the skill in WT frequency enough to compensate for this.

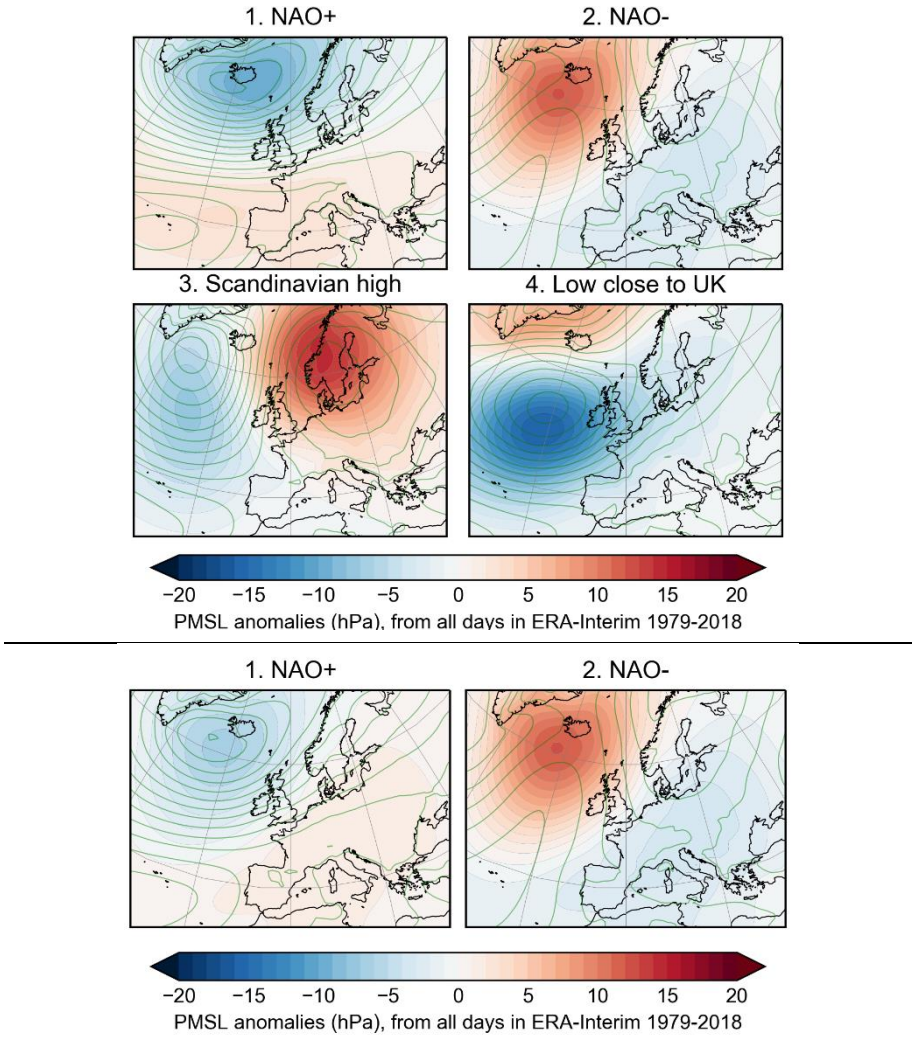


Figure 8. Top panels: The 4 weather types formed from reducing the original 30. Colours indicate PMSL anomalies, and the actual PMSL patterns are shown as green contours. Bottom panels: The 2 weather types resulting from reducing the number of types further, corresponding to NAO+ and NAO- like patterns. Note that the “NAO-” types in the set of 4 and the set of 2 are identical; going from four to two types just merges types 1, 3 and 4 into the “NAO+” pattern.

2.5 Alternative Approaches

2.5.1 Focus on the NAO

Our results have demonstrated that GloSea5 only has significant skill in the NAO-like weather types. An alternative approach would therefore be to use the NAO index itself to forecast temperature and wind speed for NG (following Clark et al., 2017), through a linear regression. Just as with the WT-based method, this involves two elements: The skill in the seasonal forecasting system in predicting the NAO index itself; and the strength of the relationship between the NAO and the quantity of interest (e.g. temperature, wind speed).

The GloSea5 hindcast we are using here has a 95% confidence interval on its NAO skill of 0.18 – 0.78, with a central estimate of 0.55. The relationships in ERA-Interim data between the NAO and the four meteorological quantities we have been considering are shown in Figure 9.

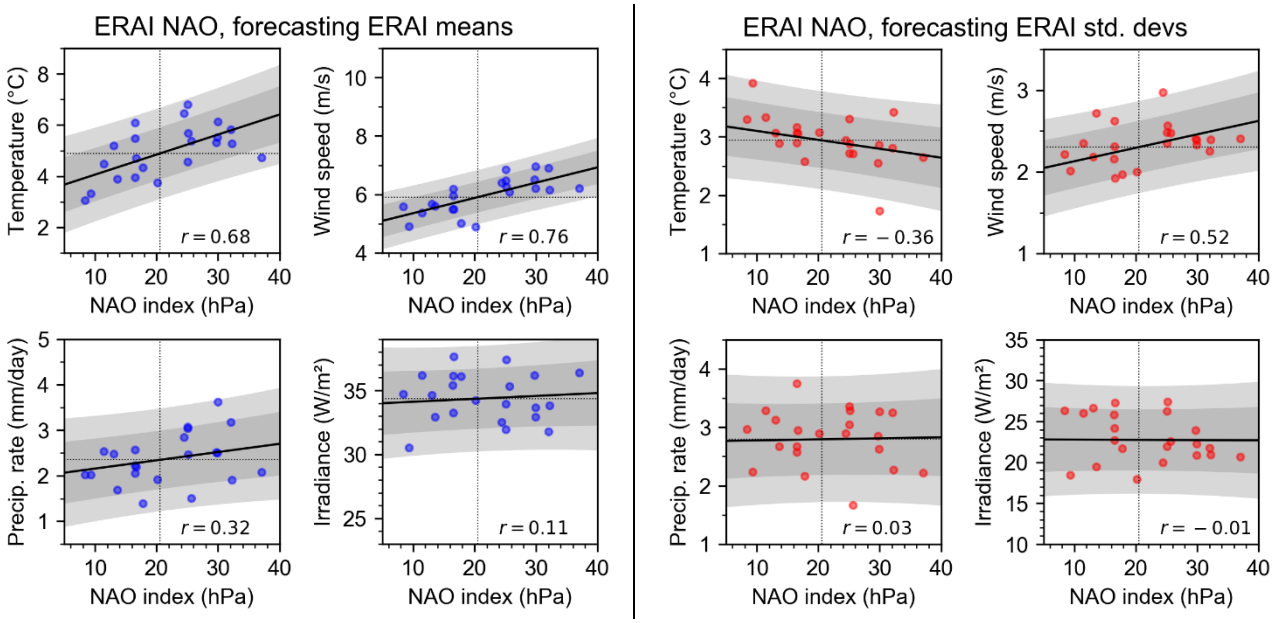


Figure 9. Left four panels (blue): Linear regression relationships between the observed winter mean NAO and observed winter mean GB-mean values of the four variables (as labelled). Right (red): Regression between the NAO and the GB-mean seasonal standard deviation of daily mean values of each variable. All data is from ERA-Interim, although we use the same 23 years as the GloSea5 hindcast for consistency. Correlations *r* between the NAO and each quantity are shown above each panel, and are statistically significant at the 5% level if they exceed ±0.41. Grey shading indicates the 75% and 95% prediction intervals for the linear regression.

Note that there is a strong relationship between the NAO and seasonal mean wind speed, and a significant relationship between the NAO and seasonal mean temperature, as well as with the seasonal standard deviations of temperature and wind speed. A more positive NAO leads

to stronger and more variable wind speeds, as more low-pressure systems pass over the UK. Low pressure systems also cause higher UK temperatures, due to the advection of warm, moist air from the Atlantic. This also tends to result in increased cloud cover, which helps to keep temperatures more consistent, reducing their variability. In contrast, a neutral or more negative NAO allows more variation between occasional storms and more blocked conditions: while winds on average are less variable (fewer high-wind extremes), a wider range of temperatures is experienced more frequently.

Replacing the ERA-Interim NAO with the GloSea5 forecast NAO (which has a prediction skill of r = 0.55) reduces the correlations by a factor of roughly 0.5. These are shown in Figure 10: The observed relationship between NAO and mean temperature is not strong enough to translate into significant skill when using the predicted NAO; however, there is skill in the mean wind speed. The seasonal standard deviations also aren't sufficiently well correlated with the NAO to be forecast using this method

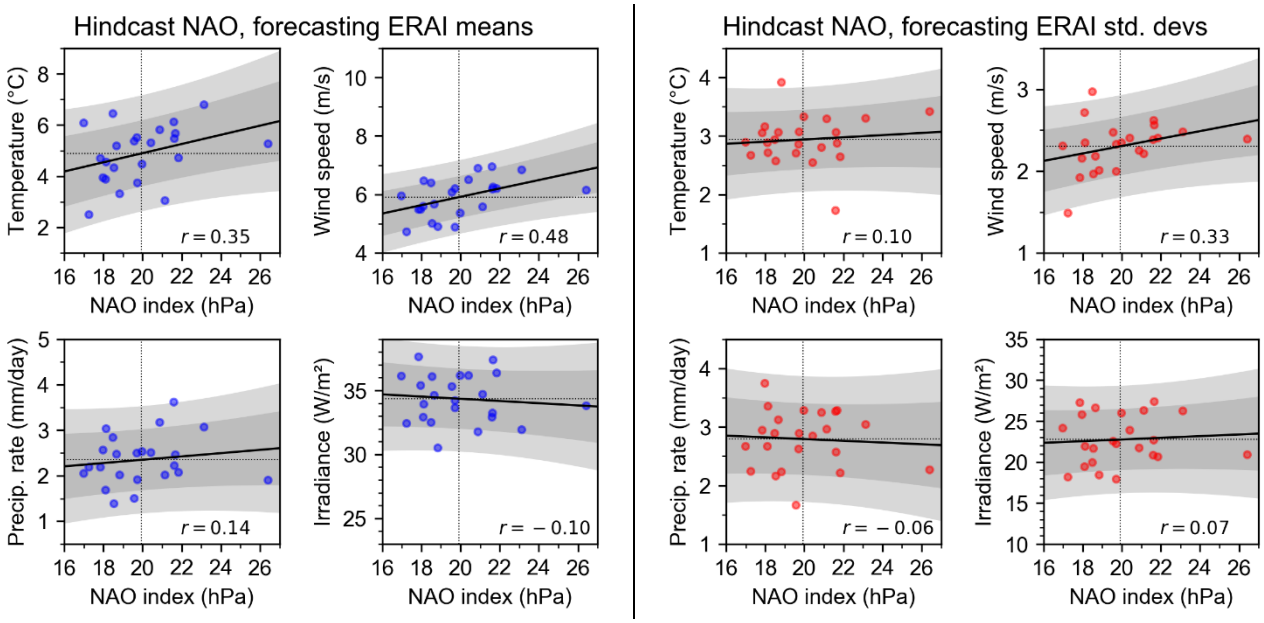


Figure 10. As Figure 9, but using the NAO index from the GloSea5 hindcast instead of the ERA-Interim NAO value.

Although the skill in forecasting mean temperature based on the NAO is low, there is promising skill in forecasting the seasonal mean wind speed. Indeed, the NAO-based wind speed skill (r = 0.48, 95% confidence interval of 0.08 – 0.74) is not significantly different from the skill from using the weather types (r = 0.51, 95% confidence interval of 0.12 – 0.76). The NAO-based method is much simpler, and requires considerably fewer choices to be made when developing the process. Therefore, there is no reason to prefer the WT-based method over the NAO-based method for winter mean wind speeds over this site-based GB region.

2.6 Conclusions and Next Steps

This section has shown that the daily weather types currently used successfully at the Met Office on sub-seasonal timescales have limited use on seasonal timescales. In particular, they do not provide significant skill in forecasting winter temperatures, neither in terms of the seasonal mean or the within-season variability. This means they are unlikely to provide a useful solution for seasonal forecasting winter electricity demand in Britain.

The weather types did show significant skill in forecasting winter mean wind speed. However, this was at the same level of skill as simply using the NAO forecast to predict wind speed.

The key benefits of forecasting using weather types are that they can provide coherent multi-variate forecasts, and forecast variability as well as the mean state. The fact that we cannot realise these benefits on a seasonal forecasting timescale imply that it would be better to use the simpler NAO-based method when forecasting winter wind speed.

It should be noted as well that there may be other benefits to using daily weather types (e.g. in terms of communication or processing) that mean that they are an appropriate tool for forecasting, even though they do not offer a skill increase compared to traditional methods.

Following these results, an alternative method for producing trial seasonal forecasts for National Grid ESO has been developed. As in this report, it is based on modifying the climatological distributions of daily temperature and wind speeds, but uses forecasts of the NAO and UK mean temperature to provide the seasonal mean distributions, rather than weather types. A full description of this method and our evaluation process, in conjunction with National Grid, is described in SECLI-FIRM Deliverables D3.12 and D3.9.

There is still considerable demand for seasonal forecasts that go beyond predictions of the mean stated: developing methods to extend the skill of seasonal forecasts to predict within-season variability, including the risk of extremes. Although the weather types considered here, based on clusters of daily PMSL patterns, have not been successful at these long timescales, our work points towards alternative approaches. One possibility is to use patterns defined on longer timescales, rather than focusing on synoptic-scale variability. For example, looking at pentad-mean PMSL patterns, or even monthly means across the ensemble. Another possibility is to use larger-scale indices of the within-season variability. We have seen how the NAO index is correlated to wind variability, and anticorrelated to temperature variability, although this was not strong enough to be forecast with current levels of skill. It might instead be possible to design other measures of North Atlantic circulation that focus specifically on predicting variability. For example, a measure of the “waviness” of the jet stream, smoothed to remove synoptic variability but retaining a measure of sub-seasonal variability, might (if skilful) fit this important gap between the blunt instrument of the NAO and the over-detailed daily weather types.

3 Seasonal forecasting waves in the North Sea

3.1 Introduction

This section summarises the work done in the SECLI-FIRM project to investigate the scope for a seasonal forecast of significant wave height in the North Sea, in the shoulder months around summer.

We have been working with Shell, who have assets in the North Sea that they have to maintain. We understand that a common approach to asset maintenance in the North Sea is to plan operations on the assumption that there will be sufficient calm periods during a particular interval of time. This interval, which we will call an operational window, will depend on the type of operation and the location in the North Sea, and might be calculated on the basis of the climatological mean winds and waves. For example, Shell currently use operational windows based on the recent climatology of wind and waves. We propose that a seasonal forecast might potentially provide a better prediction of the operational window in a coming season, for example allowing it to be extended.

Skilfully forecasting waves for locations in the North Sea for a single month, at a lead time of 2.5 months, is a challenging target. The possibility that we have in mind is to use the GloSea5 seasonal forecasting system to predict, with a lead time of months, that the forecast probability of calm conditions in May is similar to the climatological probability of calm conditions in June. If the forecast on 1st March is that the range of probabilities for the coming May is similar to the climatology for June, then an operational window that is usually June–August might be planned instead to open on 1st May.

We should emphasise that this would not be a prediction that the coming May will be calm enough for the operation, any more than merely knowing the climatology for June could allow us to predict that a particular June will be sufficiently calm. Rather, knowing the climatology for June means that we know that the *balance of probabilities* is that June will be sufficiently calm. This seasonal forecast would similarly indicate that the balance of probabilities is that May would also be sufficiently calm.

We describe the data used in Section 3.2, our key methods in Section 3.3, followed by our results in Section 3.4. The subsequent discussion in Section 3.5 considers how this work might be exploited. Our conclusions are given in Section 3.6.

3.2 Data

The predictand for the forecast methodology that we will describe below is the monthly mean significant wave height in the North Sea from NORA. NORA is a high-resolution reanalysis product of wind and waves for the North Sea, Norwegian Sea and Barents Sea produced by the Norwegian Meteorological Institute. It is reported to be widely used by the North Sea energy

industry. NORA10 (Reistad et al., 2011), the original NORA data set, covered September 1957 to August 2002. It used the ERA-40 reanalysis for initial and boundary conditions, and downscaled atmospheric fields using an NWP model (HIRLAM10 version 6.4.2) on a 10–11 km resolution for a domain that included the entire North Sea. The wave hindcast was generated using a wave model forced with the reanalysis winds. NORA10 was subsequently extended by Aarnes, Breivik, and Reistad (2012) into a second period (September 2002–) using ECMWF operational analyses for initial and boundary conditions. NORA10EI[4] (Haakenstad et al., 2020) is a new edition that uses the ERA-Interim reanalysis for initial and boundary conditions, and covers 1979–2017. It retains the domain of NORA10 and a closely-matching configuration. It has the advantage over NORA10 of using a single consistent method throughout the period. We use significant wave heights for May (1993–2016) from the NORA10EI data set hereafter.

The predictor variable in our forecast methodology is 10 m wind speed from GloSea5 (MacLachlan et al., 2015). We used a 28-member hindcast ensemble comprised of the predictions for May 1993–2016 from the 9, 17, 25 February and 1 March start dates, using the 2020 hindcast runs. These start dates correspond to a real-time forecast that could be issued on 1 March from the 42-member forecast ensemble, which would be assembled by combining the forecast runs from the previous 21 days.

As part of our investigations, we also examined other fields, other months that might be predicted, and other lead times. For GloSea5 we also used mean sea level pressure, and 850 hPa geopotential height; we used hindcast sets for May predictions corresponding to 1 April and 1 May start dates; and we used a hindcast set for September predictions corresponding to a 1 July start date. We also used data for May (1993–2016) from the ERA5 reanalysis (Hersbach et al., 2020) for mean sea level pressure (MSLP), geopotential height at 850 hPa, 10 m wind speed, and significant wave height (SWH).

3.3 Methods

Our forecast methodology uses 10m wind speeds in May, as forecast by GloSea5 on 1st March, to predict the significant wave heights in May in the North Sea from NORA. However, rather than directly relating the winds and waves at each point in space, we first express the interannual variability in each independently, as a set of paired time series and spatial patterns through a principal component analysis (PCA; Wilks, 2020, chap. 13). We identified the major orthogonal patterns of interannual variability of both wind and waves over a common period (1993–2016) using the `eofs` python package[5] (Dawson, 2016). We used 24 years of standardised anomalies from GloSea5 and 23 years from NORA (omitting the year to be predicted).

[4] https://thredds.met.no/thredds/projects/nora10ei.html
[5] https://ajdawson.github.io/eofs/latest/

Because wind speeds are an important influence on wave heights, these expressions of interannual variability, although calculated independently, have a substantial amount of information in common. We can express this covariability between the winds and waves by relating their principal components through a canonical correlation analysis (CCA; Wilks, 2020, chap. 14), which is the second step in our methodology. The CCA relates the two sets of principal components from the common 23-year period, yielding pairs of so-called canonical variates that relate the principal components from the winds to those from the waves. The canonical correlation analysis was based on the `cross_decomposition` module[6] in the scikit-learn python package (Pedregosa et al., 2011), although it required some further manipulation to implement the method of Wilks (2006). Our implementation was verified against worked examples in Wilks (2006).

Taken together, the PCA and CCA allow us to build a forecast algorithm (Wilks, 2020, sec. 14.2.3). The values of the wind principal components for the year to be predicted have been calculated as part of the PCA. The canonical variates for the wind for that year may be obtained directly from these values. A forecast is made of the paired canonical variates for the waves by weighting the wind canonical variates by the canonical correlations (Wilks, 2020, equation 14.23). From these forecast canonical variates we may construct first the principal components for the waves, then combine these with the original EOF patterns to forecast the wind anomaly field.

A further step must be taken to adjust the magnitude of the forecast anomaly field, as the step of weighting by the canonical correlations has the effect of dampening the variability in the eventual forecast. The forecast anomaly is therefore 'inflated' (Barnston, 1995, p422) by making an estimate for the other 23 years using the same algorithm, calculating the variance, and multiplying the forecast anomaly by the ratio of the forecast and reanalysis standard deviations.

We adopted a leave-one-out approach to evaluate the performance of these forecasts. A full analysis was performed for each year in turn (1993–2016), as it would be for a real-time forecast. We excluded the reanalysis for that year until the forecast had been made. Thus we used 23 years of the 28-member hindcast ensemble to build a predictive model, and a 28-member ensemble for the year to forecast. Note that in a real-time forecast, we would have an additional year for model-building, and a 42-member ensemble for the year to forecast.

Once the forecast was made, we compared the forecast with the reanalysis. We used the Pearson correlation coefficient between the full set of forecasts and outcomes to measure the potential skill of a forecast (Jolliffe and Stephenson, 2012, sec. 5.4.4).

[6] https://scikit-learn.org/stable/modules/generated/sklearn.cross_decomposition.CCA.html

3.4 Results

We are looking to forecast significant wave heights in the North Sea. We therefore start with their climatology. Figure 11 shows the climatological mean significant wave height for May, from ERA5 and NORA. The patterns are similar, but the NORA reanalysis is approximately 0.2 m higher across much of the domain. The coastal diminution of wave heights is much more tightly defined in NORA than in ERA5, due to NORA's much greater horizontal resolution of 0.1° compared to 0.5° in ERA5.

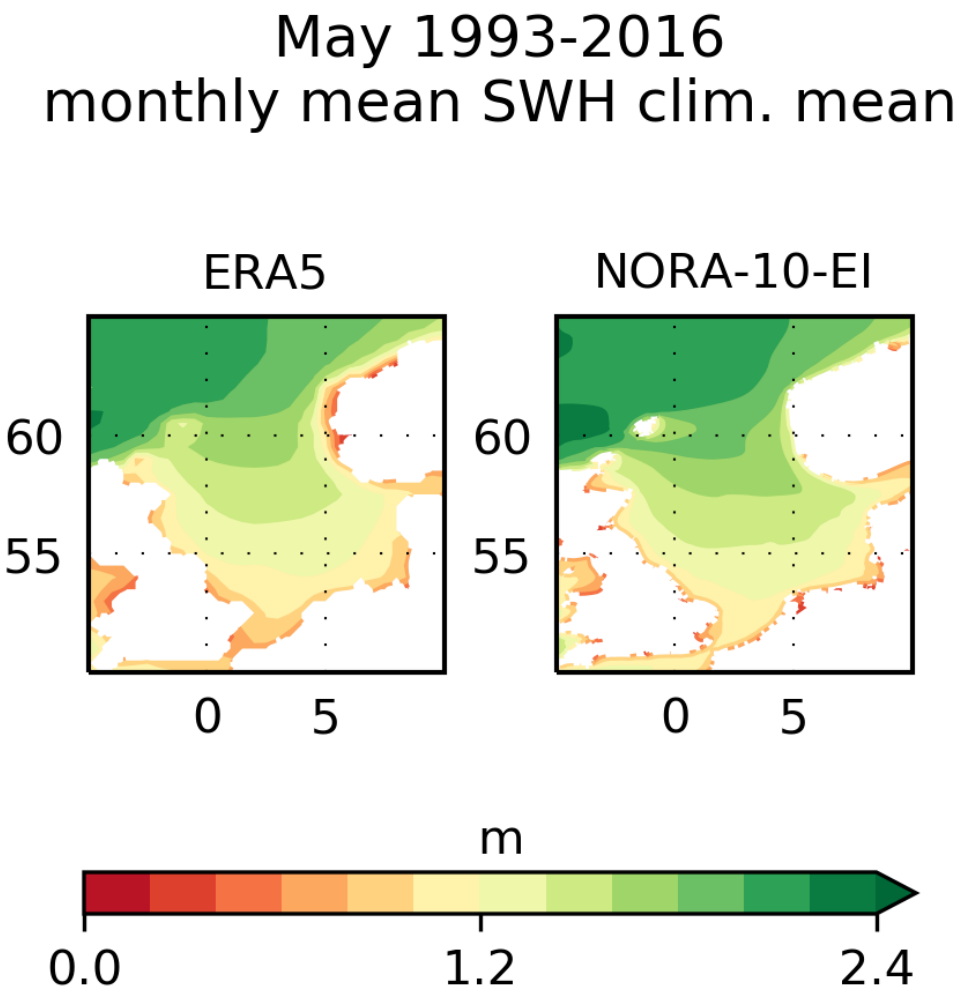


Figure 11. The climatological significant wave heights for May, from ERA5 (left) and NORA (right), based on the 1993–2016 period.

The basis for the hoped-for seasonal forecast capability is that the climatological significant wave heights in May and June are sufficiently similar for some Mays to approximate to some Junes. Figure 12 shows that this is the case. The difference in means between May and June is a similar magnitude to the interannual standard deviation in May. Therefore it is not unusual for a May to have a similar mean significant wave height to the climatological mean in June.

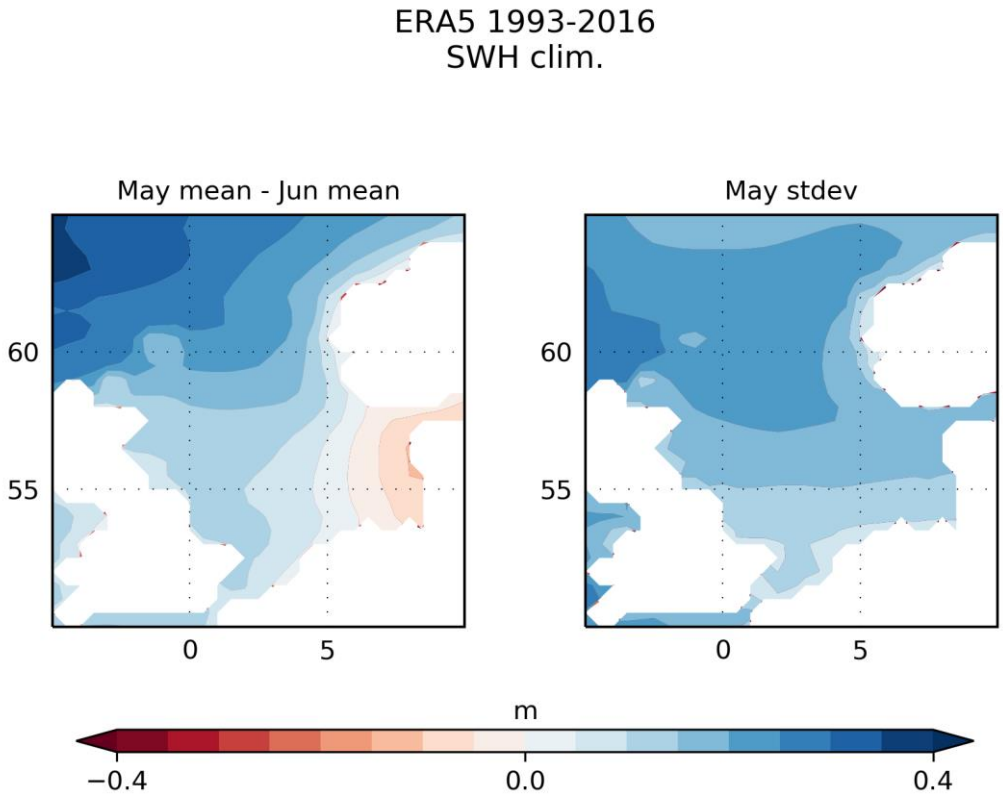


Figure 12. Climatological statistics of significant wave height in ERA5 (1993–2016). The left panel shows the difference in climatological mean between May and June, and the right panel shows the interannual standard deviation for May.

When we attempt to forecast SWHs in the North Sea from GloSea5 we will need to use some variable other than SWH, as this is not directly represented in that model. As described above, we have investigated mean sea-level pressure (MSLP), 850 hPa geopotential height and 10m wind speeds as possible predictors. The first two offer potential ways to capture the influence of the large-scale atmospheric flow on the SWHs. The 10 m wind speeds are likely to capture the more immediate influence on wave heights from nearby winds.

We found a locally-strong relationship in the Baltic with the large-scale flow (not shown), but not in the North Sea area required here. Therefore the results that we show here focus on the

10 m wind speeds, which we found to be more useful for the North Sea. Figure 13 shows the climatology of the 10 m wind speed in the North Sea. GloSea5 captures well the pattern of land–sea contrast seen in the reanalysis.

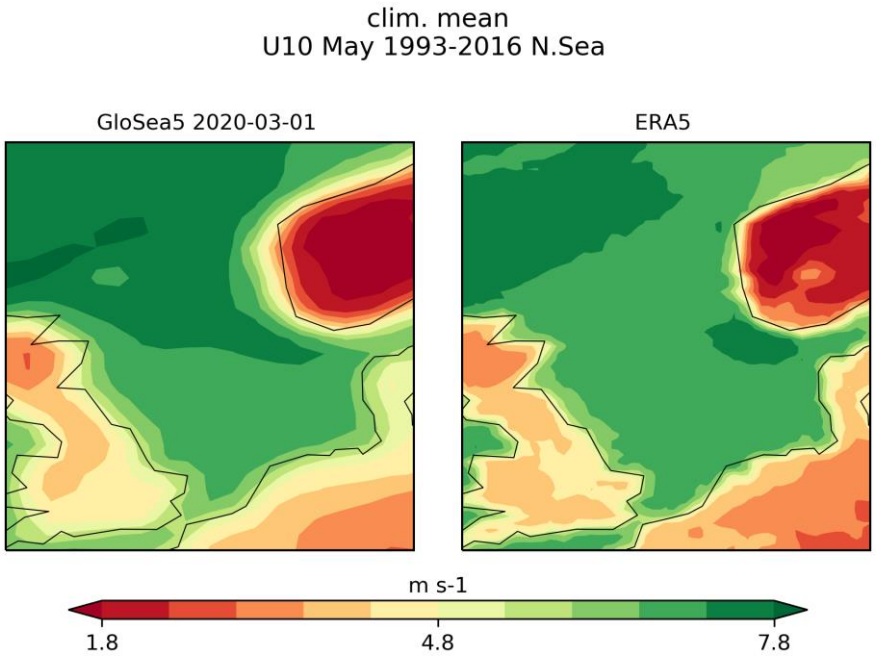


Figure 13. The climatology of 10 m wind speed in GloSea5 (left) and ERA5 (right) for May 1993–2016.

Figures 14 and 15 show the year-to-year variations in the 10 m wind speed from GloSea5 and ERA5. The averaging over 28 ensemble members in GloSea5 yields a narrower range of interannual variability than in the reanalysis.

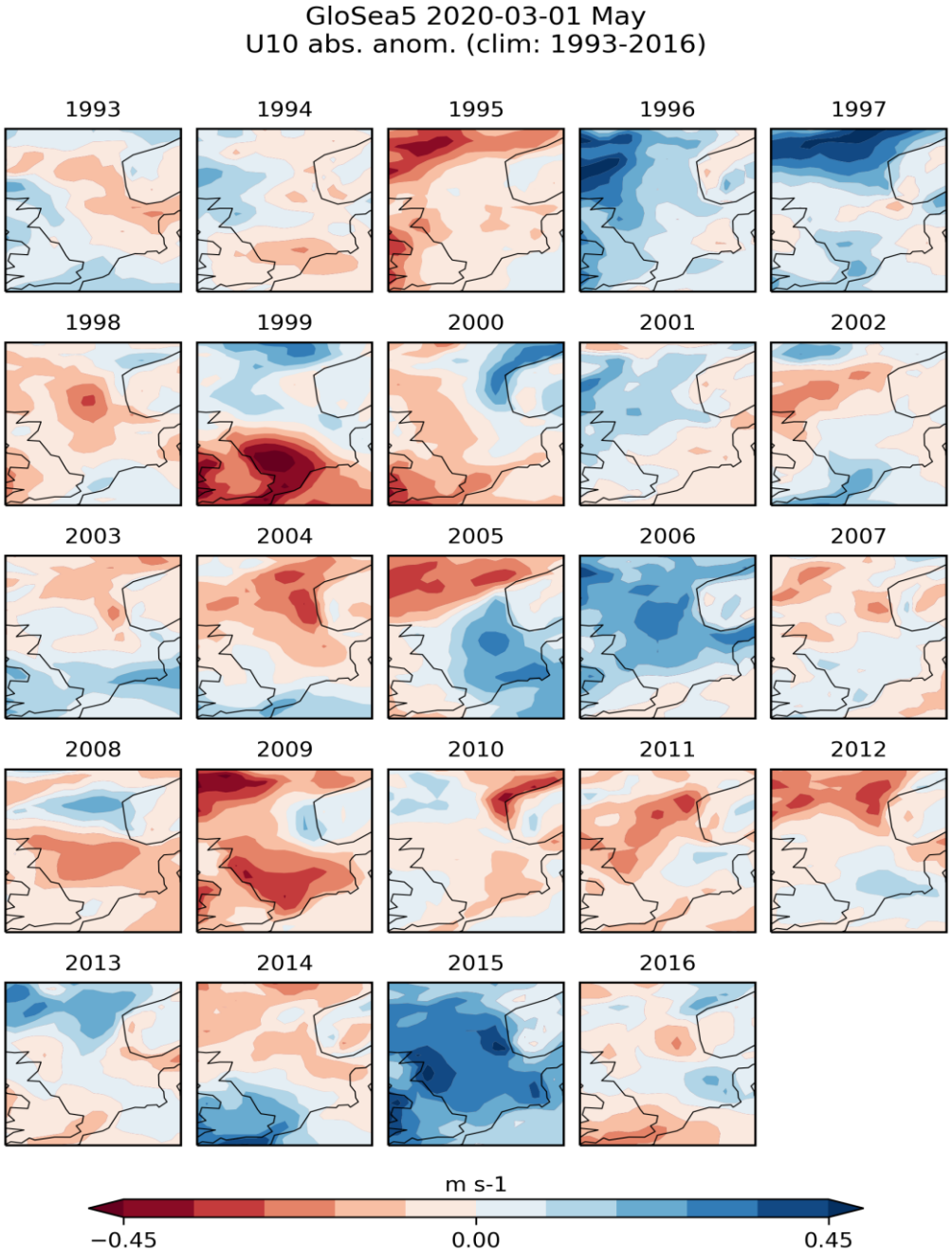


Figure 14. The ensemble-mean anomalies of the GloSea5 10 m wind speed in May from the 28 hindcast runs of GloSea5 up to 1 March 2020.

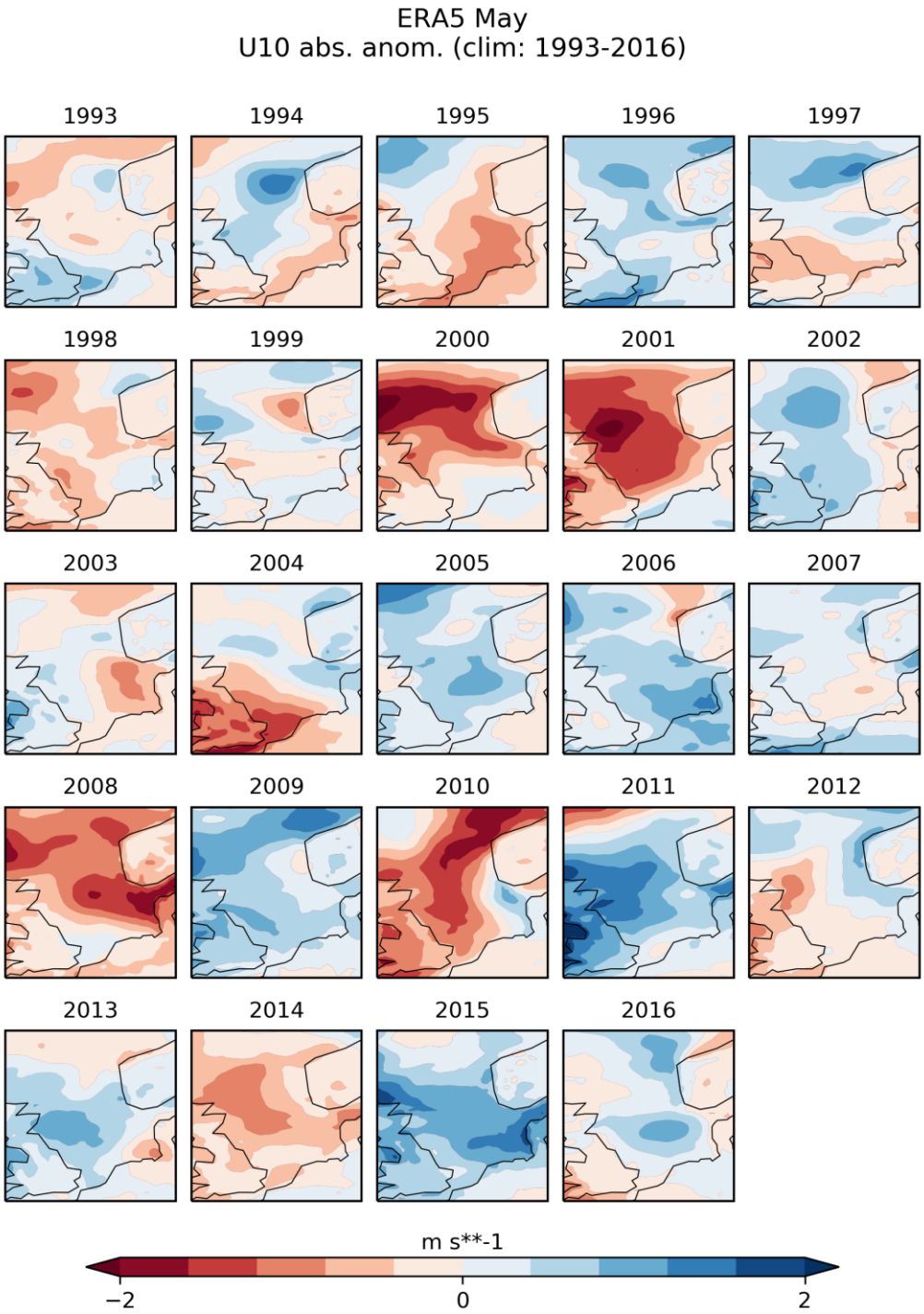


Figure 15. The anomalies of the ERA5 10 m wind speed in May.

The left-hand panel of Figure 16 shows the correlation between these two sets of year-to-year 10 m wind speed variations. There is a 'bulls-eye' of strong correlation (>0.6) in the eastern North Sea that should be treated with some caution and, this aside, the point-by-point correlations are not strong. If instead we forecast the 10 m wind speed using the CCA-on-PCA method, we obtain the correlations shown in the right-hand panel of Figure 16. In detail, a principal component analysis was performed on both the ERA5 and GloSea5 wind speeds, omitting one year. The time-matched parts of the first three PCs were fed into either side of a CCA. The omitted year of the GloSea5 PCs was supplied to the CCA relationships to estimate the GloSea5 canonical variates (CVs) for that year, and these were then used to estimate the CVs applicable to ERA5. The first three PCs for ERA5 were estimated using the estimated CVs for ERA5. The ERA5 field for the omitted year was estimated using the estimated PCs for ERA5. This was repeated, omitting each year in turn.

The right-hand panel of Figure 16 shows a sizeable area of positive correlation covering much of the southern North Sea, peaking at over 0.5. Since these numbers are obtained using a leave-one-out approach, there is good reason to think that GloSea5 has some skill in forecasting 10 m wind speed in May in the southern North Sea, and that our method offers an improvement on direct wind speed forecasts.

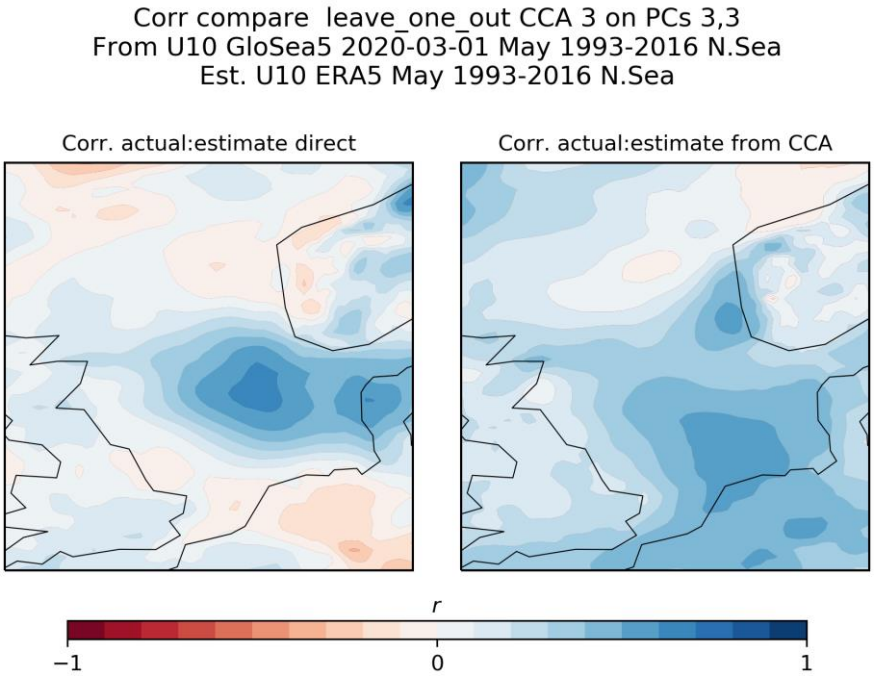


Figure 16. Pearson correlation coefficient between 10 m wind speeds from ERA5 and our forecast estimates. The left panel uses GloSea5 wind speed anomalies directly, and the right panel uses wind speeds estimated from out CCA-on-PCA analysis. We use wind speed data for May 1993–2016 in both cases.

Figure 17 shows that there is a strong relationship between the 10 m wind speeds and significant wave heights in the North Sea, whether from point-by-point correlations of SWH data (left panel) or through a CCA-on-PCA analysis (right panel). Taken together, Figures 16 and 17 suggest that it is worth examining whether the 10 m wind speeds from GloSea5 might be used to forecast significant wave heights in the southern North Sea.

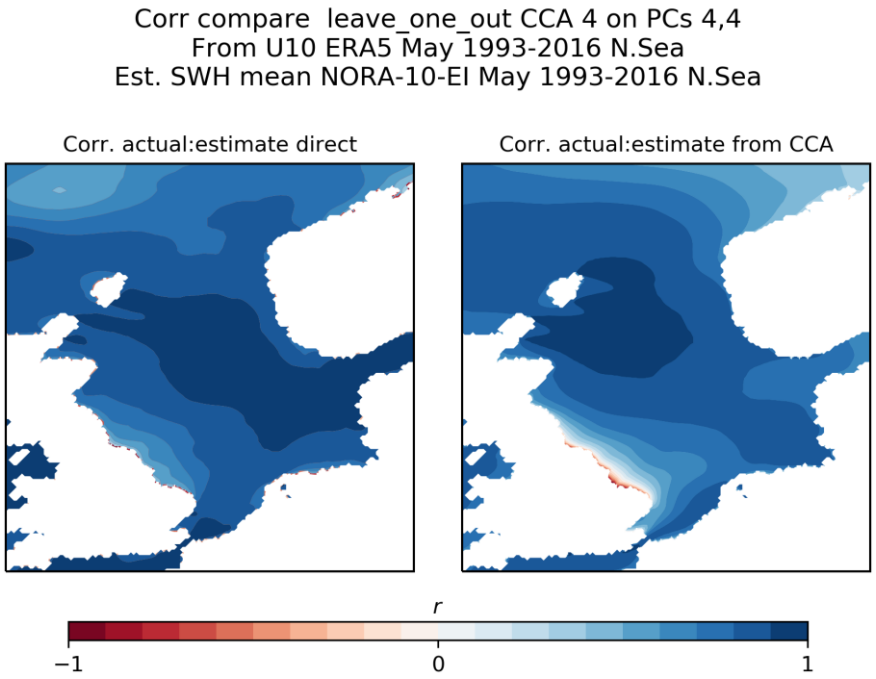


Figure 17. Pearson correlation coefficient between significant wave height (SWH) standardized anomalies from NORA, and estimates from ERA5. The left panel uses standardized 10 m wind speed anomalies from ERA5, and the right panel uses the estimates of SWH using the CCA-on-PCA method, with ERA5 standardized wind speed anomalies as the predictor.

Figures 18 and 19 show the first four EOFs from principal component analyses performed on the GloSea5 10 m wind speeds and the NORA significant wave heights. Over 70% of the NORA significant wave height variability (Figure 19) is represented by a domain-wide amplitude pattern that is centred on the North Sea, and a further 20% by SW–NE and SE–NW patterns. The leading EOF from the GloSea5 10 m wind speeds (Figure 18) is also an amplitude pattern centred on the North Sea. The next two EOFs are also SE–NW and SW–NE patterns. Together these first three patterns account for three quarters of the variance in GloSea5. The similarity of the patterns seen in the two data sets, the amount of the variance that they explain, and the fact that the reanalysis wave heights are driven by reanalysis winds, all suggest that it is reasonable to use one set of patterns to predict the other.

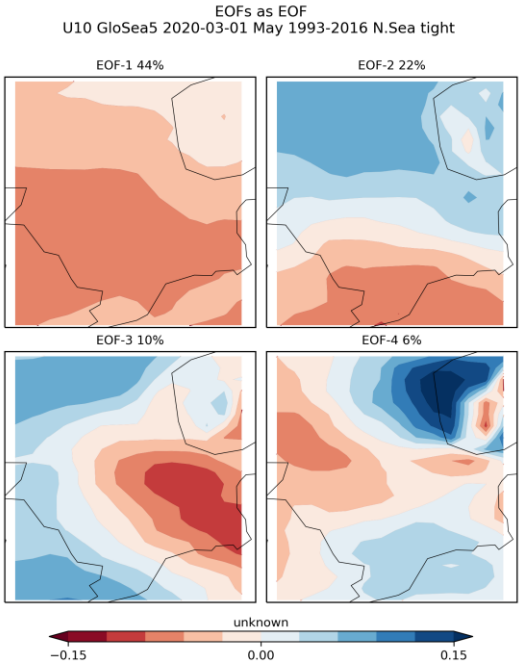


Figure 18. The first four EOFs from a PCA on the mean 10 m wind speed in May 1993–2016, from the 28 hindcast runs from GloSea5 up to 1 March 2020.

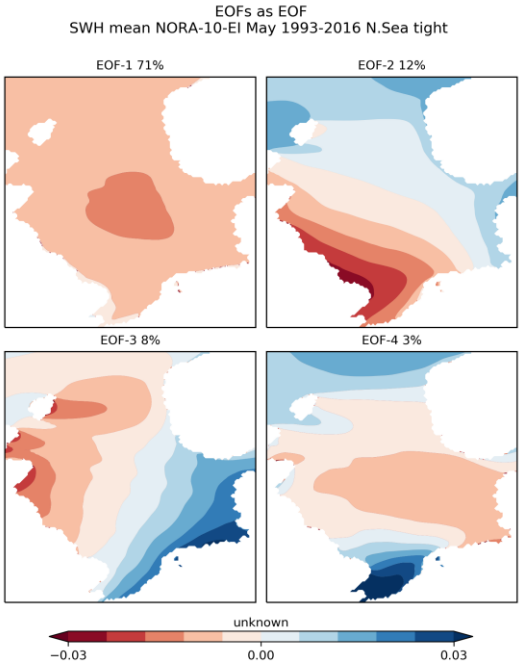


Figure 19. The first four EOFs from a PCA on the NORA mean significant wave height in May 1993–2016.

Figure 20 shows the results of a canonical correlation analysis performed on the leading three principal components from the data shown in Figures 18 and 19. The leading canonical correlation (0.67) is an indication of the strength of the relationship between these patterns.

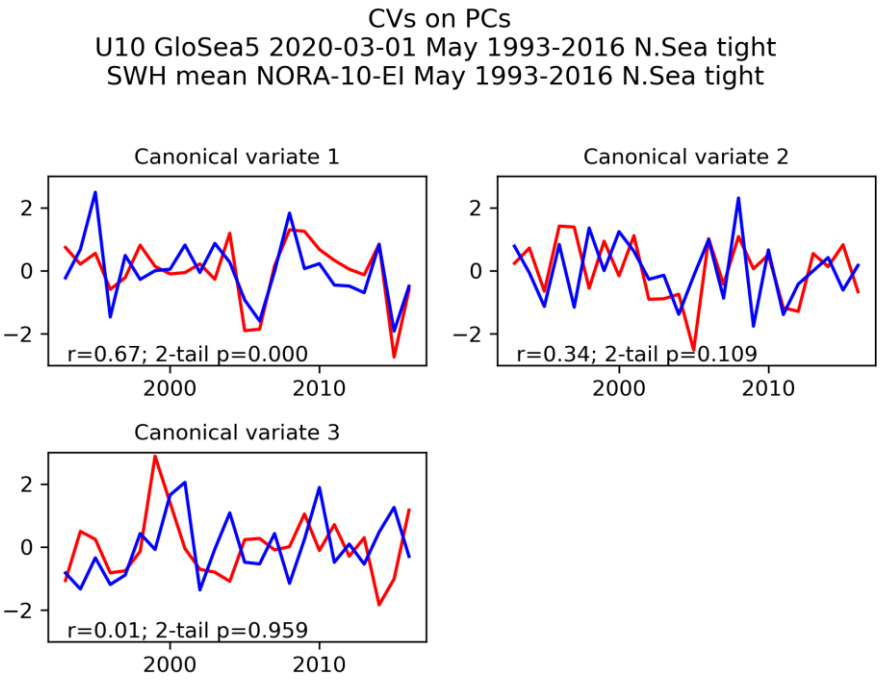


Figure 20. The three pairs of canonical variates from a CCA performed on the first three PCs from Figures 18–19.

The left-hand panel of Figure 21 shows, as a point of reference for the forecast that we will make, the point-by-point correlations between the GloSea5 10m wind speeds and the NORA significant wave heights. The pattern is broadly similar to that between the GloSea5 and ERA5 wind speeds (Figure 16, left), albeit somewhat stronger. We use the relationship between the principal components established by the canonical correlation analysis to predict significant wave heights using 10 m wind speeds from GloSea5. Figure 21 (right) shows the correlation between these predictions and the reanalysis. As when forecasting the wind speed (Figure 16, right), there is a sizeable area of positive correlation covering much of the southern North Sea, peaking at over 0.5.

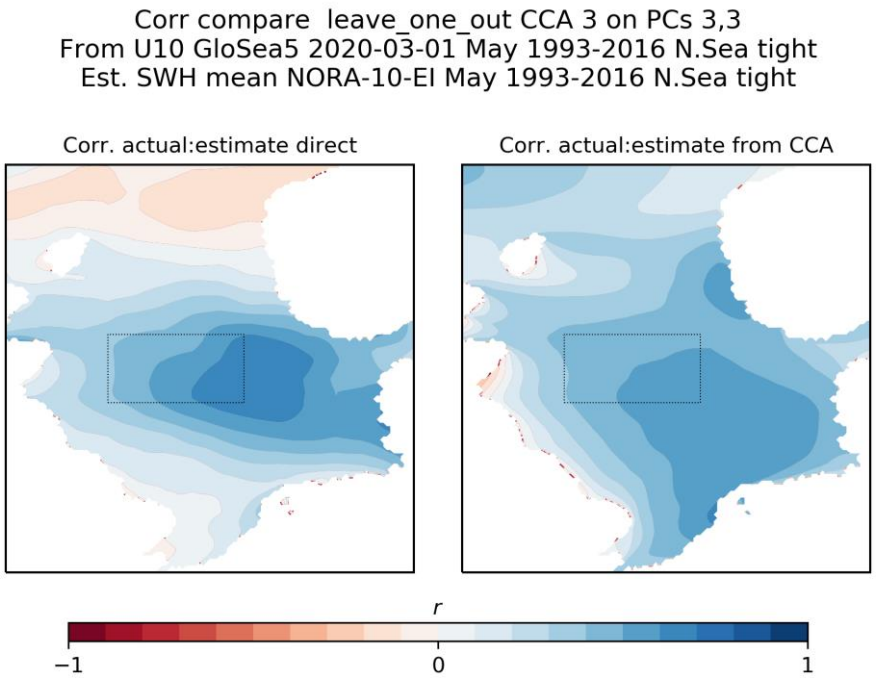


Figure 21. Pearson correlation coefficient between standardized significant wave height anomalies from NORA, and estimates from GloSea5, for May 1993–2016. The left panel uses standardized 10 m wind speed anomalies from GloSea5, and the right panel uses estimates of SWH using the CCA-on-PCA method, with GloSea5 standardized wind speeds as the predictor. The dotted outline indicates the "Gannet" box used elsewhere.

We define a box (shown in Figure 21) that we call "Gannet", which extends over a sizeable area of the North Sea[7] and includes Shell assets east of Aberdeen. Figures 22 and 23 show the forecast for this box obtained by averaging the point forecasts from the CCA-on-PCA analysis described above. The correlation of 0.49 should be treated with caution, as the two very strongly positive forecast years make a major contribution to it. However, it should also be noted that in none of the 10 years with the calmest forecasts was there an outcome of wave heights more than 0.2 standard deviations above the average for May. In other words, when the predictive model forecast a higher probability of relative calm, such that an earlier opening of the operational window might be encouraged, the outcome was indeed a near-average or relatively calm May.

[7] 56–58° North, 0–4° East

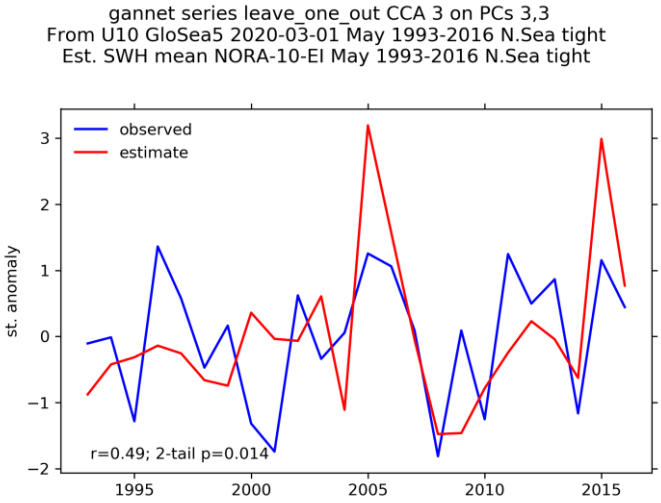


Figure 22. Forecast and reanalysis mean May SWHs for the Gannet box. The forecasts and reanalysis described in Figure 21 were averaged into box-wide means.

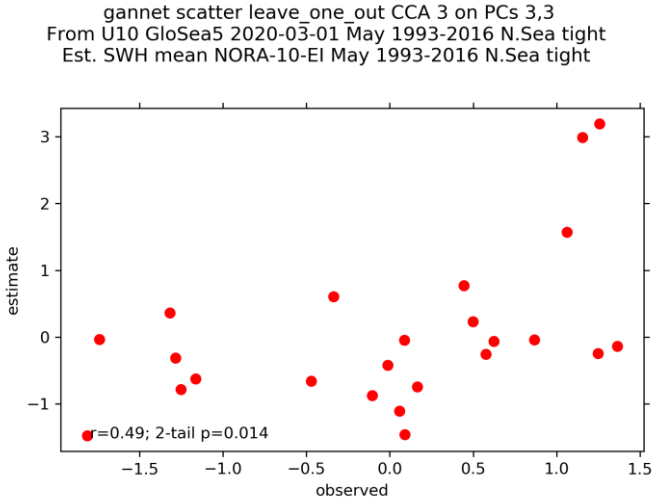


Figure 23. Scatter plot of the two time series of standardized significant wave height anomalies from Figure 22.

3.5 Discussion

The results reported here suggest that there may be some scope for making a forecast on 1st March of mean significant wave height in May in the southern North Sea. Such a forecast

would not predict a particular height, but would rather inform the recipient that the probability of a calmer-than-average May was higher/normal/lower compared to recent Mays. In certain years, the forecast might be that the probability distribution for the mean significant height in May from the seasonal forecast is similar to the probability distribution for the mean significant height in June from climatology.

The plots above are not likely to be the form in which a forecast product would be presented to a recipient. We anticipate developing the form of the product in collaboration with the eventual recipient to ensure that it corresponds to their decision-making processes. We anticipate that further statistical manipulation of the results might be valuable. In particular, it might be useful to express a May forecast as a probability distribution that might be compared with the climatological probability distributions for May and June. The skill of the forecast might be assessed through a ranked probability skill score.

If the potential skill investigated here is eventually to be used in a product, some additional steps will be required, as the NORA reanalysis only extends to 2017. The significant wave heights from ERA5 can be used for forecast evaluation, as they are updated in near-real-time. This will require model-building with ERA5, or possibly model-building with NORA and separately transforming from a NORA-based forecast to an ERA5 equivalent.

We have investigated conditions for May in this project in part because we understand that the shoulder months around summer might be of particular interest when looking to extend the operational window for activities in the North Sea. For May we found skill in the southern North Sea, but not in the northern North Sea. A similar approach might be taken for any calendar month or area if required. However, we should caution that our ability to issue a forecast depends on there being skill in the seasonal forecast model for that particular month and location.

This investigation relies on a single seasonal forecast model. It is not unusual to find that an ensemble of such models can improve the skill of a forecast. If a forecast such as that proposed here would be useful, then it may be worth extending the basis of the forecast to include multiple seasonal forecast models.

The results reported here represent only a small part of the analysis that was done in this study. Since these delivered ‘null’ results they are omitted for clarity. We attempted to find predictive skill by using mean sea level pressure and 850 hPa geopotential height fields from GloSea5, and over spatial domains extending to the entire north Atlantic–European sector. We tried an approach in which a canonical correlation analysis was performed at each grid point, so as to relate the predictor principal components to the series at that grid point. (The intention there was to eliminate any loss of skill arising from performing a principal component analysis on the predictand.) We looked at forecasting for later dates (1 April, 1 May), i.e. shorter lead times, than the 1 March forecasts reported above. We looked at what predictive skill there might be for a September forecast on 1 July. In no case did we find potential predictive skill to

match what we have reported above, or indeed at any level that might potentially support a forecast product.

When investigating potential predictability, it is always important to make trial forecasts under conditions comparable to those of a real-time forecast. It is particularly important when using the approach taken here of relating principal components using a canonical correlation analysis, as there is very considerable scope for overfitting. Indeed, if enough principal components are used, then by construction a near-perfect prediction is guaranteed when forecasting with data that has already been employed in building the predictive model. To avoid over-fitting and hence misplaced optimism regarding skill levels, a leave-one-out cross-validation approach was adopted to assess skill. This means that the skill presented in this report is a good estimate of the real-world potential skill of the forecast described.

3.6 Conclusion

We have found some scope for making a forecast on 1st March of mean significant wave height in May in the southern North Sea. In order to take this forward, we need to understand from potential users whether such a forecast might influence a decision, and if so, what decision this would be. This process has been initiated with a discussion with a Senior Metocean Engineer at Shell. Here the most relevant location, time period ad format of the forecast information was discussed. This discussion opened up the possibility of extending the forecast range for decision-making even further than investigated in D3.8. However, further work to identify skill specific to these choices would be required to take this opportunity further.

4 The application of weather types using ADAMONT

4.1 Overview

This section reports experiments conducted with ADAMONT, a tool developed by Météo-France, which is used here to correct seasonal forecast data. It is a fact that raw seasonal forecast data are affected by systematic errors and drifts in time that need to be corrected before being used as input to impact models.

ADAMONT relies on quantile mapping corrections conditioned on weather regimes. The impact of using weather regimes has not been precisely documented in previous studies, although we could reasonably suppose it leads to a finer correction. One of the main objectives of this study was to better understand the role weather regimes play, if/when/where/for which parameter we gain in predictability. Another focus of research was to test some strategies to optimize the positive impact of the introduction of weather regimes.

A first series of experiments were done on an idealized case of “perfect forecasts” of weather regimes. It allowed us to point out the maximum benefits we could expect from ADAMONT, for each parameter, and which regions in Europe could benefit the most.

Another set of experiments, with hindcast data from a seasonal forecast model, was conducted with two versions of ADAMONT: “with” and “without” weather regimes. These “real forecast” experiments were useful to measure the actual added value of introducing weather regimes in the correction process, in an operational context. They helped to demonstrate that the poor predictability of weather regime frequencies could annihilate the potential benefit of using weather regimes.

In order to overcome this weakness, some alternative approaches have been tested. The first one consisted of sub-sampling the ensemble forecast set, and to keep only the “best members” in terms of weather regime prediction. The second one relied on a multi-model combination. A third approach used a mix of the first two approaches. It opened up operational perspectives.

4.2 Introduction

For many strategic decisions where meteorological forecasts are needed, the time horizon of a week classically offered by meteorological forecasts is not long enough. Thanks to major projects (such as the WWRP/WCRP Sub-seasonal to Seasonal Prediction project or the EU FP7 EUPORIAS project), and easier access to state-of-the-art model outputs (for instance through the Copernicus Climate Change Service programme), sub-seasonal to seasonal forecast demonstrators have been set up in recent years in many sectors like hydrology, agriculture or energy. SECLI-FIRM follows this track, in associating climatologists and sectoral users to make the best use of seasonal forecast data and measure their utility.

In the framework of a work-package dedicated to optimization of climate prediction performance, we have investigated opportunities offered by a correction tool developed by Météo-France called ADAMONT. This tool, originally designed to correct climate projection scenarios over mountainous regions, has been adapted for seasonal forecast data (in the H2020 PROSNOW and MEDSCOPE projects) to produce corrected data at an hourly timescale in input to snowpack models. In this study, we have used the first steps of this adaptation of ADAMONT, to obtain daily corrected data of 10-metre wind-speed, precipitation and 2-metre temperature over Europe, at 0.25° resolution. From these daily data, we have computed monthly and three-month mean datasets expected for SECLI-FIRM case-studies.

ADAMONT relies on a classical quantile-mapping correction. One of its main innovations comes from the introduction of Weather Regimes (WR, as defined in Michelangeli et al., 1995) in the process: the quantile-mapping correction is conditioned on the WR. Thus local meteorological conditions, which depend on large-scale circulation (especially over regions with complex orography), should be better corrected.

In this report, we present in section 4.3 the datasets, ADAMONT and the Weather Regimes. The performance of ADAMONT applied in a “perfect forecast” experiment and on a Seasonal Forecast model are presented in section 4.4, as well as the WR predictability of the Seasonal Forecast model. In section 4.5, we explore some strategies of utilisation, such as member selection and a multi-model approach, and measure their impact on prediction scores. And in section 4.6 we conclude with some operational perspectives.

4.3 Data and Methodology

4.3.1 Data sets

This study relies on two types of data, freely available on the Climate Data Store of the Copernicus Climate Change programme (https://cds.climate.copernicus.eu/):

- seasonal forecast data. Four models have been chosen: **ECMWF** SEAS5, **Météo-France** (MF) system 7, Centro Euro-Mediterraneo sui Cambiamenti Climatici (**CMCC**) SPS3 and Deutscher Wetterdienst (**DWD**) GCFS 2.0. We have mainly studied the target season December–January–February (DJF), with a start date in November. We have used a historic period of 24 years (DJF 1993–1994 to DJF 2016–2017). The spatial resolution is 1°, and the temporal resolution is 6h for temperature and wind-speed, 24h for precipitation.

- **ERA-5** reanalysis data, at 0.25° resolution, at the same temporal resolution as the seasonal forecast data.

The domain of interest is Europe (11° W to 20° E, 35°S to 60° N), in order to cover areas of interest of several case-studies of SECLI-FIRM.

4.3.2 Weather Regimes methodology

The weather regimes (Michelangeli et al., 1995) or patterns of equivalent circulation, aim to provide a classification of daily recurrent large-scale states in the circulation over a wide box spanning the North Atlantic and Europe. The patterns are found in terms of geopotential or pressure fields. As for weather types, the patterns are identified by grouping together daily fields close to each other to create clusters of fields different from each other. In our case the classification is made by clustering the daily Mean Sea Level Pressure (MSLP) anomalies (compared to the 1981–2010 monthly climatology) of an observed/reanalysis dataset, in the case here presented ERA5. To obtain winter regimes, the clustering is done on December, January and February daily anomalies of ERA5 MSLP, over the period 1981–2010. The considered geographic area spans the North Atlantic and Europe (85.5° W – 45° E, 27°S – 81° N).

The algorithm proceeds as follows: after weighting the MSLP anomaly gridpoints according to the latitude, the algorithm performs a Principal Component Analysis (PCA) that retains the first 9 principal components (~80% of the variance). Then the centroids are found in two successive steps.

1. Preliminary centres are determined by repeating (1000 times) a k-means classification on a sub-sample (1000 dates) and by fixing 4 target centres. The final centres are the means of these 1000 iterations.
2. From these centres, we proceed with a k-means classification over the other dates (1000 iterations), in such a way that each daily anomaly field is uniquely attributed to one centre.

The centroids are the mean fields of each of these 4 clusters. Figure 24 shows the MSLP anomaly fields that represent their centroids.

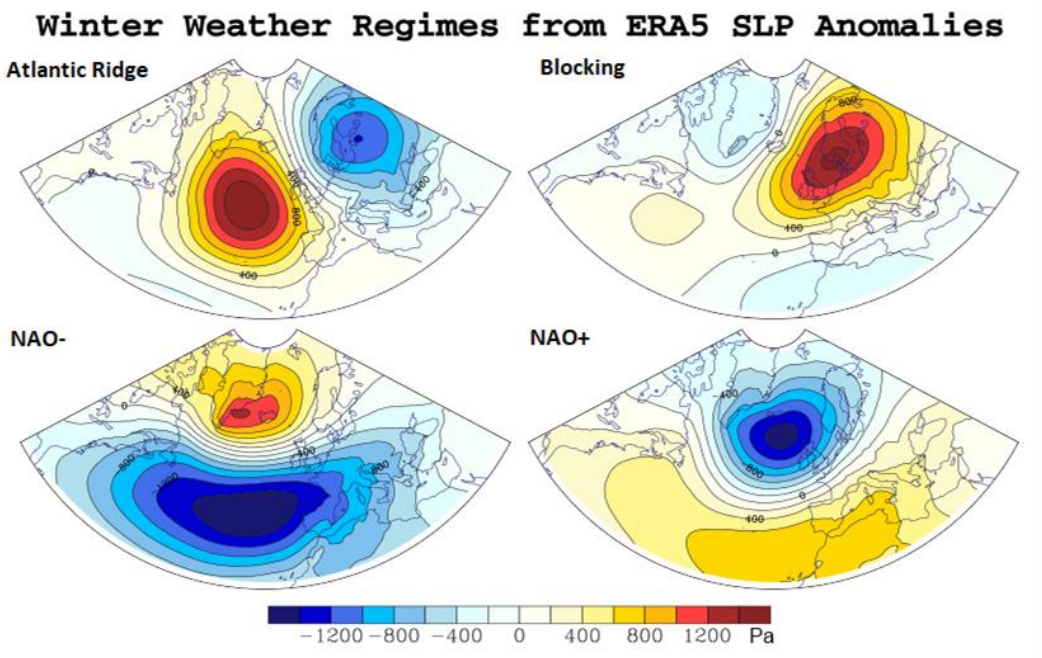


Figure 24. Anomaly fields of MSLP corresponding to the centroids of the 4 classical winter weather regimes found applying the algorithm presented in section 4.3.2. Data is from ERA5 1980–2010 in winter months DJF.

Daily MSLP anomaly fields from the same reanalysis on other dates as well as fields from any seasonal forecast model over the same geographic area can thus be attributed to one of the 4 centres identified in the previous steps: each field is attributed to the closest centroid, where the proximity is calculated using the Euclidean distance on the area-weighted fields.

The interest in using the weather regimes lies in the known impact that those regimes of circulation have on environmental variables such as 2m temperature, precipitation and wind. Figures 25, 26 and 27 show the daily anomaly fields of DJF precipitation, 2m temperature and wind respectively.

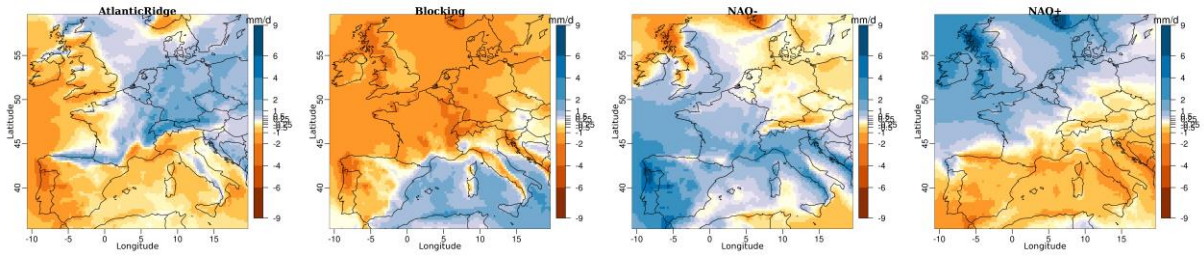


Figure 25. Maps of the impacts in DJF over Europe of the 4 classical weather regimes on precipitation. The maps show the anomalies with respect to the complete fields

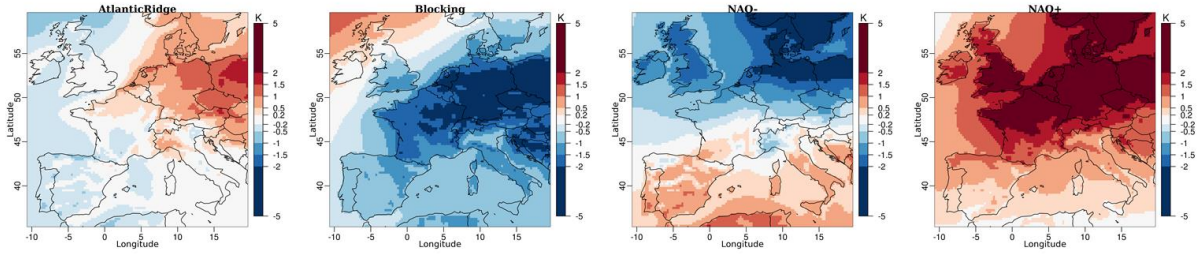


Figure 26. As Figure 25, but for temperature.

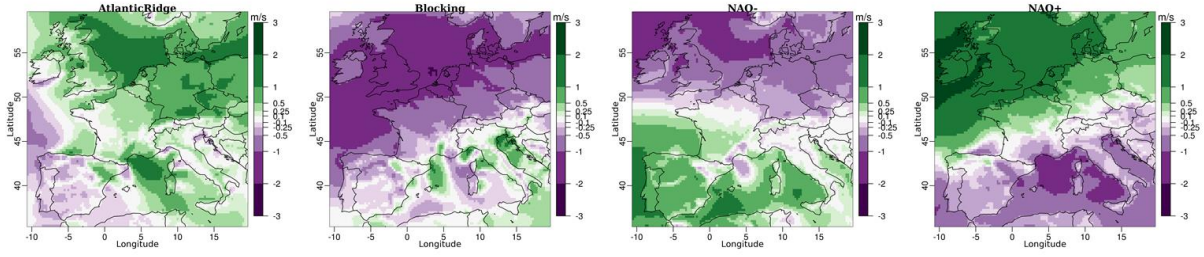


Figure 27. As Figure 25, but for wind speed.

The knowledge of those impacts seems beneficial in improving the usability of seasonal forecasts (SF). Indeed, by conditioning a correction of seasonal forecasts on WR (i.e. applying a different correction, depending on the large-scale circulation), we can expect to refine the correction, benefiting from the knowledge of the link between WR and local impact.

4.3.3 ADAMONT: a quantile mapping based on weather regimes

One way to include WR in a bias correction/downscaling procedure is the one proposed in the ADAMONT method (Verfaillie et al., 2017).

ADAMONT is a statistical method that

i. adjusts a set of parameters of daily climate projections from regional climate models (RCM)
ii. disaggregates the daily corrected values at sub-daily time steps.

Figure 28 details the main steps of ADAMONT according to Verfaillie et al. (2017); the reader is invited to refer to Verfaillie et al. (2017) for a complete treatment of the methodology. We adapted the method in order to make it suitable for the correction/downscaling of parameters from SFs. We focus on precipitation, 2m temperature and wind, as these are the relevant parameters in the case study experiments for energy production within SECLI-FIRM.

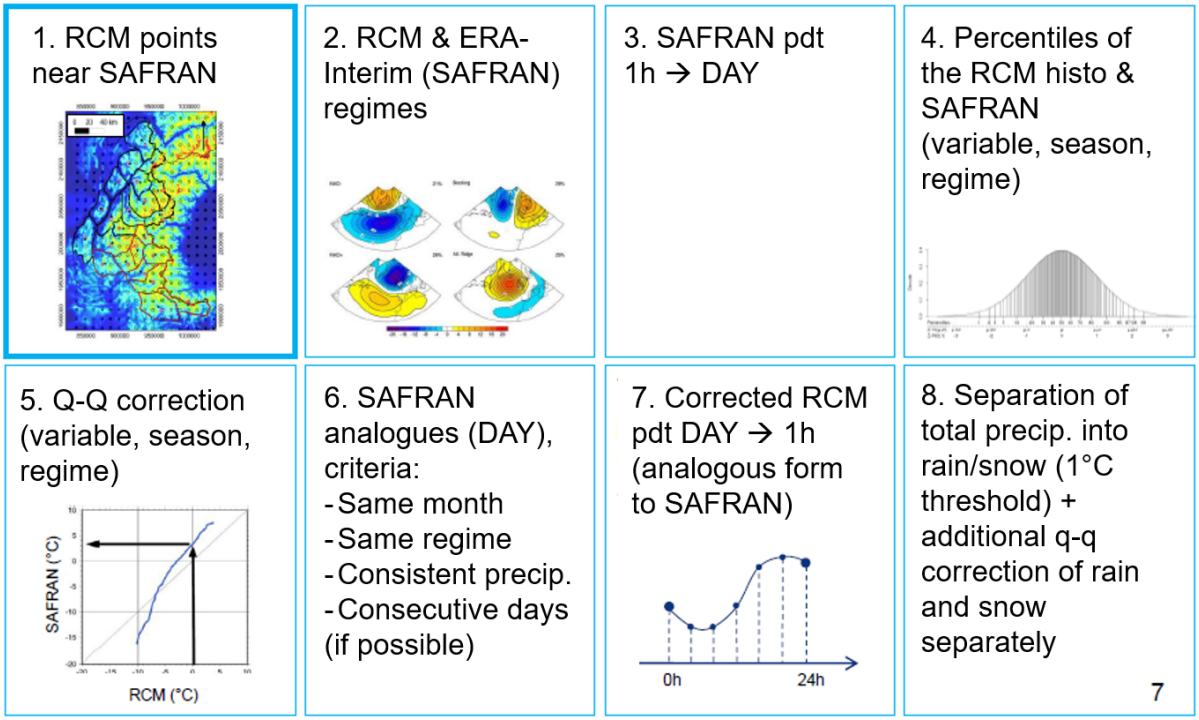


Figure 28. The treatment steps of ADAMONT method (according to Verfaillie et al., 2017).

The adjustment phase of ADAMONT, (i), applies a quantile mapping (QM) (Themeßl et al., 2012, in the context of bias correction of climate models) which is refined to be conditioned on

weather regimes. We identified indeed at least 2 main reasons to confidently proceed in refining the simple QM taking advantage of the information given by the WRs:

1) SF model bias is regime-dependent, meaning that the pattern of bias over a geographical area can present both different magnitude and sign depending on the concurrent weather regime. (see Figure 29). For example, concurrently to the NAO-, the model is incapable of reproducing the humid impact of the regime over the Mediterranean Sea, thus showing a negative bias. North to the Alps we similarly see an opposite pattern with respect to the bias during the Atlantic Ridge. At the same time, there are areas such as the Alps, where the bias is constant independent of the WR.

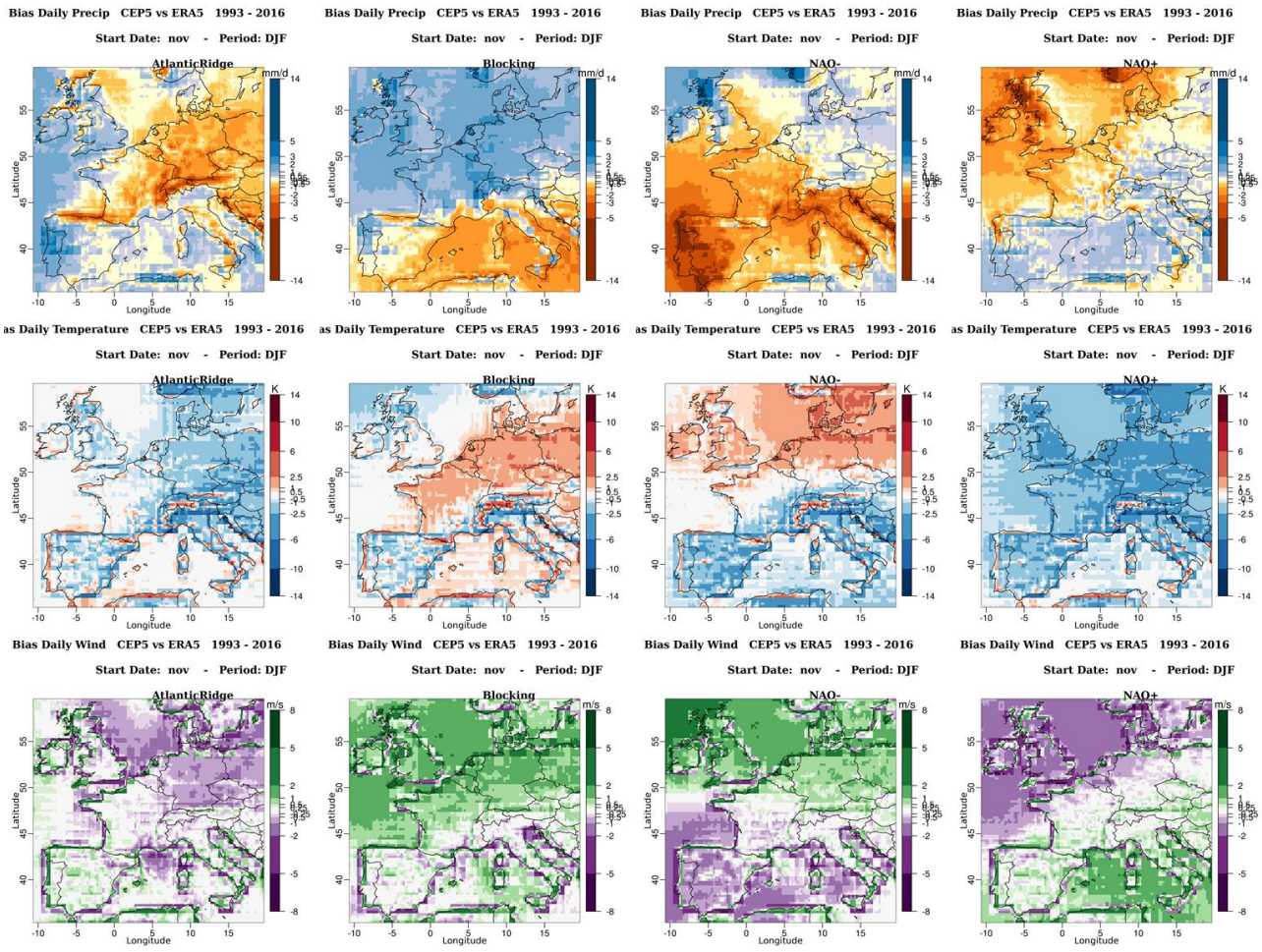


Figure 29. Bias of daily DJF precipitation (first row), temperature (second row) and wind (third row) in SEAS5 compared to ERA5 concurrently to the 4 winter WR (Atlantic Ridge, Blocking, NAO-, NAO+).

2) The known impacts WRs have on the environmental variables in both space and time. This is explained and shown in Figure 25 and following figures. Moreover, those impacts are visible on the entire probability distribution of the environmental variables, not only on their mean. As an example, Figures 30 and 31 report the 4 CDFs of Temperature in Toulouse in Januaries 1993–2016 for the 4 WR (colours and names). The distributions are not only shifted, but their shape is different for each regime. This is particularly evident in the NAO- case.

Temperature ERA5 – Toulouse

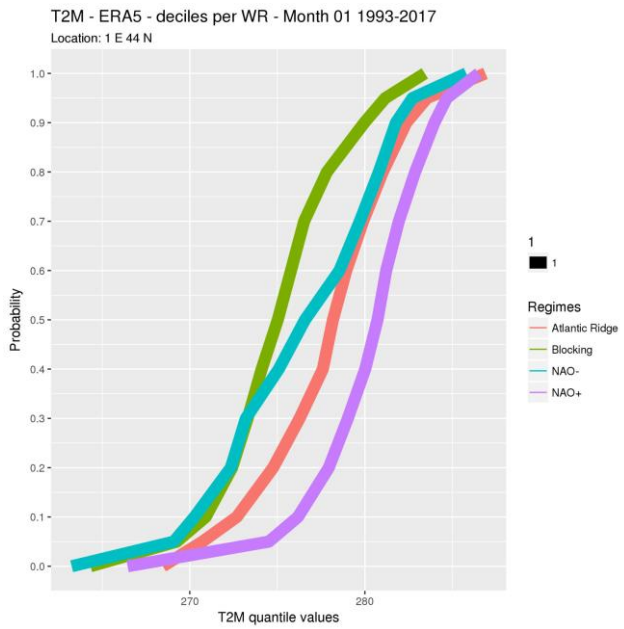


Figure 30. Deciles of observed daily mean Temperature in January 1993–2016 in Toulouse (ERA5) according to the 4 weather regimes.

Temperature SEAS5 – Toulouse

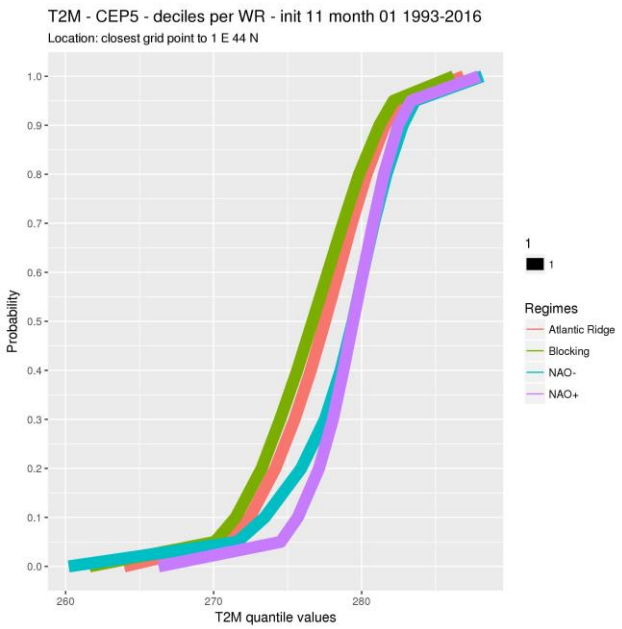


Figure 31. Deciles of daily mean Temperature in January 1993–2016 in SEAS5, start date of November according to the 4 weather regimes on the grid point closest to Toulouse.

At the same time, we are aware that the integration of weather regimes adds a level of uncertainty to the correction process: how their frequencies are reproduced in climate models might have an impact on the results. We expect the advantage of the method to be more evident when the GCMs have a better skill in forecasting weather regimes

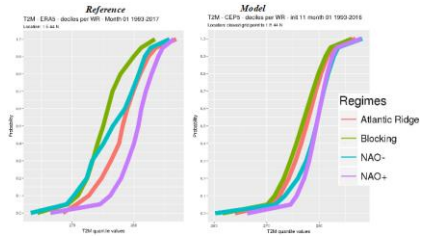
To integrate WRs, in the ADAMONT training phase, the quantiles of both model data and reference data (observations or reanalysis) are separately calculated on groups of days to which the same weather regime is attributed (e.g.: for the classical 4 weather regimes, we will obtain 4 sets of quantiles for one parameter of interest). More in particular, ADAMONT treats

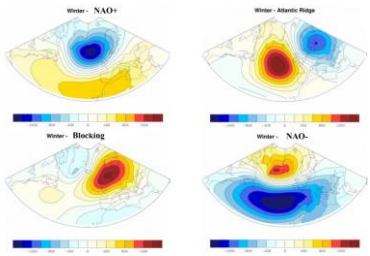
a month at a time. For example, if our aim is the correction of temperature data of all the Januaries covering the period 1993–2016, ADAMONT groups the 744 days composing that time series in the reference dataset into 4 clusters according to their correspondent observed weather regimes. Within each cluster of days, ADAMONT then identifies the quantiles. The same procedure is applied on model data merging together all the runs of the ensemble, thus producing a time series of 744×25 days to be split into 4 clusters.

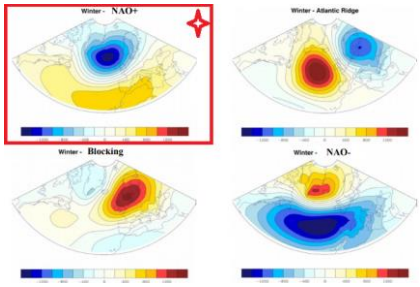
In the application phase, ADAMONT applies the quantile-quantile correction to day *t* according to the WR attributed to that day

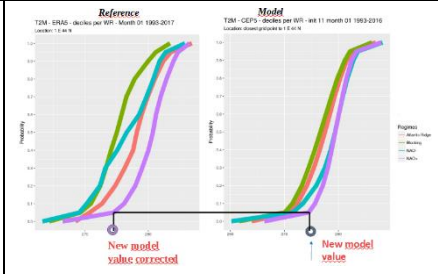
The following schematic (Table 1) traces the steps of the correction. As an example, the figure reports the correction of the forecasted temperature in Toulouse (closest point to 1°W 44°N) on a generic date *t*.

Table 1. Schematic representing the main steps of the correction implemented by ADAMONT. The exemple is focused on correcting forecasted Temperature data in Toulouse on a generic date *t*.

<table>
<tr>
<td>
<table>
<tr><th>NAO+</th><th>NAO-</th><th>Atl. Ridge</th><th>Blocking</th></tr>
<tr><td>Q 0.95</td><td>Q 0.95</td><td>Q 0.95</td><td>Q 0.95</td></tr>
<tr><td>Q 0.9</td><td>Q 0.9</td><td>Q 0.9</td><td>Q 0.9</td></tr>
<tr><td>...</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>Q 0.1</td><td>Q 0.1</td><td>Q 0.1</td><td>Q 0.1</td></tr>
<tr><td>Q 0.05</td><td>Q 0.05</td><td>Q 0.05</td><td>Q 0.05</td></tr>
</table>
1) Computation of quantiles on the reference dataset on the point of interest (e.g. 1°W 44°N, Toulouse)
</td>
<td>
<table>
<tr><th>NAO+</th><th>NAO-</th><th>Atl. Ridge</th><th>Blocking</th></tr>
<tr><td>Q 0.95</td><td>Q 0.95</td><td>Q 0.95</td><td>Q 0.95</td></tr>
<tr><td>Q 0.9</td><td>Q 0.9</td><td>Q 0.9</td><td>Q 0.9</td></tr>
<tr><td>...</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>Q 0.1</td><td>Q 0.1</td><td>Q 0.1</td><td>Q 0.1</td></tr>
<tr><td>Q 0.05</td><td>Q 0.05</td><td>Q 0.05</td><td>Q 0.05</td></tr>
</table>
2) Computation of quantiles on model data on the closest grid point to the point on interest
</td>
<td>


We obtain two approximated CDF to use in the correction:
</td>
</tr>
<tr>
<td>


3) Computation of the distance of the forecasted MSLP anomaly field on day *t* to each of the four Weather Regimes centroid.
</td>
<td>


4) Attribution of the MSLP field on day *t* to the closest Weather Regime centroid (e.g. NAO+)
</td>
<td>


5) The model value is corrected applying the correspondence between the quantiles of the reference and the quantiles of the model data under the attributed regime (e.g. NAO+)
</td>
</tr>
</table>

The geometry of the reference dataset (here: ERA5 0.25° regular grid) becomes the output geometry of ADAMONT. To each output point, ADAMONT attributes the model value from the closest model grid point.

A possible drawback of introducing WR is the reduction of sizes of samples used to calibrate distributions associated with each WR. A particular attention is necessary, in order to avoid too small sample, so to limit uncertainties in the estimation of lowest and highest quantiles. Simple experiments coded in R (see section 4.7, Annex: Choice of quantiles in the quantile-mapping correction) show that in our case, we should better stop at quantiles Q5 and Q95.

4.4 Assessment of ADAMONT skill

4.4.1 Assessment of Method by a "perfect forecast" experiment

In order to measure the skill of ADAMONT, and especially the added-value of using WR in the application of quantile-mapping correction, we have applied the method using "observed" data as input (see Figure 32). This is achieved by replacing daily WR forecasts by "observed" WR analysed by ERA5, and parameter fields (daily 2-metre temperature, precipitation and 10-metre wind-speed) by ERA5 fields. As our verification fields are ERA5 fields at full horizontal resolution (0.25°), the input parameter fields are preliminary interpolated at 2° resolution.

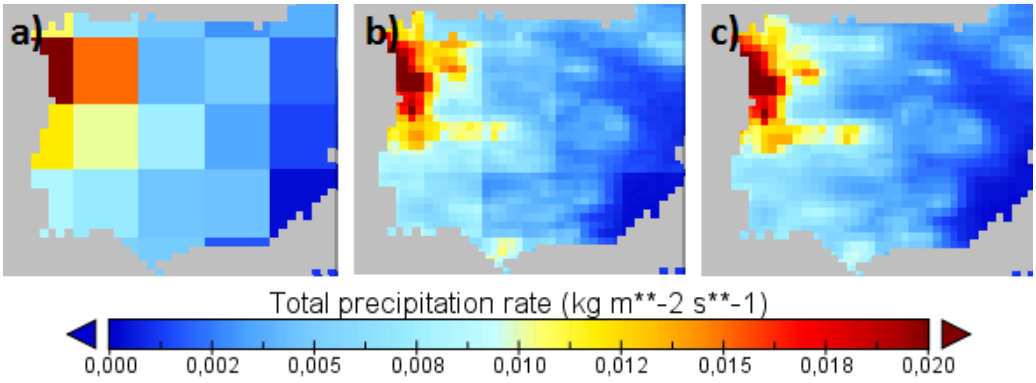


Figure 32. Example of a "perfect forecast" of Dec. 2000 monthly precipitation, zoomed in on the Iberian Peninsula. ERA5 2° resolution field (a) is corrected by ADAMONT. The output (b) can be evaluated in comparison to the ERA5 field at 0.25° resolution (c).

The skill of the method is measured on each inland grid point of the target domain (more than 7000 points over Europe at 0.25° horizontal resolution). We focussed on monthly means (the timescale of the data expected for SECLI-FIRM case studies) of bias and interannual correlation: "bias" because we expect the method to adapt large scale information to local climate, so as to produce unbiased outputs; and "interannual correlation" because a perfect forecast should reproduce the actual interannual variability, so being close to 1. Moreover, by running ADAMONT with and without WR, we aim to highlight the benefit of conditioning the quantile-mapping correction to the large scale circulation context.

4.4.1.1 *Skill of ADAMONT using WR*

Monthly wind speed and temperature biases are almost zero, whereas for precipitation a very small number of grid points show biases of up to 25% (see Figure 33).

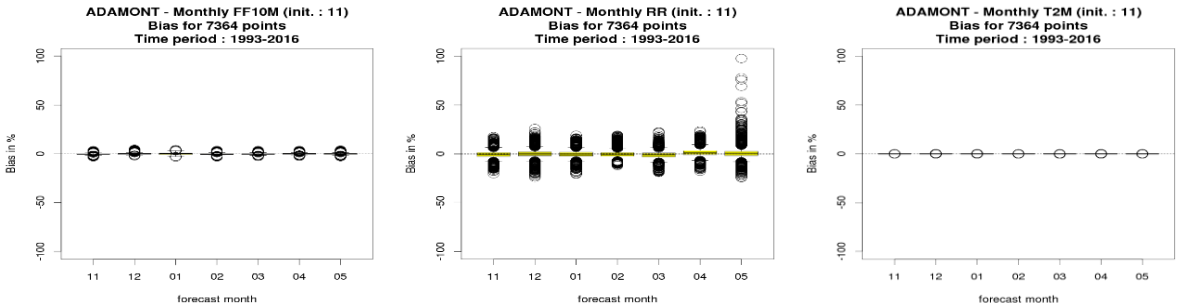


Figure 33. Boxplots of bias (in percent) of monthly wind speed (left), precipitation (middle) and temperature (right). Black circles represent outliers (points which are more/less than 1.5 times the interquartile range from the box).

Looking at the location of these outliers for precipitation (Figure 34, where errors above 20% in absolute value are plotted in dark blue and dark red), it appears that they are located is the southern part of the domain: in (or close to) mountainous and coastal regions. They are mainly regrouped in 2° squares, corresponding to the 2° input grid: this is probably linked to the fact that for each target grid point, ADAMONT makes use of the closest grid point of the input field.

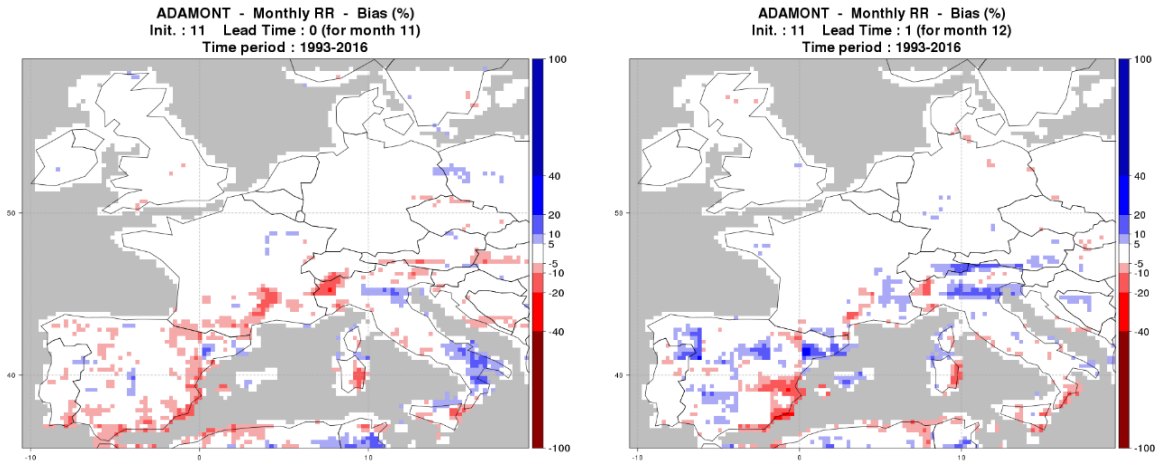


Figure 34. Bias (in percent) for precipitation at lead-time 0 and 1 month (November and December respectively for November 1st start date).

Concerning interannual correlation, the skill is excellent (close to 1) for a large majority of points for temperature, and very good for wind speed and precipitation (Figure 35).

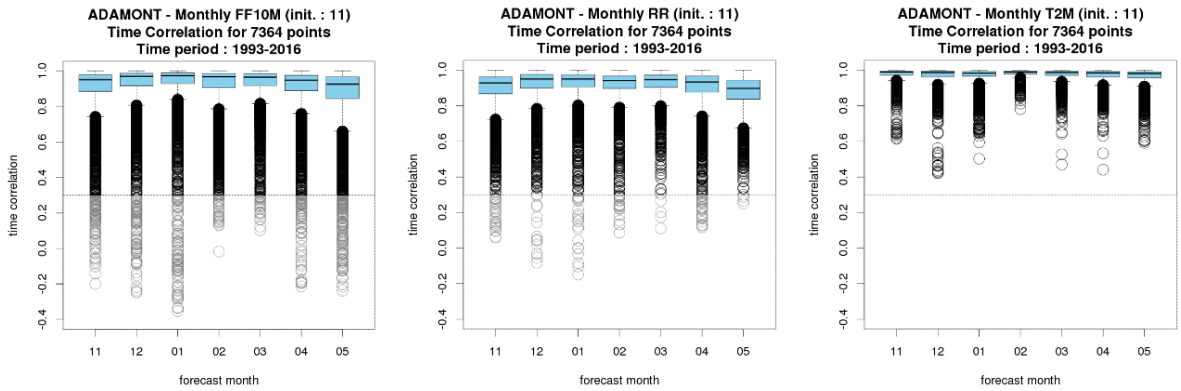


Figure 35. Boxplots of interannual correlation of monthly wind speed (left), precipitation (middle) and temperature (right). Black circles represent outliers (points which are more/less than 1.5 times the interquartile range from the box).

As with bias, points with the lowest skill correspond to mountainous or coastal regions, mainly in the southern part of the domain (Figure 36). As already noticed for bias, patterns corresponding to 2° squares are visible (north-west of Italy for instance).

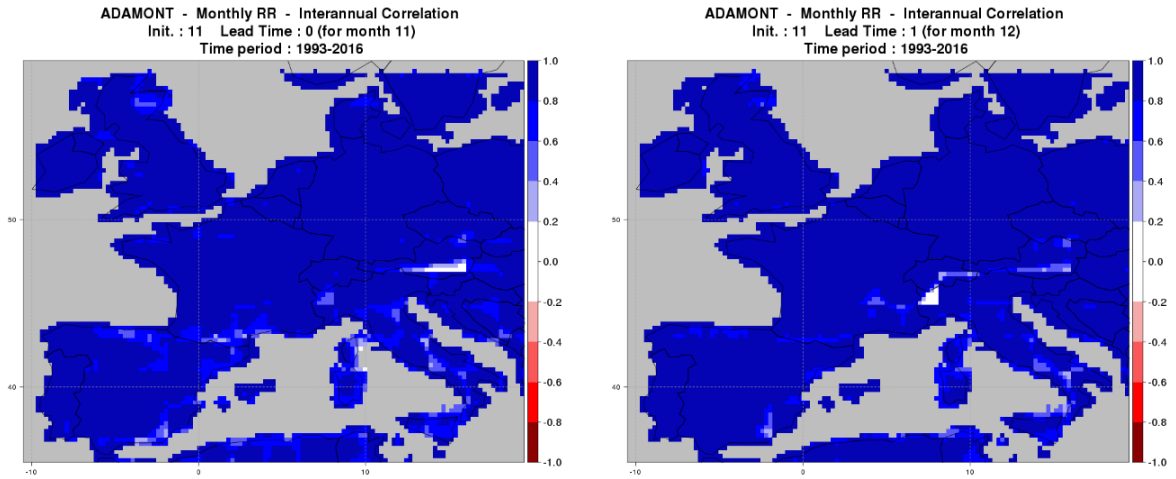


Figure 36. Interannual correlation for precipitation at lead-time 0 and 1 month (November and December respectively for November 1st start date)

The method is achieving good results in most parts of the domain, for the three parameters. Some defects have been identified in mountainous and coastal areas, where we have also

noticed square prints corresponding to input data grid-points. In section 4.4.1.3, we will test an interpolation before applying ADAMONT, in order to try to reduce these artefacts. But first of all, we will check the impact of WR in the method.

4.4.1.2 ADAMONT using/not using WR

A parallel experiment has been run without using WR. It means that we have applied a classical quantile-mapping correction that doesn't depend on large scale circulation information brought by WR.

Looking at monthly score charts, the two outputs exhibit very similar results to the previous one. The skill of the "no WR" experiment is slightly lower than the "with WR" one, in terms of bias and in terms of interannual correlation, and for the three variables. For example for precipitation (Figure 37), there are more points with correlation lower than 0.9 for the "no WR" experiment, and these points are mainly located over mountains (Alps, Pyrenees, Apennines), where the "with WR" experiment has slightly better skill. One can notice that for both experiments, there are 2° square patterns (see discussion of Figure 36).

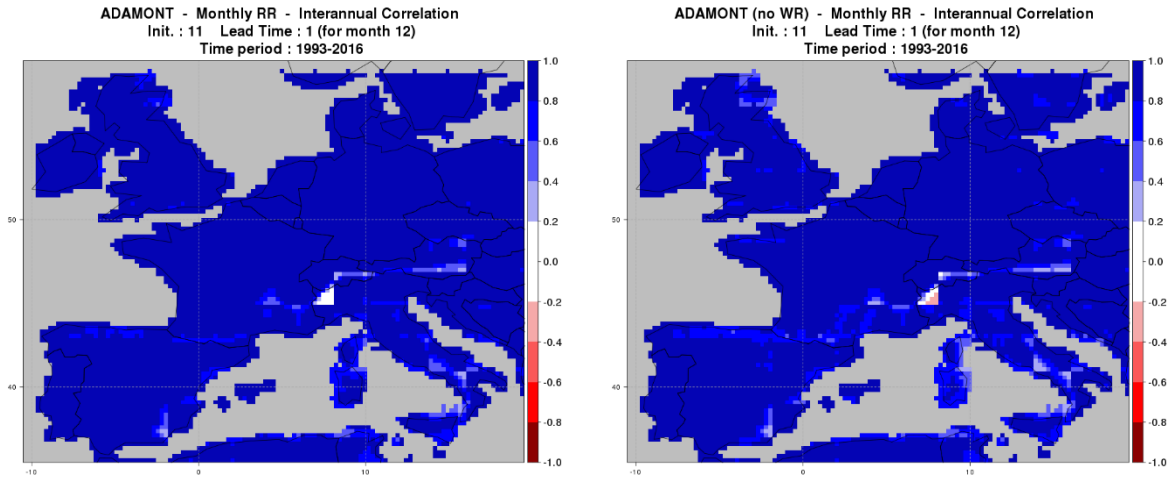


Figure 37. Interannual correlation for precipitation at 1 month lead-time for the experiment with WR (left) and without WR (right)

We are tempted to conclude that using WR doesn't bring additional skill. In fact, the correction "without WR" already provides very good results in most of the grid points: indeed, when correlation is greater than 0.9, there is little room for improvement. To refine the description of WR impact in the method, we have isolated a sub-domain over the Alps (44°N – 48°N, 5°E – 10° E) where the skill is globally lower than over lowland areas, and potentially WR could help to prove its potential. Computing a significance test on the difference of interannual correlations (Williams' test for dependant samples, Williams 1959), it appears that the introduction of WR is significantly beneficial, for example in December (Figure 38). On average, for winter months

(Table 2), there is an improvement on 15% of grid-points for wind speed (mainly over mountains and coastal areas), 7% of grid-points for precipitation (mainly over lowland) and 10% of for temperature.

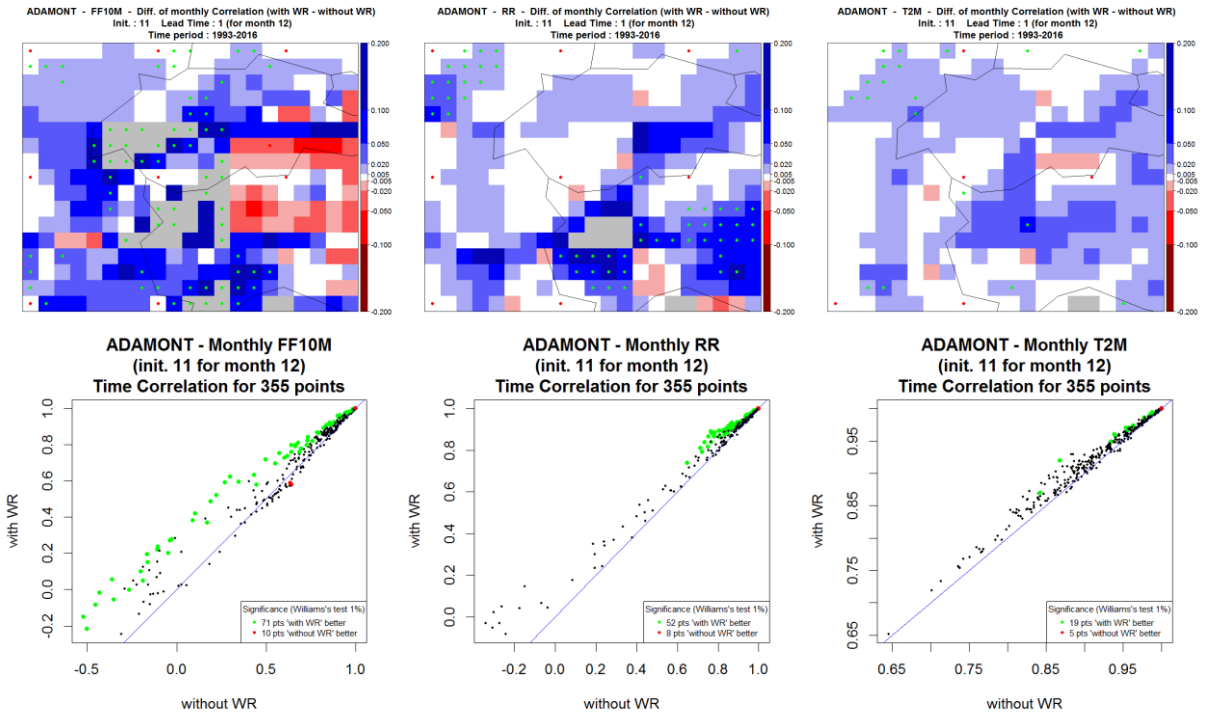


Figure 38. Comparison of interannual correlation for wind speed, precipitation and temperature at 1 month lead time, for experiments with/without WR. Green dots correspond to grid points where the "with WR" experiment is significantly better than the "without WR" experiment; red dots show where "with WR" experiment is significantly worse; black dots show where there are no significant differences.

Table 2. Percentages of grid points with significant improvement of interannual correlation, calculated on monthly mean forecasts. Values are shown for each forecast month and the mean of the four months.

	Wind speed	Precipitation	Temperature
November	15%	3%	16%
December	20%	15%	5%
January	16%	5%	10%
February	11%	6%	8%
Mean	**15 %**	**7 %**	**10 %**

4.4.1.3 *Impact on using WR on daily skill*

Even if the majority of SECLI-FIRM applications require monthly corrected data as input, ADAMONT's daily output can be useful in the construction of monthly indices based on daily (or sub-monthly) data: for instance an index of the number of days below/above a given threshold. In that respect, we calculated some skill metrics on daily corrected values, for both experiments (with/without using WR). The analysis on daily data is carried out in two dimensions: time and space.

4.4.1.3.1 Analysis in time

The analysis along the time dimension focussed on skill assessment in terms of accuracy and variability. The first is assessed by RMSE (root mean square error), the second by SD (standard deviation). RMSE measures the total magnitude of the discrepancy between corrected data and observations regardless of their sign. In contrast to bias, RSME is not a centred metric, and two equal errors but with opposite sign will not cancel out.

Figure 39 compares RMSE calculated on the corrected data with respect to ERA5 reanalysis for the three parameters, wind speed, precipitation and temperature. The score is computed over 744 days spanning Decembers 1993–2016. Each point in each panel corresponds to one grid point of the European domain (7364 grid points) and shows simultaneously the RMSE found in the experiment with WR (y-axes) and without WR (x-axes). For points below the diagonal, the skill with WR is higher than the skill without WR. Colours indicate the significance of the discrepancy of the two RMSE assessed by the Diebold-Mariano test (Diebold and Mariano, 1995) with a threshold of 0.01.

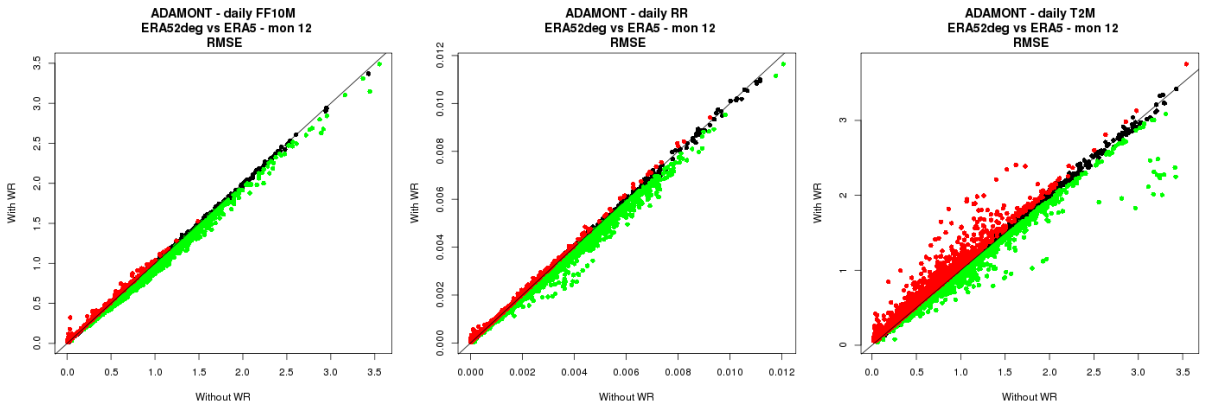


Figure 39. RMSE from the experiments with and without WR for wind speed, precipitation and temperature. RMSE is calculated on 744 days, all the December months 1993–2016. Each point is a grid point of the European domain. Green: skill is significantly better with WR. Red: skill is significantly worse with WR. Significance assessed with a Diebold–Mariano test (0.01 threshold)

Table 3 reports the percentage of points where the skill using WR is significantly better or worse for months from November to February.

Table 3. Percentage of points where the skill using WR is significantly better or worse for the months from November to February for wind speed, precipitation and temperature.

	Wind Speed		Precipitation		Temperature	
November	**16.5**	2.6	**25.7**	5.0	8.7	**11.1**
December	**23.3**	2.7	**30.6**	4.7	11.7	**15.9**
January	**23.4**	3.8	**22.4**	7.5	11.9	**19.5**
February	**16.8**	3.7	**19.4**	8.2	8.7	**15.0**

Overall, the experiment with WR outperforms the one without WR. For wind-speed and Precipitation, not only the number of points where RSME is smaller are greatly the majority, but also the points when, in the opposite, the RMSE is worse, the value of this latter is always very small, thus not impacting too much the overall results. More controversial is the interpretation for temperature, where the number of points where the performance with WR is worse are the majority. More broadly we could expect that WR might not be beneficial in delineating spatial patterns of smooth variables such as temperature.

The second analysis deals with temporal variability of corrected data. For example Figure 40 reports a comparison of the two experiments by ratios of standard deviations of daily precipitation, temperature and wind speed in Decembers (first row) and Januarys (second row) in 1993–2016, for a total of 744 days. Boxplots are created from SD computed on 355 spatial points over the Alps (the domain used in the analysis of Figure 38). The closer the ratio is to one, the better daily variability is reproduced.

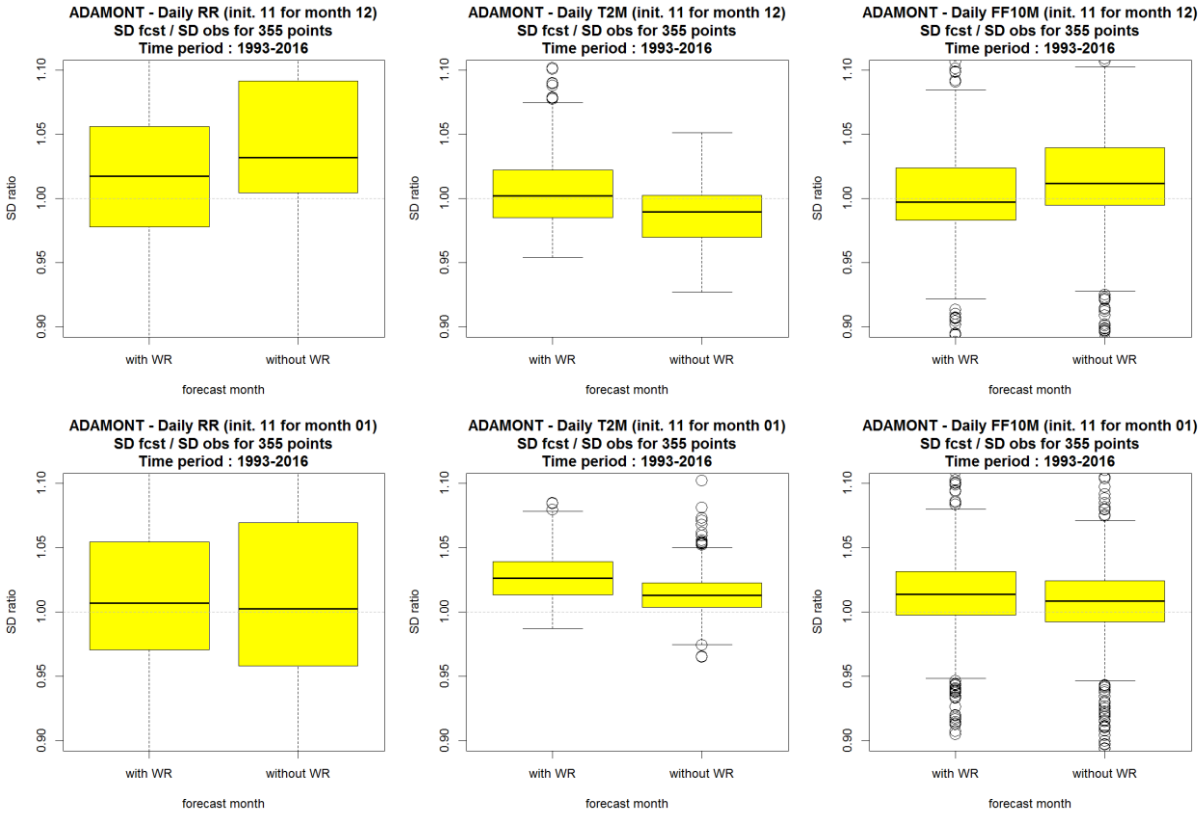


Figure 40. Ratio of standard deviation of daily corrected data by ADAMONT in the experiment with WR and without WR. 744 days spanning Decembers (first row) and Januaries (second row) in 1993–2016. Boxplots are created from the SD computed on 355 spatial points over the Alps. First column: precipitation; second column: temperature; third column: wind-speed.

For precipitation, ratios are more accurate in the experiment with WR in December, but less accurate in January. For the three parameters and considering November to February months (not shown), neither of the two experiments surpasses the other. As SD errors rarely exceed 10%, ADAMONT using or not WR can be considered as able to correctly forecast daily variability.

4.4.1.3.2 Analysis in space

As a first analysis, on the wide European domain (as in Figure 34), we explore whether the inclusion of WR in ADAMONT improves spatial coherence of the results. It is assessed by spatial correlation between daily corrected data in each experiment (with and without WR) and ERA5.

Figure 41 compares the spatial correlation of wind speed, precipitation and temperature on daily data in all December days of the period 1993–2016. In each plot, each point corresponds to one date (744 dates are plotted) and shows simultaneously spatial correlation for that date found on corrected data with WR (y-axes) and without WR (x-axes). Colours indicate when the discrepancy between the two results is significant: green dots when the experiment with WR is significantly better and red dots when it is significantly worse. The significance is assessed by the Williams test (Williams, 1959).

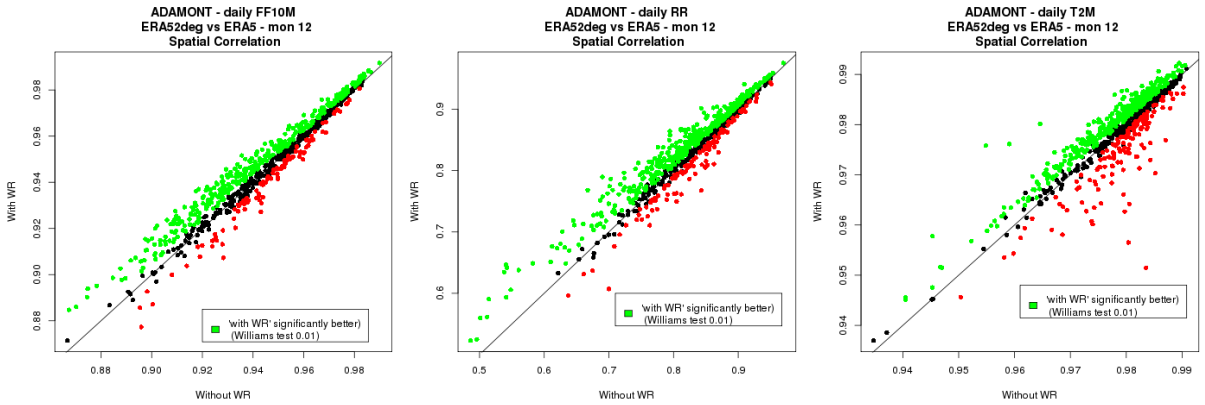


Figure 41. Comparison of spatial correlation for wind speed, precipitation and temperature for December, for experiments with/withour WR. The correlation is computed on 7364 inland points for each of the 744 days spanning Decembers 1993–2016. Green dots correspond to dates when the “with WR” experiment is significantly better than the “without WR” experiment; red dots correspond to dates where the “without WR” experiment is significantly better. Black dots indicate no significant differences. Significance is assessed by the Williams test.

Table 4. Percentages of dates when the spatial correlation is significanlty better or worse in the experiment with WR with respect to the experiment without WR, for wind speed, precipitation and temperature. This table completes Figure 41.

	Wind Speed		Precipitation		Temperature	
November	**43.2**	13.5	**50.4**	15.3	**39.3**	18.8
December	**48.3**	13.6	**52.4**	14.4	**44.6**	19.2
January	**50.7**	13.7	**53.1**	15.5	**41.9**	21.6
February	**47.2**	15.3	**51.5**	12.1	**31.0**	15.9

Complementary to correlation, we calculated RMSE in space, in order to evaluate the magnitude of the spatial coherence discrepancy. Figure 42 compares the RMSE from the experiments with and without WR calculated on 7364 spatial points along 744 days. Thus, they are the time series of the difference between spatial RMSE with WR and spatial RMSE without WR. Negative values indicate better skill with WR, colours indicate discrepancy significance of the two RMSE assessed by the Diebold–Mariano test (Diebold and Mariano, 1995) with 0.01 threshold. Table 5 complements Figure 42 with the exact number of green and red points.

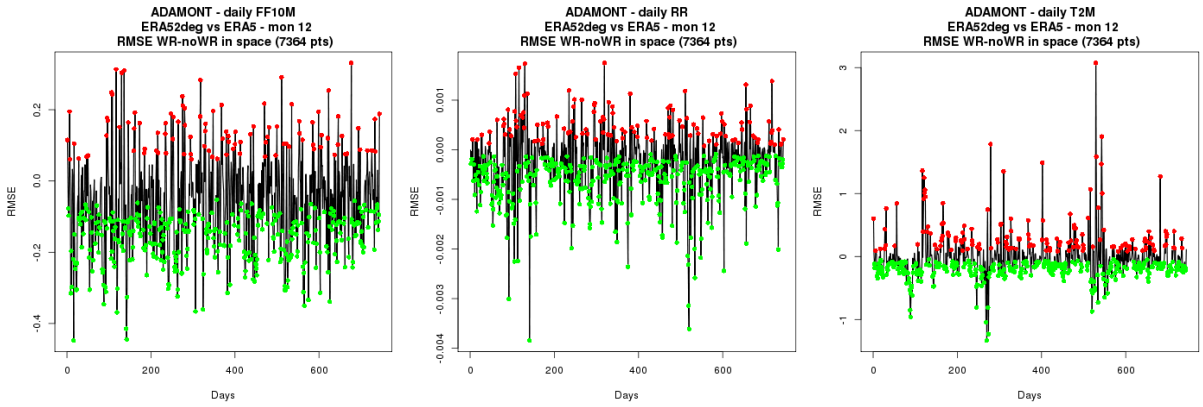


Figure 42. Time series of the Difference RMSE with WR minus RMSE without WR in daily corrected data for Wind-speed, Precipitation and Temperature. Each point is one of the 744 days spanning Decembers 1993–2016. RMSE is computed over 7364 spatial points. Green dots: skill with WR is significanlty better; Red dots: skill with WR is significantly worse. Significance assessed by Diebold–Mariano test at a threshold of 0.01.

Table 5. Percentages of dates when spatial coherence assessed by the RMSE is significantly better and worse in the experiment with WR with respect to the experiment without WR, for Wind Speed, Precipitation and Temperature. This table completes Figure 42.

	Wind Speed		Precipitation		Temperature	
November	**42.1**	14.7	**51.8**	16.7	**39.0**	18.5
December	**48.1**	13.4	**52.0**	16.9	**44.0**	20.2
January	**50.8**	13.8	**50.7**	18.5	**40.9**	22.6
February	**44.1**	16.5	**42.9**	17.0	**31.3**	16.5

Looking at Table 4 and Table 5, using WR in ADAMONT significantly improves spatial correlation and reduces errors for the three parameters at the daily timescale. Again, the benefit is lower for temperature than for wind speed or precipitation at the daily time scale.

4.4.1.4 Discussion

These “perfect forecast” experiments show the ability of ADAMONT to apply an efficient local adaptation of coarse resolution input data. The skill comes mainly from the quantile-mapping technique coded in ADAMONT. Slight (but significant) additional skill is obtained by using a WR constraint (i.e. quantile-mapping conditioned on WR), specifically over mountainous regions (or regions with complex orography) for wind speed and precipitation.

These experiments also highlight a defect of ADAMONT, visible on score charts as regular patterns corresponding to the input fields grid. Indeed, for each target grid point, ADAMONT applies a correction on the closest grid point of the input field. So it creates artificial discontinuities on output fields, also visible on score charts. This problem can be circumvented by interpolating input data, prior to running ADAMONT. In addition, one can obtain smoother daily outputs (otherwise daily outputs can be extremely spatially polluted, see Figure 43). Anyway, it seems that such a pre-treatment could weaken monthly output quality (see Figure 44 where the maximum in the Northwest of Spain is not high enough in the middle chart). So the choice of this option would require further investigation, and could be more or less beneficial depending of the use of the outputs. For instance, the user should address questions like their need of daily or monthly data, the importance of local or regional fields, etc.

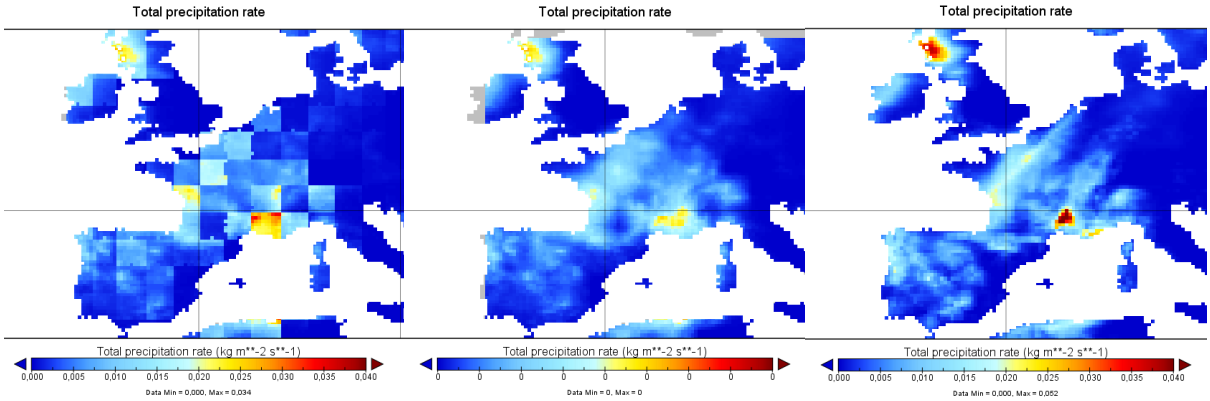


Figure 43. 2 Dec. 2000 precipitation. Example of ADAMONT applied on 2° input data (left), on 2° data bilinearly interpolated at 0.25° data (middle) and verification ERA5 field (right).

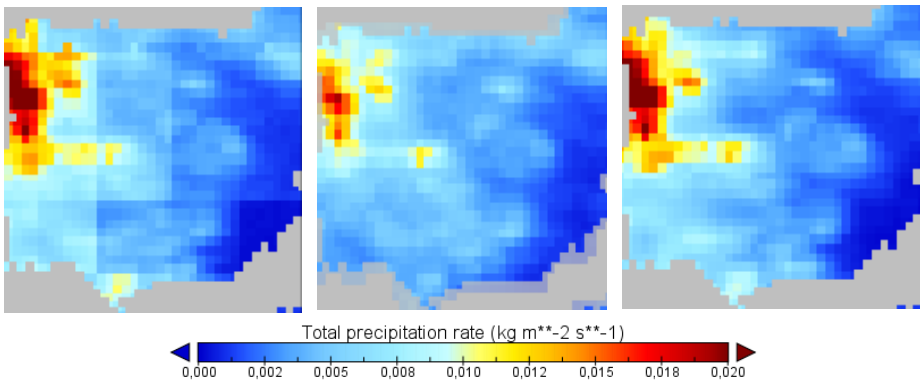


Figure 44. Dec. 2000 monthly precipitation. Example of ADAMONT applied on 2° input data (left), on 2° data bilineary interpolated at 0.25° data (middle) and verification ERA5 field (right).

4.4.2 Skill in forecasting the Weather Regimes

It seems relevant at this point to assess the skill in forecasting the Weather Regimes. Indeed, we are aware that their inclusion in the correction process adds a level of uncertainty, and we thus expect that how well their observed frequency is forecasted in the climate model will have an impact on the ADAMONT skill.

We have used WR computing from ERA5[8] as the reference, both for centroids (spatial patterns) and daily timeseries (see section 4.3.2). Each daily MSLP anomaly field from each member of a seasonal forecast model are subsequently attributed to the closest of the ERA5 WR centroid. Applying this attribution to the model hindcast, we can calculate scores to assess WR predictability. Because seasonal forecast models are not able to reproduce chronology, scores are calculated on WR frequencies within a 3-month period. For example in Figure 45, one can visualize SEAS5 forecasts for DJF (start date November) over its hindcast, for the NAO+ WR, and its verification.

[8] Some of the scores presented in this document were already available at Météo-France, based on regimes computed with ERA-Interim. As ERA-Interim regimes are very similar to ERA5 regimes, both for their centroids and for their timeseries, we have not recalculated scores and consider them to be still relevant for this study.

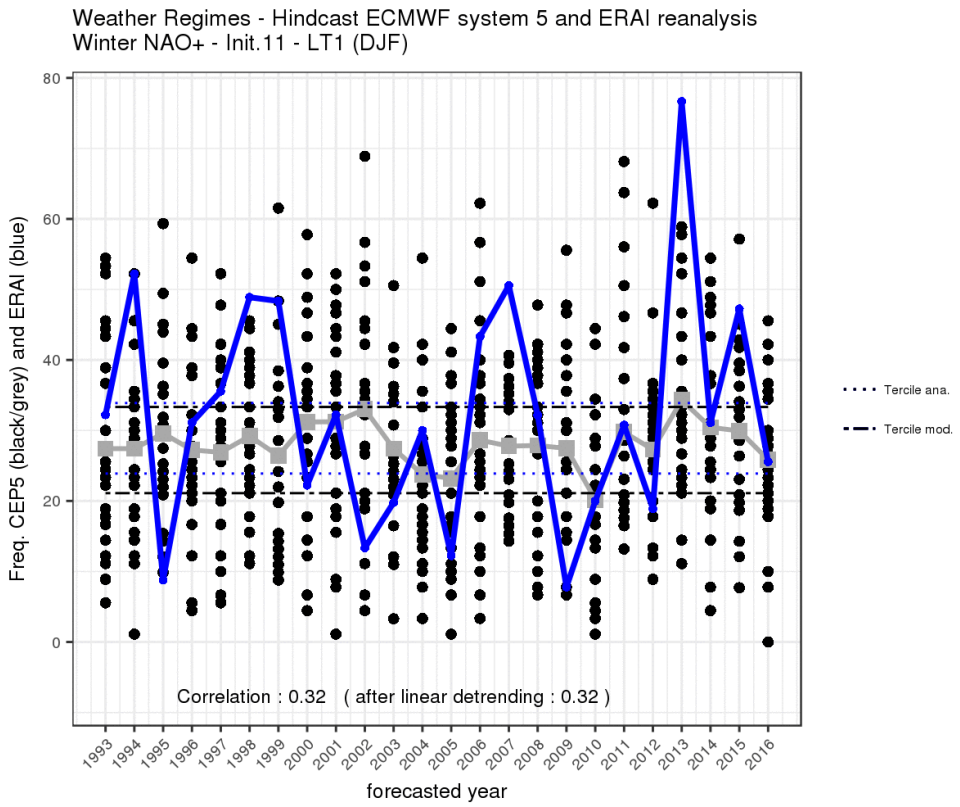


Figure 45. DJF forecasts of the NAO+ WR by SEAS5. The x-axis is hindcast years, the y-axis is regime frequency. Black dots are the frequency of each member. The grey line is the ensemble-mean frequency. The blue line is the observed frequency. Thick dashed lines are model terciles and thin dashed lines are observed terciles.

Two scores have been computed for each WR: firstly, interannual correlation between observed and forecasted frequencies (for the forecast: mean of frequencies of the ensemble); and secondly, ROC scores for terciles and for two categories (below/above average). This work has been repeated for the twelve start dates, for the different lead-times and synthesized in charts (see Figure 46). All these charts are visible on http://seasonal.meteo.fr/content/PS-scores-synthese-regimes for SEAS5 and for MF-S6 and MF-S7.

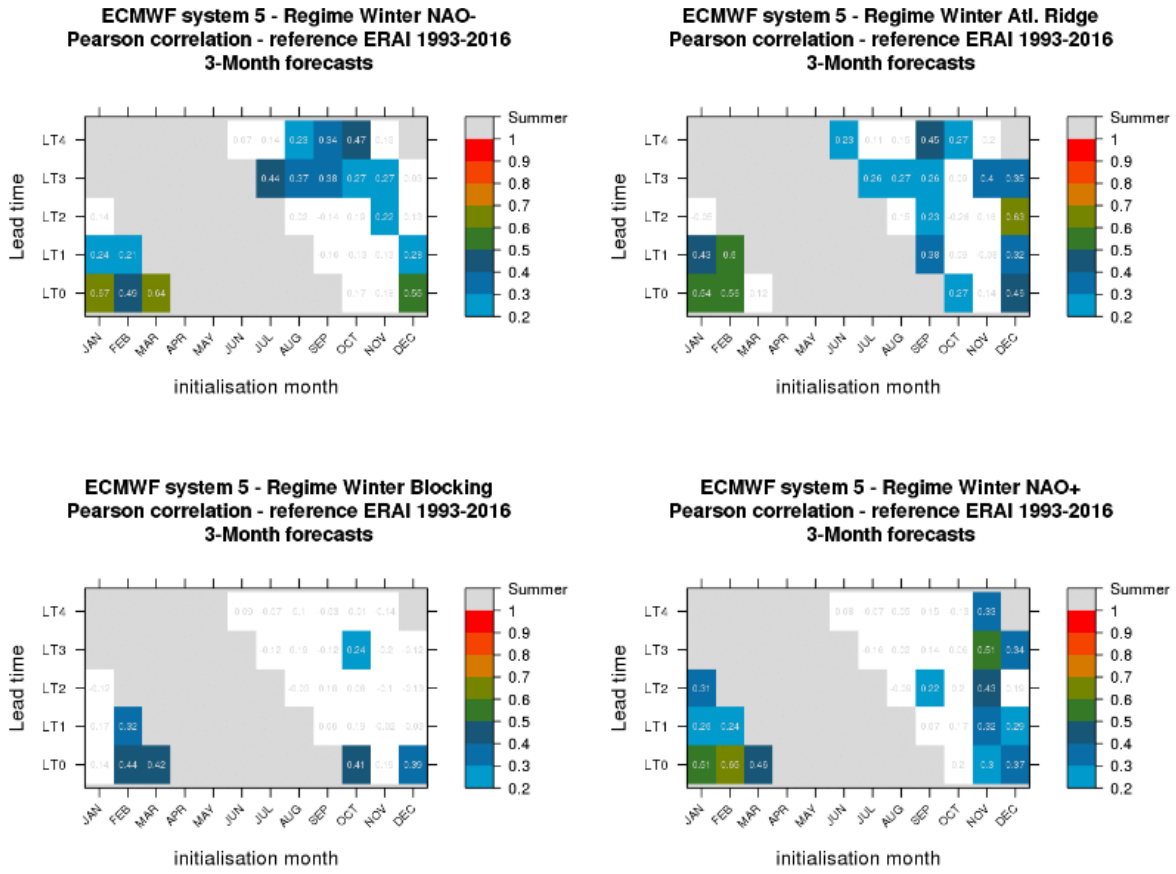


Figure 46. Synthesis of correlation scores on winter WR frequency prediction by SEAS5 versus ERA-Interim over the period 1993–2016. Each panel shows the frequencies of one WR.

What emerges from the analysis of scores for the 3 models (SEAS5, MF-S6 and MF-S7) is a poor predictability of WR frequencies in general, with differences between WR. The “Blocking” regime seems less predictable than the 3 others. For all the WR, interannual correlations are rarely higher than 0.4, which means a limited capacity to detect interannual variability when looking at the mean frequency (which is the aggregation of information from the whole ensemble). Examining AUC for 2 categories (below/above average) or 3 categories (terciles), scores are a little bit better: AUC higher than 0.5 means that SF brings predictability. For instance at lead-time 1 for SEAS5, AUC is higher than 0.5 three times out of four. The challenge is of course to be able to take advantage of this weak predictability, something we have to check in ADAMONT.

4.4.3 Skill of the ADAMONT method applied on SF data

We have applied ADAMONT to SEAS5 SF data over a European domain for Precipitation, 2-metre temperature and 10-metre wind speed, using SEAS5 forecasts of WR. We downscaled all the 7 lead-months of the start date of November, spanning the hindcast period 1993–2016. The calibration phase relies on daily ERA5 data over the same period, at 0.25° resolution. We obtained corrected daily fields at 0.25° resolution. The quantiles identified and used in the correction are (0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95). The refinement at the beginning and end of the distribution serves to refine the description of the tails, which is mostly beneficial for skewed variables as precipitation and wind. The 24 years covering the hindcast period are used for both calibration and verification.

For each parameter, each member of the ensemble and each start date, ADAMONT provides a netcdf file of daily gridded data merging the entire forecast period (7 months).

The skill of the method is assessed on running seasonal means (2-metre temperature and 10-metre wind speed) or seasonal accumulation (precipitation). Lead months are grouped as follows: NDJ, DJF, JFM, FMA. The scores considered are bias and interannual correlation, both computed on each inland grid-point of the target domain (7572 points). The skill scores RPSS and CRPSS were also computed, the globally lead to the same conclusions as interannual correlation. Bias assesses whether a residual discrepancy is still present on average between the observed and reproduced local climate after ADAMONT is applied. Interannual correlation, in this context, shows how the signal of the raw SF is maintained (or degraded, or improved) in the corrected data.

Figure 47 reports seasonal bias spread over the grid points for seasonal accumulated precipitation, seasonal mean 2-metre temperature, and seasonal mean 10-metre wind speed, for different lead-time. For temperature and wind-speed, seasonal biases are lower than a few percent everywhere. For precipitation the spread of biases reach ±20% for extreme outliers. Overall, the performance does not seem to decrease in farther lead seasons.

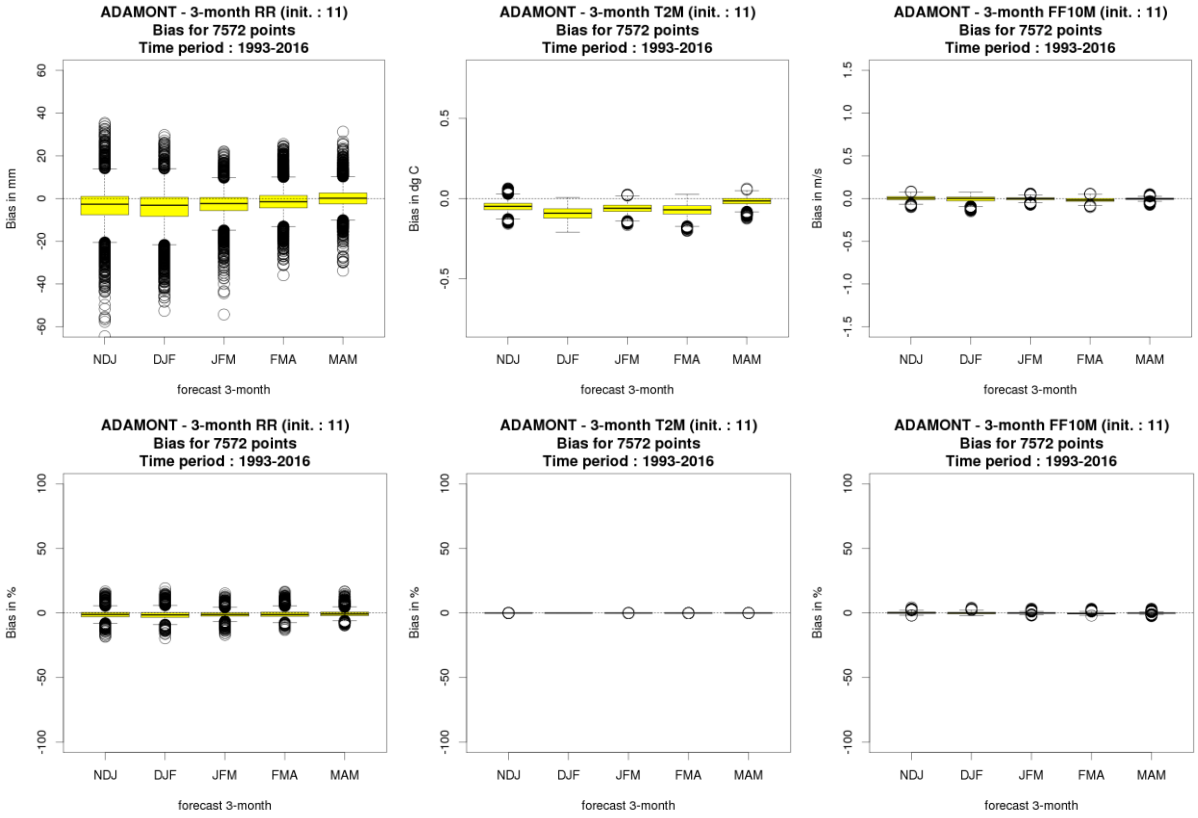


Figure 47. Box plots of seasonal bias (in native units and percentages, as labelled) over the 7572 inland points for the hindcast period (1993–2016) for seasonal accumulated precipitation, 2-metre temperature and10-metre wind speed. Start date: November. Each panel reports 5 lead seasons, thus from NDJ to MAM.

Bias distribution within geographic areas is plotted on charts in Figure 48. We focus on precipitation, the parameter with noticeable bias. The residual bias is region-dependent, spatial patterns are mainly constant across lead-seasons. Extensive areas have the same bias sign. It seems that overestimation mainly concentrates over mountainous areas, whereas there is an underestimation on most of lowland regions.

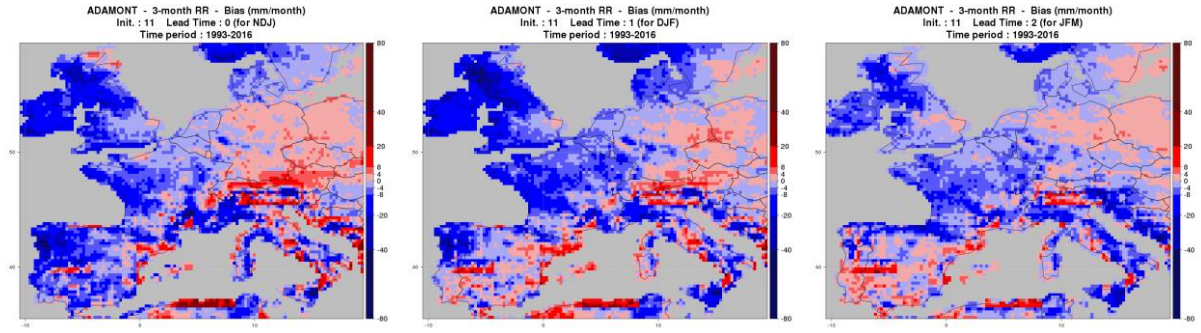


Figure 48. Residual bias in seasonal precipitation across Europe. SEAS5 corrected by ADAMONT is compared to ERA5. Left panel: NDJ (lead-season 0); middle panel: DJF (lead-season 1); right panel: JFM (lead-season 2).

Interannual correlation of corrected data (Figure 49) is rarely significant (the significance threshold is around 0.4 for this sample of 24 years) except at lead-time 0 (NDJ). Comparing to raw SF skill (Figure 50) one can find the same large-scale patterns

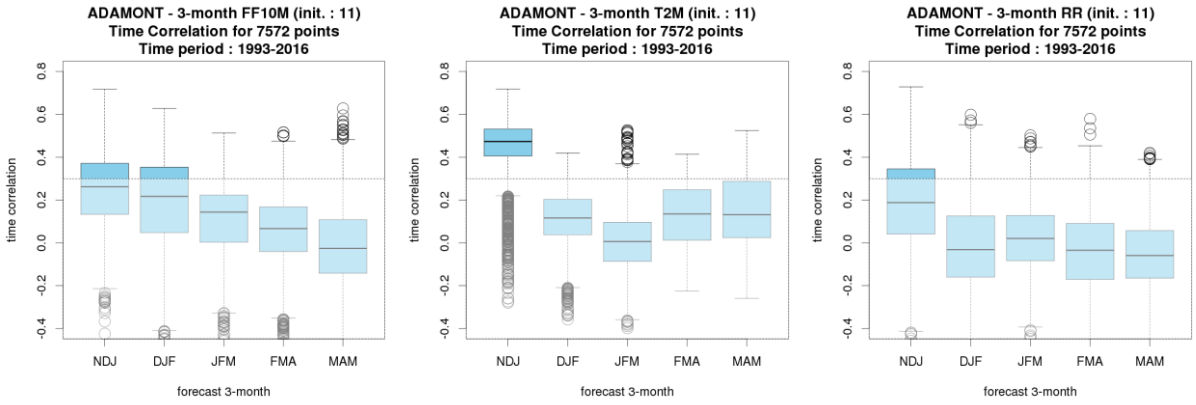


Figure 49. Boxplots of the interannual correlation computed on seasonal data, after the correction by ADAMONT. (1) Precipitation, (2) 2-metre temperature, (3) wind-speed. The masked (lighter) area is non-significant correlation. Start date: November. Each panel reports the 5 lead seasons, thus from NDJ to MAM.

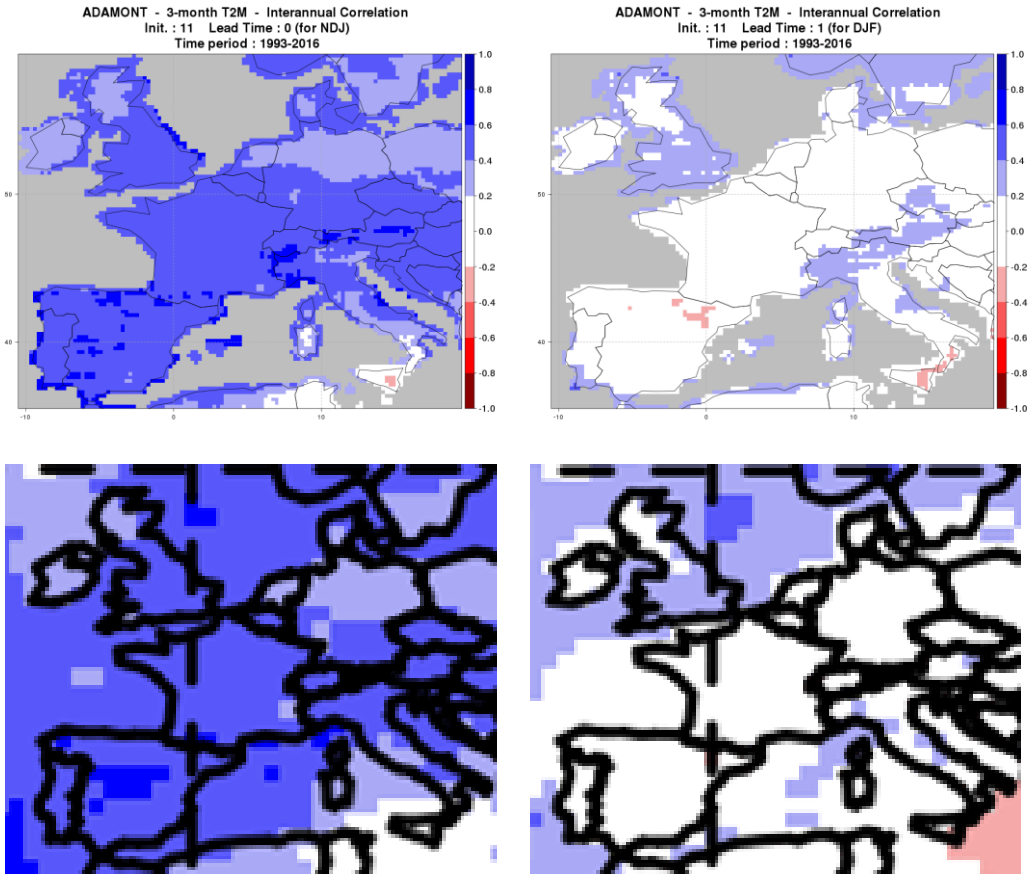


Figure 50. Interannual correlation of SEAS5 temperature, start date November, for NDJ and DJF. First line for after the correction by ADAMONT, second line for raw SF (extracted from http://seasonal.meteo.fr/content/PS-scores-cartes)

Although corrected forecasts generally have similar performance to raw forecasts in terms of interannual correlation, by construction they have a distribution consistent with the local climate. This is ensured by the quantile mapping technique, which provides corrected data following the distribution of the reference data. Thus it allows the provision of usable datasets for feeding impact models (hydrological models, energy consumption, etc). In particular, extremes are realistic and the main biases are removed. These properties are necessary for the data to be used as forcing to impact models.

Furthermore, by adding WR in the quantile-mapping process, we expect to add a better consistency between parameters. Indeed, for a specific member and a specific day, ADAMONT “imposes” a correction based on the same WR for all the parameters. So the quantile-mapping correction relies on distributions consistent with this very WR. The

advantage of this characteristic is unfortunately difficult to assess without calculating its consequences on an impact model.

4.5 Alternative approaches

Despite the reasonable intuitions that sustain the integration of WR in the correction, as explained in section 4.3.3, and measurable benefits in “ideal” conditions seen in section 4.4.1.2, the comparison of ADAMONT’s skill with and without WR show very similar results. For instance in terms of interannual correlation for wind-speed (Figure 51), where boxplots are identical; similar results are seen for the two other parameters.

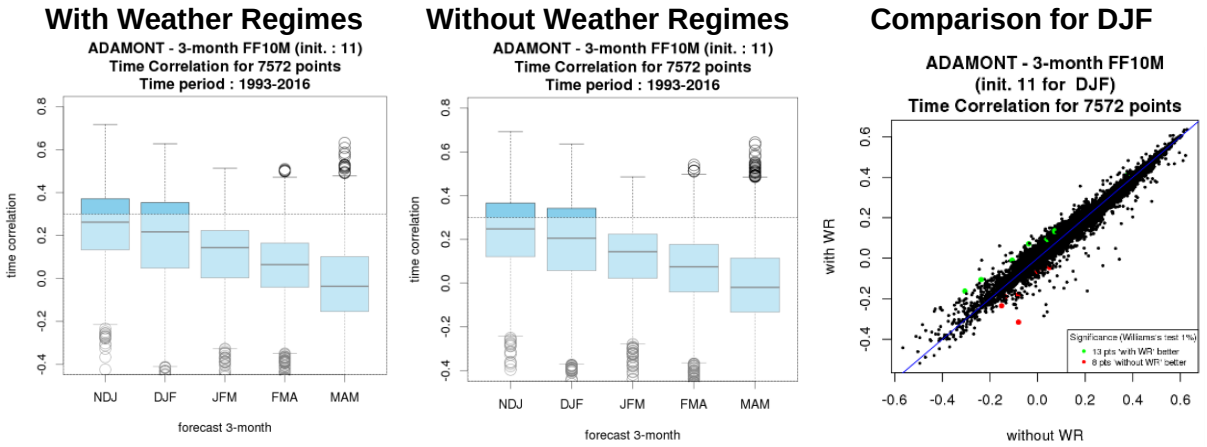


Figure 51. Box plots of interannual correlation for seasonal mean wind-speed, over 5 lead seasons, start-date November, model SEAS5 corrected by ADAMONT with (left) and without (middle) using WR. On the right, comparison of DJF. interannual correlation.

The poor predictability of WR frequency could explain these results. In order to try to “boost” ADAMONT, we will test alternative approaches. The first approach is an experiment where we select some members from the ensemble, using a criterion of “good predictability” of WR (knowing the WR effectively observed, thanks to ERA5). So we could measure if it results in a gain in predictability. In the second approach, we build a multi-model ensemble from several C3S models (SEAS5, MF-S7, CMCC and DWD), and we will check if the combination has better results than each model separately. In the third experiment, we combine the first two approaches: by a better knowledge of WR prediction (thanks to the multi-model ensemble), we select members and check if this selection can outperform a combination of all the members.

4.5.1 Member selection

In all the previous sections, we have applied ADAMONT on all ensemble members of the ECMWF-SEAS5 model hindcast (25 members). Now we would like to use only those members

that have good predictability of WR frequencies. We presume that if large scale circulation is correctly foreseen, ADAMONT should bring additional skill in the local forecasts of temperature, wind-speed and precipitation. Of course, for a “fair” comparison we have to measure the impact of WRs on the same sub-ensemble: meaning if we calculate scores of ADAMONT “with WR” with 3 members, it has to be compared to ADAMONT “without WR” using the same 3 members.

WR predictability can be measured by comparing the number of days of each regime in the forecast and in the reanalysis. Thus we calculate for each member and each season of each year the absolute error in terms of number of days. From this we propose two selection methods:

Selection 1: we impose a maximum value *M* for the absolute error. For instance in Figure 52, we keep only the members that have no more than 20 days of error within the season. Of course, this could lead to putting aside some years of the hindcast (we test several *M*), so in what follows we present results for *M* that lead to retaining at least one member per year.

Selection 2: we keep only the *N* best members of each year (we test several *N*)

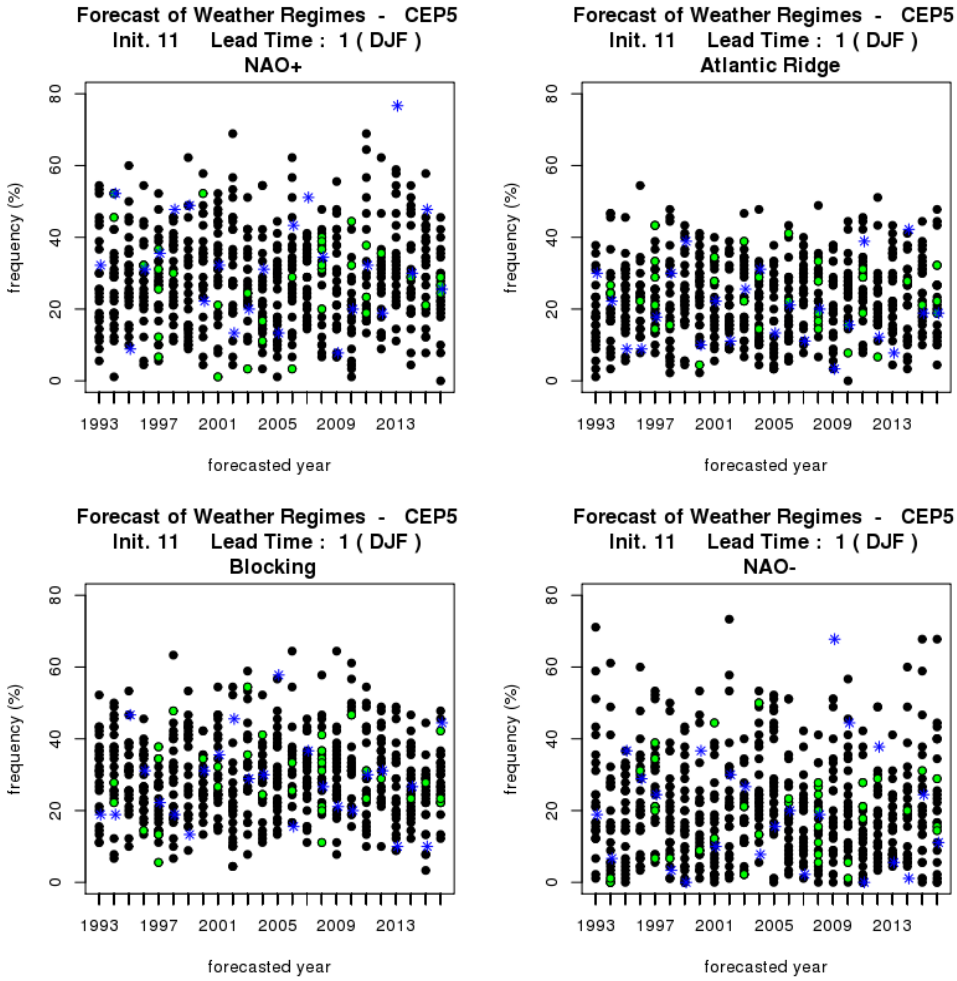


Figure 52. Example of selection. Each panel represents ECMWF-SEAS5 forecasts of one WR (black points) over its hindcast, the corresponding analysis from ERA5 (blue stars) and members selected for their "good" WR forecast (green points). For each year we keep only the members that have no more than 20 days of absolute errors. For a given year, this error is the sum (in absolute value) of the differences between the number of days foreseen and observed in each regime. In 1997, five members meet this requirement (green points); in 1993, there are none.

Because the target season is generally the 3-month period that follows the start date month, the absolute error is calculated on the 1st lead-time season of the forecast (i.e. DJF for November start-date). The chosen members are used for all the lead-times (i.e. NDJFMA for November start-date).

In applying selection 1 with different values of *M* (Figure 53), it appears that we can effectively improve interannual correlation scores of the three parameters by a proper selection of

members. The highest benefit is not obtained with a strong constraint (i.e. a small value of *M*), probably because the sub-ensemble is then too small: for instance, for *M* = 40 days, the sub-ensemble is as large as 2 to 18 members (25 for the full hindcast ensemble), depending on the hindcast year. The optimal value of *M* (taking into account the 3 parameters) is around 70 days (8 to 25 members, depending on forecast year), as the best compromise to have the largest number of improved grid-points (green bars and/or blue bars on plots) considering the 3 parameters.

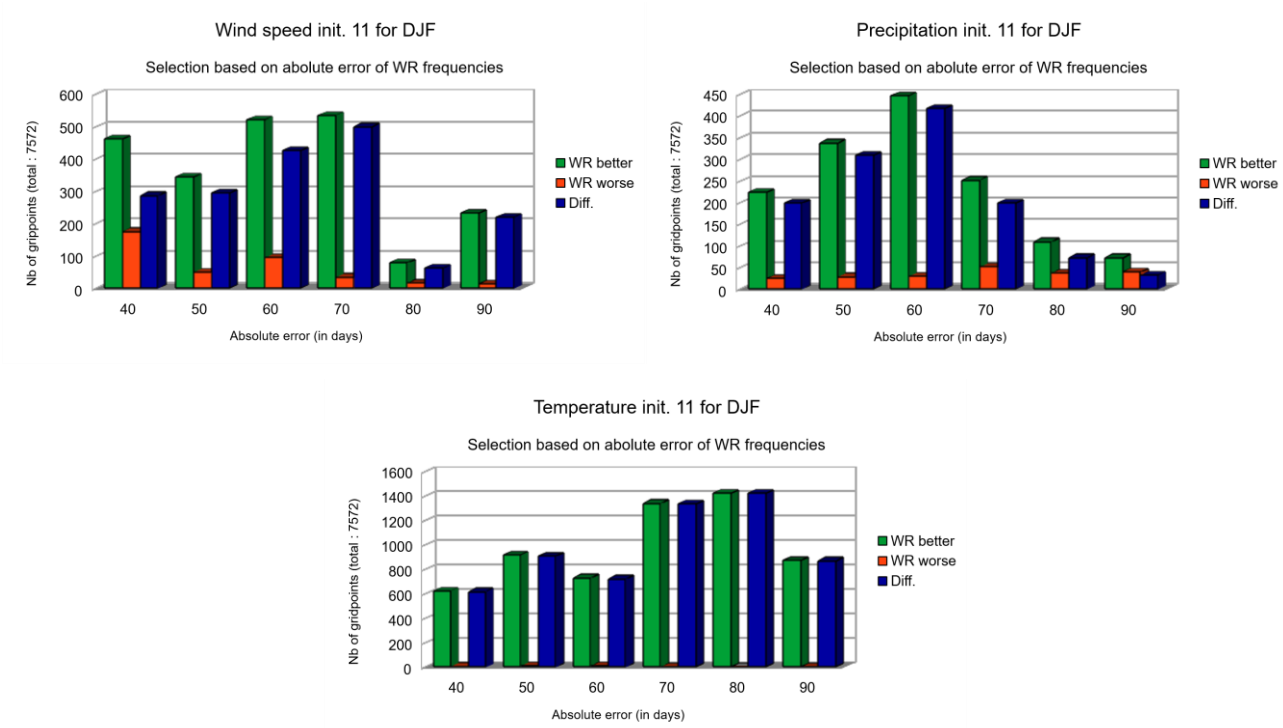


Figure 53. On the European domain (7572 grid-points), in green/red, number of grid-points where the skill of ADAMONT "with WR" is significantly better/worse than ADAMONT "without WR". Several experiments are presented, with a selection of members that have no more than *M* days of error in WR frequencies, with *M* varying from 40 days to 90 days.

The second selection, based on a pre-determined number of members (*N*), confirms that keeping the "best" members in terms of WR frequencies allows a "boost" in interannual correlation scores (see Figure 54). Interestingly, the optimal *N* is 15 members (to maximize grid-points with improvement), i.e. 60% of the ensemble size: the sub-ensemble size needs to be relatively large, as we can't count on a few members to capture the seasonal forecast signal. In Figure 55 one can visualise the increase in interannual correlation when selecting the 15 best members: it is clearly higher for wind speed and precipitation than for temperature.

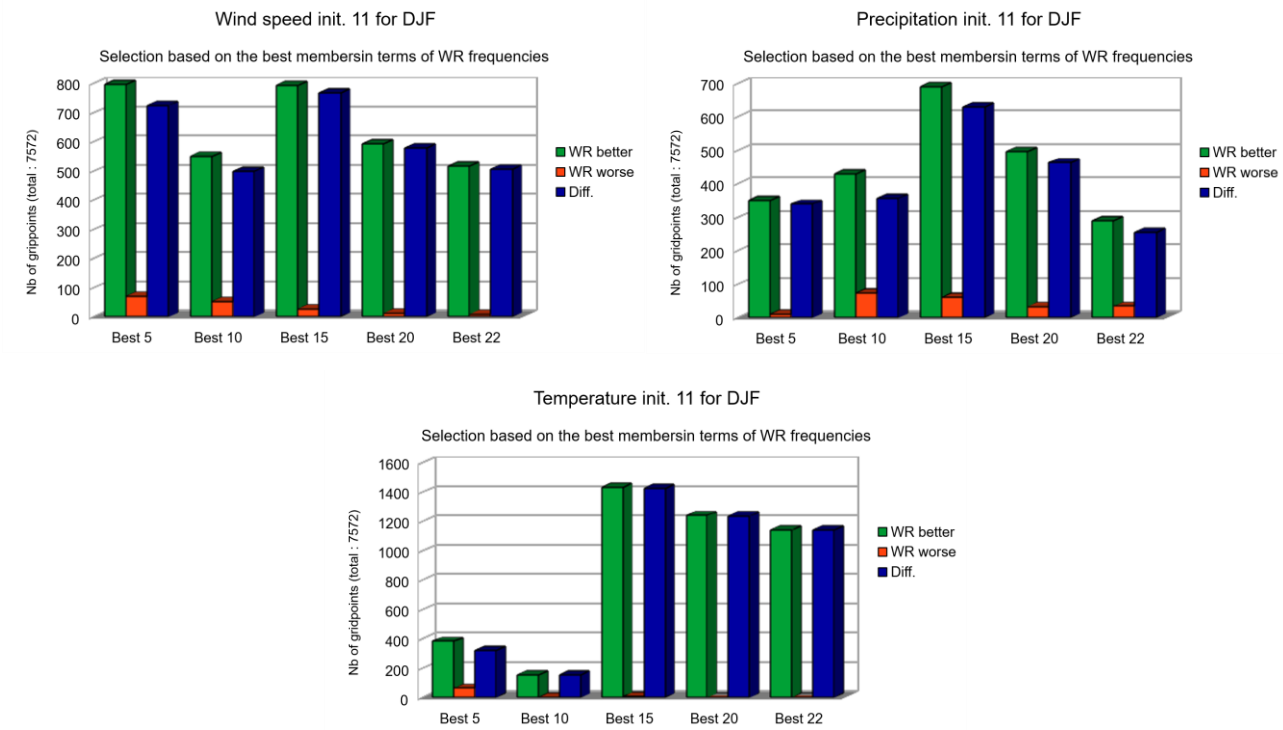


Figure 54. On the European domain (7572 grid-points), in green/red, number of grid-points where the skill of ADAMONT "with WR" is significantly better/worse than ADAMONT "without WR". Several experiments are presented, with sub-ensemble of the *N* best members in terms of WR frequencies, with *N* varying from 5 to 22 members.

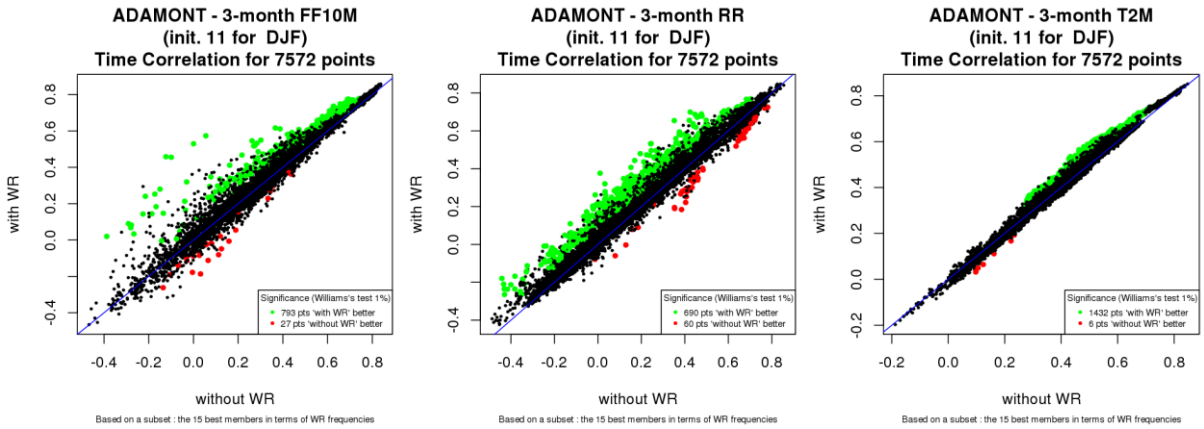


Figure 55. Comparison of ADAMONT with/without WR for DJF interannual correlation, for the 3 parameters, when selecting the 15 best members ("best" in terms of forecasting WR frequencies).

Comparing the two selection methods, the second approach seems slightly more efficient if we consider the number of grid points where there is a significant improvement. In detail, it is clearly better for precipitation, slightly worse for wind speed and they are equivalent for temperature.

So it seems that benefitting from WR in ADAMONT relies on better predictability of WR frequencies. These results open up a perspective of improvement in using WR in ADAMONT: if we were able to have an “external” forecast of WR frequencies, with enough skill, we could select members (or exclude members not compatible with these WR frequencies). This “external” WR forecast (by “external”, we mean external to ADAMONT) could come from a set of other models for example, assuming that a multi-model prediction of WR would be more skillful than one model prediction. This assumption should be addressed when dealing with multi-model skill.

4.5.2 Multi-model experiment

In seasonal forecasting, a multi-model approach is known to enhanced forecast skill: multi-model combinations generally reduce errors and improve consistency and reliability (Hagedorn et al. 2005).

We treated three C3S models (see Climate Data Store dataset documentation in Copernicus Knowledge Base https://confluence.ecmwf.int/display/CKB) in addition to ECMWF-SEAS5:

- Météo-France (MF) system 7, 25 hindcast members;
- Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) SPS3, 40 hindcast members;
- Deutscher Wetterdienst (DWD) GCFS 2.0, 30 hindcast members.

We computed each model’s climatology over the 1993–2016 hindcast period, for November start dates, and launched ADAMONT model by model, on each of their ensemble members. Indeed, as each seasonal model has a different climatology, it does not make sense to run ADAMONT on the super-ensemble over all models. The multi-model ensemble is built by collating the individual ADAMONT outputs. It is therefore composed of 120 members. As all members have the same weight, some models contribute more weight than others (for instance CMCC accounts for 33% of total weight, ECMWF only 20%).

As in the previous sections, we ran several experiments, with and without WR, and compared multi-model versus individual model scores. Before this, we analysed WR predictability of the multi-model ensemble in comparison to individual models.

4.5.2.1 WR prediction skill

First of all, we needed to check WR predictability of the multi-model ensemble, in comparison to individual model skill. We used the same metrics as in section 4.4.2. The main results are

summarized in Figure 56. The use of the multi-model ensemble does not bring noticeable additional skill. In terms of AUC (for 2 categories) and interannual correlation, the results are similar to the means of individual models. Scores are globally positive (AUC > 0.5 and correlation > 0) but weak. Globally, NAO+ and NAO- remain the most predictable regimes, and Blocking is the least predictable. Thanks to a greater number of members, multi-model scores are probably more robust than individual model scores.

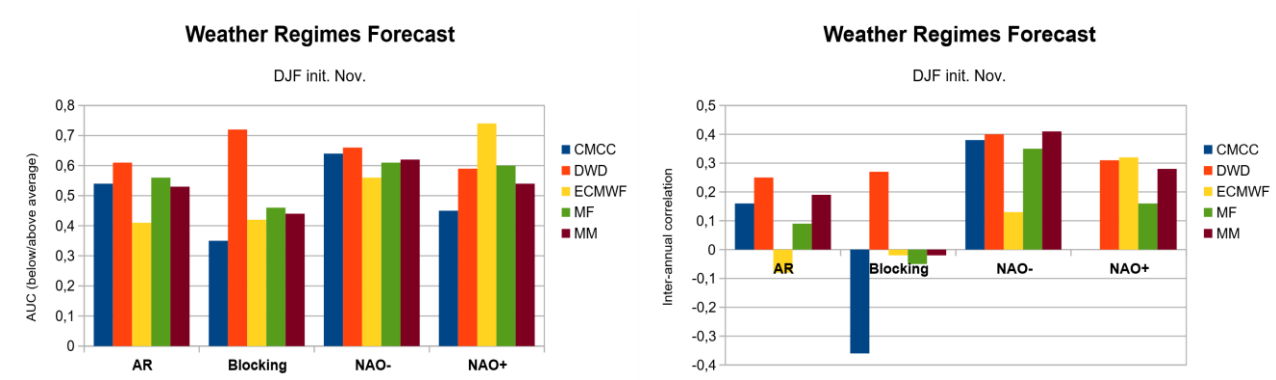


Figure 56. Scores of DJF (start date November) WR frequency prediction, for four C3S models and for a multi-model combination.

4.5.2.2 Comparison of ADAMONT applied on individual models versus multi model

Given how the multi-model ensemble is built (i.e. an aggregation of individual model outputs), we don't expect a great improvement in terms of skill in comparison to individual models.

We have computed several probabilistic and deterministic scores on each grid-point of the European domain, on three-month forecast means, as in the previous sections: bias, CRPSS (in reference to climatology) and Interannual correlation.

Considering bias (not shown), the multi-model ensemble performs very close to individual models. Looking at CRPSS and interannual correlation boxplots (made of more than 7000 grid-points, not shown), it appears that the interquartile intervals are a little bit smaller in the multi-model ensemble than for each individual models. This means that spatially, the multi-model forecast skill is slightly more spatially homogeneous. This is confirmed on maps, particularly for wind-speed (Figure 57) where the area where correlation is higher than 0.2 or 0.4 is larger for the multi-model results than for each individual model.

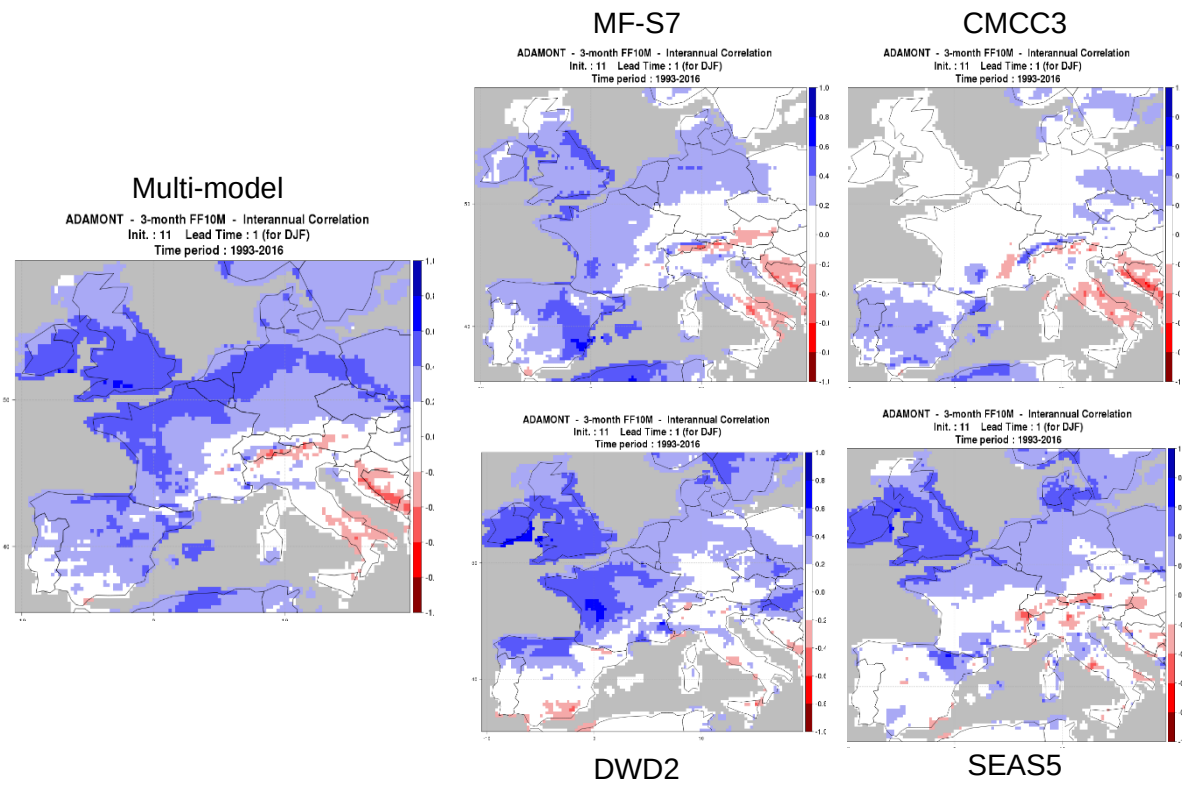


Figure 57. Interannual correlation for 10m wind speed (FF10M), for the multi-model ensemble, and individual models MF-S7, CMCC3, DWD2 and SEAS5, using ADAMONT computed with WR. The models are initialized in November, and these results are for DJF averages over the 1993–2016 hindcast period.

4.5.2.3 Added-value of using WR in ADAMONT for the multi-model

In this section, we check if WR improve predictability in the multi-model combination of ADAMONT outputs.

Figure 58 shows CRPSS boxplots built on the entire European domain. The differences between using and not using WR are too little to be noticeable, for both NDJ and DJF.

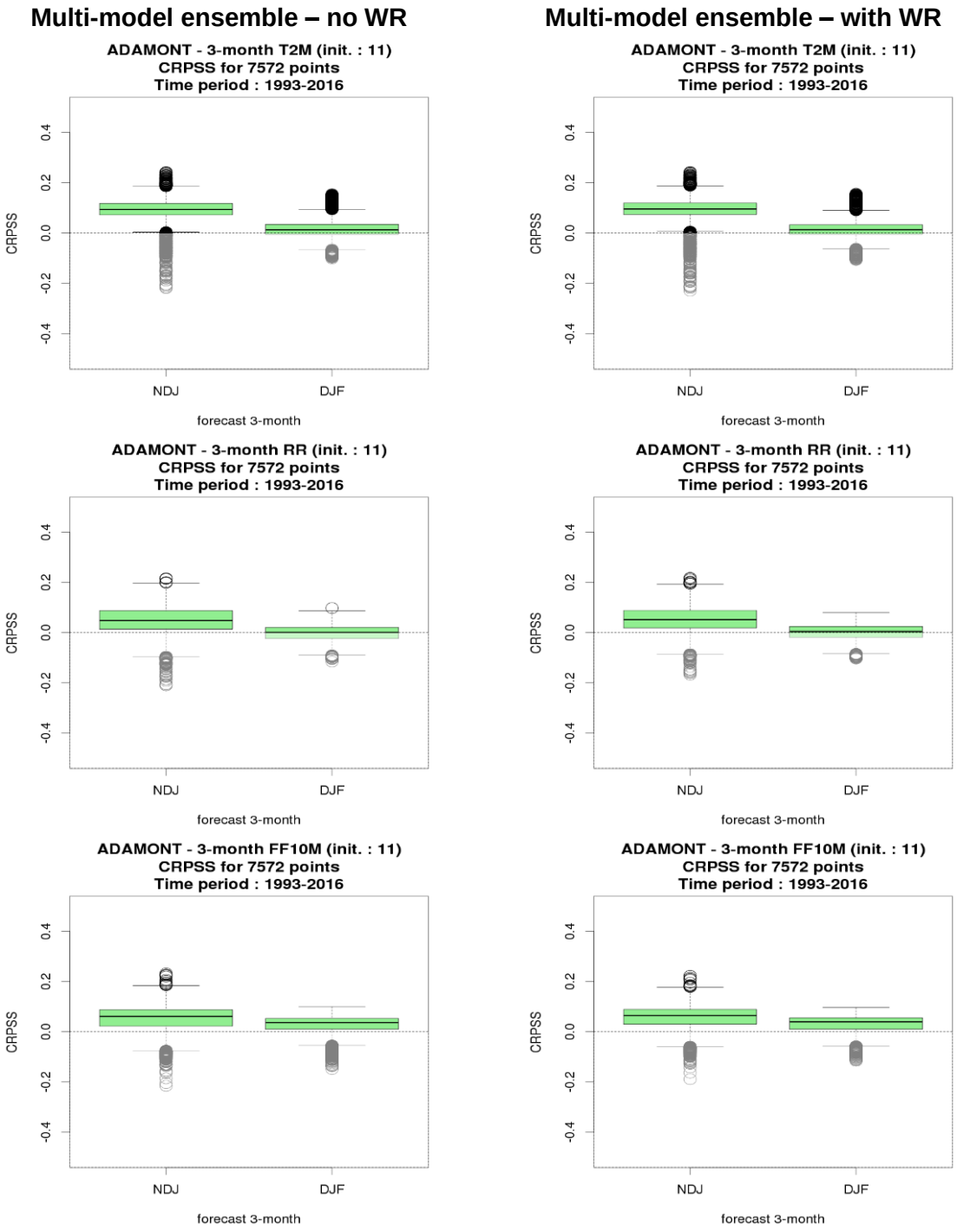


Figure 58. Boxplots of CRPSS, for temperature (upper line), precipitation (middle line) and wind-speed (lower line), from the multi-model ensemble without WR (left) and with WR (right). The models were initialized in November, and the results are for NDJ and DJF averages over the 1993–2016 hindcast period.

The results are similar when looking at maps of interannual correlation (Figure 59).

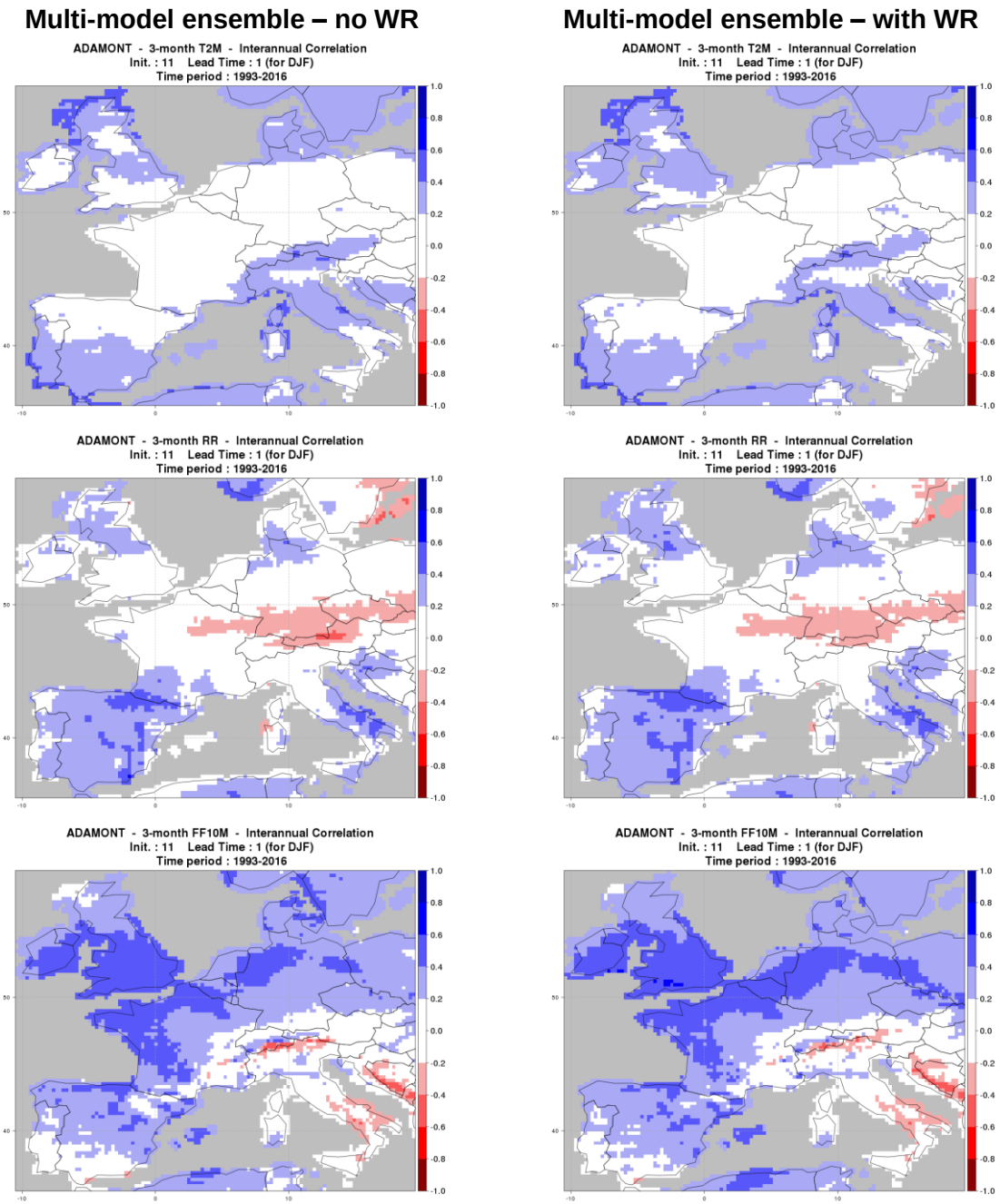


Figure 59. Interannual correlation for the multi-model ensemble, using ADAMONT without WR (left) and with WR (right), for temperature, precipitation and wind-speed. The models were initialized in November, and the results are for DJF averages over the 1993–2016 hindcast period.

4.5.3 Member selection in the multi-model super ensemble

We continued the methodology developed in section 4.5.1 to investigate the size of the ensemble (*N* members over 120), to consider how to maximize the benefit of using WR in ADAMONT. Beforehand, the 120 members have been ranked on the basis of their WR frequency forecast over DJF. The selection consists here in keeping the *N* best members in terms of WR frequency forecast.

We have tested several values of *N*. For each *N* we have compared ADAMONT "with" and "without" WR. This comparison relies on 3-month mean interannual correlations at each point of the domain: we have counted the number of significantly better or worse grid-points over the European domain, using the Williams test. Our results are summarized in Figure 60.

Considering the three parameters, the best compromise is around 90 members. We don't consider the "peak" visible for 5 members (possibly even higher for less than 5 members, not calculated), because such a small ensemble would not be realistic in operation: we need quite large ensembles to represent uncertainty in seasonal forecasting. Wind speed exhibits a broader range of improvement (between 50 and 100 members) than precipitation and temperature. Temperature even shows a net drawback for *N* between 40 to 70: clearly this parameter benefits the least from WR.

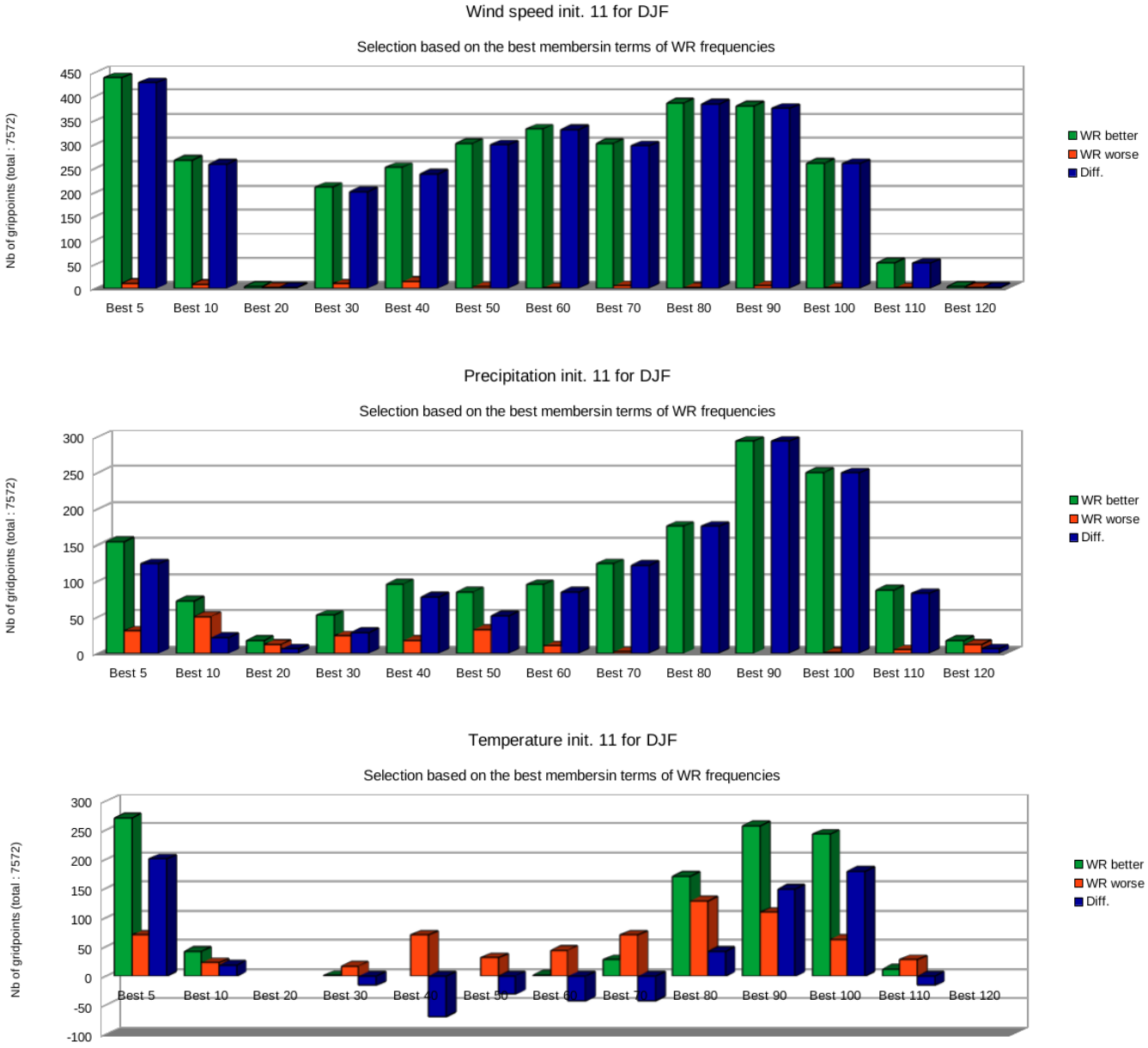


Figure 60. For different sizes of multi-model ensemble (from 5 to 120 members), the number of grid-points with significant improvement (green) or degradation (orange), and the difference (blue), of the 3-month-mean interannual correlation. Calculated for the three parameters on the European domain, for DJF (start date November).

Results for $N = 90$ members are presented in Figure 61 and Figure 62. Wind speed and precipitation show clear improvement in regions where the 120-member multi-model ensemble has low skill. Considering wind speed, the improvement is mainly visible on mountainous regions and can reach +0.6 in correlation. For precipitation the improvement is mainly located on the northern part of the domain, and benefits are also seen in lowland regions (eastern part of Germany, and Poland). For temperature, changes are much smaller and are confined on lowland regions like Southern England and North-Western parts of France.

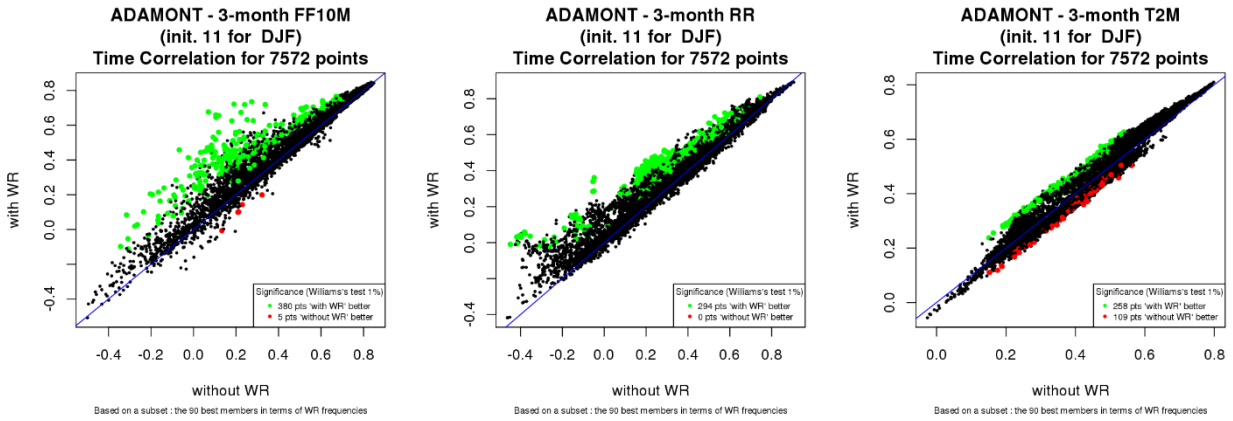


Figure 61. Comparison of interannual correlations with and without WR for wind speed (left), precipitation (middle) and temperature (right), using multi-model sub-ensembles of the the 90 better members (in terms of WR frequency forecast). Results are for DJF forecasts initialised in November.

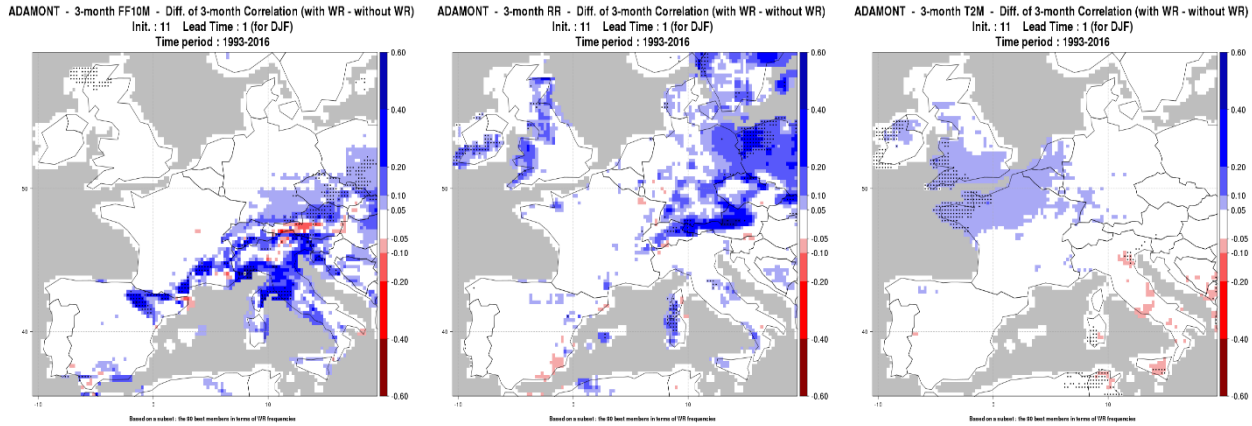


Figure 62. Maps of the difference between interannual correlations ("with-WR" minus "no-WR") for wind speed (left), precipitation (middle) and temperature (right), for multi-model sub-ensembles of the the 90 better members (in terms of WR frequency forecast). Results are for DJF forecasts initialised in November.

4.6 Conclusions and operational perspectives

In this study, we have investigated the impact of using WR in ADAMONT, through different sensitivity experiments. In the framework of SECLI-FIRM, we have applied ADAMONT on a European domain and on three variables of interest for the case studies of the project: temperature, precipitation and wind-speed.

By applying ADAMONT (with or without WR), the main defects of seasonal forecast outputs are corrected at daily timescale. Indeed daily, monthly and three-month corrected values are consistent with the “local climate”, represented by reference data (here ERA-5). Moreover, interannual variability of raw seasonal forecast data is preserved.

The introduction of WR in ADAMONT aims to enhance locally the quantile mapping correction. This is justified by systematic errors of seasonal forecast models, these errors being dependent on WR (section 4.3.3). Anyway the predictability of WR frequencies at seasonal scale by state of the art Seasonal Forecast models is poor, whether by individual models or by combining several models (sections 4.4.2 and 4.5.2.1). A priori, this is a major handicap for ADAMONT, because one of its main principles is precisely to rely on WR to apply a daily quantile mapping correction to different atmospheric parameters conditioned on WR.

It appeared that in a context of “perfect WR forecast” (section 4.4.1) there is significant added value from conditioning the quantile mapping correction on WR, for wind speed (mainly over mountains and coastal areas) and precipitation (mainly over lowlands), whereas the impact on temperature is low. This is measurable for monthly means, but also for daily data both in space and time.

Unfortunately, in “real forecast” conditions (section 4.4.3 and the introduction of section 4.5), the skill of ADAMONT “with WR” and “without WR” are almost identical when applied to the ECMWF SEAS5 model. In order to try to “boost” the effect of WR, we have tested three approaches. The first approach is to select a sub-sample of members within the ensemble (section 4.5.1), using a criterion of “good predictability” of WR. The second approach consists of using several seasonal forecast models to gain predictability in WR (section 4.5.2). The third approach is a mix of the first two (section 4.5.3). Results confirm what we have seen in the “perfect forecast” experiment: if WR are correctly foreseen, ADAMONT with WR helps to increase skill both for individual models and the multi-model combination, especially for wind speed over mountainous regions. For precipitation the improvement is mainly located in the northern part of the European domain, and mainly over lowland regions. For temperature, changes are much smaller and confined to lowland regions.

These results demonstrate the interest of using WR to refine a classical quantile mapping correction, when WR predictability is good enough. In an operational perspective, we identify two possible options:

- the value added from using WR in ADAMONT is not systematic: it is limited to situations of good predictability of the large-scale circulation. Because WR are built to "summarize" large scale circulation over the Northern Atlantic region, these situations of "higher seasonal predictability" generally occur when strong climate drivers are present, such as ENSO events (El Niño or La Niña), or in the case of strong oceanic anomalies in general (Viel, 2019; Mariotti, 2020). Identifying such windows of opportunity could lead users to choose to use a version "with WR" in the case of strong drivers, and a version "without WR" in all other situations. This strategy should be tested and evaluated in further studies.

- Statistical forecasts of the winter NAO, based on climate drivers such as ocean surface temperature, sea ice field and stratospheric circulation can outperform current dynamical models (Wang, 2017; Hall, 2017). The same methods could be used to forecast WR frequencies, and provide an "external" forecast of WR (as suggested in section 4.5.1). This forecast would help to select a subset of members, as was done in sections 4.5.1 and 4.5.3)

4.7 Annex: Choice of quantiles in the quantile-mapping correction

The question of how many quantiles we choose in the quantile-mapping correction is crucial in the application of such a statistical correction. Using too many quantiles could lead to over-fitting to samples used to fit the quantiles, which could then affect the efficiency of the correction in its application to new data. In contrast, too few quantiles could lead to a rough description of distribution.

In ADAMONT, we make an adjustment of daily data, with a differentiation per regime and per month. Using the ECMWF-SEAS5 seasonal forecast model, we can use a historical period of 24 years to calculate quantiles. For instance, if we are interested in correcting a grid-point for a specific day in December, with a forecast issued in November, quantiles can be calculated with:

- 25 members × 31 days × 24 years =18,600 elements for SEAS5; i.e. fewer than 5,000 elements per regime.
- 31 days × 24 years = 744 elements for ERA5; i.e. fewer than 200 elements per regime.

Taking the example of daily precipitation in Toulouse, its distribution can reasonably be approximated by a log-normal distribution (Figure 63).

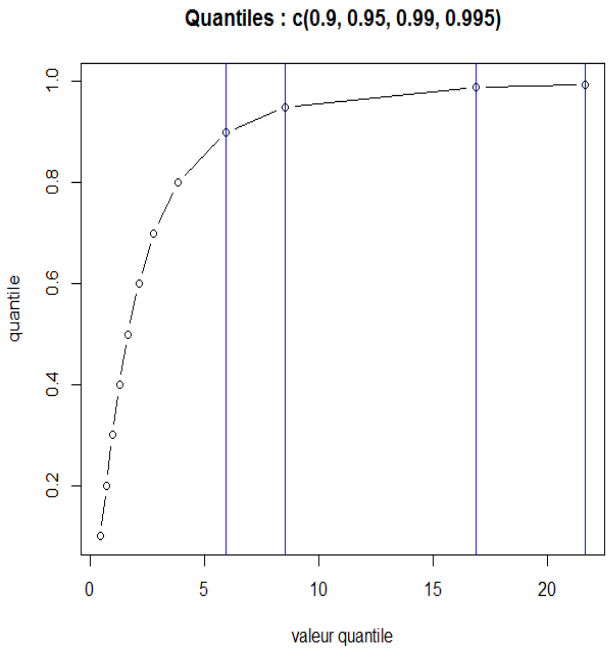


Figure 63. Log-normal law with *m* = 0.5 and sd = 1 (a good fit for Toulouse precipitation in May). Blue lines are the theoretical quantiles 0.9, 0.95, 0.95 and 0.99.

For such a log-normal law, with *m* = 0.5 and sd = 1, we know the theoretical values of any quantile (e.g. the blue lines in Figure 63).

Using a script written in *R*, from a random sample of 10,000 elements from this log-normal law (*m* = 0.5, sd = 1), we can estimate quantiles using a sub-sample of *n* elements (*n* from 10 to 10,000) randomly extracted from those 10,000 elements. We can then visualise these quantile estimates for the successive values of *n*. Two examples of this test are presented in Figure 64. They clearly illustrate that quantiles 0.9 and 0.95 can be reasonably estimated even with small samples, but that the most extreme quantiles (0.99 and higher) require far larger samples to avoid incorrect estimation.

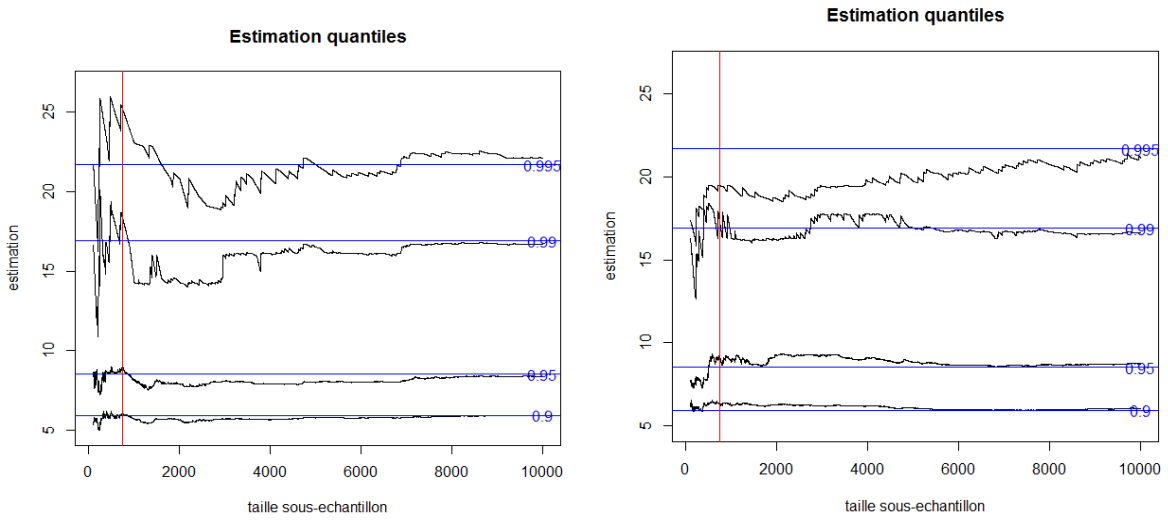


Figure 64. Two examples of an iterative estimation of 4 quantiles (0.9, 0.95, 0.95 and 0.99) of a log-normal law (m = 0.5, sd = 1).The red vertical bar corresponds to a sample size of 744 elements.

5 Use of weather types for wind prediction in Spain and Italy

5.1 Introduction

This work explores the use of weather types for the wind industry in Spain and Italy. This is motivated by SECLI-FIRM Case Study 4, which focuses on the anomalous wind production observed in Spain in 2014 and 2015. We generalized the problem from the particular period of the case study to a wider period, 1979–2019, and increased our regions of interest to Europe and Italy. The increase in the period is chosen to understand general connections between the weather types and the wind production. The increase in the regions considered is to assess the efficiency of this approach to regions that can be of interest for UL.

We present here our initial work, using daily data of sea level pressure and wind speed from ERA5 as a first approach to find this link. The rationale behind this was to start with short time scales and move later to longer time scales (weekly and monthly), understanding the point at which we have a representative signal between a classification of weather types and wind speed in these regions. Identifying this timescale could lead to a delimitation of the use of this approach, and the development of other statistical approaches for the shorter timescales where this approach could not provide any signal. For example, if we knew that a certain weather type provides high winds on a monthly time scale, and is also skilfully predicted by seasonal forecast models, we could adapt weekly wind distributions to the match the predicted value, based on the past distributions of similar months. This would allow us to test different approaches for modifying the weekly distributions depending on the predicted monthly value. This extension of the approach to longer time scales represents a possible plan for future work.

Section 5.2 describes the data used and the methodology for our investigation on daily timescales. We then present the weather regimes obtained and their impact on wind speed in Europe (section 5.3) and for the chosen regions of Spain and Italy (section 5.4; we divide Italy into northern and southern regions). The impact of increasing the number of weather types considered for the classification in the representation of the wind regimes is discussed in 5.5. Finally, in section 5.6 we summarize our conclusions and describe potential directions that we could follow to continue this study.

5.2 Data used and weather type methodology

The data used in this analysis is ERA5 (Hersbach et al. 2020) over the period 1979–2019. In particular we used PMSL daily data, removing the monthly annual cycle to obtain anomalies, and applying a 3-day running mean to smooth the variability. We also applied the same smoothing technique to daily 10 m wind speed data (WS).

We performed a *k*-means clustering on the smoothed PMSL anomalies, initially considering 8 clusters, and later extending this method to 16 and 32 clusters. The PMSL data is limited to the region 30°W – 20°E, 35°N – 70°N (Figure 65), following the definitions in Neal et al. (2016).

The centroids we obtained differ from the ones defined in Figure 2, as methodology is slightly different, and the PMSL data has different smoothing and covers a different period. Despite this, we can still see some similarities. For example, Cluster 1 is the Scandinavian high pattern, Cluster 3 corresponds to NAO-, and Cluster 5 is the Azores high pattern. We will refer to these clusters as weather types or regimes for this analysis.

We initially consider 8 clusters, to assess how well they are able to describe WS variability in the chosen regions of Spain and Italy. We divided Italy into two subregions: northern and southern, as we assumed the wind in these subregions would be subject to different climate drivers. We considered the PMSL classification obtained previously and considered the regional values of PMSL and WS locally for each of the clusters.

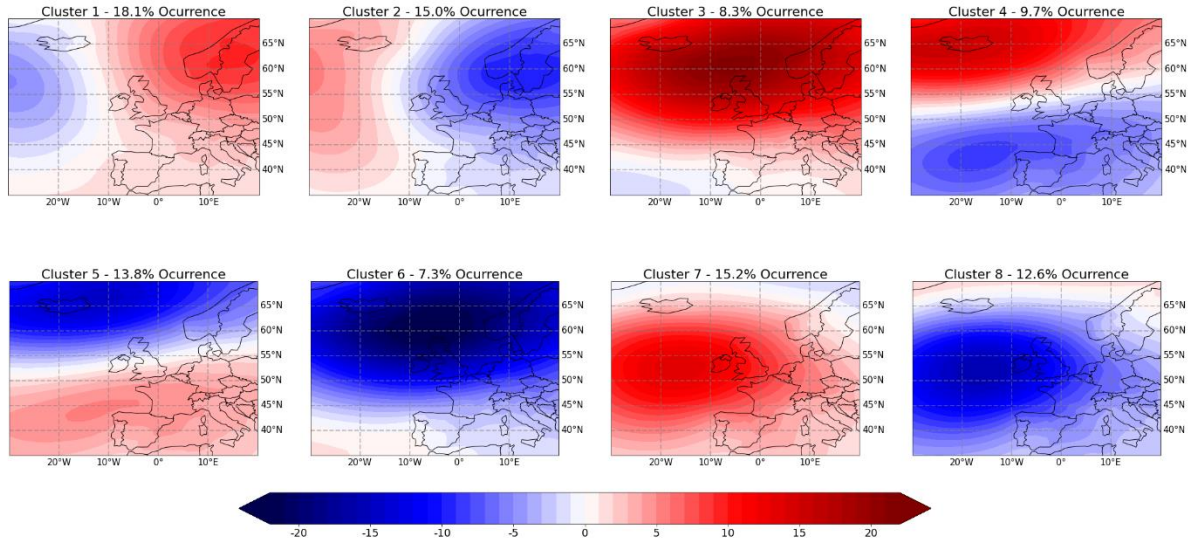


Figure 65. The 8 cluster centroids in terms of PMSL anomalies (hPa), from ERA5 in the 30°W – 20°E, 35°N – 70°N region, with respect to monthly means from 1979 to 2019.

5.3 Distribution of WS across Europe under different weather types

We have computed the mean wind speed maps from the daily data in each of the clusters (Figure 66). They show that there is a distinct spatial pattern associated with each cluster, with different patterns of strong and weak wind anomalies over Europe. Despite the information in these patterns, they are only the means of each cluster: they don't show any information about the variability of the wind speeds associated with these weather types.

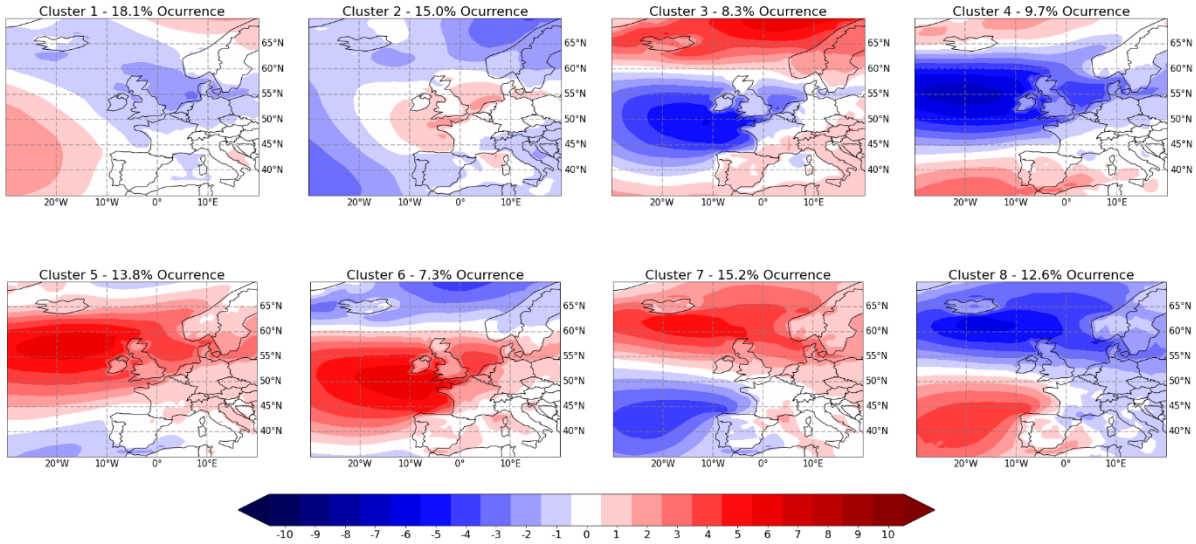


Figure 66. Mean WS anomalies (m/s) for the 8 clusters, using data from ERA5 in the 30°W – 20°E, 35°N – 70°N region, with respect to monthly means from 1979 to 2019.

The variability of the daily wind speeds, averaged over the European region, shows a very wide range of values for each of the clusters. Whereas the spatial average for the PMSL (Figure 67 top row) has values similar to the cluster mean (vertical line), the wind speeds instead have very widely-spread values with respect to the mean. This makes it difficult to distinguish days with certain wind values as belonging to one cluster or another. In order to check that the choice of the spatial average was a representative metric for the spatial pattern, we also analysed the cumulative distribution function (CDF, Figure 67 bottom row). The CDFs support the result obtained from the spatial means, that indeed the different clusters have almost indistinguishable wind regimes.

Figure 67 Top: Histograms of the spatial mean over Europe of daily PMSL values for each of the 8 clusters. Middle: As before, but for wind speed over the region. Bottom: cumulative distribution function of the daily wind speed.

5.4 Distribution of WS for weather types in Spain and Italy

We repeated the same analysis, but focusing on Spain, northern Italy and southern Italy. Italy was divided as the wind drivers of the regions could be very different due to their geographical differences. We examined in detail the patterns and distributions of the clusters obtained previously, but only considering the values in these regions. Hence the cluster-means of both PMSL and WS shown in the first two rows of Figures 68, 69 and 70 are local version of the maps shown in Figures 65 and 66 respectively. On average, we can see that the clusters do differentiate between different mean levels, showing high/low PMSL and WS anomalies with respect to the monthly means. We can use these cluster mean maps as being representative of the cluster, but the errors resulting from this assumption have to be assessed.

We therefore computed maps of the average error on each grid point for PMSL and WS (3rd and 4th rows respectively for Figures 68, 69 and 70). This is computed as the mean of the absolute differences, at each grid point, between the daily anomalous value and the cluster mean value for each cluster. The results for the regions show that, in the case of PMSL, the error would be between roughly 3–6 hPa, whereas for WS these values are between 8–11 m/s.

Figure 71 shows the result of the clustering classification in terms of the distribution of spatially averaged PMSL (even rows) and WS (odd rows) daily smoothed values. The PMSL distribution is still centred around the cluster mean values (vertical lines). The WS distributions are slightly more concentrated (less spread) compared to the European average case. However, they still exhibit a wide range of WS values for each cluster. The plots show that the peaks of the WS distributions do not correspond to the cluster mean values, due to the spread and skewness of the distributions. We still have to analyse the performance of these distribution peaks as being representative of the clusters, instead of the cluster means. Therefore, we can conclude that, similarly to the European averages, we find that, using 8 clusters of daily European PMSL, the WS cluster means do not provide a good representation of the WS daily anomalies in any of the three regions.

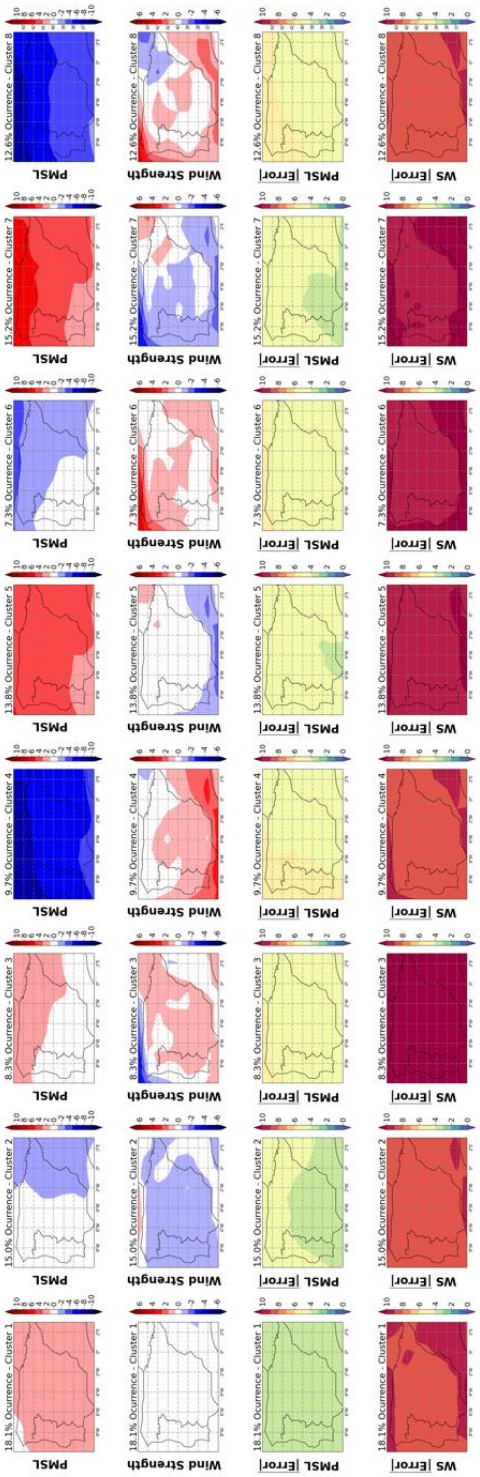


Figure 68. For Spain, for each of clusters we show the centroid PMSL, centroid WS, the averaged anomaly of each grid point to the PMSL centroid for all the days of the cluster, and the same but for WS.

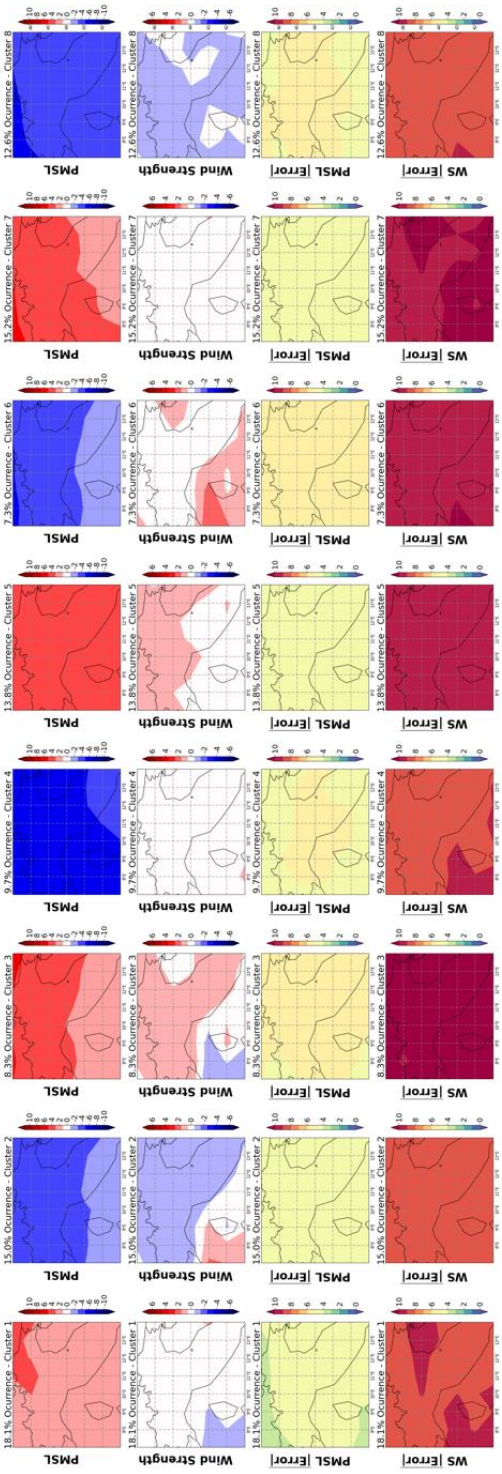


Figure 69. For northern Italy, for each of clusters we show the centroid PMSL, centroid WS, the averaged anomaly of each grid point to the PMSL centroid for all the days of the cluster, and the same but for WS

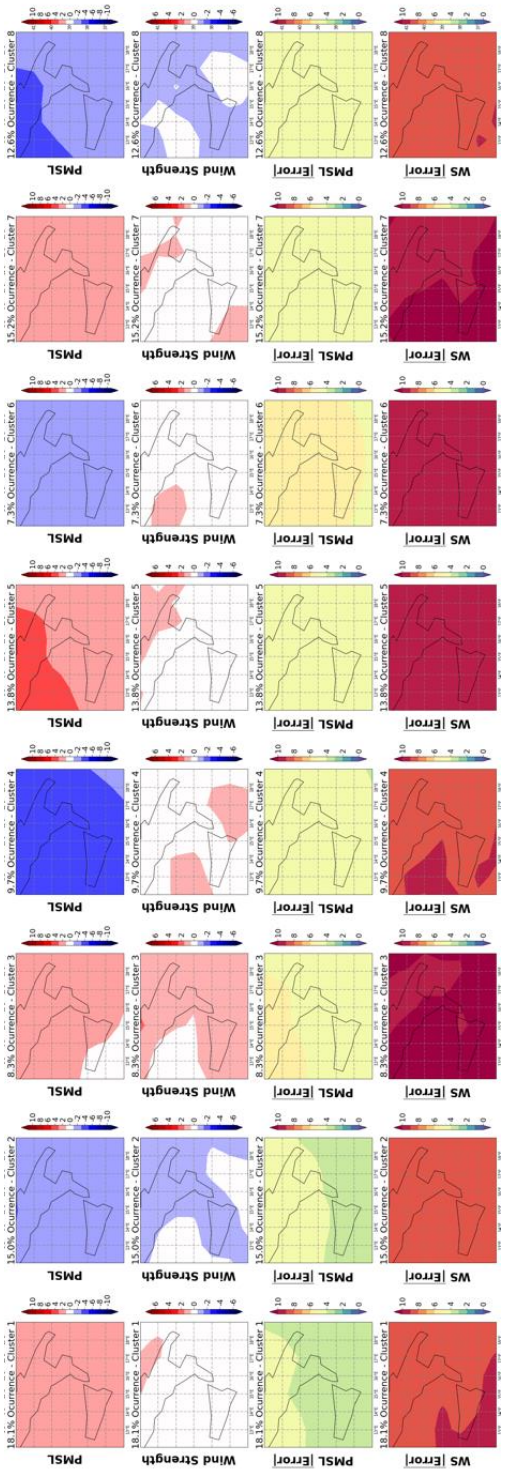


Figure 70. For southern Italy, for each of clusters we show the centroid PMSL, centroid WS, the averaged anomaly of each grid point to the PMSL centroid for all the days of the cluster, and the same but for WS

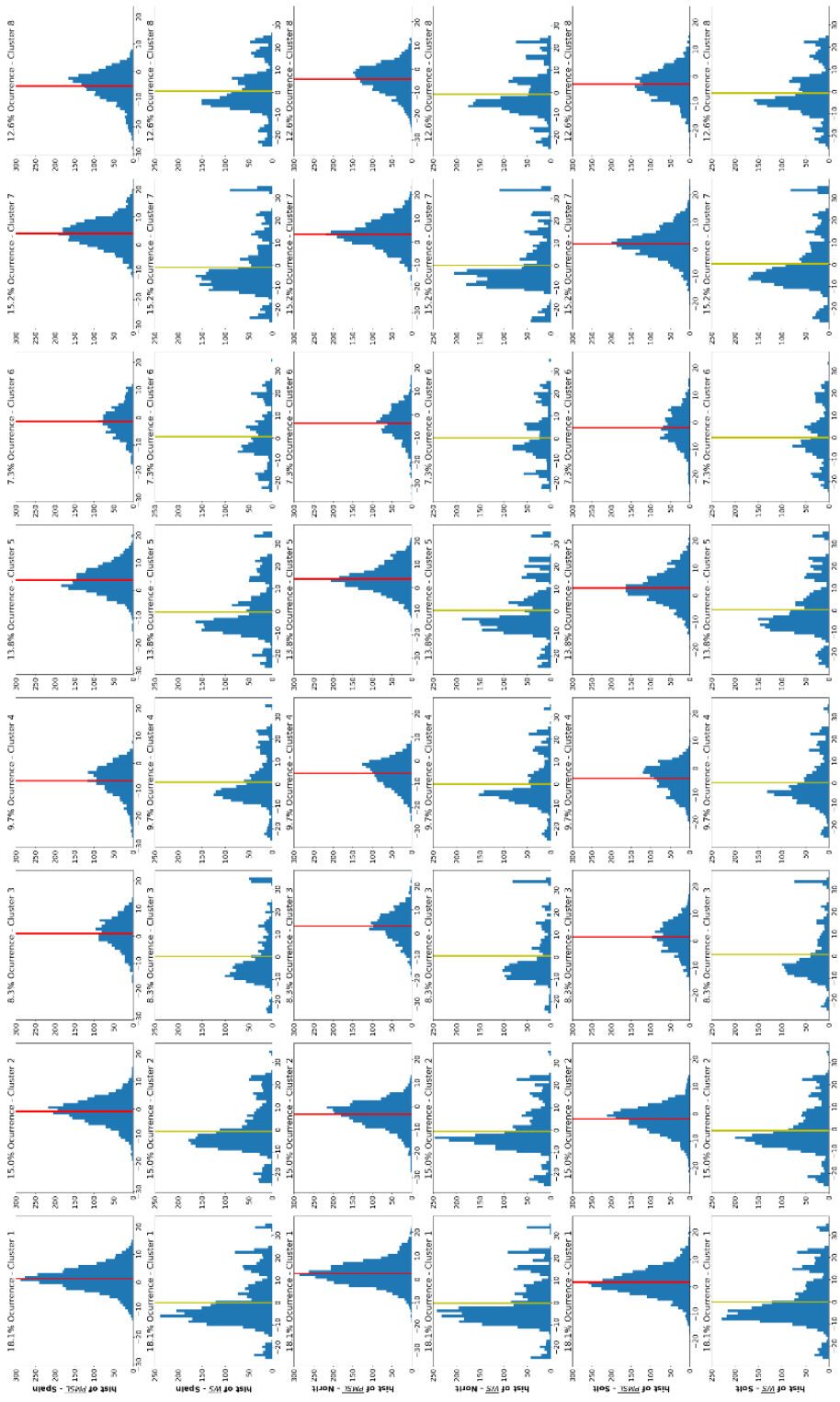


Figure 71. Histograms of spatially-averaged PMSL (odd rows, with red lines for the mean) and WS (even rows, with yellow lines for the mean), for Spain (top two rows), northern Italy (NorIt, middle rows) and southern Italy (SoIt, bottom two rows), for each cluster (columns).

5.5 Impact of the number of clusters in the representation of PMSL and WS

The choice of the number of clusters was also investigated with the hope that increasing it from 8 to 16 or 32, would improve the significant representation of WS in the chosen regions. We have compared for each of the months the distances of the spatial average of daily PMSL and WS values to the associated cluster mean. Although in Figure 72 we only show the analysis for Europe, the following results still hold for the smaller regions of Spain and northern and southern Italy (not shown).

The boxplots for PMSL (Figure 72 top) show indeed that, as expected, increasing the number of clusters provides a reduction in the distances between the monthly PMSL pattern and the centroid (cluster mean pattern). However, for WS this is not the case. Distances are maintained through increasing the number of clusters (Figure 72 bottom). This result also highlights the inability of this approach to represent daily wind speed in these regions with the WS cluster means, despite considering an increased number of clusters.

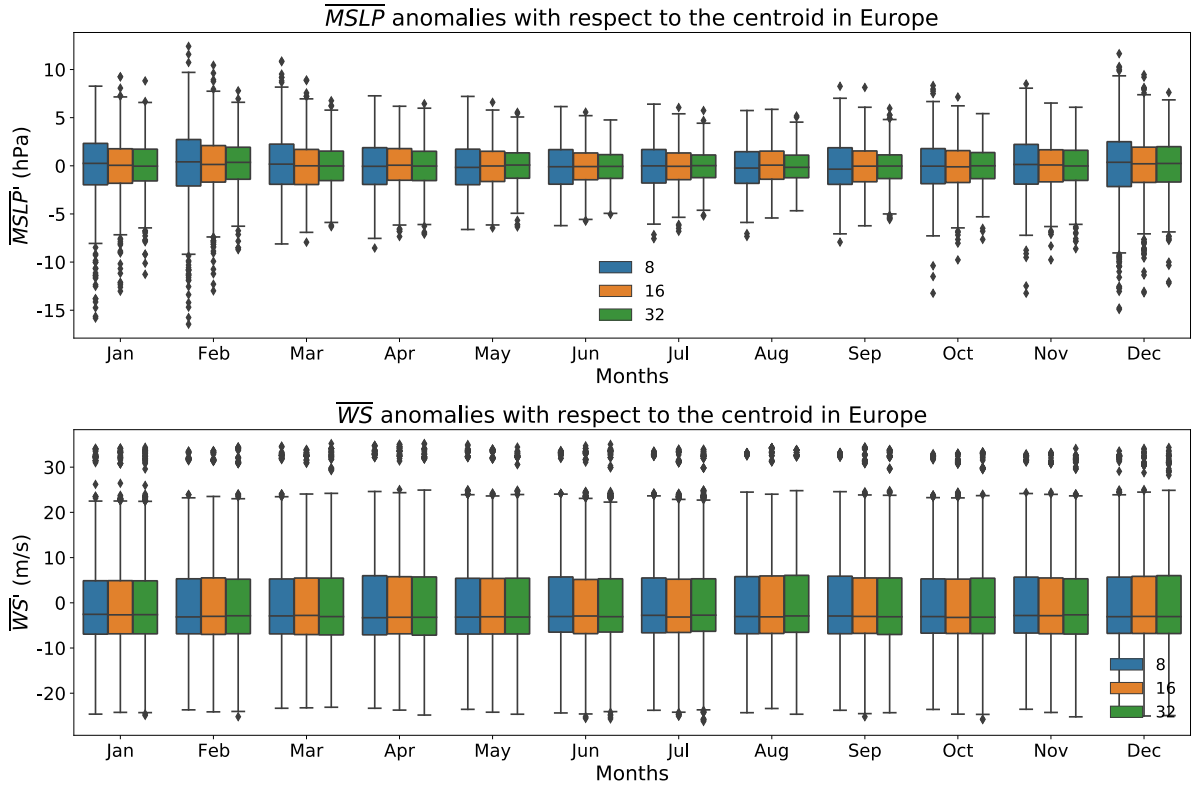


Figure 72. Box plots of the distributions each month of the distances between spatial averages of each daily PMSL (top) and WS (bottom) with respect to their associated cluster mean pattern.

5.6 Conclusions and Next Steps

In this section we have presented the use of weather types using PMSL clustering to assess the representation of WS in Spain and Italy. The analysis has been done by taking daily data from ERA5, both PMSL and 10-m WS smoothed with a 3-day running mean, and performing a clustering analysis on the PMSL data. The *k*-means algorithm used associates each of the days of the dataset to a cluster. We maintained this classification for the WS patterns, obtaining an associated WS classification and corresponding cluster mean patterns. The main conclusion of the work presented here is the lack of ability of the weather regimes approach to represent the wind speed on a daily time scale with the associated WS cluster means in the chosen regions.

Additionally, we repeated the analysis with an increased number of centres (16 and 32) to assess the impact of the number of clusters. Unfortunately, the increase in clusters didn't bring any improvement on the representation of WS variability in the regions of the study.

Regarding that daily distributions of WS cannot be well represented by the cluster mean patterns obtained by this approach, we have to test whether other metrics of the WS clusters instead of the mean could be a more suitable representative. We will also test the same approach with monthly values of ERA5 to understand whether by reducing the variability by considering monthly data, the new classification provides better results. Also, we would like to understand if we are able to infer anything about the distributions of daily values from monthly values. As daily WS is not well represented in seasonal prediction systems, it would be beneficial if a statistical approach could be developed using the monthly outputs of PMSL data provided by the prediction systems. We could explore alternative approaches to provide improved WS predictions based on the weather types approach using statistical relationships of daily and monthly wind distributions.

6 References

Aarnes, Ole Johan, Øyvind Breivik, and Magnar Reistad, 2012. Wave Extremes in the Northeast Atlantic. *J. Climate* **25,** 1529–1543, https://doi.org/10.1175/JCLI-D-11-00132.1

Ansell T.J., P. D. Jones, R. J. Allan, D. Lister, D. E. Parker, M. Brunet, et al., 2006. Daily mean sea level pressure reconstructions for the European–North Atlantic region for the period 1850–2003. *J. Clim.* **19**: 2717–2742, https://doi.org/10.1175/JCLI3775.1

Barnston, Anthony G., 1995. Our Improving Capability in ENSO Forecasting. *Weather* **50,** 419–430, https://doi.org/10.1002/j.1477-8696.1995.tb06068.x

Bett, P., H. Thornton, A. Troccoli, 2018. Skill assessment of energy-relevant climate variables in a selection of seasonal forecast models. ECEM Deliverable D2.2.1 v2, https://doi.org/10.5281/zenodo.1293862

Cassou, C, 2008. Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature* **455**, 523–527, https://doi.org/10.1038/nature07286

Clark, R, P. E. Bett, H. E. Thornton, A. A. Scaife, 2017. Skilful seasonal predictions for the European energy industry. *Environ. Res. Lett.*, **12**, 024002, https://doi.org/10.1088/1748-9326/aa57ab

Dawson, A., T. N. Palmer, S. Corti, 2012. Simulating regime structures in weather and climate prediction models. *Geophys. Res. Lett.*, **39**, L21805, https://doi.org/10.1029/2012GL053284

Dawson, Andrew, 2016. eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data. *J. Open Research Software*, **4**, e14, https://doi.org/10.5334/jors.122

Diebold, F. X., and R. S. Mariano, 1995. Comparing Predictive Accuracy. *J. Business & Econ. Stat.*, **13**, 253-263, https://doi.org/10.1080/07350015.1995.10524599.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, F. Vitart, 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828

Fereday, D. R., J. R. Knight, A. A. Scaife, C. K. Folland, and A. Philipp, 2008. Cluster Analysis of North Atlantic–European Circulation Types and Links with Tropical Pacific Sea Surface Temperatures. *J. Climate*, **21**, 3687–3703, https://doi.org/10.1175/2007JCLI1875.1

Haakenstad, Hilde, Øyvind Breivik, Magnar Reistad, and Ole J. Aarnes, 2020. NORA10EI: A Revised Regional Atmosphere-Wave Hindcast for the North Sea, the Norwegian Sea and the Barents Sea. *Int. J. Climatol.,* **40**, 4347–4373, https://doi.org/10.1002/joc.6458

Hagedorn R, Doblas-Reyes FJ, Palmer T, 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus A: Dyn. Meteorol. Oceanogr.* **57**, 219–233. https://doi.org/10.3402/tellusa.v57i3.14657.

Hall, R. J., Scaife, A. A., Hanna, E., Jones, J. M., and Erdélyi, R., 2017. Simple Statistical Probabilistic Forecasts of the Winter NAO. *Weather and Forecasting*, **32**, 1585–1601, https://doi.org/10.1175/WAF-D-16-0124.1.

Hersbach, H, B. Bell, P. Berrisford, et al., 2020. The ERA5 global reanalysis. *Q J R Meteorol Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Huth, R., C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kyselý, and O. E. Tveito, 2008. Classifications of Atmospheric Circulation Patterns. Recent Advances and Applications. *Annals of the New York Academy of Sciences*, **1146**, 105–152, https://doi.org/10.1196/annals.1446.019

Jolliffe, Ian T., and David B. Stephenson, eds. 2012. *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* 2nd ed, Wiley-Blackwell, https://doi.org/10.1002/9781119960003

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P. & Madec, G., 2015. Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Q.J.R. Meteorol. Soc.*, **141**, 1072–1084, https://doi.org/10.1002/qj.2396

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E., & Albers, J., 2020. Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond, *Bull. Am. Meterol. Soc.*, **101**, E608–E62, https://doi.org/10.1175/BAMS-D-18-0326.1.

Matsueda, M and T. M. Palmer, 2018. Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Q.J.R. Meteorol. Soc.*, **144**, 1012–1027, https://doi.org/10.1002/qj.3265

Michelangeli, P., R. Vautard, and B. Legras, 1995. Weather Regimes: Recurrence and Quasi Stationarity. *J. Atmos. Sci.* **52**, 1237–1256, https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2.

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

Pedregosa et al., 2011. Scikit-learn: Machine Learning in Python. *J. Machine Learning Res.*, **12**, 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Reistad, Magnar, Øyvind Breivik, Hilde Haakenstad, Ole Johan Aarnes, Birgitte R. Furevik, and Jean-Raymond Bidlot, 2011. A High-Resolution Hindcast of Wind and Waves for the North Sea, the Norwegian Sea, and the Barents Sea. *J. Geophys. Res.: Oceans,* **116**, C05019, https://doi.org/10.1029/2010JC006402.

Scaife, A. A. & D. Smith, 2018. A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, **1**, 28, https://doi.org/10.1038/s41612-018-0038-4

Themeßl, M.J., A. Gobiet, and G. Heinrich, 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Clim. Change*, **112**, 449–468, https://doi.org/10.1007/s10584-011-0224-4.

Thornton, H. E., A. A. Scaife, B. J. Hoskins, D. J. Brayshaw, D. M. Smith, N. Dunstone, N. Stringer, P. E. Bett, 2019. Skilful seasonal prediction of winter gas demand. *Environ. Res. Lett.* **14**, 024009, https://doi.org/10.1088/1748-9326/aaf338

Van der Wiel, K., H. C. Bloomfield, R. W. Lee, L. P. Stoop, R. Blackport, J. A. Screen, F. M. Selten, 2019. The influence of weather regimes on European renewable energy production and demand. *Environ. Res. Lett.*, **14**, 094010, https://doi.org/10.1088/1748-9326/ab38d3

Verfaillie, D., M. Déqué, S. Morin, and M. Lafaysse, 2017. The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models. *Geosci. Model Dev.*, **10**, 4257–4283, https://doi.org/10.5194/gmd-10-4257-2017

Viel, C., Richon, J., Gayrard, F., Soubeyroux, J.M., Batté, L., Guérémy, J.F., 2019. Influence of ENSO and consistency between models on seasonal forecast skill: perspectives to identify windows of enhanced predictability. Poster presented at the 14th International Meeting on Statistical Climatology, Toulouse, France, http://www.meteo.fr/cic/meetings/2019/IMSC/. Available at https://doi.org/10.13140/RG.2.2.19692.90244.

Wang, L., Ting, M. & Kushner, P.J., 2017. A robust empirical seasonal prediction of winter NAO and surface climate. *Sci. Rep.* **7**, 279, https://doi.org/10.1038/s41598-017-00353-y.

Wilks, Daniel S, 2006. *Statistical Methods in the Atmospheric Sciences.* 2nd ed, Elsevier, https://www.elsevier.com/books/statistical-methods-in-the-atmospheric-sciences/wilks/978-0-12-751966-1.

Wilks, Daniel S., 2020. *Statistical Methods in the Atmospheric Sciences.* 4th ed, Elsevier. ISBN 978-0-12-815823-4, https://doi.org/10.1016/C2017-0-03921-6.

Williams, E. J., 1959. *Regression Analysis*. Wiley Series in Probability and Statistics, ISBN 9780471946779.

Williams, K. D., C. M. Harris, A. Bodas-Salcedo, J. Camp, R. E. Comer, D. Copsey, D. Fereday, T. Graham, R. Hill, T. Hinton, P. Hyder, S. Ineson, G. Masato, S. F. Milton, M. J. Roberts, D. P. Rowell, C. Sanchez, A. Shelly, B. Sinha, D. N. Walters, A. West, T. Woollings, and P. K. Xavier, 2015. The Met Office Global Coupled model 2.0 (GC2) configuration. *Geosci. Model Dev.*, **8**, 1509–1524, https://doi.org/10.5194/gmd-8-1509-2015

Zubiate, L., F. McDermott, C. Sweeney, M. O'Malley, 2017. Spatial variability in winter NAO–wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns. *Q.J.R. Meteorol. Soc.*, **143**, 552–562, https://doi.org/10.1002/qj.2943

The Added Value of Seasonal Climate Forecasting for Integrated Risk Management (SECLI-FIRM)

For more information visit

www.secli-firm.eu

or contact the SECLI-FIRM team at

info@secli-firm.eu

Grant Agreement n. 776868