On the use of Weather Regimes to forecast meteorological drought over Europe

Christophe Lavaysse [1,2], Jürgen Vogt [1], Andrea Toreti [1], Marco L. Carrera [3], and Florian Pappenberger [4]

[1]European Commission, Joint Research Centre (JRC), 21027 Ispra (VA), Italy
[2]Univ. Grenoble Alpes, CNRS, IRD, G-INP, IGE, F-38000 Grenoble, France
[3]Environment and Climate Change Canada, Dorval, QC, Canada
[4]ECMWF, Reading, United Kingdom

Correspondence: Christophe Lavaysse (christophe.lavaysse@ird.fr)

Abstract. An early warning system for drought events can provide valuable information for decision makers dealing with water resources management and international aid. However, predicting such extreme events is still a big challenge. In this study, we compare two approaches for drought predictions based, respectively, on forecasted precipitation derived from the extended ENSemble system of the ECMWF, and on forecasted Monthly Occurrence Anomaly of Weather Regimes (MOAWRs) also

5   derived from the ECMWF model.

Results show that the MOAWRs approach outperforms the one based on forecasted precipitation in winter in the north-eastern parts of the European continent, where more than 65% of droughts are detected one month in advance. While the approach based on forecasted precipitation achieves better performance in predicting drought events in central and eastern Europe in both spring and summer, when the local atmospheric forcing could be the key driver of the precipitation. Sensitivity tests also

10   reveal the challenges in predicting small-scales and drought onsets at longer lead times.

Finally, the results show that the ENSemble system of the ECMWF successfully represents most of the observed linkages between large scale atmospheric patterns, depicted by the weather regimes, and drought events over Europe.

1   Introduction

15   Developing a robust early warning system for drought events is a key challenge for modelers and forecasters. The timescale of these events (generally from one to several months) requires accurate numerical weather forecasts with long lead times. Due to the uncertainties of the models, the chaotic nature of the atmospheric circulation and the errors in the initial conditions, the reliability of precipitation forecasts is close to climatology beyond 2-week lead time (Haiden et al., 2017; Vigaud et al., 2017). In a recent study (Lavaysse et al., 2015), it has been shown that about 40% of the meteorological droughts, defined by

20   an anomaly of the standardized precipitation index (SPI), can be detected one month in advance by using the forecasted precip-

itation provided by the ECMWF extended ensemble. These forecasts might be improved by using post-processing techniques or predictors that are better simulated by atmospheric models (Lavers et al., 2016a, b; Ferranti et al., 2018).

The concept of Weather Regimes (WRs) was first introduced in the early 1950s, on the assumption that the atmosphere evolves between a finite number of large-scale circulation states. It is based on recurrent, persistent and/or quasi stationary states of the atmosphere, generally diagnosed with anomalies of geopotential heights at 500hPa (Michelangeli et al., 1995; Stephenson et al., 2004). First and principally studied in Winter time, when they are more stronger, 4 main states have been defined namely the positive North Atlantic Oscillation phases (NAO+), the negative NAO (NAO-), the Blocking regime and the Atlantic Ridge regime. They are well known to play an important role in creating large-scale conditions that either favor or inhibit precipitation in Europe (Plaut and Simonnet, 2001; Yiou and Nogaj, 2004), especially extreme events (Cattiaux et al., 2010; Toreti et al., 2010; Guérémy et al., 2012; Boé, 2013; Yiou and Cattiaux, 2013). These impacts can be observed in both winter and summer (Pfahl, 2014). In Europe, well identifiable spatial patterns of surface temperature and precipitation are associated with each regime. For instance, the NAO+ in winter are linked to above-normal temperature and precipitation over Northern Europe and below-normal precipitation over Southern and Central Europe (Wanner et al., 2001; Hurrell et al., 2003). Opposite results of surface temperature and precipitation anomalies are generally observed during NAO- phases. The WRs can also drive extreme events. The NAO+ regime favors heavy precipitation in northern Europe and periods of drought in the Mediterranean area. The Blocking regime determines the occurrence of dry periods in large parts of southern Scandinavia and central Europe (Yiou and Nogaj, 2004), and influences the heatwave occurrences in Russia and Northern Europe (Schaller et al., 2018). The use of WRs is also interesting since their occurrence and variability are connected to SST anomalies (Häkkinen et al., 2011; Peings and Magnusdottir, 2014; Zampieri et al., 2017) and thus somehow takes implicitly into account the Atlantic ocean influence. The practical interest in classifying large-scale geopotential anomalies into a few pre-defined patterns relies on the fact that local weather conditions depend on large-scale atmospheric flows. If WRs can be better represented and forecasted by general circulation models (GCMs), they would provide additional information for local weather anomalies via statistical downscaling techniques, which derives linkages in between large scale geopotential anomalies (i.e., the WRs) and local weather phenomena (i.e., precipitation anomalies) for medium-range lead times.

Since geopotential and temperature fields are generally better forecasted than precipitation in Numerical Weather Prediction systems (Vitart, 2014), and since long-term droughts are mainly driven by large scale forcing (Kingston et al., 2015), the benefit of using WR occurrences as predictor of meteorological drought is analysed w.r.t the precipitation forecasts. The paper is organized into six sections. The datasets and the methods are presented in section 2, the different forecast methods in section 3. The comparison of predictability scores obtained by using precipitation and MOAWRs is provided in section 4. The sources of uncertainties are then discussed in section 5 and the main conclusions are drawn in section 6.

2 Data and methods

2.1 Datasets

The observed daily cumulated precipitation data are retrieved from the Climate Assessment & Dataset (ECA&D) and the ENSEMBLES gridded dataset (E-OBS) version 12, which provides daily station-based precipitation and temperature data on regular grids (Haylock et al., 2008). While the full E-OBS resolution is 0.25 degrees, data have been here upscaled by averaging to 1 degree due to the specific focus on large-scale drought with significant socio-economic impacts. E-OBS data are available from 1950 to the present.

Atmospheric predictors are identified by using the geopotential height at 500 hPa. The daily geopotential is derived from the ERA Interim reanalysis (ERAI, Dee et al. 2011) with a spatial resolution of 1.125 degrees covering the period from 1979 to the present.

The forecast products (precipitation and 500 hPa geopotential height) are derived from the extended ensemble system of the ECMWF (ENS hereafter, Molteni et al. 1996). The ENS is the latest version of the ECMWF ensemble model and is composed of one unperturbed member and 50 perturbed members, distinguished by different initial conditions and representations of model uncertainties. In addition to these forecasts, the ECMWF produces hindcasts that are initialized the same date of the ENS for the last 20 years with five members only. In 2012, it was extended once a week to a 32-day lead time and the horizontal resolution varies from Tl639 (32 km) from t+0 to t+10days to Tl319 (64 km) from t+11 to t+32days. All of these datasets have been re-gridded onto a regular grid of 1 degree resolution based on an averaging up-scaling method. The rationale of using coarser resolution is i) to detect and so focus on larger scale precipitation deficits and ii) to take into account spatial bias in the model that could detect the right precipitation signal but with a slight spatial phasing error.

In order to build the baseline (following a normalization technique) and to have a time series long enough to calculate the scores, 21 years of hindcasts (November 1992 to November 2013) of the forecasts (November 2012 to November 2014) are used. To be coherent, the datasets of observed precipitation and the WR calculations (from ERAI) are restricted to the same period.

2.2 Weather regimes

In order to build the forecasts based on predictors (illustrated in Fig. S1 in supp. materials), the WR classification (i.e. definition of the WR patterns) is done exclusively using ERAI, by a K-means method nested within a genetic algorithm to avoid dependence on the initial conditions and the trap of the local minima (Toreti et al., 2010). The four meteorological seasons are treated independently: winter (December to February), spring (March to May), summer (June to August) and autumn (September to November) but to avoid inconsistency moving from one season to another, each season is extended by adding the last month of the previous season. This method of classification has been extensively used (Michelangeli et al., 1995; Robertson and Ghil, 1999; Santos et al., 2005), but because in this study the WR classification needs to fit with specific requests (20-year moving period of the hindcast, the four seasons), it is important to regenerate this classification. Nevertheless, the patterns of the geopotential anomalies (shown in Fig. S2 in supp. material) are strongly similar to those obtained in aforementioned

studies. The choice of using only the ERAI classification and not ENS is justified by: i) looking at previous studies that have shown the relatively similar behavior of ERAI and ENS forecasts (Ferranti et al., 2015); ii) considering that this choice avoids inconsistency (or the impossibility to derive a coherent classification) due to the continuous evolution of the ENS model. Four WRs are identified in winter and spring, while three WRs are detected in summer and autumn (see Fig. S2 in the supplementary material). The number of WRs is estimated by following Toreti et al. (2010) and depends both on the period (here from 1992 to 2013) and the region (North Atlantic) studied. This is why the number of summer WRs is different with respect to, e.g., Cassou et al. (2005).

Then, an assignation procedure is run to identify the closest WR to a given daily geopotential anomaly of ERAI and a given ENS-member. To this aim, the method proposed by Ferranti et al. (2015) is here applied. Namely, a pattern matching algorithm based on the minimum distance from the previously identified centroids is used to assign each day and individual forecast member to the closest weather regime. The climatology of the forecasted WRs is then calculated by summing the daily classification of each WR for all the members and all the days inside a 30-day window. The same climatology is derived by using ERAI. The Monthly Occurrence Anomaly of WRs (MOAWRs) is then calculated with respect to the climatological occurrences based on the hindcast period (1992-2013) and obtained by using independently both ERAI and ENS.

To potentially increase the signal emerging from the linkage between MOAWRs and precipitation anomalies, different combinations (additions and subtractions of WR occurrences) of two WRs are tested. This could be useful when two WRs have the same or opposite impacts on precipitation. For example, in the case of two regimes WRa and WRb that are associated with, respectively, dry and wet conditions over a certain region, the occurrence difference of the two WRs could be more linked to the drought events over that region (this example will be discussed later in the document). In total, a set of 6 to 12 combinations (see the list in Table 1) is tested when 3 or 4 WRs (depending the season) are detected, respectively.

2.3 Drought metrics

As suggested by the World Meteorological Organization (Svoboda et al., 2012), the Standardized Precipitation Index (SPI) is one of the most relevant indicators providing a clear and robust characterization of precipitation deficiencies and it is a good proxy for assessing meteorological droughts. The SPI calculation is relatively simple and it is performed independently at each grid point of the domain. This method is robust and has the advantage of being flexible in time, for the accumulation period studied, and in space, for the resolutions used. It also provides an unbiased product, which is important for comparing observational datasets and model simulations. In the SPI calculation, a gamma distribution is first fitted to monthly cumulated precipitation data. Then this distribution is transformed into a standard normal distribution (McKee et al., 1993, 1995). The choice of the statistical distribution has been verified in Lavaysse et al. (2015) and shown that this assumption is valid over a large proportion of Europe. Nevertheless, over the driest regions and in summer, some grid points (mainly in Spain and South Italy) the significant tests are not verified. Both the observed and modeled daily precipitation values are accumulated over a period of 30 days (i.e. we use the SPI-1, where 1 refers to the accumulation period of one month). The choice of analyzing relatively short meteorological droughts is based on two main constraints: i) a technical one connected to the limitation of the extended ENS that provides forecasts up to 33 days in the version here analyzed, and ii) the chaotic nature of the atmosphere

that limits the predictability of precipitation and geopotential forecasts after several weeks (Vigaud et al., 2018). This relative short term drought information is also relevant for users and decision makers since it provides valuable information about the onset, continuation or end of longer droughts (Svoboda et al., 2012; Stagge et al., 2015).

Based on this approach, both observed and forecasted SPI-1 values are calculated for the period 1992-2013. Here, a meteorological drought is defined as having SPI-1 values less than -1. According to the normal distribution of the SPI, this threshold corresponds to about 17.5% of the driest events. Based on Lavaysse et al. (2015), the most reliable method for producing a dichotomous forecast of drought from probabilistic forecasts of precipitation, and more specifically from the extended ENS of the ECMWF, is to predict a drought as soon as more than 40% of the ENS members are associated with a drought forecast (i.e. SPI-1< -1).

2.4 Validation tools

To assess the forecasts of drought events, traditional scores for dichotomous products are applied. These scores make use of the contingency table (Table 2) that shows the types of agreement of observed and forecasted variables.

The percentage of observed events that had been correctly forecasted are provided by the Probability Of Detection score (POD) whereas the percentage of events that had been forecasted but did not occur are indicated by the False Alarm Rate (FAR). Finally, to take into account the hits, misses and false alarms and to neglect the correct negative forecasts that will boost the scores for rare events, the Gilbert Skill Score (GSS, Jolliffe and Stephenson 2003) is used. For rare events, such as droughts, it is more relevant to use this score than the Pierce's skill score for instance. The GSS indicates how well the forecasted droughts correspond to the observed ones. This skill score is compared to the score obtained by the climatology. It is calculated as follows:

$$GSS = \frac{(hits - hits_c)}{(hits + misses + false\ alarms - hits_c)} \tag{1}$$

where $hits_c = \frac{(hits+misses)(hits+false\ alarms)}{total}$. Based on these equations, a perfect forecast achieves a score equal to 1, while a score equal to 0 is assigned to the climatology (i.e. no forecast skill). All these scores are calculated independently for each season.

3 Configuration of the drought forecasts

To forecast droughts using the MOAWRs approach, 3 steps are needed (see also Fig. S1 in the SI): 1) the WR classification, to determine the main patterns of 500hPa geopotential anomalies; 2) the daily WR attribution, to determine which is the closest WR classified previously for a forecasted (or reanalyzed) geopotential anomaly for each day and member, and the calculation of the MOAWRs; 3) the predictor assignation, to determine which WR, or combination of WRs, is the best predictor of droughts for each grid point. These different steps are now detailed.

3.1 WR classification and MOAWRs calculation

The WR classification (step 1 in Fig. S1) is detailed previously. The patterns of the geopotential height anomalies for the four seasons are displayed in Fig. S2. The best known WRs occur in winter, namely the NAO- (Fig. S2a), the NAO+ (Fig. S2e), the Atlantic Ridge (Fig. S2i) and the Blocking regime (Fig. S2m). Once the WR classification done, the closest WR to the daily geopotential of each ENS-member is attributed for both ERAI and ENS. From daily attribution, the climatology and the anomalies of occurrence of each WR can be done (illustrated by the step 2 in Fig. S1) using ERAI or ENS depending the configuration of the forecast experiment (see list in Table 3). These two steps allow producing of daily attribution of WRs and then the MOAWRs for both the ERAI and ENS datasets.

3.2 Assignation of predictors

The objective of this step is to identify the best SPI-1-predictor within the three or four WRs identified for each season and their 6 to 12 possible combinations (step 3 in Fig. S1). This is done by using the temporal correlation between the MOAWRs (deriving from ERAI or ENS depending the forecast experiment) and the SPI-1 for each grid point and each MOAWRs. This allows highlighting the large-scale impacts of the WRs on precipitation (see Fig. S3 and comments in SI). An automatic attribution is then applied based on the maximum of the absolute values of the correlations. The sign of the correlation is recorded to keep track of the type of linkage. An example of linkage over Scandinavia in provided in SI (see SI and Fig. S4).

3.3 Forecast configurations

The potential benefits and limitations of using these predictors are assessed thanks to 5 different drought forecasting approaches. These approaches are differentiated by the methodologies employed for the three steps listed previously and are illustrated in Fig. S1 and summarized in Table 3.

The first method of drought forecasting, called 'Reference', is based on forecasted precipitation (Lavaysse et al., 2015). The skill scores of this forecasting approach are here used as benchmark.

The second method of forecasting, called 'Idealized' (red arrows in Fig. S1), uses exclusively the MOAWRs derived from ERAI and does not take into account the uncertainties related to the forecasts of WRs. In this method, the assignation of predictors is based on a best correlation between MOAWRs from ERAI and SPI-1 observed. It is interesting to note that following this approach, the large majority of SPI-1 in Europe is associated with a combination of WRs (Fig. 1). The highest absolute values of these correlations show significant spatial differences (Fig. 2). Throughout the year, there are generally higher linkages in northern Europe than in southern Europe. There is also a strong seasonal difference. In winter, the mean correlation is about 0.55 whereas it is about 0.28 in summer. The origin of precipitation, which is more synoptically-driven in winter and more local in summer, can explain these results.

The third forecasting method, called 'Operational', computes MOAWRs derived from the ENS forecasts, but the WR classification (definition of the different regimes) and the WR assignation (best WR over each grid point) are still derived from ERAI and observed SPI-1 (see Table 3 and green arrows in Fig S1). The advantage of this method is a real assessment of the

model ability to forecast both the WRs and the relationship between the SPI and the WRs. As this assignation is constant (i.e. derived from ERAI and observations), it is also easier to set up operationally and there is no problem when the version of the operational ENS model changes. The disadvantage is the non-optimization of the forecast, i.e. there is no correction in case of bias in the forecasted WRs.

5 The fourth forecasting method, called 'Optimized' (see Table 3 and blue arrows in Fig. S1), is relatively close to the previous one but uses a different assignation procedure. It is defined as the best correlation between the observed SPI-1 and the forecasted MOAWRs by the ENS as predictor (instead of those derived from ERAI). This method derives the best relationships between forecasted MOAWRs and observed precipitation, and by definition will obtain the best scores even if the WRs are not correctly forecasted. This methodology tends then to optimize the forecasts by correcting some bias in the forecasted MOAWRs. The

10 correlation values for observed SPI with MOAWRs derived from ENS for winter are provided in SI (Fig. S5) and depicts patterns very similar to those obtained with ERAI (Fig. S3), illustrating the good representation of the linkages in ENS.

Finally the fifth forecasting method, called 'Process' (purple arrows in Fig. S1 and table 3), can be used to investigate the skill of the model in representing observed processes. Except for the WR classification procedures, which is still derived from ERAI, this method uses ENS forecasts. The forecasted MOAWRs are then linked to the forecasted SPI-1 (instead of observed SPI

15 for the other configurations) in the assignation procedure. This configuration allows to analyze the modeled linkage between MOAWRs and precipitation, that will be compared to the observed ones provided by the 'Idealized' configuration.

4 Results

4.1 Skill scores

The skill scores of the forecasted precipitation, called 'Reference', is used as benchmark. It is derived from Lavaysse et al.

20 (2015), where a drought is forecasted when at least 40% of members forecasts $SPI < -1$. The best achieved performance (for winter in Central Europe) shows how slightly more than 40% of the observed drought events are correctly predicted with a 30-day lead time (Fig. 3a, d, g, j) with about 60% of false alarms (Fig. 3b, e, h, k). For both POD and FAR, the spatial variability is small (standard deviation lower than 0.2) especially during spring and fall. In winter, high scores can be noticed in Germany, Poland, Spain and Norway; whereas in summer, drought seems to be more predictable in eastern Europe. When the

25 POD and the FAR are combined in the integrated GSS (Fig. 3c, f, i, l), higher seasonal and spatial differences appear. Overall, the score reaches up to 0.3 in winter, especially in northern Germany [50deg.N, 10deg.E]; while the worst value is reached in spring and summer, especially in western Europe (France, Belgium). Due to the impacts of the local forcing on precipitation, the drought forecasts based on large-scale predictors are better in continental than in coastal regions (more details in Lavaysse et al. (2015)).

30 The forecasts using predictors, using the 'Operational' forecast, can now be assessed with respect to the 'Reference'. In order to detect the same number of drought events when using the predictors and the precipitation, the threshold of the MOAWRs is chosen equal to 0.176 (0.824 for negative correlations). The POD, FAR and GSS anomalies w.r.t. the 'Reference' forecast (i.e. Fig. 3) for the four seasons are shown in Fig. 4. This is done using 20 years with a leave-one-out technique, which is a

cross-validation method for small samples sizes enabling to validate results by simply partitioning the series into a training and a test part. The 'Operational' forecast is more spatially variable. As for winter in the northern part of Europe, this forecast is significantly better in terms of both POD and GSS; whereas in central Europe the 'Reference' forecast is more reliable. Despite the fact that the patterns are less homogeneous for the other seasons, some positive impacts of this 'Operational' forecast appear, e.g. in: northern Russia in spring, western Europe in summer, central Europe during fall. The same results have been plotted in SI (Fig. S6) for the 'Optimized' forecasts and do not show significant differences with Fig. 4.

These results are consistent with the intensity of the linkage measured during the assignation procedure between the SPI and the WRs (Fig. 2) and highlight the regions where the large-scale atmospheric patterns associated with the WRs could better explain strong precipitation deficits when compared to local drivers (e.g. orography, soil moisture, coastline).

4.2 Intensity and initial conditions

To better understand the potential performance of the approach, sensitivity tests are conducted. In the previous section, the SPI-1 intensity threshold to define a drought was fixed to -1. The previously used skill scores are here derived for SPI lower than -1.5 and -2 (respectively $\sim 7\%$ and $\sim 2.5\%$ of the most extreme cases). A second sensitivity test is done on the initial conditions influencing all results but also bringing useful information on drought onset and persistence. Most of the studies on drought focuses on 3-month (or longer) cumulated precipitation that could have more severe impacts on, e.g., agriculture and water resources. Due to the unpredictable nature of the weather and the limitation of the lead time of the ENS model, the assessment of drought forecasting is limited to 1-month lead time in our study. Nevertheless, the information of the two previous months (observed SPI-2 with a threshold defined as -1) is taken into account to measure the impacts of these initial conditions and the ability to forecast drought persistence/onset.

In Fig. 5 the GSS scores in winter for the whole domain shown in the previous Figures are synthesized by using boxplots. The results shown in Fig. 3a and Fig. 4a are represented by the black boxplots plotted for SPI < -1 in Fig. 5a and b respectively. Overall, the predictability decreases with the drought intensity. In winter, dry initial conditions generate a favorable environment to better forecast droughts. In other words, the persistence of drought is better predicted than the onset. Finally, the last main result concerns the improvement of the 'Operational' forecasts. For the SPI lower than -1, all GSS values shown in Fig. 5 are quite close. But, as also highlighted by Fig. 4, there is a larger spatial variability with the MOAWRs approach. For more intense droughts, there is a global and significant improvement by using 'Operational' forecast. Indeed for drought intensities with SPI lower than -2, the median of the GSS scores goes up from close to 0 (using the precipitation-based method) to 0.05 (using the MOAWRs).

The same sensitivity tests are conducted for the others seasons (Figs. S7, S8 and S9 in SI), and the decrease of predictability with increasing drought intensity is found for all of them. Nevertheless, the conclusions on the role of the initial conditions depend on the season. For instance in summer, drought onsets are slightly better predicted than drought persistence. The reason could be the higher temporal variability of the monthly precipitation deficits in summer than in winter due to the larger impact of local forcings. Finally in all the seasons, the use of atmospheric predictors (i.e. 'Operational' forecast) leads to better performance when looking at the most extreme events (SPI < -2).

5 Sources of uncertainty

To better discuss and understand the results and their uncertainties, additional tests are here reported. The main objective is to quantify the contribution of the uncertainties in WR predictions and the linkage between the SPI and the WRs.

5.1 Validation of the WR forecasts

The first question to address is about the quality of the forecasts of MOAWRs. The purpose here is not to provide a complete evaluation of the WRs forecasts, that has been already studied (Ferranti et al., 2015; Matsueda and Palmer, 2015), but to focus on errors that could impact the drought forecasts.

To validate the forecast of the WRs, first the comparison of the frequency of occurrence of each daily WR is performed (Fig. 6). To do so, the climatology of the total occurrence of each WR among all the members and the entire lead time (5 members∗ 30-day LT) is calculated to extract the monthly anomalies. The forecasted anomalies are divided by the number of ensemble members to create comparable results with the data provided by ERAI. The WR-distributions as given by the forecasts are characterized by a higher degree of similarity than the ones given by ERAI, with a peak of occurrence at around 5-8 days in winter (blue bars, Fig. 6). The same holds for the other seasons (not shown). The lower spread of the forecasted WR occurrences, associated with reduced tails (i.e. reduced occurrences for durations exceeding 20 days), could be explained by the underestimation of the long-term persistence of regimes. A further comparison of the MOAWRs from ERAI and ENS (scatter plots in Fig. 7) suggests that : i) the distribution of forecasted drought occurrences could be explained by the overestimation of low occurrences using ENS than the reanalysis (i.e., larger number of forecasted events compared to those derived from ERAI with durations shorter than 5 days) and, ii) the underestimation of longer duration events (i.e., lower events with durations longer than 15 days using ENS than ERAI, red dotted lines in Fig. 7). Despite this behavior, the correlations appear significant with a maximum of 0.65 for the WRa (significance with 90% of confidence at 0.58). These significant scores are obtained in winter, while for the other seasons the correlations are lower (see Table 4). In summer, they are not significant for two-thirds of the WRs.

5.2 Strength of MOAWR-precipitation linkage

According to the previous subsection, the WR forecast could be improved. Thus, it is important to assess the limitation of the method using predictors and so assessing the strength of the MOAWR and precipitation linkage. To this aim, the procedure 'Idealized' forecasts of MOAWRs, i.e. geopotential anomalies provided by ERAI, without uncertainties on forecast, are compared to the forecast of precipitation discussed and shown in Fig. 3.

The POD scores are strongly improved between seasons and regions (Fig. 8a, d, g, and j). These results are strongly connected to the correlation values obtained and shown in Fig. 2 with the same North-South and seasonal variabilities being observed. However, almost all the northern part of Europe show a better POD with the 'Idealized' than 'Reference' forecasts. Up to 70% of observed drought events are correctly detected during winter. This percentage falls to about 17.5% in summer (i.e., the climatological value) in the southern part of the domain. The results in terms of FAR are more variable depending on

both the season and the region. On average, there is a small decrease of the FAR. However, the GSS shows a clear and significant improvement in the drought forecast when using the WR predictors. Compared to the scores using 'Operational' forecasts in Fig. 4, the bigger difference is more in terms of magnitudes than in the spatial distribution. For instance, in winter a high improvement is observed in northern Europe (up to 0.2 for 'Idealized' against 0.1 for 'Operational' forecast over Scandinavia)

5 whereas a decrease/close score is obtained in central Europe. Based on this sensitive analysis, the linkage between the SPI-1<-1 and the MOAWRs is strong enough to provide significant improvements of the prediction scores in most of the regions. Nevertheless, this analysis also highlights the limitations of the methods used in this study when and where the influence of the WR on drought is lower (e.g. Germany and Poland in winter; eastern Europe in summer; southern Europe in fall).

5.3   Modeled linkage

10 Some additional tests are also conducted on the predictor assignation procedures (definition of the best predictant for SPI-1 < -1 at each grid point) to see the impacts of using either ERAI or ENS (the latter could potentially correct bias of the ENS). This is done by the 'Optimized' forecast. Due to the errors associated with the WR forecasts, the procedures using WRs from ERAI or ENS provide different results (Fig. 9a compared to Fig. 1a). The assignation patterns done by using ERAI (Fig. 1a), for the 'Operational' and 'Idealized' forecasts, have less homogeneous large-scale structures (i.e., more spatial variability) than the

15 'Optimized' forecast (Fig. 9a) showing a more complex linkages using ERAI than ENS. Nevertheless over continental regions, there are similarities (impact of WRs b, b-a, a, c-d), illustrating the relative good representation of the impacts of specific WR on precipitation by ENS. The correlations between the WRs forecasted and the observed precipitation are then plotted (Fig. 9b). The correlation values, which can be compared to the correlation shown in Fig. 2a, are low as a result of the relatively low predictability of the WRs previously discussed. The values are also sensitive to the strength of the linkage between WRs and

20 precipitation (i.e. highest scores in southern Norway and the northern part of the U.K, lowest scores in Central Europe.).

The last analysis is focused on the modeled linkage between the SPI and the WRs both provided by the ENS (i.e. using the 'Process' forecast, Fig. 9c and 9d). It is remarkable the great similarities of the maps for the assignations and the observed linkage by using 'Process' or the 'Idealized' forecasts (correlation values greater than 0.65, Fig. 1a and 2a). This is especially true over the U.K, Ireland, Scandinavia, Spain and north-western Russia. Despite some differences observed in southern France and Italy where 'Process' overestimates the large-scale forcing on precipitation (i.e. with stronger correlation with WRs than

25 observed), the patterns obtained when comparing the correlation values between SPI-1 (observed or forecasted) and MOAWRs (from ERAI or ENS) in Figs. S3, S5 and S10 are very similar. This highlights the overall good representation by ENS of the processes linking large scale circulation and local precipitation deficits. So the ENS model succeeds in capturing the impacts of the WR occurrences on the precipitation anomalies as shown with observations and ERAI over a large part of Europe. These

30 results could suggest limitations in using such predictors as the lack of skill score could result from a failure in forecasting the large-scale atmospheric circulation rather than from a misrepresentation of the physical processes from the large scale forcing to local weather.

6 Conclusions

In this study, a drought forecasting method based on large-scale atmospheric predictors is proposed in order to improve the early warning of atmospheric drought events. The method is based on the Monthly Occurrence Anomalies of Weather Regimes (MOAWRs) within a 30-day lead-time. The methodology used to select the predictors is based on a three step procedure.

First, WRs (described by daily 500 hPa geopotential anomalies) are identified by using a Genetic K-means algorithm for each season separately and for both ERAI and extended ENS forecasts. The climatological occurrences are calculated for each WR. The identified three/four WRs (depending on the season) are combined (added or subtracted) with each other to enhance the potential signal of their impacts. Second, the MOAWRs is used as a predictor of meteorological droughts at each grid point. The predictor assignation procedure is based on the correlation between the MOAWRs and the SPI-1. To select the best predictor, the MOAWR associated with the strongest absolute value of correlation is selected. The last step involves the forecasting of the SPI-1 lower than -1. Two approaches are derived and compared. The first one is based on the index developed by Lavaysse et al. (2015) for drought events and derived from the forecasted precipitation provided by the ENS and called 'Reference'. This represents a benchmark for the early warning of drought forecasting. At most around 40% of drought events are detected one month in advance with 65% of false alarms. The second forecasting approach, called 'Operational', is based on MOAWRs. In the north-eastern parts of the European continent, an improvement of the Gilbert Skill Score (GSS) is observed using 'Operational' w.r.t. the 'Reference'. Nevertheless, this is balanced by other regions where the forecast skills is clearly lower (central Europe in winter, eastern Europe in summer) than the reference. The origin of this spatial and temporal variability in the skill scores are linked to the atmospheric dynamic associated with the precipitation. In winter, precipitation is much more related to large-scale atmospheric forcing mainly captured by the MOAWRs. On the contrary in summer, precipitation is more affected by local forcings that could influence, for instance, the trajectory and the occurrence of convective systems. In this study, this behavior is captured by the better correlation between MOAWRs and precipitation in winter than in summer. The spatially variable skill scores are mainly controlled by the intensity of the linkage between the MOAWRs and SPI-1. Due to the distance of the geopotential anomalies to some target regions, or because of some local effects that could be predominant to the large scale forcing, the impacts of these MOAWRs on precipitation could be low, as observed in winter over central Europe. According to these scores, the most reliable forecast could result from choosing the best method for each grid point independently. The influence of the initial conditions and the intensity of the drought highlight i) the losses of predictability with increasing drought intensity and ii) the better scores in predicting persistency rather than the onset of drought especially in winter. Also, the benefits of using the WRs to predict droughts appear to be more important when the most intense droughts (i.e. SPI < -2) are forecasted.

This study shows the importance of improving the prediction of the WR occurrences. The methodology applied here could be compared to more complex methodologies using clustering of the members to define the most probable scenario, or by taking into account the transition between WRs. Future work should also take into account the uncertainties in WR prediction, as also suggested by Magnusson (2017); Weisheimer et al. (2017). Recent studies (Matsueda and Palmer, 2014, 2015; Vigaud et al., 2018) have shown that WR prediction is still a big challenge for lead-times greater than 15 days. Some improvements

could be also done by using a multi-model ensemble such as the one recently developed in the framework of the Seasonal to Sub-Seasonal (S2S) project (Vitart et al., 2016). Finally, the physical drivers should be analysed in detail to better understand why the predictors are more useful when predicting the most extreme events.

Most of the weather services provide new forecasts up to several months. For users, it appears essential to scientifically and statistically evaluate the added values of these forecasts for specific extreme events such as meteorological droughts. This is the main objective of this study. Nevertheless, evaluating the practical usefulness of this operational forecast is difficult without taking into account the costs for each case of the contingency table (hits, misses, false alarms) that strongly vary depending on their applications (i.e. civil protection, water management services, farmers' decision supporting systems, etc.). The statement provided in this study is based on statistical scores independent of these costs. According to the GSS, there is a significant improvement of using forecasts in relation to the climatology. Moreover, the forecasts using predictors generate, in some regions and some seasons, significant improvements of these forecasts by using the same score. To evaluate these improvement for specific users, the costs should be taken into account that is a major perspective of this study.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This study has been carried out with the support of ECMWF, which provided the forecasted precipitation and 500hPa geopotential heights. The authors would like to thank the two anonymous reviewers for their fruitful comments.

References

Boé, J.: Modulation of soil moisture–precipitation interactions over France by large scale circulation, Climate dynamics, 40, 875–892, 2013.

Cassou, C., Terray, L., and Phillips, A. S.: Tropical Atlantic influence on European heat waves, Journal of climate, 18, 2805–2811, 2005.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F.: Winter 2010 in Europe: a cold extreme in a warming climate, Geophysical Research Letters, 37, 2010.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Ferranti, L., Corti, S., and Janousek, M.: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector, Quarterly Journal of the Royal Meteorological Society, 141, 916–924, 2015.

Ferranti, L., Magnusson, L., Vitart, F., and Richardson, D. S.: How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe?, Quarterly Journal of the Royal Meteorological Society, 2018.

Guérémy, J.-F., Laanaia, N., and Céron, J.-P.: Seasonal forecast of French Mediterranean heavy precipitating events linked to weather regimes, Natural Hazards and Earth System Science, 12, 2389–2398, 2012.

Haiden, T., Janousek, M., Bidlot, J., Ferranti, L., Prates, F., Vitart, F., Bauer, P., and Richardson, D.: Evaluation of ECMWF forecasts, including 2016-2017 upgrades, ECMWF - Technical Memorandum, 2017.

Häkkinen, S., Rhines, P. B., and Worthen, D. L.: Atmospheric blocking and Atlantic multidecadal ocean variability, Science, 334, 655–659, 2011.

Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M.: An overview of the North Atlantic Oscillation, Wiley Online Library, 2003.

Jolliffe, I. and Stephenson, D.: Forecast Verification: A Practitioner's Guide in Atmospheric Science, 2003.

Kingston, D. G., Stagge, J. H., Tallaksen, L. M., and Hannah, D. M.: European-scale drought: understanding connections between atmospheric circulation and meteorological drought indices, Journal of Climate, 28, 505–516, 2015.

Lavaysse, C., Vogt, J., and Pappenberger, F.: Early warning of drought in Europe using the monthly ensemble system from ECMWF, Hydrology and Earth System Sciences Discussions, 12, 1973–2009, 2015.

Lavers, D. A., Pappenberger, F., Richardson, D. S., and Zsoter, E.: ECMWF Extreme Forecast Index for water vapor transport: A forecast tool for atmospheric rivers and extreme precipitation, Geophysical Research Letters, 43, 2016a.

Lavers, D. A., Waliser, D. E., Ralph, F. M., and Dettinger, M. D.: Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western US extreme precipitation and flooding, Geophysical Research Letters, 43, 2275–2282, 2016b.

Magnusson, L.: Diagnostic methods for understanding the origin of forecast errors, Quarterly Journal of the Royal Meteorological Society, 143, 2129–2142, 2017.

Matsueda, M. and Palmer, T.: Predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts, in: EGU General Assembly Conference Abstracts, vol. 16, p. 12945, 2014.

Matsueda, M. and Palmer, T.: Predictability of summertime Euro-Atlantic weather regimes in medium-range forecasts, in: EGU General Assembly Conference Abstracts, vol. 17, p. 11809, 2015.

McKee, T. B., Doesken, N. J., Kleist, J., et al.: The relationship of drought frequency and duration to time scales, in: Proceedings of the 8th Conference on Applied Climatology, vol. 17, pp. 179–183, American Meteorological Society Boston, MA, USA, 1993.

McKee, T. B., Doesken, N. J., and Kleist, J.: Drought monitoring with multiple time scales, in: Proceedings of the 9th Conference on Applied Climatology, pp. 233–236, American Meteorological Society Dallas, Boston, MA, 1995.

5 Michelangeli, P.-A., Vautard, R., and Legras, B.: Weather regimes: Recurrence and quasi stationarity, Journal of the Atmospheric Sciences, 52, 1237–1256, 1995.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The ECMWF ensemble prediction system: Methodology and validation, Quarterly Journal of the Royal Meteorological Society, 122, 73–119, 1996.

Peings, Y. and Magnusdottir, G.: Forcing of the wintertime atmospheric circulation by the multidecadal fluctuations of the North Atlantic
10   ocean, Environmental Research Letters, 9, 034 018, 2014.

Pfahl, S.: Characterising the relationship between weather extremes in Europe and synoptic circulation features, Natural Hazards and Earth System Science, 14, 1461–1475, 2014.

Plaut, G. and Simonnet, E.: Large-scale circulation classification, weather regimes, and local climate over France, the Alps and western Europe, Climate Research, 17, 303–324, 2001.

15 Robertson, A. W. and Ghil, M.: Large-scale weather regimes and local climate over the western United States, Journal of Climate, 12, 1796–1813, 1999.

Santos, J., Corte-Real, J., and Leite, S.: Weather regimes and their connection to the winter rainfall in Portugal, International Journal of Climatology, 25, 33–50, 2005.

Schaller, N., Sillmann, J., Anstey, J., Fischer, E., Grams, C., and Russo, S.: Influence of blocking on Northern European and Western Russian
20   heatwaves in large climate model ensembles, Environmental Research Letters, 13, 054 015, 2018.

Stagge, J. H., Kohn, I., Tallaksen, L. M., and Stahl, K.: Modeling drought impact occurrence based on meteorological drought indices in Europe, Journal of Hydrology, 530, 37–50, 2015.

Stephenson, D., Hannachi, A., and O'Neill, A.: On the existence of multiple climate regimes, Quarterly Journal of the Royal Meteorological Society, 130, 583–605, 2004.

25 Svoboda, M., Hayes, M., , and Hood, D.: Standardized Precipitation Index User Guide, World Meteorological Organization, pp. Report WMO–No. 1090, Geneva, 2012.

Toreti, A., Xoplaki, E., Maraun, D., Kuglitsch, F., Wanner, H., Luterbacher, J., et al.: Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns, Nat. Hazards Earth Syst. Sci, 10, 1037–1050, 2010.

30 Vigaud, N., Robertson, A. W., and Tippett, M. K.: Multimodel ensembling of subseasonal precipitation forecasts over North America, Monthly Weather Review, 145, 3913–3928, 2017.

Vigaud, N., Robertson, A. W., and Tippett, M. K.: Predictability of recurrent weather regimes over North America during winter from submonthly reforecasts, Monthly Weather Review, 146, 2559–2577, 2018.

Vitart, F.: Evolution of ECMWF sub-seasonal forecast skill scores, Quarterly Journal of the Royal Meteorological Society, 140, 1889–1899,
35   2014.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., et al.: The sub-seasonal to seasonal prediction (S2S) project database, Bulletin of the American Meteorological Society, 2016.

Wanner, H., Brönnimann, S., Casty, C., Gyalistras, D., Luterbacher, J., Schmutz, C., Stephenson, D. B., and Xoplaki, E.: North Atlantic Oscillation–concepts and studies, Surveys in geophysics, 22, 321–381, 2001.

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

5   Quarterly Journal of the Royal Meteorological Society, 143, 917–926, 2017.

Yiou, P. and Cattiaux, J.: Contribution of atmospheric circulation to wet north European summer precipitation of 2012, Bulletin of the American Meteorological Society, 94, S39, 2013.

Yiou, P. and Nogaj, M.: Extreme climatic events and weather regimes over the North Atlantic: When and where?, Geophysical Research Letters, 31, 2004.

10   Zampieri, M., Toreti, A., Schindler, A., Scoccimarro, E., and Gualdi, S.: Atlantic multi-decadal oscillation influence on weather regimes over Europe and the Mediterranean in spring and summer, Global and Planetary Change, 151, 92–100, 2017.
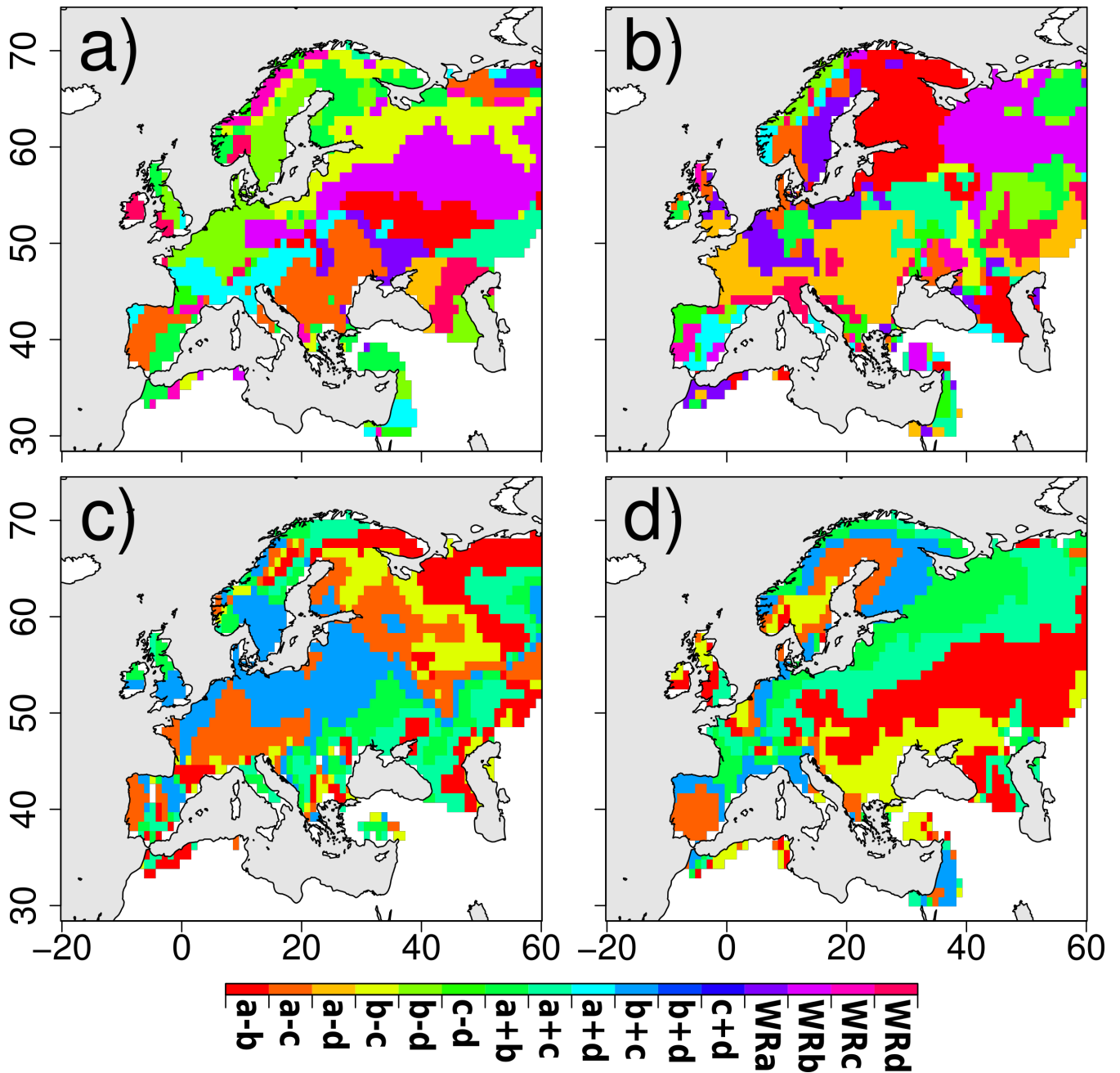
Figure 1. Automatic attribution of the best predictors in winter (a), spring (b), summer (c), and autumn (d) based on the occurrence anomalies of WRs of ERAI and the observed precipitation (used for the 'Operational' and 'Idealized' forecasts). The names of the predictors are indicated in the color scale.
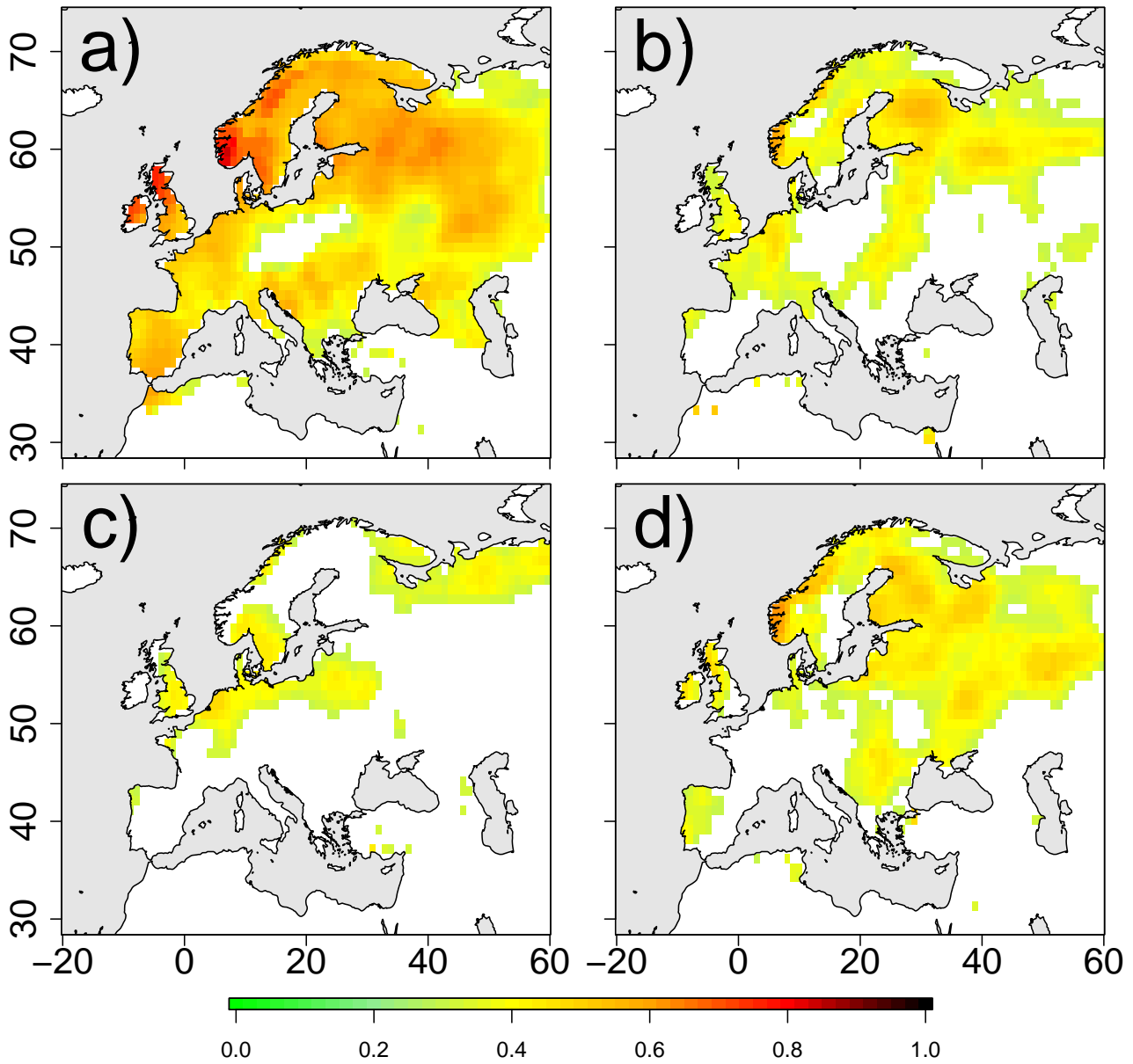
Figure 2. Absolute values of temporal correlation between SPI-1 and MOAWR derived from ERAI (used for the 'Operational' and 'Idealized' forecasts) attributed from the 16 combinations in winter (a), spring (b), summer (c) and autumn (d). Only values with a confidence level larger than 90% are plotted.
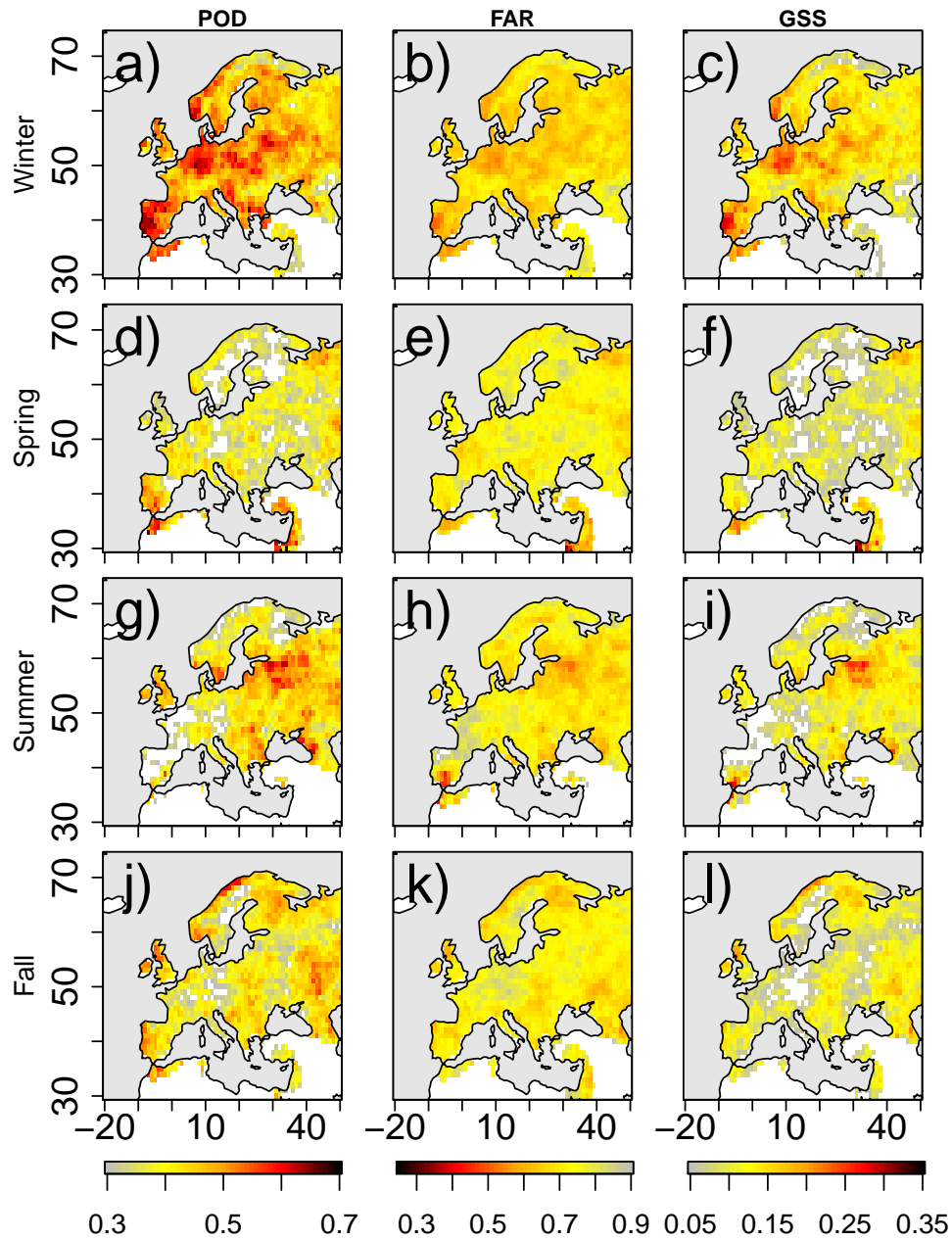
Figure 3. POD (left), FAR (centre, with reverse colours) and GSS (right panels) scores of droughts prediction calculated using the 'Reference' forecast. The scores are calculated for (from top to bottom) winter (DJF, first), spring (MAM, second), summer (JJA, third) and autumn (SON, fourth line).
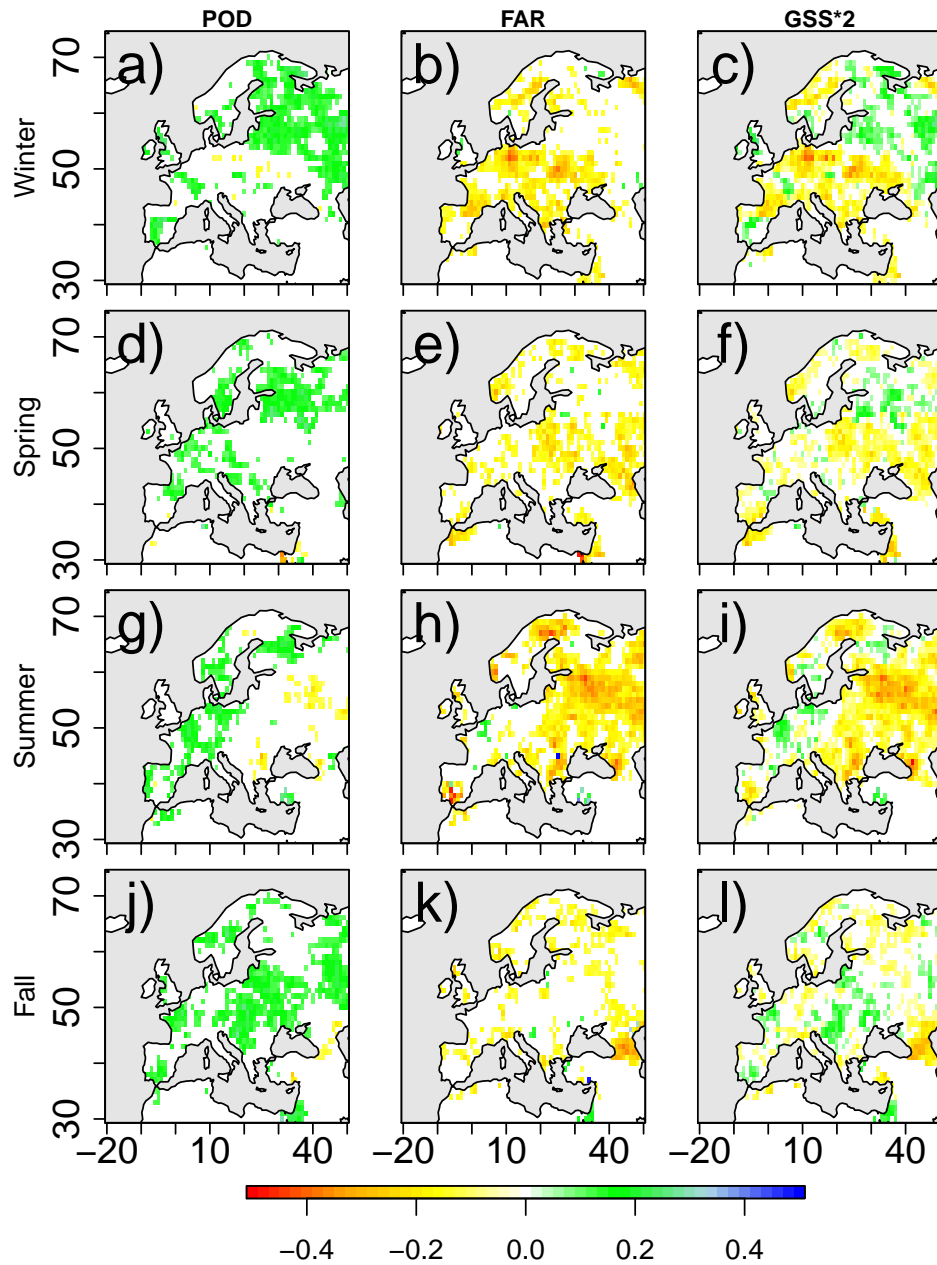
Figure 4. Anomalies of POD (left), FAR (centre) and GSS*2 (right panels) scores of drought prediction using the 'Operational' w.r.t. the 'Reference' forecast. The scores are calculated for (from top to bottom) winter (first), spring (second), summer (third) and autumn (fourth line). Improvement scores by using the predictors are indicated in green (inverse scale for FAR). Only difference with confidence interval larger than 90% are plotted. GSS is multiplied by 2 to use the same scale as the other metrics.
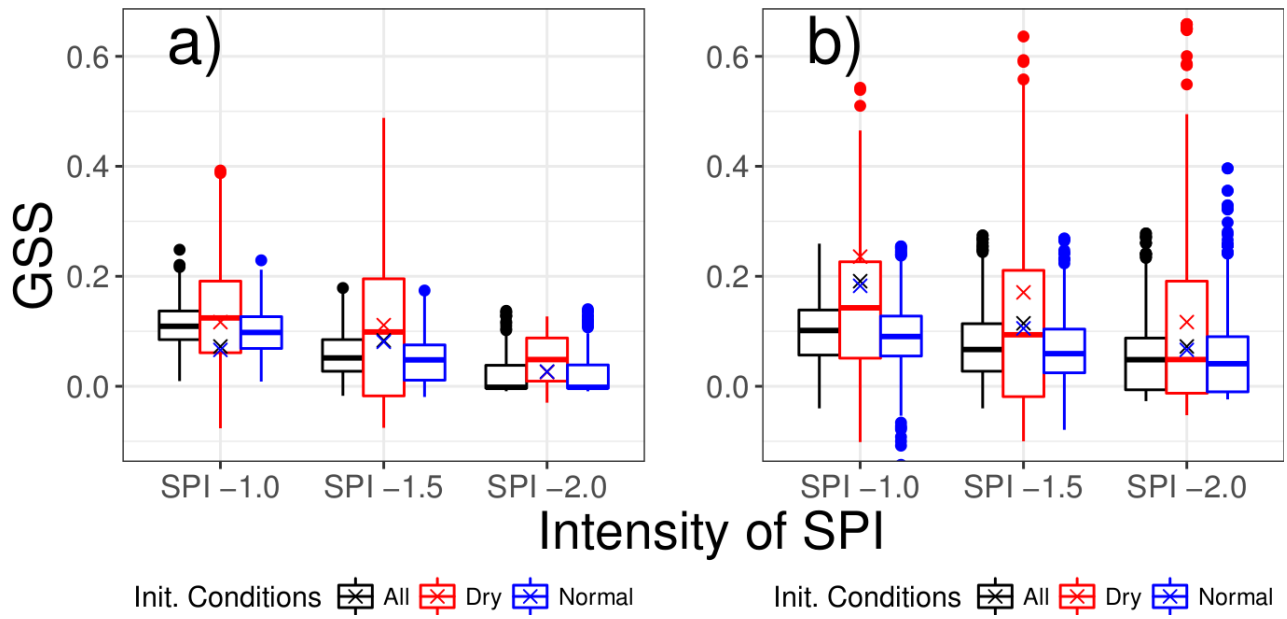
Figure 5. Boxplot of the GSS scores in winter using the 'Reference' forecast (a) and the 'Operational' forecast (b). The scores are calculated over the entire domain and the boxes display the spatial variability. The scores are depending to the SPI intensities (-1, -1.5 and -2, x-axis) and the initial conditions defined by the previous observed SPI-2 conditions (see text for more details). Crosses indicate the scores but calculated by merging all the grid cells.
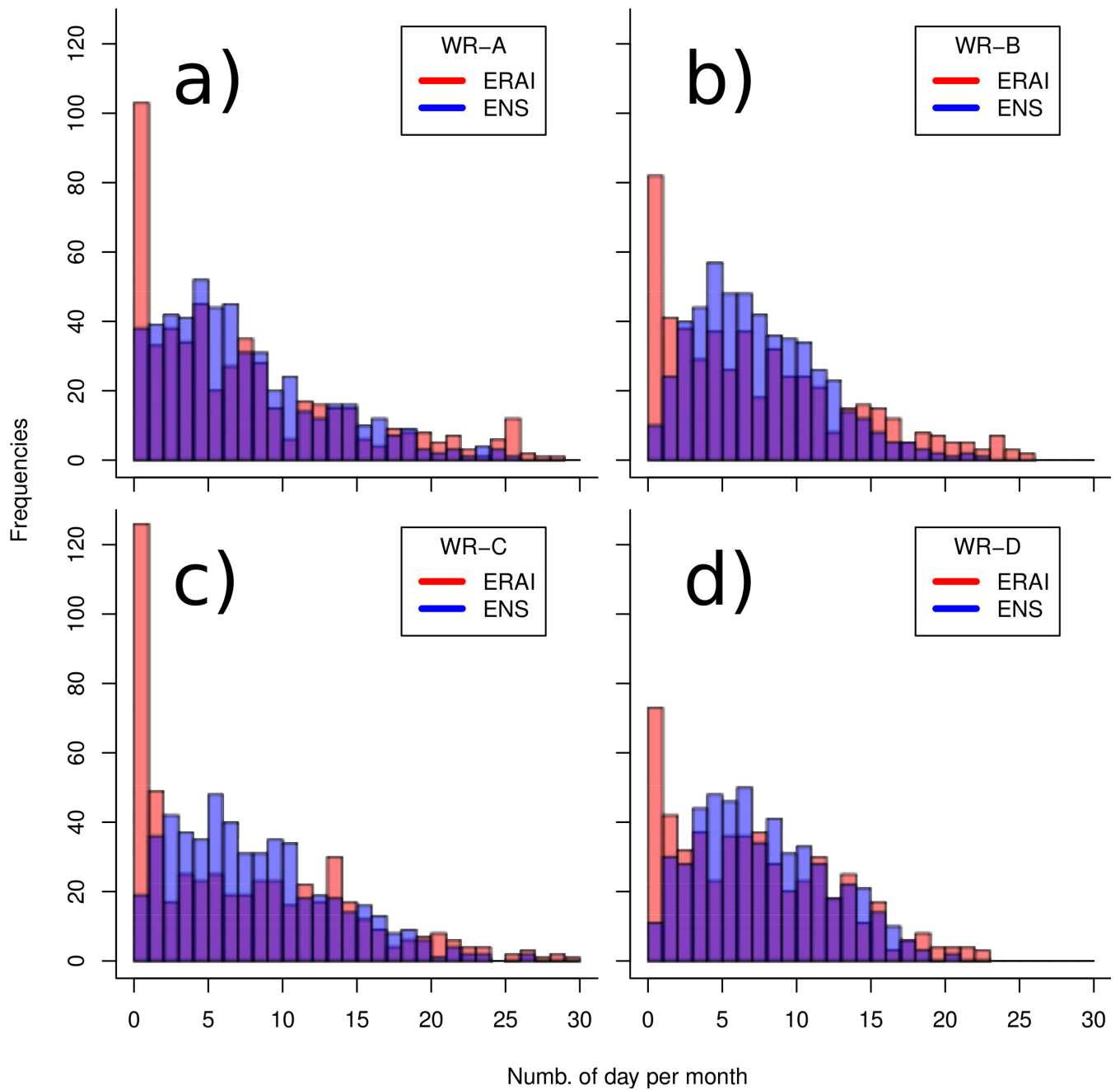
Figure 6. Example of frequency distribution of WR occurrences (in days per 30-day windows) in winter for WR-A (a), WR-B (b), WR-C (c) and WR-D (d) using ERAI and ENS (red and blue bars respectively, purple when the two overlap).
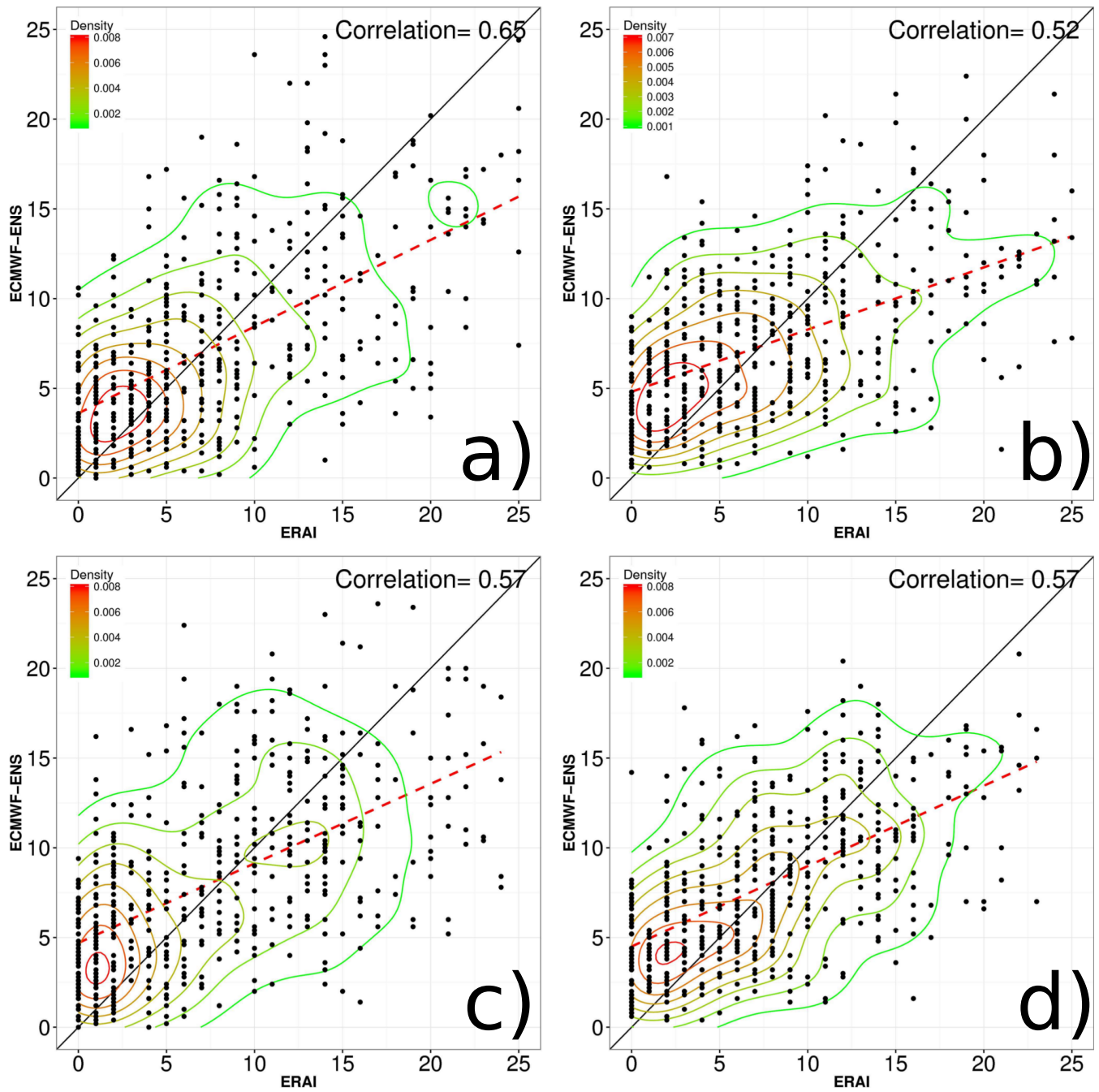
Figure 7. Scatter plots of the occurrence of the four winter WRs provided by ERAI (x axis) and provided by ENS (y axis). The linear least square regression are indicated with red dashed lines and the correspondent correlation on the top right of each panel.
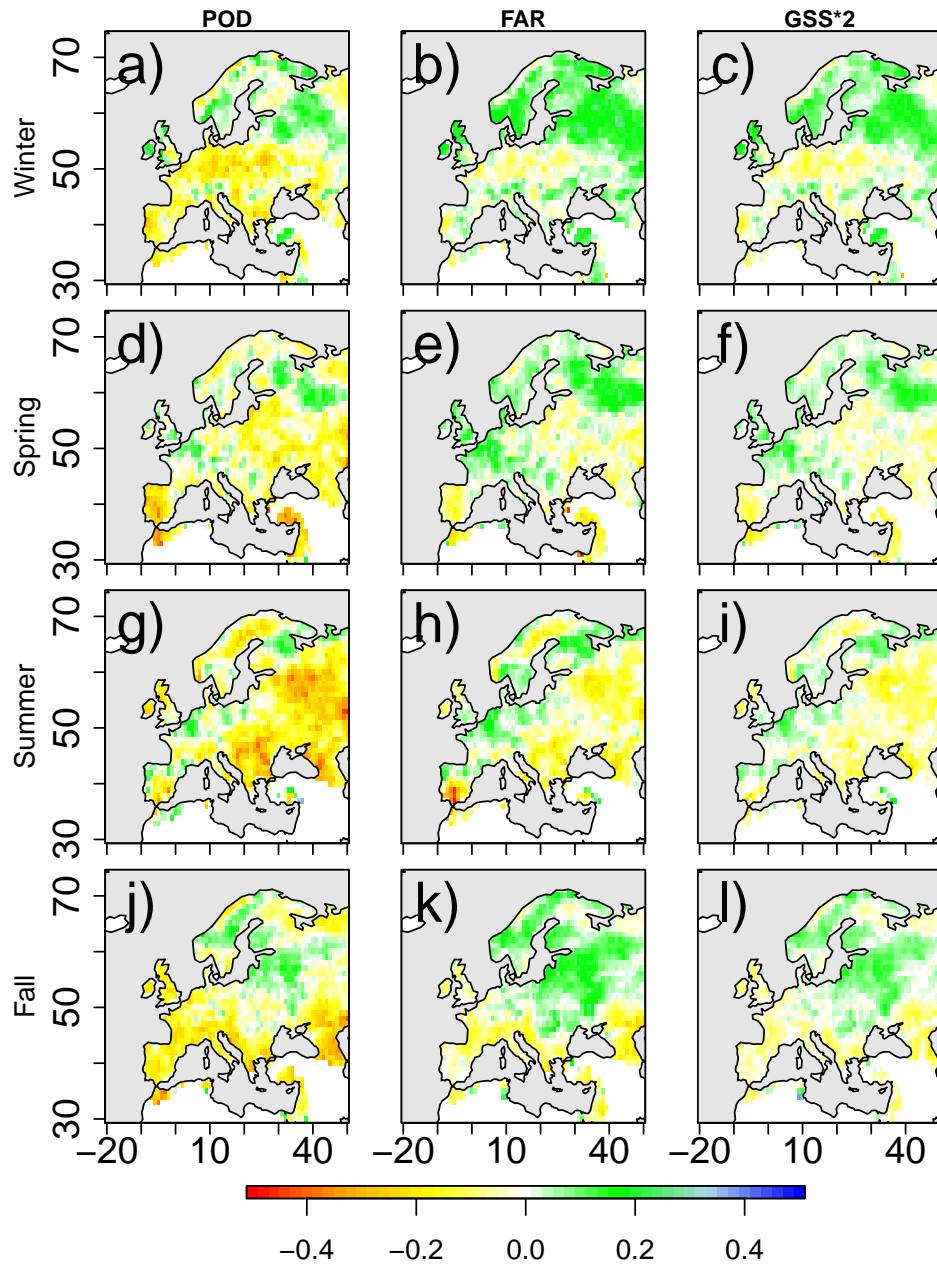
Figure 8. Anomalies of POD (left panels), FAR (middle panels) and GSS*2 (right panels) of the drought prediction based on the 'Idealized' forecasts w.r.t. the 'Reference' forecast, in winter (first), spring (second), summer (third) and fall (fourth line). Improvement scores by using the predictors are indicated in green (inverse scale for FAR). Only difference with confidence interval larger than 90% are plotted. GSS is multiplied by 2 to use the same scale as the other metrics.
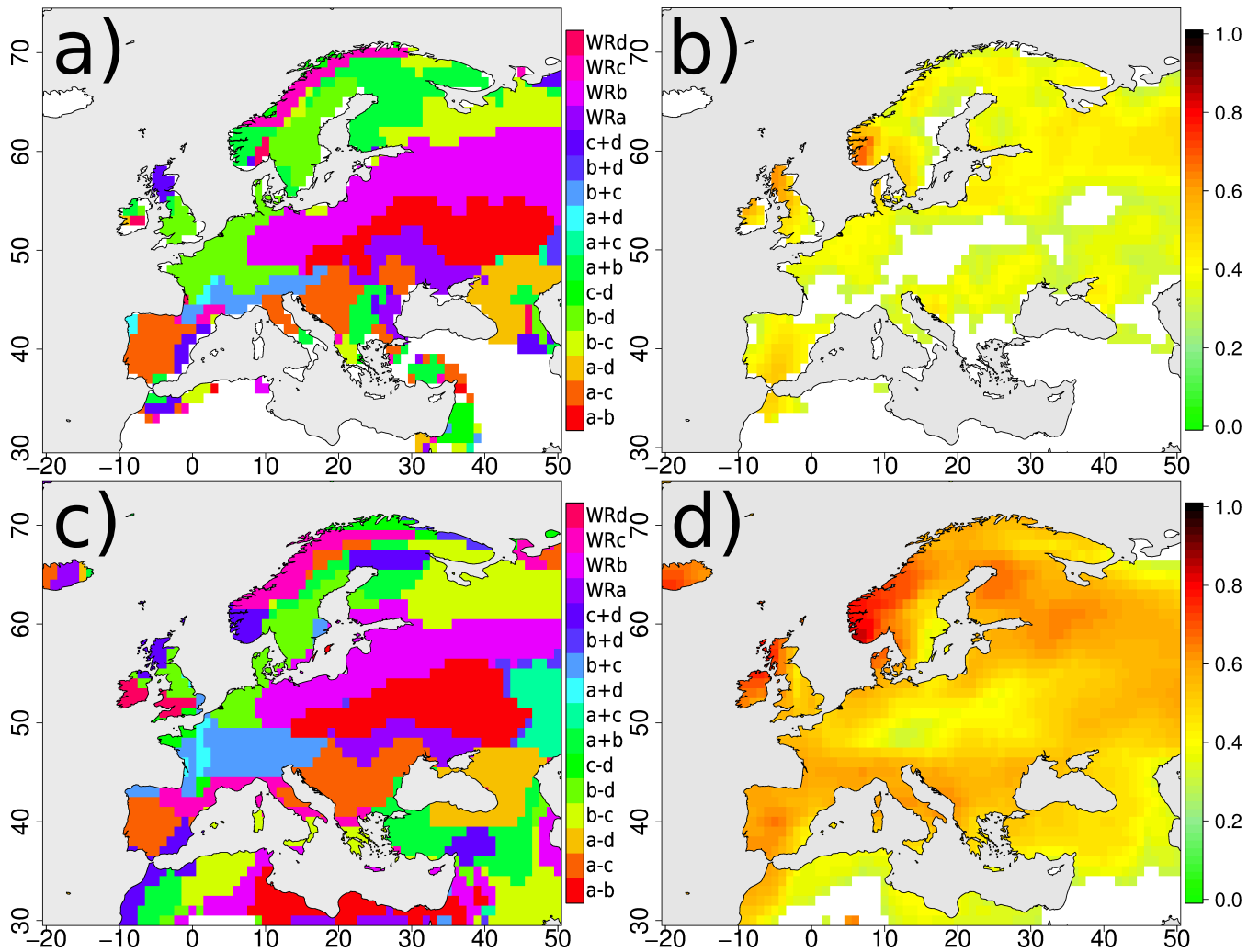
Figure 9. Assigned winter WR (a) and associated absolute correlation values (b) for the 'Optimized' Forecast (i.e. predictors defined using MOAWRs from ENS and observed SPI-1). c) and d), same as a) and b) for the 'Process' Forecast (i.e. predictors defined using MOAWRs from ENS and forecasted SPI-1).

WR	A	B	C	D
A	a	a+b	a+c	a+d*
B	a-b	b	b+c	b+d*
C	a-c	b-c	c	c+d*
D	a-d*	b-d*	c-d*	d*

Table 1. Definition of WRs and WR combinations. WR combinations are defined as either additions or subtractions of monthly WR frequencies. * indicate regimes that exist only in winter and spring when 4 WRs are detected.

		Event observed	
		Yes	No
Event	Yes	hits	false alarms
Forecasted	No	misses	correct negative

Table 2. Contingency table of dichotomous events illustrating the four types of classification between observed and forecasted events.

Name	WR classification	MOAWR predictor	WR for assignation	SPI for assignation
Reference	no predictor (Precip. forecast)	-	-	-
Idealized	ERAI	ERAI	ERAI	Observed
Operational	ERAI	ENS	ERAI	Observed
Optimized	ERAI	ENS	ENS	Observed
Process	ERAI	ENS	ENS	Forecasted (ENS)

Table 3. Definition of the five sets of forecasts compared in that study. The differences are based on the use of predictor or not, the use of predictors derived from reanalyse or forecasted and for the assignation procedure, the use of observed or forecasted SPI.

Season	WR	Correlation	Season	WR	Correlation
Winter	A	**0.65**	Spring	A	**0.48**
	B	**0.52**		B	0.43
	C	**0.57**		C	**0.47**
	D	**0.57**		D	**0.62**
Summer	A	0.45	Autumn	A	**0.51**
	B	0.42		B	**0.47**
	C	**0.47**		C	**0.47**

Table 4. Correlation values between the forecasted and observed MOAWR for each WRs and the four seasons. Values indicated in bold have a significance level above 0.9.