European Commission

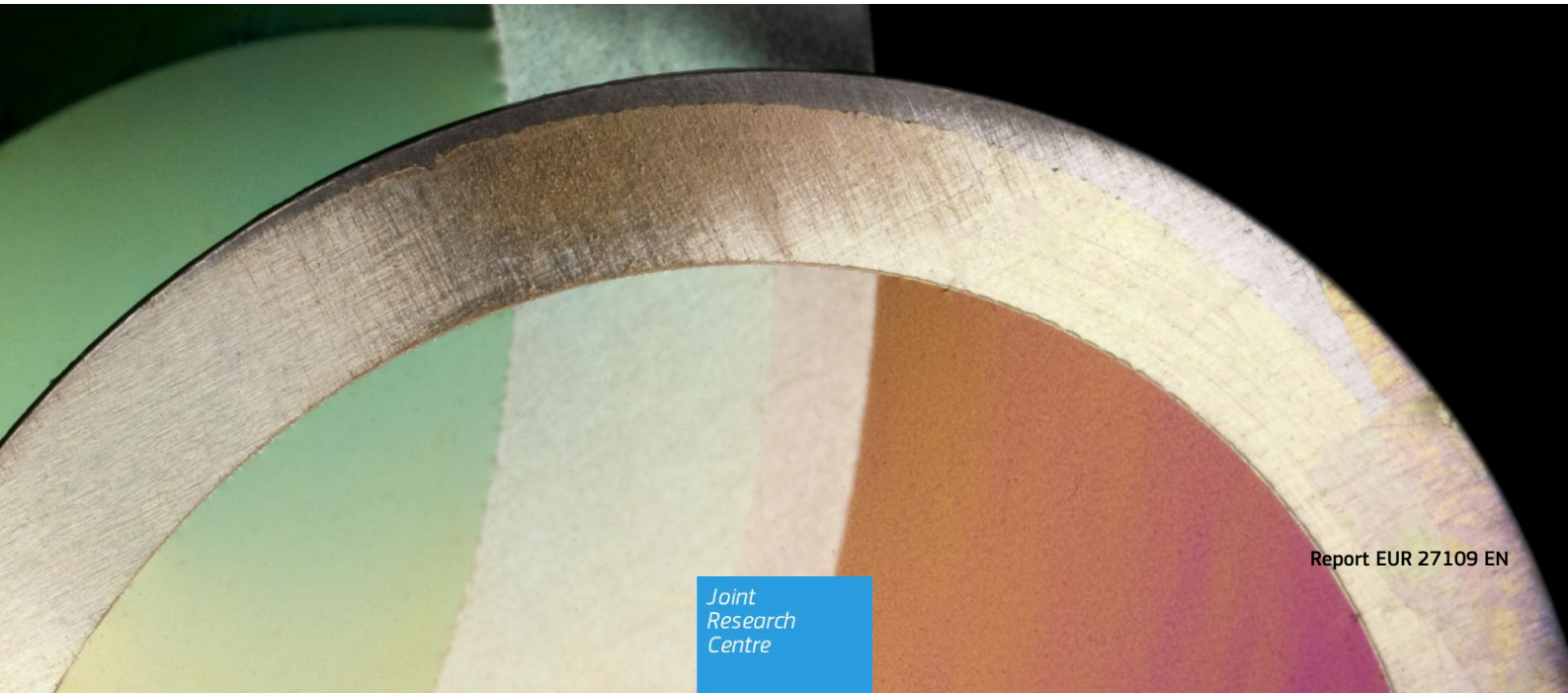
JRC SCIENCE AND POLICY REPORT

Early warning of drought in Europe

On the use of probabilistic forecasts to produce drought early

Christophe Lavaysse

2015



Report EUR 27109 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Christophe Lavaysse
Address: Joint Research Centre, via Enrico Fermi 2749, 21027 Ispra (VA), ITALY
E-mail: christophe.lavaysse@jrc.ec.europa.eu
Tel.: +39 0332 78 3640

JRC Science Hub
https://ec.europa.eu/jrc

JRC94423

EUR 27109 EN

ISBN 978-92-79-45597-1 (PDF)

ISSN 1831-9424 (online)

doi: 10.2788/093828

Luxembourg: Publications Office of the European Union, 2015

Abstract

The performances of the first month of the ECMWF probabilistic extended forecast and the seasonal forecast to predict droughts over Europe are compared. The Standardized Precipitation Index is used to quantify droughts.
It can be shown that on average the extended forecast has higher skill than the seasonal forecast whilst both outperform climatology. No significant spatial or temporal patterns can be observed but the scores are improved when we focus on large scale droughts.
This report further analyses several different methods to convert the probabilistic forecasts of SPI into a Boolean drought warning. It can be demonstrated that methodologies which convert low percentiles of the forecasted cumulative distribution function of SPI into warnings are superior in comparison to alternatives such as the mean or the median of the ensemble. This work demonstrates that around 40% of droughts in Europe are correctly forecasted one month in advance.
Nevertheless, due to the lack of the significant difference between the distributions of the ensemble members for false alarms or misses on one hand side and correct forecasts on the other hand side, it is not yet possible to quantify the uncertainty of the drought forecasts.

Contents

1 Introduction

Droughts impact many human activities and environmental processes. They often spread over vast geographical regions and last for many months or even years (Lloyd-Hughes and Saunders, 2002). This makes them one of the costliest of all natural disasters (Below et al., 2007). Droughts significantly impact economic sectors, such as agricultural activities or water resources management, especially in vulnerable areas (Fraser et al., 2013). In particular, forecasts are needed to anticipate droughts and mitigate their effects. Decision makers and end users require simple and robust forecast indices which can detect the onset, maintenance and end of the drought conditions. Droughts can be classified under several categories (Wilhite and Glantz, 1985): (i) hydrological drought, which is associated with the effects of periods of precipitation deficits on surface or subsurface water supply; (ii) agricultural drought, which links meteorological (or hydrological) drought to agricultural impacts, focusing on the plant water stress, and (iii) meteorological drought, which is defined as a large-scale and prolonged rainfall deficit over one or several months.

In this study, we will focus on meteorological droughts based on monthly precipitation. This timescale is a key challenge because it is considered to be a difficult time range, which falls between medium-range forecasting (which is strongly related to initial conditions) and the seasonal time-scale (which is mainly driven by oceanic variables) (Vitart, 2014).

Meteorological droughts can be analysed using the Standardized Precipitation Index (SPI, McKee et al. (1993), recommended by WMO (2012)), which is a normalised quantification of the precipitation deficit (Vicente-Serrano, 2006; Dutra et al., 2013).

It has been demonstrated that droughts can be forecasted using stochastic or neural networks (Kim and Valdés, 2003; Mishra et al., 2007). These forecasts can provide "reasonably good agreement for forecasting with 1-2 month lead time" (Mishra and Desai, 2005). But a large part of these studies do not compare the score of these forecasts with the forecasts provided by the precipitation fields of the probabilistic models. Forecasts of droughts can also be produced using Numerical Weather Prediction Models (Dutra et al., 2013, 2014). Such forecasts are highly uncertain due to the chaotic nature of the atmosphere, which is particularly strong at the sub-seasonal time scale (Stockdale et al., 1998; Vitart, 2014). Therefore, ensemble prediction systems are developed which forecast multiple scenarios of future weather conditions. Probabilistic forecasts become particularly important in assessing the risks associated with high-impact and rare weather events such as tropical cyclones (Hamill et al., 2012), or for identifying the uncertainties of

the forecasts (Buizza et al., 2005). Forecasts on the subseasonal and seasonal timescales using dynamic models have evolved considerably over recent years, and demonstrate potential usefulness for predicting large-scale features and teleconnections (Barnston et al., 2012; Arribas et al., 2011). The latter can be used in statistical downscaling methods using weather types, for example, Eshel et al. (2000) used the North Atlantic sea level pressure precursors to forecast drought over the eastern Mediterranean region. However, even though, the forecasts were statistically significant, with a lead time of several months, they were made for a limited region that is one of the most sensitive to weather types in Europe. even if the forecasts are statistically significant for a lead time of several months, the region studied is restricted and is one of the most sensitive to weather types in Europe.

Numerical Weather Predictions also produce forecasts of precipitation. In general, the published literature indicates that the accuracy of these forecasts of precipitation in Europe is low skill of these precipitation fields over Europe is low (Richardson et al., 2013; Weisheimer and Palmer, 2014) although there are considerable spatial variations. However, such analysis tends to be performed from the viewpoint of a meteorologist and does not incorporate the non-linear transformations or specific properties that are relevant for drought forecasting. Drought forecasts can be based on different lead times ranging from a few weeks to several months, and the accuracy of any forecast will decrease with increasing lead time. The monthly time scale poses a particular challenge because it represents the transition between medium-range forecasting (up to 14 days), which is strongly related to initial conditions, and seasonal forecasting, which is largely driven by oceanic variability (Vitart, 2014).

The European Centre for Medium-Range Weather Forecasts (ECMWF) provides two different types of forecasts for this time range: an extended forecast with a lead time of up to 32 days, which is issued twice a week, and a seasonal forecast with a lead time of up to 12 months, issued once a month. The extended range forecast incorporates more recent model developments and is usually more accurate (Vitart et al., 2008). Moreover, the seasonal forecasting system is based on an older model cycle (Molteni et al., 2011). In order to exploit such methodologies, one needs to understand and analyse the property and skill differences between the two systems in the context of the particular application. Such an analysis must be performed not only based on the numerical skill of forecasting droughts, but also within the context of the binary decision (drought forecasted or not) to issue drought warnings. The latter poses a particularly challenge if such decisions are based on probabilistic forecasts.

The objectives of this report are to analyse the predictability of monthly

drought forecasts based on Numerical Weather Predictions and the Standardized Precipitation Index (SPI). The extended range and seasonal forecasting systems will be compared directly and within the context of decision-making frameworks. Multiple scores and multiple methodologies which allow probabilistic forecasts to be transformed into binary decisions will be developed and tested.

The main underlying issues are: what is the predictability of a drought occurrence based on the SPI-1, which is the most useful model - the Seasonal (SEAS) or the monthly ENSemble system (ENS) for the 30-day cumulative precipitation, and what are the spatial and temporal variabilities of each model? Adapted skill scores will provide information about the ability of the probabilistic models to accurately forecast such kinds of extreme events.

The report is organised as follows: the datasets and method are presented in section 1; the tools and methods used are explained in section 2; the results are discussed in section 3; and Conclusions are drawn in section 4.

2 Data and methods

2.1 Precipitation

Observations

This study used a gridded precipitation dataset from the ENSEMBLES project and ECA & D (Haylock et al. 2008; Van den Besselaar et al. 2011, E-OBS Version 5), which includes continuously updated data from 1950 onwards. As this analysis focuses on large-scale droughts, the spatial resolution of this dataset (0.25 degrees) was upscaled by averaging the cumulative precipitation to a 1-degree grid.

The data was validated by Pereira et al. (2013); Sunyer et al. (2013) who found that datasets from ECA & D show higher values for extreme precipitation, and E-OBS tends to smooth the data too much. This can generate problems when analysing intense precipitation events, but appears to be of secondary import for drought analysis. Daily precipitation values were aggregated to monthly accumulates to be able to compare with monthly forecasts. These data are available from 1950 and are regularly updated. Nevertheless, to be consistent with the data provided by the ensembles from the ECMWF, the hindcast period of 1992 to 2013 was used to calculate and analyse the precipitation anomalies.

Table 1: ENS and SEAS configurations for the hindcast and the forecast periods.

Periods	Evaluation Period	ENS	SEAS
Hindcasts	11/1992 to 10/2012	5 members	15/51 members
Forecasts	01/11/2012 to 31/10/2013	51 members	51 members

Forecasts

Two sets of coupled ensemble forecasting systems are provided by the ECMWF to forecast one month ahead: an extended-range monthly forecast and a seasonal forecast (Table 1).

The results of the ECMWF monthly (32-day) extended-range ensemble forecasting system (Vitart, 2004), hereafter ENS) have been issued twice a week since October 2011. This model is the latest version of the ECMWF Integrated Forecasting System. For lead times up to day 10 the model is not coupled to the ocean and has a resolution of about 32 km (T639). It is forced by persistent sea-surface temperature anomalies. While beyond a lead time of 10 days the resolution of the model is coarser (T319, 64 km), it is coupled to an ocean model. The vertical resolution remains unchanged during the entire simulation at 62 vertical levels. The ECMWF provides a back statistic (Hindcasts) for ENS which is a 5-member ensemble starting on the same day and month as each Thursday's real time forecast for each of the past 20 years. For a more detailed description see Vitart (2014).

The second ECMWF ensemble system used in this study is a seasonal forecast called System 4 (Molteni et al. (2011), hereafter SEAS) that is launched on the first day of each month. It has lead times up to 13 months and a resolution of T255 (80 km). This model is the 2011 version of the Integrated Forecast System with 91 vertical levels. SEAS provides a back statistic, which is a 15/51 member ensemble (the number depends on the month) identical to SEAS for every month from 1980 onwards. In this study only the first forecast month was used.

SEAS and ENS are composed of 50 members, which are generated by perturbing initial conditions and physical tendencies (Molteni et al., 1996; Weisheimer et al., 2014)) and one unperturbed member. Both datasets were re-gridded to a one-degree resolution using a mass conservative interpolation. The two systems were compared over their hindcast periods and a forecast period, as can be seen in Table 1. This allows for a larger sample size and a more significant comparison.

However, while this technique is robust and commonly used, it has a few disadvantages: there are only five members in the ensemble of the reforecasts,

compared to 51 members used for the realtime forecasts. The ensemble size can have an impact on skill scores, which needs to be corrected for. Weigel et al. (2008) faced the same issue when they scored the ECMWF reforecasts produced in 2006 - they used a correction of the probabilistic skill score which takes into account the ensemble size.

2.2 Drought detection

In this study the Standardized Precipitation Index (SPI) is used to detect droughts solely based on precipitation data. The SPI was developed by McKee et al. (1993) and is currently used in scientific studies and operational systems (Guttman, 1999; Khan et al., 2008; Dutra et al., 2013, 2014). The SPI has the advantage that it is very simple to use and provides information about precipitation anomalies. It is also very flexible, allowing calculations to be aggregated over different spatial scales (from station data to large-scale areas) as well as temporal domains (from 10 daysâĂŹ to several monthsâĂŹ cumulative precipitation, Mishra and Desai (2006); Cacciamani et al. (2007)).

As this study focuses on the monthly timescale, the SPI was calculated using monthly accumulated precipitation data (SPI-1). The SPI is usually computed by fitting a probability density function (often a Gamma distribution) to the data (Lloyd-Hughes and Saunders, 2002; Edossa et al., 2010; Dutra et al., 2013; Guy Merlin and Kamga, 2014) as illustrated in Figure 1. Through the application of an inverse normal (Gaussian) function, data are transformed into normal space with a mean of 0 and a standard deviation of 1. The hypothesis that the data can be approximated by a Gamma distribution must be tested to ensure that all conclusions are valid, by comparing the reconstructed distribution to the empirical one. The Gamma function cannot be fitted when only a few data points (events) or very low data values (precipitation) exist, because in such cases numerical convergence of the optimisation process cannot be achieved. Therefore, the SPI methodology cannot be applied in very arid regions.

The method of the SPI was performed for each grid point of the domain and built from forecasts and hindcasts based on the ECMWF system, as shown in Figure 2

The SPI value can be broken down into different classes (WMO, 2012): normal conditions from -1 to 1; moderate drought with SPI $<$ -1; severe drought with SPI $<$ -1.5; and extreme drought with SPI $<$ -2. The time series of the analysed forecasts in this report are too short to justify focusing on an SPI lower than -2 (last 2.3% of the distribution). This is illustrated in Figure 3, which shows the significant spatial variability of drought occurrence using this threshold. Based on the method used, this occurrence should be

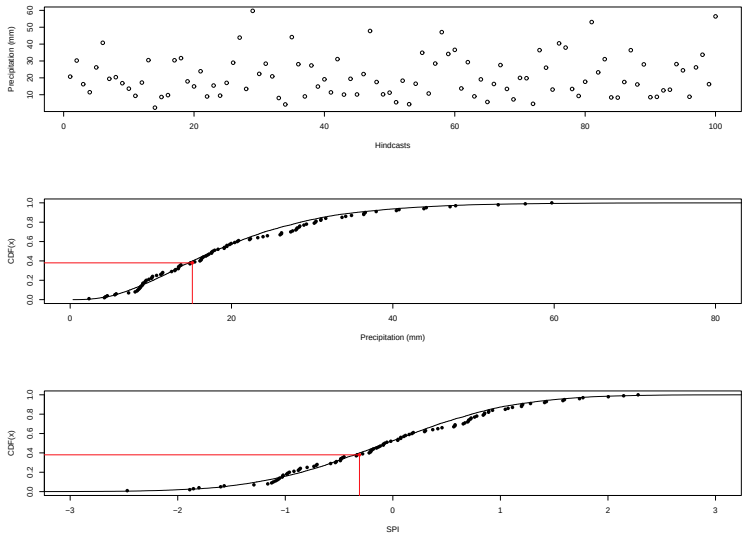


Figure 1: Three steps of the SPI calculation: i) monthly cumulated precipitation; ii) empirical CDF and fitting of the Gamma distribution; iii) transformation into a normal CDF with mean=0 and SD=1. Red points indicate an example of an SPI forecast of about 15 mm that becomes an SPI-1 of -0.27.

equal to 2.3% over Europe. Therefore, this study focuses on moderate and severe droughts only.

One strong advantage of this method is that it produces an unbiased product with a homogeneous rank histogram (Talagrand Diagram) of the observed precipitation ranked onto the forecasted precipitation (Figure 4).

2.3 Deriving decision support from probabilistic forecasts

One of the main objectives of this work is to provide decision makers and end users with a simple and robust boolean index to forecast the occurrence of drought based on a probabilistic forecasting system. It is therefore important to select appropriate tools to characterise the quality of the forecasts. It is difficult to classify these forecasts using a very simple score due to the three dimensions of the forecasts, as illustrated in Figure 5. Several methods for selecting a boolean solution were tested and compared to a deterministic model (defined here as the unperturbed member of the Ensemble). A comparison was also made with a climatological forecast. Methods to derive this index are given in Table 2 and can be categorised into three types: in-

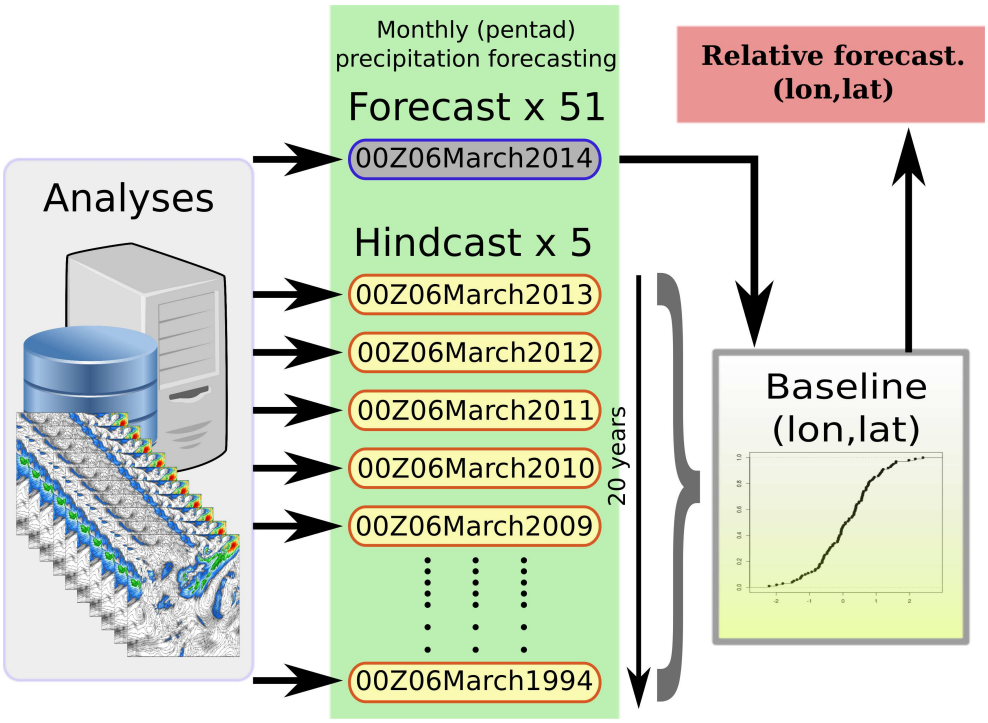


Figure 2: Methodology to generate SPI with the ENS operational model. The hindcasts generate the baseline and the forecast is placed in this distribution. This allows for the provision of an SPI index relative to the grid cell or station studied.

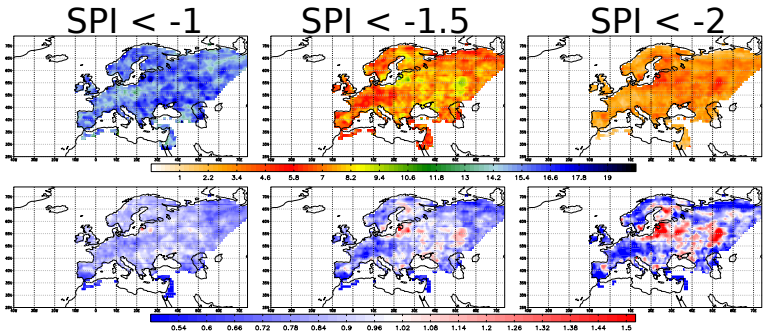


Figure 3: Drought occurrence forecasts (as a percentage - top panels, or relative to the theoretical distribution - bottom panels) calculated using different thresholds, SPI-1 $<$ -1 (left panels), SPI-1 $<$ -1.5 (central panels), and SPI-1 $<$ -2 (right panels)

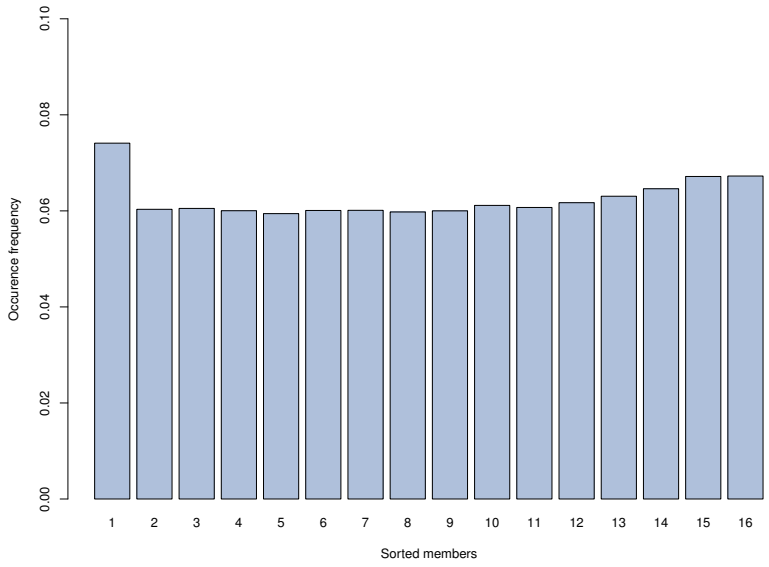


Figure 4: Rank diagram of observed SPI-1 into the sorted SPI-1 forecasted for the period of hindcast using SEAS.

dividual, where the index is based on an individual member or percentile; partially integrative, where the sum of individual members or percentiles are used; and integrative, which is represented by the ensemble mean. The individual types should be seen as providing complementary information about the intensity of the SPI-1 and the distribution of the members.

The individual types have been subdivided into five classes representing a strong dry member (Q13), a strong wet member, (Q88) or the median. The extreme members of the distribution are not used, so as to avoid the outliers which are generally associated with ensemble systems (Lavaysse et al., 2013).

A threshold was defined for each method. An SPI less than -1 (-1.5) will select 16% (6.7%, respectively) of the normalised series. Therefore, to be coherent, the thresholds were defined in such a way as to select the same number of events for all the methods.

2.4 Evaluation scores

There are a plethora of scores to evaluate probabilistic forecasts (Nurmi, 2003). In this study, we have chosen only those scores which are suitable for drought forecasting.

The Relative Operating Characteristic (ROC) score was proposed by Mason (1982) to plot the false alarm rate against the hit rate. The objective of the ROC score is to calculate the ability of the forecast to discriminate

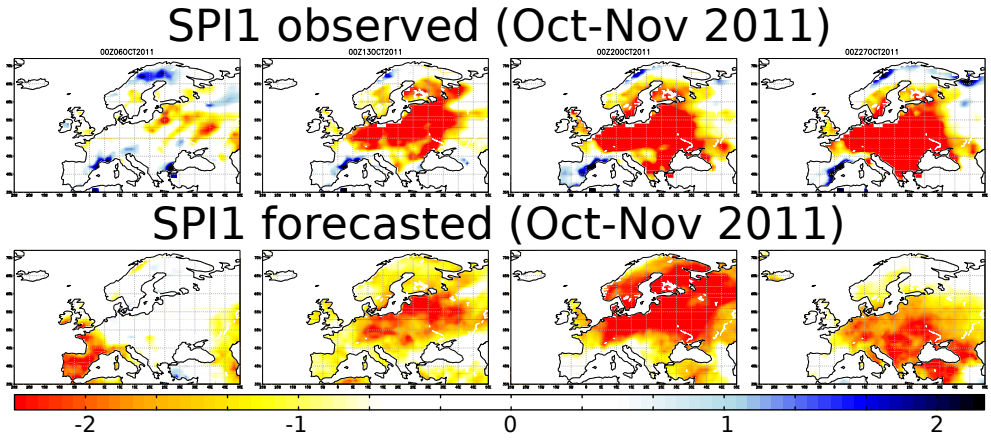


Figure 5: Example of temporal evolution of four consecutive weeks of SPI-1 observed (top panels) and forecasted using ENS (bottom panels) during the period from October 2011 to November 2011.

Table 2: List of the 10 methods used to provide a boolean index for drought forecasting using an ensemble system

Name	Definition
13 percentile (Q13)	member located at the 13% of the CDF
23 percentile (Q23)	member located at the 23% of the CDF
median (MED)	member located at the 50% of the CDF
77 percentile (Q77)	member located at the 77% of the CDF
88 percentile (Q88)	member located at the 88% of the CDF
Large spread (SpL)	sum of the extreme members (Q13 + Q88)
low spread (Spl)	sum of the members (Q23 + Q78)
Dry spread (SpD)	sum of the dry members (Q13 + Q23)
Flood spread (SpF)	sum of the wet members (Q77 + Q88)
Mean	ensemble mean

between events and non-events. It is not bias-sensitive to the forecast and can be considered as a potentially useful measure because it is conditioned by observations (i.e. given that a drought occurred, what was the corresponding forecast?). The area under the ROC curve can be calculated, with a range between 0 and 1. Higher numbers indicate a better forecast.

The reliability diagram, which is conditioned on the forecasts, is a good complementary score to the ROC because it assesses the average agreement between the forecast values and the observed values. In a reliability diagram the forecast probability is plotted against the observed relative frequency (Nurmi, 2003). A perfect score is associated with the 1:1 line. The climatology score (i.e. no resolution) corresponds to the mean observed frequency (i.e. observed relative frequency of y=0.159 for SPI $<$ -1).

The accuracy of the probability forecasts is assessed using the Brier Skill Score (Brier, 1950). A skill score can be derived by comparing the Brier Skill Score to climatology. The Brier Skill Score ranges from -infinity to 1. The higher the score the more skillful the forecast, and any negative values indicate that the climatological forecast outperforms the probabilistic forecast.

The abovementioned scores are complemented by the correlation of the ensemble mean and the Root Mean Squared Error of the ensemble mean, which are frequently used in the evaluation of seasonal forecasts.

Several scores deal with the contingency table. Using this representation, both the forecasted and observed solutions are booleans. In this study, we have used five such scores. The Probability Of Detection (POD, perfect = 1) is the ratio of the observed to the forecasted events. The False Alarm Rate (FAR, perfect = 0) is the fraction of the forecasted events that did not actually occur. The extreme dependency score (EDS) integrates the POD and the FAR (Ferro and Stephenson, 2011). Finally, the Gilbert score balances the POD and correct percentage of cases (Jolliffe and Stephenson, 2003; Hogan et al., 2010), and measures the fraction of observed and/or forecasted events that were correctly predicted, adjusted for hits associated with random chance.

3 Results

3.1 Evaluation of the SPI calculation

The sensitive part of the SPI calculation is the step of the fitting of a distribution to the empirical distribution. In this study, the Gamma distribution is fitted to the probability density function of monthly precipitation. It is

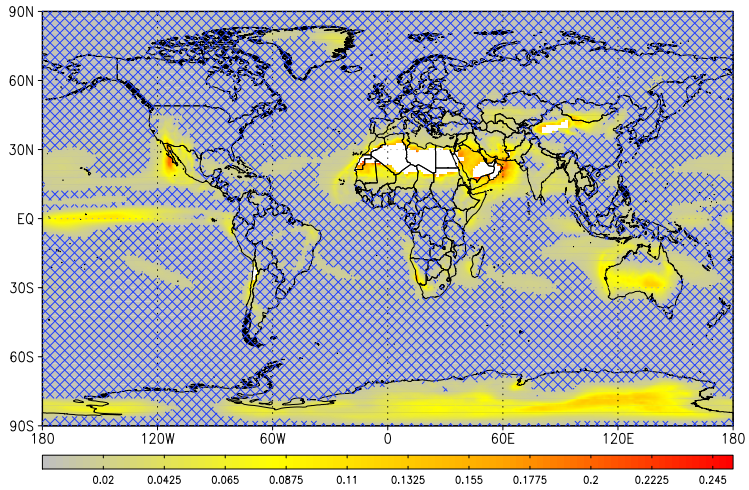
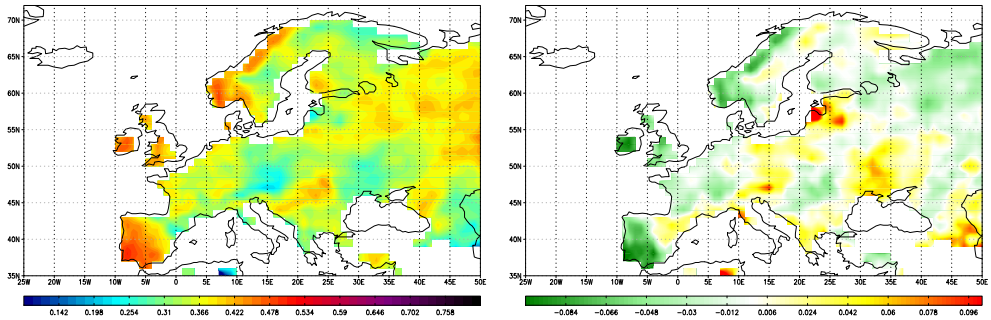


Figure 6: Bias of the SPI-1 calculated between the fitted Gamma distribution and the observed monthly cumulative precipitation (see text for details). Regions in white are considered to be too dry to fit this distribution. Regions where the bias deviates significantly from 0 (non-hatched areas) could generate bias in the SPI calculation.

therefore necessary to set a threshold at which minimum cumulative precipitation can be considered significant.

Different thresholds were tested (0 mm, 1 mm, 5 mm, 10 mm and 20 mm, not shown), and it was decided that only monthly precipitation levels greater than 10 mm are significant. This threshold allows us to keep a large number of events and to ignore events or regions with insignificant monthly accumulated rainfall. As outlined in the methodology, fitting a Gamma distribution to precipitation data relies on an adequate sample size (with respect to the variability of the data). The Gamma distribution was fitted to the distribution if at least 66% of a gridpointâĂŹs values were significantly larger than 0 (i.e. larger than 10mm). This ensures the inclusions of a minimum number of events to fit the distribution. These thresholds facilitated the removal of arid areas for which the fitting of the Gamma distribution presented biased values due to the low spread and low sampling of the time series.

The performance of the fitting procedure and all assumptions can be analysed by investigating the resulting SPI-1 distribution. This was done by calculating the integral of the differences between the fitted Gamma distribution and the empirical distribution. Zero values are considered to be perfect values (no bias of the SPI-1), whereas positive or negative values indicate bias and therefore give rise to doubts about the validity of the fitting procedure. In Figure 6, the bias of the Gamma distribution is shown. It can



(a) Correlation (b) RMSE

Figure 7: (a) Correlation of the forecasted (using the mean of the ensemble) and observed scores for the hindcast period (from November 1992 to November 2012) for (a) SPI-1 and (b) RMSE.

be seen that the method is adapted in a large parts of Europe.

However, the low precipitation levels in southern Spain can create some bias in the fitting process. This is especially true during the summer season, and therefore the assumptions for fitting the Gamma distribution are not validated for the entire year. This analysis shows that it will be necessary to adapt the method, particularly over dry areas, for example by carrying out the study only during the rainy seasons.

3.2 Validation during the hindcast period

This evaluation is based on the hindcast period (see table 1) of ENS and SEAS. It allows for the long-term evaluation using the same version of the model. The correlation and root mean square error of the ensemble means are displayed in Figure 7. The mean correlation (0.32) and the mean Root Mean Square Error (RMSE, 1.02) for ENS are better than those for SEAS (0.05 and 1.45 respectively, not shown). Neither the correlation nor the RMSE deviate significantly from 0, suggesting that a mean monthly forecast has no skill. In addition, the spatial variability is low, which indicates there is no significant spatial difference in the ability of the model to predict the SPI-1 on average.

The SPI-1 values of individual ensemble members and observations were analysed in bins to assess whether these results are also valid for extreme events. Here, the individual ability (for each member independently) was assessed by breaking down the forecasted and observed SPI-1 over Europe during the hindcast period into 10 classes (from SPI-1 lower than -2, to SPI-1 larger than +2, at intervals of 0.5). The frequency in each bin naturally

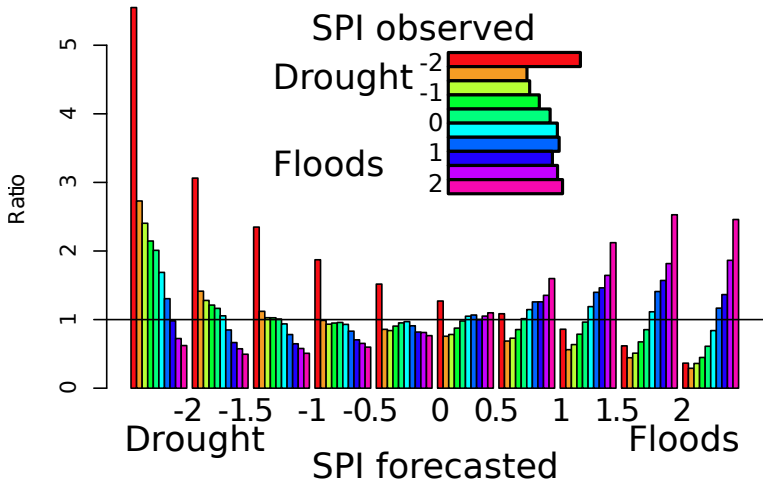


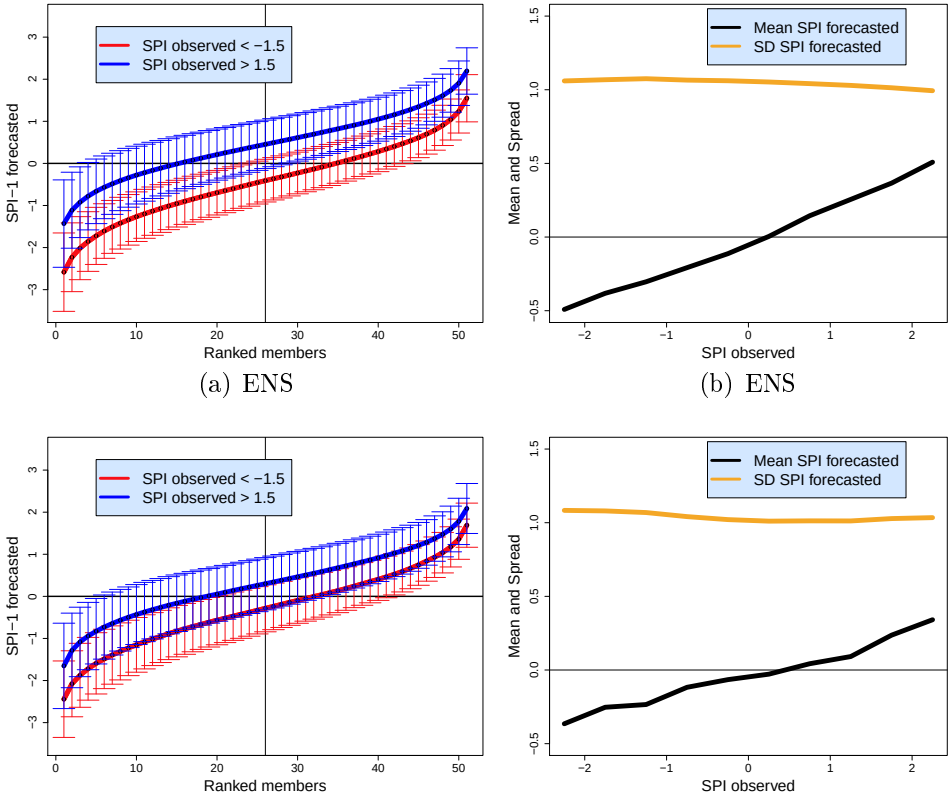
Figure 8: Ratio of events following the forecasted (x-axis) and observed SPI-1 (colour bars) over Europe using the hindcast period in relation to the theoretical distribution. Results are standardised by the theoretical normal distribution of events.

follows the Gamma distribution, which generates a large number of cases centred around 0. This distribution was normalised by computing the ratio between the empirical and the theoretical distributions. The result is shown in Figure 8. The figure shows that the more a drought is forecasted, the more it is observed (red bars). In addition, it should be noted that the distribution is highly unsymmetric. This indicates that the forecasts of extreme dry events are more accurate than the forecasts of extreme wet events. This result could be due to the spatial and temporal characteristics of drought events that are better simulated in a global model one month ahead.

3.3 Validation during the forecast period

The analysis of the forecast period from November 2012 to November 2013 largely confirms earlier findings of the forecasts over a significantly longer time period, but allows for a more detailed investigation of the distributions due to the larger ensemble number (see table 1).

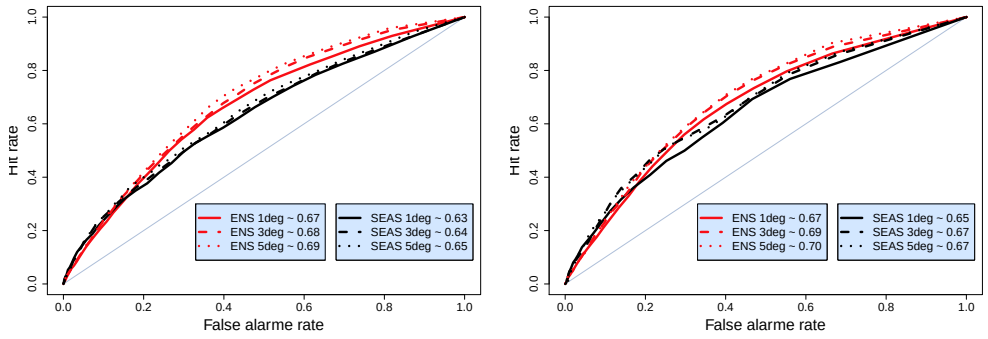
Figure 9a compares the behaviour of the ENS members during observed extreme wet and dry events. In both cases, the normal distributions of the ranked ensemble members are quite similar. The only difference is the shift of forecasted SPIs towards negative values when a drought is observed (red line) compared to when wet events are observed (blue line). Nevertheless, the standard deviation (indicated by the barlines) highlights that there is no



(a) ENS (b) ENS

(c) SEAS (d) SEAS

Figure 9: (a) Mean SPI-1 forecasted of ranked members using ENS during observed drought or floods (SPI-1 < -1.5 and SPI-1 > 1.5 respectively). (b) Ensemble mean and standard deviation of the SPI-1 forecasted using ENS following the associated observed SPI-1. (c) and (d) as (a) and (b) but using SEAS.



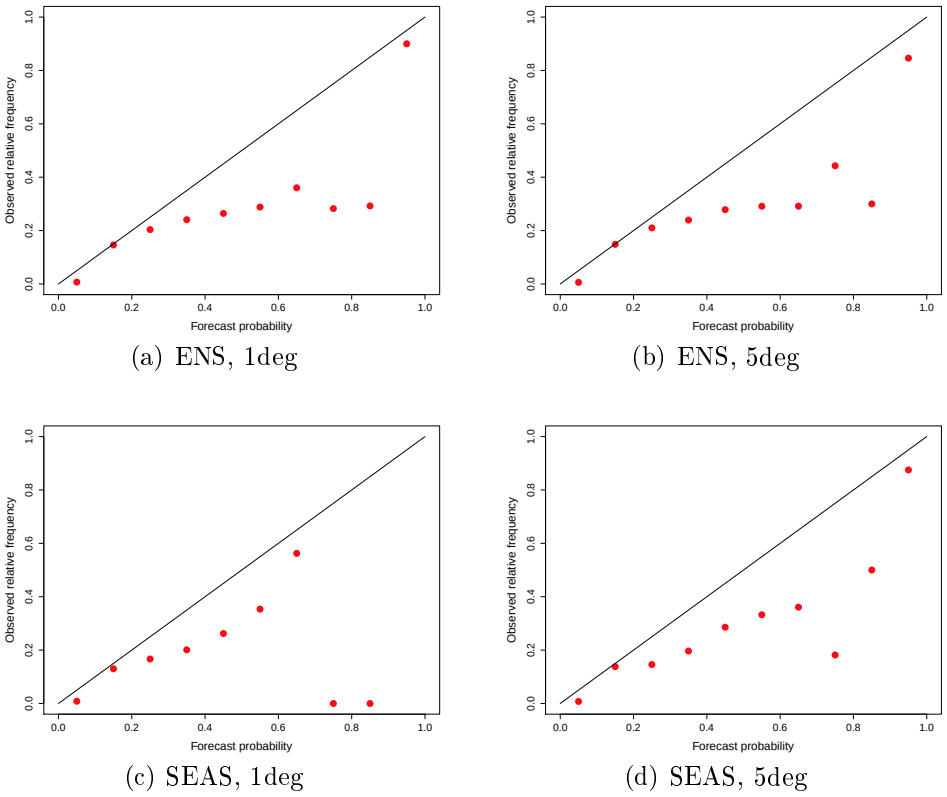
(a) SPI-1 $<$ -1 (b) SPI-1 $<$ -1.5

Figure 10: ROC curve using ENS and SEAS (red and black lines respectively) for the period November 2012 to November 2013 over Europe, calculated to detect a drought defined as having an SPI lower than -1 (a) or -1.5 (b). The ROC area values for the different spatial resolutions are indicated.

significant difference (significance level of 0.9) between the two events. It is interesting to observe that the value of the ensemble mean increases in line with the observed SPI-1 (black line in figure 9b), whereas the spread of the ensemble (defined as the standard deviation) shows little sensitivity (yellow line in figure 9b). It can be concluded that only the ensemble mean displays a significant difference between wet and dry anomalies, whilst there is no relation in the standard deviation. Similar trends are observed for SEAS, but the difference between the two conditional distributions are reduced (figure 9c and 9d). This indicates that ENS has a stronger resolution than SEAS, and is therefore better able to distinguish events with a better frequency distributions.

These results are confirmed by analysing the ROC curve. For the European continent, the ROC curves display an improvement in relation to the âĂŸno skillâĂŹ curve (1:1 in figure 10). The ROC area is slightly better for ENS than for SEAS (+0.4 and +0.2 for SPI-1 $<$ -1 and SPI-1 $<$ -1.5, respectively).

Both ENS and SEAS present an ascending but low reliability in detecting SPI-1 $<$ -1 (Figure 11). Indeed, the observed relative frequencies increase in line with the forecast probabilities. The percentage distribution of cases (not shown) indicates more events with ENS showing a larger percentage of members associated with a drought than SEAS. This result indicates that ENS members are more consistent in forecasting extreme rainfall deficits than are SEAS member. Using ENS, several events are forecasted with more than 93% of members associated with a drought forecasting, whereas using SEAS,



(a) ENS, 1deg (b) ENS, 5deg

(c) SEAS, 1deg (d) SEAS, 5deg

Figure 11: Reliability diagrams for drought detection defined as an SPI-1 lower than -1 using ENS (toppanels) and SEAS (bottom panels) in the period from November 2012 to November 2013. The spatial resolution is one degree (left panels) and 5 degrees (right panels).

the maximum is 81%.

The ENS and SEAS systems are better than climatological forecasts of drought events, achieving values of 0.14 and 0.12 respectively. In this case, the difference between ENS and SEAS is not significant.

3.4 Sensitivity to drought scales

Although the analysis so far has been performed on a scale of 1 by 1 degree, the sensitivity to different resolutions needs to be analysed, because the impacts of large-scale droughts will be stronger. Figure 10 shows SPI-1 values smoothed to 3 and 5 degrees using a simple upscaling method based on the average of the values. The resolution of about 1 degree has been kept to compare the impact of the resolution on the native grid. The results show a

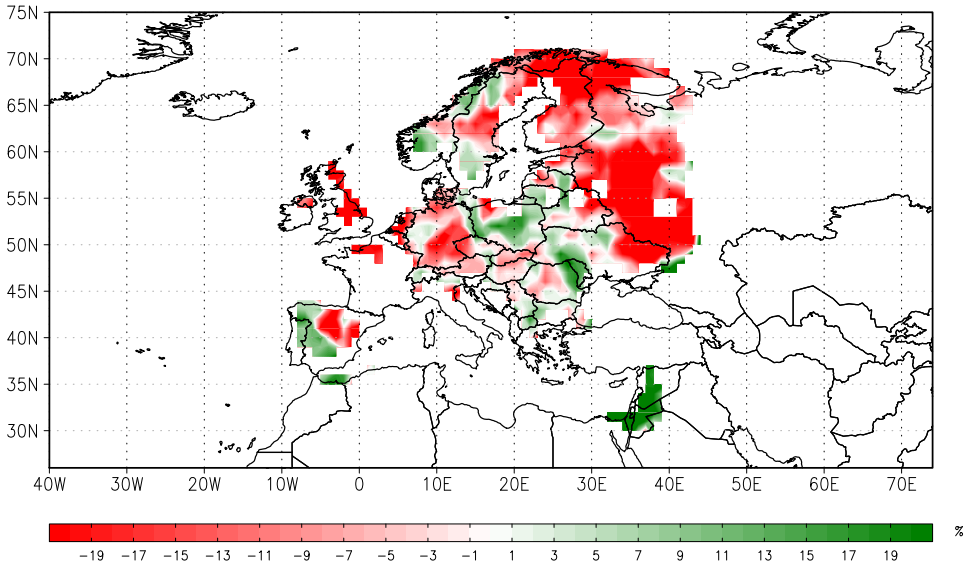


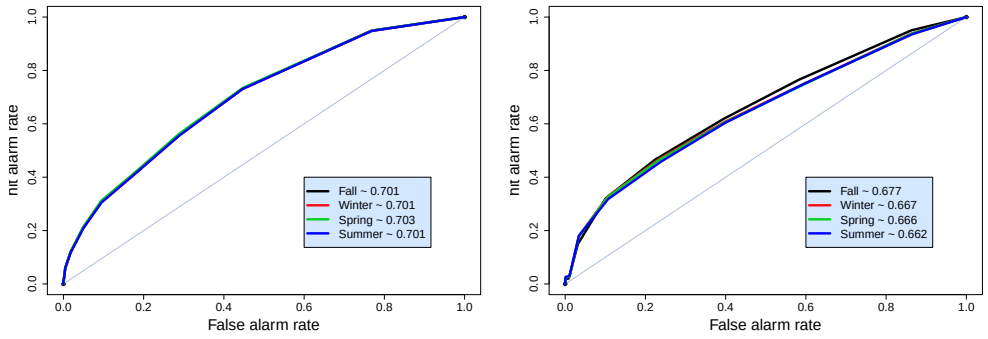
Figure 12: ROC anomaly (in %) in relation to the mean value of the ROC over the domain (equal to 0.67) for the period from November 2012 to November 2013 with drought defined as a SPI-1 < -1

slight improvement of the ROC area with a coarser resolution (broken and dotted lines in Fig. 10). The smoothed signal favours the large-scale signatures that are better represented in models than the small-scale structures of droughts. The effect of spatial upscaling can also be seen in the ROC results as a little positive impact for SEAS for the largest forecast probabilities (Fig. 11d). However, as mentioned previously, the number of events in these cases is low. This effect has been quantified onto the Brier Skill Score (BSS) that goes up to respectively 0.17 and 0.14 for the 5-degree smoothed signal.

3.5 Spatial and seasonal variabilities

Spatial variability

The analysis so far has ignored the spatial and seasonal scales. Figure 12 shows the ROC anomaly for the forecast period, which is the ROC area for each grid cell in relation to the average (0.67 for ENS). The anomaly is preferred to the raw value to highlight regions where the ROC is improved or reduced. A maximum variability of 20% can be observed. For the hindcast period (not shown) this variability is much lower (at around 6%). There is a difference in spatial patterns between the two periods, which seems to indicate that the spatial patterns are not significant and are mainly driven



(a) ENS (b) SEAS

Figure 13: Seasonal decomposition of the ROC curves for drought forecasting (with the 5 degree smoothing) using ENS (left) and SEAS (right) over Europe for the period from November 2012 to November 2013, with drought defined as a SPI-1 $<$ -1

by the extreme cases encountered during the period.

Seasonal variability

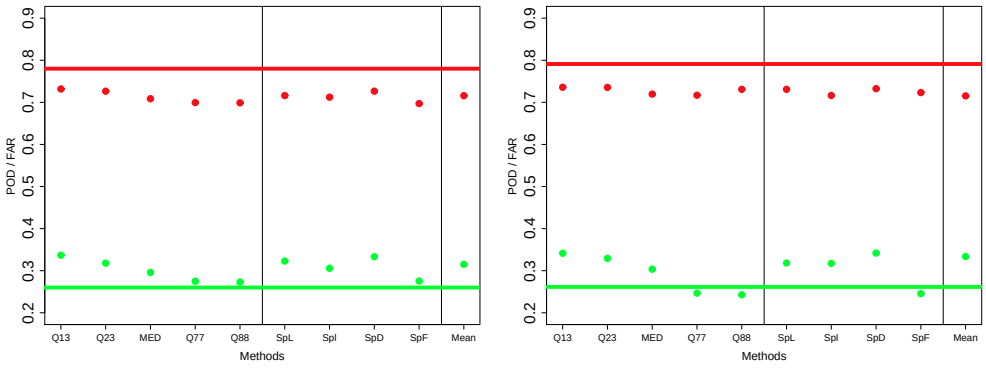
A seasonal decomposition is used to highlight the temporal variabilities. ROC scores and curves were independently calculated for autumn (September to November), winter (December to February), spring (March to May) and summer (June to August) seasons - see figure 13 (for SPI-1 $<$ -1).

The four ROC areas are very similar, and the four distributions are identical for ENS, which means that the ability to forecast droughts is identical throughout the year. By contrast, SEAS shows some differences between the seasons, with a small improvement of the forecast during autumn. As identical interpretations can be derived for the SPI-1 $<$ -1.5, they are not shown.

3.6 Index performance

Figure 14 shows the POD and the FAR for ENS and SEAS. POD indicates that, on average, one out of three drought events in Europe are correctly forecasted one month in advance. This is significantly better than both the climatology forecast (16%) and the deterministic forecast (around 25%, green line in Fig. 14).

The highest POD is achieved by using the 13 percentile (7^{th} member of the ranked ensemble distribution), and the product using Q13 and Q23 (noted



(a) ENS (b) SEAS

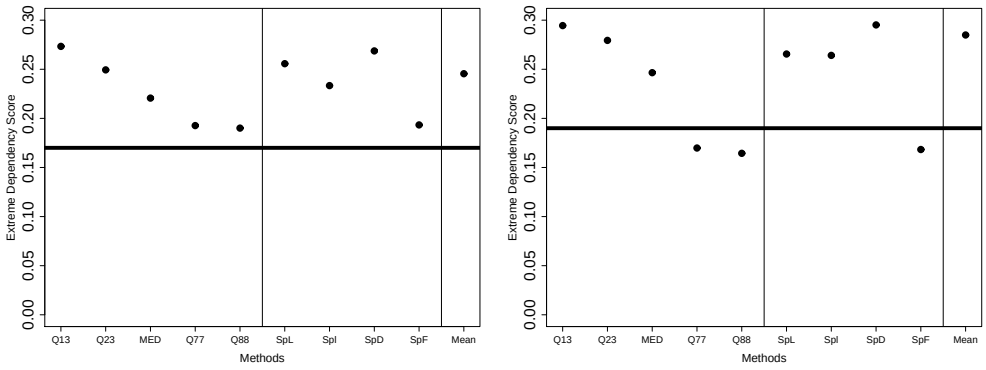
Figure 14: Probability of detection (POD, in green) and False alarm ratio (FAR, in red, per fect=0) for different methods used to detect drought (x-axis), using ENS (left) and SEAS (right). Lines indicate the scores of the deterministic model (unperturbed member of the ensemble).

SpD). The mean of the ensemble (last point on the right of each graph), which is used widely, is not the best method to detect droughts.

The POD values of the wettest members of the ranked distribution (Q77 and Q88 in Fig. 14) give the worst results of all methods, which indicates that there is little consistency between the extreme dry and wet members. The FAR displays low variability between the methods, but each of these are better than the deterministic solution (red lines). It is also worth noting that, using the ENS, the driest members are associated with a decrease of FAR in relation to the dry members. This could be explained by the previous scores that show a larger consistency between the members. However, it can also be due to a technical effect; because the number of events selected is constant, these scores could be dependent on each other.

The highest EDS is achieved for the driest members (Q13 and Q23, Fig. 15), whereas the wettest members (Q77 and Q88) have the lowest scores. The score of the ensemble mean is better than that of the median. Even if the POD and FAR differences are partially statistically significant, the improvement of the EDS for the driest members is significant for all differences larger than 0.04.

ENS and SEAS are reliable (see Fig. 11), and hence a method of detection could be simply based on the percentage of ensembles that predict a drought. In total, ten different percentage thresholds were selected. Figure 16 shows the rate of 'percent correct' increases in line with the percentage used for both models (black points in Fig. 16a and 16c). The trend is in agreement with the positive reliability found previously. The percent correct increases



(a) ENS (b) SEAS

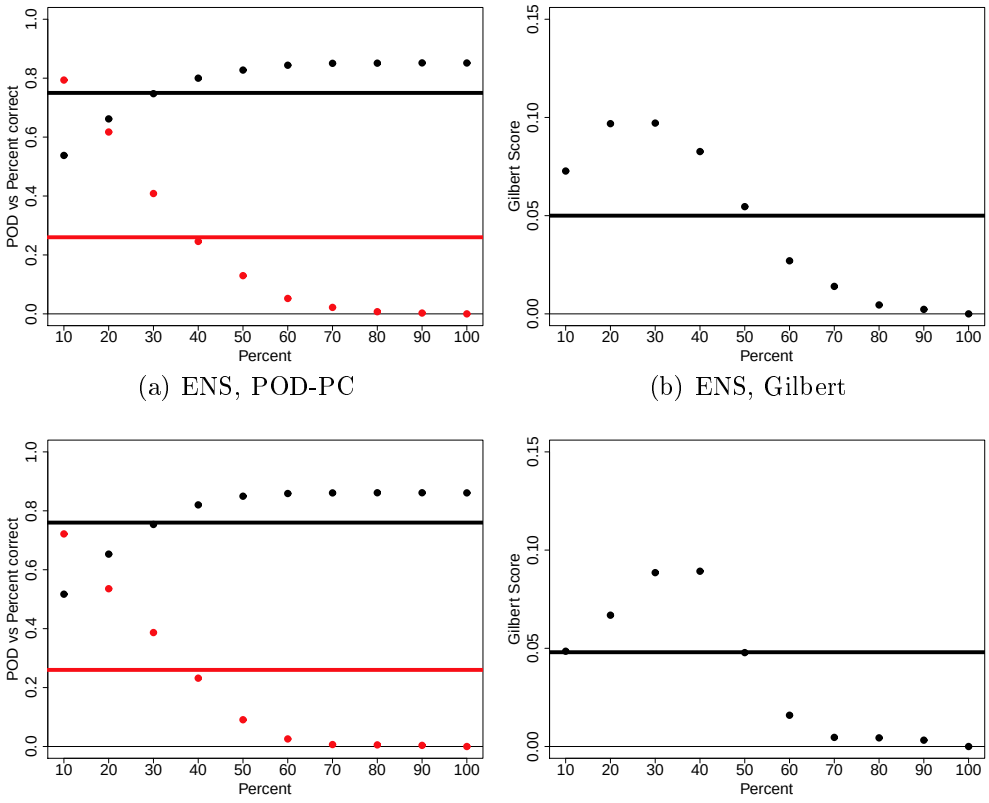
Figure 15: Extreme Dependency Score (EDS) for the 10 methods used to forecast a drought (x-axis, see table 1 for more details) using the ENS (left) and SEAS (right) ensemble system. Black lines indicate the score of the unperturbed member.

in line with the percentage used for both models (black points in Fig. 16a and 16c). That means that the greater the number of members forecasting a drought, the greater the possibilty of observing a drought. However, with an increasing threshold, the number of misses also increases (provided by the POD value, red lines in Fig. 16a and 16c). For example, if the threshold to determine a drought is defined with the 10% of members associated with a drought forecasting, around 80% of droughts that occurred were correctly detected (red points), but more than 50% of the forecasted droughts are false alarms. On the other hand, if the threshold of detection is defined with a percentage larger than 70%, the percentage correct is about 85%, but the POD is close to 3%. Based on this result, the user can tune the percentage to the false alarm ratio of misses cases that is acceptable.

The maximum Gilbert score (Fig. 16b and 16d) is achieved for a threshold of 30% for ENS and 40% for SEAS. The number of missed events becomes too high with a larger percentage threshold, whereas for lower percentage thresholds the errors are associated with false alarms.

3.7 Assessment the uncertainties of the forecasts

Several previous studies (He et al., 2009; Palmer, 2000; Georgakakos et al., 2004; Doblas-Reyes et al., 2009) have shown that probabilistic simulations can provide additional information to assess the uncertainties of the simulation. The idea here is to estimate the quality of the forecast based on a specific behaviour of the simulation. The characteristics of the ensemble in



(a) ENS, POD-PC (b) ENS, Gilbert

(c) SEAS, POD-PC (d) SEAS, Gilbert

Figure 16: (a) POD (red) and percent correct (black) using different percentage of members to forecast a drought event using ENS. (b) Gilbert score (see text for more details) for different percentage used to forecast a drought using ENS. Lines indicate the score of the deterministic model (unperturbed member). (c) and (d) same as (a) and (b) using SEAS.

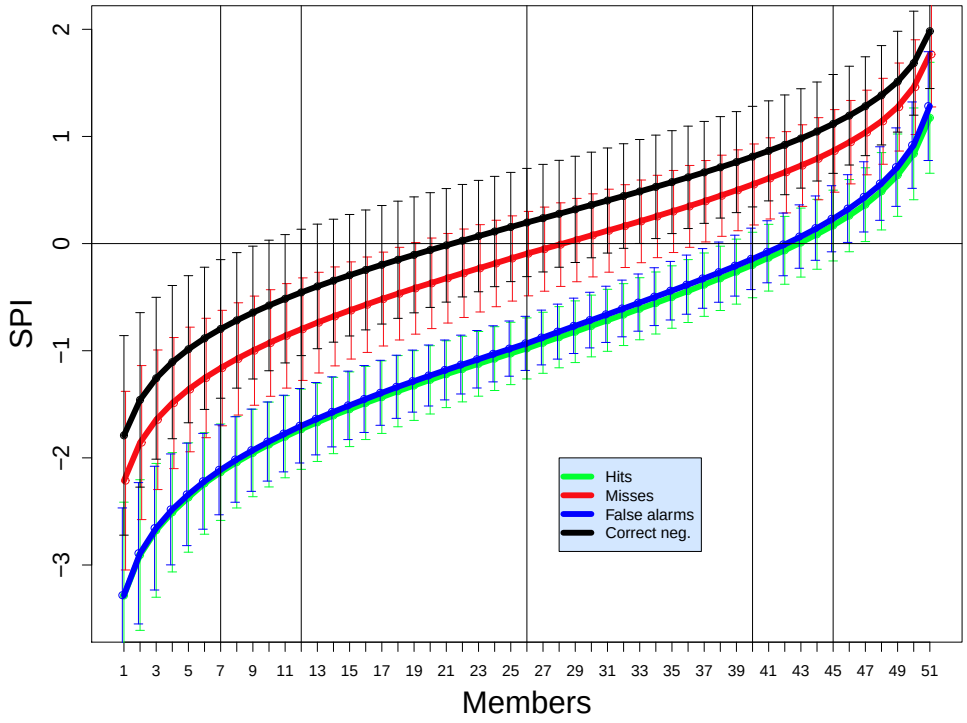


Figure 17: Mean SPI-1 and standard deviation of the ranked members following the four conditions in the contingency table (see table 2 and text for more details): hits (green), false alarm (red), misses (blue) and correct negative (black line), using ENS. Vertical lines indicate the members used for the boolean drought detection methods.

the four different cases of the contingency table were analysed. This table was compiled using the threshold of SPI-1 < -1 to detect a drought, and the forecast method was based on the median of the members.

The mean SPI-1 of the 51 ranked members for the four cases are illustrated in Fig. 17. During correct negative events (i.e. where droughts were neither forecasted nor observed), where more than 70% of the events are located, a normal distribution is observed with a mean slightly larger than 0. During the missed cases, the median is very close to 0 and the distribution of the ranked members is very close to the ensemble mean. In addition, the spread of the members is displayed (barbed lines) and shows the increase of the spread for extreme members and the fact that the two distributions become indistinguishable. That means that the response of the model is no different to a normal distribution. So it is not significant to find a specific behaviour of the model to assess the missed events. Finally, the distribution of

Table 3: Contingency table (in percentages) obtained using the median of ENS to forecast a drought. An observed drought is defined as having an SPI-1 lower than -1, and a forecasted drought is defined as having an ensemble median that is lower than the 16th percentile. The second values of each case indicate the ensemble spread, and its standard deviation is given in brackets

		drought	observed
		yes	no
drought	yes	4.4% / 2.31 (0.4)	10.7% /2.37 (0.4)
forecasted	no	10.4% / 1.99 (0.4)	74.5% / 1.88 (0..3)

the members during hits and false alarms are compared. In that case, there is no significant difference. The average and the distribution of the mean SPI-1 of the ensemble is the same. These results are in agreement with Table 3, which quantifies the ensemble spread for each case of the contingency table. Based on these results, it appears impossible to evaluate the uncertainties of the ensemble simulation associated with a boolean decision.

4 Spatial and temporal variabilities of SPI-1

A model of the early warning of drought over Europe will be tested in 2015 using atmospherical predictors (such as geopotential and temperature in the free troposphere), which are better represented in the model than precipitation. In order to achieve this objective, a preliminary study was carried out to characterise the spatial and temporal variabilities of SPI-1 over Europe. The principle component analysis (PCA), using the empirical orthogonal functions (EOF) is the most appropriate tool to perform this kind of study.

To illustrate this first ongoing step, Figure 18 and 19 illustrates the two main components of the SPI in Europe during the coldest period of the year (from November to March) that spans 1992 to 2012. In the first mode, the pattern is associated with a high variability of SPI-1 located over Denmark and northern Germany. Droughts occurred in this area during the beginning of the period (i.e. around 1995) and more recently (strong rainfall deficit in 2012). The second mode is centred in the northern part of Finland. Here we recorded at least 8 episodes of strong rainfall deficits (SPI-1 lower than -1). Based on these identified modes and temporal variabilities, predictors will be identified by finding atmospherical variabilities that are highly correlated to them.

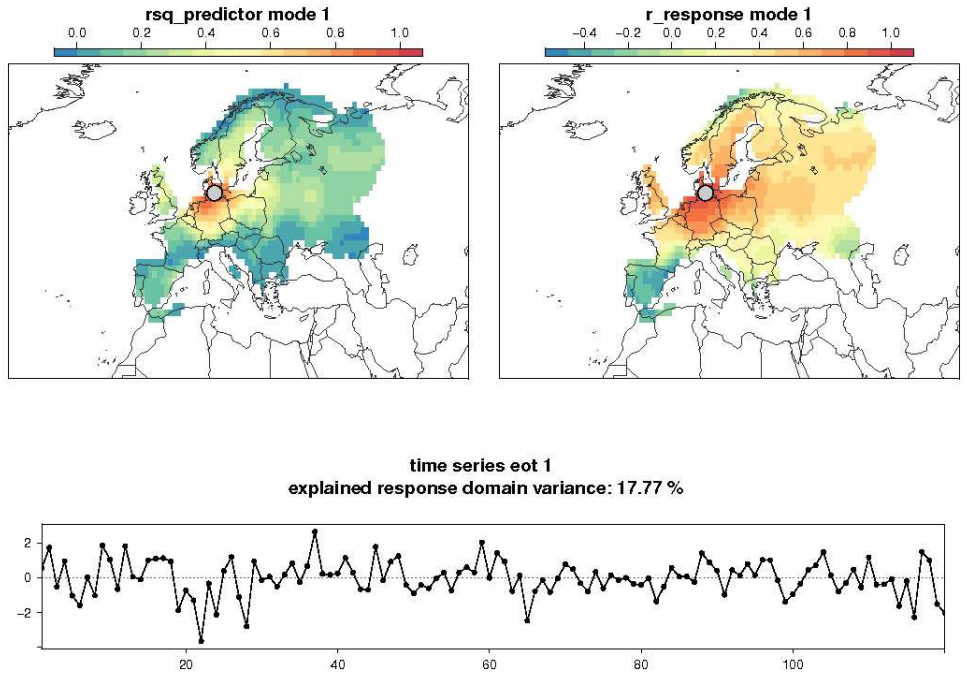


Figure 18: Coefficient of determination (top left), correlation coefficient (top right) and temporal evolution of the principal component (PC, bottom) of the first mode of the SPI-1 observed in Europe during the cold season (from November to March) from 1982 to 2012.

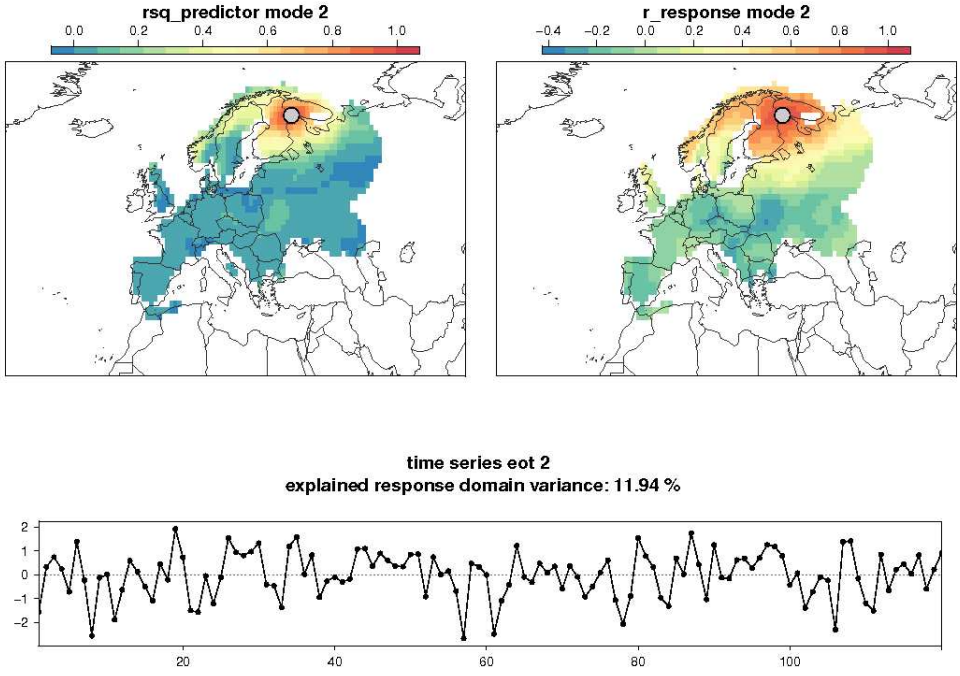


Figure 19: Same as Figure 18 for the second mode of the EOF. TThis mode explains 10% of the total variance.

5 Conclusions

This study provides the first assessment of the predictability of meteorological droughts over Europe and of the ability to issue an early warning of such droughts with a one month lead time. The analysis is based on the one month forecast of the SPI-1 from the precipitation outputs provided by two ECMWF ensemble systems. In a first step the ability to forecast SPI-1 from the ensemble outputs was tested, showing that

- The reliability of the ensemble is better than the climatology,
- The spatial variability of the scores can reach up to 20% over Europe and the seasonal variability is not significant. Nevertheless, we note a large variability of the ensemble score depending on the events that occurred,
- Ensemble models are better at forecasting large-scale droughts, using a spatial smoothing up to five square degrees.

In a second step the ability to provide a robust Boolean index for drought forecasting was analyzed. The best method is defined by using a threshold of 30% of ensemble members associated with a drought. In that case, slightly more than 40% of the droughts observed are forecasted correctly one month ahead, with only 25% of false alarms. This is significantly better than using the climatology (16%) or the deterministic models (around 25%). Finally, this study has shown that there is no possibility to provide uncertainties associated with the boolean index.

By providing the first global assessment of meteorological drought forecasting in Europe, this work will be particularly useful by providing a benchmark comparison for future studies that could be developed using others methods, such as those based on atmospheric predictors, which are better represented in the seasonal models. It could farther be useful to investigate the use of moving windows of 10-day cumulative precipitation to detail the temporal behaviour of the forecasted SPI-1. As the forecast skills are better for short lead times, an SPI-1 lower than -1, explained by a strong decrease in precipitation at the beginning of the period, should be more reliable.

Acknowledgements

The author would like to thanks Jürgen Vogt (JRC) and Florian Pappenberger (ECMWF) for their valuable comments and suggestions. This document has been improved by Gráinne Mulhern (JRC) by correcting the text.

Bibliography

Arribas, A., Glover, M., Maidens, A., Peterson, K., Gordon, M., MacLachlan, C., Graham, R., Fereday, D., Camp, J., Scaife, A., et al.: The GloSea4 ensemble prediction system for seasonal forecasting, Monthly Weather Review, 139, 1891–1910, 2011.

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., and DeWitt, D. G.: Skill of Real-Time Seasonal ENSO Model Predictions during 2002-11: Is Our Capability Increasing?, Bulletin of the American Meteorological Society, 93, 631–651, 2012.

Below, R., Grover-Kopec, E., and Dilley, M.: Documenting drought-related disasters a global reassessment, The Journal of Environment & Development, 16, 328–344, 2007.

Brier, G. W.: Verification of forecasts expressed in terms of probability, Monthly weather review, 78, 1–3, 1950.

Buizza, R., Houtekamer, P., Pellerin, G., Toth, Z., Zhu, Y., and Wei, M.: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems, Monthly Weather Review, 133, 1076–1097, 2005.

Cacciamani, C., Morgillo, A., Marchesi, S., and Pavan, V.: Monitoring and forecasting drought on a regional scale: Emilia-Romagna Region, in: Methods and tools for drought analysis and management, pp. 29–48, Springer, 2007.

Doblas-Reyes, F., Weisheimer, A., Déqué, M., Keenlyside, N., McVean, M., Murphy, J., Rogel, P., Smith, D., and Palmer, T.: Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts, Quarterly Journal of the Royal Meteorological Society, 135, 1538–1559, 2009.

Dutra, E., Giuseppe, F. D., Wetterhall, F., and Pappenberger, F.: Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index, Hydrology and Earth System Sciences, 17, 2359–2373, 2013.

Dutra, E., Pozzi, W., Wetterhall, F., Di Giuseppe, F., Magnusson, L., Naumann, G., Barbosa, P., Vogt, J., and Pappenberger, F.: Global meteorological drought–Part 2: Seasonal forecasts, Hydrology and Earth System Sciences Discussions, 11, 919–944, 2014.

Edossa, D. C., Babel, M. S., and Gupta, A. D.: Drought analysis in the Awash river basin, Ethiopia, Water resources management, 24, 1441–1460, 2010.

Eshel, G., Cane, M. A., and Farrell, B. F.: Forecasting eastern Mediterranean droughts, Monthly weather review, 128, 3618–3630, 2000.

Ferro, C. A. and Stephenson, D. B.: Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events, Weather and Forecasting, 26, 699–713, 2011.

Fraser, E. D., Simelton, E., Termansen, M., Gosling, S. N., and South, A.: âĂIJVulnerability hotspotsâĂİ: Integrating socio-economic and hydrological models to identify where cereal production may decline in the future due to climate change induced drought, Agricultural and Forest Meteorology, 170, 195–205, 2013.

Georgakakos, K. P., Seo, D.-J., Gupta, H., Schaake, J., and Butts, M. B.: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles, Journal of Hydrology, 298, 222–241, 2004.

Guttman, N. B.: Accepting the Standardized Precipitation Index: A calculation algorithm, 1999.

Guy Merlin, G. and Kamga, F. M.: Computation of the Standardized Precipitation Index (SPI) and Its Use to Assess Drought Occurrences in Cameroon over Recent Decades, Journal of Applied Meteorology and Climatology, 2014.

Hamill, T. M., Brennan, M. J., Brown, B., DeMaria, M., Rappaport, E. N., and Toth, Z.: NOAA's Future Ensemble-Based Hurricane Forecast Products, Bulletin of the American Meteorological Society, 93, 209–220, 2012.

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.

He, Y., Wetterhall, F., Cloke, H., Pappenberger, F., Wilson, M., Freer, J., and McGregor, G.: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, Meteorological Applications, 16, 91–101, 2009.

Hogan, R. J., Ferro, C. A., Jolliffe, I. T., and Stephenson, D. B.: Equitability revisited: Why the âĂIJequitable threat scoreâĂİ is not equitable, Weather and Forecasting, 25, 710–726, 2010.

Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification, A Practtioners Guide in Atmospheric Science, 2003.

Khan, S., Gabriel, H., and Rana, T.: Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas, Irrigation and Drainage Systems, 22, 159–177, 2008.

Kim, T.-W. and Valdés, J. B.: Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks, Journal of Hydrologic Engineering, 8, 319–328, 2003.

Lavaysse, C., Carrera, M., Bélair, S., Gagnon, N., Frenette, R., Charron, M., and Yau, M.: Impact of Surface Parameter Uncertainties within the Canadian Regional Ensemble Prediction System, Monthly Weather Review, 141, 1506–1526, 2013.

Lloyd-Hughes, B. and Saunders, M. A.: A drought climatology for Europe, International Journal of climatology, 22, 1571–1592, 2002.

Mason, I.: A model for assessment of weather forecasts, Aust. Meteor. Mag, 30, 291–303, 1982.

McKee, T. B., Doesken, N. J., and Kleist, J.: 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, 1993.

Mishra, A. and Desai, V.: Drought forecasting using stochastic models, Stochastic Environmental Research and Risk Assessment, 19, 326–339, 2005.

Mishra, A. and Desai, V.: Drought forecasting using feed-forward recursive neural network, Ecological Modelling, 198, 127–138, 2006.

Mishra, A., Desai, V., and Singh, V.: Drought forecasting using a hybrid stochastic and neural network model, Journal of Hydrologic Engineering, 12, 626–638, 2007.

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.

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., and Vitart, F.: The new ECMWF seasonal forecast system (System 4), European Centre for Medium-Range Weather Forecasts, 2011.

Nurmi, P.: Recommendations on the verification of local weather forecasts, 2003.

Palmer, T. N.: Predicting uncertainty in forecasts of weather and climate, Reports on Progress in Physics, 63, 71, 2000.

Pereira, S., Carvalho, A., Ferreira, J., Nunes, J., Keizer, J., and Rocha, A.: Simulation of a persistent medium-term precipitation event over the western Iberian Peninsula, Hydrology and Earth System Sciences, 17, 3741–3758, 2013.

Richardson, D., Bidlot, J., Ferranti, L., Haiden, T., Hewson, T., Janousek, M., Prates, F., and Vitart, F.: Evaluation of ECMWF forecasts, including 2012–2013 upgrades, Tech. rep., ECMWF Technical Memo, 2013.

Stockdale, T., Anderson, D., Alves, J., and Balmaseda, M.: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model, Nature, 392, 370–373, 1998.

Sunyer, M., Sørup, H. J. D., Christensen, O. B., Madsen, H., Rosbjerg, D., Mikkelsen, P. S., and Arnbjerg-Nielsen, K.: On the importance of observational data properties when assessing regional climate model performance of extreme precipitation, Hydrology and Earth System Sciences Discussions, 10, 7003–7043, 2013.

Van den Besselaar, E., Haylock, M., Van der Schrier, G., and Klein Tank, A.: A European daily high-resolution observational gridded data set of sea level pressure, Journal of Geophysical Research: Atmospheres (1984–2012), 116, 2011.

Vicente-Serrano, S. M.: Differences in spatial patterns of drought on different time scales: an analysis of the Iberian Peninsula, Water Resources Management, 20, 37–60, 2006.

Vitart, F.: Monthly forecasting at ECMWF, Monthly Weather Review, 132, 2761–2779, 2004.

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

Vitart, F., Buizza, R., Alonso Balmaseda, M., Balsamo, G., Bidlot, J.-R., Bonet, A., Fuentes, M., Hofstadler, A., Molteni, F., and Palmer, T. N.: The new VAREPS-monthly forecasting system: A first step towards seamless prediction, Quarterly Journal of the Royal Meteorological Society, 134, 1789–1799, 2008.

Weigel, A. P., Baggenstos, D., Liniger, M. A., Vitart, F., and Appenzeller, C.: Probabilistic verification of monthly temperature forecasts, Monthly Weather Review, 136, 5162–5182, 2008.

Weisheimer, A. and Palmer, T.: On the reliability of seasonal climate forecasts, Journal of The Royal Society Interface, 11, 20131 162, 2014.

Weisheimer, A., Corti, S., Palmer, T., and Vitart, F.: Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372, 20130 290, 2014.

Wilhite, D. A. and Glantz, M. H.: Understanding: the drought phenomenon: the role of definitions, Water international, 10, 111–120, 1985.

WMO: Standardized Precipitation Index, User Guide, Tech. Rep. 1090, 2012.

European Commission
EUR 27109 EN – Joint Research Centre – Institute for Environment and Sustainability

Title: **Early warming of drought in Europe**

Author: Christophe Lavaysse

Luxembourg: Publications Office of the European Union

2015 – 35 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online),

ISBN 978-92-79-45597-1 (PDF)

doi: 10.2788/093828

LB-NA-27109-EN-N

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Serving society
Stimulating innovation
Supporting legislation

doi: 10.2788/093828

ISBN 978-92-79-45597-1

Publications Office