Evaluation of a New Precipitation-Based Index for Global Seasonal Forecasting of Unusually Wet and Dry Periods

CHRISTOPHE LAVAYSSE

Institut des Geosciences de l'Environnement, UGA, CNRS, INPG, IRD, Grenoble, France, and European Commission, Joint Research Centre, Ispra, Italy

TIM STOCKDALE

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

NIAL McCORMICK AND JÜRGEN VOGT

European Commission, Joint Research Centre, Ispra, Italy

(Manuscript received 27 September 2019, in final form 5 April 2020)

ABSTRACT

This paper describes the assessment of the performance of a method for providing early warnings of unusually wet and dry precipitation conditions globally. The indicator that is used for forecasting these conditions is computed from forecasted standardized precipitation index (SPI) values for accumulation periods of 1, 3, and 6 months. The SPI forecasts are derived from forecasted precipitation produced by the latest probabilistic seasonal forecast of ECMWF. Early warnings of unusual precipitation periods are shown only when and where the forecast is considered robust (i.e., with at least 40% of ensemble members associated with intense forecasts), and corresponding with significant SPI values (i.e., below -1 for dry, or above $+1$ for wet conditions). The intensity of the forecasted events is derived based on the extreme forecast index and associated shift of tails products developed by ECMWF. Different warning levels are then assessed, depending on the return period of the forecast intensity, and the coherence of the ensemble forecast members. The assessment of the indicators performance is based on the 25-member ensemble forecast system that is carried out every month during the 36 years of the hindcast period (1981–2016). The results show that significant information is provided even for the longest lead time, albeit with a large variability across the globe with the highest scores over central Russia, Southeast Asia, and the northern part of South America or Australia. Because of the loss of predictability, each SPI is based on the first lead time. A sensitivity test highlights the influence on the robustness of the forecasts of the warning levels used, as well as the effects of prior conditions and of seasonality.

1. Introduction

Meteorological droughts, which are defined as periods with an abnormal precipitation deficit relative to the long-term average conditions for a region (Wilhite et al. 2014), can have a serious impact on human activities

such as agriculture and water resource management, energy production, tourism, and recreation (Fraser et al. 2013). Droughts are among the costliest natural disasters, particularly when the drought impacts are spread over wide regions, and last for several months (Lloyd-Hughes and Saunders 2002; Below et al. 2007). In that context, the capability to forecast droughts for different time periods represents an important decision support tool for triggering actions to mitigate the negative impacts of droughts. Similarly, there is an urgent need for improved forecasting of extreme wet periods, which also have major socioeconomic and environmental impacts. Floods and landslides, for example, are common consequences of extreme rainfall events, with river floods in particular being recognized as one of the major causes of

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-19-0196.s1>.

Corresponding author: C. Lavaysse, christophe.lavaysse@ird.fr

DOI: 10.1175/WAF-D-19-0196.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

economic damages and loss of human lives worldwide (CRED/UNDRR 2015).

Decision-makers require clear, robust forecast information that can communicate the probability and likely severity of the expected events. However, the chaotic nature of atmospheric processes tends to generate large uncertainties in deterministic (i.e., categorical) forecasts derived from numerical weather prediction (NWP) models, particularly when dealing with subseasonal and seasonal time scales (Stockdale et al. 1998; Vitart et al. 2014). Therefore, ensemble prediction systems (EPS) have been developed in order to forecast also the uncertainties that are associated with NWP models. These “probabilistic” forecasts become particularly important when assessing the risks associated with rare weather events such as droughts and tropical cyclones (Hamill et al. 2012; Dutra et al. 2013, 2014), and for identifying uncertainties in the forecasts (Buizza et al. 2005; Lavaysse et al. 2015).

Climate forecasts at seasonal time scales using dynamical models—which, in contrast with traditional statistical models, use basic physical principles to calculate changes in climate features—have evolved considerably over recent years, and have been shown to be potentially useful for predicting large-scale climate features and “teleconnections” (i.e., climate variability links over long distances, typically thousands of kilometers) (Barnston et al. 2012; Arribas et al. 2011; Lyon et al. 2012). The latest generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) real-time seasonal forecast system—version 5 of the seasonal prediction system (SEAS5)—is based on an upgrade (Cycle 43r1) of ECMWF’s state-of-the-art Integrated Forecasting System (IFS) (Johnson et al. 2019). Recent studies have demonstrated the capability of SEAS5 to drive seasonal forecasts of hydrological models (Emerton et al. 2018) and to represent different physical elements, such as large-scale circulation and climate teleconnection patterns (Weisheimer et al. 2018), sea ice cover (Palmer et al. 2019), precipitation (Ratri and Schmeits 2018), and tropical and extratropical cyclones (Angus and Leckebusch 2018).

A 2014 study on the reliability of seasonal climate forecasts indicated a low forecast “skill” (or reliability performance) of the ECMWF seasonal prediction system 4 (the predecessor of SEAS5) for summer forecasts of rainfall over northern Europe (Weisheimer and Palmer 2014). However, these analyses did not incorporate the impact of “persistence” on the perceived forecast skill, and indeed in other regions in the world this predictability could be higher. In fact, seasonal forecasts are mainly driven by changes in slowly varying forcings such as sea surface temperature (SST) anomalies—most

notably those associated with the El Niño–Southern Oscillation (ENSO) cycle but also others such as the Indian Ocean dipole (IOD). Such slow variations of the climate system are often a source of predictability on seasonal time scales (Stockdale et al. 2010). The impact of these signals is strongly variable, and may differ from one region to another. Therefore, good forecast skill tends to be feasible, especially over regions where there is a large impact from these oscillations.

In this context, the aim of the present study is to assess the skill of a precipitation-based index for predicting unusually wet and dry periods, which are defined as periods with abnormally high and low amounts of precipitation—in our case about 16% of the most intense events of the sampling. Such events are of crucial importance to many decision-makers. Because the deficit or exceedance of rainfall may have different impacts depending on the accumulation period of interest, several forecasts are needed, ranging from a few weeks to several months of cumulative rainfall. The accuracy of the forecasts is therefore quantified according to the forecasting lead times, ranging from 1 to 6 months, and forecast skill is assessed, in order to characterize the ability of the model to predict the considered climatic features (Boer et al. 2013).

In the remainder of this paper, the forecasted precipitation datasets on which the study is based are first described, and the metrics that are used to detect forecasted precipitation anomalies and to calculate the severity of the forecasted events, are outlined. Then, the method for estimating the main forecast skill scores is explained, and the sensitivity of these skill scores to various factors (e.g., climatic trends, initial model conditions, warning levels, and seasons) is discussed. Finally the main conclusions, in terms of the ability and limitations of the indicator for predicting unusually wet and dry periods, are presented.

2. Data and methodology

a. Forecasted precipitation datasets

The indicator for forecasting unusually wet and dry conditions is computed based on forecasted precipitation derived from the long-range (i.e., seasonal) forecast system (SEAS5) of ECMWF. The long-range probabilistic forecast of SEAS5 consists of a 51-member ensemble forecast (ENS), which is integrated for a total forecast length of 13 months. A reforecast ensemble is also available for the period 1981–2016, which comprises an ensemble size of 25 members with up to 7-month lead time, and 15 members with up to 13-month lead time.

FIG. 1. Illustration of the definition of the alert levels: (i) Calculation of the SPI. (ii) Calculation of the EFI–SOT index of SPI values. The red and blue curves indicate the ensemble cumulated distribution of the 51 sorted SPI values forecasted over a grid point during dry and wet conditions, respectively. The surface area from the member at the 40th and 60th percentiles (y axis) to the extreme ones is then calculated (shaded areas) and provides the intensity of the event. (iii) Intensity is transformed to an equivalent return period value. See text for more details.

The SEAS5 seasonal forecast system is described in detail by Johnson et al. (2019).

While SEAS5 has a spatial native resolution (grid spacing) of approximately 36 km, for our purposes all of the results are regridded to a resolution of $1^\circ \times 1^\circ$ (i.e., approximately 110 km over the tropics), in order to focus on large-scale events, which are better represented and forecasted in NWP models. This upscaling is done using an averaging method weighted by the overlapping surfaces.

A big challenge for validating forecasts at the global scale, is to define global datasets representing the true values with a similar robustness everywhere. Due to the wide variations in density of weather stations around the world, the robustness of ground observations is considered to be too variable in space. In this study, therefore, the ground observations are derived from the gridded Global Precipitation Climatology Centre (GPCC) precipitation data—developed by the German Weather Service (DWD) for the World Meteorological Organization (WMO)—which are currently an important input for the Global Drought Observatory (GDO) of the European Commission’s Copernicus Emergency Management Service (<https://emergency.copernicus.eu/>). The precipitation data are estimated thanks to a network of ground observations—namely Global Telecommunication System (GTS) data and meteorological synoptic data (SYNOP)—and satellite image data, for spatialization and homogenization. The precipitation data, which are subject to manual quality control, are interpolated to a 1° grid resolution, and are available from 1986 to the present, with a time lag of 2 months. The ground observation data for the 1-, 3- and 6-month standardized precipitation index (SPI) accumulation periods are used, as described later in this paper, as the basis for

verifying the SEAS5 global forecast data for the corresponding lead times.

b. Detection of forecasted precipitation anomalies and calculation of forecast intensity

There are three main steps for transforming the probabilistic forecasts of precipitation into simplified and robust deterministic alerts. First, the forecasted precipitation is transformed into values of the standardized precipitation index (SPI) (McKee et al. 1993; Edwards 1997). Then, a detection is applied according to defined thresholds of exceedance. Based on this detection, the intensity calculation is transformed into a return period, based on the hindcast period. The three steps are illustrated in Fig. 1.

The SEAS5 forecasted precipitation is transformed to SPI values for three accumulation periods, which are relevant in terms of their impacts: 1 month (SPI-1), 3 months (SPI-3), and 6 months (SPI-6). For the SPI calculations, the continuous probability function for each of the accumulation periods is derived from the entire set of precipitation values during the 30-yr SEAS5 reforecast period (1981–2010), in line with WMO guidelines (WMO 2012).

According to Lavaysse et al. (2015), one of the most reliable ways to provide a dichotomous (i.e., two-class) forecast of extreme dry conditions, based on a probabilistic (ensemble) system, is to use the 40th percentile (P40) of the ensemble members, which have been sorted from driest to wettest for each grid point, month and SPI accumulation period. When the SPI of the P40 member is below -1 , the forecast of unusually dry conditions is considered reliable or robust. Conversely, when the SPI of the 60th percentile (P60) ensemble member is greater than $+1$, unusually wet conditions are reliably

forecasted. The reason behind these empirical thresholds of members is related to the balance between the need for coherence of the members that is associated with the reliability of the forecast, and the uncertainties of forecasting extreme events, which are located in the tails of the member distribution. This sorting of the ensemble members, which is carried out independently for each grid cell, enables the direct derivation of thresholds, and calculation of the forecast severity (i.e., the intensity of the most extreme ensemble members), as described below.

After local standardization as SPI values, the results of analysis carried out as part of this study (not shown here) revealed a large spatial variability of the internal (i.e., between members) and interannual variance of the SPI, for all accumulation periods. Regarding the spatial variability of the SPI variance between members, this can be over three times lower for some regions (e.g., northern Russia) than for other regions (e.g., Pacific Ocean). In general, this variability depends on the latitude. Regarding the spatial variability of the interannual variance of the mean ensemble SPI values, this appears to be generally higher than the aforementioned internal variance, and less dependent on latitude, with lowest values mainly over dry regions (e.g., Arabian Peninsula, the northern part of continental Australia). This latter variability is less dependent on latitude, with lowest values mainly over dry regions (e.g., Arabian Peninsula, the northern part of continental Australia). To compensate for spatial variability of the model variance, which will affect the detection of extreme events, the forecast thresholds are adjusted to give a statistically constant number of detected events (around 16% of the sample size). The corrected forecast thresholds are defined independently for each model run, for each grid cell and for each SPI accumulation period. Therefore, the forecast of unusually dry conditions is robust when the P40 ensemble member is below the adjusted dry threshold (or -1 in case of no correction), while the forecast of unusually wet conditions is robust when the P60 ensemble member is above the adjusted wet threshold (or $+1$ in case of no correction).

Once there is a reliable forecast of unusually wet or dry conditions, the intensity of the forecast is derived based on the extreme forecast index (EFI)–shift of tails (SOT) products, which were developed by ECMWF (Lalaurette 2003; Zsótér 2006; Owens and Hewson 2018). The EFI was developed by ECMWF to establish the severity of forecasted events, by indicating where the cumulative distribution function (CDF) of the ensemble forecast (ENS) substantially differs from climatology (represented by the CDF of the model climate). For our purposes, the mean forecasted SPI is computed for the

tails (i.e., the 40%) of the ensemble forecast's CDF. This adapted EFI method, which takes account of both the degree of coherence of ENS members (derived from the ensemble spread) to predict unusual anomalies, and their intensities, is particularly appropriate for extreme events, since by definition the method looks at the 40% of the ensemble forecasts CDF, and thus is less affected by anomalies closer to the ensemble median (i.e., the middle of the CDF).

In Fig. 1 (center panel), the red curve indicates the CDF of the SPI ensemble forecast, for the 51 sorted members, forecasted for a grid point during dry conditions, while the blue curve indicates the same but during wet conditions. As explained earlier, for the red curve the forecast of unusually dry conditions is considered robust when the 40th percentile (P40) is below -1 , while for the blue curve the forecast of unusually wet conditions is considered robust when the 60th percentile (P60) is above $+1$. The surface area bounded by the P40 or P60 members is then calculated (shaded red and blue areas), and provides the intensity of the forecasted dry or wet event, respectively.

The forecast product, which is derived from the integral of members in the tails of the CDF of the SPI ensemble forecast, is not straightforward to interpret nor to use, and this issue may be amplified by the spatial variability of the model. To address this point, the same calculation has been done for the full reforecast period of 36 years (i.e., 1981–2016) of the SEAS5 seasonal forecast system. Forecast intensity is then transformed into an equivalent return period (i.e., recurrence interval) for the event, highlighting regions where significant unusually precipitation conditions are forecasted. Three different warning levels are then derived, associated with return periods of 6 years (minimum detection level), 10 years, and 20 years (Fig. 1, right panel). These levels are sensitive to both the level of significance of the forecast (i.e., the coherence of the members) and the severity of the extreme condition (i.e., the intensity of the extreme ensemble members—defined as the second driest or wettest members, to remove potential outliers), as shown in Fig. 2. The results depict the increase of both the level of agreement (from less than 35% to over 45% of members associated with the forecast of an unusually event) and the intensity of the events, with the increase of the alert levels.

Validation of the forecast results is done using observed precipitations that are also standardized as SPI values. The reference period for the observed SPI values is similar to that for the forecast datasets [i.e., from 1981 to 2010 (although they are available up to 2017)]. Observed unusually dry conditions are considered to begin at an SPI value of -1 or less, and observed

FIG. 2. Relationship, for SPI-3, between (top) the different warning levels and the coherency of the ensemble members defined by the percentage of members below -1 or above 1 , for dry and wet conditions, respectively, and (bottom) the intensity defined as the SPI of the second driest or wettest member (defined as P90). Similar results are obtained for SPI-1 and SPI-6 but with increased (for SPI-1) and decreased (for SPI-6) coherency of members.

unusually wet conditions are considered to begin at an SPI value of $+1$ or more. These thresholds result in detection of a sufficiently robust number of events of important precipitation anomalies (i.e., about 16% of the sample size). Note that over dry regions, the low amount of precipitation tends to generate artifacts in the SPI computation. To keep only the robust SPI values, all grid cells with less than 50 mm of precipitation during a specific period have been excluded.

The extreme dependency score (EDS; Stephenson et al. 2008) is an integrated score derived from the dichotomous forecasts and observations. This score, adapted to rare events (Ghelli and Primo 2009), does not converge to zero when the number of events is low, but nevertheless ignores the false alarms (i.e., events predicted but not observed). Values above zero are considered better than the climatology, and reflect the

relevance of using these forecasts. The calculation is done as follows:

$$\text{EDS} = \frac{2 \log\left(\frac{a+c}{n}\right)}{\log\left(\frac{a}{n}\right)} - 1, \quad (1)$$

where a represents the “hits,” c is the “misses,” and n is the total number of events.

3. Results

In this section, the reliability of the ensemble forecasts is assessed through probabilistic skill scores, the dichotomous forecasts (i.e., alerts) of unusually wet and dry periods are evaluated, and the sensitivity of the forecast skill scores to different parameters—such as

FIG. 3. Reliability diagram and sharpness histogram of the seasonal forecast of (top) unusually dry conditions and (bottom) unusually wet conditions, for (left) SPI-1, (center) SPI-3, and (right) SPI-6. The color lines represent the latitudes. Midlatitudes: 30° – 60° N/S and tropical: 0° – 30° N/S.

climatic trends, initial conditions, warning levels, and seasons—is analyzed.

a. Assessment of forecast reliability

The performance of the ensemble system is assessed using reliability diagrams (Wilks 2011; Hamill 1997). According to Fig. 3, SEAS5 reveals good reliability of the forecasted SPI-1, with a linear relationship between the observed relative frequency and the forecast probability of unusually conditions. In general, this is particularly true for short lead times (SPI-1 and SPI-3), where the curves are closer to the “perfect” reliability (shown as gray lines). A slight overconfidence of the forecasts—whereby members converge too quickly to the same forecast compared to the percentage of observed events—can be observed. Despite the loss of predictability for the longest lead time (SPI-6), the ensemble system appears significantly reliable over the tropics. Nevertheless, some outliers with dry conditions over midlatitudes are evident.

The “sharpness” of the forecasts (i.e., the number of forecasts that fall in each forecast bin), which is illustrated in the bar plots of each panel in Fig. 3, highlights the main issue of these forecasts. The number of events associated with high consistency of the ensemble

members (i.e., high percentages) is low. Thus, according to the reliability diagram, forecasters could expect a high success ratio when the ensemble forecasts converge to an extreme event. Nevertheless, the low number of events indicates a high chance of the forecast system to record misses.

The impacts of the different latitudes are visible. Indeed, southern midlatitudes (30° – 60° S) are associated with the highest reliability for both wet and dry conditions. This result could be related to the fact that climates of continental in midlatitudes in Southern Hemisphere are more linked to Oceanographic oscillations. In the Northern Hemisphere Continental climates are wider. Nevertheless, the significance test, using resampling, shows low significance levels (about 80%). The differences are much larger when the decomposition for each season is done (see the online supplemental material). The seasonality of the skill scores is much larger for the predictability of dry conditions than that of wet conditions. Indeed, in the Northern Hemisphere winter [i.e., December–February (DJF)], the Northern Hemisphere has better skill for forecasted dry conditions. Opposite results are found in the Northern Hemisphere summer [June–August (JJA)]. The differences are not significant for forecasted wet conditions.

FIG. 4. ROC diagrams of the SEAS5 forecasts for (top) dry ($SPI < -1$) and (bottom) wet ($SPI > 1$), for (left) SPI-1, (center) SPI-3, and (right) SPI-6. Line colors represent the latitudes (see Fig. 3).

The results are relatively close from one season to another (not shown).

To quantify the capacity of the ensemble system to discriminate between events and nonevents, the relative operating characteristic (ROC)—which is a measure of the quality of probability forecasts that relates the hit rate to the corresponding false-alarm rate—is then calculated. In all of the results (depending on latitudes, wet or dry conditions, and the SPI accumulation period), the obtained ROC is better than the climatology (Fig. 4). Some general conclusions may be also drawn. The loss of skill with increasing SPI accumulation period that is evident for all cases, is to be expected because of the corresponding increase of lead time. The ROC results also highlight the better predictability for dry than wet conditions, especially for shorter accumulation periods (SPI-1 and SPI-3), and the influence of the latitudes. For dry conditions, the ROC area is higher for tropical

latitudes. The differences are lower for wet conditions, but the northern midlatitudes record the lowest scores for the wet SPI-3 and SPI-6. The ROC areas do not show an influence according to the seasons (see supplemental material).

b. Evaluation of warning levels for forecasted unusually wet and dry periods

To assess the forecast skill scores for the different warning levels, a second validation has been carried out using EDS. Results for the predictability of unusually dry SPI-1 conditions (Fig. 5, first line) display better scores over eastern Europe, Russia, the United States, and the eastern side of South Africa and Australia. In contrast, the corresponding predictability appears lower over Canada, Africa, central China, and the western part of Australia. Not surprisingly, the scores decrease for longer lead times (SPI-3 and SPI-6). For these

FIG. 5. Extreme dependency score (EDS) of the first level of warning of (top) unusually dry conditions and (bottom) unusually wet conditions, for (left) SPI-1, (center) SPI-3, and (right) SPI-6. Blank areas highlight dry regions that have been masked out due to insufficient statistical robustness of the SPI calculations.

accumulation periods, the spatial variability is lower, except in South America and Southeast Asia. It is worth noting that some spatial patterns can be similar to those obtained in previous studies using multimodel seasonal forecasts especially the increase of the skill scores over Australia, the northern part of South America, Mexico, western Russia, or Argentina (Yuan and Wood 2013).

The corresponding results for unusually wet SPI-1 conditions (Fig. 5, second line) reveal relatively similar scores in terms of intensity (i.e., slightly better scores for wet than dry conditions) and spatial variability. In eastern Africa and Australia, the predictability is larger for wet than for dry conditions, while for longer lead times (SPI-3 and SPI-6), the loss of predictability appears larger than for dry conditions. Some regions (i.e., Australia, South America, Southeastern Asia, and western Mexico) display a relative high predictability.

The origins of this higher predictability for both wet and dry conditions over certain regions may be related to a higher influence of tropical climatic variabilities that are more predictable. Indeed, the regions with higher predictability are those that are most affected by the cold and warm phases of the ENSO cycle (i.e., El Niño and La Niña, respectively): namely, the southern part of North America (Pan et al. 2018), northeastern and southern South America (Tedeschi et al. 2016), eastern and southern Africa (Manatsa et al. 2017), central and Southeastern Asia (Mariotti 2007), and Australia (Vicente-Serrano et al. 2011). In contrast, Europe, the northern part of North America, Africa, and Russia are less affected by these events.

Finally, it is worth highlighting the improvement of the new ECMWF seasonal forecast system (SEAS5) compared with the older version (S4). Indeed, the patterns found over South America look relatively similar to those obtained in (Carrão et al. 2018), but with a clear

increase of significance. The same scores as those used by these authors reveal this improvement between the two versions (not shown).

c. Sensitivity tests of forecast skill scores

One potential problem arising when using hindcasting (i.e., historical reforecasting) to build climatology, relates to the possible presence of “nonstationarity” in the precipitation data time series, which is manifested as a trend component and/or a sudden jump in the data’s statistical characteristics during the period, due to climate change effects. Indeed, in the context of climate change, the climatological trends of precipitation may introduce bias in the calculation of the forecasted SPI and in the return periods of very intense and extreme conditions.

To assess this effect, the evolution of forecasted and observed wet or dry conditions (i.e., SPI values above 1 or below -1 , respectively) was compared with the theoretical values (i.e., around 16% of the sampling used), for the period from 1981 to 2018 (Fig. 6). Due to the relatively low number of events, a small change may create large differences. Nevertheless, thanks to the number of SPI values calculated (i.e., each month during 36 years of hindcasts and 2 years of forecasts) and the number of grid cells ($1^\circ \times 1^\circ$ resolution) over the globe, these anomalies should be stabilized in the signal. To exclude arid regions, and focus on grid cells where the fitting of the gamma probability distribution of the cumulative rainfall is more reliable, the trends are calculated over grid points where more than 50 mm of rainfall were recorded during the accumulation period. As was already explained, the forecasted conditions are defined using the sorted ensemble members located at the 40th percentile of the cumulative distribution function (for dry conditions), and at the 60th percentile for wet

FIG. 6. Trend of occurrence of wet (SPI-3 > 1; blue) and dry (SPI-3 < -1; red) (a)–(l) forecasted and (m)–(p) observed conditions vs theoretical distribution (i.e., 16% of cases). Points display the monthly values of fraction and lines the filtered trend. Results are displayed according to the latitudes (lines). Tropics are 30°S–30°N, midlatitudes are 30°–60°N/S, polar north above 60°N, and polar south below 60°S. For the forecasted results, columns distinguish all grid points in (a), (d), (g), and (j), sea in (b), (e), (h), and (k), and land in (c), (f), (i), and (l).

FIG. 7. Seasonal differences of EDS scores (JJA–DJF) of the first warning level of (top) unusually dry or (bottom) wet conditions, using (a),(d) SPI-1, (b),(e) SPI-3, and (c),(f) SPI-6. Positive values indicate better scores during the JJA months. Grid cells with less than 50 mm of cumulated rainfall are excluded.

conditions. Regarding the grid cells over land, the forecasted grid cells (Fig. 6, third column) can be directly compared with those observed (Fig. 6, fourth column).

As can be seen in Fig. 6 (fourth row), globally there is a trend toward an increased occurrence of unusually wet conditions (blue lines), with a relatively linear annual increase since the 1990s. The occurrence of unusually dry conditions (red lines) appears relatively stable in time. These trends are very dependent on the latitude. For polar regions, over land there is a linear increase of unusually wet conditions, and a linear decrease of unusually dry conditions (Figs. 6c,m). Over the tropics, the variability is strongly associated with El Niño events over land (Figs. 6i,o), and especially over sea (Fig. 6h). For these latitudes no significant interdecadal trends are observed. The variability of the detected dry and wet events is closely linked with the specific 10-yr interval. These results also enable us to draw conclusions about the ability of the model to represent the climatological trend of precipitation. First, the trends are generally less pronounced in the model, and the variance of the monthly SPI values is reduced. Both of these differences could be explained due to the use of ensemble forecasts, which tend to reduce the variability from one month to another.

Some errors in the trends are also evident in Fig. 6. During the last 15 years, over midlatitudes, for unusually wet conditions a linear increase is forecasted, in agreement with the observed trends. However, during the same period, for unusually dry conditions the forecasted and observed trends are contradictory. Furthermore, over polar regions, for unusually dry conditions a linear decrease is forecast whereas no significant trend is observed. While this region is generally more difficult to analyze, because of the lack of observations and the uncertainties of some numerical schemes of the models

(e.g., convective scheme) over polar regions, this issue should nonetheless be studied in more detail, as the differences are quite large, and suggest an overestimation of the model regarding the impacts of the initial conditions. For other regions and trends, both the forecast and the observations are in quite good agreement.

It should be noted that some of the results underestimate unusually dry events, especially in the tropics (i.e., between 30°S and 30°N). This issue is mainly due to the quality of the SPI calculations when fitting the gamma distribution, which has intrinsically more uncertainty over dry regions and more sensitivity to small anomalies. The underestimation could be also due to the choice of the gamma distribution itself, as it is known that the cumulative density function of precipitation does not fit everywhere with this probability distribution (Lavaysse et al. 2015).

The climatological trends bring uncertainties in the forecasts because (i) as for the observations, the return period calculation is affected, and (ii) when the forecasted and observed trends are different, a bias is inserted. Nevertheless, quantification of the impact of this bias is impossible without making strong hypotheses that are not verified.

The seasonality in the predictability, based on the EDS, of unusually wet and dry periods has been analyzed for the three SPIs. To provide robust statistics, the focus is on the largest seasonal difference (i.e., JJA in relation to DJF). The differences of predictability reveal large-scale features (Fig. 7). Hence, all the scores are generally higher in DJF over the Northern Hemisphere (especially over central Russia, North America, and Greenland), and in JJA over the Southern Hemisphere (especially over Australia, Brazil, Southeast Asia). Over the tropics, seasonality is less pronounced, with a strong

FIG. 8. Boxplot of the EDS scores computed globally over the continental land grid cells of the different warning levels (x axis) of unusually (left) dry or (right) wet conditions, using SPI-1. Colors indicate observed initial conditions (given by the observed SPI-1 of the month prior the forecasts).

contrast between regions in both northern and southern hemispheres.

The impact of the initial conditions has also been quantified. For each forecast, the observed SPI-1 of the previous month is taken into account, with initial conditions considered dry for observed SPI lower than -1 , and wet for observed SPI larger than $+1$. Each of the scores for predictability of dry and wet conditions, depending on these initial conditions, is displayed in Fig. 8 (in red and blue boxes, respectively). The persistence of unusually conditions (i.e., forecasted dry events with dry initial conditions, and forecasted wet events with wet initial conditions) increases the predictability of the event, compared with normal or opposite initial conditions.

With the increase of the warning levels (Fig. 8), a loss of predictability by using the EDS is evident. This was also observed using another classical score—the equitable threat score, also called the Gilbert skill score (not shown here). This is mainly due to the increased number of misses. Nevertheless, the fraction of correctly forecast conditions—computed as the percentage correct (PC) or hit rate—increases, implying a decrease of false alarms (not shown). The previous conclusion regarding best predictability for persistence conditions is still observed.

These results highlight the influence of warning levels on the occurrence of observed events. Figure 9 displays the probability of distribution of the observed SPI-1, when the different warning levels are triggered. It shows that the occurrence of unusually dry (SPI-1 below -1) and wet (SPI-1 above 1) conditions is more frequent, as discussed previously. It also shows the increase of intensity, depicted by the increase of the frequencies of the most extreme cases, represented by SPI-1 below -2 for dry conditions, and above 2 for wet conditions. These differences are more pronounced for wet than dry warning levels, and similar results are found for the other SPI accumulation periods (not shown). These results show that the warning levels are associated with a higher percent correct (i.e., less false alarms) and stronger events (i.e., more intense dry or wet conditions).

Finally, analysis of the dependence of warning levels with regard to the different lead times (SPI-1, SPI-3, and SPI-6), indicates that there is a good chance of triggering warnings due to a large anomaly at the beginning of the lead time, which is similar for the different SPI accumulation periods: namely, the first month for all SPIs, and the first 3 months for SPI-3 and SPI-6. This is mainly

FIG. 9. Probability distribution functions of observed SPI-1 during different forecasted warning levels of SPI-1 over land.

related to the increase in ensemble spread with lead time, resulting in the detection of less significant anomalies for the end of the long lead time, and increasing the weight of coherent anomalies at the beginning of the lead time.

The associated comparison tables are provided in Tables 1–3, which indicate that, when the warning exists for the shorter lead time, the model increases the chance of event detection compared with the observed values (i.e., the numbers in parentheses in the tables). This is particularly true when comparing the shortest and longest lead times (i.e., SPI-1 and SPI-6). For example, in Table 2 it can be seen that, of the total of 16.6% of grid points that are forecast as unusually dry conditions based on SPI-6, 7.1% (i.e., almost half) are forecast as unusually dry conditions based on SPI-1. Similarly, of the total of 16.6% of grid points that are forecast as unusually wet conditions based on SPI-6, 8.0% (again, almost half) are forecast as unusually wet conditions based on forecasted SPI-1. In other words, around half of the unusually dry and wet conditions that are detected with the shortest lead time (1 month) are also detected with the longest lead time (6 months) over the same locations. According to Tables 1–3, it is also evident that the model overestimates the relationship between the long and short lead times compared with the observed values, and where the relationship does exist in the observed values, it is lower than in the forecast values.

4. Conclusions

In this study, the skill of a developed method for forecasting unusually wet and dry periods, for lead times

of 1–3 and 6 months, has been assessed. The forecasting methodology is derived from the ECMWF long-range (i.e., seasonal) forecast system (SEAS5), and is based on the extreme forecast index (EFI) of the SPI values, for three 1-, 3-, and 6-month accumulation periods (i.e., SPI-1, SPI-3, SPI-6). The forecasted SPI values are ranked, and an unusually dry or wet event is detected, based on a peak over threshold (POT) method, with the respective thresholds located at the 40th and 60th percentiles of the ensemble distribution. The EFI–shift of tails (SOT) index is then calculated as the mean SPI value for very intense wet and dry ensemble members. These values are transformed to return periods based on the 36-yr hindcast period (1981–2016). Three warning levels for unusually wet and dry conditions are defined, each linked with a return period of the forecast intensities: level 1, 6-yr return period (i.e., minimum alert level); level 2, 10-yr return period; level 3, 20-yr return period (i.e., maximum alert level).

TABLE 1. Comparison of events (i.e., unusually dry, normal, and unusually wet) detected with a forecast lead time of 3 months (SPI-3) vs with a forecast lead time of 1 month (SPI-1). Values are expressed as percentages of all grid points over land. The numbers in parentheses show the corresponding values based on observed SPI.

		SPI-3			
		Dry	Normal	Wet	Total
SPI-1	Dry	8.9 (6.0)	7.2 (6.8)	0.1 (0.5)	16.2 (13.3)
	Normal	7.5 (8.6)	53.5 (54.6)	6.4 (7.6)	67.5 (70.7)
	Wet	0.1 (0.1)	6.2 (8.0)	10.0 (7.9)	16.3 (16)
	Total	16.5 (14.7)	67 (69.3)	16.5 (16)	100 (100)

TABLE 2. Comparison of events (i.e., unusually dry, normal, and unusually wet) detected with a forecast lead time of 6 months (SPI-6) vs with a forecast lead time of 1 month (SPI-1). Values are expressed as percentages of all grid points over land. The numbers in parentheses show the corresponding values based on observed SPI.

		SPI-6			
		Dry	Normal	Wet	Total
SPI-1	Dry	7.1 (4.7)	8.7 (7.7)	0.5 (0.8)	16.2 (13.3)
	Normal	9.2 (9.9)	50.1 (51.8)	8.1 (9.1)	67.5 (70.7)
	Wet	0.3 (0.5)	7.9 (9.3)	8.0 (6.2)	16.3 (16)
	Total	16.6 (15.1)	66.8 (68.8)	16.6 (16.1)	100 (100)

Assessment of the ensemble skill scores shows the forecasting method to be reliable, with significant and robust information provided by SEAS5 for predicting unusually wet and dry events, even for longer lead times. Despite some uncertainties, validation of the alert levels based on hindcasting shows that the different forecast warnings of unusually wet and dry conditions—which are currently implemented within the European Commission’s Copernicus Emergency Management Service (EMS)—provide important and relevant information for end users and decision-makers. As highlighted earlier, this information is especially valid over regions that are strongly influenced by El Niño and La Niña climate events.

Based on an analysis of different parameters, some sensitivity issues were assessed, which indicate that the forecast skill scores appear to be dependent on specific conditions. Generally speaking, predictability is higher for unusually wet than for unusually dry conditions, for all SPI lead times. For the larger alarm levels, the percentage of forecast correct (PC) scores are higher, but globally the scores are lower than the first alarm level, because of the misses. Generally, the predictability is better during the local winter season in the midlatitudes, and when persistence occurs (i.e., dry events forecasted during dry initial conditions).

TABLE 3. Comparison of events (i.e., unusually dry, normal, and unusually wet) detected with a forecast lead time of 6 months (SPI-6) vs with a forecast lead time of 3 months (SPI-3). Values are expressed as percentages of all grid points over land. The numbers in parentheses show the corresponding values based on observed SPI.

		SPI-6			
		Dry	Normal	Wet	Total
SPI-3	Dry	9.8 (8.2)	6.6 (6.1)	0.1 (0.3)	16.5 (14.7)
	Normal	6.7 (6.8)	54.2 (56.3)	6.0 (6.4)	67 (69.3)
	Wet	0.1 (0.1)	5.9 (6.5)	10.5 (9.4)	16.5 (16)
	Total	16.6 (15.5)	66.8 (68.8)	16.6 (16.1)	100 (100)

Nevertheless, especially over the midlatitudes, the benefic in predictability is present but appears quite low compared to the climatology. The forecasting of unusually conditions using atmospheric predictors (i.e., weather regimes)—which are more predictable, and could drive extreme precipitation events (Lavayasse et al. 2018)—could be tested and verified against these results at global scale.

Acknowledgments. The authors thank S. Johnson (ECMWF) and A. Robertson (IRI) for the discussions and their fruitful comments. The authors thank also the two anonymous reviewers for their evaluation of this paper and their positive comments and suggestions.

REFERENCES

- Angus, M., and G. Leckebusch, 2018: Assessing the relationship between Atlantic hurricanes and European winter windstorms in a seasonal forecast ensemble. *2018 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract A23H-2939.
- Arribas, A., and Coauthors, 2011: The GloSea4 ensemble prediction system for seasonal forecasting. *Mon. Wea. Rev.*, **139**, 1891–1910, <https://doi.org/10.1175/2010MWR3615.1>.
- Barnston, A. G., M. K. Tippett, M. L. L’Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, <https://doi.org/10.1175/BAMS-D-11-00111.1>.
- Below, R., E. Grover-Kopce, and M. Dille, 2007: Documenting drought-related disasters: A global reassessment. *J. Environ. Dev.*, **16**, 328–344, <https://doi.org/10.1177/1070496507306222>.
- Boer, G., V. Kharin, and W. Merryfield, 2013: Decadal predictability and forecast skill. *Climate Dyn.*, **41**, 1817–1833, <https://doi.org/10.1007/s00382-013-1705-0>.
- Buizza, R., P. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, <https://doi.org/10.1175/MWR2905.1>.
- Carrão, H., G. Naumann, E. Dutra, C. Lavayasse, and P. Barbosa, 2018: Seasonal drought forecasting for Latin America using the ECMWF S4 forecast system. *Climate*, **6**, 48, <https://doi.org/10.3390/cli6020048>.
- CRED/UNDRR, 2015: The human cost of weather-related disasters, 1995–2015. United Nations, 30 pp.
- Dutra, E., F. D. Giuseppe, F. Wetterhall, and F. Pappenberger, 2013: Seasonal forecasts of droughts in African basins using the standardized precipitation index. *Hydrol. Earth Syst. Sci.*, **17**, 2359–2373, <https://doi.org/10.5194/hess-17-2359-2013>.
- , and Coauthors, 2014: Global meteorological drought—Part 2: Seasonal forecasts. *Hydrol. Earth Syst. Sci.*, **18**, 2669–2678, <https://doi.org/10.5194/hess-18-2669-2014>.
- Edwards, D. C., 1997: Characteristics of 20th century drought in the United States at multiple time scales. M.S. thesis, Department of Atmospheric Science, Colorado State University, 155 pp.
- Emerton, R., and Coauthors, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0. *Geosci. Model Dev.*, **11**, 3327–3346, <https://doi.org/10.5194/gmd-11-3327-2018>.

- Fraser, E. D., E. Simelton, M. Termansen, S. N. Gosling, and A. South, 2013: "Vulnerability hotspots": Integrating socio-economic and hydrological models to identify where cereal production may decline in the future due to climate change induced drought. *Agric. For. Meteorol.*, **170**, 195–205, <https://doi.org/10.1016/j.agrformet.2012.04.008>.
- Ghelli, A., and C. Primo, 2009: On the use of the extreme dependency score to investigate the performance of an NPW model for rare events. *Meteor. Appl.*, **16**, 537–544, <https://doi.org/10.1002/met.153>.
- Hamill, T. M., 1997: Reliability diagrams for multicategory probabilistic forecasts. *Wea. Forecasting*, **12**, 736–741, [https://doi.org/10.1175/1520-0434\(1997\)012<0736:RDFMPF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2).
- , M. J. Brennan, B. Brown, M. DeMaria, E. N. Rappaport, and Z. Toth, 2012: NOAA's future ensemble-based hurricane forecast products. *Bull. Amer. Meteor. Soc.*, **93**, 209–220, <https://doi.org/10.1175/2011BAMS3106.1>.
- Johnson, S. J., and Coauthors, 2019: SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.*, **12**, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>.
- Lalurette, F., 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quart. J. Roy. Meteor. Soc.*, **129**, 3037–3057, <https://doi.org/10.1256/qj.02.152>.
- Lavaysse, C., J. Vogt, and F. Pappenberger, 2015: Early warning of drought in Europe using the monthly ensemble system from ECMWF. *Hydrol. Earth Syst. Sci.*, **19**, 3273–3286, <https://doi.org/10.5194/hess-19-3273-2015>.
- , —, A. Toreti, M. L. Carrera, and F. Pappenberger, 2018: On the use of weather regimes to forecast meteorological drought over Europe. *Nat. Hazards Earth Syst. Sci.*, **18**, 3297–3309, <https://doi.org/10.5194/nhess-18-3297-2018>.
- Lloyd-Hughes, B., and M. A. Saunders, 2002: A drought climatology for Europe. *Int. J. Climatol.*, **22**, 1571–1592, <https://doi.org/10.1002/joc.846>.
- Lyon, B., M. A. Bell, M. K. Tippett, A. Kumar, M. P. Hoerling, X.-W. Quan, and H. Wang, 2012: Baseline probabilities for the seasonal prediction of meteorological drought. *J. Appl. Meteor. Climatol.*, **51**, 1222–1237, <https://doi.org/10.1175/JAMC-D-11-0132.1>.
- Manatsa, D., T. Mushore, and A. Lenouo, 2017: Improved predictability of droughts over southern Africa using the standardized precipitation evapotranspiration index and ENSO. *Theor. Appl. Climatol.*, **127**, 259–274, <https://doi.org/10.1007/s00704-015-1632-6>.
- Mariotti, A., 2007: How ENSO impacts precipitation in southwest Central Asia. *Geophys. Res. Lett.*, **34**, L16706, <https://doi.org/10.1029/2007GL030078>.
- McKee, T. B., and Coauthors, 1993: The relationship of drought frequency and duration to time scales. *Proc. Eighth Conf. on Applied Climatology*, Boston, MA, Amer. Meteor. Soc., 179–183.
- Owens, R., and T. Hewson, 2018: ECMWF forecast user guide. ECMWF, Reading, United Kingdom, <https://doi.org/10.21957/m1cs7h>.
- Palmer, C., M. Müller, and A. Melsom, 2019: An intercomparison of verification scores for evaluating the sea ice edge position in seasonal forecasts. *Geophys. Res. Lett.*, **46**, 4757–4763, <https://doi.org/10.1029/2019GL082482>.
- Pan, Y., N. Zeng, A. Mariotti, H. Wang, A. Kumar, R. L. Sánchez, and B. Jha, 2018: Covariability of central America/Mexico winter precipitation and tropical sea surface temperatures. *Climate Dyn.*, **50**, 4335–4346, <https://doi.org/10.1007/s00382-017-3878-4>.
- Ratri, D., and M. Schmeits, 2018: A comparative verification of raw and bias-corrected ECMWF seasonal ensemble precipitation forecasts in Java (Indonesia). *20th EGU General Assembly Conf.*, Vienna, Austria, EGU, Vol. 20, EGU2018-17339, <https://meetingorganizer.copernicus.org/EGU2018/EGU2018-17339.pdf>.
- Stephenson, D. B., B. Casati, C. Ferro, and C. Wilson, 2008: The extreme dependency score: A non-vanishing measure for forecasts of rare events. *Meteor. Appl.*, **15**, 41–50, <https://doi.org/10.1002/met.53>.
- Stockdale, T. N., D. L. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**, 370–373, <https://doi.org/10.1038/32861>.
- , and Coauthors, 2010: Understanding and predicting seasonal-to-interannual climate variability—the producer perspective. *Procedia Environ. Sci.*, **1**, 55–80, <https://doi.org/10.1016/j.proenv.2010.09.006>.
- Tedeschi, R. G., A. M. Grimm, and I. F. Cavalcanti, 2016: Influence of central and east ENSO on precipitation and its extreme events in South America during austral autumn and winter. *Int. J. Climatol.*, **36**, 4797–4814, <https://doi.org/10.1002/joc.4670>.
- Vicente-Serrano, S. M., J. I. López-Moreno, L. Gimeno, R. Nieto, E. Morán-Tejeda, J. Lorenzo-Lacruz, S. Beguería, and C. Azorin-Molina, 2011: A multiscale global evaluation of the impact of ENSO on droughts. *J. Geophys. Res.*, **116**, D20109, <https://doi.org/10.1029/2011JD016039>.
- Vitart, F., G. Balsamo, R. Buizza, L. Ferranti, S. Keeley, L. Magnusson, F. Molteni, and A. Weisheimer, 2014: Sub-seasonal predictions. ECMWF Tech. Memo. 738, ECMWF, 47 pp.
- Weisheimer, A., and T. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 20131162, <https://doi.org/10.1098/rsif.2013.1162>.
- , D. Decremier, D. MacLeod, C. O'Reilly, T. N. Stockdale, S. Johnson, and T. N. Palmer, 2018: How confident are predictability estimates of the winter North Atlantic oscillation? *Quart. J. Roy. Meteor. Soc.*, **145**, 140–159, <https://doi.org/10.1002/qj.3446>.
- Wilhite, D. A., M. V. Sivakumar, and R. Pulwarty, 2014: Managing drought risk in a changing climate: The role of national drought policy. *Wea. Climate Extremes*, **3**, 4–13, <https://doi.org/10.1016/j.wace.2014.01.002>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp..
- WMO, 2012: Standardized precipitation index, user guide. Tech. Rep. 1090, 16 pp.
- Yuan, X., and E. F. Wood, 2013: Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett.*, **40**, 4900–4905, <https://doi.org/10.1002/grl.50949>.
- Zsótér, E., 2006: Recent developments in extreme weather forecasting. *ECMWF Newsletter*, No. 107, ECMWF, Reading, United Kingdom, 8–17, <https://doi.org/10.21957/ki9821hnc7>.