European Commission

JRC TECHNICAL REPORTS

Drought Forecasting for Latin America

Desertification, Land Degradation and Drought (DLDD), and bio-physical modelling for crop yield estimation in Latin America under a changing climate

Deliverable No. 6

Carrão, H., Naumann, G., and Barbosa, P.

2016

Joint Research Centre

EUR 28260 EN

Contact information
Name: Paulo Barbosa
Email: paulo.barbosa@jrc.ec.europa.eu

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

JRC104401

EUR 28260 EN

PDF	ISBN 978-92-79-64025-4	ISSN 1831-9424	doi:10.2788/826147
Print	ISBN 978-92-79-64024-7	ISSN 1018-5593	doi:10.2788/928

Luxembourg: Publications Office of the European Union, 2016

How to cite this report: Carrão, H., Naumann, G., and Barbosa, P., *Drought Forecasting for Latin America*, EUR 28260 EN; doi:10.2788/826147

Contents

List of Figures

List of Tables

1 Executive Summary

This Technical Report was developed in the framework of Component 3 of the second phase of the Programme EUROCLIMA: “Sustainable Agriculture, Food Security and Climate Change in Latin America: Strengthening the capacities of key stakeholders to adapt agriculture to climate change and mitigate its effects”. EUROCLIMA is a regional cooperation program between the European Union and Latin America aiming at facilitating the integration of mitigation and adaptation strategies into climate change public policies and development plans in Latin America. In the framework of EUROCLIMA, EU development assistance funding has been provided through the Commission’s Directorate General for International Cooperation and Development (AA JRC No. 2013/332-909) to work on the topics of Desertification, Land Degradation and Drought (DLDD), as well as on bio-physical modelling for crop yield estimation in Latin America.

In this study, precipitation predictions from the European Centre for Medium Range Weather (ECMWF) seasonal forecast system (System 4) are combined with observed precipitation data to generate forecasts of the standardized precipitation index (SPI) for Latin America, and their skill is evaluated over the hincast period 1981–2010. The value-added utility in using the ensemble S4 forecast to predict the SPI is identified by comparing the skill of its forecasts with a baseline skill based solely on the climatological characteristics of the SPI itsdf. As expected, skill of the S4-generated SPI forecasts depends on the month, location, and specific index considered (the 3- and 6-month SPI were evaluated). Added skill from the S4 for lead times equaling the SPI accumulation periods is primarily in regions with high intra-annual precipitation variability, and is found mostly for the months at the end of the dry seasons for 3-month SPI, and half-yearly periods for 6-month SPI. Thus, in the near term, the largest advances in the prediction of meteorological drought for Latin America are obtainable from improvements in near-real-time precipitation observations for the region. In the longer term, improvements in precipitation forecast skill from dynamical models will be essential in this effort.

2 Introduction

Drought is a recurring and extreme climate event that is originated by a temporary water deficit and may be related to a lack of precipitation, soil moisture, streamflow, or any combination of the three taking place at the same time [3]. Drought differs from other hazard types in several ways. First, unlike earthquakes, floods or tsunamis that occur along generally well-defined fault lines, river valleys or coastlines, drought can occur anywhere (with the exception of desert regions and cold areas where it does not have meaning) [4, 5]. Secondly, drought develops slowly, resulting from a prolonged period (from weeks to years) of precipitation that is below the average, or expected, value at a particular location [6, 7].

To improve drought mitigation, different indicators are used to trigger a drought [8, 9]. While an indicator is a derived variable for identifying and assessing different drought types, a trigger is a threshold value of the indicator used to determine the onset, intensity or end of a drought, as well as the timing to implement proper drought response actions [10, 11]. Since precipitation is one of the most important inputs to a watershed system and provides a direct measurement of water supply conditions over different timescales, several commonly used drought indicators rely on precipitation measurements only [6, 10]. Among them, the Standardized Precipitation Index (SPI) of [1] is certainly the most prominent and has been recommended by the World Meteorological Organization (WMO) for characterizing the onset, end, duration and severity of drought events deriving from precipitation deficiencies taking place at different accumulation periods and occurring at different stages of a same hydro-meteorological anomaly [12].

The immediate consequences of short-term droughts (i.e. a few weeks duration) are, for example, a fall in crop production, poor pasture growth and a decline in fodder supplies from crop residues, whereas prolonged water shortages (e.g. of several months or years duration) may, among others, lead to a reduction on hydro-electrical production and an increase of forest fire occurrences [13]. Therefore, predicting the onset and end of a drought a few months in advance will benefit a variety of sectors by allowing sufficient lead time for drought mitigation efforts. Indeed, drought forecasting is nowadays a critical component of drought hydrology science, which plays a major role in drought risk management, preparedness and mitigation. It has been demonstrated that droughts can be forecast using stochastic or neural networks [14, 15]. While [16] demonstrated that these type of forecast can provide "reasonably good agreement for forecasting with 1 to 2 months lead times", they do not quantify the improvement of these methods with respect to using probabilistic forecasts of the precipitation fields. Forecasts of droughts can also be produced using deterministic numerical weather prediction models. However, such forecasts are highly uncertain due to the chaotic nature of the atmosphere, which is particularly strong on a sub-seasonal timescale [17].

As an alternative, ensemble prediction systems that forecast multiple scenarios of future weather have considerably evolved over recent years. Indeed, the routine generation of global seasonal climate forecasts coupled with advances in near-real-time monitoring of the global climate has now allowed for testing the feasibility of generating global drought forecasts operationally. Systems to monitor drought around the globe are described in [11] for meteorological drought and [18] for hydrologic and agricultural conditions. For example, [19] used seasonal precipitation forecasts from the North American Multi-Model Ensemble (NMME) and other coupled ocean-land-atmosphere general circulation models (CGCMs) to examine the predictability of drought onset around the globe based on the SPI, applying an onset definition used by [20]. For the global domain, they found only a modest increase in the forecast probability of onset relative to baseline expectations when using the GCM forecasts. [21] described the Global Integrated Drought Monitoring and Prediction System (GIDMaPS) that uses three drought indicators. The forecasting component of their system relies on a statistical approach based on an ensemble streamflow prediction (ESP) methodology. More recently, [22, 23] generated global forecasts of 3-, 6-, and 12-month SPI by combining seasonal precipitation reforecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) System 4 (S4) with precipitation observations

from the Global Precipitation Climatology Centre (GPCC) and, alternatively, the ECMWF Interim Reanalysis. They reported on several verification metrics for the SPI forecasts for 18 regions around the globe. Using the same definition as [19], they found that the ECMWF S4 provides useful skill in predicting drought onset in several regions, and the skill is largely derived from *El Ninõ*-Southern Oscillation (ENSO) teleconnections. However, they also found that it is difficult to improve on "climatological" forecasts generally.

In this deliverable, we build on the work of [22, 23] by considering a ECMWF S4 ensemble framework to generate seasonal forecasts of the SPI, and perform their verification against corresponding SPI from precipitation observations of the GPCC over Latin America. Drought is viewed from a meteorological perspective, and seasonal forecasts of the 3- and 6-month SPI (SPI3 and SPI6) are generated and verified on a monthly basis for the hindcast period of 1981–2010. While the focus of the work is on the prediction of meteorological drought, the study assesses two fundamental constraints in generating reliable regional drought predictions that will arise whether using the reported method or any other approach (e.g. land surface modeling): 1) the accuracy of summary statistics (e.g. mean, median, percentile) at predicting a seasonal drought from the members of the ensemble forecasting system and 2) the skill of probabilistic categorical predictions of seasonal drought from the members of the ensemble forecasting system.

This report is organized as follows: section 3 describes the study area, data and methods used in this study. In Section 4, we perform the verification of the seasonal forecast of drought from SPI3 and SPI6 computed with the ECMWF S4, namely their monthly and overall accuracy, as well as the skill of categorical predictions for South-Central America. We conclude the deliverable with a summary of main achievements and their implications for drought risk management on Section 5.

3 Study area, Datasets and Statistical Methods

3.1 Study area

The study area covers the whole South–Central America region (the domain of analysis is limited to land surface grid points between 56°S–35°N, 33°–120°W). South–Central America spans a vast range of latitudes and has a wide variety of climates. It is characterized largely by humid and tropical conditions, but important areas have been extremely affected by meteorological droughts in the past [24, 25, 26, 27] and the climate change scenarios foresee an increased frequency of these events for the region [28, 29]. Given the significant reliance of South–Central American economies on rainfed agricultural yields (rainfed crops contribute more than 80% of the total crop production in South-Central America [30]), and the exposure of agriculture to a variable climate, there is a large concern in the region about present and future climate and climate-related impacts [31]. South–Central American countries have an important percentage of their GDP in agriculture (10% average [30]), and the region is a net exporter of food globally, accounting for 11% of the global value [32]. According to the agricultural statistics supplied by the United Nations Food and Agriculture Organization (UN FAO) [30], 65% of the world production of corn and more than 90% of the world production of soybeans are grown in Argentina, Brazil, the United States and China. The productivity of these crops is expected to decrease in the extensive plains located in middle and subtropical latitudes of South–America America (e.g. Brazil and Argentina), leading to a reduction in the worldwide productivity of cattle farming and having adverse consequences to global food security [28, 33].

3.2 Drought Indicator: The Standardized Precipitation Index (SPI)

In this study, we selected the standardized precipitation index (SPI) [1] as a meteorological drought indicator. The SPI is a statistical indicator that compares the total precipitation received at a particular location during a period of time with the long-term precipitation distribution for the same period of time at that location. In order to allow for the statistical comparison of wetter and drier climates, the SPI is based on a transformation of the accumulated precipitation into a standard normal variable with zero mean and variance equal to one. SPI results are given in units of standard deviation from the long-term mean of the standardized precipitation distribution. Negative values, therefore, correspond to drier periods than normal and positive values correspond to wetter periods than normal. The fundamental strength of the SPI is that it can be calculated for a variety of precipitation timescales (e.g. weekly, monthly, seasonal or yearly accumulation periods) and updated on various time steps (e.g. daily, weekly, monthly), enabling water supply anomalies relevant to a range of end users to be readily identified and monitored. SPI is typically calculated on a monthly basis for a moving window of n months, where n indicates the precipitation accumulation period. For example, the SPI computed from 1-month precipitation totals (SPI-1) is mainly a meteorological drought indicator [34, 12], while the SPI-3 (based on 3-month precipitation totals) is suitable to monitor seasonal dryness in the surface layers (root zone) that are important for agricultural management [10, 35]. Longer aggregation periods (e.g. SPI-6, SPI-12, SPI-24) are important, for example, to monitor groundwater levels and surface water supplies [15, 35].

The magnitude of negative SPI values correspond to percentiles of a probability distribution that are frequently used as threshold levels (triggers) to classify drought intensity [1, 36, 37, 38]. Several classification systems of meteorological drought intensity basedon fixed threshold levels of the SPI have been presented in the literature. The most widely known is that proposed by [1], which maps precipitation totals below the 50^{th} percentile into four fixed categories of drought intensity (Table 3.1). For example, a "moderate" drought event starts at SPI=-1.0 (units of standard deviation), which corresponds to a cumulative probability of 15.9 %, i.e. approximately the 16^{th} percentile. [1] determined that every region is in "mild" drought 34% of the time, in

"moderate" drought 9.2% of the time, in "severe" drought 4.4% of the time, and in "extreme" drought 2.3% of the time (Table 3.1). The threshold levels of drought intensity proposed by [1] have been used worldwide in numerous applications at different timescales of precipitation accumulation, such as to monitor drought in the United States [39, 40, 41] and Europe [42], to calculate a drought climatology for Europe [43], for detecting droughts in East Africa [44], to monitor drought conditions and their uncertainty in Africa using data from the Tropical Rainfall Measuring Mission (TRMM) [45], to evaluate the transition probabilities of drought events in the Kansabati River basin in India [15] and in the Alentejo region in Portugal [46], for improving the fire danger forecast in the Iberian Peninsula [47] and to assess drought intensity in low and high precipitation districts of Andhra Pradesh in India [48], to cite but a few.

Table 3.1: *SPI Classification following [1].*

SPI Value	Class	Cumulative Probability	Probability of Event [%]
SPI>2.00	Extreme wet	0.977 - 1.000	2.3%
1.50<SPI≤2.00	Severe wet	0.933 - 0.977	4.4%
1.00<SPI≤1.50	Moderate wet	0.841 - 0.933	9.2%
-1.00<SPI≤1.00	Near normal	0.159 - 0.841	68.2%
-1.50<SPI≤-1.00	Moderate dry	0.067 - 0.159	9.2%
-2.00<SPI≤-1.50	Severe dry	0.023 - 0.067	4.4%
SPI<-2.00	Extreme dry	0.000 0.023	2.3%

The computation of the SPI follows a three-stage process (Figure 3.1):

1. a parametric statistical distribution is fitted to the long-term record of precipitation observations x_{t1}, x_{t2},..., x_{tn}; this is performed for an averaging timescale of t months (where t is typically 1-, 3-, 6-, or 12-months) collected over n years. In the classical SPI definition, precipitation amount records are fitted to a Gamma distribution. A Gamma distributed variable X is continuous and positive and has a probability density function defined by two parameters as follows:

$$f(x) = \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-\frac{x}{s}}, \quad \text{for } x \geq 0 \text{ and } a, s > 0, \tag{3.1}$$

   where s and a are respectively the scale and shape parameters, and $\Gamma(a)$ is the mathematical Gamma function. The Gamma distribution with parameters s and a is denoted as Gamma(s, a). The expectation and variance of a $X \sim$ Gamma(s, a) variable are:

$$E(X) = \mu = a \cdot s; \; Var(X) = a \cdot s^2. \tag{3.2}$$

2. the non-exceedance probability of a precipitation observation x_t is computed related to the respective Gamma distribution: this is simply done by estimating the cumulative probability $\hat{F}(x_t)$ of the precipitation observation x_t;

3. the non-exceedance probability is transformed to the standard normal variable Z ($mean = 0$ and $variance = 1$) and the SPI value is found.

In arid regions with many months with zero precipitation, the SPI needs to be interpreted with care [41]. Therefore, for locations where observations of zero precipitation occur, the fitting of the Gamma distribution becomes problematic since it is not defined for zero. The zero values must be cut out from the data and in this case the cumulative probability $F(x)$ becomes:

$$G(x) = q + (1 - q)F(x) \tag{3.3}$$

where q is the probability of zero precipitation calculated from the frequency of observations of zero, and $F(x)$ is the cumulative probability derived from the Gamma distributi on fitted to the non-zero precipitation data.

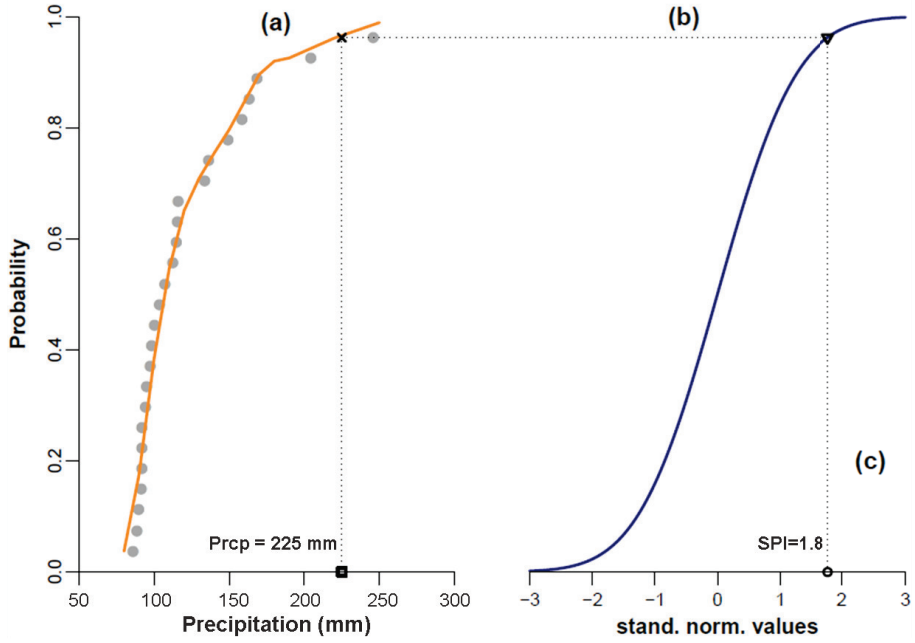


Figure 3.1: *The three step computation process of the SPI. The example is for November values at a grid point in Bahia, Brazil (38.875° W, 10.875° S). The orange curve on the left represents the $\hat{F}(x_t)$ related to the Gamma density function fitted by means of a ML to GPCC precipitation obervations collected in the period between 1981 and 2010 (gray filled circles). The blue curve on the right represents the standard normal cumulative distribution function (CDF). The non-exceedance probability is estimated in Step (a) for a precipitation amount of 225 mm, transformed to the standard normal variable Z in Step (b), and the SPI value (1.8) is found in Step (c).*

3.3 Precipitation datasets

3.3.1 Observations: The GPCC Full Data Reanalysis Version 6.0

In this study, monthly precipitation totals at 1.0° latitude/longitude grid spacing from the Full Data Reanalysis Monthly Product Version 6.0 of the Global Precipitation Climatology Centre (GPCC) are used as a reference data set (for the forecast verification). The GPCC was established in 1989 on request of the World Meteorological Organization (WMO) and provides a global gridded analysis of monthly precipitation over land from operational *in situ* rain-gauges based on the Global Telecommunications System (GTS) and historic precipitation data measured at global stations. The data supplies from 190 worldwide national weather serv ices to the GPCC are regarded as primary data source, comprising observed monthly totals from 10,700 to more than 47,000 stations since 1901. The monthly gridded data sets are spatially interpolated with a spherical adaptation of the robust Shepard's empirical weighting method [49].

Validation of the original data sets for drought monitoring has been performed by [50, 45, 22], who found that GPCC data sets show higher values for extreme precipitation, and tend to over-smooth the data. This can generate some problems when analysing intense precipitation events but appears of secondary importance in drought analysis. Therefore, to be consistent with the data provided by the ensembles from ECMWF, a common period of the hindcast that covers the period from 1981 to 2010 is used to calculate the SPI.

3.3.2 Forecasts: The ECMWF seasonal forecast system

In this study, we use the ECMWF seasonal forecast system 4 (hereafter S4; [51]) to forecast 3- and 6-month SPI. The S4 is a dynamical forecast system based on an atmospheric-ocean coupled model, which has been operational at ECMWF since 2011 and is launched once a month (on the first day of the month). The 2011 version of the forecast model has 91 vertical levels, lead times up to 13 months, and a resolution of T255 ($\approx$ 80 km). It provides back integrations (hindcasts) with 15/51 member ensemble (number depends on month) for every month from 1980 onwards. [51] provide a detailed overview of S4 performance.

For the comparison with the GPCC observations, the S4 has been re-gridded to 1.0° latitude/longitude grid spacing, and daily precipitation values over its hindcast period (1981–2010) have been aggregated to monthly values. The ability of the probabilistic model to accurately forecast seasonal drought conditions has been evaluated up to 6 months of lead time. In addition to the dynamical seasonal forecasts, climatological forecasts (CLM) were also generated by randomly sampling past years from the reference data set to match the number of ensemble members in the hindcast.

3.4 Verification methods

Forecast verification is the process of assessing the quality of forecasts. The usefulness of forecasts to support decision making clearly depends on their error characteristics, which are elucidated through forecast verification methods. A wide variety of forecast verification methods exist, and all involve measures of the relationship between a forecast or set of forecasts, and the corresponding observation(s) of the predictand. In this study, the forecasts correspond to the monthly SPI3 and SPI6 values computed with the ECMWF S4 for the period between 1981–2010; the observations correspond to the SPI3 and SPI6 values computed with the GPCC for the same historical period.

3.4.1 Nonprobabilistic Forecasts of Continuous SPI Values

We first verify the scalar accuracy of the SPI values for the multimodel ensemble mean at 3 and 6 months lead time (respectively for SPI3 and SPI6). Ensemble mean SPI values are verified against observations for the hindcast period (i.e. from 1981 to 2010). In this case, the SPI magnitude can take any value in a specified segment of the real line, rather than being limited to a finite number of discrete classes (see Table 3.1). We perform an independent verification of drought forecasts for each month, by using four common accuracy measures of continuous nonprobabilistic forecasts, namely: the Pearson product-moment correlation coefficient, r; the Mean Error, ME; the Root Mean Squared Error, RMSE; and the Willmott's index of agreement [52, 53]. To be considered statistically significant at the 5% (10%) confidence level, the r between forecast SPI values and those in the verifying GPCC data needs to be greater than 0.37 (0.31), as defined by [54] for $N_{years} = 29$ observations (i.e. after subtracting 1 year from the total number of available years in the dataset).

Although the correlation does reflect linear association between two variables (in this case, forecasts and observations), it is not sensitive to biases that may be present in the forecasts. On the other hand, the ME, which is simply the difference between the average forecast and average observation, expresses the bias of the forecasts [55]. Forecasts that are, on average, too high will exhibit ME$>$ 0 and forecasts that are, on average, too low will exhibit ME$<$ 0. It is important to note that the bias gives no information about the typical magnitude of individual forecast errors, and is therefore not in itself an accuracy measure. To complement the ME, we have computed the RMSE, which has the same physical dimensions as the forecasts and observations, and can be thought of as a typical mean magnitude for individual forecast errors. In addition, we also compute the Willmott's index of agreement. The advantage over the RMSE is that the errors and differences are given their appropriate weighting, and therefore are not inflated by their

squared values [56]. The index of agreement is dimensionless quantity that varies from 0 to 1, with higher index values indicating that the modelled values have better agreement with the observations.

We also verify the skill score for the multimodel ensemble mean at 3 and 6 months lead time (respectively for SPI3 and SPI6). Skill score refers to the relative accuracy of an ensemble set of forecasts and is interpreted as the improvement over a reference forecast [55]. Therefore, if the ECMWF S4 is providing value-added skill to the SPI forecasts, it will first be manifested by temporal correlations with observations, r_1, that exceed the expected correlation of the same observations with the climatological SPI baseline value (0), r_2. Under the assumption that the sets of forecasts are normally distributed, to assess the statistical significance of the difference between two correlations r_1 and r_2, we used Fisher's Z transformation, as explained in [57]. We define Z_i as

$$Z_i = \frac{1}{2}\ln(\frac{1+r_i}{1-r_i}) \tag{3.4}$$

for $i = 1$ and 2. The transformation Z is assumed to be normally distributed with variance $(N-3)-1$, where $N = 29$ observations (i.e. after subtracting 1 year from the total number of available years in the dataset). We then transformed r_1 and r_2 to Z_1 and Z_2, and computed the statistical significance for the difference in correlations using the Z statistics:

$$Z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}} \tag{3.5}$$

where $N_1 - 3$ and $N_2 - 3$ are the degrees of freedom for r_1 and r_2, respectively. Here $N_1 = N_2 = 29$. Using a null hypothesis of equal correlation and a nondirectional alternative hypothesis of unequal correlation, if Z is greater than 1.96, the difference in correlations is statistically significant at the 5% confidence level.

A complementary skill score measure was constructed using the RMSE as the underlying accuracy statistic. The reference RMSE is based on the climatological average $\overline{SPI}$, and is computed as:

$$RMSE_{Clim} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\overline{SPI} - SPI_k)^2} \tag{3.6}$$

For the SPI, the climatological average $\overline{SPI}$ does not change from forecast occasion to forecast occasion (i.e. as a function of the yearly index k). This implies that the $RMSE_{Clim}$ is an estimate of the sample variance of the predictand. For the RMSE using climatology as the control forecasts, the skill score becomes

$$SS_{Clim} = 1 - \frac{RMSE}{RMSE_{Clim}}. \tag{3.7}$$

Because of the arrangement of the skill score in Equation 3.7, the SS_{Clim} based on RMSE is sometimes called the reduction of variance (RV), because the quotient being subtracted is the average squared error (or residual, in the nomenclature of regression) divided by the climatological variance.

3.4.2 Nonprobabilistic Forecasts of Categorical SPI Values

The temporal correlation between forecast and observed values of the SPI provides an overall measure of forecast accuracy and skill, one that is not limited to the case of drought alone. Therefore, we also evaluated SPI forecasts in the context of being able to detect drought, i.e. when the SPI drops below a particular threshold. Here, we identified a drought event as occurring when

the SPI value for a given month was ≤ -1, which corresponds to a "moderate drought" category or higher in the classification system presented in Table 3.1. Ensemble drought detection was based on several methods (Table 3.2) and can be categorized into three types [2]: individual, where the index is based on an individual member or percentile; partially integrative, where the sum of particular 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, but also about the distribution of the members. The individual type of drought detection have been subdivided into five classes representing dry members (Q13, Q23), wet ones (Q77, Q88) or the median. The extreme members of the distribution are not used to avoid outliers generally associated with ensemble systems [58].

Table 3.2: *Methods to detect drought from the S4 ensemble system. Adapted from [2].*

Name	Definition	Type
13th percentile (Q13)	Member located at the 13% of the CDF	Individual
23th percentile (Q23)	Member located at the 23% of the CDF	Individual
Median (MED)	Member located at the 50% of the CDF	Individual
77th percentile (Q77)	Member located at the 77% of the CDF	Individual
88th percentile (Q88)	Member located at the 88% of the CDF	Individual
Large spread (SpL)	Sum of the extreme members (Q13+Q88)	Partially Integrative
Low spread (Spl)	Sum of the members (Q23+Q78)	Partially Integrative
Dry spread (SpD)	Sum of the dry members (Q13+Q23)	Partially Integrative
Flood spread (SpF)	Sum of the wet members (Q77+Q88)	Partially Integrative
Mean (EM_RES)	Ensemble mean	Integrative

For 3- and 6-month lead times (respectively for SPI3 and SPI6), we computed several verification measures for the categorical forecasts (i.e. below the SPI "-1" threshold) identified with the methods described in Table 3.2. All verification measures are based on a contingency table approach, which is applied at each grid point in the study area. The entries in the table are defined as follows: "A" is the number of drought events that are forecast and occur; "B" is the number of drought events that are forecast but do not occur; "C" is the number of drought events that are not forecast but do occur; and "D" is the number of drought events that are not forecast and do not occur. The variable N is the total number of cases analyzed from 1981 to 2010. Based on these values, the percentage correct (PC, perfect = 1, see 3.8) is the ratio of good forecasting events in relation to the total number of events.

$$PC = \frac{A + D}{N} \tag{3.8}$$

The extreme dependency score (EDS, see 3.9) provides a skill score in the range $[-1, 1]$ that can be used to find the hit-rate exponent [59]. The EDS takes the value of 1 for perfect forecasts and 0 for random forecasts, and is greater than zero for forecasts that have hit rates that converge slower than those of random forecasts.

$$EDS = \frac{2log\frac{A+B}{N}}{log\frac{A}{N}} - 1 \tag{3.9}$$

The Gilbert score (GSS, see 3.10) measures the fraction of forecast events that were correctly predicted, adjusted for the frequency of hits that would be expected to occur simply by random chance [2].

$$GSS = \frac{A + A^*}{A + B + C - A^*} \tag{3.10}$$

where A^* is the number of random hits, computed as:

$$A^* = \frac{(A+B).(A+C)}{N} \tag{3.11}$$

The GSS is often used in the verification of rainfall forecasts because its "equitability" allows scores to be compared more fairly across different regimes (for example, it is easier to correctly forecast rain occurrence in a wet climate than in a dry climate). However, because it penalizes both misses and false alarms in the same way, it does not distinguish the source of forecast error. Therefore, it should be used in combination with at least one other contingency table statistic, e.g. bias [59]. Here, we compute bias as [55]:

$$Bias = \frac{A+B}{A+C} \tag{3.12}$$

The probability of detection (POD, perfect=1) is the ratio of the total number of observed events that have been forecasted [61].

$$POD = \frac{A}{A+C} \tag{3.13}$$

The false alarm rate (FAR, perfect = 0) is the fraction of the forecasted events which actually did not occur [61].

$$FAR = \frac{B}{A+B} \tag{3.14}$$

3.4.3 Probabilistic Forecast of Categorical SPI Values

Verification of probability forecasts is somewhat more subtle than verification of nonprobabilistic forecasts. Since nonprobabilistic forecasts contain no expression of uncertainty, it is clear whether an individual forecast is correct or not. On the other hand, unless a probabilistic forecast is either 0.0 or 1.0, the situation is less clear-cut. For probability values between these two (certainty) extremes, a forecast is neither right nor wrong, so that meaningful assessments can only be made using collections of multiple forecast members and observation pairs. A number of accuracy measures for verification of probabilistic forecasts of dichotomous events exist, but by far the most common is the Brier score (BS) [55]. The Brier score is essentially the mean squared error of the probability forecasts, considering that the GPCC drought observation at time k is $o_k = 1$ if a drought event occurs (i.e. SPI ≤ -1), and that the GPCC observation at time k is $o_k = 0$ if a drought event does not occur (i.e. SPI > -1). The BS averages the squared differences between pairs of forecast probabilities, $fcst_k$, and the subsequent binary reference observations,

$$BS = \frac{1}{n}\sum_{k=1}^{N}(fcst_k - o_k)^2 \tag{3.15}$$

where the index k again denotes a numbering of the N forecast-event pairs. Comparing the BS with Equation 3.6 for the root-mean squared error, it can be seen that the two are completely analogous. As a mean-squared-error measure of accuracy, the BS is negatively oriented, with perfect forecasts exhibiting BS = 0. Less accurate forecasts receive higher BS values, but since individual forecasts and observations are both bounded by zero and one, the score can take on values only in the range $0 \leq BS \leq 1$.

Skill scores of the form of Equation 3.7 are also often computed for the BS [61], yielding the Brier Skill Score (BSS)

$$BSS = 1 - \frac{BS}{BS_{ref}}. \tag{3.16}$$

The BSS is the conventional skill-score form using the BS as the underlying accuracy measure. Usually, for the SPI, the reference forecasts are the relevant climatological probabilities of a drought event taking place with a certain severity (Table 3.1). For example, the frequency of "Moderate" drought events is approximately the 16%. The BSS ranges between minus infinity and 1; 0 indicates no skill when compared to the reference forecast; the perfect score is 1.

A good companion to the BSS is the Relative Operating Characteristic (ROC) of the forecast [61]. ROC is conditioned on the observations, answering the question "given that a drought occurred, what was the corresponding forecast probability?". It therefore measures the ability of the probabilistic forecasting system to discriminate between drought events and non-events of different frequencies, i.e. the resolution of the forecast. ROC is not sensitive to bias in the forecast - a biased forecast could give a good ROC. However, the ROC is a measure of potential usefulness of the probabilistic forecast, and the area under the ROC curve gives a measure of its skill [61]. Since ROC curves for perfect forecasts pass through the upper-left corner, the area under a perfect ROC curve includes the entire unit square, so $A_{perf} = 1$. Similarly ROC curves for random forecasts lie along the 45° diagonal of the unit square, yielding the area $A_{rand} = 0.5$. The area A under a ROC curve of interest can also be expressed in standard skill-score form SS_{ROC}, as

$$SS_{ROC} = \frac{A - 1/2}{1 - 1/2}. \tag{3.17}$$

[55] states that SS_{ROC} is a reasonably good discriminator among relatively low-quality forecasts, but that relatively good forecasts tend to be characterized by quite similar (near-unit) areas under their ROC curves. The SS_{ROC} ranges between 0 and 1; 0.5 indicates no skill; the perfect score is 1.

4 Results and Discussion

First, we assess the ability of the ECMWF S4 ensemble system to seasonal forecast the spatial distribution of SPI in South-Central America, and assess its monthly s calar accuracy and skill score at each location with 3- and 6-month lead time (respectively for the SPI3 and SPI6). In the sequence, we verify the nonprobabilistic forecast of discrete droughts events (i.e. SPI≤ -1), by means of individual, partially integrative and integrative ensemble statistics. Finally, we look into the probabilistic identification of drought events by means of the ECMWF S4 system for South-Central America.

4.1 Nonprobabilistic Forecasts of Continuous SPI Values

In Fig. 4.1, we present the monthly correlation between observed and forecast ensemble mean (a) SPI3 and (b) SPI6 at, respectively, 3- and 6-month lead time for the hindcast period of 1981-2010. The maps depicted in Fig. 4.1(a) show that there is a positive correlation between SPI3 forecast and observations at all months and for most of the study area. Overall, the forecast SPI3 values follow the trends (increases or decreases) of the observed SPI3 values. Notwithstanding, the statistical significance between observed and forecast SPI3 varies across regions and months: for example, the correlation along the East Pacific coast is almost never statistical significant during the year, it is mostly statistical significant during the whole year for Northeast of South-America, and significant patterns only verify for Central America during the months between December and May. On the other hand, SPI6 forecasts present extensive geographic areas that are negatively correlated with SPI6 observations at 6-month lead time (Fig. 4.1(b)). These large forecast errors are not systematic but occur mainly for the Amazon and Central East part of South America, and are most evident during the months of January–April and June–August. Surprisingly, and as similar as for the SPI3, the correlation is statistical significant during almost the whole year for Northeast of South-America and for larges parts of Central America from March to May. [54] suggest that the statistical significant correlation patterns in Central America and Northeast of South America are likely contributed by ENSO: these regions are known to have a strong ENSO signal, and the seasonal skillfull of precipitation forecasts contribute to the SPI3 and SPI6 seasonal forecasts. Since the correlation is statistical significant for some regions at some months, then it suggests that the forecast has some skill at 3 and 6-months lead time.

The scalar skill score was also analyzed to assess the ability of the forecasts to improve SPI prediction over the expected climatological values (i.e. SPI=0). The differences between the ECMWF-based forecasts and the climatological forecasts will indicate whether there is additional skill obtained from the dynamical model drought forecasts. In Fig. 4.2(a), we present the monthly SPI3 forecast skill (using the mean of the ensemble) at 3-month lead time relative to baseline skill for the hindcast period 1981-2010, which shows the difference in correlation between the ECMWF S4 SPI3 forecasts and the baseline SPI3 forecasts based on climatological probabilities. Our results confirm that the forecasts have higher skill than the baseline, but the differences are often not significant at the 5% level based on the Fishers Z test. Indeed, although the correlation with observations is extensively significant over the study area, it does not extensively improve over the climatological SPI values. Marked improvements are observed for Northeast Brazil during the months of April-July, Mexico during the months of December-April, and North of South America between January-April. Overall, our results are consistent with [22, 23, 54], namely, that it is difficult to improve on SPI forecasts that are based on climatology and persistence.

Interestingly, scalar skill score results suggest that SPI3 forecasts match the observations in dry regions mainly during the beginning of the dry seasons, while at re gions with high rainfall variability and/or during the wet seasons the forecasts are usually less skillful. Therefore, we believe that the ECMWF S4 ensemble mean might underestimate monthly rainfall and thus increase the intensity of dry periods and lessen the forecast values of SPI3 for the study region. The bias of the scalar forecast, as depicted by the monthly mean error (ME) between observed

and forecast SPI3 values at 3-month lead time for the hindcast period 1981-2010 (Fig. 4.3(a)), suggests a spatially underestimation of the SPI values by the forecast ensemble mean (positive errors are representative of larger observed SPI values). This forecast underestimation is not systematic, but predominates over overestimated cases that match mostly the Amazon region.

On the other hand, the 6-month seasonal forecasts are less skillful than the 3-month forecasts (Fig. 4.2(b)). Indeed, and as expected from the correlation analysis, skill scores for SPI6 forecasts are generally lower than for SPI3 and almost not statistically significant at the 5% level. In Fig. 4.2(b), it is perceptible that regions with meaningful SPI6 forecasts are also depicted as skillful for the SPI3. The monthly skill scores clearly show that meaningful forecasts are concentrated over the eastern Amazon, namely in most of the states of AP (Amapá), PA (Pará), and MA (Maranhão). [51] states that there were introduced some important reductions in S4 bias, as compared to S3, particularly in the tropical Atlantic and Indian Oceans, and some improvements over land areas e.g. in East Asia and over the Amazon basin. It is possible that these improvements over the bias of the ECMWF S4 precipitation forecasts will reduce the residual errors between observed and predict seasonal SPI values. As similar as for the SPI3, this is not outstanding for monthly MEs of SPI6 forecasts (Fig. 4.3(b): mean errors are not systematic, although seem to underestimate predicted SPI values and overestimate forecasts of monthly drought.

Let us now look at the individual residual errors for predicted seasonal SPI values. In Fig. 4.4, we present the monthly RMSE values between observed and forecast (a) SPI3 and (b) SPI6 at 3- and 6-month lead time, respectively, for the hindcast period 1981-2010. Curiously, the results suggest that the predicted SPI is less consistent with the observations derived from GPCC for those regions placed in the subtropical subsidence zones around 10^o and 30^o N/S, such as subtropical southeast and central Brazil, Paraguay, and Bolivia, as well as large areas of Peru. It is noticeable a match between the geographic distribution of statistical disagreements among observed and predicted SPI, and the arid regions with highly marked precipitation seasonality, as measured by the relative entropy indicator proposed by [62]. They found that the precipitation models participating to the CMIP5 project, systematically overestimate the distribution of monthly precipitation throughout the year in arid and semiarid regions with intermittent precipitation regimes due to, in most cases, an excess of rainfall during the premonsoonal months. Although ECMWF S4 bias is not systematically positive for the region (Fig. 4.3), it is plausible that the high variability of precipitation regimes within those latitudes makes it difficult to predict drought at a seasonal scale. The results based on the analysis of residual errors also suggest that locations with monthly forecast errors inferior to $\approx$ 0.2 have significant skill, whereas those superior to $\approx$ 0.5 have negative correlation and are unskillful. This output is confirmed by the monthly skill score measured in terms of the RMSE (Fig. 4.5). RMSE skill score approximates the skill score computed with the correlation index (Fig. 4.2) and its spatial patterns: overall, seasonal SPI3 and SPI6 forecasts are monthly skillful for a small region in the eastern part of the Amazon Basin, located in the Northern part of Brazil.

Finally, an analysis of the Willmott's index of agreement (Fig. 4.6) between observed and forecast (a) SPI3 and (b) SPI6 values at, respectively, 3- and 6-month lead times for the hindcast period 1981-2010, reveals that the spatial patterns of error are matching the geographic distribution of RMSE values. The major advantage of the Willmott's index over the RMSE is that it is dimensionless: generally speaking, we could easily verify the ability of ECMWF S4 precipitation data at predicting drought indexes that are measured at different scales.

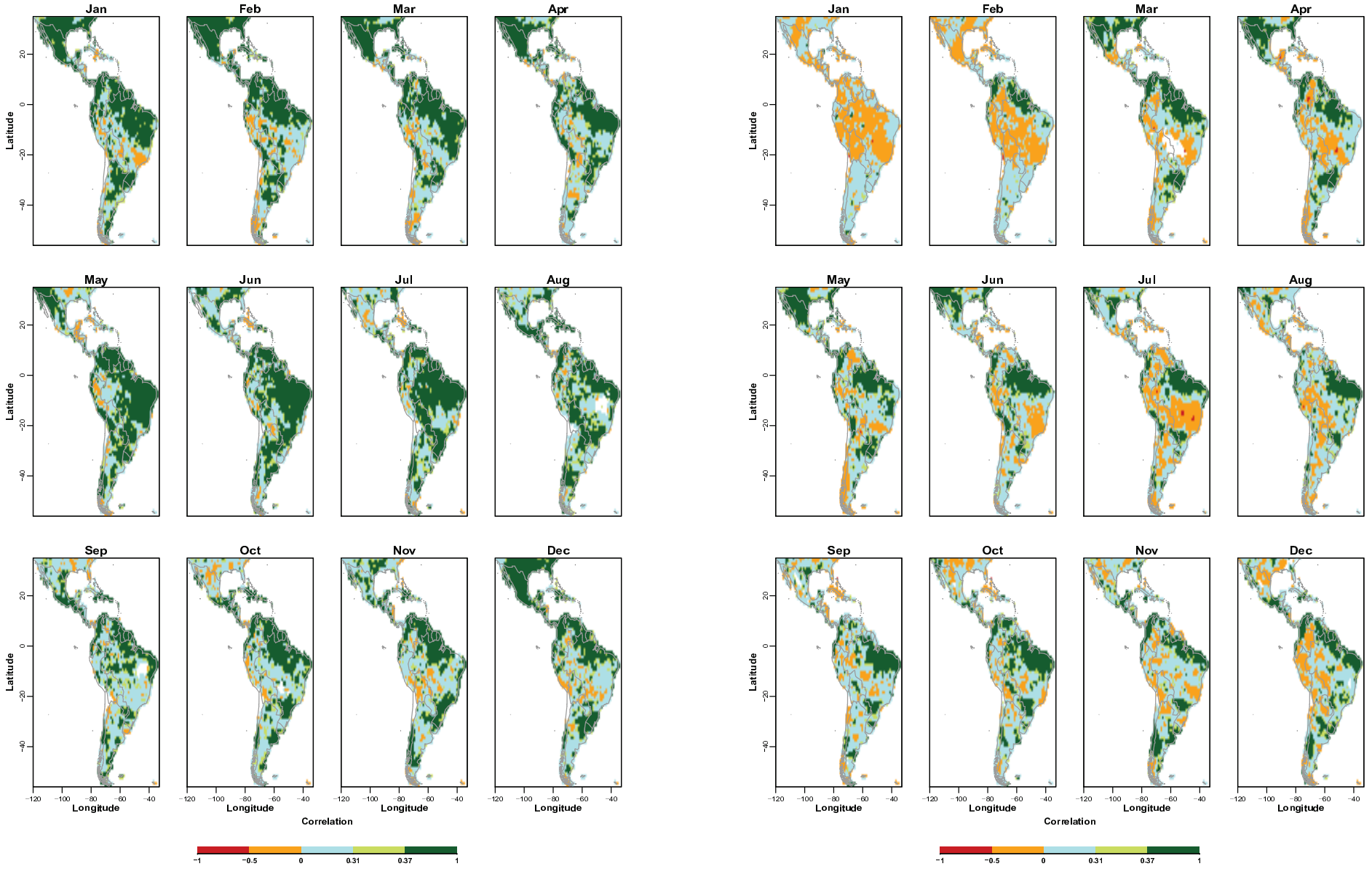
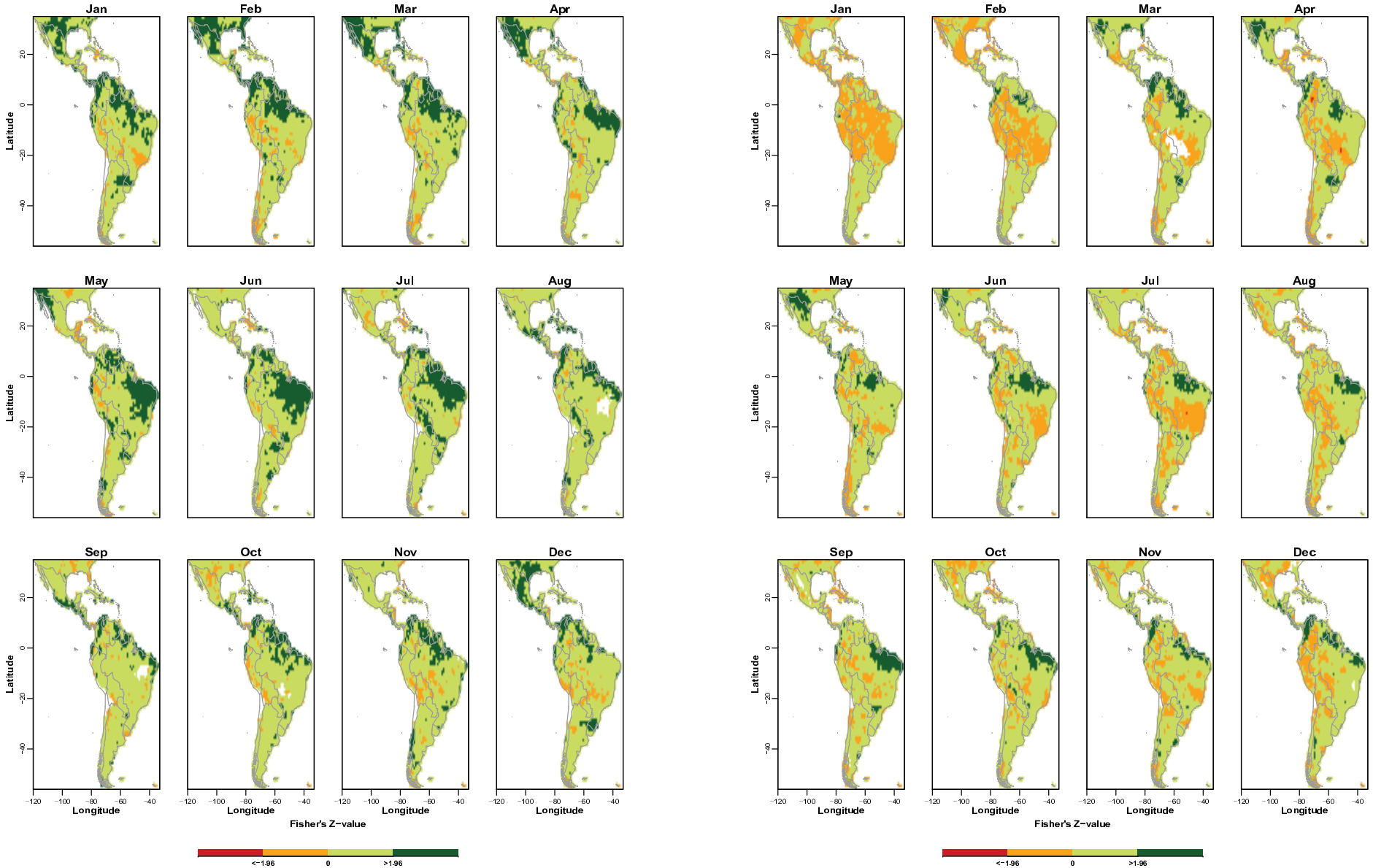


Figure 4.1: *Monthly correlation of the observed and forecast (using the mean of the ensemble) for the hindcast period (from January 1981 to December 2010). Values are indicated in the color bar: 0.31 (0.37) is statistical significat at 10% (5%) significance level. (a) SPI3 at 3-month lead time; (b) SPI6 at 6-month lead time.*



(a) (b)

Figure 4.2: *Monthly difference in forecast skill (Pearson correlation) between the forecast SPI (using the mean of the ensemble) and climatological SPI for the hindcast period (from January 1981 to December 2010). Values are indicated in the color bar: 1.96 is statistical significant at the 5% significance level. (a) SPI3 at 3-month lead time; (b) SPI6 at 6-month lead time.*

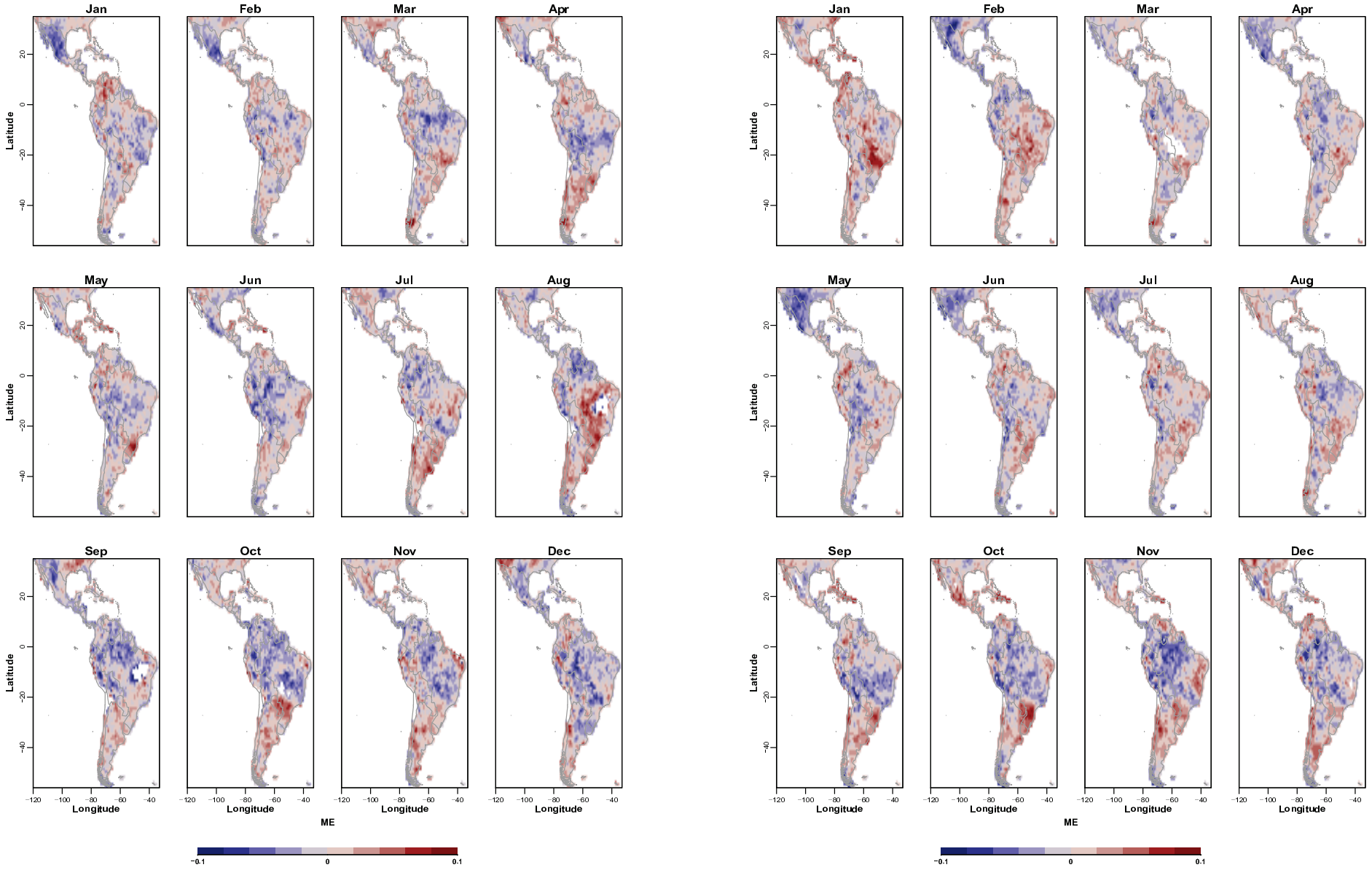


Figure 4.3: *Mean Error (ME) between the observed and forecast SPI (using the mean of the ensemble) for the hindcast period (from January 1981 to December 2010). Values in difference of percentile magnitude are indicated in the color bar. (a) SPI3 at 3-month lead time; (b) SPI6 at 6-month lead time.*

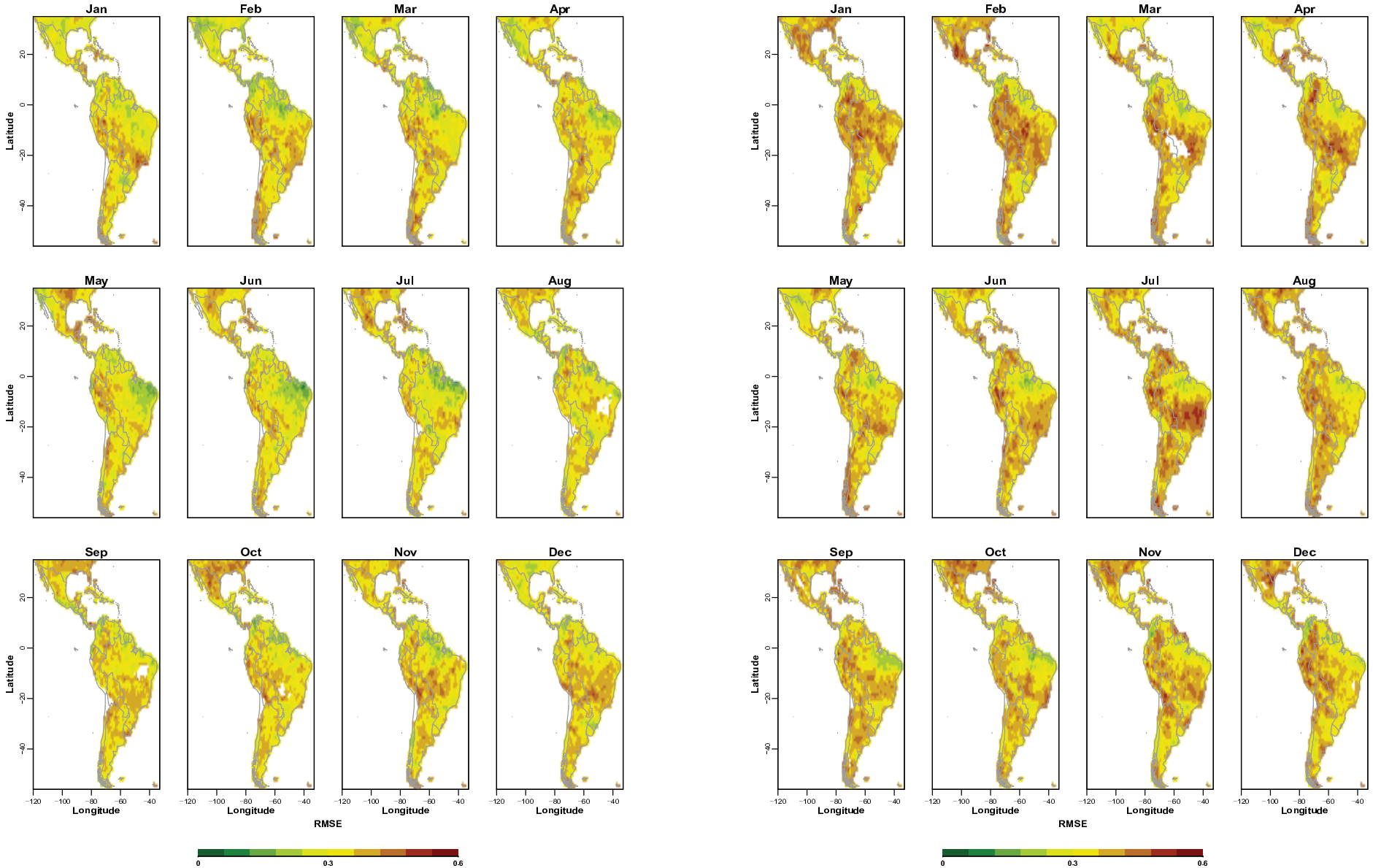


Figure 4.4: *RMSE between the observed and forecast SPI (using the mean of the ensemble) for the hindcast period (from January 1981 to December 2010). Values in difference of percentile magnitude are indicated in the color bar. (a) SPI3 at 3-month lead time; (b) SPI6 at 6-month lead time.*

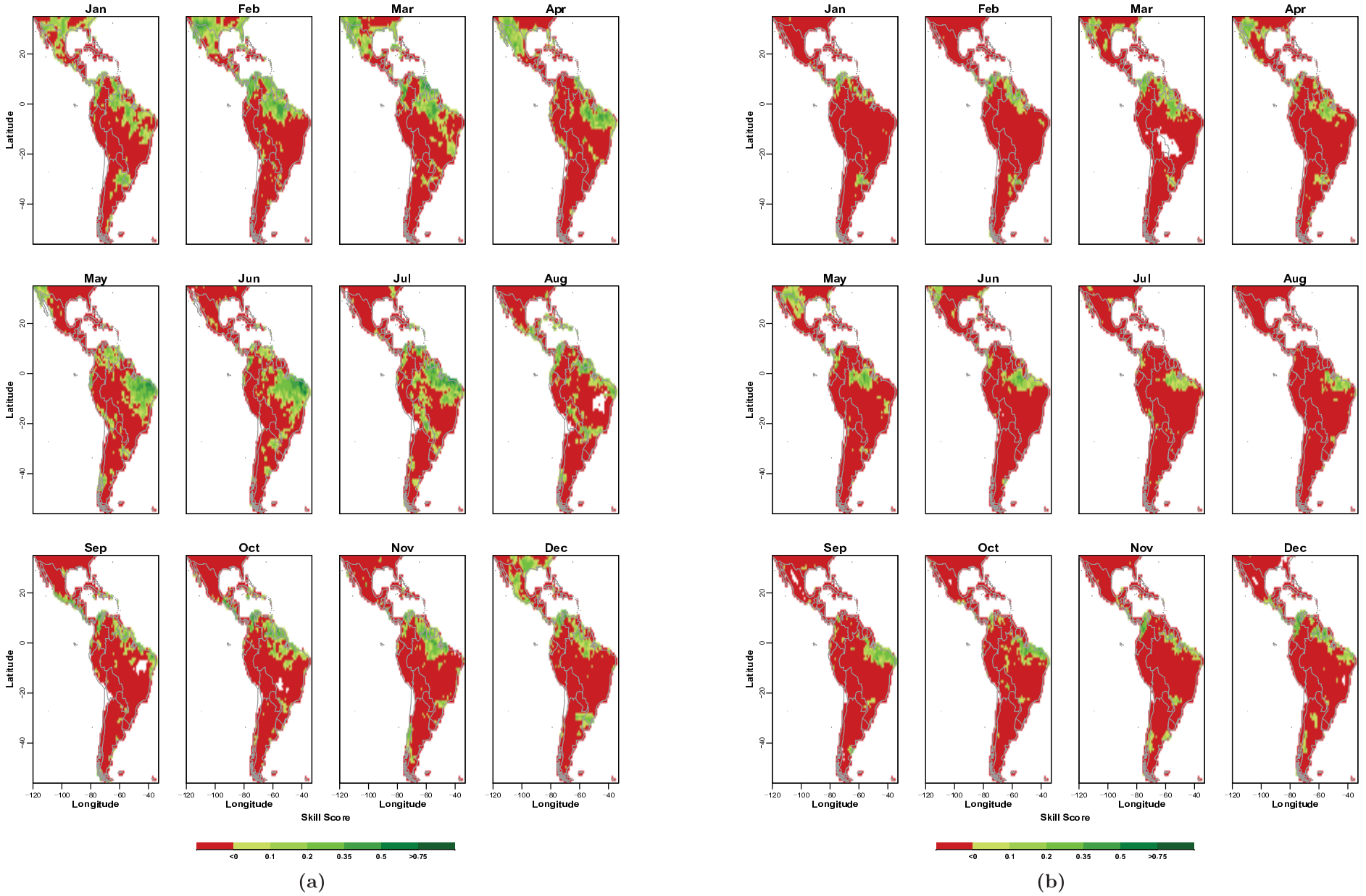


Figure 4.5: *Skill Score of the SPI forecast, measured in terms of the RMSE relative to climatological RMSE for the hindcast period (from January 1981 to December 2010). Values are indicated in the color bar. (a) SPI3 at 3-month lead time; (b) SPI6 at 6-month lead time.*

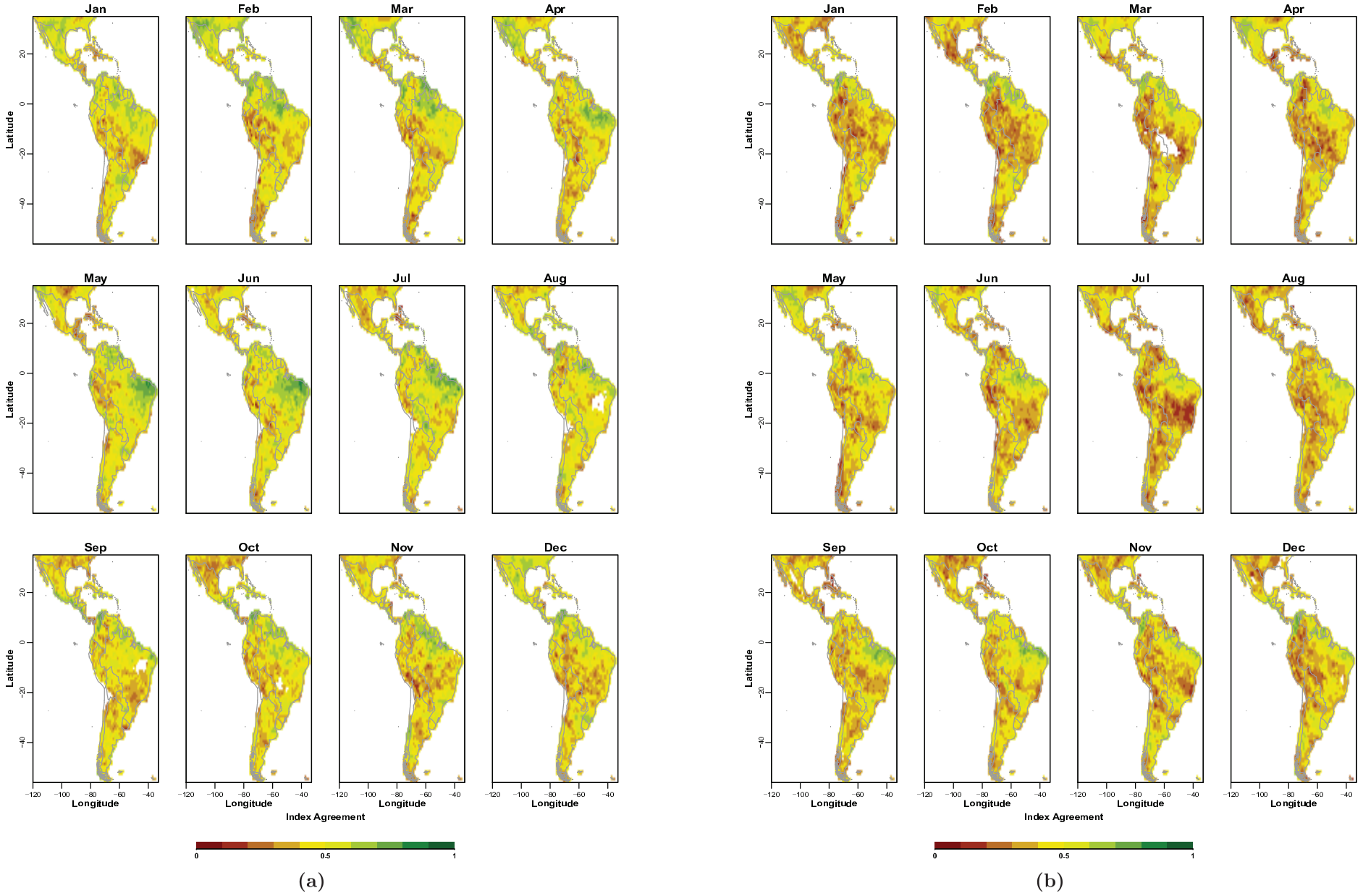


Figure 4.6: *Willmott's index of agreement between the observed and forecasted SPI (using the mean of the ensemble) for the hindcast period (from January 1981 to December 2010). (a) SPI3 at 3-month lead time; (b) SPI6 at 6-month lead time.*

4.2 Nonprobabilistic Forecasts of Categorical SPI Values

In Figures 4.7 and 4.8, we present, respectively, the score values of categorical drought forecasts (i.e. below the SPI "-1" threshold, Table 3.1), estimated with the methods described in Table 3.2 for SPI3 at 3-month lead time and for SPI6 at 6-month lead time. We have pooled together all seasons and locations at the study area in generating Figures 4.7 and 4.8. S urprisingly, the distribution of score values for SPI3 and SPI6 are alike for all methods and all verification measures. The detailed analysis of this phenomenon is outside the scope of this study, but is sensible to present a plausible explanation for it. We hypothesize that because the boundary conditions of seasonal dynamical model forecasts are often characterized by a low frequency variability, then these may lead to similar predictability of medium range climate conditions that extend from a few to several months lead time. In general, precipitation is a result of complex and interacting atmospheric phenomena at different spatial and temporal scales, but regional climate conditions that are actively involved in its extreme large-scale seasonal patterns (such as long-term drought conditions) are persistent and influenced b y predictors that can be accurately estimated at large lead times, such as sea surface temperature and solar radiation [63]. Therefore, we believe that precipitation anomalies over extreme peak thresholds (i.e. drought events) might be similarly predicted for different accumulation periods and seasonal lead times, although the accuracy of their scalar values, as measured by the SPI3 and SPI6, may vary regionally and intra-annually (please see subsection 4.1 for further details). Moreover, given the similar distribution of score values for different methods of categorical drought identification with the SPI3 and SPI6, we decided to perform a joint analysis of the results presented in Figs. 4.7 and 4.8.

For categorical drought events predicted with both the SPI3 and SPI6 computed with the ECMWF S4 ensemble mean (EM_RES), POD values indicate that for at least 50% of the locations in South-Central America, one in each three seasonal drought events is correctly predicted. This is better than the respective climatology (16% of drought events are correctly identified, Table 3.1), and extends over a geographic area larger than that with statistical significant scalar skill scores (Figs. 4.2 and 4.5). Although the ensemble mean performs better than the climatology, POD values are still higher for the methods Q13 (60% of detection) and SpD ($\approx$ 80% of detection); the worst results of all methods are given by the wettest members of the ranked distributions (Q77 and Q88). This means that drier members are better than the mean at detecting drought onset, but also that there is a low consistency between the extreme and dry members of the ECMWF S4 ensemble set.

Looking now at the FAR scores, we perceive that by using the ensemble mean SPI values to detect the onset of a drought (EM_RES), there will be at median 70% of commission errors, i.e. events that are not a drought, but still identified as a drought. Median FAR values are even larger for dryer members ($\approx$ 10% more for Q13 and SpD), and the inter-quantile range of the wettest members is about six times greater than that of the mean ($\approx$ 60%), which indicates a large spread of FAR values. Based on these results, it is difficult to select the method that better optimizes between the number of drought hits and the number of drought misses. Indeed, while the mean of the ensemble always shows an average number of hits and misses (as similar as for the Spl and SpL, which represent the mean of ensemble extreme and opposite members), the dryer and wetter members of the ensemble attain, respectively, extreme number of hits or misses.

This situation becomes even more difficult to evaluate by taking into account the values of PC and EDS. Looking at PC, we might suggest that the Q13 of the ensemble is the worst method to detect between drought and non-drought events. On the other hand, by looking at EDS, we might suggest that Q13 is the best method to detect the onset and end of a drought. Because of the non-dependency of the EDS on the false alarms and the correct rejections when the sample size is fixed, [60] have suggested to not use the EDS alone to assess a model's performance on forecasting rare events (e.g. drought occurrences). Those authors have shown that the EDS equation results in an increased freedom for false alarms and correct negatives, which can freely

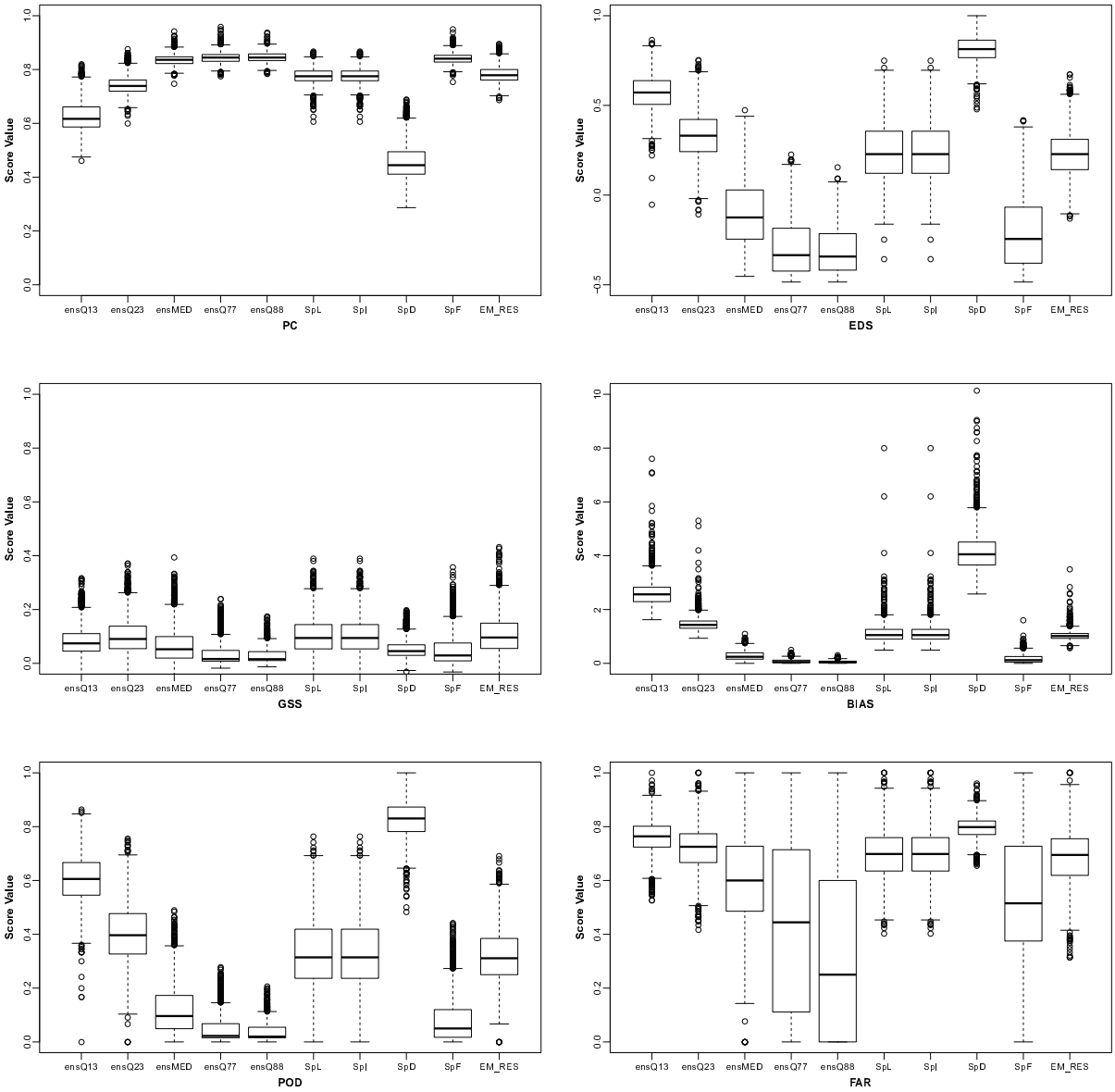


Figure 4.7: *Verification measures of categorical drought forecasts (i.e. below the SPI3 "-1" threshold) estimated with the methods described in Table 3.2.*

vary with the only restriction that their sum has to be constant. This feature encourages hedging, that is, forecasting the event all the time to guarantee a hit, and thus to ensure a higher success rate. Forecasting the event all the time will not only increase EDS, but also the false alarm rate and bias. Therefore, it is paramount to use the EDS in combination with other scores that include the right hand side of the contingency table, as the false alarm rate and/or the bias. Indeed, both FAR and BIAS show that SpD is not an accurate method to detect drought, as it forecasts a large number of drought events that do not occur.

[2] proposes to use the maximum Gilbert score (GSS) as trigger-point to find the method that better optimizes among the number of false alarm ratio, misses and hits of drought events identified with the SPI. Looking at Figures 4.7 and 4.8, we perceive that the ensemble mean (EM_RES) is the best choice for discriminating among seasonal drought and non-drought events at 3- and 6-month lead time, whilst keeping a minor number of false alarms. Although the SpD gives the best POD, it also increases the ratio of false alarms and diminishes the overall skill score of the method. Following the approach proposed by [2], we are of the opinion that the

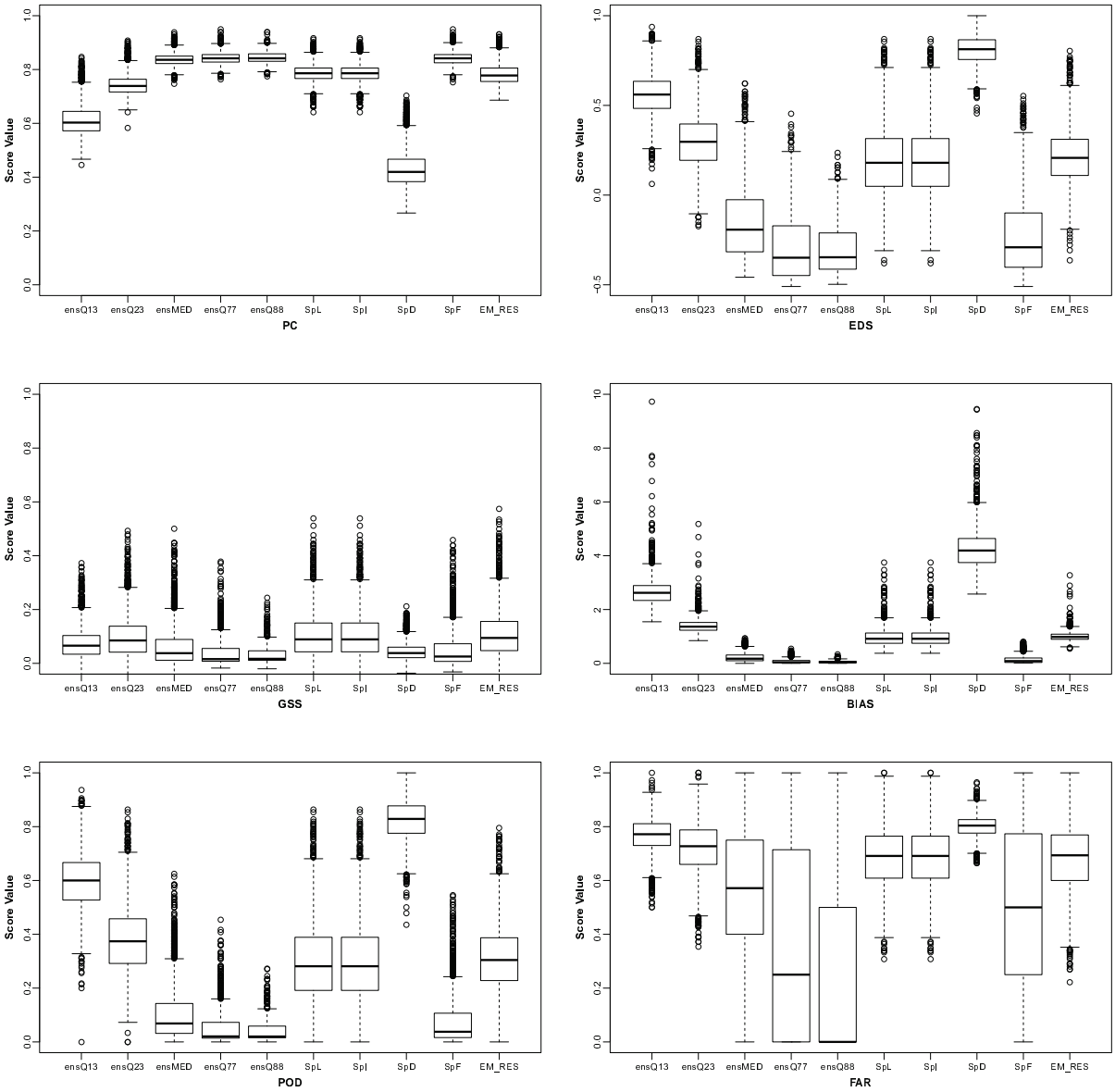


Figure 4.8: *Verification measures of categorical drought forecasts (i.e. below the SPI6 "-1" threshold) estimated with the methods described in Table 3.2.*

ensemble mean should be used to trigger the onset of seasonal drought events for South-Central America by means of the SPI3 and SPI6 for, respectively 3- and 6-month lead times.

4.3 Probabilistic Forecasts of Categorical SPI Values

In addition to having skillful forecasts of scalar SPI3 and SPI6 derived with the ECMWF S4 ensemble mean at seasonal lead times, a second fundamental challenge to generate reliable drought forecasts for the region is associated with uncertainties in the ensemble used. Therefore, to further quantify the uncertainties arising from the spread of the ensemble when computing the SPI, we computed the overall Brier Skill Score (BSS, see section 3.4.3), based upon the climatological frequency of "moderate", "severe" and "extreme" drought events (Table 3.1) as a reference. In Fig. 4.9 (4.10), we map the spatial distribution of BSS for the ECMWF S4 SPI3 (SPI6) forecast, measured in terms of the BS relative to climatological BS (for SPI ≤ -1.0, ≤ -1.5, and ≤ -2.0), at a lead time of 3 (6) months for the hindcast period 1981-2010. We have pooled together all seasons at each 1dd grid point in generating the maps of Figures 4.9 and 4.10.

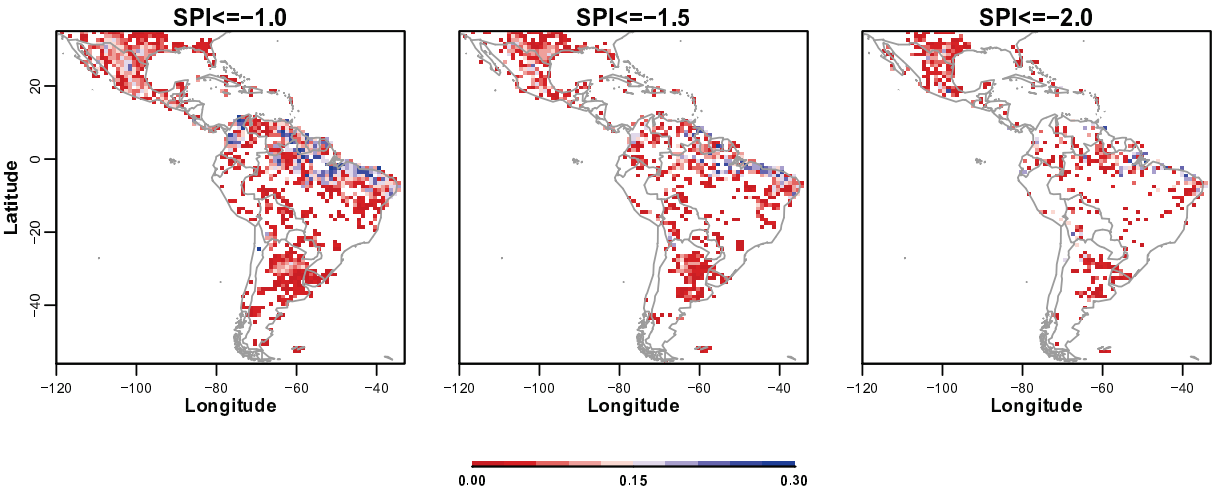


Figure 4.9: *Brier Skill Score (BSS) of the ECMWF S4 SPI3 forecast for different probabilities of SPI occurrence, at a lead time of 3 months for the hindcast period 1981-2010. Values are indicated in the color bar; land grid points colored in white indicate th at the forecasting system is no more skillful than the reference forecast.*

The spatial distribution of BSS for SPI3 and SPI6 at different frequencies of drought occurrence, suggest that the skill of the forecasting system is very similar for both accumulation periods and decreases with the increasing intensity of drought. Looking at the skill for predicting "moderate" drought events, the maps introduced in Figures 4.9 and 4.10 show that the forecasting system behaves better than the climatology for large clustered grid points at the North of South America, Northeast of Argentina and Mexico. On the other hand, the system skill for predicting "extreme" drought events is limited to few locations in Northeast Brazil, Northeast Mexico, Northeast Amazon, and Northeast of Argentina. These results are encouraging, but only Northeast of Mexico shows some spatial clustering with positive BSS for extreme drought events, while positive BSS is spatially scattered for the other regions. On combining these results, it can thus be reasonably assumed that forecasting different magnitudes of meteorological drought intensity on seasonal time scales remains quite challenging, but the ECMWF S4 forecasting system does at least a promising job in capturing the onset of drought events (i.e. "moderate" drought) for some regions.

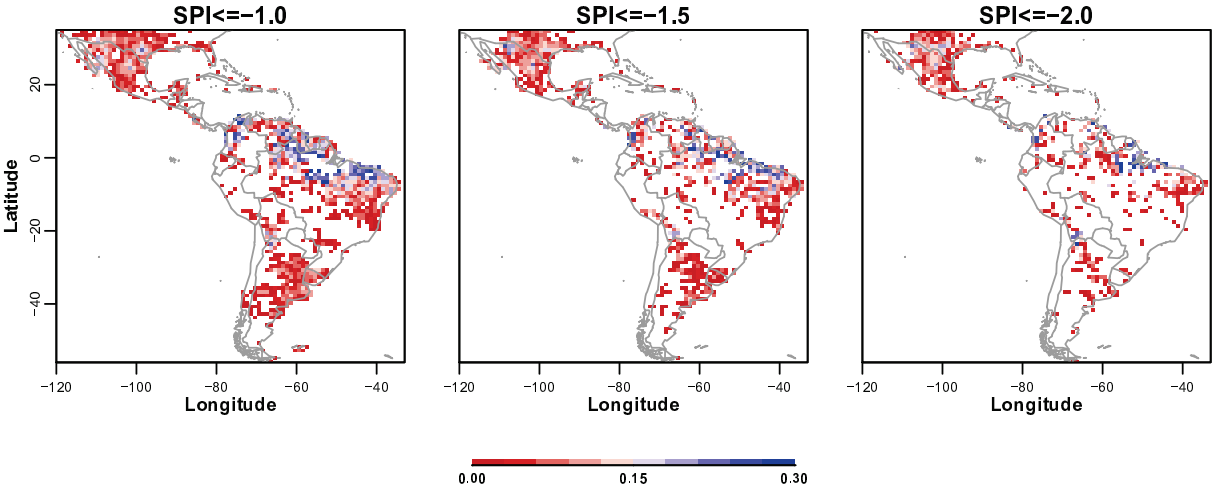


Figure 4.10: *Brier Skill Score (BSS) of the ECMWF S4 SPI6 forecast for different probabilities of SPI occurrence, at a lead time of 6 months for the hindcast period 1981-2010. Values are indicated in the color bar; land grid points colored in white indicate that the forecasting system is no more skillful than the reference forecast.*

It is interesting to note that the spatial pattern of positive BSS at different SPI categories is closely matching the regions that show significant skill scores for nonprobabilistic drought forecasts (subsection 4.1, Figs. 4.2 and 4.5), as well as the geographic grid points that have the lowest monthly RMSEs (Fig. 4.4). As expected, the BSS is lower for the locations where the scalar mismatch between the forecast and observations is larger, which implies more categorical misses and/or false alarms at any SPI intensity. Notwithstanding, since the increase of SPI intensity is accompanied by a decrease of the respective cumulative probability, it was expected that the BSS would decrease with an increase of the SPI drought category because there is a larger probability for mismatching.

To finalize the evaluation of seasonal drought forecasts with the ECMWF S4 data set for South-Central America, we now proceed with the analysis of the Relative Operating Characteristic (ROC) of the forecasts. In Fig. 4.11 (4.12), we present the spatial distribution of the area under the ROC curve for the probability of drought detection at different SPI frequencies. The values are estimated considering the ECMWF S4 SPI3 (SPI6) forecasts at a lead time of 3 (6) months for the hindcast period 1981-2010. We have pooled together all seasons at each 1dd grid point in generating the maps of Figures 4.11 and 4.12.

For the SPI3 and SPI6, for the "moderate" drought threshold, the area under the ROC curve at all grid points in South-Central America is well above the *no skill* line indicating that, despite the poor reliability measured by the BSS, the forecasting system does have some skill. Nevertheless, as similar as for the BSS, we perceive that the regions in the North of South America, Northeast of Argentina and Mexico are more skillful than the remaining locations. As the intensity of drought increases, the usefulness of the forecasting system decreases both in magnitude and area. For "extreme" drought events, the grid-points located in South, Central and Northeast of South America are not skillful, as the area under ROC curve is below the 0.5.

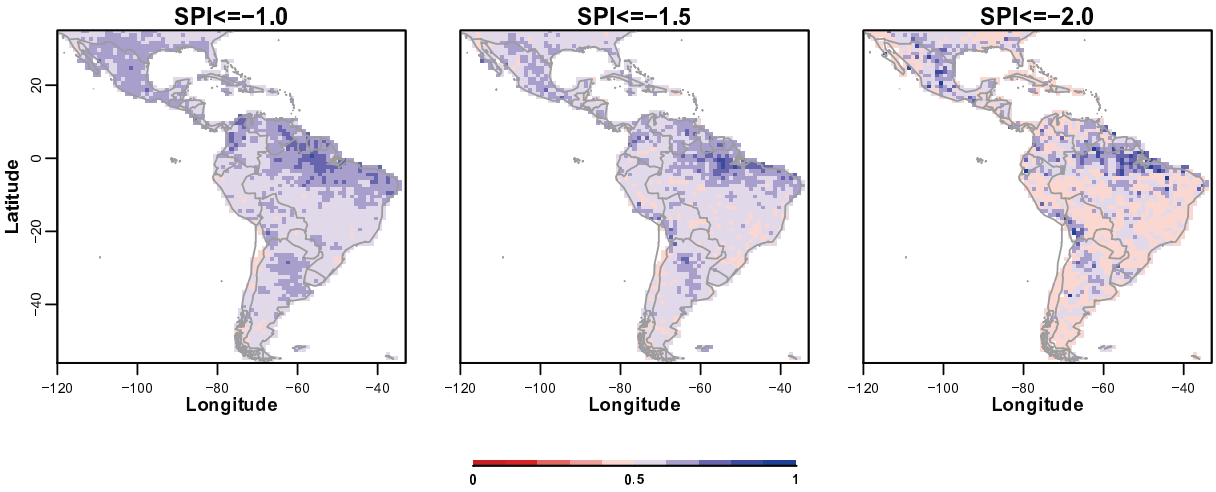


Figure 4.11: *Area under the ROC curve for the probability of drought detection at different SPI3 frequencies. Values indicated in the color bar are estimated at a lead time of 3 months for the hindcast period 1981-2010.*

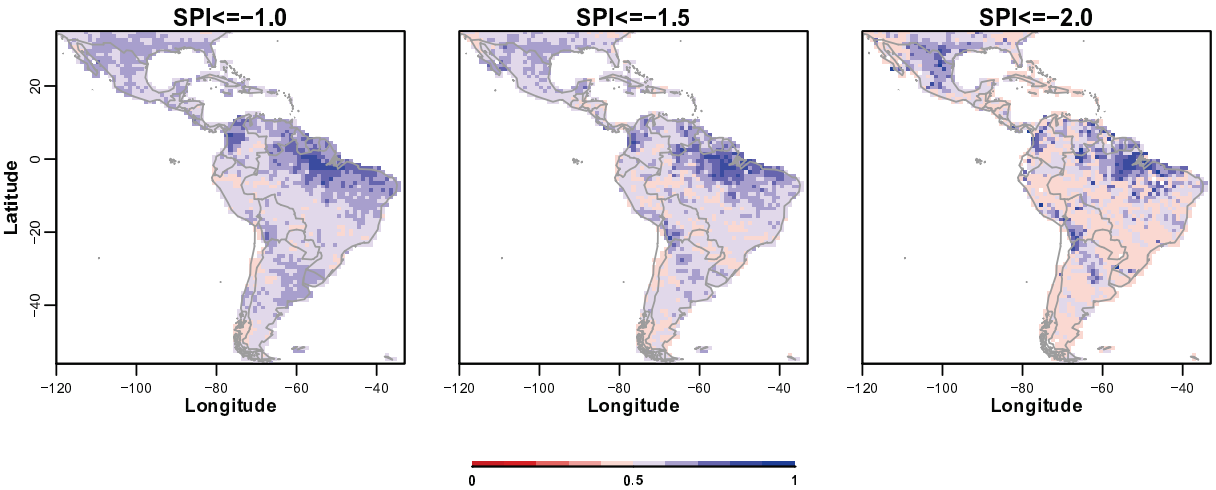


Figure 4.12: *Area under the ROC curve for the probability of drought detection at different SPI6 frequencies. Values indicated in the color bar are estimated at a lead time of 6 months for the hindcast period 1981-2010.*

5 Conclusions

This technical report presents an assessment of seasonal drought forecasts, as characterized by the standardized precipitation index (SPI) at 3- and 6-month accumulation periods for, respectively, 3- and 6-month lead times. An advantage of using the SPI for drought monitoring and prediction is that it is already used in many countries around the globe and it is a drought index endorsed by the World Meteorological Organization (WMO). Here, we have produced SPI forecasts with ECMWF S4 ensemble system and have validated the match between observed (as depicted with the GPCC precipitation dataset) and predicted droughts for the hindcast period of 1981-2010 for South-Central America. We have followed a rigorous analysis to evaluate the scalar accuracy of SPI forecasts, considering both the dry and the wet anomalies at each month. In the sequence, we have verified the skill of nonprobabilistic forecast of discrete droughts events (i.e. SPI≤ -1), by means of individual, partially integrative and integrative ensemble statistics. Finally, we have also assessed the skill of probabilistic drought identification with the SPI. This study is part of the activities developed in the framework of Component 3 of the second phase of the Programme EUROCLIMA: "Sustainable Agriculture, Food Security and Climate Change in Latin America: Strengthening the capacities of key stakeholders to adapt agriculture to climate change and mitigate its effects".

The scalar skill of the SPI3 and SPI6 forecasts was found to be seasonally and regionally dependent, but for some locations, SPI3 predictions at a lead of 3 months and SPI6 predictions at a lead of 6 months are found to have "useful" skill (monthly correlation with observations is statistically significant at the 5% significance level). The difference in skill between the ECMWF S4 SPI forecasts for South-Central America and a baseline forecast based on the climatological characteristics of the SPI, while positive in many areas and for many months, is largely statistically insignificant. Nevertheless, for the SPI3, our results show that the skill of the dynamic seasonal forecast is always equal to or above the climatological forecasts. On the other hand, for SPI6, our results indicate that it is more difficult to improve the climatological forecasts.

In a second step, we have evaluated several methods to forecast the onset of a drought event from an ensemble system when the SPI value at a given lead month is below the "-1" threshold. Ensemble drought detection was based on several methods (Table 3.2) and can be categorized into three types [2]: individual, where the index is based on an individual member or percentile; partially integrative, where the sum of particular individual members or percentiles are used; and integrative which is represented by the ensemble mean. Although individual dry members and partially integrative methods were providing an outstanding accuracy for seasonal drought detection, our results have shown that the spread of the ensemble is too large and these methods have also large bias and false alarm ratio. The best (or more consistent) method is defined by using the ensemble mean SPI values, both for SPI3 and SPI6, at three and six month lead times. Our decision was based on the highest threat threshold (or GSS index), which according to many authors provides an optimum solution for selecting a classification method based on the number of hits, misses and false alarm ratio. The ensemble mean achieves an overall accuracy of about 80%, with POD above 30% for at least 75% of the study area, and a false alarm ratio that is overall below the 70%. Although the ECMWF S4 forecast system often overestimates the drought onset, it is significantly better than using the climatology ($\approx$16%).

Finally, standard verification measures for probabilistic forecasts were used to assess the accuracy of drought predictions based on the SPI values for "moderate", "severe" and "extreme" categories. The Brier Skill Score, which measures the probabilistic forecast skill against a forecast derived from climatology, showed that both the SPI3 and SPI6 were for some regions slightly more skillful than climatology. The ECMWF S4 forecasting system behaves better than the climatology for clustered grid points at the North of South America, Northeast of Argentina and Mexico. The skillful regions are similar for the SPI3 and SPI6, but become reduced in extent for severest SPI categories. We hypothesize that because an increase of SPI intensity is accompanied

by a decrease of the respective cumulative probability, then the likelihood of mismatching is larger. As expected, the BSS is lower for the locations where the scal ar mismatch between the forecast and observations is larger, which implies more categorical misses and/or false alarms at any SPI intensity.

On combining these results, it can thus be reasonably assumed that forecasting different magnitudes of meteorological drought intensity on seasonal time scales remains quite challenging, but the ECMWF S4 forecasting system does at least a promising job in capturing the onset of drought events (i.e. "moderate" drought) for some regions and months. It is noticeable a match between observed and predicted SPI for dry months in arid region s with highly marked precipitation seasonality. Although the performance of Numerical WeatherPrediction models is always improving and advances in the representation of physical processes in the models is an area of intense active research, the performance is not sufficient to provide useful guidance on months with high precipitation amounts, but provides information that is more skillful than the climatology for dry periods. Therefore, it is currently recommended to use seasonal predictions of SPI3 and SPI6 based on the ECMWF S4 precipitation forecasts to detect drought events at 3- and 6-month lead time with some cautiousness. They are likely to give mi sleading information that could result in numerous costly false alarms and missed events. At present, an accurate drought monitoring system is still the best tool for aiding strategic decisions and mitigation procedures.

References

[1] T. B. McKee, N. J. Doeskin, J. Kleist, The relationship of drought frequency and duration to time scales, in: 8th Conf. on Applied Climatology, Amer. Meteor. Soc., Anaheim, Canada, 1993, pp. 179–184.

[2] C. Lavaysse, J. Vogt, F. Pappenberger, Early warning of drought in europe using the monthly ensemble system from ecmwf, Hydrology and Earth System Sciences 19 (7) (2015) 3273–3286. doi:10.5194/hess-19-3273-2015.
URL http://www.hydrol-earth-syst-sci.net/19/3273/2015/

[3] H. Carrão, A. Singleton, G. Naumann, P. Barbosa, J. Vogt, An optimized system for the classification of meteorological drought intensity with applications in frequency analysis, J. Appl. Meteor. Climatol. 53 (2014) 1943–1960.

[4] S. Goddard, S. K. Harms, S. E. Reichenbach, T. Tadesse, W. J. Waltman, Geospatial decision support for drought risk management, Commun. ACM 46 (2003) 35–37.

[5] A. Dai, Drought under global warming: a review, Wiley Interdiscip. Rev . Clim. Change 2 (2011) 45–65.

[6] J. A. Dracup, K. S. Lee, E. G. P. Jr., On the definition of droughts, WaterResour. Res. 16 (1980) 297–302.

[7] D. A. Wilhite, M. H. Glantz, Understanding the drought phenomenon: The role of definitions, Water Int. 10 (1985) 111–120.

[8] A. Steinemann, L. Cavalcanti, Developing multiple indicators and triggers for drought plans, WRPM 132 (2006) 164–174.

[9] G. B. Goodrich, A. W. Ellis, Climatological drought in arizona: An analysis of indicators for guiding the governor's drought task force, Prof. Geogr. 58 (2006) 460–469.

[10] R. R. Heim, A review of twentieth-century drought indices used in the united states, Bull. Am. Meteorol. Soc. 83 (2002) 1149–1165.

[11] S. M. Vicente-Serrano, S. Beguería, J. I. López-Moreno, A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index, J. Climate 23 (2009) 1696–1718.

[12] M. Svoboda, M. Hayes, D. Wood, Standardized precipitation index user guide, Tech. Rep. WMO-No. 1090, World Meteorological Organization (WMO), Geneva (2012).

[13] A. K. Mishra, V. P. Singh, A review of drought concepts, J. Hydrol. 391 (2009) 202–216.

[14] T.-W. Kim, J. B. Valds, Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks, Journal of Hydrologic Engineering 8 (6) (2003) 319–328. doi:10.1061/(ASCE)1084-0699(2003)8:6(319).

[15] A. K. Mishra, V. R. Desai, V. P. Singh, Drought forecasting using a hybrid stochastic and neural network model, Journal of Hydrologic Engineering 12 (6) (2007) 626–638. doi:10.1061/(ASCE)1084-0699(2007)12:6(626).

[16] A. K. Mishra, V. R. Desai, Drought forecasting using stochastic models, Stochastic Environmental Research and Risk Assessment 19 (5) (2005) 326–339. doi:10.1007/s00477-005-0238-4.

[17] F. Vitart, R. Buizza, M. Alonso Balmaseda, G. Balsamo, J.-R. Bidlot, A. Bonet, M. Fuentes, A. Hofstadler, F. Molteni, T. N. Palmer, The new vareps-monthly forecasting system: A first step towards seamless prediction, Quarterly Journal of the Royal Meteorological Society 134 (636) (2008) 1789–1799. doi:10.1002/qj.322.
URL http://dx.doi.org/10.1002/qj.322

[18] B. Nijssen, S. Shukla, C. Lin, H. Gao, T. Zhou, Ishottama, J. Sheffield, E. F. Wood, D. P. Lettenmaier, A prototype global drought information system based on multiple land surface models, Journal of Hydrometeorology 15 (4) (2014) 1661–1676. doi:10.1175/JHM-D-13-090.1.

[19] T. Yuan, L. Oreopoulos, On the global character of overlap between low and high clouds, Geophysical Research Letters 40 (19) (2013) 5320–5326. doi:10.1002/grl.50871.
URL http://dx.doi.org/10.1002/grl.50871

[20] K. C. Mo, L.-C. Chen, S. Shukla, T. J. Bohn, D. P. Lettenmaier, Uncertainties in north american land data assimilation systems over the contiguous united states, Journal of Hydrometeorology 13 (3) (2012) 996–1009. doi:10.1175/JHM-D-11-0132.1.

[21] Z. Hao, A. AghaKouchak, N. Nakhjiri, A. Farahmand, Global integrated drought monitoring and prediction system, Sci. Data 1 (2014) 140001.

[22] E. Dutra, F. Wetterhall, F. Di Giuseppe, G. Naumann, P. Barbosa, J. Vogt, W. Pozzi, F. Pappenberger, Global meteorological drought part 1: Probabilistic monitoring, Hydrology and Earth System Sciences 18 (7) (2014) 2657–2667. doi:10.5194/hess-18-2657-2014.
URL http://www.hydrol-earth-syst-sci.net/18/2657/2014/

[23] E. Dutra, W. Pozzi, F. Wetterhall, F. Di Giuseppe, L. Magnusson, G. Naumann, P. Barbosa, J. Vogt, F. Pappenberger, Global meteorological drought part 2: Seasonal forecasts, Hydrology and Earth System Sciences 18 (7) (2014) 2669–2678. doi:10.5194/hess-18-2669-2014.
URL http://www.hydrol-earth-syst-sci.net/18/2669/2014/

[24] O. L. Phillips, L. E. O. C. Arago, S. L. Lewis, J. B. Fisher, J. Lloyd, G. Lpez-Gonzlez, Y. Malhi, A. Monteagudo, J. Peacock, C. A. Quesada, G. van der Heijden, S. Almeida, I. Amaral, L. Arroyo, G. Aymard, T. R. Baker, O. Bnki, L. Blanc, D. Bonal, P. Brando, J. Chave, . C. A. de Oliveira, N. D. Cardozo, C. I. Czimczik, T. R. Feldpauschand, M. A. Freitas, E. Gloor, N. Higuchi, E. Jimnez, G. Lloyd, P. Meir, C. Mendoza, A. Morel, D. A. Neill, D. Nepstad, S. Patio, M. C. Peuela, A. Prieto, F. Ramrez, M. Schwarz, J. Silva, M. Silveira, A. S. Thomas, H. Steege, J. Stropp, R. Vsquez, P. Zelazowski, E. A. Dvila, S. Andelman, A. Andrade, K.-J. Chao, T. Erwin, A. D. Fiore, E. H. C., H. Keeling, T. Killeen, W. F. Laurance, A. P. Cruz, N. C. A. Pitman, P. N. Vargas, H. Ramrez-Angulo, A. Rudas, R. Salamo, N. Silva, J. Terborgh, A. Torres-Lezama, Drought sensitivity of the amazon rainforest, Science 323 (2009) 1344–1347.

[25] J. Sheffield, K. M. Andreadis, E. F. Wood, D. P. Lettenmaier, Global and continental drought in the second half of the twentieth century: Severityareaduration analysis and temporal variability of large-scale events, J. Clim. 22 (2009) 1962–1981.

[26] M. Zhao, S. W. Running, Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, Science 329 (2010) 940–943.

[27] W. M. Vargas, G. Naumann, J. L. Minetti, Dry spells in the river plata bas in: an approximation of the diagnosis of droughts using daily data, Theor. Appl. Climatol. 104 (2011) 159–173.

[28] G. Magrin, C. G. García, D. C. Choque, J. C. Giménez, A. R. Moreno, G. J. Nagy, C. Nobre, A. Villamizar, Climate change 2007: Impacts, adaptation and vulnerability, in: M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, C. E. Hanson (Eds.), Contribution of Working Group II to the Fourth Assessment Report of the Intergovern mental Panel on Climate Change, Cambridge University Press, 2007, pp. 581–615.

[29] IPCC, Summary for policymakers, in: C. B. Field, V. Barros, T. F. Stoc ker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, P. Midgley (Eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, Cambridge University Press, 2012, pp. 1–19.

[30] FAO, Aquastat database - food and agriculture organization of the united nations (fao), [Available online at http://www.fao.org/nr/water/aquastat/main/index.st m.] (cited 18/11/2014).

[31] K. E. Trenberth, D. Stepaniak, Indices of el Niño evolution, J. Clim. 14 (2011) 1697–1701.

[32] S. S. Yadav, R. Redden, J. L. Hatfield, H. Lotze-Campen, A. J. W. Hall, Crop Adaptation to Climate Change, John Wiley & Sons, 2011.

[33] M. P. Llano, W. Vargas, G. Naumann, Climate variability in areas of the world wit h high production of soya beans and corn: its relationship to crop yields, Meteorol. Appl. 19 (4) (2012) 385–396.

[34] J. F. Santos, I. Pulido-Calvo, M. M. Portela, Spatial and temporal variability of droughts in portugal, Water Resour. Res. 46 (2010) n/a–n/a.

[35] A. K. Mishra, V. P. Singh, Drought modeling – a review, J. Hydrol. 403 (2011) 157–175.

[36] C. T. Agnew, Using the spi to identify drought, Drought Network News 12 (2000) 6–12.

[37] M. Svoboda, D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, S. Stephens, The drought monitor, Bull. Am. Meteorol. Soc. 83 (2002) 1181–1190.

[38] A. Steinemann, Drought indicators and triggers: a stochastic approach to evaluation, JAWRA 39 (2003) 1217–1233.

[39] M. J. Hayes, M. D. Svoboda, D. A. Wilhite, O. V. Vanyarkho, Monitoring the 1996 drought using the standardized precipitation index, Bull. Am. Meteorol. Soc. 80 (1999) 429–438.

[40] N. B. Guttman, Accepting the standardized precipitation index: a calculation algorithm, JAWRA 35 (1999) 311–322.

[41] H. Wu, M. J. Hayes, D. A. Wilhite, M. D. Svoboda, The effect of the length of record on the standardized precipitation index calculation, Int. J. Climatol. 25 (2005) 505–520.

[42] G. Sepulcre-Canto, S. Horion, A. Singleton, H. Carraõ, J. Vogt, Development of a combined drought indicator to detect agricultural drought in europe, Earth Syst. Sci. 12 (2012) 3519–3531.

[43] B. Lloyd-Hughes, M. A. Saunders, A drought climatology for europe, Int. J. Climatol. 22 (13) (2002) 1571–1592. doi:10.1002/joc.846.

[44] H. K. Ntale, T. Y. Gan, Drought indices and their application to east africa, Int. J. Climatol. 23 (2003) 1335–1357.

[45] G. Naumann, P. Barbosa, H. Carrão, A. Singleton, J. Vogt, Monitoring drought conditions and their uncertainties in africa using trmm data, J. Appl. Meteor. Climatol. 51 (2012) 1867–1874.

[46] A. A. Paulo, E. Ferreira, C. Coelho, L. S. Pereira, Drought class transition analysis through markov and loglinear models, an approach to early warning, Agr. Water Manag. 77 (2005) 59 – 81.

[47] B. Hofer, H. Carrão, D. Mcinerney, Multi-disciplinary forest fire danger assessment in europe: The potential to integrate long-term drought information, IJSDIR 7 (2012) 300–322.

[48] M. N. Kumar, C. S. Murthy, M. V. R. S. Sai, P. S. Roy, On the use of standardized precipitation index (spi) for drought intensity assessment, Meteorol. Appl. 16 (2009) 381–389.

[49] B. Rudolf, A. Becker, U. Schneider, A. Meyer-Christoffer, M. Ziese, New full data reanalysis version 5 provides high-quality gridded monthly precipitation data, GEWEX News 21 (2011) 4–5.

[50] R. I. N. Jurez, W. Li, R. Fu, K. Fernandes, A. de Oliveira Cardoso, Comparison of precipitation datasets over the tropical south american and african continents, Journal of Hydrometeorology 10 (1) (2009) 289–299. doi:10.1175/2008JHM1023.1.

[51] F. Molteni, T. Stockdale, M. Balmaseda, G. BALSAMO, R. Buizza, L. Ferranti, L. Magnunson, K. Mogensen, T. Palmer, F. Vitart, The new ecmwf seasonal forecast system (system 4), ECMWF Tech. Memo. 656, ECMWF, Reading, UK (2011).

[52] W. C.J., On the validation of models, Physical Geography 2 (1981) 184194.

[53] W. C.J., Some comments on the evaluation of model performance, Bulletin of the American Meteorological Society 63 (1982) 13091313.

[54] K. C. Mo, B. Lyon, Global meteorological drought prediction using the north american multi-model ensemble, Journal of Hydrometeorology 16 (3) (2015) 1409–1424. doi:10.1175/JHM-D-14-0192.1.

[55] D. S. Wilks, Statistical Methods in the Atmospheric Sciences, 2nd Edition, Academic Press, 2005.

[56] G. Duveiller, D. Fasbender, M. Meroni, Revisiting the concept of a symmetric index of agreement for continuous datasets, Sci. Rep. 6 (2016) 19401.

[57] X.-W. Quan, M. P. Hoerling, B. Lyon, A. Kumar, M. A. Bell, M. K. Tippett, H. Wang, Prospects for dynamical prediction of meteorological drought, Journal of Applied Meteorology and Climatology 51 (7) (2012) 1238–1252. doi:10.1175/JAMC-D-11-0194.1.

[58] C. Lavaysse, M. Carrera, S. Blair, N. Gagnon, R. Frenette, M. Charron, M. K. Yau, Impact of surface parameter uncertainties within the canadian regional ensemble prediction system, Monthly Weather Review 141 (5) (2013) 1506–1526. doi:10.1175/MWR-D-11-00354.1.

[59] D. B. Stephenson, B. Casati, C. A. T. Ferro, C. A. Wilson, The extreme dependency score: a non-vanishing measure for forecasts of rare events, Meteorological Applications 15 (1) (2008) 41–50. doi:10.1002/met.53.
URL http://dx.doi.org/10.1002/met.53

[60] A. Ghelli, C. Primo, On the use of the extreme dependency score to investigate the performance of an nwp model for rare events, Meteorological Applications 16 (4) (2009) 537–544. doi:10.1002/met.153.
URL http://dx.doi.org/10.1002/met.153

[61] A. Singleton, Forecasting drought in europe with the standardized precipitation index, JRC Scientific and Technical Reports EU 25254 EN, European Commission, Ispra, Italy (2012).

[62] S. Pascale, V. Lucarini, X. Feng, A. Porporato, u. S. Hasson, Analysis of rainfall seasonality from observations and climate models, Clim. Dyn. 44 (11) (2015) 3281–3301. doi:10.1007/s00382-014-2278-2.

[63] D. M. Smith, A. A. Scaife, B. P. Kirtman, What is the current state of scientific knowledge with regard to seasonal and decadal forecasting?, Environmental Research Letters 7 (1) (2012) 015602.

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (http://europa.eu).

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

LB-NA-28260-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2788/826147

ISBN 978-92-79-64025-4