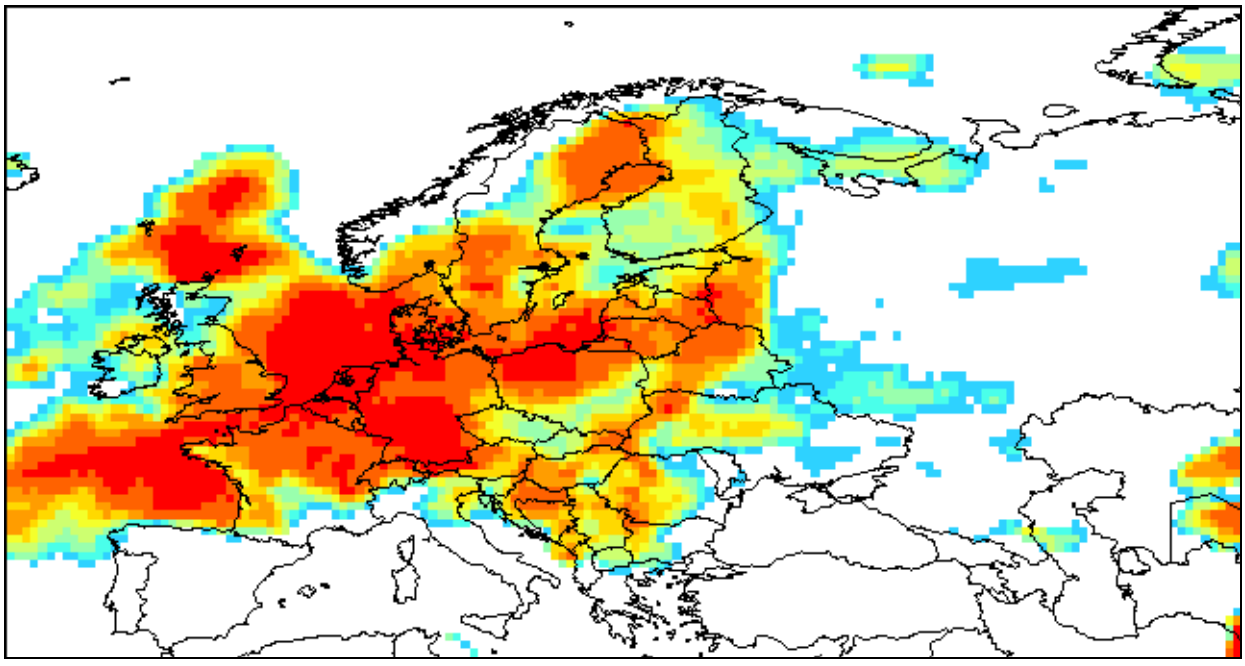
JRC Scientific and Technical Reports

Forecasting Drought in Europe with the Standardized Precipitation Index

An assessment of the performance of the European Centre for Medium Range Weather Forecasts Variable Resolution Ensemble Prediction System.

A. Singleton



EUR 25254 EN - 2012

JRC EUROPEAN COMMISSION

ies Institute for Environment and Sustainability

The mission of the JRC-IES is to provide scientific-technical support to the European Union's policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Address: TP280, I-21027 Ispra (VA), ITALY
E-mail: juergen.vogt@jrc.ec.europa.eu
Tel.: +39 033278 5481
Fax: +39 033278 9803

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 688839

EUR 25254 EN
ISBN 978-92-79-23394-4 (PDF)
ISBN 978-92-79-23393-7 (print)

ISSN 1018-5593 (print)
ISSN 1831-9424 (online)

doi:10.2788/16522

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

CONTENTS

1. Introduction

Drought is a natural hazard that can have serious consequences for a range of human activities. At the most catastrophic end of the scale, particularly in areas of the world that do not have the infrastructure to effectively mitigate its affects, drought can lead to famine and numerous loss of life. In Europe the impacts of drought are more felt economically and socially: agricultural production and water resource availability for industry and households being the most affected sectors. It is estimated that the annual average economic cost of drought events in Europe amounts to €5.3 billion (European Commission, 2006).

Drought is often seen as a "creeping" phenomenon with slow onset and cessation. As a result, an effective drought monitoring system is the most important tool for developing and implementing efficient mitigation strategies. However, not only can the onset of drought conditions be rapid, an indication of how long drought conditions may continue will enable improved planning and resource allocation. For this reason, a capability to accurately forecast the onset, persistence and cessation of drought conditions will enable more effective drought mitigation strategies to be developed.

Drought can be defined as a period of time with water availability less than some specified amount at a particular location. It is primarily driven by a shortage of precipitation, the effects of which can be enhanced or reduced at any stage of the water cycle. Therefore, as a means towards developing a drought forecasting system, this study concentrates on forecasting the precipitation contribution, or lack thereof, towards drought conditions.

There are two main indicators used for monitoring precipitation in terms of drought: the Standardized Precipitation Index (SPI; McKee *et al*., 1993) and the Palmer Drought Severity Index (PDSI; Palmer, 1965). The SPI is a probabilistic indicator based purely on precipitation, whereas the PDSI uses empirical relationships to estimate the effect of precipitation and temperature on the soil moisture. Both of these indicators have been widely used in a drought monitoring context (see Mirsha and Singh, 2010 for a review). Guttman (1998) compared the applicability of SPI and PDSI to drought events in the contiguous USA and found that the SPI had the advantage being statistically consistent, in both time and space, as well as having the ability to monitor drought at any timescale. The PDSI on the other hand was found to be location specific with an inherent timescale of between six and twelve months.

The SPI has been defined as a key indicator for monitoring drought by the World Meteorological Organization (WMO; 2006) and has been widely applied as an operational (e.g. Wilhite *et al*., 2000; Heim, 2002; Svoboda *et al*., 2002; Quiring 2009; McRoberts and Nielsen-Gammon, 2011) and analysis tool. For example, Lloyd-Hughes and Saunders (2002) used the SPI to develop a drought climatology for Europe; Santos *et al*. (2010) examined both the temporal and spatial variability of drought in Portugal using the SPI; and Hannaford *et al*. (2011) used a regionally aggregated SPI, amongst other indicators, to analyse spatial coherence patterns of drought in Europe.

The probabilistic nature of the SPI makes it an ideal candidate for a drought risk analysis tool (Guttman, 1998). However, there have only been limited attempts to forecast drought using SPI. Bordi *et al*. (2005) compared an autoregressive (AR) model with what they termed the Gamma Highest Probability (GAHP) method for forecasting the 1-month SPI. The GAHP method uses the mode of a gamma distribution fit to the precipitation record as the forecast variable. They found that the GAHP method generally performed better than the AR model, although the mean-squared error (MSE) was relatively high for both approaches. Cancelliere *et al*. (2007) proposed methods for forecasting transition probabilities from one drought class to another and for forecasting SPI. They showed that an approach based on the analytical derivation of the auto-covariance matrix of SPI based on the underlying precipitation statistics displayed some skill in predicting the transitional probability from one drought class to the next. Furthermore, they demonstrated that the SPI could be forecast with a reasonable degree of accuracy, in terms of the MSE, using conditional expectation based on past values of monthly precipitation. Hwang and Carbone (2009) used a conditional resampling technique to generate ensemble forecasts of SPI and found reasonable forecast performance for 1-month lead time. Hannaford *et al*. (2011) made us of spatial coherence patterns in SPI to propose a method for forecasting drought in the United Kingdom based on current occurrence of drought elsewhere.

Thus far, methods used to forecast SPI are based purely on statistics. There is no evidence in the literature of an assessment of the performance of Numerical Weather Prediction (NWP) models in forecasting drought. It is well known that as the lead time increases, the skill of NWP models in forecasting hydrological processes diminishes rapidly. However, with the rapid development of NWP models, particularly in terms of spatial resolution and ensemble methods that estimate the uncertainty, an objective assessment of NWP model performance in forecasting SPI is timely. In this report the performance of the European Centre for Medium Range Weather Forecasts (ECMWF) variable resolution Ensemble Prediction System (varEPS) in predicting the probability of SPI for a range of drought thresholds at 1-month lead time is assessed.

Ensemble forecasts are attractive in that through the computation of the probability of an event from the ensemble members the uncertainty of the forecast can be communicated. If this information is used properly, probabilistic forecasts will provide improved guidance in making the most effective decisions.

2. The Standardized Precipitation Index (SPI)

The SPI was introduced by McKee *et al*. (1993) as measure of the precipitation deficit that is uniquely related to probability. It can be calculated for any accumulation timescale, usually from monthly precipitation observations, and is typically expressed as SPI-n, where n is the number of months of accumulation. The time-series is analogous to a moving average in the sense that a new value is calculated each month and is auto-correlated to previous months depending on the accumulation timescale.

The computation of SPI is based on an equi-probability transformation of the probability of observed precipitation to the standard normal variable with mean 0 and variance 1. SPI is therefore expressed in units of the number of standard deviations from the mean, with negative (positive) values denoting drier (wetter) conditions than "expected" for the timescale and location. The standardization procedure to the standard normal variable means that the SPI is spatially and temporally invariant. This characteristic enables precipitation anomalies to be objectively compared between locations and times.

The computation of the SPI is a three stage process:

i. a parametric statistical distribution is fitted to the observed record from a reference period for precipitation accumulations over n months (where n is e.g. 1, 3, 6, 9, 12, 24, or 48 months);
ii. the non-exceedance probability of a precipitation observation is computed related to the parametric distribution;
iii. the non-exceedance probability is transformed to the standard normal variable with mean = 0 and variance = 1.

For stage (i) of the process, the selection of the parametric distribution is a key decision for the accuracy of the final SPI value. There is some debate as to which parametric distribution should be selected. McKee et el. (1993, 1995) recommend the gamma distribution. The gamma distribution is described by only two parameters, but offers considerable flexibility in describing the shape of the distribution, from an exponential to a Gaussian form. It has the advantage that it is bounded on the left at zero and therefore excludes the possibility of negative precipitation. Additionally, it is positively skewed with an extended tail to the right, which is especially important for dry areas with low mean and a high variability in precipitation. Guttman (1999) suggests that the Pearson-III distribution is the "best" universal model to adopt since its three parameters give it more flexibility than the gamma distribution. Lloyd-Hughes and Saunders (2002) suggest that the gamma distribution is the most appropriate for Europe, although they do not include the Pearson-III distribution in their analysis. Furthermore, the method used to estimate the parameters of the distribution is of importance. McKee *et al*. (1993, 1995), Lloyd-Hughes and Saunders (2002) are among those who have use Maximum Likelihood Estimation (MLE) to estimate the parameters of the fitted distribution. However, Guttman (1999) showed that, for the gamma distribution at least, the L-moments method (Hosking and Wallis, 1997) for estimating the distribution parameters results

in more accurate SPI values. In addition, knowledge of the L-moments of a precipitation record has the advantage of enabling regional frequency analyses to be performed (Hosking and Wallis, 1997).

The computation of SPI is illustrated in Figure 2.1 for the gamma distribution, but the methodology equally applies for whichever distribution is selected.

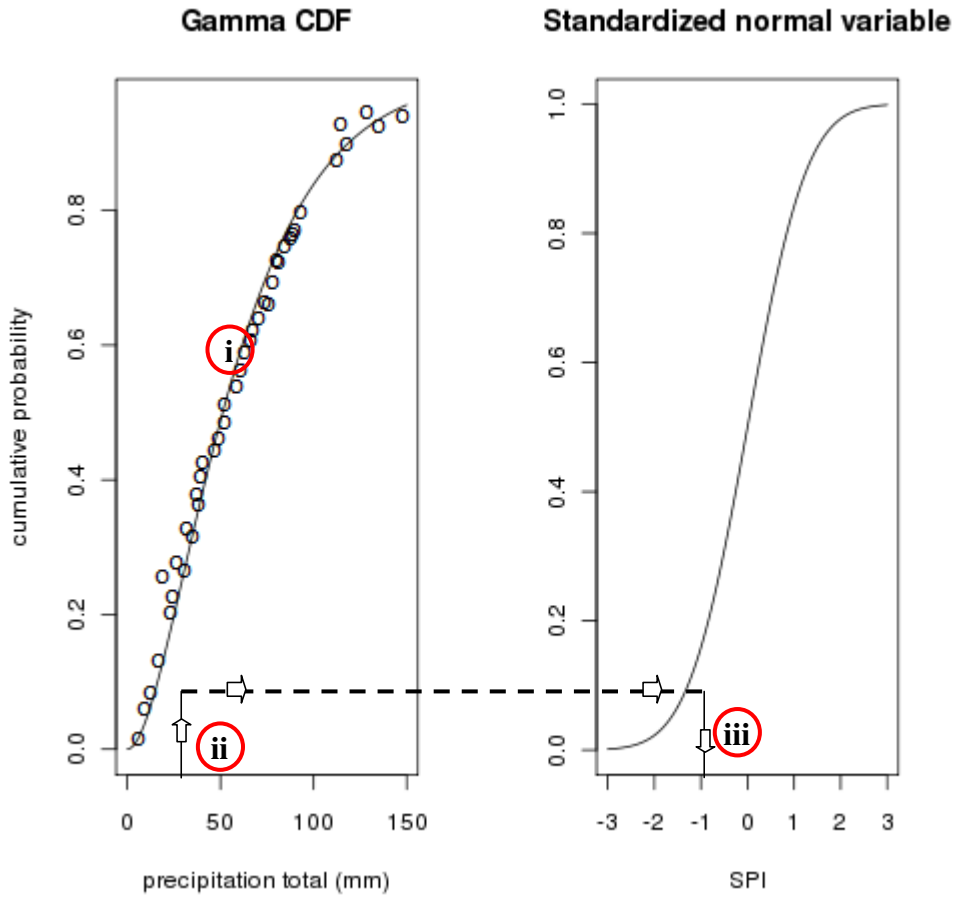


Figure 2.1 – Illustration of the method to compute SPI – (i) the parametric gamma cumulative distribution function (CDF) is fitted to the observed record for the reference period; (ii) the non-exceedance probability of the observed precipitation is computed relative to the gamma distribution; (iii) the non-exceedance probability is transformed to the standard normal variable and the SPI is found.

In order to obtain a parametric distribution that accurately describes the frequency distribution of precipitation for a location McKee *et al*. (1993) recommend that at least 30 years of continuous observations are used. For locations where observations of zero precipitation occur, the fitting of a gamma distribution becomes problematic since it is not defined for zero. In this case the cumulative probability *H(x)* becomes

$$H(x) = q + (1 - q)G(x), \tag{2.1}$$

where *q* is the probability of zero precipitation calculated from the frequency of observations of zero, and *G(x)* is the cumulative probability derived from the gamma distribution. In dry seasons, or arid regions, where the probability of zero becomes

significant, care should be taken in interpreting the SPI (Wu *et al*., 2007). In this case the probability of observed precipitation has a minimum of q, which means that the SPI has an equivalent minimum value. For example, if q = 0.5, the minimum SPI possible is 0.

McKee *et al*. (1995) proposed a classification of the SPI that divides the SPI into moderate, severe and extreme classes for both negative (dry) and positive (wet) SPI as shown in Table 2.1. This classification has been adopted by most centres that use the SPI operationally to define the severity of the precipitation deficit. Whilst this classification is somewhat arbitrary, the probabilistic nature of the SPI means that return periods can be assigned to SPI values: SPI < -1 (~6 years), SPI < -1.5 (~15 years), SPI < -2 (~44 years).

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%

Table 2.1 – SPI classification following McKee *et al*. (1995)

3. The ECMWF Variable Resolution Ensemble Prediction System (varEPS)

3.1. An Introduction to ensemble forecasting

Most atmospheric processes can be considered as chaotic. As a result, errors in the initial conditions of a forecasting system will propagate and likely grow as the forecast is integrated. This will likely have a major impact on a single deterministic forecast in terms of the intensity, spatial location and timing of a weather event. Atmospheric observations are limited in number and have an uneven spatial distribution around the globe, which means that there will always be some uncertainty in an estimate of the current state of the atmosphere. As forecast lead time increases, the initial state has diminishing influence on the atmosphere with physical and dynamical processes having an increasing influence. Our ability to model these processes, based on the governing equations of momentum and thermodynamics, for example, and approximations based on empirical relationships is limited by the spatial and temporal resolution of the model and the accuracy of the parameterizations of processes that occur at finer spatial scales than the model.

Ensemble forecasting takes account of the uncertainty in the current atmospheric state by generating a set, or ensemble, of different, but similar, atmospheric states. Uncertainty in the physical processes is taken account of by applying stochastic perturbations to the model physics. This results in an ensemble of forecasts that provide an estimate of the uncertainty of the forecast. An ensemble forecast may be post-processed into an average forecast (the ensemble mean), a smaller number of alternative forecasts (clusters), or the probability of occurrence of a particular weather event.

Forecasts of probability can provide improved guidance to decision makers over deterministic forecasts as the uncertainty of the forecast is communicated. When a probabilistic forecast is used in conjunction with a cost-loss model, the user will make decisions that over time have smallest economic impact. The process can be illustrated by summarising the decisions from a forecast in a 2x2 contingency table (Table 3.1).

		Event forecast	
		Yes	**No**
Event observed	**Yes**	Hit Mitigated Loss (L_m)	Miss Loss (L)
	No	False Alarm Cost (C)	Correct Rejection No Cost or Loss

Table 3.1 – 2x2 Contingency table to illustrate the cost-loss decision model used with probabilistic forecasts

For the forecast to have value, it is a requirement that $C \leq L_m \leq L$, i.e. the cost of protective action must be less than the losses incurred if no action is taken. Furthermore, the mitigated loss incurred when protective action is taken should be less than the loss incurred without that action (i.e. the preventative action must serve to reduce the loss). Over many events, the strategy that serves ti minimise the

economic impact is to take action when the probability of the event occurring is greater than the ratio of costs to losses, $\frac{C}{L}$ (Mylne, 2002).

3.2. Description of the model

The ECMWF varEPS is a global model discretized horizontally onto a spectral grid and vertically onto a terrain following hybrid coordinate system. The grid resolution is expressed in terms of the wavenumber of the shortest wavelength that the model is able to resolve. For example, T639 L62 refers to 639 wavenumbers in the horizontal (approximately 30km) and 62 vertical levels.

For short lead-times (< 10 days), the forecast is essentially an initial value problem. For longer lead times, ocean variability exerts an increasing influence on atmospheric dynamics. For this reason after 10 days the atmospheric varEPS is coupled with the Hamburg Ocean Primitive Equation (HOPE) model (developed at the Max Planck Institute for Meteorology, Hamburg, Germany) daily at 00 UTC via the OASIS coupler (developed by CERFACS: Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique).

The model is run twice daily with 50 perturbed ensemble members plus the control member at 00 and 12 UTC. For day 1-10 the model is run at T639 L62 (~30km) resolution. For day 11-15 the resolution is degraded to T319 L62 (~60km) resolution with ocean coupling at 00 UTC. Once a week (on Thursday) an extension of the 00 UTC varEPS is run with ocean coupling at 00 UTC and resolution degraded to T639 L62 from day 11.

As forecast lead time increases, model drift becomes a problem. In order to reduce the effects of model drift in the monthly forecasts, 5-member ensemble hindcasts are run for the same initialisation time for the previous 18 years. The hindcasts provide model climatology that is used to calibrate the monthly forecast.

3.3. Generation of initial conditions and model perturbations

The initial atmospheric state for one ensemble, the control member, is provided by the ECMWF operational analysis. 50 perturbed realisations of the operational analysis are used to provide the initial state to the other ensemble members. The perturbations added to the operational analysis are computed by a combination of three methods:

i. The singular vector (SV) method computes perturbations on wind, temperature and pressure that will maximize the impact on total energy in the hemisphere away from the tropics. The SV method is computationally costly, so the computation is done at low resolution (T42). In the tropics (30° N - 30° S) moisture processes have different properties to those at mid-latitudes so a version of the SV is created using a linearised diabatic version of the model.

ii. The perturbations generated by the SV are modified using differences between members of an ensemble data assimilation system. The ensemble data assimilation system is a 10-member set of six hour forecasts generated from initial states with small perturbations added to atmospheric observations and sea surface temperature with stochastic model physics.

iii. To represent the uncertainty in the model dynamics, stochastic physics and stochastic backscatter are used. Stochastic physics perturbs the tendencies in the parametrization schemes that model sub-grid scale processes, and stochastic backscatter models the uncertainty in the sub-grid scale kinetic energy by perturbing the vorticity tendencies.

The perturbations to be applied to the initial state are linearly combined and multiplied by 25 coefficients randomly sampled from a Gaussian distribution. A second set of 25 perturbations are obtained by reversing the sign of the first set resulting in 50 perturbed initial states plus the unperturbed initial state for the control forecast.

3.4. SPI from the varEPS

For the computation of SPI, model precipitation is interpolated from the spectral grid to a regular latitude-longitude grid with 0.5° x 0.5° horizontal resolution using a 4-point bi -linear interpolation. Before interpolation trace amounts of precipitation are adjusted to zero.

3.4.1. Reference statistics

The varEPS has been running operationally since 2004. The most recent operational cycle has been running since 2008. This means that the time-series of precipitation forecasts does not meet the requirement of 30 years of continuous data (McKee *et al*., 1993) needed to fit a parametric distribution. In order to generate the reference statistics, monthly precipitation from the ECMWF Interim Reanalysis (ERA-I) (Dee *et al.*, 2011) is used. ERA-I is a reanalysis of the atmospheric state with horizontal resolution T255 (~75km) and 6-hour temporal resolution and is part of the same family of models as the varEPS. It uses a state-of-the-art 4-dimensional variational (4DVAR) data assimilation system and extends from 1979 to the present day. 24-hour forecasts of precipitation totals are available daily at 00 UTC and are aggregated to monthly totals to provide a time-series of monthly precipitation.

Time-series of n-month precipitation totals, where n is the SPI timescale, from 1981-2010 are used to generate reference statistics. Following Lloyd-Hughes and Saunders (2002) the two-parameter gamma distribution is fitted to the data. However, as recommended by Guttman (1999), the parameters of the gamma distribution are derived from the L-moments. It should be noted that although ERA-I is from the same family of models as varEPS, the reference statistics are generated from daily precipitation forecasts aggregated to monthly totals that may not belong to the

same distribution as the monthly forecasts. However, due to the short time-series of varEPS data, ERA-I is the closest precipitation dataset available for data homogeneity.

3.4.2. SPI Forecasts

varEPS forecasts to 32 days ahead once a week (on Thursday). As a result, the precipitation total on the last day of the month is the total from the forecasts initialisation date to the end of month. When the first Thursday in the month is not on the first day of month, the total precipitation for each ensemble member on the last day of the month is added to the total precipitation at the initialization time of the 32-day forecast from the control member initialised at 00 UTC on the first day of the month. The control member is used as the random nature of ensemble generation means that the same ensemble member from a later forecast will not necessarily be following the same evolution as a previous forecast.

To compute the SPI-1, the cumulative probability of the forecast monthly precipitation from each ensemble member is computed relative to the gamma distribution for the 1-month ERA-I precipitation and transformed to the standard normal variable as described in section 2. For longer timescale SPI-*n*, forecast totals need to be generated. This is done by adding each ensemble member of the 1-month forecast to the ERA-I precipitation totals for the previous *n*-1 months, where *n* is the SPI timescale. In this way, the forecast is conditioned on the observations from the previous n-1 months. As with the SPI-1, the SPI-*n* is computed relative to the gamma distribution for *n*-month ERA-I precipitation:

$$SPIn_{i,j+1}(n > 1) = f(P(EPS_{i,j+1} + \sum_{k=0}^{n-2} ERA_{j-k})) \quad (3.1)$$

In other words, the forecast SPI-*n* for ensemble member *i* and month *j+1* is a function of the probability the forecast precipitation for ensemble member *i* for month *j+1*, the first term on the right hand side of equation (3.1) plus the total precipitation from ERA-I from month *j-(n-2)* to month *j*, the second term on the right hand side of equation (3.1).

4. Forecast Verification

Forecasts of SPI from varEPS were generated for the latest operational cycle of the model implementation. This provided 41 forecasts of SPI from March 2004 to August 2011. Forecasts of the probabilities of SPI-1 and SPI-3 for SPI thresholds of -1 (moderate drought), -1.5 (severe drought), and -2 (extreme drought) are verified against SPI derived from independent observations. The SPI-1 represents a pure forecast in that the precipitation totals used to calculate the SPI come from the forecasting system alone. The SPI-3 forecast, however, is a forecast conditioned on observations in that the first two months of the precipitation are from the ERA-I and only the third month is from the forecasting system.

The forecast verification methods used are the Brier Score (BS), the Brier Skill Score (BSS), reliability and Relative Operating Characteristics (ROC). These methods are described in the following sections.

4.1. Observational data

Observed SPIs are computed from the E-OBS dataset (Haylock *et al.*, 2008). E-OBS is an interpolation of daily precipitation observations from approximately 2500 rain gauges throughout Europe onto 0.25° x 0.25° and 0.5° x 0.5° grids at land points only. For the purposes of forecast verification, the coarser 0.5° x 0.5° resolution dataset is used as it is the same resolution as the forecast. The precipitation observations are fully quality controlled and homogenised. Gaps in the time-series for a station are filled with data from neighbouring stations if the neighbouring stations are within 12.5 km and have a height difference of no more than 25m. Observed SPIs are calculated following the method described in section 2 using a gamma distribution with parameters estimated from L-moments and the same reference period as the forecasts: 1981 – 2010.

The verification is done for a domain stretching from 15.5°W to 60.5°E longitude and 35.5°N to 75.5°N latitude for pixels where observations exist (i.e. only land points).

4.2. Brier Score

The Brier Score (BS) is used to measure the Mean Squared Error (MSE) of a probabilistic forecast (Brier, 1950). For *N* forecasts it is calculated as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2 \,, \tag{4.1}$$

where p_i is the forecast probability and o_i is the observation: o_i = 1 when the event occurs and o_i = 0 when the event does not occur. The BS has a range from 0 to 1, with 0 representing the perfect score. It is sensitive to the climatological frequency of the event, in that the more rare an event the easier it is to achieve a good BS without having any real forecast skill.

Murphy (1973) showed that the BS can be partitioned into three separate terms: reliability, resolution and uncertainty for *K* probability categories as follows

$$BS = \frac{1}{N}\sum_{k=1}^{K} n_k \left(p_k - \bar{o}_k\right)^2 - \frac{1}{N}\sum_{k=1}^{K} n_k \left(\bar{o}_k - \bar{o}\right)^2 + \bar{o}\left(1 - \bar{o}\right), \qquad (4.2)$$

BS = RELIABILITY - RESOLUTION + UNCERTAINTY

where *k* is the probability category and *n* is the number of forecasts in the probability category. The reliability measures how close the forecast probabilities are to the observed frequency, with 0 representing perfect reliability. The reliability can also be illustrated with a reliability diagram (see section 4.4). The resolution measures the ability of the forecasting system to forecast events in different probability categories with larger values indicating better resolution. The uncertainty measures the inherent uncertainty of the event (not the uncertainty in the forecast) and has a maximum when the event occurs 50% of the time is zero when the event always or never occurs.

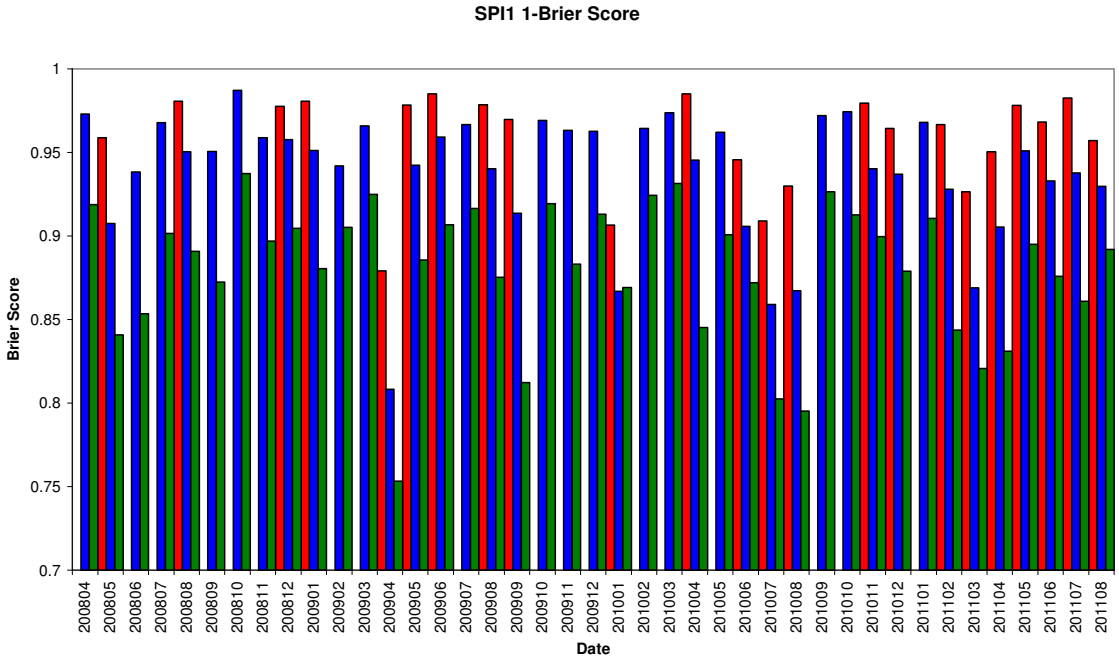
For Brier Score and its components for all 41 months of the forecast are summarised in Table 4.1, and time-series of 1-BS (the BS is negatively oriented so 1-BS is shown for clarity) for each forecast are shown in Figure 4.1.

	SPI-1<-1	SPI-1<-1.5	SPI-1<-2	SPI-3<-1	SPI-3<-1.5	SPI-3<-2
BS	0.119	0.061	0.027	0.108	0.054	0.025
Reliability	0.011	0.007	0.004	0.014	0.008	0.004
Resolution	0.009	0.003	0.001	0.026	0.010	0.004
Uncertainty	0.118	0.057	0.024	0.120	0.057	0.025

Table 4.1 – The Brier Score and its components for probabilistic forecasts of SPI-1 and SPI-3 for thresholds of -1 (moderate drought), -1.5 (severe drought) and -2 (extreme drought). The statistics are generated from forecasts for the period March 2004 to August 2008.

Table 4.1 shows that as the event becomes rarer (event uncertainty closer to zero), the BS and reliability seem to approve as the values get closer to zero. However, the resolution of the forecast worsens. The SPI-3 for all thresholds has a better BS than the SPI-1 with improved resolution, although for thresholds of -1 and -1.5 the reliability of the SPI-3 forecast is not so good. The better BS for the SPI-3 is to be expected as the forecasts are conditioned on two months of reanalysis precipitation.

(a)



(b)

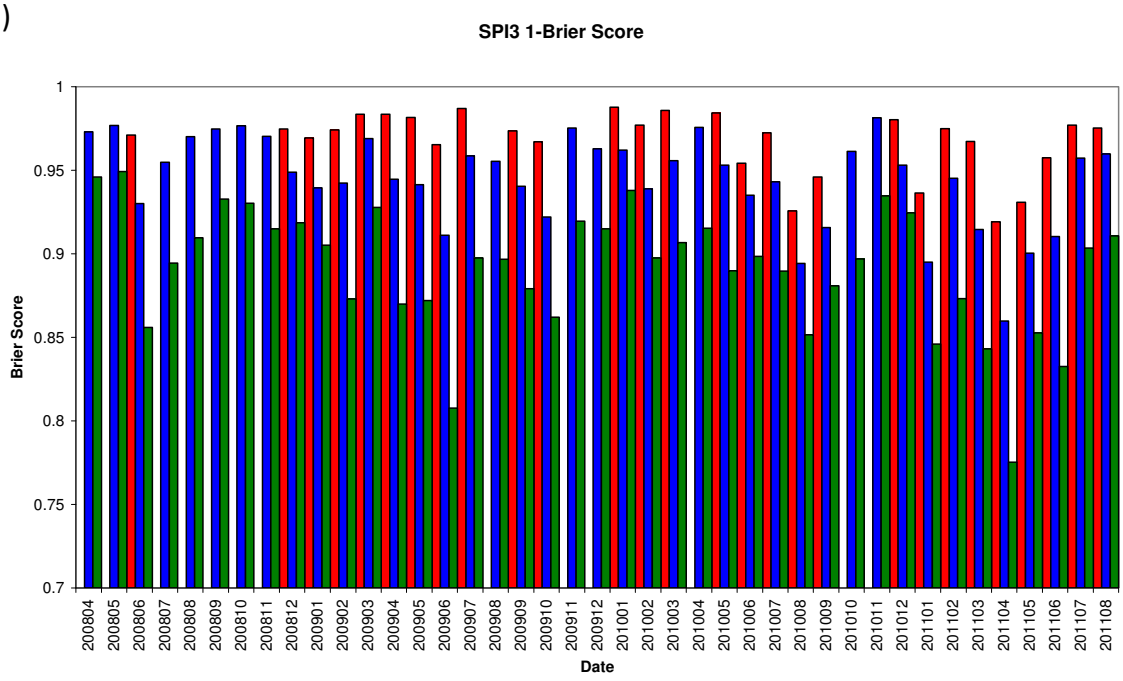


Figure 4.1 – 1 - Brier Score for each monthly forecast of (a) SPI-1 and (b) SPI-3. The thresholds are SPI<-2 (red), SPI<-1.5 (blue) and SPI<-1 (green). The Brier Score was only calculated fore forecasts where more than 1% of the verification area met the threshold.

Similar to the total BS, the BS for each forecast is better for the more rare events of severe (SPI<-1.5) and extreme (SPI<-2) droughts. The BS is on the whole marginally better for the SPI-3. There are some forecasts that stand out as having less skill than the others. For the SPI-1, the BS is noticeably inferior for the April 2009 forecast. At this time much of eastern Europe was experiencing extreme drought according to the SPI-1 derived from E-OBS date (Fig. 4.2(a)). Extreme drought conditions were most prevalent in Poland, the Ukraine, Belarus, the Baltic countries, north-western Romania, eastern Hungary and north-eastern Serbia. The forecast probabilities for this event show that the spatial extent of the SPI-1 drought was reasonably well predicted for the SPI-1<-1 threshold (Fig. 4.2(b)), albeit with relatively low probability. For the SPI-1<-1.5 (Fig. 4.2(c)) and SPI-1<-2 (Fig. 4.2(d)) thresholds the drought area was forecast to be concentrated around the area of Serbia and southern Hungary, again with relatively low probabilities. The failure of the forecast to predict a high probability of the SPI-1 moderate drought, and the forecast underestimating the spatial extent of the extreme SPI-1 drought contributed to the relatively poor Brier Score for this event. For the SPI-3, the BS is relatively poor for April 2011. This event is discussed in more detail in section 6.2.

4.3. Brier Skill Score

The BS alone as a verification method can often give misleading information. For rare events it is easier to achieve a good BS as the computation is dominated by non-events. The Brier Skill Score (BSS) is used to put the BS into context by comparing the BS from the forecasting system to the BS from a reference forecast. The climatological frequency of the event is normally used the reference forecast, which in the case of the SPI is simply the cumulative probability associated with the SPI class (Table 2.1). BSS is calculated as

$$BSS = 1 - \frac{BS_{fcst}}{BS_{ref}}, \tag{4.3}$$

Where BS_{fcst} is the BS for the forecasting system and BS_{ref} is the BS for the reference forecast. Values of BSS ≤ 0 indicate no skill in the forecasting system when compared to the reference forecast. A value of BSS = 1 indicates a perfect forecast. BSS for all thresholds for SPI-1 is negative indicating that the forecasting system is no more skilful than the reference forecast (Table 4.2). However, for SPI-3, where the forecasts are conditioned on two months of observations BSS is positive for the moderate and severe drought classes, but negative for the extreme drought class (Table 4.2). This suggests that the forecasting system has some skill for the SPI-3 compared to the reference forecast.

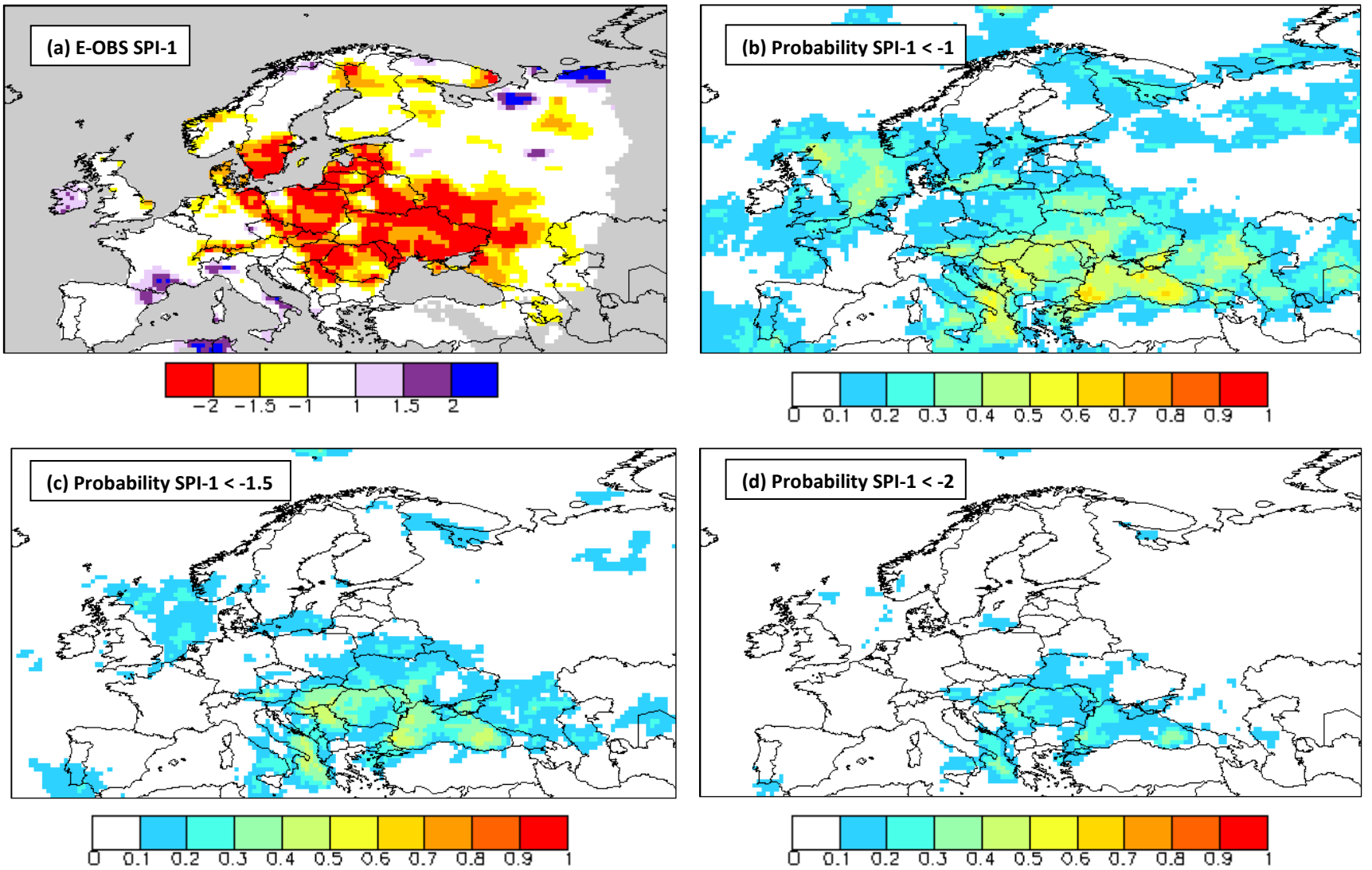


Figure 4.2 – SPI-1 April 2009. (a) Observed SPI-1 (grey areas indicate missing data); (b) Forecast probability of SPI-1<-1; (c) Forecast probability of SPI-1<-1.5 and (d) Forecast probability of SPI-1<-2.

	SPI-1<-1	SPI-1<-1.5	SPI-1<-2	SPI-3<-1	SPI-3<-1.5	SPI-3<-2
BSS	-0.008	-0.060	-0.112	0.108	0.054	0.025

Table 4.2 - The Brier Skill Score for probabilistic forecasts of SPI-1 and SPI-3 for thresholds of -1 (moderate drought), -1.5 (severe drought) and -2 (extreme drought). The statistics are generated from forecasts for the period March 2004 to August 2008 and from the probabilities of the SPI thresholds.

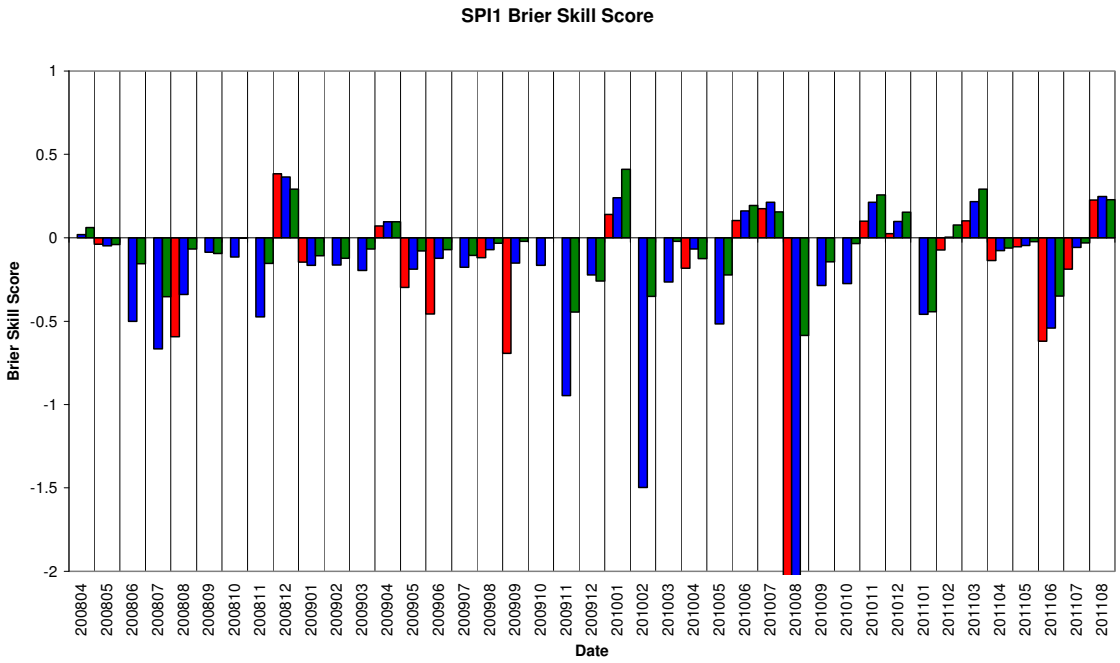
Time-series of BSS for SPI-1 and SPI-3 for each of the forecast months are shown in Fig. 4.3. For SPI-1 (Fig. 4.3(a)), the BSS is negative for most forecasts. The SPI-1 BSS is particularly poor for August 2010, a case which is discussed in section 6.1. In a limited number of forecasts, the SPI-1 BSS is positive. For example, in December 2008 the BSS for all thresholds is approaching 0.5. In this case, extreme drought relating to the SPI-1 was observed over south-western Russia, near to the border with Kazakhstan, the Netherlands and north-western Germany (Fig. 4.4(a)). The probabilistic forecasts did a reasonably good job of predicting the drought in Russia with relatively high probabilities for all thresholds (Fig. 4.4(a-d)). However, for the Netherlands and northern Germany the probabilities were lower and central and southern England were forecast to have similar conditions, which was not observed.

For the SPI-3, more than half of the forecasts achieved a positive BSS. The conditioning of the forecasts on 2 months of reanalysis data clearly improves performance against the reference forecast.

4.4. Reliability

As mentioned in section 4.2, an important feature of a probabilistic forecast is reliability. This is the ability of the forecast to predict probabilities that correspond with observed frequencies of the events. For example, if an event is forecast with a probability of 80% then for a perfectly reliable forecasting system the event would be observed 8 times out of every 10 that this forecast is issued. The reliability of a forecast can be illustrated in a reliability diagram which is a plot of the observed frequency as a function of the forecast probability. As well as the forecast reliability, the reliability diagram shows the *perfect reliability* line, the *no skill* line and the *no resolution* line (Fig. 4.5(a)). The *perfect reliability* line is simply the diagonal where the observed frequency is equal to the forecast probability. The closer the measured reliability is to this line, the more the reliable the forecast. If the measured reliability is below (above) the *perfect reliability* line, the forecasting system over (under) forecasts the probabilities i.e. the probabilities are too high (low). The *no resolution* line is simply the climatological frequency, or the sample frequency. Since this only has 1 value the forecast has no resolution i.e. it cannot discriminate between events and non-events. The *no skill* line is half way between the *no resolution* and *perfect reliability* lines. Measured reliability between the *no skill* line and the *perfect reliability* line indicates that the forecast has skill compared to climatology i.e. a positive contribution to the BSS.

(a)



(b)

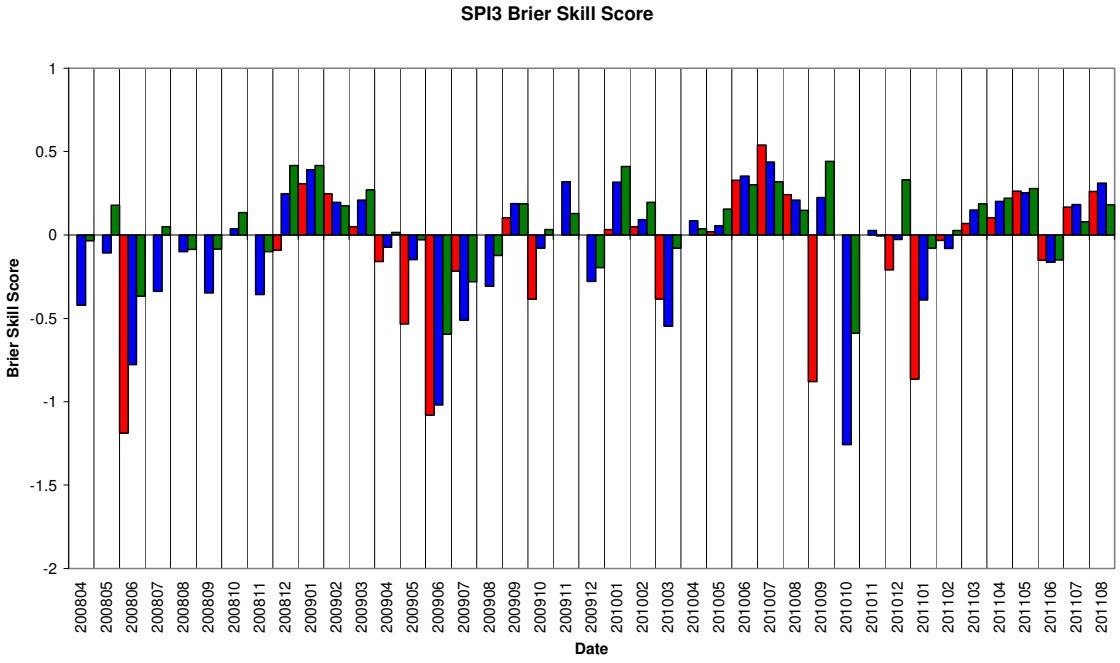


Figure 4.3 – Brier Skill Score for each monthly forecast of (a) SPI-1 and (b) SPI-3. The thresholds are SPI<-2 (red), SPI<-1.5 (blue) and SPI<-1 (green). The Brier Skill Score was only calculated fore forecasts where more than 1% of the verification area met the threshold.

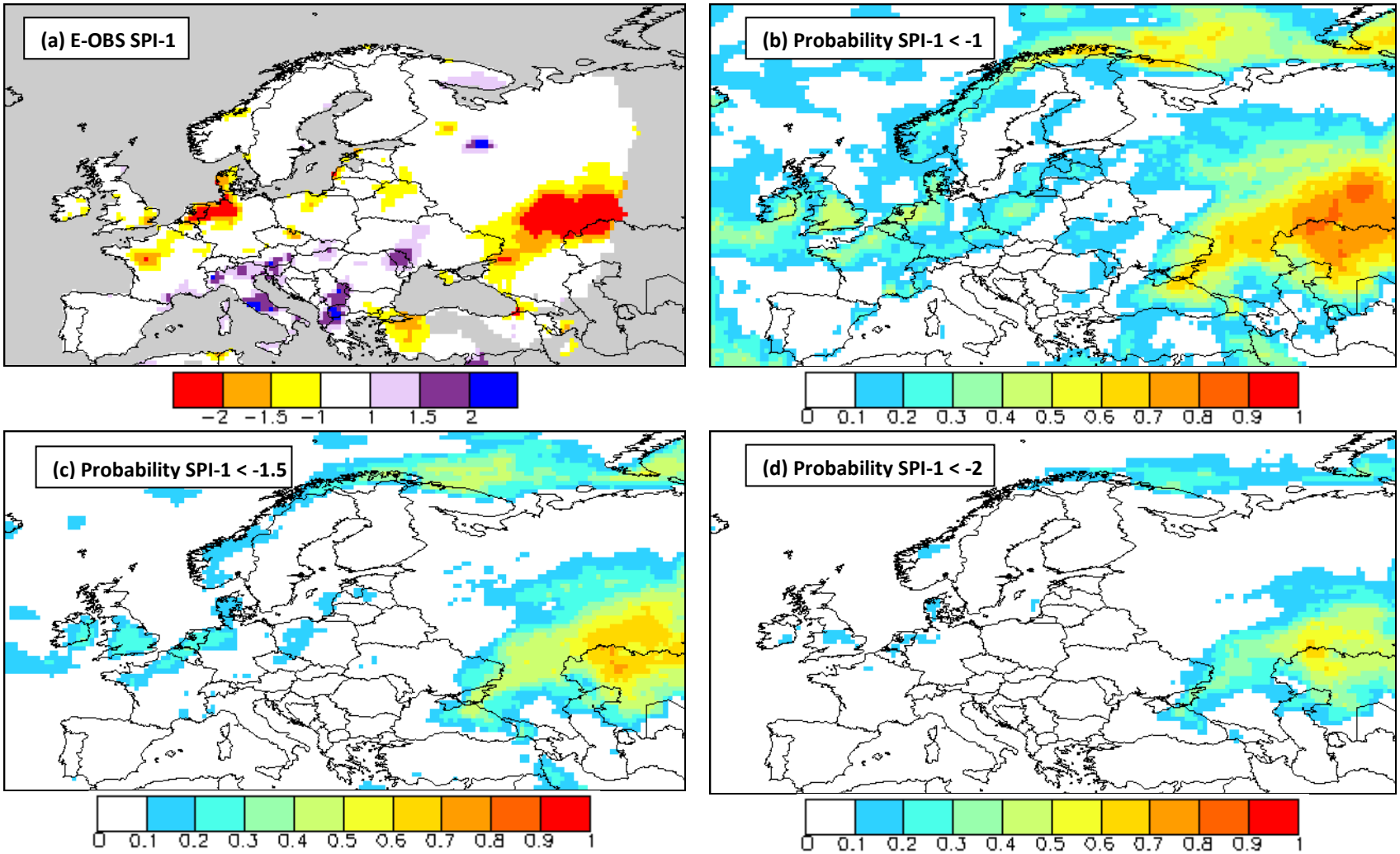


Figure 4.4 – SPI-1 December 2008. (a) Observed SPI-1 (grey areas indicate missing data); (b) Forecast probability of SPI-1<-1; (c) Forecast probability of SPI-1<-1.5 and (d) Forecast probability of SPI-1<-2.

The reliability diagrams for the SPI-1 are shown in Fig. 4.5. For all thresholds, the reliability is around the same level as the no-skill line before dropping off as the probability increases. As the SPI threshold goes towards the extreme the drop off occurs at a lower probability level. The fact that the measured reliability is below the *perfect reliability* line suggests a bias towards overforecasting the probabilities.

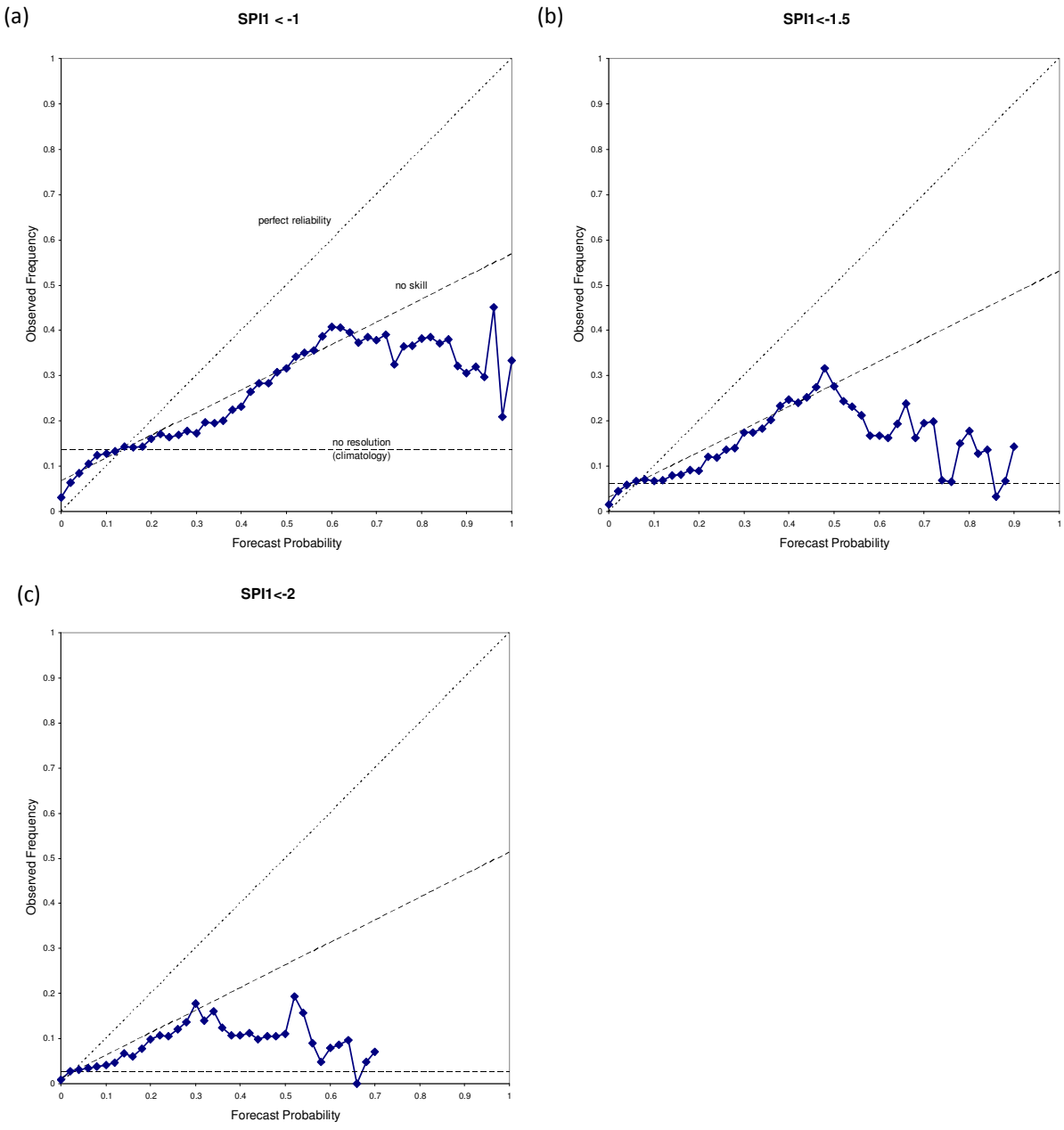


Figure 4.5 – Reliability diagrams for SPI-1 for all forecasts from March 2008 to August 2011 for thresholds of (a) SPI-1 < -1 (moderate drought), (b) SPI-1 < -1.5 (severe drought) and (c) SPI-1 < -2 (extreme drought). The *perfect reliability*, *no skill* and *no resolution* lines are labelled in (a).

Furthermore, the fact that for most probabilities the reliability is below the *no skill* line confirms the negative BSS values for the SPI-1 (Table 4.2). It should be noted that the *no resolution* lines, and therefore the *no skill* lines, in Figure 4.5 are derived from the

sample climatology for the verification period of March 2008 to August 2011. However, the sample climatologies for each threshold are broadly similar to the expected probabilities in Table 2.1.

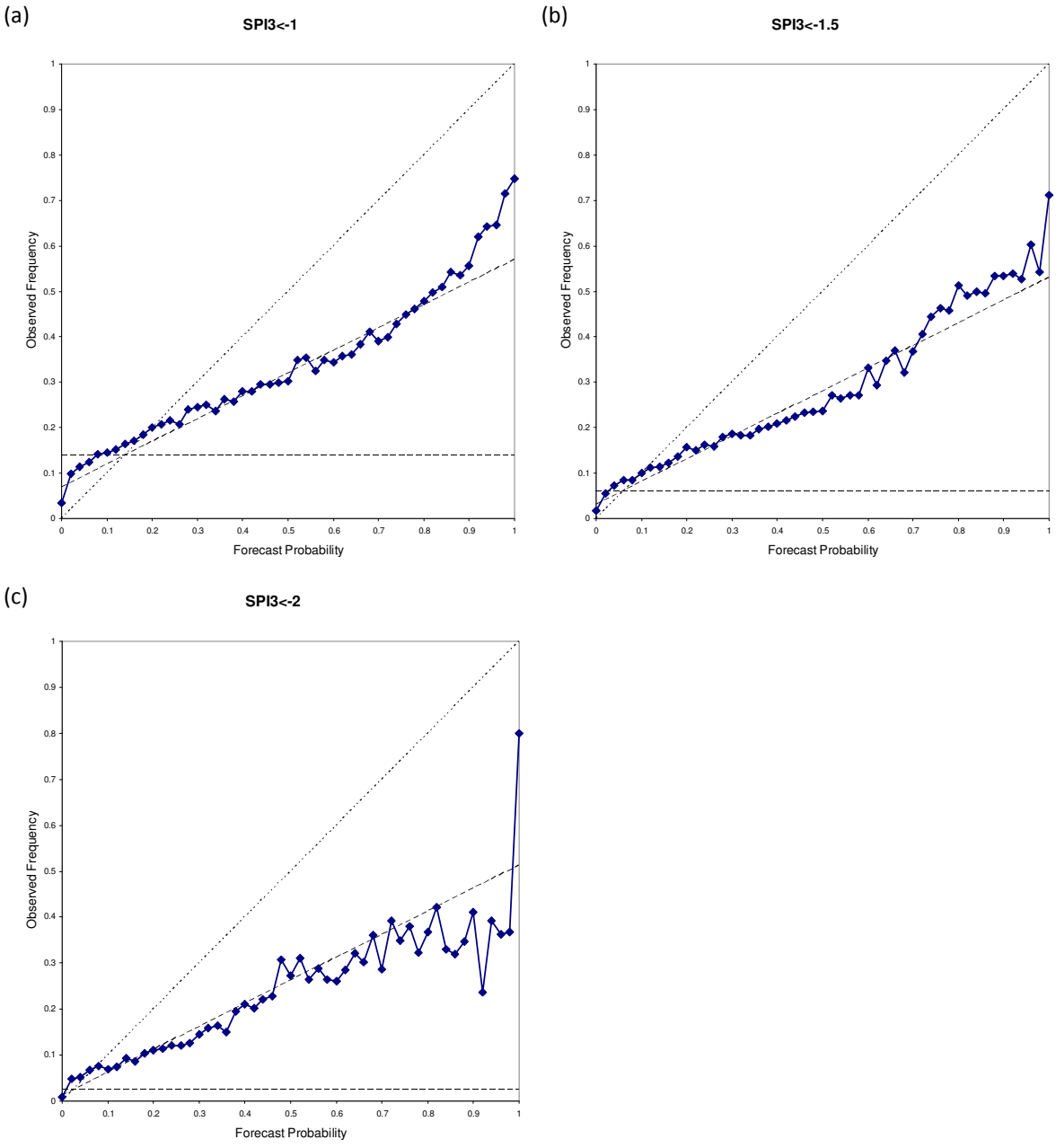


Figure 4.6 – Reliability diagrams for SPI-3 for all forecasts from March 2008 to August 2011 for thresholds of (a) SPI-3 < -1 (moderate drought), (b) SPI-3 < -1.5 (severe drought) and (c) SPI-3 < -2 (extreme drought).

The reliability diagrams for SPI-3 are shown in Figure 4.6. Unlike for the SPI-1 (Fig. 4.5) there is no drop off in the reliability as the forecast probability increases. This is likely due to the forecasts being conditioned on the 2-months of reanalysis precipitation. For all thresholds, the reliability generally follows the *no skill* line, although for the less extreme thresholds the reliability appears to be slightly better for the higher probabilities.

4.5. Relative Operating Characteristic

The reliability diagram (section 4.4) is conditioned on the forecast. In other words it answers the question *given X was predicted, what was the outcome?* It gives information about the real meaning of the forecast. A good companion to the reliability diagram is the Relative Operating Characteristic (ROC) of the forecast. ROC is conditioned on the observations, answering the question *given that Y occurred, what was the corresponding forecast?* It therefore measures the ability of the forecasting system to discriminate between events and non-events, i.e. the resolution of the forecast. ROC is not sensitive to bias in the forecast – a biased forecast could give a good ROC, but it will say nothing about the reliability of the forecast. This means that ROC is a measure of potential usefulness of the forecast, and with good ROC it may be possible to improve the forecast through calibration.

ROC is calculated by means of a 2x2 contingency table for each probability as in Table 4.3, which counts the number of forecast hits (*H*), the number of misses (*M*), the number of false alarms (*FA*) and the number of correct rejections (*CR*).

		Event forecast	
		Yes	**No**
Event observed	**Yes**	Hit (H)	Miss (M)
	No	False Alarm (FA)	Correct Rejection (CR)

Table 4.3 – 2 x 2 Contingency table for ROC calculation

ROC is then probability of detection (PoD) as a function of the false alarm rate (FAR), where

$$PoD = \frac{H}{H + M} \tag{4.4}$$

$$FAR = \frac{FA}{FA + H} \tag{4.5}$$

A ROC curve is plotted as a curve joining the PoD as a function of the FAR for all forecast probabilities. The area under the ROC curve gives a measure of the skill of the forecast. ROC curves for each threshold are shown in Figure 4.7 for (a) SPI-1 and (b) SPI-3.

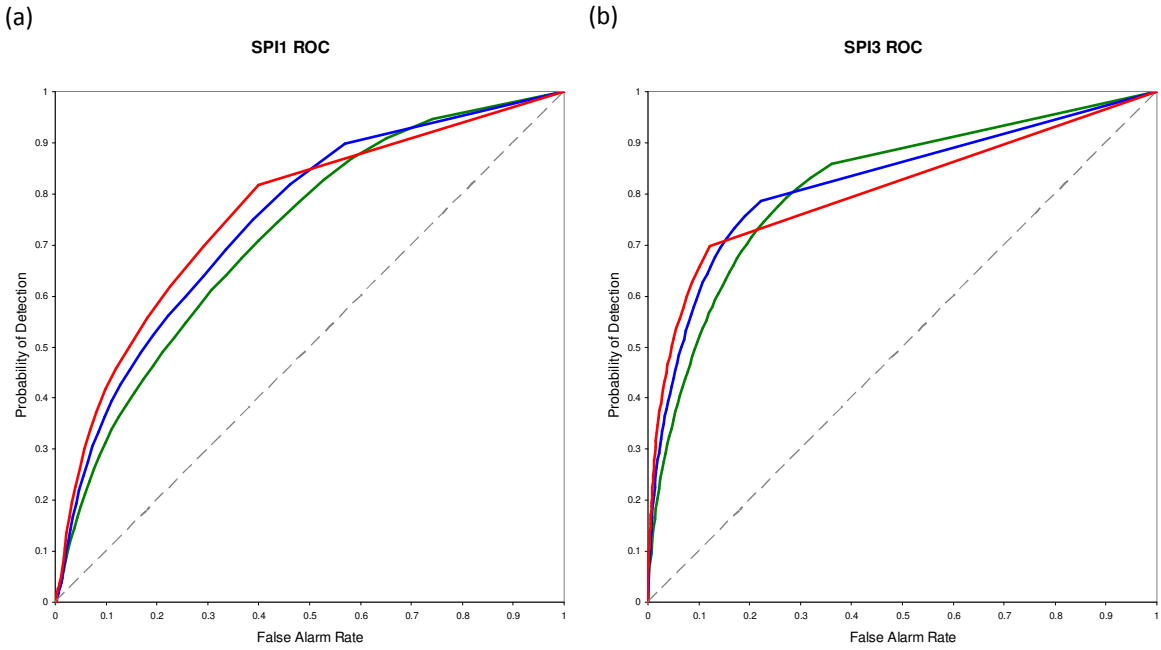


Figure 4.7 – Relative operating characteristic (ROC) curves for forecasts of (a) SPI-1 and (b) SPI-3 for all forecasts from March 2008 to August 2011. The green curves are for the SPI < -1 threshold, the blue curves are for the SPI < -1.5 threshold and the red curves are for the SPI < -2 threshold. The grey dashed diagonal line is the no skill line – ROC below this line indicates no skill.

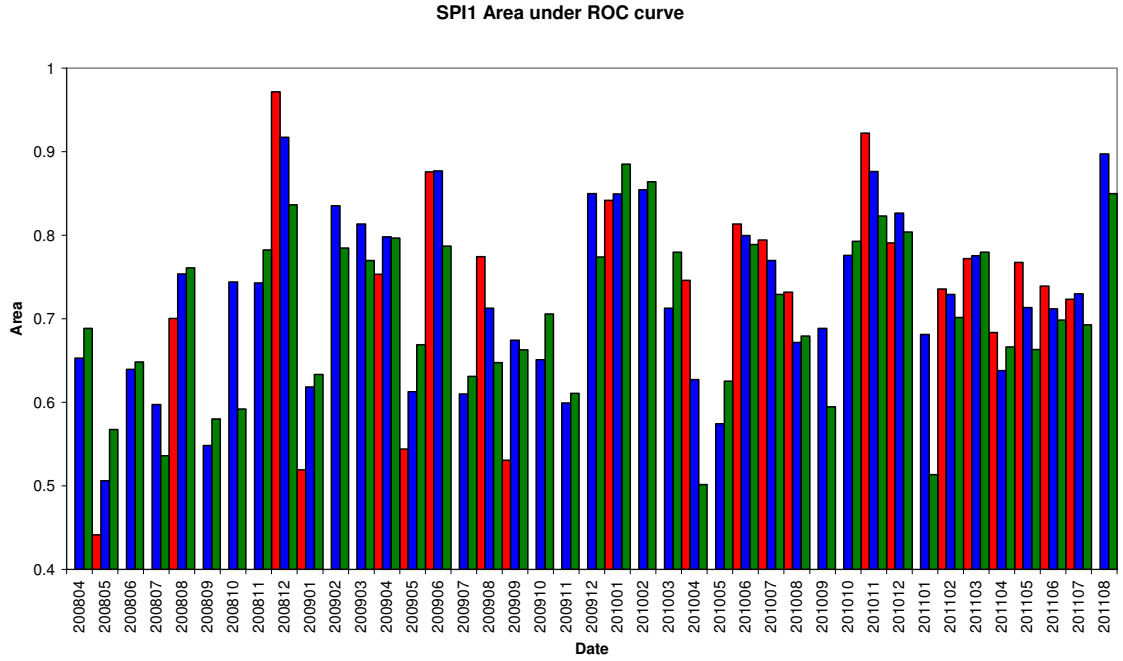
For the SPI-1 and the SPI-3 for all thresholds the ROC curves are well above the *no skill* line indicating that, despite the poor reliability, the forecasting system does have some skill. The areas under the ROC curves are shown in Table 4.4. They show that the SPI-3 forecast is more skilful than the SPI-1 forecast, as would be expected, and that the forecasting system has potentially greater skill for the rarer events with more extreme thresholds for SPI.

	SPI < -1	SPI < -1.5	SPI < -2
SPI-1	0.725	0.774	0.814
SPI-3	0.838	0.871	0.891

Table 4.4 – Area under the ROC curves for SPI-1 and SPI-3

Time series of the area under the ROC curves for each threshold are shown in Figure 4.8 for (a) SPI-1 and (b) SPI-3.

(a)



(b)

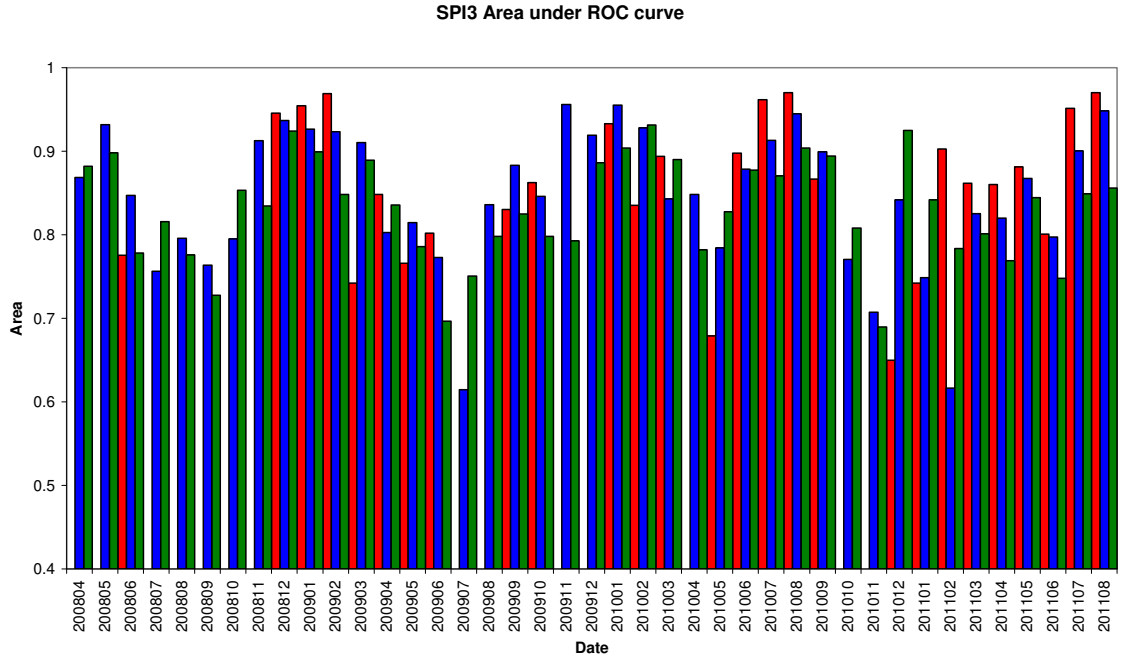


Figure 4.8 – Area under ROC curve for each monthly forecast of (a) SPI-1 and (b) SPI-3. The thresholds are SPI<-2 (red), SPI<-1.5 (blue) and SPI<-1 (green). The ROC was only calculated fore forecasts where more than 1% of the verification area met the threshold.

The area under the ROC curve is quite variable from one forecast to the next, but is in almost all cases greater than 0.5 indicating that the forecast has some skill. Only the SPI-1 in May 2008 for the SPI-1 < -2 threshold has an area under the ROC curve less than 0.5. This appears to be due to an extreme drought in the SPI-1 over coastal areas around the Baltic Sea as well as central Germany, western Norway and parts of Scotland (Fig. 4.9) that was not forecast at all with any probability, even for the SPI-1 < -1 threshold.

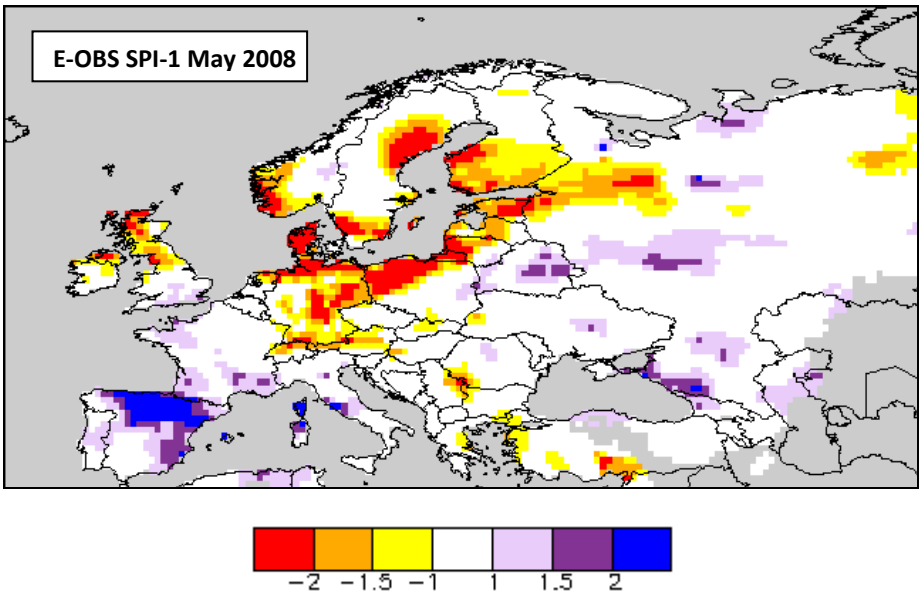


Figure 4.9 – Observed SPI-1 (grey areas indicate missing data) for May 2008.

4.6. Forecast Calibration

The ROC statistics suggest that the forecasting system is potentially skilful in that it is able to discriminate between events and non-events. However, the reliability diagrams show that the system is biased towards overforecasting. Calibration of the forecast aims to remove the bias and forecast probabilities that are comparable with the observed frequencies of the events.

The calibration strategy adopted uses the reliability diagrams to adjust the forecast probabilities towards their associated observed frequencies. One of the outcomes of this strategy is that the maximum probability that the calibrated system can forecast is limited by the observed frequency of the event for the uncalibrated forecast probability. For example, if the reliability shows the that events that are forecast with 100% probability are only observed 30% of the time, the maximum probability forecast of the calibrated system will be 30%. In essence, the calibrated system puts a limit on the certainty that the forecast can communicate. Care will need to be taken in interpreting such forecasts as low probabilities will be less likely to trigger a response when perhaps one is required.

Figure 4.10 shows the effect of forecast calibration on BSS for (a) SPI-1 and (b) SPI-3. The BSS for SPI-1 for all thresholds has become positive for the majority of forecasts, and for SPI-3 the BSS has increased in value for all forecasts. This suggests that the calibrated forecast is more skilful than a forecast based on climatology, unlike that uncalibrated

forecast. Table 4.5 shows the verification statistics (BS and its decomposition, BSS and area under the ROC curve) aggregated over the 41 forecasts from March 2004 to August 2008, and Table 4.6 shows the change in those statistics compared with the uncalibrated forecast (Tables 4.1, 4.2, 4.4).

	SPI-1<-1	SPI-1<-1.5	SPI-1<-2	SPI-3<-1	SPI-3<-1.5	SPI-3<-2
BS	0.110	0.055	0.024	0.094	0.047	0.021
Reliability	0.000	0.000	0.000	0.000	0.000	0.000
Resolution	0.010	0.004	0.001	0.026	0.010	0.004
Uncertainty	0.118	0.057	0.024	0.120	0.057	0.025
BSS	0.086	0.064	0.040	0.220	0.183	0.145
ROC Area	0.723	0.752	0.766	0.817	0.822	0.808

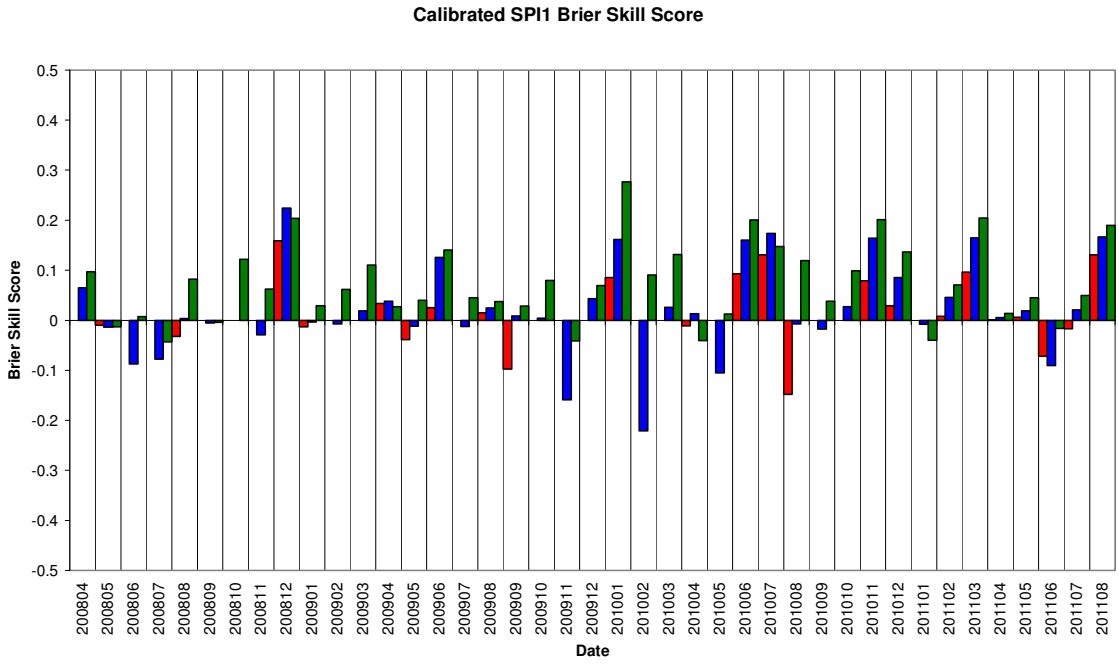
Table 4.5 – Verification statistics for calibrated forecast aggregated over all forecasts from March 2004 to August 2011.

	SPI-1<-1	SPI-1<-1.5	SPI-1<-2	SPI-3<-1	SPI-3<-1.5	SPI-3<-2
BS	-0.009	-0.006	-0.003	-0.014	-0.008	-0.004
Reliability	-0.011	-0.007	-0.004	-0.014	-0.008	-0.004
Resolution	0.001	0.000	0.000	0.000	0.000	0.000
Uncertainty	0.000	0.000	0.000	0.000	0.000	0.000
BSS	0.094	0.124	0.152	0.119	0.135	0.152
ROC Area	-0.002	-0.022	-0.049	-0.021	-0.048	-0.084

Table 4.6 – Difference in verification statistics between the calibrated forecast and the uncalibrated forecast aggregated over all forecasts from March 2004 to August 2011.

The calibration has reduced the value of the BS to closer to zero for all thresholds with the smallest improvements for the most extreme SPI < -2 threshold. Since the calibrated forecast has forced the probabilities to the values of the observed frequency (i.e. perfect reliability), the reliability component of the BS is reduced to zero, as would be expected. There is no appreciable change in the resolution component of the BS of the forecast due to calibration meaning that the skill of the forecasting system in discriminating between events and non-events has not been lost. As previously discussed, the BSS is improved for all of the individual forecasts and this is reflected in the increases in BSS aggregated over the forecasts. The area under the ROC curves is reduced, more so for the more extreme thresholds suggesting a loss of resolution. This is most likely due to a smaller number of probability categories in the calibrated forecast affecting the shape of the ROC curves.

(a)



(b)

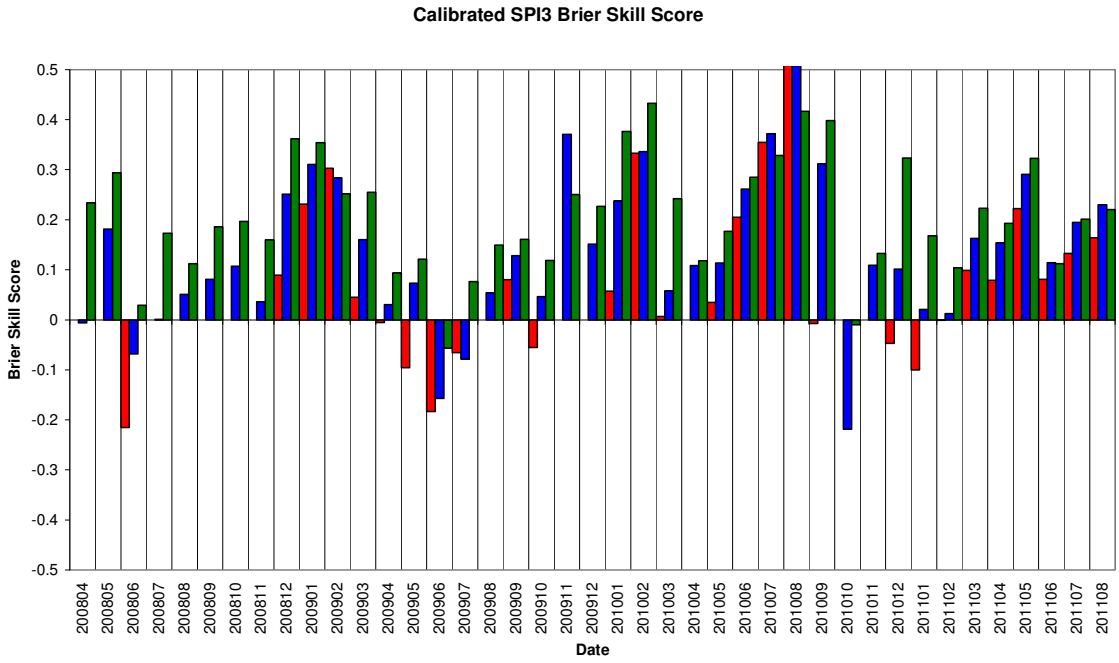


Figure 4.10 – Brier Skill Score for each monthly calibrated forecast of (a) SPI-1 and (b) SPI-3. The thresholds are SPI<-2 (red), SPI<-1.5 (blue) and SPI<-1 (green). The Brier Skill Score was only calculated fore forecasts where more than 1% of the verification area met the threshold.

5. Case Studies

The verification statistics show that varEPS has some skill in forecasting drought events, which is improved through calibration. The two months of reanalysis precipitation used in the computation of SPI-3 forecasts mean that these forecasts are more skilful than those for the SPI-1. In order to further investigate the varEPS performance, two case studies of extreme drought events were investigated: summer 2010 in Russia, and spring 2011 in north-western Europe. For each case study, the performance of both the uncalibrated and the calibrated forecasts was analysed.

5.1. Summer 2010 – Russia

An extreme drought affected the central part of European Russia in the summer of 2010. Much of the affected area was the Volga River basin, which is a major source of water for the Caspian Sea (Rodionov, 1994; Golitsyn, 1995; Arpe *et al.*, 2011). The Caspian Sea can be subject to major changes in sea level with magnitudes of variability of up to 3m in the last century affecting industry, infrastructure and livelihoods in the region. The drought resulted from a blocking anticyclone that remained in place over the region for 55 days bringing hot, dry air from the Middle East into the region (Arpe *et al.*, 2011). The hot, dry conditions led to widespread crop losses and extensive forest and grassland fires (Arpe *et al.*, 2011).

Figure 5.1 shows the evolution of the observed SPI-1 from E-OBS for the months from June to September 2010. In June 2010 (Fig. 5.1(a)) a sizeable area in south-western Russia to the north and northwest of the Caspian Sea was affected by extreme drought according the SPI-1. In July 2010 (Fig. 5.1(b)) much of western Russia was affected by extreme SPI-1 drought, but by August 2010 (Fig. 5(c)) only a small area to the northwest of the Caspian Sea was affected. By September 2010 (Fig. 5.1(d)) the majority of the region had returned to near normal conditions for the SPI-1.

The uncalibrated forecasts of the probability of SPI-1 < -1 is shown in Fig. 5.2 for June – September 2010. In June 2010 (Fig. 5.2(a)) varEPS forecast an area to north of the Caspian Sea extending west through the Ukraine into Romania with relatively high probabilities of SPI-1 < -1. The location the north of the Caspian was well forecast, but the westward extent of the relatively high probabilities was not observed, and in fact extreme wet conditions were observed in Romania. In July 2010, when the extreme SPI-1 drought was most widespread in western Russia (Fig. 5.1(b)), relatively high probabilities of SPI-1 < -1 were forecast in this area, but similar probabilities were forecast for much of continental Europe where near normal conditions were observed. In August 2010, when the spatial extent of the SPI-1 drought was restricted to the south-western corner of Russia (Fig. 5.1(c)), the highest probabilities of SPI-1 < -1 were forecast, suggesting that for a large number of ensemble members the atmospheric state in the initial conditions were forecast to persist through the month. In September 2010, when SPI-1 drought conditions had completely ceased in the region (Fig. 5.1(d)) much lower probabilities of SPI-1 < -1 were forecast, but western Russia, to the north of Kazakhstan was still forecast to have some probability of drought conditions. Similar

patterns are forecast for the SPI-1 < -1.5 (Fig. 5.3) and SPI-1 < -2 (Fig. 5.4) probabilities. However, for the more extreme thresholds the forecast probabilities are lower.

The main problems with the SPI-1 forecast are that similar probabilities were predicted in western Europe in July 2010 (Fig. 5.2(b), 5.3(b), 5.4(b)), where there was little evidence of widespread drought conditions (Fig. 5.1(b)), to those in Russia where extreme drought occurred; and that the forecast for August 2010 (Fig. 5.2(c), 5.3(c), 5.4(c)) predicted extreme drought conditions with the highest level of certainty in Russia when observations showed that the drought had become less extreme. The calibrated forecast did not improve on these problems – the main effect of calibration was to lower the probabilities (Fig. 5.5, 5.6, 5.7) – with a probability of SPI-1 < -2 of only 10-20% when the SPI-1 was at its most extreme in July 2010.

For the SPI-3, observations from E-OBS showed extreme drought in the SPI-3 had started in southern Russia to the north of Kazakhstan in June 2010 (Fig. 5.8(a)) with the spatial extent spreading through July 2010 (Fig. 5.8(b)) and August 2010 (Fig. 5.8(c)). Extreme drought SPI-3 conditions had mostly finished by September 2010 (Fig. 5.8(d)). The SPI-3 forecast was somewhat better than the SPI-1 forecast. The probabilities for SPI-3 < -1 (Fig. 5.9) showed with very high probabilities (90-100%) that the drought would begin over southern Russia to the north of Kazakhstan in June 2010 (Fig. 5.9(a)) becoming more widespread through July 2010 (Fig. 5.9(b)) and August 2010 (Fig. 5.9(c)) with the spatial extent reducing in September 2010 (Fig. 5.9(d)). For the more extreme SPI-3 thresholds the probabilities remained high (Fig. 5.10, 5.11) with the extent of the SPI-3 < -2 drought in August 2010 particularly well forecast (Fig. 5.11(c)). The effect of calibration was to reduce the probabilities (Fig. 5.12, 5.13, 5.14), but not by a large amount. The calibrated forecast predicted SPI-3 <-2 in southern Russia with 80-90% certainty (Fig. 5.14(c)), which can be considered to be a very good forecast.

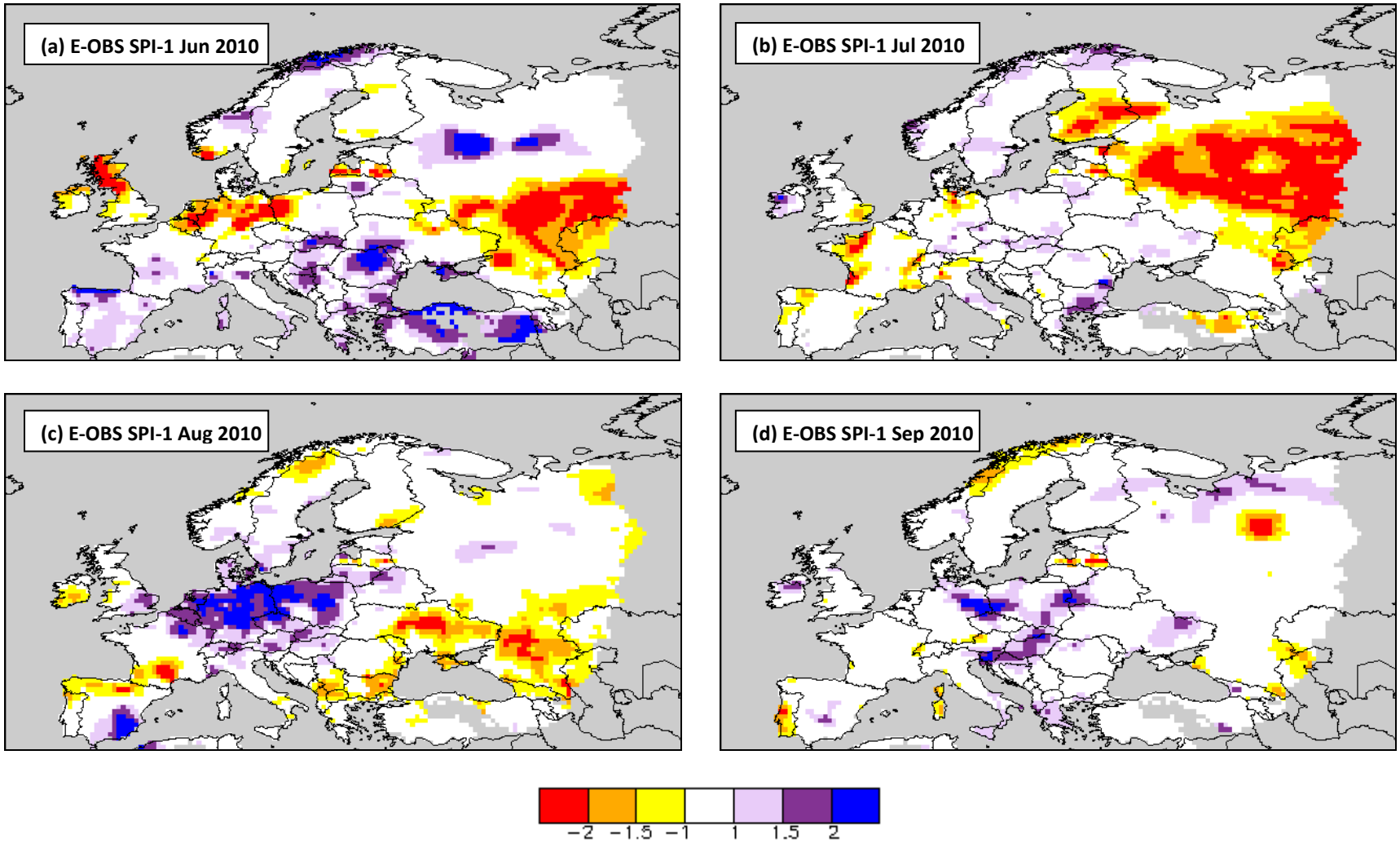


Figure 5.1 – Observed SPI-1 (grey areas indicate missing data) for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

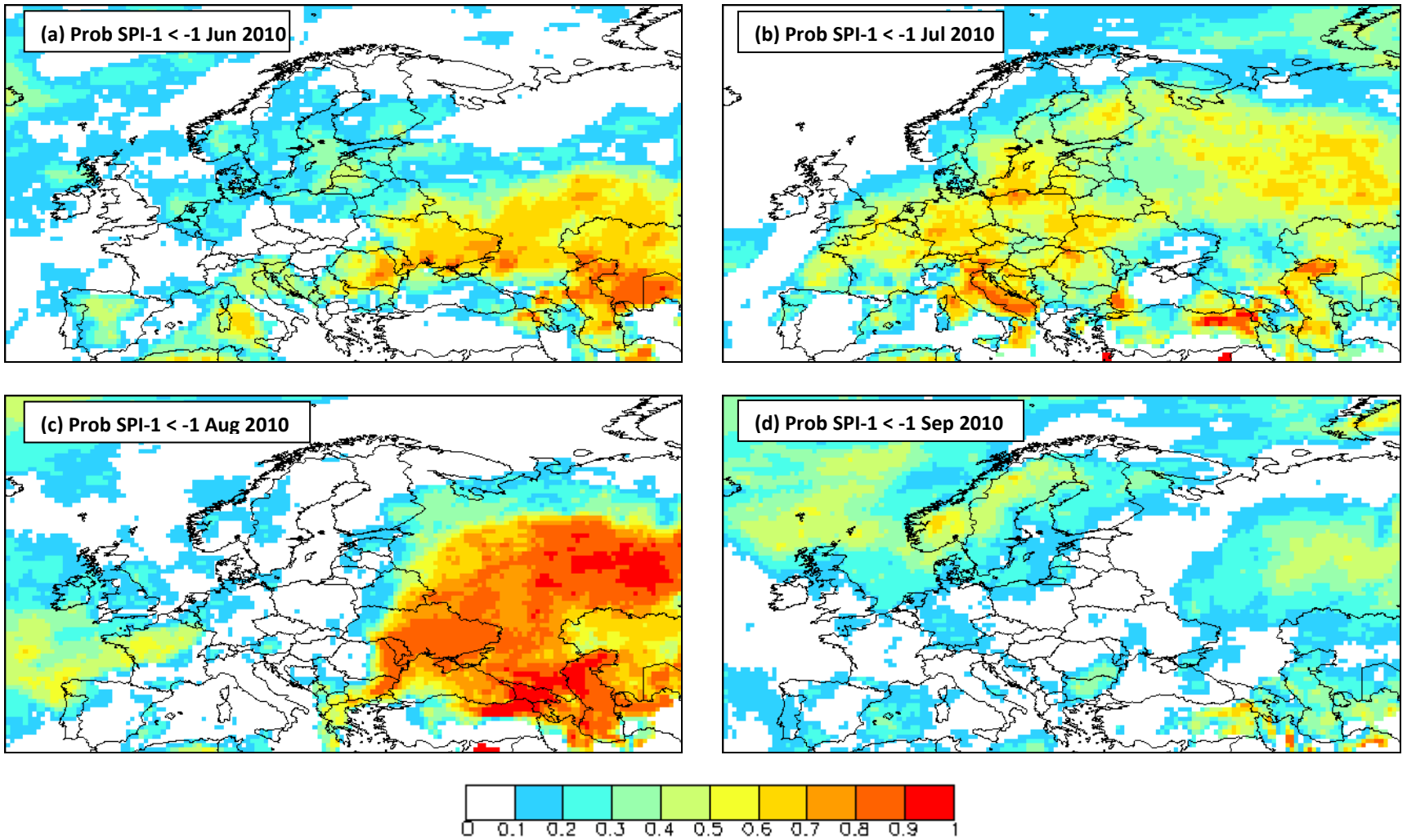


Figure 5.2 – Uncalibrated forecast probability SPI-1 < -1 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

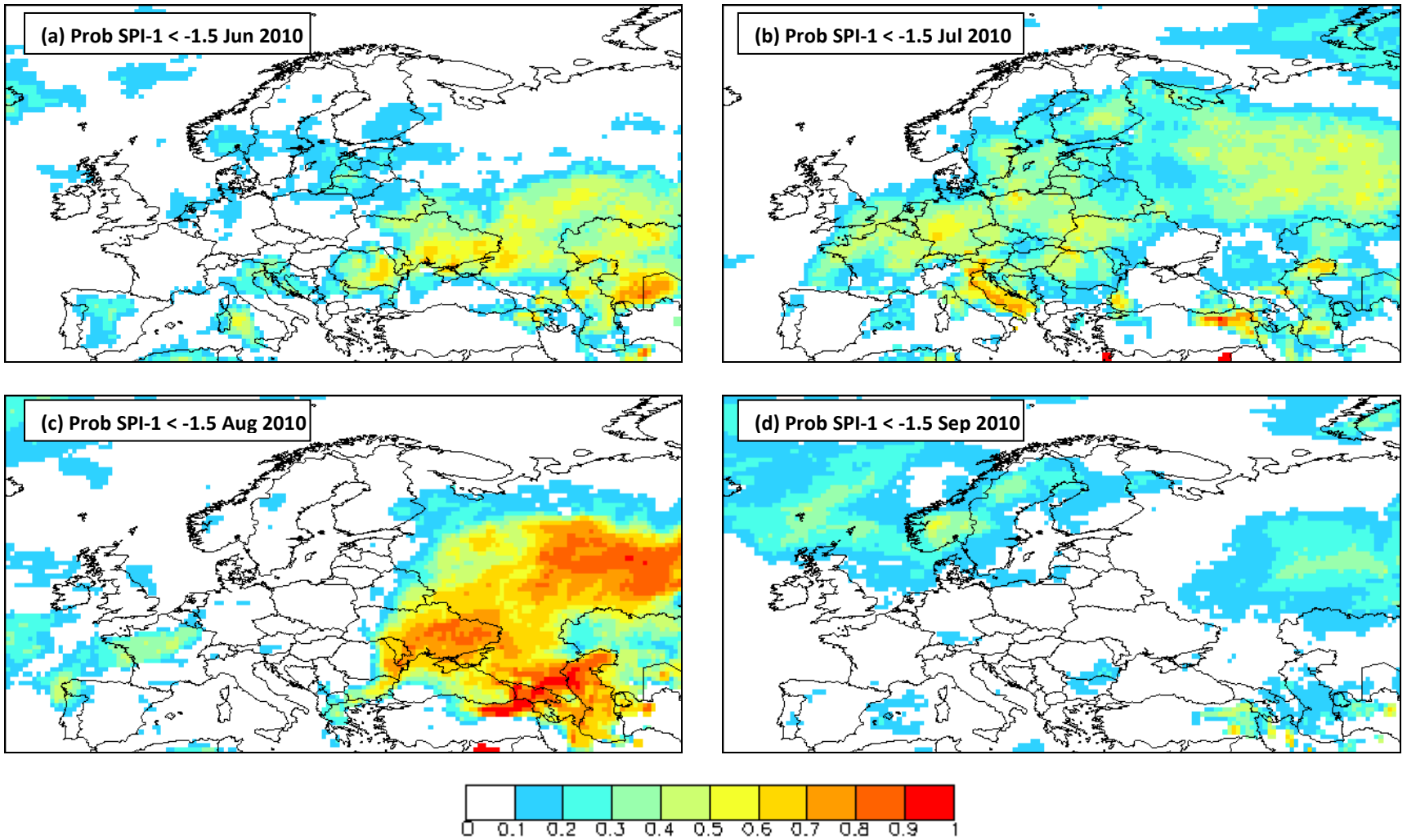


Figure 5.3 – Uncalibrated forecast probability SPI-1 < -1.5 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

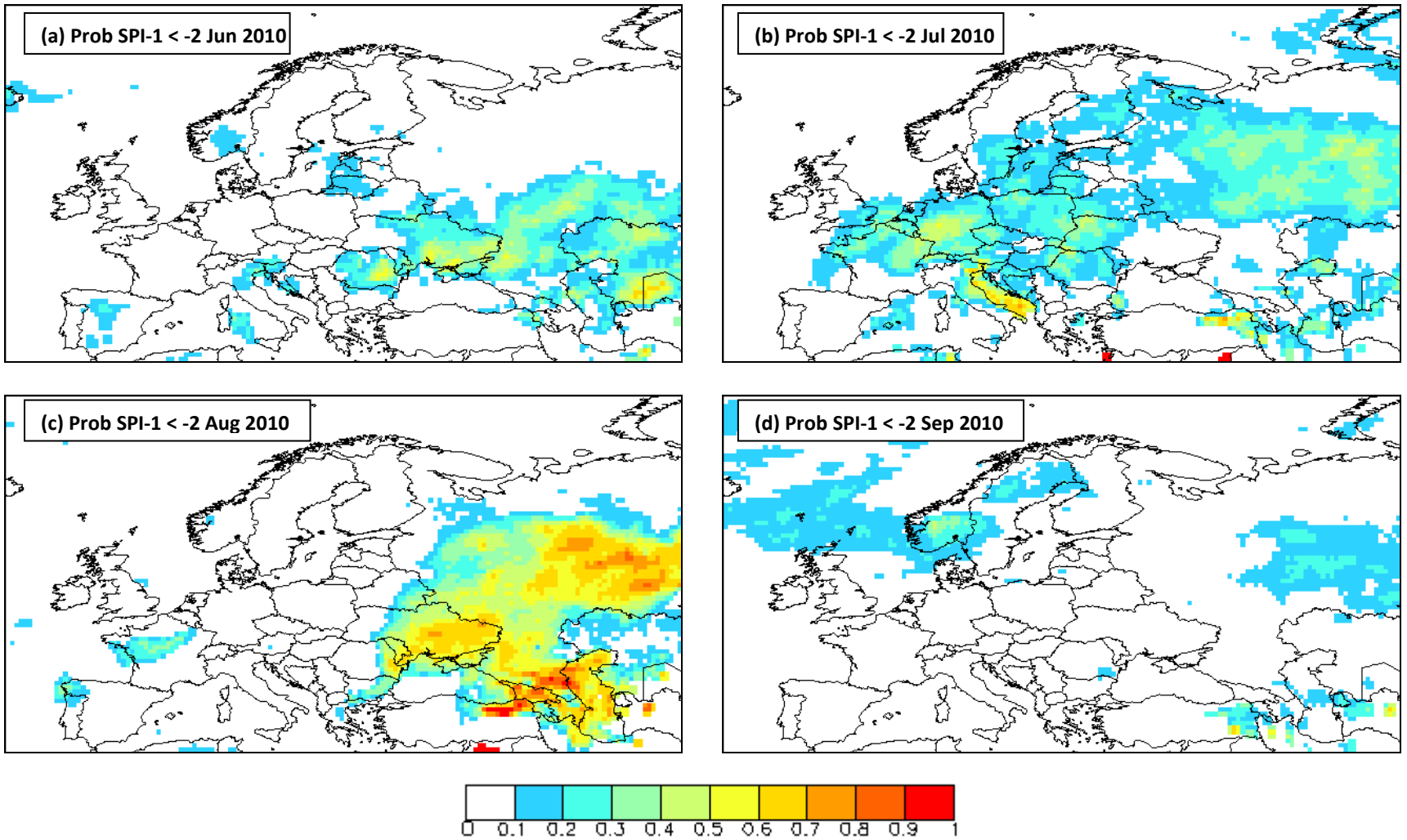


Figure 5.4 – Uncalibrated forecast probability SPI-1 < -2 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

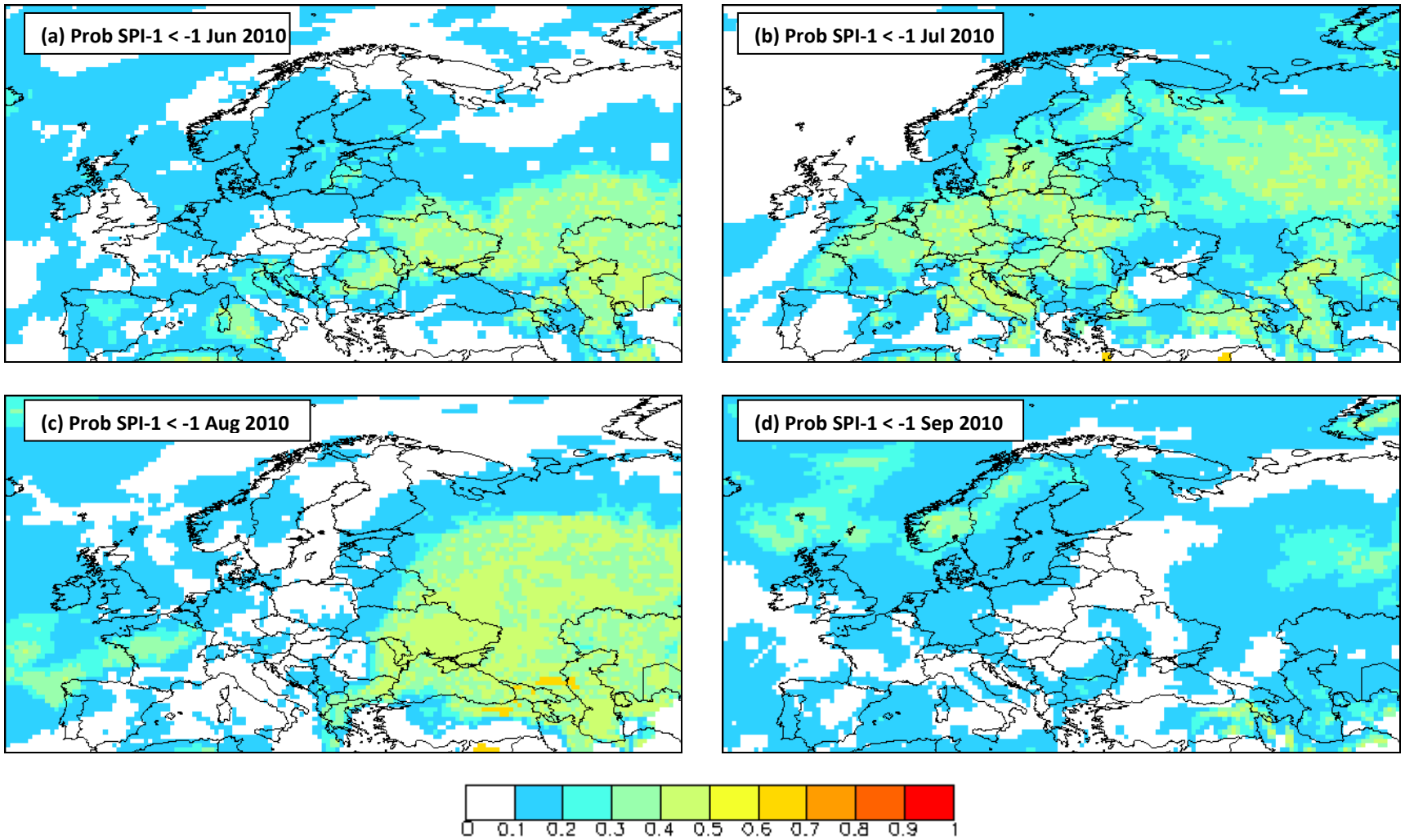


Figure 5.5 – Calibrated forecast probability SPI-1 < -1 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

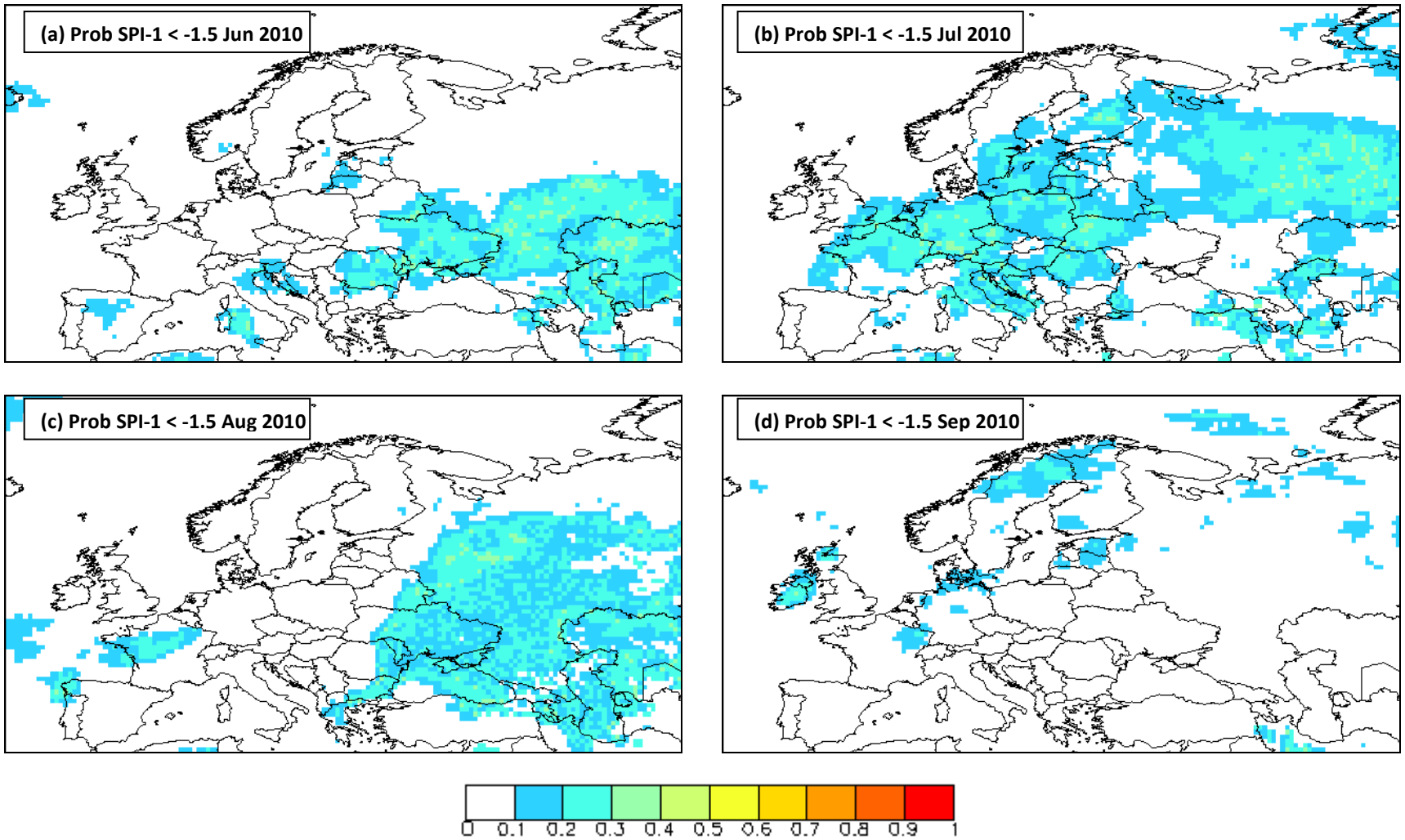


Figure 5.6 – Calibrated forecast probability SPI-1 < -1.5 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

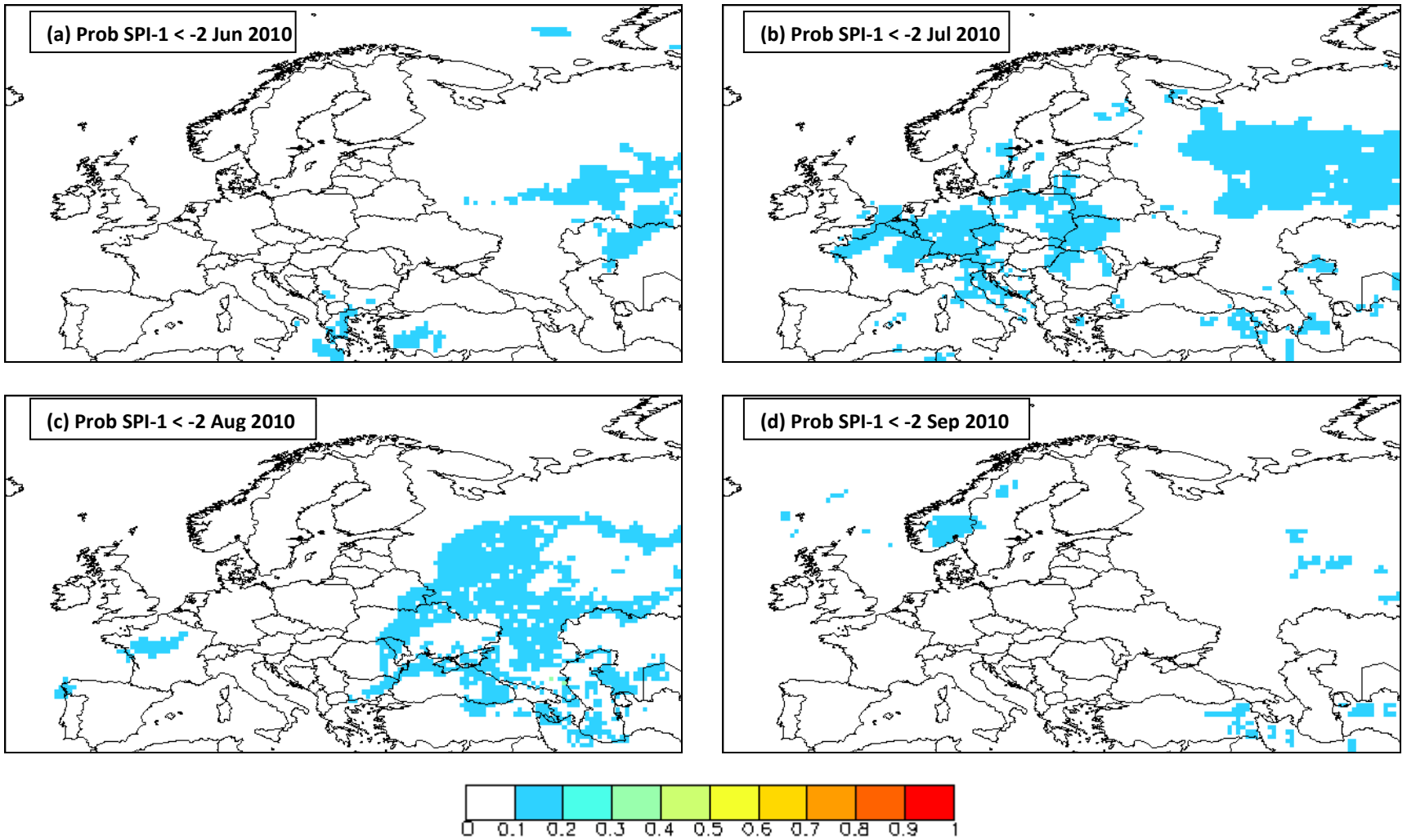


Figure 5.7 – Calibrated forecast probability SPI-1 < -2 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

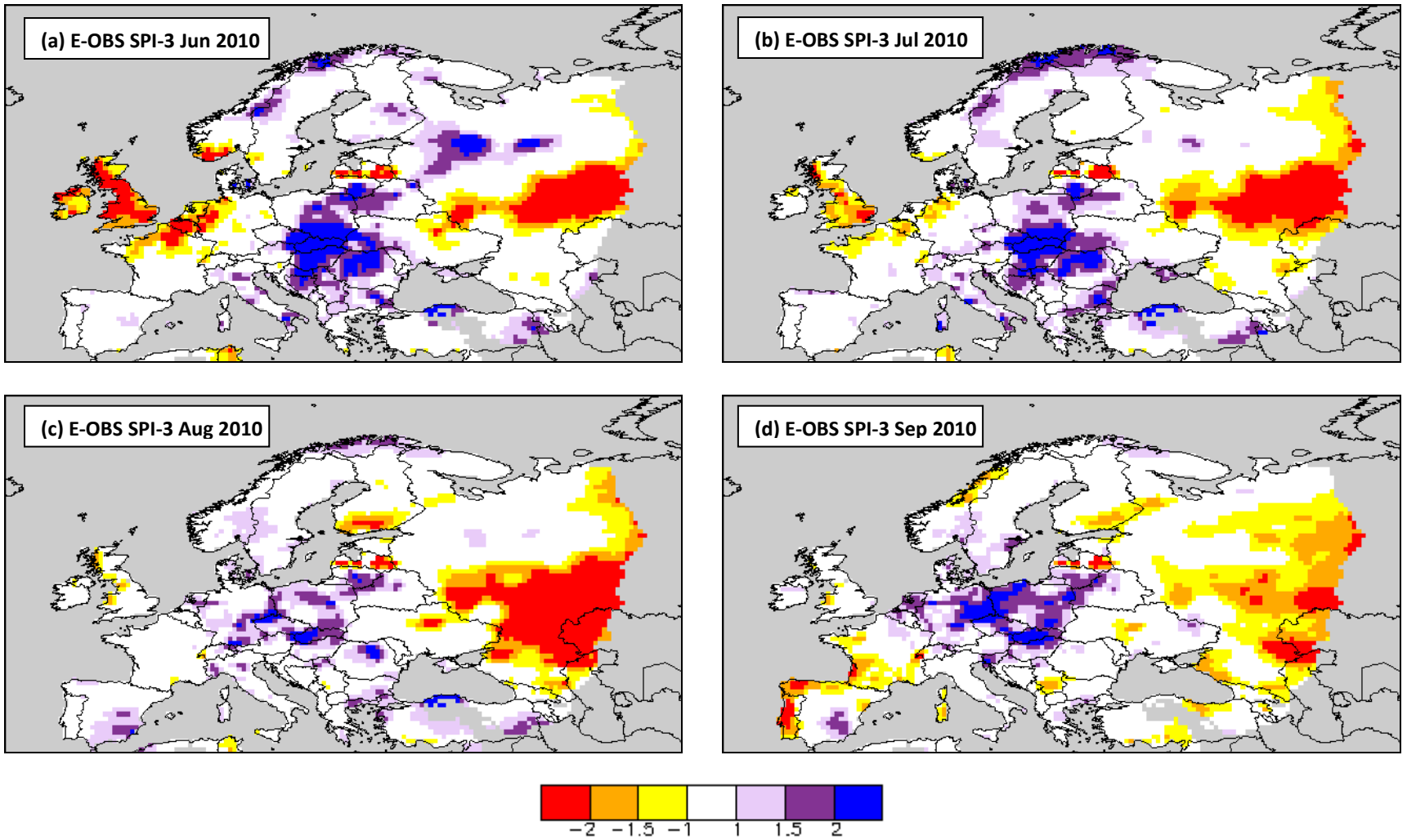


Figure 5.8 – Observed SPI-3 (grey areas indicate missing data) for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

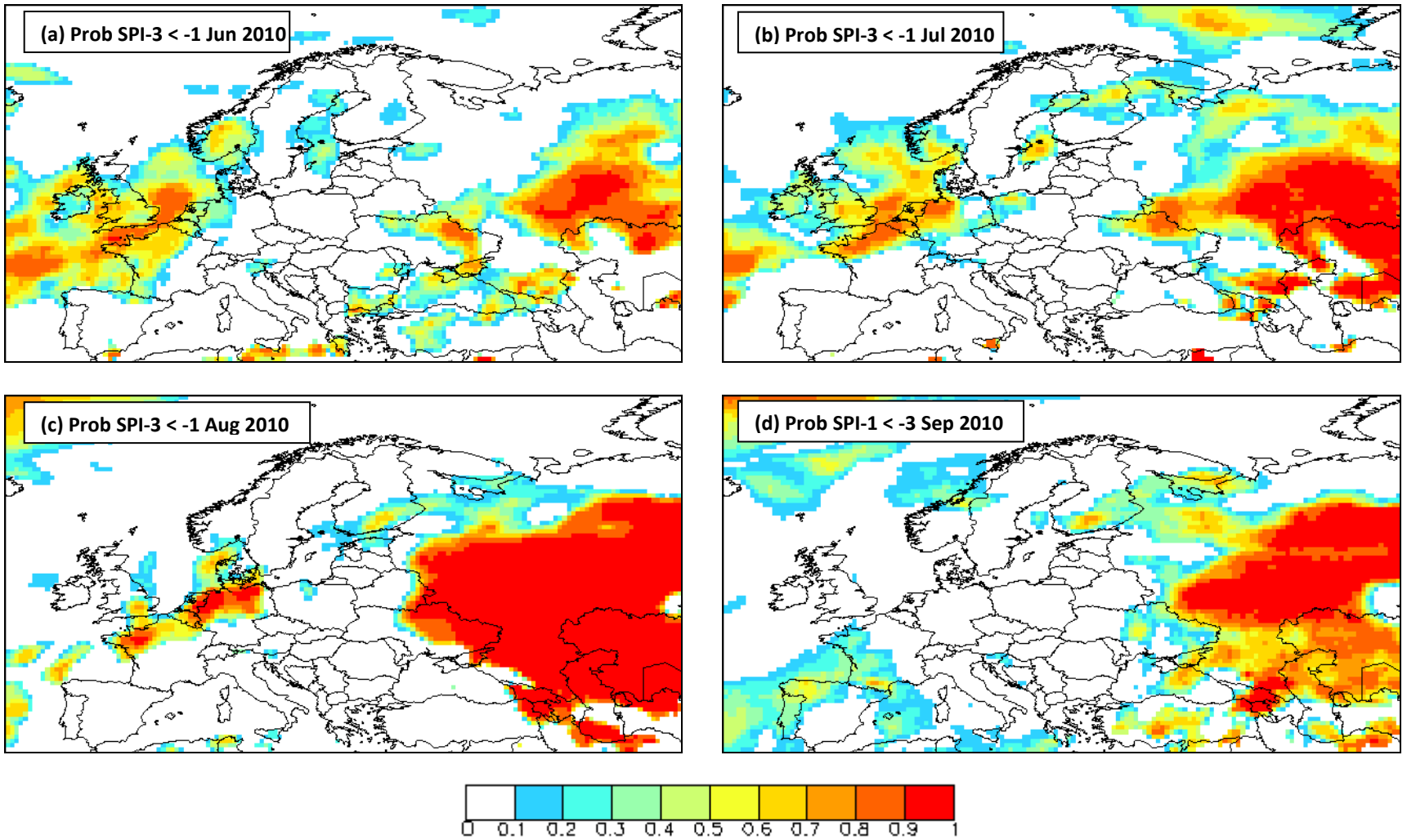


Figure 5.9 – Uncalibrated forecast probability SPI-3 < -1 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

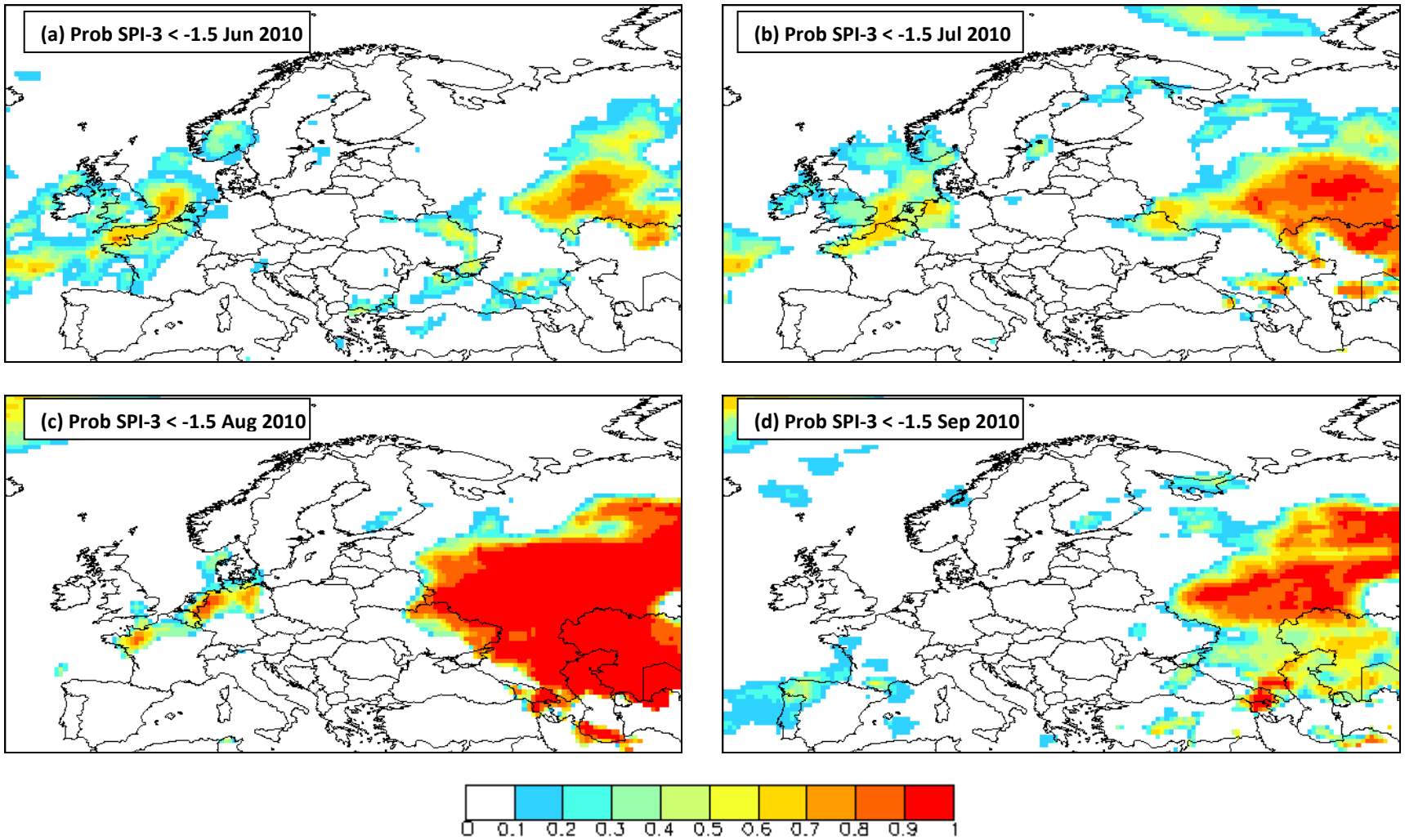


Figure 5.10 – Uncalibrated forecast probability SPI-3 < -1.5 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

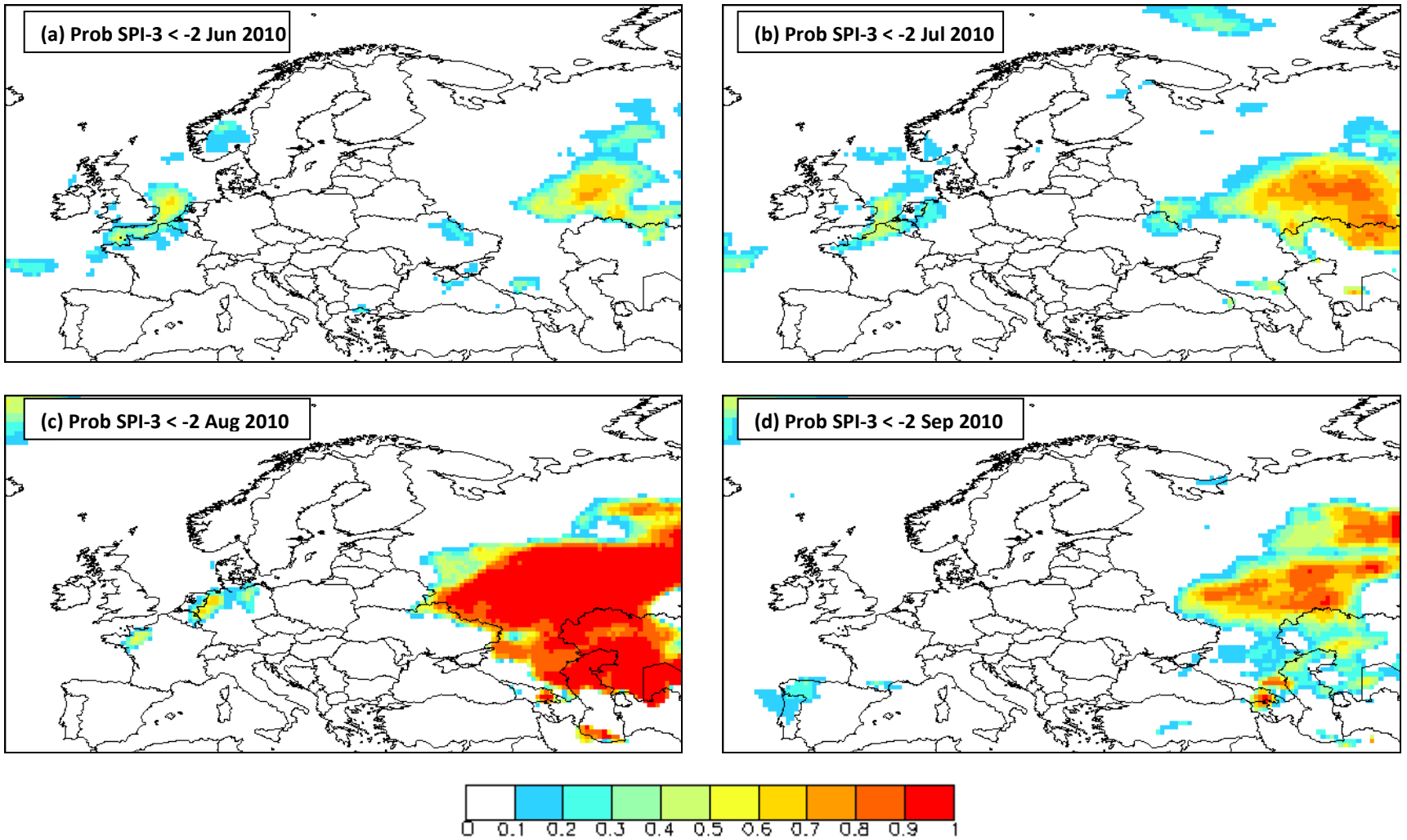


Figure 5.11 – Uncalibrated forecast probability SPI-3 < -2 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

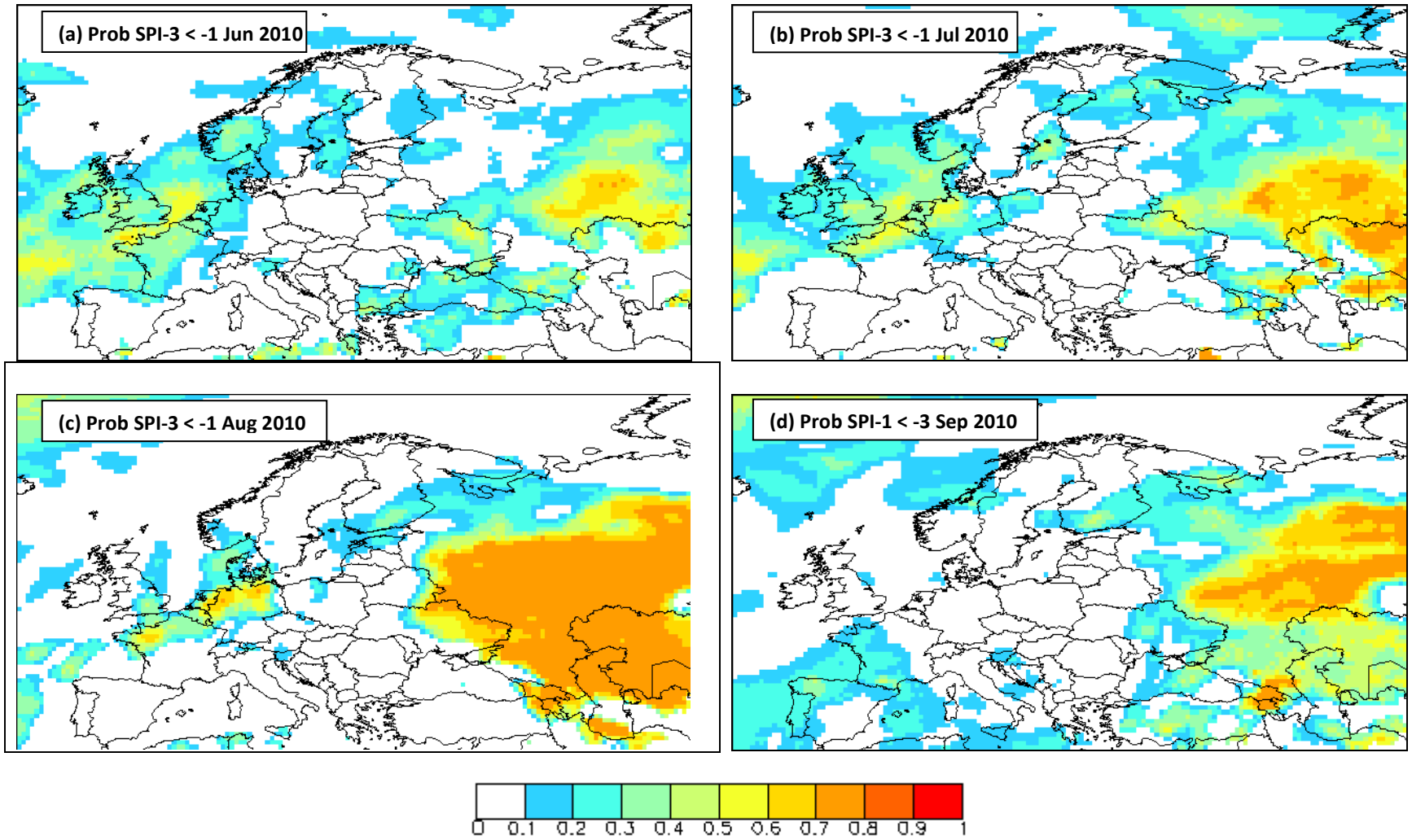


Figure 5.12– Calibrated forecast probability SPI-3 < -1 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

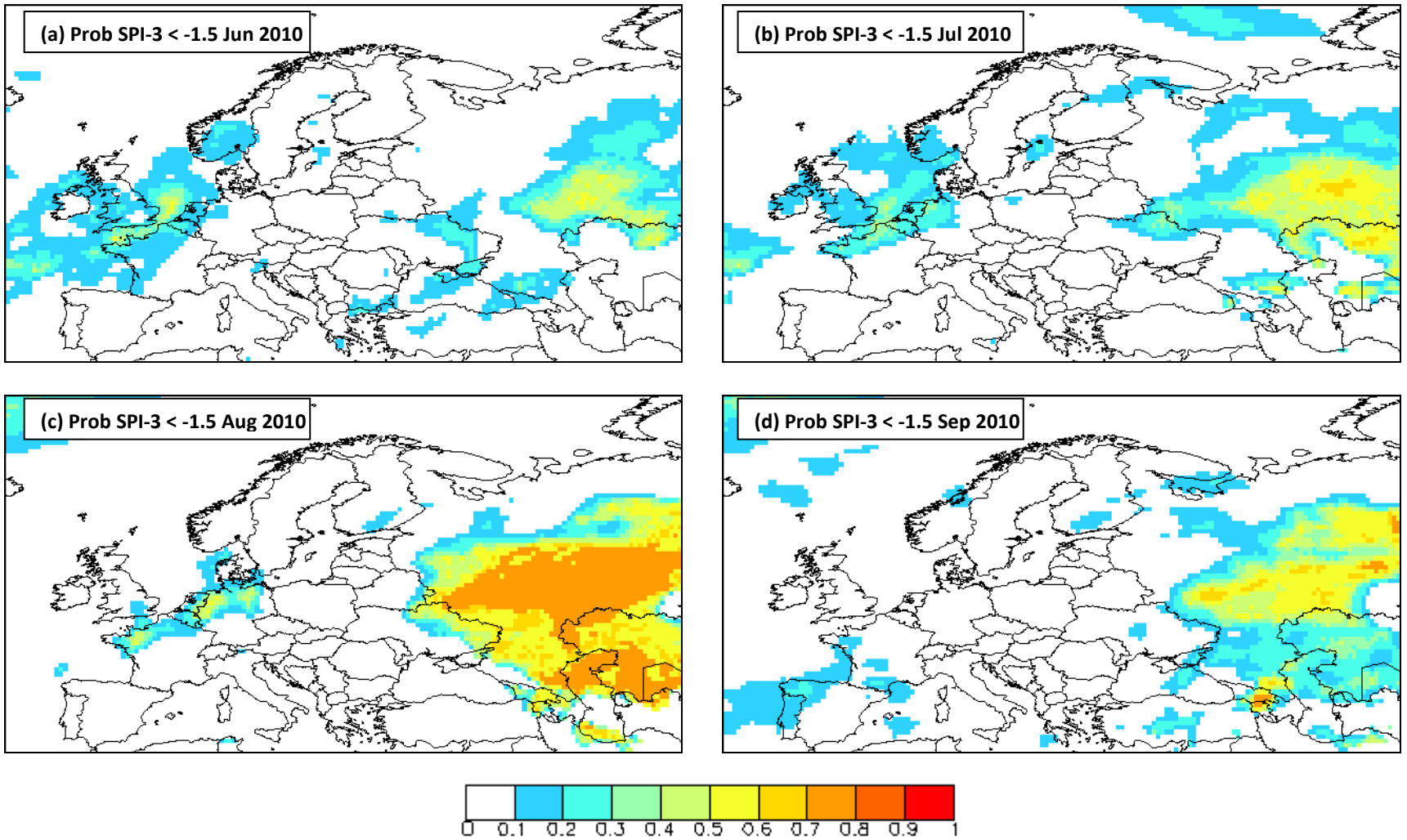


Figure 5.13 – Calibrated forecast probability SPI-3 < -1.5 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

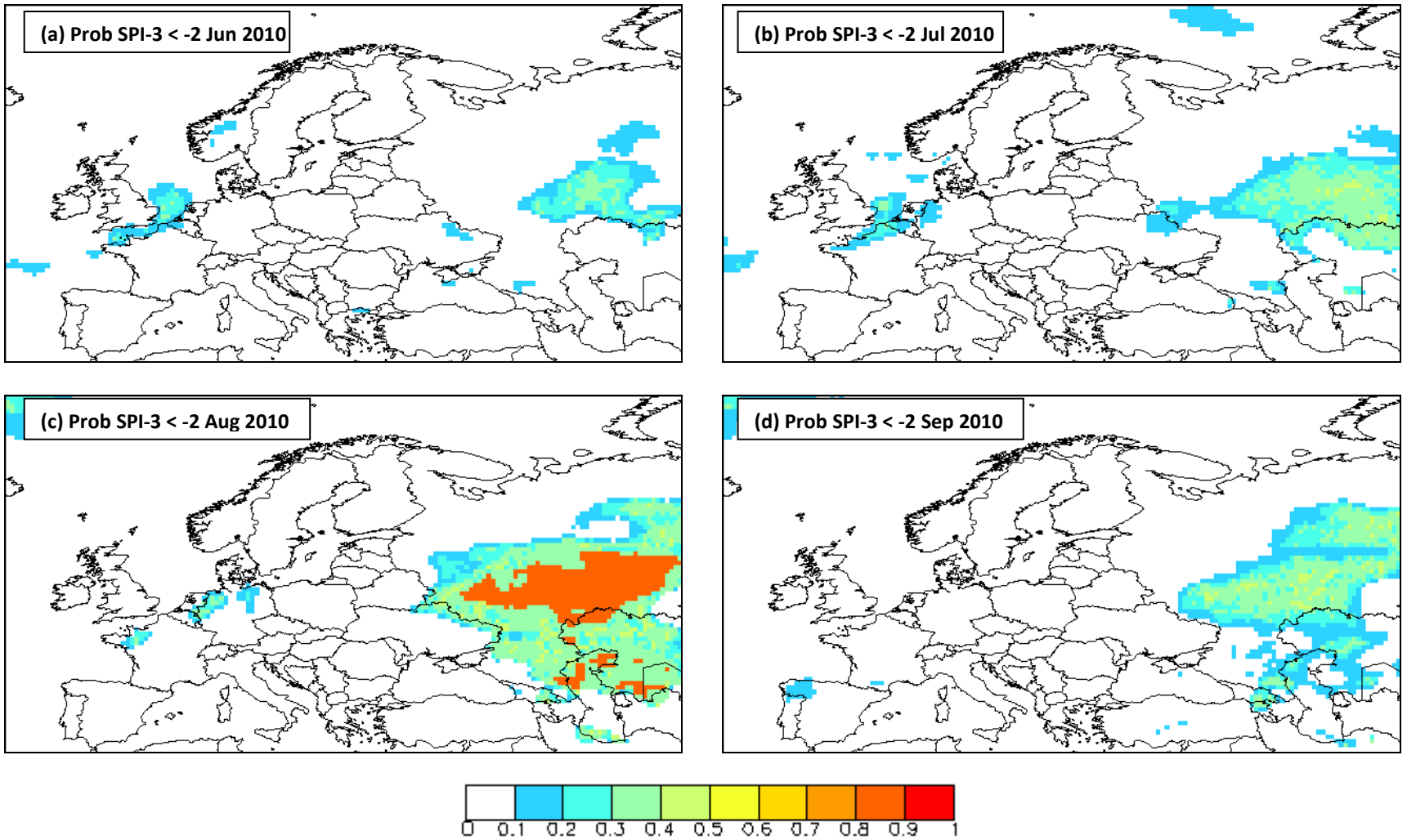


Figure 5.14 – Calibrated forecast probability SPI-3 < -2 for (a) June 2010, (b) July 2010, (c) August 2010, (d) September 2010.

5.2. Spring 2011 – north-western Europe

From March to May 2011, north-western Europe experienced extreme dry conditions. The most affected countries were England, France, Belgium, The Netherlands, Germany and northern Italy. The SPI-1 derived from E-OBS shows extreme drought conditions over southern England in March 2011 (Fig. 5.15(a)) that persisted into April 2011 (Fig. 5.15(b)). North-western Germany, the Ukraine, Belarus and the Baltic countries were similarly affected in March 2011 (Fig. 5.15(a)) with northern France and northern Italy becoming affected in April 2011 (Fig. 515(b)). By May 2011, conditions had returned to near normal in England, but much of northern and western France had become affected by the extreme dry conditions (Fig. 5.15(c)). By June 2011, most of western Europe had returned to near normal conditions (Fig. 5.15(d)).

The forecast for SPI-1 suggested relatively high probabilities of widespread dry conditions over Europe in March 2011 mostly to the east of France (Fig. 5.16(a), 5.17(a), 5.18(a)). For April 2011, the forecast suggested the most extreme dry conditions would be over southern Europe, particularly over an area stretching from the Mediterranean to the Black Sea (Fig. 5.16(b), 5.17(b), 5.18(b)) with zero probability of severe or extreme dry conditions over northern parts of Europe (Fig. 5.17(b), 5.18(b)) where these conditions occurred (Fig. 5.15(b)). For May 2011, lower probabilities of dry SPI-1 were forecast over western Europe (Fig. 5.16(c), 5.17(c), 5.18(c)), with no indication of severe or extreme SPI-1 drought over western France (Fig. 5.17(c), 5.18(c)). The forecast for June 2011 suggested relatively high probabilities of severe or extreme dry conditions over eastern Europe (Fig. 5.16(d), 5.17(d), 5.18(d)), which were not observed except over Latvia (Fig. 5.15(d)), and in fact extreme wet conditions were observed over the Ukraine (Fig. 5.15(d)). Calibration did not improve the forecast for SPI-1, but only reduced the probabilities (Fig. 5.19, 5.20, 5.21).

The observations for SPI-3 (Fig. 5.22) were similar to those for SPI-1 (Fig. 5.15) except that they showed the extreme conditions to start in April 2011 (Fig. 5.22(b)) and to be more widespread in May 2011 (Fig. 5.22(c)) and to persist in France into June 2011 (Fig. 5.22(d)). The forecast for SPI-3 showed relatively high probabilities of widespread SPI-3 drought from France in the west stretching east across Russia in Mar 2011 (Fig. 5.23 (a)). Highest probabilities of SPI-3 <-1 were forecast over southern Germany, Austria, the Czech Republic, Slovakia and Hungary (Fig. 5.23(a)) where severe (Fig. 5.24(a)) and extreme (Fig. 5.25(a)) were predicted with lower probabilities. For April 2011, the highest probabilities of dry SPI-3 were forecast over eastern France and Germany (Fig. 5.23(b), 5.24(b), 5.25(b)) becoming more widespread in May 2011 (Fig. 5.23(c), 5.24(c), 5.25(c)) with the most extreme conditions forecast with the highest probability over western central Germany (Fig. 5.25(c)). The June 2011 forecast suggested the dry SPI-3 to be less widespread (Fig. 5.23(d), 5.24(d), 5.25(d)) with the most extreme SPI-3 forecast with the highest probability to be over northern France adjacent to Belgium (Fig. 5.25(d)). The SPI-3 forecast was not able to predict the extreme dry SPI-3 observed over southern

England and western France (Fig. 5.22) with a high level of probability (Fig. 5.24, 5.25).

The effect of calibration on the forecast was to reduce the probabilities of passing each SPI-3 threshold everywhere (Fig. 5.26, 5.27, 5.28) without improving the spatial characteristics of the forecasts.

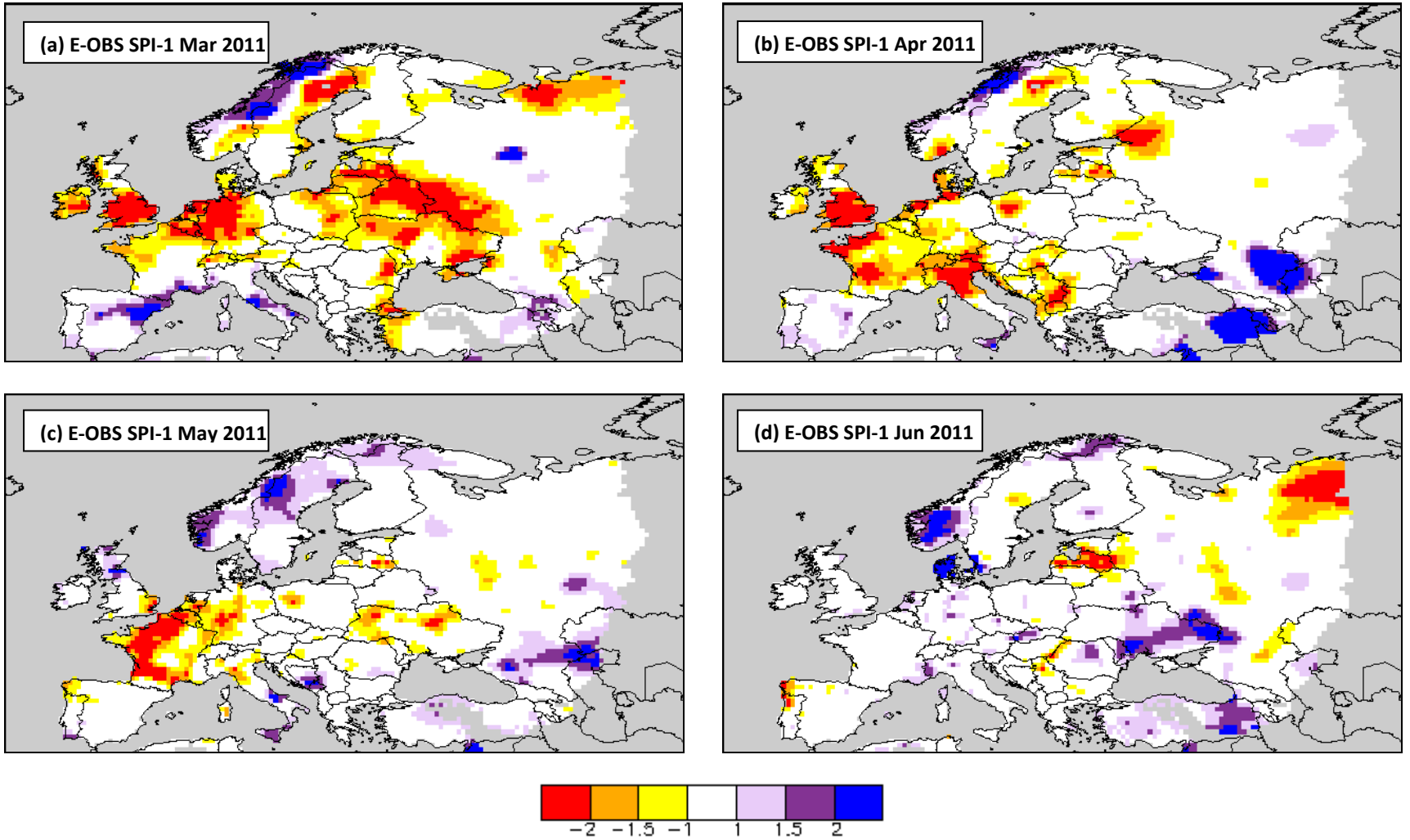


Figure 5.15 – Observed SPI-1 (grey areas indicate missing data) for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

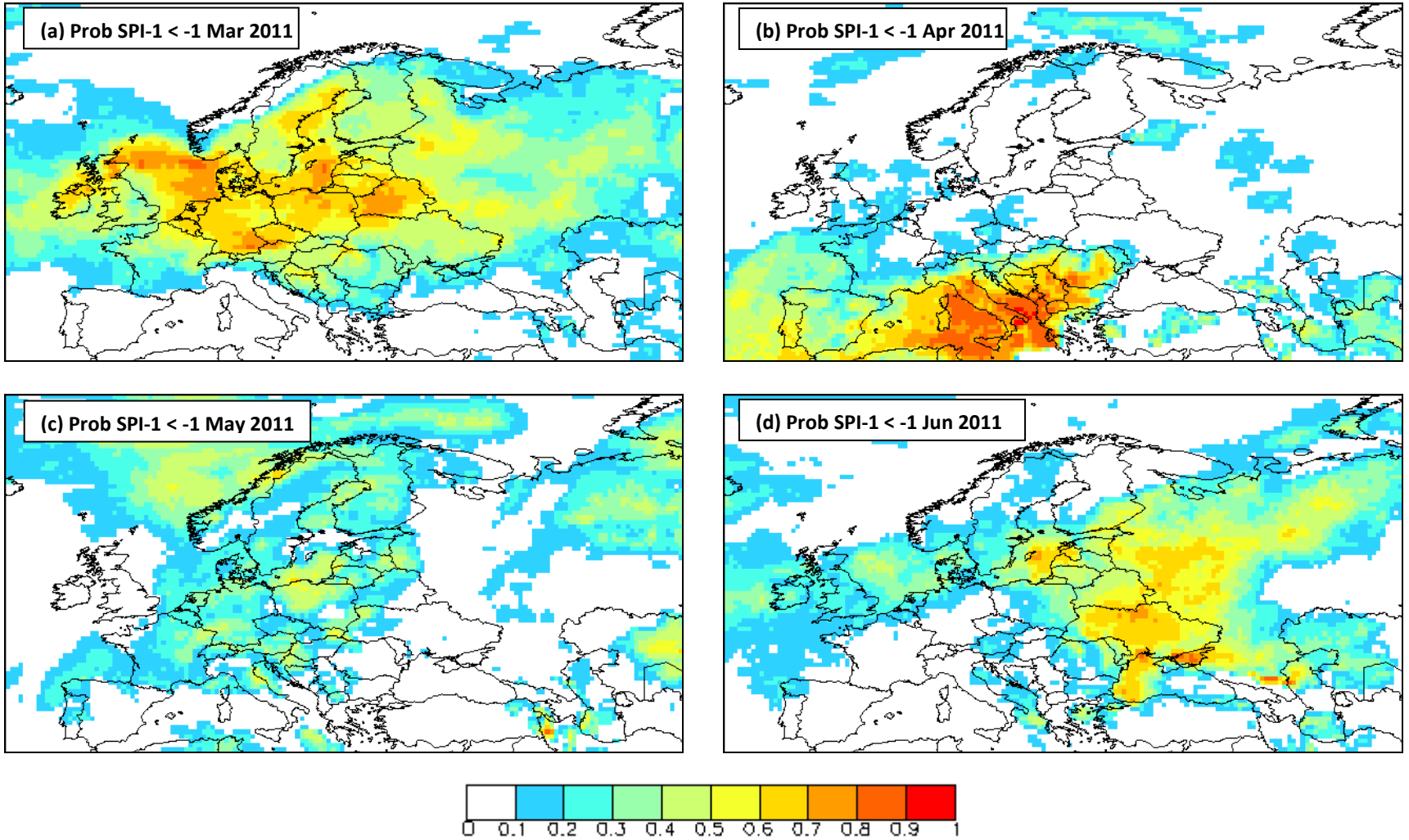


Figure 5.16 – Uncalibrated forecast probability SPI-1 < -1 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

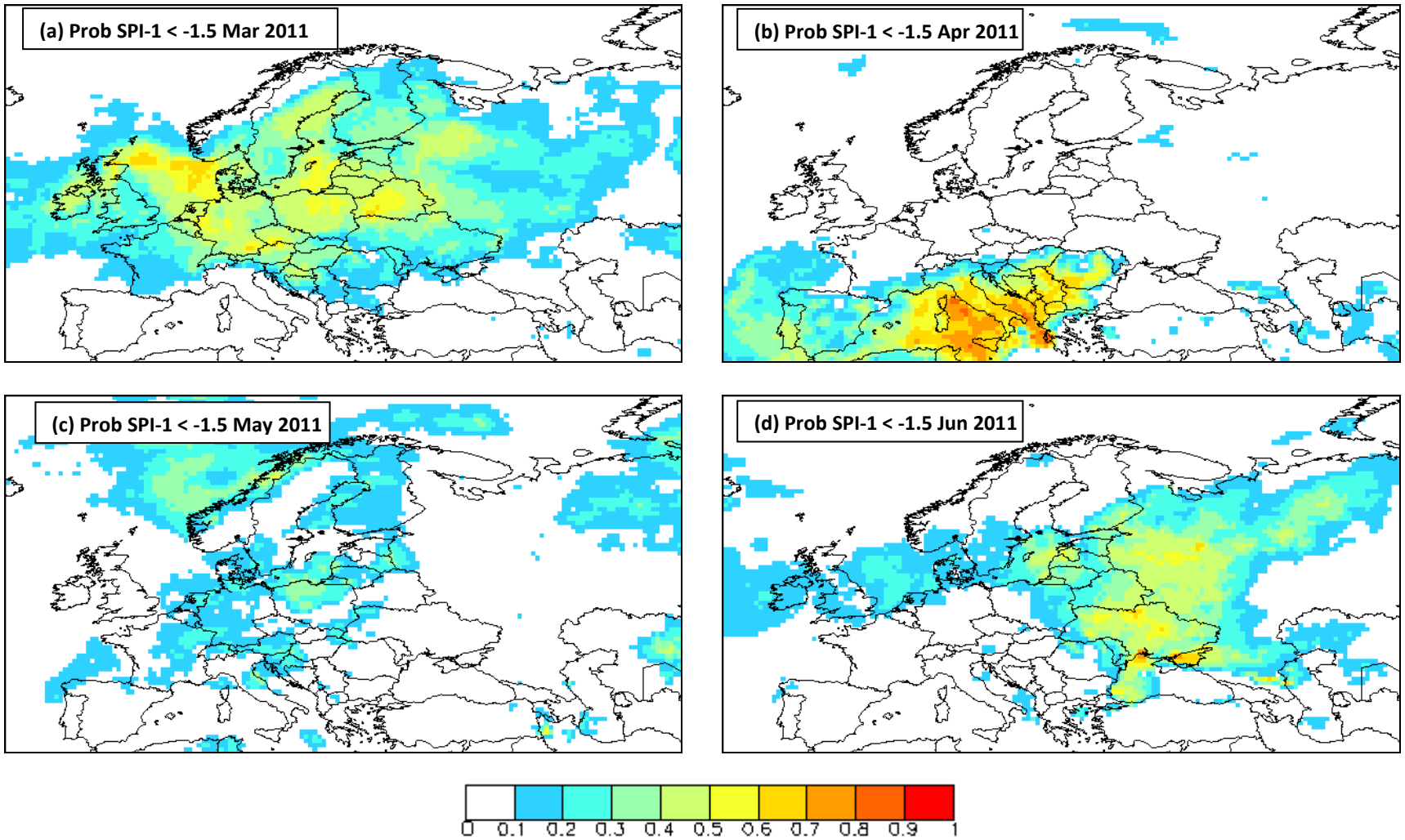


Figure 5.17 – Uncalibrated forecast probability SPI-1 < -1.5 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

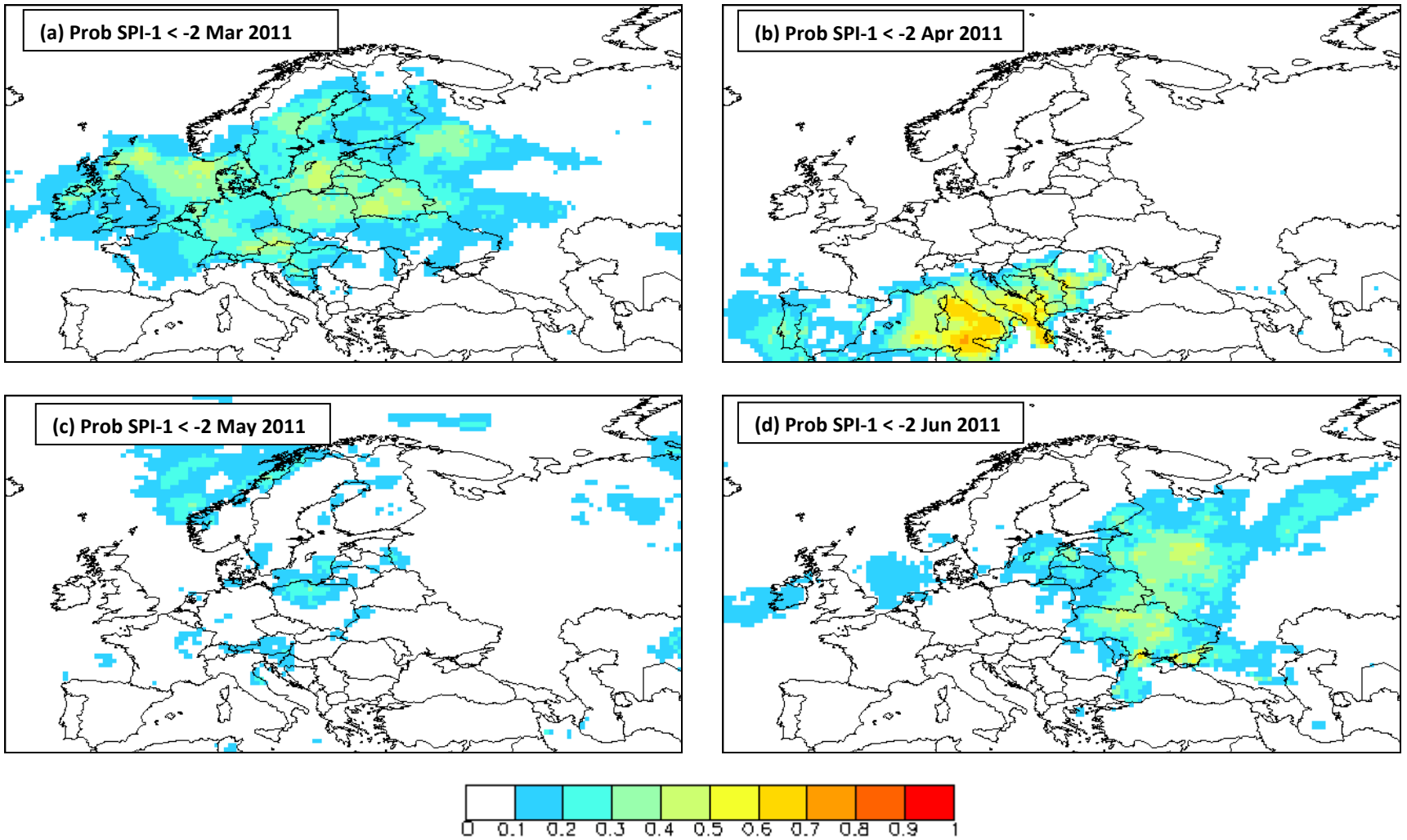


Figure 5.18 – Uncalibrated forecast probability SPI-1 < -2 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

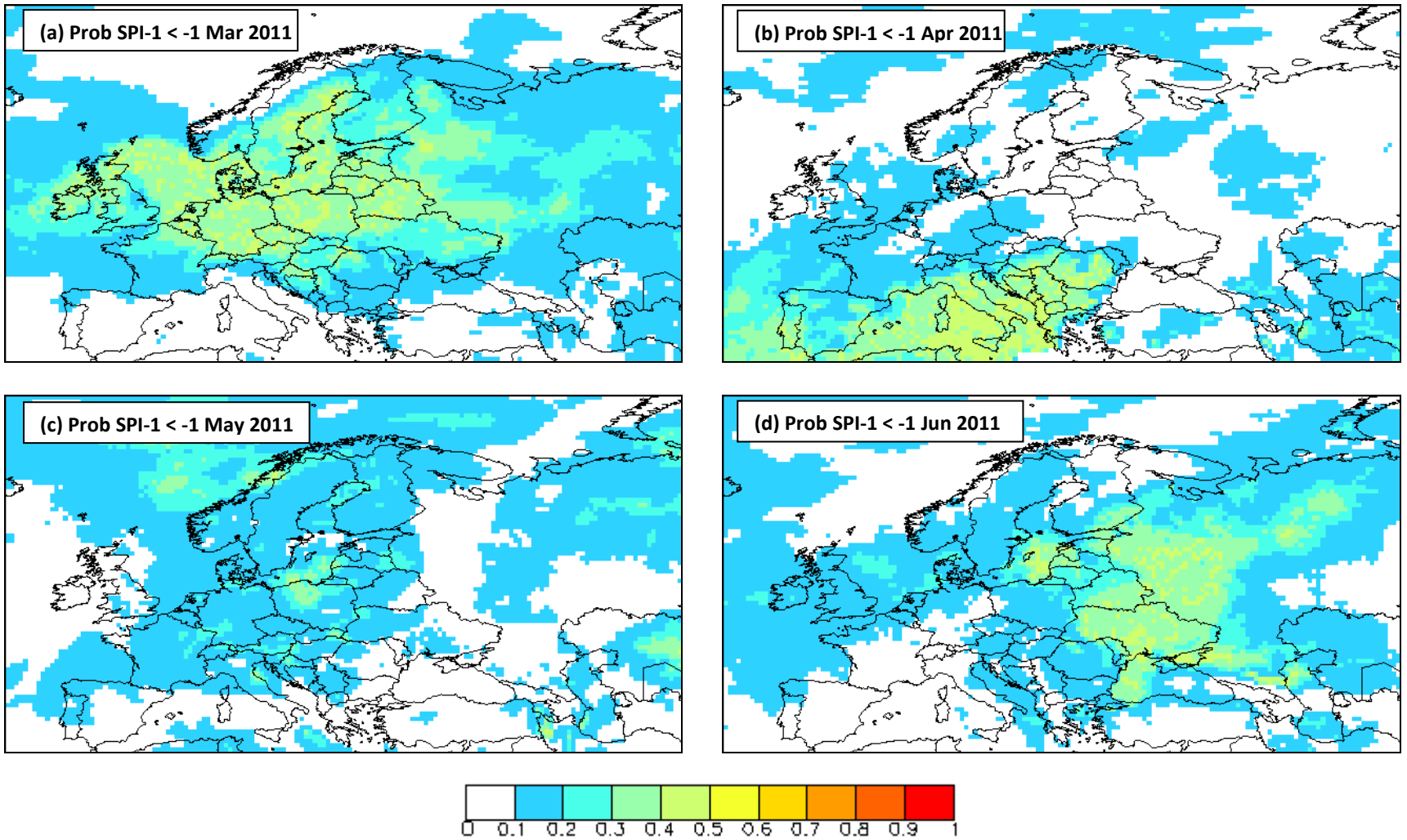


Figure 5.19 – Calibrated forecast probability SPI-1 < -1 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

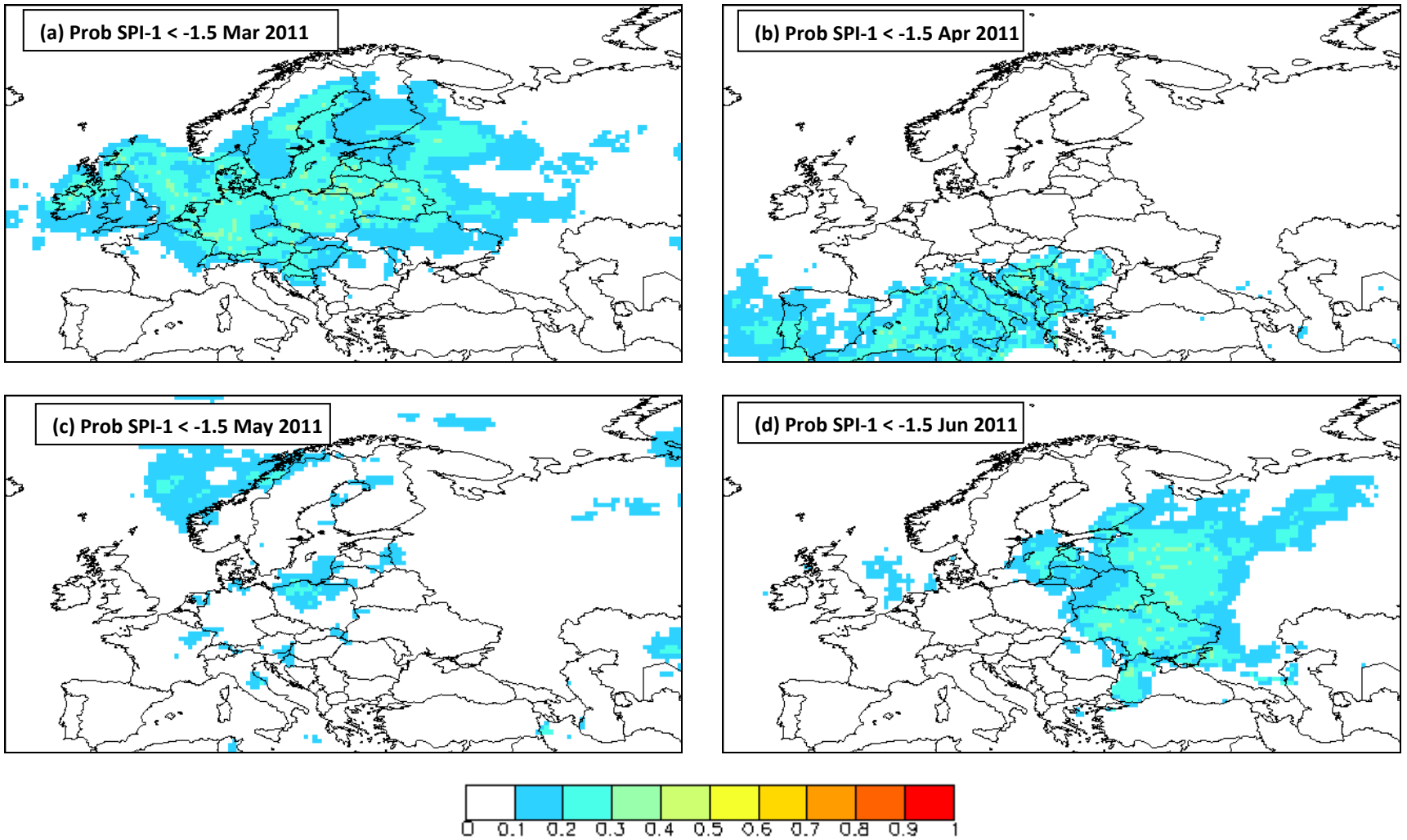


Figure 5.20 – Calibrated forecast probability SPI-1 < -1.5 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

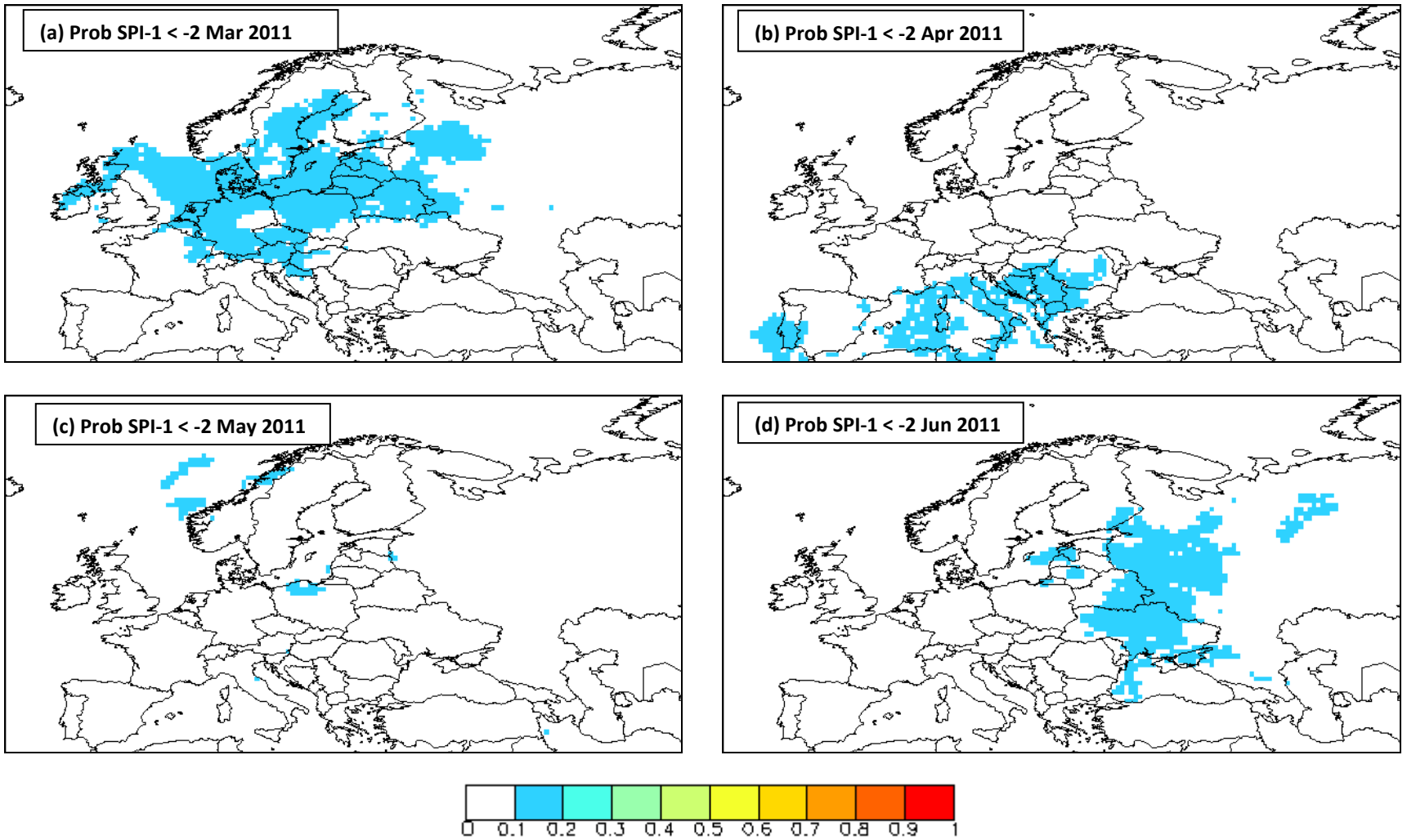


Figure 5.21 – Calibrated forecast probability SPI-1 < -2 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

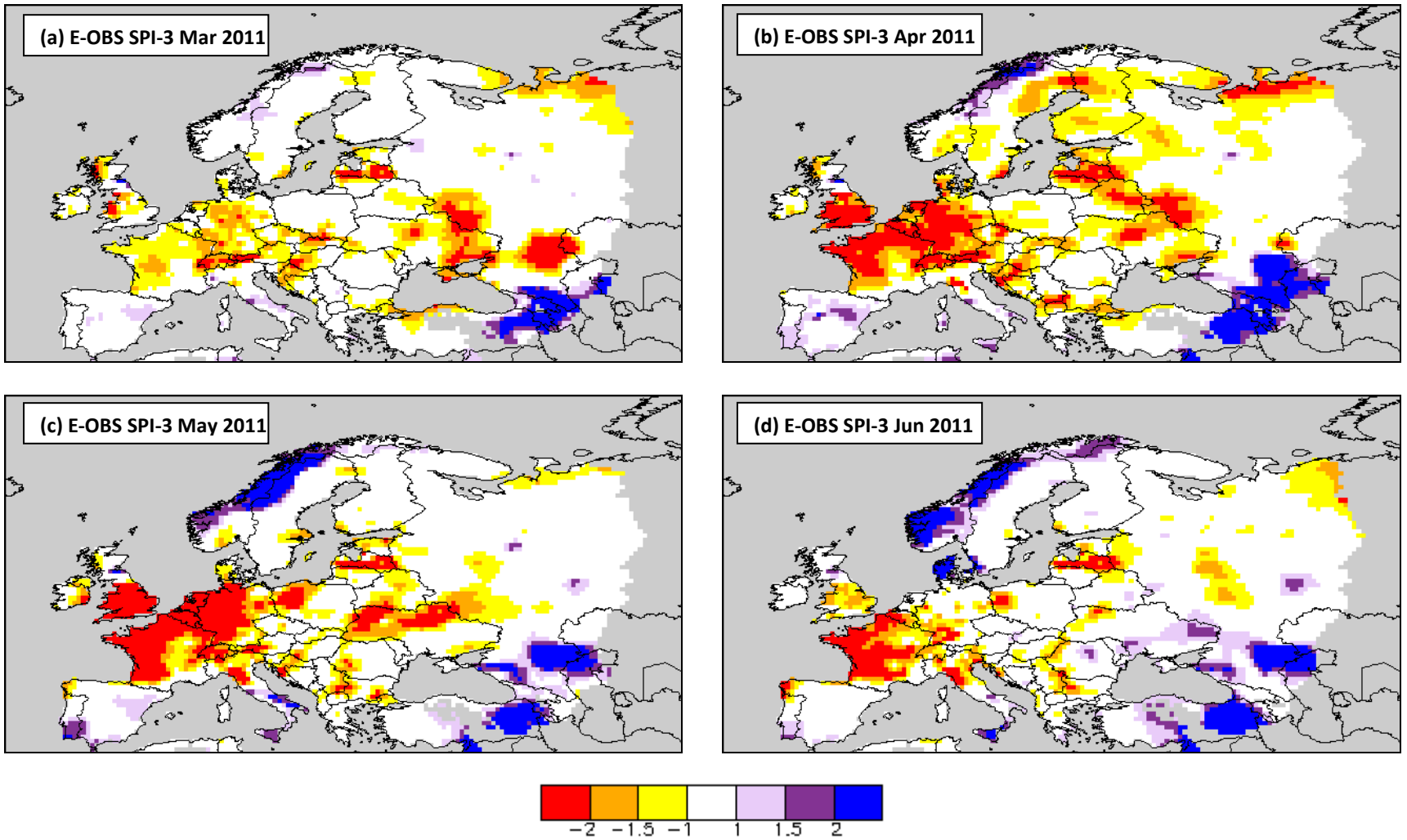


Figure 5.22 – Observed SPI-3 (grey areas indicate missing data) for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

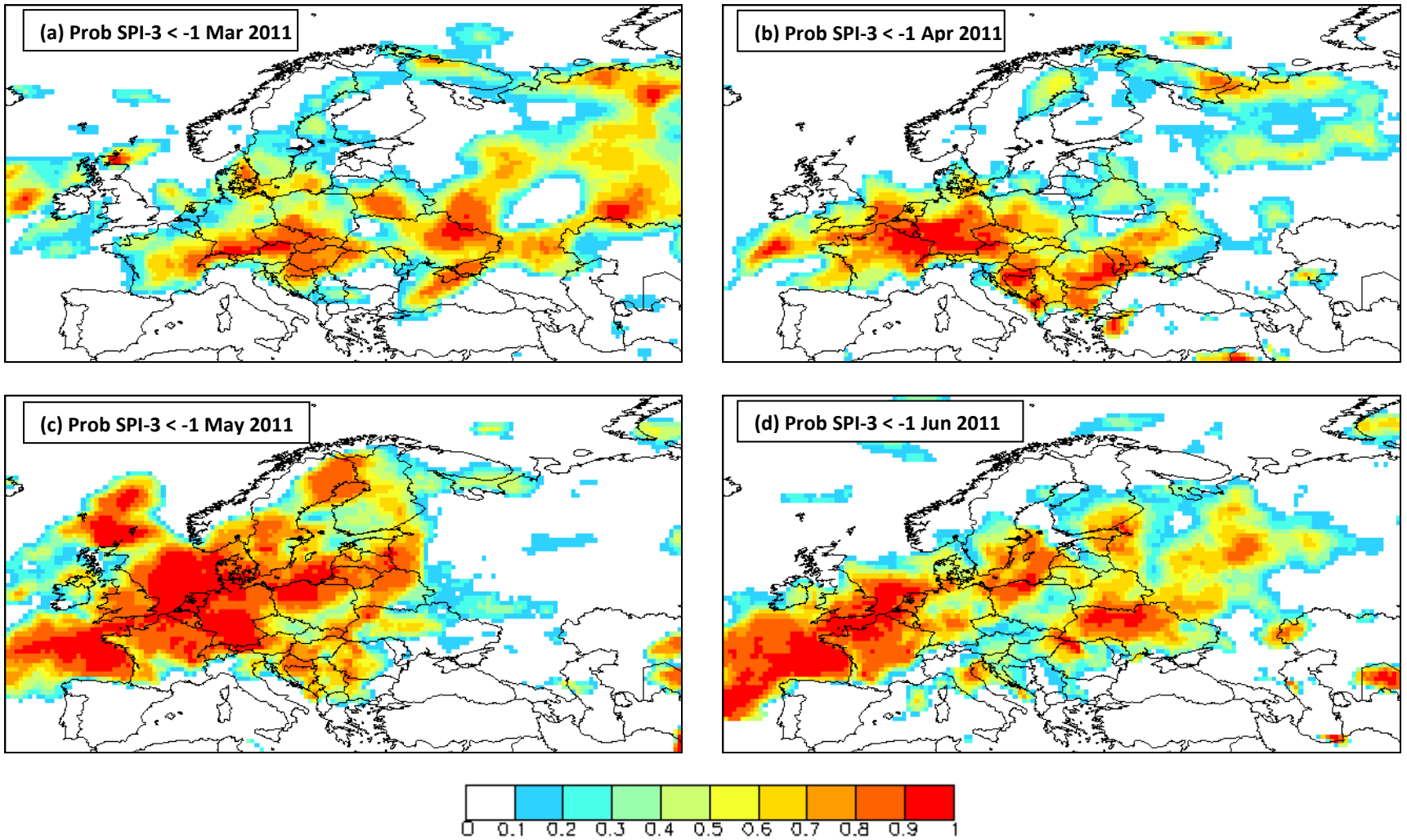


Figure 5.23 – Uncalibrated forecast probability SPI-3 < -1 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

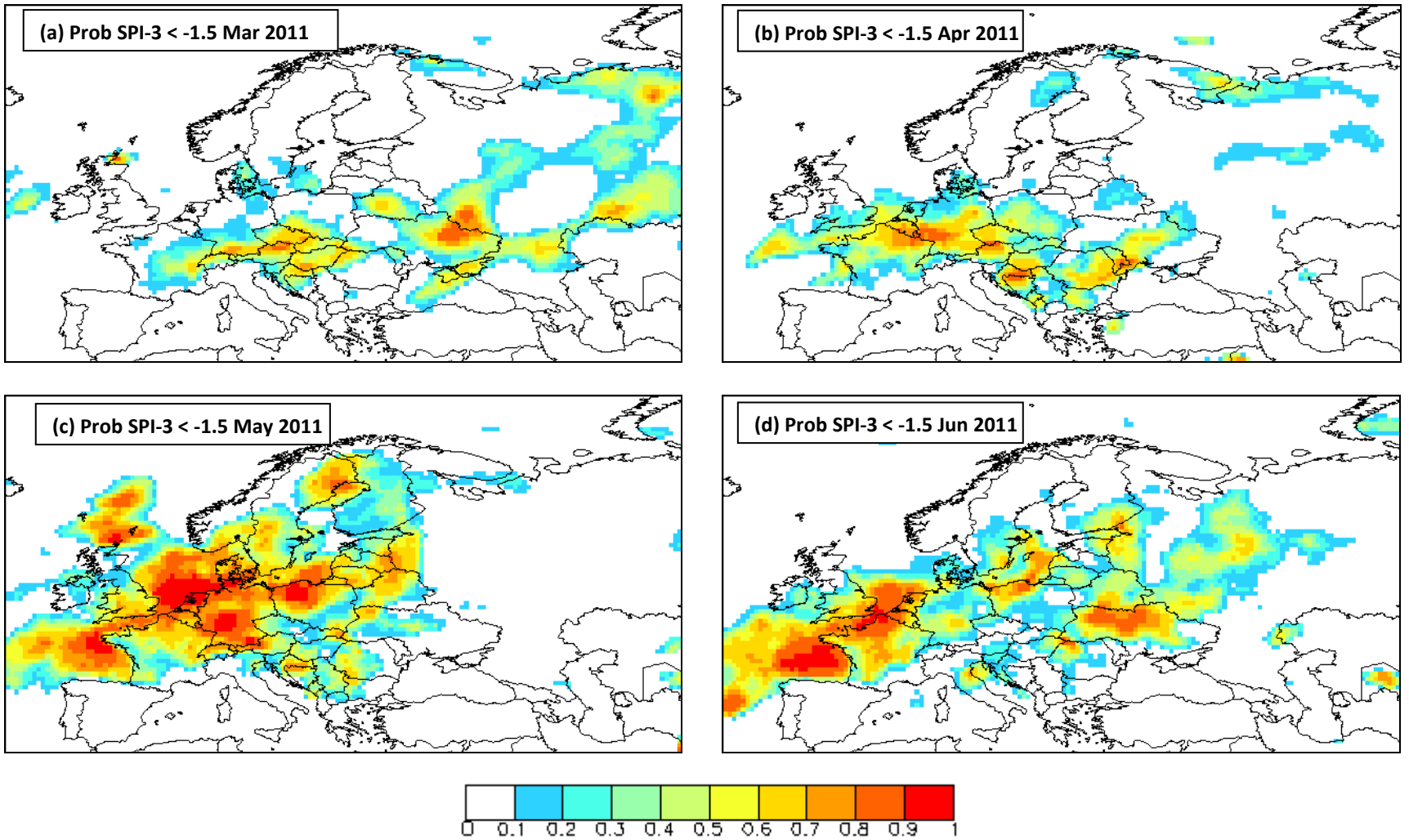


Figure 5.24 – Uncalibrated forecast probability SPI-3 < -1.5 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

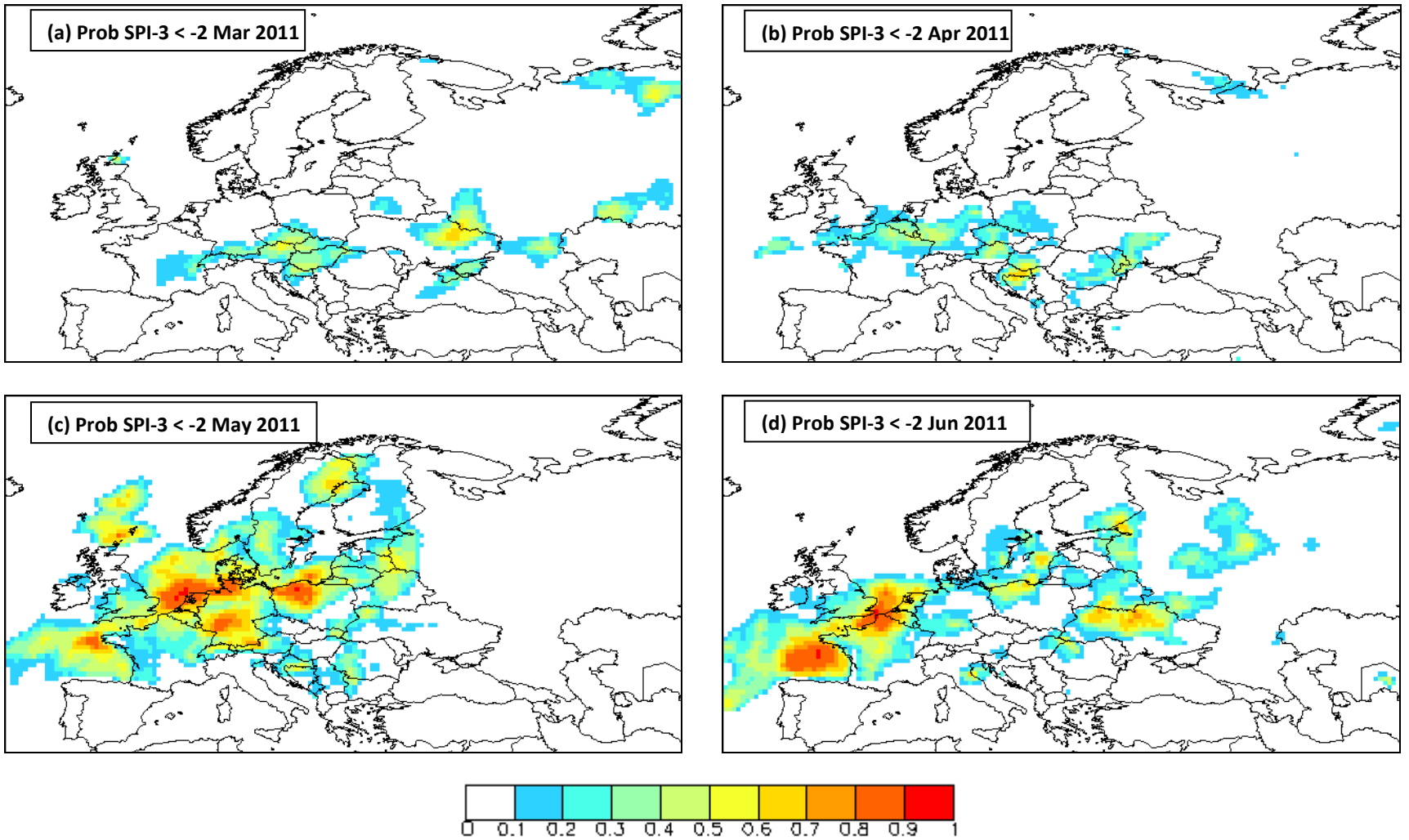


Figure 5.25 – Uncalibrated forecast probability SPI-3 < -2 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

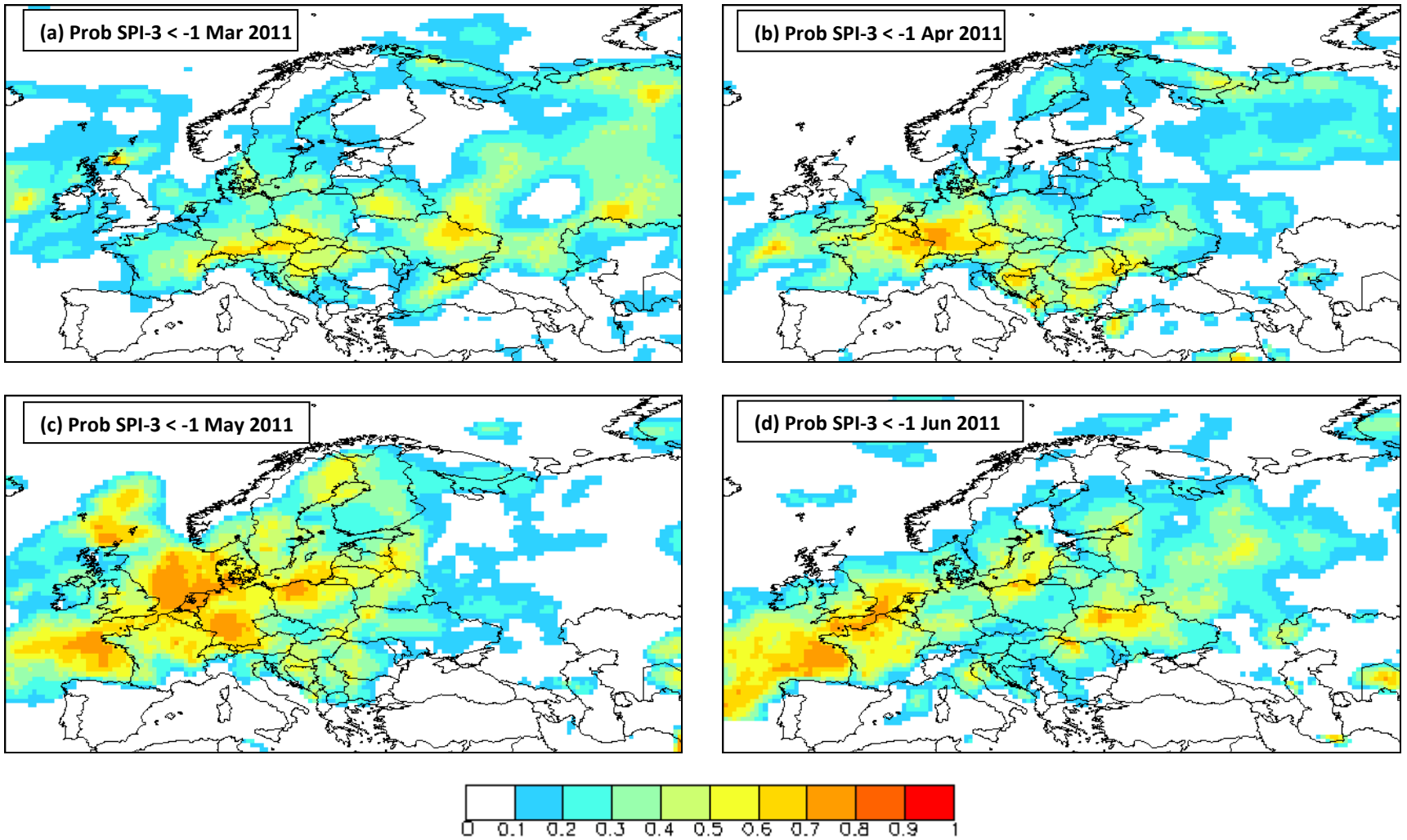


Figure 5.26 – Calibrated forecast probability SPI-3 < -1 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

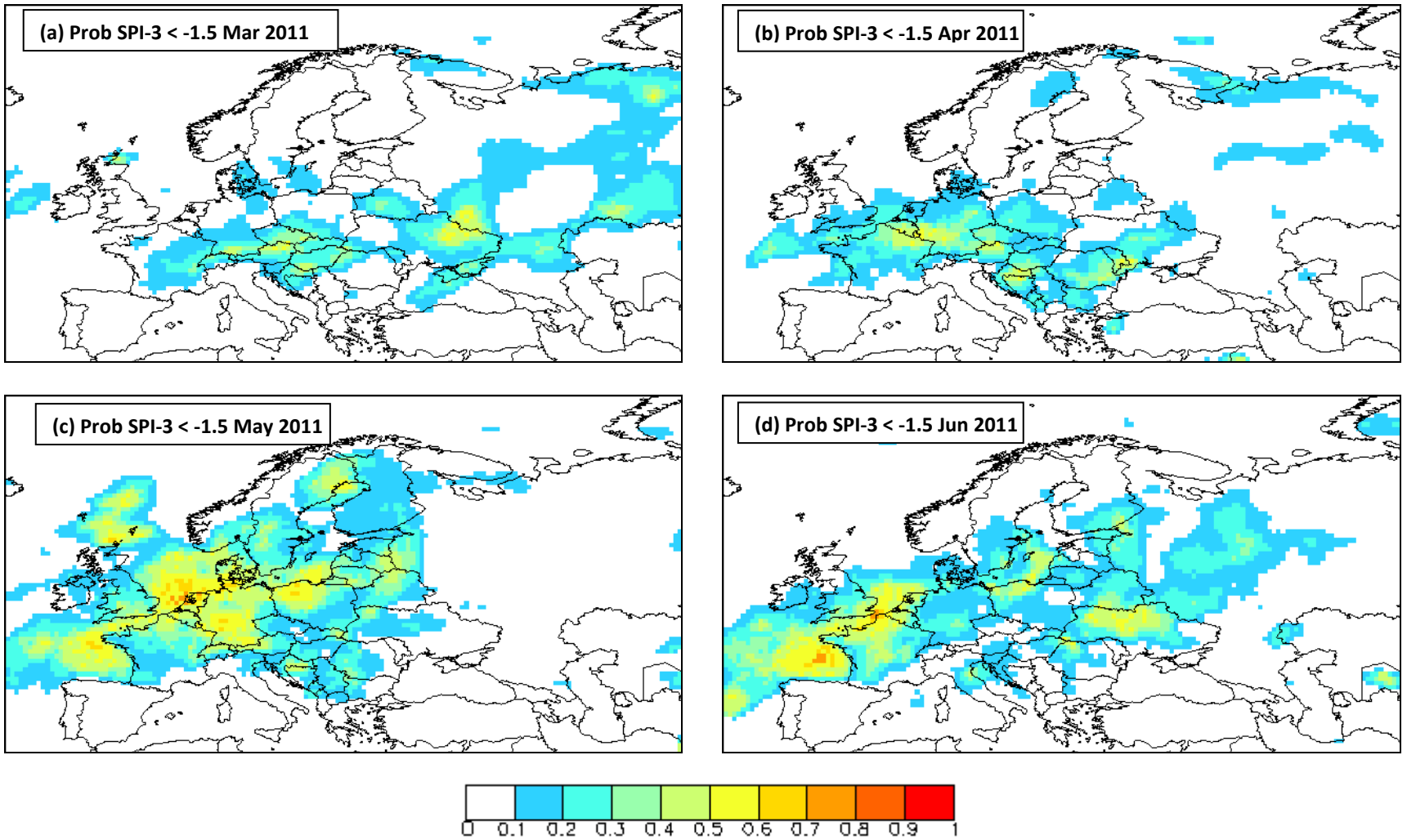


Figure 5.27 – Calibrated forecast probability SPI-3 < -1.5 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

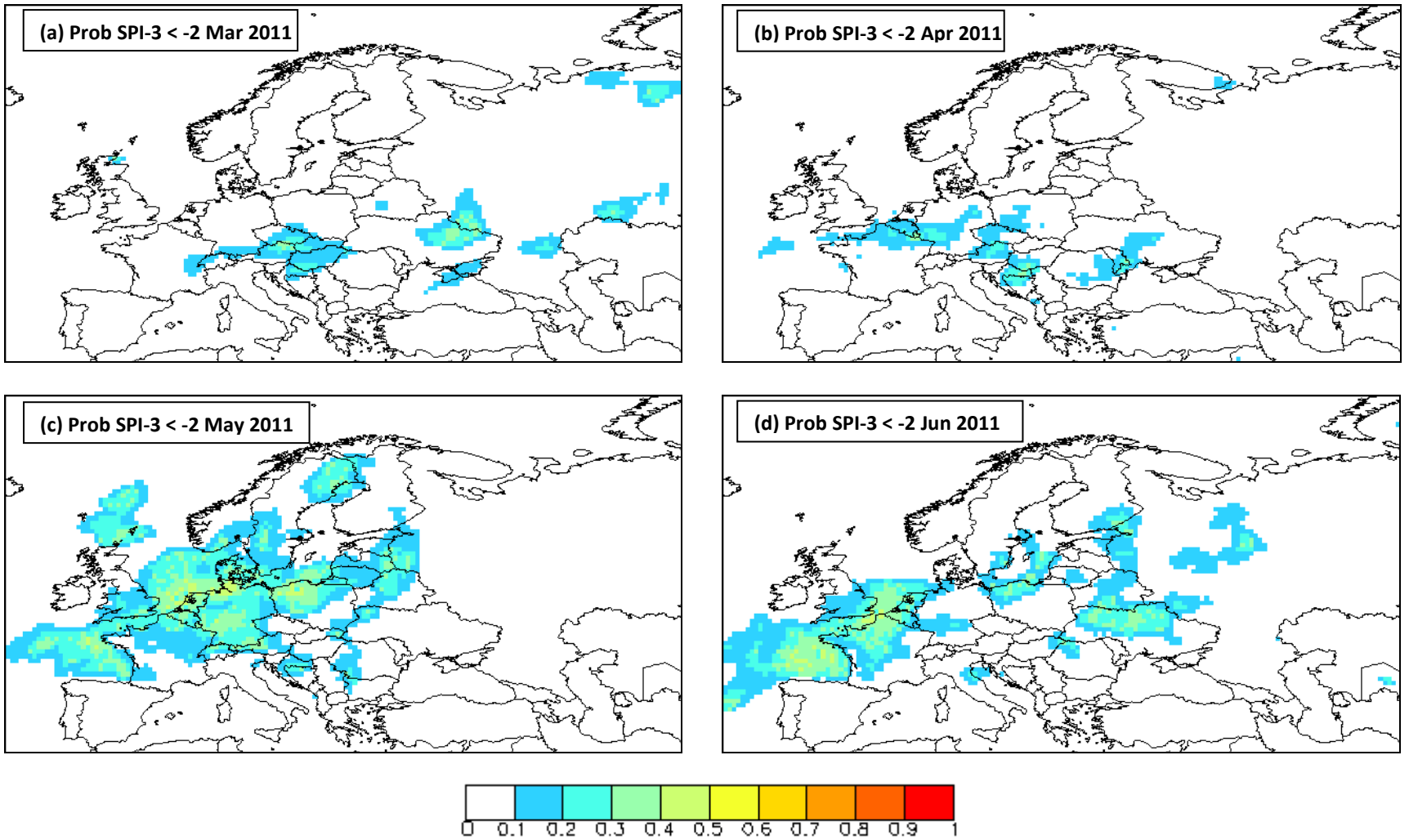


Figure 5.28 – Calibrated forecast probability SPI-3 < -2 for (a) March 2011, (b) April 2011, (c) May 2011, (d) June 2011.

6. Conclusions

The performance of the ECMWF varEPS monthly forecasts for predicting the probability of meteorological drought has been analysed for the most recent operational cycle of the model from March 2004 to August 2011. varEPS forecasts are made up of 50 ensemble members, each starting from slightly different atmospheric states to represent the observation uncertainty. Drought intensity was measured by the SPI and forecasts of SPI-1 and SPI-3 were verified against independent observations from the E-OBS dataset. The SPI-1 forecasts were pure forecasts whereas the SPI-3 forecasts were derived from 2 months of precipitation from the ECMWF ERA-Interim reanalysis and 1 month of forecasts.

Standard verification measures for probabilistic forecasts were used to assess the accuracy of the forecasts. The Brier Score, which measures the mean squared error in the forecast probabilities suggested that varEPS has some skill for the SPI-1 and is more skilful for the SPI-3. However, the Brier Skill Score, which measures the forecast skill against a forecast derived from climatology, showed that the SPI-1 forecast was no more skilful than climatology, but that the SPI-3 forecast was on the whole slightly more skilful than climatology. Furthermore the reliability of both the SPI-1 and SPI-3 forecasts was shown to be weak, with the system tending to forecast probabilities that were too high. However, an analysis of the relative operating characteristics of the forecasting system suggested that it is able to discriminate between events and non-events relatively well. This being the case, improved forecasts may be achieved simply through calibration. Calibration of the forecast showed improved Brier Skill Scores for both the SPI-1 and SPI-3, with calibrated SPI-1 forecasts proving to be more skilful than climatology for most forecasts.

The guidance provided by the forecasts was assessed for two case studies of extreme drought – in Russia in the summer of 2010 and in north-western Europe in the spring of 2011. The forecasts of SPI-1 were found to provide poor information, in both cases predicting dry conditions where they were not observed and in the case for spring 2011 not predicting extreme conditions in the correct locations. Calibration of the forecasts led to lower probabilities of the event being forecast, which could, in areas where the uncalibrated forecast correctly gave high probabilities, be regarded as reducing the skill of the forecast. The SPI-3 forecasts were much better than those for SPI-1, generally predicting the timing and location of the drought events quite well. However, some locations of extreme drought were not predicted at all in the SPI-3 forecasts.

The poor performance of the SPI-1 forecasts is mostly due to the model's inability to predict the hydrological cycle with any degree of skill at long lead times. This is a well known weakness in Numerical Weather Prediction models. Although the performance of NWP 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 for drought prediction at one month lead time, even when using ensemble techniques to model the uncertainty. SPI-3 forecasts however, do provide useful guidance, but most of the skill in these forecasts is due to the 2-months of reanalysis data used to compute the SPI-3.

Another source of error in the forecasts is in the computation of the SPI itself. The precipitation used to model the reference distribution comes from daily forecasts from the slightly coarser resolution ERA-Interim reanalysis model. Therefore to compute the monthly SPI, monthly forecasts compared with aggregated daily forecasts. The climatology of the aggregated daily forecasts is unlikely to be the same as that from monthly forecasts, so the anomaly may not be accurate. This problem may be lessened by using statistics from the varEPS model itself as the reference climatology. ECMWF runs 5-member hindcasts of the varEPS for 18 years into the past for each forecast, which is not considered to be a long enough time period to produce a representative distribution. However, it is expected that these hindcasts will be extended further back in the near future and could be used as reference statistics for the SPI forecast.

It is currently not recommended to use forecasts of SPI with one month lead time to provide guidance. They are likely to give misleading 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

Arpe, K., S. A. G. Leroy, H. Lahijani, and V. Khan, 2011. Impact of the European Russia drought in 2010 on the Caspian Sea level. *Hydrol. Earth Syst. Sci. Discuss.,* **8,** 7781–7803.

Brier, G. W., 1950. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.,* **78,** 1–3.

Bordi I, K. Fraedrich, M. Petitta, and A. Sutera, 2005. Methods for predicting drought occurrences. *Proc. Sixth international Conf. of the European water resources association,* Menton, France, 7–10.

Cancelliere A., G. Di Mauro, B. Bonaccorso, and G. Rossi, 2007. Drought forecasting using the Standardized Precipitation Index. *Water Resour. Manage.* **21,** 801–819.

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

European Commission, 2006. *Water Scarcity and Drought. First Interim Report*, November 2006.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008. A European daily high-resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res (Atmospheres),* **113,** D20119, doi:10.1029/2008JD10201

Hosking, J. R. M., and J. R. Wallis, 1997. *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.

Hwang Y., and G. J. Carbone, 2009. Ensemble forecasts of drought indices using a conditional residual resampling technique. *J. Appl. Meteor. Climatol.*, **48,** 1289–1301.

Golitsyn, G. S., 1995. The Caspian Sea level as a problem of diagnosis and prognosis of the regional climate change. *Izv. Russ. Acad. Sci. Atmos. Oceanic Phys., Engl. Trans.,* **31,** 366–372..

Guttman, N., 1998. Comparing the Palmer drought index and the standardized precipitation index. *J. Am. Water Resour. As.,* **34,** 113-121.

Guttman, N. B., 1999. Accepting the Standandardized Precipitation Index: a calculation algorithm. *J. Am. Water Resour. As.,* **35,** 311-322.

Hannaford, J., B. Lloyd-Hughes, C. Keef, S. Parry, C. Prudhomme, 2011. Examining the large-scale spatial coherence of European drought using regional indicators of rainfall and streamflow deficit. *Hydrol. Process.*, **25,** 1146-1162.

Heim, R. R., Jr, 2002. A review of twentieth-century drought indices used in the United States. *Bull. Amer. Meteor. Soc.*, **83,** 1149-1165.

Lloyd-Hughes, B., and M. A. Saunders, 2002. A drought climatology for Europe. *Int. J. Climatol.*, **22,** 1571–1592.

McKee, T. B., N. J. Doesken, and J. Kleist, 1993. The relationship of drought frequency and duration to time scales. *Proc. Eighth Conf. on Applied Climatology,* Anaheim, CA, Amer. Meteor. Soc. 179-184.

McKee, T. B., N. J. Doesken, and J. Kleist, 1993. Drought monitoring with multiple time scales. *Proc. Ninth Conf. on Applied Climatology,* Dallas, TX, Amer. Meteor. Soc. 233-236.

McRoberts, D. B., and J. W. Nielsen-Gammon, 2011. The Use of a High-Resolution SPI for Drought Monitoring and Assessment. *J. Appl. Meteor. Climatol.*, doi: http://dx.doi.org/10.1175/JAMC-D-10-05015.1

Mishra A. K. and V. P. Singh (2010). A review of drought concepts. *J. Hydrol.*, **391,** 202- 216.

Murphy, A. H., 1973. A new vector partition of the probability score. *J. Appl. Meteor.,* **12,** 595–600.

Mylne K. R., 2002. Decision-making from probability forecasts based on forecast value. *Meteorol. Appl,* **9,** 307–315.

Palmer, W. C., 1965. Meteorological drought. Research Paper 45, U.S. Weather Bureau, Washington, DC, 58pp.

Rodionov, S. N., 1994. *Global and Regional Climate Interactions: The Caspian Sea Experience.* Kluwer Academic Publishers, Dordrecht/Boston/London, 241 pp.

Quiring, S. M., 2009: Developing objective operational definitions for monitoring drought. *J. Appl. Meteor. Climatol.*, **48,** 1217–1229.

Santos, J. F., I. Pulido-Calvo, and M. M. Portela, 2010. Spatial and temporal variability of droughts in Portugal. *Water Resour. Res.*, **46,** W03503, doi:10.1029/2009WR008071.

Svoboda, M., D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, and S. Stephens, 2002. The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83,** 1181-1190.

Wilhite, D. A., M. J. Hayes, and M. D. Svoboda, 2000. Drought monitoring and assessment: status and trends in the United States. *Drought and drought mitigation in Europe.* Vogt, J. V., and F Somma Eds., Kluwer Academic Publishers, 149-160.

World Meteorological Organization – WMO, 2006. Drought monitoring and early warning: Concepts, progress and future challenges, WMO No. 1006.

Wu, H., M. D. Svoboda, M. J. Hayes, D. A. Wilhite and F. Wen, 2007. Appropriate application of the standardized precipitation index in arid locations and dry seasons. *Int. J. Climatol.,* **27,** 65-79.

European Commission

EUR 25254 EN – Joint Research Centre – Institute for Environment and Sustainability
Title: Forecasting Drought in Europe with the Standardized Precipitation Index
Author(s): Andrew Singleton
Luxembourg: Publications Office of the European Union
2012 – 68 pp. – 21 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593 (print), ISSN 1831-9424 (online)

ISBN 978-92-79-23394-4 (PDF)
ISBN 978-92-79-23393-7 (print)

doi:10.2788/16522

Abstract
This report describes an assessment of the performance of the European Centre for Medium Range Weather Forecasts variable resolution ensemble prediction system as a tool for forecasting drought using the Standardized Precipitation Index (SPI) with one month lead time. The model is verified using standard verification measures of the Brier Score, the Brier Skill Score, reliability and relative operating characteristics. It is found that for the 1-month SPI, the model has little skill in forecasting drought events and the forecast is generally unreliable. For the 3-month SPI the model has more skill, but this skill comes from the use of 2 months of reanalysis precipitation in and 1-month of forecast precipitation in building in the 3-month SPI. Calibration of the forecasts through adjusting the forecast probabilities to observed frequencies improved the verification statistics. Two case studies using the model were analysed and it was found that the model did not give useful guidance, and in fact calibration had the effect of underestimating the probability of extreme events where the model had some skill. It is recommended that ensemble probabilistic forecasts not be used as a tool for decision making with regard to drought without further improvement in the model performance.

Acknowledgements
We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://eca.knmi.nl). We acknowledge the ECMWF varEPS and ERA-Interim datasets from the Meteorological Archiving System (MARS) of the European Centre for Medium Range Weather Forecasts (ECMWF).

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-25254-EN-N

Publications Office

ISBN 978-92-79-23394-4