

The extreme forecast index applied to seasonal forecasts

E. Dutra, M. Diamantakis,
I. Tsonevsky, E. Zsoter, F. Wetterhall,
T. Stockdale, D. Richardson and
F. Pappenberger

Research Department

June 2013

To appear in Atmospheric Science Letters

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/publications/>

Contact: library@ecmwf.int

© Copyright 2013

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

The Extreme Forecast index (EFI) concept has been applied to the ECMWF seasonal forecasts (S4) of 2-meter temperature and total precipitation using a novel semi-analytical technique. Results derived from synthetic data highlight the importance of large ensemble sizes to reduce the EFI calculation uncertainty due to sampling. This new diagnostic, complements current diagnostics, as exemplified for the 2012 warm summer in South Central and Eastern Europe. The EFI provides an integrated measure of the difference between a particular seasonal forecast ensemble and the underlying model climate which can be used as an early warning indicator.

1 Introduction

The quality of ensemble forecast systems has steadily improved over the last decades and several measures of extreme or anomalous situations that can potentially provide the forecast users with early warning systems have been developed. The Extreme Forecast Index (EFI, Lalaurette, 2003, Zsoter, 2006), developed at ECMWF, is an example of an index that was designed to identify situations where the medium-range ensemble prediction system (EPS) forecasts are detecting extreme situations. Detection of extremes can be accomplished by comparing model forecasts to the underlying model climatology (Lalaurette, 2003, Thielen et al., 2009, Bartholmes et al., 2009, Cloke et al, 2010, Alfieri et al., 2011). The major advantage of such an approach is that it can be applied everywhere including in areas where observations are sparse or unavailable and that it inherently accounts for the need of forecast calibration as it is based on relative difference between forecasts and model climatology. This does not overcome the problems associated with sparse or unavailable observations, in particular surface observations of temperature and precipitation which are limitation in any evaluation of the model skill.

On the seasonal time scales (up to 6 months lead time) comparable extreme indexes have not been applied. Operational detection of “extremes” has been mainly focused at the probability of exceeding percentiles. Seasonal forecasting is not a weather forecast: weather can be considered as a snapshot of continually changing atmospheric conditions. Seasonal forecasts provide a range of possible climate changes that are likely to occur in the season ahead. Due to the chaotic nature of the atmospheric circulation, it is not possible to predict the daily weather variations at a specific location months in advance. However, in some parts of the world, and in some circumstances, it may be possible to give a relatively narrow range within which weather values are expected to occur. Such forecast can easily be understood and acted upon, some of the forecasts associated with strong El Nino events fall into this category (e.g. Stockdale et al. 2011). More typically, the probable ranges of the atmospheric conditions differ only slightly from year to year. Forecasts of these modest shifts might be useful for some users for example as a first warning and could support further decision making involved in drought risks and water resources management.

In this paper, we describe the development and application of the EFI methodology to seasonal forecasts (on the monthly to seasonal time scales). While on the medium-range the EFI provides information on extreme events on a daily/local scale (e.g. storms), on the seasonal scale it will measure how the mean model climate and the actual forecast differ on a monthly/large scale. The extraction and analysis of large volumes of ensemble data is a complex and difficult task. The EFI is a possible and efficient way of summarizing the available information by scaling the ensemble forecast with respect to the model climate. The real advantage of using the EFI lies in the fact that it is an integral measure referenced to the model climate that contains all the information regarding variability

of a parameter in location and time. Therefore, the user can recognize anomalous situations without defining different space – and – time dependent thresholds. This will summarize the probabilistic forecasts and can highlight potential anomalies on the long-range that should be analyzed in detail by the user/forecast provider using the full range of the ensembles forecasts. In the following section the data and methods are presented followed by the results; the main conclusions are summarized in the last section.

2 Data and Methods

2.1 ECMWF Seasonal forecast

ECMWF seasonal forecasts, based on an atmosphere-ocean coupled model, were used for the EFI calculations. We evaluate the recently implemented System 4, which became available in November 2011 (S4; Molteni et al. 2011). The horizontal resolution of the atmospheric model is about 0.7° in the grid-point space (spectral truncation TL255) with 91 vertical levels in the atmosphere. The ocean model has 42 vertical levels with a horizontal resolution of approximately 1° . The seasonal forecasts consist of a 51-member ensemble with 7 months lead time, including the month of issue, referred as 0 month lead time. S4 also has a set of re-forecasts starting on the 1st of every month for the years 1981 to 2010. These hindcasts are identical to the real-time forecasts in every way, except that the ensemble size is only 15 rather than 51. The data from these hindcasts create the “model climate” that can be used for calibration of forecasts products, and are used here to define the model climate from which the EFI forecasts are calculated. Molteni et al. (2011) present an overview of the model biases and forecast scores of S4.

The EFI calculations were performed for 2-metre temperature (T2M) and total precipitation (TP) considering the model climate as the full hindcast period with 450 values (30 years x 15 ensemble members). Prior to the calculations, the T2M and TP fields were spatially interpolated from TL255 resolution to a regular grid of $2.5^\circ \times 2.5^\circ$ (mass conservative for TP and bilinear for T2M), reducing the amount of data to process and smoothing the spatial fields.

2.2 EFI calculation

The EFI formulation scales departures from the reference climate Cumulative Distribution Function (CDF) and is defined as:

$$EFI = \frac{2}{\pi} \int_0^1 \frac{p - F(p)}{\sqrt{p(1-p)}} dp \quad (1)$$

$F(p)$ is a function denoting the proportion of ensemble members lying below the p quantile of the climate record. The term $1/\sqrt{p(1-p)}$, which takes its minimum for $p=0.5$ and its maximum at both ends of the probability range, is used to give more weight to the tails of the distribution. This can be also interpreted as using the statistical Anderson-Darling (Anderson and Darling 1952) test as a modification of the known Kolmogorov-Smirnov test. Given that $0 \leq F(p) \leq 1$ EFI values will lie in the same interval with unit values obtained when all the ensemble members are above (positive) or below (negative) the climate distribution.

The numerical integration of Eq. (1), where $F(p) \in [0,1]$, is problematic near 0 or 1 where the integrated function $\rightarrow \infty$. To circumvent this problem, the endpoints can be excluded from the integration interval and the function is integrated in a slightly smaller domain $[\varepsilon, 1-\varepsilon]$. However, this has several disadvantages: (i) loss of accuracy, (ii) EFI calculation is sensitive to the chosen numerical integration method and (iii) the EFI values can overshoot/undershoot the interval $[-1, 1]$.

To deal effectively with these problems, a new technique, semi-analytical in nature, has been developed which is stable and produces results in the desired interval $[-1, 1]$. The main idea is to do an analytical integration thus avoiding the numerical problems associated with the singularity of the integrand. However, for an analytical calculation, a continuous representation of the model data dependent function $F(p)$ is needed. This can be done using a monotone interpolation formula such as linear interpolation:

$$F(p_i) \cong \hat{F}(p_i) = f_i + \frac{p - p_i}{p_{i+1} - p_i} (f_i - f_{i+1}), \quad p \in [p_i, p_{i+1}], \quad f_i \equiv F(p_i) \quad (2)$$

where the index i refers to the sorted climate record within a set of N samples.

The resulting EFI formula is a second order accurate finite series expansion:

$$EFI = \frac{2}{\pi} \sum_{i=0}^{N-1} \left\{ \left[(2f_i - 1) - (2p_i - 1) \frac{\Delta f_i}{\Delta p_i} \right] [\arccos(\sqrt{p_i + 1}) - \arccos(\sqrt{p_i})] + \left(\sqrt{p_{i+1}(1 - p_{i+1})} - \sqrt{p_i(1 - p_i)} \right) \frac{\Delta f_i}{\Delta p_i} \right\} \quad (3)$$

where $\Delta f_{i+1} = f_{i+1} - f_i$; $\Delta p_i = p_{i+1} - p_i$; $p_0 = 0$; $p_N = 1$

In the results that follow, the EFI calculations were performed using Eq. (3). The numerical accuracy of Eq. (3) is mainly dependent on the number of samples N in the climate distribution. Numerical experiments show that for a setup similar to the one used by ECMWF seasonal forecasts ($N=450$ samples: 15 ensemble members x 30 years), the absolute value of the numerical error will be less than 10^{-2} . The good numerical accuracy of Eq. (3) can be attributed to two facts: (i) the formula integrates the singular part of the integral exactly and (ii) the piecewise linear approximation used for $F(p)$ turns to be accurate given that $F(p)$ is smooth and monotonically increasing at each interval $[p_i, p_{i+1}]$. Further details of the numerical accuracy of the EFI calculation are presented in the appendix.

3 Results

In this section we present the EFI sensitivity to the forecast ensemble size and the EFI relation with the changes in the forecast ensemble mean and standard deviation using synthetic data, which allows a broad testing and understanding of the EFI behaviour.

3.1 Idealized EFI

3.1.1 EFI sensitivity for forecast ensemble size

Each seasonal forecast of S4 is composed of 51 ensemble members in real time and 15 ensemble members in the hindcast period. The decision on the number of ensemble members is mainly constrained by the available computational resources. Since the real time forecasts of S4 are only available since May 2011, EFI calculations previous to that period will be comparing CDF from the model climate with 450 samples against forecasts of only 15 samples. The reduced size in the hindcast period will impact the uncertainty of the EFI values. To evaluate the impact of the ensemble size on the EFI calculation, we performed a sensitivity analysis by producing synthetic data with the following characteristics:

- Model climate: 100000 samples each with 450 values (to represent 30 years with 15 ensemble members) randomly sampled from a normal distribution;
- Forecasts: 100000 forecasts with ensemble sizes ranging from 10 to 300 ensemble members. The forecasts were generated by randomly sampling data from the model climate.

Since the forecasts are samples of the model climate, i.e. drawn for the same distribution, if the ensemble size would not have impact, the EFI values should be very close to zero. Figure 1 displays EFI values for each ensemble size, where the boxplots represent the EFI values distribution of the 100000 forecasts with the same ensemble size. As expected, the median of the EFI values is zero, since the forecasts are sub-samples of the model climate. The vertical extension of the boxplots (between the percentiles 10 to 90 - 80% of the EFI values) gives an indication of the associated uncertainty of the EFI calculation due to the ensemble size, or sampling.

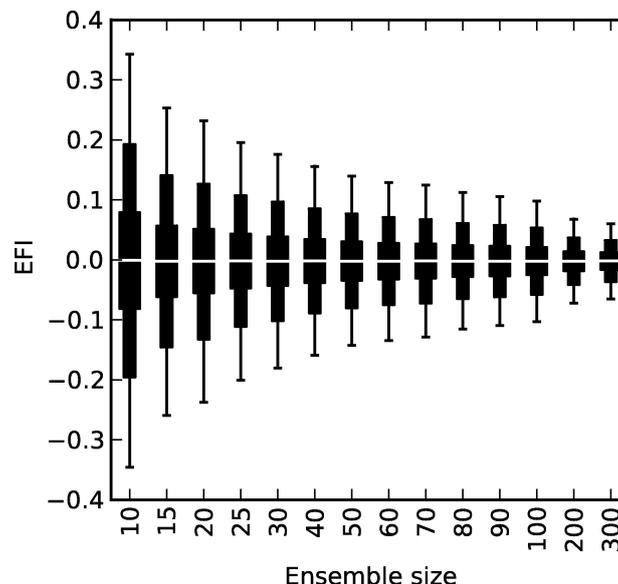


Figure 1: EFI values distribution comparing 100000 forecasts with different ensemble sizes (horizontal axis) sampled from the same distribution as the model climate (with 450 values). The boxplots represent the percentiles 10, 30, 50 (white line), 70 and 90 and the lines extend from percentiles 1 to 99.

For forecasts with 100 members this uncertainty is +/- 0.05, dropping to +/- 0.03 with 300 members. However, for ensemble sizes of 50 members the uncertainty is +/- 0.08 increasing for 0.14 for 15 ensemble members. Similar results were found when increasing the sample sizes, using a uniform random distribution, or changing the mean of the forecast (mean EFI is different from zero, but the uncertainty bounds remained the same). Although these results were derived from synthetic data they highlight the importance of large ensemble sizes to properly capture the forecast distribution and to allow the EFI calculation with a low uncertainty. The EFI sensitivity to the ensemble size is not caused by the numerical calculation (Eq. 3), but by sampling errors of the forecast distribution ($F(p)$ in Eq. 1)(due to small ensemble sizes). Similar results were also found when analyzing the fraction of ensemble members below or above the lower or upper tercile, which is a common metric used in seasonal forecasts. This should be considered when analyzing the EFI fields of S4 for dates previous to May 2011 (based on 15 ensemble members) that will have almost twice the uncertainty of the real time EFI fields (based on 51 ensemble members).

3.1.2 *EFI relation with the forecast mean and standard deviation*

The EFI measures departures of the forecast CDF with respect to the climate CDF. These values are not intuitive, apart from 1 (or -1) when all the forecast ensemble members are above (or below) the climate distribution, or close to zero when the forecast distribution is very similar to the climate. To evaluate the relation between the EFI values and the changes in the ensemble mean and standard deviation of the forecast in the climate we performed the following EFI synthetic calculations:

- Model climate: 1000 samples each with 450 values randomly sampled from a normal distribution of mean X and standard deviation Y ;
- Forecasts: 1000 random forecasts with 51 ensemble members with normal distribution of mean $X + \Delta \times Y$ (with Δ varying from -5 to 5) and standard deviation $\Delta \times Y$ (with Δ varying from 0.2 to 4).

The changes in the EFI due to changes in the ensemble mean and standard deviation are represented in Figure 2. The results show that EFI is sensitive to changes in both the ensemble mean and spread: the same change in the ensemble mean in a sharper forecasts will result in a higher EFI. Note that in practice it is unlikely that a forecast will have a larger uncertainty than the climate, i.e. the scaled forecast uncertainty can be expected to be less than or approximately equal to one. A forecast with the same standard deviation as the climate, with changes in the ensemble mean of 0.5, 1, 1.5, 2 and 2.5 standard deviation of the climate will have 0.2, 0.42, 0.60, 0.74 and 0.85 of EFI values, respectively (see bottom panel of Figure 2). Different computations were performed (changing the baseline model climate, from normal to uniform distribution and sample sizes) and the results were similar. The information in Figure 2 can be used as a simple lookup table to connect EFI values to changes in the ensemble mean/ spread that can be further analyzed by examining in detail the different ensemble members.

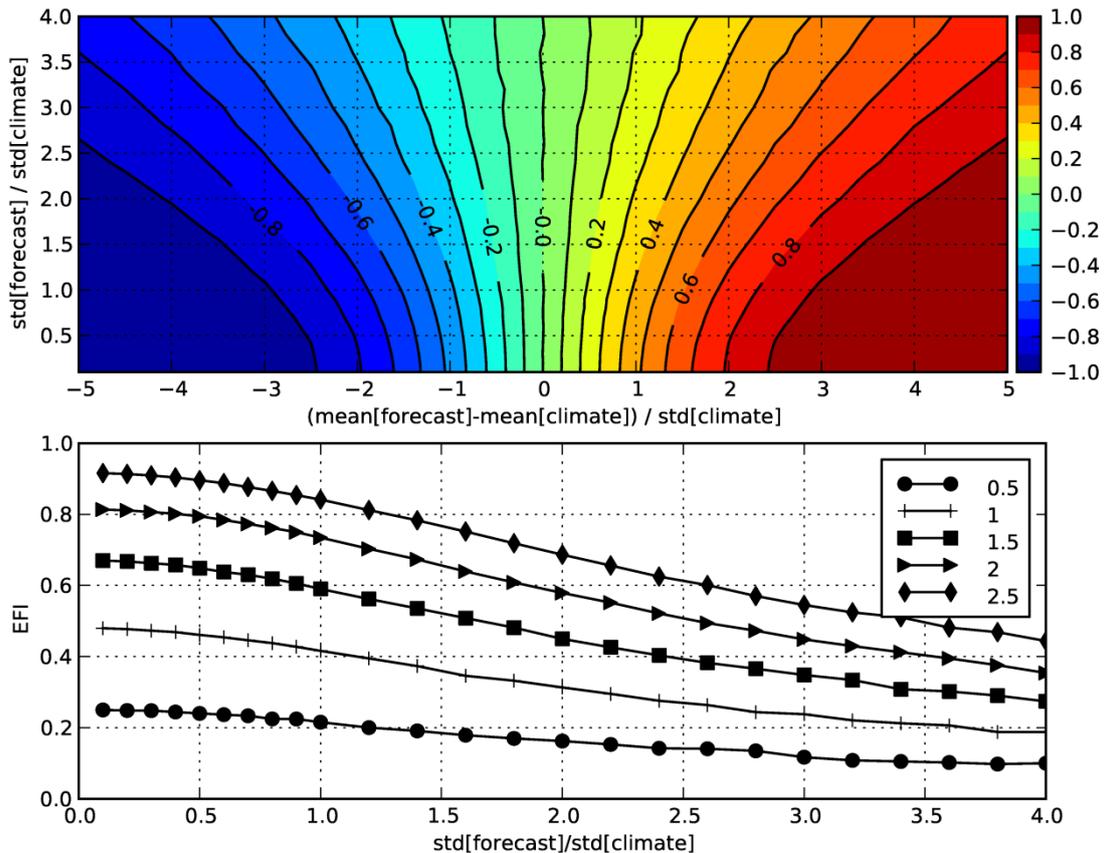


Figure 2: Top: EFI values as a function of changes in the ensemble mean and standard deviation of the forecast in respect to the climate. The changes in the mean (horizontal axis) are rescaled as the climate standard deviation anomalies added/subtracted to the climate mean, and the changes in the standard deviation (vertical axis) are given as the ration between the forecast and the climate standard deviation. Bottom: sub-sample of the contours in the top panel of the EFI for changes in the forecast ensemble mean (plus or minus) of 0.5 (circles), 1 (plus), 1.5 (square), 2 (right triangle) and 2.5 (diamond) standard deviations of the climate as a function of the changes in the standard deviation (horizontal axis).

3.2 EFI in the seasonal forecasts

Following the previous results using idealized distributions of the model climate and forecasts, this section presents the behavior of the EFI applied to the ECMWF S4 seasonal forecasts. This paper is only focused on the EFI development for seasonal forecasts and its behavior, while the forecasts skill is not addressed. For a particular application, the skill of the forecasts, i.e. comparing with actual observations should also be performed. The EFI will only have potential benefit to users in case the forecasts is skilful.

The distribution of EFI over all land points and ocean points for T2M and TP is represented in Figure 3 as a function of forecast month lead time. For both TP and T2M (and also land/ocean points), there is a decrease in the number of high (or low) values of the EFI with lead time. This decrease is mainly from the first/second month of forecast to the remaining forecast months. This behavior can be primarily attributed to the loss of predictability. While the first month of forecast still has some predictability on the medium-range associated with the initial conditions, with increasing lead time the

predictability is reduced along with the forecasts sharpness. This is similar to what was shown in the previous section (3.1.2), where an increase of the forecasts standard deviation for the same change in the forecast mean leads to a reduction of the EFI. Note that in the later months, the EFI is not much bigger than sampling errors would give for an identical forecast/climate distribution. There is also a remarkable difference between TP and T2M, where the EFI of the latter has larger range in particular over the ocean. To further investigate this feature, Figure 4 represents the 90th percentile of TP and T2M EFI forecasts issued in January for different lead times from the hindcast period (1981 to 2010). The spatial distribution of the 90th percentile highlights the differences between the first forecast month and the remaining as well as the differences between T2M and TP and over land and ocean, resembling a map of predictability -higher 90th percentile values associated with higher predictability. The EFI in the medium-range forecasts is typically used as a warning for severe weather when its values are close to 1 (or -1). Applied to these long range forecasts, the thresholds to define warnings cannot be so extreme, since it will be very unlikely for a forecast distribution to differ greatly for the baseline climatology. An option is to use the hindcasts EFI to calculate different thresholds based on the percentiles, varying spatially, for each initial forecast data and lead time. The example given in Figure 4 could be used to define warning levels as well as to identify areas and lead times where the EFI will have limitation, for example if the 90th percentile is below the sampling uncertainty, estimated as ± 0.08 for 50 ensemble members in section 3.1.1.

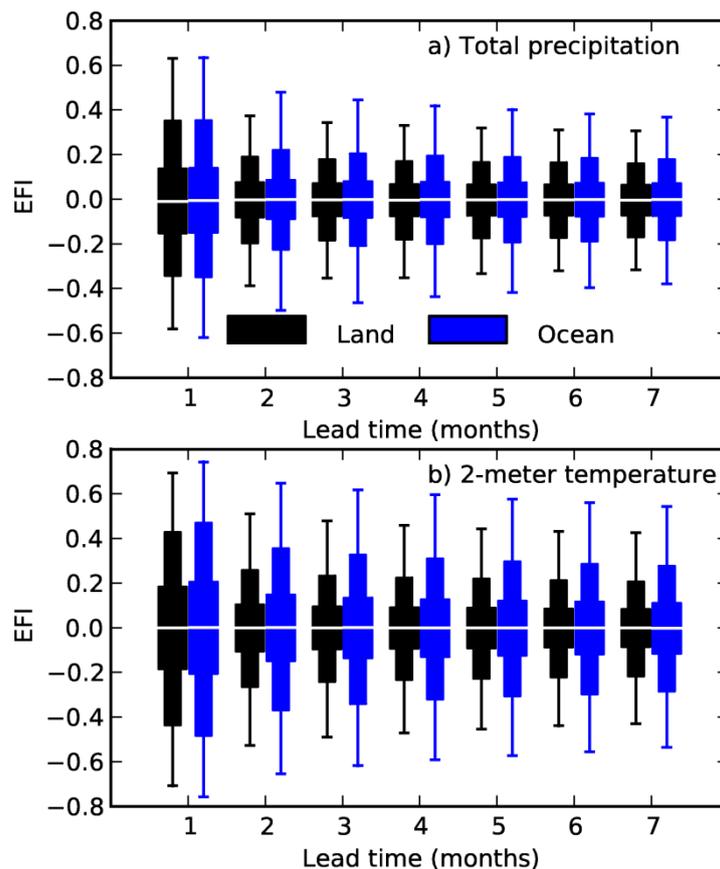


Figure 3: Total precipitation (a) and 2-meter temperature (b) EFI distribution over all land points (black) and ocean points (blue) for the full S4 hindcast period (1981 to 2010, all months) as a function of lead time.

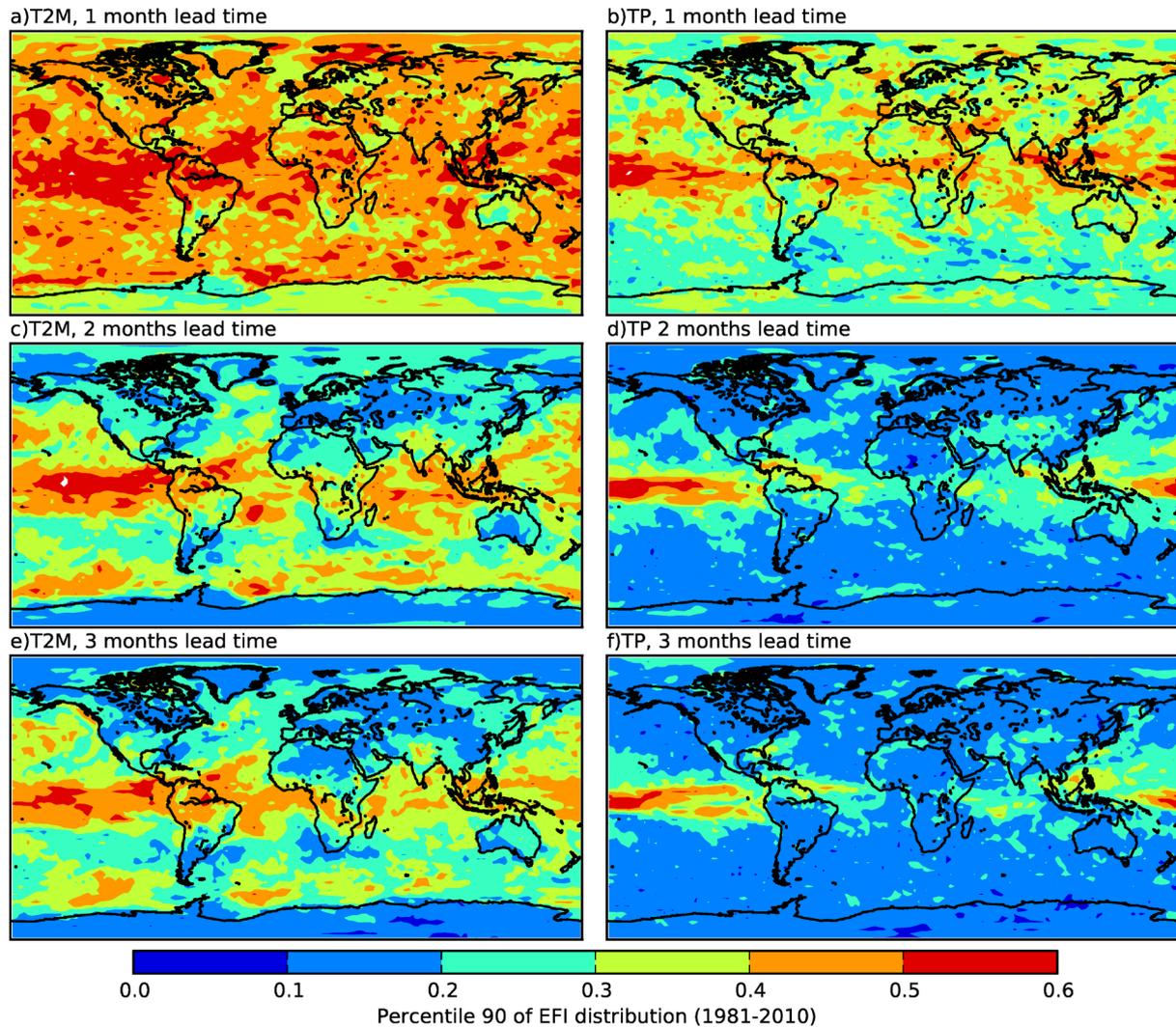


Figure 4: Spatial distribution of the 90 percentile of total precipitation (b,d,f) and 2-meters temperature (a,c,e) EFI calculated between 1981 and 2010 for the forecasts initialized in January for lead times of 1 (a,b), 2 (c,d) and 3 (e,f) months.

Figure 5 compares three standard products: probabilities below the lower tercile, and above the median and upper tercile with the EFI for the S4 T2M forecasts initialized in May 2012 and valid for JJA 2012. Summer 2012 had above normal temperature anomalies in south central and Eastern Europe (Figure 5e). This warm anomaly was partially detected by S4 forecasts issued in May 2012, with high probabilities of T2M above the median and upper tercile, low probabilities of T2M below the lower tercile, and positive values of the EFI. An example of the use of thresholds to define EFI warning levels is presented in Figure 5d where the grid points with EFI values above the 90th percentile are highlighted (Figure 5d). The EFI map resumes the information contained in the other three products. Additionally, user defined warning levels (e.g. values above below a certain percentile based on the hindcasts) could be used as early warning/detection of forecasts anomalies that should be further analyzed using remaining diagnostics.

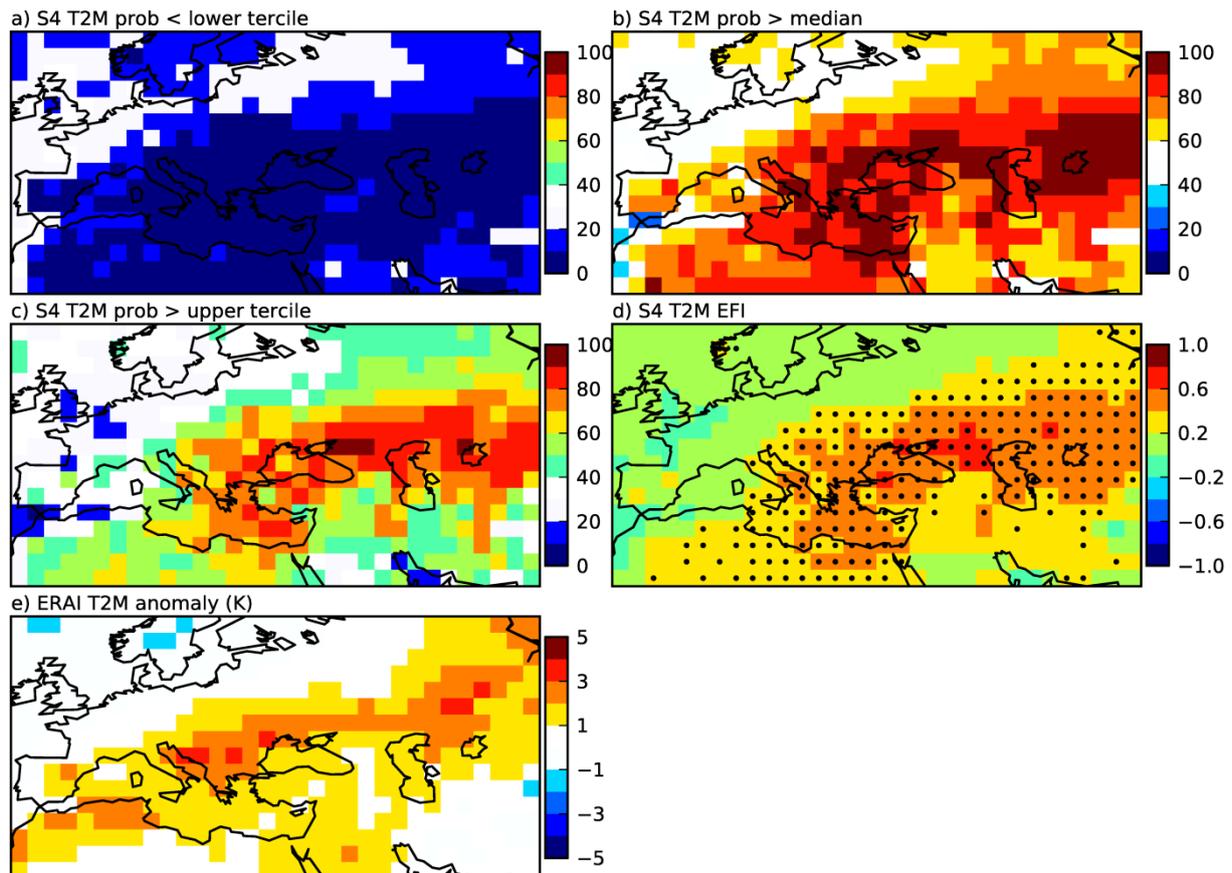


Figure 5: S4 2-meter temperature (T2M) forecasts valid for JJA 2012 initialized in May 2012. a) probability of T2M below the lower tercile; b) probability of T2M exceeding the median; c) probability of T2M above the upper tercile; d) T2M EFI and; e) ERA-Interim JJA 2012 T2M anomaly. In panel d) the black dots indicate EFI values above the 90th percentile of the EFI values between 1981 and 2010.

4 Conclusions

The EFI concept, mainly used in medium-range ensemble forecasts, has been extended and implemented on seasonal forecasts. A new semi-analytical formulation is presented that allows an accurate calculation of the EFI. An assessment of the EFI behaviour using synthetic data showed the variation of EFI with changes in the ensemble mean and spread: similar changes in the ensemble mean will result in higher or lower EFI values for low and high ensemble spread, respectively. This information can be used as a guide for the interpretation of the EFI. Furthermore, we also show the importance of large ensembles to reduce the uncertainty of the EFI due to sampling errors of the forecast distribution.

The EFI was applied to the ECMWF seasonal forecasts of monthly means of 2-metre temperature and total precipitation up to 6 months lead time. It was found that the EFI distribution changes with lead time, with a reduction in the occurrence of high/low values. This is associated with smaller changes of the ensemble mean in respect to the model climate, and to an increase of the ensemble. In this situation, the distribution of the forecast is similar to the underlying model climate. This was mainly visible for total precipitation over land. On the other hand, the EFI of monthly 2-meters temperature in the tropical regions shows a higher range, even on long lead times. These results are associated with

the low predictability of total precipitation over land when compared with sea surface temperature in the tropical regions. These results are coherent with the synthetic data tests showing that an increase in the ensemble spread (can be associated with a reduction of predictability) leads to a decrease of the EFI for similar changes in the ensemble mean.

We have successfully implemented the EFI applied to seasonal prediction, and examined its behaviour. It is clearly sensitive to ensemble size, which makes a detailed study difficult with only 15 member hindcasts. Our results do not include an evaluation of the skill of the EFI, since it is not possible to derive an observed EFI for verification. Such skill assessment should be performed on the original fields, and the EFI can be evaluated on a case study basis, as it was shown for the 2012 summer in Southern Europe. With the currently available data, the EFI does not bring additional information to the standard products, such as tercile or median probabilities, but resumes such information in a single indicator, complementing currently used diagnostics. Further investigations can be carried out when larger sample sizes become available (which is planned for selected start dates).

Acknowledgements

This work was funded by the FP7 EU projects GLOWASIS (<http://www.glowasis.eu>) and DEWFORA (<http://www.dewfora.net>).

Appendix: Numerical accuracy of the EFI calculations

To demonstrate that errors in the EFI calculation are mainly due to sampling and not due to numerics, the new formula given by Eq. (3) was applied to a case with increasing number of ensemble members sampled from a randomly generate climate distribution. In this case, the exact value of EFI is known and equal to 0. The climate distribution was generated with 1025, 513 and 257 samples to allow an exact subdivision of the forecasts sampling from the climate, i.e. with 1025 samples: 1025 (climate == forecast), 513, 257, 129, 65, 33, 17, 9, 5, and 3. The climate distribution was randomly generated using different distributions (uniform, normal, gamma, and several combinations) with no impact on the results. The results in figure A1 show that with only 3 ensemble members homogenously sampled from the climate (N=513) the EFI is very close to zero (0.01), and this value drops to 0.002 for 65 ensemble members.

The numerical accuracy of Eq. (3) is mainly dependent on the number of samples N in the climate distribution (Figure A2). For a setup similar to the one used by ECMWF seasonal forecasts (N=450 samples: 15 ensemble members x 30 years), the numerical accuracy will be higher than 10^{-2} . The good numerical accuracy of Eq. (3) can be attributed to two facts: (i) the formula integrates the singular part of the integral exactly and (ii) the piecewise linear approximation used for F(p) turns to be accurate given that F(p) is smooth and monotonically increasing at each interval $[p_i, p_{i+1}]$.

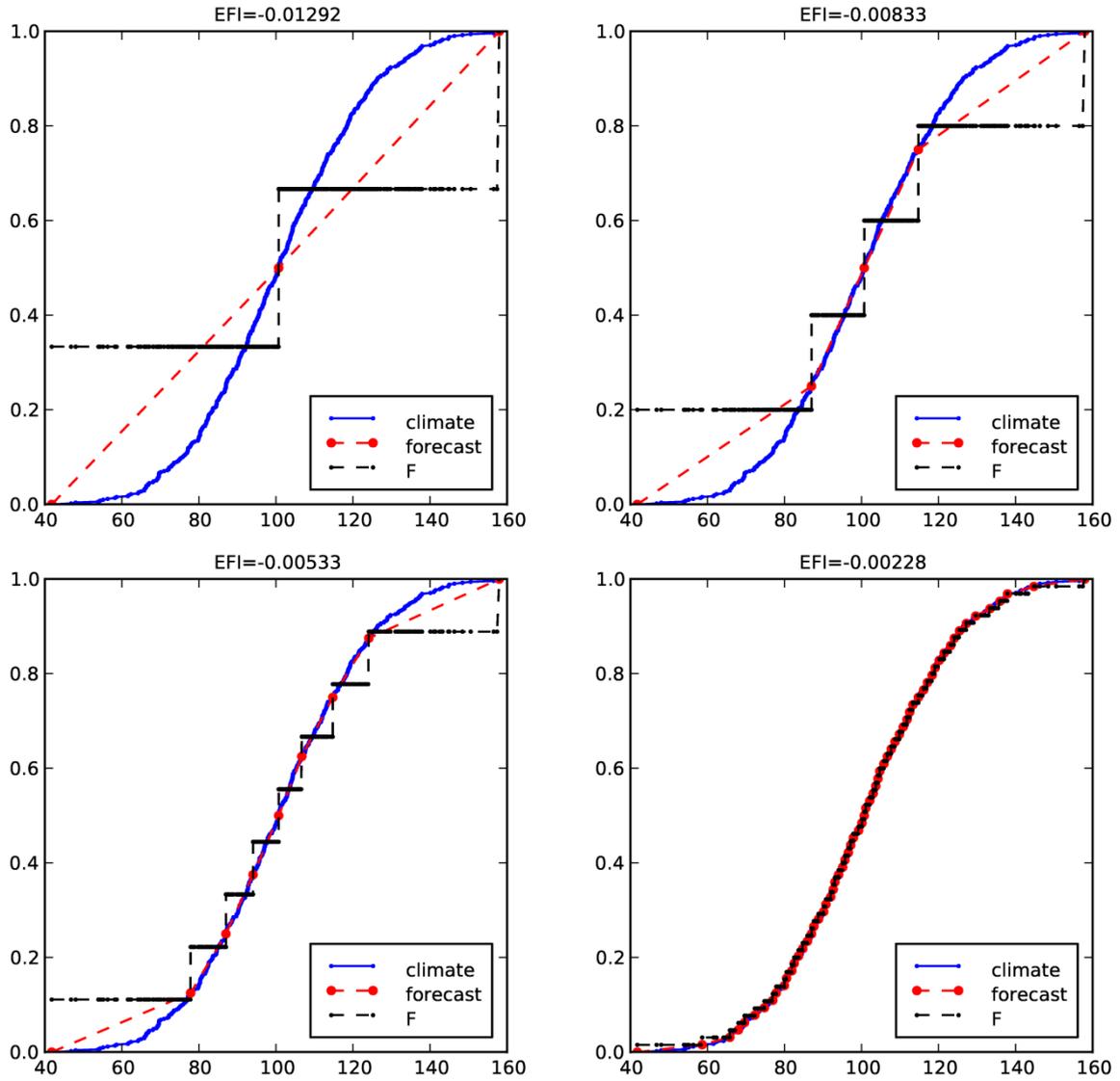


Figure A1: Numerical accuracy of the EFI calculation based on a climate with 513 samples (blue) and forecasts (red) with 3 (top left), 5 (top right) 9 (bottom left) and 65 (bottom right) ensemble members homogeneously sampled from the climate. The integration function $F(p)$ in equation (1): proportion of ensemble members lying below the p quantile of the climate record is represented by the black lines.

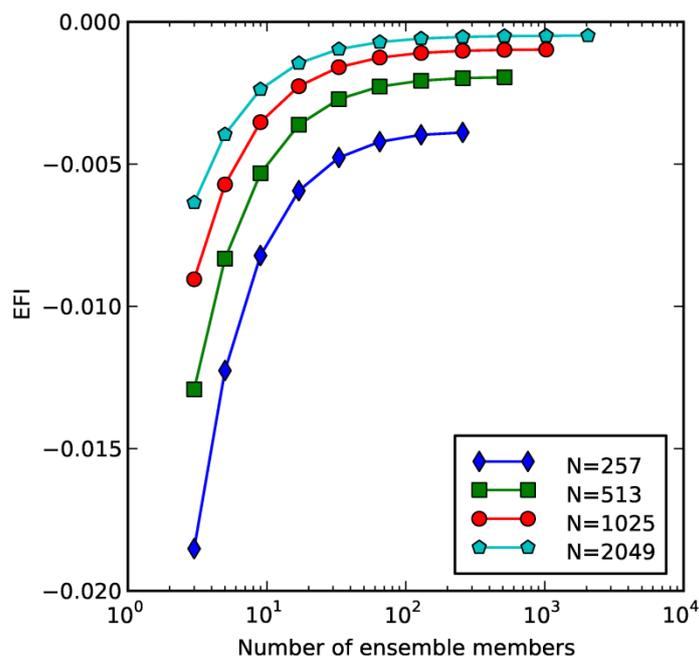


Figure A2: Numerical accuracy of the EFI calculation for different ensemble members (horizontal axis – logarithm scale) based on a climate with 2049 (cyan) 1025 (red), 513 (green) and 257 (blue) samples.

References

- Alfieri, L., Velasco, D., and Thielen, J., 2011: Flash flood detection through a multi-stage probabilistic warning system for heavy precipitation events, *Adv. Geosci.*, **29**, 69–75, doi:10.5194/adgeo-29-69-2011.
- Anderson, T. and D. Darling, 1952: Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193-212.
- Cloke, H. L., Jeffers, C., Wetterhall, F., Byrne, T., Lowe J., Pappenberger, F., 2010: Climate impacts on river flow: projections for the Medway catchment, UK, with UKCP09 and CATCHMOD, *Hydrological Processes*, DOI : 10.1002/hyp.776
- Lalauette, F. 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, **129** (594):3037-3057.
- Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (System 4), *ECMWF Tech. Memo.*, **656**, 49 pp.
- Stockdale, T., and Coauthors, 2011: ECMWF seasonal forecast system 3 and its prediction of sea surface temperature, *Climate Dyn.*, **37**(3), 455-471, doi: 10.1007/s00382-010-0947-3.
- Thielen-del Pozo, J., J. Bartholmes, M.-H. Ramos, and A. de Roo, 2009: The European Flood Alert System. Part 1: Concept and development, *Hydrology and Earth System Sciences*, **13**, 125-140, 2009
- Zsoter, E., 2006: Recent developments in extreme weather forecasting, *ECMWF Newsletter*, **107**, 8-17.