TECHNICAL MEMORANDUM

373

Early Detection of Abnormal Weather Using a Probabilistic Extreme Forecast Index

François Lalaurette

Operations Department

July 2002

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

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading
RG2 9AX
library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Abstract

A new method to extract information related to unusual forecast distributions from ensemble prediction systems (EPS) is presented. It consists in an Extreme Forecast Index (EFI) that ranks the departure between the EPS forecast and the model climate between -1 (all members are breaking records for low values in the model climate) and +1 (record-breaking high values). First the new index is derived, related to other measures of statistical significance, and a few properties are given. It is then demonstrated how the accumulation of ensemble forecasts everyday allows to quickly build up a model pseudo-climate that is representative enough to detect large departures from normal conditions while being representative of the latest developments in terms of resolution or physical parameterisations. The EFI is then subjected to several severe weather events that did happen in Europe over the last few years, and it is shown to be potentially useful to get forecasters alerted of the risk of severe weather up to three of four days in advance. Objective verification over five months in 2001-2002 is finally presented. Although the results confirm that the model pseudo-climate is good enough to set-up thresholds for severe weather evenly throughout Europe, the false alarm rates are much larger than usually expected by forecasters or users. It is argued however that operating characteristics in the early medium range for severe weather have largely been ignored in the past. Although the signal is weak in the 3 to 5 days range, it is undoubtedly associated with a forecast skill that might be used for setting up pre-alerts to be used either internally by Meteorological services, or externally by advanced users aware of its error characteristics.

1 Introduction

Both the quality improvement of numerical forecasts of weather parameters (precipitation, wind, temperatures, cloud cover) and the development of Ensemble Prediction Systems (EPS) have risen the expectations in terms of providing the forecast users with early warnings of extreme events. Direct comparison between observed, local weather parameters and direct model output is however difficult (Ghelli and Lalaurette, 2000). When dealing with extreme events, this usually means that the large scale models would forecast on very rare (if any) occasions the events that are of interest for the users. Although the usual approach to overcome this problem is to post-process the direct model output using past knowledge of error characteristics (Model Output Statistics, Perfect Prognostic methods, Kalman filters, etc....), such methods are usually only globally optimum: they have a tendency to miss extreme cases that have only a small weight in the global optimisation procedure. The proposal is made here instead to qualify extreme events not with respect to the observed, local climate, but rather to the model climate. In this way we will keep control of the relative frequency of occurrence of our alerts.

Common sense seems to dictate the use of very high resolution models in order to cope with extreme weather events. It is true indeed that the higher the resolution, the wider is the range of weather phenomenons that can be described explicitly, including potentially damaging ones such as squall lines, tornadoes or tropical cyclones. It is now widely recognised however that not all phenomena that can be represented with realism in high resolution models are predictable. This is because of the very demanding error constraints on the prescription of the initial state that cannot be achieved with the observing systems currently in operations (Gall and Shapiro, 2000). On the synoptic scales, optimal control theory offers a convenient framework where the sensitivity of the forecast to small errors in the initial conditions can be demonstrated using adjoint models (Rabier et al., 1996; Hello et al., 2000). Recently, the Fronts and Atlantic Storm-Track EXperiment (FASTEX) has been a large scale cooperative effort to bring over a two-months period the supplementary data needed to improve the reliability of deterministic forecast of rapidly developing perturbations over the Northern Atlantic (Joly et al., 1999).

Such efforts to improve the subsynoptic and mesoscale deterministic forecasts of extreme events in the mid-latitudes are of considerable importance to improve the service brought by the meteorological community to our societies, most notably in terms of the protection of human lives and properties. It is argued however in this paper that a complementary approach can be followed that could provide early warnings of some value

up to three or four days in advance. Although the predictability of extreme weather events decreases sharply with time, the EPS in operations in ECMWF as well as in other global numerical weather prediction centres (Molteni et al., 1996; Toth and Kalnay, 1997; Szunyogh and Toth, 2002; Houtekamer et al., 1996; Kyouda, 2002) develop a smooth, flow dependent transition from an almost purely deterministic forecast in the very short ranges to an almost purely climatic probability distribution in the late medium range. Instead of basing the warnings on a single realisation of a highly chaotic process, it is argued here that a more reliable procedure in the early medium range is to make the warning decision on the basis of the full forecast probability distribution. A nice demonstration of the potential benefit of using probabilities rather than deterministic forecasts in decision making processes can be found in Richardson (2000); Palmer (2002); Zhu et al. (2002). Rather than fine tuning the level of probability upon which action should be taken for a well defined event, the approach we present here is one where a continuous measure of the distance between the probability distribution of the day and the probability distribution of the model climate is used.

The paper is organised as follows: the new Extreme Forecast Index is defined in Section 2, where a few basic properties are derived. Section 3 is about the computation of the model climate information needed to evaluate the index. A few cases when the potential use of the index are shown can be found in Section 4, first verification results are presented in Section 5 and a short summary with perspectives for further developments are given in Section 6.

2 The Extreme Forecast Index

2.1 A new index: why?

The Extreme Forecast Index is a measure of the difference between a probabilistic forecast and a model climate distribution. The problem of measuring distances between probability distributions is a classical one that is addressed in a series of statistical test procedures: a well known example is the Kolmogorov test that aims at accepting or rejecting the hypothesis "Is the given empirical sample of values likely to be drawn from a given (e.g.: Gaussian) distribution?" (Der Megreditchian, 1992; Wilks, 1995). The Kolmogorov test gives the answer to this question on the basis of statistics of the distance:

$$K = \max_{-\infty<x<+\infty} |F(x) - F_e(x)| \tag{1}$$

with $F(x)$ being the tested (e.g. Gaussian) distribution, and $F_e(x)$ the empirical (ordered sample) one. What we want to achieve is in a way exactly the opposite: we want to know whether or not the EPS forecast is very different from the usual (climate) distribution of events. Also, rather than a yes/no answer, we would like to build an index that gives a continuous scale of values. This has been achieved already in the context of the verification of probabilistic forecasts: the Continuous Ranked Probability Score (CRPS) is a measure of how far the verification (the observed value x_o of a given meteorological parameter) is from a probabilistic forecast (Bouttier, 1994; Hersbach, 2000). The CRPS writes:

$$CRPS = \int_{-\infty}^{+\infty} (F_e(x) - F_o(x))^2 dx \tag{2}$$

with $F_o(x)$ the observed (one-stepped, Heavyside) distribution. If the empirical distribution itself is one-stepped (sampling a unique value x_e), then it can be shown that the CRPS reduces to the Mean Absolute Error (MAE). Although we could easily extent the CRPS to the measure of distances between two distributions without necessarily having one of them stepwise, the CRPS has the default of scaling like the forecast parameter. It is not an index, and therefore it would be difficult to compare values for regions with very different climates, and each parameter would have its own range of critical values. For that reason, the proposal is to formulate the index in probability space rather than meteorological parameter space as has been proposed recently for

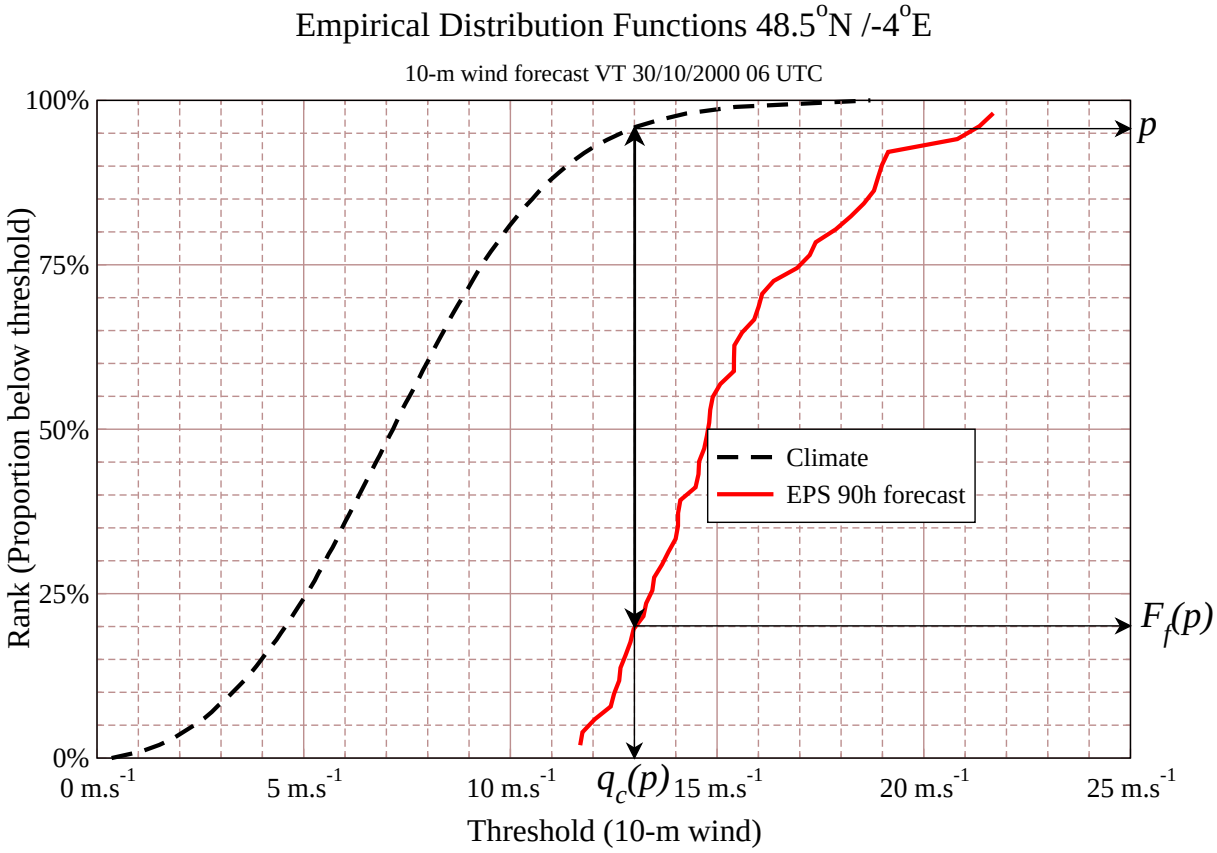

Figure 1: Climate (dashed) and EPS 90h forecast (full) 10m wind speed distributions near Brest, France (30 October 2000 06UTC, "Halloween" storm). A large EFI_2 value (84.1%) is generated by the the unusually large number of EPS members forecasting strong winds ($F_f(p) \ll p$)

the Linear Error in Probability Space (LEPS, Ward and Folland (1991)). Another shortcoming of the CRPS for our application is that it is (like the Kolmogorov statistics) definite positive: it gives zero values only if the two distributions are identical, and is positive otherwise. For detecting severe weather conditions, we not only want to know if the EPS distribution deviates a lot from the climate, but also if it deviates in a direction that may be dangerous for human activities: to know that the wind will be unusually calm is very different from knowing that it will blow with unusual strength! Therefore we need to have a signed index, which is achieved neither by the Kolmogorov, nor by the CRPS measures.

2.2 Definition and properties

For a given location on Earth and a given meteorological parameter, one can associate to each proportion p of the ranked climate records a parameter threshold $q_c(p)$. This is known as a *quantile* or *percentile* of the climate distribution: $q_c(0)$ is the absolute minimum and $q_c(1)$ is the absolute maximum than can be found in the climate history, while $q_c(\frac{1}{2})$ is the median, $q_c(\frac{1}{4})$ the first quartile, etc.... If then we define $F_f(p)$ as the probability with which a given EPS forecast predicts that the observation will be below $q_c(p)$ (see Fig. 1), we can define:

$$EFI_{2m+1} = 2(m+1)\int_0^1 \Big(p - F_f(p)\Big)^{2m+1} dp \tag{3}$$

which is an index than cumulates differences between the climate and the forecast distribution: $F_f(p) = p$ only in the case when EPS members are below $q_c(p)$ in exactly the same proportion as climate records are.

$F_f(p) > p$ means that the EPS gives a larger probability to lie below $q_c(p)$ than the climate proportion p: if this is true for most of the quantiles, then the EFI gets negative values, otherwise it has positive ones.

Simple properties can be derived from the definition above:

1. the EFI takes values between -1(if all EPS members forecast values that are below the absolute minimum: $F_f(p) = 1$ for all percentiles) and +1 (all EPS members forecast values above the absolute maximum: $F_f(p) = 0$)

2. if all EPS members forecast the same value ranking as quantile p_0 in the climate records, then:

$$EFI_{2m+1} = p_0^{2m+2} - (p_0 - 1)^{2m+2} \tag{4}$$

   and therefore EFI_{2m+1}=0 if $p_0 = \frac{1}{2}$ (all EPS members forecast the climate median value)

Although the EFI is definite positive for even values, it can easily be transformed into a signed index in such cases, depending on which side of the climate distribution the forecast distribution mostly lies:

$$EFI_{2m} = (2m+1)\frac{EFI_1}{|EFI_1|}\int_0^1 \Big(p - F_f(p)\Big)^{2m} dp \tag{5}$$

Finally, using the identity $EFI_1 = 1 - 2\int_0^1 F_f(p)dp$, the expressions for all possible values of EFI_n are:

$$EFI_n = -(n+1)\int_0^1 \Big(p - F_f(p)\Big)^n dp \quad \text{if } n \text{ is even and } \int_0^1 F_f(p)dp > \tfrac{1}{2} \tag{6}$$

$$EFI_n = (n+1)\int_0^1 \Big(p - F_f(p)\Big)^n dp \qquad \text{in all other cases} \tag{7}$$

In the same way, Eq.4 (deterministic case when the forecast reduces to a single value) can be generalised to all values of n as:

$$EFI_n = (p_0 - 1)^{n+1} - p_0^{n+1} \quad \text{if } n \text{ is even and } p_0 < \tfrac{1}{2} \tag{8}$$

$$EFI_n = p_0^{n+1} - (p_0 - 1)^{n+1} \qquad \text{in all other cases} \tag{9}$$

In practice, only $n = 2$ and $n = 3$ have been tested: the corresponding variations are plotted in Fig.2 which can be used to compute an Extreme Forecast Index from a deterministic forecast. The index has its maximum selectivity near both extremes: this is a desirable property, as the objective is to identify severe weather conditions. Although first tests where using $n = 2$, since December 2001, daily indices have been produced using $n = 3$ instead. Higher values could prove more selective, but also give increasing weight to the tails that are poorly sampled portions of the distribution, adding undesired levels of noise. It may be useful to note that in the deterministic case the following identities apply:

$$EFI_1 = 2p'_0 \tag{10}$$

$$EFI_2 = (3p'^2_0 + \frac{1}{4})\frac{p'_0}{|p'_0|} \tag{11}$$

$$EFI_3 = p'_0(1 + 4p'^2_0) \tag{12}$$

where $p'_0 = p_0 - \frac{1}{2}$ is the departure from the median ($-\frac{1}{2} \le p'_0 \le \frac{1}{2}$). The case when rainfall are the forecast parameter is singular: the zero value cannot be ranked in the climate as there is a full range of cases with no

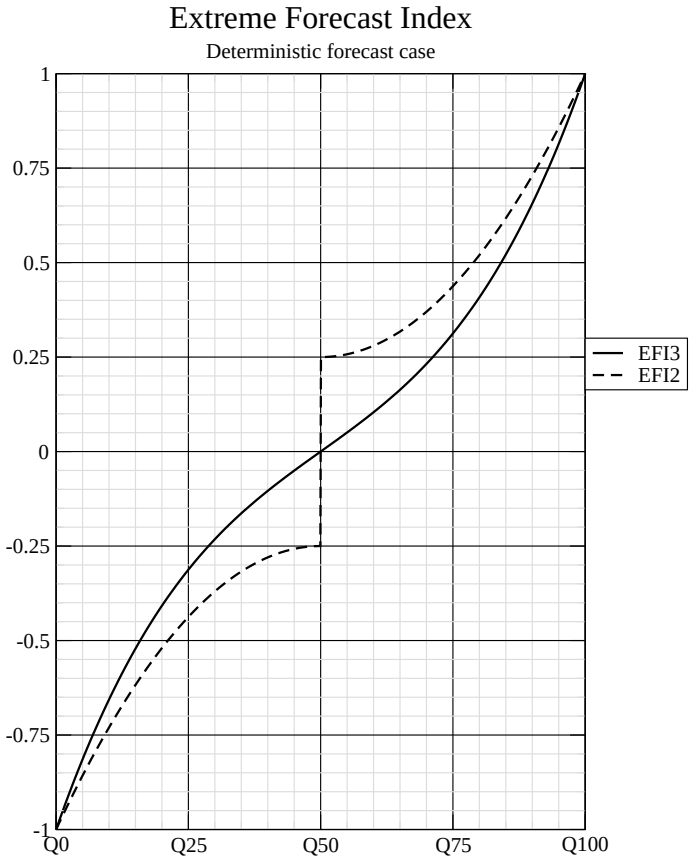

Figure 2: EFI variations in the case of a deterministic forecast (see Eq. (4)); the x-axis indicates the rank of the forecast in the climate (quantile); the EFI_2 gets the sign from the EFI_1 (+ if > Q50, - if < Q50

rainfall. If the proportion of dry days in the climate is p_1, there is therefore no unique way to define $F_f(p)$ for $0 \leq p \leq p_1$. The proposed definition for the EFI in this case will be:

$$EFI_{2m+1} = \frac{2m+2}{1-p_1^{2m+2}} \int_{p_1}^{1} \Big(p - F_f(p)\Big)^{2m+1} dp \tag{13}$$

Variation with (p_0, p_1) of the index for $n = 3$ in the case of a deterministic forecast (or of an ensemble with all members forecasting the same value) can be found in Fig.3.

The design of the index in any case is strongly dependant on the prescription of a reference climate distribution from which the measure of abnormality is taken. The following section describes how this climate distribution is derived in practice.

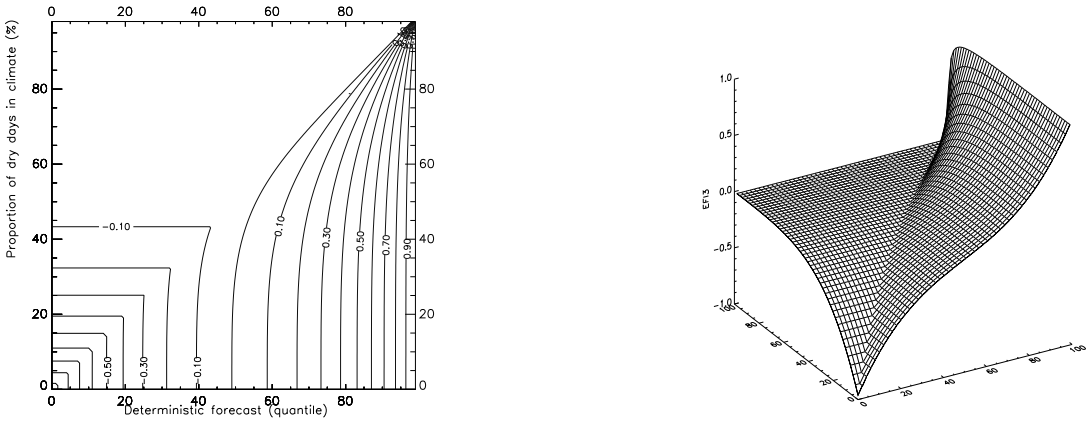

Figure 3: Two representations of the EFI_3 index variations for a deterministic forecast of rain; in both cases, the x-axis is the quantile value forecasted, and the y-axis refers to the proportion of dry cases in the climate - respectively p_0 and p_1 in Eq. (13)

3 Deriving climate information from EPS forecasts

3.1 Three strategies to gather the model climate information

WMO defines a climate as a collection of statistics over three decades (30-years period). Although using ECMWF model analyses has been shown to give a valuable insight into the climatology of a variety of atmospheric phenomena, including extratropical cyclones (Ayrault et al., 1995), such a reference period is not currently available at ECMWF: operations only started in 1979, and the 40-years reanalysis project (ERA-40, Simmons and Gibson (2000)) has not been completed yet. It should also be realised that the climate we need for the computation of the EFI must provide a good estimate of the probability that the model, not necessarily the atmosphere, will lie in a particular area of the space of meteorological variables. We know for example that the model orography has only a limited accuracy, and this will strongly affect the wind regimes near the ground. The same could be said for the influence that changes in the land-sea mask might have on the diurnal cycle of 2m-temperature. It is therefore important that the "climate" used for the EFI computation and the EPS model have a consistent representation of the orography. It is also desirable that they have the same representation of physical processes, so that the limitations, biases, etc.... are offset in a way by being given the same weight in the daily forecast and in the climate reference. We therefore have to do with two conflicting objectives: the first is to gather a large sample of atmospheric situations to generate a realistic distribution of possible outcomes, while the other is to use an up-to-date model version to generate this sample. Three different strategies can be designed to achieve such objectives:

1. the best but most costly solution is to re-run a consistent, long set of forecasts starting from re-analyses (e.g. ERA-40) using the model version used for the EPS (at the time of writing IFS Cy24r3 at T255L40 resolution); this should be repeated each time a major upgrade of the model (e.g. a resolution change, or a major revision in the representation of surface processes) is implemented;

2. to gather information from model forecasts valid for the same calendar months over the previous three years

3. to gather information bootstrapping model forecasts valid for the previous calendar month on the current year ($YYYY/MM_{-1}$), and from the previous year, the current calendar month ($YYYY_{-1}/MM$) and the following one ($YYYY_{-1}/MM_{+1}$)

Option 1. is not ruled out: when ERA-40 data are available, the cost and benefits of running such a model climatology on a regular basis will be evaluated, and it is currently considered as a serious option. In order to go ahead with the testing of the EFI, "pseudo-climate" strategies 2. and 3. are however the only ones that can be considered currently. Although three years (option 2.) are far shorter than the agreed length required to gather climate data, we can try to turn the atmospheric unpredictability into our favour for once, by using the wide range of inaccurate but realistic patterns found in the ensemble of medium range forecasts run since 1996 as information on the climate variability. The underlying assumption is that most of the climate variability over Europe is of unforced, chaotic type - which obviously excludes decadal modes of variability such as those involved in the North-Atlantic Oscillation (NAO) for example. The transient patterns however are losing their predictability over less than a week, and therefore their variability is expected to be explored in our super-ensemble of 10 days forecast/ 51 members much quicker than in the real world. It is difficult however to put numbers on how quicker the climate variability is explored by such a super ensemble. An upper bound is 510 times quicker: this could be achieved only if all scales lost all correlation from one day to the next and from one ensemble member to the other. This is however not likely to be the case, as the ensemble forecasts usually cluster into similar types of supra-synoptic scale scenarios, when a perfect decorrelation would result in an ensemble mean looking like climatology. This may look like a desirable property for our climate application, but would also result in a perfectly useless ensemble from a forecast point of view.

Rather arbitrarily, it has been decided to keep from each ensemble forecast only the initial condition, the 5-days and the 10-days forecast. They have all been kept as independent realisations, which once again is not a valid assumption - among other implications, this means that we are making the assumption that model systematic errors are developing much more slowly than the differences between ensemble members, so that initial conditions, day 5 and day 10 forecasts can be considered as part of the same probability distribution. Although this would obviously not be true on seasonal scales or in the tropics, we expect this to be a reasonable gross hypothesis in the mid latitudes for 10-days forecasts.

Both strategies 2. and 3. have been tested so far: in both cases, the pseudo-climate information was gathered for each calendar month and each geographical location in the following way:

- selected parameters are 2m-temperature (over land only), 10-m wind speed (both sustained and daily maximum gust), accumulation of precipitation over the last 24, 120 and 240 hours;
- analysis and EPS forecasts (lead time 5 and 10 days, control and 50 members) valid for 12UTC any day of the month were extracted on the model reduced Gaussian grid using the ECMWF Meteorological Archiving and Retrieval System (MARS, Raoult (1996)); this is referred later as our super-ensemble: its size is about 10000 members (it varies with the number of days in the calendar month), with the exception of the 10-days accumulation of precipitation whose size is only half (5000) this size;
- for every grid point, forecasts were ranked and percentiles were archived; the extreme per-millennium values (0.1% and 99.9%) were also archived, the result being used as empirical distribution functions;

The difference between strategies 2. and 3. is that in the former case, the period 1997-1999 was used as a consistent period of three years with the model run at consistent resolution (T159) while in the latter, more recent, T255 data were used. Empirical distribution functions from the T159 pseudo-climate for parameters retrieved over two very contrasted locations in Europe (Iceland and Greece) are shown in Fig. 4 and Fig. 5. As expected, the pseudo-climate gives a clear indication that Iceland has a more windy, cold and rainy climate that Greece! It reads from these curves that you can find a period of 5 days without rain over Greece in more than 85% of July cases, while with the same probability you will get more than 5mm of rain in Iceland. Another way to look at the pseudo-climate information is to look at the geographical variations of a given quantile of the climate distribution. This has been done in Fig. 6 for 12UTC 2m temperature, where both unusually cold (Q20) and warm (Q80) temperatures have been mapped for January. A crude, constant lapse rate correction for altitude has been applied. The continental effect on winter temperatures is clearly noticeable there. Both the

oceanic and Mediterranean influences are also found, mainly in the reduction of very cold weather (Q20) along the coastlines; the sharp contrast in temperatures between the west coast of Norway and inland Scandinavia is also detected in these maps. Similar information on daily rainfall in January (Q80 and Q95) is shown in Fig. 7. The most noticeable features there are the maxima over North Portugal and South of Norway.

Using these data as the reference from which the EFI is computed results in maps such as Fig. 8 that was generated when a good proportion of EPS members went for a strong wind scenario - what has later been referred to as the ”Halloween storm”, the first in a series of events that have caused major floods over the Midlands and South-Eastern regions of England in November-December 2000. The wind distribution both in the climate and in the EPS 90h forecast are shown in Fig. 1. On this figure, how the large EFI values have been generated can be seen as a result of the large departures between what is the frequency of occurrence in the pseudo-climate (e.g. only 4% of cases exceed 13m s^{-1}) and in this 90h EPS forecast (41 members out of 50 exceed this value). This type of information is generally a very useful addition to the EFI maps, as it allows the identification of the value reached by individual EPS members. Although ideally the forecaster would like to embrace the full distribution of EPS scenarios in one picture, it is not possible in practice to get such a comprehensive view for a wide range of parameters and forecast ranges. The EFI maps therefore may be a useful way to focus the attention on a small selection of areas, parameters and forecast ranges where significant events might occur that deserve more attention from the human expert.

3.2 The impact of the EPS resolution change in November 2000

In November 2000, only a few months after the first tests of EFI map generation, the EPS resolution was increased from T159 up to T255, which roughly translates into a mesh size reduction from 120km to 80km (Buizza and Hollingsworth, 2002). Using the 1997-1999 climate proved to generate spurious signals on many occasions, which can be understood as being at least partly the result of ”moving the posts” between the model used to gather the pseudo-climate and the one used for the current forecast. One particular example is shown in Fig. 8, where the EFI_2 for 12UTC average wind is computed both using the then operational T159 EPS forecast and the T255 forecast that at the time was undergoing final, pre-operational testing. In both cases, the T159, 1997-1999 pseudo-climate is used. Although small scale features are expected to be handled better at T255 than at T159, some suspicious ”bull eye” features appear, most notably in Denmark and French Brittany where the land sea mask has been changed markedly. An ad-hoc correction procedure was developed to correct from observed differences in the parameters distributions during periods when both versions of the model were run in parallel. Although the method proved a good first order correction, it was still generating a high level of noise, and therefore strategy 3. was implemented from December 2001, i.e. when a 12-months period was covered running the EPS at T255 resolution. On that occasion, the pseudo-climate was extended to the globe on the model, reduced Gaussian Grid (it was restricted to Europe before, and was retrieved on a 0.5x0.5 degree lat lon grid). The maximum gust wind of the day was another addition introduced in December 2001: this parameter was only introduced in the model post-processing in June 2000, and therefore was not available for the 1997-1999 pseudo-climate.

A comparison between strategy 2. and 3. for January can be found in Fig. 9 for the same locations in Greece and Iceland used in Fig. 4 and Fig. 5. The differences found here can be both attributed to model changes and non-representativeness (or instability) of the pseudo-climate. The clear increase in the number of strong wind cases (lower right panel) is to a certain extend to be credited to positive changes in the model: higher horizontal resolution (November 2000), increased boundary layer resolution and different post-processing of the 10m-wind speed (October 1999) were introduced in the period separating the two sample climates. Differences in the 2m-temperature distributions are however likely to be partly pointing at shortcomings inherent to strategy 3: using both forecasts from February and December as representative of the —January climate has the effect of broadening the distributions unrealistically. This effect clearly shows in the distributions for Greece, when the new distributions have much more warm cases than the old one.

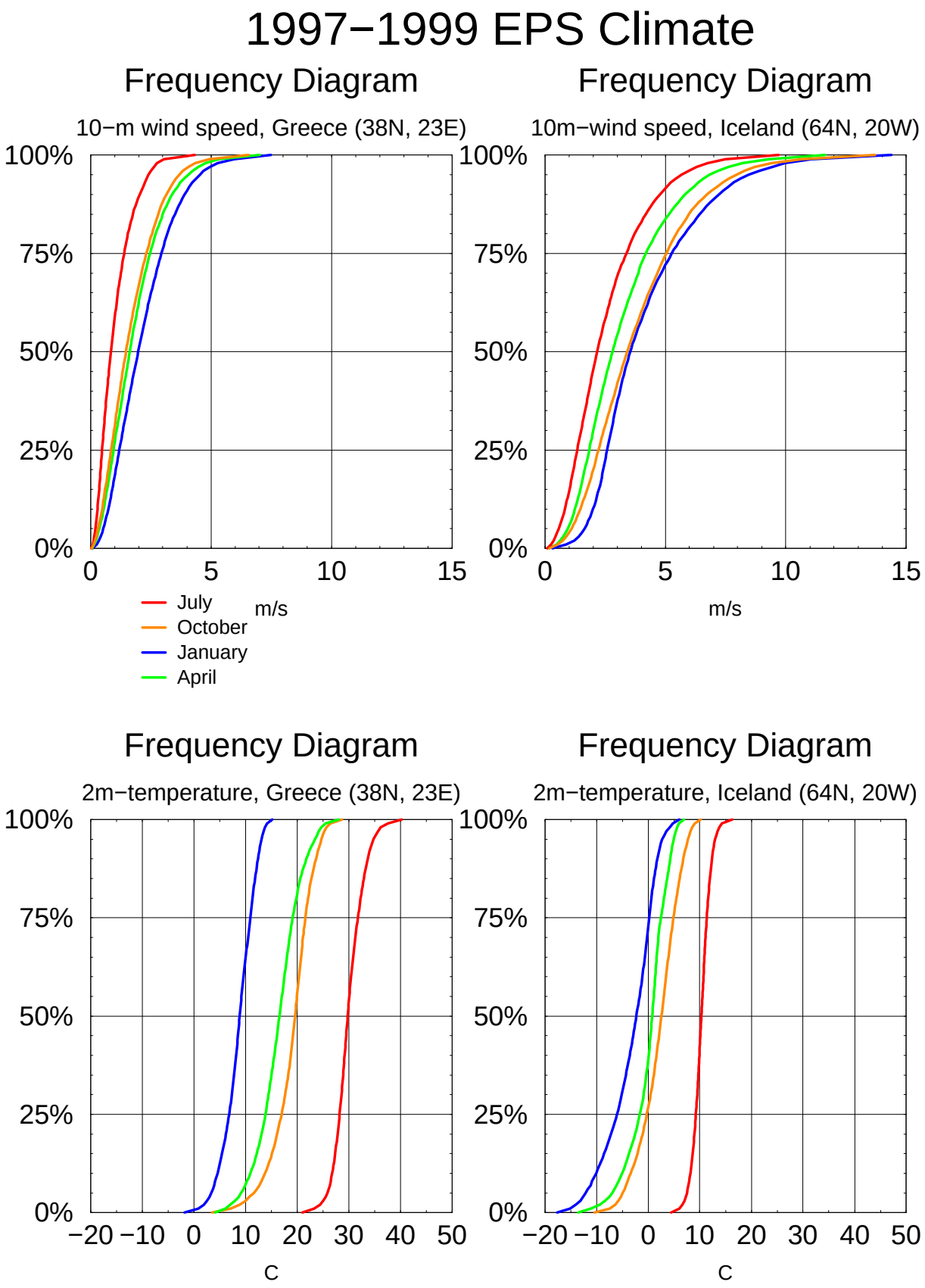

Figure 4: Pseudo-climate distributions derived from EPS T159 operational forecasts in 1997-1999. In all panels, red is for July, orange for October, blue for January and green for April. Left column is for Greece, right for Iceland. Upper Row: 10-m average wind speed at 12UTC, lower : 12UTC 2m-temperature

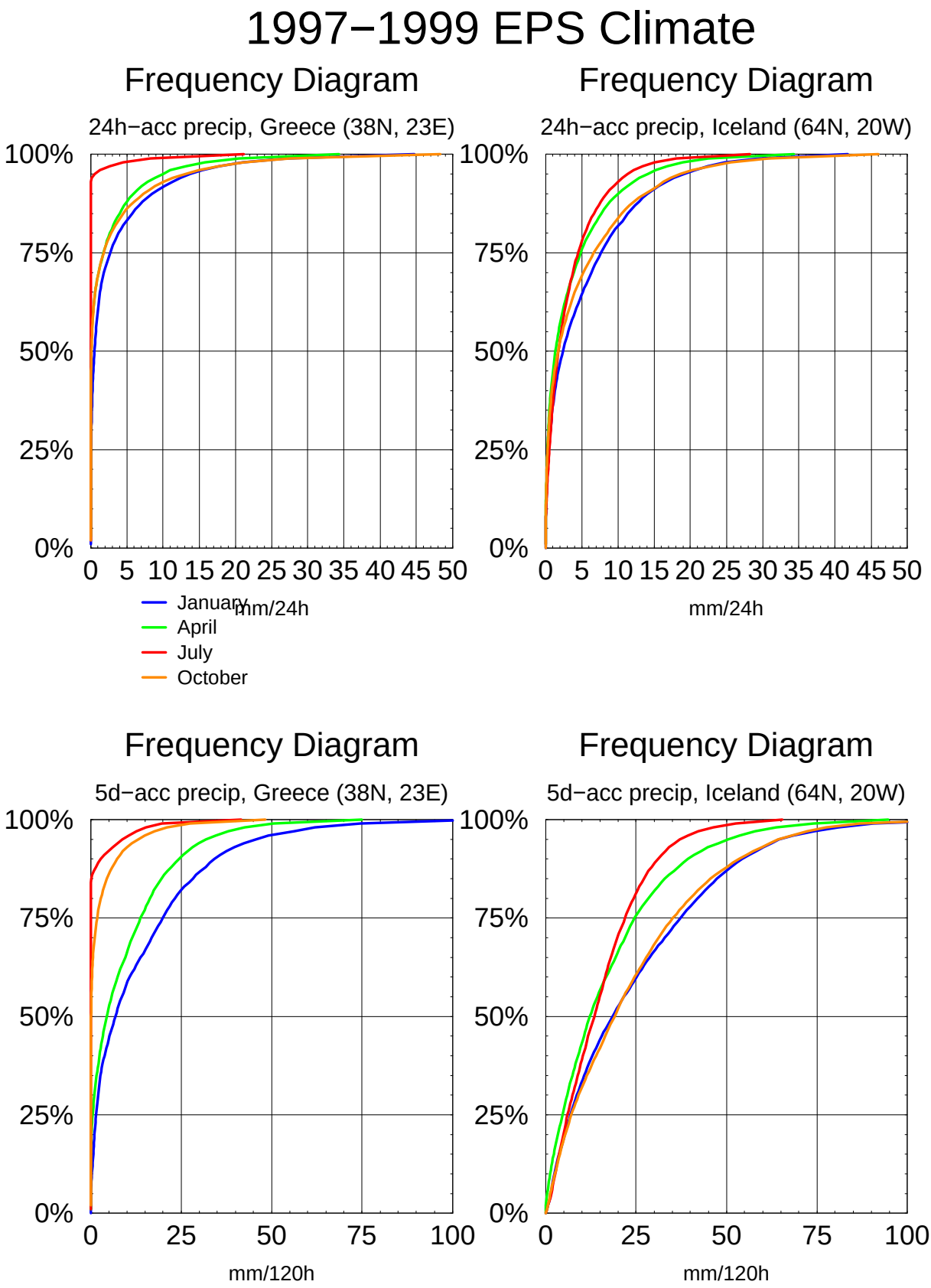

Figure 5: Pseudo-climate distributions derived from EPS T159 operational forecasts in 1997-1999. In all panels, red is for July, orange for October, blue for January and green for April. Left column is for Greece , right for Iceland. Upper Row: 24h-precipitation, lower : 120h-precipitation

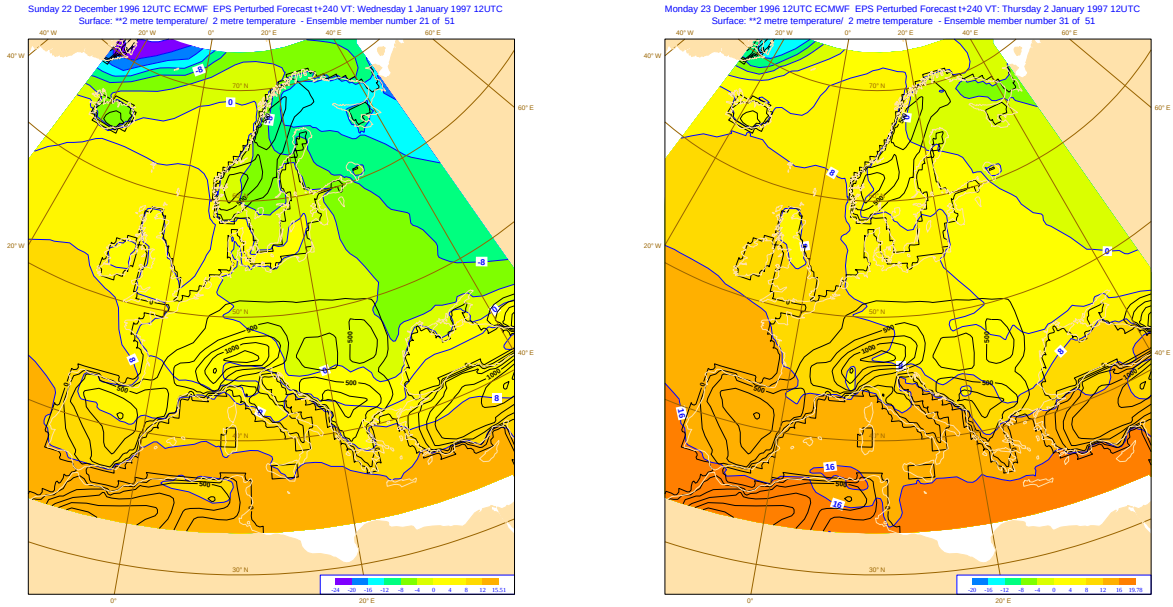

Figure 6: 1997-1999 EPS pseudo-climate quantiles for 2m temperature at 12UTC in January: left is Q20 (unusually cold), right Q80 (unusually warm); contours are every 4K; the model T159 orography is also reported every 250m (black contours). 2m temperature is brought to mean sea level by applying a standard correction based on altitude (6.5K.km-1).

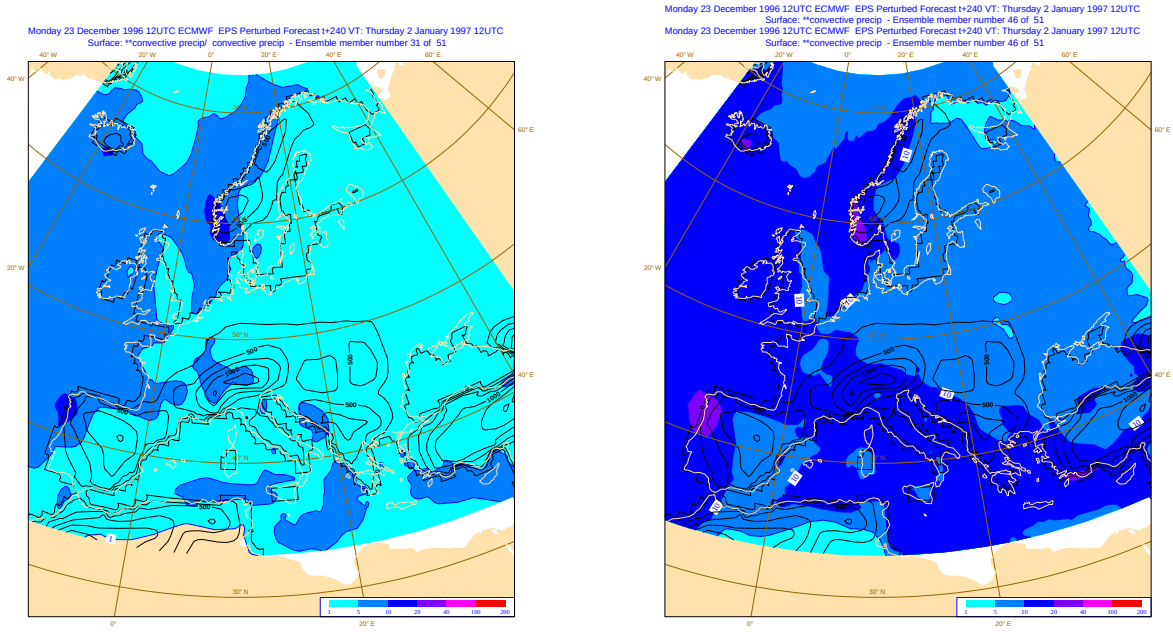

Figure 7: 1997-1999 EPS pseudo-climate quantiles for daily precipitation in January: left is Q80, right Q95; contours are 1, 5, 10 and 20mm; the model orography is also reported every 250m (black contours)

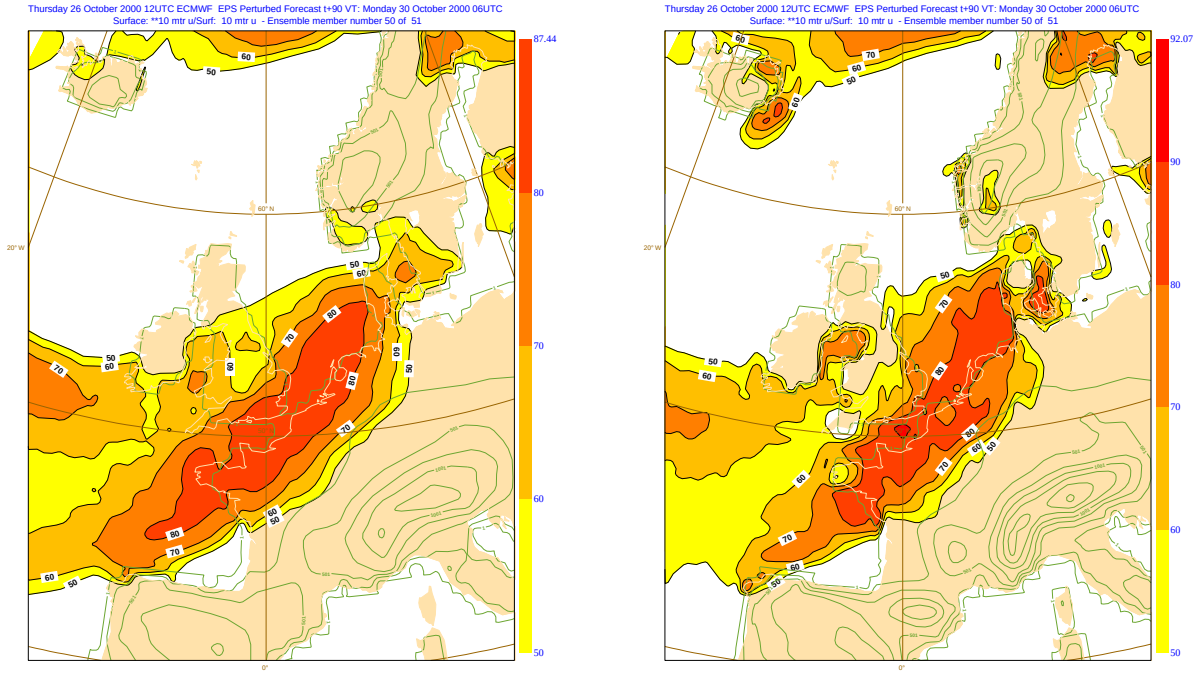

Figure 8: Comparison between two EFI_2 fields from two 90h forecasts valid on the same date (30 October 2000 06UTC, "Halloween Storm"); left panel: T159 EPS forecast; right panel: T255 EPS forecast; both use the same 1997-1999 T159 pseudo-climate reference

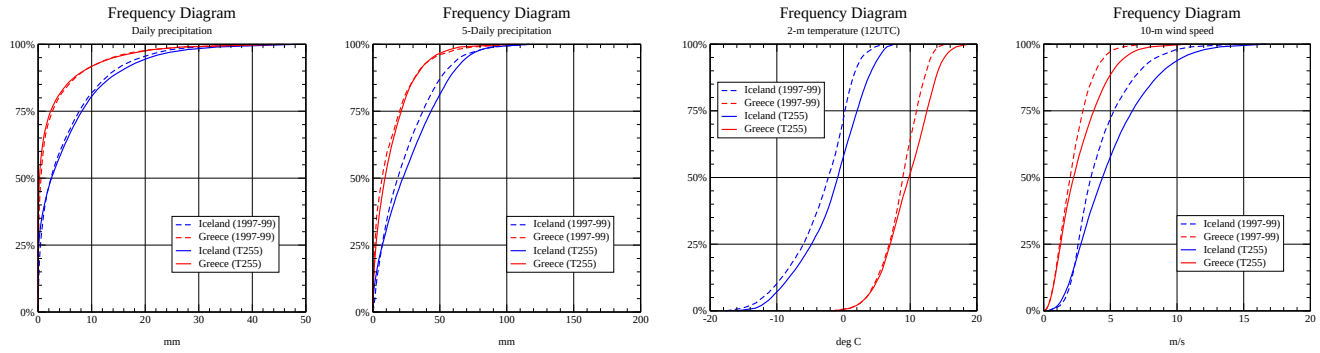

Figure 9: Changes in the pseudo-climate distributions resulting from the move from T159, 1997-1999 January samples to the T255, December-January-February 2001 samples. From left to right: Daily and 5-daily precipitation, 12UTC 2m-temperature and 10-m wind speed

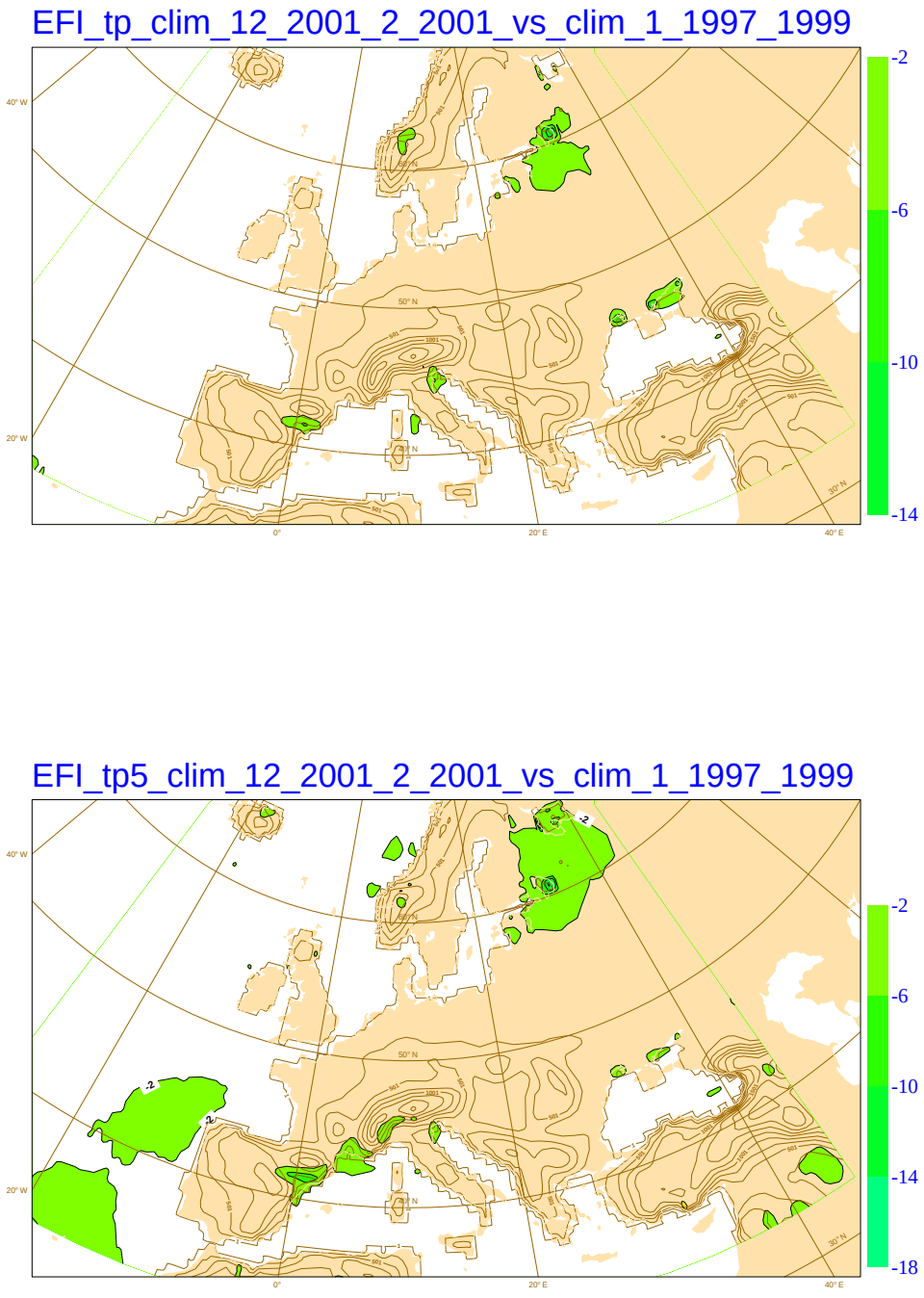

Figure 10: EFI-like *difference between the 1997-1999, T159 January distribution and the January-February-December 2001, T255 ones; daily precipitation (upper), 5-days accumulation (lower), orography is at T255 resolution every 250m*

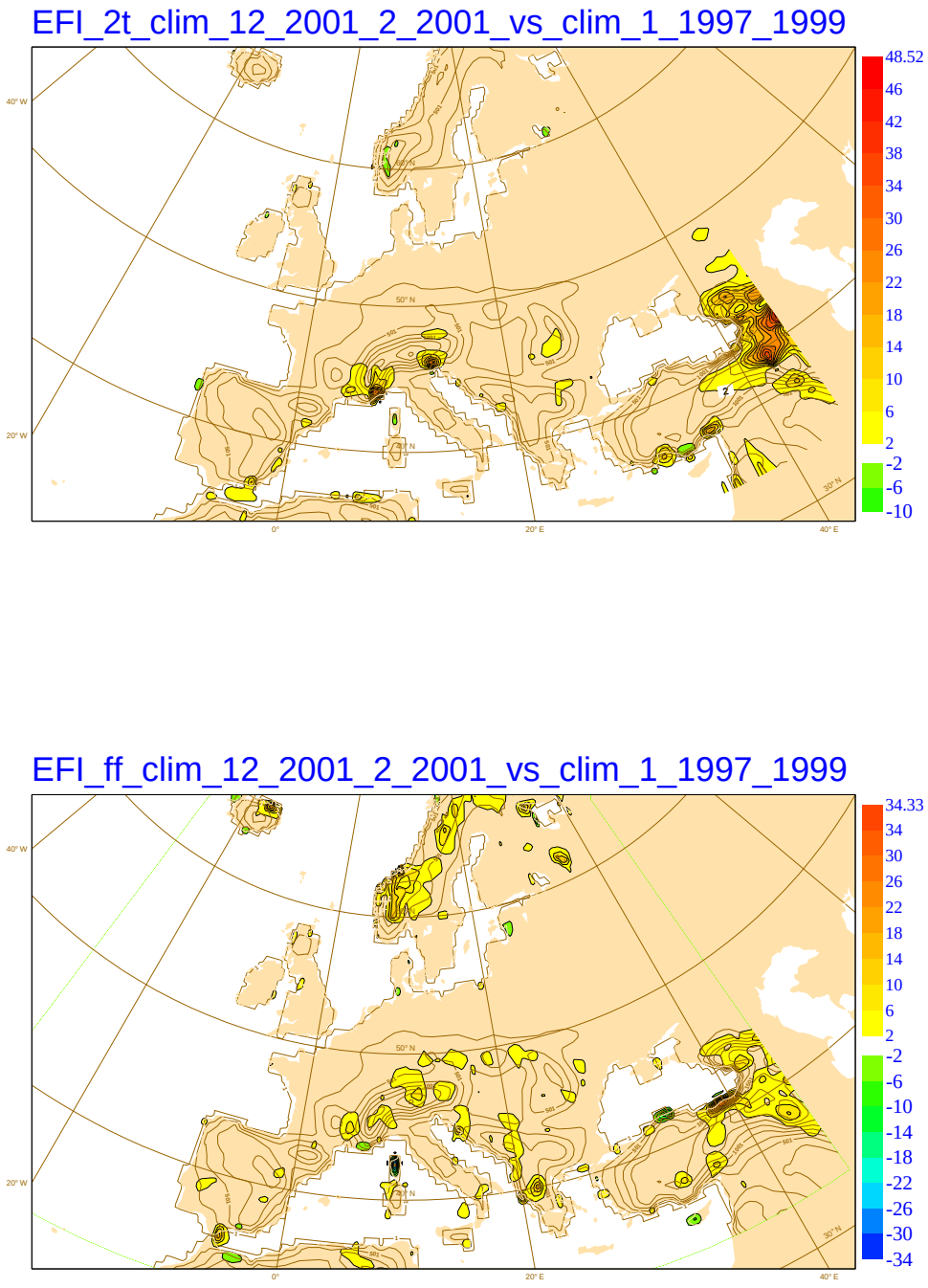

Figure 11: EFI-like *difference between the 1997-1999, T159 January distribution and the January-February-December 2001, T255 ones; 12UTC 2m-temperature (upper, next page) and 12UTC, 10m average wind speed (lower, next page); contours are every 4; orography is at T255 resolution every 250m*

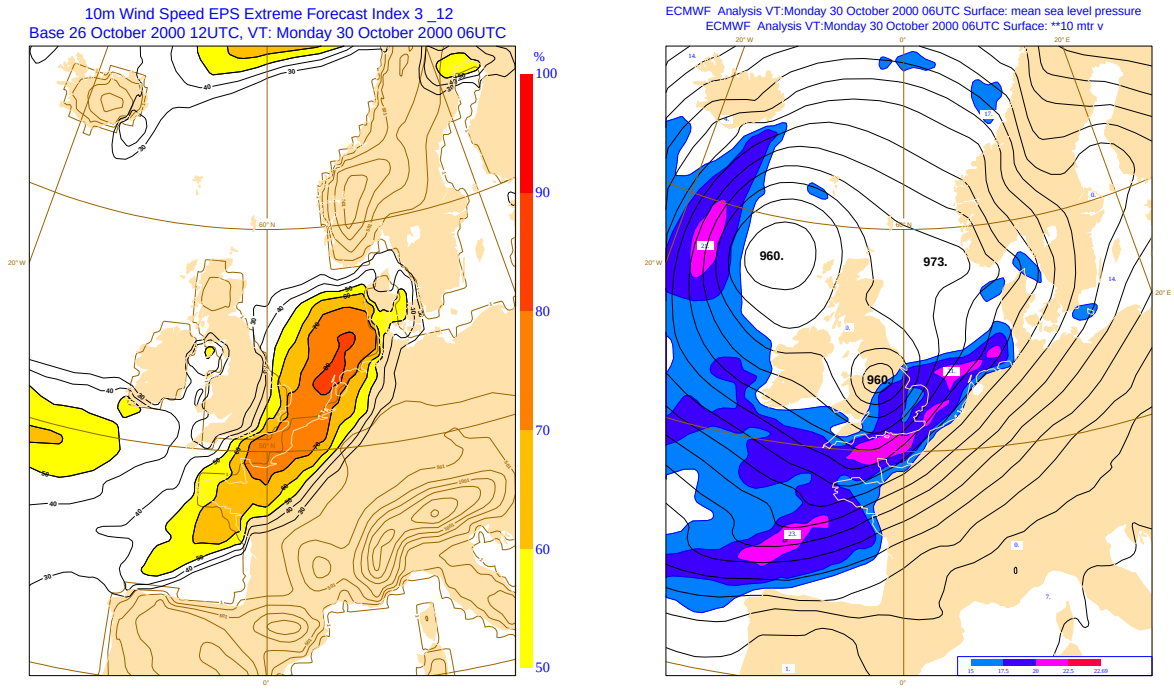

Figure 12: Upper panel: EFI_3 field for 10m wind speed on the 30 October 2000 06UTC from the T255, 90h EPS forecast using the T255 pseudo-climate as a reference ("Halloween Storm", same case as Fig. 8); lower panel: Analysis (mean sea level pressure in black contours every 4Pa and 10m wind speed shaded beyond 15 m/s every 2.5 m/s)

The procedure used here by no way is expected to be optimum. Indeed, we hope that we will be able to move from pseudo-climate to "true model climate" distributions rather soon using ERA-40 data. The strategy used here is mainly used as our best estimate that we found achievable at a reasonable data handling cost. In order to find a broad estimate of the uncertainty attached to the pseudo-climate data used here, EFI-type distances can be used. Instead of applying Eq. (3) to one climate and one EPS forecast distribution, we can apply it for two different estimates of the climate distribution. This has been done in Fig. 10 and 11. Although differences exist, they mostly are in the range +/- 2% (only values out of this range are shaded). The differences for rain (Fig. 10) are mostly in the direction of a dryer model climate, which happens mostly at the foothill of mountain ranges that are better resolved in the T255 than T159 climate. The main signature is in the Ebro valley in Spain, although this can also be found in the Po (Italy) and the Rhone valleys (France). Differences in 2m-temperature (Fig. 11) show noisy patterns that are located over steep orography and some coastlines that are differently handled in the two model resolutions. The same can be said for wind speed, although meteorological signatures show up over the straight of Gibraltar, in the French Rhone valley where steeper orography generates more reallistic channelling effects; on the opposite, Corsica and Cyprus appear in T255 that were missing at T159, and the local wind distribution is therefore slower in the T255 climate.

The model pseudo-climate is anyway used here only to detect conditions when the forecast deviates markedly from it. The previous figures indicate that the differences that can be spotted from different periods and model resolution are usually small compared to such deviations - and when they are not, it is our belief that this is mostly due to true differences in the model climate resulting from different orography and land sea mask, more than interannual variability. This of course is probably true only in first approximation, and the collection of more representative samples of the model climate remains a goal to achieve in the future. Fig. 12 illustrates

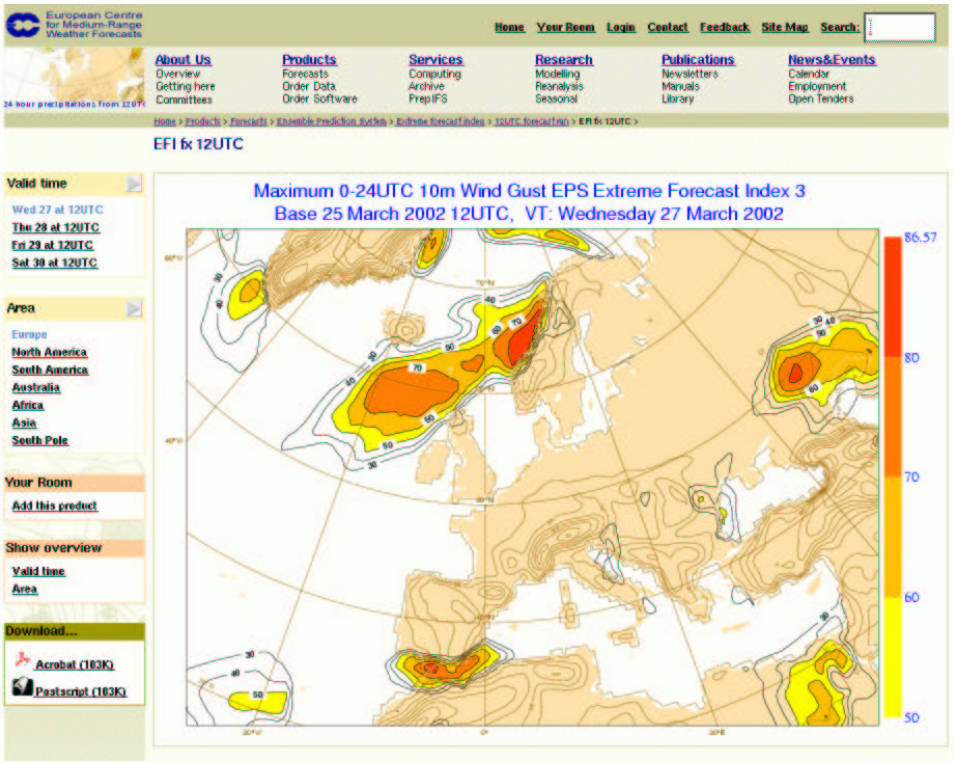

Figure 13: Web page showing EFI-related information provided to ECMWF Member States (`http://www.ecmwf.int/products/forecasts/d/charts/eps/efi`*)*

the result of the new processing (T255 pseudo-climate and use of EFI_3 instead of EFI_2) for the Halloween storm case already discussed in Fig. 8: although both the climate reference, the forecast model and the index definition have been modified, Fig. 8 and Fig. 12 bring essentially the same signature showing the robustness of the procedure. It can be stressed however that the move of the maximum of the index towards the Northern Sea and the Netherlands appears to be more reallistic, as the storm was heading for the Netherlands at that time.

4 EFI production : examples

EFI_2 data have been generated daily over Europe ($33°N - 73.5°N, 27°W - 45°E$) since April 2000 on a 0.5° resolution grid using the T159, 1997-1999 pseudo-climate. Since December 2001, the experimental production is global on the model reduced Gaussian grid using the EFI_3 formulation and the T255 "bootstrapped" pseudo-climate. Since that time, the data are generated twice per day using the experimental 00UTC run in addition to the 12UTC operational run. Products have first been made available to Member States on an informal, bilateral agreement basis, and since January 2002, they have been pushed on ECMWF web server with an access restricted to ECMWF Member States and Co-Operating states (Fig. 13). All results have been archived as Grib files, although technical work still needs to be done to give them a proper description (GRIB headers), make use of the operational (MARS) archive and include the products in the list for operational daily dissemination to ECMWF Member States and Co-operating states.

4.1 Christmas 1999 storms

Storms that did hit France, Germany, Switzerland and Italy between Christmas 1999 and New year have been the most devastating since several years, and are likely to have return period of several decades in France

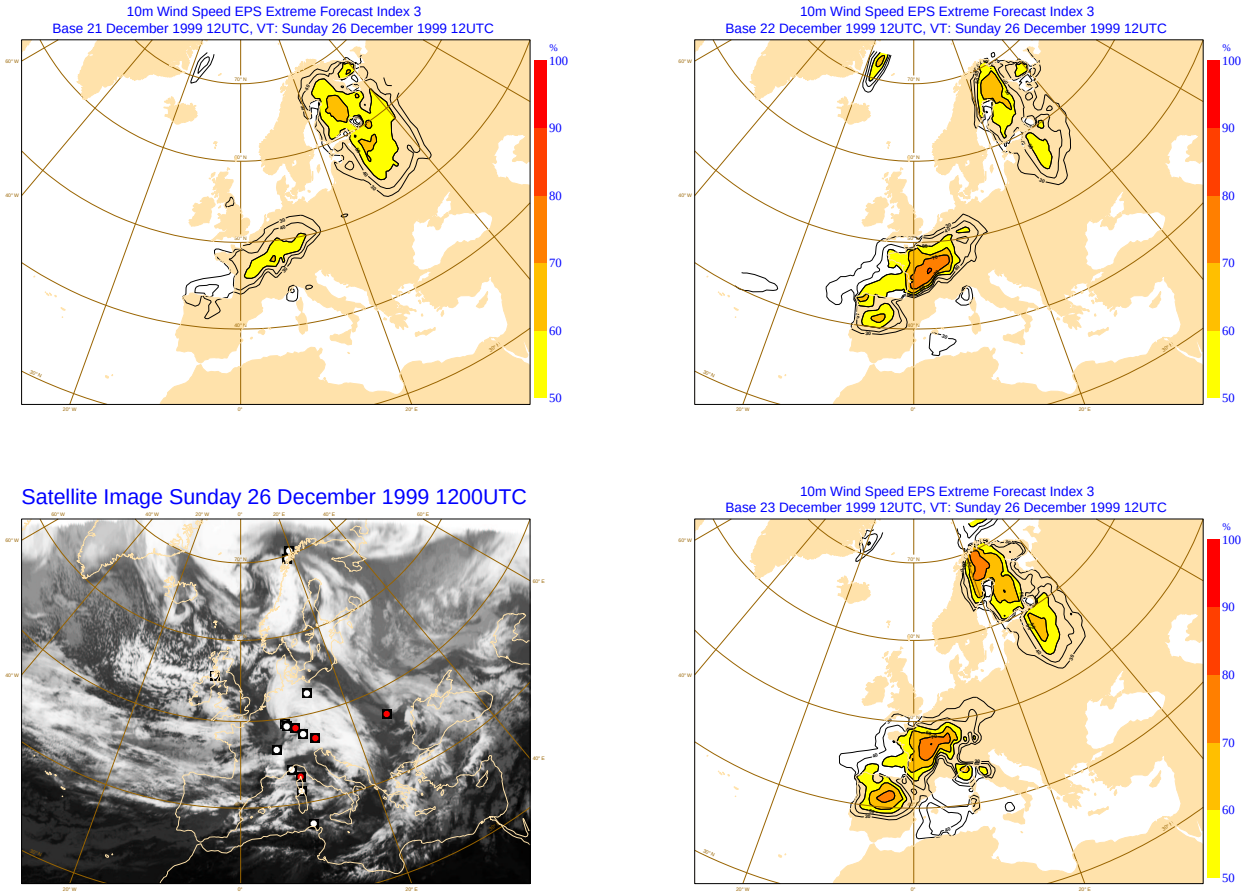

Figure 14: 10m wind speed Extreme Forecast Index (EFI_3) maps valid for the first December Storm (Lothar) on 26 December 1999 12UTC (clockwise from top left: 120, 96 and 72h forecast); lower right: Meteosat infra-red image together with SYNOP observations reporting wind beyond 25m/s (white) and 30m/s (red)

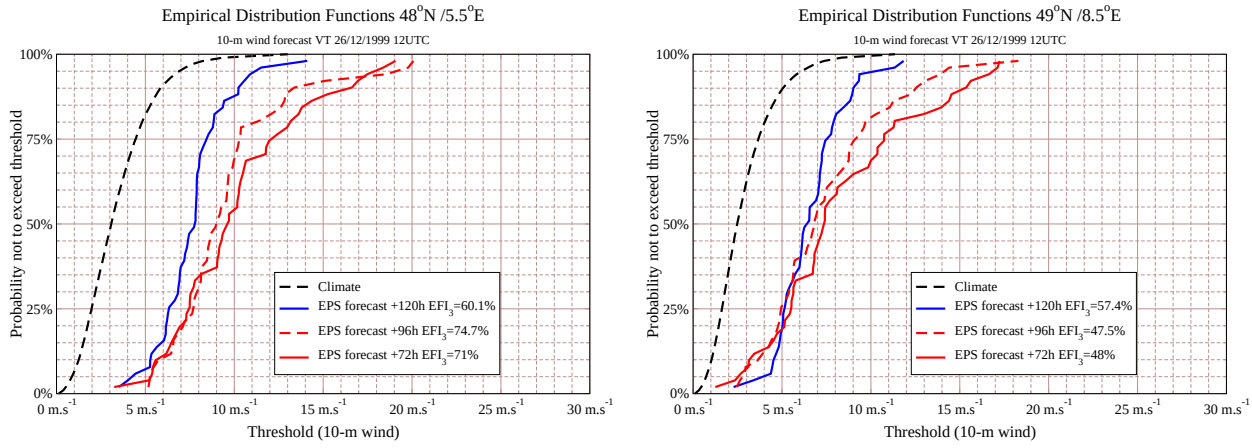

Figure 15: Pseudo-climate (black dashed) and EPS wind forecast distributions valid for 26 December 1999 12UTC near Langres (France, upper panel) and Karlsruhe (Germany, lower panel)

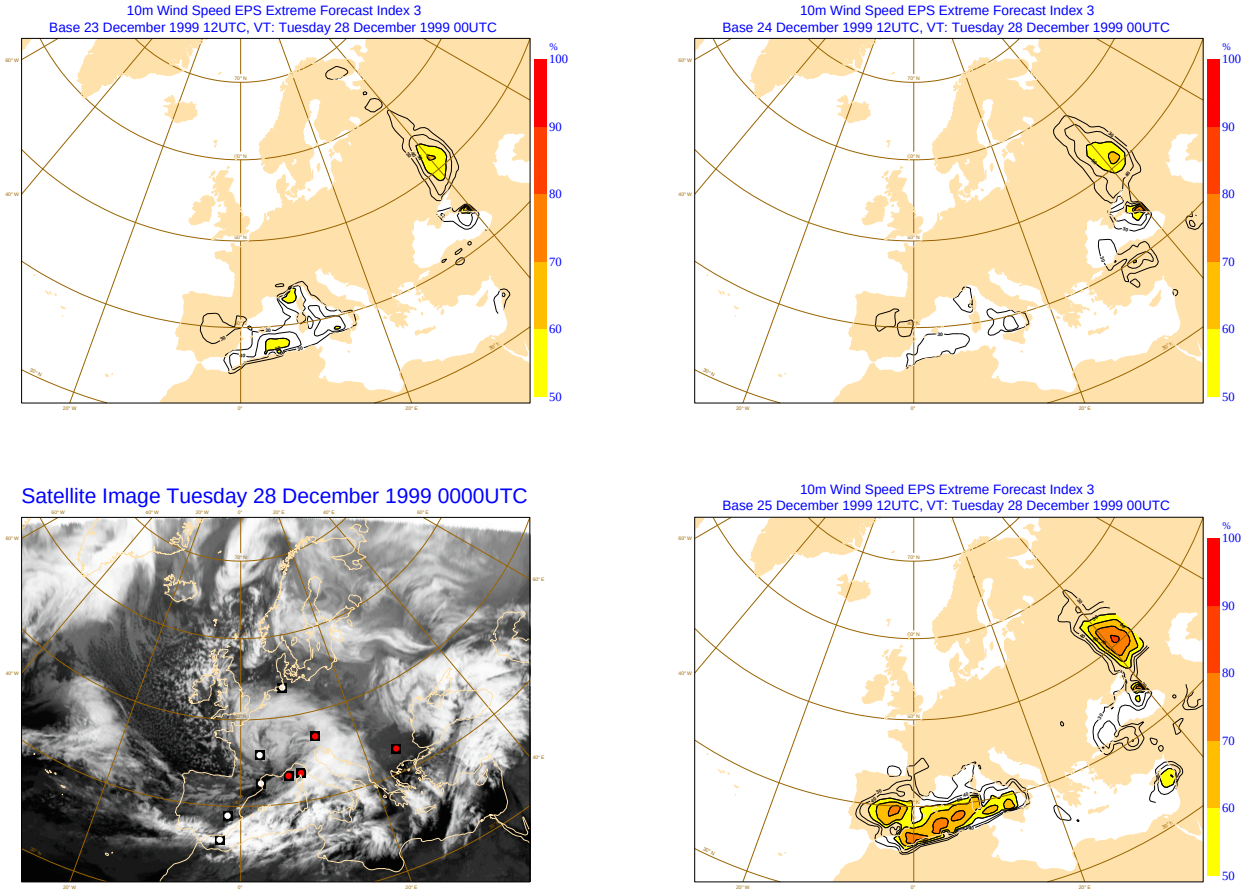

Figure 16: 10m wind speed Extreme Forecast Index (EFI_3) maps valid for the second December Storm (Martin) on 28 December 1999 00UTC (clockwise from top left: 114, 90 and 66h forecast); lower left: Meteosat infra-red image together with SYNOP observations reporting wind beyond 25m/s (white) and 30m/s (red)

(Balleste et al., 2001). The EPS at this time was run with resolution T159, and post-processing of EPS fields were only available every 12h (currently, surface parameters are retrieved every 6h). Wind gusts were not available at this time either. Maps to be shown for this case use the 1997-1999 version of the pseudo-climate, and therefore values are restricted to Europe ($33°N - 73.5°N, 27°W - 45°E$). The series of D5, D4 and D3 EFI maps valid for 26 December 12UTC are presented in Fig. 14. On this figure, both the Meteosat image and wind SYNOP reports when exceeding 25m/s are also shown. It can be seen that the areas signalled as potentially at risk from unusually strong winds were reasonably well identified as far as 120h hours in advance, although they are too far to the the West compared to the one that was actually hit: by 12UTC , the strongest sustained winds were to be found in the Karlsruhe area in Germany. Looking in more details at the distribution of EPS winds near Langres and Karlsruhe (Fig. 15) shows that sustained winds were always below 20m/s in the EPS when the value found at this time in Karlsruhe SYNOP report was 25m/s. However the proportion of EPS members exceeding all values found in the 1997-1999 pseudo-climate were around 10% in the forecast from 22 December and 20% in the one from 23 December. This translates into EFI values that although not as unusual as the storms themselves would have been potentially useful as an early warning.

Not all cases are as successful - indeed, the forecast of the second storm (Martin) two days later clearly missed the event (Fig. 16). Although very strong winds have been associated to this storm during the evening of 27 and the following night over the southern part of France, northern Italy and Switzerland (25m/s sustained wind speed was reported at 00UTC on the 28th in Clermont-Ferrand SYNOP message), there was no clear signal to

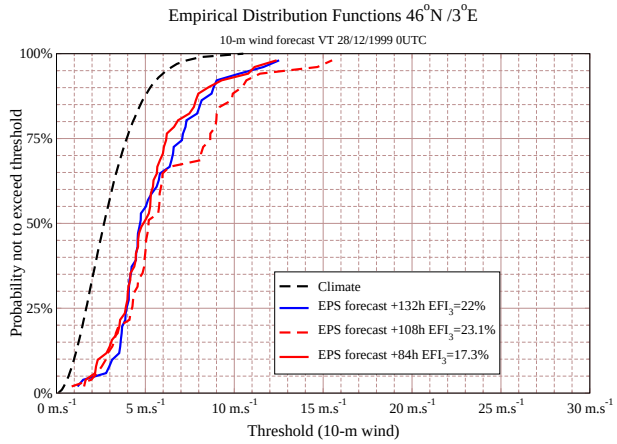

Figure 17: Pseudo-climate (black dashed) and EPS wind forecast distributions valid for 28 December 1999 00UTC near Clermont-Ferrand (France).

be seen in the EFI maps - only the active cold front associated with the system did generate EFI values beyond 50%. EPS wind forecast distribution near Clermont Ferrand valid for the same time are also shown in Fig. 17, which confirm that only very few EPS members predicted any significant wind in this area.

4.2 Piedmont floods

Severe floods in the Italian Alps Piedmont are relatively common and precipitating events leading to them have been investigated recently during the MAP field experiment (Bougeault et al., 2001). The event that happened in mid-October 2000 has however been one of the most severe experienced in the area for years: it had led to the overflow of the Maggiore lake, the Po river reaching historic heights, and severe damages were caused to a very large portion of the Po valley (Hollingsworth et al., 2002). This period happened to be during the pre-operational testing of the then new, now operational version of the EPS that among other things, brought the horizontal resolution from T159 (around 120km) up to T255 (around 80km). More details on the testing of this version with application to severe weather prediction can be found in Buizza and Hollingsworth (2002).

Both EFI maps generated from the T159 and T255 version are shown in Fig. 18. Because the EPS used both in operations and in pre-operational testing suffered from two numerical problems affecting the stochastic tendencies and the initial spread of the ensemble perturbations (Lalaurette and Ferranti, 2001), the forecasts have been rerun at both T159 and T255 resolution in their ”bug-corrected” version. Clearly the new, higher resolution model is better at capturing the sharp enhancement of precipitation generated by the south-easterly, warm and humid air mass when lifted by the Alps. A closer look to the distributions of 5-days accumulation of rain near Locarno (Switzerland, Fig. 19) reveals that the enhancement of the EFI signal from the T159 to theT255 version of the EPS is not only due to an increase in the amount of predicted rain, but also in a different climate. Although part of these differences may be related to the different time samples used, the dryer T255 climate is to be expected in that area. Indeed westerly weather regimes are likely to generate a better defined Foehn effect on the lee side of the Alps in the high resolution EPS - an effect that has been found on the lee side of most mountain ranges (not shown). Another feature is that although almost the entire T255 climate distribution is drier, the tail of the distribution extends into larger amounts of rain with T255 than T159: the 0.1% highest cumulative amounts (around 10 cases) exceed 144mm in the T255 climate instaed of 128mm in the T159 one. Such a signature is an indication that the sharper orography has allowed a more vigourousforcing on the Souhthern side of the Alps in more that the one case ilustrated here.

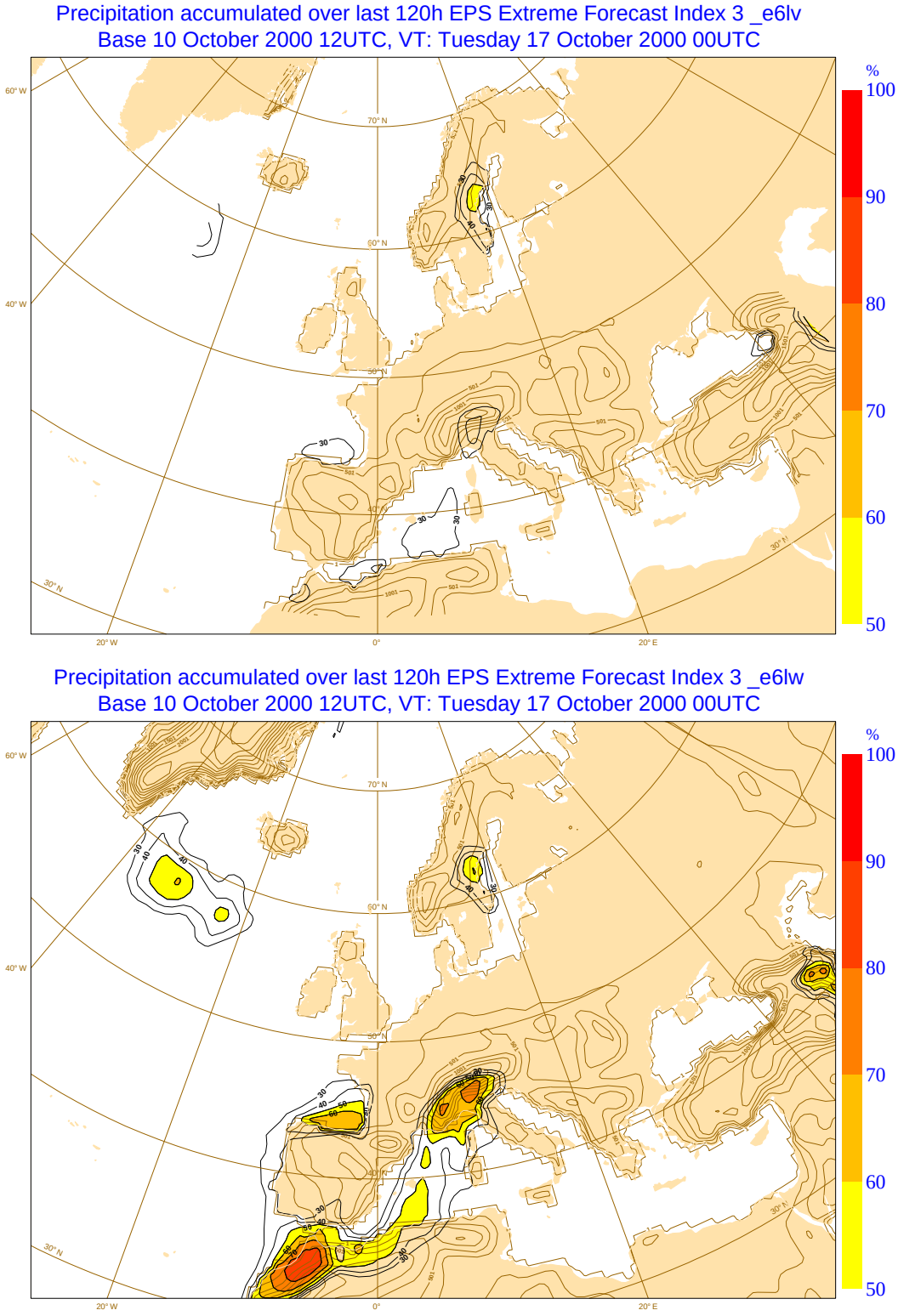

Figure 18: EFI_3 for accumulation of rain from 20001012 00UTC to 20001017 00UTC. Upper is T159 (120km) , lower T255 (80km) version. In both cases, the model orography is reported

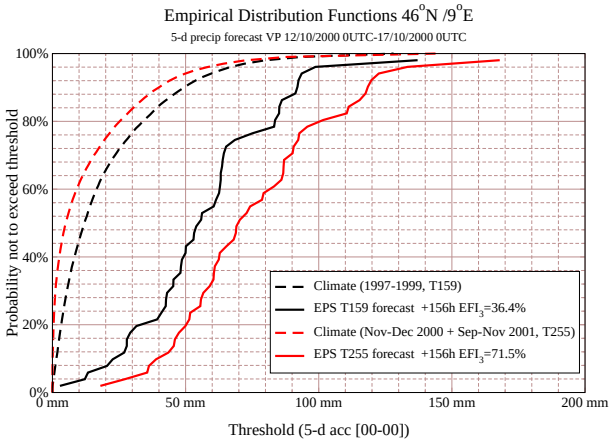

Figure 19: Pseudo-climate (dashed) and EPS five-days rain accumulation distributions valid for 20001012 00UTC to 20001017 00UTC near Locarno (Switzerland). Black: T159 (120km) , red: T255 (80km)

5 Preliminary Verification results

It is out of the scope of this paper to address the issue of the verification of the EFI comprehensively. Nevertheless it is needed to go beyond the subjective assesment of a small number of cases that has been made in the previous section, however representative these cases were of severe weather conditions. Because only a small sample could be illustrated, we had to focus them on events that did occurr, therefore biasing the view towards hit rates rather than false alarm rates. We will try to provide here some indications of the rates of success and failure of early warnings of severe weather that would be only based on the value reached by the EFI. The difficult challenge in dealing with the verification of severe weather conditions is to collect verification data. Although current global model analyses provide very good estimates of the mass and temperature fields in the troposphere and lower stratosphere, they usually fail to provide a reallistic, quantitative picture of the most active weather systems. On the other hand, fine scale analyses of the weather parameter geographical distributions are still a goal to achieve, even in non-real time.

As a first attempt to quantify the medium range forecast skill that can be achieved using the type of probabilistic index introduced here, and keeping in mind the limitation of the approach, we will use the model analysis or short range forecast as our verification. During the 2001-2002 cold season (December to April), we have used the EPS Control 2m-temperature at step 0 (12UTC run), the maximum wind gust between step 0 and step 24h (00UTC run), and the accumulation of rain beween step 6 and 30h (00UTC run) as our verification. These events have been ranked with respect to the local pseudo-climate. Rather arbitrarily, we have set-up the event we want to forecast as being in the top 5% of the climate records (Q95). This is certainly not enough to be qualified as an extreme event - this is an event that, at any given location, should happen on average every 3 weeks. We have kept the definition on a rather conservative side both to build a verification sample more quickly, and because for the most extreme events the use of model proxies instead of observations would be totally innapropriate. Because negative anomalies also make sense for 2m-temperatures, we have also defined a cold event as being in the first 5% (Q05).

The information related to the verification of categorical forecasts can conveniently be summarised in the form of contingency tables giving the number h of hits (yes/yes), f of false alarms (yes/no), m of missed events (no/yes) and z of zeros (no/no) where the first value refers the forecast, the second to the observation. Although such information is usually only shown averaged over an area, we show in Fig. 20 the geographical maps of occurrence, hit, false alarms and misses associated with forecasts based on the $EFI_3 > 50\%$ and 30% respectively, in the 90-114h range. The frequency of occurrence (upper left panel) shows to be around the expected value (5% of the 151 days used for verification means that the event should have happened between 7

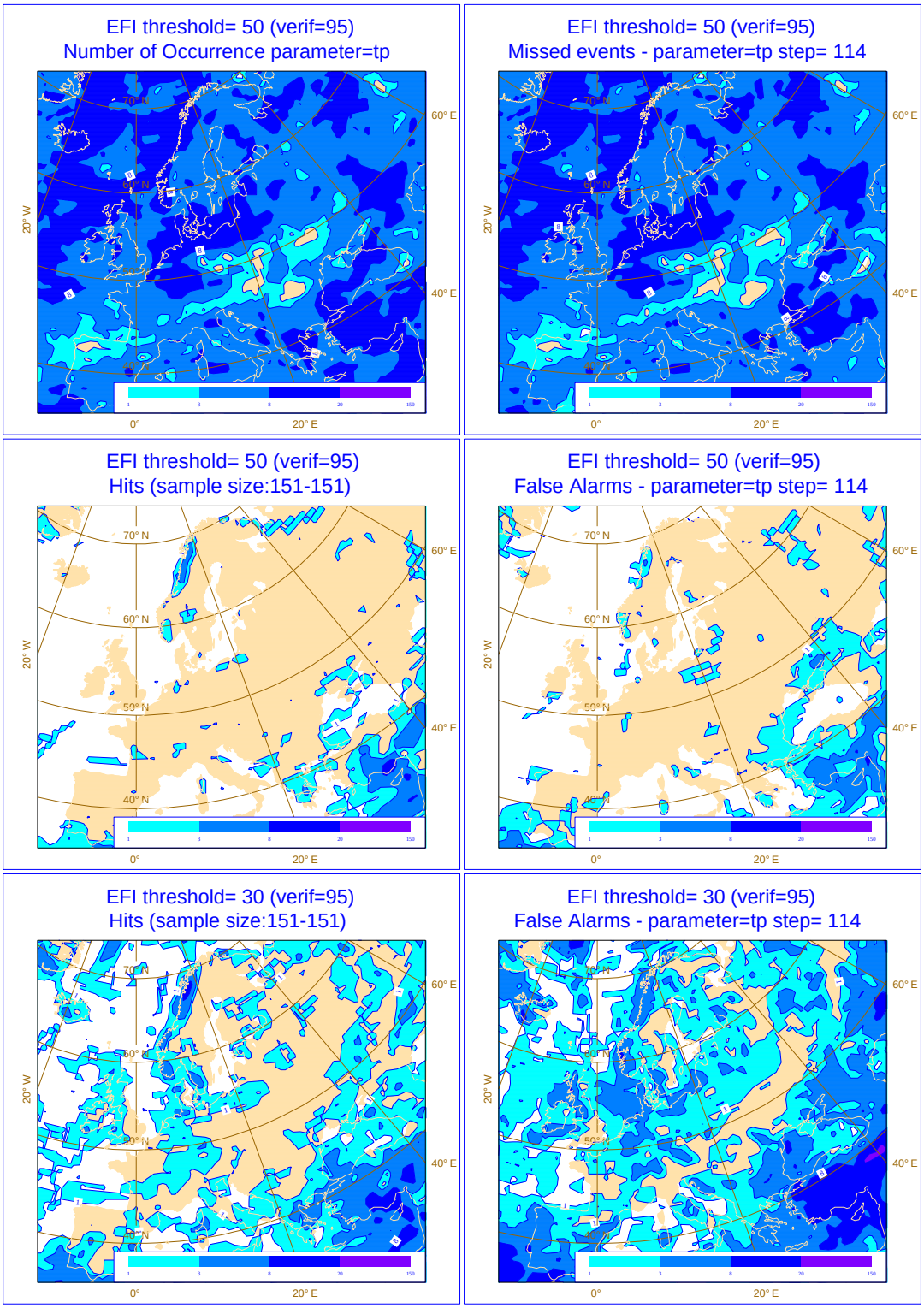

Figure 20: Verification of 12UTC EPS EFI warnings that daily precipitation will exceed Q95 (5% largest climate amounts) in the 90-114h range; Top left: number of occurrences in the verification (00UTC EPS control, 6-30h accumulation, December 2001-April 2002); middle row: number of hits (left) and false alarms (right) when the event is forecasted whenever the EFI_3 exceeds 50%; lower row: same as middle row, but forecasts are issued whenever $EFI_3 > 30\%$; Top right: number of misses associated to $EFI_3 > 50\%$; shading classes are 1-3, 3-8, 8-20 and 20-150 cases.

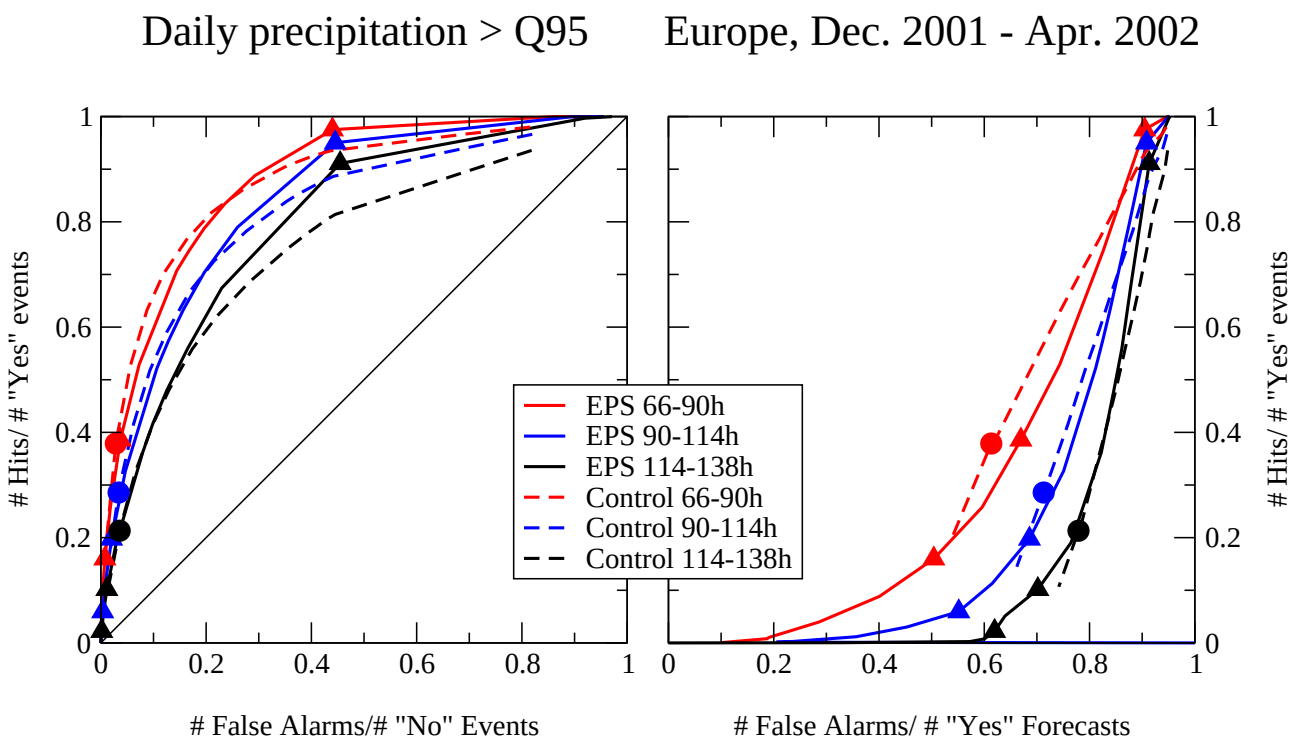

Figure 21: ROC (left) and modified x-axis ROC (right) curves for the forecast that daily precipitation amounts will exceed 95% of the pseudo-climate records; full lines: EPS, dashed lines: Control; each point of the curve corresponds to a different EFI threshold to discriminate a "yes" from a "no" forecast (in the case of the Control, these different thresholds correspond to different precipitation amounts of the single-value forecast -see Fig. 2). Triangles on the EPS curves mark 50%, 30% and 0% EFI_3 thresholds, while circles on the Control curve mark the EFI_3 deterministic threshold corresponding to Q95 ($\simeq$ 81.45% -see Fig. 2 or Eq. 4)

and 8 times). That this value is pretty uniform validates our pseudo-climate approach: using the EPS forecasts over only two years before the verification period seems to provide a reasonably accurate estimate of the Q95 threshold. The other maps are rather disappointing: the number of hits (middle row, left pannel) is quite small compared to the number of events, which is reflected by the large number of misses (top right). The number of false alarms (middle right) however is kept to a small number, which is approximately the same order of magnitude than the number of hits. In order to improve on the detection of events (hits), the threshold on which the forecast is based can be relaxed: results can be seen in the lower row of the same Fig. 20 , where instead of 50%, the threshold has been lowered to 30%. As expected, more events are detected at this level, a result that is achieved however at the expense of a higher number of false alarms.

A convenient way to summarise the information related to the quality of a forecast system that can be adjusted to the user's requirements with a variable decision-making parameter is to compute Relative Operating Characteristics (ROC) curves (see Mason and Graham (2002) and references hereof). ROC curves for the verification of the EPS forecasting 90-114h precipitation beyond Q95 in December-April 2002 are in Fig. 21 (full lines, left panel): each point on this diagram corresponds to a different threshold of the EFI_3. The performance of a pure deterministic forecast (EPS Control) is also reported in this diagram (dashed lines, left panel). This diagram is therefore a clean comparison of a purely deterministic versus a probabilistic approach: only the distance between the forecast distribution (single valued for the deterministic, multiple value for the ensemble) and the climate distribution are used to discriminate "yes" and "no" forecasts. In the deterministic case, this distance is only depending on how the unique forecast value ranks in the climate (see Fig. 2 or Eq. 4), while in the ensemble, probabilistic case, the distance depends on all members (both their median value and spread are important). Such verification procedures are needed to demonstrate the benefit of a dynamical-statistical approach compared to a purely dynamical, deterministic one - more on this subject can be found in (Atger, 1999, 2001).

The ROC curves show that there is a good level of skill in both the deterministic and ensemble forecasts - and the most the forecast goes into the medium range, the most beneficial the ensemble is: the dashed line lies entirely below the full one at day 5, while at day 3 there is a range of thresholds when the control has higher hit rates (for a given false alarm rate) than the EPS. However to achieve such a high hit rate, the EFI threshold has to be very low: the three triangles in the top part of the figure are all related to $EFI_3 \geq 0$, which means that the "yes" forecast is issued every time the forecast distribution is more weighted towards positive than negative anomalies!

The ROC curves exhibit a lot of desirable properties, one of the most remakable being that the distance of each point from the diagonal is exactly the Hansen-Kuipers discriminant or True Skill Statistics (Richardson, 2000) which is itself quite a remarkable, equitable deterministic score (Wilks, 1995). In the case of extreme weather events however, that the x-axis is in terms of False Alarms per "No" event ($\frac{f}{f+z}$) instead of the most common False Alarms per "Yes" forecast ($\frac{f}{f+h}$, e.g. Wilks, 1995) is a severe limitation to the readiness of these ROC diagrams. This is because extreme events are such that $z >> h$ (much more "no" events than "yes" forecasts), and therefore the FAR axis on the ROC curves is emphasizing False Alarms ranges that most users will consider impratical. To illustrate this point, ROC curves are shown with False Alarm per "Yes" forecast ($\frac{f}{f+h}$) as the x-axis in the right panel of Fig. 21. They confirm that high hit rates are only achieved at the expense of very high numbers of false alarms - although it remains true that the forecasts remain more skillful than a pure random forecast that would be a false alarm in 95% of the "yes" forecasts by definition of the Q95 event. False alarms are more in the range 60 to 80% when using the model forecasts instead. Another property of the probabilistic forecast can be assessed by looking at how the forecast characteristics change with the forecast range for a given EFI threshold (triangles for the EPS, circles for the Control in Fig. 21). Using deterministic forecasts leads to a sharper increase of false alarms and a slower decrease of hit rates with forecast ranges than using the EPS: the number of false alarms decreases with forecast range. This is because the EPS forecast probability distribution is converging towards climatology, and therefore cases with large departures from the climate are less frequent when the forecast range increases. On the other hand, a good deterministic forecast system should generate severe weather events with the same frequency at all forecast ranges, but the longer the range, the less skillful these forecasts are. It remains true however that for all ranges of thersholds by day 5, the EPS has better characteristics than the Control. The situation becomes more threshold dependant when going into the shorter ranges.

6 Summary and Perspectives

A new procedure to detect large departures of the early medium range ensemble forecast from normal model behaviour has been presented in this paper. It has involved both the design of a new Extreme Forecast Index to scale objectively these departures, and the collection of pseudo-climate records from a large sample of model forecasts. The objective is twofold:

1. to provide forecasters and users of the medium range forecasts with an objective early warning.
2. to set-up a fully objective early warning procedure that can be used as a poor man reference against which any operational procedure can be assessed

As a tool for the forecaster, this of course is only a starting point for further investigations which should involve a closer look at ensemble distributions in the area where and when the EFI gets high values. This not only means looking at the full EPS local distributions, but also at meteorological scenarios in a more traditional way, e.g. by selecting the most extreme forecasts and to follow them in time in order to be prepared to the occurrence of "worse case scenarios" - e.g. by looking at satellite imagery to detect the characteristic signatures of such scenarios.

Work obviously needs to be done in the direction of a more consistent evaluation of the EFI forecast skill. A strong incentive for this work has been the common criticism among users of the numerical forecast models that too much time is spent on broad scale, average features of the forecast quality and not enough on the ability of the model to detect ”outliers” of the climate distribution. To push it a bit, one thing is to be good on average, which means mostly in cases when no significant weather happens, and another to provide useful guidance when human lives or properties are at risk. This paper is expected to be a contribution into the direction of more stringent procedures of verification of numerical forecasts in severe weather context. The use of the Extreme Forecast Index is however not expected to pave the way for fully automated early warning procedures: results shown in Figure 21 show that error characteristics are such that most users would find it difficult to stand the false alarm rates generated, even if the same results show a significant reduction of the false alarms compared to a random, no-skill system. The fact that a level of skill has been demonstrated indicates that there are potentially applications that could derive a benefit from its knowledge, and indeed some of ECMWF Member States have already considered to use it as part of their internal early warning procedures.

The perspectives for improvement and development of the system in the future are many. The weakest link in the process described here is almost certainly the pseudo-climate. We hope that in 2003, a ”real” EPS climate will be built using ERA-40 analyses over a period of 30 years (1970-2000). Although tests have been conducted that indicate that the EFI is not too sensitive to changes in the climate reference (not shown), this is to some extend the result of some kind of censorship of our own. If the ERA-40 climate delivers distributions that can usefully be fitted near the tails where the most extreme weather events are, then using more selective formulations of the EFI can be tried by increasing the parameter n in the definition (Eq. (1)). This however is likely to request some kind of fitting of the EPS distributions themselves.

As any statistical post-processing method, and the EFI is rather heavily relying on statistical post-processing of long records of model forecasts, any development of new EFI parameters is demanding a significant amount of both computer and human resources. Although extensions such as minimum or maximum temperatures, oceanic wave heights or accumulation of rain over large hydrological basins have been mentioned by users as potentially valuable, it has not yet been decided to provide them.

Finally, the status of the EFI at the time of writing this paper is still very experimental. Operational production is however expected that will involve both archiving in MARS and dissemination of the products in real time. Ideally, forecasters should be provided not only with the EFI maps, but also with interactive applications that would provide them with pseudo-climate and EPS local distribution (as in Fig. 8 for example), and maybe with animations showing them how the most extreme scenarios highlighted by these distributions are occurring. In a nutshell, time seems to be ripe now to extract from the EPS not only information on the large scale weather regimes in the late medium range, but also more detailed information on the potentially hazardous weather phenomena that may occur in the early medium range. It is hoped that this paper will have opened a few perspectives in this direction for the near future of operational forecasting.

References

Atger, F.: 1999, The skill of ensemble prediction systems. *Mon. Weather Rev.*, **127**, 1941–1953.

— 2001, Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Proc. in Geophys.*, **8**, 401–417.

Ayrault, F., F. Lalaurette, A. Joly, and C. Loo: 1995, North atlantic ultra high frequency variability: an introductory survey. *Tellus*, **47A**, 671–696.

Balleste, M.-C., H. Brunet, A. Mougel, J. Coiffier, N. Bourdette, and P. Bessemoulin: 2001, *Les tempêtes exceptionnelles de Noël 1999*, volume 8 of *Phénomènes remarquables*. Météo-France.

Bougeault, P., P. Binder, A. Buzzi, R. Dirks, R. Houze, J. Kuettner, R. Smith, R. Steinacker, and H. Volkert: 2001, The MAP special observing period. *Bull. Am. Meteorol. Soc.*, **82**, 433–462.

Bouttier, F.: 1994, *Sur la prévision de la qualité des prévisions météorologiques*. Ph.D. thesis, Université Paul Sabatier, Toulouse.

Buizza, R. and A. Hollingsworth: 2002, Storm prediction over europe using the ECMWF ensemble prediction system. *Meteorol. Appl.*.

Der Megreditchian, G.: 1992, *Le Traitement Statistique des Données Multidimensionnelles*, volume 8 of *Cours et Manuels de l'École Nationale de La Météorologie*. Météo-France.

Gall, R. and M. Shapiro: 2000, The influence of Carl-Gustaf Rossby on mesoscale weather prediction and an outlook for the future. *Bull. Am. Meteorol. Soc.*, **81**, 1507–1523.

Ghelli, A. and F. Lalaurette: 2000, Verifying precipitation forecasts using upscaled observations. *ECMWF Newsletter*, **87**, 9–17.

Hello, G., F. Lalaurette, and J.-N. Thépaut: 2000, Combined use of sensitivity information and observations to improve meteorological forecasts: A feasibility study applied to the "Christmas storm" case. *Q. J. R. Meteorol. Soc.*, **126**, 621–647.

Hersbach, H.: 2000, Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecasting*, **16**, 559–570.

Hollingsworth, A., P. Viterbo, and S. A. J.: 2002, The relevance of numerical weather prediction for forecasting natural hazards and for monitoring the global environment. *A Half Century of Progress in Meteorology: A Tribute to Richard Reed*, R. H. Johnson and R. A. H. Jr, eds., ????

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell: 1996, Simulation approach to ensemble prediction. *Mon. Weather Rev.*, **124**, 1225–1242.

Joly, A., K. A. Browning, P. Bessemoulin, J.-P. Cammas, G. Caniaux, J.-P. Chalon, S. A. Clough, R. Dirks, K. A. Emanuel, L. Eymard, R. Gall, T. D. Hewson, P. H. Hildebrand, D. Jorgensen, F. Lalaurette, R. H. Langland, Y. Lemaî tre, P. Mascart, J. A. Moore, P. O. G. Persson, F. Roux, M. A. Shapiro, C. Snyder, Z. Toth, and R. M. Wakimoto: 1999, Overview of the field phase of the Fronts and Atlantic Storm-Track EXperiment (FASTEX) project. *Q. J. R. Meteorol. Soc.*, **125**, 3131–3163.

Kyouda, M.: 2002, Ensemble prediction system. *Outline of the operational numerical weather prediction at the Japan Meteorological Agency*, T. Tsuyuki and T. Fujita, eds., Japanese Meteorological Agency, Japan Meteorological Agency 1-3-4 Ote-machi Chiyoda-ku, Tokyo 100-8122 JAPAN, 59–63.

Lalaurette, F. and L. Ferranti: 2001, Verification statistics and evaluations of ECMWF forecasts in 2000-2001. *ECMWF Tech. Memorandum*, **346**, 55pp.

Mason, S. J. and N. E. Graham: 2002, Areas beneath th relative operating characteristics (ROC) and levels (ROL) curves: statistical significance and interpretation. *Q. J. R. Meteorol. Soc.*, submitted.

Molteni, F., R. Buizza, T. Palmer, and T. Petroliagis: 1996, The new ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73–119.

Palmer, T. N.: 2002, The economic value of ensemble forecasts as a tool for risk assesment: From days to decades. *Q. J. R. Meteorol. Soc.*, **128**, 747–774.

Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth: 1996, Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.*, **122**, 121–150.

Raoult, B.: 1996, Data handling via MARS at ECMWF. *ECMWF Newsletter*, **72**, 15–19.

Richardson, D. S.: 2000, Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **126**, 649–668.

Simmons, A. J. and J. K. Gibson, eds.: 2000, *The ERA-40 Project Plan*, volume 1 of *ERA-40 Project Report Series*. ECMWF, 63pp.

Szunyogh, I. and Z. Toth: 2002, The effect of increased horizontal resolution on the NCEP global ensemble mean forecasts. *Mon. Weather Rev.*, **130**, 1125–1143.

Toth, Z. and E. Kalnay: 1997, Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, **125**, 3297–3319.

Ward, M. N. and C. K. Folland: 1991, Prediction of seasonal rainfall in the north nordeste of brazil using eigenvectors of sea surface temperature. *Int. J. Climatol.*, **11**, 711–743.

Wilks, D. S.: 1995, *Statistical Methods in the Atmospheric Sciences*, volume 59. Academic Press.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne: 2002, On the economic value of ensemble based weather forecasts. *Bull. Am. Meteorol. Soc.*, submitted.