# Early indication of extreme winds utilising the Extreme Forecast Index

THOMAS I. PETROLIAGIS, PIERRE PINSON

The Extreme Forecast Index (EFI) was developed at ECMWF as a tool to provide forecasters with an indication of potential extreme weather events based on information from the ensemble predictions. Verification results (*Richardson et al.*, 2011) show that the EFI has substantial skill in forecasting extreme events several days in advance, confirming the subjective experience of forecasters in the Member States where the EFI is widely used. EFI skill is one of the six headline scores used to monitor long-term trends in performance of the ECMWF forecasting system (*Andersson & Richardson*, 2011).

The typical forecast lead time for the EFI has been the early medium-range (3 to 5 days). During this period, EFI predictions of an extreme weather event can be considered as an 'early indication'. Beyond day 5, the EFI may serve as 'alarm bells' resulting from the ability of the ensemble to capture the risk of very intense weather systems (possible windstorms) at medium- and late medium-range. Box A contains a description of various terms used in this study: 'alarms', 'early indication' and 'alarm bells'.

This article considers the process by which forecasters could make use of the EFI to extract information about future extreme weather events. The concepts are illustrated by studying the extreme winds affecting three airports in Germany. Results are presented for a synoptic study of extremes, skill assessment of the EFI and the possibility of setting optimal EFI thresholds for an early indication of windstorms. Finally some examples of utilising the EFI are given.

It is intended that the results presented here will assist forecasters in providing warnings of high wind speeds.

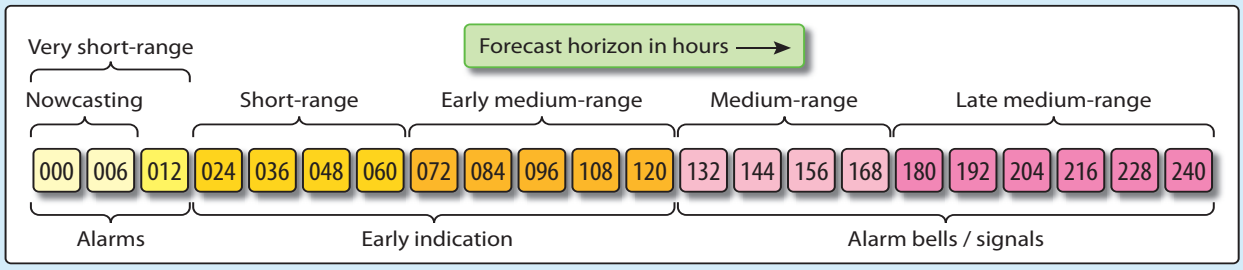
## Rare Severe Events

National Meteorological Services provide warnings about severe or high-impact events that can result in considerable damage and large losses. It is expected that much of the benefit to society through improved weather forecasts will come from advances in our capability to forecast such events so that mitigating actions can be taken. Indeed, one of the principal goals of ECMWF in the next ten years is to provide Member States' National Meteorological Services with reliable forecasts of severe weather across the medium-range while meeting Member States' requirements for high quality near-surface weather forecast products such as precipitation, wind and temperature.

Fortunately severe events tend to be rare, hence the use of the term 'Rare Severe Event' (RSE) by *Murphy* (1991). Such events are also loosely referred to as 'Extreme Events'

### Description of various terms used in the study

A




- ***'Alarms'*** refers to information concerning severe weather being anticipated in the very short-range. This type of information is based on methodologies or models capable of providing estimates about the level of predictability in the very short-term (mainly 0 to 6 hours while sometimes extending to 12 hours). Near-real time online observations are utilised in conjunction with immediate very short-term forecast updates on regional and local scales.
- ***'Early indication'*** refers to information about the occurrence of severe weather in the short range and early medium-term, i.e. in the next 12 to 60 hours (short-range) and 60 to 120 hours (early medium-range). Such tools, based mainly on the ability of the EFI to provide an early indication of extremes, can be used for issuing a warning of a moderate risks and thereby allow users to prepare an effective response.
- ***'Alarm bells'*** refers to those cases for which very low probability extreme events can be captured by some members (sometimes only one) of the ensemble in the medium- or even in the late medium-range. As such 'signals' become stronger and stronger, they should be considered as the basis (necessary elements) of issuing a more specific type of alert (i.e. an early warning).

in atmospheric science. RSEs can come in many forms, associated for example with very intense winds, heavy rain, extreme heat and cold, floods and droughts.

Forecasting RSEs poses specific problems because they are infrequent, poorly documented by observations, and at the limit of predictability. Quantitative verification of RSEs is therefore difficult and the statistical significance of verification results is mostly difficult to establish. At the same time, it is recognized that an imperfect numerical forecast in absolute terms can be of great value if it is well interpreted by an experienced forecaster. This means that a forecast error of given amplitude may have varying significance depending on where the forecast is placed with respect to the climatological distribution.

### Predictability limitations concerning extremes

In operational forecasting, a 'gap' seems to exist between some of the events for which forecasters need to issue warnings and the guidance available from the numerical model. A study of past extreme wind events (such as windstorms) reveals that only a small proportion of ensemble members (or of single deterministic forecasts from different NWP centres) succeeded in predicting their true severity, even about 24 hours in advance. Some types of damaging or disruptive weather, such as lightning, wind gusts and fog, are not explicitly predicted by the models, and must therefore be inferred. Even if a type of weather can be explicitly predicted (e.g. heavy rain), the model resolution might be insufficient to capture its peak intensity; this could be because the associated processes are sub-grid scale. Several mesoscale models are being run experimentally at resolutions of 1–2 km, but most operational mesoscale models have grid scales of 5–15 km, and global models are even coarser.

Therefore we should not expect the current models always to reproduce the maximum values of weather parameters observed in extreme events because their resolution is relatively low. We should, however, design methods to diagnose severe weather based on the existing models, and thoroughly verify the validity of these diagnostics (*Bougeault*, 2003).

### Extreme events and the EFI

The ability of models to generate extreme/severe storms with realistic frequency has improved significantly in recent years. Furthermore the development of ensemble prediction techniques has enabled the explicit representation of uncertainty in the forecast, both in the synoptic-scale evolution and in the development of associated severe weather events. This means that models can now be used to provide information about the likelihood of extreme events occurring.

The Extreme Forecast Index (EFI) (*Lalaurette*, 2003) has been developed to identify the risk of extreme events depending on location and season. It measures the difference between the probability distribution of the ensemble forecast and that of the model climate. The underlying assumption is that if a forecast is extreme relative to the model climate, the real weather is also likely to be extreme compared to the real climate. The EFI is defined such that it lies between –1 and +1.

The EFI allows the forecaster to identify a possible future extreme weather situation without having to define specific thresholds for an extreme event. If the EFI indicates potential for a severe weather event, the forecaster can examine more detailed information from the forecast to make a more thorough assessment of the risk to the public.

Note that during the period covered by this study the resolution of ECMWF's Ensemble Prediction System (EPS) has changed. Up to February 2006 it had a resolution of 80 km, while up to January 2010 it had a resolution of 50 km out to ten days – it then increased to ~30 km.

### Dealing with extremes

Ensemble forecasts provide information on the uncertainty of forecasts. It is desirable to communicate this information, particularly for events that can induce large losses. Probabilistic forecasts can also be used for decision-making by quantitatively assessing risk for specific users using a cost-loss model (for example). However, in the medium range, prediction of severe weather is likely to be associated with relatively low levels of confidence. Bearing this in mind, medium-range 'alarm bells' can ensure that potentially dangerous events do not go unnoticed by the forecasters.

In this study we consider events for which daily wind speed extremes exceed the 99$^{th}$ percentile of the model and station (synoptic) climate records. We will show that the EFI provides a useful indication of extreme events: high EFI values are generally associated with more extreme winds. By selecting an appropriate EFI threshold value, a user can tune their alert system to provide an optimal balance between hits and false alarms.

### Case study for Bremen, Hamburg and Hannover airports

The link between extreme wind events and the EFI has been investigated for three synoptic stations based at airports in North Germany: Bremen, Hamburg and Hannover (as shown in Figure 1).

Two methods are used to define the wind speed extremes.

- *'Reanalysis' mode.* The ECMWF ERA-Interim (Simmons et al., 2007) was used to construct a time series of daily maximum wind speeds for each station, spanning 2,374 days from 1 December 2003 to 31 May 2010. The maximum wind speed for each day was defined as the maximum value of the wind at the five synoptic hours: 00, 06, 12, 18 and 24 UTC.
- *'Observation' mode.* A time series was constructed based on each station's observations of maximum wind speed. In this case the daily maximum values are defined by considering 8 reported observations at 00, 03, 06, 09, 12, 15, 18 and 21 UTC.

The next step was to construct a time series of the daily maximum anomaly for each station in both 'Reanalysis' and 'Observation' modes. For each station and for all cases exceed-


**Figure 1** Geographical position of Bremen, Hamburg and Hannover airports/synoptic stations in North Germany (denoted by red circles)

ing the 99th percentile, the synoptic meteorological environment was investigated. The extremes were found to be linked to deep surface pressure lows, on most occasions affecting all three stations on the same day, as shown in Table 1.

## Utilisation of the DWD Objective Weather Type Classification

The synoptic situation associated with the extremes has been investigated by examining the large-scale atmospheric circulation on the one hand and surface climate and environmental variables on the other. The Objective Weather Type Classification (OWTC) methodology of the Deutscher Wetterdienst (DWD) (*Bissolli & Dittmann*, 2001) uses meteorological criteria such as:

- 700 hPa advection ('No advection', 'Northeast', 'Southeast', 'Southwest' and 'Northeast')
- Cyclonicity at 950 ('Cyclonic', 'Anticyclonic')
- Cyclonicity at 500 hPa ('Cyclonic', 'Anticyclonic')
- Humidity from 950 to 300 hPa ('Wet', 'Dry')

from which a total of 40 weather types are derived. The classification used in this study, however, is based only on the 700 hPa advection. A time series of weather type for North Germany was constructed to correspond to the 'Reanalysis' and 'Observation' time series described above.

It was found that all >99% extremes belonged to weather systems being advected by the 'Southwest' or 'Northwest' flow regimes with 50% falling into each category. It is interesting that no extremes belong to the 'Northeast' or 'Southeast' regimes or to the 'No advection' category. These results seem to agree quite well with those by *Donat* (2010) who found that about 80% of storms affecting Central Europe are associated with westerly flow regimes.

Though this synoptic approach is of value in making forecasters aware of the possibility of extreme winds, it is advantageous for forecasters to base warnings of extreme events at short- and early medium-range on more objective criteria. A probabilistic approach is desirable in order to tailor the signal from the numerical forecasts to the specific needs of users. We investigate identification of extremes based on the value (i.e. a critical threshold) of the EFI.

| Date | Surface Low Identifier | Bremen | Hamburg | Hannover |
|---|---|---|---|---|
| 21/12/03 | Jan | * | | * |
| 13/01/04 | Hanne | * | * | * |
| 14/01/04 | | * | * | * |
| 31/01/04 | Pia & Quinne | * | * | * |
| 01/02/04 | | * | * | * |
| 20/03/04 | Melita & Nina | * | * | * |
| 01/03/04 | Oralie & Paloma | * | * | * |
| 17/11/04 | Pia (New) | | * | |
| 18/11/04 | | | * | |
| 02/01/05 | Alloys | * | | |
| 08/01/05 | Dimitri & Erwin | * | * | * |
| 12/02/05 | Ulf | * | * | * |
| 17/03/05 | Heijo & Iradj | * | * | * |
| 30/12/06 | Karla & Lotte | * | * | * |
| 31/12/06 | | * | * | * |
| 11/01/07 | Franz & Anonym | * | * | * |
| 12/01/07 | Gerhard & Hanno | | * | |
| 13/01/07 | | | * | |
| 18/01/07 | Kyrill | * | * | * |
| 19/01/07 | Kyrill & Lancelot | * | | * |
| 21/01/07 | Lancelot | | | * |
| 10/04/07 | Xenophon | | * | |
| 11/05/07 | Ewald I & II | | | * |
| 26/06/07 | Uriah & Vanni | | | * |
| 27/06/07 | | | * | |
| 26/01/08 | Paula | | * | |
| 31/01/08 | Resi | * | * | * |
| 01/02/08 | | * | * | |
| 01/03/08 | Emma | * | * | * |
| 02/03/08 | | * | | * |
| 12/03/08 | Johanna & Kirsten | * | | * |
| 23/03/09 | Herbert | * | | * |
| 03/10/09 | Ralf & Soeren | * | * | |
| 16/10/09 | Vimar & Xavier | * | | |
| 18/11/09 | Ingmar & Jurgen | * | * | * |
| 01/03/10 | Xynthia | | | * |

**Table 1** Dates and names of intense surface lows linked to >99% daily extremes in 'Reanalysis' mode for Bremen, Hamburg and Hannover. An asterisk is used when a storm is hitting one of the airports.

## Detecting extreme events based on the EFI

The EFI is not only sensitive to a shift in the tails of the frequency distribution (i.e. in the extremes) but also to the median.

In this study, the EFIs for two variables were utilised:

- EFI-10FGI based on a maximum wind gust
- EFI-10WSI based on daily average of instantaneous 10-metre wind speed.

For each of these, EFI forecasts based on both initialisation times (i.e. 00 and 12 UTC) were considered in 'Reanalysis' and 'Observation' modes.

Clear signs that EFI values are closely linked to daily maximum wind speeds are contained in Figure 2. The 24-hour forecast is used in this example, but similar results apply for the other lead times. These results reveal beyond any doubt that all reanalysis daily extremes (falling in the >99th percentile category) for Hannover correspond to strong positive EFI-10FGI values based on 00 UTC runs.

## Skill assessment of the EFI

In addition to assessing point wise EFI values (Bremen, Hamburg and Hannover), their average wind maxima were also considered.

Results in terms of hit rates and false alarm rates for different EFI thresholds are studied by utilizing ROC (Relative Operating Characteristic) diagrams and more specifically ROCA (Area under the ROC Curve) values.

In terms of ROCA, the EFI-10FGI gust factors based on 00 UTC and 12 UTC data are comparable in skill in 'Reanalysis' mode, both comprising high values. Furthermore, the skill of the EFI forecasts over single points seems to be the same as that for the average of the three points.

For EFI-10WSI no significant difference in skill was detected between forecasts based on 00 UTC and 12 UTC data in 'Reanalysis' mode. Also the skill of EFI-10WSI for selected points was found to be comparable to that obtained over the area covering Bremen, Hamburg and Hannover.

In the 'Observation' mode there were no significant differences between using 00 and 12 UTC data in EFI-10FGI. The same applied to EFI-10WSI. However, for both the EFI-10FGI and EFI-10WSI the forecasts are found to be less skilful in 'Observation' mode. This is not surprising: the model wind (representative of a 50×50 km grid box) is not directly comparable to the observations at individual points. Another reason might be that the model has an easier task verifying against its own analysis (reanalysis for our case) extremes than against synoptic observations.

Overall EFI-10WSI was found to be less skilful as a forecast for maximum wind than EFI-10FGI. This could be anticipated since we constructed daily series of extreme wind values that are different from mean (daily averaged) wind time series in both 'Reanalysis' and 'Observation' modes. Going after such extremes, the EFI-10FGI formulation being based on model's 'gusty' components seems a more appropriate option than the EFI-10WSI formulation that is based on 'normal' instantaneous 10-metre wind components.

Results in predicting extremes by utilising the EFI indicate significant skill in both the short- and early medium-range. It should be pointed out that to achieve high hit rates for all forecast lead times (as in the example shown in Figure 3), a significant number of false alarms would be generated as well. This behaviour is somewhat hidden by the rarity of the rare severe events represented in ROC curves and the associated ROCA scores (*Choo*, 2009). However, early indications of potential extreme events allow users to take appropriate

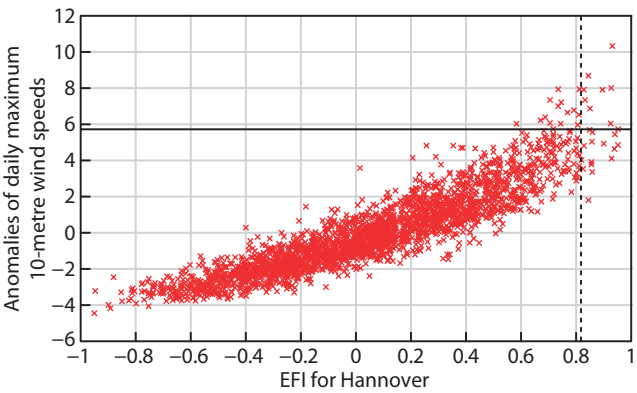



**Figure 2** Example of anomalies of daily maximum 10-metre wind speeds in 'Reanalysis' mode against the 24-hour forecasts of EFI-10FGI (based on 00 UTC) values for Hannover. The dashed black vertical line represents the 99th percentile EFI threshold, while the solid black horizontal line is the 99th percentile of maximum daily wind speed anomalies.

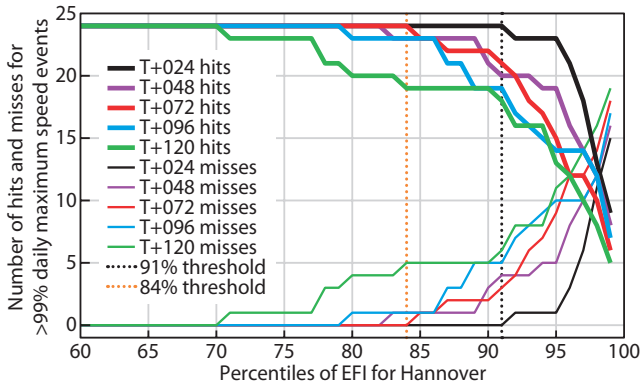



**Figure 3** Hits and misses for the >99th percentile wind extremes based on different EFI-10FGI (00 UTC) thresholds for various lead times (Hannover). Also shown are the EFI thresholds for the 91st percentile (zero misses for day 1; black vertical line) and the 84th percentile (zero misses for day 3; red vertical line).

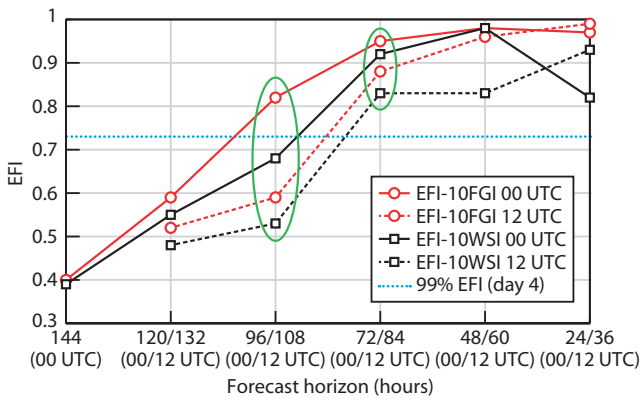



**Figure 4** EFI-10FGI and EFI-WSI (based on 00 and 12 UTC) values for the area of Xynthia's maximum impact at the borders of Luxembourg and France (28 February 2010).

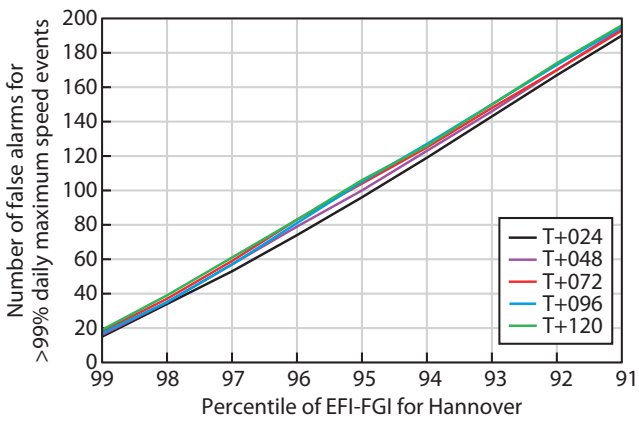



**Figure 5** Number of False Alarms for different EFI-10FGI percentile thresholds for lead times from 24 to 120 hours. It is obvious that the 91st percentile (resulting to zero misses for T+24) also introduces 190 false alarms.

mitigating action. Depending on their sensitivity to the event, different users will take action at different levels of risk. A user who is especially vulnerable to an extreme event may decide to act even at a relatively low risk threshold, while others may prefer to wait until the event is more certain.

## Setting an optimal EFI threshold

The usefulness of early indications of severe weather based on the EFI can be seen in Figure 4. This shows the EFI values for the maximum impact location (borders of Luxembourg and France) of storm Xynthia on 28 February 2010. It is clear that the EFI-10FGI is capable of providing an early indication of high winds four days in advance. The same holds for the other EFI variables but there is a delay of 24 hours.

Using the 99th percentile of EFI, very high (skilful) ROCA values were found for all three airports. This threshold is capable of providing an early indication for some extremes, but not for all (as displayed in Figure 3). By lowering this threshold, the number of hits can be increased till eventually all extremes are captured, but the number of false alarms is then increased significantly. This unavoidable drawback can be seen in Figure 5 where the number of false alarms is plotted against different EFI-10FGI thresholds for Hannover airport corresponding to the hits contained in Table 2.

The number of hits for the 24-hour forecast is equal to 9, but there are also 15 misses and 15 false alarms (Table 2). The 'zero misses' EFI threshold (i.e. the one corresponding to the 91st percentile), highlighted by yellow shading in Table 2, is able to predict all 24 hits (i.e. zero misses), although by doing so the number of false alarms is increased significantly and reaches 190. This limitation becomes more pronounced when different (longer) lead times are considered, as easily seen by examining the results for days 1 to 5 in Table 2. For instance, the day 5 'zero misses' for the 99th percentile extreme wind anomalies corresponds to a considerably lower threshold of EFI, equal to the 70th percentile (resulting in 688 false alarms).

Overall, it is clear that all observed extremes (falling in the >99th percentile category) are linked to high positive EFI values. The highest skill in providing an early indication is from the EFI-10FGI.

| EFI threshold (%) | Day 1 T+24 | Day 2 T+48 | Day 3 T+72 | Day 4 T+96 | Day 5 T+120 |
|---|---|---|---|---|---|
| 70 | 24 | 24 | 24 | 24 | 24 |
| 71 | 24 | 24 | 24 | 24 | 23 |
| 72 | 24 | 24 | 24 | 24 | 23 |
| 73 | 24 | 24 | 24 | 24 | 23 |
| 74 | 24 | 24 | 24 | 24 | 23 |
| 75 | 24 | 24 | 24 | 24 | 23 |
| 76 | 24 | 24 | 24 | 24 | 23 |
| 77 | 24 | 24 | 24 | 24 | 23 |
| 78 | 24 | 24 | 24 | 24 | 21 |
| 79 | 24 | 24 | 24 | 24 | 21 |
| 80 | 24 | 24 | 24 | 23 | 20 |
| 81 | 24 | 24 | 24 | 23 | 20 |
| 82 | 24 | 24 | 24 | 23 | 20 |
| 83 | 24 | 23 | 24 | 23 | 20 |
| 84 | 24 | 23 | 24 | 23 | 19 |
| 85 | 24 | 23 | 23 | 23 | 19 |
| 86 | 24 | 23 | 23 | 23 | 19 |
| 87 | 24 | 23 | 22 | 21 | 19 |
| 88 | 24 | 23 | 22 | 21 | 19 |
| 89 | 24 | 23 | 22 | 19 | 19 |
| 90 | 24 | 21 | 22 | 19 | 19 |
| 91 | 24 | 20 | 21 | 19 | 18 |
| 92 | 23 | 20 | 20 | 17 | 16 |
| 93 | 23 | 20 | 18 | 16 | 16 |
| 94 | 23 | 19 | 17 | 15 | 16 |
| 95 | 23 | 19 | 15 | 14 | 13 |
| 96 | 21 | 16 | 12 | 14 | 12 |
| 97 | 18 | 14 | 12 | 14 | 10 |
| 98 | 13 | 12 | 10 | 12 | 8 |
| 99 | 9 | 8 | 6 | 7 | 5 |

**Table 2** Number of Hits for >99th percentile extremes based on various EFI-10FGI (00 UTC) thresholds for different lead times valid for Hannover (maximum number of hits: 24). The red cells indicate the 'zero misses' EFI thresholds for the various lead times.

## Examples of utilising the EFI

The setting of optimal EFI thresholds is further investigated for extreme events over Hannover. All daily maximum wind speed values for Hannover ('Reanalysis' mode) over a period of 2,374 days are plotted in Figure 6. A selection of the four most recent spikes has been made (highlighted by a red circle). These spikes indicate the following storms: Kyrill (18 January 2007), Emma (1 March 2008), Herbert (23 March 2009) and Xynthia (1 March 2010).

As an example the various EFI-10FGI maps valid for Emma storm are displayed in Figure 7 for the forecast period from

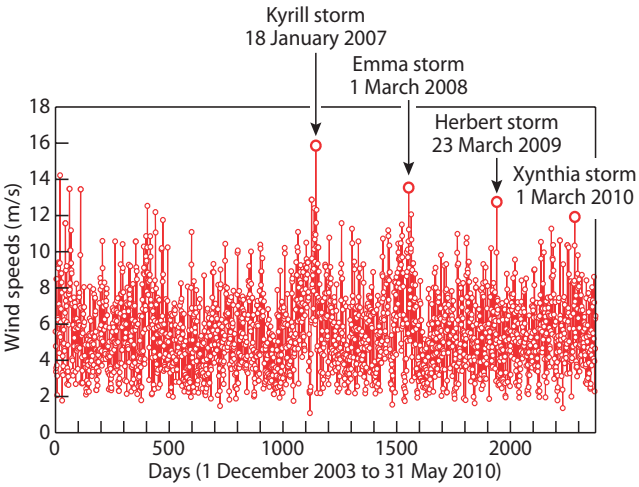



**Figure 6** Time series of daily maximum wind speed values for Hannover over a period of 2,374 days from 1 December 2003 to 31 May 2010 ('Reanalysis' mode).

24 to 132 hours with 12-hour intervals. It is clear that both the 95% and 98% EFI thresholds (highlighted by a yellow line) are able to provide an early indication of the Emma windstorm from day 5.5 (T+132 h) onwards.

To investigate whether these thresholds can provide an early indication of the other storms considered here, Figure 8 is constructed. Clearly both the 95th and 98th percentile thresholds work quite well for the Kyrill and Emma storms, but they seem to be inadequate for Herbert and Xynthia. More specifically, for Herbert, using the 98th percentile threshold fails to give an indication of high winds, while use of the 95th percentile seems to do a better job for lead times shorter than 84 hours. As for Xynthia, the 98th percentile seems to work only for the 96-hour lead time, while the 95th percentile threshold works for all lead times shorter than 120 hours (except for the 36-hour one). For both Herbert and Xynthia, a slightly lower threshold (say between 90th and 95th) could have resulted in forecasters having an early indication of the severity of the winds associated with the approaching storms.

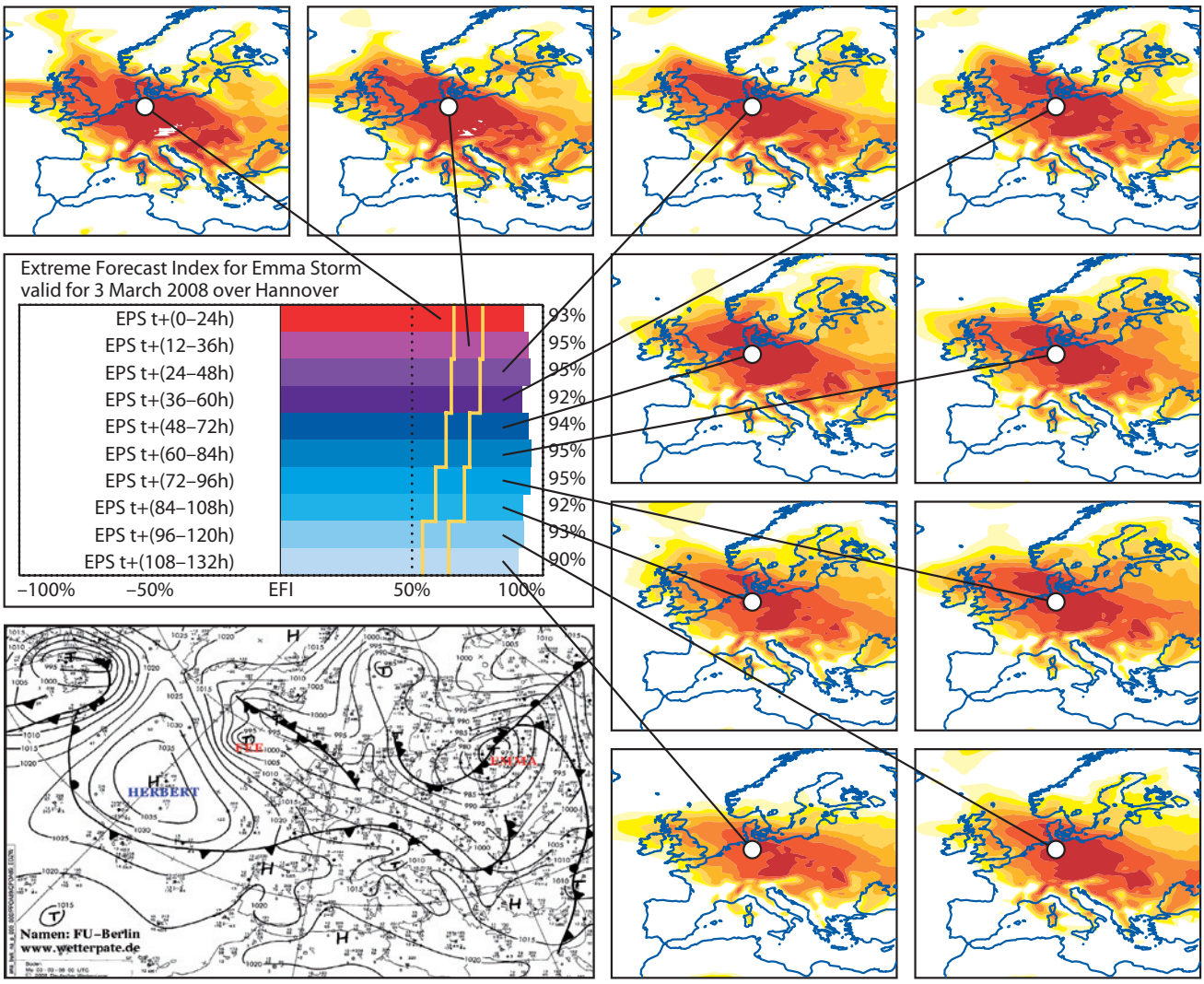



**Figure 7** Example of different EFI-10FGI maps ('EFI-GRAM') valid for the Emma storm hitting Hannover airport on 1 March 2008. The arrows from each map (initiating from Hannover's position) point to the part of the central graph constituting the currently operational 'EFI-GRAM'. The different forecast steps are displayed on the left of the diagram while the exact EFI values over Hannover are displayed on the right. Forecast lead times span from 24 to 132 hours with 12-hour intervals. A near-crash incident of an Airbus A320 took place at the nearby Hamburg airport.

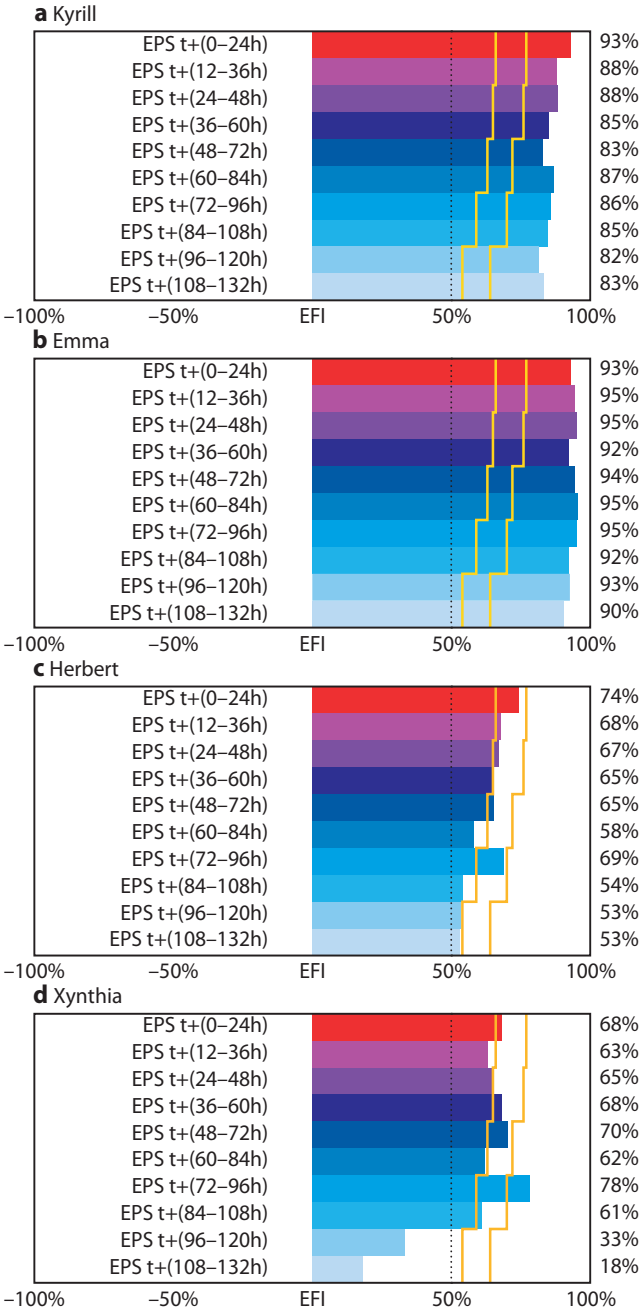



**Figure 8** EFI-10FGI values over Hannover for four windstorms (a) Kyrill (18 January 2007), (b) Emma (1 March 2008), (c) Herbert (23 March 2009) and (d) Xynthia (1 March 2010). The 95th and 98th percentile thresholds are plotted using a yellow line.

## Overview

This study is focused on the early indication of extreme winds in the short- and early medium-range using the EFI. For the assessment of the quality of the EFI, three synoptic stations at airports in North Germany (i.e. Bremen, Hamburg and Hannover) were considered. An investigation of synoptic weather type for each station indicated that all wind extremes (exceeding the 99th percentile) were linked to surface pressure lows being advected in south-westerly and north-westerly flow regimes.

For the objective evaluation of early indications of an extreme weather event, the EFI for wind gusts and mean wind speed were compared to daily maximum wind speeds (in both 'Reanalysis' and 'Observation' modes). The highest skill in detecting extremes is given by the EFI-10FGI. Extreme observed events are clearly linked to higher values of the EFI.

Although the EFI is designed to be used qualitatively as a general 'alarm bell' for potential extreme weather, it is also possible to use the EFI in a more quantitative way. The user can select a specific EFI threshold and take appropriate action whenever the EFI exceeds this threshold. The examples shown in this article illustrate some possible uses of this objective approach. There is no direct mathematical correspondence between percentiles of the EFI distribution and those of the climate distribution. However, in general selecting a high EFI threshold (e.g. the 99th percentile) focuses on the strongest warnings and will have fewest false alarms.

By lowering this threshold the number of hits is increased until all extremes are captured (i.e. zero misses), but by doing so the number of false alarms is increased significantly. Some users will be especially sensitive to missed events while others will be interested in limiting the number of false alarms. As this study has shown, each user is able to choose an appropriate EFI threshold for their own requirements, to provide an optimal trade-off between hits and false alarms.

## FURTHER READING

**Andersson, E.** & **D. Richardson**, 2011: Forecast performance 2011. *ECMWF Newsletter No. 130*, 15–16.

**Bissolli, P.** & **E. Dittmann**, 2001: The objective weather type classification of the German Weather Service and its possibilities of application to environmental and meteorological investigations. *Meteorologische Zeitung*, **10**, S. 253–260.

**Bougeault, P.**, 2003: WGNE survey of verification methods for numerical prediction of weather elements and severe weather events. *Report 18 CAS/JSC WGNE*.

**Choo, C.W.**, 2009: Information use and early warning effectiveness: Perspectives and prospects. *J. Am. Soc. Information Sci. Technol.*, **60**, 1071–1082.

**Donat, M.G.**, 2010: *European wind storms, related loss potentials and changes in multi-model climate simulations.* Doktorarbeit am Institut für Meteorologie, Freie Universität Berlin, 175 pp.

**Lalaurette, F.**, 2003: Early detection of abnormal weather using a probabilistic Extreme Forecast Index. *Q.J.R. Meteorol. Soc.*, **129**, 3037–3057.

**Lalaurette, F.** & **G. v.d. Grijn**, 2005: Predictability of Weather and Climate: Ensemble forecasts: can they provide useful early warnings? In *Predictability of Weather and Climate*, Palmer, T.N. and R. Hagedorn, editors, Cambridge University Press.

**Murphy, A.H.**, 1991: Probabilities, odds and forecasts of rare events. *Wea. Forecasting*, **6**, 302–307.

**Richardson, D.S.**, **J. Bidlot**, **L. Ferranti**, **A. Ghelli**, **T. Haiden**, **T. Hewson**, **M. Janousek**, **F. Prates** & **F. Vitart**, 2011: Verification statistics and evaluations of ECMWF forecasts in 2010-2011. *ECMWF Tech. Memo. No. 654*.

**Simmons, A.**, **S. Uppala**, **D. Dee** & **S. Kobayashi**, 2007: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter No. 110*, 25–35.