Detection of Severe Weather Events in a High-Resolution Ensemble Prediction System Using the Extreme Forecast Index (EFI) and Shift of Tails (SOT)

LAURE RAYNAUD, BENOÎT TOUZÉ, AND PHILIPPE ARBOGAST
CNRM/GMAP/RECYF, Météo-France, Toulouse, France

(Manuscript received 13 December 2017, in final form 29 March 2018)

ABSTRACT

The extreme forecast index (EFI) and shift of tails (SOT) are commonly used to compare an ensemble forecast to a reference model climatology, in order to measure the severity of the current weather forecast. In this study, the feasibility and the relevance of EFI and SOT computations are examined within the convection-permitting Application of Research to Operations at Mesoscale (AROME-France) ensemble prediction system (EPS). First, different climate configurations are proposed and discussed, in order to overcome the small size of the ensemble and the short climate sampling length. Subjective and objective evaluations of EFI and SOT for wind gusts and precipitation forecasts are then presented. It is shown that these indices can provide relevant early warnings and, based on a trade-off between hits and false alarms, optimal EFI thresholds can be determined for decision-making.

1. Introduction

Kilometer-scale ensemble prediction systems (EPSs) are currently being developed at several operational centers for the short-range prediction of high-impact weather (Kühnlein et al. 2014; Tennant 2015; Raynaud and Bouttier 2016). To improve the detection of extreme events, Lalaurette (2003) introduced the extreme forecast index (EFI), which compares the current EPS distribution to the model climatology. Zsoter (2006) then complemented the EFI with the shift of tails (SOT) index, which provides information about the extremes of the EPS distribution.

Several studies demonstrated the usefulness of the EFI and SOT indices to extract relevant information from large-scale EPSs (e.g., Zsoter 2006; Petroliagis and Pinson 2014; Boisserie et al. 2016b). However, to the authors’ knowledge no similar works have been performed with high-resolution EPSs yet. Hence, this paper aims at presenting a first application and evaluation of EFI and SOT within the convective-scale Application of Research to Operations at Mesoscale EPS (AROME-EPS) operational at Météo-France, based on a limited climate sample.

This paper is organized as follows. After some reminders about EFI and SOT definitions in section 2, section 3 presents the dataset used. Different strategies for computing the model climate are presented in section 4, and their impact on EFI and SOT is discussed in section 5. An objective evaluation of the EFI forecast skill is then assessed in section 6. Finally, conclusions and future work are provided in section 7.

2. EFI and SOT definitions

The definition of the EFI used in this paper has been introduced by Zsoter (2006):

\[
\text{EFI} = \frac{2}{\pi} \int_0^1 \frac{p - F_I(p)}{\sqrt{p(1-p)}} dp,
\]

where \( F_I(p) \) is the proportion of EPS members lying below the climate \( p \) quantile. EFI values lie between \(-1\) and \(1\): the value \(1\) (\(-1\)) is reached when all EPS members are beyond the high (low) record of the model climate.

In addition to the EFI, the SOT index has been developed to focus on the tails of the distribution and is expressed as (for the upper tails)

\[
\text{SOT}_+(p) = \frac{Q_f(p) - Q_c(100)}{Q_c(p) - Q_c(100)},
\]

where \( Q_f(p) \) refers to the EPS \( p \) quantile, and \( Q_c(100) \) is the high record of the climate distribution. Positive
values of SOT then indicate that the \( p \) quantile in the forecast is above the climate maximum. A similar expression is obtained for the lower tails of the EPS distribution \([SOT_{-}(p)]\) by replacing the maximum \( Q_{e}(100) \) with the minimum \( Q_{e}(0) \).

3. Experimental setup

a. The AROME-France model

AROME is a spectral, nonhydrostatic convective-scale model operational at Météo-France since December 2008 and covering the French territory (Seity et al. 2011). It is designed for the short-range prediction of severe weather events, such as heavy precipitation and strong winds. The current operational configuration uses 1.3-km horizontal resolution and 90 vertical levels, and the analysis step is performed with an hourly three-dimensional variational data assimilation (3D-Var) scheme (Brousseau et al. 2016).

b. The AROME-EPS

In addition to the deterministic AROME forecasts, a high-resolution EPS has been built upon the AROME-France model to provide short-range probabilistic weather forecasts. In its present operational configuration, AROME-EPS is a 12-member ensemble running with 2.5-km horizontal resolution and 90 vertical levels, which provides 45-h forecasts twice a day, starting at 0900 and 2100 UTC. Each AROME-EPS member is perturbed in order to account for the uncertainty in the initial and lateral boundary conditions, in the surface conditions, and in the model formulation, as described in Bouttier et al. (2012, 2016).

4. The AROME-EPS climate

a. Climate configuration setup

As indicated by the definitions of EFI and SOT, the calculation of these indices requires a prior estimate of the model climate. The World Meteorological Organization (WMO) defines climate as a collection of statistics over a 30-year period. However, such a period is not available for AROME-EPS, and numerical resources do not allow for the computation of a comprehensive reforecast dataset. An alternative is then to use the operational AROME-EPS archive, which is available from 10 August 2015 for the 2100 UTC run. This choice is justified by the fact that the AROME-EPS has not undergone any major upgrades in terms of horizontal resolution and perturbation strategy during this two-year period. EFI/SOT values computed for dates between August 2016 and July 2017 will then rely on a climatology based only on the 12-member ensemble forecasts of the corresponding day the year before. Since this two-year sampling length is not sufficient to provide a reliable climate (Zsoter et al. 2015), additional temporal and spatial sampling methods are proposed in order to improve the model climate estimate.

The temporal tolerance, as suggested by Zsoter (2006), assumes that the climatology is similar within a temporal window of \( \pm N_d \) days. Hence, forecasts within this temporal window can be considered to be samples from the same climate distribution. In the present study, model forecasts of the current year (i.e., the previous \( N_d \) days) are also included in this temporal sample, but the last three days are removed to account for a possible correlation with the situation examined.

The spatial tolerance, also called the neighborhood method, assumes that, under the hypothesis of local homogeneity of the weather, forecast values in a neighborhood of radius \( r \) can be considered to be samples from the same distribution. Hence, for each ensemble member, forecasts are accumulated from the \( N \) grid points of the neighborhood centered around the target grid point. As detailed in Bourtier and Kriat (2015), a random sampling of \( N \) values is then performed (among the \( 12 \times N \) available values), in order to reduce the amount of data to be processed and the correlations between neighboring points. The initial 12-member ensemble is thus transformed into a larger ensemble of \( N \) members.

Different model climatologies can then be obtained, depending on the temporal and spatial samplings setup. Table 1 provides the effective sample sizes corresponding to different values of \( N_d, r, \) and \( N \). In the next sections, the analysis of different climate configurations is presented, along with their impact on the EFI and SOT values.

b. Analysis of the AROME-EPS climate

The effect of climate sampling on the 10-m wind gust field for the case of the Zeus windstorm that hit France on 6 March 2017 is presented. It is worth mentioning that, since the “true” climatology is not known, the climate configuration \((N_d, r, N) = (10, 0, 0)\) will be taken as the reference climate for comparison purposes.

The impacts of spatial (Figs. 1a,b) and temporal (Figs. 1a,c) tolerances on the climate median [50th percentile (Q50)] are shown. As expected, the spatial tolerance has a smoothing effect that reduces the small-scale details, which are likely to be unrealistic to some extent because of undersampling, while the relevant larger-scale signal is mostly preserved. However, in regions with spatial heterogeneities such as reliefs and coasts, this spatial tolerance is to be considered with caution. The impact of temporal tolerance appears to be larger on average, with the median generally being
weaker when using \( N_d = 20 \). Since the sensitivity of the climate estimate to the limited sample is likely to be more visible on the tails of the distribution, the 99th climate percentile (Q99) is presented in Figs. 2a–c. Similar impacts of spatial and temporal samplings are observed. It is also interesting to notice that some local well-known meteorological features are present in the different maps, for instance the area of strong winds in the Rhône valley.

Finally, Q99 for 24-h accumulated precipitation (Figs. 2d–f; valid on 25 November 2016) shows very similar sensitivity to the sampling configurations. It may be noted that the Mediterranean coastal regions and the Cévennes area, which are regularly affected by sudden heavy precipitation events in autumn, seem correctly emphasized.

5. Analysis of EFI and SOT

Following the sensitivity of the model climate to different sampling configurations, the impact on EFI and SOT is now presented. Figure 3 shows the EFI and SOT values computed with three climate configurations for the Zeus windstorm. In addition, since sampling errors can also affect the EPS distribution \( F(p) \) [due to the small EPS size; Dutra et al. (2013)], the impact of applying a spatial tolerance in its computation is examined.

It is first observed that all configurations, despite the short climate hindcast length, provide areas of high EFI and positive SOT that are in good agreement with the observations of strong winds in the vicinity of dynamically active regions (e.g., southwestern France, Brittany, and northeastern Corsica; Fig. 3f). The different model climates lead, however, to some amplitude differences. In particular, the larger EFI values obtained when \( N_d = 20 \) (Fig. 3b) are consistent with the associated lower climate values. Regarding the spatial tolerance, its application to the climate calculation appears to have a rather limited impact on the resulting EFI (Fig. 3c), whereas its application to the EPS distribution (Fig. 3d) leads to a smoother map of EFI and SOT, where small-scale less-predictable details are removed while the useful large-scale signal is preserved (with the exception of relief areas). These first results thus suggest the following:

- Among the different configuration parameters, the length of the temporal window used for the climate computation appears to have the largest impact in general, for both the climate and the EFI.
- The spatial tolerance applied to the climate sample seems to have only a marginal impact on the EFI and SOT values. This result can be explained as follows: since the spatial sampling improves the smoothness of the model climate without significantly affecting its distribution, the expected impact on the EFI is mainly a gain in numerical accuracy. In the present case, adding a spatial sampling on top of a temporal sampling of ±20 days allows for an increase from 708 to 2950 members (as indicated in Table 1), which

#### Table 1. Effective size \( N_{\text{eff}} \) of the climatological sample for different spatial and temporal configurations. The neighborhood radius \( r \) is given in grid points, and \( r = 0 \) corresponds to no spatial tolerance.

<table>
<thead>
<tr>
<th>((N_d, r, N_r))</th>
<th>((10, 0, 0))</th>
<th>((20, 0, 0))</th>
<th>((10, 10, 50))</th>
<th>((20, 10, 50))</th>
<th>((10, 10, 100))</th>
<th>((20, 10, 100))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_{\text{eff}})</td>
<td>348</td>
<td>708</td>
<td>1450</td>
<td>2950</td>
<td>2900</td>
<td>5900</td>
</tr>
</tbody>
</table>

Fig. 1. Q50 of 10-m wind gusts for different climate configurations, valid at 0600 UTC 6 Mar 2017 and computed from 9-h AROME-EPS forecasts.
merely affects the resulting EFI (but contributes to slightly smoothing the SOT).

- The spatial tolerance applied to the EPS sample has a noticeable impact on the reduction of the sampling noise in the EFI and SOT indices. This highlights the fact that, with a sufficiently large climate sample, the numerical accuracy of the EFI in mainly limited by the sampling of the EPS distribution.

In the remainder of this paper, the model climate is computed with the configuration $N_d = 20$, $r = 10$, $N_r = 50$, and a spatial tolerance with setup $r = 10$, $N_r = 100$ is applied to the EPS distribution.

As a first step toward the validation of the AROME-EPS EFI, Fig. 3e presents the EFI computed with the French operational large-scale EPS, the Prévision d’Ensemble Action de Recherche Petite Echelle Grande Echelle (PEARP) system (Descamps et al. 2015), which provides the boundary conditions to the AROME-EPS. The advantage of PEARP is the availability of a 30-yr reforecast dataset that allows for a robust climate estimate (Boisserie et al. 2016a,b). The comparison of the EFI and SOT computed from PEARP and AROME-EPS indicates relatively good agreement in terms of both position and intensity. This result then contributes to demonstrating the relevance of the EFI/SOT computation from the AROME-EPS.

6. AROME-EPS EFI scores

This section presents an objective evaluation of precipitation and wind gusts forecasts based only on the EFI value. The verification is performed over a set of about 20 relevant cases for each parameter, selected between September 2016 and May 2017. Each targeted event corresponds to a day for which a significant part of the domain is associated with an episode of severe weather. The aim is first to verify if the AROME-EPS EFI provides a relevant signal for the detection of observed extreme events and, second, to determine optimal EFI thresholds for issuing early warnings.

The approach adopted here is to evaluate EFI on the basis of the warnings that are operationally issued at
Météo-France (and corrected upon observations). The French “vigilance” watch map, operational since October 2001, defines for the next 24 h the level of high-impact weather at the scale of a French *département* (an administrative division of the national territory). Each department is assigned a level of warning between 1 and 4 (the higher the level, the more severe the event), associated with a color, and defined according to the exceedance of some thresholds. It is produced twice a day at 0600 and 1600 local time. An example is given below (see Fig. 6b).

The scores presented in this section have been computed using the vigilance map issued at 0600 local time (and valid for the next 24 h) as the verification. EFI values over the warning period are computed based on AROME-EPS forecasts from the latest 2100 UTC run. For the verification of precipitation, EFIs are directly computed from the 24-h accumulated precipitation forecast. Regarding 10-m wind gusts, the maximum EFI value over the 24-h period is considered. In addition, since EFI maps are defined at the grid scale while the vigilance map is defined at the departmental scale, an upscaling method has been developed in order to obtain departmental EFI maps. This simple procedure assigns to each department the median of the EFI distribution within that department.

### a. Scores for 24-h rainfall rate

Figure 4a shows a scatterplot of EFI values against the level of warning. It appears that, as the severity of the observed event increases, the EFI distribution is shifted toward higher values. For instance, the median EFI values are equal to 0.27, 0.65, and 0.78 for warning levels 1, 2, and 3, respectively. This is a first objective indication of a close link between the EFI computed with the AROME-EPS and the issued warnings. This also suggests that this index could provide a relevant signal for severe weather, despite the limited climatological sample.

In addition, hit rates $H$ and false alarms $F$ have been computed for different EFI thresholds: a correct forecast...
is obtained when the EFI is greater than or equal to a given threshold and the warning level is greater than or equal to 3. Results are summarized with the relative operating characteristic (ROC) curve shown in Fig. 5a. Each point on this curve corresponds to the $H$ and $F$ values obtained with a given EFI threshold. As expected, the lower the threshold, the higher the hit rate and the number of false alarms. The area under the ROC curve is equal to 0.91, which indicates that the AROME-EPS EFI is able to properly discriminate between severe and nonsevere events. Moreover, it is possible to find an optimal trade-off between a sufficient hit rate and an acceptable false alarm rate, for instance based on the user’s cost/loss ratio. In the present case, the EFI threshold value that offers the best compromise between the minimization of $F$ and the maximization of $H$ is 0.65. For this threshold, the nondetection rate is 18% and the false alarm rate is 75%. This cost/loss ratio of 0.24 is close to the values used by users interested in extreme event forecasts, who usually tolerate more false alarms than nondetections.

b. Scores for 10-m wind gusts

The scatterplot (Fig. 4b) indicates a close connection between EFI values and the operational wind vigilance, with the median EFI values equal to 0.37, 0.73, and 0.86 for warning levels 1, 2, and 3, respectively. The area under the ROC curve (Fig. 5b) is equal to 0.85, which is slightly lower than for 24-h precipitation. This lesser skill of wind EFI could be related to the slightly lower performance of AROME-EPS for wind forecasts. The optimal EFI threshold for wind gusts is equal to 0.8 and is associated with nondetection and false alarm rates of 35% and 53%, respectively.

c. Application: Prealert maps

Following these verification and calibration results, we are interested in providing prealert maps for remarkable
events based on the AROME-EPS EFI. For this purpose, if the departmental EFI is higher than the optimal threshold previously obtained, the department is highlighted to inform of a risk of severe weather. An example is shown in Fig. 6 for the Zeus storm. For this case the operational vigilance map indicates that several departments are concerned by a warning of level 3 (in orange). The prealert map derived from the EFI appears to be quite well correlated with the vigilance map. In particular there is no miss: the prealert procedure correctly activates (in violet) all departments whose warning level is superior or equal to 3 (we point out that the prealert EFI threshold has been chosen to detect severe events corresponding to warning levels greater than 3). Some departments in southwestern France are, however, activated while the warning level is only 2 (in yellow), but some of them were at the limit of the level-3 warning. It is also worth recalling that the choice of the optimal threshold allows for more false alarms than nondetections. Nevertheless, the two departments highlighted in northeastern France while the warning level is only 1 (in green) constitute a more important false alarm.

7. Conclusions

EFI and SOT indices have been successfully implemented within a convective-scale ensemble prediction system, based on spatial and temporal methods that allow for significantly reducing sampling errors due to the small sample sizes of the currently available EPS and climate archive. The results obtained over a few cases of severe weather are encouraging: areas of strong EFI and SOT are often consistent with the strongest observations and the Météo-France “vigilance” maps. In the future, it is expected that these indices will become even more accurate owing to a larger climate sample.

Objective verification scores computed for EFI-based forecasts relative to the Météo-France warning procedure also show that this index could be relevant within an operational context for anticipating remarkable weather events. However, a more accurate calibration of optimal thresholds is necessary in order to improve its utilization. In the short term, a regular monitoring of EFI by forecasters could be suggested in order to further explore this aspect.

REFERENCES


