—A statistical method based on the Ensemble probability density function for the prediction of "Wind Days"

A. Tateo[a], M. M. Miglietta[d], F. Fedele[b], M. Menegotto[b], A. Pollice[e], R. Bellotti[a,c]

[a]Dipartimento Interateneo di Fisica, Università deli Studi di Bari "A. Moro", Via G. Amendola 173, 70126, Bari, Italy,
[b]Apulia Region Environmental Protection Agency (ARPA Puglia), C.so Trieste 27, 70126, Bari, Italy,
[c]Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Bari, Via Orabona 4, 70125 Bari, Italy,
[d]National Research Council of Italy (CNR), Institute of Atmospheric Sciences and Climate (ISAC), Lecce s.p. Lecce-Monteroni km 1.2 73100, Lecce, Italy,
[e]Dipartimento di Economia e Finanza, Università degli Studi di Bari, "A. Moro", Largo Abbazia S. Scolastica (già via C. Rosalba, 53) 70124 - Bari, Italy

Abstract

Numerical Weather Prediction (NWP) models are often used to predict meteorological events in a deterministic way. In recent years, operational Ensemble Prediction Systems are able to take into account some of the errors affecting the NWP models, and allow to estimate the probability of occurrence. In the traditional approach, this probability is given by the percentage of ensemble members predicting the event. In this study, we propose an alternative method to estimate the probability of occurrence, based on the ensemble probability density function (PDF), which takes into account only random errors unavoidably affecting the model. To estimate its reliability, we compare this method with classical categorical and probabilistic approaches by using different global models: ECMWF, GFS, and GEFS.
In particular, we focus on wind speed forecasts in the area around the city of Taranto, located in Apulia region (southeastern Italy), to simulate the events called "Wind Days", i.e. northwesterly wind above 7 m/s for 3 consecutive hours. Our analysis concerns 34 case studies covering 2016, opportunely chosen to have a balanced dataset of WD and no WD, the latter category mainly including cases that are very difficult to predict, at the border of the two categories. The results show that the probabilistic approaches have a better skill than the categorical ones. Among the probabilistic approaches, the best result (accuracy of 82%) is obtained using the method proposed here, with the control run of GEFS used to estimate the true value and the gamma distribution to model the error distribution.
To reduce the systematic error, we test different thresholds and numbers of consecutive hours when the definition of WD is applied to model outputs. All the models show remarkably better performances after these parameters are changed. In particular, our method shows the best performance, with an accuracy of 94%. The analysis on test (leave-one-out strategy in 2016) and validation datasets (66 cases in 2017) confirms the previous outcomes. We test our procedures considering the forecast time intervals of 49-72 and 25-48 hours, where similar performances are found. In conclusion, our analysis show that the proposed method presents better performances compared to the traditional approaches for different statistical performance indicators.

Keywords: probabilistic prediction approaches, GEFS, wind day, heavy events prediction

1.1 Introduction

Despite the exponential growth of computational power in the last years, the numerical models for weather prediction are still affected by errors (Boisserie et al., 2014; Wang, 2015). These errors are due to different causes, such as the inability of the models to represent correctly both the atmospheric dynamics and the relevant physical processes, and the sensitivity of simulations to the initial conditions, which is unavoidable.

Although this kind of errors cannot be prevented, due to the intrinsic limitation in the observing systems and in numerical models, some efforts can be done to reduce the limitation of deterministic numerical systems. A typical way to surmount this problem is to develop a probabilistic approach, by using ensembles of simulations, which somehow take into account the errors in the initial conditions and/or in the model formulation.

To deal with the first kind of problem, a simple approach is the use of a multiphysics ensemble, where relevant physical parameters in a fixed model are varied within a range of plausible values, as in Berner et al. (2011). An alternative simple approach is the use of a poor man's ensemble, where the outputs of different models (Ebert E. E., 2001; Corazza, 2018), or same model but with different implementations (e.g., starting at different initial times; time-lagged forecast (Miglietta et al., 2015; Miglietta et al., 2016)), are considered as independent members of the ensemble (Ebert E. E., 2001; García-Ortega, 2017).

About the second kind of problem, it is known that the chaotic nature of the atmosphere amplifies the unavoidable errors in the initial conditions. Thus, small variations in the initial state can lead to large variations in the final state (Lorenz, 2000). For these reasons, in recent years weather centers all over the world have developed operational Ensemble Prediction Systems (EPSs), to take into account errors associated with parameterizations (Buizza et al., 1999) and with the uncertainty in the initial condition (Magnusson et al., 2008).

As an example, the US National Oceanic and Atmospheric Administration (NOAA; http://www.noaa.gov/) has developed the Global Ensemble Forecast System (GEFS) (Guan et al., 2015), while the European Center for Medium-range Weather Forecasts (ECMWF) has developed its own Ensemble Prediction System, operational since 1992 (Molteni et al., 1996). Many studies have shown that the ensemble mean is more accurate than a deterministic forecast (Leith, 1974; Zhang and Krishnamurti, 1997), while the prediction using the ensemble mean is better than individual member forecasts (Murphy, 1988). However, the key added value in using the ensemble systems is that each member provides a different scenario that should be taken into account within a probabilistic approach. The ensemble forecasts of Lothar windstorm affecting northern Europe on December 24, 1999 represent a paradigmatic example of the relevance each single member may have especially in the prediction of extreme events (Palmer and Hagedorn, 2006).

Whitaker and Loughe (Whitaker and Loughe, 1998) evaluated the relationship between ensemble spread and ensemble mean skill, showing that the ensemble spread can be used to estimate the forecast uncertainty (Zacharov, 2009), although they are correlated only in a limited way (Stensrud et al., 1999). Also, for an EPS to be reliable, it is expected that the future atmospheric state should fall within the predicted ensemble spread; however, as shown in Wilks (2011), the outputs of NWP models could be systematically biased with respect to local observations (provided, for example, by ground weather stations). Thus, in the case of site specific applications, reducing this systematic error component is of primary importance (Perera et al., 2014; Pelosi et al., 2016; Cassola and Burlando, 2012).

To take into account these model limitations, several statistical postprocessing techniques have been implemented to improve the model output (Vannitsem, 2008). These statistical approaches consist of estimating the model correction during a training period, in a statistical or dynamical way. In the first case, we need a large training set, consisting of couples of forecasts and observations, in order to fix, once and for all, the correction parameters. An overview of this topic is exposed in Wilks (2011) and in Schefzik (2017).

In the dynamic training, the correction parameters are continuously updated by considering the most recent observations. Many of these methods are based on the Kalman Filter (Libonati, 2008). Pelosi et al. (2017) provides a rich overview and presents an original adaptive kalman filtering procedure for single-model ensemble forecasts. The skill of all these methods is evaluated with respect to their ability to reduce the error of the ensemble mean.

Another important aspect of an EPS is the possibility to associate the probability of occurrence of an event to a numerical prediction. This probability can be estimated as the fraction of ensemble members predicting the event (Sokol, 2017), which thus allows an immediate evaluation of the reliability of a given forecast (Alpert and Wang, 2005). In this way, an EPS may provide not only a probabilistic forecast (by means of the ensemble mean) but also the probability of occurrence.
In this paper, we propose an alternative way to estimate the probability of occurrence of a local event (i.e., the presence of moderate-intense northwesterly wind in a fixed station), by combining probabilistic and deterministic information in a statistical way. In particular, we focus our attention on 10 m wind forecasts in the area around the city of Taranto, located in Apulia region (southeastern Italy), comparing the method we propose with both a categorical approach, based on deterministic weather forecasts, and a traditional probabilistic procedure, based on an ensemble model.
After a short section, describing the motivation of the present study, in the following two sections the analyzed data and a detailed description of the proposed new approach are provided. In the discussion section, we show that the proposed approach has better performances with respect to both the categorical and probabilistic approach in the prediction of WD. For the categorical approach, deterministic model outputs, such as GFS and ECMWF Integrated Forecasting System, the control member of GEFS, and the ensemble average are considered. For the probabilistic approach, we compare three alternative methods based on GEFS data. An additional correction, taking into account the model bias, is proposed. Discussion and conclusions are drawn in the final section.

1.2 Environmental problems in the area of interest

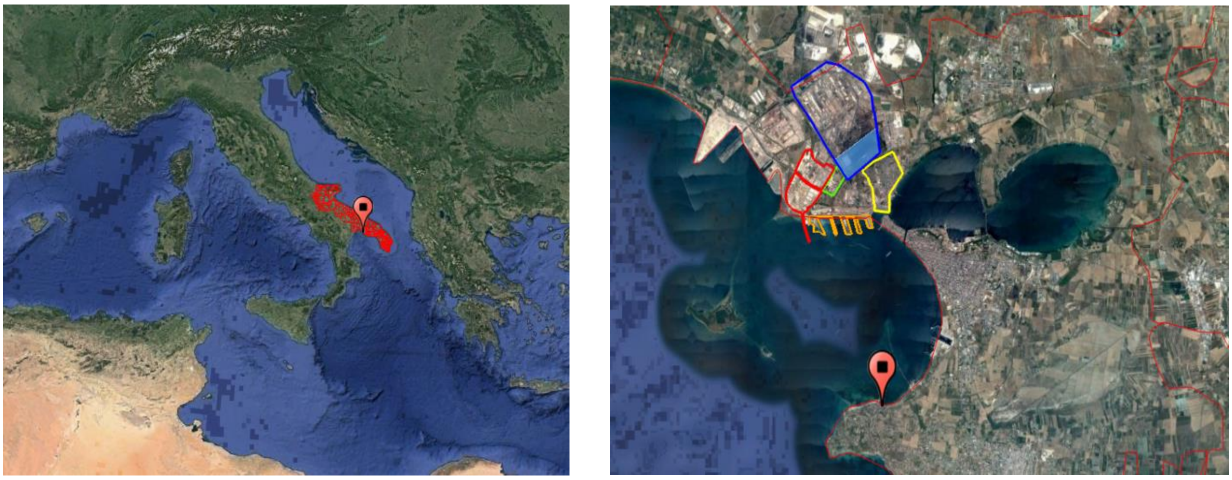

Fig. 1: (Left) Geographical position of Apulia region (red area) in the Mediterranean. Red marker identifies the location of Taranto city in Apulia region. (Right) The industrial area: the largest European steel plant, ILVA (blue-bordered area); oil refinery (red-bordered area); cement plant (green-bordered area); port area (orange area); Tamburi neighborhood (yellow-bordered are), located less than 1 km far from the industrial area. Red marker identifies the location of the ground weather station, located in San Vito neighborhood.

The present analysis is relevant for environmental purposes: in fact, as shown in fig. 1, the city of Taranto, and in particular the neighborhood named Tamburi, is in close proximity of a large industrial complex, including the largest integrated steel plant in Europe (ILVA), an oil refinery, a big cement plant and a port area. The ILVA plant covers a surface of 15 million square meters, with the presence of a large open air mineral stockyard. The Tamburi neighborhood is located less than 1 km far from the stockyard, downwind of the plant with respect to the prevailing northwesterly winds.
The air quality network of the Apulia Region Environmental Protection Agency (ARPA) has registered, in the last few years, exceedances of the limit permitted by law (European Air Quality Directive 2008/50/EC) both for PM10 (suspended particles with diameter under 10 µm) and for benzo(a)pyrene (B(a)P) concentrations in the Tamburi neighborhood (Trizio et al., 2016). Amodio et

al. (2013) have demonstrated a close correlation of limit exceedances with wind conditions, due to short-range transport of air pollution from the industrial site to the adjacent urban area.
In order to improve the air quality in the Tamburi neighborhood, in 2012 the Apulia Government adopted a Regional Air Quality Plan (Regione, 2012). This Act constrains industrial plants to reduce the mean daily B(a)P and PM10 emissions by diffuse and point sources by 10%, during events named as "wind days" (WD) (Fedele et al., 2014). WD are characterized by at least 3 consecutive hours of wind coming from the NW quadrant and speed higher than 7 m/s; the Air Quality Plan requires that WDs must be forecasted 72 hours in advance. For these reasons, it is necessary to accurately simulate wind speed and direction in order to predict the occurrence of WD several hours ahead.
Studies aimed at reducing the error of numerical models in simulating the wind field near Taranto were performed, showing that a significant improvement in the model predictions can be reached using post-processing techniques (Fedele et al., 2015). In Tateo et al. (2017), the outputs of a multi-physics ensemble using different boundary layer parameterization schemes in the WRF model were post-processed by means of Artificial Neural Networks in order to improve the forecast of 10 m wind speed. Other postprocessing techniques have been proposed in Mastrantonio et al. (Mastrantonio et al., 2018), and in Tateo et al. (2015).

2. Material

The skill of large scale models in the prediction of WD is analyzed using the Global Forecast System (GFS) and the Integrated Forecasting System (IFS) of the European Center for Medium-range Weather Forecasting (ECMWF) as deterministic models. The horizontal grid spacing of these models is respectively about 50 km and about 16 km (the actual resolution for both models is better than this; however, the considerations descending from our method do not depend on the resolution of the data we used). The initial time of the runs we consider in our analysis is 00:00 UTC, and the model outputs are available every 3 hours.
Additionally, we use the weather forecast dataset generated by the 2012 version of NCEP's Global Ensemble Forecasting System (GEFS, Version 10) in the ESRL/PSD 2nd-generation Reforecast Project. This Reforecast V2 dataset consists of an 11-member ensemble; the forecasts are produced every day (00 UTC as initial time) from December 1984 to present. The horizontal grid spacing of GEFS is T254 (about 50 km) out to 8 days, and T190 (about 70 km) from day 8 to day 16 (Hamill et al., 2013). The use of this dataset will allow to compare the deterministic simulations with the probabilistic approach, and to draw some considerations on their use for operational purposes.
WD predictions are compared with the observed WD at a ground weather station, which is located in San Vito neighborhood (marker icon in fig.1), in the neighborhood of Taranto, and belongs to ARPA Puglia. Starting from the definition of WD (wind speed greater than 7 m/s for at least three consecutive hours blowing from the north-western quadrant) and from the consideration that systematic model errors may affect the results, we will look for a different wind speed threshold and number of consecutive hours in the model output in order to maximize the performance in WD predictions. Considering that WDs are characterized by large wind speed, dominated by synoptic forcing that the model is able to reproduce properly, we do not consider the wind direction in the evaluation of WDs (the wind is always predicted from the north-western quadrant for each observed WD event), but we focus only on wind speed.
We perform our analysis considering 34 case studies covering 2016. We built a balanced dataset, by including 16 observed WDs and 18 cases with no WDs. The latter set includes situations difficult to predict, characterized either by a wind speed above the threshold only for two consecutive hours, or by a wind speed slightly below the threshold. In fig. 2, as an example, we report the wind speed measured on July 17, 2016 by the ground station in San Vito: the wind is above the threshold of 7 m/s for only two consecutive hours, thus it is not a WD, although very close to it.

In contrast, among the observed WDs, the dataset includes, in addition to days when the WD criterion is satisfied sharply, also days characterized by wind speed just above the threshold either with long persistence (fig. 3A) or for exactly three consecutive hours (fig. 3B).

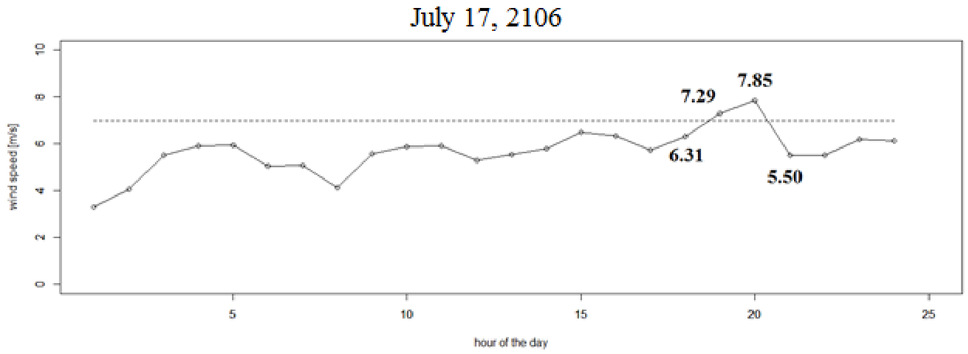

Fig 2: No WD case: there are only two consecutive hours with wind speed above the threshold, while in the previous hour the wind was just below the threshold. The horizontal dashed line represents the threshold of 7 m/s.

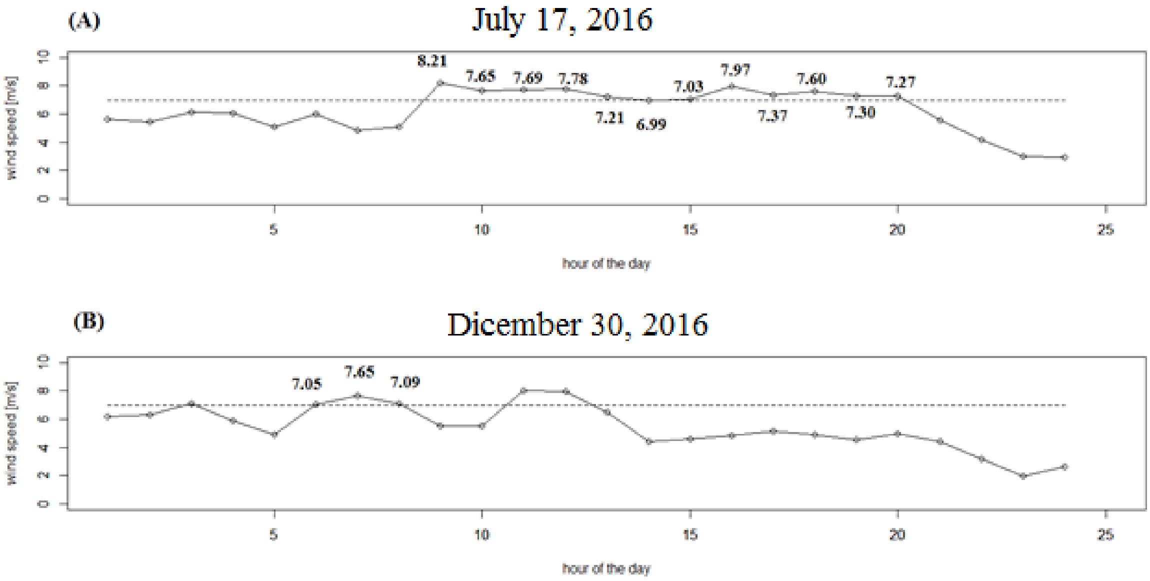

Fig. 3: Two observed WD cases: (A) persistence of wind speed just above the threshold for several hours; (B) only three consecutive hours of wind, just above threshold. The horizontal dashed line represents the wind speed threshold of 7 m/s.

In the full list of WDs, there are both isolated events and sequences of several consecutive wind days (up to 3 consecutive WD events). Particular attention will be devoted to the latter group of events, for which we noted that the first and the last of several consecutive wind days are often predicted inaccurately. In the following analysis, the predictions in the time window 25-48 hour and 49-72 hour are considered separately.

3. Methods

Our study deals with two different approaches, respectively a categorical and a probabilistic one. The former refers to deterministic outputs, such as the GFS and ECMWF global forecast data, the control member of GEFS, and the GEFS mean (we found similar performances for the median). Since the WD definition requires hourly data, the three-hourly global data have been interpolated to provide hourly outputs using a spline interpolation method.

The probabilistic approach is based on GEFS data. In our study, we consider three different methodologies to treat with ensembles. The first is the traditional approach (Ebert E. E., 2001), based on the percentage of ensemble members that predict the occurrence of the event (WD). The second method is based on the quantiles of the distribution. The third approach uses the ensemble probability density function, where we suppose that the prediction error of a reference run (e.g., a deterministic run or the ensemble control member) can be estimated from the ensemble distribution at the same time. Unlike the categorical approach, the probabilistic approach assigns a probability to each prediction.

For each of the 34 cases considered here, four categorical WD predictions (GFS, ECMWF, the GEFS control run starting at 00 (c00), and GEFS mean) and the probabilistic predictions based on GEFS, using the three different approaches mentioned above, are compared.

The categorical approach is based on the definition of WD adopted for the observed data: a wind day event is predicted 72 (48) hours in advance if, within the interval 49-72 (25-48) hours, there are at least 3 consecutive hours with wind speed greater than 7 m/s. From now on, we will indicate 72 (48) hours in advance to represent the time window from 49 to 72 (25 to 48) hours.

The first probabilistic approach considered here is the traditional way to use the ensemble members of a generic Ensemble Prediction System (EPS). This method, here called as percentage method ($WDP_{\%}$), consists in counting the ensemble members predicting a WD event. This is equivalent to apply the categorical approach to each ensemble member. The probability of occurrence of a WD for this method is given by the percentage of ensemble members predicting a wind day:

$$WDP_{\%} = \frac{(\text{number of ensembles predicting WD})}{(\text{number of ensemble members})} \cdot 100\ \% \qquad (1)$$

The second approach proposed here is based on the percentiles of the hourly ensemble distributions, with percentile ranks ranging from 0 to 100, every 1. For each hour of the day, we evaluate the percentiles of the ensemble distribution. Then, instead of ensemble members, we consider the hourly percentile curves. The lowest hourly percentile curve predicting the wind day in accordance with the categorical approach provides the wind day probability for the quantile method (WDP_q):

$$WDP_q = 100 - [\text{lowest hourly percentile curve predicting the wind day}]\ \% \qquad (2)$$

As an example, in fig. 4 we show the estimation of WDP_q on January 5, 2016, at the forecast range +72h. The lowest percentile that meets the WD conditions is the 70th percentile (starting from the bottom), so WDP_q for January 5, 2016 is 30%.

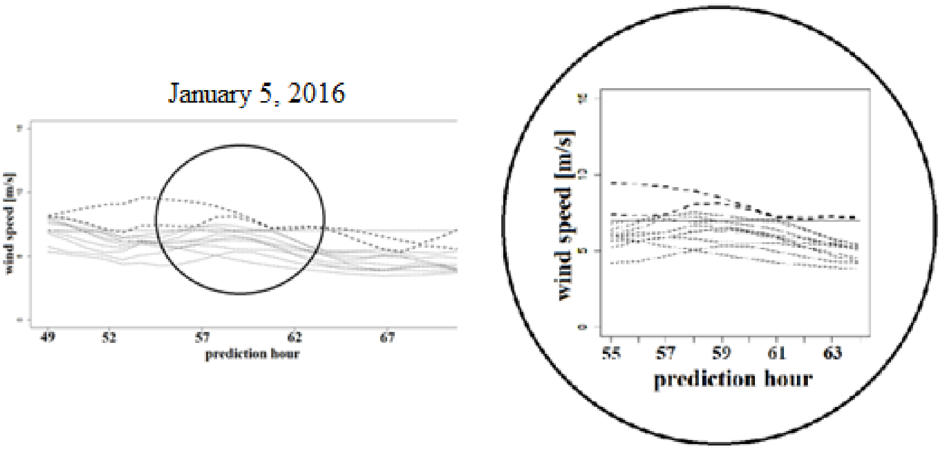

Fig. 4: The lowest percentile meeting the WD conditions is the 70th percentile, so the wind day probability for 05 January, 2016, is 30%. The horizontal dashed line represents the wind speed threshold of 7 m/s. On the right, only the WD hours are shown.

Note that the latter approach comes out to be useful in cases similar to ours, i.e. when the temporal evolution plays an important role. In fact, in case of events defined considering the value of the variables only at one time, the percentage and percentile approaches are equivalent. Since the wind day prediction concerns three consecutive hours, the two approaches may differ. In order to better explain this idea, figures 5 and 6 show an idealized example, considering five ensemble members, where the differences between the two methods are evident. None of the members - shown in the panels from (B) to (F) of fig. 5 - predicts a wind day in accordance with the categorical approach. This means that $WDP_{\%}$ is equal to 0.

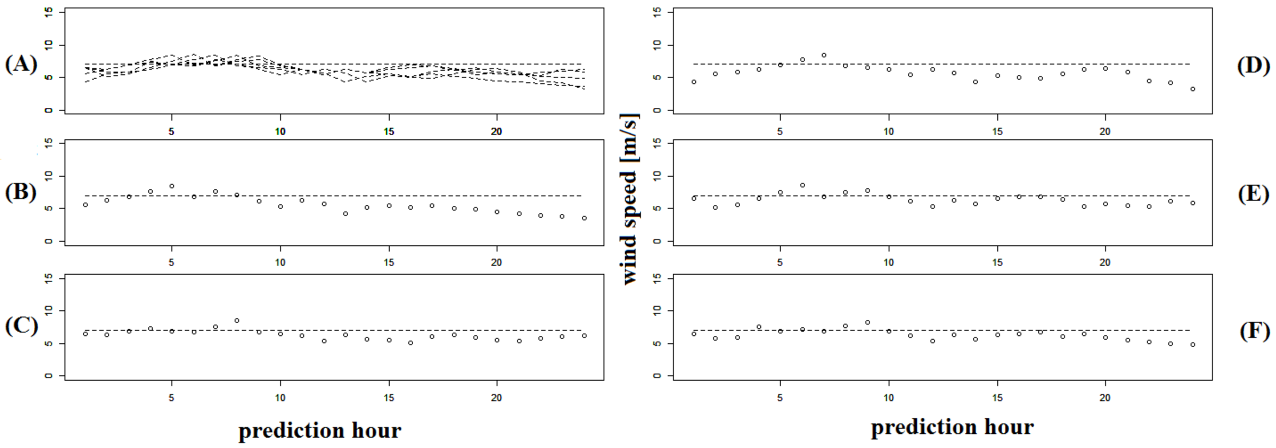

Fig. 5: An idealized example composed of five ensemble members. (A) five ensemble members all together; (B)-(F) single prediction ensemble member. No member predicts wind day in accordance with the categorical approach. The prediction hours refer to the third day of forecast. The horizontal dashed line represents the wind speed threshold of 7 m/s.

In contrast, with reference to fig. 6, the lowest percentile meeting the conditions for WD in accordance with the categorical approach is the 40th percentile, thus WDP_q for this idealized example is 60%.

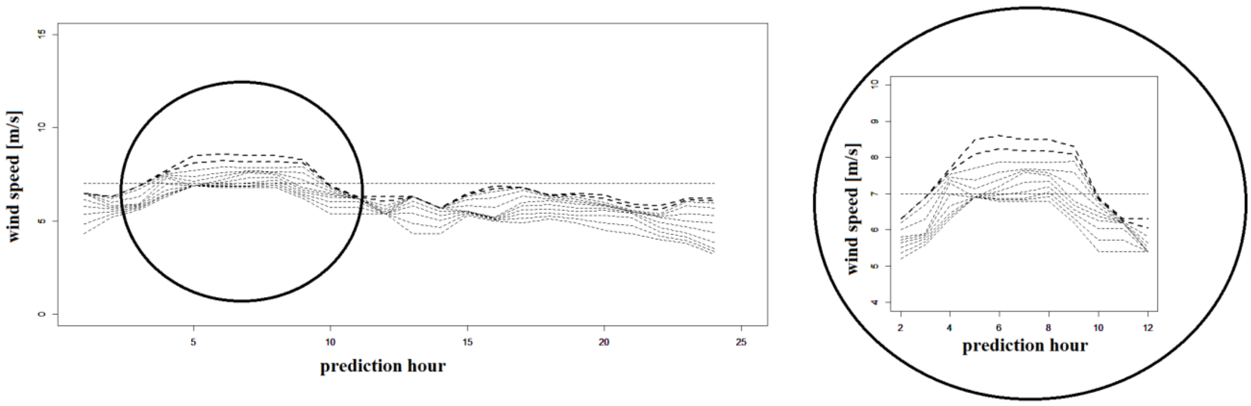

Fig. 6: The quantile temporal evolution for an idealized example. The lowest percentile meeting the WD conditions in accordance with the categorical approach is the 40th percentile. The prediction hours refer to the third day of forecast. The horizontal dashed line represents the wind speed threshold of 7 m/s.

The third approach is based on the measurement theory and in particular on the error theory (Ramirez et al., 2001; Taylor, 1997), which states that every physical measurement is affected by some error or

uncertainty. Although there may be different sources of errors, they can be summarized in two main classes: random and systematic errors.
In the framework of this analysis, the systematic errors may be due to an improper calibration of the global forecast model, whereas the random errors may be due to the uncertainty in the initial conditions and to numerical approximations.
The third proposed approach is based on the ensemble Probability Density Function (PDF); each ensemble member is supposed to be a specific measure of the observed field, assuming that the members are distributed according to a normal distribution (in our analysis, the gamma distribution has been evaluated too). Since the ensemble members include the random uncertainty, only random errors are taken into account in our third proposed approach. A further analysis, described in the result section, is performed to reduce the systematic error.
In general, the ensemble mean cannot be considered as the "best" estimate of the observations since, in our analysis, we found some cases where the ensemble members do not include the measured values within the range of possible outcomes. In some cases, the deterministic GFS run is much closer to the observed value than the ensemble mean. As an example, in fig. 7 we compare the predictions from 49^{th} to 72^{th} hour of the GEFS members (dashed lines) and the deterministic GFS (points and segments) with the observed values (thick line) for December 30, 2016. One can see that the observation is not included within the ensemble range and, at the same time, the deterministic GFS is closer to observations than the ensemble mean.

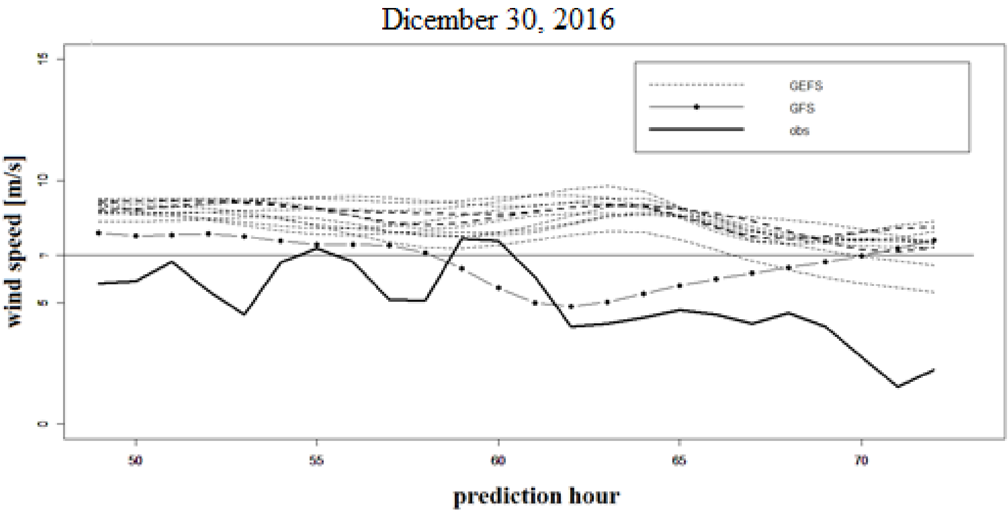

Fig. 7: Comparison of predictions (72 hours in advance) of the GEFS members (dashed lines) and the deterministic GFS (points and segments) with the observed values (thick line) for 30 December, 2016.

In the situation shown in fig. 7, no member "includes" the true values; thus, since all ensemble members are greater than 7 m/s for most of the forecast time, the two previous probabilistic approaches will fail to predict correctly the occurrence of a WD (all members predict WD condition, while the observations do not show WD). In order to take into account this limitation, in the third proposed method we extract only the information on the shape of the random error distribution from the ensemble PDF. The estimate of the "true" value is taken from different alternative approaches: the deterministic GFS, the control member of GEFS, the mean of GEFS, the average between the deterministic GFS and the control member of GEFS, and the average between the deterministic GFS and the mean of GEFS. We assume that the shape of the ensemble distribution is similar to the shape of the forecast error distribution; as a consequence, the ensemble distribution is equivalent to the error distribution around the estimate of the true value.
Hereafter, we describe the proposed procedure in detail. First, we consider an empirical ensemble distribution having the same shape of the original ensemble distribution and mean coincident with the estimate of the true value. From now on, all quantities refer to the empirical distribution.
In this way, the probability to observe the event E_h, that is the observation of a value greater than or equal to the reference threshold $\tilde{x}$ at a specific forecast time h, is:

$$P(E_h) = 1 - C_h(\tilde{x}) \qquad (5)$$

where $C_h(\tilde{x})$ is the Cumulative Distribution Function (CDF) related to the empirical ensemble distribution at time h estimated in $\tilde{x}$ (Brownlee, 165). Since the CDF is the probability that the variable u will take a value less than or equal to $\tilde{x}$, eq. (6) provides its definition

$$C_h(\tilde{x}) = \int_{-\infty}^{\tilde{x}} x_h(u)du \qquad (6)$$

where $x_h(u)du$ is the probability to have, at a specific forecast time h, a value between u and $u+du$. The joint probability to have values greater than or equal to $\tilde{x}$ for three consecutive hours, h, $h+1$ and $+2$, $P(E_h \cap E_{h+1} \cap E_{h+2})$, is obtained from the probability theory (Jeffreys, 1961), specifically from the conditional probability definition. Indeed, given two generic events A and B, we define as conditional probability of A given B - $P(A|B)$ - the probability that the event A occurs when we know that B has already occurred:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (7)$$

where $P(A \cap B)$ is the joint probability of A and B, the probability that both events will occur. Through this concept, from the "compound probability theorem" (eq. 8), we can calculate the probability of the intersection of the two events, the joint probability. We have

$$P(A \cap B) = P(A|B) \cdot P(B) \qquad (8)$$

Then, we get:

$$P(E_h \cap E_{h+1} \cap E_{h+2}) =$$

$$P((E_h \cap E_{h+1}) \cap E_{h+2}) =$$
$$P(E_{h+2}|E_h \cap E_{h+1}) \cdot P(E_h \cap E_{h+1}) =$$
$$P(E_{h+2}|E_h \cap E_{h+1}) \cdot P(E_{h+1}|E_h) \cdot P(E_h)$$

By summarizing,

$$P(E_h \cap E_{h+1} \cap E_{h+2}) = P(E_{h+2}|E_h \cap E_{h+1}) \cdot P(E_{h+1}|E_h) \cdot P(E_h) \qquad (9)$$

The conditional probability value of E_{h+2} given $E_h \cap E_{h+1}$ and the conditional probability value of E_{h+1} given E_h are empirically estimated. Referring to the empirical ensemble distribution, the first conditional probability in eq. (9) is the percentage of ensemble members greater than or equal to the reference threshold at time $h+2$ among all those greater than or equal to the reference threshold at both h and $h+1$ forecast times. Similarly, the second term is the percentage of ensemble members greater than or equal to the reference threshold at the time $h+1$ among all those greater than or equal to the reference threshold at h forecast time.
Once this probability is estimated for each group of three consecutive hours from the 49^{th} to 72^{th} hour, we can calculate the wind day probability for this third method based on PDF, WDP_{pdf}. Eq. (10) provides the probability of wind day occurrence 72 hours in advance.

$$WDP_{pdf} = max_{h=49}^{70}[P(E_{h+2}|E_h \cap E_{h+1}) \cdot P(E_{h+1}|E_h) \cdot P(E_h)] \qquad (10)$$

As an alternative to the Gaussian distribution, we also considered the gamma distribution to represent the error distribution. Gamma is a positive distribution, defined by the shape and scale coefficients, but, differently from the normal distribution, because of its asymmetry, the average does not necessarily coincide with the median (Hogg et al., 2005). Due to these properties, it is suitable to represent the wind speed distribution.
Differently from the categorical prediction, the probabilistic prediction needs to compute the best cut-off in the available set of cases. In other terms, an additional analysis can be applied to define the probability threshold that maximizes the skill of the numerical system. This point will be discussed later in detail.
For the estimation of reliability in both approaches, categorical and probabilistic, we have considered some indices based on the contingency table (see tab. 1) consisting of True positives (*Tp*), True negatives (*Tn*), False positives (*Fp*), and False negatives (*Fn*).

Contingency Table		Event Observed	
		1	0
Event Forecasted	1	*Tp*	*Fp*
	0	*Fn*	*Tn*

Tab. 1: Contingency Table

From this table, we can derive some statistical performance indicators, such as bias (*BIAS*), probability of detection (*POD*), false alarm ratio (*FAR*), threat score (*TS*, also known as Critical Success Index, *CSI*), equitable threat score (*ETS*), and the accuracy. Only for the probabilistic approaches, we have determined the cut-off that maximizes the accuracy index, and we have considered also the Brier Score (*BS* - also known as mean square probability error) and the area under the ROC (relative operating characteristic) curve (*AUC*) representing the hit rate versus the false alarm rate.
For a graphical comparison of the different approaches, we consider the Performance Diagram (Roebber, 2009), which is widely used to visualize multiple measures of forecast quality. This method allows to represent simultaneously and simply four typical performance measures of dichotomous forecasts: *BIAS*, *TS*, *POD*, and success ratio (*SR*), defined as *1-FAR*. In tab. 2, we report the formulas of the aforementioned indices.

BIAS	***POD***	***FAR***	***TS***	***ETS***	***accuracy***
$\frac{Tp+Fp}{Tp+Fn}$	$\frac{Tp}{Tp+Fn}$	$\frac{Fp}{Tp+Fp}$	$\frac{Tp}{Tp+Fp+Fn}$	$\frac{Tp+H_Z}{Tp+Fp+Fn-H_Z}$ $H_Z=\frac{(Tp+Fn)(Tp+Fp)}{Tp+Tn+Fp+Fn}$	$\frac{Tp+Tn}{Tp+Tn+Fp+Fn}$

Brier Score
$\frac{1}{N}\sum_{i=1}^{N}(p_i-o_i)^2$

Tab. 2: Definition of the indicators used for the statistical evaluation: bias (BIAS), probability of detection (POD), false alarm ratio (FAR), threat score (TS), equitable threat score (ETS), accuracy, and Brier Score.

4.Results
4.1 Results on Training dataset

Different statistical methods are here used to evaluate the skill of probabilistic forecasts in comparison with the categorical approach for the prediction of particular meteorological events, such as Wind Days. Table 3 shows the statistical indicators (*BIAS*, *POD*, *FAR*, *TS*, *ETS*, and *accuracy*) when the categorical approach is considered for: GFS, ECMWF, control member of GEFS (c00), ensemble mean (GEFS). *POD* is equal to 1 for GFS and the ensemble mean, meaning that all the observed WD events have been correctly identified (no *Fn*). At the same time, however, they have a higher *FAR* than the other simulations, thus more *FAs* (False Alarms or *Fp*) are present. In particular, the *BIAS* for GFS is equal to 1.88, which means that the number of WD forecasts is almost twice as much the observed WDs.
In general, a *BIAS* greater than 1 does not imply that necessarily all cases of WD are correctly identified and that a null number of *Fn* is present; for example, the control member of the GEFS (c00) shows a *BIAS* of 1.50 (> 1) and a *POD* of 0.94, which means that *Fn* are also present. Despite the

presence of some *Fn*, the run c00 is characterized by a better performance in terms of accuracy due to fewer misclassified (*Fp* + *Fn*) events than the other model runs.

Table 3 shows that, considering the categorical approach, the GEFS model has the best overall performance using either c00 or the ensemble mean. For a graphical, more immediate comparison, fig. 8 shows the Performance Diagram for the four categorical approaches. Since the best model is the closest to the top right corner of the diagram, c00 appears as the best.

DATA	***BIAS***	***POD***	***FAR***	***TS***	***ETS***	***accuracy***
GFS	1.88	1.00	0.47	0.53	0.12	0.59
ECMWF	0.75	0.50	0.33	0.40	0.16	0.65
control member (GEFS)	1.50	0.94	0.38	0.60	0.27	**0.71**
ensemble mean (GEFS)	1.69	1.00	0.41	0.59	0.23	**0.68**
expectation value	**1**	**1**	**0**	**1**	**1**	**1**

Tab. 3: Evaluation indices for the four categorical approaches. The last row shows the best possible value for each index.

Tables 4 and 5 show the statistical performance indicators (*BIAS, POD, FAR, TS, ETS, accuracy, BS,* and *AUC*) obtained for the three probabilistic approaches based on the GEFS model. In particular, tab. 4 shows the results for the percentile and quantile approaches, while tab. 5 shows the result for the approach based on the ensemble PDF using different reference values (GFS, c00, average of GFS and c00, and average of GFS and ensemble mean). For each reference value, we have considered both the normal and gamma distributions to model the error distribution.

By comparing the results shown in tables 3, 4, and 5, it is apparent that, overall, the results using the probabilistic approaches are better than those using the categorical approach. In fact, except for one case (the second row in tab. 5), *TS* and *accuracy* are better for probabilistic approaches than for the best categorical approach (c00), which means that a lower number of misclassified events and a higher number of events classified correctly are simulated. Among the statistical indicators considered here, the *FAR* is one of the most significant indicators of performances because it concerns the false positives ("*false alarms*"). Similar to TS and the accuracy, the *FAR* performs better with probabilistic approaches. Although the categorical approach for ECMWF presents a *FAR* comparable with that shown by probabilistic methods, the *POD* value is the worst among all simulations, meaning that categorical approach for ECMWF shows few *Fp* but more *Fn*. Among the probabilistic approaches, the best result is obtained using the method based on the ensemble PDF, with c00 as reference and the gamma distribution to model the error distribution. In addition to the better accuracy, this approach shows improvements in almost all the evaluation indexes. In particular, its *FAR* and *POD* values are a compromise between the results of percentile and quantile methods. In fact, the percentile method shows a better *POD* while the quantile method shows a better *FAR*. The selected approach based on PDF presents improvements for both indexes. Figure 9 shows the Performance Diagram summarizing the performance of the three probabilistic approaches (for the method based on the PDF, only the best configuration is shown).

Probabilistic method	***BIAS***	***POD***	***FAR***	***TS***	***ETS***	***accuracy***	***BS***	***AUC***
percentile	1.31	0.94	0.29	0.68	0.42	0.79	0.24	0.80
quantile	1.19	0.88	0.26	0.67	0.42	0.79	0.22	0.78
expectation value	**1**	**1**	**0**	**1**	**1**	**1**	**0**	1

Tab. 4: Evaluation indices for the two probabilistic approaches based on the percentile and quantile. The last row includes the best possible value for each index.

reference of true value	***distribution***	***BIAS***	***POD***	***FAR***	***TS***	***ETS***	***accuracy***	***BS***	***AUC***
deterministic GFS	Normal	1.63	1.00	0.38	0.62	0.27	0.71	0.33	0.69
deterministic GFS	Gamma	1.19	0.75	0.37	0.52	0.22	0.68	0.35	0.69
control member (c00)	Normal	1.31	0.94	0.29	0.68	0.42	0.79	0.25	0.77

control member (c00)	**Gamma**	**1.25**	**0.94**	**0.25**	**0.71**	**0.48**	**0.82**	**0.25**	**0.80**
Average(GFS,c00)	Normal	1.56	1.00	0.36	0.64	0.32	0.74	0.29	0.73
Average(GFS,c00)	Gamma	1.44	1.00	0.30	0.70	0.43	0.79	0.29	0.78
Average(GFS,ensMean)	Normal	1.50	1.00	0.33	0.67	0.37	0.76	0.30	0.73
Average(GFS,ensMean)	Gamma	1.38	0.94	0.32	0.65	0.37	0.76	0.27	0.75
expectation value	// //	1	1	0	1	1	1	**0**	1

Tab. 5: Evaluation indices for the probabilistic method based on the PDF, for different distribution models and reference true values. The last row includes the best possible value for each index. The best configuration is bolded.

The prediction of WD depends critically on the ability of the models to simulate correctly the wind speed. In this framework, it is relevant to analyze tab. 3 again. The latter shows that the ECMWF model, differently from other runs, has a *BIAS* less than one and a smaller *POD*, meaning that the forecasted WDs are less than those actually observed. However, the ECMWF *accuracy* is comparable with the other cases (and is even better than GFS): this means that the ECMWF model has a minor number of *Tp* and *Fp*, but more *Tn*. This is a consequence of the underestimation of wind speed in the ECMWF run, which is apparent in fig. 10, where the boxplots of the differences between the wind speed predicted within 49-72 hours (using respectively GFS, ECMWF, c00, and the ensemble mean) and the observed wind speed are shown. A positive mean error leads to a higher number of *Tp* and *Fp*; conversely, a negative mean error leads to more *Tn* and less *Fp* (as in the ECMWF run).
In order to consider the error in the prediction of WD due to the wind speed mean error (i.e., the model bias) and to maximize the model *accuracy*, one can try to change the wind speed threshold and the number of consecutive hours used in the definition of WD when this is applied to the model data. For this purpose, we tested different wind speed thresholds (between 5 m/s and 9 m/s) and changed the number of consecutive hours (from 1 to 5). Table 6 shows the best results in terms of accuracy obtained for each approach. As a consequence of the generally positive model bias, all models need thresholds greater than 7 m/s, apart from the ECMWF model, which needs a reduced threshold due to its average underestimation. Also, the accuracy gain of the ECMWF run, which is lower than for the other deterministic runs, as shown in tab. 6, depends on the larger boxplot representing the wind speed error for ECMWF in fig. 10; because of the larger errors, there is no wind speed threshold suitable for most cases.
Again, the method based on the ensemble PDF shows the best performance, with an accuracy of 94% obtained when we take the average of the deterministic GFS and probabilistic GEFS (ensemble mean or control) as reference value. The Performance Diagrams in Figures 11 and 12 show the improvement of the categorical and probabilistic approaches after the wind speed threshold and the number of consecutive hours are changed, as shown in tab. 6. In particular, fig. 12 shows the methods based on the ensemble percentiles, on the ensemble quantiles, the three best configurations in tab. 6 based on the PDF (obtained when the average between GFS and c00 or the ensemble mean is adopted as reference value), and the case showing the best performances in tab. 5 (c00 member as reference value and the gamma distribution to model the error distribution). All the models show remarkably better performances after the changes in the threshold and in the number of hours are considered.

4.2 Validation

Since our proposed approach requires the evaluation of one or more external parameters (i.e., the cut-off and the probability threshold), it has become settled practice to apply the method to a test and to a validation dataset in order to assess how well it performs in comparison with real data different from those used for the analysis. In our work, we used a single dataset of 34 events (including data recorded in 2016). For the test analysis, we adopted the leave-one-out cross-validation strategy, i.e. using one case as the test dataset and the remaining cases as the training set. In other terms, we have fixed the threshold by considering 33 cases (used as training set) and we have estimated the WD probability on the remaining case (used as validation set). The statistical analysis is complete when all cases were considered in the validation process. In this way, the result is more reliable because the cut-off is chosen independently from the validation data.

For validation analysis, we have considered an independent dataset of 66 case studies covering 2017, including 29 observed WDs and 37 no WDs, representing situations difficult to predict. The analysis on test and on validation dataset was carried out only for the GFS and GEFS data by comparing the categorical and probabilistic approaches (we do not have considered the ECMWF data in this additional analysis). For the analysis on the validation dataset (2017 cases), we have used the best cut-off estimated using the 34 case studies in 2016; in the leave-one-out analysis, we have adopted for each case study a specific threshold estimated from the 33 remaining case studies in 2016.

Table 7 shows the result in terms of accuracy for both classical and no classical WD definition, i.e. changing the wind speed threshold and the number of consecutive hours. For each method and for both test and validation analysis, we have used the values reported in tab. 6. As expected, the performance on the test set is similar (although not equal) to the performance on the validation set and both are lower than the performance on the training set.

Furthermore, the results in tab. 7 confirms that, in general, the performance based on probabilistic approach is better than the categorical approach. Among probabilistic approaches, the best result is obtained using the ensemble PDF method, using as true value a combination of deterministic (GFS) and probabilistic information (ensemble mean or control member of GEFS). The validation analysis confirms that the improvement due to the use of probabilistic methods consists in the reduction of false positives (not shown).

We test the above procedures also considering the forecast time interval in the range 25-48 hours. Similar performances are found (not shown).

data or method		**accuracy 7m/s-3h**	**new threshold [m/s]**	**new # hours**	**Best accuracy**	**accuracy gain [%]**
GFS		0.59	9	4	0.82	39
ECMWF		0.65	6	3	0.76	17
control member (GEFS)		0.71	8	5	0.88	24
Ensemble mean (GEFS)		0.68	8	3	0.85	25
percentile		0.79	8	3	0.88	11
quantile		0.79	9	5	0.88	11
PDF-deterministic GFS	**Normal**	0.71	9	3	0.88	24
PDF-deterministic GFS	**Gamma**	0.68	9	4	0.88	29
PDF-control member (c00)	**Normal**	0.79	8	5	0.88	11
PDF-control member (c00)	**Gamma**	0.82	9	2	0.88	7
PDF-Average(GFS,c00)	**Normal**	0.74	8	5	0.94	27
PDF-Average(GFS,c00)	**Gamma**	0.79	8	5	0.94	19
PDF-Average(GFS,ensMean)	**Normal**	0.76	8	5	0.94	24
PDF-Average (GFS,ensMean)	**Gamma**	0.76	9	5	0.91	20

Tab. 6: For all considered methods and models: the accuracy related to the classical WD definition and the accuracy when a new value for wind speed threshold and number of consecutive hours is used.

data or method	**Accuracy 7m/s-3h (classical WD definition)**			**Best accuracy (no classical WD definition)**		
	Training (34 case studies - 2016)	**Test (Leave-one-out 2016)**	**Validation (66 case studies - 2017)**	**Training (34 case studies - 2016)**	**Test (Leave-one-out 2016)**	**Validation (66 case studies - 2017)**
GFS	0.59	0.59	0.68	0.82	0.82	0.79
control member (GEFS)	0.71	0.71	0.70	0.88	0.88	0.79

Ensemble mean (GEFS)		0.68	0.68	0.68	0.85	0.85	0.80
Percentile		0.79	0.72	0.71	0.88	0.82	0.79
Quantile		0.79	0.74	0.77	0.88	0.85	0.80
PDF-deterministic GFS	Normal	0.71	0.68	0.76	0.88	0.85	0.86
PDF-deterministic GFS	Gamma	0.68	0.64	0.74	0.88	0.82	0.86
PDF-control member (c00)	Normal	0.79	0.79	0.80	0.88	0.85	0.82
PDF-control member (c00)	Gamma	0.82	0.79	0.80	0.88	0.85	0.82
PDF-Average(GFS,c00)	Normal	0.74	0.75	0.79	0.94	0.85	0.85
PDF-Average(GFS,c00)	Gamma	0.79	0.77	0.79	0.94	0.85	0.85
PDF-Average(GFS,ensMean)	Normal	0.76	0.75	0.80	0.94	0.85	0.85
PDF-Average (GFS,ensMean)	Gamma	0.76	0.75	0.80	0.91	0.88	0.83

Tab. 7: For all considered methods (categorical and probabilistic), models (except ECMWF data), and dataset (training, test, and validation set): the accuracy for the classical WD definition and the accuracy when a new value for wind speed threshold and number of consecutive hours is used.

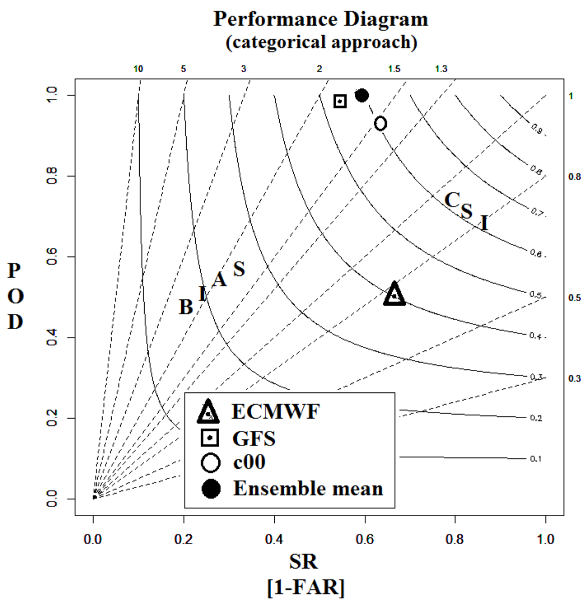

Fig. 8: Performance Diagram of the four models from the categorical approach: ECMWF, GFS, c00 (control member of the GEFS), and Ensemble mean of the GEFS.

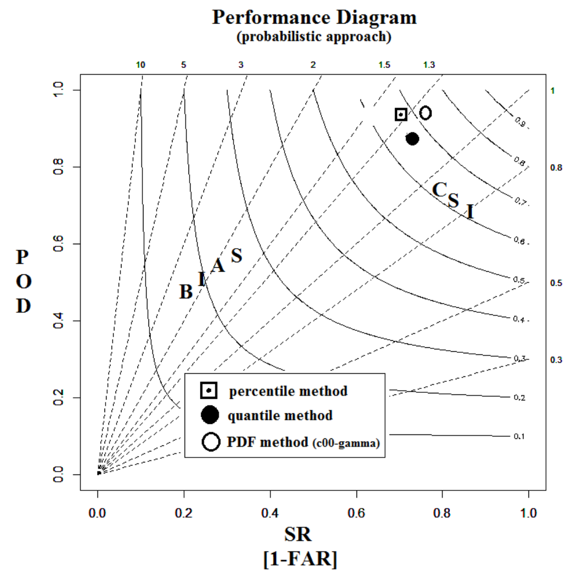

Fig. 9: Performance Diagram of the three considered probabilistic approaches: method based on ensemble percentiles, method based on the ensemble quantiles, and method based on the PDF. For the last method, we report only the best configuration, obtained with the c00 member as reference value and the gamma distribution to model the error distribution.

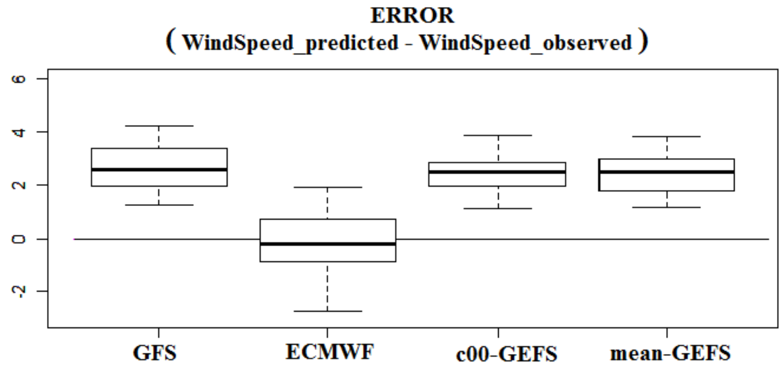

Fig. 10: Boxplots of the differences between the wind speed predicted within 49-72 hours respectively using GFS, ECMWF, c00, and the ensemble mean, and the observed wind speed.

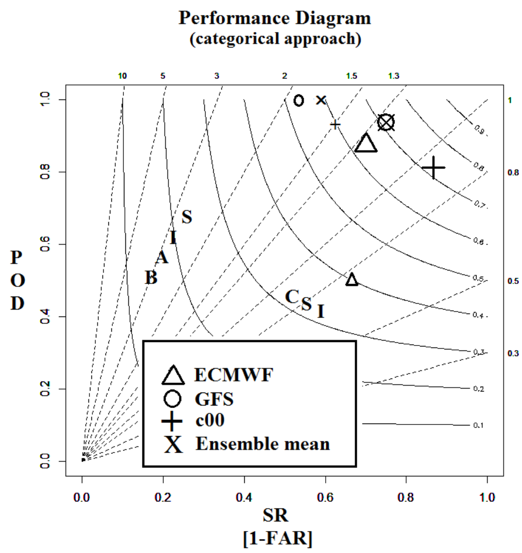

Fig. 11: Performance Diagram after the wind speed threshold and the number of consecutive hours are changed as shown in Tab. 5 in the WD definition applied to the model results. The smaller symbols refer to the previous results while the larger symbols refer to the improved results.

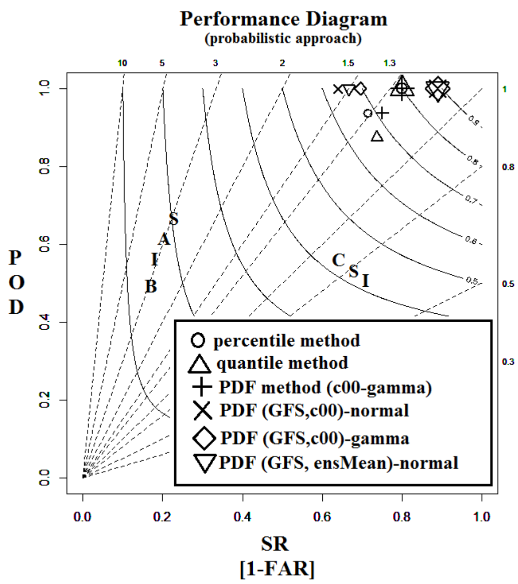

Fig. 12: Performance Diagram after the wind speed threshold and the number of consecutive hours are changed as shown in tab. 5 in the WD definition applied to the model results. The smaller symbols refer to the previous results while the larger symbols refer to the improved results.

5. Conclusion and discussion

In this study, we propose an alternative approach to estimate the probability of occurrence of a local meteorological event covering a time period of a few hours. To estimate its skill, we compare this method with classical categorical and probabilistic approaches using different global models: ECMWF, GFS, and GEFS.
In particular, we focus our attention on wind speed forecasts (range 49-72 hours) in the area around the city of Taranto, located in Apulia region (southeastern Italy), to forecast the so called "Wind Days", i.e. days with observed northwesterly wind above the threshold of 7 m/s for at least 3 hours; these conditions are generally associated with an increase in the concentration of pollutants, and may strongly affect the air quality in town. Our analysis includes 34 case studies covering 2016. These cases are opportunely chosen in order to have a balanced dataset of 16 WD (all those occurring in 2016) and 18 no WD, including in the latter category the events that are very difficult to predict, i.e. at the border between the two categories.
Our proposed method aims to reduce the error of a forecast based on an ensemble system. We assume that the ensemble spread provides information on the error distribution (ensemble PDF) around the estimate of the true value provided by a deterministic run or by a combination of a deterministic run and the ensemble mean). To compare this new approach with other more traditional methods, various statistical indices are evaluated.
Results show that, among the categorical approaches based on deterministic model outputs, the GEFS model shows the overall best performance, using either the control member or the ensemble mean. In general, results using the ensemble systems are better than those using the deterministic runs. Among the probabilistic approaches, the best result (accuracy of 82%) is obtained using the ensemble PDF method, with the control run as true value and the gamma distribution to model the error distribution.
Since the systematic error in the WD prediction is related to the wind speed mean error of the model, one can try to change the wind speed threshold and the number of consecutive hours used in the definition of WD when this is applied to the model data. Apart from the ECMWF model, which needs a reduced threshold due to its average underestimation, all models need thresholds greater than 7 m/s and a larger number of consecutive hours to improve their skill, as a consequence of the generally positive model bias.
The skill of all the models is remarkably better after the wind speed threshold and the number of consecutive hours are changed. In particular, the method based on the ensemble PDF shows the best performance (accuracy of 94%) when we take the average of the deterministic GFS and probabilistic GEFS (ensemble mean or control) as reference value.
The analysis on test (leave-one-out strategy in 2016) and validation datasets (66 cases in 2017) confirms the previous outcomes. The results presented here concern a dataset of selected WD and no WD cases. We select cases at the border of the two categories, which are more difficult to predict. For this reason, better results are expected using a more general dataset, consisting also of WD cases predictable in a clear-cut manner.
We test our approach by considering the reliability in the prediction of WD, which is closely related to the wind speed and wind direction prediction. The new approach proposed here could be extended to other situations where the occurrence of an event, associated with the overcoming of a threshold in a time window, should be predicted. Our method could be easily applied to the prediction of other extreme events, e.g. to estimate the probability of heavy precipitation at a fixed point.
The authors are planning to validate this methodology for future real-time predictions of WD, in order to check how the value of the statistical parameters changes considering a different sample of events.

Acknowledgment:

M.M. Miglietta gratefully acknowledges the funding from theEuropean Commission (Project “CEASELESS”, grant agr eement no. 730030).

Pollice A. and Tateo A. were supported by the PRIN2015 project "Environmental processes and human activities: capturing their interactions via statistical methods (EPHASTAT)" funded by MIUR - Italian Ministry of University and Research.

References

Alpert, J., & Wang, J. (2005, January). A client application for real time Nomads at NCEP todissemination NOAA's information data base. *In 21st International Conference on Interactive Information Processing Systems.*

Amodio, M., Andriani, E., de Gennaro, G., Di Gilio, A., Ielpo, P., & ... and Tutino, M. (2013). How a Steel Plant Affects Air Quality of a Nearby Urban Area: A Study on Metals and PAH concentrations. *Aerosol Air Qual. Res, 13(2)*, pp. 497–508.

Berner, J., Ha, S. Y., Hacker, J. P., Fournier, A., & ... Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Monthly Weather Review, 139(6)*, pp. 1972-1995.

Boisserie, M., Arbogast, P., Descamps, L., Pannekoucke, O., & & Raynaud, L. (2014). Estimating and diagnosing model error variances in the Météo-France global NWP model. *Quarterly Journal of the Royal Meteorological Society, 140(680)*, pp. 846-854.

Brownlee, K. A. (165). *Statistical theory and methodology in science and engineering (Vol. 150).* New York: Wiley.

Buizza, R., Milleer, M., & Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society, 125(560)*, pp. 2887-2908.

Cassola, F., & Burlando, M. (2012). Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Applied energy, 99*, pp. 154-166.

Corazza, M. S. (2018). The ARPAL operational high resolution Poor Man’s Ensemble, description and validation. *Atmospheric Research, 203*, pp. 1-15.

Ebert, E. E. (2001). Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Monthly Weather Review, 129(10), 2461-2480., 129*(10), pp. 2461-2480.

Ebert, E. E. (2001). Ability of a poor man's ensemble to predict the probability and distribution of precipitation. pp. 129.10: 2461-2480.

Fedele, F., Menegotto, M., Trizio, L., Angiuli, L., Guarnieri Calò Carducci, A., Bellotti, R., . . . and Assennato, G. (2014). Meteorological effects on pm10 concentrations in an urban industrial site: a statistical analysis. *In Conference proceedings: 1st International Conference on Atmospheric Dust*, pp. 162-167.

Fedele, F., Miglietta, M. M., Perrone, M. R., Burlizzi, P., Bellotti, R., Conte, D., & ... and Guarnieri Calò Carducci, A. (2015). Numerical simulations with the WRF model of water vapour vertical profiles: A comparison with LIDAR and radiosounding measurements. *Atmospheric Research, 166*, pp. 110-119.

García-Ortega, E. L. G. (2017). Performance of multi-physics ensembles in convective precipitation events over northeastern Spain. *Atmospheric Research, 190*, pp. 55-67.

Guan, H., Cui, B., & Zhu, Y. (2015). Improvement of statistical postprocessing using GEFS reforecast information. *Weather and Forecasting, 30(4)*, pp. 841-854.

Hamill, T. M., Bates, G. T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneau Jr, T. J., & ... & Lapenta, W. (2013). NOAA's second-generation global medium-range ensemble reforecast dataset. *Bulletin of the American Meteorological Society, 94(10),* , pp. 1553-1565.

Hogg, R. V., McKean, J., & ... Craig, A. T. (2005). *Introduction to mathematical statistics.* Pearson Education.

Jeffreys, H. (1961). *Theory of probability (3rd edt.).* Oxford university press.

Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts. pp. 409-418.

Libonati, R. T. (2008). Correction of 2 m-temperature forecasts using Kalman filtering technique. *Atmospheric Research, 87(2),* , pp. 183-197.

Lorenz, E. (2000). The butterfly effect. pp. 91-94.

Magnusson, L., Leutbecher, M., & Källén, E. (2008). Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF ensemble prediction system. *Monthly Weather Review, 136(11),* pp. 4092-4104.

Mastrantonio, G., Pollice, A., & Fedele, F. (2018). Distributions-oriented wind forecast verification by a hidden Markov model for multivariate circular–linear data. *Stochastic Environmental Research and Risk Assessment, 32(1),* pp. 169-181.

Miglietta, M. M., Mastrangelo, D., & Conte, D. (2015). Influence of physics parameterization schemes on the simulation of a tropical-like cyclone in the Mediterranean Sea. *Atmospheric Research, 153, 153,* pp. 360-375.

Miglietta, M., Manzato, A., & Rotunno, R. (2016). Characteristics and Predictability of a Supercell during HyMeX SOP1. *Q. J. Roy. Meteor. Soc, 142,* pp. 2839-2853.

Molteni, F., Buizza, R., Palmer, T. N., & Petroliagis, T. (1996). The ECMWF ensemble prediction system: Methodology and validation. *Quarterly journal of the royal meteorological society, 122(529),* pp. 73-119.

Murphy, J. M. (1988). The impact of ensemble forecasts on predictability. *Quarterly Journal of the Royal Meteorological Society, 114(480),* pp. 463-493.

Palmer, T., & Hagedorn, R. (. (2006). Predictability of weather and climate. *Cambridge University Press.*

Pelosi, A., Medina, H., Van den Bergh, J., Vannitsem, S., & Chirico, G. B. (2017). Adaptive Kalman Filtering for Postprocessing Ensemble Numerical Weather Predictions. *Monthly Weather Review, 145(12),* pp. 4837-4854.

Pelosi, A., Medina, H., Villani, P., D'Urso, G., & Chirico, G. B. (2016). Probabilistic forecasting of reference evapotranspiration with a limited area ensemble prediction system. *Agricultural water management, 178,* pp. 106-118.

Perera, K. C., Western, A. W., Nawarathna, B., & George, B. (2014). Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. *Agricultural and forest meteorology, 194,* pp. 50-63.

Ramirez, S., Seidel, R., & ... Hiebert, J. (2001). *Physics 218 Lab Manual 7th Ed. by.* AMU, College Station: Hayden-McNeil Publishing.

Regione, P. (2012). Piano contenente le prime misure di intervento per il risanamento della qualità dell'aria nel quartiere Tamburi (Ta) per gli inquinanti PM10 e Benzo(a)Pirene ai sensi del D.lgs.155/2010 art. 9 comma 1 e comma 2.

Roebber, P. (2009). Visualizing Multiple Measures of Forecast Quality. *Weather and Forecasting, 24(2),* pp. 601-608.

Schefzik, R. (2017). Ensemble calibration with preserved correlations: unifying and comparing ensemble copula coupling and member-by-member postprocessing. *Quarterly Journal of the Royal Meteorological Society*, *143(703)*, pp. 999-1008.

Sokol, Z. M. (2017). Probabilistic precipitation nowcasting based on an extrapolation of radar reflectivity and an ensemble approach. *Atmospheric Research*, *194*, pp. 245-257.

Stensrud, D. J., Brooks, H. E., Du, J. T., & Rogers, E. (1999). Using ensembles for short-range forecasting. *Monthly Weather Review, 127(4)*, pp. 433-446.

Tateo, A., Bellotti, R., Fedele, F., Carducci Guarnieri Calò, A., & ... and Pollice, A. (2015). Post-processing of the Weather Research and Forecasting (WRF) mesoscale model by Artificial Neural Networks. *GRASPA-SIS Biennial Conference.* Bari.

Tateo, A., Miglietta, M. M., Fedele, F., Menegotto, M., Monaco, A., & Bellotti, R. (2017). Ensemble using different Planetary Boundary Layer schemes in WRF model for wind speed and direction prediction over Apulia region. *Advances in Science and Research, 14*, pp. 95-102.

Taylor, J. (April 1997). *An Introduction to Error Analysis 2nd Ed.* University Science Books.

Trizio, L., Angiuli, L., Menegotto, M., Fedele, F., Giua, R. M., & ... Assennato, G. (2016). Effect of the Apulia air quality plan on PM10 and benzo (a) pyrene exceedances. *Global Journal of Environmental Science and Management*, pp. 95-104.

Vannitsem, S. (2008). Dynamical properties of MOS forecasts: Analysis of the ECMWF operational forecasting system. *Weather and Forecasting, 23(5)*, pp. 1032-1043.

Wang, G. W. (2015). Improvement of forecast skill for severe weather by merging radar-based extrapolation and storm-scale NWP corrected forecast. *Atmospheric Research, 154*, pp. 14-24.

Whitaker, J. S., & Loughe, A. F. (1998). The relationship between ensemble spread and ensemble mean skill. *Monthly weather review*, *126(12)*, pp. 3292-3302.

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100).

Zacharov, P. &. (2009). Using the fractions skill score to assess the relationship between an ensemble QPF spread and skill. *Atmospheric Research*, *94(4)*, pp. 684-693.

Zhang, Z. H., & Krishnamurti, T. N. (1997). Ensemble forecasting of hurricane tracks. *Bulletin of the American Meteorological Society*, *78(12)*, pp. 2785-2796.