Série des Documents de Travail

n° 2017-88
Probabilistic forecasting of the wind energy resource at the monthly to seasonal scale

B. ALONZO[1]
P. DROBINSKI[2]
R. PLOUGONVEN[3]
P. TANKOV[4]

[1] IPSL; LMD; CNRS; Ecole Polytechnique; Université de Paris-Saclay; Laboratoire de Probabilités et Modéles Aléatoires, Université Paris Diderot - Paris 7, E-mail: alonzo@math.univ-paris-diderot.fr

[2] IPSL; LMD; CNRS; Ecole Polytechnique; Université de Paris-Saclay, E-mail: philippe.drobinski@lmd.polytechnique.fr

[3] IPSL; LMD; CNRS; Ecole Polytechnique; Université de Paris-Saclay, E-mail: riwal.plougonven@polytechnique.org

[4] CREST; ENSAE ParisTech, E-mail: peter.tankov@ensae.fr

Probabilistic forecasting of the wind energy resource at the monthly to seasonal scale

Bastien Alonzo[a,b], Philippe Drobinski[a], Riwal Plougonven[a], Peter Tankov[c]

[a] *IPSL/LMD, CNRS, Ecole Polytechnique, Université de Paris-Saclay, Palaiseau, France*
[b] *Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot - Paris 7, Paris, France.*
[c] *CREST - ENSAE ParisTech, Palaiseau, France*

Abstract

We build and evaluate a probabilistic model designed for forecasting the distribution of the daily mean wind speed at the seasonal timescale in France. On such long-term timescales, the variability of the surface wind speed is strongly influenced by the atmosphere large-scale situation. Our aim is to predict the daily mean wind speed distribution at a specific location using the information on the atmosphere large-scale situation, summarized by an index. To this end, we estimate, over 20 years of daily data, the conditional probability density function of the wind speed given the index. We next use the ECMWF seasonal forecast ensemble to predict the atmosphere large-scale situation and the index at the seasonal timescale. We show that the model is sharper than the climatology at the monthly horizon, even if it displays a strong loss of precision after 15 days. Using a statistical postprocessing method to recalibrate the ensemble forecast leads to further improvement of our probabilistic forecast, which then remains sharper than the climatology at the seasonal horizon.

Keywords: Wind energy, Wind speed forecasting, Seasonal forecasting, Probabilistic forecasting, Ensemble forecasts, Ensemble model output statistics

Email addresses: `bastien.alonzo@lmd.polytechnique.fr` (Bastien Alonzo), `philippe.drobinski@lmd.polytechnique.fr` (Philippe Drobinski), `riwal.plougonven@lmd.polytechnique.fr` (Riwal Plougonven), `peter.tankov@ensae.fr` (Peter Tankov)

1. Introduction

In the recent years, energy transition has been on the forefront of political and societal issues, mainly due to the increasing awareness of the need to act against the climate disruption. This has led many countries to encourage the use of renewable energy. Since 2008, the European Union (EU) targets 20% of renewable energy contribution to the total energy mix by 2020, and 27% by 2030. Consequently, wind energy has seen a major growth in Europe. To give an idea of this sharp increase, the total installed wind power capacity in the EU has changed from 12.9 GW in 2000 to 141.6 GW in 2015 (EWEA (2016)). The actual share in the final consumption met by wind energy in the EU was 11.4% in 2015 (EWEA (2016)).

At such scales, the variability of the wind power production due to the natural intermittency of the wind resource becomes a critical issue for successful network integration of this source of energy (Albadi and El-Saadany (2010)). As a consequence, the interest and demand for near-surface wind speed forecasts has seen a major boost. Numerous methods exist for forecasting wind speed at different horizons, motivated by different applications (Chang (2014); Soman et al. (2010)). Many studies focus on the short-term scale ranging from several minutes to one day (Carpinone et al. (2015); Gomes and Castro (2012); Stesfos (2002)). Medium-term forecasting methods, ranging typically from 3 days up to 10 days, have also been investigated in depth (Barbounis et al. (2006); Taylor et al. (2009); Wytock and Kolter (2013)). On much longer timescales and with very different motivations, the impact of the climate change on wind speeds has also been adressed (Najac et al. (2009); Pryor and Barthelmie (2010); Sailor and M. Smith (2008)).

Whereas both relatively short and very long timescales have been thoroughly studied, the intermediate timescale going from monthly to seasonal horizon is a research topic for which not so many studies exist. This timescale is becoming very important for the transmission system operators (TSOs) as the proportion of intermittent ressources in the energy mix increases. The TSOs are responsible for balancing the supply and demand of energy and they are required to make seasonal projections, e.g., to guarantee the security of energy supply during the coming winter, which becomes more difficult with the the increased variability of energy production. The risk of not being able to satisfy the energy demand may be quantified in terms of the notion of Loss of load expectation (LOLE). Quoting from NationalGrid (2016), the LOLE is a “measure of the risk across the whole winter of demand exceeding supply under normal operation. It gives an indication of the amount of time across the whole winter that the System Operator may need to call on a range of emergency balancing tools to increase supply or reduce demand." For instance, a cold winter characterized by weaker winds than normal may in some cases lead to a lack of energy if not enough other production means have been made available upstream to meet the energy demand.

Among the few existing studies of long term wind speed forecasting, (Azad et al. (2014); Bilgili et al. (2007)) advocate the use of Artificial Neural Networks (ANN) for forecasting average monthly wind speed. These studies give an accurate estimate of the trend of the wind speed at the yearly horizon but provide limited information on the wind variability at higher frequencies. ANN models have also been used for forecasting daily mean wind speed at the seasonal scale providing more information on the wind variability within a given season for energy production evaluation (Guo et al. (2012); J. Wang et al. (2015); More and Deo (2003)).

These studies provide 'point forecasts', which give one value for the wind energy production at the specified horizon, but do not consider the uncertainty on the forecast (as a rule, forecast uncertainty is difficult to quantify with neural networks since the underlying probabilistic model is not easy to define). At such timescales, the idea of point forecast can be very questionable due to the dominant chaotic nature of the atmospheric system at the timescales exceeding typically 10 days. At this long-term horizon, the idea of probabilistic forecasting therefore gains sense. Indeed, forecasting centers such as the European Center for Medium-range Weather Forecasts (ECMWF) use ensemble forecasts to take into account the uncertainty due to the growing small errors in the atmospheric system. Surface wind speed is a variable that is not provided by such prediction models because of its complexity and interaction with the surface, but valuable information on the general circulation of the atmosphere can be retrieved from such forecasts. Indeed, several works have confirmed the predictability at seasonal timescales of recurrent oscillating patterns in the atmosphere, such as the El Nino (Cassou (2008); Owen and Palmer (1987)) or the North Atlantic Oscillation (NAO) (Davies et al. (1997); Rodwell et al. (1999)). On the other hand, large-scale atmospheric patterns have already been shown to partly explain the surface wind speed in France at such timescales (Alonzo et al. (2017)).

In this paper, our aim is to use long-term forecasts of large-scale circulation patterns provided by ECMWF to obtain probabilistic long-term forecasts of local surface wind speeds. Our approach is to build a probabilistic model describing the relationship between the local surface wind speed and the large-scale circulation of the atmosphere, summarized by a single purpose-built index. To this end we estimate the conditional probability density function of the wind speed given the index by gaussian kernel density estimation over 20 years of daily data. We next use the ECMWF seasonal forecast ensemble to predict the large-scale situation of the atmosphere and the index at the seasonal timescale. The prediction of the index is then plugged into our model, to obtain probabilistic forecasts of the surface wind speed. The ensemble forecast displays a growing uncertainty with time leading to an increase of the confidence interval width predicted by the probabilistic model. We show that the model is sharper than the climatology at the horizon of one month, even if it displays a strong loss of precision after 15 days. Using the statistical postprocessing method EMOS (Ensemble Model Output Statistics) to recalibrate the ensemble forecast leads to a further improvement of our probabilistic forecast, which then remains sharper than the climatology at the seasonal horizon.

This paper is structured as follows. Section 2 describes the method to build the probabilistic model as well as the data used in this study. In section 3, the performance of the model is assessed. In section 4, the probabilistic model is used to forecast the wind speed at the monthly and seasonal horizon by applying it to seasonal ensemble forecasts of large scale circulation patterns of the atmosphere.

2. Data & methods

2.1. Data : ECMWF reanalysis and forecasts

In this paper, we use the so called "perfect model" approach meaning that the ECMWF ERA-I reanalysis is considered as the reality. This is justified by the comparison of ECMWF products and observation, in particular for surface wind speed (Jourdier (2015)). The model is estimated and evaluated the surface wind speed retrieved from this data. We use ECMWF reanalysis for 37 years between 1 January 1979 and 31 December 2015[1].

The basic idea of this work is to link the large-scale circulation of the atmosphere with the daily mean surface wind speed distribution in France. The

[1]ECMWF Data are available at http://apps.ecmwf.int/datasets/

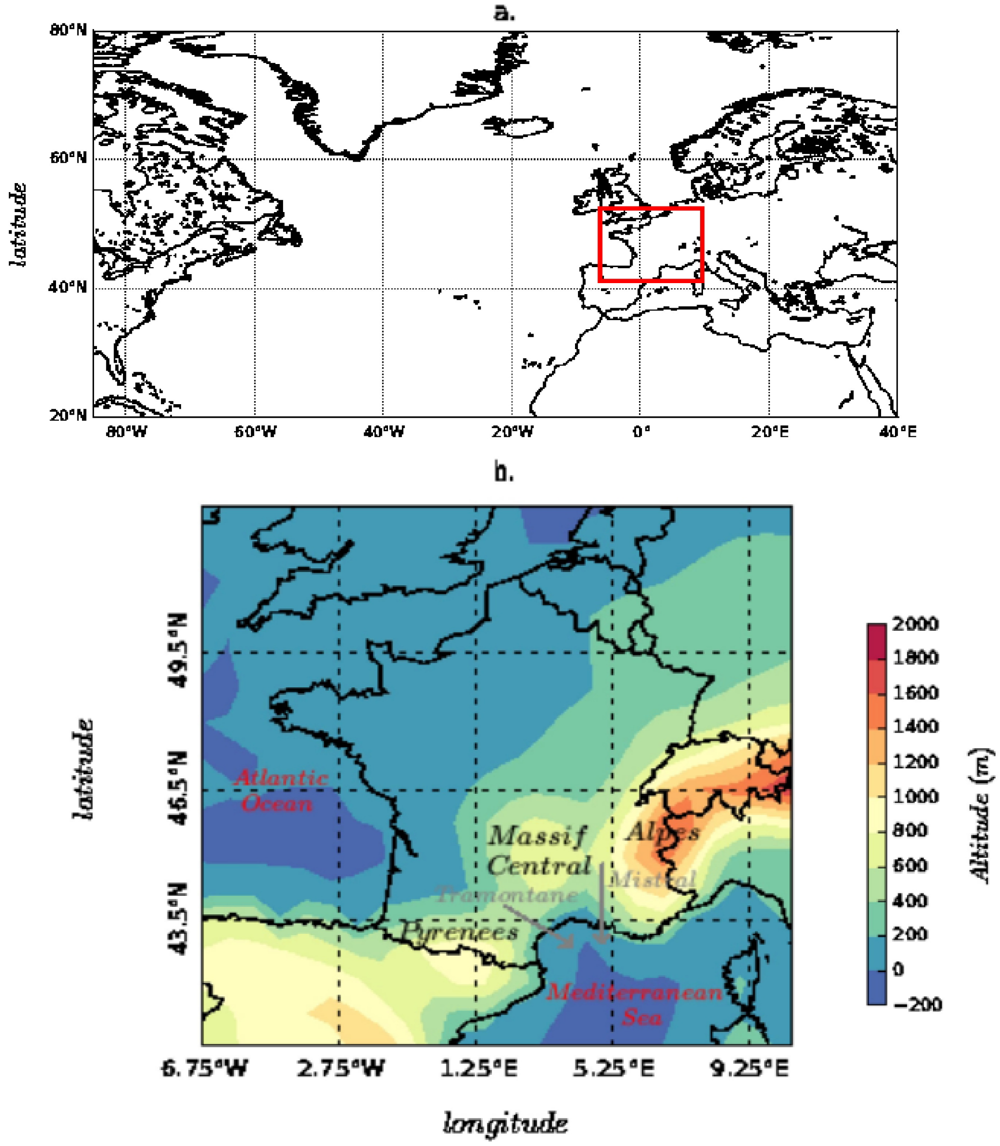

Figure 1: a. Domain used to perform the PCA on the Z500. The red box corresponds to the domain covering France on which the model is built and assessed. b. Domain covering France and part of its neighbouring countries. The colors represents the altitude above the sea level.

large-scale circulation is well described by the 500-hPa geopotential height (Z500) over the North Atlantic/European region (Michelangeli et al. (1995)).

We therefore retreive from the ECMWF reanalysis the daily time series of 500-hPa geopotential height (Z500) over a large domain that spans over North Atlantic Ocean and Europe (20°N to 80°N and 90°W to 40°E), (Figure 1, a), with grid size of 0.75°. The daily surface wind speed used to build and evaluate the probabilistic model is also retrieved from ECMWF reanalysis. The data spans the same period, but over a smaller domain which covers France and parts of neighbouring countries (40.5°N to 52.5°N and -6.75°W to 10.5°E), (Figure 1, b).

The 37 years of data are split into three periods. The first 20-year period (1 January 1979 to 31 December 1998), is used to build and estimate the model. This period is referred to as the fitting period. On the subsequent 13-year period (1 January 1999 to 31 December 2011), the probabilistic model is evaluated and compared to the past seasonal climatology of the wind speed, considered as the benchmark for wind speed forecasting at such long-term horizon. The seasonal climatology is defined as the empirical distribution of the daily average wind speed computed over all days in a given season of the fitting period. This period is referred to as the validation period. The results of the validation of our model are described in section 3.

On the 4 remaining years, we use the model to build probabilistic forecasts of the surface wind speed at the seasonal horizon. 48 ECMWF seasonal ensemble forecasts of the Z500 field over the large domain are retrieved. Sets of forecasts are retrieved from 2012 to 2015, beginning on every first day of each month. A major change of the assimilation system and forecast model limits the use of seasonal forecasts before November 2011. Seasonal forecasts provide a prediction of the Z500 at more than three months horizon, allowing to predict the surface wind speed at either monthly or seasonal horizon. The seasonal ensemble forecasts consist of 41 members. Each member has a slightly different initial state, so that the uncertainty on the atmospheric circulation grows with the forecast horizon giving a range of different possible states of the atmosphere. The forecasting performance of our model is analyzed in section 4.

2.2. Statistical methods

In the first step of building our model we apply the Principal Component Analysis (PCA) to the Z500 variable to reduce its dimension. The outputs of the PCA are the Empirical Orthogonal Functions (EOF) describing the prevalent spatial patterns in the data, and the associated Principal Components (PC) time series which show how the state of the atmosphere projects onto these patterns. The first EOFs may be identified with the classical large-scale weather patterns (NAO, SCA, ...) which control the European climate variability (Casanueva et al. (2014); Folland et al. (2008); Wallace and Gutzler (1980)). We expect the PCs to be well predicted in the seasonal ensemble forecasts.

In the second step, we build a model giving the probability distribution of the daily mean wind speed knowing the first n PCs. In other words, we want to compute the conditional density $p(y|X_1, ..., X_n)$ of the daily mean surface wind speed Y given the PCs of Z500 X_1 to X_n. Computing this conditional density directly is difficult due to the high dimension of the vector $(X_1, \ldots, X_n)$. To overcome this issue, we use the single index approximation (Delacroix et al. (2003)): we assume that the information about the PCs $(X_1, \ldots, X_n)$ may be summarized by a single scalar index

$$I = \beta_0 + \sum_{i=1}^{N} \beta_i X_i + \sum_{i=1}^{N} \beta_{ii} X_i^2 + \sum_{i=1}^{N-1} \sum_{j>i}^{N} \beta_{ij} X_i X_j, \tag{1}$$

where the coefficients β_0, β_i and β_{ij} are computed by least-squares regression of the surface wind speed Y on the principal components $X_1, \ldots, X_n$ for each location. A test of optimization of the index parameters β_i by minimization of the continuous ranked probability score (CRPS – see below) has been performed at several locations, but did not produce a significant improvement (only of the order of 0.1% of the initial CRPS).

The conditional probability density function $p(y|I)$ is given by the standard formula

$$p(y|I) = \frac{p(y, I)}{p(I)}, \tag{2}$$

where $p(y, I)$ is the joint density of the surface wind speed Y and the index I and $p(I)$ is the marginal density of the index. A gaussian kernel density estimator (KDE) is used to estimate the joint density and the marginal density over the period of length T :

$$\widehat{p}(y|I = i) = \frac{\sum_{t=1}^{T} K_{h_1}(y - Y_t) K_{h_2}(i - I_t)}{\sum_{t=1}^{T} K_{h_2}(i - I_t)}, \tag{3}$$

where K_h is the gaussian kernel function writen as :

$$K_h(x) = \frac{1}{h\sqrt{2\pi}} \exp\Big(- \frac{x^2}{2h^2}\Big). \tag{4}$$

While the estimated density is not very sensitive to the choice of the kernel function, the bandwidth parameters h_1 and h_2 have a significant impact on the resulting probability density function. In our study, the bandwidth parameters have been computed by cross-validation.

3. Evaluation and optimization of the model

3.1. Criteria for model evaluation: calibration and sharpness

The performance of a probabilistic forecasting model is typically assessed in terms of calibration and sharpness (Carney and Cunningham (2006); Foster and Vohra (1998); Gneiting et al. (2007); Thorarinsdottir (2013)). While calibration refers to the statistical consistency between the model and the actual values of the variable to predict, sharpness is a property of the model only and measures the width of the confidence intervals. Different modes of calibration exist and must be considered for the model to be fully calibrated. In the following, we evaluate probabilistic calibration and marginal calibration. Consider a probabilistic forecast at time t in the form of a predictive distribution function $F_t(x)$, and corresponding to the realization x_t. Probabilistic calibration (Gneiting et al. (2007)) measures the compatibility of the probabilistic forecast $F_t(x)$ with the actual realization x_t by means of the probability integral transform (PIT) defined by $p_t = F_t(x_t)$. The forecast is said to be probabilistically calibrated if the PIT follows a uniform distribution.

On the other hand, the marginal calibration (Gneiting et al. (2007)) compares the long-run distribution of the probabilistic forecast $\overline{F}(x) := \frac{1}{T}\sum_{t=1}^{T} F_t(x)$ to the long-run (climatological) distribution of the data, provided that the data is stationary. In meteorological terms, the assumption of stationarity of the data corresponds to the common assumption of the existence of a stable climate.

To evaluate the model performance we consider as the benchmark the seasonal climatology, which is often used within the wind energy industry for such long-term wind energy prediction (Pinson and Kariniotakis (2009)). Indeed, the persistence and autocorrelation of the wind disappear after 5 days at most so that the seasonal pattern is the only information that remains in absence of additional data.

Probabilistic calibration.

Probabilistic calibration is assessed using the Probability Integral Transform (Gneiting et al. (2007)). By applying, at each time step, the predicted

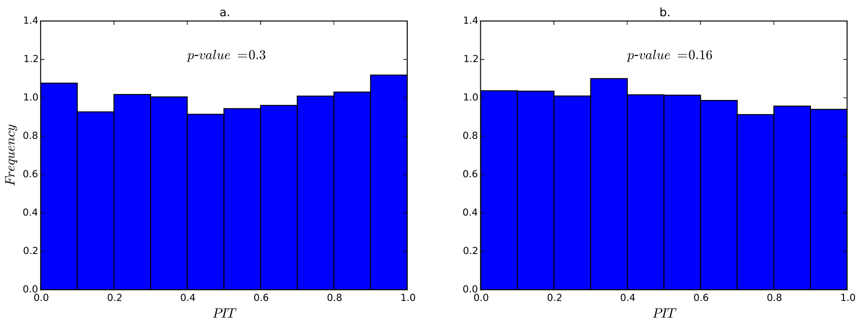

Figure 2: Example of a PIT histogram for one point in France (49.5°N/2.25°E), for the model (a.) and the climatology (b.). The p-value of the KS test performed on the 3 days sampled PIT is indicated. In this particular example the null hypothesis of uniformly distributed PIT is not rejected at the 5% confidence interval for neither the model nor the climatology.

Cumulative Distribution Function (CDF) $F_t(\cdot)$, to the actual value y_t, we obtain a sample $(F(y_t))_{t=1}^{T}$ of values in $[0,1]$, which must follow a uniform distribution on $[0,1]$ if the forecast is probabilistically calibrated. Uniformity of the PIT can be evaluated visually by ploting its histogram, usually referred to rank histogram in meteorology, (Figure 2), or more rigorously by performing a Kolmogorov-Smirnov (KS) test on the sample. The KS test is to be performed on independent and identically distributed random variables. Hamill (Hamill (2000)) shows that the correlated errors of samples can lead to misinterpretation of the PIT while testing uniformity. The samples thus have to be spaced far enough in space and time to be reasonably close to being independent. As the daily mean wind speed is autocorrelated up to time scales of about 3 to 5 days, so is the PIT. Figure 3 shows the autocorrelation of the entire sample of the PIT (Fig 3 a) and of the PIT resampled every 3 days (Fig 3 b). After 3 days, the sampled PIT shows little or no autocorrelation and the KS test is thus performed on a 3 day sampled PIT.

Marginal calibration.

Marginal calibration can be seen as a way to ensure that the actual climatology of the wind speed over the validation period is well represented by

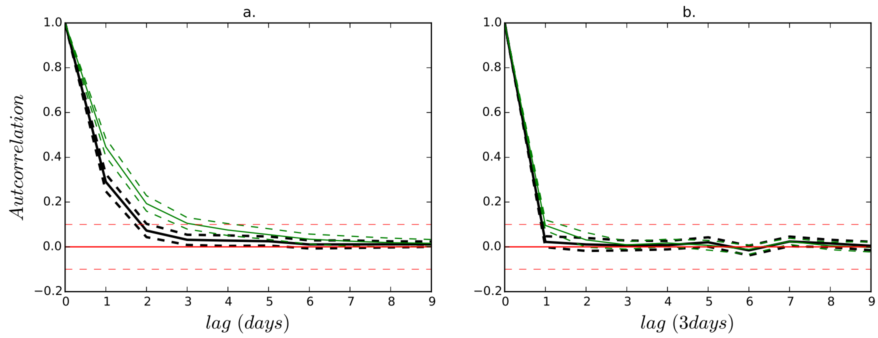

Figure 3: Autocorrelation of the PIT of the model (bold black curves) and the climatology (green curves), as function of the time lag : for the entire PIT sample (a) and for the PIT sampled every 3 days (b). The straight line represents the spatially averaged autocorrelation, and the dashed lines represent the 20^{th} and the 80^{th} percentiles.

the model. Actual climatology refers here to the probability density function of the wind speed over the validation period (and should not be confused with the past seasonal climatology computed on the fitting period, taken as a predictive distribution of reference). Marginal calibration can be assessed visually by plotting the difference between the climatological CDF on the validation period and the mean predicted CDF given by the model.

Figure 4 shows the marginal calibration computed using the probabilistic model (black dashed line) and the past seasonal climatology (black solid line) on the validation period at one grid point in the center of France (49.5°N/2.25°E).

To highlight the fact that part of the deviation comes from the statistical variations of the samples, we generate fifty random samples from the distribution of the actual wind speed over the 17 years of validation period (actual climatology) estimated by KDE. From these random samples we compute fifty different resampled actual climatologies, and calculate the difference between the distribution obtained for each one of them and that of the actual climatology. The red solid line represents the mean difference between the actual climatology and resampled actual climatologies, and red dotted lines represent the 20^{th} and 80^{th} percentiles. We see that for this particular point the curve corresponding to the past seasonal climatology is outside this bootstrap-style confidence interval, while the curve corresponding to our probabilistic foreacast is well inside it.

In order to visualize marginal calibration on a map, we compute at each

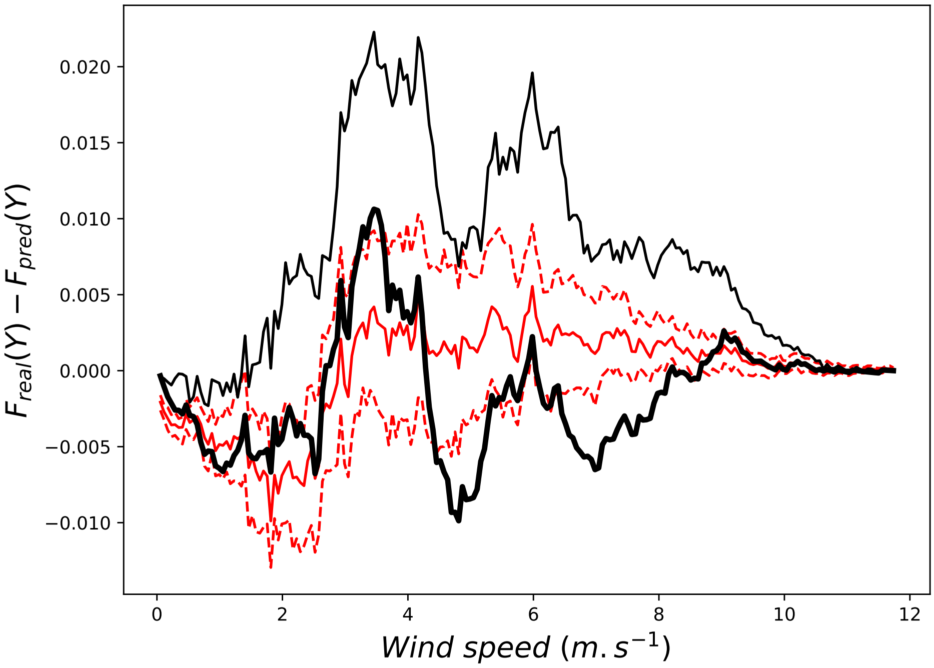

Figure 4: Difference between the actual CDF (actual climatology) on the validation period and : the mean predicted CDF (bold black solid line), the mean past seasonal climatology (black solid line), the mean of the resampled actual climatologies (red solid line), the 20^{th} and 80^{th} percentiles of the resampled actual climatologies (red dotted lines) for one point in France (49.5°N/2.25°E).

grid point the Mean Absolute Errors (MAE) between those distributions. MAE is calculated following the equation.

$$MAE = \int_{-\infty}^{\infty} |F_{real}(Y) - F_{pred}(Y)| dy \tag{5}$$

The model is considered marginally calibrated if the computed MAE defined above is less than the 95^{th} percentile of the MAE computed for the so called resampled actual climatologies.

Sharpness.

Sharpness refers to the width of the predictive distribution, that is to say, the accuracy of the forecast. Confidence interval widths are therefore good diagnostics of the sharpness of a probabilistic forecasting model. In this paper, the 90% confidence interval width is used as measure of sharpness.

Continuous Ranked Probability Score (CRPS).

The Continuous Ranked Probability Score (CRPS) is a widely used scoring rule in meteorological probabilistic forecasts (Candille et al. (2007); Candille and Talagrand (2005)). It aims to evaluate both calibration and sharpness simultaneously. The CRPS for a single predictive distribution F and realization y_t is defined by:

$$CRPS(F, y_t) = \int_{-\infty}^{\infty} (F(y) - 1_{(y \leq y_t)})^2 dy \tag{6}$$

with $1_{(y \leq y_t)}$ being defined as :

$$1_{(y \leq y_t)} = \begin{cases} 1, & \text{if } y \geq y_t. \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

For the entire sample of size T we define the CPRS by

$$CPRS = \frac{1}{T} \sum_{t=1}^{T} CPRS(F_t, Y_t).$$

3.2. Optimization of the model

In this section we discuss the choice of the number of principal components to be used in the model. By adding more PCs, the variability of the large scale circulation is better accounted for, but too many PCs can also lead to overfitting and thus poor calibration of the model. Depending on the region, the optimal number of PCs can be estimated. For example, although the onshore wind variability can be partially explained by the large-scale atmosphere circulation, smaller scale phenomena such as topography effects, can have a significant influence on the wind speed. Conversely, offshore wind speed is more regular and obviously not impacted by orography so that large-scale atmosphere circulation is the main driver of its variability at those long timescales.

To determine the optimal number of PCs we increase their number from 5 to 30 with an increment of 5 (which corresponds to 6 different models) and evaluate the probabilistic and marginal calibration and sharpness on the validation period of 17 years for the model based on the index computed with the corresponding number of PC.

Unexpectedly, marginal calibration shows no significant variation depending on the number of PC (no more than 10% of the statistical error), probably because of the CDF averaging effect. Conversely, probabilistic calibration and sharpness show high sensitivity to the number of PC (Figure 5). Unfortunately, on average, adding PCs sharpens the model, but also decalibrates it (Figure 5 a and c).

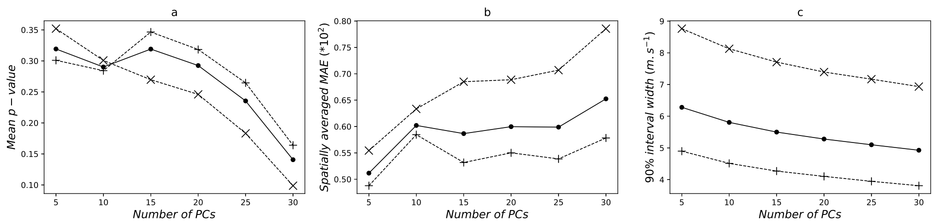

Figure 5: **a.** Spatially averaged p-value of the KS test performed on 3 days sampled PIT used to assess probabilistic calibration ; **b.** Spatially averaged MAE between actual climatology and the predicted climatology given by the model used to assess marginal calibration ; **c.** Spatially averaged 90% confidence interval width used to assess sharpness. All three graphs are plotted as function of the number of PCs used to fit the index. The black line with point markers is the average over the entire domain, the black doted line with 'x' markers is the average over the offshore part of the domain, and the black doted line with '+' markers over the onshore part of the domain.

In our final model, we use the following methodology to choose the optimal number of PCs for each location. We first test the null hypothesis that the PIT follows a uniform distribution with a 95% confidence level using the 3 days sampled PIT for each model. If the hypothesis is not rejected for any of the 6 models corresponding to different numbers of PCs, we keep the model which maximizes the sharpness. If the null hypothesis is rejected for all 6 models, we keep the model that maximizes the p-value of the KS test, with the risk to have a non-calibrated model.

Figure 6 shows the result of the choice described above. Over the northern half of the domain and along the western coast of France, a large number of PC (> 15) is required to build the index meaning that the variability can be explained by shorter scale phenomena without compromising the calibration quality of the model. This results in a sharper model than when using less PCs. Conversely, for offshore wind taking a large number of PCs reduces the calibration quality.

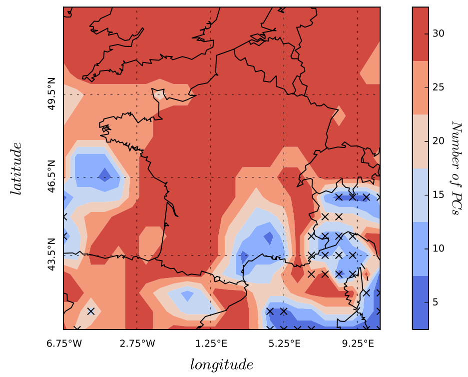

Figure 6: Optimal number of PCs used to fit the index of the model determined using the optimization process described in the text. 'x' markers show points where the model is not calibrated (Figure 7).

In the southeast of France, over the Mediterranean coast and the sea, we can find a clear signature of the orography. Offshore, the Mistral, which refers to the strong wind blowing over the Mediterranean sea after being channeled in the valley formed by the Alps and the Massif Central (Drobinski et al. (2017)), can be identified by an intermediate number of PC (20-30). The Tramontane also refers to an orographic wind blowing over the same region but channeled in the valley formed by the Pyrenees and the Massif Central (Brossier and Drobinski (2009)) (Fig 1 b). South of the Alps, the model is not calibrated, and south of the Massif Central only 5 PCs are used resulting in model that is less sharp.

3.3. Evaluation of the optimized model

Figure 7 shows the results of the KS test performed on the 3 days sampled PIT given by the optimized model (Fig 7 a and b) and the climatology (Fig 7 c and d). The p-value for the climatology ranges between 0 and 0.8, while it ranges between 0 and 0.5 for the model. The null hypothesis of adequate calibration is not rejected in the North part of the domain for the model, while for the climatology this hypothesis is rejected over the North part of the domain. The climatology does not represent the law of the wind well in those regions but the probabilistic model represents it quite well (Fig 7 b and d). This can be surprising as the climatology is built using 20 years of data which may seem to be sufficient to ensure calibration over a period of the same length. Nevertheless, it has been shown that annual wind trends can be significant over 1 to 2 decades in this region (Jourdier (2015)). Using only the past five years of wind speed data to build the empirical seasonal CDF allows to follow those trends. This sliding CDF displays a null hypothsis of adequate calibration which is not rejected over the entire domain. Nevertheless, it performs as well as the seasonal climatology in terms of sharpness *(Not shown)*. In the South of the domain, the model and the climatology perform similarly in terms of probabilistic calibration showing large non-calibrated areas. Indeed, the region is very complex and strongly influenced by orography. This complexity seems to be hard to recover with the information on the season only (climatology) or the information on the large-scale circulation (model).

Figure 8 shows the MAE between the real climatological CDF over the 13-year validation period and, on the one hand, the averaged CDF predicted by the model (Fig 8 a.) and on the other hand the climatological CDF based on the 20 year fitting period (Fig 8 b). We can clearly see a strong correlation between marginal calibration and probabilistic calibration. The model is considered marginally calibrated if the computed MAE is inferior to the 95^{th} percentile of the MAE computed for the resampled actual climatologies. Applying this criterion to MAE computed for the model and for the past seasonal climatology gives a map (not shown) which is very similar to those in Figures 7c and 7. If the model or the climatology is probabilistically calibrated in a given location, it is also marginally calibrated there. Overall, for both calibration criteria, the calibration of the model is at least as good as that of the climatology and often much better.

Figure 9 displays the 90% confidence interval width averaged over the validation period for the model (IC_{90mod}) (Fig 9 a) and the climatology (IC_{90clim}) (Fig 9 b), and the ratio of IC_{90clim} to IC_{90mod} (Fig 9 c). The first striking observation is that the 90% confidence interval is much larger offshore than onshore for both the model and the climatology. This highlights the fact that even if the wind may be more regular, it can also be much stronger because of the low roughness, so that the difference between weak and strong wind events is by far larger than onshore. The signature of the Mistral and Tramontane is clear (Fig 9 a, b), with even larger interval width that may come from the bimodal distribution of the wind speed in this region (Drobinski et al. (2015)).

Over the entire domain, on average, the model is sharper than the climatology. The model does not perform more than 50% better than the clima-

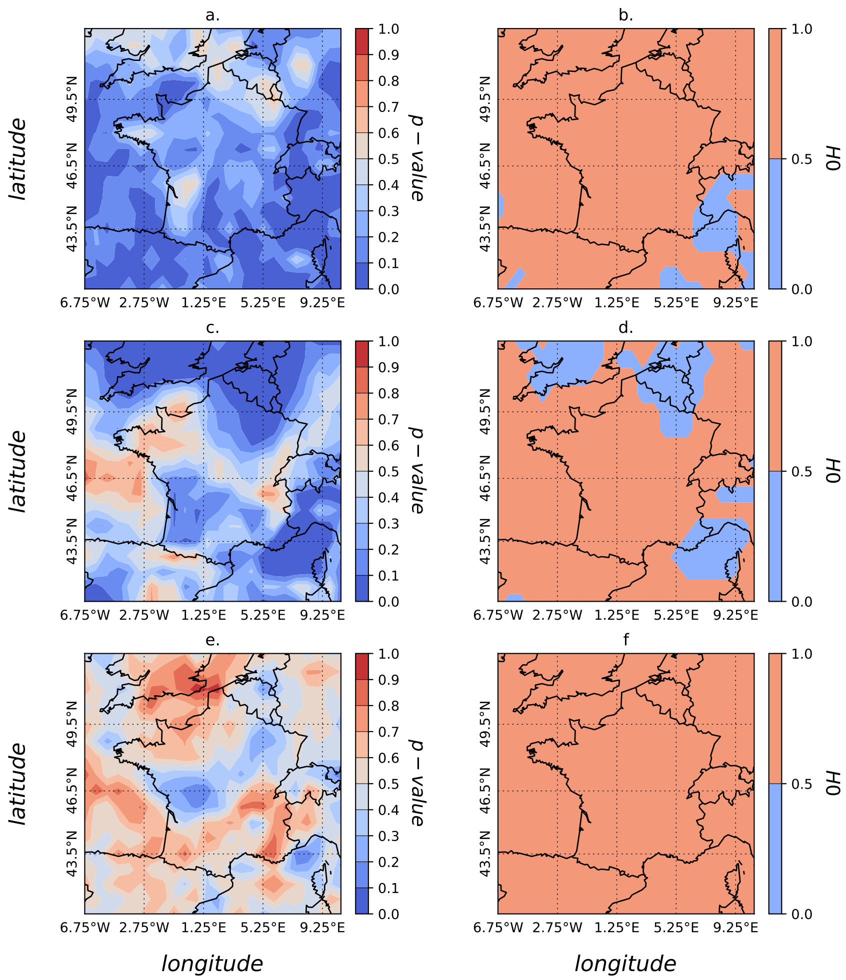

Figure 7: Left graphs: p-value of the KS test performed on the 3 days sampled PIT of the model (a.) and the climatology (b.) and for the empirical seasonal CDF based on the last five years of wind speed (e). Right graphs: the blue area (0 value) shows the regions where the null hypothesis of adequate calibration is rejected for the model (b) and the climatology (d) and for the empirical seasonal CDF based on the last five years of wind speed (f).

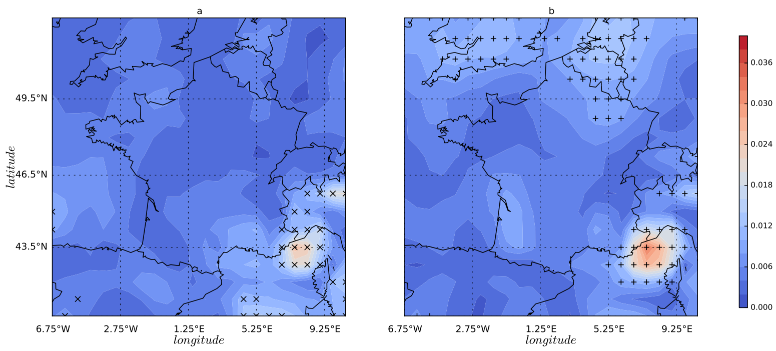

Figure 8: Mean Absolute Error between the real climatological CDF over the 15 years of validation and the averaged CDFs predicted by the model (a.) and the climatological predictive CDFs based on the 20 years of calibration period (b.). 'x' marker on panel a. and '+' markers on panel b. show respectively points where the model and the climatology are not probabilistically calibrated.

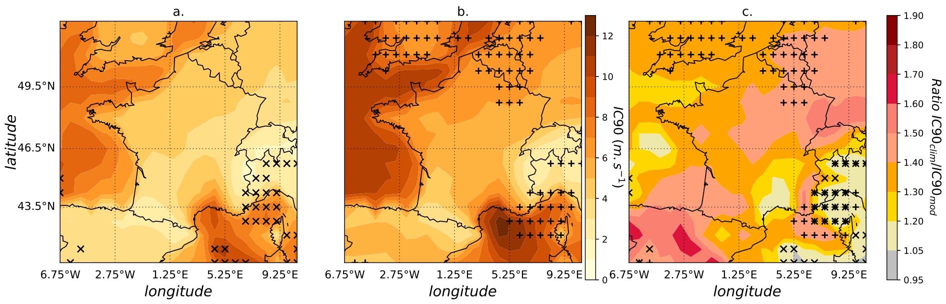

Figure 9: 90% confidence interval width averaged over the validation period, for the model (a.) and the climatology (b.). Panel c. displays the ratio of the confidence interval width of the climatology over the model. 'x' and '+' markers indicate places where respectively the model or the climatology are not calibrated.

tology, except in the northeast of the domain. Over the north of France, the model is sharper than the climatology by more than 40% which is encouraging because of the high wind energy potential in those regions. Unfortunately, over the west Atlantic ocean, the model is not as sharp as expected compared to climatology, but still, it performs 20% to 30% better. Over the south of France, in addition to the bad calibration of the model, its performance in terms of sharpness is not as spectacular as in other regions. Again, this can be due to the complexity of the wind variability in this region.

-	-	All seasons	Winter	Spring	Summer	Fall
Model	Mean IC90	5.2	5.5	5.1	4.8	5.4
	σ	2.1	2.3	2.0	1.9	2.2
Climatology	Mean IC90	6.9	7.9	6.4	5.8	7.7
	σ	2.4	2.7	2.1	2.1	2.6
Ratio	IC_{90clim}/IC_{90mod}	1.3	1.4	1.2	1.2	1.4

Table 1: 90% confidence interval width (IC90) ($m.s^{-1}$) averaged on the validation period and on the whole domain, for all seasons, and every season separately, for the model and the climatology.

By averaging the interval width separately for each season, we can highlight a strong seasonal variability of the interval given by the climatology, which is not so noticeable for the model (Table 1). Thus, the model shows even better performance compared to the climatology in winter and fall (40% sharper than climatology on average over the domain) which are the seasons when the risk of high LOLE may be larger because of low temperature. The model is 80% sharper than the climatology in the northeast regions in winter and fall *(not shown)*. Differences between land and sea are present for all seasons, and the Mediterranean region is always more problematic.

The CRPS of the model and the climatology should inform us on both the calibration and sharpness. It is expressed in the same units as the predicted quantity ($m.s^{-1}$ in the case of wind speeds) and reduces to the MAE for point forecasts. Figure 10 shows the mean CRPS on the validation period, for the model (Fig 10a), the climatology (Fig 10b) and the ratio of $CRPS_{clim}/CRPS_{mod}$ (Fig 10c). All panels are very comparable to those of Figure 9. Even if there is no doubt that the CRPS adresses both calibration and sharpness, on average, it appears to put too much weight on sharpness. For instance, over the Alps, the model is not calibrated, nevertheless the CRPS has very low values, indicating that the model has good performance. The CRPS values do display a very significant difference between land and sea, and so does the confidence interval width. It is thus clear by comparing figures 9 and 10 that the average CRPS is informative about the sharpness of the model more than about its calibration quality.

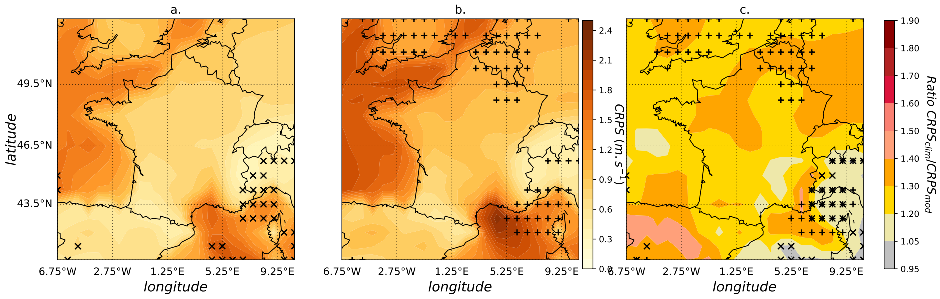

Figure 10: CRPS of the model (a.) and the seasonal climatology (b.). 'x' marker on panel a. and '+' markers on panel b. show respectively points where the model and the climatology are not calibrated. c. is ratio of the CRPS

4. Forecasting the wind at the monthly and seasonal horizon

4.1. Methodology

To make monthly / seasonal forecasts with our model, we must take into account the uncertainty of the Z500 forecast, and thus also of the index. The seasonal ensemble forecasts of ECMWF are based on 41 members displaying a large range of possible Z500 fields. For each member, we first calculate the values of the principal components by projecting the corresponding Z500 field onto the EOFs identified during the stage of model calibration. Next, for each member of the ensemble forecast, and for each location where surface wind speed forecast is needed, we compute the corresponding index value using equation (1), where the coefficients β_i were identified during the stage of model calibration. This gives us an ensemble of index values $I_1, \ldots, I_n$. From this ensemble we construct the predictive distribution of index values, denoted by μ. This can be done in two different ways. The first method (raw forecast) consists in taking simply the empirical distribution of $I_1, \ldots, I_n$, that is, $\mu = \frac{1}{n}\sum_{k=1}^{n} \delta_{I_k}$, where δ_x is the point mass at point x. The second method uses statistical post-processing of the ensemble forecast to construct a distribution μ with better calibration / sharpness properties than the raw forecast. In this paper, we use the Ensemble Model Output Statistics (EMOS) method, described below, for forecast post-processing.

Once the predictive distribution for the index has been constructed, the density of the predictive distribution for the surface wind speed given the forecast $p(y|F)$ is obtained by integrating the density of the conditional distribution of the wind speed given the index with respect to the predictive distribution of the index:

$$p(y|F) = \int_{-\infty}^{\infty} p(y|I = x)\mu(dx), \tag{8}$$

This should produce a less sharp model with a higher chance to be calibrated than if only the mean of the forecast ensemble is used.

Ensemble Model Output Statistics - EMOS.

To recalibrate and sharpen a forecast ensemble different statistical post-processing methods exist such as the Bayesian Model Averaging (BMA) (Möller et al. (2013); Raftery et al. (2005); Sloughter et al. (2013)) or the Ensemble Model Output Statistics (EMOS) (Gneiting et al. (2005); N.Schuhen et al. (2012); Thorarinsdottir and Gneiting (2010)). EMOS aims at recalibrating the distribution of ensemble forecasts, but also at sharpening it. This method is inspired by Gneiting et al. (2005) apart from the optimization algorithm. This method is based on the assumption that μ has a normal distribution $N(m_I, \sigma_I)$, where m_I is a weighted linear combination of the index values of the ensemble,

$$m_I = b_0 + \sum_{m=1}^{n} b_m I_m, \tag{9}$$

and σ_I is parameterized by

$$\sigma_I = c + d\,\mathrm{Var}(I), \tag{10}$$

where $\mathrm{Var}(I)$ is the empirical variance of the ensemble.

The parameters of the EMOS method $b_0, \ldots, b_n$, c and d are estimated as follows. In the first step of the estimation procedure, on the training period, set to three years in this study, we perform a linear regression of the index I computed from the actual ERAI-reanalysis on the index values $I_1, \ldots, I_n$ computed from the ECMWF seasonal forecasts. This gives us a first estimate of $b_0, \ldots, b_n$. In this first step we set $c = 0$ and $d = 1$. Then, in the second step, we improve the first-step estimates by minimizing the Continuous Ranked Probability Score (CRPS) of the forecasts, averaged over the training period, seen as function of the parameters $b_0, \ldots, b_n$, c and d using the Powell algorithm (Powell (1964)).

In the end, we obtain a set of parameters b_1, ..., b_m, c and d that minimize the CRPS score. The minimization of the CRPS must optimizes the calibration and the sharpness. We apply the obtained parameters on the remaining year of forecasts to estimate the gaussian distribution $N(m_I, \sigma_I)$ of the index and then integrate over this distribution as in eq (8). The procedure is repeated 4 times by training on three different years and testing on the remaining year. This results in 48 EMOS forecasts of the daily mean wind speed distribution at the seasonal horizon.

4.2. Results

Figure 11 displays the p-value of the KS test performed on the 3 days sampled PIT of the 4 years of forecasts[2], for the climatology (Fig 11 a), raw forecasts (Fig 11 b), and EMOS forecasts (Fig 11 c). The colorbar is designed to test the null hypothesis at 95% confidence. Regarding this test, forecasts are not calibrated in the northeast part of France which disagrees with the test performed on the validation period. As this behaviour is quite comparable to the climatology, this suggests that the fitting period of 20 years used to build the model and climatology may not be representative enough of the wind in the forecasting years in those regions.

Figure 12 shows the ratio IC_{90clim}/IC_{90mod} averaged at 15 days (Fig 12 a.), monthly (Fig 12 b.) and seasonal horizon (Fig 12 c.) for raw forecasts (black dots) and EMOS forecasts (green dots), for all forecast years.

The accuracy decreases with the forecast horizon. It appears that the forecast is quite sharp within 15 days and deteriorates significantly for larger horizons. The largest deterioration occurs between November and February. In spring and summer, forecasting performance does not seem to be highly sensitive to the forecast horizon.

Moreover, EMOS forecasts only significantly improves the accuracy with respect to raw forecasts, in winter and fall, especially at monthly and seasonal horizon. Raw and EMOS forecasts at 15 days horizon are almost always

[2]Note that we only have 48 independent forecasts which is not enough to test calibration. To get around this difficulty, we use forecasts obtained at the same date for different horizons (3 days, 6 days, 9 days, ..., 30 days) as if they were independent. Autocorrelation analysis shows that they are indeed uncorrelated. Each forecast thus corresponds to 10 data points.

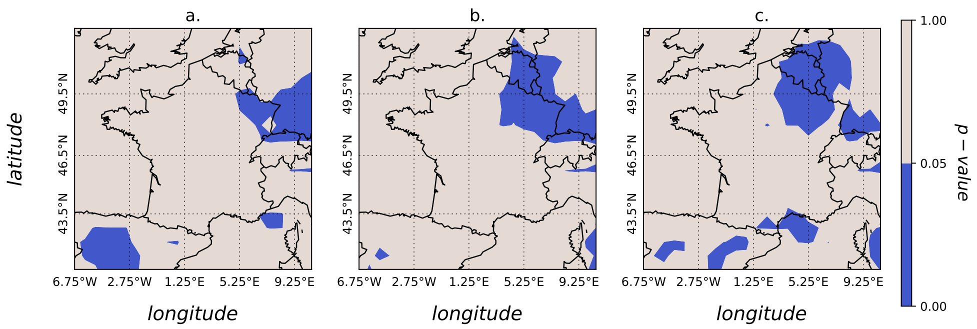

Figure 11: p-value of the KS test performed on the 3 days sampled PIT of the 4 years forecasts, for the climatology (a.), raw forecasts (b.), and EMOS forecasts (c.). Blue areas correspond to regions where the null hypothesis of calibration is rejected at the 5% confidence level.

better than the climatology. The improvement with EMOS optimization does not seem to be very large at this horizon, most probably because the distribution of the index is already very sharp. EMOS forecasts still slightly improve raw forecasts by about 5% in the beggining and at the end of the year.

The seasonal variability described in Table 1 is recovered at the 15 days and monthly horizon, which is encouraging. At the seasonal horizon, raw forecasts performance does not display a strong seasonal variability and the ratio is close to one, so the model does not perform better than the climatology. EMOS forecasts performance displays an even lower intra-annual variability but the ratio is systematically around 1.10. This is a very interesting result as it shows that there is a valuable statistical information on the local surface wind speeds in the seasonal forecasts of the large-scale circulation post-processed using the EMOS method, which leads to a 10% improvement over the climatology on average even at this long timescale.

Figures 13, 14 and 15 show the ratio IC_{90clim}/IC_{90mod} for each year of forecasts, respectively at 15 days, monthly and seasonal horizons, for raw forecasts (top) and EMOS forecasts (bottom). A sharp decrease of the ratio can be seen between figures 13 and 14. Comparatively, the acuracy decrease

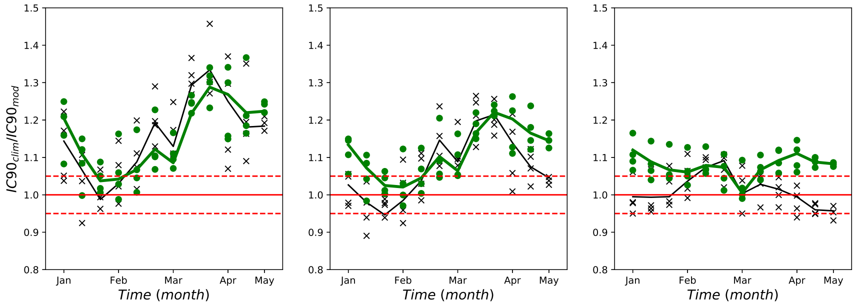

Figure 12: Ratio IC_{90clim}/IC_{90mod} at 15 days (a.), monthly (b.) and seasonal horizon (c.) for every raw forecasts (black cross) - the black slight line represents the mean ratio ; and EMOS forecasts (green dots) - the green bold line represents the mean ratio. Four forecasted years are 2012, 2013, 2014, 2015.

is less pronounced, between figure 14 and 15. In the northeast of the domain, the ratio is the highest so that the model seems to be very sharp, especially at the 15 days horizon for the year 2015. This could be the cause of the decalibration of the model (Fig 11).

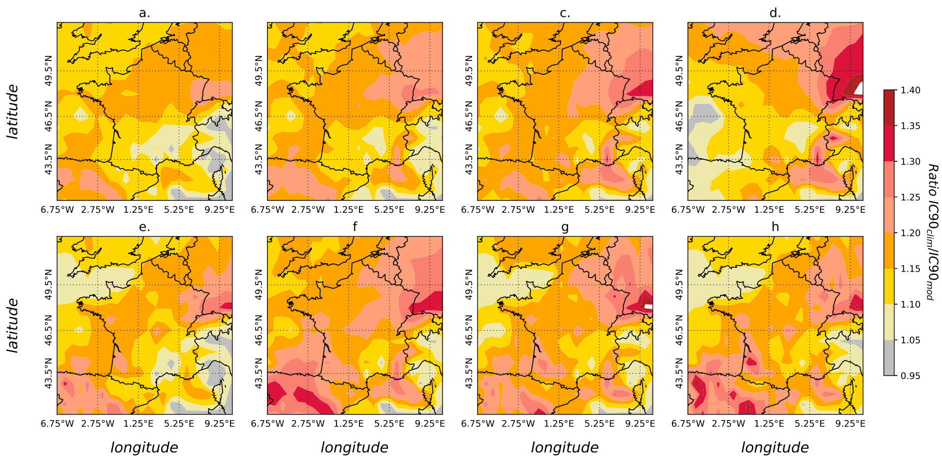

Figure 13: Ratio of IC_{90clim} over IC_{90mod} averaged over 12 seasonal forecasts, for forecasted years 2012 (a., e.), 2013 (b., f.), 2014 (c., g.), and 2015 (d., h.) at 15 days horizon for raw forecasts (top) and EMOS forecasts (bottom)

Those figures also show the efficiency of the EMOS method to reduce

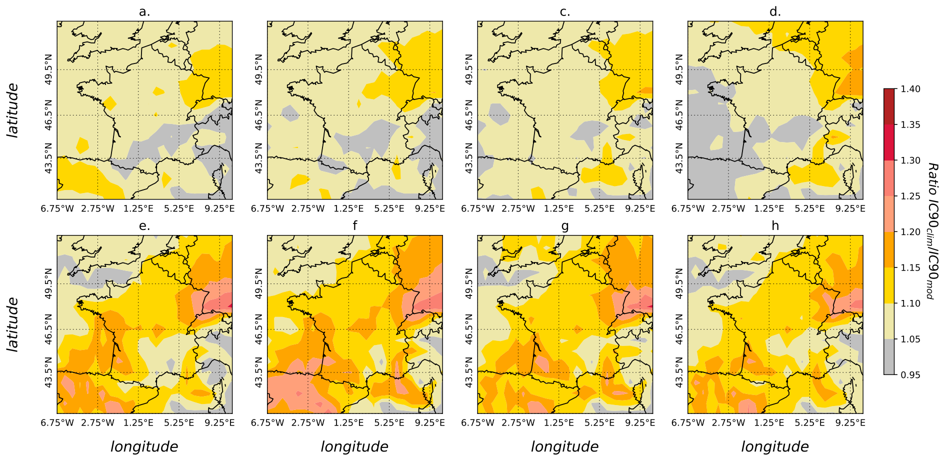

Figure 14: Same as figure 13 for monthly horizon forecasts.

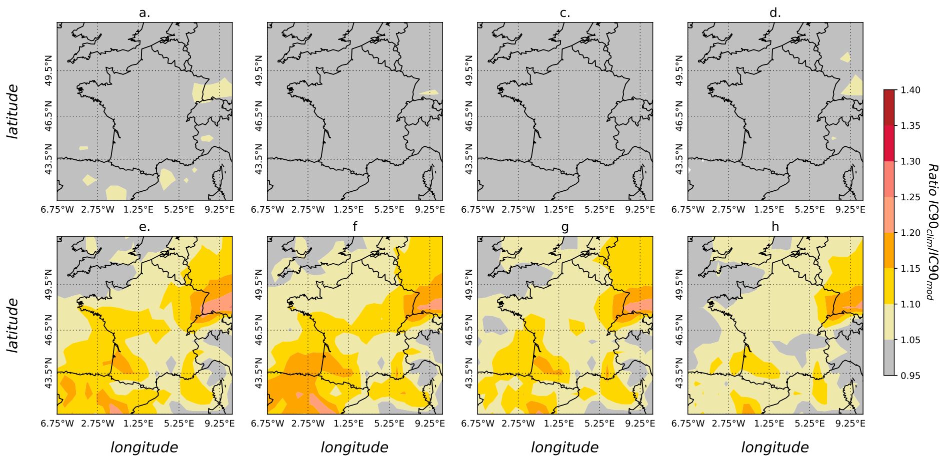

Figure 15: Same as figure 13 for seasonal horizon forecasts.

the uncertainty on the index and thus to highly sharpen the model so that forecasts at the seasonal horizon give more information on the wind than the climatology which is at this moment widely used for such long-term wind energy evaluation. The EMOS forecasts display a consistent spatial pattern in terms of accuracy for any forecast horizon which is not the case for raw forecasts. Indeed, the performance of raw forecasts displays a noticeable inter-annual variability. For instance, at the monthly horizon, forecasted year 2015 is comparable to the climatology in the east of the domain while year 2012, 2013, and 2014 are sharper than the climatology in this part of France. It again highlights the added value of EMOS forecasts compared to raw forecasting method, even if a larger sample of forecasted years should be analysed to confirm this behaviour. The Mediterranean region is the region where the model performs the worst compared to the climatology. This result confirms what was found on the validation period. For EMOS forecasts, the spatial pattern of the ratio is very comparable to the Fig 9 for any forecast horizon and for all years. This is not as clear for raw forecasts. It means that the uncertainty on the ensemble forecast is highly reduced by EMOS method, but moreover that this method reduces the inter-annual variability of the uncertainty of the ensemble.

5. Conclusion

A probabilistic model is proposed to predict daily wind speed distribution from a few days to seasonal timescale. It is compared to the climatology which is often the reference used as the best seasonal forecast for energy management. The study shows that the model is better statistically calibrated than the climatology and is able to follow very long-term trends of the wind speed. On average over France, the model is shown to be 30% sharper than the climatology. It is shown to be more accurate than the climatology especially onshore, in the northwest regions and in winter and fall.

We apply the probabilistic model to the seasonal forecast ensemble of ECMWF. We test two methods to forecast wind speed with these ensembles. The first method uses the empirical density of the raw calculated index, and the second estimates the density of the calculated index by optimizing calibration and sharpness of the ensemble using the EMOS statsitical post-processing technique (Gneiting et al. (2005)). We show that the model is able to be more precise than the climatology at 15 days and monthly horizon using both methods and that at the seasonal horizon, the EMOS method is systematically more precise than climatology.

Acknowledgement.

This research was supported by the ANR project FOREWER (ANR-14-CE05- 0028). This work also contributes to TREND-X program on energy transition at Ecole Polytechnique as well as to the HyMeX program (HYdrological cycle in The Mediterranean EXperiment ((Drobinski et al., 2014))) through the working group Renewable Energy.

Albadi, M., El-Saadany, E., 2010. Overview of wind power intermittency impacts on power systems. Electric Power Systems Research 80, 627–632.

Alonzo, B., Ringkjob, H.K., Jourdier, B., Drobinski, P., Plougonven, R., Tankov, P., 2017. Modelling the variability of the wind energy resource on monthly and seasonal timescales. Renewable energy 113, 1434–1446.

Azad, H.B., Mekhilef, S., Ganapathy, V.G., 2014. Long-term wind speed forecasting and general pattern recognition using neural networks. IEEE Transaction on Sustainable Energy 5, 546–553.

Barbounis, T.G., Theocharis, J.B., Alexiadis, M.C., , Dokopoulos, P.S., 2006. Long-term Wind Speed and Power Forecasting Using Local Recurrent Neural Network Models. IEEE Transaction on Energy Conversion 21, 273–284.

Bilgili, M., Sahin, B., Yasar, A., 2007. Application of Artificial Neural Networks for the wind speed prediction of target station using reference stations data. Renewable Energy 32, 2350–2360.

Brossier, C.L., Drobinski, P., 2009. Numerical high-resolution air-sea coupling over the Gulf of Lions during two tramontane/mistral events. Journal of Geophysical Reseasrch 114, D10110.

Candille, G., Cote, C., Houtekamer, P., Pellerin, G., 2007. Verification of an ensemble prediction system against observations. Monthly Weather Review 135, 2688–2699.

Candille, G., Talagrand, O., 2005. Evaluation of probabilistic prediction systems for a scalar variable. Q. J. R. Meteorol. Soc. 131, 2131–2150.

Carney, M., Cunningham, P., 2006. Evaluating density forecasting models. Trinity College Dublin, Department of Computer Science .

Carpinone, A., Giorgio, M., Langella, R., Testa, A., 2015. Markov chain modeling for very-short-term wind power forecasting. Electric Power Systems Research 122, 152–158.

Casanueva, A., Rodrìguez-Puebla, C., Frìas, M.D., Gonzàlez-Reviriego, N., 2014. Variability of extreme precipitation over Europe and its relationships with teleconnection patterns. Hydrol. Earth Syst. Sci. 18, 709–725.

Cassou, C., 2008. Intraseasonal interaction between Madden-Julian Oscillation and the North Atlantic Oscillation. Nature 455, 523–597.

Chang, W., 2014. A literature review of wind forecasting methods. Journal of Power and Energy Engineering 2, 161–168.

Davies, J.R., Rowell, D.P., Folland, C.K., 1997. North atlantic and European seasonal predictability using an ensemble of multidecadal atmospheric GCM simulations. International Journal of Climatology 17, 1263–1284.

Delacroix, M., Hardle, W., Hristachea, M., 2003. Efficient estimation in conditional single-index regression. Journal of Multivariate Analysis 86, 213–226.

Drobinski, P., Alonzo, B., Basdevant, C., Cocquerez, P., Doerenbecher, A., Fourri, N., Nure, M., 2017. Lagrangian dynamics of the mistral during the Hymex SOP2. Journal of Geophysical Research: Atmospheres 122, 1387–1402.

Drobinski, P., Coulais, C., Jourdier, B., 2015. Surface wind-speed statistics modelling: Alternatives to the Weibull distribution and performance evaluation. Boundary-Layer Meteorol 157, 97–123.

Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Branger, K., Borga, M., Braud, I., Chanzy, A., Davolio, S., Delrieu, G., Estournel, C., Boubrahmi, N.F., Font, J., Grubisic, V., Gualdi, S., Homar, V., Ivancan-Picek, B., Kottmeier, C., Kotroni, V., Lagouvardos, K., Lionello, P., Llasat, M., Ludwig, W., Lutoff, C., Mariotti, A., Richard, E., Romero, R., Rotunno, R., Roussot, O., Ruin, I., Somot, S., Taupier-Letage, I., Tintore, J., Uijlenhoet, R., , H.Wernli, 2014. a 10-year multidisciplinary program on the Mediterranean water cycle. Meteorol. Soc. 95, 1063–1082.

EWEA, 2016. Wind in Power: 2015 European Statistics. Technical Report. European Wind Energy Association.

Folland, C.K., Knight, J., Linderholm, H.W., Fereday, D., Ineson, S., Hurrell, J.W., 2008. The Summer North Atlantic Oscillation: Past, Present, and Future. Journal of Climate 22, 1082–1103.

Foster, D.P., Vohra, R.V., 1998. Asymptotic calibration. Biometrika, , 85, 379–390.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Statist. Soc. B 69, 243–268.

Gneiting, T., Raftery, A.E., III, A.H.W., Goldman, T., 2005. Calibrated probabilistic forecasting using Ensemble Model Output Statistics and minimum CRPS estimation. Monthly Weather Review 133, 1098–1118.

Gomes, P., Castro, R., 2012. Wind speed and wind speed forecasting using statistical models: Autoregressive Moving Average (ARMA) and Artificial Neural Networks (ANN). International Journal of Sustainable Energy Development (IJSED) 1.

Guo, Z., Zhao, W., H.Lu, J.Wang, 2012. Multi step forecasting for wind speed using a modified EMD based Artificial Neural Network model. Renewable Energy 37, 241–249.

Hamill, T.M., 2000. Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review 129, 550–560.

J. Wang, S.Q., Zhou, Q., Jiang, H., 2015. Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang China. Renewable Energy 76, 91–101.

Jourdier, B., 2015. Wind resource in metropolitan France: assessment methods, variability and trends. Ph.D. thesis. Ecole Polytechnique.

Michelangeli, P.A., Vautard, R., Legras, B., 1995. Weather regimes: Recurrence and quasi stationarity. Journal of Atmospheric Sciences 52, 1237–1256.

Möller, A., Lenkoski, A., Thorarinsdottir, T.L., 2013. Multivariate probabilistic forecasting using Bayesian model averaging and copulas. Q. J. R. Meteorol. Soc. 139, 982–991.

More, A., Deo, M., 2003. Forecasting wind with neural networks. Marine Structures 16, 35–49.

Najac, J., Boe, J., Terray, L., 2009. A multi model ensemble approach for assessment of climate change impact on surface winds in France. Climate Dynamics 32, 615–634.

NationalGrid, 2016. Winter Outlook Report 2016/2017. Technical Report. National Grid.

N.Schuhen, Thorarinsdottir, T.L., Gneiting, T., 2012. Ensemble Model Output Statistics for wind vectors. Monthly Weather Review 140, 3204–3219.

Owen, J., Palmer, T., 1987. The impact of El Niño on an ensemble of extended range forecasts. Monthly Weather Review 115, 2103–2117.

Pinson, P., Kariniotakis, G., 2009. Conditional prediction intervals of wind power generation. IEEE Transaction On Power Systems 25, 1845–1856.

Powell, M.J.D., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. The Computer Journal 7, 155–162.

Pryor, S., Barthelmie, R., 2010. Climate change impacts on wind energy: A review. Renewable and Sustainable Energy Reviews 14, 430–437.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review 133, 1155–1174.

Rodwell, M.J., Rowell, D.P., Folland, C.K., 1999. Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. Nature 398, 320–323.

Sailor, D.J., M. Smith, M.H., 2008. Climate change implications for wind power resources in the northwest United States. Renewable Energy 33, 2393–2406.

Sloughter, J.M., Gneiting, T., Raftery, A.E., 2013. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. Journal of the American Statistical Association 141, 2107–2118.

Soman, S.S., Zareipour, H., O. Malik, P.M., 2010. A review of wind power and wind speed forecasting methods with different time horizons. North American Power Symposium (NAPS) , 1–8.

Stesfos, A., 2002. A novel approach for the forecasting of mean hourly wind speed time series. Renewable Energy 27, 163–174.

Taylor, J., McScharry, P., Buizza, R., 2009. Wind power density forecasting using ensemble prediction and time series model. IEEE, Transactions on Energy conversion 34.

Thorarinsdottir, T.L., 2013. Calibration diagnostics for point process models via the Probability Integral Transform. Stat 2, 150–158.

Thorarinsdottir, T.L., Gneiting, T., 2010. Probabilistic forecasts of wind speed: Ensemble Model Output Statistics using heteroskedastic censored regression. J. R. Statist. Soc. A 173, 371–388.

Wallace, J.M., Gutzler, D.S., 1980. Teleconnection in the geopotential height field during the northern hemisphere winter. Monthly Weather Review 109, 784–812.

Wytock, M., Kolter, J.Z., 2013. Large-scale Probabilistic Forecasting in Energy Systems using Sparse Gaussian Conditional Random Fields. Proceedings of the IEEE Conference on Decision and Control (CDC) , 1019–1024.