Renewable Energy 113 (2017) 1434-1446

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Review

Modelling the variability of the wind energy resource on monthly and seasonal timescales



Renewable Energy

用

Bastien Alonzo ^{a, b, *}, Hans-Kristian Ringkjob ^{a, b, d}, Benedicte Jourdier ^{a, c, e}, Philippe Drobinski ^a, Riwal Plougonven ^a, Peter Tankov ^{b, f}

^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 EDF R & D - MFFE, Applied Meteorology Group, Chatou, France

^d Geophysical Institute, University of Bergen, Bergen, Norway

^e French Environment and Energy Management Agency, Angers, France

f CREST - ENSAE ParisTech, Malakoff, France

A R T I C L E I N F O

Article history: Received 11 July 2016 Received in revised form 6 June 2017 Accepted 3 July 2017 Available online 5 July 2017

Keywords: Seasonal modelling Wind distribution Variability Large-scale circulation Forecasts Wind energy

ABSTRACT

An avenue for modelling part of the long-term variability of the wind energy resource from knowledge of the large-scale state of the atmosphere is investigated. The timescales considered are monthly to seasonal, and the focus is on France and its vicinity. On such timescales, one may obtain information on likely surface winds from the large-scale state of the atmosphere, determining for instance the most likely paths for storms impinging on Europe. In a first part, we reconstruct surface wind distributions on monthly and seasonal timescales from the knowledge of the large-scale state of the atmosphere, which is summarized using a principal components analysis. We then apply a multi-polynomial regression to model surface wind speed distributions in the parametric context of the Weibull distribution. Several methods are tested for the reconstruction of the parameters of the Weibull distribution, and some of them show good performance. This proves that there is a significant potential for information in the relation between the synoptic circulation and the surface wind speed. In the second part of the paper, the knowledge obtained on the relationship between the large-scale situation of the atmosphere and surface wind speeds is used in an attempt to forecast wind speeds distributions on a monthly horizon. The forecast results are promising but they also indicate that the Numerical Weather Prediction seasonal forecasts on which they are based, are not yet mature enough to provide reliable information for timescales exceeding one month.

© 2017 Elsevier Ltd. All rights reserved.

Contents

| 1. 2. | Introdu Data an | ıction | 1435 1435 |
|----------|--------------------|---|--------------|
| | 2.1. I | Data | 1435 |
| | : | 2.1.1. ERAI reanalysis | 1435 |
| | 2 | 2.1.2. ECMWF forecasts | 1436 |
| | 2.2. I | Methods | 1436 |
| | | 2.2.1. Principal component analysis | 1436 |
| | 2 | 2.2.2. Weibull distribution | 1436 |
| | 2 | 2.2.3. Multi-polynomial regression | 1437 |
| | 2 | 2.2.4. Optimizing the number of principal components through cross validation | 1439 |
| 3. | Evaluat | ting the reconstruction methods | 1439 |

* Corresponding author. IPSL/LMD, CNRS, Ecole Polytechnique, Université de Paris-Saclay, Palaiseau, France.

E-mail address: bastien.alonzo@lmd.polytechnique.fr (B. Alonzo).



| | 3.1. Performance of methods for wind speed distribution reconstruction | . 1439 |
|----|--|--------|
| | 3.2. Performance of the methods for estimating the capacity factor | . 1440 |
| 4. | Towards monthly and seasonal forecast of the wind speed distribution | . 1444 |
| 5. | Conclusion | . 1445 |
| | Acknowledgments | . 1445 |
| | Cramer-Von Mises score | . 1445 |
| | References | . 1445 |
| | | |

1. Introduction

Owing to a well-established technology and the ever stronger push towards replacing fossil fuels with clean renewable power, wind energy has seen a dramatic growth in the recent years. According to the European Wind Energy Association, about 12.8 GW of wind power was installed in the European Union (EU) in 2015, bringing EUs total installed capacity to 141.6 GW. This corresponds to an electricity generation sufficient to cover 11.4% of the EUs electricity consumption during an average year [1].

With the growing importance of wind energy, the interest and demand for forecasts of the wind speed near the surface has seen a major boost. Numerous methods exist for forecasting the wind speeds at different forecast horizons implying different applications [2,3]. Many studies focus on the short-term scale ranging from several minutes to 1 day [4–6]. Medium-term forecast methods, ranging from several days up to 10 days, have also been well investigated [7–9]. On much longer timescales and with very different implications and motivations, the impact of climate change on wind speeds has also been addressed [10–12].

By contrast, the intermediate timescale ranging from one month to a season (hereafter referred to as long-term) has received only little attention. Monthly and seasonal forecasts can be very useful for example in maintenance planning, financial estimates and predictions of electricity generation for network management. Some studies showed good results in forecasting the monthly mean wind speed at several observation sites by using Artificial Neural Network models (ANN) [13,14], giving an accurate trend of the wind speed at the yearly horizon, but a limited information on the wind variability at higher frequency. Other authors forecasted daily mean wind speed at the seasonal scale using ANN [15–17] allowing to gather more information on the wind variability inside a given season and which would allow to evaluate the energy production. The ANN output is a predicted wind time serie. They calculate the error regarding the real wind speed and compare the results to other ANN [15,16] or other statistical methods namely ARIMA models [17]. As ANN behaves like black box fed with data, the results are difficult to explain physically. Moreover, each methods focuses on different observation sites, thus giving a limited idea of the spatial variations of the method performance. Even though there are very few works on seasonal forecasts of wind speeds, seasonal forecasting of other meteorological quantities is a popular research topic with continuous improvement. For example, there have been many works on seasonal forecasts of recurrent oscillating patterns in the atmosphere, such as the El Nino [18,19].

This paper focuses on modelling the wind variability on the long-term timescale and makes an attempt of long-term wind speed distribution forecasting. The method proposed in this work aims to use the information found in the large-scale configuration of the atmosphere in order to reconstruct expected distribution of surface winds. This paper answers some questions that arise from this topic:

- How much information on the monthly or seasonal distribution of surface winds can we obtain from knowledge of only the large-scale state of the atmosphere?
- Are the proposed methods performing better than the climatology in reproducing the surface wind speed distribution, and in estimating the electricity generation?
- Do seasonal forecasts from an operational center of weather production contain relevant information for an attempt of forecasting wind speed distributions and electricity generation?

To address these questions in a consistent framework, we use data from the European Center for Medium-Range Weather Forecasts (ECMWF). Indeed, surface winds from ECMWF reanalysis have been shown to well reproduce the observed surface winds in France [20]. Using reanalysis data allows a better investigation of the statistical relation between local surface winds and the large-scale circulation variability, especially because they provide a continuous description of surface wind speed over a wide domain and over long time period. We focus on France and its vicinity not only because the reanalyzed winds had been assessed there, but also because France has a significant wind energy potential and interestingly includes regions with different wind regimes. In Northern France the wind energy potential stems from the storm tracks, whereas local orographic effects and channeling play a major role in strong wind events of Southern France [21].

In the first part of this paper, the data and methodology used to link the large scale circulation with the surface wind speed and to reconstruct its monthly/seasonal distributions is described. Then, the performance of the proposed methods is evaluated by comparing their results to the climatological distributions. The performance is evaluated in terms of reconstructed electricity generation as well. In the last part of the paper, an attempt in forecasting wind speed distributions and electricity generation is discussed.

2. Data and methods

2.1. Data

2.1.1. ERAI reanalysis

Wind speed, geopotential height at 500 hPa (Z500) and Mean Sea Level Pressure (MSLP) are collected from ERA-Interim reanalysis (ERAI, [22]) with a time-step of six hours during 35 years between 01/01/1979 and 12/31/2013, and then averaged to daily data in order to remove the intra-day variability of the wind. Indeed, large-scale circulation of the atmosphere cannot explain intra-day variability such as the diurnal cycle. Moreover, at the seasonal timescale, 6-hourly wind speed variance amounts to about 15% of the total variance. The horizontal resolution of ERAI is 0.75° in latitude and longitude. Z500 and MSLP span the North Atlantic and European grid (20° N to 80° N and 90° W to 40° E), and the surface wind speeds are obtained for a domain encompassing France (40.5° N to 52.5° N and -6.75° W to 10.5° E). The ERAI reanalysis data act as the reference data for wind speed. B. Jourdier [20] showed that the ERA-Interim reanalysis has a good skill for wind speeds in France, and is the best in comparison to two other reanalyses: MERRA and the NCEP/NCAR. To reconstruct the distribution of the wind, a 20-year calibration period, on which we train our methods, has been defined from 1 January 1979 to 31 December 1998. Then a validation period lasting 15 years from 01 January 1999 to 31 December 2013 follows. This period is used to assess the reconstructed distributions with respect to seasonal and monthly wind speed distributions based on ERA-I.

2.1.2. ECMWF forecasts

In the forecast section, the full 35 years period of ERAI is used as a calibration period, while the period of forecast is always of 3 months, allowing to predict either monthly or seasonal distribution of the surface wind speed. We retrieve twelve seasonal forecast sets of ECMWFs numerical weather prediction model [23], from the years 2012, 2013 and 2014, each lasting three months, starting from January, April, July and October. Each set is composed of 41 seasonal forecast members from which we compute the most likely scenario. This scenario is used as the only forecasted state of the atmosphere. We apply the same methods using the 35 years of ERAI to learn the relation between the surface wind speed and the large-scale circulation of the atmosphere, and apply this relation to the forecasted state of the atmosphere to predict wind speed distribution. In this part of the paper, we use ECMWF Analysis as reference for wind speed. Indeed, ECMWF analysis allows to confront the methods to data which are very close to the actual wind speed. The assimilation system and model may not differ to much from the one used for ERA-I reanalysis.

2.2. Methods

In this paper, we follow an approach similar to the perfect model approach ([24]) by using the surface wind speed from ECMWF reanalysis or analysis as the reference against which reconstructed wind speed distributions must be evaluated. By construction, the surface wind speed and 500-hPa geopotential height from ECMWF reanalysis or analysis are consistent between each other. This approach ensures to isolate the errors associated with the reconstruction methods only. Evaluating against surface wind speed measurements would require to quantify the various sources of errors, including representativity and instrument errors, which is out of the scope of this work and would not help in quantifying the ability of our method for monthly/seasonal forecast of wind speed.

At a monthly to seasonal timescale, the surface wind speed is mainly explained by the large scale circulation of the atmosphere. The geopotential height at 500 hPa (Z500) and the Mean Sea Level Pressure (MSLP) are variables that well summarize this circulation. In this paper, we only present the results of reconstruction using the Z500 variable as a predictor of the surface wind speed. Indeed, results found when adding MSLP to Z500 predictor were comparable and the improvement was neither systematic nor significant.

In the following paragraphs, we describe in detail the reconstruction methodology which is summarized in Fig. 1.

Our attempt aims at reconstructing the distribution of winds on the monthly to seasonal timescales, but not at reconstructing daily timeseries of winds. Indeed, our reconstruction methodology is based on the principal components analysis of the Z500 predictor which informs about the large-scale state of the atmosphere. This knowledge will constrain the likely distribution of surface winds on timescales larger than the lifetime of individual synoptic systems (fronts, storms) and thus will not allow to reconstruct such high frequency timeseries. Following the common practice, we use the Weibull distribution to summarize the surface wind speed distribution [25,26].

2.2.1. Principal component analysis

To obtain a more compact representation of the large-scale situation we perform a Principal Component Analysis (PCA) on Z500. It results in a set of Empirical Orthogonal Functions (EOF), which represent the typical oscillation patterns spanning the North Atlantic domain. Each EOF is associated with one scalar timeseries (the corresponding PC) which describes how each pattern evolves in time. Fig. 2 shows the five first EOFs and their associated PCs.

The first PC corresponds to the seasonal cycle (Fig. 2 a,b), explaining as much as 54.1% of the variance in the dataset: in winter the meridional pressure gradient strengthens, leading to stronger winds and more intense synoptic systems. The following four PCs have a clear physical interpretation [27,28], they all be related to teleconnection patterns, respectively the North Atlantic Oscillation (NAO) (Fig. 2 c,d), the Eastern Atlantic Pattern (EA) (Fig. 2 e,f), the Scandinavian pattern (SCA) (Fig. 2 g,h) and the 2nd European pattern (EU2) (Fig. 2 i,j). These five first PCs explain 76.8% of the variance in the entire dataset (Table 1).

2.2.2. Weibull distribution

To summarize the wind distributions, we choose the Weibull distribution as the parametric representation for monthly and seasonal distribution of the surface wind speed at a given location. This theoretical distribution is widely used in the wind energy industry [29,30,25]. It provides a simple way to represent the wind distribution as it is based on only two parameters: the shape parameter and the scale parameter. We must highlight the fact that other theoretical distributions better capture the shape of the real wind distribution. In particular, the Rayleigh-Rice distribution can have two modes, which is not the case for the Weibull [21,31].

The probability density function (PDF) and the cumulative distribution function (CDF) of the Weibull distribution are expressed as follows.

$$f(u;k,c) = \frac{k}{u} \left(\frac{u}{c}\right)^k e^{-(u/c)^k}$$
(1)

$$F(u;k,c) = 1 - e^{-(u/c)^k},$$
(2)

where *u* is the wind speed, *k* and *c* are respectively the shape and the scale parameter.

We now define three ways to reconstruct the parameters k and c from the data. The WAsP method, referred in the following as WAsP [32], computes these parameters from the moments \overline{U} and $\overline{U^3}$, as well as the probability of exceeding the mean wind speed $1 - P(\overline{U})$ (which must be estimated from the data). The method focuses on the right-hand tail of the Weibull distribution, which is an important part of the distribution in terms of energy [33]. This is why the WAsP method is preferred amongst the wind energy industry. In this method, k and c are calculated by solving the following equations.

$$\frac{\overline{U}^3}{\overline{U}^3}\Gamma\left(1+\frac{3}{k}\right)^{\frac{k}{3}} = -\ln(1-P(\overline{U}))$$
(3)

$$c = \sqrt[3]{\frac{\overline{U^3}}{\Gamma\left(1 + \frac{3}{k}\right)}} \tag{4}$$

In a second method, referred in the following as KCrec, we take



Fig. 1. Reconstruction Method: Flow chart describing the reconstruction methodology.

advantage of the fact that the Weibull distribution is given by two parameters, k and c, and straightforwardly reconstruct these: they are fitted by the Maximum Likelihood Estimator (MLE) [34] on the calibration period. The MLE of the Weibull parameters is defined by the following equations.

$$\frac{\sum_{i=1}^{n} u_{i}^{k} \ln(u_{i})}{\sum_{i=1}^{n} u_{i}^{k}} - \frac{1}{k} - \frac{1}{n} \sum_{i=1}^{n} \ln(u_{i}) = 0,$$
(5)

$$c = \frac{\sum_{i=1}^{n} u_i^k}{n}.$$
(6)

A last method was introduced in order to take into account how spread out the wind distribution is. This method, referred in the following as Perc, uses two values, $F(u_1)$ and $F(u_2)$, of the Weibull distribution function, corresponding to wind speeds u_1 and u_2 . The Weibull k and c parameters are then given explicitly by:

$$c = \frac{\ln \ln\left(\frac{1}{1-F(u_2)}\right) \ln(u_1) - \ln \ln\left(\frac{1}{1-F(u_1)}\right) \ln(u_2)}{\ln \ln\left(\frac{1}{1-F(u_2)}\right) - \ln \ln\left(\frac{1}{1-F(u_1)}\right)},$$
(7)

$$k = \frac{c}{u_1} \ln \ln \left(\frac{1}{1 - F(u_1)} \right).$$
(8)

In order to determine the optimal values of u_1 and u_2 , a synthetic test was performed. First, we generated 30 (one month) or 90 (one season) samples from the reference Weibull distribution with parameters k = 2 and c = 3.5. Next, we determined the two Weibull parameters from the simulated samples using the Perc method, using different combinations (u_1, u_2) . To find the best combination, we compared the resulting distributions with the reference

distribution using the Cramer-von Mises (CvM) score (see Appendix). It was found that the best combination on a monthly scale is the 11th and 83rd percentile. On the seasonal scale, the optimal combination is the 17th and the 92nd percentile. The combination of the percentiles was not found to be very sensitive, as there was a small region around the optimum combination with very similar scores.

2.2.3. Multi-polynomial regression

We propose to link the large-scale situation and surface wind speed distribution by a multi-polynomial regression taking the monthly mean PCs as explanatory variables and the parameters of the Weibull distribution as dependent variables:

$$\tilde{P} = \beta_0 + \sum_{n=1}^{N} \beta_n C_n(t) + \sum_{n=1}^{N} \beta_{n,n} C_n(t)^2 + \sum_{n=1}^{N-1} \times \sum_{m=n+1}^{N} \beta_{n,m} C_n(t) C_m(t).$$
(9)

Here, \tilde{P} is the dependent variable (Weibull parameter *k* or *c* for a given location), *C*_n are the principal components and $\beta_{n,n}$ and $\beta_{n,m}$ are the regression weights found by least squares. The number *N* of principal components is determined by cross validation as explained below. We perform the regression on a calibration period of 20 years between 1979 and 1998. This results in weights quantifying the relationship between the large-scale circulation and the Weibull parameters for each individual location. These weights can be combined with the known PC values on the reconstruction period of 15 years between 1999 and 2013 to reconstruct the monthly/seasonal Weibull distribution.



Fig. 2. EOFs & associated PCs: Five firsts EOFs (left side) and five first PCs (right side) of the PCA performed on the 35 years and on the entire domain of ERAI Z500 dataset.

Table 1

Variance Explained: Five first Principal Components of the Z500 and the percentage of variance explained by each of them.

| Pattern | РС | Variance explained (%) | Cumulative variance explained (%) |
|----------------|----|---------------------------|-----------------------------------|
| Seasonal cycle | 1 | 54.1 | 54.1 |
| NAO | 2 | 8.1 | 62.2 |
| EA | 3 | 5.6 | 67.8 |
| SCA | 4 | 5.2 | 72.9 |
| EU2 | 5 | 3.9 | 76.8 |

2.2.4. Optimizing the number of principal components through cross validation

The first five PCs of the Z500 can be easily interpreted as predictors of the wind. Still, to a certain extent, the following PCs can also explain the variability of the wind at the monthly/seasonal scale. To check whether taking five PCs is really optimal, we performed a cross-validation procedure. For this purpose, we calculated the temporally and spatially averaged CvM score (see Appendix) of 7 reconstructions of 5 years each, taking the remaining 30 years of the data set as calibration period. Fig. 3 plots the CvM scores as function of the number of PCs used. The minimum mean CvM is clearly apparent for all three methods for both monthly (Fig. 3 a, b, c) and seasonal (Fig. 3 d, e, f) reconstruction. The minimum is found for 5 PCs for monthly distribution reconstructed by Wasp method (Fig 3 a), 7 PCs for monthly distribution reconstructed by KCrec method (Fig 3 c), and 6 PCs for the others (Fig. 3 b,d,e,f). As the 5 first PCs are all related to well understood teleconnection patterns explaining the large-scale circulation of the atmosphere, we expect them to be accurately linked to the wind speed variability at the monthly and seasonal timescales. Results obtained by cross-validation confirms this relation, as the optimal amount of PCs used is close to 5 (Fig. 2).

3. Evaluating the reconstruction methods

As mentionned in the Introduction, we use the wind speed from the ERAI-reanalysis as the reference wind speed. To assess the reconstruction quality, the CvM score (see Appendix) is calculated between the reconstructed CDF and the ERA-I wind CDF. The CvM scores of the reconstructed wind speed distributions are then compared to the CvM scores computed between the ERA-I wind CDF and the ERA-I climatological CDF. In simple terms, the climatological distribution is the distribution of all values of wind for each month or season in one specific location, based on all reanalysis data from this location and the specific month or season. The climatological distributions are usually used by the industry to have a first assessment of the wind energy production at a seasonal timescale. An example of ERA-I, ERA-I climatological and reconstructed wind speed CDFs is shown in Fig. 4. In the following, the monthly and seasonal ERA-I climatological distributions (averaged over the 20-year reference period) will be denoted 'climatology'.

3.1. Performance of methods for wind speed distribution reconstruction

The CvM score allows to test the null hypothesis (H0) that the two samples come from the same distribution. Assuming that the reconstructed distributions and the ERA-I distributions are based on samples large enough to say that the corresponding CvM scores follow the limiting distribution, we can define the p-value corresponding to 95% confidence (see Appendix). If the calculated CvM score is below this value, we can say at 95% confidence that the two compared samples come from the same distribution. We compare results of the tests for the climatology and the reconstruction methods. We can define five different cases:

- Case A: H0 is not rejected for the method and rejected for the climatology
- Case B: H0 is not rejected for both and the CvM of the method is smaller than the CvM of the climatology
- Case C: H0 is not rejected for both and the CvM of the method is larger than the CvM of the climatology
- Case D: H0 is rejected for the method and not rejected for the climatology
- Case E: H0 is rejected for both the method and the climatology



Fig. 3. *Cross validation:* Mean CvM score obtained by cross validation in function of the number of PCs used to reconstruct the distribution of the surface wind speed. From left to right: Wasp (a,d), Perc (b,e), and KCrec (c,f) methods; top: CvM score for monthly wind distribution reconstruction (a,b,c); bottom: CvM score for seasonal wind distribution reconstruction (d,e,f).



Fig. 4. CDFs example: ERA-I, ERA-I climatological, and reconstructed seasonal CDFs for winter 2012 at 48.5° N 3.0° W.

The hypothesis of the Weibull distribution introduces a bias in the reconstructed CDF. We compare results obtained for reconstructed CDF to ERA-I climatology (a) and to the parametric climatology (b) which follows a Weibull distribution fitted by MLE on the ERA-I wind speed. Results over the whole domain in all different cases are given in Tables 2 and 3 for monthly and seasonal reconstruction respectively.

The first lines of Tables 2 and 3 show the fraction of time each method gives a distribution not discernable from the ERA-I distribution at 95% confidence level. It shows that all methods, appart from Wasp, have a good ability to reconstruct the real wind distribution. We can also see that fitting a Weibull distribution on the climatology reduces by about 10% this percentage. Cases A and B summarize the number of time each method is doing better than the climatology (non-parametric (a) or parametric (b)). On the contrary, Cases C and D summarize the number of time the climatology is doing better than the method. On average, on the all domain and for monthly and seasonal timescales, the nonparametric climatology (a) do better than every methods more than 60% of the time (78.1% against Wasp at the seasonal scale, to 62.3% against KCrec at the monthly scale). Nevertheless, when comparing to the parametric climatology, for monthly and seasonal reconstruction, the KCrec method performs 49.1% of the time better at monthly scale, and 48.1% at the seasonal scale. This shows again the error brought by the Weibull distribution reconstruction. In all cases, methods perform better at the monthly scale than at the seasonal scale. It is interesting to notice that the cases for which the percentage is increased at the seasonal scale are cases D and E, corresponding to times when reconstructed distribution cannot be believed to come from the same distribution as the ERA-I sample, at 95% confidence level. (Tables 2 and 3).

Figs. 5 and 6 show on average on the validation the number of time each method behaves better than the classical climatology

(Cases A and B). It can be seen that the Perc and KCrec methods do better than the Wasp method. Indeed, at monthly timescale, the Perc and KCrec methods can do better than the climatology in average more than 30% of times, while the Wasp method does better than the climatology about 25% of times on average displaying a clear difference between north and south (Fig. 5 and Table 2). On a seasonal scale, the Wasp method performs clearly worse than at a monthly scale. The Perc and KCrec methods at a seasonal scale display an interesting spatial variability. Indeed, they do more than 40% of times better than the climatology in the north of France, whereas in the south, this percentage is about 20%–25% (Fig. 6). When comparing to the parametric climatology, all methods display the same pattern, but all percentages are increased by more than 10% (Not shown).

Fig. 7 shows the ratio of the number of times each method is doing better than the climatology for seasonal distributions, by taking each season separately. We can clearly see on this figure that the performance regarding the climatology of the Perc and KCrec methods, and to a certain extent the Wasp method, depends on the season and on the region. Indeed, both the Perc and the KCrec methods display a high percentage of times (up to 70% at some points) when they do better than the climatology in the north of France for the winter and autumn seasons. In the north of France, the storm track variability is highly responsible for the inter-annual variability of the wind speed distribution. In winter and fall, it is very active compared to spring and summer. This explains why reconstruction methods perform better than the climatology in this region, in those seasons. The comparison analysis mainly highlights the fact that, by construction, the climatology captures only a mean state of the atmosphere over 20 year (and is not significantly affected by long-term wind trends), while the model captures the variability of the wind at timescales from intra-annual to decadal.

3.2. Performance of the methods for estimating the capacity factor

For wind energy purposes, it is not exactly the full wind distribution that needs to be estimated. For a given turbine, once the wind is beyond the cut-out wind speed or below the cut-in speed, the precise value does not matter. In the present section we take this into account and reevaluate each method. A preliminary step consists in designing a procedure which imitates the weighting of wind values by a power curve, in a manner which accounts for the considerable geographical variations of the wind (a single, generic power curve would not make sense).

Each wind turbine is characterized by its power curve which gives the output power as a function of the wind speed. The energy produced during a given period can be expressed as:

$$E = T \int_{0}^{\infty} P_{out}(u) dU, \tag{10}$$

where *T* is the period considered (month or season) and $P_{out}(u)$ is

Table 2

CvM test results: Percentage of time the result of the CvM test gives Cases A,B,C,D, or E on the whole domain, for the entire validation period, for monthly reconstructed distribution compared to the classical climatology (a) and to the parametric climatology (b). The p-value, p, is 0.46136 for 95% confidence level (see Appendix).

| | | 05 () | 1 | 65 () | 1 . 1 . | | | 11 2 |
|-----------------|------|--------|------|--------|---------|------|------|-----------------|
| Methods | Wasp | | Perc | | KCrec | | Clim | Parametric Clim |
| CvM < p | 69.1 | | 82.1 | | 85.2 | | 89.3 | 81.7 |
| Comparison with | a | b | a | b | a | b | - | _ |
| Case A | 5.8 | 11.5 | 6.5 | 11.9 | 6.9 | 13.0 | - | _ |
| Case B | 17.8 | 24.1 | 25.0 | 34.0 | 27.3 | 37.1 | - | _ |
| Case C | 45.5 | 33.5 | 50.5 | 36.2 | 51.1 | 35.1 | _ | _ |
| Case D | 26.0 | 24.1 | 13.8 | 11.6 | 11.2 | 9.5 | - | - |
| Case E | 4.8 | 6.8 | 4.1 | 6.3 | 3.7 | 5.3 | - | - |
| | | | | | | | | |

| Table 3 | | | | |
|---------|---------|---------|----------|--------------|
| Same as | Table 2 | but for | seasonal | distribution |

| Methods | Wasp | | Perc | | KCrec | | Clim | Parametric Clim |
|-----------------|------|------|------|------|-------|------|------|-----------------|
| CvM < p | 44.3 | | 73.3 | | 79.8 | | 88.6 | 77.3 |
| Comparison with | a | b | a | b | a | b | _ | _ |
| Case A | 3.8 | 8.9 | 5.5 | 11.0 | 6.1 | 13.9 | _ | _ |
| Case B | 10.5 | 13.6 | 22.8 | 31.2 | 23.8 | 34.2 | _ | _ |
| Case C | 30.0 | 21.9 | 45.0 | 31.1 | 49.9 | 31.7 | _ | _ |
| Case D | 48.1 | 41.9 | 20.8 | 15.0 | 14.9 | 11.4 | _ | _ |
| Case E | 7.5 | 13.8 | 5.9 | 11.7 | 5.3 | 8.8 | _ | _ |



Fig. 5. Reconstructed CDFs vs climatology: Fraction of times each method does better than the climatology (cases A and B) for monthly distribution reconstruction. From left to right: Wasp (a), Perc (b), KCrec (c).

the output power given the wind speed *u*. The capacity factor (CF) is defined as the ratio between the actual energy produced during a given period and the energy that would have been produced if the wind turbine had run at its maximum power during the entire period:

$$CF = \frac{E}{P_n T},\tag{11}$$

where P_n is the nominal power of the wind turbine.

In order to take into account the fact that the data used are at 10m height and the mean wind speed is highly varying among different locations, we use a location-adapted power curve, proposed by Ref. [31]. In this curve, the wind speed is divided by a location-dependent parameter *a*, chosen so that the modified power curve has a capacity factor of 23% on the calibration period. This corresponds to the average capacity factor in France in 2014 [35]. This procedure is illustrated in Fig. 8.

To assess the accuracy of the reconstructed capacity factor, the relative error between the reconstructed capacity factor and the capacity factor from the reanalysis is computed:

$$\Delta CF = \frac{CF - CF_{real}}{CF_{real}} \tag{12}$$

Figs. 9 and 10 show the relative error on the calculated capacity factor for monthly and seasonal reconstructions respectively. At both timescales, the Perc method overestimates it mostly onshore



Fig. 6. Same as Fig. 5 but for seasonal distribution reconstruction.



Fig. 7. Reconstructed CDFs vs climatology, Seasonal fraction: Fraction of time each method do better than the climatology (cases A and B) for seasonal distribution reconstruction based on Z500 for each season. From left to right: Winter, Spring, Summer and Autumn; From top to bottom: Wasp, Perc, and KCrec methods.

by about 25% on average. The KCrec method behaves like the Perc methods at the monthly scale, but is performing better at the seasonal scale with an overestimation of about 10% onshore. As expected, the Wasp method shows good performance in estimating



Fig. 8. *Power curve*: Example of the location-adapted power curve. In solid black: the real power curve for wind speed at 80 m height; in dashed blue: the adapted power curve. It has the same shape, but the wind speed is divided by a number *a* to achieve a capacity factor of 23%. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the capacity factor as its reconstruction focuses on the right tail of the Weibull distribution. Nevertheless, it overestimates the energy production in the northern part of France at a monthly scale and underestimates it in the southern part of France at a seasonal scale. The non-parametric climatology behaves very well at the seasonal scale even though it displays a slight overestimation in the north of France. At the monthly scale, on average, on the entire domain it overestimates the capacity factor by about 25%. By contrast, the parametric climatology behaves very badly at the monthly scale, overestimating the energy production by 50% in average. At a seasonal scale, this overestimation decreases but is still high, highlighting again the error induced by the Weibull distribution hypothesis.

In any case, there is a tendency of all methods to overestimate the capacity factor, mostly onshore. The climatology acts as a filter of high frequency variation of the wind, meaning that it does not describe well the tails of the distribution. As the power curve is designed so that the wind turbine works at its nominal power near the mean wind speed, this results in an overestimation of the capacity factor.

On the other hand, Drobinski et al. [21] showed that a Weibull distribution fitted by MLE describes well the center of the distribution (near the mean wind speed), but tends to underestimate the tails of the distribution. This leads to the same consequence. This explains why the parametric climatology acts worse than the non-



Fig. 9. Reconstruction of the capacity factor: Relative error on the capacity factor (%) for monthly distributions given by: non parametric climatology (a), parametric climatology (b), Wasp (c), Perc (d), and KCrec (e).



Fig. 10. Same as Fig. 9 but for seasonal distributions.

1443

parametric climatology, but also why KCrec overestimates the capacity factor. This has no such effect offshore because the wind above sea is steadier so that the distribution is more peaked around the mean. Regarding the Perc method, the Weibull reconstruction is based on two percentiles defined to minimize the CvM score. It may results in the same effect of underestimating the tails of the distribution. Future work could focus on a sensitivity analysis to the percentiles definition by minimizing the error on capacity factor.

At the seasonal scale, the real distribution is based on a larger sample which implies that the center of the distribution has a much larger weight than the tails at this scale than at the monthly scale. The effect of underestimating the tails is thus less visible.

4. Towards monthly and seasonal forecast of the wind speed distribution

The analysis described above has shown that the large-scale state of the atmosphere contains information on the likely distribution of surface winds, and our proposed methods allow to recover at least part of this information. A long-term perspective will be to use this to build forecasts of surface wind distributions. Below we present a preliminary attempt based on seasonal ensemble forecasts of the Z500 from ECMWF [23], to assess the potential of this method for monthly or seasonal forecasts of the wind speed distribution.

A first step is to assess the skill in seasonal forecasts for predicting the large-scale state of the atmosphere in our region of interest. The root mean square error (RMSE) between the daily PCs of Era-Interim and those of the seasonal forecast is shown in Fig. 11. This figure gives an idea of the lead-time of such a forecast. It shows that the error increases rapidly until it levels off after 20 days indicating that there is no more valuable information on the largescale circulation in the data. As a consequence, it will not be possible to have an accurate wind distribution forecast at more than the monthly horizon.

One technical difficulty arises: the monthly distribution of wind coming from the ECMWF analysis stands for the real distribution. As the analysis does not come from the same model as the ERA-Interim data, a bias exists between the distributions coming from the analysis and the distributions based on ERA-Interim data. We thus apply a classical quantile/quantile correction between the 4 years based distributions of the analysis and of ERA-Interim between 2012 and 2015 at each point of the gridded domain. We apply this correction to the monthly wind distribution of the analysis. Because of the small amount of forecasts and of the uncertainties due to the bias, we will not be able to have the same deep analysis as in the reconstruction part of the paper. The corrected monthly distribution of the wind speed coming from the analysis is compared to the climatology of ERA-Interim and to the forecast distributions using the CvM score.

The percentage of time each method does better than the climatology, averaged over the entire domain, for the 1st month of the 12 forecasts, is summarized in Table 4. The results for the Perc and KCrec methods are comparable to the reconstruction results. On the contrary, the Wasp method shows a very high score when evaluating the entire distribution and a lesser score when evaluating the energy production, which is not consistent with the reconstruction results. When calculating the error on the capacity factor, the forecast methods always highly overestimate the wind energy production onshore (more than 100% at some points), and



Fig. 11. Errors in raw seasonal forecasts: RMSE calculated between the PCs of Era-Interim and the PCs of the seasonal forecast. The solid line represents the median of the error, dashed lines represent the 60th percentile (top) and the 40th percentile (bottom). a. Seasonal, b. NAO, c. EA, d. SCA, e. EU2.

Table 4

Forecasts results: Percentage of the number of times each method does better than the climatology on the whole domain for the 3 years of forecasts. First values correspond to the evaluation of the entire distribution; values in parenthesis corresponds to the evaluation of the distribution between the cut in and the cut out.

| Forecast method | Wasp | Perc | KCrec |
|----------------------|---|---|---|
| total 1st month | 46.4 (31.2) | 20.2 (25.5) | 28.8 (27.5) |
| 2012 2013 2014 | 41.0 (35.0) 44.1 (25.7) 54.0 (33.0) | 15.1 (20.5) 22.9 (25.4) 22.5 (30.4) | 22.9 (23.6) 32.4 (26.5) 31.3 (32.3) |

slightly underestimate it offshore (more than 10%). The nonparametric climatology overestimates the capacity factor by more than 10% onshore and underestimates it offshore, whereas the parametric climatology highly overestimates the energy production on the whole domain as it was the case in the evaluation part.

Regarding the large uncertainty due to the limited number of forecasts, the robustness can be inferred from the consistency of the forecasts results with those obtained in the previous section.

Still, work must be continued to evaluate the forecasts performance of such methods, by using larger sets of numerical seasonal weather forecast, but also by testing methods based on nonparametric distribution estimation.

5. Conclusion

In this paper, a new approach for modelling the wind speed at the seasonal scale has been proposed. We suggest to model not only the mean wind speed but the entire monthly/seasonal distribution of the wind. Linking the wind to its synoptic predictors we have shown that there is valuable information in the large-scale circulation variability that can explain the wind speed distribution at such long timescales. The proposed methods show good performances in reconstructing the monthly and seasonal wind speed distributions even if the climatology is still a good predictor. Moreover, reconstruction methods performances display an interesting spatial and seasonal variability. Indeed, in the north of France in winter and fall, the proposed methods showed better ability to model strong winds than the climatology. Nevertheless, the attempt of forecasting also highlights the fact that seasonal forecasts of ECMWF are not yet mature enough to give valuable information on the large-scale circulation variability at the horizons exceeding a month. Future work will be dedicated to the evaluation of the method against in-situ data.

Acknowledgments

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 [36]) through the working group Renewable Energy.

Appendix. Cramer-Von Mises score

To assess the reconstruction quality, we use the Cramer-Von-Mises score defined in Anderson et al. [37]:

$$C\nu M = \frac{MN}{M+N} \int_{-\infty}^{\infty} \left[F_N(x) - F_M(x)\right]^2 dH_{M+N}(x)$$
(.1)

Here, *M* and *N* are the sample sizes in each of the distributions, $F_N(x)$ and $F_M(x)$ are the CDFs of the two samples and $H_{M+N}(x)$ is the

combined distribution of the two samples together. The smaller the CvM score, the better the goodness of fit between the two tested distributions. Anderson et al. [37] showed that Equation (.1) is equivalent to

$$CvM = \frac{U}{NM(M+N)} - \frac{4NM - 1}{6(N+M)},$$
 (.2)

where $U = N \sum_{i=1}^{N} (r_i - i)^2 + M \sum_{j=1}^{M} (r_j - j)^2$, r_i are the ranks of the elements of the sample of size N in the combined sample and r_j are the ranks of the sample of size M in the combined sample.

The CvM score allows to test the null hypothesis H0:"the two samples come from the same distribution". When $M \rightarrow \infty$ and $N \rightarrow \infty$, under the null hypothesis, the CvM score follows the limiting distribution with mean $\frac{1}{6}$ and variance $\frac{1}{45}$. In this configuration, the p-value giving 95% confidence that the null hypothesis is true is p = 0.46136, [37].

References

- EWEA, Wind in Power: 2014 European Statistics, Tech. rep., European Wind Energy Association, 2015.
- [2] S.S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: North American Power Symposium (NAPS), 2010, pp. 1–8.
- [3] W. Chang, A literature review of wind forecasting methods, J. Power Energy Eng. 2 (2014) 161–168.
- [4] A. Sfetsos, A novel approach for the forecasting of mean hourly wind speed time series, Renew. Energy 27 (2002) 163–174.
- [5] P. Gomes, R. Castro, Wind speed and wind speed forecasting using statistical models: Autoregressive Moving Average (ARMA) and Artificial Neural Networks (ANN), Int. J. Sustain. Energy Dev. (IJSED) 1 (2012) 41–50.
- [6] A. Carpinone, M. Giorgio, R. Langella, A. Testa, Markov chain modeling for very-short-term wind power forecasting, Electr. Power Syst. Res. 122 (2015) 152–158.
- [7] J. Taylor, P. McScharry, R. Buizza, Wind power density forecasting using ensemble prediction and time series model, IEEE Trans. Energy Convers. 34 (2009) 775–782.
- [8] M. Wytock, J.Z. Kolter, Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields, in: Proceedings of the IEEE Conference on Decision and Control, 2013, pp. 1019–1024.
- [9] T.G. Barbounis, J.B. Theocharis, M.C. Alexiadis, P.S. Dokopoulos, Long-termwind speed and power forecasting using local recurrent neural network models, IEEE Trans. Energy Convers. 21 (2006) 273–284.
- [10] J. Najac, J. Boe, L. Terray, A multi model ensemble approach for assessment of climate change impact on surface winds in France, Clim. Dyn. 32 (2009) 615–634.
- [11] D.J. Sailor, M. Smith, M. Hart, Climate change implications for wind power resources in the northwest United States, Renew. Energy 33 (2008) 2393–2406.
- [12] S. Pryor, R. Barthelmie, Climate change impacts on wind energy: a review, Renew. Sustain. Energy Rev. 14 (2010) 430–437.
- [13] M. Bilgili, B. Sahin, A. Yasar, Application of artificial neural networks for the wind speed prediction of target station using reference stations data, Renew. Energy 32 (2007) 2350–2360.
- [14] H.B. Azad, S. Mekhilef, V.G. Ganapathy, Long-term wind speed forecasting and general pattern recognition using neural networks, IEEE Trans. Sustain. Energy 5 (2014) 546–553.
- [15] J. Wang, S. Qin, Q. Zhou, H. Jiang, Medium-term wind speeds forecasting utilizing hybrid models for three different sites in xinjiang, China, Renew. Energy 76 (2015) 91–101.
- [16] Z. Guo, W. Zhao, H. Lu, J. Wang, Multi step forecasting for wind speed using a modified EMD based artificial neural network model, Renew. Energy 37 (2012) 241–249.
- [17] A. More, M. Deo, Forecasting wind with neural networks, Mar. Struct. 16 (2003) 35–49.
- [18] J. Owen, T. Palmer, The impact of El Nino on an ensemble of extended range forecasts, Mon. Weather Rev. 115 (1987) 2103–2117.
- [19] C. Cassou, Intraseasonal interaction between Madden-Julian oscillation and the North Atlantic oscillation, Nature 455 (2008) 523–597.
- [20] B. Jourdier, Wind Resource in Metropolitan France: Assessment Methods, Variability and Trends, Ph.D. thesis, Ecole Polytechnique, 2015.
- [21] P. Drobinski, C. Coulais, B. Jourdier, Surface wind-speed statistic modelling: alternatives to the Weibull distribution and performance evaluation, Boundary Layer Meteorol. 157 (2015) 97–123.
- [22] D.P. Dee, S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer,

L. Haimberger, S.B. Healy, H. Hersbach, E.V. Holm, L. Isaksen, P. Kallberg, M. Kohler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thepaut, F. Vitart, The erainterim reanalysis: conguration and performance of the data assimilation system, Q. J. R. Meteorol. Soc. 137 (2011) 553–597.

- [23] F. Molteni, T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer, F. Vitart, The New Ecmwf Seasonal Forecast System (System 4), Tech. rep., ECMWF Technical Memorandum, 2011.
- [24] R.D. Elia, R. Laprise, B. Denis, Forecasting skill limits of nested, limited-area models: a perfect-model approach, Mon. Weather Rev. 130 (2002) 2006–2023.
- [25] T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi, Wind Energy Handbook, Wiley, 2011.
- [26] J. Manwell, J. McGowan, A. Rogers, Wind Energy Explained. Theory, Design and Application, Wiley, 2009.
- [27] M. Vrac, P.V. Ayar, P. Yiou, Trends and variability of seasonal weather regimes, Int. J. Climatol. 34 (2013) 472–480.
- [28] C. Cassou, L. Terray, J.W. Hurrel, C. Deser, North atlantic winter climate regimes: spatial asymmetry, stationarity with time, and oceanic forcing, J. Clim. 17 (2004) 1055–1068.
- [29] I. Lun, J. Lam, A study of Weibull parameters using long-term wind observation, Renew. Energy 20 (2000) 145–153.
- [30] C. Justus, W. Hargreaves, A. Yalcin, Nationwide assessment of potential output

from wind-powered generators, J. Appl. Meteorol. 15 (1976) 673–678. [31] B. Jourdier, P. Drobinski, Errors in wind resource and energy yield assessments

- based on the weibull distribution, Ann. Geophys. 35 (2017) 691–700.
 [32] I.T.N.G. Mortensen, L. Landberg, E. Petersen, Wind Atlas Analysis and Application Program (wasp), vol. 1: Getting Started. vol. 2: Users Guide, Tech. rep., Ris National Laboratory, 1998.
- [33] S. Pryor, M. Nielsen, R. Barthelme, J. Mann, Can satellite sampling of offshore wind speeds realistically represent wind speed distributions? Part II: Quantifying uncertainties associated with distribution fitting methods, J. Appl. Meteorol. 43 (2004) 739–750.
- [34] A. Cohen, Maximum likelihood estimation in the Weibull distribution based on complete and censored samples, Technometrics 7 (1965) 579–588.
- [35] RTE, Syndicat des Energies Renouvelables, ERDF and ADEef, Panorama de l'electricite renouvelable 2014, Tech. rep., RTE, 2015.
- [36] P. Drobinski, V. Ducrocq, P. Alpert, E. Anagnostou, K. Branger, M. Borga, I. Braud, A. Chanzy, S. Davolio, G. Delrieu, C. Estournel, N.F. Boubrahmi, J. Font, V. Grubisic, S. Gualdi, V. Homar, B. Ivancan-Picek, C. Kottmeier, V. Kotroni, K. Lagouvardos, P. Lionello, M. Llasat, W. Ludwig, C. Lutoff, A. Mariotti, E. Richard, R. Romero, R. Rotunno, O. Roussot, I. Ruin, S. Somot, I. Taupier-Letage, J. Tintore, R. Uijlenhoet, H. Wernli, A 10-year multidisciplinary program on the Mediterranean water cycle, Bull. Am. Meteorol. Soc. 95 (2014) 1063–1082.
- [37] T. Anderson, On the distribution of two sample Cramer Von Mises criterion, Ann. Math. Stat. 33 (1962) 1148–1159.