Meteorologische Zeitschrift, PrePub DOI 10.1127/metz/2015/0659
© 2015 The authors

Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble

Constantin Junk[1*], Luca Delle Monache[2], Stefano Alessandrini[2], Guido Cervone[3] and Lueder von Bremen[1]

[1]ForWind – Center for Wind Energy Research, University of Oldenburg, Oldenburg, Germany
[2]National Center for Atmospheric Research, Boulder, CO, United States of America
[3]Department of Geography and Institute for CyberScience, The Pennsylvania State University, University Park, PA, United States of America

(Manuscript received November 12, 2014; in revised form February 26, 2015; accepted March 9, 2015)

Abstract

Unlike deterministic forecasts, probabilistic predictions provide estimates of uncertainty, which is an additional value for decision-making. Previous studies have proposed the analog ensemble (AnEn), which is a technique to generate uncertainty information from a purely deterministic forecast. The objective of this study is to improve the AnEn performance for wind power forecasts by developing static and dynamic weighting strategies, which optimize the predictor combination with a brute-force continuous ranked probability score (CRPS) minimization and a principal component analysis (PCA) of the predictors. Predictors are taken from the high-resolution deterministic forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF), including forecasts of wind at several heights, geopotential height, pressure, and temperature, among others. The weighting strategies are compared at five wind farms in Europe and the U.S. situated in regions with different terrain complexity, both on and offshore, and significantly improve the deterministic and probabilistic AnEn forecast performance compared to the AnEn with 10-m wind speed and direction as predictors and compared to PCA-based approaches. The AnEn methodology also provides reliable estimation of the forecast uncertainty. The optimized predictor combinations are strongly dependent on terrain complexity, local wind regimes, and atmospheric stratification. Since the proposed predictor-weighting strategies can accomplish both the selection of relevant predictors as well as finding their optimal weights, the AnEn performance is improved by up to 20 % at on and offshore sites.

Keywords: energy meteorology, wind power forecasting, analog ensemble, uncertainty quantification, probabilistic verification, predictor-weighting strategies

1 Introduction

Accurate wind power forecasts, which can be categorized into deterministic and probabilistic, are important for a safe and efficient operation of wind farms and a cost-effective integration of wind generation into grids (Pinson, 2013, and references therein). While deterministic forecasts are easy-to-understand and single-valued for each forecast lead time and grid box, probabilistic forecasts provide estimates of the forecast uncertainty. Although forecast uncertainty might be reduced due to improved forecast systems, it can never be eliminated because of imperfect knowledge of the initial state of the atmosphere and of some physical processes determining its evolution, numerical approximations, and the atmosphere chaotic nature (Lorenz, 1963). Thus, knowledge of forecast uncertainty can lead to additional value for decision-making (Hirschberg et al., 2011). In particular, the value of probabilistic wind power forecasts has been explored for trading wind generation on the energy market in several studies (Roulston et al., 2003; Zugno et al., 2013; Alessandrini et al., 2014).

Analog-based methods have been successfully implemented for probabilistic forecasts of precipitation, 850-hPa temperature, 2-m temperature, and 10-m wind speed (Van den Dool, 1989; Hamill and Whitaker, 2006; Delle Monache et al., 2011; Panziera et al., 2011; Delle Monache et al., 2013, among others). Recently, the analog ensemble (AnEn) approach proposed by Delle Monache et al. (2013) has been applied to wind power forecasts (Alessandrini et al., 2015), resulting in reliable quantification of the forecast uncertainty. The AnEn is a mean to generate uncertainty (i.e., probabilistic) information from a purely deterministic forecast, rather than being a calibration technique of an existing ensemble, e.g., as proposed by Hamill and Whitaker (2006).

A key aspect of the AnEn algorithm is the search of analogs (past forecasts), which is based on a multivariate metric that estimates the degree of analogy between the current deterministic prediction and past forecasts from a historical data set. As Delle Monache

*Corresponding author: Constantin Junk, ForWind – Center for Wind Energy Research, University of Oldenburg, Ammerländer Heerstr. 136, 26129 Oldenburg, Germany, e-mail: constantin.junk@forwind.de

© 2015 The authors

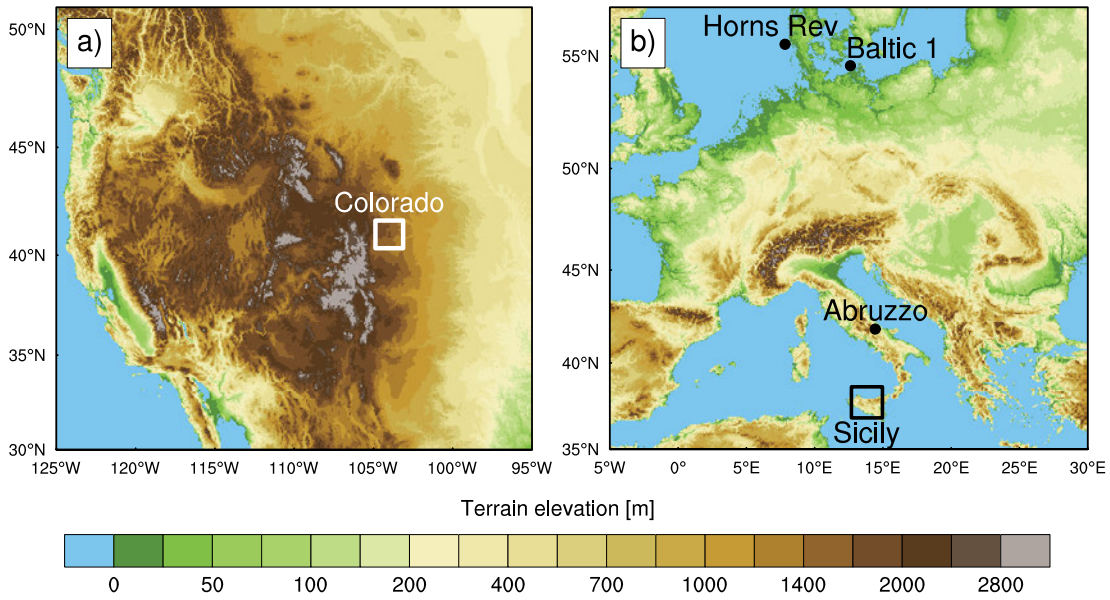

Figure 1: Geographical location and terrain elevation [m] of a) the Colorado wind farm and b) the Abruzzo, Sicily, Horns Rev, and Baltic 1 wind farms. The position of the Colorado and Sicily wind farms is indicated by a rectangle for confidentiality reasons.

et al. (2013) put it, "including multiple predictor variables that exhibit correlations to the predictand further helps distinguish the analogs by perhaps identifying specific weather regimes." However, Delle Monache et al. (2013) and Alessandrini et al. (2015) do not optimize the predictor selection and assign equal weights to the predictors, neglecting relationships between individual predictors and the variable to be predicted, the predictand. For instance, Timbal et al. (2008) and Cavazos and Hewitson (2005) have underlined the importance of identifying these relationships for multiple, physically meaningful predictors in statistical downscaling procedures.

Therefore, the main objective of this paper is to explore several predictor-weighting techniques with the goal to further improve the AnEn performance for wind power forecasting, by:

- Developing static and dynamic weighting strategies where the optimized combination of multiple predictor variables is found by minimizing a probabilistic score over all possible combinations of weight values. Compared to the static strategy, the dynamic strategy updates the optimized predictor weights each month to explore their seasonal dependency.
- Identifying the AnEn predictors that are relevant for probabilistic wind power forecasting. The weighting strategies are applied to a range of analog predictors including wind speed and direction at several heights in the atmospheric boundary layer, geopotential height at 850 hPa, 925 hPa, and 500 hPa, mean sea level pressure, 2-m temperature, and boundary layer height.
- Applying the weighting strategies after a principal component analysis (PCA) of the multiple predictor data set. This approach has the advantage that redundant information in the predictor data set are removed and that the dimensionality of the predictor data set is reduced.

We compare the weighting strategies at five wind farms in Europe and the U.S. situated in regions with different terrain complexity, both on and offshore, and over a period of two years or longer, to study site-dependent differences between the strategies. In Section 2 we provide a description of the available wind and power data at each wind farm and the deterministic numerical weather predictions. Section 3 presents the verification methods, while the AnEn method and the proposed weighting techniques are introduced in Section 4, and their performance is compared in Section 5. The main findings of this study are discussed in Section 6 and a conclusion is provided in Section 7.

2 Data

2.1 Wind farm observations

In this section, a description of the available data at each wind farm is provided. The geographic locations of the wind farms are shown in Figure 1. Because the data are considered sensitive with respect to commercial competition, we are unable to reveal specific locations of some of the wind farms below. Nevertheless, a rigorous evaluation of the weighting approaches is performed, and full details of the error characteristics relative to the measurements are shown in the reminder of this paper.

2.1.1 Offshore wind farms

The offshore wind farm EnBW *Baltic 1*, which started operation in May 2011, is located 16 km north of the

peninsula Darß-Zingst in the Baltic Sea. The wind farm consists of 21 Siemens wind turbines with a hub height of 67 m with a total installed capacity of 48.3 MW. Wind power and nacelle wind speed observations are available at each turbine of the wind farm with a resolution of 10 min for the 3-year period May 2011 to April 2014.

The wind farm *Horns Rev* is located in the North Sea 14 km west of Denmark and includes 80 turbines capable of producing 160 MW. The hub height of the turbines is 80 m and observations are collected for 10-min wind power and nacelle wind speed. They are available for each turbine during the 2-year period from February 2010 to December 2011.

2.1.2 Onshore wind farms

The wind farm located in Colorado in the U.S. Great Plains consists of several hundred of wind turbines. The terrain in the direct vicinity of the wind farm is hilly. Wind speed observations and wind power data at each turbine are available with a time resolution of 15 min for the 3-year period from February 2010 to December 2012.

The *Sicily* wind farm, which has been previously studied in Alessandrini et al. (2013) and Alessandrini et al. (2015), is situated in Northern Sicily, Italy, in a complex-terrain area. Heights in the area around the wind farm range between 400 m and 1800 m. The wind farm with a nominal power of 7.6 MW consists of nine Vestas turbines with 50-m hub height. Available data are wind power and wind speed observations at each turbine and 50-m wind speed at one measurement mast in the center of the wind farm. The data with a time resolution of 10 min are provided for the 2.5-year period from November 2011 to March 2013.

The *Abruzzo* wind farm has a nominal power of about 100 MW and is located in Central Italy in the Abruzzo region. Heights in the complex-terrain area around the wind farm range between 500 m and 1400 m. The turbines with hub heights between 46 m and 50 m are mainly situated on mountain ridges. Available data are hourly averages of farm-wide wind power and wind speed observations at four 10-m masts for the 2-year period from February 2010 to December 2011.

2.1.3 Data handling

For each wind farm except Abruzzo, quality control of the data is done by carefully comparing wind speed and power observations at each turbine, which should follow the typical shape of a turbine power curve. Apparent outliers from this power curve are removed from the data set, for example, data points of wind speed below cut-in but positive power and wind speed above cut-in and below cut-off but zero power. Curtailments, which we define as constant power values below rated power over several consecutive time steps but fluctuating wind speed observations, are also removed. Additionally, the datasets for Horns Rev and Baltic 1 are provided with turbine status signals and – in case of Horns Rev – also with quality flags, which are used to further clean the data.

The quality control filters approximately 10 % of the power values at Baltic 1, 8 % at Horns Rev and Sicily, and 3 % at Colorado. Finally, a normalized wind farm time series is achieved by calculating the mean power signal over all turbines in the wind farm, which have a quality controlled and valid signal at each time step, and by normalizing the time series with rated power of the turbines. The 10-min and 15-min values are averaged to hourly values to obtain the same time resolution of the observation data at each wind farm.

For the Abruzzo wind farm, the available farm-wide power time series do not allow to filter curtailment periods and down times at each turbine. There is no information available about the number of properly working turbines at each time step, and therefore the shutdown of single turbines due to icing or curtailment is not easily detectable. For this reason, the Abruzzo data are particularly challenging since the generation of skillful analogs depends on the quality of the power time series. However, to remove obvious non-coherent observations, the 10-m wind speed observations and aggregated power values are compared against each other. This procedure filters approximately 3 % of the power values.

Table 1: List of the ECMWF IFS meteorological variables used in this study.

ECMWF IFS variables		
Long name	Short name	Unit
mean sea level pressure	SLP	Pa
2-m temperature	2-m T	K
10-m wind direction	10-m WD	°
10-m wind speed	10-m WS	m/s
hub-height wind direction	hub WD	°
hub-height wind speed	hub WS	m/s
100-m wind direction	100-m WD	°
100-m wind speed	100-m WS	m/s
300-m wind direction	300-m WD	°
300-m wind speed	300-m WS	m/s
boundary layer height	BLH	m
925-hPa geopotential height	925-hPa GH	m
850-hPa geopotential height	850-hPa GH	m
500-hPa geopotential height	500-hPa GH	m

2.2 Deterministic forecasts

The AnEn predictor data set is composed of the high-resolution deterministic forecasts of the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) and considers 0000 UTC forecasts of meteorological variables every three hours (Table 1). Note that wind speed and direction at 10-m and 100-m height are diagnostic model output of the ECMWF IFS and that hub-height and 300-m wind speed and direction are vertically interpo-

Table 2: Terrain specification, training and test period at each wind farm.

Wind farm	Terrain	(Initial) training period	Test period
Baltic 1	offshore	05/2011–04/2013	05/2013–04/2014
Horns Rev	offshore	02/2010–12/2010	01/2011–12/2011
Colorado	hilly terrain	02/2010–12/2011	01/2012–12/2012
Abruzzo	complex terrain	02/2010–12/2010	01/2011–12/2011
Sicily	complex terrain	11/2010–02/2012	03/2012–02/2013

lated from neighboring model levels making use of pressure, temperature, and humidity profiles.

On the 26th January 2010, the horizontal resolution of the ECMWF deterministic forecast has been increased to a spectral truncation at wave number 1279, which corresponds to $0.125\,^\circ \times 0.125\,^\circ$ (Miller et al., 2010). Therefore, the horizontal resolution remains unchanged for the training and test periods employed in this study (Table 2). A horizontal bilinear interpolation of the forecasts to the geographic coordinates of each wind farm center is applied.

2.3 COSMO-DE analysis

To understand stability-dependent forecast errors at Baltic 1, data from the consortium for small-scale modeling (COSMO)-DE analysis are used. COSMO-DE is a high-resolution, convection-permitting configuration of the numerical weather prediction model COSMO and the 2.8-km analysis is available across Germany (Baldauf et al., 2011). Hourly potential temperatures from the COSMO-DE analysis are used at the surface and at model levels 47, 48, 49, and 50 to discuss the stability dependence of the forecast errors in Section 5.3. The heights at full model levels of the COSMO-DE analysis are 122.32 m (level 47), 73.03 m (level 48), 35.72 m (level 49), and 10 m (level 50) at Baltic 1.

3 Verification methods

To verify the probabilistic attributes of the ensemble forecasts, the joint distribution of ensemble forecasts and observations can be investigated (Murphy and Winkler, 1987). In this section, approaches for the evaluation of ensemble forecasts are introduced to facilitate the presentation of the results.

A reliability diagram includes the reliability curve and the sharpness histogram. The former evaluates the reliability (also called *calibration*) for an event threshold. A reliability curve plots the observed relative frequency of an event against the binned forecast probability for any given level of probability. An ensemble forecast is reliable if the reliability curve lies on the diagonal. Sharpness is an attribute of the forecast only, and is displayed in the sharpness histogram (relative frequency of use of each probability level).

The relative operating characteristics (ROC) diagram is a discrimination-based display of forecast verification. The ROC diagram does not evaluate the full joint distribution at an event threshold, but evaluates the ability of the ensemble forecast to discriminate between the occurrence and non-occurrence of the event. The ROC diagram is generated by plotting the false alarm rate (i.e., false alarms divided by the total of non-occurrences of the event) against the hit rate (i.e., the correct forecasts divided by total occurrences of the event). The ROC discrimination can be summarized using the area A under the ROC curve as a single scalar value, with $A = 1$ for a perfect forecast and $A = 0.5$ for the sample climatology (Wilks, 2011). The ROC skill score (ROCSS) is then defined as

$$\mathrm{ROCSS} = \frac{A - 0.5}{1 - 0.5} = 2A - 1. \tag{3.1}$$

The reliability and ROC diagram are both graphical displays of forecast verification for event thresholds. The continuous ranked probability score (CRPS), however, is computed across the entire variable range and can be seen as the integral of the Brier score over all possible threshold values (Hersbach, 2000). It is a proper scoring rule for the evaluation of ensemble forecasts and can be evaluated as

$$\mathrm{CRPS}(F_{\mathrm{ens}}, y) = \frac{1}{M}\sum_{m=1}^{M}|x_m - y| - \frac{1}{2M^2}\sum_{n=1}^{M}\sum_{m=1}^{M}|x_n - x_m|, \tag{3.2}$$

where F_{ens} is the ensemble forecast with ensemble members $x_1, \ldots, x_M \in \mathbb{R}$ and $y \in \mathbb{R}$ is the observation (i.e., verifying wind power value) (Gneiting and Raftery, 2007). Over N prediction/observation pairs, the CRPS values are given by

$$\mathrm{CRPS} = \frac{1}{N}\sum_{n=1}^{N}\mathrm{CRPS}(F_{ens,n}, y_n). \tag{3.3}$$

To assess the statistical consistency of the ensemble spread, the spread-skill relationship is analyzed by comparing the root-mean-square error (RMSE) of the ensemble mean to the square root of average ensemble variance (S) (Wilks, 2011; Fortin et al., 2014, among

others), which are defined as

$$\mathrm{RMSE} = \left(\frac{1}{N}\sum_{n=1}^{N}|\overline{x}_n - y_n|^2\right)^{\frac{1}{2}}, \tag{3.4}$$

$$\mathrm{S} = \left(\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{M-1}\sum_{m=1}^{M}(x_{m,n} - \overline{x}_n)^2\right)\right)^{\frac{1}{2}}, \tag{3.5}$$

where $\overline{x}$ is the ensemble mean. Note that the correction factor $M/(M+1)$, where M is the number of ensemble members, has to be applied to the RMSE when compared to S (Fortin et al., 2014). The interpretation of the spread-skill relationship relies on the assumption of a Gaussian distribution of ensemble mean errors and ensemble members. Since this assumption might not be met for wind power, the matching of spread and skill is only a necessary condition for statistical consistency. To provide an in-depth analysis of the spread-skill relationship, the so-called binned-spread/skill diagram is chosen, which compares S and RMSE over small class intervals of spread (Van den Dool, 1989; Alessandrini et al., 2015, among others).

To evaluate the deterministic skill of the ensemble mean forecasts, the RMSE of the ensemble mean is chosen (Eq. 3.4). To test the statistical significance of the scoring rules (RMSE, CRPS, ROCSS) used in this study, we calculate confidence intervals, generated with the bootstrap resampling technique (Efron, 1979; Bröcker and Smith, 2007; Pinson et al., 2010). We repeat the procedure 1,000 times and calculate 5 %–95 % confidence intervals from the resulting bootstrap. The consistency bars in the reliability curve are calculated with a quantile function for a binomial distribution (Bröcker and Smith, 2007; Pinson et al., 2010).

4 The analog ensemble (AnEn) method

4.1 Generalities

The AnEn method as proposed by Delle Monache et al. (2013) is a technique to generate an uncertainty forecast from a purely deterministic prediction. The uncertainty information is estimated using a set of M past verifying observations (i.e., wind power) that correspond to the M past forecasts (analogs), which are most similar to a current deterministic forecast. Since this approach directly uses the verifying observations as ensemble members, the AnEn method automatically accounts for observational errors in the verification. The multivariate metric used to estimate the degree of analogy between the current deterministic forecast and past predictions from a historical data set is defined in terms of a cost function (Delle Monache et al., 2011; Delle Monache et al., 2013):

$$\| F_t, A_{t'} \| = \sum_{i=1}^{N_v}\frac{w_i}{\sigma_i}\left(\sum_{j=-\tilde{t}}^{\tilde{t}}(F_{i,t+j} - A_{i,t'+j})^2\right)^{\frac{1}{2}}. \tag{4.1}$$

F_t is the current deterministic forecast valid at future time t and $A_{t'}$ an analog forecast with the same forecast lead time but valid at a time t' before F_t was issued; $\tilde{t}$ corresponds to half the number of additional forecast lead times around the future time t over which the metric is computed; $F_{i,t+j}$, $A_{i,t'+j}$ are the values of the current and past forecast, respectively, of the meteorological predictor i within the time window. The consideration of a time window is important since the metric accounts for the similarity of a temporal trend between a past and current deterministic forecast at a specific location. The N_v is the number of meteorological predictors used in the analog search and w_i the weight assigned to each predictor. Each meteorological predictor is normalized with its standard deviation σ_i over the training period of past forecasts to include predictors with different units. Circular meteorological variables such as wind direction are treated with circular statistics (Jammalamadaka and Sengupta, 2001).

4.2 Predictor-weighting strategies

Previous studies on AnEn do not optimize the predictor selection and assign equal weights $w_i = 100\,\%$ to each variable (Delle Monache et al., 2011; Delle Monache et al., 2013; Alessandrini et al., 2015; Vanvyve et al., 2015). Alessandrini et al. (2015) use wind speed and wind direction at 10-m model height as predictors for generating AnEn for wind power. By assigning $w_i = 100\,\%$, however, these studies neglect the strength of the relationships between individual predictors and the variable to be predicted. Furthermore, there might be more potential predictors for wind power predictability in terms of the AnEn method. For this reason, static and dynamic predictor-weighting strategies are proposed here to find optimized predictor weights w_i for a range of predictors. The usage of several and independent predictors as input to the predictor-weighting strategies is desirable to capture different sources of wind power predictability.

4.2.1 CRPS minimization

The static and dynamic weighting strategies have in common that the optimal predictor weights are found by minimizing the wind power CRPS over all possible combinations. To limit the number of possible combinations, the predictor weights w_i are restricted by

$$\sum_{i=1}^{N_v} w_i = 100\,\% \text{ and } w_i \in \{0, 10, \ldots, 100\}\%. \tag{4.2}$$

The reason for the latter restriction is that a finer granularity of w_i substantially increases the number of possible combinations and therefore computational costs. Just adding $w_i = 5\,\%$ as another possible weight for $N_v = 14$ increases the number of combinations by an order of magnitude.

Note that we implement the minimization as a brute force, given that each weight combination is tested by

computing the resulting performance in terms of the CRPS. This expensive calculation can be performed offline, and it allows to do a thorough exploration to find the optimal predictor weights that maximize AnEn performance with respect to CRPS over the training period. The possibility of carrying out the minimization with numerical optimization algorithms is discussed in Section 6.

4.2.2 Static weighting

The static predictor-weighting strategy optimizes the predictor combination by minimizing the wind power CRPS of AnEn over the entire training period of say 12–24 months. Thus, the optimization period and the training period for searching the analog dates are identical. For a given day of the training period, the analogs are searched over the entire training period excluding that particular day. The best combination is then applied to the test period that does not overlap with the training. Since the optimized predictor combination might be lead-time dependent, the following variants of the static-weighting strategy are considered: A lead-time independent strategy, where the CRPS is minimized over all forecast lead times together, and a lead-time dependent strategy, where the CRPS is minimized either for each lead time separately or for each time of the day. The latter groups forecast lead times that correspond to the same time of the day, e.g. lead times 3 h, 27 h, and 51 h of a 0000 UTC forecast correspond to 0300 UTC and are grouped together. Minimizing the CRPS for all lead times together has the advantage of a larger sample size available for the optimization procedure compared to optimizing the predictor combination for each lead time separately. The time-of-the-day dependent optimization might be a good compromise solution as the sample size increases when the lead times are grouped, while a possible daily cycle of predictor combinations is taken into account.

4.2.3 Dynamic weighting

Compared to the static strategy, the dynamic predictor-weighting approach updates the optimized predictor combination to explore the dependency of the weights on the season. The predictor combination is updated each month by optimizing the CRPS of the wind power AnEn only over the k-months (later called *optimization period*) that precede the 1-month test period. The analog dates are still searched over the entire training period, which is only initially the same compared to the static-weighting strategy. However, the CRPS minimization is done only for the optimization period. Lead-time dependent and independent variants of the dynamic predictor-weighting strategy are also considered.

4.2.4 Principal component weighting

The static and dynamic strategies optimize the weights for multiple predictors. These predictors are deterministic forecasts of meteorological variables from numerical weather predictions. If there is redundant information in the data set due to correlations among the predictor variables, principal component analysis (PCA) can be an appropriate statistical tool to reduce the dimensionality of the original predictor data set to fewer variables, which are linear combinations of the original ones (Jolliffe, 2005; Wilks, 2011, among others).

We apply PCA to the standardized predictor data set, which means that the mean of each predictor time series is subtracted and that the time series is divided by their respective standard deviation. Furthermore, wind speed and wind components are used instead of wind speed and wind direction predictors as input to the PCA. The new predictors or principal components are obtained by projecting the original predictors on each eigenvector.

To reduce the dimensionality of the original data set, the principal components need to be truncated. One subjective approach is based on the eigenvalue spectrum, which is the eigenvalue magnitude as a function of the principal component number (Wilks, 2011). The eigenvalue number where a slope separation of the spectrum occurs, is taken as the principal-component cutoff.

4.3 Implementation specifics

Alessandrini et al. (2015) carried out a sensitivity study of AnEn to different parameters in Eq. 4.1 at the Sicily wind farm. The best AnEn performance was achieved for $\tilde{t} = 1$ and an ensemble consisting of $M = 20$ members. The same configuration is selected here since it also yields the best AnEn performance in terms of the CRPS with the data set analyzed in this study (not shown). Note that $\tilde{t} = 1$ in this study corresponds to a time window of six hours since ECMWF 3-hourly forecasts are used, which means that similar temporal trends of the predictors over six hours are matched and not just the predictor values at one lead time.

5 Results

In Section 5.1 the selected predictors from each predictor-weighting strategy based on the optimization over the training period are presented, while in Sections 5.2 and 5.3 the performance of the 20-member AnEn associated with the different weightings is evaluated over the test period.

5.1 Optimized predictor weights

The predictor-weighting strategies are applied to the $N_v = 14$ predictors from the high-resolution ECMWF deterministic forecast (Table 1). Considering the weight restriction (Eq. 4.2), 14 predictors lead to 1,144,066 possible combinations for which the AnEn-based wind power predictions are generated.

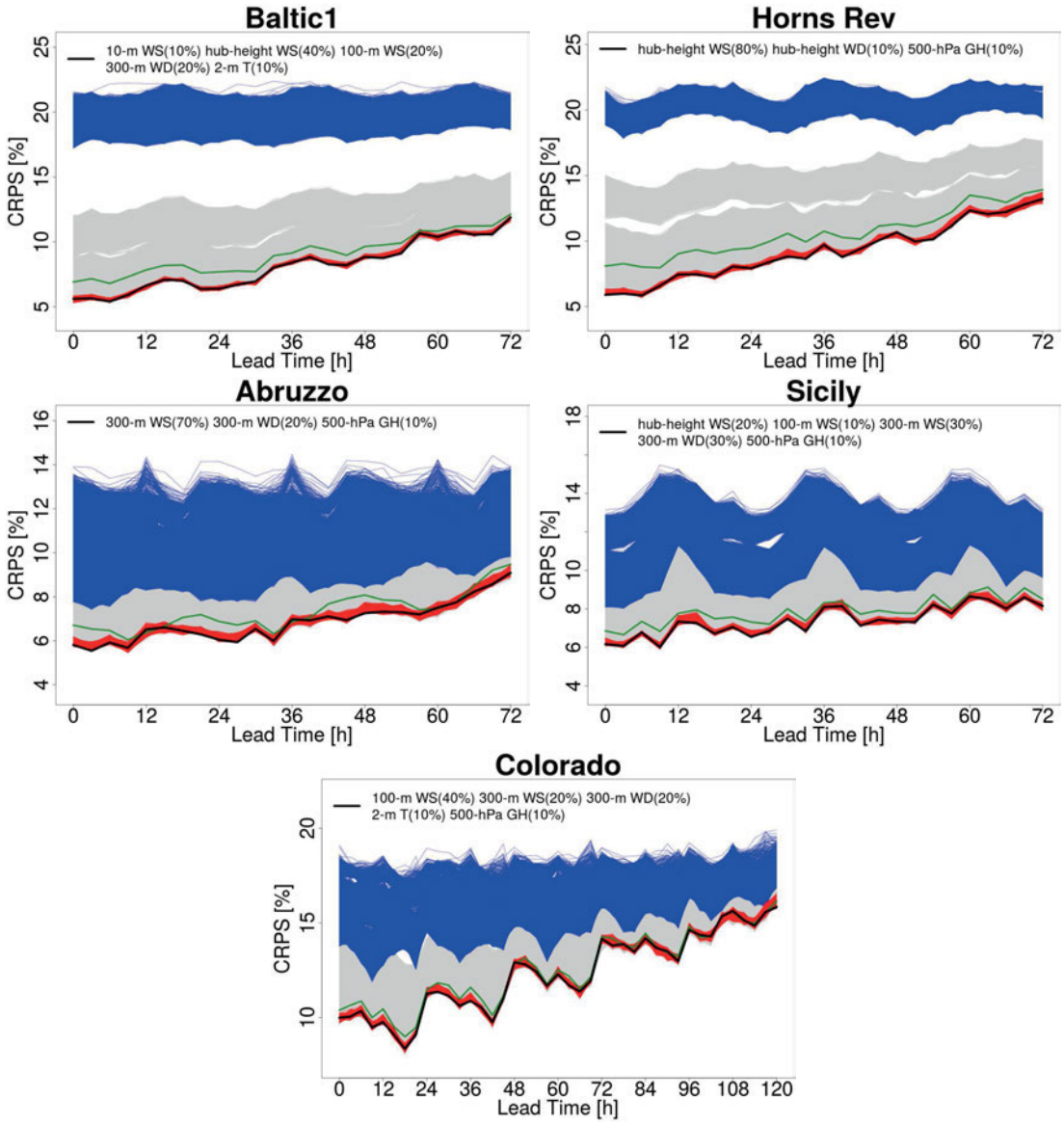

Figure 2: CRPS of normalized wind power [%] as a function of the forecast lead time [h] at Baltic 1, Horns Rev, Abruzzo, Sicily, and Colorado. The gray and blue lines indicate the CRPS of the 1,144,066 possible predictor combinations, the red lines the 5000 best predictor combinations, and the black line the best predictor combination of the lead-time independent static-weighting strategy. While the blue-colored lines show the CRPS of AnEn excluding wind speed at all heights as predictors, the gray lines show the CPRS of AnEn including wind speed predictors at all heights with at least 10 %. The green line additionally shows the CPRS of the AnEn based on all 14 predictors given the same weight (i.e., non-weighted). Each predictor combination is evaluated over the entire training period.

5.1.1 Static weighting

The results of the lead-time independent static predictor-weighting strategy (later called *static weighting*) are shown in Figure 2. As indicated by the 1,144,066 lines, the normalized wind power CRPS covers a wide range (i.e., 5–22 %) at Baltic 1 and Horns Rev, which emphasizes the need for optimizing the predictor combination. The solutions at those offshore wind farms are divided into two main regimes. The regime with the highest CRPS (blue-colored) occurs when wind speed predictors are excluded. Once the sum of the wind speed predictors (10 m, hub height, 100 m, and 300 m) is equal to 10 % or higher, the CRPS falls to lower values. Including wind speed as a predictor is therefore crucial for increasing the AnEn forecast performance. The clear separation of the AnEn solutions is not evident at Abruzzo, Sicily, and Colorado since probabilistic wind power predictability might be less dominated by wind speed predictors at these onshore sites.

As expected the best predictor combination is dominated by wind speed predictors at all sites (see top of each panel in Figure 2). Wind speed predictors in the proximity of the turbine swept area (10 m, hub height, 100 m) are selected at offshore wind farms. However, the wind power CRPS is barely degraded when the hub-height wind speed predictor is exchanged by 300-m wind speed since the correlation between wind fore-

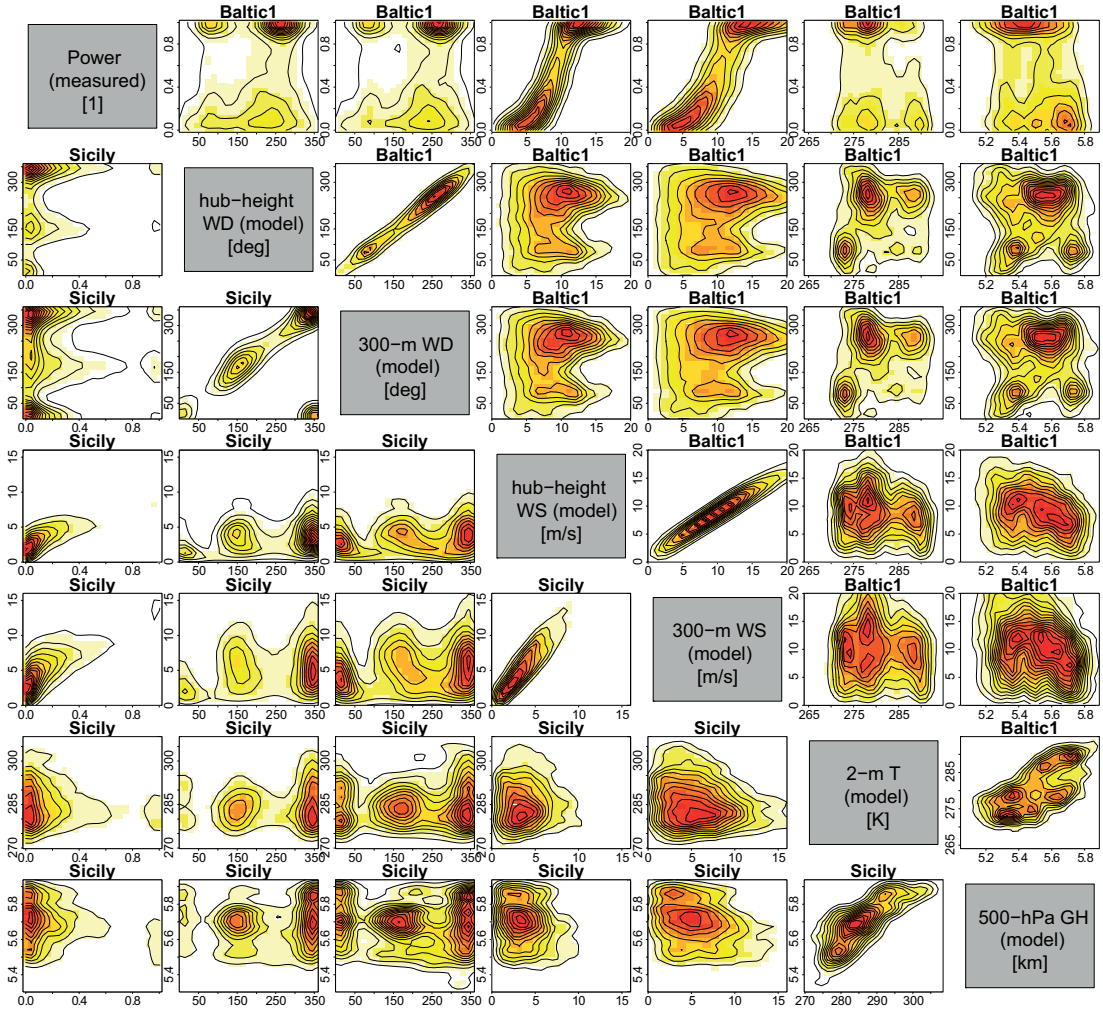

Figure 3: Contours of pairwise marginal densities between each couple of the following variables: wind speed (hub height, 300 m), wind direction (hub height, 300 m), 2-m temperature, and 500-hPa geopotential height (for intraday and day-ahead forecasts of the ECMWF IFS) as well as wind power (measured). The data are shown for Sicily (lower left plots, time period November 2010 to March 2012) and for Baltic 1 (upper right plots, time period November 2011 to March 2013). Red colors indicate high densities and yellow colors lower densities.

casts at both heights is high at the offshore sites (see marginal density plots for Baltic 1 in Figure 3). At the complex-terrain wind farms in Abruzzo and Sicily, predictors at 300 m dominate the static weights (Figure 2). Since 300-m wind speed forecasts resolve low and high wind power events by a larger wind speed range (see marginal density plots for Sicily in Figure 3), the metric can better distinguish the analog quality using 300-m wind speed forecasts.

In the best predictor combinations, wind direction is assigned 10–30 % weight. Wind direction might be important if observed wind power variability is strongly connected to certain wind direction regimes. The density plots at Sicily indicate that high wind power events occur for southerly and northerly wind direction forecasts (Figure 3), which could explain the assignment of 30 % weight to 300-m wind direction forecasts at this complex-terrain site. Other predictors such as geopotential height at 500 hPa are of less importance for wind power predictability, but still receive 10 % weight at Horns Rev and the onshore wind farms.

5.1.2 Dynamic weighting

Compared to the static strategy, the dynamic predictor-weighting approach updates the predictor combination each month and optimizes over a k-month optimization period instead of the entire training period. We tested the sensitivity of the dynamic strategy to the length of the k-month optimization period with $k \in \{1, 2, 3, 4\}$. Overall the lowest forecast errors are achieved with a 3-month optimization period (not shown) for which reason the dynamic weighting is presented with $k = 3$. Figure 4 illustrates the weights of the dynamic predictor-weighting strategy with 3-month optimization periods and with the lead-time independent approach (later called *dynamic weighting*).

For certain months, the predictor combinations differ substantially from the predictor combination of the

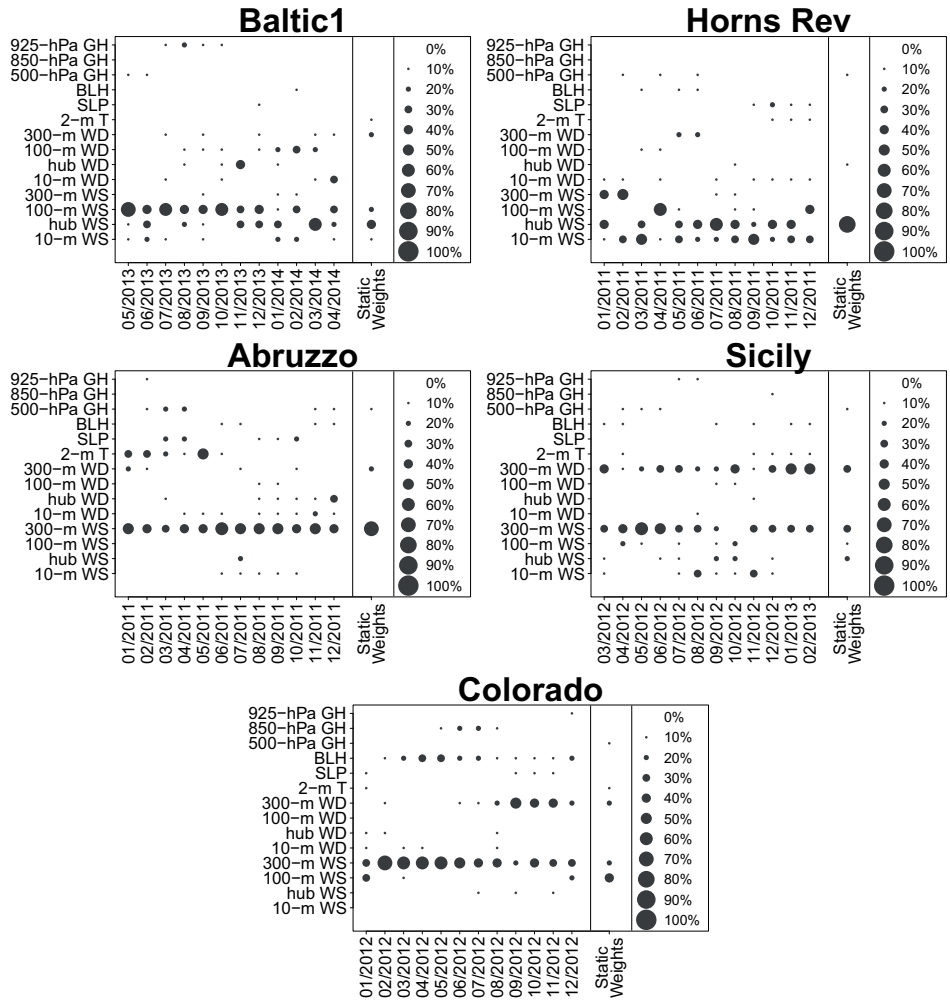

Figure 4: Optimized predictor combinations for the dynamic predictor-weighting strategy with sliding 3-month optimization periods and the lead-time independent approach. The optimized predictor combinations of the static strategy are shown for comparison. The higher the weight of a predictor, the larger the filled circle. The date on the abscissa indicates the month for which the optimization is carried out, for instance 05/2013 gives the information that the 3-month optimization period is 02/2013–04/2013 and that the optimized predictor combination is applied to 05/2013.

static strategy. For instance, at the Abruzzo wind farm, 2-m temperature is assigned up to 60 % weight in the first half of 2011, while assigned 0 % for the static-weighting strategy. Other examples are the dominance of the 300-m wind speed predictor in the beginning of 2012 or the 300-m wind direction predictor from September to November 2012 at the Colorado wind farm. We will evaluate in Section 5.2 and 5.3 how those and other differences between the static and dynamic predictor-weighting strategy impact the AnEn performance over the test period.

5.1.3 Principal component (PC) weighting

The weighting strategies are applied to principal components, given that the marginal density plots in Figure 3 indicate high correlations between certain predictors such as wind speed at hub height and 300 m. Since the PC cutoff occurs at PC number six as indicated by the eigenvalue spectrum (Figure 5), the PC static-weighting strategy is based on six PCs. The total variance explained by the leading six PCs is between 97–99 %, indicating that there is not an excessive information loss when truncating. Computational costs of the weighting optimization procedure are considerably lower when the number of predictors is reduced from $N_v = 14$ to $N_v = 6$ since the number of possible predictor combinations is decreased from 1,144,066 to 3,003. To optimize the static-weighting AnEn over 1,144,066 combinations at Horns Rev, our Fortran code runs about 55 hours on 12 cores (Intel Westmere-EP) with each 2.66 GHz. In comparison, the optimization over 3,003 combinations just requires about 0.16 hours. Note that the Fortran code used in this study is not optimized for performance, and the above computational costs could be significantly reduced with it. An alternative PC-weighting approach to brute-force could be to set the PC weight proportional to the variance explained by each component (not shown).

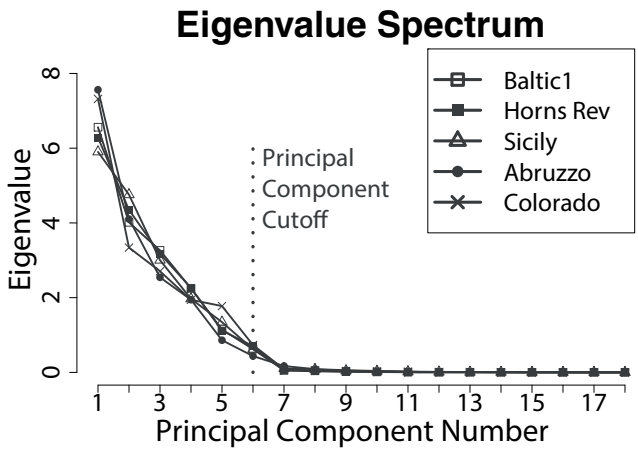

Figure 5: Eigenvalue magnitudes as a function of the principal component number (eigenvalue spectrum) for an 18-dimensional principal component analysis at all wind farm sites. The principal component number six, where a slope separation of the spectrum occurs, is taken as the principal component cutoff.

Table 3: Optimized weights [%] for the (lead-time independent) static predictor-weighting strategy based on six leading principal components at each wind farm.

	PC1	PC2	PC3	PC4	PC5	PC6
Baltic 1	50	10	10	20	0	10
Horns Rev	50	10	0	30	0	10
Abruzzo	40	30	10	0	10	10
Colorado	40	10	20	10	10	10
Sicily	20	40	20	10	0	10

The weights for each PC are optimized by applying the lead-time independent static-weighting strategy to the PC predictors (later called *PC static weighting*). The weights in Table 3 indicate that the leading PC is assigned 40–50 % weights at Baltic 1, Horns Rev, Abruzzo, and Colorado, but 20 % weight at Sicily where the second PC is assigned the largest weight. The latter can be explained by the second eigenvector that primarily points in the direction of meridional wind and in the direction of wind speed at all four heights (not shown). Thus, the variance explained by the meridional wind and wind speed is most important for wind power predictability at Sicily, which confirms the previous discussion on the importance of the wind direction predictor at that location.

5.1.4 Predictor importance

The previous analysis of the best predictor combination at each site led to a first insight into the importance of each predictor. However, many other predictor combinations have similar CRPS values compared to the best predictor combination (Figure 2, red lines). Thus, to refine the previous analysis and to rank the predictors according to their importance, predictor combinations with slightly higher CRPS values should be considered as well. This can be achieved by defining the importance Q_i of each predictor i:

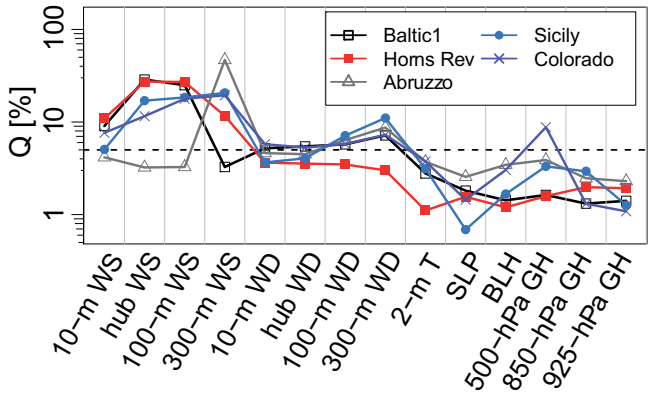

Figure 6: Importance Q [%] of each predictor for the static-weighting strategy at all wind farm sites plotted on a logarithmic y-axis. The horizontal dashed line indicates the 5 % value of Q.

$$Q_i = \frac{\sum_{k=1}^{K} w_{ik} e^{-\alpha k}[1 - \frac{C_k - C_B}{C_W - C_B}]}{\sum_{k=1}^{K} e^{-\alpha k}[1 - \frac{C_k - C_B}{C_W - C_B}]} = \frac{\sum_{k=1}^{K} w_{ik} e^{-\alpha k}(C_W - C_k)}{\sum_{k=1}^{K} e^{-\alpha k}(C_W - C_k)}, \quad (5.1)$$

where C_B is the CRPS of the best predictor combination, C_W the CRPS of the worst predictor combination, C_k the CRPS of the k-th possible combination, and $\sum_{i=1}^{N_v} Q_i = 1$. The weight w_{ik} is the weight of each predictor i for the given combination k. To emphasize predictor combinations with low CRPS values, we add the decay function $e^{-\alpha k}$ with an exponential decay constant α. To achieve an e-folding after the 5000 best predictor combinations, which are highlighted in Figure 2 for the static-weighting strategy, we set $\alpha = 2{\cdot}10^{-4}$ (i.e., $1/\alpha = 5000$).

The importance values of each predictor are shown in Figure 6. At all wind farms except Colorado, predictors such as temperature, pressure, boundary layer height, and geopotential height at different pressure levels have low importance values (below 5 %), while wind speed and direction predictors have considerably higher importance depending on the wind farm site. At Colorado, however, upper-level geopotential height at 500-hPa reaches an importance value of ~ 10 %. At the offshore wind farms, wind speed predictors are important at several heights. The 300-m wind speed predictor has clearly the highest importance (~ 40 %) at the Abruzzo wind farm. At Horns Rev, wind direction predictors are of low importance, while 100-m and 300-m wind direction predictors receive importance values clearly above 5 % at the remaining wind farms. At Sicily, wind direction at 300-m is of highest importance, which is in line with the analysis of the best predictor combination.

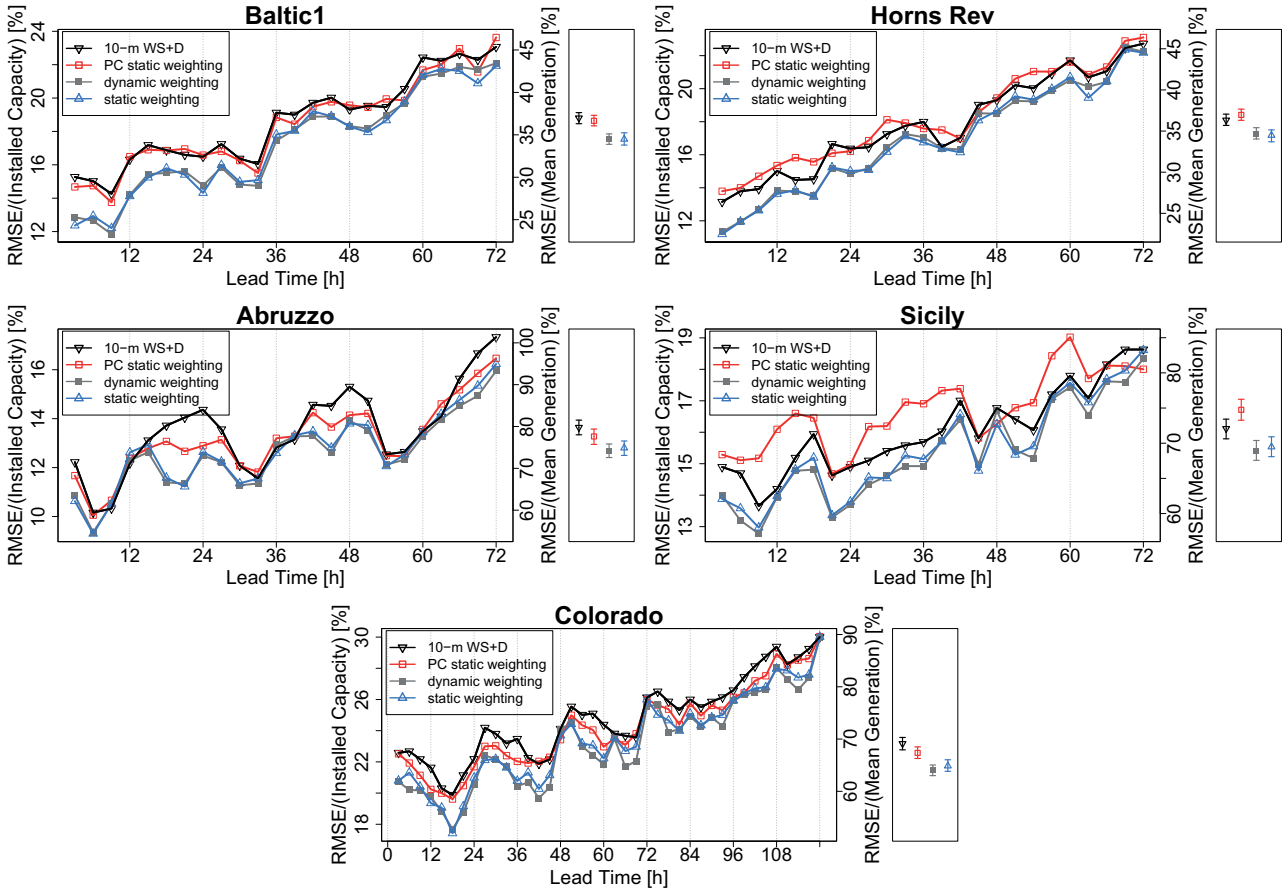

Figure 7: RMSE [%] of the ensemble mean of wind power normalized with either the installed capacity (left ordinate) or the mean generation (right ordinate) as a function of the forecast horizon for the test period at each site. Different weighting strategies have been applied (see legend box). The same plot but calculated for all forecast lead times is shown to the right of each figure. The 95 % bootstrap confidence intervals are indicated by the errors bars.

5.2 Verification of the ensemble mean

The 20-member AnEn with different weighting strategies are first compared in terms of the RMSE over the test period to evaluate the deterministic skill of the ensemble mean. The RMSE is normalized with both the installed capacity and the mean generation at each wind farm over the test period (see left and right ordinates in Figure 7). The comparison of static and dynamic weighting is carried out against the AnEn generated with 50 % weight on both wind speed and direction (WS+D) at 10-m height (later called *10-m WS+D*). The latter predictor combination was applied in Alessandrini et al. (2015), which is the only study where the AnEn formulation adopted here has been tested for wind power predictions, and as such is used as a baseline performance.

The RMSE normalization with mean generation allows a comparison of forecast errors between wind farms. The intraday (3–21 h) and day-ahead (24–45 h) RMSE is between 25–40 % at offshore sites and 55–100 % at onshore sites (Figure 7). The RMSE is considerably higher for the onshore sites where diurnal and terrain effects reduce wind power predictability when compared to offshore sites. This is underlined by the pronounced diurnal cycle of the RMSE at Abruzzo, Sicily, and Colorado.

In addition to the lead-time dependent RMSE, the RMSE is calculated for all forecast lead times together with 90 % bootstrap confidence intervals to test the statistical significance of differences between the weighting strategies (Figure 7). The RMSE values of the static and dynamic weighting AnEn are significantly lower than the RMSE of the 10-m WS+D AnEn at all wind farms considered in this study. The confidence intervals indicate that the PC static-weighting has an overall similar RMSE compared to the 10-m WS+D and a higher RMSE than the static and dynamic weighting based on the original ECMWF predictors. We also applied the dynamic-weighting strategy to the PCs as well as an equal weighting of each PC to generate the AnEn. Both strategies yield higher RMSE values than the presented PC static-weighting strategy (not shown).

At offshore wind farms, the static and dynamic weighting update strategies improve over the 10-m WS+D weighting for all lead times, while improvements are strongly lead-time dependent at Abruzzo. Largest improvements occur during evening and nighttime hours, while there are almost no improvements dur-

ing noon and afternoon, which may suggest that near-surface wind predictors exhibit a similar wind power predictability as the 300-m predictor during situations where the atmospheric boundary layer is well-mixed at this complex-terrain site.

Note that we present the weighting strategies with a lead-time independent approach, i.e., the predictor combination is optimized for all forecast lead times together. Although an optimization for each lead time separately or time of the day appears reasonable, the lead-time dependent strategy yields similar or higher forecast errors for both deterministic and probabilistic scores over the test period (not shown). This might be explained by the reduced sample size available for the optimization procedure, which decreases the robustness of the weighting strategy.

Furthermore, the forecast verification in this study is done over the test period, which does not overlap with the training period. This aims to simulate a real-time situation, when the observations are not available for the period covered by the forecast. It is also interesting to derive the optimized predictor combination over the test period to detect the highest possible AnEn performance. For the static weighting AnEn, the approach with identical training and test periods slightly outperforms the approach with independent training and test periods (not shown). Since improvements are statistically not significant, the real-time optimization is effective in approaching the best achievable weight configuration.

5.3 Probabilistic Verification

5.3.1 CRPS

Figure 8 shows the relative improvement of the CRPS of the static and dynamic weighting AnEn over the 10-m WS+D AnEn. The improvements are statistically significant at all wind farms and up to 20 % for intraday forecast horizons at offshore sites, but gradually decrease towards larger forecast horizons. At the Abruzzo wind farm, improvements in the CRPS have a strong diurnal cycle similar to the ensemble mean RMSE.

The static and dynamic weighting are again superior to the PC static weighting at all wind farms. An analysis of first and second eigenvectors at each site indicates that the eigenvectors point in the direction of wind speed and wind components at all heights at the same time. The forecast skill at Abruzzo, however, emphasizes the particular importance of the 300-m wind speed predictor. The PCA removes the ability to differentiate between predictors from certain heights that could explain the lower forecast skill of the PC static weighting and that would justify the computationally expensive brute-force approach based on the original ECMWF predictors.

The static and dynamic weighting have overall the same forecast skill. At the Colorado wind farm, however, the dynamic-weighting strategy is superior to the static-weighting strategy at certain lead times. As evident from Figure 4, the dynamic-weighting strategy puts more emphasis on the 300-m wind predictors for certain months that might be advantageous in terms of wind power predictability. However, differences between both weighting strategies are statistically not significant as indicated by the confidence intervals.

The similarity of the skill of the static and dynamic weighting AnEn raises the question if the selection of appropriate predictors is most important, while the weights applied to each predictor are less important. To answer this quesion, we generate an analog ensemble based on the predictors (that have at least 10 % weight) found with the static-weighting strategy at each site, but instead of applying the optimized weights the same weight is given to each predictor (i.e., non-weighted). We refer to this approach as *static predictors (non-weighted)*. This approach performs worse than the static-weighting strategy at all sites (Figure 8), with particularly low skill at Horns Rev. The latter might be explained by considerably increasing the weight of the 500-hPa geopotential height predictor in the non-weighted approach compared to the 10 % weight in static-weighting approach.

To analyze how much further the AnEn performance decreases when all available predictors are equally weighted, we generate a *14-predictors (non-weighted) AnEn*. Its performance, which is shown in Figure 2 for the training period and in Figure 8 for the test period, is lower than the static-predictors (non-weighted) AnEn at all wind farm sites. Thus, the comparison of 14-predictors (non-weighted) AnEn, static-predictors (non-weighted) AnEn and static-weighting AnEn emphasizes that both the selection of relevant predictors and the predictor weighting are important to improve AnEn skill.

Although the overall forecast skill of the static and dynamic weighting AnEn are similar, we want to analyze existing differences to highlight the potential of the dynamic-weighting strategy in terms of wind power predictability. Figure 9 shows the CRPS for each month, but for intraday and day-ahead lead times together, which are most important in terms of wind power trading on the energy market. At Baltic 1, forecast improvements with the static and dynamic weighting strategies are 3–4 % in January 2014 and 20–21 % in February 2014. An analysis of the potential temperature profiles provides an explanation (Figure 10). While potential temperature differences between the surface and upper model levels of the COSMO-DE analysis are on average negative in January 2014, they are clearly positive in February 2014, which implies stable stratification. Since stably-stratified atmospheric boundary layers are characterized by reduced vertical transport, the 10-m near-surface wind predictors are decoupled from hub-height winds, and therefore less relevant for the predictability of wind power during stable stratification.

At the Abruzzo wind farm, the dynamic-weighting strategy puts considerable weight to the 2-m predictor

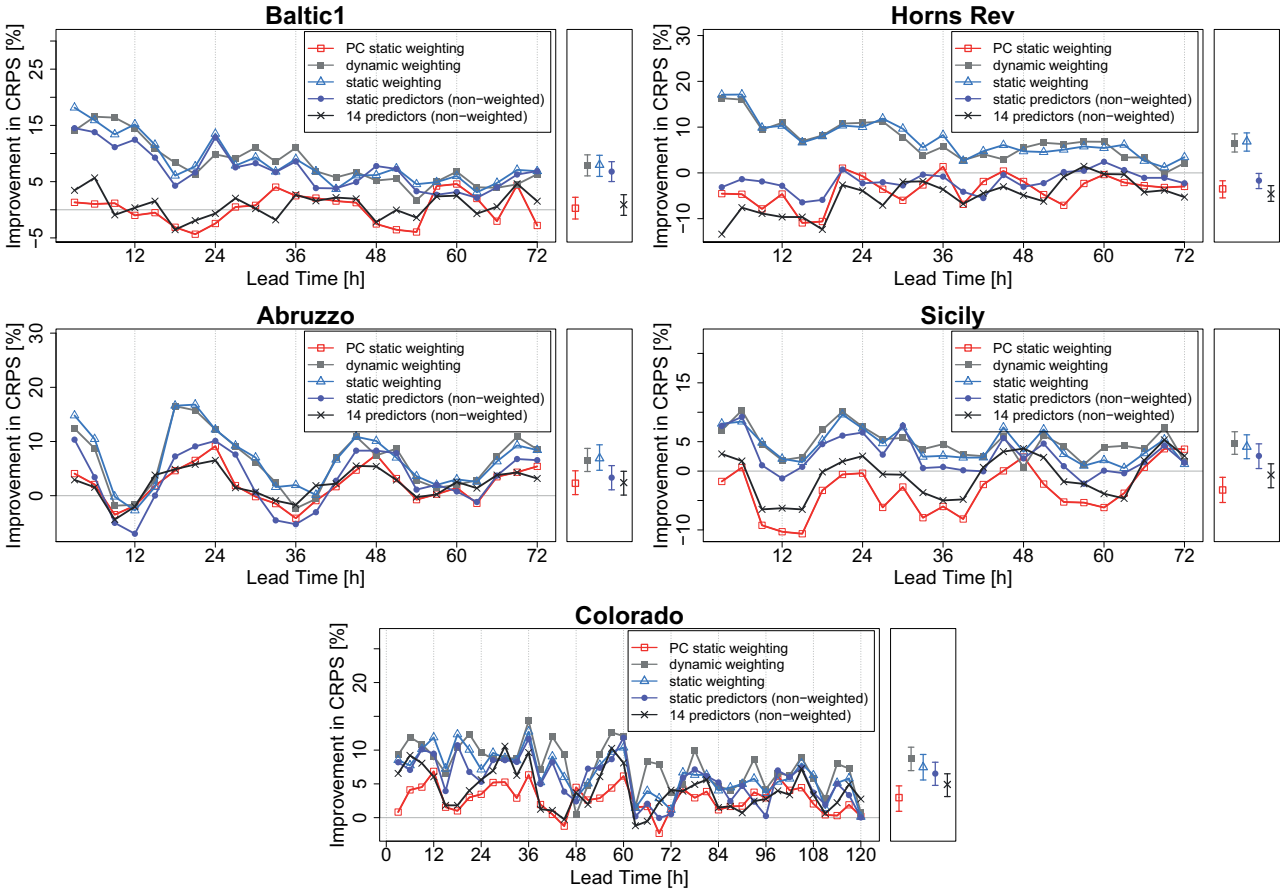

Figure 8: Improvement in the CRPS [%] as a function of the forecast horizon for the test period at each site for different weighting strategies with respect to the use of 10-m wind speed and direction as predictors. The same plot but calculated over all forecast lead times is shown to the right of each figure. The 90 % bootstrap confidence intervals are indicated by the errors bars.

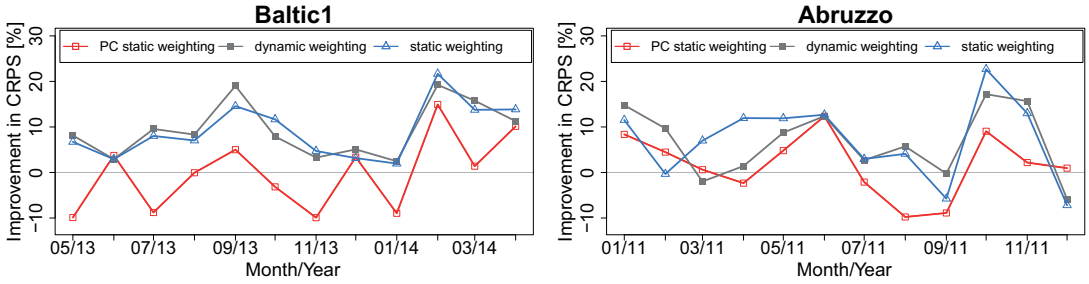

Figure 9: As in Figure 8, but as a function of time at Baltic 1 (left) and Abruzzo (right) and for intraday and day-ahead lead times grouped together.

early in 2011 (Figure 4). This appears to have a positive effect on forecast skill compared to the static-weighting strategy in January and February 2011 (Figure 9). In March and April 2011, the static weighting AnEn has a higher forecast skill. During this period, the dynamic-weighting AnEn puts 20 % weights to sea level pressure and 500-hPa geopotential height, which leads to lower performance.

5.3.2 Binned-spread/skill diagrams

To evaluate the statistical consistency of the ensemble spread, Figure 11 shows binned-spread/skill diagrams. If the binned-spread/skill diagram is on the 1:1 diagonal, the binned spread matches the RMSE, which is a necessary condition for statistical consistency. The analysis of the binned-spread skill diagrams of Horns Rev, Baltic 1, and Colorado as well as of Sicily and Abruzzo lead to very similar conclusions. Thus, for brevity reasons we select Baltic 1 and Sicily as representative sites. For similar reasons, we only show the reliability and ROC diagrams of Baltic 1 and Sicily. Furthermore, intraday and day-ahead forecast horizons are jointly considered.

The static and dynamic weighting AnEn and the 10-m WS+D AnEn exhibit a good statistical con-

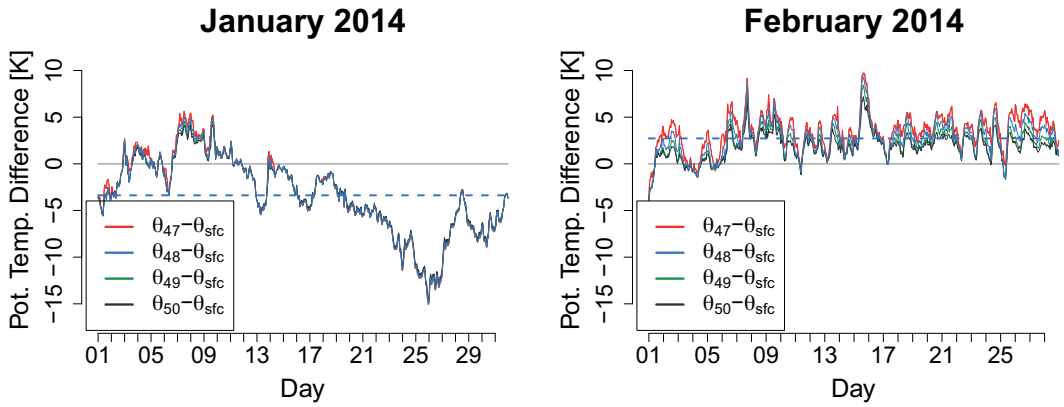

Figure 10: Hourly potential temperature differences [K] between model levels (47, 48, 49, 50) and the surface of COSMO-DE at the wind farm Baltic 1 for January 2014 (left) and February 2014 (right). The dashed line indicates the mean potential temperature difference between level 48 and the surface.

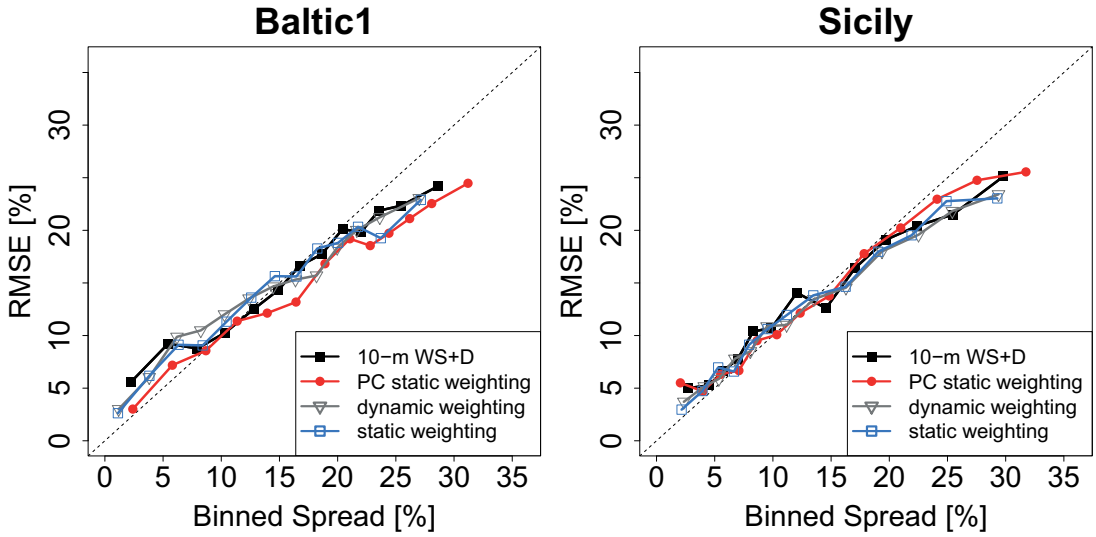

Figure 11: Binned ensemble spread [%] versus RMSE of the ensemble mean [%] of normalized wind power for intraday and day-ahead forecasts at Baltic 1 (left) and Sicily (right).

sistency particularly for mean spread values between 10–20 % since the spread-skill curve is close to the 1:1 diagonal. For the high-spread class, however, the AnEn's of all weighting-strategies are overdispersive. A closer analysis indicates that high wind power spread and relatively low RMSE (i.e., overdispersion) in this high-spread class mainly occurs for observed wind power events in the steep part of the power curve. In this part of the power curve, slight deviations from the observed weather regime in the analog selection process can easily lead to an overestimate of the variance of the 20 ensemble members.

At Baltic 1, the spread-skill relationship of the low-spread class indicates an underdispersive static and dynamic weighting AnEn and 10-m WS+D AnEn. This situation mainly occurs if all ensemble members predict either zero wind power or rated wind power while the observation slightly differs from that. Although the spread of the PC static weighting AnEn is statistically consistent for the low-spread class at Baltic 1, it is characterized by a stronger overdispersion for the other classes.

The binned-spread/skill diagrams of the static and dynamic weighting AnEn are both shifted towards lower RMSE and spread values compared to the PC static weighting AnEn and 10-m WS+D AnEn. This implies lower forecast errors of the static and dynamic weighting AnEn but still statistical consistency since the spread is reduced at the same time. Thus, a shift of the binned-spread/skill diagram towards the lower left is preferable.

5.3.3 Reliability and ROC diagrams

To assess reliability, sharpness, and discrimination, we assess the different weighting strategies in terms of the reliability and ROC diagram for certain event thresholds (Figures 12–13). The 50th and 90th percentile of observed wind power over the test period is chosen as event thresholds.

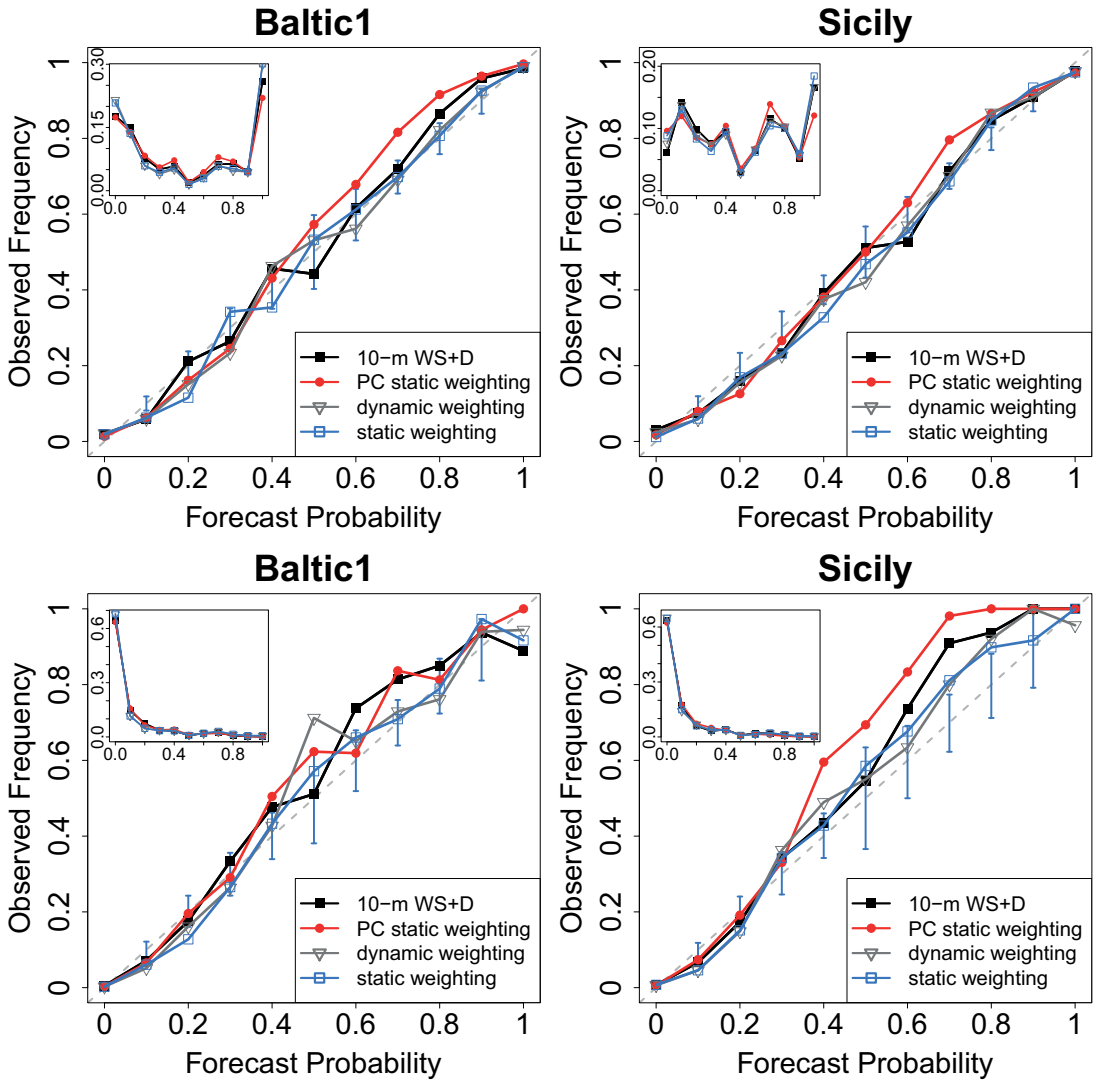

Figure 12: Reliability diagram and sharpness histogram for intraday and day-ahead forecasts at Baltic 1 (left) and Sicily (right). Results are shown for events larger than the 50th percentile (top row) and 90th percentile (bottom row) of observed wind power. The sharpness histogram displays the relative frequency of events in each forecast probability bin. The vertical bars of the static-weighting AnEn represent 90 % consistency bars that have been calculated with a quantile function for a binomial distribution.

The reliability curve and sharpness histogram of the static and dynamic weighting indicate a similar reliability and sharpness at both thresholds and both sites (Figure 12). At Baltic 1, the static and dynamic weighting AnEn of the 50th percentile threshold are more reliable for forecast probabilities >0.5 compared to the 10-m WS+D and particularly the PC static weighting. Furthermore, the sharpness of the static and dynamic weighting AnEn is higher since the lowest and highest forecast probabilities are more populated. The higher sharpness at the 50th percentile threshold is in agreement with the previous finding that the binned-spread/skill diagrams of the static and dynamic weighting AnEn are shifted towards lower spread values.

The PC static weighting AnEn appears to be the least reliable with a tendency to underconfidence for high forecast probabilities. This is particularly clear for the high threshold at Sicily. The sharpness of the AnEn's at the 90th percentile threshold is fairly similar for all weighting strategies although there is a tendency of sharper static and dynamic weighting AnEn's.

The ROC diagram and ROC skill score (ROCSS) evaluate the ability of the AnEn to discriminate between the occurrence and non-occurrence of an event (Figure 13). The higher the ROCSS is, the better its discrimination ability. We also provide the 90 % confidence intervals of the ROCSS to evaluate the statistical significance of differences between the predictor-weighting approaches.

The ROCSS of the static and dynamic weighting AnEn are higher compared to the 10-m WS+D and PC static weighting AnEn at all sites and for both thresholds. However, the confidence intervals indicate that the differences are statistically not significant except for the 90th percentile threshold at Baltic 1 where the curvature of the static and dynamic weighting ROC curve is clearly higher.

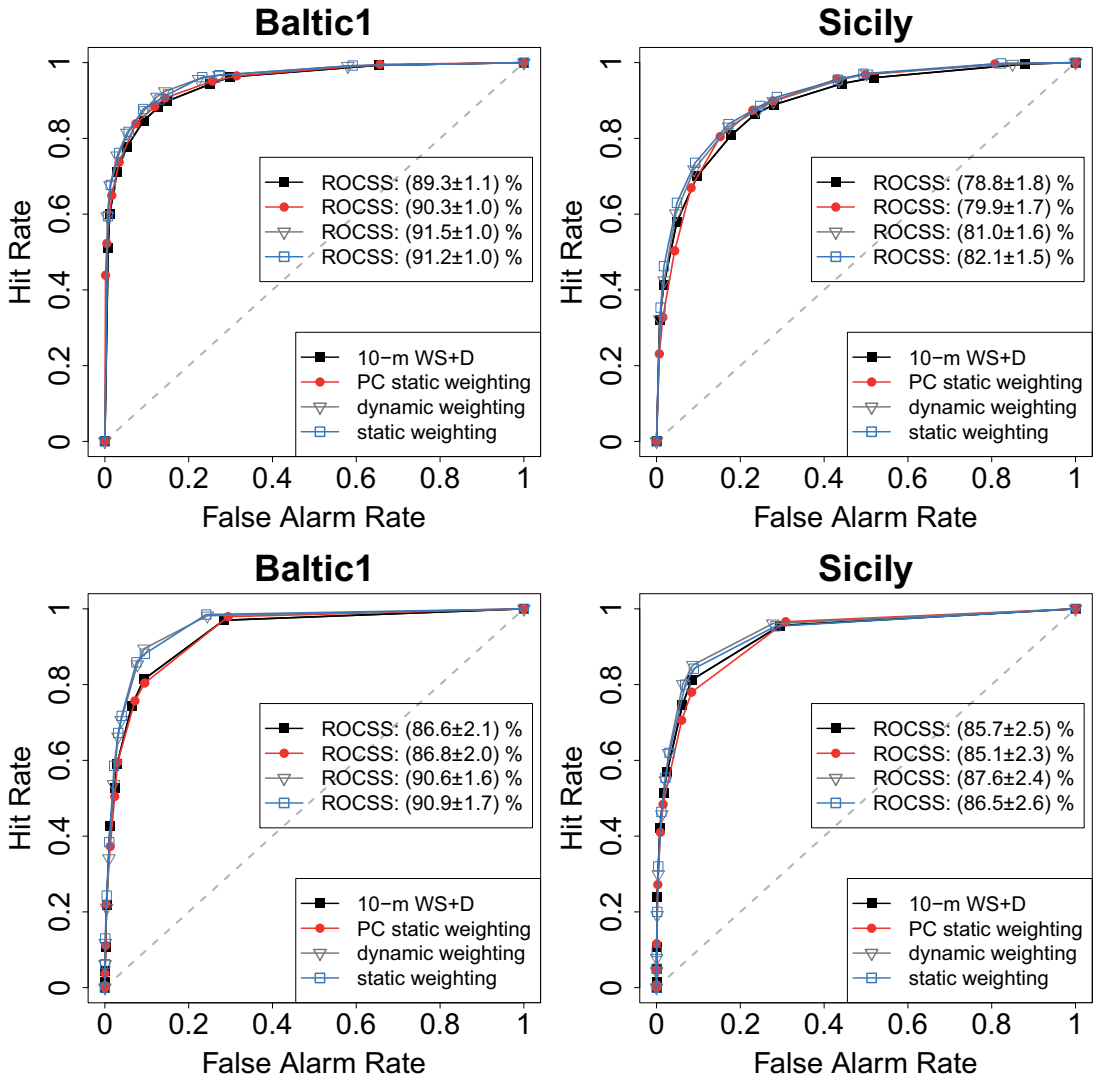

Figure 13: ROC diagram for intraday and day-ahead forecasts at Baltic 1 (left) and Sicily (right). Results are shown for events larger than the 50th percentile (top row) and 90th percentile (bottom row) of observed wind power over the test period. The area under the ROC curve relative to a climatological forecast is the ROC skill score (ROCSS). The confidence range of the ROCSS are 90 % bootstrap intervals.

6 Discussion

In the previous section, we have shown that the static and dynamic predictor-weighting strategies increase the AnEn performance over the 10-m WS+D AnEn significantly. The optimized predictor combination strongly depends on influencing factors such as terrain complexity and atmospheric stratification. Assuming that transmission system operators or wind power traders require ensemble forecasts for several wind farms within a portfolio, attention should thus be turned to optimize the predictor combination in case the analog ensemble method is applied. The strong dependence of improvements on for example the terrain complexity also justifies optimizing the predictor combinations with the brute-force approach at each wind farm. Future studies, however, could develop optimization procedures based on numerical optimization algorithms, which may be appropriate for optimizing the predictor combination of the analog ensemble, but are more efficient than the brute-force approach. The CRPS minimization with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm as proposed for the ensemble model output statistics by Gneiting et al. (2005) could serve as a basis. Some preliminary work on CRPS minimization based on the BFGS algorithm indicated that the shape of the surface on which the minimum is searched, is characterized by very low gradients and several local minima, which are different from the absolute minimum found with the brute force.

To decrease the computational costs of the brute-force CRPS minimization by reducing the dimension of the multivariate predictor data set, we applied the weighting strategies to the leading principal components instead of the original predictor data set. The AnEn performance significantly decreased with this approach at all wind farms. One reason might be that PCA transforms the dataset into a new variable space spanned by

leading eigenvectors, which removes the important ability to differentiate between predictors for example from certain heights. An alternative approach, which reduces the dimension of the dataset, but avoids data transformation, might be to find a predictor subset from all possible subsets based on the Bayesian information criterium (Schwarz, 1978) with forward selection, backward elimination, or stepwise selection. However, PCA could be an appropriate statistical tool for regional wind power forecasting with the analog ensemble. In that context, PCA was recently applied to large-scale sea level pressure fields from reanalyses with a coarse horizontal resolution of $2.5\,^\circ \times 2.5\,^\circ$ to estimate regional wind power output with an analog ensemble (Martín et al., 2014).

As shown in the previous Section, the dynamic-weighting strategy does not outperform the static strategy although it is designed to continuously adapt the predictor combination to seasonal changes. One explanation might be the statistical variability, which increases as the optimization period is reduced (i.e., less available samples) compared to the static strategy. The problem of statistical variability could be reduced if several years of wind power observations and predictions were available. In that case the optimization period could correspond for example to the same month as the test period, but taken from all previous years. A long training history, however, brings along the problem of possibly several changes in the numerical weather prediction (NWP) model version leading to changing forecast error characteristics that prevents the generation of skillful analogs. The usage of reforecasts, which are retrospective weather forecasts generated with a fixed numerical NWP model, could overcome this problem (Hamill and Whitaker, 2006; Hamill et al., 2006; Hagedorn et al., 2008, among others).

In future studies, the increase of the AnEn forecast performance by optimizing the predictor combinations with weighting strategies should be compared with other state-of-the-science approaches for estimating wind power forecast uncertainty. Alessandrini et al. (2015) took a first step in this direction by comparing the 10-m WS+D AnEn against reference wind power ensembles at the Sicily wind farm where the reference forecasts are based on quantile regression and postprocessing (calibration) of wind predictions from NWP-based ensembles. An in-depth extension of such a comparison of statistical and dynamical approaches would lead to valuable insights about the strengths and weaknesses of the differing approaches depending on the complexity of the terrain, lead time, and atmospheric stratification. In this context, statistical approaches could be the analog ensemble with predictor-weighting strategies and probabilistic wind power forecasts generated with for example quantile regression (Nielsen et al., 2006). In dynamical approaches, first an existing NWP-based wind ensemble is calibrated with state-of-the-science post-processing methods (Thorarinsdottir and Gneiting, 2010; Pinson, 2012; Junk et al., 2014), and the calibrated ensemble is transformed to wind power by means of a power curve.

7 Conclusions

In this study, we implemented and tested predictor-weighting strategies with the goal to improve the analog ensemble (AnEn) performance for wind power forecasting at on and offshore wind farms. The optimized combination of multiple predictor variables from the high-resolution ECMWF deterministic forecast are found by a brute-force CRPS minimization over all possible predictor combinations given discrete weight values. The static and dynamic weighting strategies increase deterministic and probabilistic AnEn performance up to 20 % compared to the AnEn with 10-m wind speed and direction or principal components as predictors at both on and offshore wind farms. They can accomplish both the selection of relevant predictors as well as finding their optimal weights, and also provide reliable estimates of the forecast uncertainty.

The static and dynamic weighting strategies lead to strongly site-dependent predictor combinations. For complex-terrain sites, the terrain is not well represented by the numerical weather prediction model and wind speed predictors above the rotor swept area in 300 m are important to increase wind power predictability. The optimized predictor combination at offshore wind farms is less sensitive to the height of the wind speed predictors. In fact, in neutral stability conditions as those commonly found above the sea, near-surface and hub-height wind predictors are well correlated and lead to similar AnEn performance. For stably-stratified atmospheric boundary layers, however, near-surface wind predictors are more often decoupled from hub-height winds and decrease the AnEn performance considerably. Wind direction predictors are particularly important for sites where wind regimes cause a wind direction dependency of generated wind power. Predictors such as temperature, pressure, and geopotential height are of less importance for wind power predictability compared to wind speed and direction predictors.

Compared to the static strategy, the dynamic-weighting strategy continuously updates the weights to consider the seasonal dependency of the weights. This strategy, however, does not improve the AnEn performance over the static strategy. Furthermore, an alternative strategy is developed, where principal component analysis reduces the dimension of the multiple ECMWF predictor data set and therefore decreases computational costs of the brute-force approach. The principal-component weighting strategy has significantly lower forecast skill, and therefore the computationally more demanding brute-force optimization approach is the preferred choice.

Acknowledgments

The work presented in this study has been funded by the research project Baltic I (FKZ 0325215A, Federal

Ministry for Economic Affairs and Energy) and the Ministry for Education, Science and Culture of Lower Saxony. We thank EnBW for providing the Baltic 1 data, the Public Service Company of Colorado for data from Colorado, DONG Energy, Vattenfall and DTU Wind Energy for data from Horns Rev, Enel for data from the Sicily wind farm, and Edison for the Abruzzo data. Numerical weather prediction data has been provided by ECMWF and the COSMO-DE analysis by the German Meteorological Service. The topography data are used by permission of D.T. Sandwell, W.H.F. Smith, and J.J. Becker (copyright 2008, The Regents of the University of California. All Rights Reserved). This paper has been improved by valuable comments and suggestions of Martin Kühn, Detlev Heinemann and Martin Dörenkämper (ForWind); and Jakob Messner (University of Innsbruck). Furthermore, we thank the reviewers for their valuable comments and suggestions.

References

Alessandrini, S., S. Sperati, P. Pinson, 2013: A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data. – Appl. Energ. **107**, 271–280, DOI: 10.1016/j.apenergy.2013.02.041.

Alessandrini, S., F. Davò, S. Sperati, M. Benini, L. Delle Monache, 2014: Comparison of the economic impact of different wind power forecast systems for producers. – Adv. Sci. Res. **11**, 49–53, DOI: 10.5194/asr-11-49-2014.

Alessandrini, S., L. Delle Monache, S. Sperati, J. Nissen, 2015: A novel application of an analog ensemble for short-term wind power forecasting. – Renew. Energ. **76**, 768–781, DOI: 10.1016/j.renene.2014.11.061.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. – Mon. Wea. Rev. **139**, 3887–3905, DOI: 10.1175/MWR-D-10-05013.1.

Bröcker, J., L. Smith, 2007: Increasing the reliability of reliability diagrams. – Wea. Forecasting **22**, 651–661, DOI: 10.1175/WAF993.1.

Cavazos, T., B.C. Hewitson, 2005: Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation. – Climate Res **28**, 95–107, DOI: 10.3354/cr028095.

Delle Monache, L., T. Nipen, Y. Liu, G. Roux, R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. – Mon. Wea. Rev. **139**, 3554–3570, DOI: 10.1175/2011MWR3653.1.

Delle Monache, L., F.A. Eckel, D.L. Rife, B. Nagarajan, K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. – Mon. Wea. Rev. **141**, 3498–3516, DOI: 10.1175/MWR-D-12-00281.1.

Efron, B., 1979: Bootstrap methods: another look at the jackknife. – Ann. Stat. **7**, 1–26, DOI: 10.1214/aos/1176344552.

Fortin, V., M. Abaza, F. Anctil, R. Turcotte, 2014: Why should ensemble spread match the rmse of the ensemble mean?. – J. Hydrometeorol. **15**, 1708–1713, DOI: 10.1175/JHM-D-14-0008.1.

Gneiting, T., A.E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. – J. Am. Stat. Assoc. **102**, 359–378, DOI: 10.1198/016214506000001437.

Gneiting, T., A. Raftery, A. Westveld, T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. – Mon. Wea. Rev. **133**, 1098–1118, DOI: 10.1175/MWR2904.1.

Hagedorn, R., T. Hamill, J. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. part I: Two-meter temperatures. – Mon. Wea. Rev. **136**, 2608–2619, DOI: 10.1175/2007MWR2410.1.

Hamill, T., J. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. – Mon. Wea. Rev. **134**, 3209–3229, DOI: 10.1175/MWR3237.1.

Hamill, T.M., J.S. Whitaker, S.L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. – Bull. Amer. Meteor. Soc. **87**, 33–46, DOI: 10.1175/BAMS-87-1-33.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. – Wea. Forecast. **15**, 559–570, DOI: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hirschberg, P.A., E. Abrams, A. Bleistein, W. Bua, L.D. Monache, T.W. Dulong, J.E. Gaynor, B. Glahn, T.M. Hamill, J.A. Hansen, others, 2011: A weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information. – Bull. Amer. Meteor. Soc. **92**, 1651–1666, DOI: 10.1175/BAMS-D-11-00073.1.

Jammalamadaka, S.R., A. Sengupta, 2001: Topics in circular statistics, volume 5. – World Scientific Publishing Company.

Jolliffe, I., 2005: Principal component analysis. – Wiley Online Library.

Junk, C., L. von Bremen, M. Kühn, S. Späth, D. Heinemann, 2014: Comparison of post-processing methods for the calibration of 100 m wind ensemble forecasts at off-and onshore sites. – J. Appl. Meteorol. Climatol. **53**, 950–969, DOI: 10.1175/JAMC-D-13-0162.1.

Lorenz, E., 1963: Deterministic nonperiodic flow. – J. Atmos. Sci. **20**, 130–141, DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Martín, M., F. Valero, A. Pascual, J. Sanz, L. Frias, 2014: Analysis of wind power productions by means of an analog model. – Atmos. Res. **143**, 238–249, DOI: 10.1016/j.atmosres.2014.02.012.

Miller, M., R. Buizza, J. Haseler, M. Hortal, P. Janssen, A. Untch, 2010: Increased resolution in the ECMWF deterministic and ensemble prediction systems. – ECMWF newsletter **124**, 12.18.

Murphy, A., R. Winkler, 1987: A general framework for forecast verification. – Mon. Wea. Rev. **115**, 1330–1338, DOI: 10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2.

Nielsen, H., H. Madsen, T. Nielsen, 2006: Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. – Wind Energy **9**, 95–108, DOI: 10.1002/we.180.

Panziera, L., U. Germann, M. Gabella, P. Mandapaka, 2011: NORA–nowcasting of orographic rainfall by means of analogues. – Quart. J. Roy. Meteor. Soc. **137**, 2106–2123, DOI: 10.1002/qj.878.

Pinson, P., 2012: Adaptive calibration of (u, v)-wind ensemble forecasts. – Quart. J. Roy. Meteor. Soc. **138**, 1273–1284, DOI: 10.1002/qj.1873.

Pinson, P., 2013: Wind energy: Forecasting challenges for its operational management. – Stat. Sci. **28**, 564–585, DOI: 10.1214/13-STS445.

Pinson, P., P. McSharry, H. Madsen, 2010: Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation. – Quart. J. Roy. Meteor. Soc. **136**, 77–90, DOI: 10.1002/qj.559.

Roulston, M.S., D.T. Kaplan, J. Hardenberg, L.A. Smith, 2003: Using medium-range weather forcasts to improve the value of wind energy production. – Renew. Energ. **28**, 585–602, DOI: 10.1016/S0960-1481(02)00054-X.

Schwarz, G., 1978: Estimating the dimension of a model. – Ann. Stat. **6**, 461–464.

Thorarinsdottir, T., T. Gneiting, 2010: Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. – J. R. Stat. Soc.: Series A(Statistics in Society) **173**, 371–388, DOI: 10.1016/S0960-1481(02)00054-X.

Timbal, B., P. Hope, S. Charles, 2008: Evaluating the consistency between statistically downscaled and global dynamical model climate change projections. – J. Climate **21**, 6052–6059, DOI: 10.1175/2008JCLI2379.1.

Van den Dool, H., 1989: A new look at weather forecasting through analogues. – Mon. Wea. Rev. **117**, 2230–2247, DOI: 10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2.

Vanvyve, E., L. Delle Monache, A.J. Monaghan, J.O. Pinto, 2015: Wind resource estimates with an analog ensemble approach. – Renew. Energ. **74**, 761–773. DOI: 10.1016/j.renene.2014.08.060.

Wilks, D.S., 2011: Statistical methods in the atmospheric sciences. – Academic press, 676 pp.

Zugno, M., P. Pinson, T. Jonsson, 2013: Trading wind energy on the basis of probabilistic forecasts both of wind generation and of market quantities. – Wind Energy **16**, 909–926, DOI: 10.1002/we.1531.