

Forecasting ramps of wind power production with numerical weather prediction ensembles

Arthur Bossavy, Robin Girard, Georges Kariniotakis

► **To cite this version:**

Arthur Bossavy, Robin Girard, Georges Kariniotakis. Forecasting ramps of wind power production with numerical weather prediction ensembles. *Wind Energy*, Wiley, 2013, 16 (1), pp.51-63. <10.1002/we.526>. <hal-00682772>

HAL Id: hal-00682772

<https://hal-mines-paristech.archives-ouvertes.fr/hal-00682772>

Submitted on 7 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Forecasting Ramps of Wind Power Production with Numerical Weather Prediction Ensembles

Arthur Bossavy¹, Robin Girard², and George Kariniotakis³

^{1,2,3}MINES ParisTech, PSL - Research University,, PERSEE - Center for Processes, Renewable Energies & Energy Systems,, CS 10207 rue Claude Daunesse, 06904 Sophia Antipolis Cedex, France.

This is a pre-print version. Please refer to as: Bossavy et al. Forecasting Ramps of Wind Power Production with Numerical Weather Prediction Ensembles. *Wind Energy*. 2013. Vol. 16, issue 1, p. 51-63

Corresponding author email: abossavy@gmail.com

Abstract

Today, there is a growing interest in developing short-term wind power forecasting tools able to provide reliable information about particular, so-called "extreme" situations. One of them is the large and sharp variation of the production a wind farm can experience within a few hours called *ramp event*. Developing forecast information specially dedicated to ramps is of primary interest both because of the difficulties usual models have to predict them, and the potential risk they represent in the management of a power system. First, we propose a methodology to characterize ramps of wind power production with a derivative filtering approach derived from the edge detection literature. Then, we investigate the skill of numerical weather prediction ensembles to make probabilistic forecasts of ramp occurrence. Through conditioning probability forecasts of ramp occurrence to the number of ensemble members forecasting a ramp in time intervals, we show how ensembles can be used to provide reliable forecasts of ramps with sharpness. Our study relies on 18 months of wind power measures from a 8 MW wind farm located in France and forecasts ensemble of 51 members from the Ensemble Prediction System (EPS) of the European Center for Medium-Range Weather Forecasts (ECMWF).

1 Introduction

Wind energy is the fastest growing renewable energy source in the world. The total installed capacity has increased impressively during the last decade, passing from 10 GW in 1998 to 158 GW in 2009¹. This is an encouraging development if we consider the issue of managing power systems with high penetration of wind power. Indeed, the variable nature of wind generation makes it difficult to reach the basic requirement of balancing the electricity demand by an equal production. Short-term forecasts of wind power production up to 2 or 3 days ahead can facilitate the management of power systems by operators. Wind power forecasts are useful for performing various power system management functions, like the dynamic quantification of reserves or the optimization of combined wind-hydro power plant scheduling [1, 2]. They also prove valuable when incorporated into bidding strategies for participating in electricity markets [3, 4].

Most existing wind power forecasting models are designed to provide point forecasts of expected future wind farm production. In recent years, research work has focused on associating uncertainty estimation with this type of point forecast. New forecasting approaches (often called probabilistic forecasting methods) have recently emerged that provide estimations of the entire distribution of future production. In such methods, forecasts may take the form of either quantile estimations [5, 6, 7] or density estimations [8, 9]. Two extensive reviews of the state of the art in wind power forecasting are given in [10, 11].

¹<http://www.ewea.org>

Nowadays, wind power forecasting systems are used operationally by end-users. It is increasingly apparent, however, that current forecasting technology cannot properly handle extreme situations related to wind generation. Extreme situations may take different forms, corresponding either to extreme weather phenomena or to critical periods for power system operation [12]. An example is the steep and high increase or drop in production that a wind farm can experience in the space of a few hours, commonly known as a *ramp event*. Operational systems’ performance in forecasting large variations of wind power production is often reported to be significantly lower than usual [13, 14]. This means that we either need to improve forecasting tools’ performance in these situations, or develop new dedicated forecasting tools. This paper investigates the development of a new tool to forecast ramps and the associated uncertainty.

In literature related to wind energy, the term “ramp” may refer to wind power variations taking place over different time scales. It sometimes refers to intra-hourly variations, e.g. 10 to 60 minute variations [15, 16, 17]. In this paper, we use it to refer to significant changes in wind generation over one or several hours. An accurate prediction of this type of variation for the next few hours up to several days ahead therefore relies on weather forecasts provided by Numerical Weather Prediction (NWP) models [13, 14, 18]. Making a reliable forecast of exactly when such events will occur is a significant challenge. Indeed, errors in predicting the underlying weather conditions made by NWP models often result in inaccurate forecasting of ramp timing [19]. An error in forecasting ramp timing is referred to as a *phase error*. There is a growing interest in developing forecasting approaches dedicated to ramps. A survey of these approaches is presented in [20].

The way standard forecasting approaches provide uncertainty estimations cannot be adapted when focusing on the temporal uncertainty of ramp events. For instance, prediction intervals are commonly provided on a per-horizon basis, and are dedicated to uncertainty estimations associated to power level forecasts (see [21] for instance). While such forecasts may be easily integrated into a decision-making process, they do not provide an explicit estimation of the temporal uncertainty associated with a particular ramp event. Pioneer work in [19] associates an estimation of the phase error distribution with point forecasts of wind power production. These estimations take the form of a probability density function with finite support. This kind of support is designed to contain all ramps. However, evaluation results indicate poor reliability since numerous ramps are observed outside the support. Another limitation of the methodology in [19] is the unconditional estimation of the phase error distribution.

In this paper, we propose to forecast ramps on the basis of a specific timing, plus the probability of observing a ramp within a set of time intervals (hereafter referred to as “prediction intervals”) centered on that timing. Thus, our approach aims at providing a suitable uncertainty estimation associated with the forecast of a ramp timing. The proposed methodology relies on numerical weather prediction ensembles. Ensembles are alternative numerical weather predictions that are produced by perturbing the initial conditions, or a different parameterization of a numerical weather prediction model. The unperturbed prediction is referred to as control forecast and, in general, provides the most accurate wind production forecasts. Ensembles have already been used to derive probabilistic forecasts of wind generation [22, 23], and to assess future large-scale weather patterns [24]. Forecasts of wind power variations made by high-resolution ensembles have been evaluated to some extent in offshore conditions [25]. They have also been considered in a prototype ramp forecasting system [26]. In the latter, NWP ensembles are used to derive hourly distribution forecasts of wind generation. Then, statistical random sampling on two consecutive distribution forecasts makes it possible to estimate the distribution of production changes from one hour to the next. The main limitation of this methodology is that the correlation between two consecutive wind generation distribution forecasts is not properly captured, resulting in unrealistic forecasts of hourly wind power changes. The methodology proposed here is completely different, since it is based on collecting ramp forecasts from the members of the wind power ensembles. These forecasts are then clustered into coherent groups, each of them resulting in a unique forecast event. The use of the ensembles from an event-based point of view allows us to go further in evaluating the merits of the ensembles themselves as weather forecast products [27].

The paper is organized as follows. In Section 2, we discuss the definition of a ramp and introduce a methodology to detect and characterize a ramp event from a wind power time series. The definition proposed relies on considerations from the signal processing field and uses a filtering/thresholding approach.

In Section 3, we start by discussing how to identify a forecast ramp event among an ensemble of wind power forecast time series. Then, we introduce the proposed ramp forecasting methodology. Section 4 gives an evaluation of the proposed methodology. Conclusions and perspectives are given in Section 5.

2 Definition and characterization of a ramp event

In the related literature, a change in wind power production is considered as a ramp if it exceeds a given threshold in a period shorter than a given maximum duration. This threshold is sometimes expressed as a percentage of the nominal capacity of the wind farm. Thus, detecting a ramp relies on computing the difference between the minimum and maximum production during the considered period. This computing approach seems to be followed by [13, 14, 19], where it is explained in a descriptive way rather than using an explicit mathematical formulation. In general, this approach is expected to be highly noise sensitive given the high variability of wind power production. In this section, we propose a methodology for defining and detecting a ramp in a wind power time series, based on a linear filtering approach that aims to handle the above mentioned noise issue. As a result, we achieve a ramp characterization using a set of three parameters: the *support*, *timing* and *intensity* of the ramp.

2.1 Measuring time variations of wind power

The wind power production signal is characterized by high variability. For ramps, we need to propose a definition of such events that relies on a robust and computationally efficient way of measuring high and steep power variations, while appropriately handling the inherent noise in the process. Measuring variations, and detecting edges in a signal have been widely considered in the field of signal processing (for a literature overview we refer to [28]). The most common approach consists in estimating the first order derivative of the signal through filtering. The filtering considered associates the operations of smoothing and differentiating the signal. Following standard practice, we maintained a measure that would make interpretations easy and then chose to smooth the wind power signal using a moving average linear filter before computing the first order finite differences.

$$p_t^f = \frac{1}{n} \sum_{h=1}^n p_{t+h} - \frac{1}{n} \sum_{h=0}^{n-1} p_{t+h-n} \quad (1)$$

where p_t is a wind power time series and n is both the order of the moving average filter and the time step of finite differences. The filtered power p_t^f is merely the difference between the levels of average power from both sides of the instant t . Smoothing data allows us to regularize the differentiation operation, making it robust against noise [29]. The resulting variation measure can be rewritten by convolving the power signal p_t with the commonly known *difference of boxes* edge detector f_{DOB}

$$p_t^f = p_t * f_{DOB} \quad (2)$$

$$f_{DOB}(t) = \begin{cases} 0 & \text{if } t = 0 \text{ or } |t| > n \\ 1/n & \text{if } 0 < t \leq n \\ -1/n & \text{if } 0 > t \geq -n \end{cases} \quad (3)$$

Canny [30] introduced three criteria to evaluate edge detectors based on to their ability to detect and localize edges. Considering a model composed by an ideal step edge and white Gaussian noise, he theoretically showed the good performances of the considered edge detector f_{DOB} to detect step edges such as ramps.

In the literature on signal processing, the parameter n is generally introduced as a smoothing parameter controlling the trade off between noise reduction in the signal p_t and interference mitigation in detecting consecutive edges [31]. To avoid such interferences, n should be chosen to be no greater than the time period separating the two consecutive edges we are interested in. The choice of the value of n is described in more detail in the case-study presented in Section 4.1.

2.2 Ramp detection and characterization

Variations in the power signal p_t coincide with local extrema of the filtered power signal p_t^f (see Figure 1). Such extrema can be seen as a combined measure of both the magnitude and the steepness of the wind power variations. We consider that a variation is high and steep enough to be regarded as a ramp if the absolute value of p_t^f is higher than a given threshold $\tau > 0$

$$|p_t^f| \geq \tau \quad (4)$$

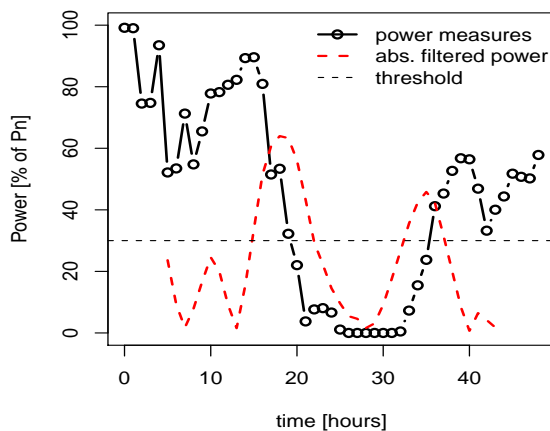


Figure 1: Example of an hourly wind power time series p_t covering 48 hours of production from a French wind farm. The measures have been normalized based on the wind farm nominal capacity P_n . The absolute value of the filtered production p_t^f (red dashed line) upwardly crosses the threshold τ (black dashed line) twice. First, a decreasing ramp is detected. Its timing is located at $t_r = 19$ hours, its intensity represents 64,0% of the nominal capacity of the wind farm and its support is 6 hours long. It is followed by an increasing ramp at $t_r = 36$ hours, with an intensity 45.8% of P_n and a support of 4 hours.

The complex geometry of a wind power variation may result in the absolute value of the filtered signal p_t^f fluctuating around the threshold τ . This phenomenon, which is responsible for breaking up edges, is known as “streaking” in literature on edge detection [30]. To palliate it, a common practice is to apply a so-called hysteresis thresholding. Rather than a single value τ , two thresholds are then considered. Filtered outputs above the highest threshold are immediately considered as associated with an edge, as are those connected to them and lying between the two thresholds. In our study, we noticed that the streaking phenomenon occurs only rarely. For the sake of simplicity, we decided not to use this type of thresholding methodology.

A ramp is a time-limited event that can be characterized by a starting time t_s and an ending time t_e . They are defined here by the instants at which the absolute value of the filtered signal p_t^f crosses upwards and downwards the threshold τ . The period between these two instants $[t_s, t_e]$ then defines the *support* of the ramp. It is clear that the length of the support depends on the value of the threshold. In fact, ramp support should be considered as a requisite characteristic for developing our forecasting approach but is not essential to the definition of the event. Even when a ramp is not localized in time, it is useful to associate a particular timestamp to a ramp. We choose to associate to a ramp event the time t_r ($t_s \leq t_r \leq t_e$), called *timing* of the ramp, for which the absolute value of the filtered signal p_t^f reaches its maximal magnitude. If there are several maxima, the one associated to the nearest instant from the starting time t_s is chosen. This maximum defines the *intensity* of the ramp. To summarize, three quantities are defined to characterize a ramp event: support, timing and intensity.

3 Forecasting ramp events using ensemble wind power forecasts

The detection methodology of ramps that we introduced in the previous section relies on a filtering/thresholding approach applied to a measured wind power signal. In subsection 3.1, we discuss its use to detect ramps in a point forecast time series and then in an ensemble of forecast time series. In the latter case, an ensemble of ramp characteristics is produced, which may be related to different ramp events. In subsection 3.2, we propose solutions to cluster ramp forecasts obtained by different ensemble members into coherent groups. This enables us to discriminate between forecasts and associate them with separate events. The characteristics of the forecast events are then exploited in subsection 3.3 to produce probability forecasts of ramp occurrence in a set of prediction intervals.

3.1 Forecasting an ensemble of ramp characteristics

The methodology presented in the previous section was introduced for detecting ramps in a time series of measured wind power. Its application to a single or ensemble of wind power forecast time series is straightforward. In the case of ensembles, it is assumed that the methodology is applied to each member separately. However, it is not obvious whether the parameters, and namely the value of the threshold, used to detect ramps in a time series of measurements can be used as such to detect ramps in a time series of forecasts. Hereafter, we denote as $\hat{\tau}$ the threshold used to detect ramps in forecast time series.

Short-term wind power point forecasting models tend to overestimate production within a low wind speed range, and underestimate it at high wind speeds. This phenomenon, which is related to the wind-to-power conversion process, was demonstrated in [32] for five power prediction models based on "statistical" and "physical" approaches (with model output statistics correction), all using spot NWP as input, and tuned to minimize the overall Root Mean Squared Error (RMSE). As a consequence, it is expected that the variations in power forecasts will on average be lower than those of the measurements. When such forecasting methods are applied to NWP ensembles, the resulting power ensembles may be poorly calibrated. In most applications, a form of recalibration is required [22, 23]. Here, in order to correctly identify ramps detected on both forecasts and observations as the same event, we can use a lower threshold value $\hat{\tau} \leq \tau$ to detect forecast ramps. One possible approach is to adjust this second threshold, so that the same amount of ramps are detected on both forecasts and observations. An alternative approach is to select the value of this threshold with respect to some other desirable properties of the considered ramp forecasting methodology. In this case, we need to keep in mind that increasing $\hat{\tau}$ decreases the number of forecast events, resulting in both less captured ramps and a reduced number of wrongly forecast events. In Section 4, we investigate the performance of the proposed approach to forecast ramps for different values of this threshold.

Once we have chosen a suitable value for the threshold parameter $\hat{\tau}$, the filtering/thresholding approach can be applied to each member of the ensemble of wind power forecast time series. This procedure results in the production of an ensemble of ramp forecast characteristics (timing, support, intensity). These characteristics then need to be clustered into coherent groups, each of which corresponds to a unique forecast event.

3.2 Clustering an ensemble of ramp characteristics

When the ramp detection procedure is applied to each member of the forecast power ensembles, it is expected that the ramps detected using each member will correspond to the same ramp event when all or part of their characteristics (intensity, support and timing) are close enough to each other. Thus, for each lead time, we propose to compute the number of ensemble members with an absolute value of filtered power higher than $\hat{\tau}$. For a given lead time, this corresponds to the number of ensemble members with overlapping ramp supports. An example of the resulting signal is given in Figure 2, for respectively the increasing and decreasing forecast ramps.

Based on this signal, we propose two alternative approaches for clustering the ramp characteristics. In the first approach, denoted hereafter as **A1**, we cluster forecast ramps with overlapping supports into a unique event. In the second approach, denoted as **A2**, any local maximum in the number of

overlapping ramp supports defines a single event. In the example of Figure 2, the first approach gives two forecast events: an increasing ramp followed by a decreasing one. On the other hand, the second approach gives three forecast events: an increasing ramp and two decreasing ones. Although the second approach may result in forecasting more ramps than observed, it may allow us to capture more events. The two approaches are evaluated in Section 4.

Once forecast events have been identified using one of the proposed clustering approaches, we count the number of forecasting members for each of them, taking care not to count each member more than once (e.g. a member could forecast the event two or more times because of the streaking phenomenon). We also forecast a mean timing \bar{t}_r of the event by averaging the timings forecast by the ensemble members. If we consider the mean intensity, we end up obtaining a full characterization of the forecast event from the ensemble of forecast time series.

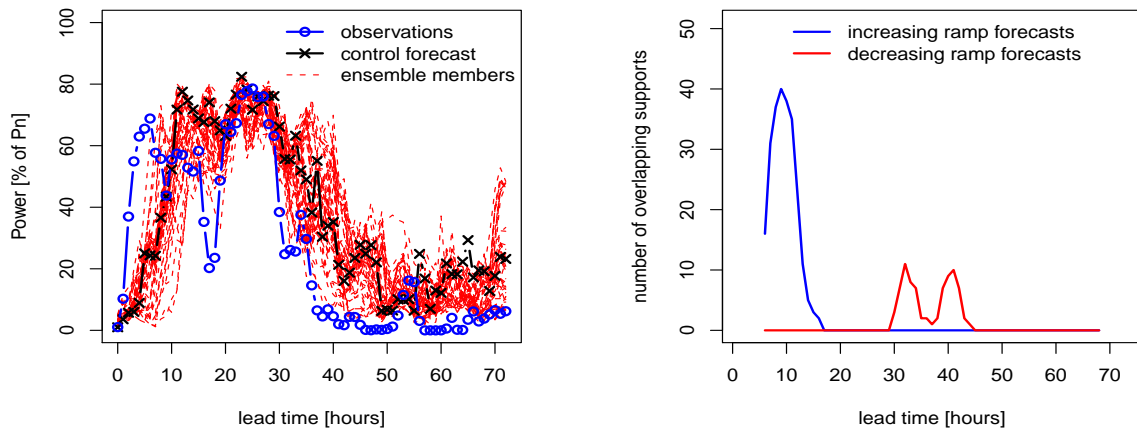


Figure 2: Example of an ensemble of wind power production forecasts composed by the control forecast and 50 members. The figure on the left depicts the hourly time series of measured and forecast wind power up to 72 hours ahead. The right-hand figure shows the number of ensemble members with overlapping ramp supports for increasing forecast ramps (blue curve) and decreasing forecast ramps (red curve). Ramps have been detected using a time scale parameter of $n = 5$ hours and a threshold parameter $\hat{\tau} = 30\%$ of the wind farm’s nominal capacity P_n .

3.3 Making probabilistic forecasts of ramp occurrence using ensembles

In the proposed approach, we produce probabilistic forecasts of ramp occurrence using a set of prediction intervals $\{I_\delta, \delta = 1, \dots, \delta_{max}\}$ centered on the average forecast timing \bar{t}_r , with associated probabilities p_δ . The intervals’ radius δ are fixed beforehand and the probabilities of observing a ramp in these intervals are estimated. The maximum interval radius considered δ_{max} is a parameter of the procedure which can be configured as a function of the end-user requirements. Unlike the proposal made in [19], characterizing the phase error distribution using a set of prediction intervals with associated probabilities allows us to account for the non-occurrence of a ramp.

To provide forecast information depending on forecast conditions, we propose to estimate p_δ conditioned to the number of ensemble members N_{mem} forecasting the ramp. For the sake of simplicity, and to produce parsimonious base-line models, we consider N_{mem} to be the only predictor. However, as a perspective of this work, one could imagine more complex models with additional inputs (e.g. average intensity forecast made by the ensemble members).

For a given interval's width 2δ , and since the center of the intervals is already forecast, it is possible to reformulate the problem of estimating p_δ in terms of a statistical regression problem. Consider the binary random variable Y_δ that equals 1 if a ramp occurs in I_δ and 0 otherwise. Thus, the probability $p_{\delta,m}$ of observing a ramp in I_δ given that m ensemble members are forecasting the ramp is given by the following regression formula

$$\begin{aligned} p_{\delta,m} &= \mathbb{P}(Y_\delta = 1 | N_{mem} = m) \\ &= \mathbb{E}[Y_\delta | N_{mem} = m] \end{aligned} \quad (5)$$

The estimation procedure Estimating $p_{\delta,m}$ as given by Equation (5) is a particular instance of a regression problem. As a consequence, there are numerous alternative estimation procedures that could be used here. In our case however, there are some particular properties that should be taken into account:

- $p_{\delta,m}$ is a probability and there are regression models dedicated to this kind of estimation problem.
- The distribution of the random variable N_{mem} used as an explanatory variable is strongly right-skewed. In other words, a ramp is more likely to be forecast by a small rather than a large number of ensemble members.

Based on these considerations, we choose to apply two alternative statistical estimation procedures, which are evaluated in Section 4.3. First, a Nadarya-Watson estimator is used with a varying bandwidth to introduce flexibility into our estimations and to adapt to the particular sampling of the explanatory variable N_{mem} . Then, we consider the alternative of a logistic regression model specifically dedicated to estimating probabilities.

Nadarya-Watson estimator A Nadarya-Watson estimator makes estimations according to the following generic formula

$$p_{\delta,m} = \frac{\sum_{i=1}^n \mathcal{K}_\lambda(m_i, m) Y_\delta^i}{\sum_{i=1}^n \mathcal{K}_\lambda(m_i, m)} \quad (6)$$

where n is the size of the data set dedicated to probability estimations and \mathcal{K}_λ is a kernel function. Regression techniques using kernels fit a local model around each target point m . The value $\mathcal{K}_\lambda(m_i, m)$ gives the weighting of points m_i in the neighborhood of the target point m . The kernel's bandwidth λ gives the size of such a neighborhood and controls the estimation bias-variance trade off. Increasing λ results in averaging more observations and lowers the variance (but increases the bias) in each target point estimation. These techniques are very popular as they do not require any parametric assumption. In our study, we consider a tricube kernel and select λ by making it vary with m according to a nearest-neighbors procedure

$$\lambda(m) = |m - m_{[k]}| + 1 \quad (7)$$

where $m_{[k]}$ is the k th closest point from the target m . The parameter k is selected from a 10-fold cross-validation procedure. A local choice of λ allows us to incorporate more information about the probability law generating data points m_i and often gives better results when estimating curves with a complex shape [33]. For an overview on kernel smoothers and bandwidth selection procedures we refer to [34, 35].

Logistic regression The logistic regression model allows us to infer ramp occurrence probabilities from a linear function of the number of ensemble members m forecasting a ramp

$$\log \frac{p_{\delta,m}}{1 - p_{\delta,m}} = \alpha_\delta m + \beta_\delta \quad (8)$$

where α_δ and β_δ are the model parameters to estimate. To ensure that the estimated probabilities remain within $[0, 1]$, the modeling requires us to express such probabilities as a linear function of m through a logit transformation. The model is fit by maximum likelihood, assuming the conditional distribution of Y_δ given N_{mem} is binomial. For all values of δ , a statistical test (the Wald test) rejects the hypothesis that assumes a null value of coefficients α_δ and β_δ with a significance level of 5%. For an introduction to logistic and generalized linear models, we refer to [34, 36].

4 Evaluation framework and results

4.1 The Case-Study

In order to evaluate the proposed methodology, we considered a wind farm with a nominal capacity of 8MW, located on a complex terrain in the south of France (a few dozen kilometers from the Mediterranean Sea). Power measurements delivered by the SCADA system of the wind farm were considered. They covered a period of 18 months from July 2004 to December 2005 with a 10-min temporal resolution. For the purpose of the study, the data was averaged to hourly values.

In our study, ramp events correspond to variations detected with a filter (half) width of $n = 5$ hours. Such a value is reasonable if variations under study are related to somewhat persistent changes in weather conditions. Appropriate choices could be made by end-users focusing on shutdown events, for instance². In this case, a lower value of n would be required. Further work could include a multi-scale analysis of the wind power signal using multi-scale edge detection techniques [28]. For a variation to be considered as a ramp, its minimum intensity τ has to be set according to which proportion of the most “extreme” (e.g. highest and steepest) variations we are interested in. We consider a value of $\tau = 30\%$ of the nominal capacity P_n of the considered wind farm. With such a value, we detected 168 ramps during the period between 1st April 2005 and 31st December 2005, that is an average 4.2 ramps a week. Note that our methodology is independent of the value of τ and a more specific choice could be made by end-users. As mentioned in Section 3.1, we will now investigate the performance of our approach to forecasting ramps considering different threshold values $\hat{\tau}$ to detect ramps on wind power forecast ensembles.

NWP ensembles of 51 members are provided by the EPS system of ECMWF with a spatial resolution of 1° in both longitude and latitude. This resolution corresponds to about 75 – 80 km in the East-West direction and 110 km in the North-South direction. The NWP data is issued twice a day with a temporal resolution of 6 hours and for a horizon of 72 hours. Hourly predictions were obtained through interpolating the 6-hour values. To forecast wind power, we considered weather forecasts generated at the closest grid point to the wind farm located 10 meters above ground level. We used the *Random Forest* nonparametric estimation procedure to produce wind power forecasts using wind speed and direction NWP as input. *Random Forest* is a machine-learning algorithm that aggregates predictions from a large number of regression trees, each tree being built from a bootstrap replica of the sample data and random selection of potential predictors. Recent work on meteorological downscaling and post-processing acknowledges the robustness of *Random Forest* against over-fitting and its ability to deal with nonlinearity and complex interactions between variables [37, 38]. For more details we refer to [34, 39]. To deal with ensembles, we applied the Random Forest procedure to each member of the NWP ensemble to obtain a corresponding power forecast ensemble.

Our wind power forecasting model was trained over a nine-month period (July 2004 - March 2005) and tested for another nine months (April to December 2005). In the testing set, the model provides a root mean squared error and a mean absolute error not greater than respectively 15% and 11% of the nominal capacity of the wind farm for the next 24h. This is a satisfactory performance given the state of the art [40], especially if one considers the complexity of the terrain in which the wind farm is embedded and the rather coarse resolution of the NWP data. The data sets used for training and testing the ramp forecasting models were four and a half months long each, covering the 9-month period between April

²In situations with very high wind speeds, some or all of a wind farm’s turbines can be temporary stopped to avoid damage due to strong mechanical constraints. Such situations can turn into high variations of decreasing and increasing wind power production within a short period of time

and December 2005. We derive and hereafter evaluate ramp forecasts taking prediction intervals with a maximal radius value of $\delta_{\max} = 8$ hours. With such a value, we expect to describe most of the phase error distribution from the considered set of prediction intervals.

4.2 Evaluating the capture of ramp events from ensemble-based forecasts

In this section we focus on evaluating the rate of observed ramps that are effectively forecast, denoted as *ramp capture ratio* [19].

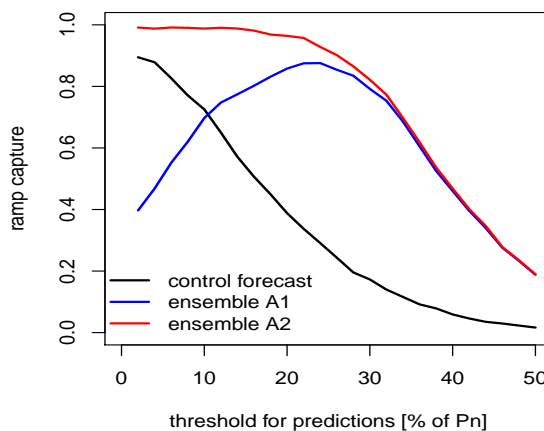


Figure 3: Ramp capture results when forecasting ramps using the control forecast (black curve) or the ensembles (blue curve is for clustering approach **A1**, red curve for approach **A2**). Results are for values of $\hat{\tau}$ ranging from 2% to 50% of the wind farm nominal capacity P_n , and a maximum delay between forecast and observed ramp events of $\delta_{\max} = 8$ hours.

Figure 3 gives this ratio for ensemble-based forecasts and forecasts derived from the control member. As noticed in Section 3.1, predicting ramps from forecast variations of lower intensity (e.g. by decreasing $\hat{\tau}$), results in capturing more events. This property is valid when predicting ramps from a single wind power scenario, but may not hold when considering ensembles. Actually, it depends on the approach used to cluster forecasts issued by different ensemble members. When using the approach **A1**, low values of $\hat{\tau}$ result in ramp supports forecast by the ensemble to overlap, and ultimately in separated forecast events to be merged. This is the reason why in Figure 3, for $\hat{\tau} \leq 20\%$ of P_n and with clustering approach **A1**, the ramp capture decreases with $\hat{\tau}$.

As seen in Figure 3, ensembles generally outperform the control member in terms of ramp capture. However, such improvements may coincide with a higher number of false alarms, e.g. events that are forecast but not observed. In fact, the use of ensembles may not result in a better ramp capture/forecast accuracy trade off. Alternative approaches to cluster ramp forecasts issued by different ensemble members could be envisaged. We have considered other clustering approaches, such as hierarchical agglomerative clustering with an euclidean distance [34]. However, such clustering approaches turn out to be far more computationally demanding when the number of ramp forecasts to cluster is increased (e.g. when decreasing $\hat{\tau}$).

4.3 Evaluating ensemble-based probability forecasts of ramp occurrence

Here we evaluate the probability forecasts for ramp occurrence. We first propose using the Brier score (BS) [41], which is a dedicated measure of accuracy. Consider $p_i = p_{\delta, m}^i$ as a probability forecast of ramp

occurrence in a prediction interval I_δ . Let $y_i = y_\delta^i$ be equal to one if a ramp is observed in I_δ and zero otherwise, then

$$\text{BS} = \frac{1}{n} \sum_{i=1}^n (p_i - y_i)^2 \quad (9)$$

where n is the test set sample size. As a reference approach, we consider the climatology which, for a given interval's width 2δ , derives unconditional probability forecasts from the empirical event frequency $\bar{y} = 1/n \sum_{i=1}^n y_i$. The relative improvement in terms of the decrease in forecasting errors, referred to as Brier skill score with respect to climatology (BSS_{clim}), is given in Table 1. Conditioning clearly improves forecast accuracy. Such improvement is paramount for low values of $\hat{\tau}$, e.g. for high ramp capture. When focusing on the approach **A1**, such improvement also increases with the size of prediction intervals I_δ , which then results in better performances compared to the approach **A2** for large values of δ (e.g. $\delta = 8$ hours in Table 1). However, the sampling uncertainty in the estimation of BSS_{clim} makes any further discrimination between approaches **A1** and **A2** difficult. For the same reason, it is not easy to conclude in favor of a particular estimation procedure between the logistic regression model and the Nadarya-Watson estimator.

Table 1: Brier skill score with respect to climatology BSS_{clim} (in percentages) of ensemble-based probability forecasts of ramp occurrence in prediction intervals I_δ (δ is in hours). Results are given for both the clustering approaches **A1** and **A2**, and for the logistic regression (log. reg.) and the Nadarya-Watson (N.W.) estimation procedures. To account for sampling uncertainty, the standard deviation of BSS_{clim} (given in parenthesis) has been estimated following the theoretical work of Bradley et al. [42].

$\hat{\tau}$ ($\%P_n$)	log. reg.			N.W.		
	$\delta = 2$	$\delta = 5$	$\delta = 8$	$\delta = 2$	$\delta = 5$	$\delta = 8$
Clustering approach A1						
10	5.5 (1.2)	11.1 (1.7)	16.6 (2.2)	5.4 (1.1)	10.8 (1.5)	16.6 (2.1)
20	3.8 (1.2)	9.9 (1.7)	14.3 (2.1)	4.5 (1.0)	9.8 (1.4)	13.8 (1.8)
30	3.7 (1.2)	8.0 (1.7)	9.5 (1.9)	4.2 (1.1)	8.2 (1.5)	9.4 (1.7)
40	2.9 (1.3)	1.5 (2.6)	3.3 (2.4)	2.8 (1.1)	3.7 (2.5)	4.8 (2.3)
Clustering approach A2						
10	6.0 (1.0)	10.2 (1.3)	9.2 (1.3)	6.2 (0.9)	10.9 (1.3)	9.9 (1.3)
20	5.8 (1.1)	9.9 (1.4)	8.9 (1.4)	6.1 (1.0)	9.7 (1.3)	8.6 (1.4)
30	3.9 (1.2)	6.9 (1.3)	6.3 (1.4)	3.9 (1.2)	7.0 (1.4)	5.9 (1.5)
40	3.2 (1.21)	1.0 (1.8)	1.2 (1.7)	3.2 (1.3)	2.8 (2.0)	0.9 (2.1)

In addition to using the Brier score, we evaluated our forecasts according to a reliability/sharpness paradigm [43, 44]. Results for a particular threshold value $\hat{\tau} = 30\%$ of the wind farm nominal capacity, using the clustering approach **A2** and the logistic regression model, are shown in Figure 4. Forecasts show reliability (top left Figure), which does not deviate from perfect reliability (first diagonal) any more than the sampling uncertainty does. The latter was estimated with 90% confidence intervals (vertical segments) derived from the resampling technique proposed by Broecker and Smith [45]. An alternative was to derive them analytically considering the Poisson-Binomial distribution of $n\bar{y}$ [46]. The results from the two approaches were found to be very close.

The top right and bottom plots of Figure 4 show the sharpness property of our forecasts. We can see that the probability of a ramp occurring in prediction intervals increases with the number of members forecasting it. At the same time, ramps are seldom forecast with high probability. We quantified reliability and sharpness using measures derived from the decomposition of the Brier score proposed by Murphy

[47]. Results showed that the decreasing skill of our forecasts (see Table 1) while $\hat{\tau}$ increases was due to the simultaneous decrease in reliability and sharpness. In any case, high values of $\hat{\tau}$ (e.g. $\hat{\tau} > 40\%$ of P_n) make it possible to mitigate false alarms, which may be desirable when making deterministic forecasts of ramps. Following our probabilistic approach, the choice should be for lower threshold values leading to (sometimes large numbers of) forecasts with good reliability and sharpness.

To illustrate the forecast product of the proposed methodology, the 3 days ahead forecast scenario given in Figure 2 is represented in Figure 5, with prediction intervals and associated probability forecasts of ramp occurrence. Forecasts were derived from the clustering approach **A2** and the logistic regression model. As we described in Section 3.2, the situation considered should present two decreasing ramp forecasts. For more visibility, however, we removed one of them.

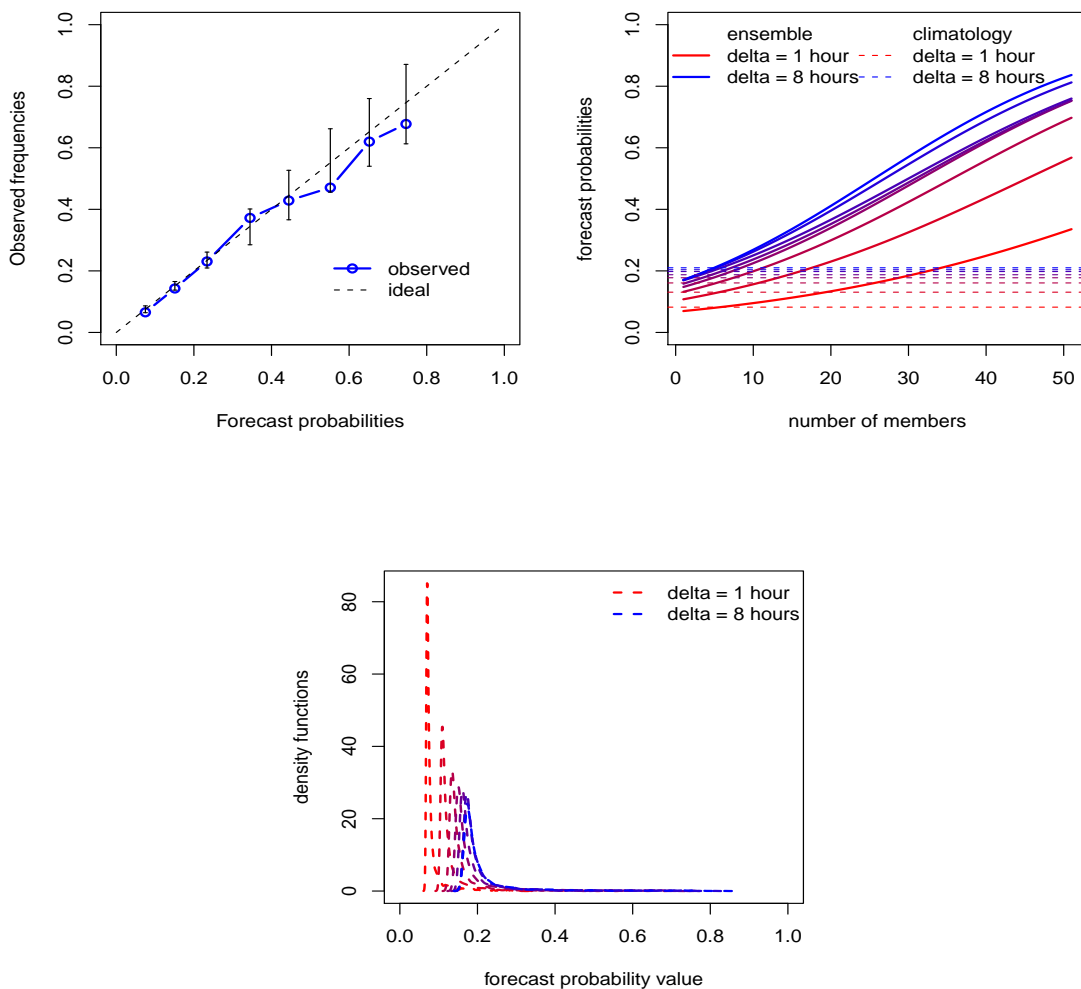


Figure 4: Reliability (top left Figure) and sharpness (top right and bottom Figures) of ensemble-based probability forecasts of ramp occurrence in prediction intervals I_δ . Results are for forecasts derived from the clustering approach **A2**, with $\hat{\tau} = 30\%$ of P_n and the logistic regression model. To take the sampling uncertainty into account when estimating the event frequencies (top left Figure), 90% centered confidence intervals were derived using the resampling technique proposed in [45].

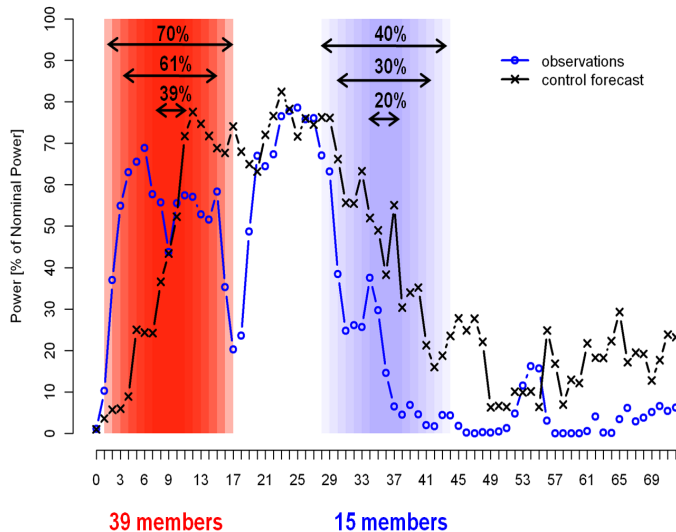


Figure 5: Prediction intervals I_δ , $\delta = 1, \dots, 8$ hours for an increasing ramp forecast by 39 members, followed by a decreasing one forecast by 15 members. Probability forecasts of ramp occurrence (given in percentages at the top of the figure) were derived using the clustering approach **A2**, $\hat{\tau} = 30\%$ of P_n and the logistic regression model. The change in forecast probability depending on the number of members forecasting a ramp provides a good illustration of the sharpness property of the proposed methodology.

5 Conclusions and Perspectives

There is growing demand for developing wind power forecasting models focusing on certain extreme situations. One well-known situation is the so-called ramp event, which involves sharp and high variations in production experienced by wind farms in the space of a few hours. These variations are critical for managing power systems with high wind penetration, as they can imply complex changes in power supply. For a power system operator, an estimation of the uncertainty of a forecast ramp should provide valuable information for making management decisions.

The first challenge in ramp forecasting is to correctly define a ramp. We proposed a definition based on filtering, guided by the significant signal processing literature on edge detection. Such a definition addresses several issues inherent to defining a ramp: at what level of power and within what duration does a variation represent a "ramp" for the end-user? We discussed these questions and proposed a detection methodology using two parameters that make it possible to set at what time scale and magnitude a variation is considered to be a ramp. Such parameters can be set depending on either the modeler's or the end-user's opinion.

A single time series of wind power forecasts up to several days may not include enough information to make secure management decisions related to the potential occurrence of a ramp. In this paper, we studied the extent to which numerical weather prediction ensembles could provide information when forecasting ramps with associated uncertainty. We showed that ensembles used with our methodology can better capture ramp events than when considering a unique wind power scenario.

Then, we set out to estimate the uncertainty associated with ramp forecasts using a set of prediction intervals with associated probabilities of ramp occurrence. We introduced sharpness into our forecasts by conditioning them to the number of ensemble members forecasting a ramp. The resulting forecasts turned out to be reliable with greater accuracy regarding climatology. Our conclusions relate to data from a multi-megawatt wind farm located on a complex terrain in the south of France, and numerical weather prediction ensembles of 51 members from the EPS system of ECMWF. In further work, an evaluation of the methodology on more sites should help to validate our conclusions under different weather regimes.

An interesting perspective would be to compare the forecasting ramp performance of different wind

power forecast ensembles. Indeed, the methodology we developed in this paper could be used to forecast ramps from either the commonly known *poor man's* ensembles, or from ensembles derived from multivariate statistical sampling, such as those proposed in [48]. The latter could provide a good alternative to numerical weather prediction ensembles for forecasting ramps, as their generation can be far less computationally expensive. Particular attention should also be paid to future potential applications of the proposed approach in forecasting wind generation ramps from a portfolio of wind farms. Finally, we observed that different ramp typologies occur, and the limits of a single ramp definition will need to be explored in further work.

Acknowledgements

The work presented in this paper is part of the European R&D Project *SafeWind* (Grant No 213740) partly funded by the European Commission under the 7th Framework Programme. One of the aims of this project is to facilitate large-scale wind integration through developing advanced wind modeling and forecasting methods, with an emphasis on extreme situations. This work is also supported by *Agence de l'Environnement et la Maitrise de l'Energie* (ADEME). We thank EDF for providing some of the data used in this work, and Erik Holmgren for his support. Finally, we would like to thank the anonymous referees who reviewed this paper for their comments and suggestions.

References

- [1] R. Doherty and M. O'Malley. A new approach to quantify reserve demand in systems with significant installed wind capacity. *Power Systems, IEEE Transactions on*, 20(2):587 – 595, 2005.
- [2] E.D. Castronuovo and J.A.P. Lopes. On the optimization of the daily operation of a wind-hydro power plant. *Power Systems, IEEE Transactions on*, 19(3):1599 – 1606, 2004.
- [3] J.M. Angarita and J.G. Usaola. Combining hydro-generation and wind energy: Biddings and operation on electricity spot markets. *Electric Power Systems Research*, 77(5-6):393 – 400, 2007.
- [4] P. Pinson, C. Chevallier, and G. Kariniotakis. Trading wind generation from short-term probabilistic forecasts of wind power. *Power Systems, IEEE Transactions on*, 22(3):1148 – 1156, 2007.
- [5] B.J. Bremnes. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, 9(1):3–11, 2006.
- [6] J.K. Moller, H.Aa. Nielsen, and H. Madsen. Time-adaptive quantile regression. *Computational Statistics & Data Analysis*, 52(3):1292 – 1303, 2008.
- [7] H.Aa. Nielsen, H. Madsen, and T.S. Nielsen. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, 9(1-2):95–108, 2006.
- [8] J. Juban, N. Siebert, and G. Kariniotakis. Probabilistic short-term wind power forecasting for the optimal management of wind generation. In *Proceedings of the IEEE Power Tech Conference, Lausanne, Swizerland*, 2007.
- [9] J.W. Taylor, P.E. McSharry, and R. Buizza. Wind power density forecasting using ensemble predictions and time series models. *Energy Conversion, IEEE Transactions on*, 24(3):775 – 782, 2009.
- [10] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725 – 1744, 2008.
- [11] G. Giebel, G. Kariniotakis, and R. Brownsword. The state of the art on short-term wind power prediction - A literature overview. Technical report, ANEMOS EU project, deliverable report D1.1. [Available online: <http://www.anemos-project.eu>], 2003.

- [12] P. Pinson. Catalogue of complex to extreme situations. Technical report, EU Project SafeWind, Deliverable Dc1.2. [Available online: <http://www.safewind.eu>], 2009.
- [13] N.J. Cutler, M. Kay, K. Jacka, and T.S. Nielsen. Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT. *Wind Energy*, 10(5):453–470, 2007.
- [14] N.J. Cutler. *Characterizing the uncertainty in potential large rapid changes in wind power generation*. PhD thesis, Electrical Engineering & Telecommunications, Faculty of Engineering, UNSW. [Available online: <http://handle.unsw.edu.au/1959.4/43570>], 2009.
- [15] H. Zheng and A. Kusiak. Prediction of wind farm power ramp rates: A data-mining approach. *Journal of solar energy engineering*, 131:031011.1–031011.8, 2009.
- [16] H. Zareipour. Wind power ramp events classification and forecasting: A data mining approach. In *Proceedings of the 2011 IEEE Power and Energy Society (PES) Annual General Meeting, Detroit, USA*, 2011.
- [17] C. Kamath. Associating weather conditions with ramp events in wind power generation. In *IEEE PES Power Systems Conference & Exposition, Phoenix, Arizona*, 2011.
- [18] N. J. Cutler, H. R. Outhred, I. F. MacGill, M. J. Kay, and J. D. Kepert. Characterizing future large, rapid changes in aggregated wind power using Numerical Weather Prediction spatial fields. *Wind Energy*, 12(6):542–555, 2009.
- [19] B. Greaves, J. Collins, J. Parkes, and A. Tindal. Temporal forecast uncertainty for ramp events. *Wind Engineering*, 33(11):309–319, 2009.
- [20] C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang. A survey on wind power ramp forecasting. A report from the Argonne U.S. Department of Energy Laboratory. [Available online at <http://www.dis.anl.gov/>], 2010.
- [21] P. Pinson and G. Kariniotakis. Conditional prediction intervals of wind power generation. *Power Systems, IEEE Transactions on*, 25(4):1845–1856, 2010.
- [22] H.Aa. Nielsen, T.S. Nielsen, H. Madsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, L. Voulund, and J. Tøfting. From wind ensembles to probabilistic information about future wind power production - Results from an actual application. In *Proceedings of the IEEE PMAPS 2006 Conference, Probabilistic Methods Applied to Power Systems, Stockholm, Sweden*, 2006.
- [23] P. Pinson and H. Madsen. Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, 12(2):137–155, 2009.
- [24] P.A. Chessa and F. Lalaurette. Verification of the ECMWF ensemble prediction system forecasts: A study of large-scale patterns. *Weather and Forecasting*, 16(5):611–619, 2001.
- [25] WEPROG ApS. High resolution ensemble for Horns Rev - HRensembleHR -. Technical report, Project funded by the Danish PSO F&U Program, Final Report. [Available online: <http://www.hrensemble.net/>], 2010.
- [26] E. Gritit and C. Potter. A prototype day-ahead forecast system for rapid wind ramp events. In *Proceedings of Windpower 2008 Conference and Exhibition, Houston, Texas*, 2008.
- [27] R. Girard and P. Pinson. Evaluation of time trajectories - Application to wind power forecasting. Submitted to Applied Energy, 2011.
- [28] D. Ziou and S. Tabbone. Edge detection techniques - An overview. *International Journal of Pattern Recognition and Image Analysis*, 8:537–559, 1998.

- [29] V. Torre and T.A. Poggio. On edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(2):147–163, 1986.
- [30] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679–698, 1986.
- [31] D. Demigny. On optimal linear filtering for edge detection. *Image Processing, IEEE Transactions on*, 11(7):728–737, 2002.
- [32] P. Pinson. *Estimation of the uncertainty in wind power forecasting*. PhD thesis, MINES ParisTech. [Available online: <http://pastel.paristech.org>], 2006.
- [33] J. Fan and I. Gijbels. Variable bandwidth and local linear regression smoothers. *Annals of Statistics*, 20:2008–2036, 1992.
- [34] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning, second edition: Data mining, inference, and prediction*. Springer Series in Statistics. Springer, 2nd ed. 2009. corr. 3rd printing edition, 2009.
- [35] W. Schucany. Kernel smoothers: An overview of curve estimators for the first graduate course in nonparametric statistics. *Statistical Sciences*, 19:663–675, 2004.
- [36] H. Madsen and P. Thyregod. *An introduction to general and generalized linear models*. Chapman & Hall, 2010.
- [37] E. Eccel, L. Ghielmi, P. Granitto, R. Barbiero, F. Grazzini, and D. Cesari. Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models. *Nonlinear Processes in Geophysics*, 14(3):211–222, 2007.
- [38] R. Davy, M. Woods, C. Russell, and P. Coppin. Statistical downscaling of wind variability from meteorological fields. *Boundary-Layer Meteorology*, 135:161–175, 2010.
- [39] L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.
- [40] G. Kariniotakis, I. Marti, D. Casas, P. Pinson, T.S. Nielsen, H. Madsen, G. Giebel, J. Usaola, I. Sanchez, A.M. Palomares, R. Brownsword, J. Tambke, U. Focken, M. Lange, P. Loucka, G. Kallos, C. Lac, G. Sideratos, and G. Descombes. What performances can be expected by short-term wind power prediction models depending on site characteristics? In *Proceedings of the 2004 European Wind Energy Conference EWEC'04, London, UK, 2004*.
- [41] G. Brier. Verification of forecast expressed in terms of probability. *Monthly weather review*, 78:1–3, 1950.
- [42] A. Bradley, S.S. Schwartz, and T. Hashino. Sampling uncertainty and confidence intervals for the Brier score and Brier skill score. *Weather and Forecasting*, 23(5):992–1006, 2008.
- [43] T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.
- [44] P. Pinson, H. Aa. Nielsen, J. K. Moller, H. Madsen, and G. Kariniotakis. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, 10(6):497–516, 2007.
- [45] J. Broecker and L.A. Smith. Increasing the reliability of reliability diagrams. *Weather and Forecasting*, 22(3):651–661, 2007.
- [46] Y. Hong. On computing the distribution function for the sum of independent and non-identical random indicators. Technical report, Department of Statistics, Virginia Tech, Blacksburg, VA, 2011.

- [47] A.H. Murphy. A new vector partition of the probability score. *Journal of Applied Meteorology*, 12:595–600, 1973.
- [48] P. Pinson, H. Madsen, H.Aa. Nielsen, G. Papaefthymiou, and B. Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009.