SafeWind


Collaborative project funded by the European Commission under the 7th Framework Program, Theme 2007-2.3.2: Energy

"Multi-scale data assimilation, advanced wind modelling & forecasting with emphasis to extreme weather situations for a safe large-scale wind power integration"

Grant Agreement N°: 213740

"Deliverable Dp-6.3 : Forecast combination conditioned to a predictability information, and benefits for controlling forecast error."

DOCUMENT TYPE	deliverable
DOCUMENT NAME:	Swind_deliverableDp-6.3_ForecastCondtionaPredictInfo.pdf
VERSION:	V1.0
DATE:	2012.01.04
CLASSIFICATION:	R0: Public
STATUS:	Released

Abstract: This document is a two part report providing the contribution of Oxford and ARMINES to Task 6.3 and Deliverable Dp6.3. Both work propose a probabilistic forecasting procedure conditionally to predictability information. In the work of Oxford (starting p3), the used information is general variability information, in the work of ARMINES (starting p33) the used information is information related to the possibility of upcoming ramp events (large and sharp variation of the production).

AUTHORS[1], REVIEWERS				
MAIN AUTHOR/EDITOR:	R.Girard			
AFFILIATION:	ARMINES			
ADDRESS:	B.P. No 207, F-06904 Sophia-Antipolis Cedex, France.			
TEL.:				
EMAIL:	robin.girard@mines-paristech.fr			
FURTHER AUTHORS:	Patrick McSharry, Georgios Anastasiades, Arthur Bossavy			
PEER REVIEWERS:				
REVIEW APPROVAL:	Approved :		Rejected (improve as indicated below) :	
SUGGESTED IMPROVEMENTS:	*For a long list of remarks make reference to another document*			

VERSION HISTORY			
VERSION[2]:	DATE:	COMMENTS, CHANGES, STATUS:	PERSON(S):
1.0	06.08.2012	Combination of reports from Oxford and ARMINES	R. Girard

STATUS, CONFIDENTIALITY, ACCESSIBILITY									
STATUS:				CONFIDENTIALITY:				ACCESSIBILITY:	
S0	Approved/Released	**X**		**R0**	General public	**X**		Private web site	
S1	Reviewed			**R1**	Restricted to project members			Public web site	**X**
S2	Pending for review			**R2**	Restricted to European Commission			Paper copy	
S3	Draft for comments			**R3**	Restricted to WP members + PL				
S4	Under preparation			**R4**	Restricted to Task members +WPL+PL				

1 **The authors of this document are solely responsible for its content, which does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of data appearing therein.**

2 **VERSION NAMING : V0.x** draft before peer-review approval, **V1.0** at the approval, **V1.x** minor revisions, **V2.0** major revision

Quantile forecasting of wind power using variability indices

Quantile forecasting of wind power using variability indices

Patrick McSharry[a,b], Georgios Anastasiades[a,b]

[a] *Oxford Centre of Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford, United Kingdom*

[b] *Smith School of Enterprise and the Environment, University of Oxford, Oxford, United Kingdom*

Abstract

Wind power forecasting techniques have received substantial attention recently due to the increasing penetration of wind energy in national power systems. While the initial focus has been on point forecasts, the need to quantify forecast uncertainty and communicate the risk of extreme ramp events has led to an interest in producing probabilistic forecasts. Using four years of wind power data from three wind farms in Denmark, we develop quantile regression models to generate short-term probabilistic forecasts from 15 minutes up to six hours ahead. More specifically, we investigate the potential of using various variability indices as explanatory variables in order to include the influence of changing weather regimes. These indices are extracted from the same univariate wind power series and optimized specifically for each quantile. The forecasting performance of this approach is compared with that of various benchmark models. Our results demonstrate that variability indices can increase the overall skill of the forecasts and that the level of improvement depends on the specific quantile.

Keywords: wind power forecasting, wind power variability, quantile forecasting, density forecasting, quantile regression, continuous ranked probability score, quantile loss function

1. Introduction

Wind power is one of the fastest growing renewable energy sources (Barton & Infield (2004)). However, due to the large variability of wind speed, and the unpredictable and dynamic nature of the earth's atmosphere,

Email addresses: patrick.mcsharry@smithschool.ox.ac.uk (Patrick McSharry), georgios.anastasiades@exeter.ox.ac.uk (Georgios Anastasiades)

there is great variability in wind power production. This inherent variability of wind speed is the main cause of the uncertainty observed in wind power generation. Recently, scientists have been attempting to, directly or indirectly, model this uncertainty and produce accurate forecasts of wind power production.

Since there is no efficient way to store wind energy, the wind power production decreases to zero if wind speed drops below a certain level. On the other hand, excessively strong winds can cause serious damage to the wind turbines, and hence they are automatically shut down at the 'disconnection speed', leading to an abrupt and quite unexpected drop of power generation. In addition, the wind power generated is limited by the capacity of each turbine. Therefore, it is quite important to produce accurate wind power forecasts for an efficient operation of wind turbines and reliable integration of wind power into the national grid.

The length of the relevant forecast horizons usually depends on the required application. For example, in order to schedule power generation, forecast horizons of several hours are usually enough, but for maintenance planning we need forecast horizons of several days or even weeks (McSharry, Pinson, & Gerard (2009)).

The literature of wind power forecasting starts with the work of Brown, Katz, & Murphy (1984) where they used autoregressive processes to model and simulate the wind speed, and then estimate the wind power by applying suitable transformations to values of wind speed. Most of the early and recent literature focuses on producing wind power point forecasts, directly, or indirectly in the sense that the focus is on modeling the wind speed and then transforming the forecasts through a power curve (Sanchez (2006), Taylor, McSharry, Member, IEEE, & Buizza (2009)).

Sanchez (2008) proposed an adaptive forecast combination procedure (AEC) that tends to be similar to the use of the best available predictor in a time varying environment. The proposed procedure was applied to two wind farms where alternative forecasts were available, showing the advantage of the proposed method. Jursa & Rohrig (2008) introduced a new short-term prediction method based on the application of evolutionary optimization algorithms for the automated specification of two well-known time series prediction models, namely, neural networks and the nearest neighbour search. The proposed method was used to produce wind power forecasts for a certain wind farm.

One of the most popular methods for producing point forecasts, is the use of time series models and especially the autoregressive integrated moving average (ARIMA) model, applied in the past by Nielsen (1999), Kariniotakis, Nogaret, & Stavrakakis (1997) and Makarov, Hawkins, Leuze, & Vidov (2002). Recent stud-

ies have also used the generalized autoregressive conditional heteroskedasticiry (GARCH) models for the conditional variance (Tol (1997), Taylor et al. (2009), Lau & McSharry (2010)). Tol (1997) fitted an autoregressive (AR)-GARCH model to daily Canadian mean wind speed series, and Taylor et al. (2009) an autoregressive moving average (ARMA)-GARCH model to the wind speed time series in the United Kingdom. Lau & McSharry (2010) used an ARIMA-GARCH process to model the aggregated wind power time series in Ireland.

Recent research has focused on producing probabilistic or density forecasts, because the point forecast methods are not able to quantify the uncertainty related to the prediction. Up to now, the number of studies on multi-step quantile/density forecasting is relatively small compared to point forecasting. Moeanaddin & Tong (1990) estimated densities using recursive numerical methods, which are quite computationally intensive. Pinson (2010), by applying a suitable transformation, managed to produce ten minutes ahead density forecasts at the Horns Rev wind farm in Denmark.

Taylor et al. (2009) used sophisticated time series models and weather ensemble[1] predictions to produce density forecasts for five wind farms in the United Kingdom. Also Lau & McSharry (2010) produced multi-step density forecast for the aggregated wind power series in Ireland, using ARIMA-GARCH models and exponential smoothing methods.

The quantile regression method, introduced by Koenker & Bassett (1978), has been extensively used to produce wind power quantile forecasts, using a variety of explanatory variables, among which, wind speed, wind direction, temperature and atmospheric pressure. Recent literature includes papers by Bremnes (2004), Nielsen, Madsen, & Nielsen (2006) and Moller, Nielsen, & Madsen (2008). Davy, Milton, Russell, & Coppin (2010), proposed a new variability index that is designed to detect rapid fluctuations of wind speed or power that are sustained for a length of time, and used it as an explanatory variable in the quantile regression model they constructed.

Bossavy, Girard, & Kariniotakis (2010) extracted two new indices that are able to recognize and predict ramp events in the wind power series, and used them to produce quantile estimates with the quantile regression forest method as their basic model. Finally, Gneiting (2011) studied the behavior of quantiles as optimal predictors and illustrated the relevance of decision theoretic guidance in the transition from a

[1] A relatively new approach for wind power forecasting is to use ensemble forecasts produced from numerical weather prediction (NWP) methods (Taylor et al. (2009) and Pinson & Madsen (2009)). Using this approach meteorological forecasts are transformed to wind power forecasts.

predictive distribution to a point forecast using the Bank of Englands density forecasts of United Kingdom inflation rates, and probabilistic predictions of wind energy resources in the Pacific Northwest.

In this paper we use univariate wind power series from three different wind farms in Denmark, to produce very short-term quantile forecasts, from 15 minutes up to six hours ahead. In order to produce quantile forecasts, we will use a quantile regression model, with explanatory variables extracted form the same wind power time series. More specifically, four new variability indices will be produced (extracted form the original wind power time series), which serve to capture the volatile nature of the wind power series. These indices, together with some lagged versions of the wind power series, will be used as explanatory variables in the quantile regression model. As for any regression model, we need predictions (point forecasts) for the future values of the explanatory variables, in order to produce future estimates. For the point forecasts we will use time-series models that are able to model both the mean and the variance of the underlying series.

The three Danish wind farms were chosen according to their monthly wind power capacity and standard deviation. We choose one high, one low, and one average variability wind farm in order to better understand the ability of each model to produce probabilistic forecasts under different circumstances.

The indices used will be independently optimized for each of the three wind farms, using a one-fold cross validation technique. In fact, two different optimizations will take place for each wind farm: The first one will aim to minimize the check function value produced by a 1-step ahead quantile regression forecast, for each of nineteen different quantiles. The second one will aim to minimize the average check function value, produced by taking the average over all 24 forecast horizons (equal to six hours), for each quantile. The final forecast results will be compared with those of some widely used benchmark models.

The remainder of the paper is presented as follows. In Section 2 we will introduce the wind power data, and in Section 3 the new variability indices will be derived. Section 4 will present the methodology behind quantile regression models with ways to evaluate the resulting quantile forecasts. In Section 5 we will optimize the variability indices using the in-sample data, and in Section 6 the out-of-sample forecast performance of the quantile regression models will be assessed. Section 7 will conclude the paper.

2. Wind power time series

We use wind power data recorded at three wind farms in Denmark summarized in Table 1. These wind farms were chosen to have different amounts of wind power variability, and have the smallest percentage of

missing values[2] compared to the one of other available wind farms.

Wind farm station name	Region	Wind power variability	No of turbines	Wind farm rated capacity (MW)
DØR	6	Low	1	1000
ALB	3	Medium	23	25500
VES	2	High	8	2195

Table 1: The Danish wind farms used in this study.

Our data sets contain wind power measurements recorded every 15 minutes for four years, from 01-Jan-2007 until 31-Dec-2010. The data of each wind farm is bounded between zero and the maximum capacity of the turbines. The zero value is attained in the case of excessively strong wind, where the turbines shut down in order to prevent them from damage. In order to facilitate comparisons between the data sets of different capacities, we normalize the wind power data of each wind farm by dividing by the total (rated) capacity. Hence, the data is now bounded within the interval [0,1].

We dissect the data of each farm into a set of exactly two years (2007 and 2008) for in-sample model training and calibration, and an out-of-sample testing set (the remaining two years) for out-of-sample testing and model evaluation. The in-sample set is dissected again into two sub-sets, a training set and a testing set. For the in-sample training set we use the first 1.5 years and for the in-sample testing set the remaining half year. This way we can use a *one-fold cross validation technique* to optimize the indices introduced in Section 3, and test the performance of our final chosen model using the out-of-sample testing set.

The time series plots for the year 2010, together with the monthly capacity and standard deviation indices are shown in Figure 1. The monthly capacity and variability indices were generated by taking the mean and standard deviation of wind power, respectively, for each month over the entire four year period. As we can clearly observe, the three wind farms have different wind power variability. More specifically, the first two wind farms have the highest and lowest possible wind power variability for all four years (from all the available wind farms in Denmark), without having any significant changes[3] in the capacity form year to year. The third wind farm was chosen to have an average (medium) variability compared to the other two farms, but with the same stationarity criterion for capacity over the four years.

[2]The missing values in the data sets were imputed using linear interpolation.

[3]Wind power variability may change form year to year by addition of new turbines or removal (maybe for maintenance) of existing ones.

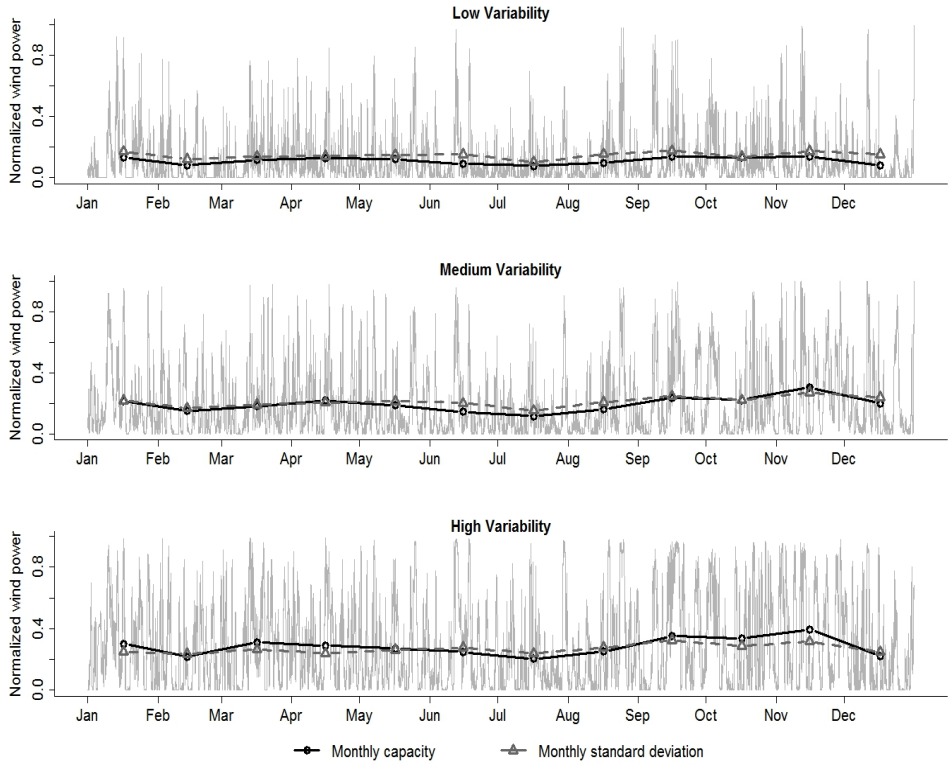


Figure 1: Time series plots of normalized power data for the three chosen Danish wind farms, for year 2010.

3. Indices of wind power variability

Davy et al. (2010), proposed a variability index that is design to detect rapid fluctuations of wind speed or power that are sustained for a length of time. They defined this variability index as the standard deviation of a band-limited signal in a moving window, and they constructed such an index for a wind speed time series. This variability index depends on four parameters: the order of the filter (integer greater than one), the upper and lower frequencies of the extracted signal, and the width of the moving window. We would like to use such an index as an explanatory variable in our quantile regression, but a proper optimization of this is too time consuming because of the number of the parameters involved.

Instead, we propose a parsimonious variability index, which depends on only two parameters, and is constructed as follows. Firstly we smooth our original wind power series using an averaging window of size m, in order to obtain the smoothed wind power series,

$$r_t = \frac{1}{m} \sum_{i=0}^{m-1} y_{t-i}, \tag{1}$$

starting from $t = m$. Note that this series behaves in a fully retrospective way, in the sense that each point of the series depends only on the historical values of the original series. Since the smoothed series is $m-1$ points smaller than the original series, we set $r_t = r_m$, for $t = 1, 2, ..., m-1$.

Finally, the new variability index is just the standard deviation of the extracted smoothed wind power series in a moving window of width n. So, if r_t is a given point of the smoothed series, we define the new index as

$$SD_t = \sqrt{Var(r_{t-n+1}, ..., r_{t-1}, r_t)}, \tag{2}$$

for $t \geq n$. Again, we impute the first $n-1$ points of the series by setting $SD_t = SD_n$, for $t = 1, 2, ..., n-1$. This index can be optimized much easier than the one proposed by Davy et al. (2010), since it has only two parameters: the smoothing parameter m, and the variability parameter n .

By similar reasoning, we create another three variability indices. We smooth the original wind power series as described above, and then instead of finding the standard deviation, we find the interquartile range (IQR), the 5% and the 95% quantiles of the smoothed series over a moving window (different for each series). These three new indices can be defined as:

$$IQR_t = IQR(r_{t-n+1}, ..., r_{t-1}, r_t), \tag{3}$$

$$\mathbb{P}\left((r_{t-n+1}, ..., r_{t-1}, r_t) < Q05_t\right) = 0.05, \tag{4}$$

$$\mathbb{P}\left((r_{t-n+1}, ..., r_{t-1}, r_t) < Q95_t\right) = 0.95, \tag{5}$$

for $t \geq n$. We also impute their values for $t = 1, ..., n-1$ as described above. An example of the construction of the three variability wind power indices is shown in Figure 2. These indices will be properly optimized, and together with some lagged values of the original power series will be used as explanatory variables in the quantile regression introduced in the next section.

4. Quantile regression, forecasting, and evaluation methodology

4.1. Quantile regression

Given a random variable[4], y_t, and a strictly increasing CDF, $F_t(y)$, the α_i-quantile, $q_t^{(\alpha_i)}(y)$, with proportion $\alpha_i \in [0, 1]$ is defined as the value for which the probability of obtaining values of y_t below $q_t^{(\alpha_i)}$ is α_i:

$$\mathbb{P}\left(y_t < q_t^{(\alpha_i)}\right) = \alpha_i \quad or \quad q_t^{(\alpha_i)} = F_t^{-1}(\alpha_i) \tag{6}$$

[4]The notation y_t, is used for denoting both the stochastic state of the random variable at time t, and the measured value at that time.

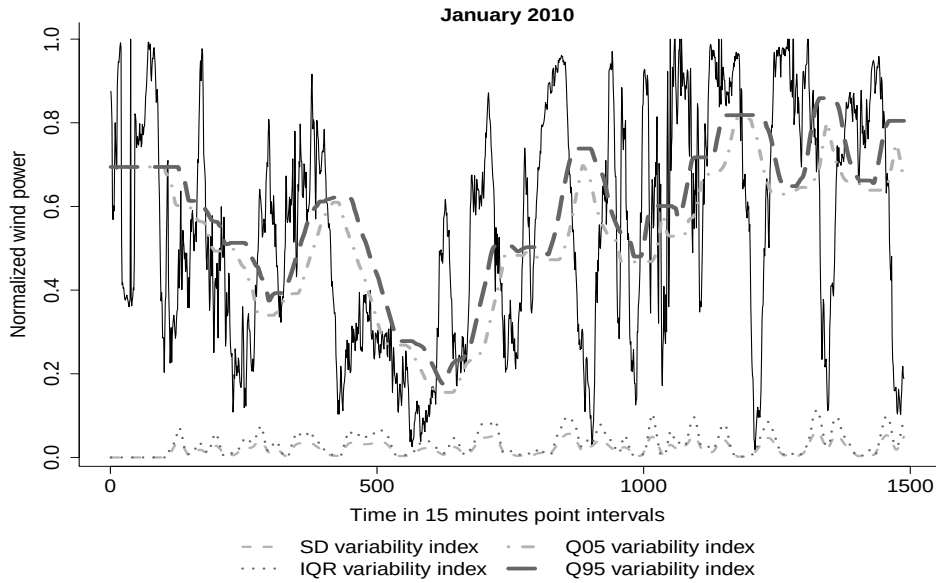


Figure 2: Wind power time series plot of the medium variability farm, together with the four variability indices. The parameters are chosen to be the same for all indices just for comparison ($m = 100$ and $n = 30$).

The quantile regression, introduced by Koenker & Bassett (1978), models $q_t^{(\alpha_i)}$ for $\alpha_i \in [0, 1]$, as a linear combination of some given explanatory variables. This is quite similar to multiple linear regression where the mean is modeled in exactly the same way. So, the α_i-quantile is modeled as:

$$\begin{aligned} q_t^{(\alpha_i)} &= \gamma_0^{(\alpha_i)} + \gamma_1^{(\alpha_i)} x_{t,1} + ... + \gamma_p^{(\alpha_i)} x_{t,p} \\ &= \gamma_0^{(\alpha_i)} + \sum_{j=1}^{p} \gamma_j^{(\alpha_i)} x_{t,j}, \end{aligned} \tag{7}$$

where $\gamma_j^{(\alpha_i)}$ are unknown coefficients depending on α_i, and $x_{t,j}$ are the p known regressors. In quantile regression, a regression coefficient estimates the change in a specified quantile of the response variable produced by a one unit change in the corresponding predictor variable.

We define the *quantile loss function*, also known as the *check function*, for a given proportion $\alpha_i \in [0, 1]$ as

$$\begin{aligned} \rho_{\alpha_i}(u) &= \big(\alpha_i - \mathbb{1}_{\{u<0\}}\big)u \\ &= \begin{cases} \alpha_i u, & u \geq 0, \\ (\alpha_i - 1)u, & u < 0, \end{cases} \end{aligned} \tag{8}$$

where u is a given function. Then, the sample α_i-quantile can be estimated by minimizing $\sum_{t=1}^{T} \rho_{\alpha_i}(y_t - q)$ with respect to q. Hence, we can estimate the unknown coefficients, $\gamma_j^{(\alpha_i)}$, by replacing q with the right-hand side of equation (7):

$$\hat{\gamma}^{(\alpha_i)} = \underset{\gamma}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_{\alpha_i}\big\{y_t - (\gamma_0 + \gamma_1 x_{t,1} + ... + \gamma_p x_{t,p})\big\}, \tag{9}$$

where $\hat{\gamma}^{(\alpha_i)}$ is a vector containing the unknown coefficients. Usually, these estimates are calculated using linear programming techniques as in Koenker & D'Orey (1987).

In this paper we will use quantile regression to forecast the values of quantiles with nominal proportion $\alpha_i = \{0.05, 0.01, ..., 0.95\}$, for forecast horizons[5] $k = 1, 2, ..., 24$. We denote the forecast for the quantile with nominal proportion α_i issued at time t for forecast time $t + k$, by $\hat{q}_{t+k|t}^{(\alpha_i)}(y)$. In order to produce these forecasts, we use equation (7), and the estimated coefficients, $\hat{\gamma}^{(\alpha_i)}$:

$$\begin{aligned} \hat{q}_{t+k|t}^{(\alpha_i)}(y) &= \gamma_0^{(\alpha_i)} + \gamma_1^{(\alpha_i)} \hat{x}_{t+k|t,1} + ... + \gamma_p^{(\alpha_i)} \hat{x}_{t+k|t,p} \\ &= \gamma_0^{(\alpha_i)} + \sum_{j=1}^{p} \gamma_j^{(\alpha_i)} \hat{x}_{t+k|t,j}, \end{aligned} \tag{10}$$

where $\hat{x}_{t+k|t,j}$ for $j = 1, ..., p$ denote the forecasts of the explanatory variables $\mathrm{x}_{t,j}$, issued at time t with lead time $t + k$.

The random variable y_t will represent the normalized wind power time series, (y_t), and the explanatory variables will be represented by time series, $(x_{t,j})$, extracted from the normalized wind power series. So, in order to produce the forecasts, $\hat{x}_{t+k|t,j}$, we will fit suitable time series models to the variables $(x_{t,j})$, and then predict from these models up to t+k values ahead.

4.2. Quantile forecast evaluation

The evaluation of the quantile forecasts, for each quantile, $\alpha_i = \{0.05, 0.01, ..., 0.95\}$, will be done using the quantile loss function:

The ***quantile loss function*** (Koenker & Bassett (1978)), also known as the ***check function*** is used to define a specific quantile of the distribution and was defined in Section 4, equation (8) .

[5]The forecast horizon, k, is measured in time steps of 15 minutes.

Hence, we can estimate a particular quantile, $\hat{q}^{(\alpha_i)}$, with proportion α_i, using

$$\hat{q}^{(\alpha_i)} = \min_q \sum_{t=1}^{N} \rho_{\alpha_i}(y_t - q), \tag{11}$$

and therefore we can evaluate a series of quantile forecasts, $\hat{q}^{(\alpha_i)}_{t+k|t}$, issued at time t with lead time $t+k$ and nominal proportion α_i, using:

$$QL(k, \alpha_i) = \frac{1}{N} \sum_{t=1}^{N} \rho_{\alpha_i}(y_{t+k} - \hat{q}^{(\alpha_i)}_{t+k|t}). \tag{12}$$

.

Using the different quantile forecasts we can also reconstruct the whole probability / cumulative forecasted distribution. We use the ***Continuous Ranked Probability Score (CRPS)*** in order to evaluate the the density forecasts for each forecast horizon:

The CRPS (Matheson & Winkler (1976)) is computed by taking the integral of the Brier scores for the associated probability forecasts at all real valued thresholds,

$$crps(\hat{F}_{t+k|t}(y), y_{t+k}) = \int_{-\infty}^{+\infty} \big(\hat{F}_{t+k|t}(y) - \mathbb{1}_{\{y \geq y_{t+k}\}}\big)^2 \, \mathrm{d}y \tag{13}$$

$$= \int_0^1 QS_{\alpha_i}\big(\hat{F}^{-1}_{t+k|t}(\alpha_i), y_{t+k}\big) \, \mathrm{d}\alpha_i, \tag{14}$$

where $\hat{F}_{t+k|t}(y)$ corresponds to the CDF forecast, and y_{t+k} to the corresponding verification. $\mathbb{1}_{\{y \geq y_{t+k}\}}$ is an indicator function that equals one if $y \geq y_{t+k}$ and zero otherwise. The quantile score, QS_{α_i}, (Gneiting & Raftery (2007)) is defined by

$$QS_{\alpha_i}(q, y) = 2\big(\alpha_i - \mathbb{1}_{\{y<q\}}\big)\big(y - q\big). \tag{15}$$

Hence, the average of these $crps$ values over each forecast-verification pair gives the CRPS for each forecast horizon k:

$$\text{CRPS}(k) = \frac{1}{N} \sum_{t=1}^{N} crps(\hat{F}_{t+k|t}(y), y_{t+k}) \tag{16}$$

$$= 2 \int_0^1 QL(k, \alpha_i) \, \mathrm{d}\alpha_i, \tag{17}$$

where $QL(k, \alpha_i)$ is the mean quantile loss function defined in equation (12). Representation (17) will be particularly useful because we will derive the CRPS using the loss function value for each quantile.

5. Optimization of the variability indices

Our main goal is to check whether or not the four variability indices (introduced in Section 3) can help to provide trustworthy quantile forecasts of wind power, when used as explanatory variables in the quantile regression model (7). For this purpose, we have to optimize the two parameters (m, n) of these indices using the following procedure.

For each index, we sample different combinatios of parameters from the range $m, n = \{0, 8, 16, ..., 192\}$, in order to produce 625 different realizations of each index, for each wind farm. A preliminary analysis showed that creating a moving window larger than 192 time-points wide (2880 minutes) did not increase the performance of the indices.

Then, for each set of parameters, we fit the following four different quantile regression models on the in-sample training set (of each wind farm), with the normalized wind power series, (y_t), as the response, for each of the 19 quantiles $\alpha_i = \{0.05, 0.1, ..., 0.95\}$:

$$\begin{aligned}
\text{SD model:} \quad & q_t = \gamma_{01} + \gamma_{11} y_{t-1} + \gamma_{21} y_{t-2} + \gamma_{31} y_{t-3} + \gamma_{41} SD_t \\
\text{IQR model:} \quad & q_t = \gamma_{02} + \gamma_{12} y_{t-1} + \gamma_{22} y_{t-2} + \gamma_{32} y_{t-3} + \gamma_{42} IQR_t \\
\text{Q05 model:} \quad & q_t = \gamma_{03} + \gamma_{13} y_{t-1} + \gamma_{23} y_{t-2} + \gamma_{33} y_{t-3} + \gamma_{43} Q05_t \\
\text{Q95 model:} \quad & q_t = \gamma_{04} + \gamma_{14} y_{t-1} + \gamma_{24} y_{t-2} + \gamma_{34} y_{t-3} + \gamma_{44} Q95_t,
\end{aligned} \tag{18}$$

where $q_t \equiv q_t^{(\alpha_i)}$ is defined in (6), $\gamma_{hl} \equiv \gamma_{hl}^{(\alpha_i)}$ are the regression coefficients, and y_{t-j} are lagged wind power series. The choice of using three lagged series of the response as explanatory variables was taken after many trials of different numbers of lags for all three wind farms.

Next, we produce point forecasts from 15 minutes up to six hours ahead[6], from each point of the in-sample testing set, by fitting ARIMA$(1, 1, 1)$ models to each realization of the four variability indices of the above regressions. The point forecasts will be used as predictors to produce the quantile forecasts for the normalized wind power time series, as explained in Section 4.1. Our choice of ARIMA$(1, 1, 1)$ model was made mainly for simplicity, after exploring the forecast performances of various time series models. Fitting a different model for each set of parameters is very computationally expensive, and in this case the sacrificed accuracy is very small.

[6] i.e. with forecast horizon $k = 1, 2, ..., 24$.

Moreover, modeling the variance of the indices using ARCH/GARCH models (in combination with an ARIMA model for the mean), does not provide a consistent and significant improvement of the RMSE[7] of the point forecasts. This is mainly because of the very small forecast horizon we have, and hence it suffices to use a simple ARIMA model.

In order to produce point forecasts of the lagged wind power series, model solution using BIC identified an ARIMA$(0, 1, 2)$ - GARCH$(1, 1)$ model for the low variability farm, an ARIMA$(1, 1, 3)$ - GARCH$(1, 1)$ for the medium variability farm, and an ARIMA$(2, 1, 1)$ - GARCH$(1, 1)$ for the high variability farm. These models have the ability to capture the heteroskedastic effects that the wind power series have, taking into account the non-linear nature of the variations. Also, these forecasts are calculated only once for all different realizations of the quantile regression models, and hence there is no point in this case to sacrifice the (small) accuracy gain for simplicity and computational efficiency.

After producing quantile forecasts for 24 different forecast horizons, we evaluate them (i) using the check function value averaged over all forecast horizons, and (ii) using the check function value of only the first step ahead forecasts. The results justify our inspection of better forecast performance for the models with small moving windows. We repeat the above procedure by restricting our parameters' range even more for each variability index, and sample every different combination of parameters from the range $m, n = \{0, 1, 2, ...50\}$.

We end up with distinct sets of parameters (for each model and wind farm) that minimize the averaged and 1-step ahead check function value of each different quantile. The check function minimization results are shown in Tables 3 and 4 of the Appendix. In general, we can not distinguish any particular parameter pattern, but there are some features that are worth mentioning. For all the models, it is more common to have the smoothing window width (m) smaller than the variability window width (n), especially for quantiles less or equal to the median. This pattern changes for the upper quantiles (larger than the median) were we do not observe a clear pattern. Also, on average, the parameters for the 24-steps optimization are smaller than the corresponding ones of the 1-step optimization.

Finally, using (17), we gather two more sets of parameters (for each model and wind farm) that minimize (i) the total CRPS averaged over all forecast horizons, and (ii) the total CRPS of the first step ahead forecasts. Using these we would like to check whether or not there is an improvement in the overall skill of the models

[7] We used the Root Mean Square Error to evaluate the point forecast performance of various time series models.

when treating each quantile separately. The CRPS minimization results are shown in Table 2 of the Appendix. Again, on average, the parameters for the 24-steps optimization are smaller than the corresponding ones of the 1-step optimization, and the smoothing window width is smaller than the variability window.

6. Out-of-sample forecast performance results

In this section we will fit the optimized models of (18) to the whole two years in-sample set (of each farm), produce quantile forecasts from 15 minutes up to six hours ahead from each point of the out-of-sample set, and assess their forecast performance using the CRPS, check function and reliability diagrams.

Firstly, it will be useful to check the R-squared value of the four fitted models for each quantile. Essentially, by using quantile regression we fit a linear regression line for each quantile, and hence the R-squared value will tell us the proportion of total variance explained by each model, for each quantile. Of course this is an in-sample calculation, but we expect to get roughly an idea of how good the models are performing. Figure 3 shows the R-squared values for all models and wind farms when we optimize the models using the check function value of each quantile. A general observation is that we have relatively high R-squared values (> 0.5) for all quantiles, confirming the ability of the models to explain a large proportion of the total variability. Also the variance explained is much larger for the upper quantiles than the lower quantiles, mainly because of the asymmetric shape of the wind power distribution.

Some models for some specific wind farms and quantiles close to the median, have their R-squared values close to one. If we also examine these models' optimal parameters (Tables 3, 4), we observe that when the R-squared is close to one, both parameters are chosen to be zero. In other words, the variability indices of these models are identical to their original wind power series. When this happens the regression coefficients of the lagged series are almost zero, and the coefficient of the corresponding variability index is almost equal to one (see Figure 8 of the Appendix). This is perfectly acceptable since for a quantile regression model, the response is a specific quantile of a series (or data points) and not the series itself. Intuitively, if the specific quantile is close to the mean value of the series, then the check function is minimized by this mean value. Hence, the optimization forces the parameters to be zero, and consequently forecasts this quantile by predicting the mean of the series itself.

A final observation is that the R-squared values of the low variability wind farms are significantly smaller than the corresponding ones of the other farms. This is reasonable as the variability indices are used to

capture the different regime changes of the series, and hence we expect our models to explain a larger proportion of the total variance when the variability is higher.

Figure 4 shows the analogous R-squared values when the models are optimized based on the CRPS (1-step or averaged 24-steps). The observations are very similar to the ones stated above, with the difference that now the plots are much smoother (since we don't treat each quantile separately), without any values close to 1.

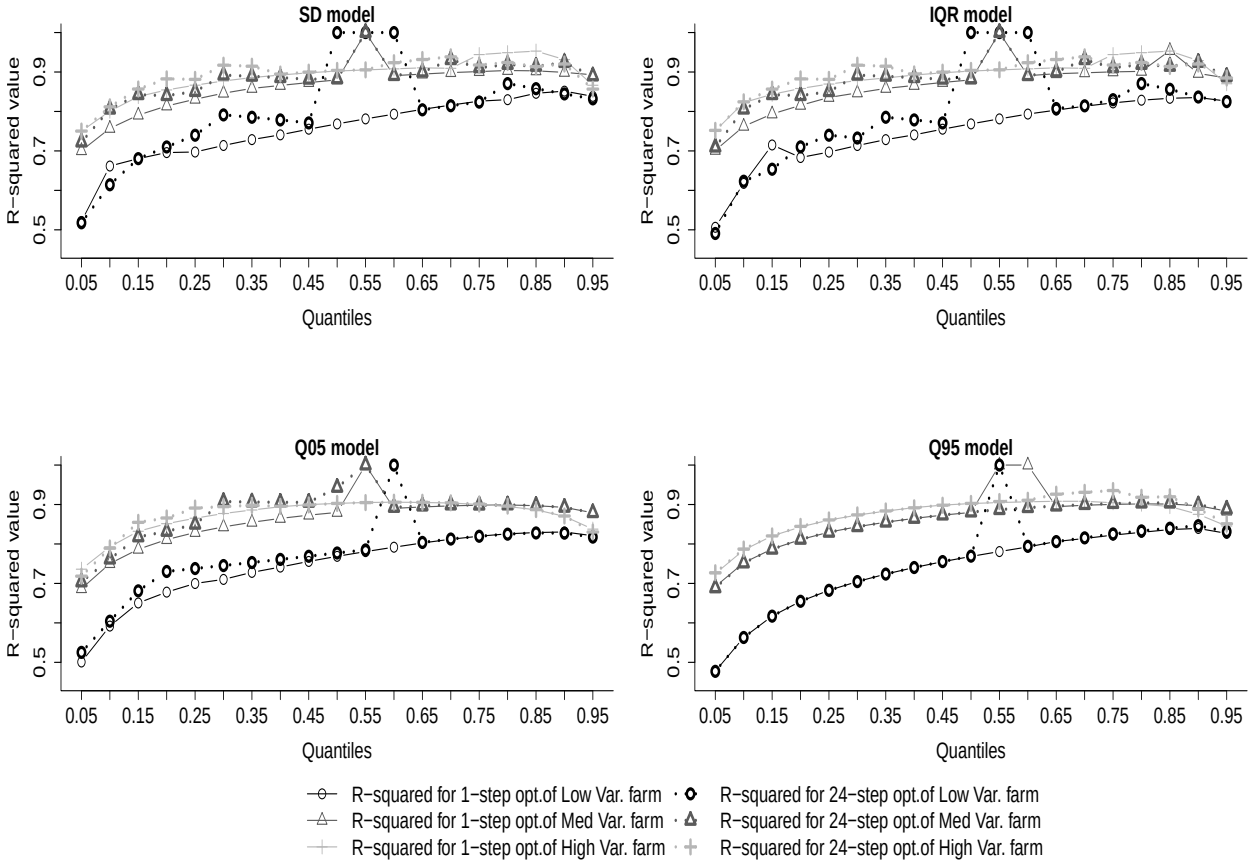


Figure 3: R-squared values for models with 1-step check function optimization and averaged over 24-steps check function optimization.

Next, we produce quantile forecasts from each point of the out-of-sample set (final two years), from 15 minutes up to six hours ahead. This is achieved for all the models and for every wind farm. The forecast performance of each model is assessed using the evaluation methods of Section 4.2. In order to facilitate the comparison of forecast performance across different models, we will introduce two widely used *probabilistic benchmarks*:

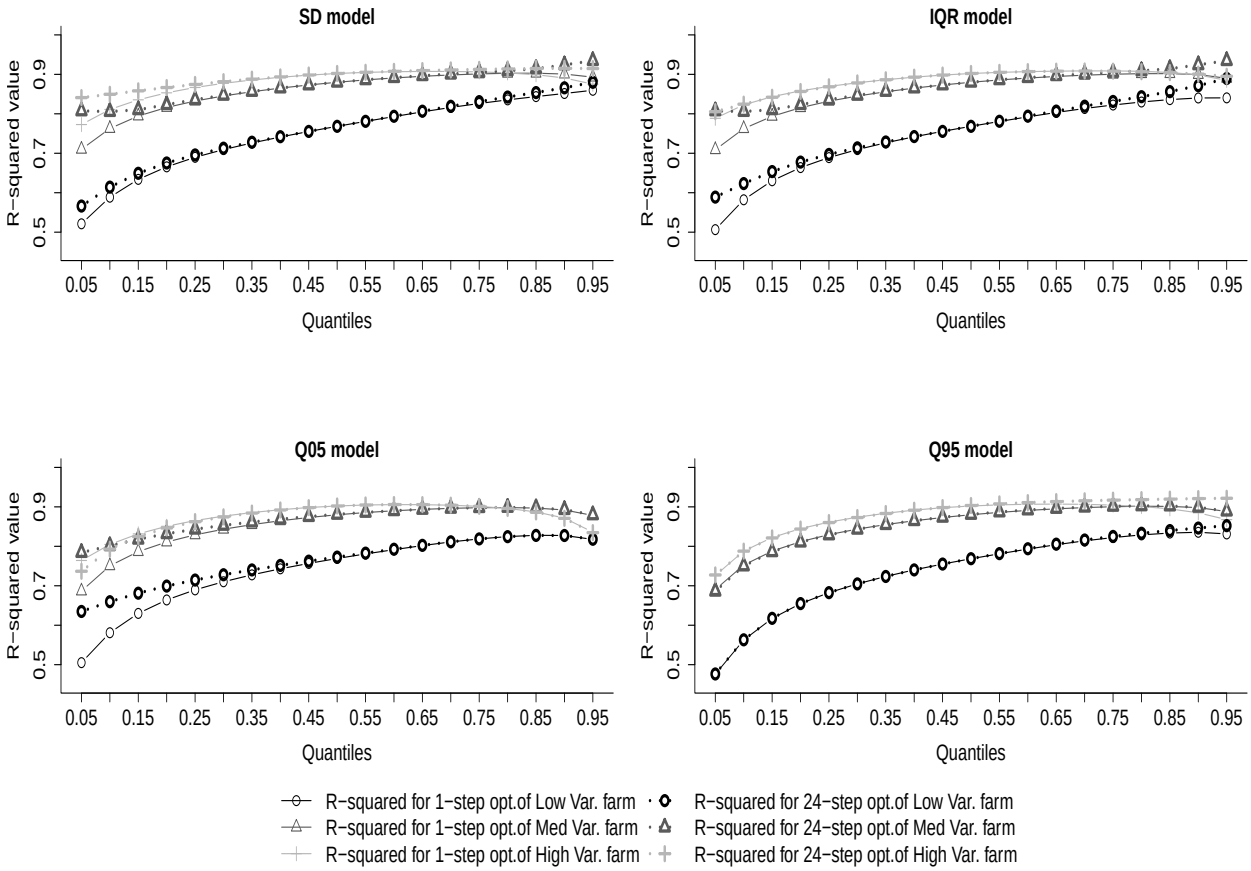


Figure 4: R-squared values for models with 1-step CRPS optimization and averaged over 24-steps CRPS optimization.

- *Persistence distribution*: It is defined as the distribution of the last n observations. The persistence benchmark is independently optimized for each wind farm, by using the same optimization methods as for the variability indices: 1-step check function, average of 24-steps check function, 1-step CRPS and average of 24-steps CRPS minimization.
- *Unconditional distribution*: We construct this benchmark by using all the past observations of the time series. This benchmark assumes that the time ordering of the observations is not relevant when attempting to predict the distribution the response. It is also referred to as *climatology*.

The third benchmark used in this paper is the quantile regression model with only the three lagged series as explanatory variables. This benchmark will help us to identify the gain in forecast performance acquired by using the four variability indices. In the next figures we will present the evaluation results of the four models, and include only the best model(s) for each evaluation criterion. In case that the results are quite similar

across wind farms we will present the averaged results[8] over the three wind farms. Also, it will be necessary to quantify the gain of some forecasting models to a chosen reference system. Following McSharry et al. (2009), this gain, denoted as an improvement with respect to the considered reference forecasts system, is called a *Skill Score* and is defined as:

$$SkillScore(k) = \frac{SCORE_{\text{ref}}(k) - SCORE(k)}{SCORE_{\text{ref}}(k)} = 1 - \frac{SCORE(k)}{SCORE_{\text{ref}}(k)}, \tag{19}$$

where k is the lead time of the forecast and $SCORE$ is considered the evaluation criterion (such as CRPS or quantile loss function score).

6.1. Out-of-sample model comparison and evaluation

Firstly we would like to check the density forecasting performance of the competing models. Figures 5 (a), (c) and (e) show the out-of-sample normalized CRPS for the three wind farms, where the competing models were optimized by minimizing the averaged over 24-steps check function. The plots show only the best two models, namely, the SD and IQR models, together with the three benchmarks. Th subplots show the Skill CRPS of the four competing models, with the three lagged series as the reference model.

A first observation is that the CRPS values are getting larger as the variability of the series increase, which is reasonable because intuitively it is easier to forecast a low variability series than a high one. On the other hand, the two best models perform better for the high variability wind farm. They perform better in the sense that they manage to outperform their competing benchmarks (especially climatology), with greater score difference for each forecast horizon, than the lower variability farms' models. The subplots inside each figure help us to compare the four models. The SD and IQR models do not produce very promising density forecasts for the first lead time, but their performance increases and outperform all the other models after approximately two time steps, for all wind farms.

Figures 5 (b), (d), (e) show the Skill CRPS of the 1-step check function optimization models, with the corresponding (averaged over) 24-steps optimization models as reference, for each wind farm. These Skill CRPS plots can help us to compare the two ways of optimizations carried out in this paper. The SD and IQR models produced by the 1-step optimization procedure perform much better when forecasting

[8]The averaging of the results is reasonable if the relative performances of the methods are similar for each of the three locations (Taylor et al. (2009)).

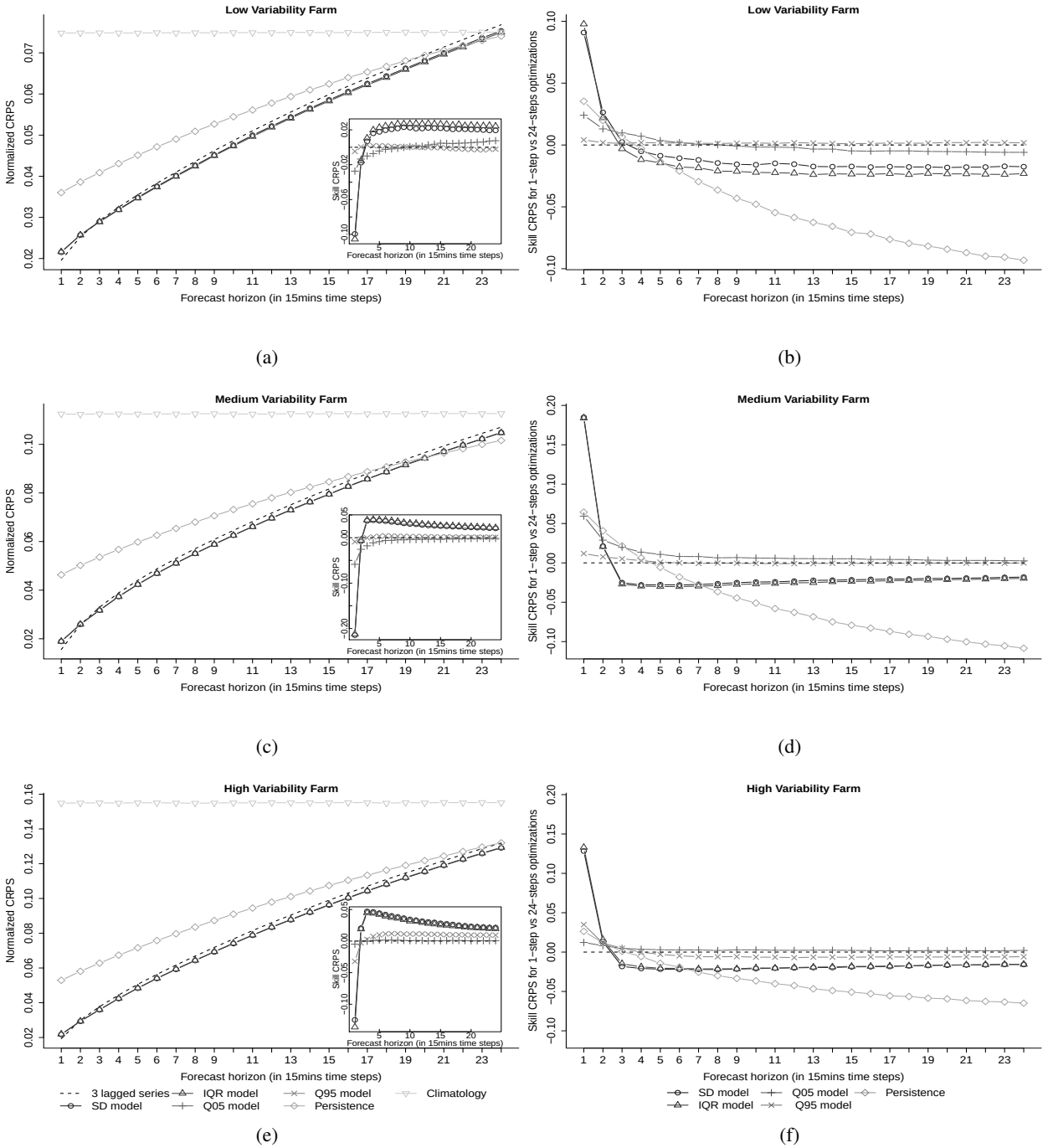


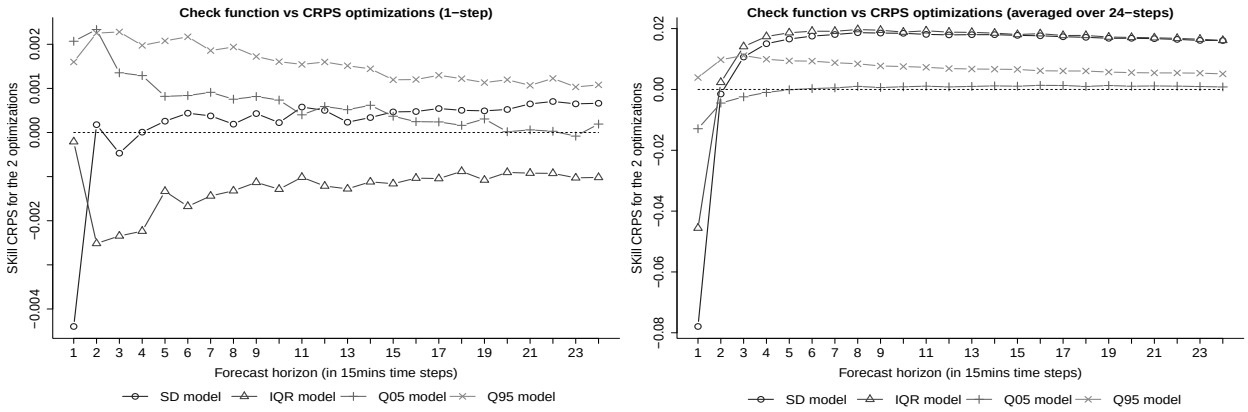
Figure 5: (a), (c), (e): Normalized CRPS for all three wind farms showing the three benchmarks and the best model(s). The subplots show the Skill CRPS of our competing models, with the three lagged series as the reference model. These results are outcome of the 24-steps check function optimization procedure, for each quantile. (b), (d), (f): Skill CRPS of the 1-step optimization models, with the corresponding 24-steps optimization models as reference.

the first two lead times. They can perform up to 17% better than the corresponding models of the 24-steps optimization. The Q05 and Q95 models also perform better when they are optimized using the 1-step optimization procedure.

For forecast horizons greater than 30 minutes (two time steps) the averaged over 24-steps SD and IQR models, produce density forecasts with skill gain up to 3%, compared to the analogous models of the 1-step optimization. On the other hand the Q05 and Q95 models behave almost identically with the corresponding 24-steps optimized models. What is also noteworthy is the huge gain in forecast performance of the persistence benchmark for large forecast horizons, when optimized using the 24-steps optimization method. The 1-step ahead optimized persistence benchmark may outperform the corresponding 24-steps optimized one for up to (maximum) three forecasted time steps, but then it is massively outperformed by the second one.

Finally we would like to check if there is any gain in the models' density forecasting performance when we treat each quantile separately. In other words, we want to check whether or not the resulting models of the (1-step or 24-steps) check function optimization outperform the corresponding ones of the CRPS optimization. Figure 6 (a) shows the Skill CRPS of the optimized models using the 1-step ahead check function optimization, with the corresponding 1-step CRPS optimized models as reference. The scores shown are averaged over all wind farms, because the relative performance of the methods are similar for each location. The performance gain (or loss) is maximum 0.2%, and hence the effort spent to optimize each models' quantile individually does not seem to be worthwhile. On the other hand, the SD and IQR models resulting from a 24-steps check function optimization, seem to outperform the corresponding CRPS optimized models by roughly 2% for forecast horizons greater than 30 minutes, and be outperformed by maximum 8% for the first lead time (Figure 6 (b)). The Q05 and Q95 models again have identical density forecasting performance for both optimization criteria.

Next, we briefly summarize the results found using the CRPS criterion. The best density forecasting models are the SD and IQR models, without any distinguishable -in terms of performance- difference between them. Both models can outperform all used benchmarks for forecast horizons up to (at least) 20 time steps ahead, for all three different variability wind farms. There are two more questions that we need to answer: what type of optimization is the best (1-step or 24-steps), and is it worthy to treat each quantile separately (by using check function instead of CRPS minimization)? The results show that if somebody is interested in forecasting only (up to) two time steps ahead ($\leq$ 30 minutes), then it is recommended to use one of the



(a) (b)

Figure 6: (a) Skill CRPS of the 1-step check function optimization models, with the corresponding 1-step CRPS optimization models as reference. (b) Skill CRPS of the 24-steps check function optimization models, with the corresponding (averaged over) 24-steps CRPS optimization models as reference. Note: Results are averaged over all wind farms.

SD or IQR models, optimized using the 1-step CRPS minimization. This is a pretty fast optimization as all parameters are the same for all the quantiles. On the other hand, in order to predict lead times larger than two time steps, it is recommended to optimize the SD or IQR models using an averaged over 24-steps check function minimization, for each quantile.

Now, we would like to compare the forecasting performance of each model, for some specific quantiles. For the purpose of these evaluations we will use the quantile loss function. Since all the models have quite similar performance for all wind farms, we will present the averaged results over the three wind farms. Moreover, for forecasting a specific quantile there is no advantage in using the CRPS optimization procedure, and hence the results presented will be only for the check function optimized models, for the specific quantiles. The chosen quantiles are the 0.05, 0.25, 0.75 and 0.95 quantiles. The 0.05 and 0.95 quantiles form the two tails of the forecasted density, and represent the rare events (such as ramps, cut offs) of a wind power series.

Figure 7 shows the Skill check function scores of the four models, with the three lagged series as the reference model. The subplots show the normalized check function score of the best model(s), together with persistence and climatology benchmarks. Note that these plots are outcomes of the averaged over 24-steps check function optimization. A first noteworthy observation is about the performances of the Q05 and Q95 models, for the lower and upper quantiles. We see that regardless of the Q95 model's structure, it is the

most appropriate for producing quantile forecasts for the 0.05 quantile for all forecasts horizons (Figure 7 (a)). In contrast, the Q05 model is the most appropriate for producing quantile forecasts (for $k > 9$) of the upper quantile, with the Q95 model performing the worst (Figure 7 (d)).

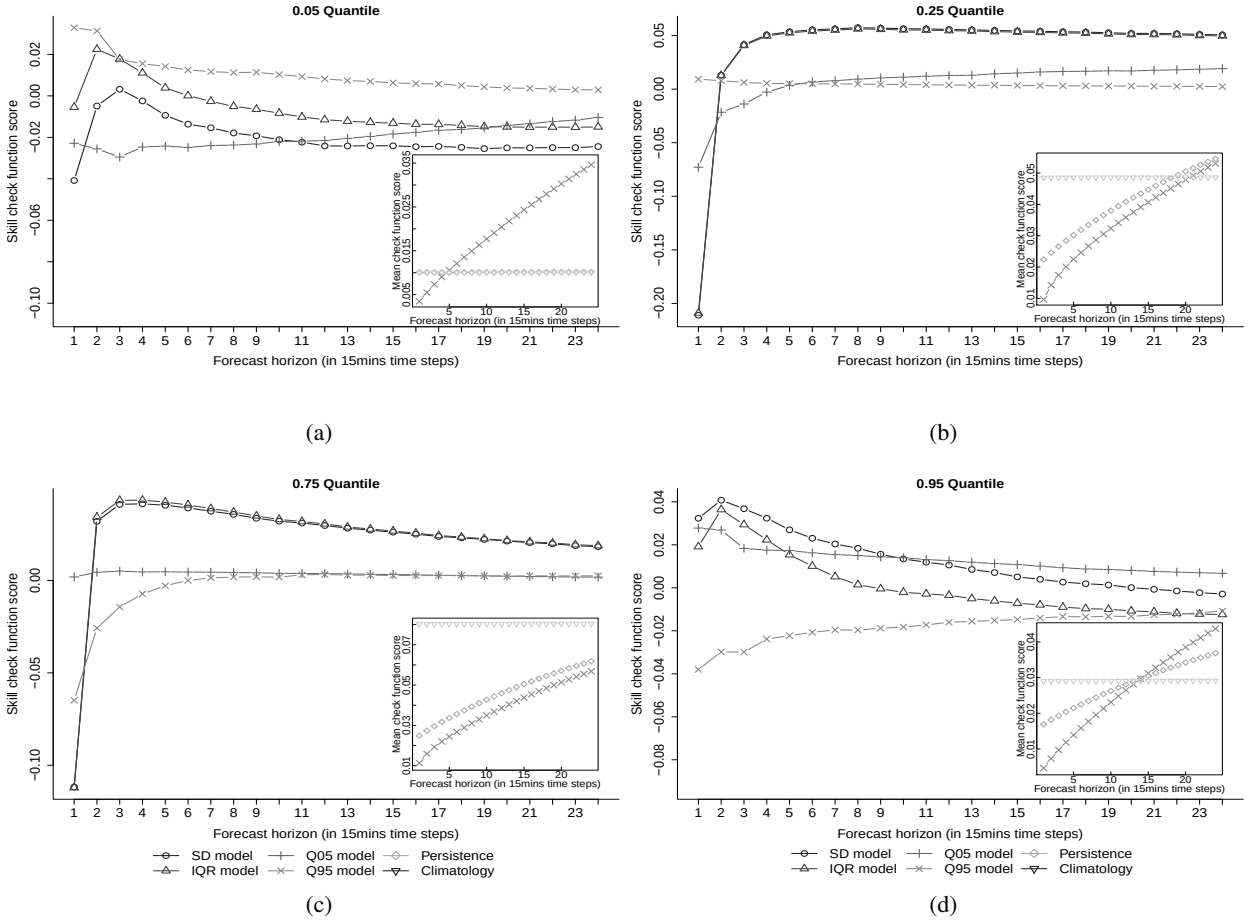


Figure 7: Skill check function scores, averaged over the three wind farms, with the three lagged series as the reference model. The subplots show the mean check function score of the best model(s), together with persistence and climatology benchmarks. These results are outcome of the (averaged over) 24-steps check function optimization procedure, for each quantile.

For forecasting all the other quantiles (including the ones whose plots are not included), the SD and IQR seem to be the most appropriate, without again any distinguishable -in terms of performance- difference between them. Figures 7 (b) and (c) show that the 1-step ahead forecast performances of these two models are not as good as the other models' (similarly for the density forecasting evaluation before). By using a 1-step check function minimization, we manage to increase their forecasting performance for the first lead

time, but produce higher check function scores for all the other forecast horizons. Since the CRPS arises from the check function scores of the different quantiles, this result was expected (and therefore there is no need to present any quantile forecast results for the 1-step ahead optimization).

Finally, by having a look at the different subplots of Figure 7 we observe that the models outperform the persistence and climatology benchmarks only for the middle quantiles. For the lower quantile, the Q05 model outperforms these two benchmarks only for the first four lead times (up to one hour). Due to the asymmetric shape of the wind power distribution, our best models for the upper quantile outperforms the two benchmarks for up to 14 time steps ahead. For the 0.30 up to 0.85 quantiles, the SD and IQR models outperform the benchmarks for all 24 forecast horizons.

7. Conclusions and discussions

In this paper we showed how to produce wind power quantile and density forecasts, for lead times from 15 minutes up to six hours ahead, using a single univariate wind power series. In order to do that we introduced some useful variability indices, which are able to capture the volatile behavior of the wind power series.

We used linear quantile regression as our main tool for producing quantile forecasts for 19 different quantiles, with three lagged versions of the wind power series as the main explanatory variables. Four models were created, each one having as a fourth explanatory variable one of the four extracted variability indices.

In order for the final results to be consistent, we used data from three wind farms in Denmark, each one chosen to have different wind power variability (low, medium and high). We used in total four years of data, with a 15 minutes resolution, for each wind farm. The first two years were used for in-sample model calibration and training, and the final two years for out-of-sample forecasting and evaluation.

All four quantile regression models were optimized in the in-sample set, in order to find their specific set of parameters which minimize the check function score at each quantile. Each optimization run that produces a check function score for a specific quantile is quite computationally expensive, so we chose to find the global minimum in a two-dimensional grid of parameter values. In fact this approach gave us the opportunity to try four different optimizations without any extra cost: 1-step ahead check function minimization, average over 24 forecast horizons check function minimization, 1-step ahead CRPS minimization, average over 24 forecast horizons CRPS optimization. The first two give a unique set of parameters for each quantile, and the last two a set of common parameters for all quantiles.

Our main goal was to check how well these models performing compared to the persistence and climatology probabilistic benchmarks. It is worth mentioning that persistence is a very strong and simple benchmark for very short forecast horizons, and was optimized using the same cost (optimization) functions as the four regression models. The density forecasts of the models were evaluated using the CRPS. The best two models found, are the SD and IQR models. They managed to outperform the benchmarks for most of the forecasts horizons. The models of the high variability wind farm, outperform the benchmarks for all forecast horizons and with greater score differences compared to the other two farms.

The above best two models, optimized using the 1-step ahead check function minimization, outperform the models of the 24-steps check function minimization only for the first lead time. Moreover, the 1-step CRPS optimized models do not seem to have different forecasting performance for the corresponding check function optimized models. If one is interested only in the first lead time, we recommend using the 1-step CRPS optimization, as it produces models with a single set of parameters for all quantiles. On the other hand, for larger forecast horizons, the 24-steps check function optimization is more appropriate.

Acknowledgments

The authors would like to thank Energinet.dk for data provision and support. This work has been partly supported by the European Commission under the SafeWind project (ENK7 - CT2008 - 213740), Her Majestys Government and an IBM Innovation Award.

References

Barton, J. & Infield, D. (2004). Energy storage and its use with intermittent renewable energy. *Energy Conversion, IEEE Transactions on*, *19*(2), 441 – 448.

Bossavy, A., Girard, R., & Kariniotakis, G. (2010). Forecasting uncertainty related to ramps of wind power production.

Bremnes, J. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind energy*, *7*, 47–54.

Brown, B. G., Katz, R. W., & Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of Applied Meteorology*, *23*, 11841195.

Davy, R., Milton, J., Russell, C., & Coppin, P. (2010). Statistical downscaling of wind variability from meteorological fields. *Boundary-Layer Meteorol*, *135*, 161–175.

Gneiting, T. (2011). Quantiles as optimal point forecasts. *International Journal of Forecasting*, *27*(2), 197 – 207.

Gneiting, T. & Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Jursa, R. & Rohrig, K. (2008). Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. *International Journal of Forecasting*, *24*(4), 694 – 709.

Kariniotakis, G., Nogaret, N., & Stavrakakis, G. (1997). Advance short term forecasting of wind power production. Energy Conference, Dublin Castle, Ireland, EWEA.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46(10)*, 33–50.

Koenker, R. & D'Orey, V. (1987). Computing regression quantiles. *Applied Statistics*, *36*, 383–393.

Lau, A. & McSharry, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *Annals of Applied Statistics*, *4*, 1311–1341.

Makarov, Y., Hawkins, D., Leuze, E., & Vidov, J. (2002). California iso wind generation forecasting service design and experience. Windpower, Portland, OR, AWEA.

Matheson, J. & Winkler, R. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*, 1087–1096.

McSharry, P., Pinson, P., & Gerard, R. (2009). Methodology for the evaluation of probabilistic forecasts. SafeWind report.

Moeanaddin, R. & Tong, H. (1990). Numerical evaluation of distributions in non-linear autoregression. *Journal of time series analysis*, *11*, 33–48.

Moller, J., Nielsen, H., & Madsen, H. (2008). Time-adaptive quantile regression. *Computational Statistics and Data Analysis*, *52(3)*, 12921303.

Nielsen, H., Madsen, H., & Nielsen, T. (2006). Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind energy*, *9*, 95–108.

Nielsen, T. (1999). Experiences with statistical methods for wind power prediction. European Wind Energy Conference, Nice, France, EWEA.

Pinson, P. (2010). On probabilistic forecasting of wind power time-series. *Wind Energy*.

Pinson, P. & Madsen, H. (2009). Ensemble-based probabilistic forecasting at horns rev. *Wind energy*, *12*, 137–155.

Sanchez, I. (2006). Short term prediction of wind energy production. *International Journal of Forecasting*, *22*, 43–56.

Sanchez, I. (2008). Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting*, *24*(4), 679 – 693.

Taylor, J., McSharry, P., Member, S., IEEE, & Buizza, R. (2009). Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, *24*, 775–782.

Tol, R. (1997). Autoregressive conditional heteroskedasticity in daily wind speed measurements. *Theoretical Applied Climatology*, *56*, 113–122.

Appendix

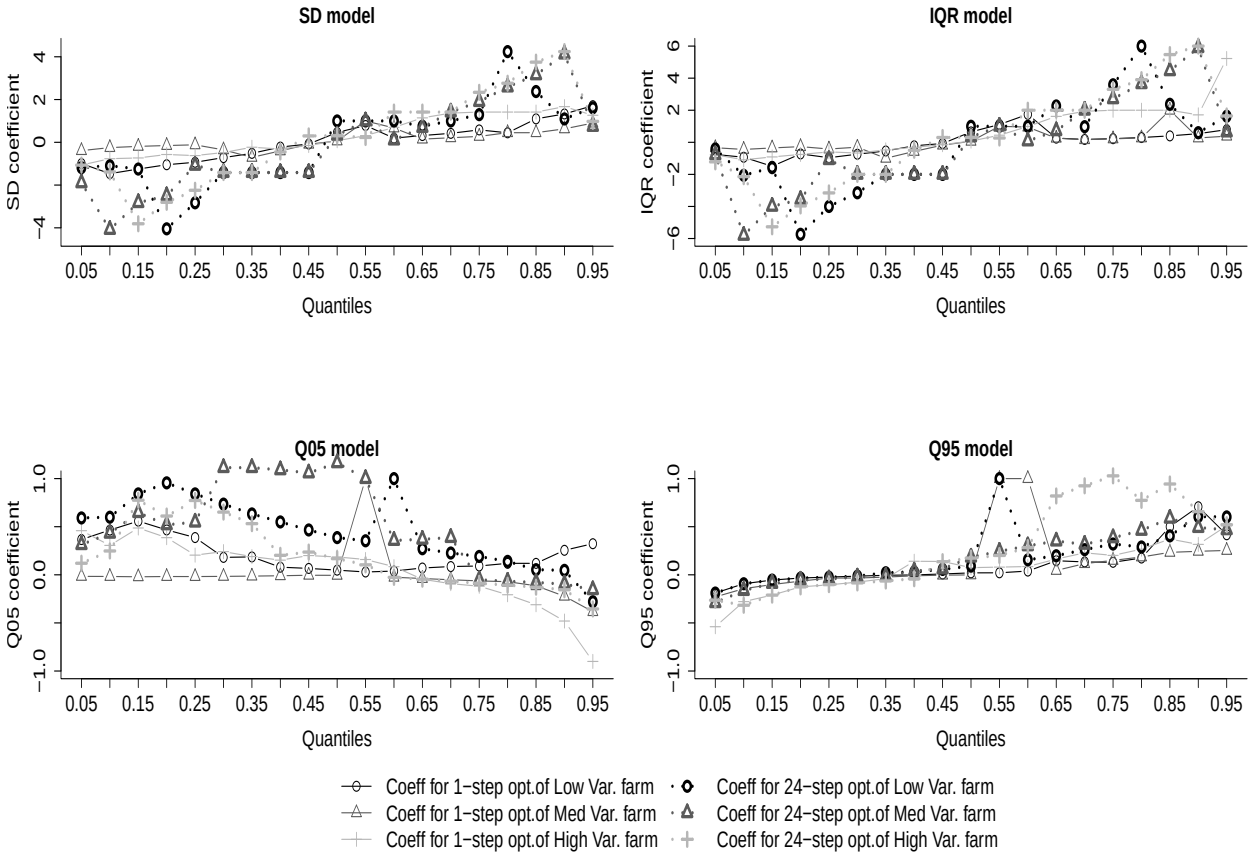


Figure 8: Variability indices' coefficients for models with 1-step check function optimization and averaged over 24-steps check function optimization.

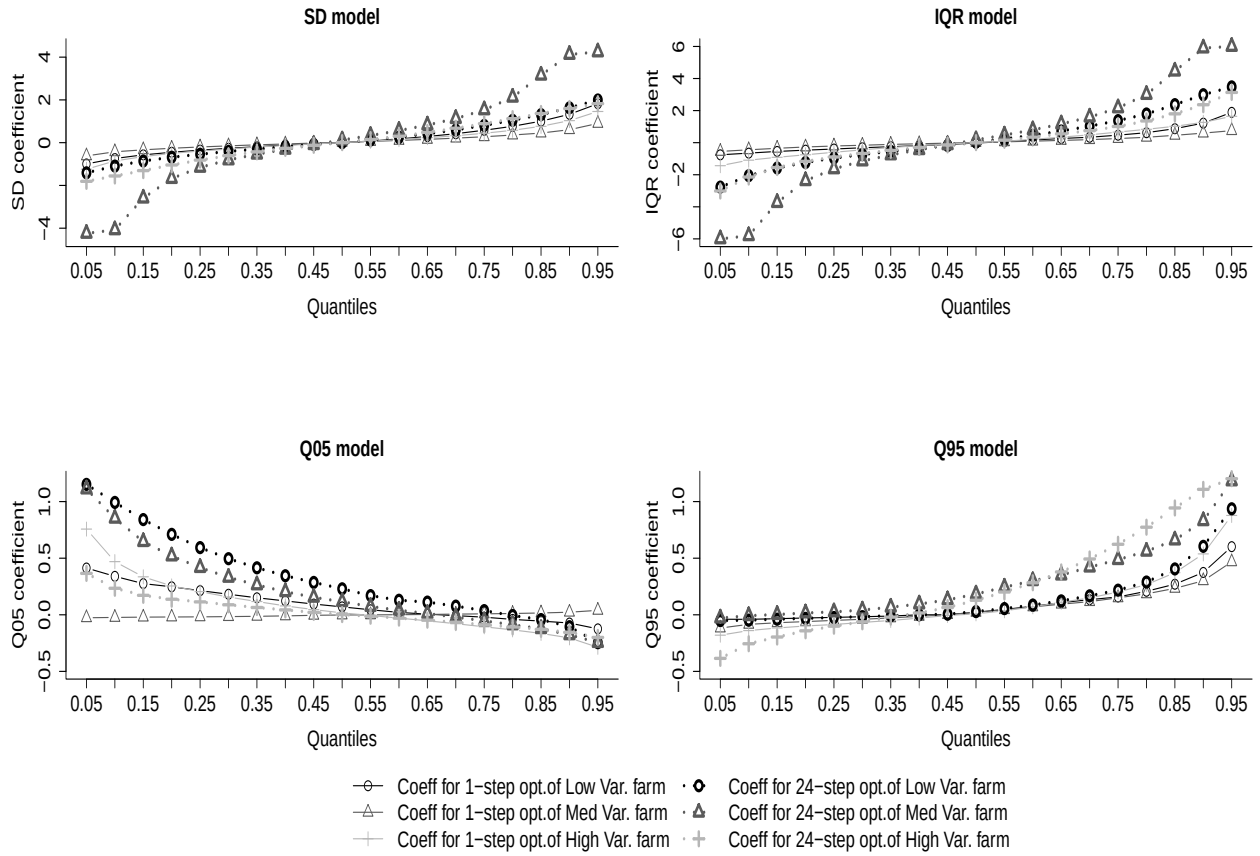


Figure 9: Variability indices' coefficients for models with 1-step CRPS optimization and averaged over 24-steps CRPS optimization.

	Low Var.		Med Var.		High Var.		Low Var.		Med Var.		High Var.	
	m	n	m	n	m	n	m	n	m	n	m	n
Model	1-step CRPS optimization						24-step CRPS optimization					
SD	0	9	0	12	0	8	0	6	3	2	0	4
IQR	2	4	0	5	0	4	2	3	3	2	2	3
Q05	0	17	24	3	0	11	0	8	0	7	0	15
Q95	2	11	2	9	0	11	0	12	5	3	0	6

Table 2: 1-step and 24-steps CRPS optimization results for all models of (18).

	Low Var.		Med Var.		High Var.		Low Var.		Med Var.		High Var.	
	m	n	m	n	m	n	m	n	m	n	m	n
Quantile	1-step optimization of **SD** model						24-steps optimization of **SD** model					
0.05	0	9	0	18	2	7	2	5	3	3	2	7
0.10	0	4	0	20	2	7	0	6	3	2	2	4
0.15	0	4	0	21	0	7	0	4	2	2	3	2
0.20	0	4	0	21	0	7	3	2	2	2	2	2
0.25	2	3	0	21	2	4	2	2	0	3	2	2
0.30	2	3	2	4	2	4	0	2	0	2	0	2
0.35	2	3	2	2	0	7	0	2	0	2	0	2
0.40	5	3	2	2	5	3	0	2	0	2	0	3
0.45	5	3	0	4	6	3	0	2	0	2	7	0
0.50	28	2	6	3	9	0	0	0	7	0	7	0
0.55	14	2	0	0	2	2	0	0	0	0	0	3
0.60	0	8	13	2	2	2	0	0	0	12	0	2
0.65	0	8	2	11	2	2	5	3	0	3	0	2
0.70	0	9	0	12	3	2	5	3	0	2	0	2
0.75	0	9	0	12	0	2	5	3	2	2	2	2
0.80	0	15	0	9	0	2	3	2	2	2	2	2
0.85	0	8	0	12	0	2	2	3	3	2	3	2
0.90	0	9	2	9	0	3	0	12	3	2	3	2
0.95	2	8	0	12	0	10	3	7	0	15	0	14
Quantile	1-step optimization of **IQR** model						24-steps optimization of **IQR** model					
0.05	2	4	0	9	0	5	2	6	2	4	2	4
0.10	0	4	0	5	0	4	2	3	3	2	2	3
0.15	0	3	0	5	0	4	2	3	2	2	3	2
0.20	0	4	0	5	0	4	3	2	2	2	2	2
0.25	2	3	0	4	0	4	2	2	0	3	2	2
0.30	2	3	0	4	2	3	2	2	0	2	0	2
0.35	2	3	2	2	5	3	0	2	0	2	0	2
0.40	5	3	2	2	5	3	0	2	0	2	0	3
0.45	5	3	0	3	6	3	0	2	0	2	7	0
0.50	28	2	6	3	9	0	0	0	7	0	7	0
0.55	14	2	0	0	2	2	0	0	0	0	0	3
0.60	11	2	13	2	2	2	0	0	0	7	0	2
0.65	2	4	0	4	2	2	5	2	0	3	0	2
0.70	0	11	0	7	3	2	5	3	0	2	0	2
0.75	0	11	0	7	0	2	5	2	2	2	2	2
0.80	0	11	0	7	0	2	3	2	2	2	2	2
0.85	0	11	0	2	0	2	2	3	3	2	3	2
0.90	0	11	0	12	0	3	0	9	3	2	3	2
0.95	2	7	0	12	4	2	5	4	0	7	0	4

Table 3: 1-step and 24-steps check function optimization results for the SD and IQR models of (18)

	Low Var.		Med Var.		High Var.		Low Var.		Med Var.		High Var.	
	m	n	m	n	m	n	m	n	m	n	m	n
Quantile	1-step optimization of **Q05** model						24-steps optimization of **Q05** model					
0.05	0	18	48	48	2	10	0	14	0	14	7	4
0.10	0	14	15	4	2	10	0	12	3	6	3	9
0.15	0	11	25	3	2	7	0	8	0	7	0	7
0.20	0	11	35	6	2	7	0	6	0	7	0	7
0.25	0	11	24	3	0	11	0	6	0	6	0	5
0.30	0	17	24	2	2	7	0	6	0	3	0	5
0.35	2	11	24	3	2	7	0	6	0	3	0	5
0.40	2	17	25	2	0	8	0	6	0	3	0	7
0.45	2	16	24	3	2	5	0	6	0	3	5	3
0.50	2	16	13	12	6	2	0	6	0	2	5	3
0.55	21	0	0	0	6	2	2	5	0	0	5	3
0.60	21	0	2	14	9	0	0	0	7	0	0	18
0.65	17	0	2	14	0	7	2	5	7	0	0	16
0.70	16	0	2	11	0	7	2	5	7	0	0	15
0.75	16	0	2	11	0	5	2	5	0	18	0	15
0.80	15	0	2	12	0	5	2	5	0	14	0	15
0.85	16	0	0	5	0	5	14	5	0	14	0	15
0.90	12	0	0	5	0	4	14	7	0	14	0	15
0.95	12	0	0	4	0	4	0	4	0	11	0	3
Quantile	1-step optimization of **Q95** model						24-steps optimization of **Q95** model					
0.05	0	4	0	6	0	4	0	4	0	4	0	8
0.10	0	6	0	6	0	4	0	6	0	5	0	5
0.15	0	6	0	6	0	5	0	6	0	5	0	5
0.20	0	8	0	6	0	5	0	10	0	10	0	8
0.25	0	9	2	18	0	7	0	10	0	10	0	8
0.30	0	10	2	13	0	7	0	10	5	19	2	8
0.35	0	45	2	13	0	7	17	0	3	22	2	8
0.40	0	45	12	14	8	0	4	6	6	3	3	7
0.45	21	0	16	11	8	0	4	6	6	3	6	2
0.50	21	0	32	3	9	0	4	6	5	3	6	2
0.55	0	18	0	0	9	0	0	0	5	3	0	6
0.60	0	18	0	0	0	10	3	7	5	3	0	6
0.65	0	11	0	16	0	9	2	8	5	3	0	4
0.70	2	11	2	9	0	9	3	7	3	5	0	4
0.75	0	15	2	9	0	11	2	8	3	5	0	4
0.80	0	15	2	9	0	11	0	12	3	5	0	6
0.85	3	7	2	9	0	11	0	12	3	5	0	6
0.90	3	7	2	10	2	11	0	12	3	6	0	10
0.95	2	13	3	10	2	11	2	11	2	9	0	14

Table 4: 1-step and 24-steps check function optimization results for the Q05 and Q95 models of (18).

Quantile forecast conditionned to ramp information

Abreviations and Mathematical Notations

MAE	Mean Absolute Error
MADFPR	Mean Absolute Deviation From Perfect Reliability
MDFPR	Mean Deviation From Perfect Reliability
MSE	Mean Square Error
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Square Error
NSDE	Normalized Standard Deviation of the Errors
NWP	Numerical Weather Prediction
RMSE	Root Mean Square Error
SDE	Standard Deviation of the Errors
α	Nominal coverage rate (alternatively referred to as nominal proportion)
a^α	Actual coverage rate of the α-quantile estimation
A_r	Rotor swept area exposed to the wind
A_w	Weibull scale parameter
δ^α	Width of the $(1-\alpha)$ centered prediction interval
e	Prediction error
ϵ	Normalized prediction error
F	Distribution function
h	Look-ahead time (alternatively referred to as prediction horizon)
$\mathbb{H}$	Set of look-ahead times
$\hat{I}^\alpha$	Prediction interval with nominal coverage rate $(1-\alpha)$
k_w	Weibull shape parameter

$\hat{L}^{\alpha}$	Lower bound of the $(1-\alpha)$-confidence prediction interval
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution, with mean μ and variance σ^2
p	Wind power
p_t	Wind power measure made at the instant t
$\hat{p}_{t+h/t}$	Wind power prediction made at the instant t for look-ahead time h
$\mathbb{P}(X)$	Probability for the occurrence of the event X
P_n	Wind farm nominal power (alternatively referred to as installed capacity)
μ	Mean of a distribution
q^{α}	α-quantile of a distribution
ρ_{air}	air density
σ	standard deviation of a distribution
t	time
$\mathbb{T}$	Set of time index
$\hat{U}^{\alpha}$	Upper bound of the $(1-\alpha)$-confidence prediction interval
z	height above ground level
z_0	Roughness length

Contents

Chapter 1

Introduction

According to the United Nations Intergovernmental Panel on Climate Change (IPCC), a target of 25-40% of greenhouse gas reduction by 2020 is necessary in the industrialized world, to give us a chance of avoiding the 2 °C temperature increase. The predicted 230GW of installed wind power could avoid the emission of 420 million tonnes of CO_2 a year, equivalent to 36% of EU's greenhouse gas reduction target. Benefits to the environment, stimulation of the economy by creating jobs and encouraging investments at home, rather than transfers to wealth to fuel-exporting nations are so many reasons that encourage a large scale integration of wind generation in power systems.

Such a large scale integration causes several difficulties in the management of a power system. Often a high level of spinning reserve is allocated to account for the intermittent nature of wind generation, thus reducing the benefits of the use of wind energy. Predictions of wind generation contribute to a secure and economic power system operation. Also, as electricity markets are gaining importance, wind predictions are helpful for wind energy producers who have to propose their bids on the market on a day-ahead basis. Increasing the value of wind generation through the improvement of prediction systems' performance is one of the current purpose of wind energy research.

The main issue that we encounter today is to give a forecast able to handle an extreme situation. What can be considered as an extreme is not an obvious task and is highly end-user-dependent. Often for power system operators, extreme events are linked to large deviations of power generation with respect to the expected power generation. The severity of the large deviation depends highly on how fast it happens, and on the timing especially if concurrently other events happen (i.e the electricity demand is also highly fluctuating). The case of high and sharp variations of power generation called *ramp* events poses many problem, one of which is the calibration of quantile forecasts around these event. Nowadays it is a challenge to improve the forecasts and in particular quantile forecast for such situations. In this report we focus on probabilistic forecasting conditionnally to *ramp* events. Two types of solutions, rather complementary can be envisaged :

1. Try to detect early the deviation between the expected and the observed power generation and correct the corresponding forecast or send an alert to the end-user. This type of solution clearly relies on improving the data

assimilation process. It is related to time scales of a few minutes to a few hours.

2. Try to provide uncertainty bounds that indicate the possibility of such deviations to happen or provide risk indices as warnings for such large deviations. This type of solution relies on improving the way we learn the uncertainty from the past. It is related to time scales of a few hours to one week.

In this part of the report we try to improve the second type of solutions.

Structure of the report: In a first chapter, we introduce the principles and the state of the art of wind power forecasting. Most of existing wind power forecasting methods provide point forecasts (so called "deterministic" ones) which stand to be estimations of the future power outcomes. A large part of recent research works have focused on *probabilistic* forecasting methods, which associate uncertainty estimation to these point forecasts. An overview of such forecasting methods is introduced, as well as the associated forecasts evaluation framework. Then we give some general results on forecasting error and uncertainty. The chapter ends with an introduction to the ramp events problematic.

In a second part, we discuss how to define completely a ramp and give a mathematical formulation allowing to detect it. We then describe a multi-stage forecasting procedure which uses the ramps detection method on *preliminary* deterministic forecasts, to provide improved uncertainty estimations at such ramp events. The procedure is tested on real world cases. Finally we give some conclusive remarks and perspectives for future research work.

Chapter 2

Introduction to wind power forecasting

2.1 Wind power forecasting.

2.1.1 Meteorological forecasts.

One can distinguish two types of wind power forecasting methods : physical and statistical methods. The great majority uses forecasts of some meteorological variables, wind speed and wind direction being the most common. These last ones are provided by using *Numerical Weather Prediction* models (NWP). Such models are based on equations governing the motions and forces affecting motion of fluids. From the knowledge of the actual state of the atmosphere, the system of equations allows to estimate what will be the evolution of state variables such as temperature, velocity, humidity and pressure at a series of grid points. The spatial resolution of a NWP, e.g the distance between two grid points, is too large and doesn't allow to take into account local effects such as local thermal effects or turbulence mixing for instance. So it is necessary to extrapolate predictions at the level of wind farms : one talks about *downscaling procedure*. Considering wind speeds, after having extrapolated them at the level of the wind farm, the next step is to convert them to the hub height of the wind turbines. This is because NWPs are commonly given in 10 m a.g.l. or other atmospheric levels which differ from the hub height of the turbines of a specific farm. To achieve the conversion, a well-known model is the logarithmic wind profile :

$$v(z) = \frac{v_\star}{k} \ln\left(\frac{z}{z_0}\right), \quad z \geq z_0 \tag{2.1}$$

where k is the von Karman constant, $v_\star$ the friction velocity and z_0 the roughness length which is related to surface roughness. Actually, by writing equation (2.1) for two different heights z_1 and z_2 and after a trivial manipulation, a relation between the two wind speeds appears :

$$v(z_2) = \frac{\ln(z_2/z_0)}{\ln(z_1/z_0)} v(z_1) \tag{2.2}$$

Equation (2.2) makes possible the conversion. With wind speed predictions at the height z_1, it is now easy to get predictions for the height z_2. For more details on NWP models and the downscaling procedure we refer to [1] and references there are into.

2.1.2 The wind power forecasting approaches.

Whatever the forecasting approach, there is a generic formulation of the forecasting problem. In this report we denote by p_t the power value (of a wind farm) corresponding to time t. It can be an instantaneous value or an average over the past time interval $[t-1, t]$. The forecaster aims to use data in order to fit a model which describes mathematically the relation between the *variable of interest* p_t and *explanatory variables*. Some models, called univariate models, use only past and present values of the variable of interest : $p_t, p_{t-1}, \ldots, p_{t-l}$ as explanatory variables. Some others, and particularly in our case, use also a set of past, present and predicted values of q other variables $\mathbf{x} \in \mathbb{R}^q$ (such as meteorological forecasts). We can write a such multivariate model as below :

$$p_{t+h} = f(p_t, p_{t-1}, \ldots, p_{t-l}, \mathbf{x}_t, \ldots, \mathbf{x}_{t-m}, \hat{\mathbf{x}}_{t+h/t}) + e_{t+h} \tag{2.3}$$

where $\hat{\mathbf{x}}_{t+h/t}$ are predicted values of the q variables $\mathbf{x} \in \mathbb{R}^q$, for the instant $t+h$, with information available until time t. e_{t+h} is the modelisation error. Whatever the type of forecast, it is important to realize that forecasting is a type of extrapolation : a prediction model is fitted with data, and then used out of that range of data. What makes the difference between statistical and physical methods is *how is fitted the model*. Finally, one gets an estimator $\hat{f}$ which provides predictions :

$$\hat{p}_{t+h/t} = \hat{f}(p_t, p_{t-1}, \ldots, p_{t-l}, \mathbf{x}_t, \ldots, \mathbf{x}_{t-m}, \hat{\mathbf{x}}_{t+h/t}) \tag{2.4}$$

Physical approaches focus on a deep refining of the NWPs. The extrapolation of meteorological forecasts at wind park location, is done relying on physical considerations about the terrain such as roughness, orography and obstacles. The energy conversion process is assumed by a wind power curve. A single turbine power curve given by the manufacturer and obtained from experiments in wind tunnels doesn't suit to the role. To improve results, a wind farm power curve has to be estimated with data from the site. A lot of aspects are taken into account, the number and types of turbines coming first. Moreover a wind farm power curve cannot be envisaged as a sole function of wind speed, but also of wind direction. Pinson goes further in the description of wind farm power curve estimation in [1].

From a statistical point of view, it is interesting to know what are the power time-series $\{p_t, t \in \mathbb{T}\}$ properties. Two main features are its *nonstationarity* and its *nonlinearity* [1]. Nonstationarity of a time-series means that its behavior, which may be described by its moments (mean, standard deviation ...), evolves with time. Nonlinearity means that wind power time-series exhibits features that cannot be explained by linear models. While nonstationarity comes from the very nature of wind, the nonlinear and bounded nature comes from the energy conversion process : wind power curve clearly reveals the nonlinear dependence of the delivered power with wind speed.

In the statistical methodology, data is separated into two independent sets : a *training* and a *test* set. While the estimator is build with data from the training set, its evaluation is realized with data from the test set. The estimator construction hold on the resolution of an optimization problem of the following form :

$$\hat{f} = \arg\min_{f \in \mathbb{F}} \underbrace{\sum_{i=1}^{n} L(p_i, f(\mathbf{y}_i))}_{1} + \underbrace{\lambda \mathfrak{F}(f)}_{2} \tag{2.5}$$

$\{(p_i, \mathbf{y}_i), i = 1, \ldots, n\}$ is the training set. L is a *loss function*. From the choice of the loss function depends what parameter of the probability law of the power p is estimated, this conditionally to the value of the explanatory variables $\mathbf{y}$. The most common loss functions are the quadratic loss function (which permits to estimate the mean of a probability distribution) or the absoltute one (to estimate the median of the distribution). $\mathbb{F}$ depends on the model and the method of estimation used. For instance, for a linear model with q regressors : $\mathbb{F} \subset \mathbb{R}^q$. Finally the term 2 is a regularization term which permits to confer some properties to the estimator (particularly to respect the bias-variance trade-off [2]). The formulation given in (2.5) can be rewritten more explicitly, precising explanatory variables and accounting for the fact one works on a time-series :

$$\hat{f}_h = \arg\min_{f_h \in \mathbb{F}} \sum_{t=0}^{N_{T_r}} L(p_{t+h}, f(p_t, \ldots, p_{t-l}, \mathbf{x}_t, \ldots, \mathbf{x}_{t-m}, \hat{\mathbf{x}}_{t+h/t})) + \lambda \mathfrak{F}(f_h) \tag{2.6}$$

N_{T_r} represents the size of the training set. One has to notice the existence of an estimator for each look-ahead time h.

2.1.3 The performance evaluation framework for deterministic forecasting methods.

A large number of usual prediction methods are *deterministic forecasting methods*. Also called point prediction methods, they give at time t and for each look-ahead time h a single value as a prediction. This value is most of the time an estimation of the mean of the power output p_{t+h}, seen as a random variable, conditionally to the *information set* Φ_t. Φ_t is the sum of all knowledges available until time t and used for the predictions.

To do a pertinent comparison between two forecasting methods, a standardized protocol of evaluation has to be introduced. Such a protocol is a long time going, and there is not one which is definitively adopted by all the wind forecasters community. In [3] Madsen et al. give the following recommendations :

- Define clearly an operational framework before giving any result of performance (installed capacity, horizons of prediction, frequency of updates, characteristics of NWP forecasts...).
- Make the performance evaluation on the test set only, to avoid some overoptimistic conclusions.

- As a minimum set of error measures use :
  - NBIAS
  - NMAE
  - NRMSE
- Use the improvement scores for comparison between models.

Here comes a mathematical description of error measures given above. The *prediction error* at time t for horizon h is defined as the difference between measured and predicted values :

$$e_{t+h/t} = p_{t+h} - \hat{p}_{t+h/t} \tag{2.7}$$

A normalized prediction error is often given in order to compare results from wind farms of different *installed capacity* :

$$\epsilon_{t+h/t} = \frac{1}{P_n} e_{t+h/t} \tag{2.8}$$

where P_n is the installed capacity (nominal power) of the wind farm. Any prediction error can be decomposed into a systematic error μ_e and a random error ξ_e :

$$e = \mu_e + \xi_e \tag{2.9}$$

where μ_e is a constant and ξ_e a zero mean random variable.

The model bias estimates the systematic part of prediction error, by computing its average on all the N_{T_e} times t corresponding to the test set period. It is calculated for each horizon h :

$$BIAS(h) = \frac{1}{N_{T_e}} \sum_{t=1}^{N_{T_e}} e_{t+h/t} = \overline{e}_h \tag{2.10}$$

The SDE criterion estimates the standard deviation of the error distribution. So only the random part of the error contributes to it :

$$SDE(h) = \frac{1}{N_{T_e}} \sum_{t=1}^{N_{T_e}} \left(e_{t+h/t} - \overline{e}_h\right)^2 \tag{2.11}$$

There are two basic criteria for illustrating a predictor performance. The *root mean squared error* RMSE and the *mean absolute error* MAE, which in a sense, represent the distance between predictions and the mean (resp. the median) of the wind power distribution.

$$MSE(h) = \frac{1}{N_{T_e}} \sum_{t=1}^{N_{T_e}} e_{t+h/t}^2 \tag{2.12}$$

$$RMSE(h) = \sqrt{MSE(h)} \tag{2.13}$$

$$MAE(h) = \frac{1}{N_{T_e}} \sum_{t=1}^{N_{T_e}} |e_{t+h/t}| \tag{2.14}$$

All these criteria can be normalized using the normalized prediction error ϵ

A skill score which permits to quantify the benefit of an advanced method on one of reference is the *improvement* score. It is calculated relatively to one reference method and one criterion (which may be the MAE, RMSE, SDE or the normalized version of one of them) :

$$Imp_{ref}^{EC}(h) = \frac{EC_{ref}(h) - EC(h)}{EC_{ref}(h)} \tag{2.15}$$

where EC is the evaluation criterion, and ref denotes the reference method. In particular the improvement score calculated with the RMSE criterion and relatively to the *climatology* method is called the R^2 criterion.

There exist a lot of operational forecasting tools, based on statistical methods or not, developed in several country of Europe such as France, Denmark, Germany or Spain. However we decide to not present them in this paper and refer to [4] and [5] for a brief overview. Nevertheless two methods have to be introduced, because they are generally used as references to evaluate the improvement of new developed forecasting methods. The *persistence* states that the future wind generation will be the same as the last measured value, i.e

$$\hat{p}_{t+h/t}^{P} = p_t \tag{2.16}$$

To generalize, one can replace the last measured value by the average of the n last measured values. One obtains a *moving average* predictor :

$$\hat{p}_{t+h/t}^{MA,n} = \frac{1}{n}\sum_{i=0}^{n-1} p_{t-i} \tag{2.17}$$

when n goes to infinity, the quantity defined above tends to the global average (also called *mean climatology*) :

$$\hat{p}_{t+h/t}^{0} = \overline{p_t} \tag{2.18}$$

where $\overline{p_t}$ is the average of all the available observations of wind power at time t. Despite their apparent simplicity, these methods have quite good results but not in the same range of horizons. While the persistence has good results for the very first horizons (because of the slow changes in the atmosphere), climatology is better for further horizons (its not dynamic nature made of it a bad forecasting methods for small horizons).

2.1.4 Probabilistic forecasts.

Deterministic forecasts immediately seduced end-users by their simple interpretation : one has a point forecast which stands for an estimation of the future power outcome. Although *probabilistic forecasts* are not so easy to understand for not initiate people, the need of uncertainty assessment they provide, has implied the constant development of such a forecasts type during the last decade. Recent research works proved the growth of wind power producers revenue, participating in electricity markets, when using advanced bidding strategies relying on such uncertainty assessment (see [6], [7] among others).

Today probabilistic forecasts, in wind power forecasting literature, usually have the form of quantiles or prediction intervals. They are produced using

either the usual deterministic NWPs as input or ensemble NWPs which consist in alternative scenario of the weather evolution in the coming period. In any case the aim of a probabilistic model is to provide the full probability density function of wind production for each time step in the considered future period. Let p_t be wind power at the instant t seen as a random variable, which real output $p_t^\star$ is a realization. Denoting P_t the probability law of p_t, which probability density function is f_t^p. A prediction interval is a range of values, in which is expected to lie the future power output, with a predefined probability called *nominal coverage rate*. Mathematically $\hat{I}^\alpha_{t+h/t} = [\hat{L}^\alpha_{t+h/t}, \hat{U}^\alpha_{t+h/t}]$ is a prediction interval given at the instant t for the instant $t+h$ of nominal coverage $1-\alpha$, if one has

$$\mathbb{P}(p_{t+h} \in \hat{I}^\alpha_{t+h/t}) = 1 - \alpha \tag{2.19}$$

Most of the time, prediction intervals are centered prediction intervals : there is the same probability for a not-covered outcome to be under or above the interval bounds, i.e :

$$\mathbb{P}(p_{t+h} < \hat{L}^\alpha_{t+h/t}) = \mathbb{P}(p_{t+h} > \hat{U}^\alpha_{t+h/t}) = \alpha/2 \tag{2.20}$$

The lower and upper bounds $\hat{L}^\alpha_{t+h/t}$, $\hat{U}^\alpha_{t+h/t}$ are estimations of respectively the $\alpha/2$ and the $1-\alpha/2$ quantiles of the wind generation probability law P_{t+h}. To estimate such prediction intervals, there are parametric and non-parametric approaches. One talks about a parametric approach when an assumption is done on the predictive distribution. A such common assumption is to assume that prediction errors are independent and identically distributed Gaussian with zero mean and σ_e^2 variance, $e_t \sim \mathcal{N}(0, \sigma_e^2)$. By using an estimate $\hat{\sigma}_e^2$ of the variance, one obtains for a h-step ahead prediction interval with nominal coverage rate $1-\alpha$:

$$\hat{I}^\alpha_{t+h/t} = [\hat{p}_{t+h/t} + z_{\alpha/2}.\hat{\sigma}_{e,h}, \hat{p}_{t+h/t} + z_{1-\alpha/2}.\hat{\sigma}_{e,h}] \tag{2.21}$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are respectively the $\alpha/2$ and $1-\alpha/2$ quantiles of the zero mean unit variance Gaussian distribution $\mathcal{N}(0,1)$.

This first simple approach leads to not very accurate results as mentioned in [1]. A better one, assuming that the predictive distribution is a β-distribution, is also briefly described in [1]. A such distribution has its shape controlled by two parameters, which are functions of the mean and the variance of the distribution. In the related study, the mean is estimated by predictions resulting from a point-deterministic forecasting method and the variance by the historical performance of the predictor (i.e the standard deviation). This approach is presented as having a significant improvement against intervals produced from a Gaussian assumption.

Another way to produce such prediction intervals, is to directly estimate quantiles of the predictive distribution. In other words one estimates the upper and lower bounds defined above : L^α_{t+h} and U^α_{t+h}. Quantile regression is a family of non-parametric methods that aims at doing it. In [8] Nielsen et al. used linear quantile regression, with point forecasts of a state-of-the-art forecasting method and others (meteorological) explanatory variables as inputs. That method is an extension to any point forecasts method, which to deterministic forecasts associate uncertainty estimation. In [9] Bremnes et al.

compared several models of quantiles estimation based on three local regression methods : a local quantile regression one, a local Gaussian model and a Nadarya-Watson estimator for conditional cumulative distribution functions. In [10] Juban et al. described a kernel based method to estimate the complete probability density function of the predictive distribution.

Evaluation of probabilistic forecasts Just as we did for deterministic forecasting methods, one has to define a performance evaluation framework for probabilistic forecasting ones. This has recently been discussed in [11] and [12]. The first essential property that have to verify probabilistic forecasts is *reliability*. A probabilistic forecast is *reliable* if the actual coverage of the estimated prediction interval (respectively the estimated quantile) coincides with the preassigned one. In literature the predictions reliability is associated to the property of forecasts being *unbiased*, or else *consistent*. Following notations given in [11], let $\hat{q}^{\alpha}_{t+h/t}$ being an estimated α-quantile, issued at time t for lead time $t+h$ and p_{t+h} the actual outcome at the same instant. Let us introduce the indicator variable $\zeta^{\alpha}_{t,h}$:

$$\zeta^{\alpha}_{t,h} = \mathbb{1}_{p_{t+h} \leq \hat{q}^{\alpha}_{t+h/t}} = \begin{cases} 1 & \text{if } p_{t+h} \leq \hat{q}^{\alpha}_{t+h/t} \\ 0 & \text{else} \end{cases} \tag{2.22}$$

The actual coverage a^{α}_{h} of the α-quantile estimated for horizon h is then $\mathbb{E}[\zeta^{\alpha}_{t,h}]$, and is estimated by the empirical average calculated on the test period $t \in \{1, \ldots, N_{T_e}\}$:

$$\hat{a}^{\alpha}_{h} = \frac{1}{N_{T_e}} \sum_{t=1}^{N_{T_e}} \zeta^{\alpha}_{t,h} = \frac{n^{\alpha}_{h,1}}{n^{\alpha}_{h,0} + n^{\alpha}_{h,1}} \tag{2.23}$$

where $n^{\alpha}_{h,1} = \#\{p_{t+h} \leq \hat{q}^{\alpha}_{t+h/t},\ t = 1, \ldots, N_{T_e}\}$ and $n^{\alpha}_{h,0}$ is the complementary event. They can be rewritten as following :

$$\begin{aligned} n^{\alpha}_{h,1} &= \#\{\zeta^{\alpha}_{t,h} = 1\} = \sum_{t=1}^{N_{T_e}} \zeta^{\alpha}_{t,h} \\ n^{\alpha}_{h,0} &= \#\{\zeta^{\alpha}_{t,h} = 0\} = N_{T_e} - n^{\alpha}_{h,1} \end{aligned}$$

This measure serves as a basis to construct the *reliability diagram* where the nominal probability is drawn against the empirical one, for several nominal coverage rates. The closer to diagonal the better. A benefit of reliability diagrams is that they permit to visualize reliability assessment for various quantiles associated to different proportions, and so to see at a glance if a given forecasting method tends to systematically under or overestimate these quantiles.

To base the evaluation on the sole reliability criterion is not enough. Indeed a basic forecasting method such as *climatology*, which relies on past observations to estimate the unconditional wind power distribution (resp. quantiles), has a perfect reliability. However this method isn't able to distinguish the distribution variability due to specific situations (e.g conditionally to specific situations). A good forecast system, first has to be reliable and then be able to do a such discrimination. In [13] and [11] this relates to the notion of *sharpness* (or *resolution*). Sharpness is only a function of forecasts and is not a

verification measure. Between two reliable forecasting methods, one will prefer the one which better catches distribution variability according to specific situations, e.g which has a better sharpness. In the case of prediction intervals, it can be linked to the interval size distribution. As Pinson et al. did in [11], we decide in this report, to only estimate the mean size of such intervals. If writing :

$$\delta_{t,h}^{\alpha} = \hat{q}_{t+h/t}^{1-\alpha/2} - \hat{q}_{t+h/t}^{\alpha/2}$$

the size of the centered prediction interval of nominal coverage rate $1-\alpha$, issued at t for lead time $t+h$. Then one estimates the sharpness of a such coverage interval for horizon h by :

$$\overline{\delta}_{h}^{\alpha} = \frac{1}{N_{T_e}} \sum_{t=1}^{N_{T_e}} \delta_{t,h}^{\alpha} \tag{2.24}$$

Of course this measure needs pairs of quantiles to be calculated. From this measure can be drawn a δ-diagram which represents the mean size $\overline{\delta}_{h}^{\alpha}$ as a function of the nominal coverage rate. It permits to have an idea of the predictive distribution spread.

A last measure of performance is the *pinball loss*. Commonly used skill from the statistical community, it comes from the knowledge that the α-quantile $q_{\alpha}(\Phi_t)$ of a random variable p_{t+h} given the information set Φ_t can be defined by its variational formula :

$$q_{\alpha}(\Phi_t) = \text{Argmin}_f \mathbb{E}[L_{\alpha}(p_{t+h}, f(\Phi_t))] \tag{2.25}$$

where the infimum is taken among all measurable functions of the information set Φ_t and L_{α} is the so called "pinball loss" :

$$L_{\alpha}(y,u) = \begin{cases} (1-\alpha)(u-y) & \text{if } y \leq u \\ \alpha(y-u) & \text{else} \end{cases} \tag{2.26}$$

Most of the recent procedures to estimate quantiles are based on this formula. A natural way to evaluate a forecasted quantile $\hat{q}_{t+h/t}^{\alpha}$ of a random variable p_{t+h} given Φ_t is so to look at the empirical version of the term to minimize in (2.25), still calculated on the test set :

$$\frac{1}{N_{T_e}} \sum_{t=1}^{N_{T_e}} L_{\alpha}(p_{t+h}, \hat{q}_{t+h/t}^{\alpha}) \tag{2.27}$$

It has to be noticed that this is the error calculated by default in all R software packages.

Other verification measures, in particular skill scores which tend to evaluate simultaneously dispersion properties and reliability (for instance the Brier score [14]) can be found in [12].

2.1.5 About forecasting error and uncertainty.

In [1] the author presents an interesting work on the wind power forecasting error. To highlight forecasting error contributions and characteristics of the

uncertainty, several state-of-the-art methods from physical and statistical approaches are used on different wind farms. The analysis is based on performance results measured with a selected set of criteria such as the root mean squared error, the mean absolute error, the bias, the standard deviation of errors, and also an analysis based on error distributions.

As one may expect, each step of the prediction process contributes to the forecasting error. First come meteorological forecasts which are the main inputs of all the state-of-the-art forecasting methods. Of course the modeling of the wind-speed-to-power conversion has also an influence. In this last contribution the error may be due to the way NWPs forecasts are refined to the level of wind farms (and so the way park effects are considered) and to the way the power curve is estimated. As it is underlined in [1] and [15] the error part due to meteorological forecasts is of major importance, because it may be potentially amplified (between cut-in-speed and rated-speed) or reduced (below cut-in-speed or between rated-speed and cut-off-speed), when passed through the power curve. More precisely the shape of the error distribution is changed by the nonlinear and bounded nature of the power curve and while wind speed distribution error can be considered as Gaussian in many cases, it is no longer the case for wind power distribution error (see [15] and [16]). One may also point out the influence of seasons and of terrain complexity where wind farms are implanted, on the performance of forecasting methods. In addition it is interesting to know that predicting the power output for a single wind farm or for a group of wind farms does not lead to the same level of forecast uncertainty, because of a geographical smoothing effect.

Studying the moments of the error distributions as functions of the look-ahead time, Pinson concluded that the almost linear increase of the standard deviation causes the high level of uncertainty for far horizons (the other moments remaining almost at the same level on the all range of look-ahead times). Still looking at the moments of the error distributions, but conditionally to the level of the predicted or measured power, he pointed out some characteristics of the error due to the power curve effect. Indeed he concluded on a general overestimation (respectively underestimation) of the predicted power for low (resp. high) levels of power output. He also concluded on a higher level of uncertainty for power outputs of medium ranges. The same conclusion is made by Lange in [15]. To approximate the uncertainty of power predictions at a wind speed u, Lange used the local derivative of the power curve :

$$\sigma_P(u) = \left| \frac{dP}{du}(u) \right| \overline{\sigma_u} \tag{2.28}$$

where σ_P is the standard deviation representing the current power prediction error, $|dP/du|(u)$ the absolute value of the derivative of the power curve at wind speed u and $\overline{\sigma_u}$ the annual mean of the wind speed prediction error. One easily understands that on the steepest part of the curve, corresponding to medium range power outputs the absolute derivative is higher, which leads according to the equation (2.28), to a higher level of uncertainty for wind power predictions.

2.2 Unadequacy of quantile forecast around ramp events.

In this work we focus on the prediction of quantiles conditionaly to *ramps* events. A ramp is a sharp variation (increase or decrease) of wind power production. Often the amplitude of such variations is rather well predicted. What seems to be more difficult to predict is the timing of ramps. Error in the timing may result to a shift in time of the prediction resulting in turn to the so-called phase errors. Such phase errors may be high as shown in the example of Figure (2.1). Improving ramp events prediction is expected to result in reduction of phase errors.

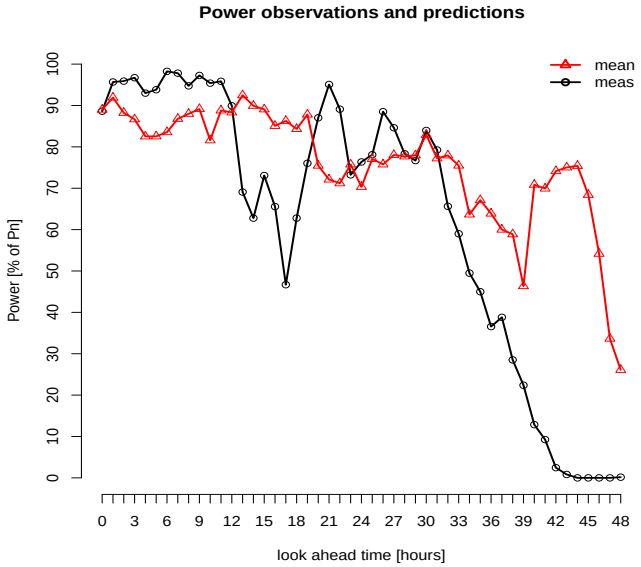


Figure 2.1: *Example of a phase error appearing at the horizon 31.*

Indeed as one can see on figure (2.1) between horizon 42 and 44 : while predicted power reaches a level near of 80% of the nominal power, the real power output is in fact of 0%, which represents an error of about 80% of the nominal power. Such error can have a high impact for end-users such as power system operators or traders in electricity markets. Working to improve these events' predictability presents unquestionably an interest.

In order to evaluate what is the contribution of phase errors in prediction error, Lange in [15] and [17] decompose the MSE in three parts. It is of common knowledge that MSE can be written as the sum of the squared bias and SDE. Lange et al. separated this last term to get :

$$\begin{aligned} MSE &= RMSE^2 \\ &= BIAS^2 + SDE^2 \\ &= BIAS^2 + sdBIAS^2 + DISP^2 \end{aligned} \tag{2.29}$$

where one has :

$$
\begin{aligned}
BIAS &= \overline{\epsilon} \\
SDE &= \sigma(\epsilon) \\
sdBIAS &= \sigma(x_p) - \sigma(x_m) \\
DISP &= \sqrt{2\sigma(x_p)\sigma(x_m)(1 - r_{p,m})}
\end{aligned}
$$

with $r_{p,m}$ the cross-correlation coefficient between the two time-series, $\sigma(x_p)$ and $\sigma(x_m)$ their standard deviations.

The $BIAS$ accounts for the difference between the mean values of prediction and measurement, whereas the standard deviation SDE, measures the fluctuation of the error around its mean. SDE has two contributions : First the $sdBIAS$ which is the difference between the standard deviations of the two time-series and represents the error due to a wrongly predicted variability. $BIAS$ and $sdBIAS$ are sensitive to *amplitude errors*, which by opposition to phase errors, are characterized by a good appreciation of the temporal evolution but with a systematic overestimation or underestimation of the real situation. Second, the dispersion $DISP$, accounts for the contribution of phase errors to the MSE. Nevertheless one has to note that here, phase errors are understood in a statistical sense in terms of the cross-correlation (which appears in its definition) and not as a well-defined phase shift between two time series.

In addition to the fact that a large part of the $RMSE$ value comes from the term of dispersion $DISP$, Lange et al. also pointed out the difficulty to remove or simply lower this error part. Indeed while $BIAS$ and $sdBIAS$ can be corrected by linear schemes, often used in post-processed methods ; it is no longer the case for the dispersion $DISP$, because the cross-correlation coefficient is invariant by linear transformations.

Phase errors come from the misevaluation of the evolution of meteorological fronts by NWP predictions. Actually these meteorological fronts move faster (respectively slower) than expected. So phase errors come from meteorological forecasts and power prediction methods, which are only advanced wind-speed to power conversion processes, don't correct them. Thus the easiest (maybe the only) way to try to remove phase errors in power forecasts, seems to do it in meteorological forecasts. That work is more from meteorology field or atmospheric sciences research, and one may only cite as a few references : [18], [19], [20] and [21]. Our purpose is rather to give information on the possible appearance of a phase error at a ramp event, by working on the uncertainty level associated to power forecasts. Typically a situation which can occur at a ramp event is presented in figure (2.2). One can see a set of eight consecutive observations, between horizons 39 and 47, lying out of the centered prediction interval of 80% nominal coverage rate, which make us feel that our uncertainty estimation is not "locally" reliable. One has to avoid any hurried conclusion, since a data sample of length eight has never been large enough to be statistically relevant. However it makes us wonder if probabilistic forecasts introduced, generated on a per look-ahead time-basis, catch well or not the apparent correlation there is between these consecutive observations at a specific ramp situation. In [22] Pinson et al. affirm that such forecasts, by neglecting the interdependence structure of forecast errors among look-ahead times, do not inform on the development of the prediction errors through the

prediction series related to a given prediction time point. They developed a method generating statistical scenarios for short-term wind generation, from the combination of probabilistic forecasts (in particular non parametric) with the interdependence structure of prediction errors. This method is based on the conversion of prediction errors through forecast series, in a multivariate Gaussian random variable which covariance matrix summarizes the mentioned interdependence structure. In order to catch the slow evolution with time of prediction error characteristics, the matrix is recursively estimated using an exponential forgetting scheme.

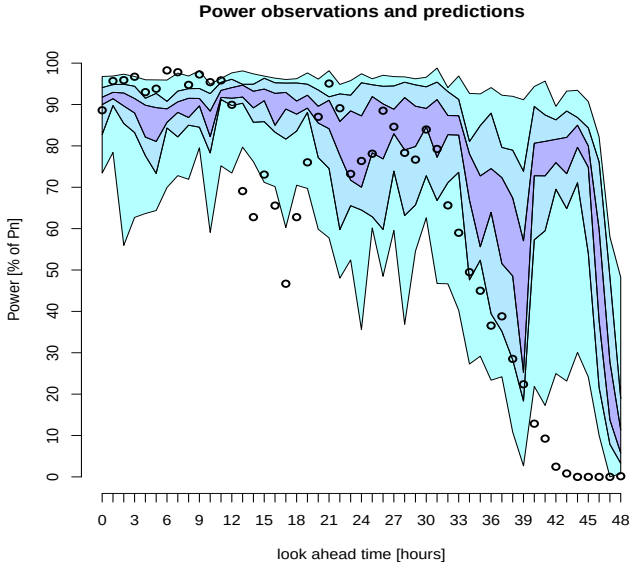


Figure 2.2: *Power observations and uncertainty estimation represented by centered prediction intervals of* 20%, 50% *and* 80% *nominal coverage rates. Between horizons 39 and 47, at a ramp event, eight consecutive observations lie out of the* 80% *prediction interval, which make us feel that uncertainty estimation is not "locally" reliable.*

Chapter 3

Development of quantile forecasting model, conditionnally to ramp information

3.1 Definition and detection of a ramp event.

In this section we discuss how to obtain a satisfying mathematical definition of a ramp event, from the basic idea we developed at the end of the previous chapter, where we introduced as a ramp : "*a high and sharp variation in the power production*". The aim is, at a first stage, to develop an automatic procedure for ramp events detection. Here the term *detection* is not used in the sense of *prediction*. The aim of ramp detection is to collect information from a wind power time series $\{p_t,\ t \in \mathbb{T}\}$ seen as a numerical signal, such as the number of ramp event occurrences in a data set, their intensity...

The best tool to detect a ramp event on a (graphically represented) numerical signal might be the human eye. Indeed with a simple glance, the eye is able to filter out extraneous noise and to decompose a signal in its principal trends, just with such simple patterns. Galati et al. noticed it in [23] and developed an algorithm named ALESDA (which stands for *Automatic Least squares Error Signal Decomposition Algorithm*), which permits to obtain a such approximation, by decomposing a signal in a sum of weighted an translated primitives (figure (3.1)).

However the huge amount of data (and the inability of the human eye to maintain a consistent interpretation for large amount of data over a long period of time) brings us to consider an automatic way to detect ramp events.

There are very few documents, in wind power forecasting literature, which deal with ramp events. In [24] Parks et al. defined a ramp event as a variation in power output exceeding a minimum size of s_{min} (in percent of the wind farm installed capacity), over a time period less than or equal to a maximum duration d_{max}. They calculated the frequency of time taken up by ramp events over a long observation time. They concluded, and that is not to surprise us,

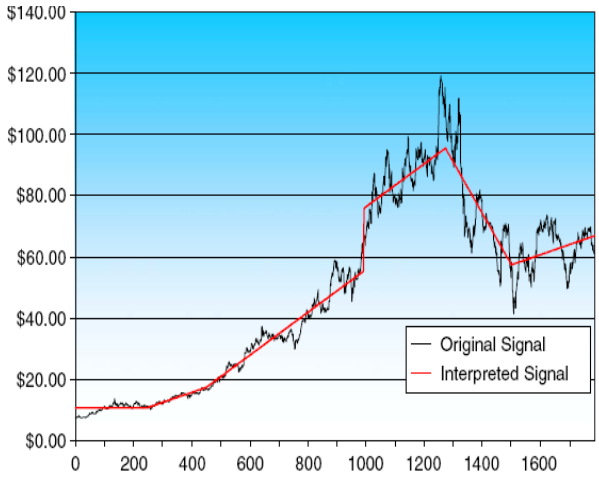


Figure 3.1: *A 1785 data points signal generated by a recording of the closing price of a share of Microsoft stock (over a period of 7 years beginning January 1, 1995 and before the 2:1 split of 2003). Superimposed on it an approximation given by the ALESDA algorithm. This figure is from [23]*

that this frequency was decreasing almost exponentially in function of s_{min}. The smaller was the parameter d_{max} the higher was the decrease.

The notion of ramp event is clearly linked to the mathematical derivative one : a ramp coincides with a period of high derivative values (positive values for increasing ramp and negative ones for decreasing ones) of the power output signal. Usually first order derivative are approximated by *divided differences* in numerical (discrete) calculus. Denoting by $f : \mathbb{R} \longrightarrow \mathbb{R}$ a real differentiable function, and $\{x_0, \ldots, x_N\}$ a sample of regularly spaced real numbers (i.e : $x_i - x_{i-1} = h > 0$, $\forall i \in \{1, \ldots, N\}$), one has for $i \in \{1, \ldots, N\}$ and $j \in \{0, \ldots, N-1\}$:

$$f'(x_i) = \frac{f(x_i) - f(x_{i-1})}{h} + O_b(h) \tag{3.1}$$

$$f'(x_j) = \frac{f(x_{j+1}) - f(x_j)}{h} + O_f(h) \tag{3.2}$$

In the second member of (3.1) and (3.2), the first term is a first order divided difference which is often denoted by $f[x_{i-1}, x_i]$ (resp. $f[x_j, x_{j+1}]$). The letter b (resp. f) in O_b (resp. O_f) stands for backward (resp. forward), one talks about first order backward (resp. forward) divided difference. Actually these equations come from the differentiation of the Newton interpolation formula given below :

$$f(x) = p_1(x) + E_1(x) \tag{3.3}$$

$$= f(x_{i-1}) + f[x_{i-1}, x_i](x - x_{i-1}) + E_1(x) \tag{3.4}$$

where p_1 is the Newton interpolation affine function associated to knots

$(x_{i-1}, f(x_{i-1})), (x_i, f(x_i))$ and E_1 the associated interpolation error . When differentiating and evaluating in x_i, p_1 becomes the first order backward divided difference $p'_1(x_i) = f[x_{i-1}, x_i]$ and E_1 the approximation term O_b. An evaluation in x_{i-1} respectively leads to the forward divided difference. More generally it is possible to approximate $f'(x_i)$ by $p'_l(x_i)$ when using $l+1 > 2$ knots (p_l being the l degree polynomial function of interpolation), the approximation error is then of the order $O(h^l)$. To have more information on numerical differentiation one can see [25].

To define a ramp event, it seems necessary to introduce two parameters : one associated to time and one to the amount of power output. For instance in [24], Parks et al. used the size s_{min} and the maximum duration d_{max}. In this paper one denotes by n_{am}[1] the first parameter, which is expressed in number of measures (and so corresponds to time) and by τ the other one expressed in percent of the wind farm nominal power. From our signal of power outputs $\{p_t,\ t \in \mathbb{T}\}$, one defines a second one as below :

$$p_t^f = \left| \frac{1}{n_{am}} \sum_{h=1}^{n_{am}} p_{t+h} - \frac{1}{n_{am}} \sum_{h=0}^{n_{am}-1} p_{t-h} \right| \tag{3.5}$$

Actually p_t^f is the gap between the mean power calculated with the n_{am} power values preceding the instant t (including it) and the n_{am} power values succeeding this moment. One can rewrite (3.5), to make appear a sum of first order divided differences :

$$\begin{aligned} p_t^f &= \left| \frac{1}{n_{am}} \sum_{h=1}^{n_{am}} p_{t+h} - \frac{1}{n_{am}} \sum_{h=0}^{n_{am}-1} p_{t-h} \right| \\ &= \left| \frac{1}{n_{am}} \sum_{h=1}^{n_{am}} p_{t+h} - \frac{1}{n_{am}} \sum_{h=1}^{n_{am}} p_{t+h-n_{am}} \right| \\ &= \left| \sum_{h=1}^{n_{am}} \frac{p_{t+h} - p_{t+h-n_{am}}}{n_{am}} \right| \end{aligned} \tag{3.6}$$

Formulation (3.6) clearly reveals the existing relation between p_t^f and the power output derivative around the instant t. Indeed p_t^f is (forgetting the absolute value) the sum of n_{am} divided differences calculated with couples of power measures spaced in time by n_{am} observation times. Finally to compute p_t^f, one needs every power measure in the time interval $[t - n_{am} + 1, t + n_{am}]$. Obviously n_{am} plays the role of a time scale parameter, by being both the number of divided differences added in the sum (3.6) and the time gap between two power measures used to calculate each of these divided differences. Small values of n_{am} lead to a localized study of the derivative around the instant t. On the other hand, higher values of n_{am} are synonym of larger time scales (see figure (3.2)).

High values of the power output derivative in a ramp, lead to high values of divided differences used in the *filtered* signal definition. From the position in a ramp, depends (among others) the number of divided differences in (3.6)

[1] "am" stands for "averaged measures"

having an important contribution to p_t^f value. Typically at a ramp, the filtered signal increases until a local maximum coinciding with an inflection point of the power output signal.

The detection of a ramp relies on the introduction of a threshold (our second parameter) τ. One will affirm that, at the instant t corresponds a ramp, if the value of the filtered signal is higher than that threshold, i.e :

$$p_t^f \geq \tau \tag{3.7}$$

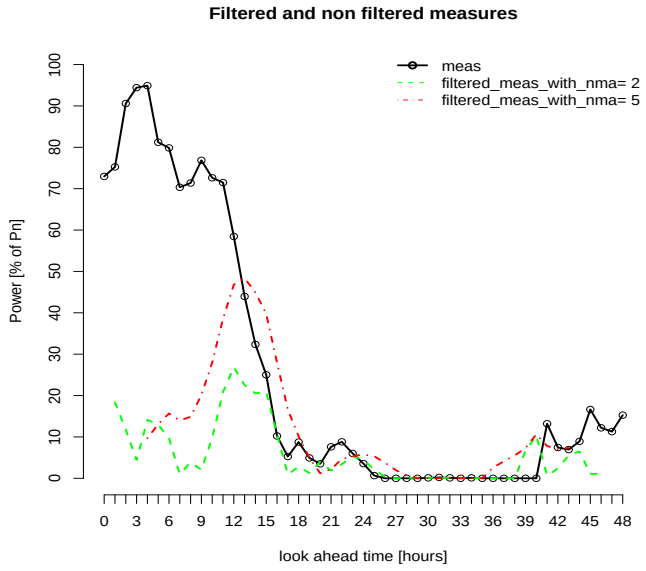


Figure 3.2: *Example of ramp detections. In black : power measures output signal and in color : filtered signals. The filtered signal with higher n_{am} value ($n_{am} = 5$, in red) seems to better detect ramps and being less sensitive to local variations.*

3.1.1 Application to case study

In this section, we used the ramp event detection method described above on a set of two wind farms WF1 and WF2 which characteristics are given in the table below.

wind farm	Pn(MW)	beginning date	ending date	#observations
WF1	$\simeq 5$	05-02-2001	14-08-2002	13330
WF2	$\simeq 20$	01-01-2003	20-07-2004	13608

In this study, and as we introduced it in the previous section, one considers that a ramp is detected when filtered signal $\mathbf{p^f} = \{p_t^f, t \in \mathbb{T}\}$ exceeds the threshold τ. More precisely since $\mathbf{p^f}$ has exceeded τ at a moment t_0 and until it goes back under τ at a moment $t_1 > t_0$, only one ramp event is counted. Even if the time interval $[t_0, t_1]$ doesn't match exactly with the ramp event duration, they clearly evolve in the same way. From now on, one calls **ramp intensity**

the local maximum reached by the filtered signal at a ramp event. Figure (3.3) shows the ramp intensities distribution, for four values of the parameter n_{am} : $n_{am} = 2, 5, 8$ and 12 and τ fixed at 25% of the wind farm installed capacity.

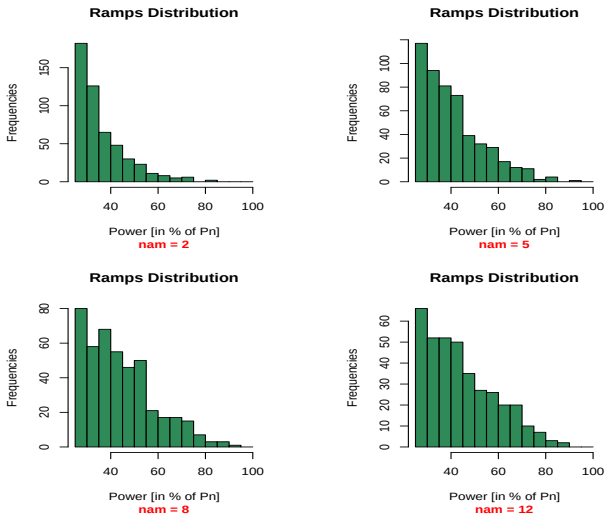


Figure 3.3: *Ramp event intensities distribution at wind farm WF1. There is one histogram for each value of* $n_{am} \in \{2, 5, 8, 12\}$ *and* τ *is fixed at* 25% *of the wind farm installed capacity.*

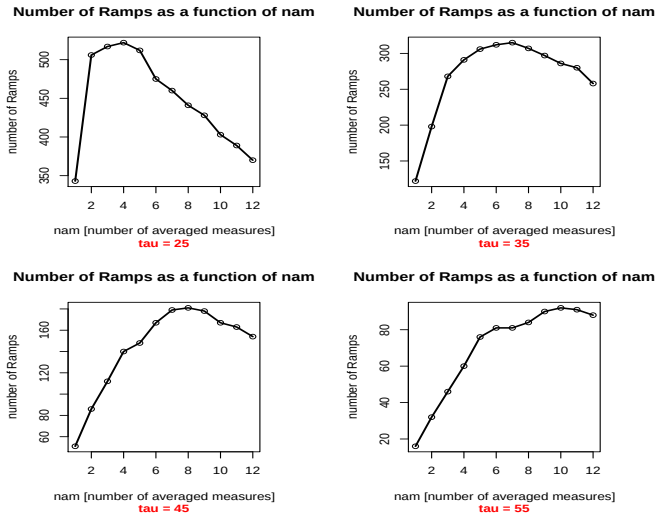


Figure 3.4: *The number of detected ramps as a function of* n_{am}*, for four values of the threshold* $\tau \in \{25, 35, 45, 55\}$ *(% of the installed capacity). For each value of* τ*, there is a maximum number of detected ramps, but this maximum is not reached for the same value of* n_{am}*.*

An unsurprising observation is that ramp events of high intensity are less common that ones with small intensity. Figure (3.4) shows the number of ramp events as a function of n_{am}, for four values of the threshold $\tau \in \{25, 35, 45, 55\}$ percent of the installed capacity. It has to be noticed that to each value of

τ, corresponds a maximum number of detected ramps, and that maximum is not reached for the same value of n_{am}. Indeed it seems that the value of n_{am}, maximizing the number of detected ramps, increases with τ. Results are for wind farm WF1, and same conclusions can be done for WF2 (see appendix 4.1).

These results don't help us to find best parameters value, to be used in coming applications. For instance, if one fixes τ, what value does one have to choose for the n_{am} parameter ? Is the value of n_{am}, maximizing the number of detected ramps, the best choice ? In any case one can already except large values of the parameter n_{am}, which seem not to be relevant, because they coincide with too long time scales for studying ramp events.

3.1.2 On the temporal uncertainty.

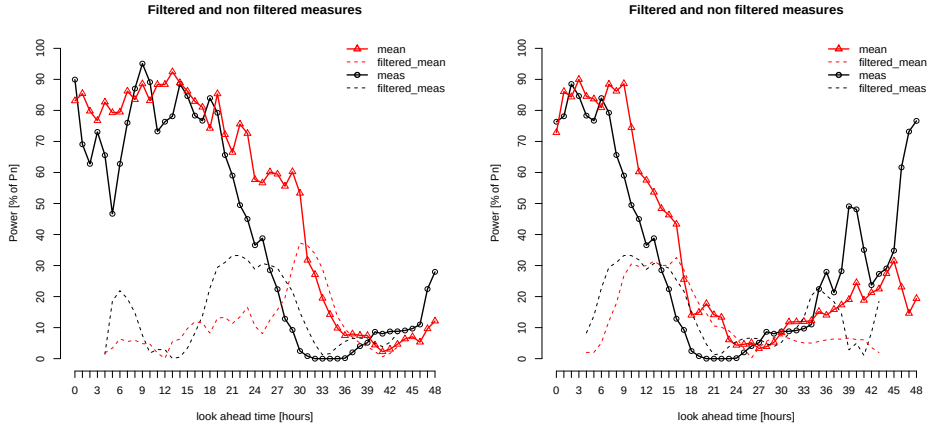
In [24] Parks et al. made an analysis of the temporal uncertainty in ramp events forecasting. Defining a ramp event as a variation of 50% or more of the wind farm nominal power in a time duration less than or equal to 4 hours; they considered that two ramps, one on observations signal and the other one on predictions signal, were associated to the same event if they occurred in a lapse of time less than or equal to 12 hours. If several ramps were detected on observations for a single one on predictions, the closest was chosen (and vice versa). In that way, only one ramp on observations could be associated to one on predictions (and vice versa). They used the instant at which a ramp occurred as a reference point, to compute separating time between observations and predictions ramps. By using fixed look-ahead time predictions, they obtained a fixed look-ahead time distribution of time-lags existing between measured and forecasted ramp, and so a measure of the temporal uncertainty in ramp events forecasting.

Using filtered signals for ramps detection, one wanted to study the phase error attenuation by actualization of predictions. Unfortunately, besides difficulties of implementation, we were faced with the problem of choosing a good point of reference to calculate time-lag existing between observations and predictions ramps associated to the same event. Indeed, to choose the local maximum of the filtered signal may lead to huge approximations, and results would be of poor interest. This lack of accuracy must be associated to ramps geometry which is not perfectly determined and evolves with actualizations. So does the filtered signals geometry (see figure (3.5)). So it appears that to measure the time-lag existing between observations and predictions ramps is maybe not the good way to evaluate a phase error.

3.2 A model to improve forecasting quantile conditionally to ramps information.

3.2.1 Introduction to the model.

In this section we focus on how to better *predict* ramp events, using the detection method developed in the previous part. While statistic inference is synonym of estimators construction from historical past data, why would not one try to "learn from the future" ? Our idea relies on extracting information



(a) before actualization **(b)** after actualization

Figure 3.5: *Example of a phase error attenuation after actualization of predictions (12 hours later). If one calculates time-lags between observations and predictions ramps, from the time separating local maxima of filtered signals, one would not catch the phase error attenuation. In (a) the time-lag would be of 9 hours with maxima located at horizons 21 for observations and 30 for predictions. In (b), the time-lag is still of 7 hours, with maxima at horizons 9 and 16 respectively.*

on eventual ramps apparition in the future from deterministic forecasts, and then try to use this information to better estimate associated uncertainty.

More precisely we first produce at the instant t and for the all range of horizons $h \in \{1, \dots, h_{max}\}$ *preliminary* deterministic forecasts. In the scope of our study, this deterministic forecasts are based on the estimation of the predictive wind power distribution mean. In a second stage the associated uncertainty estimation is provided. We use nonparametric quantile regression leading to a set of per look-ahead times quantiles estimators. To the usual explanatory variables used to construct such estimators : meteorological forecasts and past measured power values, we added two more inputs extracted from deterministic forecasts cited above : "the distance and the intensity of the nearest predicted ramp".

Let $\mathbb{T} = \mathbb{T}_1 \sqcup \mathbb{T}_2$ be the studied period split into the *training* period $\mathbb{T}_1$ followed by the *test* period $\mathbb{T}_2$. If writing $\mathcal{X}_t \in \mathbb{R}^q$ the set of q regressors used for predictions at the instant t, one has respectively the training and the test sets : $(p_{\mathbb{T}_1}, \mathcal{X}_{\mathbb{T}_1})$ and $(p_{\mathbb{T}_2}, \mathcal{X}_{\mathbb{T}_2})$. Finally the proposed model comes down to a multi-stage process of prediction, which each step is described below :

- A set of estimators $\{\hat{f}_h^d\}_{h \in \mathbb{H}}$[2], constructed from $(p_{\mathbb{T}_1}, \mathcal{X}_{\mathbb{T}_1})$ and evaluated from $(p_{\mathbb{T}_2}, \mathcal{X}_{\mathbb{T}_2})$ gives our preliminary deterministic forecasts $\hat{p}_{\mathbb{T}}^d$ for all times of the studied period.

- Keeping notations introduced in the previous section, one calculates the filtered signal associated to these predictions, in order to detect eventual predicted ramps : $\hat{p}_{\mathbb{T}}^{d,f}$

[2] $\mathbb{H}$ is the set of look-ahead times

- From the filtered predictions $\hat{p}^{d,f}_{\mathbb{T}}$, one extracts two new inputs : "the distance and the intensity of the nearest predicted ramp" :
  $D_{t,h}$:= *time between horizon h and an eventual nearest ramp in the "prediction window"* $[t, t + h_{max}]$.
  $I_{t,h}$:= *This last one intensity.*
  Let's give an illustrative example. Suppose $\hat{p}^{d,f}_{\mathbb{T}}$ has been calculated using the previously introduced parameters τ and n_{am}. If at the instant $t + h_0$ there isn't any ramp in our predictions ($\hat{p}^{d,f}_{t+h_0/t} < \tau$), but that the nearest one in the range of look-ahead times is at the instant $t + h_1$ ($h_1 = \arg\min |h_0 - h|$, $h \in \{1, \ldots, h_{max} | \hat{p}^{d,f}_{t+h/t} \geq \tau\}$), then $D_{t,h_0} = |h_0 - h_1|$ and $I_{t,h_0} = \hat{p}^{d,f}_{t+h_1/t}$.

- Then one constructs estimators providing quantiles estimation (and so the uncertainty estimation) $\{\hat{f}^p_h\}_{h\in\mathbb{H}}$ from $(p_{\mathbb{T}_1}, \{\mathcal{X}_{\mathbb{T}_1}, \mathcal{D}_{\mathbb{T}_1}, \mathcal{I}_{\mathbb{T}_1}\})$, and evaluates them from $(p_{\mathbb{T}_2}, \{\mathcal{X}_{\mathbb{T}_2}, \mathcal{D}_{\mathbb{T}_2}, \mathcal{I}_{\mathbb{T}_2}\})$.

- Finally one gets probabilistic predictions :

$$\hat{p}^p_{t+h/t} = \hat{f}^p_h(\mathcal{X}_{t+h/t}, \mathcal{D}_{t+h/t}, \mathcal{I}_{t+h/t})$$

One has to notice that the value of the two created inputs $D_{t,h}$ and $I_{t,h}$ depends on the parameters τ and n_{am}, their construction relying on the filtered deterministic predictions signal $\hat{p}^{d,f}_{\mathbb{T}}$. Actually one has $D_{t,h} = D^{\tau,n_{am}}_{t,h}$ and $I_{t,h} = I^{\tau,n_{am}}_{t,h}$.

Of course the procedure doesn't really learn from the future, the two new inputs come from deterministic predictions which are build from past information. So what could we hope to learn of more with them ? One has to notice that $\mathcal{D}_{t,h}$ and $\mathcal{I}_{t,h}$, used for predictions made at the instant t and for the horizon h, are generated looking at the predictions temporal window : $[t, t+h_{max}]$. So one may expect that they contain a part of the interdependence information there exists between future outcomes.

3.2.2 The case study.

Data

We still focus on the two wind farms already introduced in the previous section, which main characteristics are reminded in table (3.1).

wind farm	Pn(MW)	beginning date	ending date	#prediction time
WF1	$\simeq 5$	05-02-2001	14-08-2002	2219
WF2	$\simeq 20$	01-01-2003	20-07-2004	1134

Table 3.1: *Main characteristics of wind farms WF1 and WF2 of case study.*

The dates and numbers of prediction times given in (3.1) are for the entire studied period and have to be split into corresponding training and test periods. More precisely the data consist of :

- hourly power production of wind farms

- hourly weather forecasts generated four times a day for wind farm WF1 and two times a day from the NWP model Hirlam10 for wind farm WF2. For the two wind farms, lead times are from +1H to +48h. We selected three variables, commonly known as explanatory :
  - wind speed at 10m
  - wind direction at 10m
  - the last past power measure.

  These three variables were used as inputs of methods for preliminary deterministic forecasts and also in the model of reference for quantiles estimation (also called *basic* model). To these three inputs are added the two new predictors in our *advanced* model. As phase errors are more significant for larger horizons, and in order to lower computational calculus, our study focuses on look-ahead times larger than or equal to +24 hours.

Statistical methods

We used three statistical methods in our study. The preliminary deterministic forecasts were build by the *Random Forests* procedure, whereas the associated uncertainty (quantiles) estimation was made by procedures *Quantile Regression Forests* and *Kernel Quantile Regression* with and without our two new inputs : distance and intensity of eventual predicted ramps in the predictions window. Here comes a brief description of these methods. For further information, one can look at [26] for random forests, [27] for quantile regression forests and [28] for kernel quantile regression. The associated R-packages used are : `randomForest`, `quantregForest` and `kernlab`, which can be found at url : `http://cran.r-project.org/`

Random Forests : In the random forests procedure a large number of trees (regression tree in our case) are grown, each of them using a bagged version of the training data. For each tree and at each node, random forests employs a random subset of predictor variables for split point. The size of the random subset *mtry* is a parameter to tune in the procedure. For regression, the prediction of random forests for a new data point $X = x$ is the averaged response of all trees. Following the notation of Breiman, call θ the random parameter vector that determines how a tree is grown (e.g which variables are considered for split point at each node). The corresponding tree is denoted by $T(\theta)$. Let $\mathcal{B}$ be the space in which X lives, that is $X : \Omega \longrightarrow \mathcal{B} \subseteq \mathbb{R}^p$, where $p \in \mathbb{N}_+$ is the dimensionality of the predictor variable. Every leaf of a tree $l = 1, \ldots, L$ corresponds to a rectangular subspace of $\mathcal{B}$. Denote this rectangular subspace by $R_l \subseteq \mathcal{B}$ for every leaf $l = 1, \ldots, L$. For every $x \in \mathcal{B}$, there is one and only one leaf l such that $x \in R_l$ (corresponding to the leaf that is obtained when dropping x down the tree). Denote this leaf by $l(x, \theta)$ for tree $T(\theta)$.

The prediction of a single tree $T(\theta)$ for a new data point $X = x$ is obtained by averaging over the observed values in leaf $l(x, \theta)$. Let the weight vector $w_i(x, \theta)$ be given a positive constant if observation X_i is part of the leaf $l(x, \theta)$ and 0 if not. The weights sum to one, and thus

$$w_i(x,\theta) = \frac{\mathbb{1}_{\{X_i \in R_{l(x,\theta)}\}}}{\#\{j : X_j \in R_{l(x,\theta)}\}} \tag{3.8}$$

The prediction of a single tree, given covariate $X = x$, is then the weighted average of the original observations $Y_i,\ i = 1, \ldots, n$

$$\text{single tree} : \hat{\mu}(x,\theta) = \sum_{i=1}^{n} w_i(x,\theta) Y_i \tag{3.9}$$

Using random forests, the conditional mean $E(Y|X = x)$ is approximated by the averaged prediction of k single trees, each constructed with an i.i.d vector $\theta_t,\ t = 1, \ldots, k$. Let $w_i(x)$ be the averaged of $w_i(x,\theta)$ over this collection of trees,

$$w_i(x) = \frac{1}{k} \sum_{t=1}^{k} w_i(x,\theta_t) \tag{3.10}$$

The prediction of random forests is then

$$\text{Random forests} : \hat{\mu}(x) = \sum_{i=1}^{n} w_i(x) Y_i \tag{3.11}$$

The approximation of the conditional mean of Y, given $X = x$, is thus given by a weighted sum over all observations. The weights vary with the covariate $X = x$.

Quantile Regression Forests : The quantile regression forests procedure is based on the random forests procedure. Actually an estimate $\hat{Q}_\alpha(x)$ of the α-conditional-quantile $Q_\alpha(x) = \inf\{y : F(y|X = x) \geq \alpha\}$ of the probability law of Y, such that $X = x$, is obtained from the estimation the procedure gives of the conditional cumulative distribution function of Y, given below :

$$F(y|X = x) = \mathbb{P}(Y \leq y|X = x) = \mathbb{E}(\mathbb{1}_{\{Y \leq y\}}|X = x) \tag{3.12}$$

In the random forests procedure, the conditional expectation $\mathbb{E}(Y|X = x)$ is approximated by a weighted mean over the observations of Y. Similarly, in the quantile regression forests procedure, the conditional cumulative distribution function is approximated by an identically weighted mean over the observations of $\mathbb{1}_{\{Y \leq y\}}$

$$\hat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) \mathbb{1}_{\{Y_i \leq y\}} \tag{3.13}$$

where $w_i(x)$ are the same weights as for random forests, defined in (3.10). Finally the estimate $\hat{Q}_\alpha(x)$ is obtained by plugging $\hat{F}(y|X = x)$ instead of $F(y|X = x)$ in the conditional quantile definition :

$$\hat{Q}_\alpha(x) = \inf\{y : \hat{F}(y|X = x) \geq \alpha\} \tag{3.14}$$

Kernel Quantile Regression : Let $\{(y_i, \mathbf{x}_i), i = 1, \ldots, n\}$ be the training data set, where $\mathbf{x} \in \mathbb{R}^p$ is the vector of predictors and $y \in \mathbb{R}$ the real value response variable. The kernel quantile regression procedure permits to construct an estimator $\hat{f}$, which to a new point data $X = x$ associates an approximation of the conditional quantile $Q_\alpha(x)$ of Y, such that $X = x$. $\hat{f}$ comes from the resolution of the following penalized problem :

$$\min_{f \in \mathcal{H}_K} \sum_{i=1}^{n} L_\alpha(y_i, f(\mathbf{x}_i)) + \frac{\lambda}{2} ||f||^2_{\mathcal{H}_K} \tag{3.15}$$

where L_α is the pinball loss defined in (2.26), and $\mathcal{H}_K$ is a structured reproducing kernel Hilbert space (RKHS) generated by a positive definite kernel $K(\mathbf{x}, \mathbf{x}')$ (see [2] and references there are into).

It is of common knowledge that despite $\mathcal{H}_K$ is a infinite dimensional space, the solution of (3.15) has the finite form :

$$\hat{f}(\mathbf{x}) = \beta_0 + \frac{1}{\lambda} \sum_{i=1}^{n} \theta_i K(\mathbf{x}, \mathbf{x}_i) \tag{3.16}$$

One can rewrite (3.15) using the finite form defined above in (3.16) and then the problem becomes :

$$\min_{\beta_0, \theta} \sum_{i=1}^{n} L_\alpha(y_i, \beta_0 + \frac{1}{\lambda} \sum_{j=1}^{n} \theta_j K(\mathbf{x}_i, \mathbf{x}_j)) + \frac{1}{2\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} \theta_i \theta_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{3.17}$$

This last model (3.17) called kernel quantile regression (KQR) can be transformed into a quadratic programming problem; thus most commercially available packages can be used to solve the KQR. All of the resolution algorithms resolves the KQR for a prefixed regularization parameter λ. In practise, λ is selected from a predefined finite set of values covering a wide range, with a certain model selection criterion. This last point is discussed in [28].

Tuning of parameters in case study : For random forests and quantile regression forests procedures used in our study, we chose per default values of parameters, computed by the R functions : $mtry = \lceil p/3 \rceil$ (the size of the random subset of predictor variables for splitpoint), $ntree = 1000$ ($ntree$ being the number of trees grown) and $nodesize = 10$ (the minimum size of terminal nodes. Setting this number larger causes smaller trees to be grown).

The regularization parameter λ, as denoted in the above description of kernel quantile regression, has been selected by a *cross-validation* procedure (see [2]) done on a subset of the test set. The predefined range of values was : $2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 1/2, 1, 2$. We chose to preselect a set of quite sparse values, but took only few to have reasonable computational calculus times.

Introduction of a new forecasts verification framework

As our study focuses on forecasting uncertainty related to ramp events, we introduced an evaluation conditionally to the presence of such ramps. Usually the evaluation is done for each look-ahead time, considering all prediction times of the test period. In the case of a conditionally to ramp events evaluation, for

each look-ahead time, only prediction times for which was detected a ramp on observations were selected to do the evaluation. This prediction times selection is made relying on the detection method presented in the first section of the chapter, and so depends on the two parameters which were introduced : τ and n_{am}. As a result, each look-ahead time has an individual set of selected prediction times building up its individual test set. In addition these individual test sets depend on the value of τ and n_{am}. In table (3.2) is given for wind farm WF1 and studied values of parameters n_{am} and τ, the maximum Max and the minimum Min sizes of such sets of prediction times, considering the look-ahead times of case study. Sum is the total number of selected prediction times, when gathering look-ahead times.

Parameters value	Min	Max	Sum
$\tau = 25, n_{am} = 5$	181	203	3869
$\tau = 25, n_{am} = 3$	105	143	2469
$\tau = 40, n_{am} = 5$	56	66	1201

Table 3.2: *Number of selected prediction times for the conditionally to ramp evaluation. Values of parameters n_{am} and τ are the ones of case study. Results are for wind farm WF1.*

Min and Max are quite close to each other whatever the parameters value, which means that almost the same number of prediction times are selected for each look-ahead time. A significant number of consecutive look-ahead times being in a same ramp at a given prediction time, this last one belongs to their respective individual test set. So we can suppose that these sets are of nearby size because they often contain the same selected prediction times. Unfortunately we judged Min to small to envisage a conditionally to ramp evaluation for each look-ahead time, and so our evaluation brings look-ahead times together. This observation is even more true with wind farm WF2, where values of Min and Max are rather lower. Table of wind farm WF2 is given in appendix (5.1).

3.2.3 Analyze of the results.

Following results are for data and forecasting methods with tuned parameters described above. We compare so a *basic* and an *advanced* model, differentiated by the use or not of predictor variables $\mathcal{D}_{t,h}$ and $\mathcal{I}_{t,h}$. Seven quantiles were estimated : the 10, 25, 40, 50, 60, 75 and 90 percentiles. In order to study the influence of parameters n_{am} and τ we did the study for three couples of values : $(\tau, n_{am}) \in \{(25,5);(25,3);(40,5)\}$.

Preliminary deterministic forecasts Before analyzing results and giving any conclusion on our probabilistic models' performance, we first have to make sure that the preliminary deterministic forecasts from which we extracted the two new inputs of our advanced probabilistic models, reach an acceptable performance level. From these forecasts' quality depend the relevance of the information put in our new inputs. In figure (3.6) is given the NRMSE and the NMAE (normalization of the RMSE and MAE defined in (2.13) and (2.14)) of such preliminary deterministic forecasts for wind farm WF1 (see appendix (5.2) for wind farm WF2).

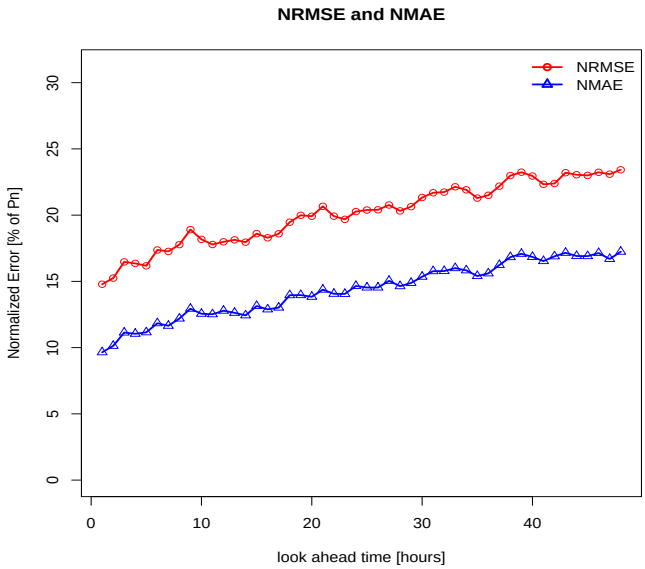


Figure 3.6: *NRMSE and NMAE of preliminary deterministic forecasts obtained with random forest method and per default parameters value. Results are for wind farm WF1.*

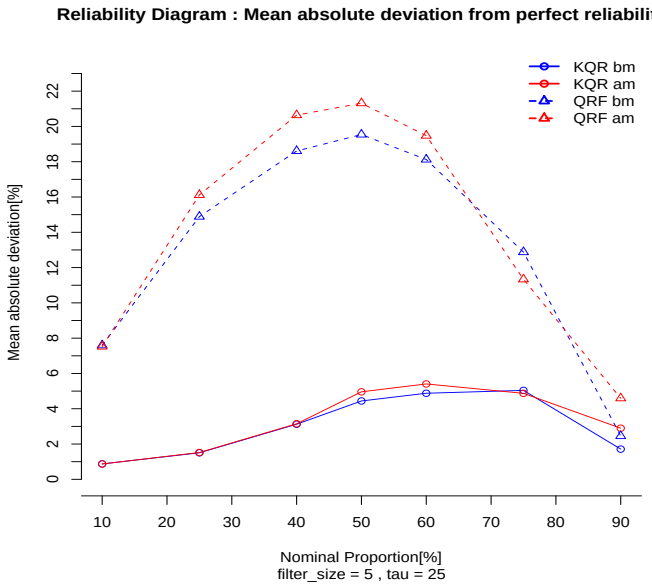
To know what performance can be expected by short-term wind power deterministic prediction models, an evaluation of operational ones used in several European countries (such as Spain, Denmark, Ireland, Germany...) have been made in the frame of the project ANEMOS [29]. The evaluation gathered a large number of case-studies with wind farms from flat, semi-complex and complex terrains, offshore ones, as well as meteorological predictions from several NWP systems. Focusing on the NMAE, our results compound with the ones given in [29], here it doesn't exceed 15% of the installed capacity. The NRMSE which is more sensitive to large prediction errors is at a higher level but results are still rather good (see [1]).

Overview on the probabilistic models' performance level We now focus on results of probabilistic models. Figure (3.7) gives the mean absolute deviation from perfect reliability (MADFPR) for the Kernel Quantile Regression method (KQR) and the Quantile Regression Forests method (QRF) for the basic model (**bm** : blue curve) and the advanced one (**am** : red curve) with parameters value $(25, 5)$. Keeping the notations introduced in the previous chapter : $\hat{a}_h^\alpha$ being the empirical coverage rate associated to the α-quantile estimation for look-ahead time $h \in \mathbb{H}$, the mean absolute deviation is then :

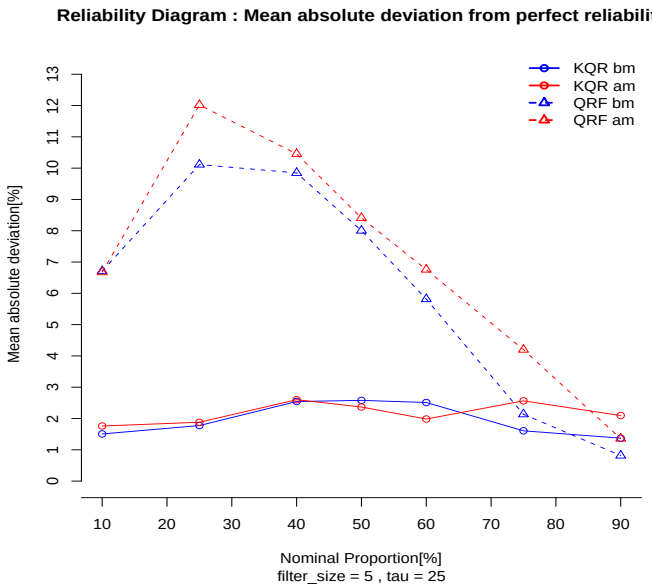
$$d^\alpha = \frac{1}{\#\mathbb{H}} \sum_{h \in \mathbb{H}} |\alpha - \hat{a}_h^\alpha| \tag{3.18}$$

Immediately one observes the poor accuracy of quantiles estimations made by the quantile regression forests procedure. Indeed when the MADFPR never exceeds $5 - 6\%$ for the kernel quantile regression method, it reaches a value of

about 20% for the quantile regression forests method when looking at quantiles estimations associated to medium range nominal proportions at wind farm WF1.



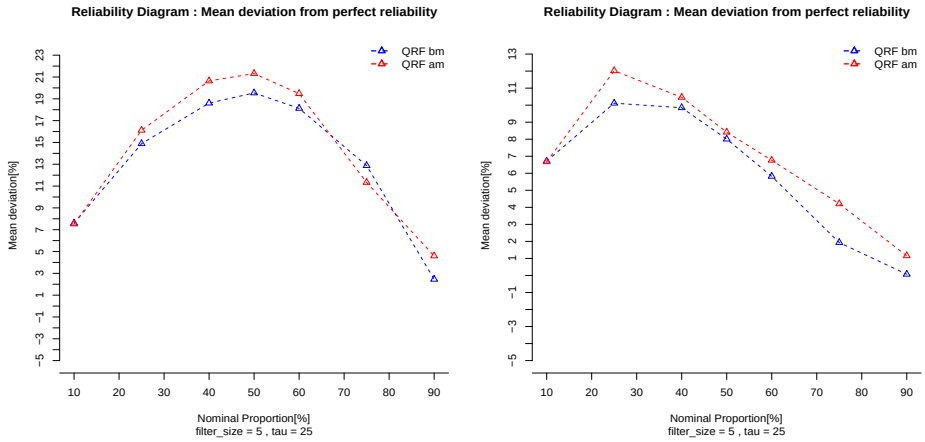
(a) wind farm WF1



(b) wind farm WF2

Figure 3.7: *Mean absolute deviation from perfect reliability for the Kernel Quantile Regression (KQR) and the Quantile Regression Forests (QRF) methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 5)$.

The general low level of performance of the QRF method is observed for all models (basic or not) and wind farm considered, even if in an overall manner results are better (but still not acceptable) for wind farm WF2 (it was already the case for deterministic forecasts and must be connected to the terrain complexity). One could notice the surprising change of the MADFPR curve shape for the advanced model with parameters value $(40, 5)$ at wind farm WF1 (see appendix (5.3)).



(a) QRF at wind farm WF1 **(b)** QRF at wind farm WF2

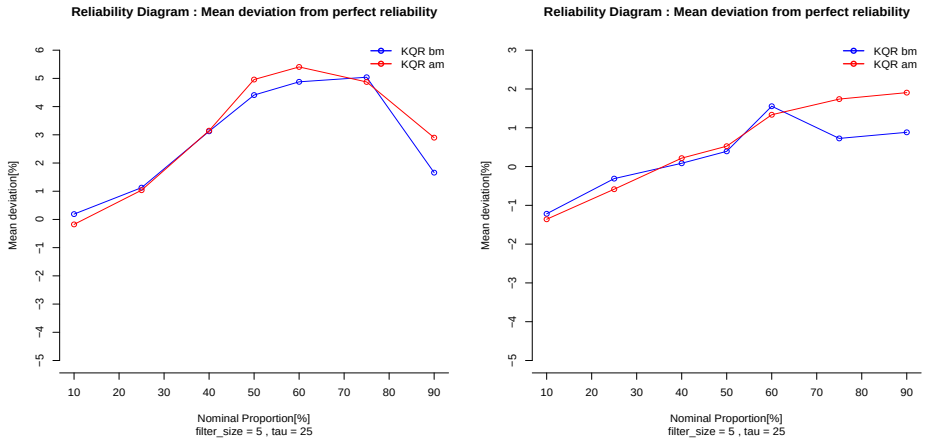
Figure 3.8: *Mean deviation from perfect reliability for the Quantile Regression Forests method (QRF). The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 5)$.

Whereas positive and negative deviations are not differenciated in the MADFPR, the mean deviation from perfect reliability (MDFPR) accounts for negative deviations balancing with positive ones. Figure (3.8) shows MDFPR results, for the QRF method, that seem very close to the ones obtained with the MADFPR. Its positivity comes from the large positive deviations predominance. Positive deviations imply that the proportions associated to estimated quantiles are lower than the nominal ones, and so that quantiles are underestimated. The observation holds for all studied parameters values (see appendix (5.4)).

There is a striking contrast between the good performance level of deterministic forecasts on a hand, and the bad performance level of the QRF method's forecasts on the other hand. Indeed the apparent similarity between the random forests and the quantile regression forests procedures (as described previously), all the more they were used with the same inputs (considering a basic model's point of view) and the same value of parameters, had lead us up to expect more statisfying results resulting from the QRF procedure.

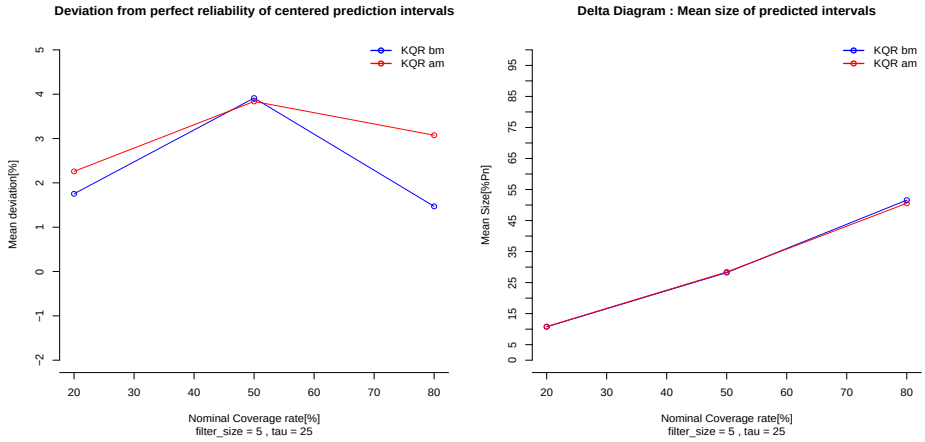
The performance level achieved by the KQR method is far better, with a positive mean deviation (figure (3.9)) which doesn't exceed $5 - 6\%$ for wind farm WF1 and a mean deviation lying between -1% and 2% for wind farm WF2. On the contrary of the QRF method, one can not conclude on a general underestimation of quantiles, for all nominal coverage rates and wind farms.

The performance level of the advanced model stays roughly the same whatever the parameters τ and n_{am} value (see appendix (5.4)), and quite close of the basic model's one.



(a) KQR at wind farm WF1 **(b)** KQR at wind farm WF2

Figure 3.9: *Mean deviation from perfect reliability for the Kernel Quantile Regression method (KQR). The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 5)$.



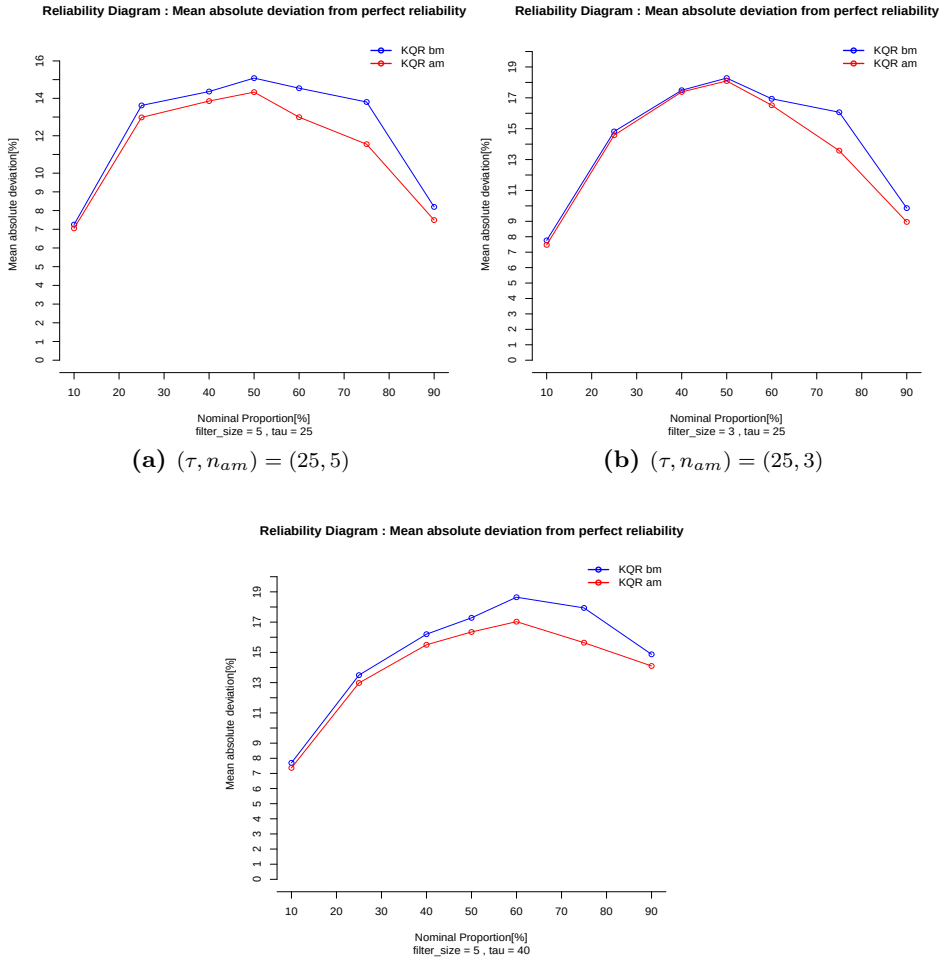
(a) KQR at wind farm WF1 **(b)** KQR at wind farm WF2

Figure 3.10: *Mean deviation from perfect reliability and mean size of centered prediction intervals, estimated with the Kernel Quantile Regression method (KQR). The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 5)$.

The overall good agreement between basic and advanced models, in the case of the KQR method, is still true when looking at the mean deviation from perfect reliability and the mean size of centered prediction intervals (figure

(3.10)). The mean size of centered prediction intervals of nominal coverage rate 80% is always of about 50% for wind farm WF1 and 35% for wind farm WF2. Here again, the higher level of prediction uncertainty for wind farm WF1, must be linked to the terrain complexity.

Conditionally to ramp events evaluation (KQR method) If the bad performance level of QRF's forecasts leaves them out of any further analyze, it is not the case for the KQR-based models. While the global evaluation didn't highlight the potential interest of using the recently introduced inputs of the advanced model, we need a conditionally to ramp events evaluation to observe a possible improvement in forecasting uncertainty at such ramps. From now, results are for a such conditional evaluation.

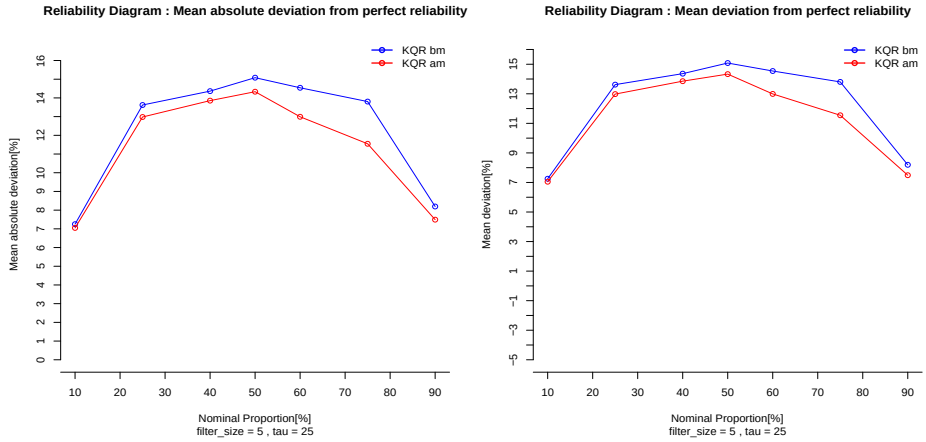


(a) $(\tau, n_{am}) = (25, 5)$ **(b)** $(\tau, n_{am}) = (25, 3)$

(c) $(\tau, n_{am}) = (40, 5)$

Figure 3.11: *Mean absolute deviation from perfect reliability for the Kernel Quantile Regression method (KQR), when conditionally to ramps evaluated. The blue curve is for the basic model and the red one for the advanced one. Results are for wind farm WF1.*

Figure (3.11) represents the mean absolute deviation from perfect reliability for the KQR method at wind farm WF1. At a glance, one sees the extremely poor accuracy of quantiles estimation, when evaluated conditionally to ramp events. Indeed the mean absolute deviation lies between proportions 7% and 19% (even 30% for wind farm WF2). The rounded form of curves shows a larger deviation for middle range proportions which are so associated to less reliable quantiles estimations. Conclusion stands for all models and wind farms (see appendix (5.7) for WF2 results).



(a) MADFPR **(b)** MDFPR

Figure 3.12: *Mean deviation and mean absolute deviation from perfect reliability for the Kernel Quantile Regression method (KQR), when conditionally to ramps evaluated. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 5)$ *and wind farm WF1.*

Comparing the mean deviation with the mean **absolute** deviation, one can't achieve to find any difference (see figure (3.12), one will be on the watch for the changes in the vertical axis thick marks). That means that for every nominal coverage α, all deviations $\{\alpha - \hat{a}_h^{\alpha},\ h \in \mathbb{H}\}$ are positive. Normally we would conlude on quantiles underestimation for each nominal coverage rate and look-ahead times, but because of the too small size of the horizons' individual test sets, we can just affirm it for each nominal proportion. The same conclusion stands for all models and wind farms. As an example, table (3.3) gives empirical proportions associated to estimated percentiles for each nominal coverage rate and for the two models. Results are for wind farm WF1, with $\tau = 25$ and $n_{am} = 5$.

Model\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
Basic	2.75	11.38	25.64	34.92	45.46	61.20	81.80
Advanced	2.95	12.02	26.15	35.67	47.01	63.45	82.51

Table 3.3: *Empirical proportions associated to estimated percentiles, for basic and advanced models and the conditionally to ramps evaluated KQR method at wind farm WF1. The value of parameters is* $(\tau, n_{am}) = (25, 5)$.

In order to evaluate the improvement of the advanced model, we calculated the improvement $Imp_b^{d^\alpha}$ (defined in (2.15)) for the mean (absolute) deviation evaluation criterion : $EC = d^\alpha$, the reference model being the basic one : $ref = b$,

$$Imp_b^{d^\alpha}(\tau, n_{am}) = \frac{Imp_b^{d^\alpha}(\tau, n_{am}) - Imp_a^{d^\alpha}(\tau, n_{am})}{Imp_b^{d^\alpha}(\tau, n_{am})} \tag{3.19}$$

Results are given in table (3.4). They are for each nominal coverage rate α and parameters values, at wind farm WF1 (see appendix (5.9) for wind farm WF2).

Parameters\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
$\tau = 25, n_{am} = 5$	.028	.047	.035	.050	.107	.164	.086
$\tau = 25, n_{am} = 3$	.037	.016	.006	.010	.025	.155	.091
$\tau = 40, n_{am} = 5$	.043	.038	.043	.054	.087	.128	.052

Table 3.4: *Improvement of the mean absolute deviation from perfect reliability : $Imp^{d^\alpha}(\tau, n_{am})$, where d^α is defined in (3.18), τ and n_{am} the parameters. Results are for wind farm WF1.*

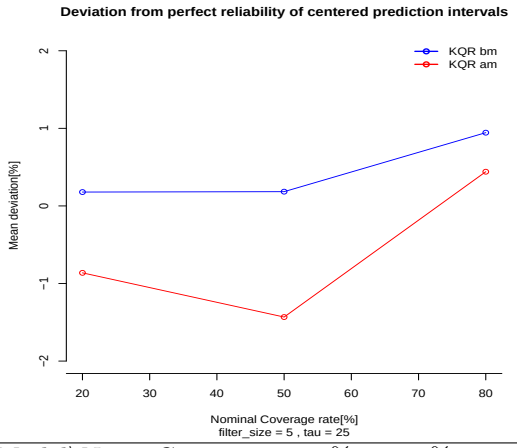
Even little, one observes an improvement for all estimated quantiles and for all parameters values (except the ninety percentile, for parameters value $(40, 5)$ at wind farm WF2). The improvement reaches a level lying between 0.6% and 16.4% (even $24,3$% for wind farm WF2). The larger improvements are for the 60 and 75 percentiles estimations for the two wind farms, which was graphically notable in figure (3.11) and in appendix (5.7). The parameters value doesn't seem to have a significant influence, in the sense that no parameters value has the highest improvement for all nominal coverage rates, at the two wind farms.

Focusing on centered prediction intervals, obtained by associating α and $1 - \alpha$ quantiles estimations, one observes that reliability results are far better than for single quantiles. An example of empirical coverage rates of such intervals is given in table (3.5). By looking at mean deviations of single quantile estimations, one observes that the positive mean deviation from perfect reliability are rather of the same order for α and $1 - \alpha$ quantiles. As a result even if our prediction intervals are not centered, they are quite reliable (see figure (3.13)).

Model\Nom. Coverage	20%	50%	80%
Basic	19.82	49.82	79.06
Advanced	20.86	51.43	79.56

Table 3.5: *Empirical coverage rates of centered prediction intervals, for basic and advanced models. Results are for the conditionally to ramps evaluated KQR method at wind farm WF1, with parameters value* $(25, 5)$.

Even if the estimations of lower and upper bounds of a prediction interval are improved by the advanced model, this is not systematically the case for prediction intervals. If the improvement is not of the same order for both of interval bounds, it can result an increasing of the mean deviation from perfect reliability of the associated interval. The deviation can even move from



Model\Nom. Coverage	20%	50%	80%
Basic	0.18	0.18	0.94
Advanced	−0.86	−1.43	0.44

Figure 3.13: *Mean deviation from perfect reliability of centered prediction intervals, for basic and advanced models. Results are for the conditionally to ramps evaluated KQR method at wind farm WF1, with parameters value* $(25, 5)$.

positive to negative values and vice versa (figure (3.13)). However the average reliability level of prediction intervals is rather good, and the mean size of intervals remains almost constant (see figure (3.14) and appendix (5.11)).

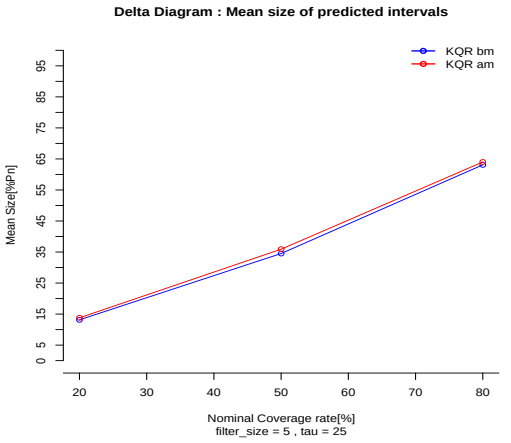


Figure 3.14: *Mean size of centered prediction intervals of nominal coverage rate :* 20%, 50% *and* 80%*, for basic and advanced models. Results are for the conditionally to ramps evaluated KQR method, at wind farm WF1 and for parameters value* $(25, 5)$.

Chapter 4

Complementary results

4.1 Wind farm WF2 : Statistics of ramp events.

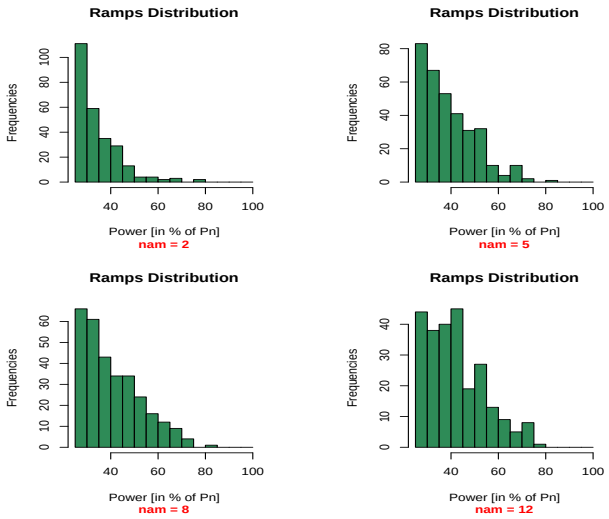


Figure 4.1: *Ramp intensities distribution. There is one graph for each value of* $n_{am} \in \{2, 5, 8, 12\}$ *and* τ *is fixed at* 25% *of the wind farm installed capacity.*

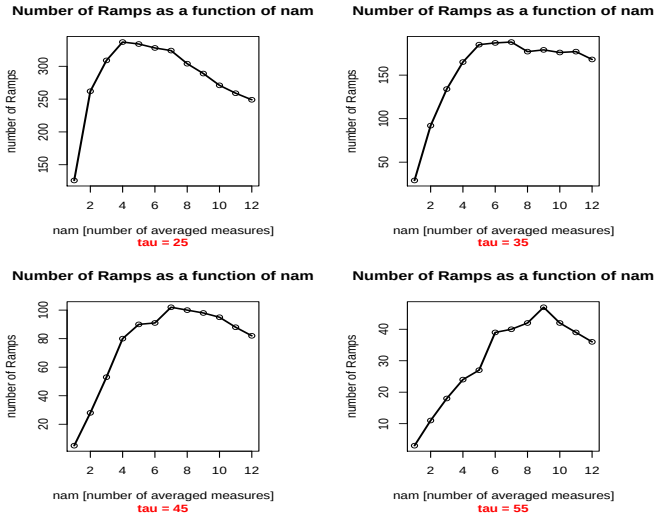


Figure 4.2: *Number of detected ramps as a function of n_{am}, for four values of threshold $\tau \in \{25, 35, 45, 55\}$ (in % of the installed capacity). For each value of τ, there is a maximum number of detected ramps, but this maximum is not reached for the same value of n_{am}.*

Chapter 5

Conclusion

A large part of recent wind power forecasting research works have put the emphasize on providing uncertainty estimation to point forecasts. If appropriately incorporated in decision-making methods, such uncertainty estimation permits to significantly increase the value of wind generation. An issue that one encounters today with these uncertainty estimation is the lack of reliability in the case when the power production encounters large and sharp variations (ramp). In this work we have proposed a quantile forecast model that uses as input informations about upcomming ramps.

The model comes down to a multi-stage forecasting procedure : it uses predicted ramps, detected on preliminary deterministic forecasts, as input information in quantiles estimation providing the associated uncertainty. The model was tested on wind farms of case study, with two nonparametric quantiles estimation methods : the kernel quantile regression method (KQR) and the quantile regression forests procedure (QRF). The influence of parameters used in ramps detection was studied. In a first time, an unconditional evaluation permitted to leave out the QRF method which obtained very low performance results, while the KQR-based models were validated. Then in the particular case of the KQR method, a conditionally to ramp evaluation lead us to conclude that quantiles estimation was of very poor accuracy at such ramp events, particularly for middle range proportions. We also concluded, still in the conditionally-to-ramps-evaluated KQR method, on a general underestimation of such quantiles leading to reliable but not centered prediction intervals. This has been observed for all parameters values. Finally if our model slightly improves the quantiles estimation for middle-high proportions, it doesn't necessarely lead to more reliable prediction intervals.

On the bounce of our study, future research works may use more inputs in our model. Indeed we only used three common explanatory variables : wind speed and direction at 10 meters height and the last past power measure. The eventual decrease of performance level it could result, would lead to a very questionable interest of the use of our model. To do the same complete study with a conditionally to increasing or decreasing ramp events, subject to a large enough size of test sets, would be interesting.

5.1 Wind Farm WF2 : Statistics about selected prediction times for the conditionally to ramp evaluation.

Parameters value	Min	Max	Sum
$\tau = 25, n_{am} = 5$	53	76	1254
$\tau = 25, n_{am} = 3$	27	44	709
$\tau = 40, n_{am} = 5$	9	22	350

Table 5.1: *Number of selected prediction times for the conditionally to ramp evaluation. Values of parameters n_{am} and τ are the ones of case study.*

5.2 Wind Farm WF2 : NRMSE and NMAE of preliminary deterministic forecasts.

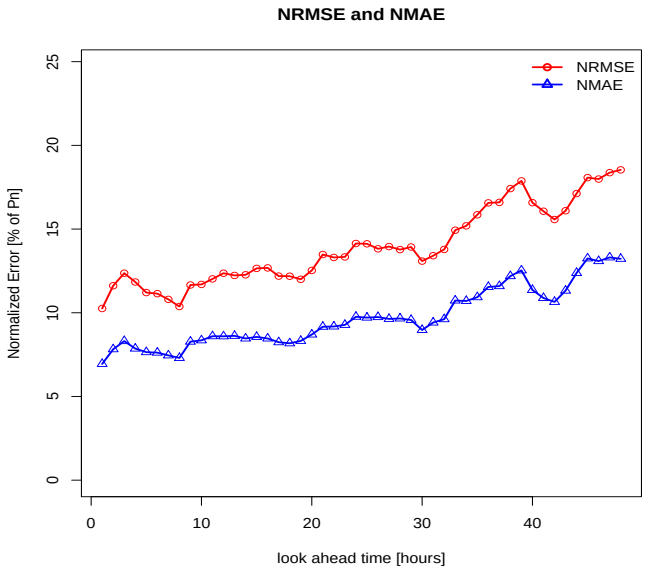
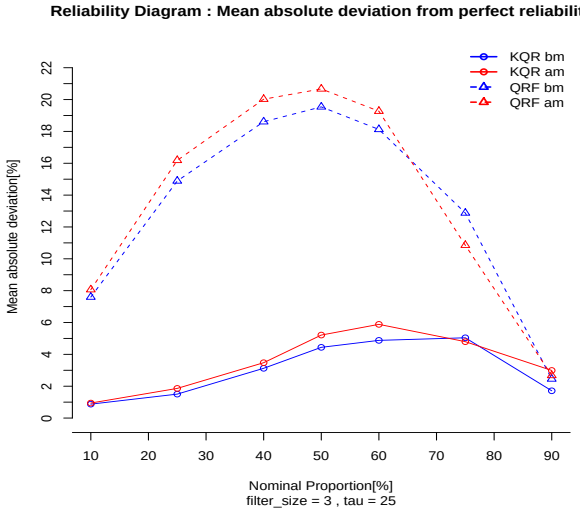
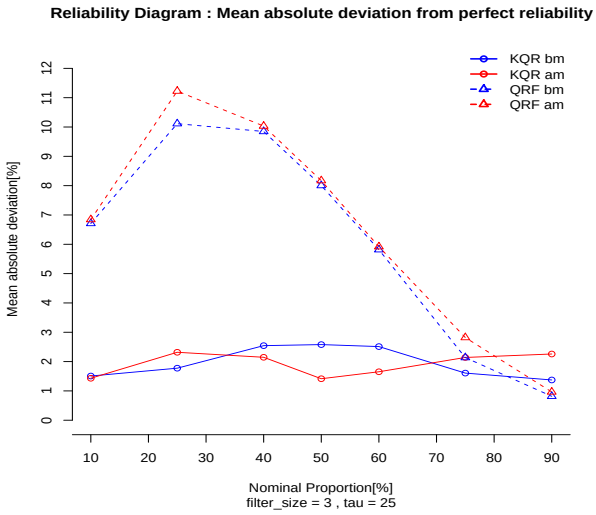


Figure 5.1: *NRMSE and NMAE of preliminary deterministic forecasts obtained from random forest method with per default parameters value.*

5.3 MADFPR for the KQR and the QRF methods used in basic and advanced models, for wind farm WF1 and WF2.

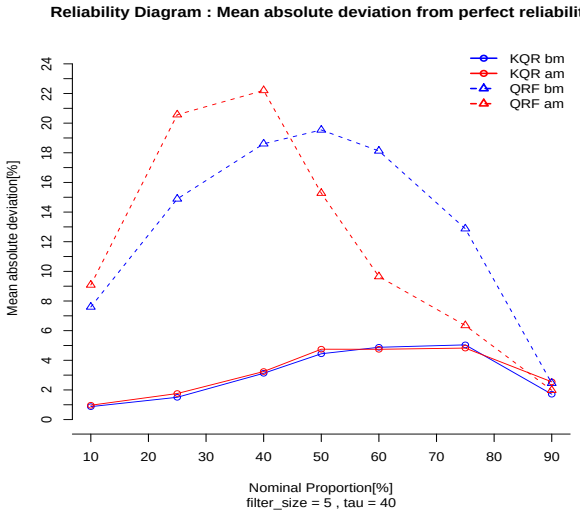


(a) wind farm WF1

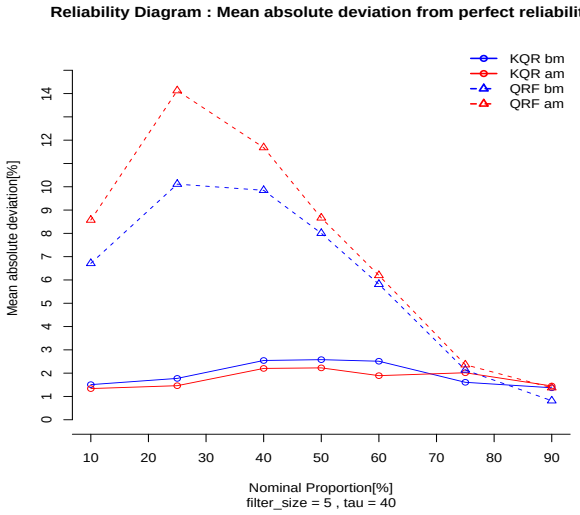


(b) wind farm WF2

Figure 5.2: *Mean absolute deviation from perfect reliability for Kernel Quantile Regression (KQR) and Quantile Regression Forests (QRF) methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 3)$.



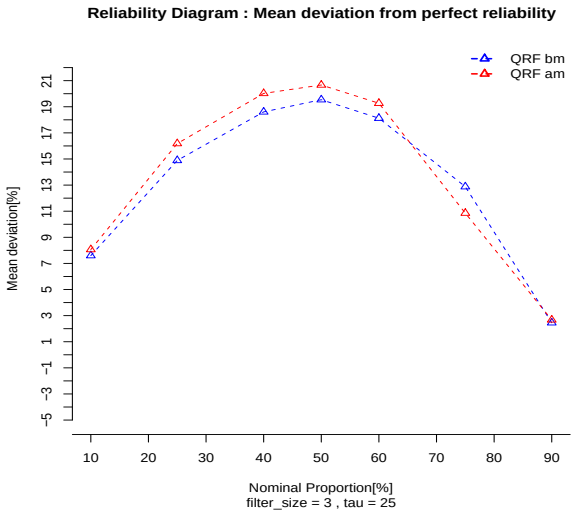
(a) wind farm WF1



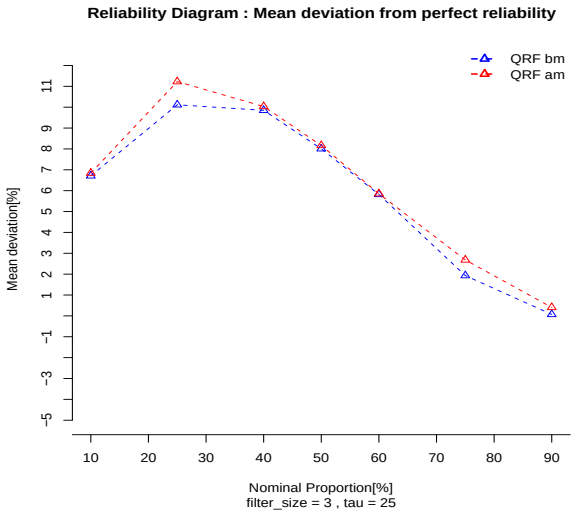
(b) wind farm WF2

Figure 5.3: *Mean absolute deviation from perfect reliability for Kernel Quantile Regression (KQR) and Quantile Regression Forests (QRF) methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (40, 5)$.

5.4 MDFPR for the KQR and the QRF methods used in basic and advanced models, for wind farm WF1 and WF2.

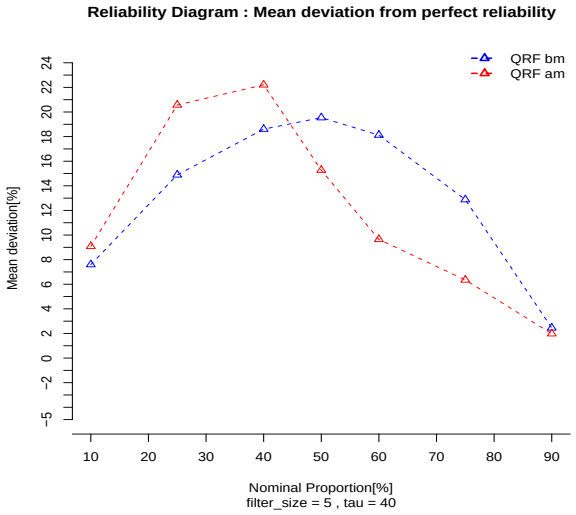


(a) wind farm WF1

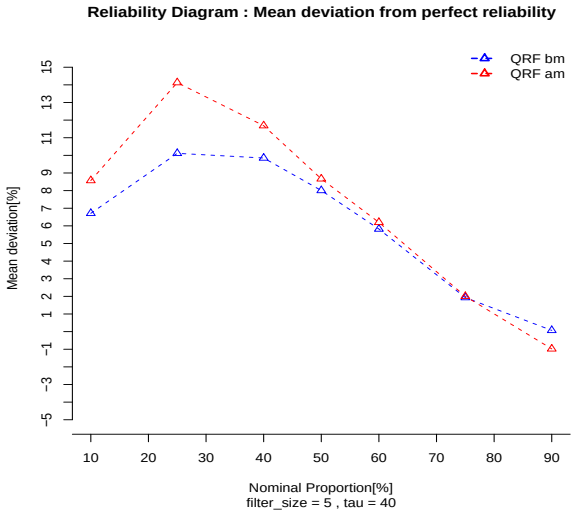


(b) wind farm WF2

Figure 5.4: *Mean deviation from perfect reliability for the Quantile Regression Forests method (QRF). The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 3)$.

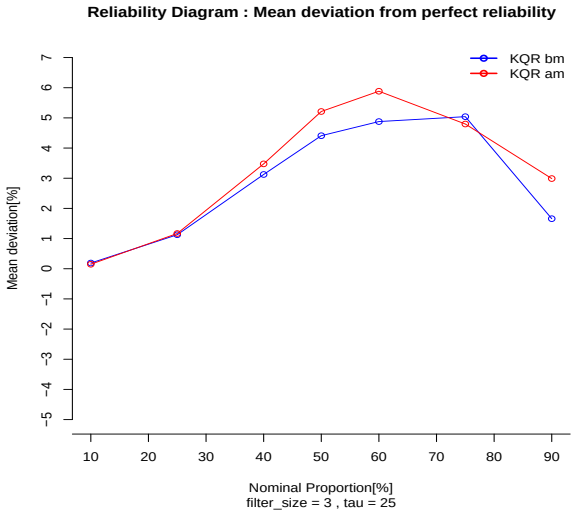


(a) wind farm WF1

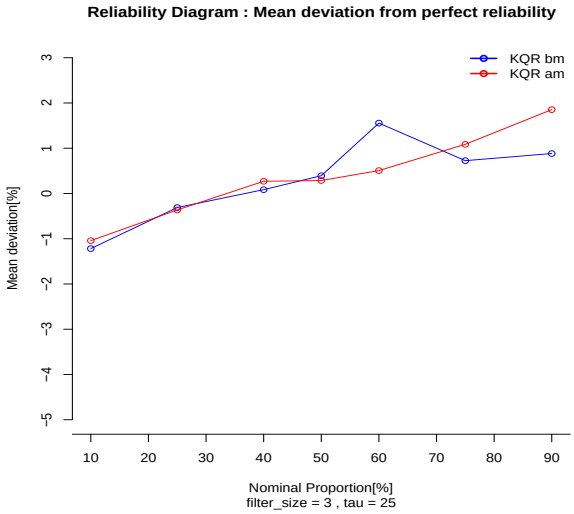


(b) wind farm WF2

Figure 5.5: *Mean deviation from perfect reliability for the Quantile Regression Forests method (QRF). The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (40, 5)$.

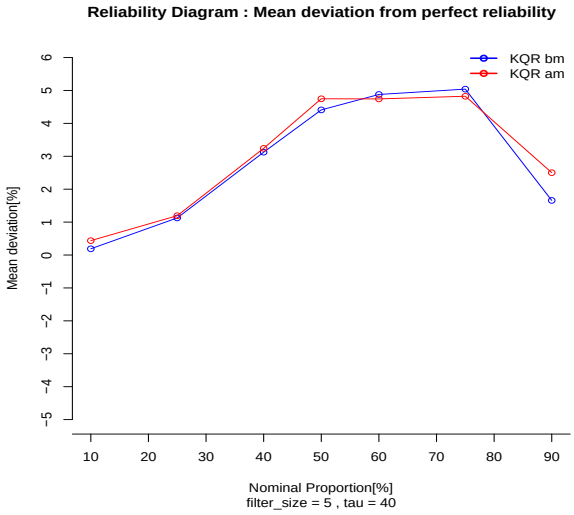


(a) wind farm WF1

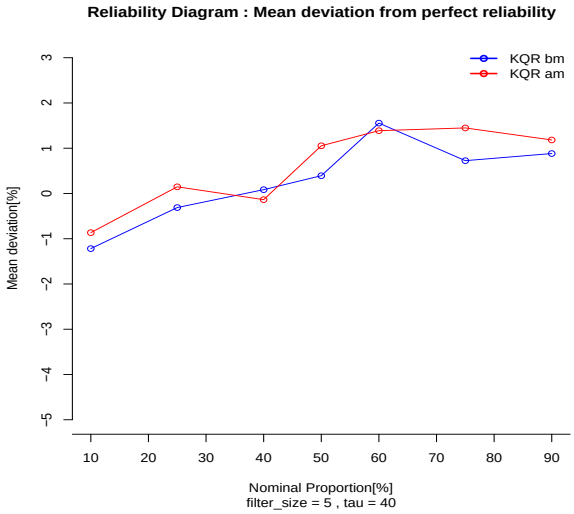


(b) wind farm WF2

Figure 5.6: *Mean deviation from perfect reliability for the Kernel Quantile Regression method (KQR). The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 3)$.



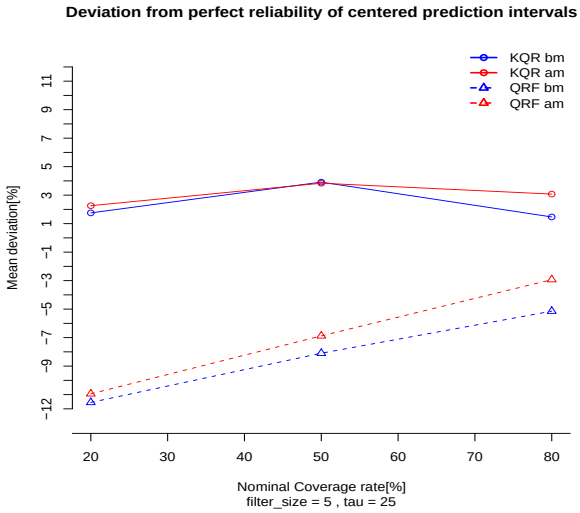
(a) wind farm WF1



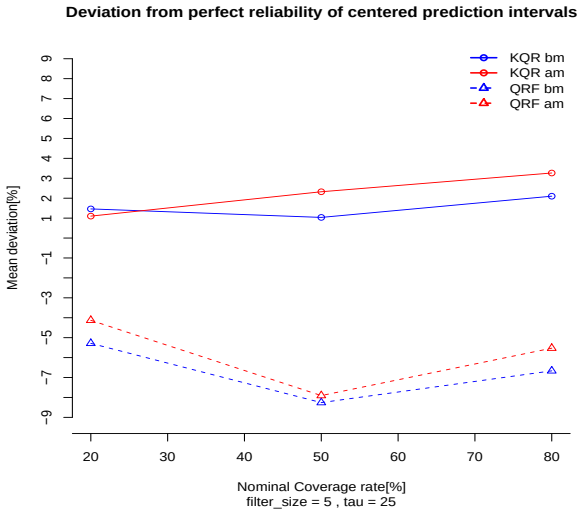
(b) wind farm WF2

Figure 5.7: *Mean deviation from perfect reliability for the Kernel Quantile Regression method (KQR). The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (40, 5)$.

5.5 Mean deviation from perfect reliability of centered prediction intervals, for the KQR and the QRF methods, at wind farm WF1 and WF2.

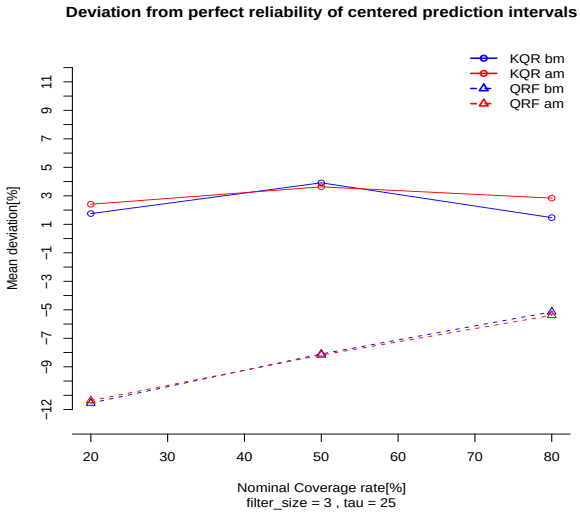


(a) wind farm WF1

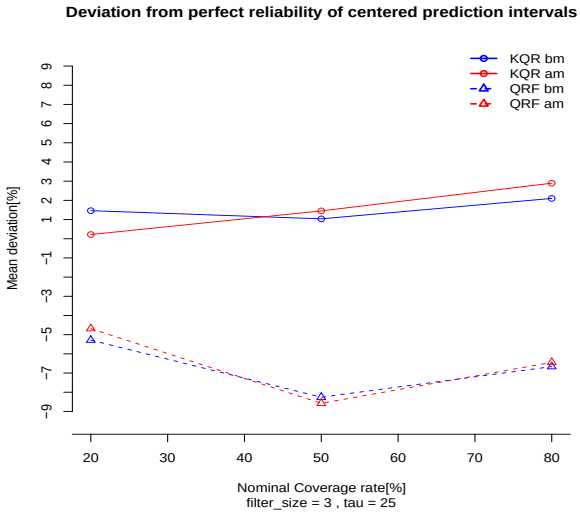


(b) wind farm WF2

Figure 5.8: *Mean deviation from perfect reliability of centered prediction intervals, for the KQR and the QRF methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 5)$.

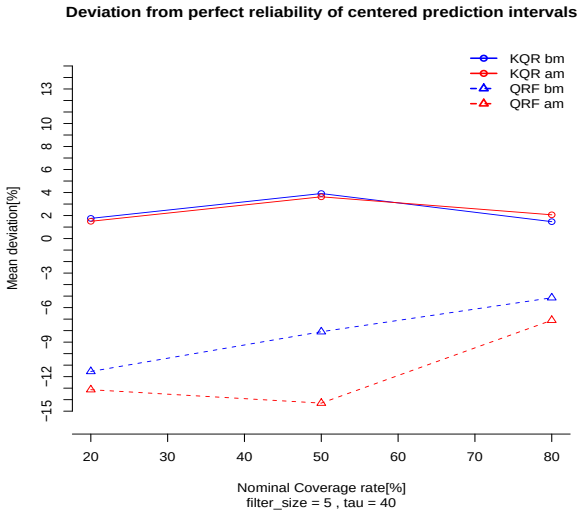


(a) wind farm WF1

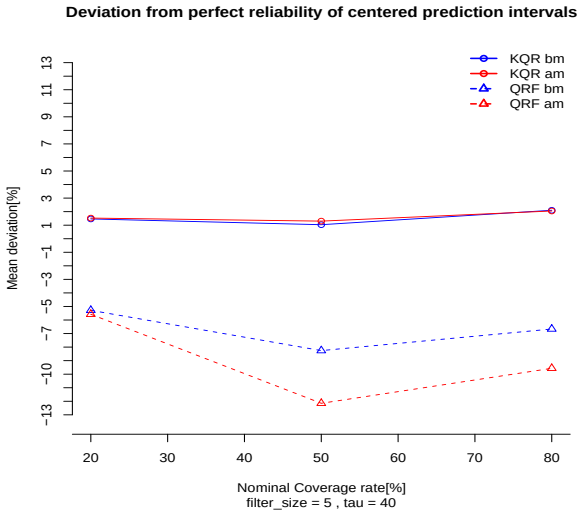


(b) wind farm WF2

Figure 5.9: *Mean deviation from perfect reliability of centered prediction intervals, for the KQR and the QRF methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 3)$.



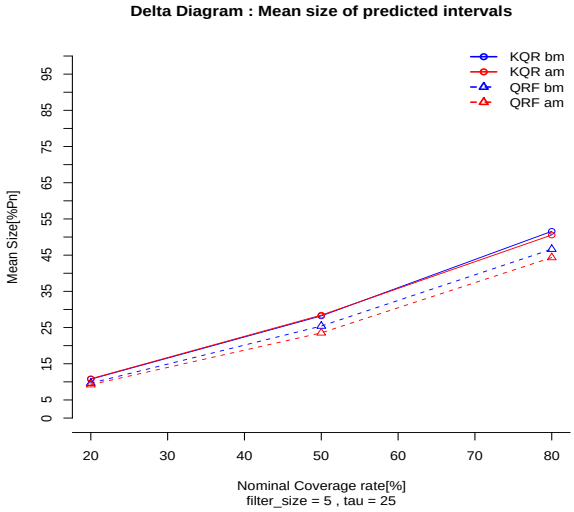
(a) wind farm WF1



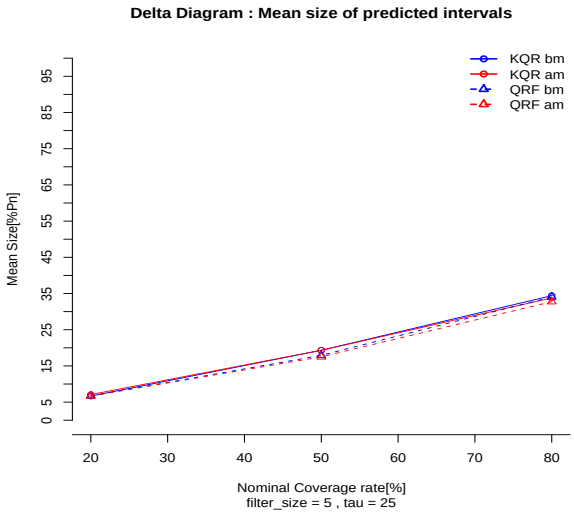
(b) wind farm WF2

Figure 5.10: *Mean deviation from perfect reliability of centered prediction intervals, for the KQR and the QRF methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (40, 5)$.

5.6 Mean size of centered prediction intervals, for the KQR and the QRF methods, at wind farm WF1 and WF2.

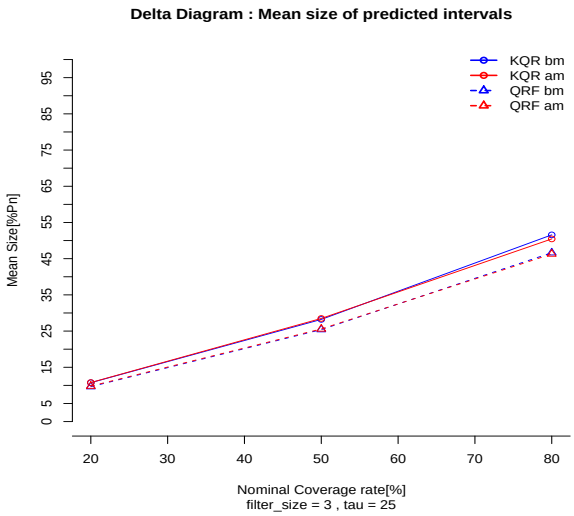


(a) wind farm WF1

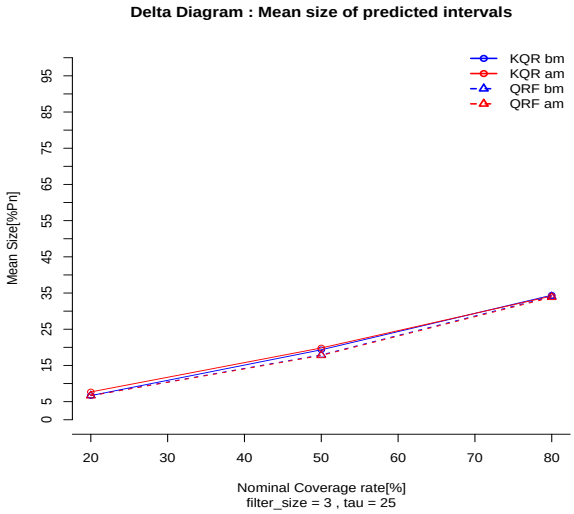


(b) wind farm WF2

Figure 5.11: *Mean size of centered prediction intervals, for the KQR and the QRF methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 5)$.

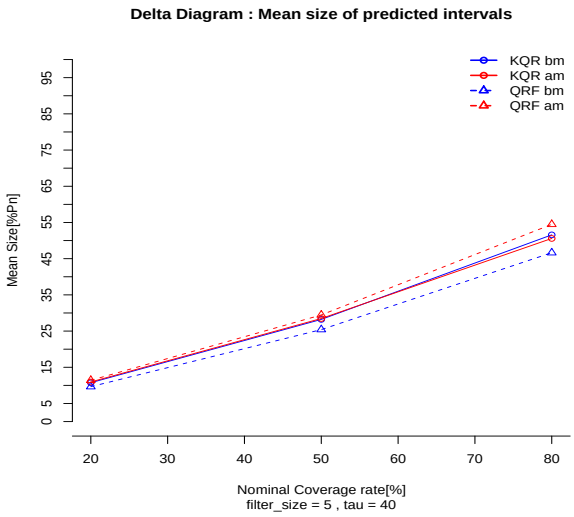


(a) wind farm WF1

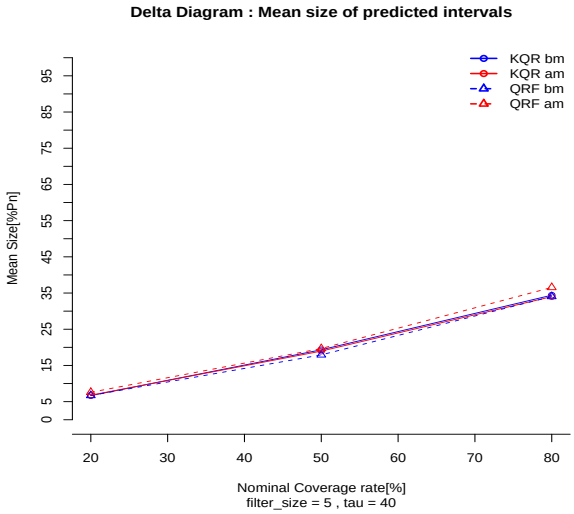


(b) wind farm WF2

Figure 5.12: *Mean size of centered prediction intervals, for the KQR and the QRF methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (25, 3)$.



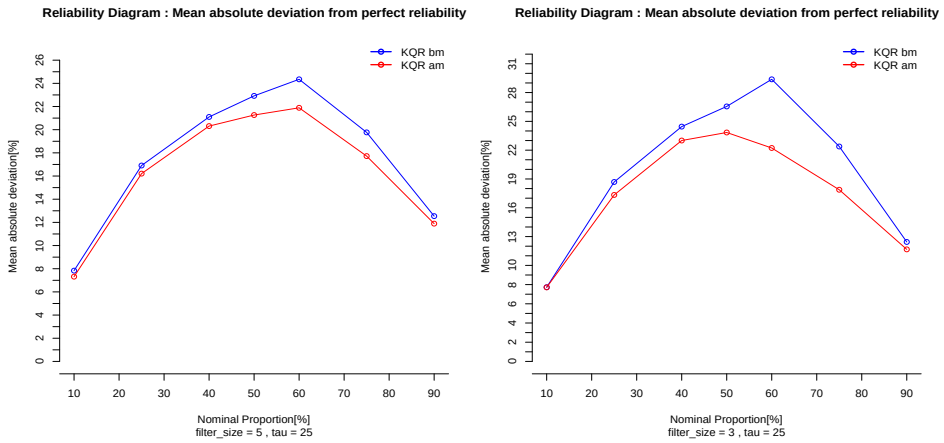
(a) wind farm WF1



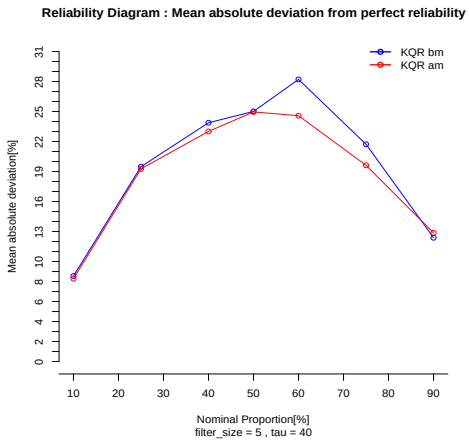
(b) wind farm WF2

Figure 5.13: *Mean size of centered prediction intervals, for the KQR and the QRF methods. The blue curve is for the basic model and the red one for the advanced one. Results are for parameters value* $(\tau, n_{am}) = (40, 5)$.

5.7 Wind Farm WF2 : Reliability diagrams representing the mean absolute deviation from perfect reliability for the KQR method , when conditionally to ramps evaluated.



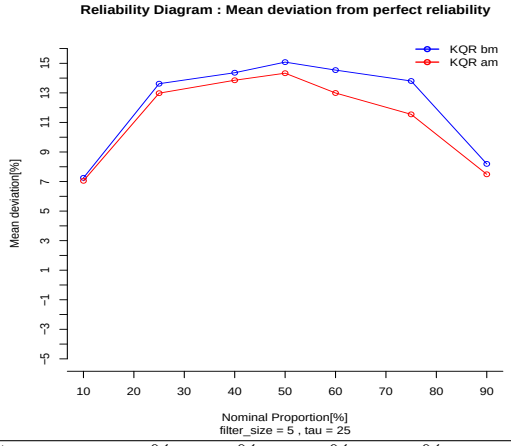
(a) $(\tau, n_{am}) = (25, 5)$ **(b)** $(\tau, n_{am}) = (25, 3)$



(c) $(\tau, n_{am}) = (40, 5)$

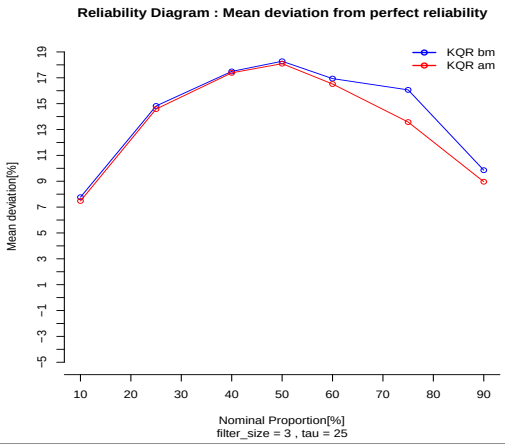
Figure 5.14: *Mean absolute deviation from perfect reliability for the Kernel Quantile Regression method (KQR), when conditionally to ramps evaluated. The blue curve is for the basic model and the red one for the advanced one. Results are for wind farm WF2.*

5.8 Mean deviation from perfect reliability for the KQR method, when conditionally to ramps evaluated.



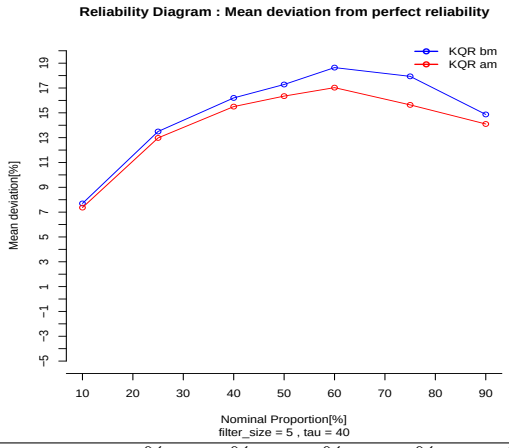
Model\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
Basic	7.25	13.62	14.36	15.08	14.54	13.80	8.20
Advanced	7.05	12.98	13.85	14.33	12.99	11.55	7.49

Figure 5.15: *Mean deviation from perfect reliability, for basic and advanced models and the conditionally to ramps evaluated KQR method at wind farm WF1. The value of parameters is* $(\tau, n_{am}) = (25, 5)$.



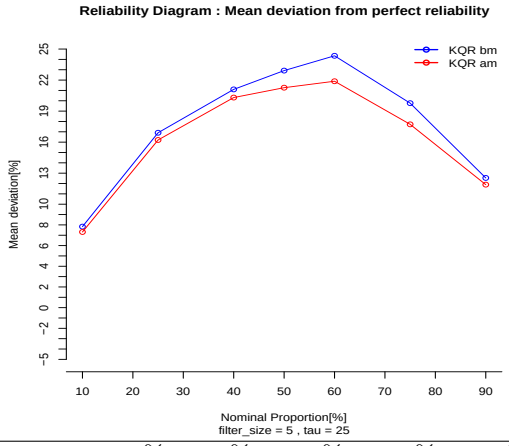
Model\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
Basic	7.76	14.82	17.49	18.28	16.94	16.07	9.86
Advanced	7.47	14.58	17.38	18.09	16.52	13.58	8.96

Figure 5.16: *Mean deviation from perfect reliability, for basic and advanced models and the conditionally to ramps evaluated KQR method at wind farm WF1. The value of parameters is* $(\tau, n_{am}) = (25, 3)$.



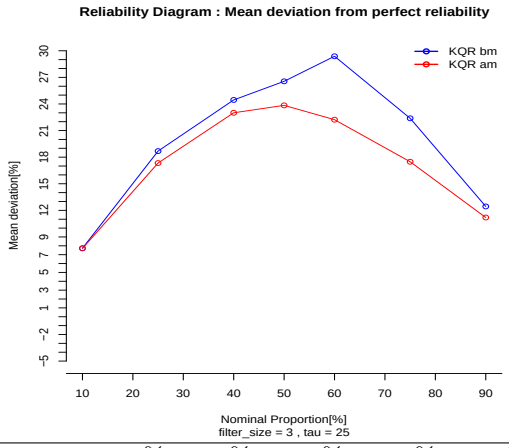
Model\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
Basic	7.70	13.50	16.20	17.28	18.64	17.94	14.87
Advanced	7.37	12.98	15.50	16.35	17.03	15.64	14.10

Figure 5.17: *Mean deviation from perfect reliability, for basic and advanced models and the conditionally to ramps evaluated KQR method at wind farm WF1. The value of parameters is* $(\tau, n_{am}) = (40, 5)$.



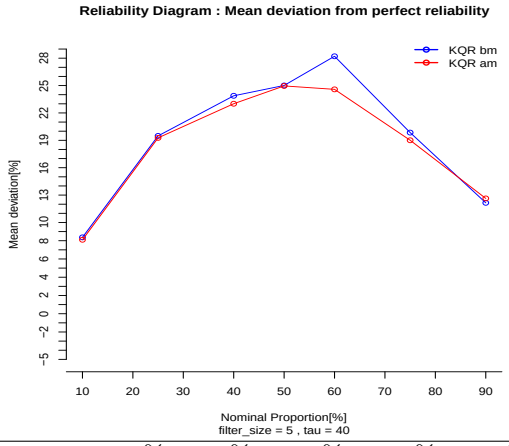
Model\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
Basic	7.84	16.91	21.10	22.92	24.35	19.77	12.54
Advanced	7.32	16.21	20.31	21.26	21.89	17.72	11.89

Figure 5.18: *Mean deviation from perfect reliability, for basic and advanced models and the conditionally to ramps evaluated KQR method at wind farm WF2. The value of parameters is* $(\tau, n_{am}) = (25, 5)$.



Model\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
Basic	7.73	18.69	24.45	26.55	29.38	22.38	12.44
Advanced	7.71	17.34	23.01	23.84	22.23	17.48	11.20

Figure 5.19: *Mean deviation from perfect reliability, for basic and advanced models and the conditionally to ramps evaluated KQR method at wind farm WF2. The value of parameters is* $(\tau, n_{am}) = (25, 3)$.



Model\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
Basic	8.37	19.48	23.88	25.01	28.21	19.85	12.15
Advanced	8.11	19.27	23.01	24.96	24.58	19.01	12.63

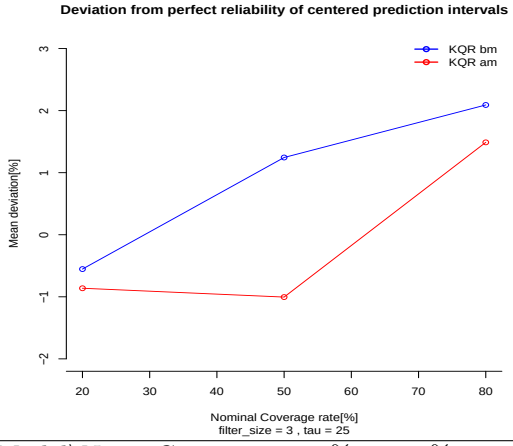
Figure 5.20: *Mean deviation from perfect reliability, for basic and advanced models and the conditionally to ramps evaluated KQR method at wind farm WF2. The value of parameters is* $(\tau, n_{am}) = (40, 5)$.

5.9 Wind farm WF2 : Reliability improvement, diminution of the mean absolute deviation in the conditionally to ramps evaluated KQR method.

Parameters\Nom. Coverage	10%	25%	40%	50%	60%	75%	90%
$\tau = 25, n_{am} = 5$	.066	.041	.037	.072	.101	.104	.052
$\tau = 25, n_{am} = 3$	.002	.072	.059	.102	.243	.201	.062
$\tau = 40, n_{am} = 5$	.031	.011	.036	.002	.128	.097	$-.039$

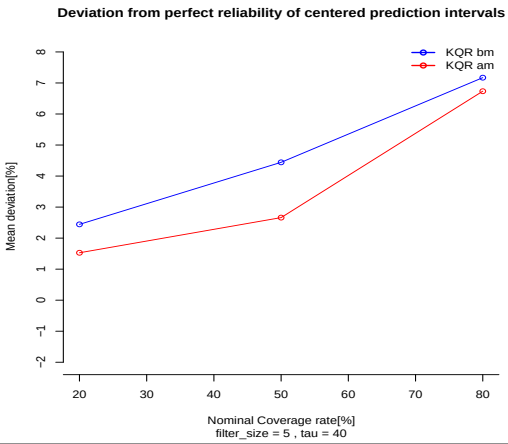
Table 5.2: *Improvement of the mean absolute deviation from perfect reliability : $Imp^{d^\alpha}(\tau, n_{am})$, where d^α is defined in (3.18), τ and n_{am} the parameters. Results are for wind farm WF2.*

5.10 MDFPR of centered prediction intervals estimated by the KQR method, conditionally to ramps evaluated.



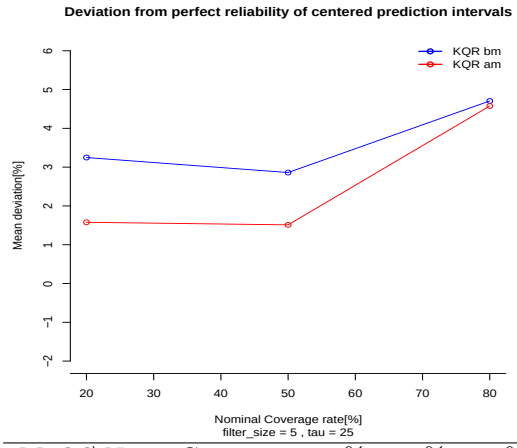
Model\Nom. Coverage	20%	50%	80%
Basic	−0.55	1.24	2.09
Advanced	−0.86	−1.00	1.49

Figure 5.21: *Mean deviation from perfect reliability of centered prediction intervals, for basic and advanced models. Results are for the KQR method, conditionally to ramps evaluated, at wind farm WF1, with parameters value* $(25, 3)$.



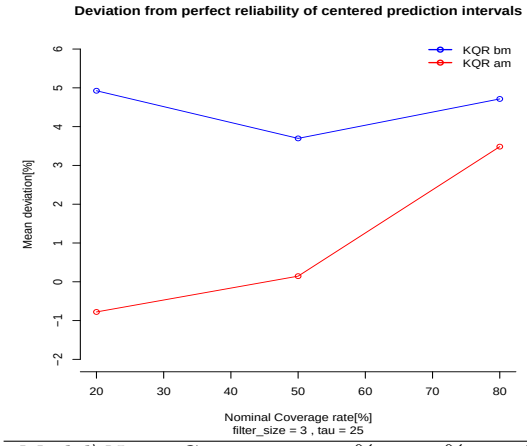
Model\Nom. Coverage	20%	50%	80%
Basic	2.44	4.44	7.17
Advanced	1.53	2.66	6.73

Figure 5.22: *Mean deviation from perfect reliability of centered prediction intervals, for basic and advanced models. Results are for the KQR method, conditionally to ramps evaluated, at wind farm WF1, with parameters value* $(40, 5)$.



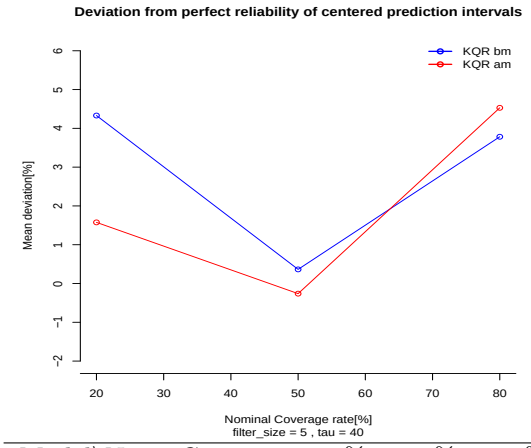
Model\Nom. Coverage	20%	50%	80%
Basic	3.25	2.86	4.71
Advanced	1.58	1.51	4.58

Figure 5.23: *Mean deviation from perfect reliability of centered prediction intervals, for basic and advanced models. Results are for the KQR method, conditionally to ramps evaluated, at wind farm WF2, with parameters value* $(25, 5)$.



Model\Nom. Coverage	20%	50%	80%
Basic	4.92	3.70	4.71
Advanced	−0.78	0.15	3.48

Figure 5.24: *Mean deviation from perfect reliability of centered prediction intervals, for basic and advanced models. Results are for the KQR method, conditionally to ramps evaluated, at wind farm WF2, with parameters value* $(25, 3)$.



Model\Nom. Coverage	20%	50%	80%
Basic	4.33	0.36	3.78
Advanced	1.58	−0.26	4.53

Figure 5.25: *Mean deviation from perfect reliability of centered prediction intervals, for basic and advanced models. Results are for the KQR method, conditionally to ramps evaluated, at wind farm WF2, with parameters value* $(40, 5)$.

5.11 Mean size of centered prediction intervals estimated by the KQR method, conditionally to ramp events evaluated

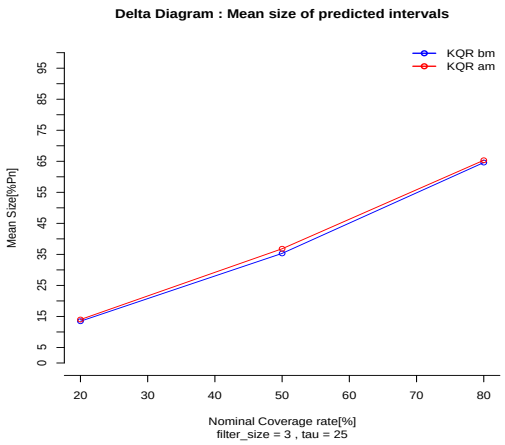


Figure 5.26: *Mean size of centered prediction intervals of nominal coverage rate :*20%, 50% *and* 80%*, for basic and advanced models. results are for the conditionally to ramps evaluated KQR method, at wind farm WF1 and for parameters value* $(25, 3)$.

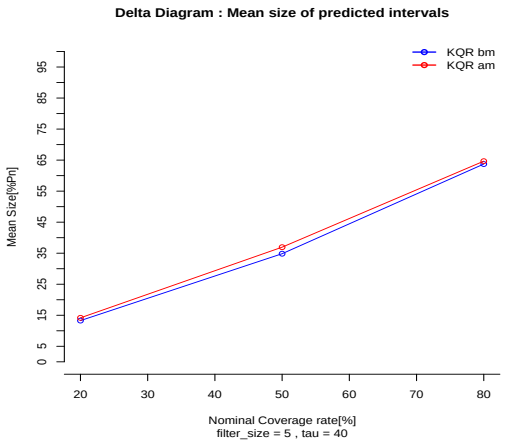


Figure 5.27: *Mean size of centered prediction intervals of nominal coverage rate :*20%, 50% *and* 80%*, for basic and advanced models. results are for the conditionally to ramps evaluated KQR method, at wind farm WF1 and for parameters value* $(40, 5)$.

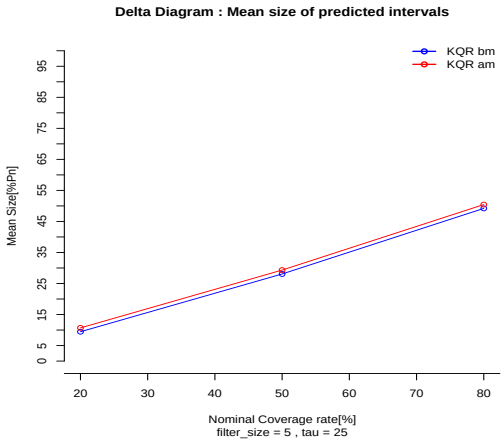


Figure 5.28: *Mean size of centered prediction intervals of nominal coverage rate :*20%, 50% *and* 80%*, for basic and advanced models. results are for the conditionally to ramps evaluated KQR method, at wind farm WF2 and for parameters value* $(25, 5)$.

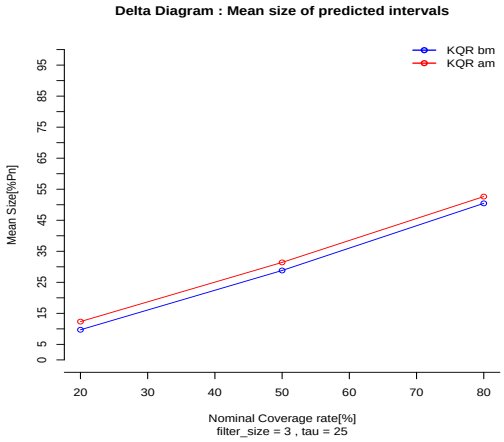


Figure 5.29: *Mean size of centered prediction intervals of nominal coverage rate :*20%, 50% *and* 80%*, for basic and advanced models. results are for the conditionally to ramps evaluated KQR method, at wind farm WF2 and for parameters value* $(25, 3)$.

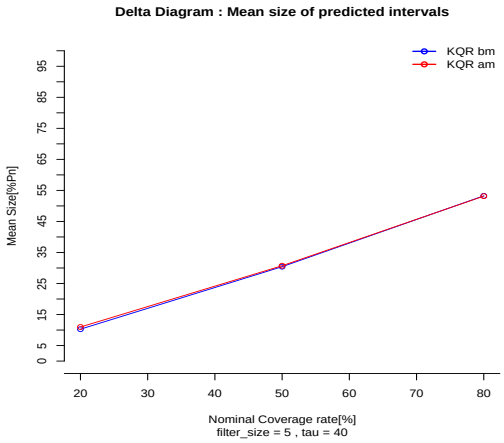


Figure 5.30: *Mean size of centered prediction intervals of nominal coverage rate :*20%, 50% *and* 80%*, for basic and advanced models. results are for the conditionally to ramps evaluated KQR method, at wind farm WF2 and for parameters value* $(40, 5)$.

Bibliography

[1] P. Pinson, "Estimation of the uncertainty in wind power forecasting," Ph.D. dissertation, MINESParisTech, 2006. [Online]. Available: http://pastel.paristech.org/2187/

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, Ed. Springer Series in Statistics, 2001.

[3] H. Madsen, P. Pinson, G. Kariniotakis, H. Nielsen, and T. Nielsen, "Standardizing the performance evaluation of short term wind power prediction models," *Wind Engineering*, vol. 29, no. 6, pp. 475–489, 2005.

[4] G. Giebel, G. Kariniotakis, and R. Brownsword, "State-of-the-art on methods and software tools for short-term prediction of wind energy production," in *EWEC Conference 2003, Madrid, Spain*, 2003.

[5] 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*, vol. 12, no. 6, pp. 1725 – 1744, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6VMY-4NGM15F-2

[6] P. Pinson, C. Chevallier, and G. Kariniotakis, "Optimizing benefits from wind power participation in electricity markets using advanced tools for wind power forecasting and uncertainty assessment," 2004.

[7] P. Pinson, J. Juban, and G. N. Kariniotakis, "On the quality and value of probabilistic forecasts of wind generation," in *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, 2006, pp. 1–7.

[8] H. Nielsen, H. Madsen, and T. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 95–108, 2006.

[9] J. Bremnes, "A comparison of a few statistical models for making quantile wind power forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 3–11, 2006.

[10] J. Juban, N. Siebert, and G. Kariniotakis, "Probabilistic short-term wind power forecasting for the optimal management of wind generation," in *Proc. of the IEEE PowerTech 2007 Conference, IEEE Conference, Lausanne, Switzerland*, 2007.

[11] P. Pinson, J. Moller, H. Nielsen, H. Madsen, and G. Kariniotakis, "Evaluation of nonparametric probabilistic forecasts of wind power," Technical University of Denmarck, Tech. Rep. IMM 2007-02, 2007.

[12] R. Girard, J. Juban, P. Pinson, and G. Kariniotakis, "Attempt of a protocol for evaluation of probabilistic forecast," 2009.

[13] D. B. Jolliffe, Ian T.; Stephenson, *Forecast Verification - A Practioner's Guide in Atmospherique Science*, D. B. Jolliffe, Ian T.; Stephenson, Ed. John Wiley & Sons, 2003. [Online]. Available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471497592.html

[14] G. Brier, "Verification of forecast expressed in terms of probability," *Monthly weather review*, vol. 78, pp. 1–3, 1950.

[15] M. Lange, "Analysis of the uncertainty of wind power predictions," Ph.D. dissertation, University Carl von Ossietzky, Oldenburg, Germany, 2003. [Online]. Available: http://www.energymeteo.de/de/downloads/

[16] H. Bludszuweit, J. Dominguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *Power Systems, IEEE Transactions on*, vol. 23, no. 3, pp. 983–991, 2008.

[17] M. Lange, "On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, vol. 127, no. 2, pp. 177–184, 2005. [Online]. Available: http://link.aip.org/link/?SLE/127/177/1

[18] K. A. Brewster, "Phase-correcting data assimilation and application to storm-scale numerical weather prediction. part i: Method description and simulation testing," *Monthly Weather Review*, vol. 131, no. 3, pp. 480–492, Mar. 2003.

[19] ——, "Phase-correcting data assimilation and application to storm-scale numerical weather prediction. part ii: Application to a severe storm outbreak," *Monthly Weather Review*, vol. 131, no. 3, pp. 493–507, 2003.

[20] R. N. Hoffman, Z. Liu, J.-F. Louis, and C. Grassoti, "Distortion representation of forecast errors," *Monthly Weather Review*, vol. 123, no. 9, pp. 2758–2770, 1995.

[21] V. Monbet, P. Ailliot, and A. Cuzol, "A state space model for wind forecast correction," *International Workshop on Applied Probability*, vol. -, pp. –, 2008.

[22] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2009. [Online]. Available: http://dx.doi.org/10.1002/we.284

[23] D. G. Galati and M. A. Simaan, "Automatic decomposition of time series into step, ramp, and impulse primitives," *Pattern Recognition*, vol. 39, no. 11, pp. 2166 – 2174, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6V14-4HNSJF9-2

[24] J. Parks, B.Greaves, J. Collins, and A. Tindal, "Temporal forecast uncertainty for ramp events," european Wind Energy Conference & exhibition 2009 - EWEC 2009.

[25] A. Fortin, *Analyse numerique pour ingenieurs*, M. Ouimet, Ed. Ecole Polytechnique de Monreal, 1995.

[26] L. B. Statistics and L. Breiman, "Random forests," *Machine Learning*, vol. ?, pp. 5–32, 2001.

[27] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, pp. 983–999, 2006.

[28] Y. Li, Y. Liu, and J. Zhu, "Quantile regression in reproducing kernel hilbert spaces," *Journal of the American Statistical Association*, vol. 102, pp. 255–268, 2007.

[29] G. Kariniotakis, I. Marti, D. Casas, P. Pinson, T. Nielsen, H. Madsen, G. Giebel, J. Usaola, I. Sanchez, A. Palomares, R. Brownsword, J. Tambke, U. Focken, M. Lange, P. Louka, G. Kallos, C. Lac, G. Sideratos, and G. Descombes, "What performance can be expected by short-term wind power prediction models depending on site characteristics ?" in *2004 EWEC Conference, London, UK*, 2004. [Online]. Available: http://epubs.cclrc.ac.uk/work-details?w=41834