On-line quantile regression in the RKHS (Reproducing Kernel Hilbert Space) for operational probabilistic forecasting of wind power

Cristobal Gallego-Castillo[a,*], Ricardo Bessa[b], Laura Cavalcante[b], Oscar Lopez-Garcia[a]

[a] *DAVE, Universidad Politécnica de Madrid, Pza. Cardenal Cisneros, 3, 28040, Madrid, Spain*
[b] *INESC Technology and Science (INESC TEC), Campus da FEUP, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal*

Abstract

Wind power probabilistic forecast is being used as input in several decision-making problems, such as stochastic unit commitment, operating reserve setting and electricity market bidding. This work introduces a new on-line quantile regression model based on the Reproducing Kernel Hilbert Space (RKHS) framework. Its application to the field of wind power forecasting involves a discussion on the choice of the bias term of the quantile models, and the consideration of the operational framework in order to mimic real conditions. Benchmark against linear and splines quantile regression models was performed for a real case study during a 18 months period. Model parameter selection was based on k-fold crossvalidation. Results showed a noticeable improvement in terms of calibration, a key criterion for the wind power industry. Modest improvements in terms of Continuous Ranked

*Corresponding author
Email address: `cristobaljose.gallego@upm.es` (Cristobal Gallego-Castillo)

Probability Score (CRPS) were also observed for prediction horizons between 6-20 hours ahead.
Keywords: Wind power, quantile regression, Reproducing Kernel Hilbert Space (RKHS), probabilistic forecast, short-term, on-line,

1. Introduction

The high integration levels of wind power in several countries demands for a paradigm shift in terms of power system management tools and operational practices, which consists in moving from deterministic to probabilistic decision-making tools [1]. In this context, probabilistic wind power forecasts with high skill is a key requirement for end-users. For Transmission System Operators (TSO), this information is vital for setting the operating reserve requirements [2, 3], stochastic unit commitment [4] and technical constraints evaluation [5]. Distribution System Operators (DSO) with high integration levels of wind power in their networks can also benefit from accurate forecasts, which can be integrated in multi-period optimal power flow problems [6]. For electricity market agents, this information can be embed in bidding optimization problems for electrical energy [7, 8] and ancillary services markets [9].

The current wind power forecasting state of the art is rich in point and probabilistic forecast methods. A detailed review can be found in [10] and [11]. Four main classes of probabilistic forecasting algorithms can be found in the literature: conditional kernel density estimation (KDE), (b) semi-parametric regression, (c) machine learning and (d) quantile regression. It is important to stress that other representations for the wind power uncertainty are also possible, such as ramp forecasting [12] and temporal trajectories (or short-term scenarios) [13, 14].

Two examples of conditional KDE algorithms are: (a) time-adaptive quantile-copula estimator that produces density forecasts for the next hours using Numerical Weather Predictions (NWPs) as inputs and explores the non-parametric copula for modelling the dependency between wind speed/direction and power (i.e., the power curve) [15]; (b) two-stage approach that, firstly, uses a vector autoregressive moving average-generalized autoregressive conditional heteroscedastic (VARMA-GARCH) model to capture wind speed and direction uncertainty forecast, secondly, employs conditional KDE to model the relationship between wind speed/direction and power [16].

One work about semi-parametric regression is presented in [17], which proposes the use of generalized logit-Normal distribution to enable a full characterization of the forecast densities by their location and scale parameters. Dynamic models based on classical time series models (e.g., autoregressive model) are proposed for the location and scale parameters.

In terms of machine learning algorithms, an online sparse Bayesian model based on warped Gaussian process is proposed in [18], and employed to generate probabilistic wind power forecasts. Furthermore, in [19] multiple radial basis function neural networks (RBFNN), combined with self-organized maps that classify the uncertainty knowledge in multiple levels, are proposed to forecast eight quantiles of wind power distribution based on point forecasts.

The majority of the methods based on quantile regression employed to model the non-linear relationship between wind speed and power use two well-known techniques, local regression (or varying coefficients) [20] and additive models with splines [21]. Local regression methods were successfully applied to model time-varying conditions, for instance the relation between wind speed and direction in very short-term forecasting [22]. The main limitation of local quantile regression is that the computational time increases significantly with the number of predictors and it is also prone to overfitting. The additive models require a correct choice of the splines for different types of variables (e.g., categorical, circular) and a hyperparameter is needed to each predictor variable.

Related to this last category, this paper proposes a new quantile regression model based on kernel methods. Kernel methods are a class of algorithms oriented to pattern analysis that have been applied to a number of problems, involving classification, regression and time series forecasting (see [23] and references therein). The presented model implements quantile regression in the Reproducing Kernel Hilbert Space (RKHS) according to the framework described in [24]. In this framework, the data from the input space is transformed to the feature space using a kernel matrix. In other words, this means transforming a non-linear space into a high dimensional linear space where the classical linear quantile regression technique can be applied. The algorithm is implemented from an on-line learning perspective. While the main advantage of this approach is to account for smooth variations in the underlying dynamics of the modelled process, other advantages as compared with the off-line approach were analysed in [25].

This paper presents a number of original contributions: it represents the first application of quantile regression in the RKHS to the wind power probabilistic forecasting problem, establishing a connection between quantile regression techniques and recent research in signal processing theory. Second, the model equations were developed for the case of including a bias term; this has an impact on the model performance since a proper choice for the initial bias allowed the model to perform at least as *climatology*, a reference model in the field. Third, the benchmark experiment relied on a detailed description of the operational framework of wind power forecasting, which was implemented to mimic real conditions characterised by meterological forecast availability each 12 hours. Finally, the observed noticeable improvement in terms of calibration (one of the criteria considered in the evaluation framework) was related to the adaptive nature of the algorithm.

The remaining of the paper is organized as follows: Section 2 provides a general description of the quantile regression models in the RKHS, an its particularization to the on-line standpoint. An overview of the operational framework in wind power forecasting is given in Section 3, outlining the interactions between the NWP delivery and models generating wind power predictions. Section 4 describes the setup of the experiment, consisting on the employed data, benchmark models and evaluation framework. The obtained results are presented and discussed in Section 5. Finally, the paper ends with concluding remarks in Section 6.

2. Quantile Regression in the RKHS

The objective of quantile regression is to model a functional relationship between a set of explanatory variables, here denoted by vector $\mathbf{x}$ in $\mathcal{X} \in \mathbb{R}^n$, and the τ-th quantile of the conditional probability density function of the objective variable y, which is assumed to be one-dimensional in the following.

In a general manner, a quantile regression model can be written as follows:

$$q^\tau(\mathbf{x}) = f(\mathbf{x}) + b, \tag{1}$$

where q^τ is the τ-th quantile, $\tau \in [0,1]$, b is a bias term and $f : \mathcal{X} \longrightarrow \mathbb{R}$, with $\mathcal{X} \in \mathbb{R}^n$, is a function to be determined. The most straightforward strategy to define f is that of linear quantile regression [26]. While linearity usually entails a number of advantages (i.e. simplicity and robustness), such hypothesis may result too restrictive when dealing with problems with complex underlying dynamics.

Regression in the RKHS allows exploiting non-linear relationships between data keeping the simplicity of the linear approach. To do so, linearity is assumed in a high-dimensional feature space given by the feature map $\varphi : \mathcal{X} \to \mathcal{H}$, where $\mathcal{H}$ is a RKHS defined by the reproducing kernel (also referred to as kernel matrix) $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$. By doing this, it holds that:

$$q^\tau(\mathbf{x}) = \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle + b, \tag{2}$$

where $\mathbf{w}$ is a vector in $\mathbb{R}^n$ containing the coefficients of the linear regression.

From the off-line standpoint, the model parameters, $\mathbf{w}$ and b, can be obtained by minimising the following regularised cost function evaluated over N samples $(\mathbf{x}_i, y_i)$ (see [24], among others):

$$R_{1:N} := \frac{1}{N} \sum_{i=1}^{N} l_\tau(y_i, q^\tau(\mathbf{x}_i)) + \frac{\lambda}{2} \|f\|^2_{\mathcal{H}}, \tag{3}$$

where $\|\cdot\|_{\mathcal{H}}^2$ is the norm in the RKHS, which measures the complexity of the function f, λ is a regularization parameter providing control on the bias/variance balance in the model estimation, and $l_\tau : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ is a loss function of the forecast error. According to quantile regression theory [24], $l_\tau(y_i, q^\tau(\mathbf{x}_i))$ is the pinball function, given by:

$$l^\tau(y_i, q^\tau(\mathbf{x}_i)) = \begin{cases} \tau \cdot (y_i - q^\tau(\mathbf{x}_i)) & \text{if } y_i \geq q^\tau(\mathbf{x}_i) \\ (\tau - 1) \cdot (y_i - q^\tau(\mathbf{x}_i)) & \text{if } y_i < q^\tau(\mathbf{x}_i) \end{cases}. \tag{4}$$

From the Support Vector Machine literature (see [27], among others), it can be demonstrated that the model that minimises Eq. (3) can be written in the form of a kernel expansion based on the available samples, as follows:

$$q^\tau(\mathbf{x}) = \sum_{t=1}^{N} \alpha_i k(\mathbf{x}_t, \mathbf{x}) + b, \tag{5}$$

where the expansion coefficients α_i and the bias term can be obtained by solving the dual problem associated to the minimisation of Eq. (3).

2.1. On-line learning in the RKHS

On-line learning is an incremental process in which a model integrates information as new observations are available. One of the main advantages of this strategy in the case of wind power forecasting is that the model is able to account for smooth variations of the underlying dynamics of the wind power output over time, which are likely to happen within a monthly scale because of meteorological seasonalities, and in the long-term due to the wind turbine aging. Since the quantile model evolves over time, it is required to index the model state to time. In this work, we define $q_t^\tau(\mathbf{x})$ as the resulting quantile regression model after integrating the available samples from $(\mathbf{x}_1, y_1)$ to $(\mathbf{x}_t, y_t)$. Mathematically:

$$q_t^\tau(\mathbf{x}) = \sum_{i=1}^{t} \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b_t. \tag{6}$$

It is noted that Eq. (6) is general in the sense that the bias term is also subject to change with the learning process.

Under the on-line standpoint, the expansion coefficients are assessed in base on the learning strategy. A number of challenges and kernel-based algorithms related to on-line learning are described in [28]. The stochastic gradient descent in Hilbert Space, also described in that work, is here adopted and generalised for the case $b \neq 0$. Stochastic gradient descent means that the model evolves in order to minimise only the most recent error, making the cost function given in (3) to collapse into $R_{t:t}$. Since the minimisation occurs in the RKHS, the gradient is computed with respect to the function q^τ. Mathematically:

$$q_{t+1}^\tau = q_t^\tau - \eta \left. \frac{\partial R_{t:t}}{\partial q^\tau} \right|_{q^\tau = q_t^\tau}, \tag{7}$$

where η is the learning rate, which is assumed to be constant.

Operating the derivative in Eq. (7), two terms can be identified:

$$\frac{\partial\, R_{t:t}}{\partial\, q^\tau} = \underbrace{\frac{\partial\, l^\tau(y_t, q^\tau(\mathbf{x}_t))}{\partial\, q^\tau}}_{(a)} + \underbrace{\frac{\lambda}{2} \frac{\partial\, \|f\|_{\mathcal{H}}^2}{\partial\, q^\tau}}_{(b)}.$$

According to the reproducing property of RKHSs, given by:

$$\langle f, k(\mathbf{x}, \cdot)\rangle_{\mathcal{H}} = f(\mathbf{x}),$$

the first term can be rewritten as:

$$(a) = \frac{\partial\, l^{\tau}(y_t, q^{\tau}(\mathbf{x}_t))}{\partial\, q^{\tau}(\mathbf{x}_t)} \cdot \frac{\partial\, q^{\tau}(\mathbf{x}_t)}{\partial\, q^{\tau}} = \frac{\partial\, l^{\tau}(y_t, q^{\tau}(\mathbf{x}_t))}{\partial\, q^{\tau}(\mathbf{x}_t)} \cdot k(\mathbf{x}_t, \cdot).$$

On the other hand, the second term can be rewritten as follows:

$$(b) = \frac{\lambda}{2} \frac{\partial\, \|q^{\tau} - b\|^2_{\mathcal{H}}}{\partial\, q^{\tau}} = \lambda q^{\tau} - \lambda b.$$

Thus,

$$q^{\tau}_{t+1} = q^{\tau}_t - \lambda\eta q^{\tau}_t + \lambda\eta b_t - \eta \frac{\partial\, l^{\tau}(y_t, q^{\tau}(\mathbf{x}_t))}{\partial\, q^{\tau}(\mathbf{x}_t)} \cdot k(\mathbf{x}_t, \cdot).$$

Making use of Eqs. (4) and (6), the following rules for updating the model from time t to time $t+1$ (i.e., once sample $(\mathbf{x}_{t+1}, y_{t+1})$ is available) are obtained:

$$\alpha_{t+1} := \begin{cases} \eta\tau & \text{if } y_{t+1} > q^{\tau}_t(\mathbf{x}_{t+1}) \\ \eta(\tau - 1) & \text{if } y_{t+1} < q^{\tau}_t(\mathbf{x}_{t+1}) \\ 0 & \text{if } y_{t+1} = q^{\tau}_t(\mathbf{x}_{t+1}) \end{cases}, \tag{8}$$

$$\alpha_i := (1 - \eta\lambda)\alpha_i \quad \text{for } i \leq t, \tag{9}$$

$$b_{t+1} := b_t. \tag{10}$$

Concerning the reproducing kernel, $k(\cdot, \cdot)$, it must meet certain conditions to be considered an admissible kernel. A number of admissible kernels are provided in [27]. From here on, we assume $k(\cdot, \cdot)$ to be the Radial Basis Function (RBF) Kernel, given by:

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\sigma \|\mathbf{x}_1 - \mathbf{x}_2\|^2\right), \tag{11}$$

where σ is a parameter related to the kernel width. According to [29], the RBF kernel is a general purpose kernel well-suited for situations in which no prior knowledge about the data is available.

Equation (8) implies that the assimilation of sample $(\mathbf{x}_{t+1}, y_{t+1})$ consists of a correction of the quantile model in the neighborhood of $\mathbf{x}_{t+1}$ of a magnitude related to η, where the notion of neighborhood in the input space $\mathcal{X} \in \mathbb{R}^n$ is linked to the aforementioned parameter σ. Out of this neighborhood, Eq. (9) states that the quantile model degrades towards the bias term with a rate related to a forgetting factor given by $\eta\lambda$. That the learning occurs locally while the forgetting occurs globally is also a consequence of the fact that the bias term is actually not updated with the new sample (see Eq. (10)), reflecting that this term must be selected with care, specially if the stream of observations $\mathbf{x}_t$ might not cover the span of the input space $\mathcal{X}$ regularly.

3. Operational framework

Wind power forecasting models are usually classified into two basic approaches. A first approach, referred to as physical approach, consists on obtaining the best wind speed forecast at the wind farm, and use it to generate the associated wind power forecast. This approach requires access to

NWP outputs delivered by one or more meteorological models. In a general manner, it can be written as follows:[1]

$$\hat{p}_{t+k|t} = f(\hat{\mathbf{m}}_{t+k|t+k-k^*}), \tag{12}$$

where $\hat{p}_{t+k|t}$ is the power forecast for time $t+k$ generated at time t, $\hat{\mathbf{m}}_{t+k|t+k-k^*}$ is a vector containing the forecast of a number of meteorological variables (typically, wind speed and direction) for time $t+k$ generated at time $t+k-k^*$, k is the prediction horizon of the power forecasting model and k^* is the prediction horizon of the meteorological forecasts. It is noted that, because the meteorological forecast must be available at time t, it holds $k^* \geq k$.

The second approach aims at seizing the inertia of the wind power output through time series analysis, so that power forecasts can be generated using recent wind power records provided online by the SCADA[2] system. It is generally accepted that time series based models outperform physical models for the very short-term (this being characterized by prediction horizons up to 3-5 hours), while meteorological information is key for generating accurate forecasts in the short-term (prediction horizons up to one-two days ahead) [30].

Because most of the applications of wind power forecasting involve prediction horizons larger than a few hours, the use of accurate NWPs is crucial in operational wind power forecasting. An important issue is that the NWP delivery scheme places conditions to operational forecast. Meteorolog-

[1]For simplicity, we present the formulation for a point-forecasting model, but this is also valid for a quantile model.

[2]Supervisory Control And Data Acquisition.

ical agencies deliver NWP outputs according to specificities of the employed meteorological model. Relevant parameters are the model cycle, (ΔT, time between two different launches of the model), the output time-step, (Δt, time between two consecutive forecasts) and the output length, (k^*_{max}, maximum prediction horizon). Figure 1 displays a generic delivery scheme for $\Delta T = 6$ h, $\Delta t = 3$ h and $k^*_{max} = 24$ h.

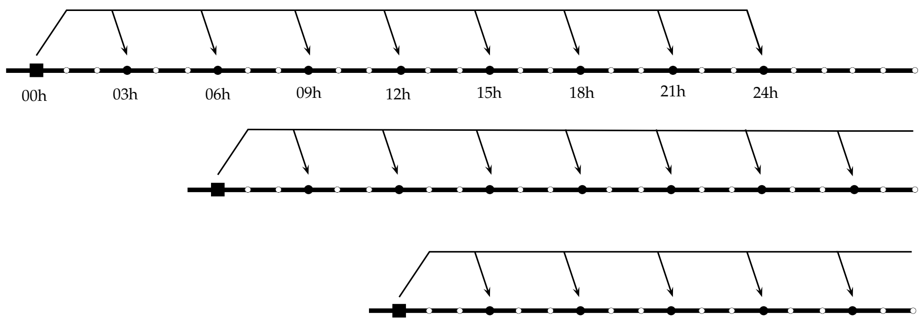

Figure 1: Scheme of a NWP delivery for $\Delta T = 6$ h, $\Delta t = 3$ h and $k^*_{max} = 24$ h. Black squares indicate different launches of the model. The forecast are generated for times indicated with black circles.

While the optimal situation corresponds to have ΔT and Δt as short as possible, these parameters typically result as a balance between computational limitations, extension of the spatial domain considered and time/spatial resolution employed by the meteorological model, among other factors. For instance, the Rapid Refresh (RAP) numerical weather model of the National Centers for Environmental Prediction (NCEP), a limited area model that covers North America with a horizontal resolution of 13 km, works with $\Delta T = 1$ h, $\Delta t = 1$ h and $k^*_{max} = 18$ h. Conversely, the global deterministic forecasting system of the European Centre for Medium-range Weather Forecasts (ECMWF) model yields forecasts according to $\Delta T = 12$ h and $\Delta t = 3$ h for $k^*_{max} = 144$ h, and $\Delta t = 6$ h for further horizons up to $k^*_{max} = 240$ h.

NWP delivery affects operational wind power forecasting in the sense that ΔT, Δt and k^*_{max} determine which meteorological forecasts (and the related prediction horizons, k^*) would be available at time t, leading to specificities and restrictions in the wind power forecasting model design.

For example, let consider the delivery scheme of Fig. 1 together with a wind power forecasting model implemented for a prediction horizon of seven hours ($k = 7$). This model, at time $t = 02$ h, would generate a power forecast for lead time $t = 09$ h in base of the meteorological forecast generated (launched) at $t = 00$ h, thus, $k^* = 9$ h. However, for time $t = 05$ h, the available meteorological forecast for the lead time $t = 12$ h (also generated at time $t = 00$ h) has a prediction horizon of $k^* = 12$ h. In summary, the considered delivery scheme of the meteorological forecasts translates on the fact that, for a wind power forecasting model with horizon k, the prediction horizon of the employed meteorological forecasts is in the range $k^* \in [k, k + \Delta T - 1]$, depending on the day time. This fact is important insofar as the performance of the power forecasting model is conditioned to the accuracy of the meteorological forecasts, which typically decreases with the increase of k^*. Thus, larger ΔT values tend to decrease the wind power performance.

Another issue is the fact that the maximum prediction horizon for a wind power forecasting model would be limited to $k^*_{max} - \Delta T + 1$ hours, because larger horizons would imply that the model is unable to generate power forecasts for some times of the day, as the related meteorological forecast is missing. For example, for the delivery scheme described above, this maximum horizon is of $k = 19$ hours; effectively, for a power forecasting model with $k = 20$ hours, there is no meteorological forecast for the related lead time at day times $t = 23, 05, 11$ and 17 h. An important problem with a power forecasting model providing forecast time series with periodically missing forecasts is that this fact is likely to distort performance assessment, which might be critical specially during a benchmark exercise.

4. Experiment setup

The case-study considered in this work consists in one real wind farm from the Global Energy Forecasting Competition dataset (GEFCOM 2012 - third wind farm), which is freely available in [31]. The employed dataset ranges from 01/07/2009 to 31/12/2010 (one year and a half) and consists of historical power measurements with hourly resolution, $\{p_t\}$, and wind speed and wind direction predictions, $\{\widehat{ws}_t\}$ and $\{\widehat{wd}_t\}$, extracted from the ECMWF model with $\Delta T = 12$ h, $\Delta t = 1$ h, and $k^*_{max} = 48$ h. The first year of data was employed to set-up the models through k-fold crossvalidation with three folds. The remaining six months of data represent the test set, employed to evaluate the performance of the models.

4.1. Proposed models

Taking into account the operational framework described in Section 3, six models are proposed. Each model actually comprises 19 quantile regression models based on the methodology described in Section 2.1 and particularised for $\tau = [0.05, 0.10, ..., 0.95]$.

- Five models, $\mathrm{M}_1^{\mathrm{RKHS}}$, ... , $\mathrm{M}_5^{\mathrm{RKHS}}$, for very-short term wind power forecasting, corresponding to prediction horizons from one hour to five hours ahead. These models put emphasis on the statistical approach, exploiting autocorrelation in wind power time series. Each quantile forecast for lead time $t+k$ is generated from the most recent power observation provided by the SCADA and the most recent meteorological forecast for time $t+k$ available at time t. Thus, data samples $(\mathbf{x}_t, y_t)$ for model M_k $(1 \leq k \leq 5)$ are in the form:

$$\begin{aligned}
\mathbf{x}_t &= [p_t, \hat{\mathbf{m}}_{t+k|t+k-k^*}] \\
\hat{\mathbf{m}} &= [\widehat{ws}, \cos(\widehat{wd}), \sin(\widehat{wd})] \\
y_t &= p_{t+k},
\end{aligned}$$

where k^*, according to Section 3, ranges between k and $k+11$ depending on the day time.

- One model, $\mathrm{M}_0^{\mathrm{RKHS}}$ for short-term wind power forecasting. This model puts emphasis on the physical approach, building optimal regressions between the meteorological forecasts and power quantile forecasts for a range of horizons up to 36 hours ahead.[3] Because the same model is employed to generate forecast time series for a wealth of prediction horizons, the prediction horizon of the meteorological forecast, k^*, is introduced as an explanatory variable. The reason why k^* is preferred as

[3]Note that the maximum horizon to have complete forecast time series, according to Section 3, is of 37 hours ahead. For convenience, we opted for a maximum horizon of one day and a half.

explanatory variable rather than k is the aforementioned implications of the NWP delivery scheme, particularly the fact that, for a given k, meteorological forecasts with different horizons k^* are involved. In addition, feeding the quantile model with k^* is deemed to be the proper approach to capture the expected decrease of accuracy of the meteorological forecast with the related prediction horizon. Data samples $(\mathbf{x}_t, y_t)$ for this model are in the form:

$$\mathbf{x}_t = [\hat{\mathbf{m}}_{t+k|t+k-k^*}, k^*]$$
$$\hat{\mathbf{m}} = [\widehat{ws}, \cos(\widehat{wd}), \sin(\widehat{wd})]$$
$$y_t = p_{t+k}.$$

The wind power variable was referred to the rated power, P_R, so that the power records belong to the interval [0, 1]. Regression variables contained in $\mathbf{x}_t$ were standardized (zero mean and unit variance) in order to put all predictors on a common scale. Standardizing is a common practice in forecasting as it helps remove the impact of the variable scale on the regression process. In addition, in view of Eq. (11), this step is deemed to be particularly important, since different predictor scales may hamper the optimisation of the parameter σ.

4.2. Benchmark Models

Linear quantile regression (Linear QR) was firstly introduced in [26], and the conditional quantile is modelled as follows:

$$q^{\tau}(\mathbf{x}; \mathbf{\Theta}) = [1\ \mathbf{x}^{T}] \cdot \mathbf{\Theta}, \tag{13}$$

where $\mathbf{x}$ are explanatory variables and $\mathbf{\Theta}$ is a vector coefficients to be determined from the historical dataset.

By considering high-order polynomials, Linear QR can be used to model non-linear relations between target and explanatory variables. However, this parametric representation is not very flexible for wind power modelling. Here, a third-order polynomial between wind speed and wind power is used as a means to model the non-linear relationship between wind and wind power output, the so-called wind power curve.

A more suitable framework is a semi-parametric QR model based on additive models theory [21, 32]. Mathematically, additive QR (or splines QR) can be expressed as follows:

$$q^{\tau}(\mathbf{x}; \theta_0) = \theta_0 + f_1(x_1) + \cdots + f_p(x_p), \tag{14}$$

where θ_0 is a constant and the functions $f_p(x_p; \tau)$ may have a parametric form (e.g., polynomial), non-parametric or semi-parametric estimated from the data. According to [32], each of the functions can be approximated by linear combinations of known basis functions of the explanatory variable, which results in linear QR model.

The R code presented in [21] is used for the benchmark model. Natural spline bases with ten degrees of freedom was used for the wind speed and the wind direction was modeled with Fourier decomposition (i.e., composition of sinusoidal functions), although periodic cubic spline basis could be also used.

4.3. Evaluation framework

Performance assessment of probabilistic models is more complex than that of deterministic models. This is so because the fitness between the observation and the predictive densities involves a higher degree of subjectivity [11]. For this reason, several metrics revealing different aspects of the forecasts are usually employed to define the evaluation framework. Three metrics widely employed in probabilistic forecasting were considered in this work: calibration, sharpness and the Continuous Ranked Probability Score (CRPS) [33, 34].

Calibration measures the difference between the nominal proportions, τ, and the empirical proportions ($\tau_{1:N}$, computed from time series $\{q_t^\tau\}$ and $\{p_t\}$, with $1 \leq t \leq N$):

$$b_{1:N}^{\tau} = \tau - \tau_{1:N}. \tag{15}$$

$b_{1:N}^{\tau} = 0$ is usually referred to as perfect calibration. A positive $b_{1:N}^{\tau}$ value means that quantile $\{q_t^\tau\}$ was infra-estimated, as it resulted higher than $\{p_t\}$ in a proportion less than τ. It is also noted that calibration is of major importance for the wind power industry and for TSOs (see discussion in [5], among others).

Sharpness is employed to assess the uncertainty conveyed by the probabilistic forecasts, regardless reliability (that is, despite how well the predictive forecast fits the observation). Sharpness is computed as the average interval size between two symmetric quantiles. Mathematically:

$$\delta^{\beta} = \frac{1}{N}\sum_{i=1}^{N}(q_t^{0.5-\beta/2} - q_t^{0.5+\beta/2}) \ , \text{ for } \beta \in (0, 1), \tag{16}$$

N being the number of samples. Given that $\delta^{\beta} = 0, \ \forall \ \beta$, means no uncertainty, as the probabilistic forecasts collapse into point-forecasts.

The CRPS is a widely employed skill score that provides an average performance of how well probabilistic forecasts compares with observations. It is considered a global skill score since it allows to jointly evaluate reliability and sharpness of the probabilistic forecasts within a single score [33], the main drawback being that bad scores are not informative about the specific aspect causing this situation. CRPS associates lower values to better performances, zero being the best mark possible. This criterion is given by:

$$CRPS = \frac{1}{N}\sum_{t=1}^{N}\int_{-\infty}^{\infty}\left(F_t(p) - H_{p_t}(p)\right)^2 \, dp, \tag{17}$$

where $F_t(p)$ is the predicted cumulative distribution function for time t, $H_{p_t}(p)$ is the Heaviside function located at the observation p_t, and N is the number of evaluated forecasts. For the case of quantile regression models, the CRPS can be estimated from a set of quantiles [35].

5. Results and discussion

5.1. Aspects of the model training

According to Section 2, there are a number of parameters to set in order to fully define a quantile regression model in the RKHS. These are the kernel parameters (for the case of the RBF kernel, the bandwidth σ), the learning rate, η and the bias term to initialize the model, b_0. Concerning σ and η, we restricted the analysis by assuming the same values for each of the nineteen quantile models enclosed within each probabilistic model. Parameter assessment was performed through k-fold crossvalidation with three folds based on the CRPS criterion obtained in the training period.

The choice for b_0 for each of the nineteen quantile models was the respective non-conditioned quantile computed from the wind power time series during the training period, Q^{τ}_{train}:

$$b_0 = Q^{\tau}_{train}. \tag{18}$$

These quantiles are depicted in Fig. 2. The rational for this choice is that, in case of a long sequence of missing data, the forgetting process makes the quantile estimates tend to the non-conditioned quantile, which actually represents a classical reference model in wind power forecasting (referred to as *climatology*). Hence, in such situations, the model would still perform as a reference. This property also applies to regions of $\mathcal{X}$ (the span of $\mathbf{x}$) where data are observed with low frequency (rare or extreme events).

Table 1 illustrates the obtained model parameters and the related performance obtained through the training period. It is observed that larger prediction horizons entail an increase of σ, which actually means, according to Eq. (11), a decrease of the kernel width. This result could be expected since the dominant underlying inputs-outputs relationship shifts from a linear pattern between consecutive power values (steaming from wind power correlation) towards the non-linear relationship between wind speed and power output (given by the power curve). Concerning the learning rate, a similar optimal value was found for models $\mathrm{M}_1^{\mathrm{RKHS}}$,..., $\mathrm{M}_5^{\mathrm{RKHS}}$, while for the case of model $\mathrm{M}_0^{\mathrm{RKHS}}$ the learning process resulted to be more effective with a lower

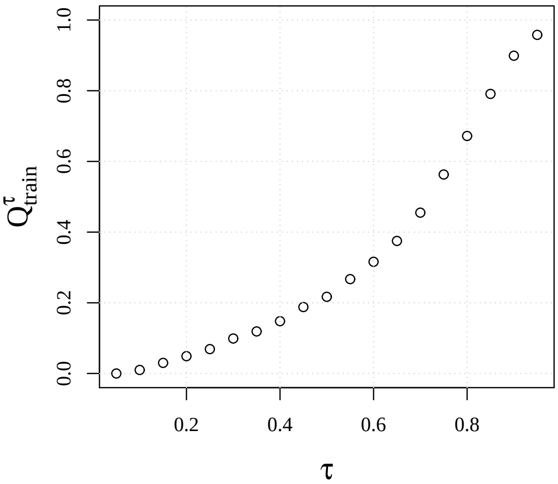

Figure 2: Non-conditioned quantiles of the wind power time series in the train set.

η. This could be explained from the fact that this model is employed for a wealth of prediction horizons, meaning that it learns as well from a higher number of samples $(\mathbf{x}_t, y_t)$. Concerning parameter λ, little variations are observed among the models. As explained, this parameter affects the forgetting factor $\eta\lambda$, the highest value obtained being 1.77e-07, which means a negligible forgetting process (a certain expansion coefficient α_i falls by 0.018% after 1 000 time steps). Finally, the decrease of the forecasting performance with the prediction horizon is a classical result in wind power forecasting [30].

Figure 3 illustrates the forecast quantiles $q_t^{0.10}$ (top) and $q_t^{0.90}$ (bottom) for the obtained model $\mathrm{M}_5^{\mathrm{RKHS}}$ at two different time instants: after two weeks ($t = 336$, left) and three months ($t = 2016$, right) of learning. In particular, the plots show the dependency of the quantiles at $t + 5$ with the forecast

Table 1: Parameter values and performance obtained for QR models in the RKHS during the training period.

	σ	λ	η	CRPS
M_1^{RKHS}	$4.64{\cdot}10^{-2}$	$1.00{\cdot}10^{-5}$	$1.00{\cdot}10^{-2}$	4.45
M_2^{RKHS}	$6.81{\cdot}10^{-2}$	$1.78{\cdot}10^{-5}$	$1.00{\cdot}10^{-2}$	6.28
M_3^{RKHS}	$1.00{\cdot}10^{-1}$	$1.00{\cdot}10^{-6}$	$1.00{\cdot}10^{-2}$	7.20
M_4^{RKHS}	$1.00{\cdot}10^{-1}$	$3.16{\cdot}10^{-6}$	$1.00{\cdot}10^{-2}$	7.76
M_5^{RKHS}	$1.00{\cdot}10^{-1}$	$1.00{\cdot}10^{-6}$	$1.00{\cdot}10^{-2}$	8.15
M_0^{RKHS}	$1.47{\cdot}10^{-1}$	$2.00{\cdot}10^{-5}$	$5.00{\cdot}10^{-3}$	9.28 $^{(a)}$

$^{(a)}$ Average value for prediction horizons from 1 to 36 hours ahead

wind speed, $\widehat{ws}_{t+5}$ (x axis), and the most recent power observation, p_t (with colours, divided into four groups for illustrative purposes). To perform each plot, the outputs of each model at the related time instant were obtained for 1000 inputs $\mathbf{x}_t$ homogeneously distributed in the input space $\mathcal{X} \in \mathbb{R}^n$. It can be seen that, after two weeks of learning, the modeled quantiles deviates little from the associated non-conditioned quantiles, specially for regions in $\mathcal{X}$ with small density of samples (shown by gray dots in the figure). After three months, the learning process has deepened, allowing the modeled quantiles to show predictions coherent with the underlying power curve and the influence of the most recent power observation (see left bottom).

5.2. Test results and discussion

Figure 4 shows the CRPS versus the prediction horizon obtained for the different models. Concerning models for very short-term forecasting ($M_1^{(\cdot)}$-$M_5^{(\cdot)}$), all the models showed relatively similar performance. This result could

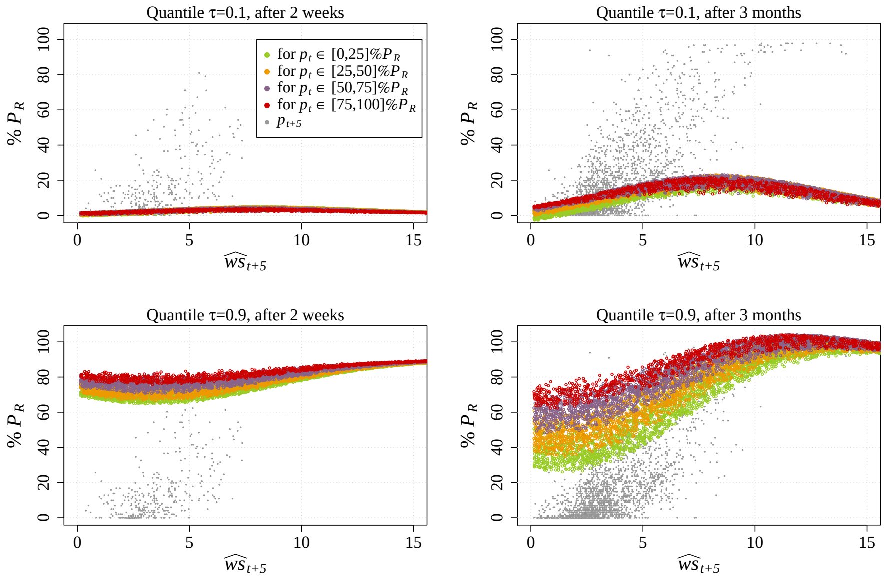

Figure 3: Quantile models for $\tau = 0.1$ (left) and $\tau = 0.9$ (right) from model $\mathrm{M}_5^{\mathrm{RKHS}}$ at two different time instants: after two weeks (top) and three months (bottom) of learning. P_R stands for rated power.

be attributed to the fact that, for such prediction horizons, the underlying relationships between inputs and outputs are dominated by autoregressive dynamics, which are typically well sized with simple linear regression. Thus, more advanced strategies, as the QR in the RKHS, contributes little or nothing to improve the probabilistic forecasts. Conversely, focusing on models $\mathrm{M}_0^{(\cdot)}$, the on-line quantile regression in the RKHS provided better results for a range of prediction horizons (namely, up to 20 hours ahead). While the improvement with respect to the spline approach implied a decrease of only up to 4.75%, this happens within a range of prediction horizons of special interest for bidding in electricity markets. For example, prediction horizons related to intra-day markets in the Iberian electricity market ranges between

3-30 hours ahead.

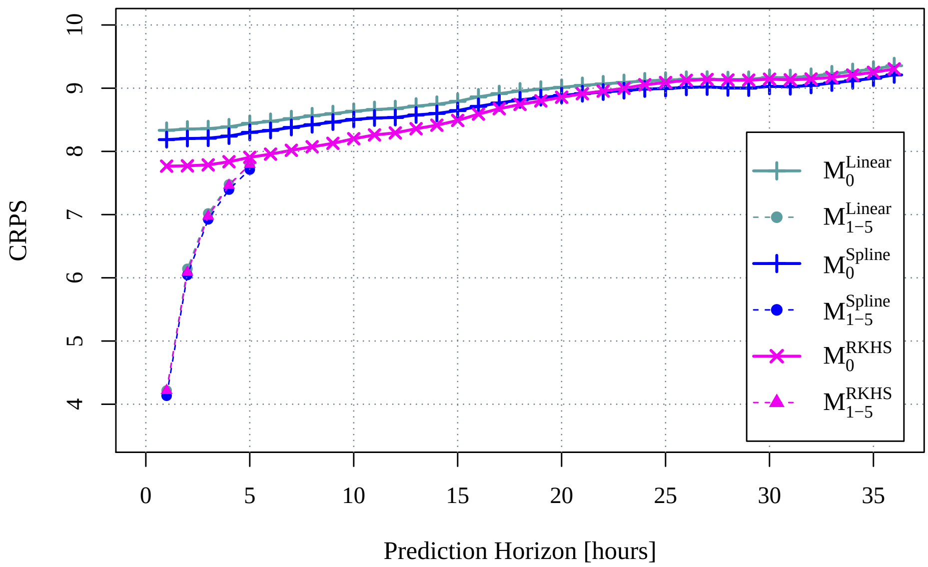

Figure 4: CRPS according to type of model and prediction horizon.

Model calibration is depicted in Fig. 5 according to type of model and prediction horizon (on top, results for models $M_1^{(\cdot)}$-$M_5^{(\cdot)}$; at the bottom, results for models $M_0^{(\cdot)}$ broken down according to prediction horizon). Results show a fairly improvement obtained by RKHS models as compared with reference models. While for the latter the absolute value of $b_{1:N}^{\tau}$ ranged between 0.01 and 0.03, the calibration of the proposed models remained within the ± 0.015 interval.

This result could be due to the adaptive nature of the algorithm, which refines the quantile regression with every new sample according to learning rules based on the pinball loss function. To confirm this hypothesis, the evo-

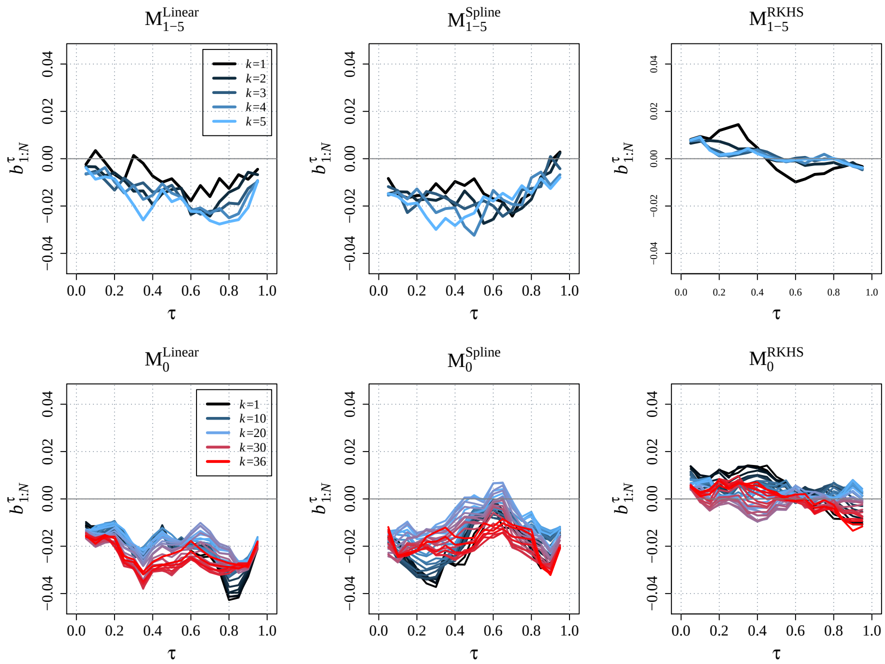

Figure 5: Calibration according to type of model and prediction horizon.

lution of the bias over time during the test set was analysed. The particular case of $k = 5$ was considered, though similar analysis were performed for other prediction horizons. Figure 6 shows the calibration bias for quantile $\tau = 0.5$ computed for an increasing size of samples from the test set, $b_{1:t}^{0.5}$, obtained for models $\mathrm{M}_5^{\mathrm{Linear}}$, $\mathrm{M}_5^{\mathrm{Spline}}$ and $\mathrm{M}_5^{\mathrm{RKHS}}$. An additional curve, labeled as $\mathrm{M}_5^{\mathrm{RKHS,frozen}}$, reflects the calibration related to model $\mathrm{M}_5^{\mathrm{RKHS}}$ without adaptivity, that is, without learning during the test set. To implement this, no more terms were added in the kernel expansion given in Eq. (6) during the test set. Thus, models $\mathrm{M}_5^{\mathrm{Linear}}$, $\mathrm{M}_5^{\mathrm{Spline}}$ and $\mathrm{M}_5^{\mathrm{RKHS,frozen}}$ generate predictive densities according uniquely to patterns captured from the training set. It can be seen that the calibration of these models evolve over time with a similar drift, suggesting that this result derives from wind power dynamic seasonalities. Conversely, model $\mathrm{M}_5^{\mathrm{RKHS}}$ performs nearly perfect calibration consistently over time, reflecting a clear improvement with respect to the previous models. Consequently, the observed result could be considered as an achievement of the adaptive nature of the model.

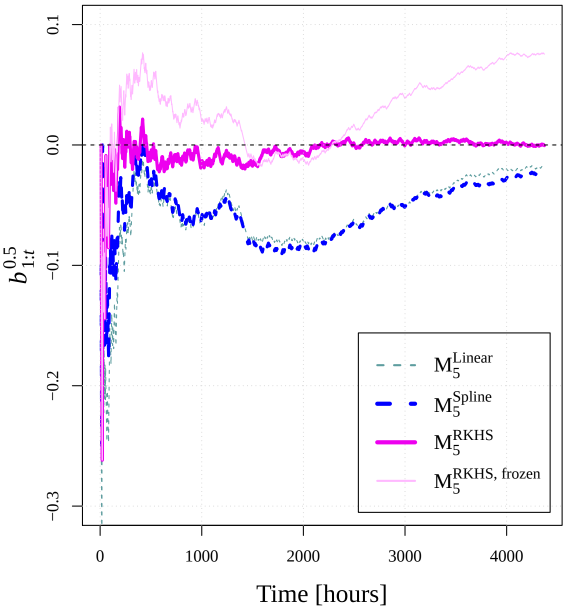

Figure 6: Calibration bias for quantile $\tau = 0.5$ evaluated over an increasing window of the test set. Case for $k = 5$.

Lastly, results concerning sharpness are illustrated in Fig. 7. As could be expected, an increase of the uncertainty conveyed by the models with the prediction horizon is observed, specially for models $\mathrm{M}_1^{(\cdot)}$-$\mathrm{M}_5^{(\cdot)}$, where the most recent power observation is key for generating predictive densities in the very short-term. For these models, the single noticeable difference is that models

$M_{(\cdot)}^{\mathrm{RKHS}}$ provided slightly higher averaged intervals between quantiles $\tau = 0.05$ and $\tau = 0.95$ (i.e. $\beta = 0.9$). Concerning models $M_0^{(\cdot)}$, the sharpness of linear and spline models were found to depend little on the prediction horizon, the latter showing slightly better marks. Model M_0^{RKHS} showed a sharpness more dependent on the prediction horizon, though the marks roughly evolved from that of splines (for the shortest horizons) to that of the linear quantile regression (for the largest horizons).

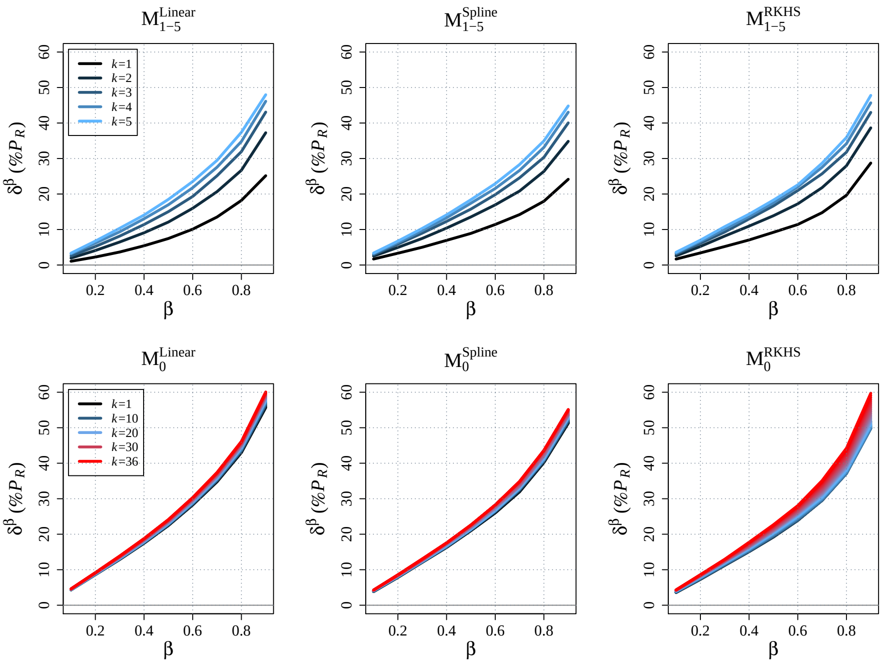

Figure 7: Sharpness according to type of model and prediction horizon. δ_β is expresses as a percentage of the rated power, P_R.

A final remark on the presented model must be done. It relates the growing sum problem. According to Eq. (6), the number of terms in the expansion grows linearly with t, increasing the computational and memory requirements. Though it did not represent a limitation in this work, as the time series were relatively short (one year and a half), this issue has been recognized as the main bottleneck in kernel adaptive algorithms [36]. To avoid this problem, [28] proposed expansion truncation by dropping the oldest samples, given that the α_i coefficients decrease with time as $(1 - \eta\lambda)^t$. This option does not represent a proper solution in cases where the optimal forgetting factor results very low (as it was the case in this work) in combination with time series spanning over several years (or less, for sub-hourly time resolution). Another option is to explore the extent to which the contribution of certain samples to the model can be approximated by linear combinations of the contribution of another samples, so that not every sample must translate into a new term in Eq. (6). This idea is the base of sparsification [37, 38]. Quantization of the feature space was presented as an alternative in [36]. According to [39], sparsification and quantization do not fully solve the growing sum problem, as they curb the rate of growth from linear to sublinear. Instead, the author proposed a new approach by approximating kernel evaluations using finite dimensional inner products in a randomized feature space. This approach was applied to the Kernel Least Mean Square (KLMS) algorithm. Thus, its application to the stochastic gradient descent algorithm employed in this work could represent a future line of research.

Finally, for illustrative purposes, the predictive densities provided by some of the quantile regression models in the RKHS during a four days period are shown in Fig. 8. In particular, results for one hour, five hours and 24 hours ahead are shown on top, middle and bottom, respectively. The impact of the prediction horizon on the wind power uncertainty can be clearly appraised: while autocorrelation in wind power allows narrow predictive densities for one hour ahead, where the contribution of the most recent power observation is paramount, uncertainty increases with the forecast horizon, especially for the case of model $\mathrm{M}_0^{\mathrm{RKHS}}$ where the forecasts are generated essentially from NWPs.

6. Conclusions

We have presented a new on-line quantile regression model based on the Reproducing Kernel Hilbert Space (RKHS) framework. This approach, based on linear regressions in the feature space, combines simplicity with non-linear modelling capabilities. In addition, since the model takes roots in kernel methods, the complexity and computational requirements remain relatively independent with the number of explanatory variables. This represents an advantage as compared with other models typically employed in probabilistic wind power forecasting.

An important feature of the model is on-line learning, which allows the predictive densities to account for smooth variations over time, typically present in wind power dynamics due to seasonalities and wind turbine aging. The developed algorithm is based on the stochastic descent gradient introduced in [28], here generalised for the case of including a bias term. The analysis of the obtained learning rules permitted us to connect this parameter with the non-conditioned quantiles of wind power output, as this allows the model to perform as a classical reference model in wind power forecasting (climatology) when facing missing input data and rare or extreme events.

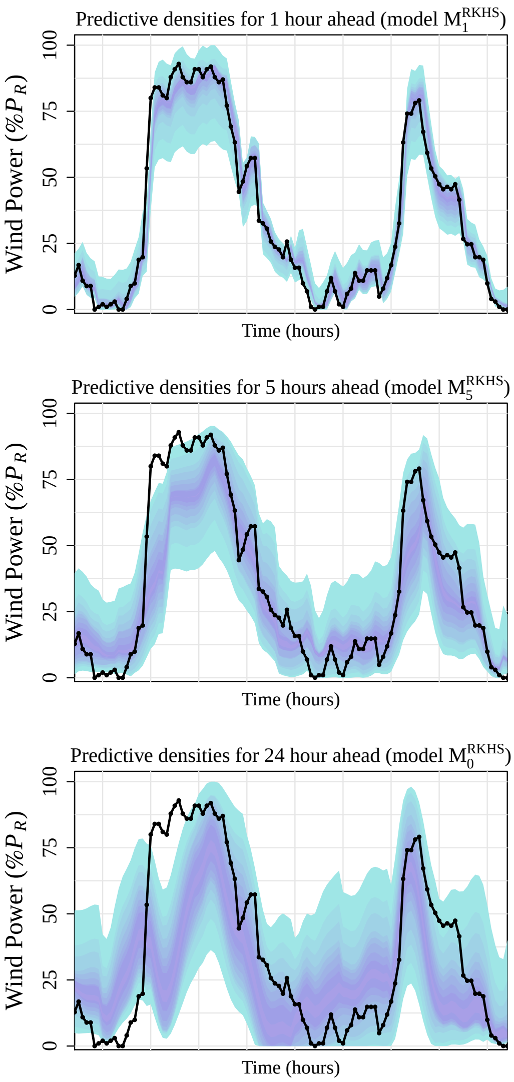

Figure 8: Probabilistic forecasts provided for $k = 1$ (top), $k = 5$ (middle) and $k = 24$ (bottom) during four days.

Specificities of the operational framework of wind power forecasting were also described. These relate the impact of the delivery scheme of a meteorological model on the generation of predictive power densities. Within this context, a benchmark exercise to predict 19 wind power quantiles from wind speed and direction forecasts delivered each 12 hours was performed for a real case study. Two model configurations were proposed: one for very-short term forecasting up to five hours ahead (including the most recent power observation gathered from the wind farm as explanatory variable) and other for short-term forecasting up to 36 hours ahead (including the prediction horizon of the meteorological model). Two benchmark models were considered, comprising linear quantile regression (with a third-order polynomial for the wind speed to better capture the wind power curve) and spline quantile regression (with ten degrees of freedom for the wind speed and using Fourier decomposition for the wind direction).

Results were based on a multi-criterion evaluation framework. The findings showed noticeable improvements in terms of calibration, this criterion being of major importance for the wind power industry and for Transmission System Operators [5]. Further analyses led us to attribute this achievement to the adaptive nature of the model. In terms of Continuous Ranked Probability Score, the RKHS approach obtained modest improvements for prediction horizons between 6 and 20 hours ahead. This result could be interesting insofar as this range of horizons is of special interest for bidding in electricity markets. Finally, concerning the sharpness criterion, no remarkable improvements were observed in reducing the uncertainty captured by the model.

In summary, the presented results support regression in the RKHS as a competitive approach for wind power probabilistic forecasting. Indeed, more studies could be performed to gain insights into its capabilities and limitations. The issue of the growing sum problem was specially remarked. While it did not represent a limitation for the considered dataset of one year and a half, this issue needs to be addressed in future studies dealing with longer time periods. In this regard, several possibilities discussed in the literature were outlined, and their application represents clear paths for improvement.

Acknowledgements

This work is partially financed by the DAVE – Departamento de Aeronaves y Vehculos Espaciales (UPM) and by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project ≪POCI-01-0145-FEDER-006961≫, National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology), as part of project UID/EEA/50014/2013.

References

[1] R. J. Bessa, C. L. Moreira, B. Silva, M. A. Matos, Handling renewable energy variability and uncertainty in power systems operation, Wiley Interdisciplinary Reviews: Energy and Environment 3 (2) (2014) 156–178.

[2] M. A. Matos, R. Bessa, Setting the operating reserve using probabilistic wind power forecasts, IEEE Transactions on Power Systems 26 (2) (2011) 594–603.

[3] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. Bessa, D. Flynn, E. Lazaro, N. Detlefsen, Methodologies to determine operating reserves due

to increased wind power, IEEE Transactions on Sustainable Energy 3 (4) (2012) 713–723.

[4] J. Wang, A. Botterud, R. Bessa, H. Keko, V. Miranda, J. Akilimali, L. Carvalho, D. Issicaba, Wind power forecasting uncertainty and unit commitment, Applied Energy 88 (11) (2011) 4014–4023.

[5] R. Bessa, M. Matos, I. Costa, L. Bremermann, I. Franchin, R. Pestana, N. Machado, H.-P. Waldl, C. Wichmann, Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study, IEEE Transactions on Sustainable Energy 3 (4) (2012) 827–837.

[6] W. A. Bukhsh, C. Zhang, P. Pinson, A integrated multiperiod opf model with demand response and renewable generation uncertainty, IEEE Transactions on Smart Grid.

[7] A. Botterud, J. Wang, Z. Zhou, R. Bessa, H. Keko, J. Akilimali, V. Miranda, Wind power trading under uncertainty in LMP markets, IEEE Transactions on Power Systems 27 (2) (2012) 894–903.

[8] I. Gonzlez-Aparicio, A. Zucker, Impact of wind power uncertainty forecasting on the market integration of wind energy in Spain, Applied Energy 159 (2015) 334–349.

[9] T. Soares, P. Pinson, T. Jensen, H. Morais, Optimal offering strategy for wind power in energy and primary reserve markets, IEEE Transactions on Sustainable Energy.

[10] C. Monteiro, R. Bessa, V.Miranda, A. Botterud, J.Wang, G. Conzelmann, Wind power forecasting: state-of-the-art 2009, Tech. Rep. Report ANL/DIS-10-1, Argonne National Laboratory (November 2009).

[11] Y. Zhang, J. Wang, X. Wang, Review on probabilistic forecasting of wind power generation, Renewable and Sustainable Energy Reviews 32 (2014) 255–270.

[12] C. Gallego-Castillo, A. Cuerva-Tejero, O. Lopez-Garcia, A review on the recent history of wind power ramp forecasting, Renewable & Sustainable Energy Reviews 52 (2015) 1148–1157. doi:10.1016/j.rser.2015.07.154.

[13] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, B. Klöckl, From probabilistic forecasts to statistical scenarios of short-term wind power production, Wind Energy 12 (1) (2009) 51–62.

[14] E. B. Iversen, J. M. Morales, J. K. Mller, H. Madsen, Short-term probabilistic forecasting of wind speed using stochastic differential equations, International Journal of Forecasting.

[15] R. Bessa, V. Miranda, A. Botterud, Z. Zhou, J. Wang, Time-adaptive quantile-copula for wind power probabilistic forecasting, Renewable Energy 40 (1) (2012) 29–39.

[16] J. Jeon, J. W. Taylor, Using conditional kernel density estimation for wind power density forecasting, Journal of the American Statistical Association 107 (497) (2012) 66–79.

[17] P. Pinson, Very short-term probabilistic forecasting of wind power with generalised logit-normal distributions, Journal of the Royal Statistical Society: Series C 61 (4) (2012) 555–576.

[18] P. Kou, F. Gao, X. Guan, Sparse online warped gaussian process for wind power probabilistic forecasting, Applied Energy 108 (2013) 410–428.

[19] G. Sideratos, N. Hatziargyriou, Probabilistic wind power forecasting using radial basis function neural networks, IEEE Transactions on Power Systems 27 (4) (2012) 1788–1796.

[20] J. B. Bremnes, Probabilistic wind power forecasts using local quantile regression, Wind Energy 7 (1) (2004) 47–54.

[21] H. A. Nielsen, H. Madsen, T. S. Nielsen, Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts, Wind Energy 9 (1-2) (2006) 95–108.

[22] C. Gallego-Castillo, P. Pinson, H. Madsen, A. Costa, A. Cuerva-Tejero, Influence of local wind speed and direction on wind power dynamics – Application to offshore very short-term forecasting, Applied Energy 88 (11) (2011) 4087–4096.

[23] G. Rubio, H. Pomares, L. J. Herrera, I. Rojas, Computational and Ambient Intelligence: 9th International Work-Conference on Artificial Neural Networks, IWANN 2007, San Sebastián, Spain, June 20-22, 2007. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, Ch. Kernel Methods Applied to Time Series Forecasting, pp. 782–789. doi:10.1007/978-3-540-73007-1_94.

[24] I. Takeuchi, Q. Le, T. Sears, A. Smola, Nonparametric quantile estimation, Journal of Machine Learning Research 7 (2006) 1231–1264.

[25] C. Gallego-Castillo, R. Bessa, A. Cuerva-Tejero, L. Cavalcante, Wind Power Probabilistic Forecast in the Reproducing Kernel Hilbert Space, in: Power Systems Computation Conference (PSCC), Genoa, Italy, 2016 (accepted presentation), 2016.

[26] R. Koenker, G. B. Jr, Regression quantiles, Econometrica 46 (1) (1978) 33–50.

[27] A. Smola, B. Schølkopf, A tutorial on support vector regression, Statistics and Computing 14 (3) (2004) 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

[28] J. Kivinen, A. J. Smola, R. C. Williamson, Online learning with kernels, IEEE Transactions on Signal Processing 52 (8) (2004) 2165–2176.

[29] A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis, kernlab – an S4 package for kernel methods in R, Journal of Statistical Software 11 (9) (2004) 1–20.

[30] G. Giebel, The state of the art in short-term prediction of wind power-A literature overview (2nd ed.), Tech. rep., ANEMOS.plus/SafeWind projects (2011).

[31] T. Hong, P. Pinson, S. Fan, Global energy forecasting competition 2012, International Journal of Forecasting 30 (2) (2014) 357–363.

[32] T. Hastie, R. Tibshirani, Generalized Additive Models, Chapman and Hall, London, 1990.

[33] H. Hersbach, Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather and Forecasting (15) (2000) 559–570.

[34] P. Pinson, H. A. Nielsen, J. K. Moller, H. Madsen, G. N. Kariniotakis, Non-parametric probabilistic forecasts of wind power: Required properties and evaluation, Wind Energy 10 (6) (2007) 497–516. doi:10.1002/we.230.

[35] G. Anastasiades, P. McSharry, Quantile forecasting of wind power using variability indices, Energies 6 (2) (2013) 662–695. doi:10.3390/en6020662.

[36] B. Chen, S. Zhao, P. Zhu, J. C. Principe, Quantized kernel least mean square algorithm, IEEE Transactions on Neural Networks and Learning Systems 23 (1) (2012) 22–32. doi:10.1109/TNNLS.2011.2178446.

[37] Y. Engel, S. Mannor, R. Meir, The kernel recursive least-squares algorithm, IEEE Transactions on Signal Processing 52 (8) (2004) 2275–2285. doi:10.1109/TSP.2004.830985.

[38] P. P. Pokharel, W. Liu, J. C. Principe, Kernel least mean square algorithm with constrained growth, Signal Processing 89 (3) (2009) 257–265. doi:http://dx.doi.org/10.1016/j.sigpro.2008.08.009.

[39] A. Singh, N. Ahuja, P. Moulin, Online learning with kernels: Overcoming the growing sum problem, in: Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on, 2012, pp. 1–6. doi:10.1109/MLSP.2012.6349811.