Energy 113 (2016) 355–365

Contents lists available at ScienceDirect

Energy

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

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

A R T I C L E I N F O

Article history:
Received 15 April 2016
Received in revised form
8 July 2016
Accepted 11 July 2016

Keywords:
Wind power
Quantile regression
Reproducing Kernel Hilbert Space (RKHS)
Probabilistic forecast
Short-term
On-line

A B S T R A C T

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 cross-validation. Results showed a noticeable improvement in terms of calibration, a key criterion for the wind power industry. Modest improvements in terms of Continuous Ranked Probability Score (CRPS) were also observed for prediction horizons between 6 and 20 h ahead.

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 Refs. [10] and [11]. Five main classes of probabilistic forecasting algorithms can be found in the literature: conditional kernel density estimation (KDE), (b) semi-parametric regression, (c) machine learning, (d) analog methods and (e) 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].

Three examples of conditional KDE algorithms are: (a) time-adaptive conditional KDE that explores the non-parametric copula for modelling the dependency between wind speed/direction and power [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]; (c) conditional KDE applied to power data transformed with a logarithmic function and that uses a new boundary kernel method [17].

One work about semi-parametric regression is presented in Ref. [18], 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

* Corresponding author.
E-mail address: cristobaljose.gallego@upm.es (C. Gallego-Castillo).

http://dx.doi.org/10.1016/j.energy.2016.07.055
0360-5442/

Bayesian model based on warped Gaussian process is proposed in Ref. [19], and employed to generate probabilistic wind power forecasts. In Ref. [20] 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. Finally, in Ref. [21] the use of gradient boosting trees and quantile random forests, combined with feature engineering techniques and post-processing with isotonic regression, resulted in high-quality probabilistic forecasts. Classical multilayer perceptron neural networks can also be extended to forecast conditional quantiles by using as cost function a hybrid of the quantile loss function and Huber norm [22]. However, this approach has two limitations: it comes with an additional parameter, the threshold magnitude of the Huber norm; the neural network training usually stops in local optima (non-convex optimization problem) and is highly affected by the random initialization of the weights.

Recent research in analog-based methods, proposes the use of a Euclidean metric to rank past forecasts' similarity to the current forecast and the probability density function is estimated using a set of N past observations (N best analogs) [23]; the results show that the analog ensemble method outperforms the linear quantile regression method. A similar concept, based on the k-nearest neighbors algorithm and KDE is proposed in Ref. [24].

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) [25] and additive models with splines [26] or radial basis functions (RBF) [27]. 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 [28]. 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; the choice of the RBF centers is not trivial (e.g., the k-means algorithm can lead to local optima [27]).

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 Ref. [29] and references therein). The presented model implements quantile regression in the Reproducing Kernel Hilbert Space (RKHS) according to the framework described in Ref. [30]. 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 Ref. [31].

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 characterized by meterological forecast availability each 12 h. 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 $\mathscr{X} \in \mathbb{R}^n$, 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 $f : \mathscr{X} \to \mathbb{R}$, with $\mathscr{X} \in \mathbb{R}^n$, is a function to be determined. The most straightforward strategy to define f is that of linear quantile regression [32]. 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 $\varphi : \mathscr{X} \to \mathscr{H}$, where $\mathscr{H}$ 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 Ref. [30], among others):

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

where $||\cdot||_{\mathscr{H}}^2$ 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 $l_\tau : \mathbb{R}^2 \to \mathbb{R}^+$ is a loss function of the forecast error. According to quantile regression theory [30], $l_\tau(y_i, q^\tau(\mathbf{x}_i))$ 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 Ref. [33], among others), it can be demonstrated that the model that