# Wind Power Scenario Generation Using a High Dimensional Nonparametric Framework

Marcelo Ruas and Alexandre Street

**Abstract**

Producing Renewable Generation scenarios is a topic of interest power system, as they are required for many different applications. In this paper, we focus on generating future scenarios of wind power. Most time series methods used to produce such forecasts rely on the assumption of Gaussian errors or Gaussian distribution, but most applications on power systems deals with nongaussian time series. Using a nonparametric approach such as Quantile Regression allows the statistical model to adapt to the data without requiring a distribution to be known *a priori*. We develop in this paper the Quantile Regularized Adaptive LASSO (QRAL), a methodology which is based on the Adaptive LASSO for quantile regression with the addition of a second derivative filter, that explores the similarity of neighbour quantiles to reduce the noise in estimation and producing better estimations and simulations. A real case study comparing with different benchmarks in the literature, the QRAL was able to outperform them in terms of minimizing the Mean Absolute Percentile Error of future scenarios in recovering the historic monthly quantiles.

**Keywords**

*Quantile Regression, Model Identification, Non-gaussian time series model*

## I. Introduction

Renewable energy power is an emergent topic which is demanding attention from the academic community. The installed capacity of renewable energy plants has been increasing in a fast pace and projections point out that wind power alone will account to 18% of global power by 2050 [1]. In spite of its virtues, several new challenges are inherent when dealing with such power source, due to its unpredictability. To overcome this lack of certainty, one has to work with many different possibilities of outcome.

New statistical models capable of handling such difficulties are an emerging field in power systems literature [2]–[8] The main objective in such literature is to propose new models capable of generating scenarios of renewable generation (RG) which are demanded in (i) energy trading, (ii) unit commitment, (iii) grid expansion planning, and (iv) investment decisions (see ([9]–[12]) and references therein). In stochastic optimization, problems such as Unit Commitment, Economic Dispatch, Transmission Expansion Planning all use scenarios as input. Such scenarios are used to characterize the probability distribution within the optimization under uncertainty framework. When working with robust optimization, bounds for probable ranges of coefficients are needed.

Conventional statistical models are often focused on estimating the conditional mean of a given random variable. By reducing the outcome to a single statistic, we loose important informations about the time series random behavior. In order to account for the process inherent variability it is important to consider probability forecasting. [2] reviews the commonly used methodologies regarding probabilistic forecasting models, splitting them in parametric and nonparametric classes. Main characteristics of **parametric models** are (i) assuming a distribution shape and (ii) low computational costs. ARIMA-GARCH, for example, model the RG series by assuming the distribution *a priori*. On the other hand, **nonparametric models** (i) don't require a distribution to be specified, (ii) needs more data to produce a good approximation and (iii) have a higher computational cost. Popular methods are Quantile Regression (QR), Kernel Density Estimation, Artificial Intelligence or a mix of them.

Most time series methods rely on the assumption of Gaussian errors. However, RG time series such as wind and solar are reported as non-Gaussian [3], [13]–[15]. To circumvent this problem, the usage of nonparametric methods - which does not rely on assuming any previously assumed distribution - is adequate. Quantile Regression (QR) is a tool for constructing a methodology for non-gaussian time series, because of its facility to implement on commercial solvers and to extend the original model. However, when estimating a distribution function, as each quantile is estimated independently, the monotonicity of the distribution function may be violated. This issue - also known as crossing quantiles - can be adressed by constraining the sequence of quantiles to be in an increasing order. Other possibility is making a transformation afterwards, as shown in [16].

The seminal work [17] defines QR as we use today. By this formulation, the conditional quantile is the solution of an optimization problem where we minimize the sum of the check function (defined formally in the next session). Instead of using the classical regression to estimate the conditional mean, the QR determines any quantile from the conditional distribution. Applications are enormous, ranging from risk measuring at financial funds (the Value-at-Risk) to a central measure robust to outliers. By estimating many quantiles on a thin grid of probabilities, one can have as many points as desired from the estimated conditional distribution function. In [18], the application of QR is extended to time series, when the covariates are lagged values of $y_t$. In our work, beyond autoregressive terms, it is also considered other exogenous variables as covariates.

In [4]–[8], QR is employed to model the conditional distribution of Wind Power Time Series. An updating quantile regression model is presented by [5]. The authors present a modified version of the simplex algorithm to incorporate new observations without restarting the optimization procedure. In [6], the authors build a quantile model from already existent independent Wind Power forecasts. The approach by [4] is to use QR with a nonparametric methodology. The authors add a penalty term based on the Reproducing Kernel Hilbert Space, which allows a nonlinear relationship between the explanatory variables and the output. This paper also develops an on-line learning technique, where the model is easily updated after each new observation. In [8], wind power probabilistic forecasts are made by using QR with a special type of Neural Network (NN) with one hidden layer, called extreme learning machine. In this setup, each quantile is a different linear combination of the features of the hidden layer. The authors of [19] use the weighted Nadaraya-Watson to estimate the conditional function in the time series.

Regularization is a topic already explored in previous QR papers. The work by [20] defines the properties and convergence rates for QR when adding a penalty proportional to the $\ell_1$-norm to perform variable selection, using the same idea as the LASSO [21]. The ADALASSO equivalent to QR has its properties investigated by [22]. In this variant, the penalty for each variable has a different weight, and this modification ensures that the oracle property is being respected. In [23], [24], the AdaLASSO is employed to QR with multiple quantile regressions at the same time, relating the estimated coefficients of different quantiles.

In this work we propose the use of a non parametric technique, which as already mentioned is free from user distribuitional specification, to simulate future scenarios of wind power generation. Such technique relies on a quantile autoregressive (QAR) framework in the same spirit as [17], [18], [25]. Notwithstanding, as an innovation from the referenced works, in order to capture which variables would improve model fit, from a set of relevant ones, we propose the use of a LASSO cost function plus an interquantile term. Such term acts as a filter to impose coefficient stability for a given covariate across the set of quantiles. In addition, interquantile dependency is explored as in [23], [24]. As a second innovative feature of this work, we propose the inclusion of a penalization parameter in the second difference. We argue that such strategy avoids extreme quantiles to be shrinked to zero as fast as the central ones. For the best of the authors knowledge, no other work has developed a methodology where regularization and estimation of the conditional distribution using QR is carried on at the same time with the objective of generating future scenarios for WPG time series.

The objective of this paper is, then, to propose a new methodology to address nonparametric time-series model focused on RG. This may be seen as a multiple quantile regression problem that specifies a time series model based on the empirical conditional distribution. The main contributions are:

- A nonparametric methodology to model the conditional distribution of RG time series to produce

scenarios.

- The proposition of a methodology that selects the global optimal solution with parsimony both on the selection of covariates as on the quantiles. This methodology is based on the Adaptive LASSO for QR (Linear Programming).
- Regularization techniques applied to an ensemble of quantile functions to estimate the conditional distribution, solving the issue of non-crossing quantiles. On regularizing quantiles, we propose a smoothness on the coefficient value across the sequence of quantiles.

The remaining of the paper is organized as follows. In section II, we present the quantile regression framework in which we build the model as well as introduce the Quantile Regression Adaptive LASSO. In section III, we discuss the estimation procedures for them. The regularization strategies are also presented on this section. Finally, in section IV, we present a few controlled studies with simulated data and a final case study using real data from wind power is presented in order to test our methodology. Section V will conclude this article.

## II. Quantile Regression based time series model

Let the $\alpha$-conditional quantile function of $Y$ for a given value $x$ of the $d$-dimensional random variable $X$, i.e., $Q_{Y|X} : [0,1] \times \mathbb{R}^d \to \mathbb{R}$, can be defined as

$$Q_{Y|X}(\alpha, x) = F^{-1}_{Y|X}(\alpha, x) = \inf\{y : F_{Y|X}(y, x) \geq \alpha\}. \tag{1}$$

Let a dataset be composed from $\{y_t, x_t\}_{t\in T}$, where $T$ is the set of time indexes, let $\rho$ be the check function

$$\rho_\alpha(x) = \begin{cases} \alpha x & \text{if } x \geq 0 \\ (1-\alpha)x & \text{if } x < 0 \end{cases}. \tag{2}$$

The sample quantile function for a given probability $\alpha$ is based on a finite number of observations and is the solution to minimizing the loss function $L_\alpha(\cdot)$:

$$\hat{Q}_{Y|X}(\alpha, \cdot) \quad \in \quad \arg\min_{q\in\mathcal{Q}} L_\alpha(q) = \sum_{t\in T} \rho_\alpha(y_t - q(x_t)), \tag{3}$$

The $\alpha$-quantile function $q(\cdot)$ belongs to a function space $\mathcal{Q}$. We might have different assumptions for space $\mathcal{Q}$, depending on the type of function we want to find for $q$. A few properties, however, must be achieved by our choice of space, such as being continuous and having limited first derivative. In this paper, we consider the case where $\mathcal{Q}$ is a linear function's space, so we have that

$$q_\alpha(x_t) = \beta_0 + \beta^T x_t.$$

The quantile function is approximated by a sequence of $|J|$ (where $J$ is an index set) quantiles $q_{\alpha_1} \leq q_{\alpha_2} \leq \cdots \leq q_{\alpha_{|J|}}$. Denote the closely related set $A = \{\alpha_j \mid j \in J\}$, whose elements lies on $[0,1]$, such that $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{|J|} < 1$.

We apply the QR model to estimate the conditional distribution, $\hat{Q}_{Y_{t+h}|X_{t+h},Y_t,Y_{t-1},\ldots}(\alpha, \cdot)$, of time series $\{y_t\}_{t\in T}$ $h$-steps ahead, where $X_{t+h}$ makes reference to a vector of exogenous variables. Once the conditional distribution is estimated, future scenarios can be obtained by simulation.

### *Regularization on selection of covariates and coefficient slope*

When dealing with many candidates of covariates, one has to deal with the problem of selecting the relevant ones out of a set of candidates which improve model fit. In practice, it means that some coefficients from the vector $\beta_j = [\beta_{1j} \cdots \beta_{pj}]$ might assume zero value. There are many ways of selecting a subset of variables among the available options. A classical approach for this problem is the Stepwise algorithm [26], [27], [21], which includes variables in sequence.

A popular way of variable selection which fits on a linear programming context are the LASSO/AdaLASSO techniques, which consists in penalize the $\ell_1$-norm. Besides shrinking coefficients towards to zero, it has also the capability of effectively pushing coefficients to zero (an effect that ridge regression cannot

achieve [21]). The usage of LASSO and AdaLASSO in QR context is the topic of study in [20], [22]–[24], [28]. We refer to the reader the work from [20], where it is possible to find specific properties and convergence rates when using the LASSO to perform model selection in a quantile regression framework.

Regarding the penalization parameter $\lambda$, which dictates the shrinkage magnitude of the linear coefficients, the level of parsimony of the model can be defined by the user through such quantity. This is due to the fact that higher values of $\lambda$ means less variables selected to be non-zero.

The single $\alpha$-quantile AdaLASSO is estimated by the following optimization problem:

$$\min_{\beta_0,\beta} \sum_{t\in T} \rho_\alpha(y_t - (\beta_0 + \sum_{p\in P} \beta_p x_{tp})) + \lambda \sum_{p\in P} w_p|\beta_p|. \tag{4}$$

What differs the AdaLASSO from the LASSO is the inclusion of the term $w_p$. Suppose the model (4) is estimated with all $w_{pj} = 1$, the output of the optimization problem are coefficients $\beta_{pj}^{LASSO}$. The AdaLASSO coefficients $\beta_{pj}^{AdaLASSO}$ are obtained When solving the same optimization problem while letting $w_{pj} = 1/|\beta_{pj}^{LASSO}|$.

As we are interested on the conditional distribution as a whole, we estimate multiple quantiles at once. In order to produce a coherent distribution function, the output of the problem must respect certain properties, such as being monotone increasing. Besides that, one can expect that the value of similar quantiles be produced by similar models. If the coefficient of a given $p$ covariate changes too abruptly with respect to a change on the probability $\alpha$, there is a high probability that the estimation did not produce good results rather than this noise belonging to the true model. In order to correct this much common behavior in QR estimation, we introduce a second derivative filter, given by the discrete approximation shown below:

$$D_{pj}^2 := \frac{\left(\frac{\beta_{p,j+1}-\beta_{pj}}{\alpha_{j+1}-\alpha_j}\right) - \left(\frac{\beta_{p,j}-\beta_{p,j-1}}{\alpha_j-\alpha_{j-1}}\right)}{\alpha_{j+1}-\alpha_{j-1}}. \tag{5}$$

With this approach, one can keep track on the crossing quantiles issue as well as using a interquartile structure as a strategy to reduce noise on estimation The works by [23], [24] also use multiple quantile regressions at once and make use of interquantile similarities to produce regularization on the quantiles. In [23], the author uses the norm $\|\beta\|_{1\infty} = \sum_{p=1}^{|P|} \max\{|\beta_j^{(k)}|\}$ as penalization. Such penalization is imposed on the maximum value among all quantiles for a given covariate. This idea is extended by [24], that uses a fused AdaLASSO mixing the LASSO penalization with the absolute interquantile difference.

### *A. Quantile Regularized Adaptive LASSO (QRAL)*

The statistical model defined by the model QRAL is defined by the vector of coefficients $\beta_0$ and the matrix of size $|P| \times |J|$ of regressor coefficients $\beta_{pj}$. These coefficients are the solution from the minimization problem given below:

$$\min_{\beta_{0j},\beta_j} \sum_{j\in J} \left( \sum_{t\in T} \rho_{\alpha_j}(y_t - (\beta_{0j} + \beta_j^T x_t)) \right.$$
$$\left. +\lambda \sum_{p\in P} w_{pj}^\delta \mid \beta_{pj} \mid \right) + \gamma \sum_{p\in P} \sum_{j\in J'} |D_{pj}^2|, \tag{6}$$

where the weights $w_{pj} = 1/\tilde{\beta}_{pj}$ and $\tilde{\beta}_{pj}$ are the coefficients from the first-step LASSO estimation. The parameter $\delta$ is an exponential parameter usually set to 1. The sum of absolute values that compose the second derivative filter $\sum_{j\in J'} \sum_{p\in P} |D_{pj}^2|$ is added on the objective function multiplied by a tuning parameter $\gamma$, where the set $J' = \{2, \ldots, |J| - 1\}$.

## III. Computational Issues

This section presents computational aspects of the estimation, such as presenting the mathematical programming formulation of the QRAL, presenting the evaluation metric and the cross-validation methodology, which is essential in the process of building the function $\hat{Q}_{y_\tau|X}(\alpha)$. The methodology is implemented in R and Julia languages (relying heavily on the packages JuMP, Gurobi, RCall and Dierckx) and using the Gurobi solver.

The QRAL model, as described in equation (6), can be implemented as a linear programming problem as shown below:

$$\min_{\beta_0,\beta,\varepsilon^+_{tj},\varepsilon^-_{tj}} \sum_{j\in J}\sum_{t\in T}(\alpha_j\varepsilon^+_{tj} + (1-\alpha_j)\varepsilon^-_{tj})$$

$$+\lambda\sum_{p\in P}\sum_{j\in J} w_{pj}(\xi^+_{pj}+\xi^-_{pj}) + \gamma\sum_{p\in P}\sum_{j\in J'}(D2^+_{pj}+D2^-_{pj}) \tag{7}$$

subject to

$$\varepsilon^+_{tj} - \varepsilon^-_{tj} = y_t - \beta_{0j} - \beta_j^T x_t, \qquad \forall t\in T, \forall j\in J, \tag{8}$$

$$\xi^+_{pj} - \xi^-_{pj} = \beta_{pj}, \qquad \forall p\in P, \forall j\in J \tag{9}$$

$$D2^+_{pj} - D2^-_{pj} = \frac{\left(\frac{\beta_{p,j+1}-\beta_{pj}}{\alpha_{j+1}-\alpha_j}\right) - \left(\frac{\beta_{p,j}-\beta_{p,j-1}}{\alpha_j-\alpha_{j-1}}\right)}{\alpha_{j+1}-\alpha_{j-1}}, \qquad \forall p\in P, \forall j\in J', \tag{10}$$

$$\beta_{0j} + \beta_j^T x_t \le \beta_{0,j+1} + \beta_{j+1}^T x_t, \qquad \forall t\in T, \forall j\in J_{(-1)}, \tag{11}$$

$$\varepsilon^+_{tj}, \varepsilon^-_{tj} \ge 0, \qquad \forall t\in T, \forall j\in J, \tag{12}$$

$$\xi^+_{pj}, \xi^-_{pj} \ge 0, \qquad \forall p\in P, \forall j\in J, \tag{13}$$

$$D2^+_{pj}, D2^-_{pj} \ge 0, \qquad \forall p\in P, \forall j\in J'. \tag{14}$$

Variables $\varepsilon_t^+$ and $\varepsilon_t^-$ represent the quantities $|y-q(\cdot)|^+$ and $|y-q(\cdot)|^-$, respectively. The first line on the objective function represents the sum of the check function over all $j$: $\rho_{\alpha_j}(y-q(\cdot)) = \alpha_j\varepsilon^+_{tj} + (1-\alpha_j)\varepsilon^-_{tj}$. The constraint (11) assures that the quantile function be monotonic by forcing that, for every $x_t$ and $\alpha_j$-quantile, $q_{\alpha_j}(x_t) \le q_{\alpha_{j+1}}(x_t)$. This second derivative term is implemented on the optimization problem by adding a penalty on the objective function to penalize the absolute value $|D^2_{pj'}|$. The tuning parameter $\gamma$ controls how rough the sequence $\{\beta_{pj}\}_{j\in J}$ can be, for a given $p$.

### A. Evaluation Metric

In this section, we present the metric for which the model fit can be evaluated. From this metric, we show two different forms - information criteria and cross validation - to determine the best tuning parameters $\lambda$ and $\gamma$.

In order to evaluate our predictions, we need to define an error metric. The minimization of this error metric is the objective in estimating the statistical model. As conditional distribution is the focus in this paper, we use a performance measurement which emphasizes the correctness across quantiles. Depending on the application, it might be interesting to put different weights on different quantiles. In this work, however, we will treat every quantile as equals concearning the error measure. The chosen error function is Mean Absolute Percentage Error (MAPE), defined by

$$MAPE^{\alpha_j} = \frac{1}{|J|}\frac{1}{|T|}\sum_{t\in T}\sum_{j\in J}\left|\frac{q_t^{\alpha_j} - y_t^{\alpha_j}}{q_t^{\alpha_j}}\right|. \tag{15}$$

where $q_t^{\alpha_j}$ is the $\alpha$-quantile from the data (in the case study, we use the monthly distribution as a good enough approximation of the true quantile, as RG series such as wind power are stationary) and $y_t^{\alpha_j}$ is the $\alpha$-quantile from these scenarios. This function has the advantage of penalizing error proportionally to the quantile value it is estimating.

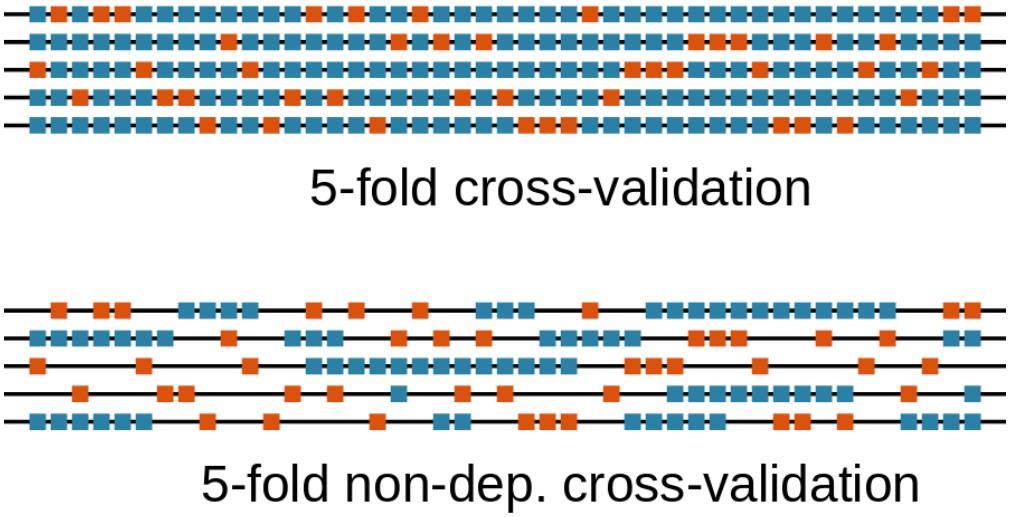



Fig. 1: $\mathcal{K}$-fold CV and $\mathcal{K}$-fold with non-dependent data. Observations in blue are used to estimation and in orange for evaluation. Note that non-dependent data does not use all dataset in each fold. Image from [29].

### *B. Time-series Cross Validation*

Estimating the QRAL involves the use of parameters $\lambda$ and $\gamma$, which should be known *a priori*. In statistics and machine learning, a popular technique is using Cross-validation (CV) to select the best set of parameters from the range of possibilities. How to select their values among this range is a crucial point in our methodology, as the estimated coefficients vary considerably with respect to parameter choice.

Out of the different possible implementations of CV, we use the $\mathcal{K}$-fold CV. It consists in first partitioning the dataset in $\mathcal{K}$ equally sized sets, which are the $\mathcal{K}$ folds. For each fold $k \in \{1, \ldots, \mathcal{K}\}$, the remaining $\mathcal{K} - 1$ folds are used to estimate the model using parameter $\theta$ (for the QRAL model, $\theta = [\gamma \quad \lambda]^T$) and predicting the values in fold $k$. The error function $MAPE_\theta^\alpha$ measures the result of this prediction. So, the CV error is given by the sum of all folds and all quantiles, for a given parameter $\theta$

$$CV(\theta) = \sum_{k \in \mathcal{K}} \sum_{j \in J} MAPE_\theta^\alpha.$$

The optimum parameter $\theta_{CV}^*$, according to this methodology, is the value of $\theta$ which minimizes the CV error

$$\theta_{CV}^* = \arg\min_\theta CV(\theta). \tag{16}$$

The usage of CV is not straightforward when data is dependent, which is the case of time series. As it is time dependent, one can be interested in using either all observations or to take the dependency away to do not interfere on the estimation. The works [29] and [30] deals specifically with the usage of CV in a time series context. They provide tests with both $\mathcal{K}$-fold CV and $\mathcal{K}$-fold with non-dependent data. Both schemes are shown of Figure 1. In both settings, the training data is randomly split into a collection of sets $S_k$, forming a $\mathcal{K}$ size partition. Each of these $S_k$ is used as test set, while the rest is used to estimate coefficients which will be used to predict values of $S_k$. As there are $\mathcal{K}$ folds, this procedure is done $\mathcal{K}$ times. So, for a given vector of tuning parameter $\theta$, the CV score is given by the sum of the error function for each fold. As the CV score is nonconvex, the optimization in (16) is done by iterating over a sequence of values in a thin grid and choosing the smallest one.

### C. Scenario generation

This section presents how to generate future scenarios of time series $y_t$ from the estimated coefficients from a QR model. To produce $S$ different future scenarios $\{\hat{y}_t\}_{t=|T|+1}^{|T|+K}$, we use the following procedure:

Procedure for simulating $S$ future scenarios of $\{y_t\}$

1) At first, initialize time index $\tau = |T| + 1$.
2) For each scenario $s \in S$, estimate coefficients from QRAL. These coefficients produce quantile estimates for a collection of discrete probabilities $\alpha \in A$. Let $\tilde{Q}_{y_\tau|X} : A \to \mathbb{R}$ be the discrete quantile function. In order to produce the continuous function $\hat{Q}_{y_\tau|X} : [0,1] \to \mathbb{R}$ from the discrete function, use linear interpolation connecting the points. As $0 < \alpha_1 < \cdots < \alpha_{|J|} < 1$, there are no quantile estimates for the intervals $[0, \alpha_1]$ and $[\alpha_{|J|}, 1]$. These gaps are filled by linearly extending the line that connects $\alpha_1$ to $\alpha_2$ on the left hand side and extending the line that connects $\alpha_{|J|-1}$ to $\alpha_{|J|}$ on the right hand side until the support $[0, 1]$ is fully mapped.
3) Let $U$ be a random variable with uniform distribution over the interval $[0, 1]$. By using the result of the Probability Integral Transform (PIT), random variable $F_{y_\tau}^{-1}(U)$ has the same distribution as $y_{tau}$. The time series value at time $\tau$ and $s$ scenario $y_{tau,s}$ is built by drawing one random observation of U and applying the transformation of the PIT.
4) Let $\tau = \tau + 1$. Go back to step 2) and repeat the procedure until all the future scenarios of size $K$ are generated.

## IV. Case Studies

### A. Controlled Studies I - Autoregressive Process

Firstly, the performance of QRAL methodology is evaluated by simulating a widely used autoregressive process and in the sequel estimating the QRAL framework to evaluate if it is capable to recover the parameters fixed at the beginning. The aforementioned autoregressive process is set as

$$y_t = \beta_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \tag{17}$$

with length 400 and $\beta_1 = 0.3$. This value was chosen due to the fact that the sample size of RG time series have such length in general.

Three different methods were employed to estimate the process given by (17):

1) **Quantile Regularized Adaptive LASSO (QRAL)**, which estimates one different model for each quantile $Q_{y_t|X}(\alpha, \cdot)$, for all $j \in J$. In practice, it means that each coefficient $\beta_{1j}$ is estimated with regularization on each quantile.
2) **Quantile Regression as Koenker (QRK)** as originally proposed by [17], where each coefficient $\beta_{1j}$ is estimated using QR.
3) A simple **Autoregressive (AR)** process.

After simulate 1000 different time series given by Equation (17), $\beta_{1j}$ was estimated for each one of the aforementioned models. Regarding parameter $\gamma$ (see Eq. (7)-(14)) it was estimated using CV, as described in section III-B. Since in this experiment the model has only one lag, model selection will not be evaluated, hence $\lambda = 0$.

The main objective of this simulation experiment was to evaluate how our nonparametric methodology can correctly recover the true AR(1) process. The model performance was evaluated by examining how closely the estimated quantiles are from the populational ones. The results for each method are depicted in Figure 2, where a boxplot containing the results for the 1000 simulations is shown. The conclusions from this experiment are: (i) Coefficient estimation errors for the central quantiles are not far from those estimated by the AR; (ii) extreme quantiles are usually harder to estimate, due to having fewer observations; as a consequence, the estimation error increases on the extremes (iii) QRAL has an advantage over QRK in terms of variance of estimators.

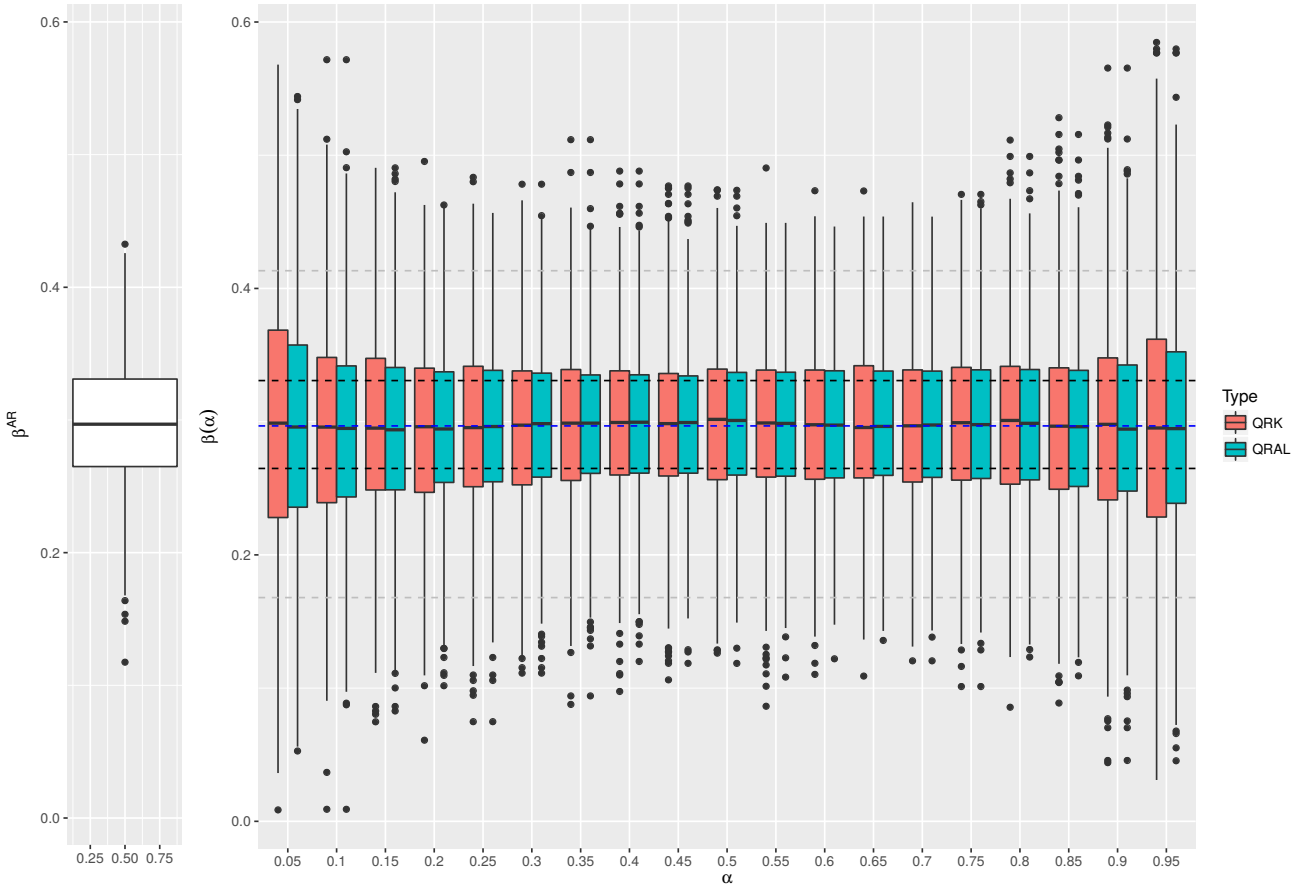



Fig. 2: Boxplot showing estimated coefficient after 1000 iterations. On the left hand side, the boxplot of the AR(1) coefficient estimation. Note that for the AR(1) the coefficient is equal for all probabilities $\alpha$. On the right hand side, the boxplot of the regular QR (where $\gamma = 0$) and the QRAL where $\gamma$ is selected using cross-validation.

### *B. Controlled Studies II - Quantile Autoregressive Process*

In the second controlled study, the performance of the QRAL is evaluated in a Quantile Autoregressive Process. While in the first study the time series $y_t$ is generated by a process with a coefficient $\beta_1$ constant across quantiles, in this experiment $\beta_1$ is a piecewise linear function of $\alpha$, described graphically on Figure 3.

The simulation of a QAR process involves two steps. The first is to determine the values of $\beta_0$ from a set of values of $\beta_1(\alpha)$; the latter given as input. These values of $\beta_{0j}$ are the output of the following optimization problem:

$$\min_{\beta_0} \ \beta_{0,|J|} - \beta_{0,1} \tag{18}$$

$$\text{subject to} \tag{19}$$

$$\beta_{0j} + \beta_j^T x_t + h \leq \beta_{0,j+1} + \beta_{j+1}^T x_t, \quad \forall t \in T, \forall j \in J_{(-1)}, \tag{20}$$

where $h$ is the minimum distance allowed between two neighbour quantiles. If $\beta(\alpha)$ is constant for an

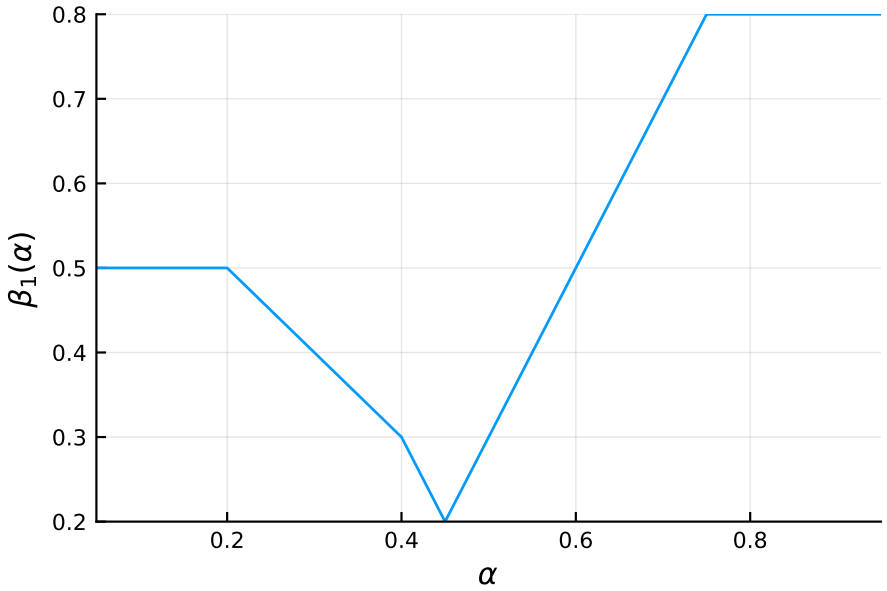



Fig. 3: Coefficient $\beta_1(\alpha)$ of the Autoregressive term $y_{t-1}$ for generating the QAR process.

interval and $h$ is set to zero, this group of quantiles would superimpose one another. Once both vectors $beta_0$ and $\beta_1$ are defined, the 1-step temporal dynamic of the process can be summarized by Figure 4.

The second step consists in simulating the values of $y_t$. For a given value of $y_{t-1}$, the next term $y_t$ in the time series follows a distribution that can be constructed by the quantiles shown on Figure 4.

This controlled study has the same sample setup as the one in the previous section: 1000 random samples of length 400 and the same three methods tested (QRAL, QRK and AR). The objective of this experiment is to test their performance with respect to their error on the estimation of each coefficient. The summary of results is given by Table I, which shows the the aggregated value of quadratic error across all $j \in J$. The conclusion from this experiment is that both QR methods are much superior than an Autoregressive process such as a AR(1) when the coefficient $\beta$ is allowed to change with probability $\alpha$.

TABLE I: Cumulated quadratic error of coefficient estimation for each method after 1000 random samples

| AR(1) | QR-AdaLASSO | QRK |
|---|---|---|
| 38.75 | 5.039 | 5.314 |

### *C. Real Case Study*

In this section, the QRAL methodology is tested in generating future scenarios of RG. A real time series of Wind Power is the input for estimating coefficients that are employed to generate scenarios by using the procedure described on section III-C. The time series is composed of 31 years (from 1981 to 2011) of monthly observations of a wind farm located in the Brazilian northeast, measured in Megawatts. The yearly series is shown on Figure 5.

The last 4 years of this period is leaved as out-of-sample test data. Six different methods are used to generate scenarios: QR-LASSO (Quantile Regularized LASSO), QRAL, QRK, LASSO (original LASSO for QR), AdaLASSO (original AdaLASSO for QR) and SARIMA. The estimated coefficients for the QR based methods are presented on Figure 6. The second derivative filter acts reducing the

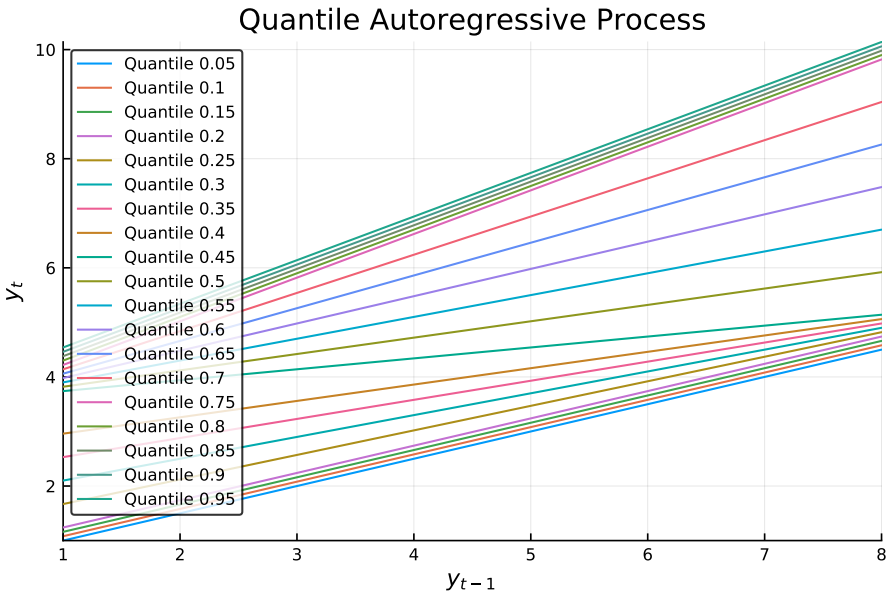



Fig. 4: Quantiles of $y_t$ vs $y_{t-1}$. For each given value of $y_{t-1}$ on the $x$-axis, the curves indicate different levels of quantiles for the values of $y_t$.

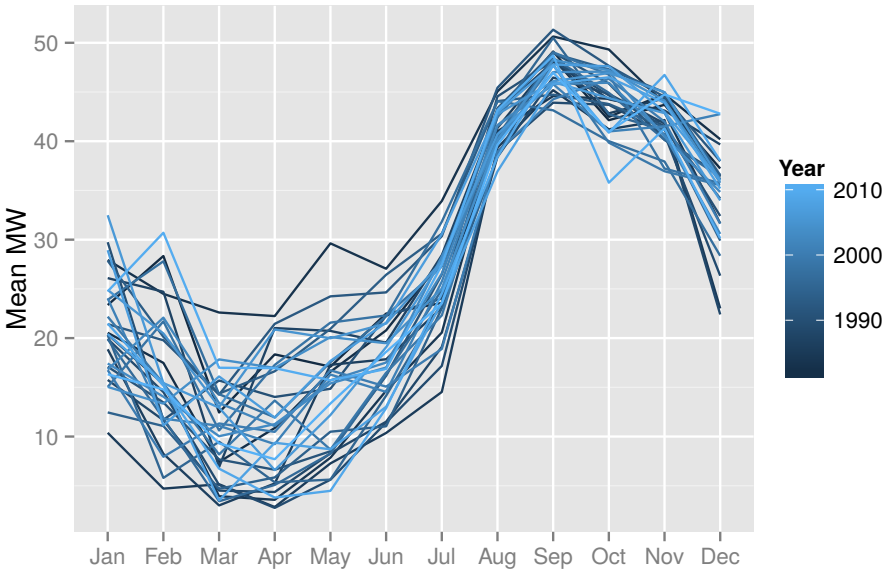



Fig. 5: Icaraizinho yearly data. Each serie consists of monthly observations for each year.

noise, and its effect can be seen clearly when comparing the estimated coefficients of QR-LASSO with the LASSO and the QRAL with the AdaLASSO.

The accuracy of the generated scenarios - considering the MAPE metric - in recovering the historic quantiles for each month is the metric used for evaluation. These results are shown on Table II. The QRAL is the method that produced the scenarios with the smaller errors.

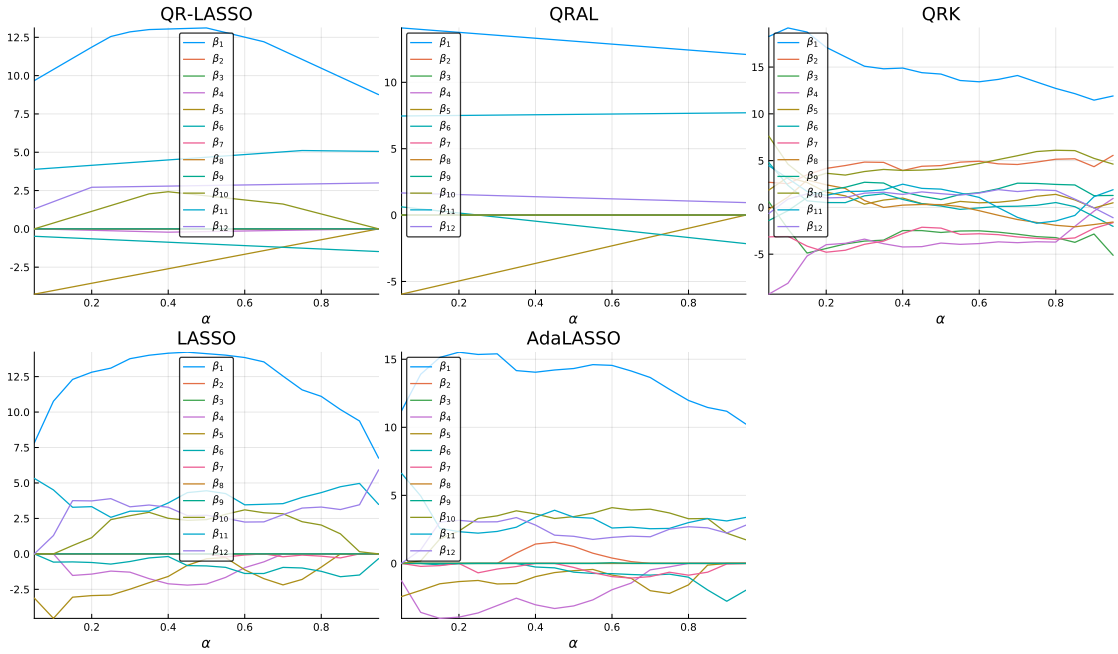



Fig. 6: Estimated coefficients for each QR based methods. Each line represents the coefficient function $\beta_p$ for a given covariate $y_{t-p}$.

TABLE II: Cumulated MAPE across all $\alpha_j$ quantiles

| QR-LASSO | QR-AdaLASSO | QRK | LASSO | AdaLASSO | SARIMA |
|---|---|---|---|---|---|
| 6.809 | 3.653 | 3.940 | 6.282 | 4.653 | 5.834 |

## V. Further Studies

The obtained results using the QR methodology give incentives to keep doing experiments and finding new aspects which could help improve the process of generating scenarios. Of the many possible paths, the following are the next steps to be investigated by the researchers.

### A. Information Criteria for Quantile Regression

Using CV can be computationally expensive, as the full estimation is done several times for each tuning parameter - in this case, $\gamma$ and $\lambda$. Other form of deciding the quantity of variables that provides a good equilibrium between in-sample prediction and parsimony is the Information Criteria.

Information criteria summarizes two aspects. One of them refers to how well the model fits the in-sample observations and the other part penalizes the quantity of covariates used in the model. By penalizing how big our model is, we prevent overfitting from happening. So, in order for a covariate to be included in the model, it must supply enough goodness of fit. In [31], it is presented a variation of the Schwarz criteria for M-estimators that includes quantile regression. The Schwarz Information Criteria (SIC), adapted to the quantile autoregression case, is presented below:

$$SIC(m) = \sum_{j \in J} \log \left( \sum_{t \in T} \rho_{\alpha_j} (y_t - \beta_0 - \beta^T x_t) \right) + \frac{\log |T|}{2|T||P|} K(m), \tag{21}$$

where $K(m)$ is the quantity of coefficients $\beta_{pj}$ greater than zero in the model $m$. By minimizing the $SIC$ function, the chosen model is the one with the best combination, according to this metric, of fit and parsimony among all models.

Even though CV is very popular and produce great results, selecting model with Information Criteria is much quicker. For the case where the selected model is very similar, it might be the case that the estimation methodology may change a little bit. It is definitely a topic that worths researching.

### *B. Quantile Autoregression with a nonparametric approach*

Fitting a linear estimator for the Quantile Auto Regression is less efficient if nonlinearity is present in the quantile functions. This nonlinearity may produce a linear estimator that underestimates the quantile for a chunk of data while overestimating for the other chunk. To prevent this issue from occurring we propose to use a nonparametric quantile regression model.

Let $\{\tilde{y}_t\}_{t=1}^n$ be the sequence of observations in time $t$ and let $\tilde{x}_t$ be the $p-$lagged time series of $\tilde{y}_t$, such that $\tilde{x}_t = L^p(\tilde{y}_t)$, where $L$ is the lag operator. Matching each observation $\tilde{y}_t$ with its $p-$lagged correspondent $\tilde{x}_t$ produces $n-p$ pairs $\{(\tilde{y}_t, \tilde{x}_t)\}_{t=p+1}^n$ (note that the first $p$ observations of $y_t$ must be discarded). When the sequence of observations of $x$ are reorder in such a way that they are in growing order

$$\tilde{x}^{(p+1)} \leq \tilde{x}^{(p+2)} \leq \cdots \leq \tilde{x}^{(n)},$$

the new sequences can be defined $\{x_i\}_{i=1}^{n-p} = \{\tilde{x}^{(t)}\}_{t=p+1}^n$ and $\{y_i\}_{i=1}^{n-p} = \{\tilde{y}^{(t)}\}_{t=p+1}^n$, where $T' = \{2, \ldots, n-p-1\}$.

The optimization model to estimate the nonparametric quantile is as follows:

$$\begin{aligned} \hat{q}_{\alpha_j}(x_t) &= \arg\min_{q_{tj}} \sum_{t \in T'} \rho_{\alpha_j}\left(y - q_{tj}\right) \\ &+\lambda_1 \sum_{t \in T'} |D^1_{x_t} q_{tj}| + \lambda_2 \sum_{t \in T'} |D^2_{x_t} q_{tj}|, \end{aligned} \tag{22}$$

where $D^1 q_t$ and $D^2 q_t$ are the first and second derivatives of the $q_\alpha(x_t)$ function, calculated as follows:

$$D^2_{x_t} q_{tj} = \frac{\left(\frac{q_{\alpha t+1} - q_{tj}}{x_{t+1} - x_t}\right) - \left(\frac{q_{tj} - q_{\alpha t-1}}{x_t - x_{t-1}}\right)}{x_{t+1} - x_{t-1}},$$

$$D^1_{tj} = \frac{q_{\alpha t+1} - q_{tj}}{x_{t+1} - x_t}.$$

The full model can be rewritten as a LP problem as bellow:

$$\min_{q_{tj},\varepsilon^+_{tj},\varepsilon^-_{tj},\xi_t} \quad \sum_{j\in J}\sum_{t\in T'}\left(\alpha_j\varepsilon^+_{tj}+(1-\alpha_j)\varepsilon^-_{tj}\right) \tag{23}$$

$$+\lambda_1\sum_{t\in T'}\gamma_{tj}+\lambda_2\sum_{t\in T'}\xi_{tj}$$

$$s.t. \quad \varepsilon^+_t-\varepsilon^-_{tj}=y_t-q_{tj}, \qquad \forall t\in T',\forall j\in J, \tag{24}$$

$$D^1_{tj}=\frac{q_{\alpha t+1}-q_{tj}}{x_{t+1}-x_t}, \qquad \forall t\in T',\forall j\in J, \tag{25}$$

$$D^2_{tj}=\frac{\left(\frac{q_{\alpha t+1}-q_{tj}}{x_{t+1}-x_t}\right)-\left(\frac{q_{tj}-q_{\alpha t-1}}{x_t-x_{t-1}}\right)}{x_{t+1}-x_{t-1}}. \qquad \forall t\in T',\forall j\in J, \tag{26}$$

$$\gamma_{t\alpha}\geq D^1_{tj}, \qquad \forall t\in T',\forall j\in J, \tag{27}$$

$$\gamma_{t\alpha}\geq -D^1_{t\alpha}, \qquad \forall t\in T',\forall j\in J, \tag{28}$$

$$\xi_{t\alpha}\geq D^2_{tj}, \qquad \forall t\in T',\forall j\in J, \tag{29}$$

$$\xi_{t\alpha}\geq -D^2_{t\alpha}, \qquad \forall t\in T',\forall j\in J, \tag{30}$$

$$\varepsilon^+_{t\alpha},\varepsilon^-_{t\alpha},\gamma_{t\alpha},\xi_{t\alpha}\geq 0, \qquad \forall t\in T',\forall j\in J, \tag{31}$$

$$q_{tj}\leq q_{tj+1}, \qquad \forall t\in T',\forall j\in J_{(-1)} \tag{32}$$

The output of this optimization problem is a sequence of ordered points $\{(x_t,q_{tj})\}_{t\in T}$, for all $j\in J$. The next step is to interpolate these points in order to provide an estimation for any other value of $x_t$. To address this issue, we propose using a linear interpolation. Note that $q_{tj}$ is a variable that represents only one point of the $\alpha_j$-quantile function $q_\alpha(x_t)$.

The quantile estimation is done for different values of $\lambda_2$. By using different levels of penalization on the second difference, the estimation can be more or less adaptive to the fluctuation. It is important to notice that the usage of the $\ell_1$-norm as penalty leads to a piecewise linear solution $q_{tj}$. Figure 7 shows the quantile estimation for a few different values of $\lambda_2$.

## VI. Concluding Remarks

In this paper, we propose a QR based methodology with regularization both on the quantile and on the covariates to produce future scenarios of RG. These scenarios are input for various applications in power systems and are essential in measuring risk in energy trading, planning the expansion of the energy systems and the dispatch problem. This nonparametric framework is specially useful when the data is not distributed according to a known distribution. The scenarios generated outperformed other known benchmarks, both in the QR litterature (the Koenker model) as in the classic time series framework (SARIMA). The results obtained in this paper brings incentives continue researching methods for nongaussian time series, as they are present in many real world applications.

## References

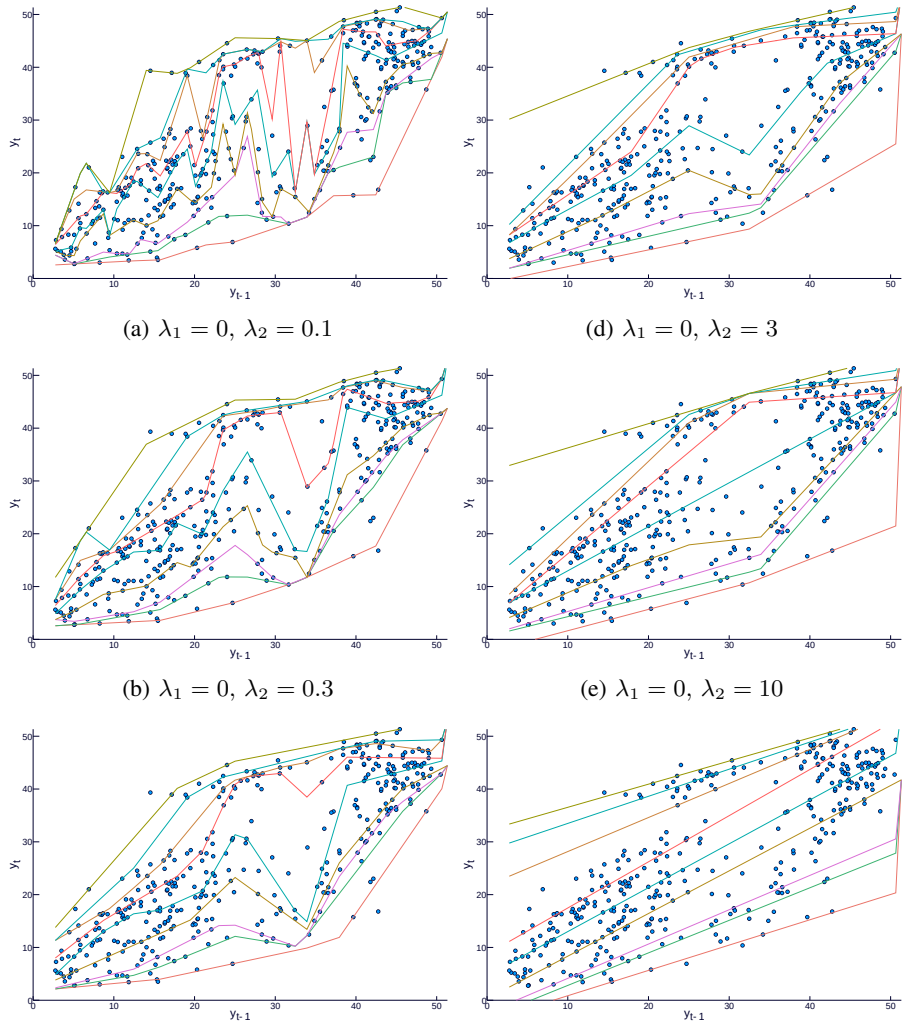

[1] International energy agency. [Online]. Available: https://www.iea.org/newsroom/news/2013/october/wind-power-seen-generating-up-to-18-of-global-power-by-2050.html

[2] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, Apr. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032114000446

[3] R. J. Bessa, V. Miranda, A. Botterud, J. Wang, and M. Constantinescu, "Time adaptive conditional kernel density estimation for wind power forecasting," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 4, pp. 660–669, 2012.

[4] C. Gallego-Castillo, R. Bessa, L. Cavalcante, and O. Lopez-Garcia, "On-line quantile regression in the rkhs (reproducing kernel hilbert space) for operational probabilistic forecasting of wind power," *Energy*, vol. 113, pp. 355–365, 2016.

[5] J. K. Møller, H. A. Nielsen, and H. Madsen, "Time-adaptive quantile regression," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292–1303, Jan. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947307002502




(a) $\lambda_1 = 0,\ \lambda_2 = 0.1$ (d) $\lambda_1 = 0,\ \lambda_2 = 3$

(b) $\lambda_1 = 0,\ \lambda_2 = 0.3$ (e) $\lambda_1 = 0,\ \lambda_2 = 10$

(c) $\lambda_1 = 0,\ \lambda_2 = 1$ (f) $\lambda_1 = 0,\ \lambda_2 = 200$

Fig. 7: Quantile estimations for a few different values of $\lambda_2$. The quantiles represented here are $\alpha = (5\%, 10\%, 25\%, 50\%, 75\%, 90\%, 95\%)$. When $\lambda_2 = 0.1$, on the upper left, we see a overfitting on the estimations. The other extreme case is also shown, when $\lambda_2 = 200$ the nonparametric estimator converges to the linear model.


[6] H. A. Nielsen, H. Madsen, and T. S. 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.

[7] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, Jan. 2004. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/we.107/abstract

[8] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct Quantile Regression for Nonparametric Probabilistic Forecasting of Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, Jul. 2017.

[9] A. Moreira, D. Pozo, A. Street, and E. Sauma, "Reliable renewable generation and transmission expansion planning: Co-optimizing system's resources for meeting renewable targets," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[10] R. Jabr, "Robust transmission network expansion planning with uncertain renewable generation and loads," *IEEE*

*Transactions on Power Systems*, vol. 28, no. 4, pp. 4558–4567, 2013.

[11] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2587–2596, July 2016.

[12] A. C. Passos, A. Street, and L. A. Barroso, "A dynamic real option-based investment model for renewable energy portfolios," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 883–895, March 2017.

[13] J. Jeon and J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66–79, 2012.

[14] J. W. Taylor and J. Jeon, "Forecasting wind power quantiles using conditional kernel estimation," *Renewable Energy*, vol. 80, pp. 370–379, 2015.

[15] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, July 2017.

[16] V. Chernozhukov, I. Fernández-Val, and A. Galichon, "Quantile and Probability Curves Without Crossing," *Econometrica*, vol. 78, no. 3, pp. 1093–1125, May 2010. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.3982/ECTA7880/abstract

[17] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

[18] R. Koenker, Z. Xiao, J. Fan, Y. Fan, M. Knight, M. Hallin, B. J. M. Werker, C. M. Hafner, O. B. Linton, and P. M. Robinson, "Quantile Autoregression [with Comments, Rejoinder]," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 980–1006, 2006. [Online]. Available: http://www.jstor.org/stable/27590777

[19] Z. Cai, "Regression Quantiles for Time Series," *Econometric Theory*, vol. 18, no. 1, pp. 169–192, 2002. [Online]. Available: http://www.jstor.org/stable/3533031

[20] A. Belloni and V. Chernozhukov, "L1-Penalized Quantile Regression in High-Dimensional Sparse Models," *arXiv:0904.2931 [math, stat]*, Apr. 2009, arXiv: 0904.2931. [Online]. Available: http://arxiv.org/abs/0904.2931

[21] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[22] G. Ciuperca, "Adaptive LASSO model selection in a multiphase quantile regression," *Statistics*, vol. 50, no. 5, pp. 1100–1131, Sep. 2016. [Online]. Available: http://dx.doi.org/10.1080/02331888.2016.1151427

[23] H. Zou and M. Yuan, "Regularized simultaneous model selection in multiple quantiles regression," *Computational Statistics & Data Analysis*, vol. 52, no. 12, pp. 5296–5304, Aug. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947308002739

[24] L. Jiang, H. D. Bondell, and H. J. Wang, "Interquantile shrinkage and variable selection in quantile regression," *Computational Statistics & Data Analysis*, vol. 69, pp. 208–219, Jan. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947313002922

[25] R. Koenker, *Quantile regression*. Cambridge university press, 2005, no. 38.

[26] M. Efroymson, "Multiple regression analysis," *Mathematical methods for digital computers*, vol. 1, pp. 191–203, 1960.

[27] R. R. Hocking and R. N. Leslie, "Selection of the Best Subset in Regression Analysis," *Technometrics*, vol. 9, no. 4, pp. 531–540, Nov. 1967. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00401706.1967.10490502

[28] Y. Li and J. Zhu, "L1-Norm Quantile Regression," *Journal of Computational and Graphical Statistics*, vol. 17, no. 1, pp. 163–185, Mar. 2008. [Online]. Available: http://dx.doi.org/10.1198/106186008X289155

[29] C. Bergmeir, R. J. Hyndman, and B. Koo, "A Note on the Validity of Cross-Validation for Evaluating Time Series Prediction," Monash University, Department of Econometrics and Business Statistics, Tech. Rep. 10/15, 2017. [Online]. Available: https://ideas.repec.org/p/msh/ebswps/2015-10.html

[30] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192–213, May 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025511006773

[31] J. A. Machado, "Robust model selection and m-estimation," *Econometric Theory*, vol. 9, pp. 478–493, 1993.