



**XXIV SNPTEE**  
**SEMINÁRIO NACIONAL DE PRODUÇÃO E**  
**TRANSMISSÃO DE ENERGIA ELÉTRICA**

CB/GOP/09

22 a 25 de outubro de 2017  
Curitiba - PR

## **GRUPO - IX**

### **GRUPO DE ESTUDO DE OPERAÇÃO DE SISTEMAS ELÉTRICOS - GOP**

#### **PROPOSTAS PARA A PREVISÃO PROBABILÍSTICA DA GERAÇÃO EÓLICA**

**José Francisco Moreira Pessanha(\*)**  
**CEPEL**

**Ricardo Bessa**  
**INESC TEC**

**João Abel Peças Lopes**  
**FEUP & INESC TEC**

## **RESUMO**

A perspectiva de uma maior participação das fontes eólicas no Sistema Interligado Nacional aponta para a necessidade de incluir a previsão de curto prazo da geração eólica nos procedimentos da operação em tempo real e da programação diária da operação. No presente trabalho apresentam-se duas abordagens para previsão probabilística da geração eólica: regressão quantílica e uma metodologia híbrida que congrega processos gaussianos, estimador de densidade de probabilidade e simulação de Monte Carlo. Uma avaliação comparativa das duas abordagens propostas foi conduzida por meio de um experimento computacional com dados reais de um parque eólico localizado na Galícia, Espanha.

## **PALAVRAS-CHAVE**

Energia eólica, Estimção de densidade, Previsão probabilística, Processo gaussiano, Regressão quantílica

### **1.0 - INTRODUÇÃO**

O comportamento intermitente da velocidade do vento impõe desafios para a integração segura e econômica dos aproveitamentos eólicos aos sistemas de potência. Para viabilizar a integração da geração eólica os centros de operação demandam novas ferramentas de gerenciamento de energia elétrica, entre as quais se destaca a previsão da geração eólica [1]. Em geral, os métodos de previsão privilegiam a predição do valor esperado da geração eólica sem, contudo, quantificar e informar apropriadamente a incerteza associada à previsão. Em virtude do comportamento aleatório e intermitente do vento, a previsão do valor esperado da geração eólica sem uma quantificação adequada da incerteza não atende completamente as necessidades da operação. Por exemplo, em sistemas com grande participação de fontes eólicas a reserva operativa deve ser dimensionada de maneira a cobrir as incertezas nas previsões de carga e de geração eólica [2]. A densidade de probabilidade contém toda a informação de uma variável aleatória, portanto, para lidar com a incerteza deve-se considerar as previsões de densidade de probabilidade da geração eólica fornecidas por métodos de previsão probabilística [3].

O presente trabalho tem por objetivo descrever e comparar duas propostas para previsão probabilística da geração eólica: a regressão quantílica [4] e uma metodologia híbrida que congrega processos gaussianos [5], estimador de densidade de probabilidade [6] e simulação de Monte Carlo [7]. A implementação e análise comparativa das duas abordagens foi conduzida por meio de um programa em linguagem R [8] aplicado em um experimento computacional que contou com dados reais provenientes de um parque eólico localizado na Galícia, Espanha. A seguir, na seção 2 tem-se uma descrição da metodologia híbrida e na seção 3 uma breve introdução à regressão quantílica. Na sequência, na seção 4 apresentam-se algumas métricas para avaliação da qualidade de previsões probabilísticas. Os principais resultados são descritos na seção 5, em especial, as previsões probabilísticas das médias horárias da geração eólica até 48 horas à frente, ao longo de uma semana. Por fim, na seção 6 tem-se as conclusões do trabalho.

(\*) Av. Horácio Macedo, n° 354 – Cidade Universitária, Ilha do Fundão, Centro de Pesquisas de Energia Elétrica - Cepel, sala 218, CEP 21941-911, Rio de Janeiro, RJ, – Brasil Tel: (+55 21) 2598-6066 – Email: francisc@cepel.br

## 2.0 – METODOLOGIA HÍBRIDA

A velocidade do vento e a geração eólica guardam entre si uma relação não linear, conforme ilustrado pela curva de potência na Figura 1 - (a). Assim, uma forma simples de calcular a previsão da geração eólica consiste em obter uma previsão  $VO$  da velocidade do vento e na sequência determinar a geração correspondente na curva de potência. Contudo, qualquer previsão pontual  $VO$  vem acompanhada de incertezas [3], representadas pela densidade de probabilidade  $f(v)$ . Adicionalmente, para uma dada velocidade  $VO$  a geração eólica  $g$  é uma variável aleatória com densidade de probabilidade condicional  $f(g|v=VO)$ .

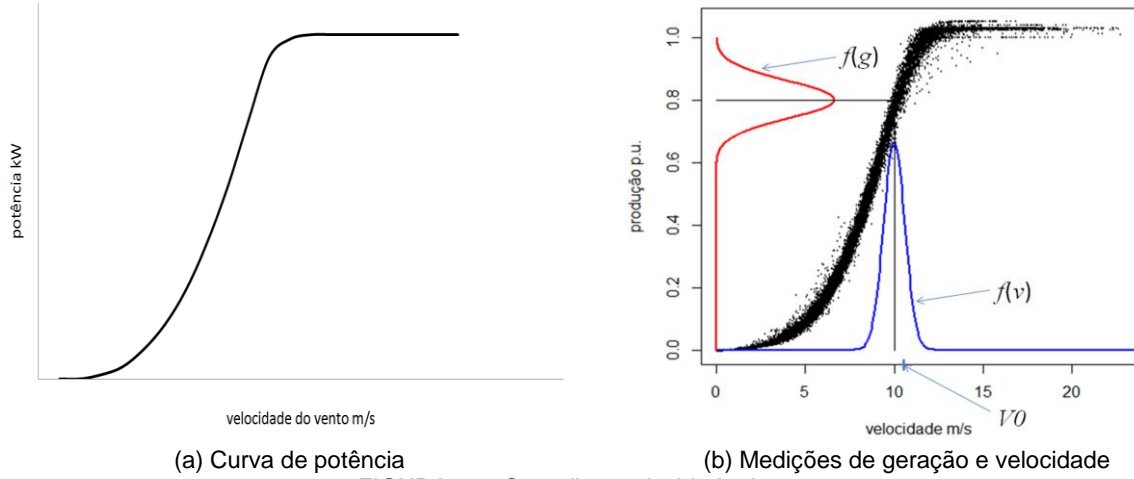


FIGURA 1 – Geração e velocidade do vento

A partir de uma previsão da densidade de probabilidade da velocidade do vento  $f(v)$  pode-se aplicar a simulação de Monte Carlo para gerar  $n$  amostras independentes  $v_1, \dots, v_n$  e na sequência buscar as densidades condicionais  $f(g|v_i)$   $\forall i=1, n$  correspondentes. A média de todas as densidades condicionais produz a previsão da densidade de probabilidade marginal da geração eólica  $f(g)$ , conforme ilustrado na Figura 1 - (b). As considerações acima descrevem as linhas gerais de uma metodologia de previsão probabilística da geração eólica [7] constituída de três etapas: 1) previsão da densidade de probabilidade da velocidade do vento  $f(v)$  em cada hora do horizonte de previsão por meio de processos gaussianos, 2) estimação da densidade de probabilidade da geração eólica condicionada à velocidade do vento  $f(g|v_i)$  e 3) cálculo da densidade de probabilidade marginal da geração eólica  $f(g)$  por simulação de Monte Carlo. Denotando por  $V$  a velocidade do vento e por  $\theta$  o ângulo que define a direção do vento, a metodologia proposta para a previsão probabilística da geração eólica parte das previsões pontuais das componentes cartesianas do vento ( $u^*$  e  $w^*$ ), calculadas a partir das médias horárias previstas  $V^*$  e  $\theta^*$  para cada hora  $h$ , fornecidas por um modelo de previsão climática (*Numerical Weather Prediction – NWP*):

$$u_h^* = V_h^* \cdot \cos(\theta_h^*) \quad (1)$$

$$w_h^* = V_h^* \cdot \sin(\theta_h^*) \quad (2)$$

Com base nos valores verificados de  $V$  e  $\theta$ , acompanhados das respectivas previsões geradas pelo modelo NWP, pode-se ajustar um processo gaussiano para cada componente cartesiana da velocidade [5]. Um processo gaussiano é uma espécie de modelo de regressão, em que as variáveis respostas correspondem aos valores verificados das componentes cartesianas  $u$  e  $w$  da velocidade do vento, enquanto as variáveis explicativas são os respectivos valores previstos  $u^*$  e  $w^*$  fornecidos pelo modelo NWP. A ideia central é que o processo gaussiano corrija o viés na previsão da velocidade do vento [14]. O processo gaussiano fornece as previsões das densidades gaussianas das componentes cartesianas da velocidade do vento em cada hora  $h$  do horizonte de previsão, cujos parâmetros previstos são as médias  $\mu_u(h)$  e  $\mu_w(h)$  e os desvios padrão  $\sigma_u(h)$  e  $\sigma_w(h)$ , conforme a seguir em (3) e (4):

$$f_u(h) = N(\mu_u(h), \sigma_u(h)) \quad (3)$$

$$f_w(h) = N(\mu_w(h), \sigma_w(h)) \quad (4)$$

Na sequência, estima-se a densidade da geração eólica condicionada  $f(g|u, w)$ , a partir de uma amostra de valores verificados da geração eólica, acompanhados dos respectivos valores verificados das componentes cartesianas  $u$  e  $w$ . A estimação de  $f(g|u, w)$  pode ser realizada por meio do *quantile-copula estimator* [6]. Assim, para cada hora  $h$  do horizonte de previsão dispõe-se das densidades gaussianas das componentes cartesianas do vento  $f_u(h)$  e  $f_w(h)$  e da densidade condicional  $f(g|u, w)$ . A previsão da densidade marginal da geração eólica  $f_g(h)$  para a hora  $h$  do horizonte de previsão é obtida pela seguinte integral, resolvida por simulação de Monte Carlo:

$$f_g(h) = \iint f(g|u, w) \cdot f_u(h) \cdot f_w(h) \cdot dudw \quad (5)$$

### 2.1 Processo gaussiano

Em uma amostra com  $n$  pares  $(y_i, x_i) \forall i=1, n$ , no qual  $y_i$  denota a variável resposta e  $x_i$  um vetor com variáveis explicativas, a relação entre  $y_i$  e  $x_i$  pode ser representada pela equação  $y_i = f(x_i) + \varepsilon_i \quad \forall i=1, n$  em que  $\varepsilon$  denota um

ruido gaussiano com média nula e desvio padrão constante  $\sigma_\varepsilon$ . Assim, a coleção  $y_1, \dots, y_n$  tem distribuição conjunta normal multivariada  $N_n(0, \Sigma)$ . A média nula na distribuição conjunta pode ser obtida após os dados serem apropriadamente escalonados ou diferenciados. Já os elementos da matriz de covariância  $\Sigma$  podem ser parametrizados por uma função núcleo (*kernel function*), por exemplo, a covariância entre as respostas  $y_p$  e  $y_q$  pode ser especificada como uma função dos respectivos vetores de variáveis de entrada  $x(p)$  e  $x(q)$  e do vetor de hiperparâmetros  $\theta = (a_1, \dots, a_d, \nu_1, \omega_1, \dots, \omega_D, \sigma_\varepsilon^2)$ .

$$\Sigma_{p,q} = K(y_p, y_q) = \nu_1 \cdot \exp\left(-\frac{1}{2} \sum_{d=1}^D \omega_d (x_d(p) - x_d(q))^2\right) + \sum_{d=1}^D a_d x_d(p) x_d(q) + \sigma_\varepsilon^2 \delta(p,q) \quad \forall p, q \in \{1, \dots, n\} \quad (6)$$

em que  $\delta(p,q)=1$  se  $p=q$ , caso contrário  $\delta(p,q)=0$ .

Para um conjunto com  $n$  observações conhecidas dos vetores de entrada  $X = \{x(1), \dots, x(n)\}$  acompanhadas das respectivas respostas  $y = \{y(1), \dots, y(n)\}$  e admitindo a priori gaussiana multivariada para  $y$  com elementos da matriz de covariância especificados em (6), deseja-se prever  $y(n+1)$  correspondente ao vetor de entrada  $x(n+1)$ . As considerações acima implicam que a distribuição conjunta de  $y = \{y(1), \dots, y(n)\}$  e  $y(n+1)$  é normal multivariada:

$$\begin{bmatrix} y \\ y(n+1) \end{bmatrix} \sim N_{n+1} \left( \begin{bmatrix} 0_n \\ 0 \end{bmatrix}, \begin{bmatrix} K_{n \times n} & k_{n \times 1} \\ k_{n \times 1}^T & k^* \end{bmatrix} \right) \quad (7)$$

Em (7),  $K_{n \times n}$  denota a matriz de covariâncias entre as  $n$  observações conhecidas e com elementos parametrizados conforme em (6). Ainda em (7), o vetor  $k_{n \times 1}$  contém as covariâncias entre as respostas das  $n$  observações conhecidas e a resposta em  $n+1$ , ou seja,  $k = [K(x(1), x(n+1)), \dots, K(x(n), x(n+1))]$ . Já  $k^*$  denota a variância da nova resposta  $k^* = K(x(n+1), x(n+1))$ . Por meio da distribuição condicional da normal multivariada [9] pode-se obter a previsão da densidade gaussiana de  $y(n+1)$ , condicionada aos valores conhecidos da variável resposta  $y$ .

$$f(y(n+1) | y) \sim N(\mu_{n+1}, \sigma_{n+1}^2) \quad (8)$$

em que  $\mu_{n+1} = k_{n \times 1}^T \cdot K_{n \times n}^{-1} \cdot y_{n \times 1}$  e  $\sigma_{n+1}^2 = k^* - k_{n \times 1}^T \cdot K_{n \times n}^{-1} \cdot k_{n \times 1}$ .

A previsão pontual para  $n+1$  é fornecida por  $\mu_{n+1}$  e a incerteza da previsão é quantificada por  $\sigma_{n+1}$ . Estes resultados dependem do vetor de hiperparâmetros  $\theta = (a_1, \dots, a_d, \nu_1, \omega_1, \dots, \omega_D, \sigma_\varepsilon^2)$ , cujos elementos são determinados pela maximização do logaritmo da verossimilhança [10] indicado a seguir:

$$\log L(\theta) = -\frac{1}{2} \log(|K|) - \frac{1}{2} y^T K^{-1} y - \frac{n}{2} \log(2\pi) \quad (9)$$

em que a matriz  $K$  corresponde à matriz  $K_{n \times n}$  em (7) e envolve apenas as  $n$  observações da amostra.

## 2.2 Quantile-copula estimator

A função densidade estimada empiricamente a partir de uma amostra aleatória com  $n$  observações  $X_1, \dots, X_n$ , consiste na soma de funções núcleos (*kernel function*)  $K(x)$  centradas em cada observação amostral, em que uma largura de banda  $h_x$  definida a priori controla a dispersão do *kernel*:

$$\hat{f}(x) = \frac{1}{n \cdot h_x} \sum_{i=1}^n K\left(\frac{x - X_i}{h_x}\right) \quad (10)$$

O *kernel*  $K(x)$  é uma função que pode assumir diferentes formas, mas sua integral deve ser unitária. Tipicamente, a função *kernel* é a densidade gaussiana padrão. A mesma idéia pode ser empregada na estimação da densidade conjunta entre  $p$  variáveis aleatórias  $x_1, \dots, x_p$ , para as quais dispõe-se de uma amostra aleatória pareada com  $n$  observações nas  $p$  variáveis  $(X_{11}, \dots, X_{1p}), \dots, (X_{n1}, \dots, X_{np})$ :

$$\hat{f}(x_1, \dots, x_p) = \frac{1}{n \cdot h_1 \cdot \dots \cdot h_p} \sum_{i=1}^n \prod_{j=1}^p K_j\left(\frac{x_{ij} - X_{ij}}{h_j}\right) \quad (11)$$

A partir do Teorema da Probabilidade Condicional [6] a densidade de probabilidade de uma variável aleatória  $y$  condicionada ao valor conhecido de uma variável  $x$  é estimada da seguinte maneira:

$$\hat{f}(y|x) = \frac{\sum_{i=1}^n K_x\left(\frac{x - X_i}{h_x}\right) K_y\left(\frac{y - Y_i}{h_y}\right)}{\sum_{i=1}^n K_x\left(\frac{x - X_i}{h_x}\right)} \quad (12)$$

em que  $h_x$  e  $h_y$  controlam os graus de suavização dos *kernels*  $K_x$  e  $K_y$  respectivamente.

Uma alternativa mais eficiente ao estimador em (12) é o *quantile-copula estimator* [6] em que  $F_x(X_i)$  e  $F_y(Y_i)$  denotam os dados  $X_i$  e  $Y_i$  transformados pelas respectivas distribuições empíricas de probabilidades acumuladas:

$$\hat{f}(y|x) = \frac{1}{n \cdot h_y} \cdot \left[ \sum_{i=1}^n K_y \left( \frac{y - Y_i}{h_y} \right) \right] \cdot \left[ \frac{1}{n \cdot h_x \cdot h_y} \sum_{i=1}^n K_x \left( \frac{F_X(x) - F_X(X_i)}{h_x} \right) K_y \left( \frac{F_Y(y) - F_Y(Y_i)}{h_y} \right) \right] \quad (13)$$

O estimador em (13) pode ser adaptado para acomodar as componentes cartesianas da velocidade do vento ( $u$  e  $w$ ) e permitir a estimação empírica da densidade condicional  $f(g|u,w)$ :

$$\hat{f}(g|u,w) = \frac{1}{n \cdot h_g} \cdot \left[ \sum_{i=1}^n K_g \left( \frac{g - G_i}{h_g} \right) \right] \cdot \left[ \frac{1}{n \cdot h_u \cdot h_w} \sum_{i=1}^n K_u \left( \frac{F_u(u) - F_u(U_i)}{h_u} \right) K_w \left( \frac{F_w(w) - F_w(W_i)}{h_w} \right) K_g \left( \frac{F_g(g) - F_g(G_i)}{h_g} \right) \right] \quad (14)$$

### 3.0 - REGRESSÃO QUANTÍLICA

Em um modelo de regressão linear a equação de regressão estimada fornece uma estimativa do valor esperado da variável resposta  $y$  em função da variável explicativa  $X$ . Já em um modelo de regressão quantílica, a equação de regressão fornece uma estimativa do quantil  $\tau$  (por exemplo,  $\tau=0,5$  para a mediana) da variável resposta em função da variável explicativa. Assim, o conjunto de equações de regressão quantílica estimadas para diferentes valores de  $\tau$  permite caracterizar a distribuição de probabilidade da variável resposta condicionada aos valores das variáveis explicativas. Em [4] apresenta-se a especificação de um modelo de regressão quantílica em que o quantil da geração eólica depende das coordenadas cartesianas ( $u,w$ ) da velocidade do vento. Os quantis estimados pelo conjunto de equações de regressão fornecem uma estimativa da densidade condicionada  $f(g|u,w)$ . Para capturar as não linearidades nas relações entre a geração eólica e as componentes  $u$  e  $w$ , os autores em [4] propõem usar como variáveis explicativas os coeficientes das previsões de  $u$  e  $w$  em dez bases *B-Splines* [11] e, portanto, o modelo de regressão tem 20 variáveis explicativas. A especificação do modelo resultante corresponde ao modelo aditivo generalizado [11] indicado a seguir:

$$\hat{Q}_\tau(\tau) = \beta_0(\tau) + \sum_{j=1}^{10} b_j(u_i^*) \beta_j^u(\tau) + \sum_{j=1}^{10} b_j(w_i^*) \beta_j^w(\tau) + \varepsilon_i \quad (15)$$

em que  $\varepsilon_i$  denota um erro aleatório,  $\beta^u(\tau)$  e  $\beta^w(\tau)$  denotam os coeficientes da regressão quantílica para o quantil  $\tau$  e  $b(u^*)$  e  $b(w^*)$  denotam os coeficientes das previsões das coordenadas  $u$  e  $w$  nas bases *B-Splines*.

De forma distinta do estimador mínimos quadrados, na regressão quantílica para um quantil  $\tau$  o estimador dos coeficientes de regressão busca minimizar a seguinte função objetivo em uma amostra com  $n$  pares  $(y_i, X_i) \forall i=1, n$ :

$$\text{Min}_\beta \sum_{i=1}^n \rho_\tau(y_i - X_i^T \beta) \quad (16)$$

em que  $\rho_\tau(z) = z \cdot [\tau - I(z < 0)]$  e  $I(z < 0)$  é uma função indicadora que assume valor unitário para  $z < 0$ , caso contrário, é igual a zero.

### 4.0 – MÉTRICAS DE DESEMPENHO

A média e a mediana de uma previsão probabilística fornecem previsões pontuais que se comparadas aos valores realizados permitem avaliar o erro e as métricas convencionais de desempenho como os erros absolutos e relativos. Contudo, em uma previsão probabilística a avaliação da qualidade das previsões não deve se restringir apenas às avaliações de performance pontual, mas deve ser ampliada e informar o grau de aderência das previsões das densidades de probabilidade aos valores verificados. Duas métricas se destacam na avaliação da qualidade de previsões probabilísticas: a calibração e a nitidez. A calibração é uma medida da aderência entre as probabilidades nominais associadas aos quantis previstos e às probabilidades empíricas extraídas das observações. Por exemplo, em uma longa série de previsões probabilísticas, cerca de metade dos valores da série deve situar-se abaixo das respectivas medianas previstas. Já a nitidez refere-se ao tamanho médio dos intervalos de previsão e intervalos mais estreitos significam maior qualidade. O *Continuous Ranked Probability Score* (CRPS) [12] é uma métrica de qualidade que combina a calibração e a nitidez em uma única estatística de desempenho. Dada uma previsão da densidade de probabilidade acumulada  $F_{t+h|t}$  e conhecido o valor da variável resposta  $y_{t+h}$ , ambos no horizonte  $h$ , o respectivo CRPS é definido conforme a seguir:

$$\text{CRPS}(F_{t+h|t}, y_{t+h}) = \int_{-\infty}^{\infty} [F_{t,h}(z) - I(z \geq y_{t+h})]^2 dz = E(|Y - y_{t+h}|) - \frac{1}{2} E(|Y - Y^*|) \quad (17)$$

Em (17) a função indicadora  $I(z \geq y_{t+h})$  assume valor unitário se  $z \geq y_{t+h}$ , caso contrário, assume valor nulo. Ainda em (17),  $Y$  e  $Y^*$  são duas amostras aleatórias independentes extraídas da densidade acumulada prevista  $F_{t+h|t}$ . O primeiro valor esperado em (17) expressa a calibração e o segundo a nitidez.

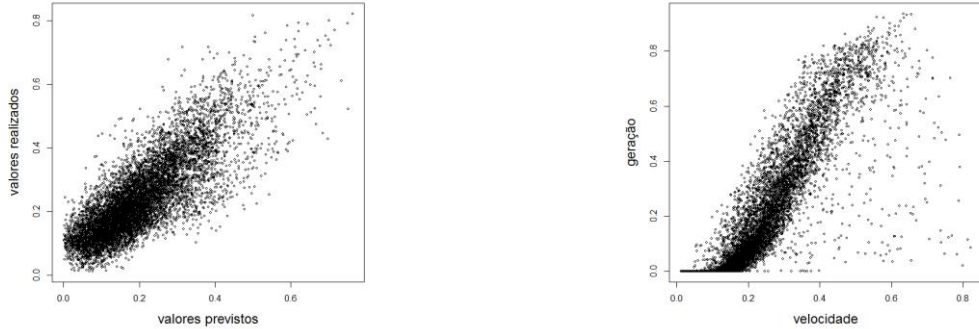
Dado um conjunto de teste de tamanho  $N$  e amostras  $y_{t+h}^{(i)}$  e  $y_{t+h}^{*(i)} \forall i=1, M$  extraídas da densidade prevista  $F_{t+h|t}$ , o CRPS para um horizonte de previsão  $h$  [13] é calculado por meio da seguinte expressão, cujos termos quantificam a calibração e a nitidez respectivamente:

$$CRPS(h) = \frac{1}{NM} \sum_{t=1}^N \sum_{i=1}^M |y_{t+h}^{(i)} - y_{t+h}| - \frac{1}{2NM} \sum_{t=1}^N \sum_{i=1}^M |y_{t+h}^{(i)} - y_{t+h}^{*(i)}| \quad (18)$$

As melhores previsões apresentam desvios pequenos entre as medianas previstas e observadas (calibração) e intervalos de previsão estreitos (nitidez), logo os melhores desempenhos correspondem à menores CRPS.

## 5.0 - EXPERIMENTO COMPUTACIONAL

Para ilustrar a aplicação das metodologias de previsão probabilística considerou-se o parque eólico Sotavento, localizado na Galícia, Espanha. O parque tem 24 aerogeradores de 5 tecnologias e 9 fabricantes diferentes e que totalizam uma capacidade instalada de 17,56 MW. A velocidade média do vento é de 6,41 m/s e a produção média anual é da ordem de 33 MWh. A metodologia proposta foi implementada em ambiente R [8] e inicialmente todos os dados de velocidade e geração foram normalizados para o intervalo [0,1], conforme ilustrado na Figura 2.



(a) Velocidade verificada x Velocidade prevista  
correlação 0,81

(b) Geração x velocidade verificada  
correlação 0,86

FIGURA 2 – Velocidades previstas e valores verificados da velocidade e da geração eólica

A partir das previsões pontuais da velocidade e da direção do vento fornecidas por um modelo NWP foram obtidas as componentes cartesianas ( $u$  e  $w$ ) e na sequência ajustados os processos gaussianos de cada componente para agregar densidades gaussianas  $f_u(h)$  e  $f_w(h)$  às previsões oriundas do NWP. Cada processo gaussiano ajusta uma regressão em que as médias horárias das componentes  $u(h)$  e  $w(h)$  são explicadas pelas respectivas médias horárias  $u^*(h)$  e  $w^*(h)$  previstas por um modelo NWP. As correlações entre as componentes cartesianas verificadas e previstas pelo NWP são da ordem de 0,57 e 0,55 para as componentes  $u$  e  $w$  respectivamente. Os hiperparâmetros dos processos gaussianos de ambas as componentes foram atualizados uma vez ao dia ao longo do período de teste (de 10/11/2014 até 16/11/2014), antes da realização das previsões até 48 à frente para os dias  $d+1$  e  $d+2$ . Em virtude do esforço computacional para ajustar os hiperparâmetros considerou-se um histórico constituído pelas médias horárias previstas e verificadas nas 96 horas (4 dias) imediatamente anteriores ao dia da previsão ( $d+1$ ), assim a matriz  $K_{NWP}$  nas expressões (7) e (9) tem dimensões 96 x 96.

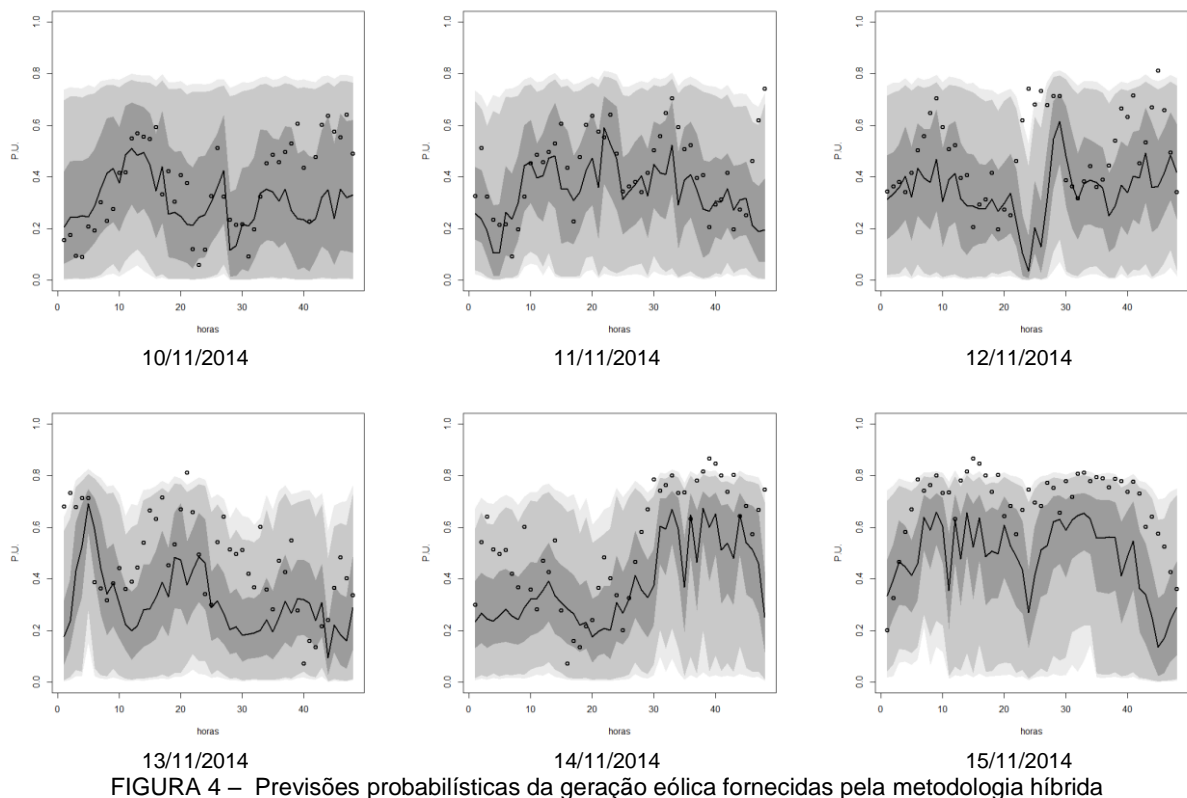
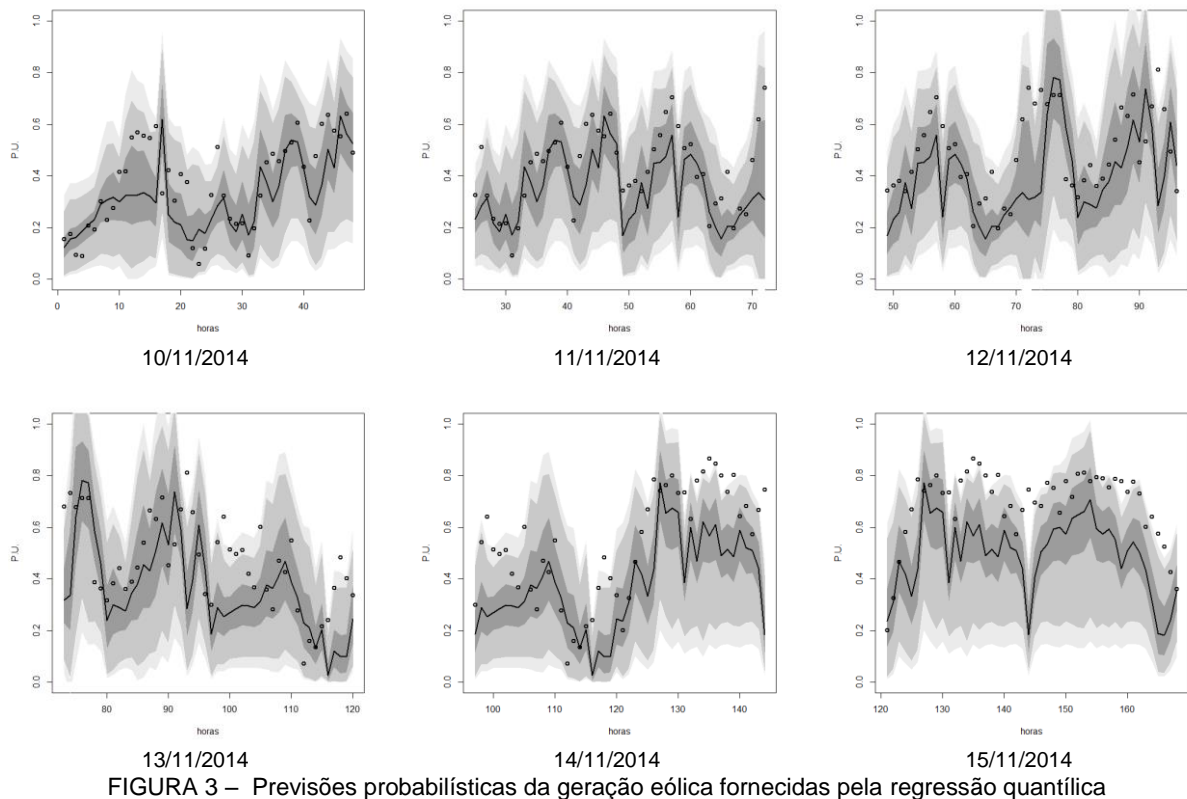
Por meio da simulação de Monte Carlo foram sorteadas 1.000 amostras aleatórias de cada uma das densidades  $f_u(h)$  e  $f_w(h)$  fornecidas pelos processos gaussianos para cada hora do horizonte de previsão. Cada par amostrado de componentes ( $u, w$ ) da velocidade do vento foi introduzido na densidade condicional  $f(g|u, w)$ , previamente estimada pelo estimador *quantile-copula conditional*, de onde cinco valores de geração eólica  $g$  foram amostrados. No *quantile-copula conditional* considerou-se uma largura de banda de 0,01 em todas as variáveis envolvidas e a estimação foi realizada uma única vez, sem atualização diária, a partir dos valores verificados de geração e da velocidade/direção do vento no período de 60 dias imediatamente anteriores ao primeiro dia do período de teste. Portanto, para cada hora  $h$  do horizonte de previsão foram extraídas 5.000 amostras da densidade de probabilidade da geração eólica  $f_g(h)$ , a partir das quais foram calculadas as previsões horárias dos quantis da geração.

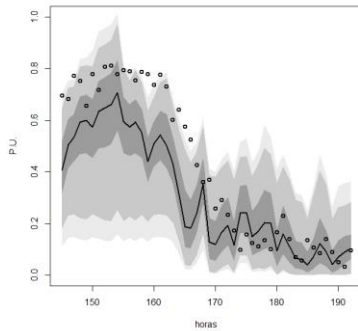
Quanto ao modelo de regressão quantílica foram ajustados separadamente 9 modelos de regressão, um modelo para cada um dos seguintes quantis: 2,5%, 5%, 10%, 25%, 50%, 75%, 90%, 95% e 97,5%. No ajuste dos modelos consideraram-se todas as 7.512 observações horárias do período *insample* (de 1/1/2014 até 9/11/2014). Os coeficientes de regressão foram ajustados uma única vez e não foram atualizados ao longo do período de teste.

A seguir, nas Figuras 3, 4 e 5 são apresentadas as previsões probabilísticas da geração eólica até 48 horas à frente realizadas para o período de teste (10/11/2014 até 16/11/2014). Os pontos denotam as médias horárias verificadas e a linha em preto a mediana prevista. A região em cinza escuro na área central é delimitada pelos intervalos de previsão com probabilidade de 50%. As previsões iniciam-se na primeira hora do dia. Os resultados indicam uma maior dispersão das densidades de probabilidade previstas pela metodologia híbrida.

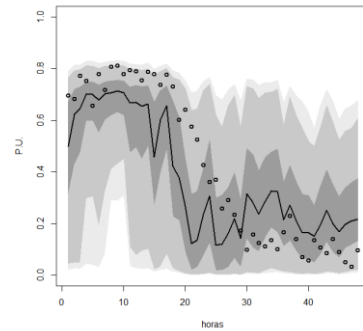
Conforme ilustrado na Figura 6 - (a), em ambos os métodos a calibração do intervalo de previsão aumenta com a proporção nominal do intervalo, sendo que nas proporções mais elevadas os intervalos produzidos pela metodologia

híbrida apresentam maior compatibilidade com os níveis nominais de probabilidade. Contudo a melhor calibração dos intervalos produzidos pela abordagem híbrida foi alcançada por meio de intervalos mais amplos (menor nitidez), conforme ilustrado na Figura 6 - (b). Uma visão agregada da performance nos quesitos calibração e nitidez é fornecida pela estatística CRPS na Figura 7. O CRPS médio alcançado pela regressão quantílica no período de teste foi da ordem de 0,10, menor que a média de 0,19 resultante da metodologia híbrida.



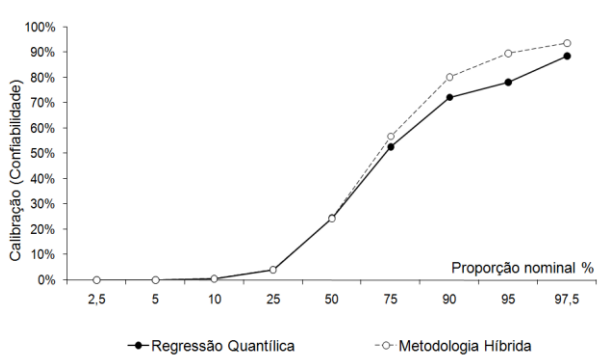


(a) Regressão quantílica

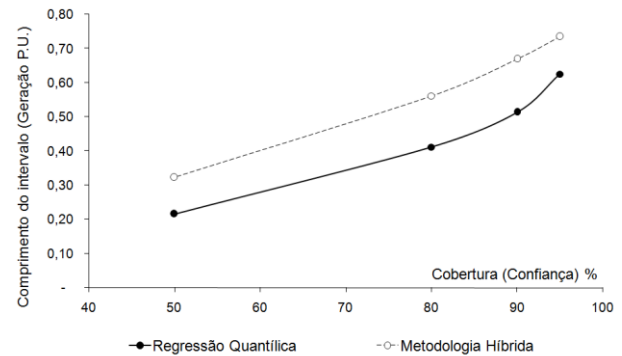


(b) Metodologia híbrida

FIGURA 5 – Previsões probabilísticas da geração eólica para o dia 16/11/2014



(a) Calibração



(b) Nitidez

FIGURA 6 – Calibração e nitidez das previsões probabilísticas

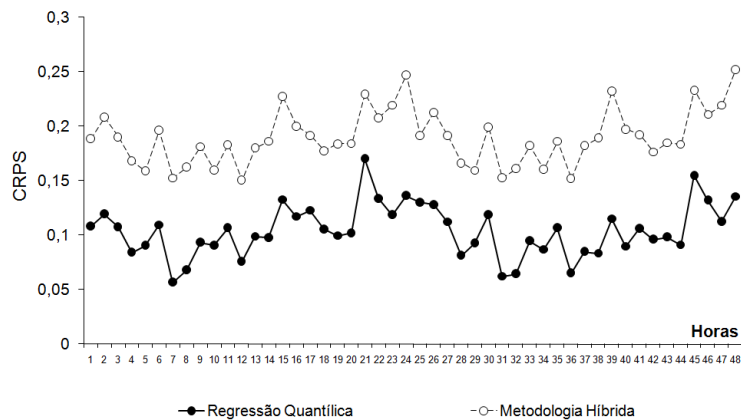


FIGURA 7 – CRPS em função do horizonte de previsão

As previsões das medianas horárias da geração eólica fornecidas pela metodologia híbrida guardam uma correlação de 0,62 com os valores verificados da geração ao longo do período de teste, enquanto nas previsões resultantes da regressão quantílica a correlação foi da ordem de 0,74. Os resultados indicam que a regressão quantílica apresentou um melhor desempenho em termos da maior precisão e nitidez das previsões probabilísticas. porém deve-se ressaltar que a metodologia híbrida oferece a possibilidade de corrigir o viés e modelar a incerteza das previsões de velocidade do vento oriundas dos modelos NWP [14]. Adicionalmente, embora os hiperparâmetros dos processos gaussianos tenham sido atualizados diariamente, a metodologia híbrida utilizou um conjunto de dados menor que o considerado no ajuste dos modelos de regressão quantílica.

## 6.0 - CONCLUSÕES

A perspectiva de uma maior participação da geração eólica no Sistema Interligado Nacional (SIN) aponta para a necessidade da operação ter meios de formular previsões da geração eólica no curto prazo. Neste trabalho apresentam-se duas abordagens possíveis para a previsão probabilística da geração eólica até 48 horas à frente:

regressão quantílica e uma metodologia híbrida que congrega o processo gaussiano, o *quantile-copula estimator* e a simulação de Monte Carlo. Ambas as abordagens foram implementadas tendo em vista as necessidades da programação diária da operação em que as previsões até 48 horas à frente são atualizadas diariamente.

Para avaliar o desempenho das duas abordagens concebeu-se um experimento computacional em ambiente R com dados provenientes de um parque eólico localizado na Galícia, Espanha. Os desempenhos das duas abordagens foram avaliados por meio da calibração e nitidez das previsões probabilísticas para a geração eólica. A regressão quantílica apresentou uma performance ligeiramente superior. Contudo, os resultados alcançados pela metodologia híbrida são satisfatórios e mostram que a metodologia é promissora e mais investigações devem ser realizadas no sentido de diminuir a dispersão das densidades previstas para a geração eólica. As discrepâncias entre os desempenhos das duas abordagens de previsão refletem as diferenças das dimensões dos conjuntos de dados utilizados na implementação de cada metodologia. A metodologia híbrida demanda um maior esforço computacional e por esta razão na sua implementação considerou-se um conjunto de dados com dimensões reduzidas, em contraste com as longas séries de dados utilizadas no ajuste dos modelos de regressão quantílica.

## 7.0 - AGRADECIMENTOS

A realização deste trabalho contou com uma bolsa para realização de pós-doutorado no Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC) na cidade do Porto, Portugal, concedida no âmbito do Programa Ciência sem Fronteiras do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com o apoio da Eletrobras (207480/2015-0 Encomenda Eletrobrás 2015 – PDE).

## 8.0 - REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Estanqueiro, A., Castro, R., Flores, P., Ricardo, J., Pinto, M., Rodrigues, R., Peças Lopes, J. How to prepare a power system for 15% wind energy penetration: the Portuguese Case Study, *Wind Energy*, v. 11, p. 75-84, 2008.
- (2) Matos, M.A., Bessa, R.J. Setting the Operating Reserve Using Probabilistic Wind Power Forecasts, *IEEE Transactions on Power Systems*, v. 26, n. 2, May, 2011.
- (3) Zhang, Y., Wang, J., Wang, X. Review on probabilistic forecasting of wind power generation, *Renewable and Sustainable Energy Reviews*, v. 32, p. 255-370, 2014.
- (4) Nielsen, H.A., Madsen H., Nielsen, T.S. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, v.9, n.1-2, p. 95-108, January-April, 2006.
- (5) Jiang, X., Dong, B., Xie, L., Sweeney, L. Adaptive Gaussian Process for Short-Term Wind Speed Forecasting, 19th European Conference on Artificial Intelligence, Lisbon, 2010.
- (6) Bessa, R.J., Miranda, V., Botterud, A., Zhou, Z., Wang, J. Time-adaptive quantile-copula for wind power probabilistic forecasting, *Renewable Energy*, v. 40, n. 1, p. 29-39, April, 2012.
- (7) Jeon, J., Taylor, J.W. Using Conditional Kernel Density Estimation for Wind Power Density Forecasting, *Journal of the American Statistical Association*, v.107, p. 66-79, 2012.
- (8) R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>, 2014.
- (9) Johnson, R.A., Wichern, D.W. Applied Multivariate Statistical Analysis, 5th edition, Prentice Hall, 2002.
- (10) Rasmussen, C.E., Williams, C.K.L. Gaussian Processes for Machine Learning, the MIT Press, <http://www.gaussianprocess.org/gpml/chapters/>, 2006.
- (11) Wood, S.N. Generalized Additive Models: an introduction with R, Chapman & Hall, 2006.
- (12) Gneiting, T., Raftery, A.E. Strictly proper scoring rules, prediction and estimation, *Journal of the American Statistical Association*, v. 102, n. 477, p. 359-378, March, 2007.
- (13) Taieb, S.B., Huser, R., Hyndman, R.J., Genton, M.G. Probabilistic time series forecasting with boosted additive models: an application to smart meter data, Monash University, Department of Econometrics and Business Statistics, Working Paper 12/15, June, 2015.
- (14) Sweeney, C.P., Lynch, P., Nolan, P. Reducing errors of wind speed forecasts by an optimal combination of post-processing methods, *Meteorological Applications*, v. 20, n. 1, p. 32-40, September, 2011.



## 9.0 - DADOS BIOGRÁFICOS



José Francisco M. Pessanha é bacharel em Estatística pela Escola Nacional de Ciências Estatísticas (Ence, 1992) e engenheiro eletricitista pela Universidade do Estado do Rio de Janeiro (UERJ, 1994). Mestre pela Universidade Federal do Rio de Janeiro (UFRJ, 1999) e Doutor pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio, 2006), ambos em engenharia elétrica, Dr. Pessanha é pesquisador do Centro de Pesquisas de Energia Elétrica (Eletrobras Cepel) e professor adjunto do Instituto de Matemática e Estatística da Uerj. Em 2016 realizou pós-doutorado no Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (Inesc Tec).

Ricardo Bessa é Engenheiro Eletricista e de Computadores pela Faculdade de Engenharia da Universidade do Porto (Feup, 2006), é Mestre em Sistemas de Apoio a Decisão e Análise de Dados pela Faculdade de Economia da Universidade do Porto (Fep, 2008) e Doutor em Sistemas de Energias Sustentáveis pela Feup em 2013. Dr. Bessa é pesquisador do Inesc Tec, onde tem atuado em diversos projetos voltados à implementação e aprimoramento da previsão da geração eólica.

João Abel Peças Lopes é Engenheiro Eletricista e de Computação pela Faculdade de Engenharia da Universidade do Porto (Feup, 1981) e Doutor em Engenharia Elétrica e de Computadores pela Feup em 1988. Dr. Peças Lopes é professor catedrático da Feup e pesquisador do Inesc Tec, onde é membro do Conselho de Administração. Dr. Peças Lopes coordenou um grande número de estudos de conexão à rede elétrica de parques eólicos localizados no Brasil, Hungria, República do Cabo Verde e Portugal.