

ÍNDICE

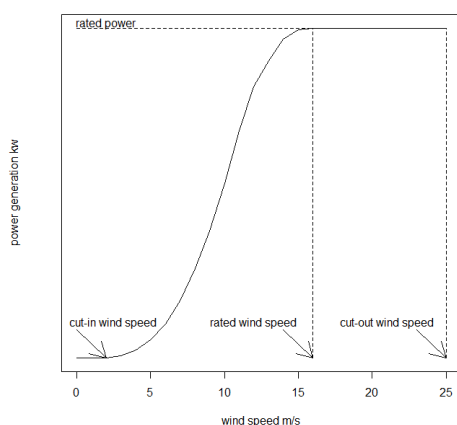
Índice	2
1. INTRODUÇÃO	3
1.1 Considerações gerais.....	3
1.2 Objetivo.....	5
1.3 Organização do texto.....	5
2. MÉTODOS DE PREVISÃO PROBABILÍSTICA	6
2.1 Abordagem com técnicas de machine learning	6
2.1.1 Estágio 1 - Previsão probabilística das componentes cartesianas da velocidade	6
2.1.2 Estágio 2 - Estimacão da densidade da produçãõ condicionada à velocidade	7
2.1.3 Estágio 3 - Previsão probabilística da produçãõ eólica	7
2.1.4 Métodos empregados na abordagem via <i>machine learning</i>	8
2.1.4.1 Estimacão não paramétrica de densidades de probabilidade... ..	8
2.1.4.2 Processo gaussiano	11
2.2 Regressão quantílica	12
3. EXECUÇÃO DO PROGRAMA	14
3.1 Dados de geração e velocidade do vento	14
3.2 Dados gerais.....	15
3.3 Execução	15
3.4 Relatório de saída.....	17
Referências bibliográficas	19

1. INTRODUÇÃO

1.1 Considerações gerais

A perspectiva de uma maior participação das fontes eólicas no SIN aponta para a necessidade de desenvolvimento de modelos capazes de prever a produção eólica em horizontes de curto prazo (até 48 horas à frente em bases horárias). Trata-se de uma informação fundamental para os modelos de apoio à decisão que suportam a operação em tempo real e a programação diária da operação do Sistema Interligado Nacional.

A relação entre a velocidade do vento v e a produção de energia P é definida pela curva de potência $P(v)$, cuja forma típica é ilustrada na Figura 1. Conforme ilustrado na Figura 4, a produção de energia inicia-se em velocidades de vento entre 2 e 3 m/s (*cut-in wind speed*). Então, a produção de energia cresce rapidamente com o aumento da velocidade do vento até encontrar a potência nominal (*rated power*) quando a velocidade alcança um valor entre 12 e 17 m/s (*rated wind speed*). Para velocidades acima de 25 m/s (*cut-out wind speed*) a turbina eólica deve ser desligada para evitar riscos de dano ao rotor.



$$P(v) = \begin{cases} 0, & v \leq V_{in} \\ 0.5C_p \cdot \eta_m \cdot \eta_g \cdot \rho \cdot A \cdot v^3, & V_{in} < v \leq V_r \\ 0.5C_p \cdot \eta_m \cdot \eta_g \cdot \rho \cdot A \cdot V_r^3, & V_r < v \leq V_{out} \\ 0, & v > V_{out} \end{cases}$$

V_{in} = cut-in wind speed (m/s)

V_r = rated wind speed (m/s)

V_{out} = cut-out wind speed (m/s)

C_p = coeficiente de performance

η_m = eficiência do gerador

η_g = eficiência da transmissão

ρ = densidade do ar (kg/m^3)

A = área de varredura do rotor (m^2)

Figura 1. Curva de potência

A curva de potência é obtida a partir de medições de velocidade do vento e da produção eólica tomadas de aerogeradores ensaiados em túnel de vento ou em operação. A seguir, na Figura 2, são ilustradas as medições de velocidade do vento e geração em um ensaio em túnel de vento de uma máquina de pequeno porte de eixo horizontal com diâmetro nominal de 1,64 m, acompanhadas da curva de potência resultante em que cada valor corresponde à produção média em cada faixa de velocidade (*binned values*).

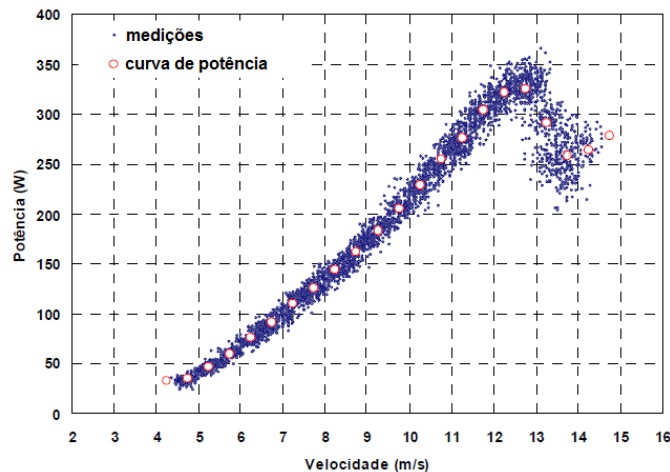


Figura 2. Curva de potência e medições de velocidade e geração [1]

Uma forma simples de gerar a previsão da produção eólica consiste em prever a velocidade do vento e buscar na curva de potência (Figuras 1) a produção correspondente. Portanto, uma previsão em dois estágios: no primeiro estágio ajusta-se um modelo estatístico (ou baseado em técnicas de *machine learning* ou mesmo híbrido) ou utiliza-se um modelo de previsão climática para prever a velocidade do vento incidente nos aerogeradores, e em seguida, por meio da curva de potência, obtém-se a previsão da produção de energia.

Contudo, conforme ilustrado na Figura 2, para um mesmo nível de velocidade do vento a produção eólica pode apresentar diferentes valores. Portanto, a relação entre a velocidade do vento e a produção eólica é estocástica e não determinística. A velocidade do vento e a produção eólica são duas variáveis aleatórias que guardam entre si uma associação não linear.

Assim, obter apenas uma previsão pontual da produção eólica sem conhecer a incerteza subjacente à previsão pode não ser suficiente para as necessidades da operação do sistema elétrico. Por exemplo, em sistemas com grande participação de fontes eólicas a reserva operativa deve ser dimensionada de maneira a cobrir as incertezas na previsão de carga e na previsão da produção eólica. Portanto, há a necessidade de conhecer a incerteza da produção eólica no horizonte de previsão. A incerteza da produção eólica pode ser percebida por meio da previsão da distribuição de probabilidade da produção eólica (previsão probabilística).

Uma previsão pontual V_0 para a velocidade do vento é apenas uma das possíveis realizações de uma variável aleatória v com densidade de probabilidade $f(v)$ que representa a incerteza da previsão anemométrica, conforme ilustrado na Figura 3. Adicionalmente, para uma dada velocidade V_0 a geração da turbina eólica pode assumir valores diferentes, i.e., a geração eólica g , correspondente à V_0 , é uma variável aleatória com densidade de probabilidade condicional $f(g|v=V_0)$.

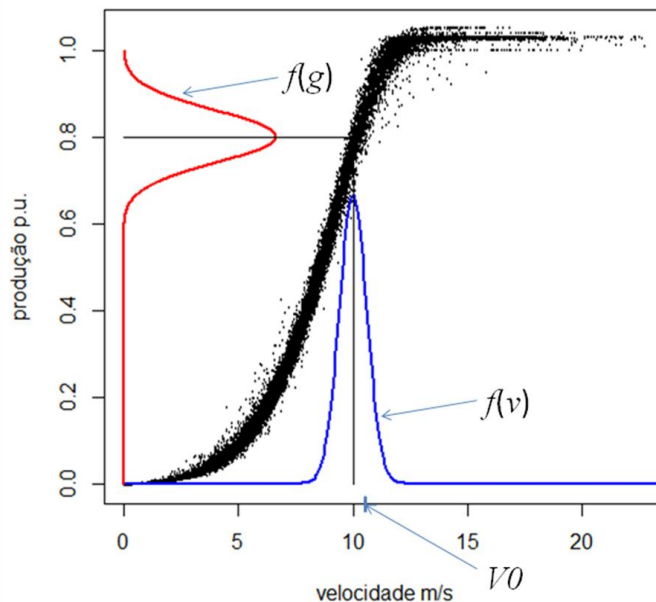


Figura 3. Densidades marginais $f(v)$ e $f(g)$

A partir de uma previsão da densidade de probabilidade da velocidade do vento $f(v)$ pode-se utilizar a simulação de Monte Carlo para gerar n amostras independentes v_1, \dots, v_n e na sequência buscar a densidade $f(g|v_i) \forall i=1, n$. A média de todas as densidades condicionais $f(g|v)$ produz a previsão da densidade de probabilidade marginal da geração eólica $f(g)$ indicada na Figura 3. As considerações acima descrevem as linhas gerais da metodologia para previsão probabilística da geração eólica.

1.2 Objetivo

Neste relatório descrevem-se as funcionalidades da primeira versão do programa VENTOS para previsão probabilística da geração eólica. O programa disponibiliza dois métodos de previsão probabilística: um modelo aditivo generalizado para regressão quantílica [2] e uma abordagem baseada em técnicas de *machine learning* que combina os métodos *Conditional Kernel Density Estimator* [3] e *Gaussian Process* [4]. O programa está implementado em C++ e nesta primeira versão não conta com interface gráfica e banco de dados.

1.3 Organização do texto

O manual está organizado em três seções, sendo a primeira esta breve introdução. Na seção 2 tem-se uma breve descrição dos métodos de previsão probabilística implementados na primeira versão do programa VENTOS. O formato dos arquivos com dados de entrada, as instruções para execução do programa e os principais relatórios de saída gerados são descritos na seção 3.

2. MÉTODOS DE PREVISÃO PROBABILÍSTICA

Nesta seção descrevem-se os dois métodos de previsão probabilística implementados no programa VENTOS para previsão probabilística da geração eólica até 48 horas à frente com resolução horária.

2.1 Abordagem baseada em técnicas de *machine learning*

Denotando por S a velocidade do vento (m/s) e por θ 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, calculadas a partir das previsões \hat{S} e $\hat{\theta}$ fornecidas por um modelo de previsão climática (*Numerical Weather Prediction* – NWP):

$$\hat{u}_h^{D+1} = \hat{S}_h^{D+1} \cdot \cos(\hat{\theta}_h^{D+1}) \quad (1)$$

$$\hat{u}_h^{D+2} = \hat{S}_h^{D+2} \cdot \cos(\hat{\theta}_h^{D+2}) \quad (2)$$

$$\hat{v}_h^{D+1} = \hat{S}_h^{D+1} \cdot \text{sen}(\hat{\theta}_h^{D+1}) \quad (3)$$

$$\hat{v}_h^{D+2} = \hat{S}_h^{D+2} \cdot \text{sen}(\hat{\theta}_h^{D+2}) \quad (4)$$

em que $D+1$ e $D+2$ indicam os dois dias do horizonte de previsão e $h=\{1,2,\dots,24\}$ corresponde ao horário dentro de um dia de previsão.

A transformação para coordenadas cartesianas se faz necessária, pois a direção θ é uma variável aleatória circular. Para obter a previsão probabilística da produção eólica propõe-se uma metodologia com três estágios. No primeiro estágio, o processo gaussiano é aplicado em uma dada previsão pontual de velocidade do vento fornecida por um modelo NWP com a finalidade de quantificar a sua incerteza e obter uma previsão probabilística da velocidade do vento $P(S \angle \theta)$, representada por distribuições gaussianas. Na sequência, no segundo estágio, por meio do estimador do núcleo de densidade de probabilidade condicional, obtém-se uma versão empírica da distribuição de probabilidade condicionada da produção G em função da velocidade, i.e., $P(G | S \angle \theta)$. Ao final, no terceiro estágio, aplica-se a simulação de Monte Carlo para calcular a convolução de $P(S \angle \theta)$ com $P(G | S \angle \theta)$ e obter como resultado a previsão da distribuição de probabilidade marginal da produção eólica $P(G)$. A seguir, tem-se uma descrição detalhada de cada um dos três estágios da metodologia baseada em técnicas de *machine learning* [5].

2.1.1 Estágio 1 - Previsão probabilística das componentes cartesianas da velocidade

Com base nos valores passados de S_h^{dia} e θ_h^{dia} ($\forall h=1,24$ e $dia \leq D$), acompanhados das respectivas previsões geradas pelo modelo NWP, pode-se ajustar um processo gaussiano (*gaussian process*) para cada componente cartesiana da velocidade.

Um processo gaussiano é uma espécie de modelo de regressão. Neste caso, as variáveis respostas correspondem aos valores verificados das componentes cartesianas u e v da velocidade do vento, enquanto as variáveis explicativas são os respectivos valores previstos \hat{u} e \hat{v} , calculados a partir das previsões fornecidas pelo modelo NWP.

O processo gaussiano fornece as previsões dos parâmetros da densidade gaussiana para cada horário $h=\{1,\dots,24\}$ ao longo dos dias $D+1$ e $D+2$, i.e., as médias $\mu_u^{D+1}(h), \mu_u^{D+2}(h), \mu_v^{D+1}(h), \mu_v^{D+2}(h)$ e os

desvios padrão $\sigma_u^{D+1}(h), \sigma_u^{D+2}(h), \sigma_v^{D+1}(h), \sigma_v^{D+2}(h)$ das componentes cartesianas da velocidade do vento.

Com base no conjunto de registros históricos verificados e previstos da velocidade e direção do vento para todo $h=1,24$ e $\text{dia} \leq D$ ajusta-se um processo gaussiano aos dados de cada componente cartesiana da velocidade do vento.

Por exemplo, no caso da componente u , ajusta-se um processo gaussiano ao conjunto de n pares entrada/saída, nos quais as saídas são os valores verificados (passados) da componente u e as entradas as respectivas estimativas (previsões) fornecidas pelo modelo NWP. O principal resultado do ajuste é o vetor de hiperparâmetros Ω_u .

Similarmente, no caso da componente v , um segundo processo gaussiano é ajustado em um conjunto contendo n pares entrada/saída, em que as saídas são os valores passados da componente v e as entradas correspondem às respectivas estimativas fornecidas pelo modelo NWP. Como no caso anterior, o resultado mais importante do ajuste é o vetor de hiperparâmetros Ω_v .

O ajuste de um processo gaussiano consiste em determinar os valores ótimos dos hiperparâmetros por meio da maximização da verossimilhança.

Na etapa de previsão, os processos gaussianos previamente ajustados são aplicados nas previsões pontuais $\hat{u}_h^{D+1}, \hat{u}_h^{D+2}, \hat{v}_h^{D+1}$ e $\hat{v}_h^{D+2} \forall h=1,24$, geradas pelo modelo NWP, com a finalidade de obter previsões das médias μ e desvios padrão σ que definem as densidades gaussianas para u e v em cada hora h dos dias $D+1$ e $D+2$:

$$f_h^{D+1}(u) = N(\mu_u^{D+1}(h), \sigma_u^{D+1}(h)) \quad (5)$$

$$f_h^{D+2}(u) = N(\mu_u^{D+2}(h), \sigma_u^{D+2}(h)) \quad (6)$$

$$f_h^{D+1}(v) = N(\mu_v^{D+1}(h), \sigma_v^{D+1}(h)) \quad (7)$$

$$f_h^{D+2}(v) = N(\mu_v^{D+2}(h), \sigma_v^{D+2}(h)) \quad (8)$$

Em (5), $\mu_u^{D+1}(h)$ e $\sigma_u^{D+1}(h)$ são funções de \hat{u}_h^{D+1} , enquanto em (6) $\mu_u^{D+2}(h)$ e $\sigma_u^{D+2}(h)$ são funções de \hat{u}_h^{D+2} . De forma similar, em (7), $\mu_v^{D+1}(h)$ e $\sigma_v^{D+1}(h)$ são funções de \hat{v}_h^{D+1} e em (8) $\mu_v^{D+2}(h)$ e $\sigma_v^{D+2}(h)$ são funções de \hat{v}_h^{D+2} . Portanto, o processo gaussino fornece as previsões probabilísticas das componentes cartesianas da velocidade do vento em função das previsões pontuais $\hat{u}_h^{D+1}, \hat{u}_h^{D+2}, \hat{v}_h^{D+1}$ e $\hat{v}_h^{D+2} \forall h=1,24$ fornecidas pelo modelo NWP.

2.1.2 Estágio 2 - Estimação da densidade da produção condicionada à velocidade

A produção eólica depende da velocidade do vento, contudo a produção guarda uma relação estocástica com a velocidade. Assim, para a previsão probabilística da produção eólica é necessário estimar a densidade da produção condicionada às componentes cartesianas da velocidade do vento $f(g | u, v)$.

A estimação empírica da densidade condicional $f(g | u, v)$ pode ser realizada por meio de estimadores não paramétricos, por exemplo, o *Conditional Kernel Density Estimator* (CKDE). A estimação requer uma amostra grande de valores passados da geração eólica e respectivos valores das componentes cartesianas u e v da velocidade do vento.

2.1.3 Estágio 3 - Previsão probabilística da produção eólica

Para cada hora do horizonte de previsão dispõe-se de densidades gaussianas das componentes cartesianas da velocidade do vento $f_h^{D+1}(u), f_h^{D+2}(u), f_h^{D+1}(v), f_h^{D+2}(v)$ nas equações (5-8). Adicionalmente, também estão disponíveis as densidades empíricas $f(g|u,v)$ estimadas pelo CKDE.

Assim, as previsões das densidades marginais da produção eólica (previsão probabilística) são obtidas pelas seguintes convoluções, resolvidas por meio de simulação de Monte Carlo para cada hora h do horizonte de previsão:

$$f_h^{D+1}(g) = \iint f(g|u,v) \cdot f_h^{D+1}(u) \cdot f_h^{D+1}(v) \cdot dudv \quad (9)$$

$$f_h^{D+2}(g) = \iint f(g|u,v) \cdot f_h^{D+2}(u) \cdot f_h^{D+2}(v) \cdot dudv \quad (10)$$

2.1.4 Métodos empregados na abordagem via *machine learning*

Na sequência apresentam-se os dois principais métodos empregados na abordagem via *machine learning*, nomeadamente a estimação não paramétrica de densidades de probabilidade [3] e o processo gaussiano [4].

2.1.4.1 Estimação não paramétrica de densidades de probabilidade

A estimação da densidade condicional de núcleo fornece uma ferramenta para construir modelos preditivos, os quais têm como saída a previsão da função densidade de probabilidade da produção eólica. Este resultado pode ser transformado em várias representações de incerteza, tais como quantis, desvio padrão e distância interquartílica, além de fornecer previsões pontuais. Trata-se de uma abordagem de previsão probabilística completa e robusta.

Dada uma amostra aleatória contendo n observações X_1, \dots, X_n . A função densidade estimada empiricamente é uma versão suavizada do histograma, conforme ilustrado pela linha azul no lado direito da Figura 4 para uma amostra com 6 observações. A estimativa não paramétrica da densidade de probabilidade (linha azul no lado direito da Figura 4) é obtida pela soma das curvas $K(x)$, denominadas núcleos de densidade (*kernel*), posicionadas em cada observação da amostra (linhas tracejadas no lado direito da Figura 4).

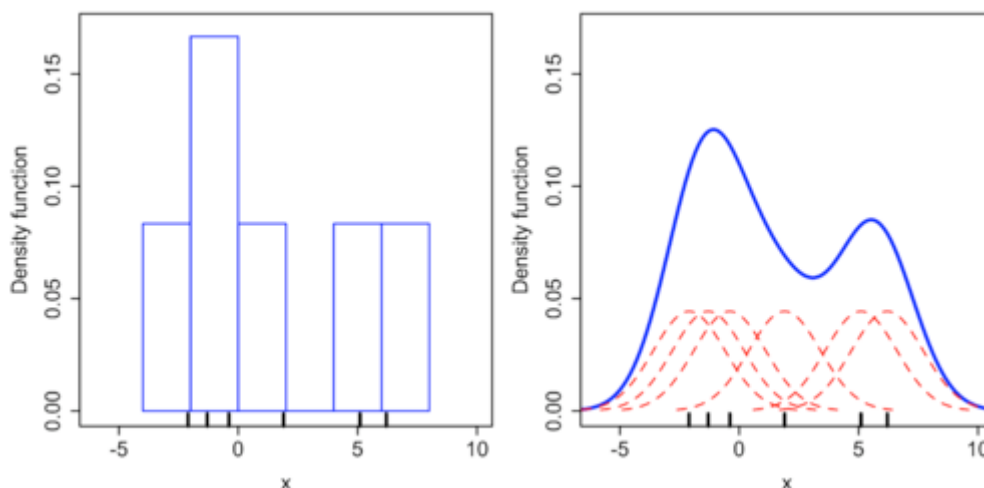


Figura 4. KDE (Fonte: Wikipedia)
https://en.wikipedia.org/wiki/Kernel_density_estimation

Formalmente, a densidade $\hat{f}(x)$ estimada pelo KDE é expressa pela seguinte soma:

$$\hat{f}(x) = \frac{1}{N \cdot h} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right) \quad (11)$$

em que h representa a largura de banda (*bandwidth*), um parâmetro previamente ajustado e que controla a dispersão do *kernel* $K(x)$.

O *kernel* $K(x)$ pode assumir diferentes formas em função da variável modelada, conforme ilustrado na Tabela 1, mas em todos os casos $\int K(x)dx = 1$. No programa VENTOS adotou-se um *Kernel* gaussiano.

Tabela 1. Especificações típicas para o *Kernel*

Tipo	Variável	<i>Kernel</i>
Variável limitada ao intervalo [0,1]	Produção eólica em p.u.	Beta
Variável que assume apenas valores não negativos	Velocidade do vento	Gama
Variável ilimitada	Temperatura	Gaussiana
Variável circular	Direção do vento	Distribuição de Von Mises

Substituindo o *kernel* $K(x)$ em (11) por um *kernel* $H(x)$, obtido a partir da integração de $K(x)$, o KDE fornece um estimador não paramétrico da função distribuição de probabilidade acumulada de uma variável aleatória x , $F(x)$:

$$\hat{F}(x) = \frac{1}{N \cdot h} \sum_{i=1}^N H\left(\frac{x - X_i}{h}\right) \tag{12}$$

O valor da largura de banda h e a função de núcleo (*Kernel*) devem ser especificados a priori. Contudo, a escolha da largura de banda é mais importante que a escolha da função núcleo. A definição da largura de banda pode ser realizada por meio de tentativa e erro, mas também há algumas heurísticas para a determinação ótima do parâmetro h . Na primeira versão do programa VENTOS a largura de banda h assume um valor fixo definido previamente.

Em geral, a distribuição de Weibull consegue fornecer uma boa modelagem da distribuição de probabilidade da velocidade do vento. Contudo, não há um modelo paramétrico conhecido para a distribuição de probabilidade conjunta da produção e da velocidade. Este problema pode ser contornado pela estimação não paramétrica de densidade conjunta. Conforme ilustrado a seguir, o KDE também pode ser empregado na estimação não paramétrica 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 $(X_{11}, \dots, X_{1p}), \dots, (X_{n1}, \dots, X_{np})$:

$$\hat{f}(x_1, \dots, x_p) = \frac{1}{N \cdot h_1 \cdot \dots \cdot h_p} \sum_{i=1}^N \prod_{j=1}^p K_j\left(\frac{x_{ij} - X_{ij}}{h_j}\right) \tag{13}$$

Em (13), o *kernel* multivariado corresponde ao produto de p *kernels* univariados com *bandwidths* h_1, \dots, h_p .

A mesma ideia empregada pelo KDE na estimação de densidades de probabilidade univariada e multivariada é utilizada pelo *Conditional Kernel Density Estimator* - CKDE na estimação da densidade de probabilidade de uma variável aleatória y condicionada ao valor conhecido de uma variável aleatória x , para as quais se dispõe de uma amostra aleatória pareada contendo n observações $(X_1, Y_1), \dots, (X_n, Y_n)$.

Pelo Teorema da Probabilidade Condicional, sabe-se que a densidade condicionada $f(y|x)$ é a razão da densidade conjunta $f_{xy}(y,x)$, pela densidade marginal $f_x(x)$. Assim, com base em (12) e (13), o CKDE para $\hat{f}(y|x)$ é definido pela seguinte razão:

$$\hat{f}(y|x) = \frac{\sum_{i=1}^n K_x\left(\frac{x-X_i}{h_x}\right) K_y\left(\frac{y-Y_i}{h_y}\right)}{\sum_{i=1}^n K_x\left(\frac{x-X_i}{h_x}\right)} \tag{14}$$

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

Para ilustrar o resultado obtido por meio estimador em (14) considere o diagrama de dispersão no lado esquerdo da Figura 5, no qual são dispostas medições do módulo da velocidade do vento e da geração eólica. A nuvem de pontos no diagrama de dispersão é compatível com o perfil típico de uma curva de potência. O estimador CKDE em (14) processa os pares de medições de velocidade do vento e geração e como resultado produz as densidades $f(\text{geração}|\text{velocidade})$ para diferentes níveis de velocidade, conforme ilustrado no lado direito da Figura 5. Uma observação mais atenta revela que as médias das densidades condicionais seguem o perfil típico de uma curva de potência.

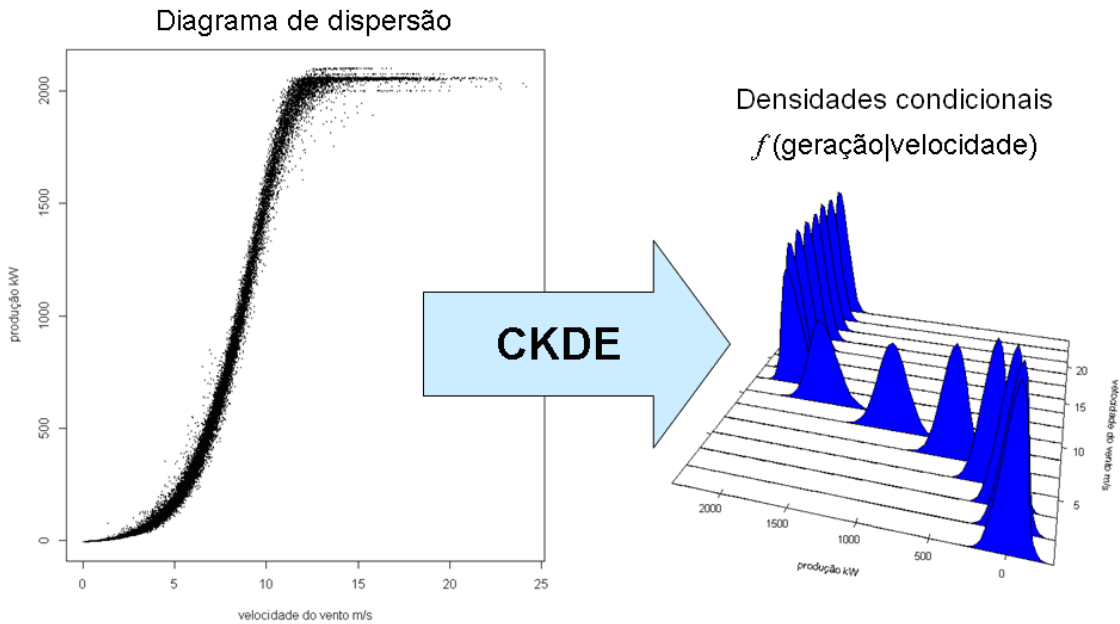


Figura 5. CKDE e as densidades condicionais

Com base no resultado em (13), o estimador CKDE em (14) pode ser facilmente adaptado para acomodar as coordenadas cartesianas da velocidade do vento (u e v) com a finalidade de incorporar o módulo e a direção da velocidade do vento, conforme indicado em (15) pelo estimador da densidade condicional da geração $f(g | u, v)$.

$$\hat{f}(g|u,v) = \frac{\sum_{i=1}^n K_u\left(\frac{u-U_i}{h_u}\right) K_v\left(\frac{v-V_i}{h_v}\right) K_g\left(\frac{g-G_i}{h_g}\right)}{\sum_{i=1}^n K_u\left(\frac{u-U_i}{h_u}\right) K_v\left(\frac{v-V_i}{h_v}\right)} \tag{15}$$

Em (15), G_i , U_i e V_i correspondem aos valores observados da geração eólica e das componentes cartesianas da velocidade do vento respectivamente.

2.1.4.2 Processo gaussiano

Em um conjunto de observações com n pares $(y_t, x_t) \forall t=1,n$, no qual y_t denota a variável resposta e x_t é um vetor com D variáveis explicativas, a relação entre y_t e x_t pode ser representada pela seguinte equação:

$$y_t = f(x_t) + \varepsilon_t \quad \forall t=1,n \tag{16}$$

em que ε denota um ruído gaussiano, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \forall t=1,n$.

Assim, a coleção y_1, \dots, y_n tem distribuição conjunta normal multivariada, $(y_1, \dots, y_n) \sim N_n(0, \Sigma)$. A média nula na distribuição conjunta pode ser obtida após os dados serem apropriadamente escalonados e diferenciados.

A matriz de covariância Σ pode ser parametrizada por meio de uma função núcleo (*Kernel function*), por exemplo, a covariância entre y_p e y_q é especificada da seguinte forma:

$$\Sigma_{p,q} = K(y_p, y_q) = v \cdot \exp\left(-\frac{1}{2} \sum_{d=1}^D \omega_d (x_d(p) - x_d(q))^2\right) + \sigma_\varepsilon^2 \delta(p,q) \quad \forall p,q \in \{1, \dots, n\} \tag{17}$$

em que $\theta = (v, \omega_1, \dots, \omega_D, \sigma_\varepsilon^2)$ é o vetor de hiperparâmetros e $\delta(p,q) = 1$ se $p = q$, caso contrário $\delta(p,q) = 0$.

O termo exponencial em (17) sugere que quando os vetores de entrada $x(p)$ e $x(q)$ são próximos, as respostas y_p e y_q são fortemente correlacionadas, enquanto vetores de entrada distantes resultam em respostas fracamente correlacionadas.

Assim, para um conjunto com n observações conhecidas do vetor 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 matriz de covariância especificada em (17), deseja-se prever $y(n+1)$ correspondente ao vetor de entrada $x(n+1)$.

Com base na priori gaussiana tem-se que a distribuição conjunta de $y = \{y(1), \dots, y(n)\}$ e $y(n+1)$ é normal multivariada:

$$\begin{bmatrix} y \\ y(n+1) \end{bmatrix} \sim N_{n+1} \left(\begin{bmatrix} 0_n \\ 0 \end{bmatrix}, \begin{bmatrix} K_{n \times n} & k_{n \times 1} \\ k_{n \times 1}^T & k^* \end{bmatrix} \right) \tag{18}$$

Na matriz de covariâncias (18), $K_{n \times n}$ denota a matriz de covariâncias entre as n observações conhecidas e parametrizadas conforme em (17). Ainda em (18), o vetor $k_{n \times 1}$ é o vetor contendo as covariâncias entre as respostas das n observações conhecidas e a resposta em $n+1$, ou seja, $k = [K(x(1), x(n+1)), K(x(2), x(n+1)), \dots, K(x(n), x(n+1))]$. Já k^* denota a variância da nova resposta $k^* = K(x(n+1), x(n+1))$. Note que as covariâncias entre as respostas nos diferentes instantes de tempo são calculadas com base nos padrões de entrada, vetor x .

Por meio da distribuição condicional da normal multivariada pode-se obter a previsão da distribuição de $y(n+1)$ que também é gaussiana.

$$f(y(n+1) | y) \sim N(\mu_{n+1}, \sigma_{n+1}^2) \tag{19}$$

em que

$$\mu_{n+1} = k_{n \times 1}^T \cdot K_{n \times n}^{-1} \cdot y_{n \times 1} \tag{20}$$

$$\sigma_{n+1}^2 = k^* - k_{n \times 1}^T \cdot K_{n \times n}^{-1} \cdot k_{n \times 1} \tag{21}$$

A previsão pontual para $n+1$ é fornecida em (20) e a medida da incerteza da previsão é obtida em (21). Note que em (20) a previsão pontual é a combinação linear das n observações conhecidas de y . As fórmulas em (20) e (21) dependem do vetor de hiperparâmetros $\theta=(v, \omega_1, \dots, \omega_D, \sigma_\varepsilon^2)$, cujos elementos devem ser estimados pela maximização do logaritmo da verossimilhança da densidade marginal de y , indicado abaixo:

$$\log L(\theta) = -\frac{1}{2} \log(|K|) - \frac{1}{2} y^T K^{-1} y - \frac{n}{2} \log(2\pi) \quad (22)$$

A matriz K em (22) corresponde à matriz $K_{n \times n}$ em (18) e envolve apenas as n observações da amostra.

A maximização do logaritmo da verossimilhança requer o cálculo do gradiente de $\log L(\theta)$, cujos elementos correspondem ao cálculo das derivadas parciais de $\log L(\theta)$ em relação a cada hiperparâmetro:

$$\frac{\partial}{\partial \theta_j} \log L(\theta) = -\frac{1}{2} \text{tr} \left(K^{-1} \frac{\partial K}{\partial \theta_j} \right) + \frac{1}{2} y^T K^{-1} \frac{\partial K}{\partial \theta_j} K^{-1} y \quad (23)$$

2.2 Regressão quantílica

Um modelo de regressão quantílica deve ser estimado para cada quantil τ que se deseja avaliar, os coeficientes de regressão β variam em função de τ . A seguir, tem-se a especificação do modelo de regressão quantílica da produção eólica, no qual as componentes u e v são modeladas por 10 funções *B-spline*, com nós posicionados nos quantis de u e v :

$$\hat{Q}_t(\tau) = \beta_0(\tau) + \sum_{j=1}^{10} b_j(\hat{u}_t) \beta_j^u(\tau) + \sum_{j=1}^{10} b_j(\hat{v}_t) \beta_j^v(\tau) + \varepsilon_t \quad (24)$$

A especificação e estimação do modelo de regressão quantílica em (24) baseou-se no *script* escrito em linguagem R disponível em [2].

Na equação (24), ε_t denota um erro aleatório e não observável na hora t , a variável resposta é a média horária da produção eólica e as variáveis explicativas $b_j(\hat{u})$ e $b_j(\hat{v}) \forall j=1,10$ correspondem aos coeficientes das previsões das componentes cartesianas u e v da velocidade do vento, oriundas do modelo NWP, em uma base com 10 funções *B-splines* e que juntas totalizam o equivalente a 20 variáveis explicativas.

De forma distinta do estimador mínimos quadrados, na regressão quantílica para um quantil $\tau \in (0,1)$ o estimador dos coeficientes de regressão busca minimizar a seguinte função objetivo em uma amostra de tamanho n com pares $(y_i, X_i) \forall i=1, n$:

$$\text{Min}_{\beta} \sum_{i=1}^n \rho_{\tau}(y_i - X_i^T \beta) \quad (25)$$

em que $\rho_{\tau}(u) = u \cdot [\tau - I(u < 0)]$ e $I(u < 0)$ é uma função indicadora que assume valor unitário para $u < 0$ e valor nulo caso contrário.

A função objetivo (25) pode ser inserida no seguinte problema de programação linear, cuja solução fornece as estimativas dos $K+1$ coeficientes de regressão:

$$\begin{aligned} \underset{u,v,\beta}{\text{Min}} \quad & \tau \sum_{i=1}^n u_i + (1-\tau) \sum_{i=1}^n v_i & (26) \\ \text{s.a.} \quad & y_i - X_i^T \beta = u_i - v_i \quad \forall i = 1, n \\ & u_i \geq 0 \quad \forall i = 1, n \\ & v_i \geq 0 \quad \forall i = 1, n \\ & \beta \in R^{K+1} \end{aligned}$$

Para $\tau = 0,5$ (mediana) o problema de programação linear (PPL) em (26) corresponde ao estimador *least absolute deviations* (LAD).

Alternativamente, a regressão quantílica pode ser ajustada por meio do algoritmo *Iteratively Reweighted Least Squares* (IRLS) [6], um procedimento no qual o estimador mínimos quadrados ponderados é aplicado iterativamente, conforme descrito a seguir:

- 1) Faça $h=0$ (contador de iterações),
 $\beta_h = \mathbf{1}_{K+1}$ (vetor com todos os elementos iguais a 1),
 $\Omega = I_{K+1}$ (matriz identidade)
- 2) Faça $h=h+1$,
 Calcule $\beta_h = (X^T \Omega X)^{-1} X^T \Omega Y$ (mínimos quadrados ponderados)
- 3) Pare se o número máximo de iterações for alcançado, caso contrário continue.
- 4) Calcule o vetor de resíduos $\hat{\epsilon} = Y - X\beta_h$
 Construa o vetor de pesos w a partir dos resíduos:
 Se $\hat{\epsilon}_i < 0$ então $w_i = \tau \cdot \hat{\epsilon}_i \quad \forall i=1, n$
 Se $\hat{\epsilon}_i \geq 0$ então $w_i = (1-\tau) \cdot \hat{\epsilon}_i \quad \forall i=1, n$
- 5) Faça $\Omega_{ii} = 1/w_i$ (diagonal da matriz Ω igual a $1/w$) e volte para o passo 2.

No programa VENTOS são ajustados modelos de regressão quantílica para a geração eólica considerando os seguintes quantis: 2,5%, 5%, 10%, 25%, 50%, 75%, 90%, 95% e 97,5%.

3. EXECUÇÃO DO PROGRAMA

Na primeira versão do programa ventos tem-se apenas um executável gerado em C++ com os métodos de previsão probabilística descritos na seção 2 deste manual. A primeira versão do programa não conta com interface gráfica e banco de dados. Assim, para executar o programa basta um duplo clique sobre o arquivo executável.

O programa requer apenas dois arquivos de dados, ambos em formato texto: um arquivo com dados gerais contendo um reduzido conjunto de parâmetros de execução, entre quais encontra-se a seleção do método de previsão, e o arquivo com dados de velocidade do vento e geração eólica.

Os resultados gerados pelo programa são disponibilizados em arquivos textos e entre eles destaca-se o arquivo contendo as previsões probabilística da geração eólica disponibilizadas na forma de quantis da geração.

3.1 Dados de geração e velocidade do vento

Os principais insumos do programa VENTOS são os dados de geração (kWh) e velocidade do vento. Os dados devem ser disponibilizados com resolução temporal horária. Cada linha do arquivo de dados deve guardar os registros de uma determinada hora e deve seguir o formato descrito na Tabela 2. Na Figura 6 tem-se um exemplo do arquivo de dados.

Tabela 2. Formato de cada linha do arquivo de dados

Campo	Colunas	Formato	Descrição
1	1 a 13	Real	média horária verificada da componente u da velocidade do vento (m/s)
2	14 a 26	Real	média horária verificada da componente v da velocidade do vento (m/s)
3	27 a 39	Real	média horária verificada da velocidade do vento (m/s)
4	30 a 52	Real	média horária verificada da geração eólica kWh
5	53 a 65	Real	média horária prevista da componente u da velocidade do vento (m/s) oriunda do modelo NWP
6	66 a 78	Real	média horária prevista da componente v da velocidade do vento (m/s) oriunda do modelo NWP
7	79 a 91	Real	média horária prevista da velocidade do vento (m/s) oriunda do modelo NWP

0	10	20	30	40	50	60	70	80	90
-8.83515	-4.30919	9.83000	9873.33000	-8.65260	-5.91651	10.48201			
-8.95527	-3.25945	9.53000	10249.02000	-9.87268	-5.18004	11.14911			
-9.49811	-4.42904	10.48000	10665.36000	-10.69125	-5.64403	12.08958			
-10.29677	-2.75901	10.66000	12298.27000	-11.55548	-6.56289	13.28911			
-10.73028	-2.28079	10.97000	12554.98000	-12.51572	-7.49623	14.58892			
-11.08335	-3.17810	11.53000	13190.88000	-13.01710	-7.62124	15.08404			
-10.92074	-3.76031	11.55000	12185.30000	-11.48625	-9.63613	14.99296			
-11.21053	-4.08030	11.93000	13382.87000	-11.82133	-10.19676	15.61146			
-13.46566	-5.99530	14.74000	11842.15000	-11.45086	-10.79467	15.73681			
-14.54624	-6.78302	16.05000	12101.23000	-13.51366	-10.40083	17.05275			
-14.13999	-9.18261	16.86000	14824.79000	-15.13897	-9.94793	18.11491			
-15.56168	-9.72402	18.35000	13230.70000	-16.01823	-9.16968	18.45716			
-15.96875	-9.97838	18.83000	10851.35000	-11.49229	-12.72447	17.14598			
-15.47688	-9.67103	18.25000	10199.29000	-10.40410	-11.87606	15.78880			
-15.60322	-8.64900	17.84000	11792.28000	-10.13732	-11.70076	15.48138			
-13.08894	-7.86463	15.27000	12630.74000	-10.14754	-11.82107	15.57915			
-14.35374	-7.00079	15.97000	12519.32000	-10.00367	-13.16075	16.53114			
-9.70239	-10.77560	14.50000	12394.06000	-8.21471	-14.33763	16.52419			
-3.70382	-14.85523	15.31000	13132.40000	-6.45423	-12.26710	13.86142			
-3.06158	-13.26118	13.61000	12121.63000	-2.58178	-11.61194	11.89549			
-4.06014	-14.15938	14.73000	13659.19000	-3.35918	-13.15016	13.57243			
-2.72156	-12.80395	13.09000	12838.20000	-4.41435	-11.54332	12.35859			
-6.03103	-10.88027	12.44000	12479.05000	-4.41736	-11.57558	12.38980			
-5.18003	-10.16638	11.41000	11764.66000	-4.18939	-10.78413	11.56929			
-5.90453	-6.79239	9.00000	7084.08000	-4.61448	-7.21072	8.56083			
-5.34799	-6.37349	8.32000	5885.90000	-4.96601	-7.52161	9.01309			
-7.08567	-4.96144	8.65000	2771.70000	-5.29722	-7.77084	9.40460			
-6.01327	-3.90506	7.17000	2488.59000	-5.75935	-7.30786	9.30457			
-6.41666	-2.99214	7.08000	3440.16000	-6.44576	-7.56917	9.94184			
-8.26891	-4.03301	9.20000	6898.17000	-6.81782	-6.74998	9.59401			
-6.50706	-2.49782	6.97000	5162.12000	-7.30917	-6.42345	9.73060			
-6.28164	-4.07935	7.49000	4996.17000	-8.12187	-6.35148	10.31048			
-5.38400	-1.64605	5.63000	3533.81000	-8.34700	-6.18978	10.39162			
-6.34237	-0.77875	6.39000	4720.64000	-8.59457	-6.01810	10.49210			
-6.85226	-0.72020	6.89000	5211.76000	-10.05939	-6.54748	12.00253			
-8.49482	3.09186	9.04000	7484.77000	-10.51350	-7.07317	12.67137			
-8.04895	0.42183	8.06000	6563.30000	-10.10209	-5.79423	11.64583			

Figura 6. Arquivo de dados eólicos

3.2 Dados gerais

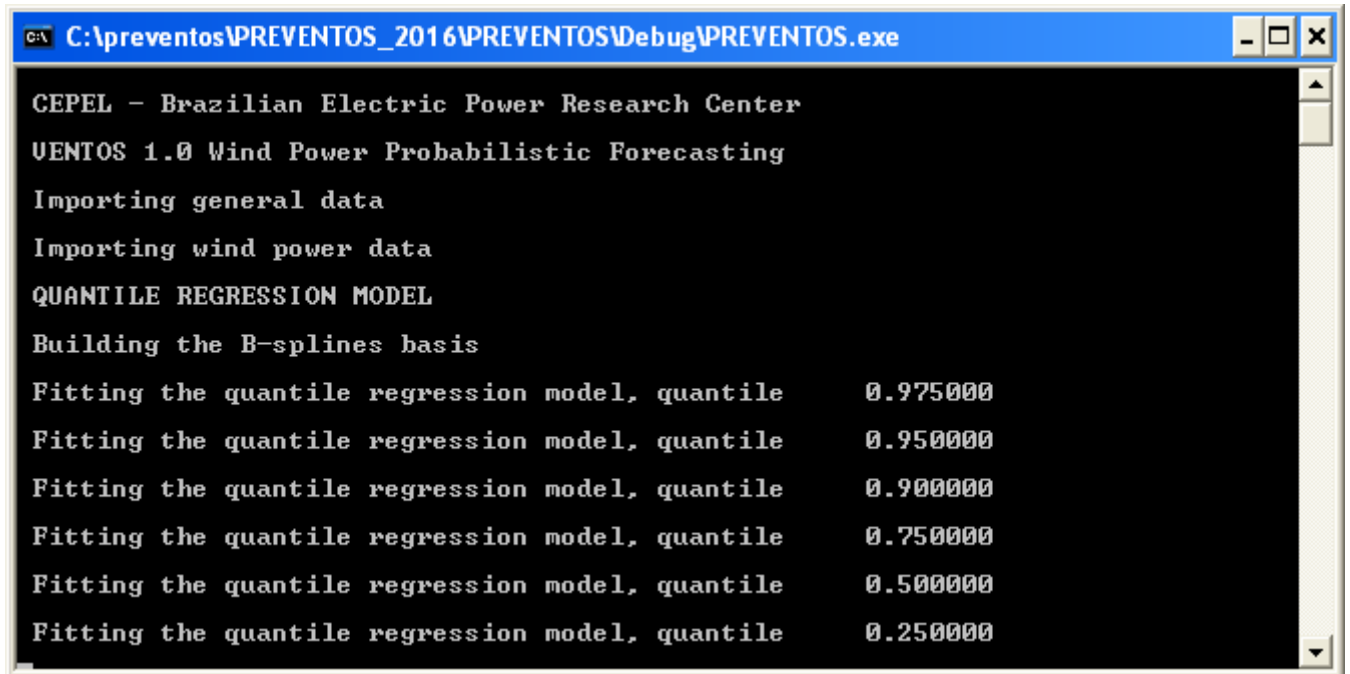
Adicionalmente, o programa lê um arquivo texto contendo dados gerais com informações de apenas quatro parâmetros de entrada, um arquivo com apenas 4 registros, conforme descrito na Tabela 3.

Tabela 3. Dados gerais

Linha	Colunas	Formato	Descrição
1	1 a 4	Inteiro	método de previsão selecionado (0 para machine learning e 1 para regressão quantílica)
2	1 a 4	Inteiro	tamanho da série para o ajuste dos modelos (conjunto de treinamento)
3	1 a 4	Inteiro	horizonte de previsão
4	1 a 4	Inteiro	número de iterações para as simulações

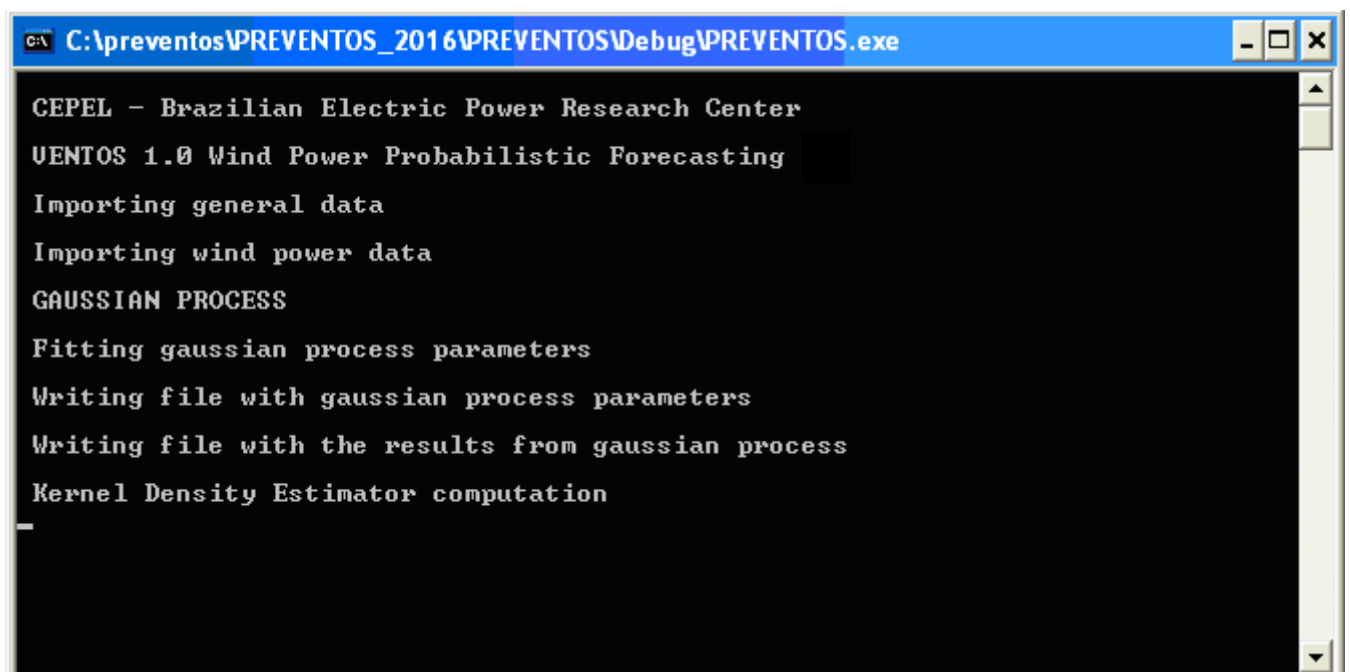
3.3 Execução

O duplo clique no arquivo executável abre a janela ilustrada na Figura 7 na execução via regressão quantílica ou a janela na Figura 8 para previsão com *machine learning*. As janelas permitem acompanhar a execução do programa. Durante a execução o programa carrega os dados de anemométricos e da geração eólica, estima o modelo de previsão e gera previsões a partir da última hora do conjunto de treinamento (Tabela 3) até o horizonte de previsão especificado nos dados gerais (Tabela 3).



```
C:\preventos\PREVENTOS_2016\PREVENTOS\Debug\PREVENTOS.exe
CEPEL - Brazilian Electric Power Research Center
VENTOS 1.0 Wind Power Probabilistic Forecasting
Importing general data
Importing wind power data
QUANTILE REGRESSION MODEL
Building the B-splines basis
Fitting the quantile regression model, quantile      0.975000
Fitting the quantile regression model, quantile      0.950000
Fitting the quantile regression model, quantile      0.900000
Fitting the quantile regression model, quantile      0.750000
Fitting the quantile regression model, quantile      0.500000
Fitting the quantile regression model, quantile      0.250000
```

Figura 7. Execução da regressão quantílica para previsão da geração eólica



```
C:\preventos\PREVENTOS_2016\PREVENTOS\Debug\PREVENTOS.exe
CEPEL - Brazilian Electric Power Research Center
VENTOS 1.0 Wind Power Probabilistic Forecasting
Importing general data
Importing wind power data
GAUSSIAN PROCESS
Fitting gaussian process parameters
Writing file with gaussian process parameters
Writing file with the results from gaussian process
Kernel Density Estimator computation
```

Figura 8. Previsão com abordagem *machine learning*

3.4 Relatório de saída

Os principais resultados gerados pelo programa são as previsões probabilística da média horária da geração eólica em cada hora do horizonte de previsão. No arquivo de saída gerado pelo programa encontram-se os quantis previstos (2,5%, 5%, 10%, 25%, 50%, 75%, 90%, 95% e 97,5%) da geração eólica em cada hora do horizonte de previsão. Conforme ilustrado nas Figuras 9 e 10, cada linha informa os quantis de 97,5% até 2.5% (dispostos em ordem decrescente para facilitar a elaboração de gráficos no MS Excel) em uma determinada hora do horizonte de previsão. A primeira coluna informa a média horária da geração eólica em cada hora do horizonte de previsão.

Arquivo	Editar	Formatar	Exibir	Ajuda					
Resultados gerados pelo modelo de regressão quantílica									
0.155017	0.334674	0.262904	0.207518	0.155040	0.119039	0.072217	0.026007		
0.174766	0.362221	0.279680	0.237098	0.181231	0.144393	0.099269	0.043462		
0.093438	0.359906	0.282127	0.240651	0.185796	0.150604	0.105844	0.047965		
0.090167	0.362749	0.294662	0.256794	0.202953	0.170141	0.126039	0.062639		
0.209145	0.371731	0.320482	0.282065	0.227272	0.193861	0.148262	0.080309		
0.193763	0.395570	0.343280	0.303813	0.251150	0.216647	0.168429	0.097533		
0.300995	0.527017	0.467243	0.402459	0.352268	0.296802	0.225056	0.154046		
0.229147	0.626523	0.547587	0.469407	0.407744	0.327307	0.238744	0.171351		
0.276650	0.642268	0.586549	0.487065	0.433407	0.348426	0.245104	0.180679		
0.414889	0.561052	0.506516	0.429489	0.379616	0.316265	0.234868	0.165648		
0.417998	0.739293	0.661807	0.550251	0.481704	0.365673	0.239321	0.179567		
0.549900	0.807921	0.723560	0.606997	0.515821	0.366981	0.217915	0.160178		
0.569411	0.792427	0.707827	0.592150	0.507937	0.371103	0.232651	0.174229		
0.556317	0.798367	0.723102	0.603685	0.516161	0.378985	0.241428	0.180471		
0.547838	0.707918	0.633157	0.535767	0.460203	0.354062	0.244936	0.179379		
0.592098	0.588142	0.536595	0.475514	0.373617	0.275166	0.183724	0.098644		
0.333405	0.934628	0.892297	0.833854	0.742509	0.601258	0.469589	0.295867		
0.421714	0.623242	0.557858	0.478168	0.364360	0.240467	0.131114	0.050120		
0.304685	0.585307	0.517906	0.438328	0.327681	0.212087	0.109203	0.037180		
0.406586	0.598582	0.529489	0.442072	0.330811	0.207887	0.100428	0.032186		
0.377271	0.490585	0.414958	0.336314	0.241114	0.149114	0.066513	0.016764		
0.120229	0.511100	0.440644	0.354301	0.252554	0.148157	0.057474	0.010146		
0.059276	0.492662	0.421225	0.357066	0.263688	0.179608	0.102619	0.040448		
0.118852	0.454720	0.376023	0.320967	0.237170	0.167488	0.103076	0.044239		

Figura 9. Relatório de saída com os quantis da geração eólica (valores em p.u.) previstos pelo modelo de regressão quantílica, cada linha informa os quantis previstos em uma determinada hora

Arquivo	Editar	Formatar	Exibir	Ajuda					
Resultados gerados pelo processo gaussiano + KDE									
0.155017	0.751266	0.698435	0.614625	0.391399	0.209482	0.091843	0.044131	0.024285	
0.174766	0.761628	0.715632	0.642163	0.422861	0.224669	0.103653	0.050896	0.029845	
0.093438	0.760572	0.714627	0.641396	0.422541	0.228789	0.111650	0.056213	0.034188	
0.090167	0.763723	0.717043	0.644365	0.429506	0.243077	0.134977	0.072373	0.047318	
0.209145	0.768817	0.721729	0.646995	0.441571	0.271680	0.165391	0.101052	0.070112	
0.193763	0.778905	0.735723	0.669850	0.472106	0.298053	0.180175	0.116572	0.082702	
0.300995	0.827428	0.794472	0.739246	0.594189	0.398396	0.264319	0.171379	0.128970	
0.229147	0.839198	0.810221	0.765678	0.644557	0.445568	0.290675	0.183836	0.138669	
0.276650	0.831046	0.797861	0.750878	0.632763	0.492423	0.370424	0.257287	0.174277	
0.414889	0.829737	0.798406	0.744324	0.609115	0.451858	0.312853	0.204839	0.146512	
0.417998	0.855131	0.811616	0.772565	0.670423	0.515592	0.404112	0.287379	0.182151	
0.549900	0.892898	0.838576	0.787686	0.693934	0.541343	0.435594	0.327855	0.186411	
0.569411	0.864727	0.818976	0.778881	0.682337	0.521361	0.404602	0.288827	0.187391	
0.556317	0.856582	0.823784	0.784748	0.686802	0.513406	0.368452	0.251345	0.176060	
0.547838	0.849073	0.820639	0.781045	0.676354	0.475625	0.300376	0.187078	0.141231	
0.592098	0.817405	0.771862	0.722735	0.594158	0.283529	0.112907	0.053044	0.030403	
0.333405	0.857563	0.832563	0.798212	0.668691	0.361791	0.108131	0.034653	0.013531	
0.421714	0.833168	0.802509	0.737965	0.575878	0.251471	0.049983	0.010226	0.000106	
0.304685	0.830712	0.799246	0.726643	0.560946	0.236545	0.050144	0.010703	0.000396	
0.406586	0.823566	0.790129	0.709505	0.539131	0.210749	0.036471	0.006047	0.000000	
0.377271	0.821715	0.784775	0.690065	0.510357	0.195989	0.044554	0.009329	0.000000	
0.120229	0.814917	0.775324	0.682808	0.496227	0.172304	0.031579	0.004853	0.000000	
0.059276	0.819821	0.779890	0.703238	0.536777	0.231049	0.069842	0.022345	0.007247	
0.118852	0.808014	0.762795	0.688311	0.519242	0.232753	0.077713	0.029054	0.011794	

Figura 10. Relatório de saída com os quantis da geração eólica (valores em p.u.) previstos pela abordagem machine learning, cada linha informa os quantis previstos em uma determinada hora

Além dos relatórios com as previsões dos quantis da geração eólica, o programa gera relatórios secundários contendo as estimativas dos parâmetros do modelo de previsão, por exemplo, a evolução do logaritmo da verossimilhança e os hiperparâmetros do *Gaussian Process*, bem como os coeficientes de regressão no caso do modelo de regressão quantílica.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ALÉ, J.A.V., SIMIONI, G.C. Procedimentos Laboratoriais para Ensaio de Aerogeradores. V Congresso Brasileiro de Metrologia, Salvador, Brasil, 2009.
- [2] 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.
- [3] BESSA, R.J., MENDES, J., MIRANDA, V., BOTTERUD, A., WANG, J., ZHOU, Z. Quantile-copula density forecast for wind power uncertainty modeling, IEEE Trondheim PowerTech, Trondheim, Norway, Jun. 2011.
- [4] RASMUSSEN, C.E., WILLIAMS, C.K.L. Gaussian Processes for Machine Learning, the MIT Press, <http://www.gaussianprocess.org/gpml/chapters/>, 2006.
- [5] 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.
- [6] WOOD, S.N. Generalized Additive Models: an introduction with R, Chapman & Hall, 2006.