



## **Análise da velocidade do ventos através da Aproximação Agregada por partes e Regressão Quantílica**

**Helen Barboza da Silva <sup>1</sup>, Ronaldo R. B. de Aquino <sup>2</sup>, Jonata C. de Albuquerque <sup>3</sup>, Manuel Herrera <sup>4</sup>, Aida A. Ferreira <sup>5</sup>**

<sup>1 2 3</sup> Universidade Federal de Pernambuco (helenbarbozafr@gmail.com),  
(rrbaquino@gmail.com), (jonata.albuquerque@ufpe.br)

<sup>4</sup> University of Bath (amhf20@bath.ac.uk)

<sup>5</sup> Instituto Federal de Pernambuco (aidaferreira@recife.ifpe.edu.br)

### **Resumo**

O alto custo da produção de energia, e o fato da energia eólica ser amplamente disponível levou vários países a estabelecer incentivos estimular a geração de energia eólica. Este trabalho propõe a implementação e comparação de dois métodos de análise de séries temporais: Piecewise Aggregate Approximation (PAA) e a Regressão Quantílica em conjunto com o PAA (PAA/RQ). O objetivo é estimar a quantidade mínima de eventos extremos da velocidade do vento, e determinar se uma região é propícia a implantação de um Parque eólico. O Brasil tem um enorme potencial de energia eólica. A diversificação da sua matriz energética está se tornando um desafio necessário hoje em dia no compromisso de usar fontes de energia renováveis. O desempenho das duas propostas baseadas no PAA é testado para os parques eólicos das regiões sul e nordeste. Os resultados indicam que o método PAA/RQ teve desempenho superior que o PAA, identificando uma maior quantidade de valores extremos.

Palavras-chave: PAA, Velocidade do Vento, Energia Eólica, Regressão Quantílica.

Área Temática: Energia e energias renováveis

## **Wind speed analysis through the Piecewise Aggregated Approach and Quantile Regression**

### **Abstract**

*The high cost of energy production, coupled with the advantages of wind power as renewable ,has led several countries to establish to wind power generation. This work proposes the implementation and comparison of two time series analysis methods: Piecewise Aggregate Approximation (PAA) and a PAA plus quantile regression process(PAA/QR). The aim is to estimate the minimum amount of extreme cut-in and cut-out events of the wind speed in the power generation process. Brazil has an enormous wind power potential. The diversification of its energy matrix is becoming a necessary challenge nowadays in the commitment of using renewable energy sources. The performance of the two PAA based proposals is tested for to the wind farms of the south and northeast regions. The results indicate that the PAA/QR method performed better than the PAA method because it identified a greater amount of extreme values.*

*Key words: PAA, wind speed, wind energy, quantile regression.*

*Theme Area: Energy and renewable energy*



## 1 Introdução

Devido às incertezas do vento e a grande quantidade de dados frequentemente associados, é necessário usar técnicas ou ferramentas especificamente adaptadas para quantificar a produção de energia. Elas visam auxiliar no melhor planejamento do aumento da capacidade de geração, a diversificação da matriz de energia através da geração distribuída, e o equilíbrio energético entre carga e geração, por exemplo. Estes são alguns benefícios a curto, médio e longo prazo, para termos um melhor gerenciamento do Sistema Interligado Nacional (SIN).

Avanços na área de eletrônica e informática melhoraram a aquisição e o armazenamento de dados. No entanto, a grande quantidade de dados disponíveis hoje em dia pode levar a dificuldades na manipulação, controle e interpretação de informações, devido a limitações humanas e/ou técnicas. A Aproximação Agregada em por Partes (PAA) representa uma melhoria no método de amostragem de dados (KEOGH et al., 2001). O processo divide os dados em segmentos ou janelas de igual tamanho. Em seguida, calcula-se a média em cada segmento, reduzindo a dimensionalidade para melhor análise de longas séries temporais. Neste artigo, o PAA é acoplado com a regressão quantílica para modelar e analisar as séries temporais de velocidade do vento. A regressão quantílica é capaz de lidar com modelos baseados em quantis inferiores e superiores em vez de usar a média que normalmente é empregada em modelos de regressão linear clássica (HERRERA et al., 2016). O objetivo da aproximação da regressão quantílica para cada intervalo de séries temporais de vento é facilitar o estudo e a representação de valores extremos. Isso é importante porque podemos identificar os locais mais favoráveis para a implementação de parques eólicos, bem como identificar os períodos em que a turbina não está em operação.

A Associação Mundial de Energia Eólica (WWEA) mostra um uso crescente da energia eólica no mundo. A capacidade mundial de vento atingiu 486.661 MW no final de 2016, dos quais 54.846 MW foram adicionados em 2016. Isso representa uma taxa de crescimento de 11,8% (17,2% em 2015). O vento é uma fonte de energia renovável. No caso do Brasil, o vento é uma das fontes da Produção de Energia Renovável (PER), que possui maior potência instalada no território nacional. Isso se deve à incentivos de programas como o Programa de Incentivos para Fontes Alternativas de Energia Elétrica (PROINFA). Espera-se que o Brasil esteja incluído nos 10 países com maior capacidade instalada de parques eólicos, com a implementação do PROINFA (SILVA, 2006).

As séries temporais utilizadas neste trabalho consistem em valores de velocidade do vento medidos a cada 10 minutos, começando às 0h de 1 de janeiro de 2014 e terminando às 23h de 31 de dezembro de 2016. Os bancos de dados correspondem a medidas obtidas do Operador Nacional do Sistema Elétrico (ONS) nas regiões Sul e Nordeste.

## 2 Regressão Quantílica

A regressão linear tradicional baseia-se na minimização da soma dos quadrados dos resíduos e permite estimar uma classe geral de modelos para a função média condicional. A regressão quantílica (RQ) é uma extensão da regressão linear tradicional (KOENKER & BASSETT, 1978). Os modelos RQ usam a função mediana condicional e também para todos os outros quantis, fazendo uso da minimização de erros absolutos ponderados.

A grande vantagem de RQ é que permite identificar variações intra e inter-quantil. RQ fornece um modelo diferente para cada quantil. Isso permite lidar com informações muito mais refinadas do que abordar um modelo em torno da média, e fornece detalhes específicos para qualquer nível de resposta variável, incluindo valores extremos. Desta forma, é possível examinar como o modelo se comporta na cauda de distribuição mais baixa (por exemplo, o



quantil de 0,05) ou na parte superior da cauda (por exemplo, o quantil de 0,95). Isso é de interesse fundamental para este artigo, que trata dos valores superiores e inferiores da velocidade do vento em um determinado período.

A função mediana condicional,  $Q_q(y|x)$ , tem o quantil  $q$  igual a 0,5. Esta expressão pode ser estendida a qualquer quantil da distribuição. O quantil  $q \in (0,1)$  para  $y$  divide os dados, deixando essa proporção de pontos abaixo e tendo  $1 - q$  dos pontos acima. Para cada  $q$ , RQ resulta em uma função ajustada do quantil condicional, de modo que  $y_q = F^{-1}(q)$  e  $F(y_q) = q$  (KOENKER & GELING, 2001) (DABROWSKA, 2005). O objetivo é encontrar os valores de  $\beta_q$  que minimizam (1).

$$Q(\beta_q) = \sum_{i: y_i \geq x'_i \beta} q |y_i - x'_i \beta_q| + \sum_{i: y_i \leq x'_i \beta} (1 - q) |y_i - x'_i \beta_q| \quad (1)$$

$$\min_{\beta \in R} \sum_{i=1}^n \rho_{\theta}(y_i - x'_i \beta) \quad (2)$$

Em que,  $\rho$  é a função *check* definida por:

$$\rho_{\theta}(u) = f(x) = \begin{cases} (\theta - 1)u, & u < 0 \\ \theta u, & u \geq 0 \end{cases} \quad (3)$$

Diferentes quantis podem ser obtidos variando  $\theta$ . Assim, esse tipo de regressão permite verificar os impactos em toda a distribuição. A função não diferenciável da Equação (1) requer métodos de programação linear para sua minimização, como o método Simplex. A adaptação deste algoritmo ao problema de regressão quantílica é encontrada em (KOENKER & D'OREY, 1987).

### 3 PAA – Aproximação Agregada por Partes

A Aproximação Agregada por Partes (PAA) (KEOGH et al., 2001) é uma técnica de redução de dimensão bastante simples para entender e implementar. Ela divide os dados da série temporal em segmentos igualmente espaçados, proporcionando resultados semelhantes aos de transformações mais elaboradas, como Transformada de Fourier e Wavelet, na tarefa de indexar séries temporais (LIN et al., 2002). Além disso, os resultados da Transformada de Wavelet Discreta (DWT) e os resultados com base no PAA são equivalentes às distâncias de cálculo entre as séries temporais (KOENKER & D'OREY, 1987). Os resultados robustos encontrados usando métodos baseados em PAA representam uma vantagem fundamental para o trabalho aqui apresentado. Por exemplo, tendo um sinal  $X = x_1, \dots, x_n$ , é possível representá-lo em um espaço  $N$ , com  $N \ll n$ , por  $X = x_1, \dots, x_N$ , onde cada elemento pode ser calculado pelo seguinte (4):

$$\bar{x}_i = \frac{N}{n} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j \quad (4)$$

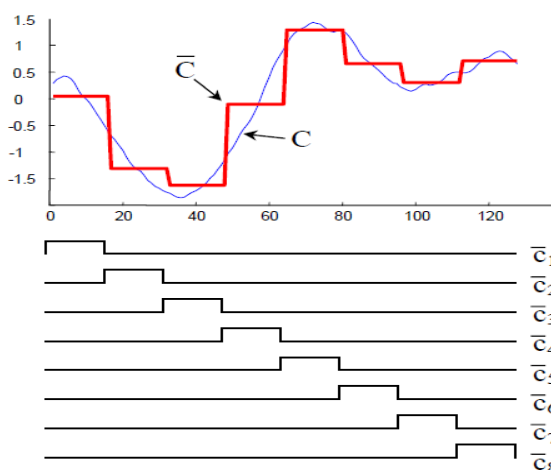
Esta equação simplesmente indica que um sinal é dividido em  $N$  janelas e que o valor médio representa todos os pontos dentro da janela. O parâmetro  $N$  pode ser encontrado traçando o erro de reconstrução como uma função de  $N$ . Desta forma, a transformação de uma série de tempo em uma representação PAA é realizada por  $(\bar{C})$ . A Fig. 1 mostra essa transformação.



Quadro 1 – Quantidade de janelas para cada conjunto de dados

<i>Quantidade de dias do mês</i>	<i>n</i>	<i>N</i>
31	4464	1116
30	4320	1080
29	4176	1044
28	4032	1008

Figura 1 – Representação PAA da Série Temporal



Fonte: (LIN et al., 2007)

## 4 Metodologia

### 4.1 Processamento dos Dados

#### 4.1.1 Descrição dos Dados

A análise foi realizada utilizando dados de duas regiões, o Parque Eólico da Região Sul (PERS), o Parque Eólico da Região Nordeste (PERNE). As bases utilizadas já levam em consideração a medida da velocidade do vento na altura do cubo do rotor da turbina eólica, ou seja, a altura de 100m. Os Parques foram escolhidos para representar melhor uma variedade de cenários para estudo dos problemas de gerenciamento de velocidade do vento no Brasil, pois os dois possuem características divergentes. Existem dois tipos principais de eventos extremos na geração de energia eólica que são conhecidos como ocorrências de cut-in e cut-out. As ocorrências de cut-out são velocidades de vento maiores ou iguais a 20 m/s (velocidade UP). Quando isso acontece, as turbinas devem ser desligadas por razões de segurança estrutural. As ocorrências de cut-in são velocidades do vento inferiores ou iguais a 3 m/s. Neste caso, as turbinas não estão em operação porque suas hélices não têm velocidade de arranque suficiente (velocidade LOW).

#### 4.1.2 Imputação de Valores Ausentes

Um problema significativo e frequentemente encontrado em bancos de dados de velocidade do vento (e especificamente nos parques eólicos do Brasil) é a ausência de medidas de vento com qualidade suficiente para uso em projetos elétricos de energia eólica (SILVA, 2006). Assim, o primeiro passo no pré-processamento dos dados é a imputação dos valores em falta. Isso é feito calculando os valores médios das medições realizadas em 10, 20,



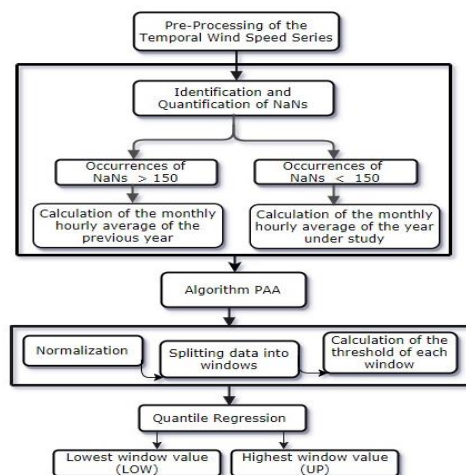
30, 40, 50 e 60 minutos em cada uma das 24 horas do dia de cada mês. Para obter um conjunto consistente de dados para que os resultados estejam mais próximos da realidade.

#### 4.1.3 Aplicação do Algoritmo Proposto

Este artigo propõe uma representação PAA das séries temporais de velocidade, com a decomposição da série que corresponde a um dia em 36 intervalos de tempo (janelas) com 4 medições de 10 minutos cada, gerando novos dados em diferentes escalas de uma fonte comum. A proposta analisa as séries temporais de velocidade do vento para modelar as características associadas aos seus valores extremos.

Em primeiro lugar, os dados obtidos com a aplicação do algoritmo PAA são analisados. A abordagem PAA é então aprimorada e a RQ é abordada sobre a nova série temporal no formato PAA para lidar com os quantil superior (0,95) e inferior (0,05) para cada janela. No presente estudo, encontram-se 4 medidas de 10 minutos em cada janela, os valores escolhidos para os quantis superior e inferior são os valores mais altos e mais baixos do PAA de cada janela. A Fig. 2 apresenta um fluxograma, das técnicas aplicadas.

Figura 2 – Fluxograma do passo a passo para aplicação do método



Fonte: Próprio autor

## 5 Resultados

Esta seção apresenta os resultados da análise experimental realizada para avaliar o desempenho da metodologia proposta. Os resultados indicam que o método PAA/RQ teve um desempenho superior que o método PAA para os dois parques eólicos.

Com relação ao PERS, para os valores UP e LOW, o método PAA identificou um número bem abaixo de ocorrências de eventos extremos quando comparado aos valores empíricos dos três anos analisados. O número de ocorrências identificadas em alguns meses não atingiu 25% das ocorrências totais, como pode ser observado nas Fig. 3 e Fig. 4. Com o método PAA/RQ, este número foi muito maior, manteve um padrão de quase 50% como pode ser observado (Ver Fig. 3 e Fig. 4). Este desempenho foi considerado satisfatório, uma vez que o método propõe a identificação do número mínimo de ocorrências, sendo que uma janela de 4 medidas de 10min é identificado valores iguais ou inferiores a 3 m/s (cut-in) e igual ou superior a 20 m/s (cut-out), como extremos. Outro ponto importante é que, para este parque, o método PAA/RQ identificou a maioria das ocorrências nos mesmos meses que o método empírico, o que não ocorreu com o método PAA, que apontou meses que não corresponderam aos identificados no método empírico. O método PAA/RQ apresentou 50% mais eventos de eventos extremos quando comparado ao método PAA.



Com relação ao PERNE, o método PAA identificou algumas ocorrências de UP, mas este parque eólico não apresentou valores de velocidade do vento superiores a 20 m/s, permanecendo em média com valores inferiores a 18 m/s. No entanto, o método PAA/RQ não identificou valores UP, conforme esperado. Em relação a ocorrências de LOW, o método PAA apresentou um número bem abaixo das ocorrências de eventos extremos quantificadas no método empírico. O número de ocorrências identificadas, em alguns meses, não atingiu 20% das ocorrências totais, como pode ser visto na Fig. 5. Com o método PAA/RQ, este número foi muito maior mantendo um padrão de quase 50% das ocorrências de identificados de eventos extremos, como pode ser visto na Fig. 5. Outro ponto importante é que o método PAA/RQ apresentou em média 60% mais eventos extremos quando comparado ao método PAA.

Figura 3 – Gráficos de múltiplas barras de ocorrências LOW nos 2014, 2015 e 2016 para SERS com (a) Valores empíricos, (b) Método PAA/RQ e (c) Método PAA



Figura 4 – Gráficos de múltiplas barras de ocorrências UP nos 2014, 2015 e 2016 para SERS com (a) Valores empíricos, (b) Método PAA/RQ e (c) Método PAA

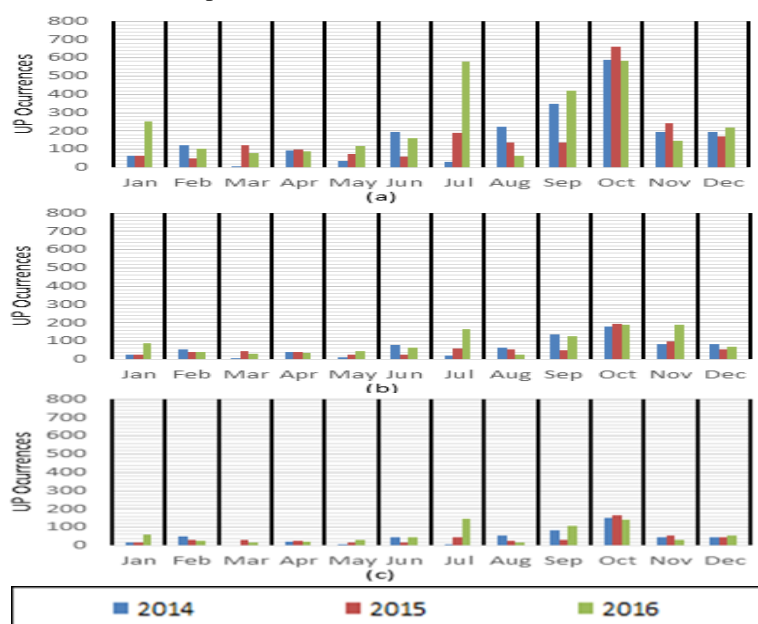
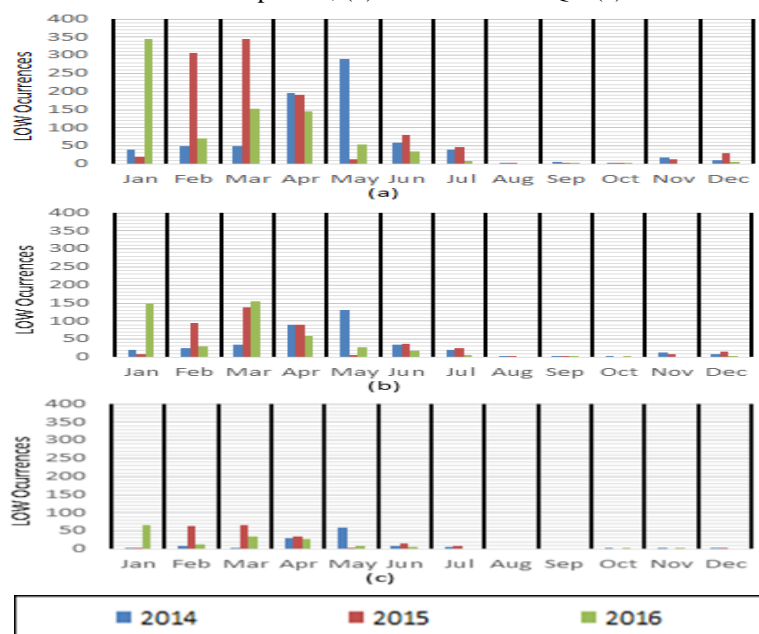






Figura 5 – Gráficos de múltiplas barras de ocorrências LOW nos 2014, 2015 e 2016 para SERNE com (a) Valores empíricos, (b) Método PAA/RQ e (c) Método PAA



## 6 Conclusões

Este trabalho analisa a geração diária de vento para dados coletados a cada 10 minutos de janeiro de 2014 até dezembro de 2016. Isso é feito usando séries temporais de dois parques eólicos localizados em diferentes regiões do país. É proposto uma comparação entre o Método PAA e o aprimoramento do método (PAA) com o uso da regressão quantílica. Os resultados iniciais são promissores, pois permitem obter aspectos descritivos rápidos sobre o comportamento dessas séries temporais devido à redução de tamanho dos dados analisados. Observou-se que este método detecta um número de ocorrências de eventos extremos do vento próximo ao dos valores empíricos observados. De modo, que é possível determinar os períodos em que as turbinas estão fora de operação, comprometendo a geração de energia dos parques eólicos. O método apresentado reduz o esforço computacional quando comparado com outros métodos de análise de séries temporais existentes na literatura.

## Referências

KEOGH E., K. CHAKRABARTI, PAZZANI M., and MEHROTRA S., **Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases**, Knowl. Inf. Syst., vol. 3, no. 3, pp. 263–286, Aug. 2001.

HERRERA M., EAMES M., RAMALLO-GONZALEZ A., LIU C., and COLEY D. Quantile regression ensemble for summer temperatures time series and its impact on built environment studies. In: **International Congress on Environmental Modelling and Software, 2016**.

SILVA N. F. **Fontes de energia renováveis complementares na expansão do setor elétrico brasileiro: o caso da energia eólica**. Rio de Janeiro, abril de 2006, Tese de Doutorado-Programa de Planejamento Energético, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006, 263p.



KOENKER R. and BASSETT G., **Regression Quantiles**, *Econometrica*, vol. 46, no. 1, p. 33–50, 1978.

KOENKER R. and GELING O., **Reappraising medfly longevity**: a quantile regression survival analysis, *J. Am. Stat. Assoc.*, vol. 96, no. 454, p. 458–468, 2001.

DABROWSKA D. M., **Quantile Regression in Transformation Models**, *J. Indian Soc. Agric. Stat.*, vol. 67, no. 2, p. 153–186, 2005.

KOENKER R. W. and D'OREY V., **Algorithm AS 229**: Computing Regression Quantiles, *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 36, no. 3, p. 383–393, 1987.

LIN J., KEOGH E., WEI L. and LONARDI S., PATEL P., Finding Motifs in Time Series. In: **Proceedings of the 2nd Workshop on Temporal Data Mining**, at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton, Alberta, Canada, p. 23-26, 2002.

CHAKRABARTI K., KEOGH E., MEHROTRA S., and PAZZANI M., **Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases**, *ACM Trans. Database Syst.*, vol. 27, no. 2, p. 188–228, Jun. 2002.

LIN J., KEOGH E., WEI L. and LONARDI S., **Experiencing SAX**: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, vol. 15, p. 107–144, 2007.

BANÑUELOS-RUEDAS F., ANGELES-CAMACHO C., RIOS-MARCUELLO S., **Analysis and validation of the methodology used in the extrapolation of wind speed data at different heights**. *Renewable and Sustainable Energy Reviews*. Vol. 14, no 8, p. 2383–2391, 2010.