

UMA METODOLOGIA PARA FILTRAGEM DE MEDIÇÕES ANEMOMÉTRICAS

José Francisco Moreira Pessanha
Valk Luiz de Oliveira Castellani
Thatiana Conceição Justino
Débora Dias Jardim Penna
Maria Elvira Piñeiro Maceira

Centro de Pesquisas de Energia Elétrica – Eletrobras Cepel
Departamento de Meio Ambiente e Otimização Energética
e-mails: {[francisc](mailto:francisc@cepel.br), [vcastell](mailto:vcastell@cepel.br), [thatiana](mailto:thatiana@cepel.br), [debora](mailto:debora@cepel.br), [elvira](mailto:elvira@cepel.br)}@cepel.br
Av. Horácio Macedo 354, Cidade Universitária
CEP 21941-911, Rio de Janeiro, RJ, Brasil

Resumo- A previsão dos recursos eólicos é um requisito fundamental para a integração segura e econômica dos parques eólicos à operação do sistema elétrico. A precisão da previsão depende da qualidade dos registros anemométricos, em geral caracterizados por lacunas de dados e observações aberrantes decorrentes de falhas no sistema de medição. Tais características comprometem a modelagem da série temporal de velocidade de vento e a precisão das previsões. Neste artigo é apresentada uma metodologia baseada em técnicas estatísticas e de mineração de dados para o preenchimento de lacunas e filtragem dos registros de velocidade do vento. A metodologia proposta tem como objetivo melhorar a qualidade dos dados para a modelagem da velocidade de vento.

Palavras-chave- Energia eólica, velocidade do vento, tratamento de dados, suavização, análise de agrupamentos.

Abstract- The prediction of wind resources has a key role for the safe and economic integration of wind farms in the operation of electrical systems. The accuracy of such predictions depends on the data quality. In general, the wind speed measurements present missing data and outliers resulting from failures in the measuring system. The presence of bad data in wind speed time series contributes to poor specification of the forecasting models and their accuracy. This paper presents a methodology based in statistical and data mining techniques to clean outliers and fill missing values in the historical wind speed records. The aim of the methodology proposed is to improve the data quality for the time series modeling.

Keywords- Wind power, wind speed, data cleaning, smoothing, cluster analysis.

1 Introdução

A integração segura e econômica dos aproveitamentos eólicos ao sistema elétrico requer previsões da disponibilidade dos recursos eólicos desde alguns minutos à frente até previsões horárias com horizontes que variam de uma hora até uma semana à frente. Na formulação de tais previsões pode-se contar com uma ampla variedade de metodologias, desde os tradicionais métodos estatísticos até métodos de inteligência computacional (Wu & Hong, 2007; Bessa et al., 2011; Catalão et al., 2011; Lorenzo et al., 2011). Em função das eventuais falhas nos sistemas de medição, as séries de registros de velocidade de vento podem apresentar lacunas e observações aberrantes (*outliers*), conforme ilustrado na Figura 1. A presença de falhas nas medições compromete o ajuste e o desempenho dos modelos de previsão e, portanto, existe a necessidade de dispor de métodos para filtragem e imputação de dados (Wettayaprasit et al., 2007).

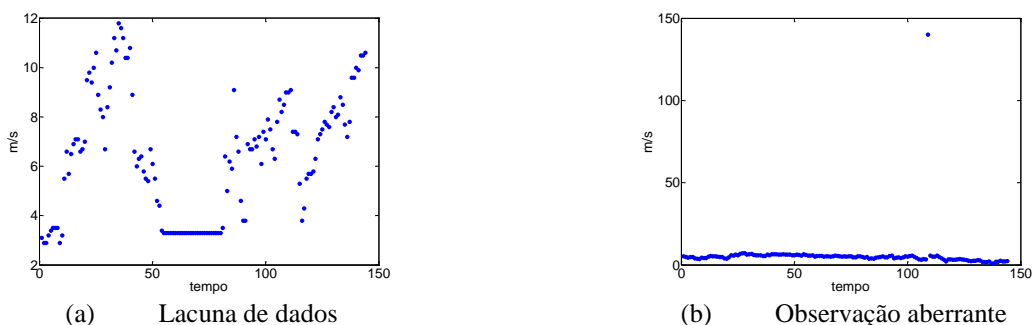


Figura 1: Exemplos de problemas encontrados nos perfis diários de velocidade de vento

Neste artigo apresenta-se uma metodologia para tratamento dos registros de velocidade de vento. A metodologia proposta baseia-se no uso combinado de técnicas para suavização de dados, em particular a regressão linear local ou *Locally Weighed Scatterplot Smoothing* - LOESS (Martinez & Martinez, 2002; Hastie et al., 2009), e a análise de agrupamentos (Jang et al., 1997; Cordón et al., 2001; Tan et al., 2009), conforme ilustrado no diagrama da Figura 2.

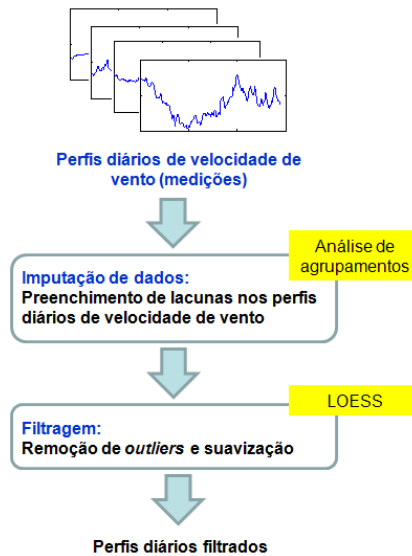


Figura 2: Esquema da metodologia proposta para o tratamento dos registros de velocidade de vento

A estratégia para imputação de dados é descrita a seguir na seção 2. Já a estratégia para remoção dos *outliers* e suavização das medições é apresentada na seção 3. Por fim, na seção 4 são resumidas as principais conclusões do trabalho. Para ilustrar a aplicação da metodologia proposta considerou-se um conjunto de medições de velocidade de vento realizadas em um parque eólico durante o período de 1 de janeiro de 2007 até 31 de dezembro de 2010, perfazendo um total de 1.461 dias. A resolução temporal das medições é de 10 minutos, portanto, cada perfil diário é um vetor com 144 elementos.

2 Imputação de dados de velocidade de vento

Os registros da velocidade de vento referem-se aos valores instantâneos tomados a cada 10 minutos e encontram-se organizados em matrizes, cujas colunas guardam os perfis diários da velocidade de vento (144 valores em cada coluna). Em função de eventuais falhas no sistema de medição, a matriz de dados pode apresentar colunas vazias (sistema de medição inoperante ao longo do dia) ou colunas com algumas lacunas de dados (medições incompletas). Destaca-se que no conjunto de perfis diários analisados há apenas dois dias sem registros de velocidade de vento.

2.1 Imputação de perfis diários nos dias sem medição

Visando minimizar o descarte de dados propõe-se atribuir um perfil diário de velocidade de vento aos dias sem medição. A imputação de um perfil diário toma como base os registros passados da velocidade do vento em um dia com registros completos, cujos dias adjacentes apresentem perfis de velocidade similares aos verificados nos dias anterior e posterior ao dia sem registros. A seguir, na Figura 3, tem-se uma ilustração desta estratégia de imputação de dados. Um exemplo do resultado desta estratégia de imputação de dados é apresentado na Figura 4.

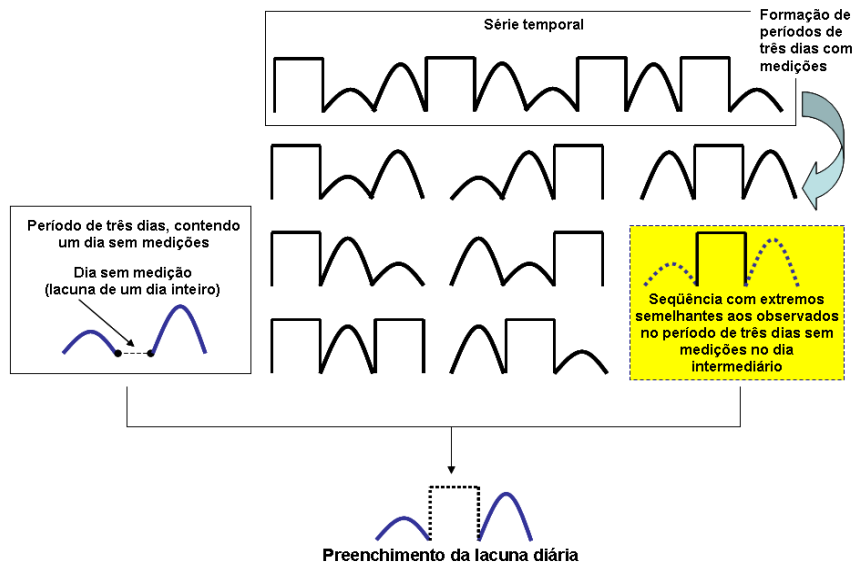


Figura 3: Estratégia para imputação de dados em dias sem medição

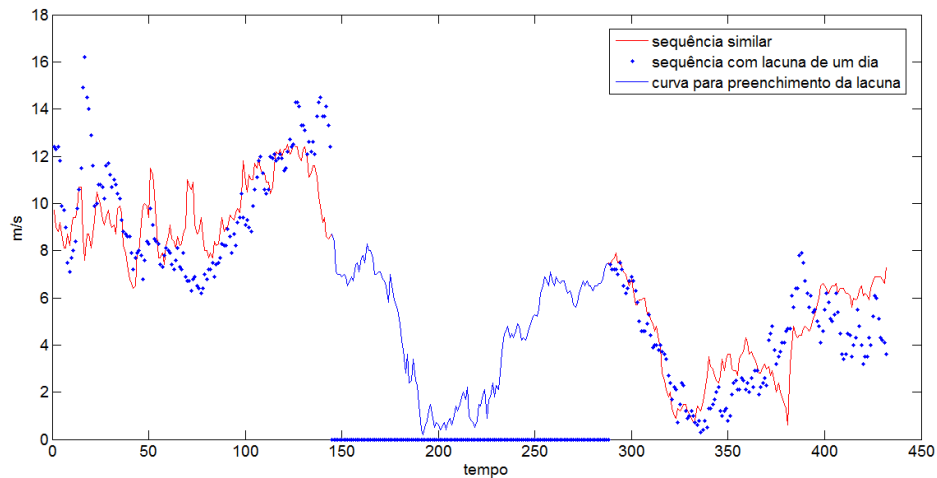


Figura 4: Resultado da imputação de dados para um dia sem medição

2.2 Imputação de dados nos perfis diários incompletos

As lacunas de dados, como as ilustradas na Figura 5, são decorrentes de falhas no sistema de medição e são caracterizadas por seqüências de valores constantes ou quase constantes. A detecção de tais seqüências inicia-se com a normalização do perfil diário de velocidade do vento pela respectiva média diária. Em seguida são calculadas as diferenças de primeira ordem $\Delta v = v(t) - v(t-1)$ em cada perfil diário normalizado $v(t)$, $t=1,144$. Uma seqüência de diferenças nulas indica que os valores de velocidade são constantes e, portanto, o perfil tem uma lacuna de dados. Neste trabalho admitiu-se que diferenças menores que 0,001 são consideradas como sendo nulas. Após alguns testes realizados com os registros históricos considerou-se que as seqüências com 3 ou mais diferenças nulas (um período mínimo de 30 minutos), constituem lacunas de dados. No conjunto de 1.461 perfis diários analisados foram observados 218 perfis com lacunas de dados.

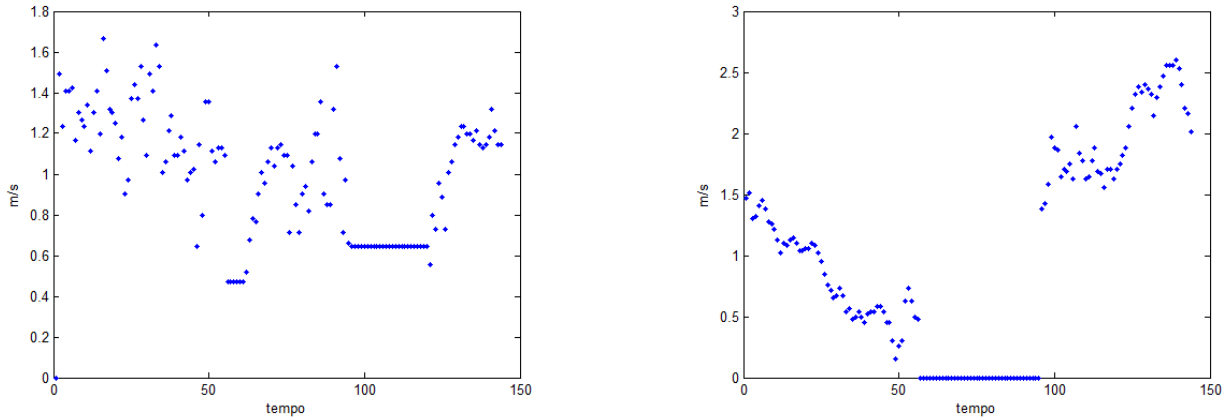


Figura 5: Exemplos de lacunas de dados nos perfis diários (perfis incompletos)

A identificação de regularidades no perfil diário da velocidade do vento permite classificar os perfis em categorias (*clusters*), de tal forma que cada categoria contenha um conjunto de perfis diários semelhantes entre si, porém distintos dos perfis nas demais categorias. Cada categoria é representada por um perfil típico definido pela média dos perfis nela classificados. Em função da regularidade da velocidade do vento, o perfil típico é uma informação valiosa no preenchimento das lacunas de dados em perfis diários.

A classificação automática dos perfis diários pode ser realizada por um algoritmo de análise de agrupamentos (*cluster analysis*). Há uma variedade de métodos para análise de agrupamentos, por exemplo, o K-Means (Tan et al., 2009) e os mapas auto-organizáveis (Braga et al., 2007), porém em função da qualidade dos dados analisados e pela facilidade de implementação optou-se pelo FCM (Jang et al., 1997; Cordón et al., 2001).

Assim, propõe-se preencher as lacunas de dados com segmentos extraídos dos perfis típicos identificados por um algoritmo de análise de agrupamentos. Naturalmente, a obtenção de perfis típicos que sejam úteis no preenchimento de lacunas de dados requer que o algoritmo de análise de agrupamentos seja aplicado em um conjunto de perfis diários sem lacunas. Neste trabalho, o FCM foi aplicado em um conjunto de 1.234 perfis diários. No FCM a identificação de uma estrutura natural de agrupamentos dos perfis diários passa pela solução do seguinte problema de programação matemática:

$$\begin{aligned} \text{Min}_{u_{ij}, c_j, \lambda_i} \quad & \sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - c_j\| + \sum_{i=1}^n \lambda_i \left(\sum_{j=1}^k u_{ij} - 1 \right) \\ \text{s.a.} \quad & \sum_{j=1}^k u_{1j} = 1 \\ & \dots \\ & \sum_{j=1}^k u_{nj} = 1 \end{aligned} \quad (1)$$

onde n é o total de perfis na amostra, x_i denota o vetor contendo os 144 pontos do i -ésimo perfil diário de velocidade do vento ($i=1, n$), m é a constante de *fuzzyficação* (em geral $m=1,25$ ou $m=2$), k é o número de agrupamentos (*clusters*), c_j é o vetor referente ao centróide do j -ésimo *cluster* e u_{ij} é o grau de pertinência do i -ésimo perfil diário no j -ésimo *cluster* ($0 \leq u_{ij} \leq 1$), se $u_{ij}=0$ o i -ésimo perfil não pertence ao j -ésimo *cluster* e se $u_{ij} = 1$ então o i -ésimo perfil pertence somente ao j -ésimo *cluster*.

Com o auxílio da função Lagrangeana, o problema de otimização em (1) pode ser escrito como:

$$\text{Min}_{u_{ij}, c_j, \lambda_i} \quad \sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - c_j\| + \sum_{i=1}^n \lambda_i \left(\sum_{j=1}^k u_{ij} - 1 \right) \quad (2)$$

Na função objetivo (2), λ_j , $j=1,n$ denotam os multiplicadores de Lagrange para as n restrições de igualdade. A partir da equação (2) são obtidas as seguintes condições de otimalidade:

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{t=1}^k \left(\frac{\|x_i - c_j\|}{\|x_i - c_t\|} \right)^{\frac{2}{m-1}}} \quad (4)$$

Usando as equações (3) e (4), pode-se implementar o algoritmo a seguir, onde a matriz U formada pelos elementos u_{ij} ($i=1,n$ e $j=1,k$) e os centróides dos k clusters ($c_j, j=1,k$) são obtidos iterativamente:

Passo 1: Inicialize a matriz U com valores entre 0 e 1, observando que, em cada linha da matriz, a soma dos valores deve ser igual a unidade. Esta etapa é conhecida como *fuzzyficação*.

Passo 2: Use a equação (3) para calcular as coordenadas dos k centróides

Passo 3: Calcule a função objetivo $\sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - c_j\|^2$. Pare se o valor da função objetivo estiver abaixo de uma tolerância ou se a melhoria em relação à iteração anterior for desprezível.

Passo 4: Use a equação (4) para atualizar os elementos da matriz U e volte para o passo 2.

Após a convergência do algoritmo, os perfis de velocidade do vento são alocados nos *clusters* onde apresentam maior grau de pertinência (um procedimento conhecido como *defuzzyficação* pelo máximo). Os perfis típicos são os k centróides, ou seja, as médias dos perfis normalizadas em cada *cluster*. Neste trabalho, inicialmente foi considerado um total de 100 agrupamentos, mas após o processo de *defuzzyficação* pelo máximo restaram somente 90 agrupamentos. Na Figura 6 são apresentados dois agrupamentos identificados pelo método FCM e os respectivos perfis típicos (curva em destaque).

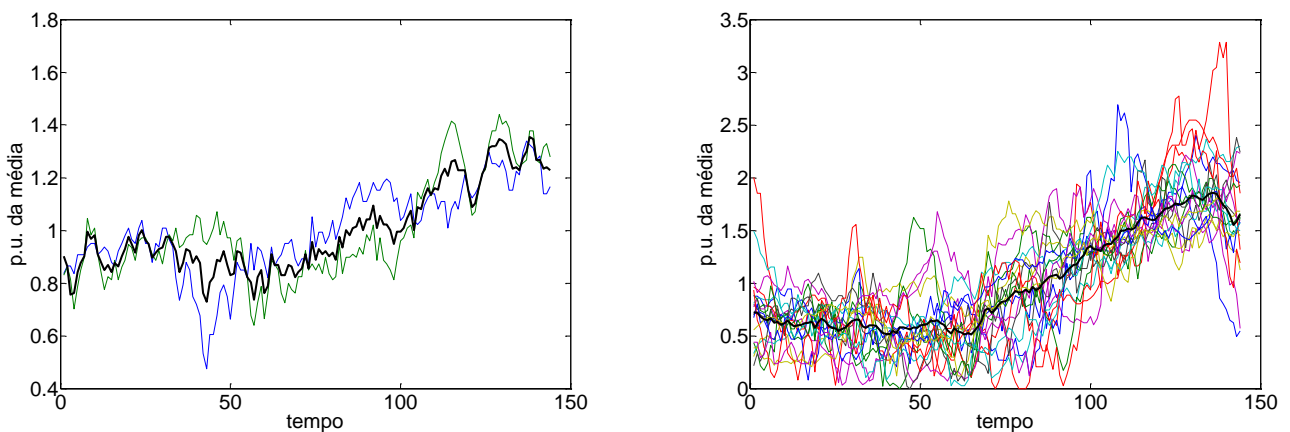


Figura 6: Exemplos de agrupamentos e perfis típicos identificados pelo FCM

A imputação de dados consiste em preencher as lacunas de um perfil diário com as estimativas determinadas pelo produto entre a velocidade média do perfil diário analisado e o perfil típico normalizado que lhe é mais similar. Alguns exemplos dos resultados obtidos por este procedimento são ilustrados na Figura 7, onde são exibidas as lacunas de dados e as respectivas propostas de preenchimento.

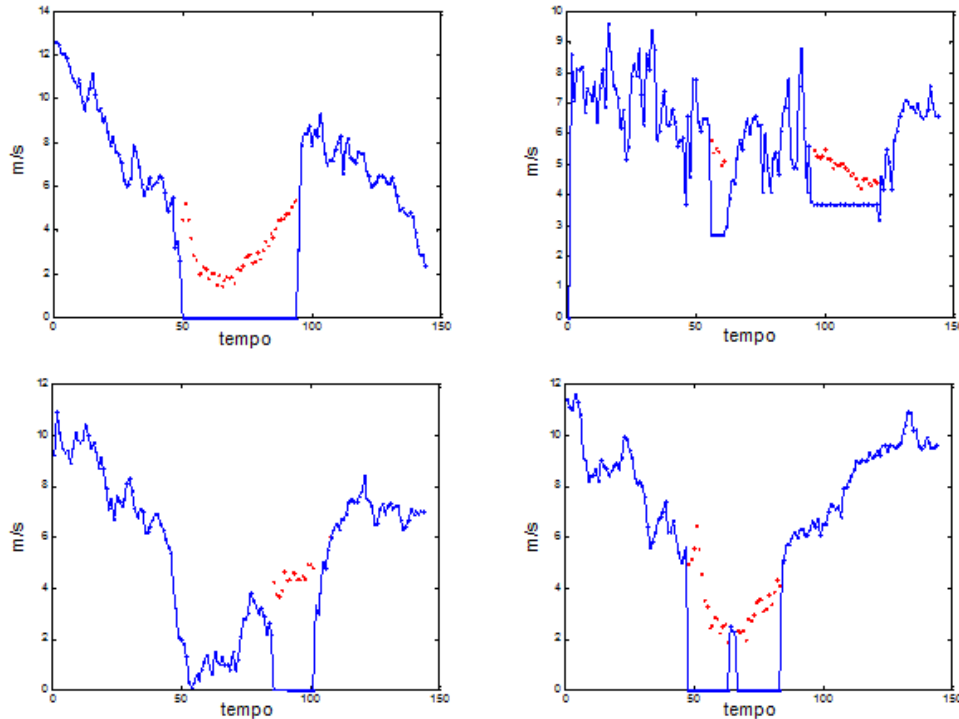


Figura 7: Perfis diários observados (em azul) e propostas de preenchimento das lacunas (em vermelho)

3 Filtragem de dados de velocidade

Os *outliers* são produzidos por erros no sistema de medição e para identificá-los propõem-se a construção de uma espécie de carta de controle para cada perfil diário, em que os *outliers* são as observações situadas do lado de fora da região delimitada pelos intervalos de confiança.

Para construir os intervalos de confiança deve-se primeiro obter uma versão suavizada do perfil diário. A regressão linear local LOESS é uma forma prática de fazer esta suavização (Martinez & Martinez, 2002; Hastie et al., 2009), conforme o algoritmo descrito a seguir:

Passo 1: Seja x o vetor com os instantes das observações e y o perfil diário da velocidade de vento.

Passo 2: Informe o tamanho k da janela de tempo. Neste trabalho foi adotado $k = 10$.

Passo 3: Para cada instante x_0 identifique os k instantes x_i ($i=1,k$) na vizinhança de x_0 e denote este conjunto por $N(x_0)$.

$$\Delta(x_0) = \text{máximo}_{x_i \in N(x_0)} \|x_0 - x_i\| \quad (5)$$

Passo 4: Calcule a maior distância entre x_0 e o ponto x_i dentro da janela $N(x_0)$.

Passo 5: Pondere cada par (x_i, y_i) , $x_i \in N(x_0)$ com base na seguinte função:

$$peso_i(x_0) = W\left(\frac{\|x_0 - x_i\|}{\Delta(x_0)}\right), \text{ onde } W(u) = \begin{cases} (1-u^3)^3 & 0 \leq u \leq 1 \\ 0 & \text{caso contrário} \end{cases} \quad (6)$$

Passo 6: Aplique mínimos quadrados ponderados para obter uma estimativa \hat{y} para y no ponto x_0 ajustado ao conjunto de observações que pertencem à vizinhança $N(x_0)$.

Passo 7: Repita os passos de 3 a 6 para cada instante de tempo no vetor x .

A curva \hat{y} obtida pela aplicação da suavização LOESS é uma referência para comparação com o perfil de velocidade $v(t)$, conforme ilustrado na Figura 8.

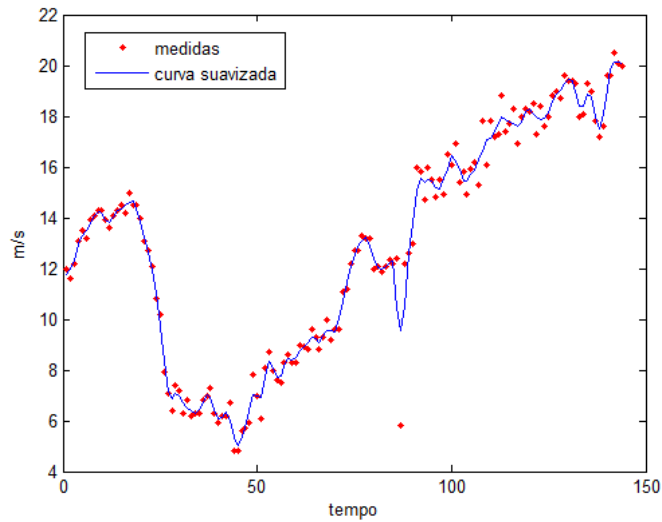


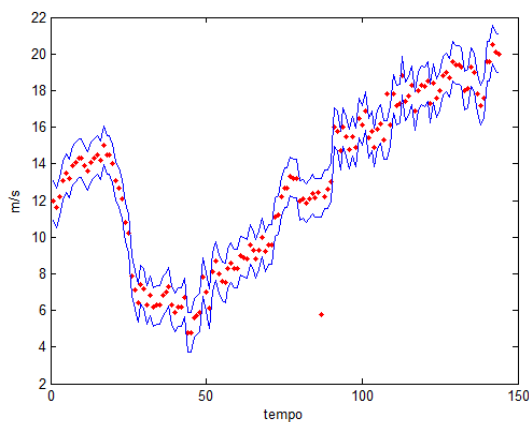
Figura 8: Perfil diário medido e perfil suavizado

Na sequência é calculado o desvio padrão (DP) dos resíduos entre o perfil diário e a sua versão suavizada e definidos os limites dos intervalos de confiança em cada instante:

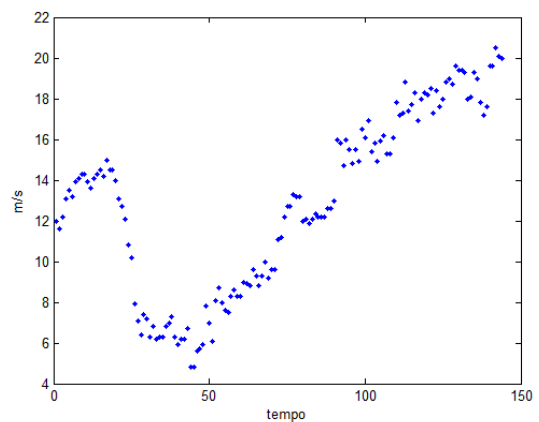
$$\text{Limite superior: } LS(t) = \hat{y}(t) + 3.5 \text{ DP} \quad (7)$$

$$\text{Limite inferior: } LI(t) = \hat{y}(t) - 3.5 \text{ DP} \quad (8)$$

As observações aberrantes (*outliers*) são os registros localizados fora dos limites definidos pelo intervalo de confiança, conforme ilustrado na Figura 9. A filtragem dos dados consiste em substituir os *outliers* pelas estimativas determinadas pela curva suavizada obtida pela LOESS.



(a) perfil medido e intervalos de confiança



(b) perfil diário após a filtragem

Figura 9: Perfil diário antes e após a remoção dos *outliers*

Por fim, os perfis diários filtrados são empilhados para formar uma única série histórica e ao final o método LOESS é aplicado novamente para suavizar toda a série. O efeito do processo de filtragem pode ser visualizado na Figura 10, na qual são apresentadas as séries de velocidade de vento antes e após a filtragem. Os histogramas na Figura 11 também indicam os bons

resultados alcançados pela filtragem, pois a distribuição da velocidade do vento após a filtragem é aderente à distribuição de Weibull, o modelo de probabilidade mais apropriado para descrever a velocidade do vento (Custódio, 2009).

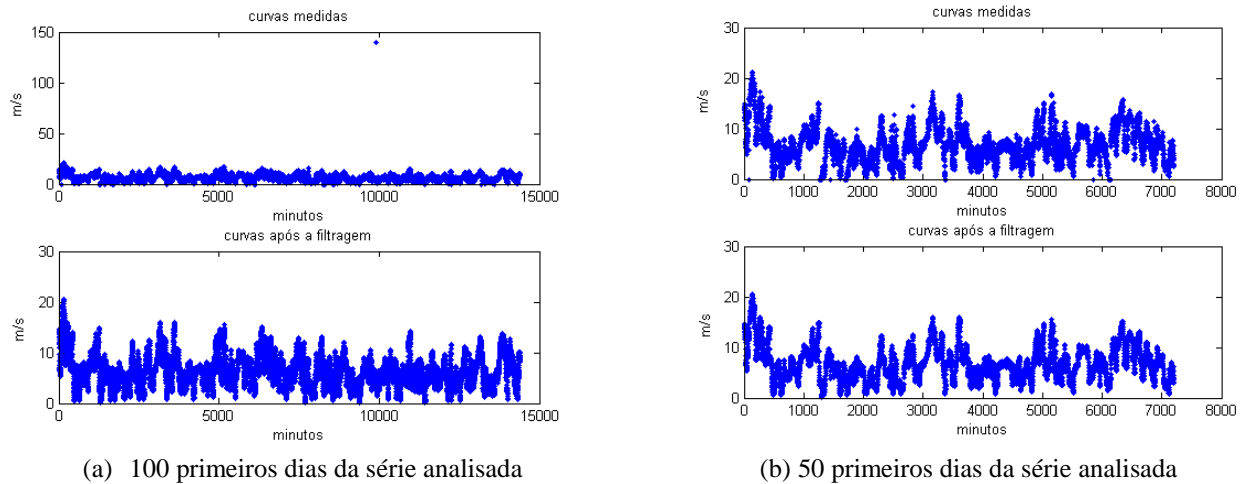


Figura 10: Séries de velocidade do vento antes e após a filtragem

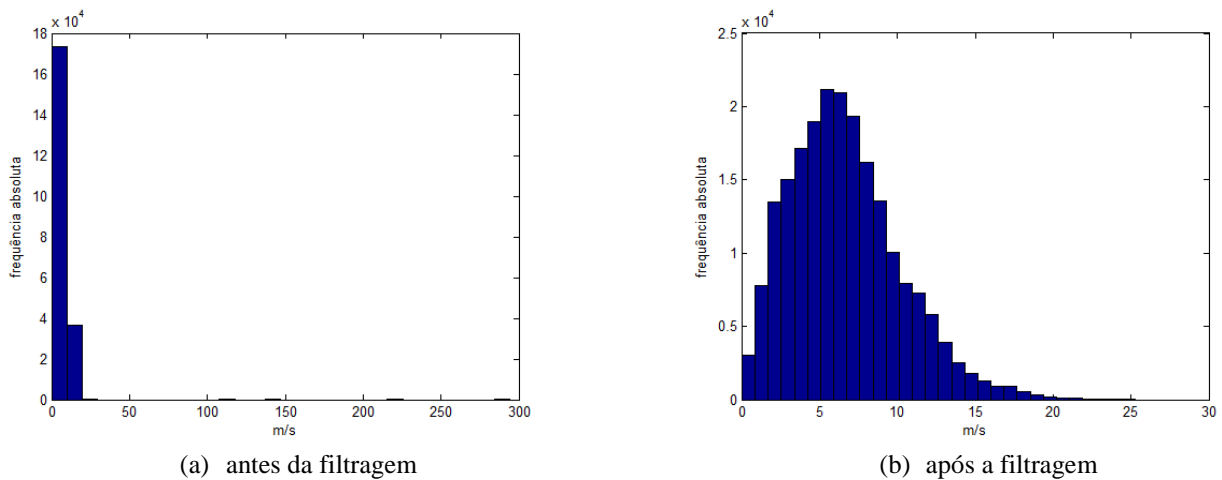


Figura 11: Histograma da velocidade do vento antes e após a filtragem

4 Conclusões

Em função de falhas no sistema de medição os registros anemométricos apresentam erros como dados aberrantes, descontinuidades e lacunas. A introdução destes dados sem um tratamento estatístico prévio compromete o ajuste dos modelos de previsão e, portanto, implica na perda da precisão das previsões de velocidade do vento. O presente artigo apresentou uma metodologia baseada em análise de agrupamentos e regressão linear local (LOESS) para a imputação e filtragem de registros de velocidade do vento. Os resultados obtidos são satisfatórios e mostram o potencial da metodologia proposta. A metodologia proposta constitui uma importante alternativa ao critério de Chauvenet, usualmente empregado no descarte de registros anemométricos inconsistentes.

5 Referências

- Bessa, R. J., Sumaili, J., Miranda, V., Botterud, A., Wang, J., Constantinescu, E. (2011). Time-adaptative kernel density forecast: a new method for wind power uncertainty modeling, 17th Power Systems Computation Conference, Stockholm, 22-26, August.
- Braga, A. P., Carvalho, A. P. L. F., Ludermit, T. B. (2007). Redes neurais artificiais: teoria e aplicações, 2 ed., LTC.
- Catalão, J. P. S., Pousinho, H. M. I., Mendes, V. M. F. (2011). New hybrid intelligent approach to forecast wind power and electricity prices in the short-term, 17th Power Systems Computation Conference, Stockholm, 22-26, August.

- Cordón, O., Herrera, F., Hoffmann, F., Magdalena, L. (2001). Genetic fuzzy systems evolutionary tuning and learning of fuzzy knowledge bases, *Advances in Fuzzy Systems - Applications and Theory*, vol. 19, World Scientific Publishing.
- Costa, A. F. B., Epprecht, E. K., Carpinelli, L. C. R. (2005). *Controle estatístico de qualidade*, 2 edição, Editora Atlas, São Paulo.
- Custódio, R. S. (2009). *Energia eólica para a produção de energia elétrica*, Eletrobrás, Rio de Janeiro.
- Jang, J. S. R., Sun, C. T., Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall Inc.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction*, second edition, Springer.
- Lorenzo, J., Mendez, J., Castrillon, M., Hernandez, D. (2011). Short-term wind Power forecast based on cluster analysis and artificial neural network, 11th International Work-Conference on Artificial Neural Networks, Torremolinos, Málaga, Spain, 8-10, June.
- Martinez, W. L., Martinez, A. R. (2002). *Computational statistics handbook with matlab*, Chapman & Hall/CRC.
- Tan, P. N., Steinbach, M., Kumar, V. (2009). *Introdução ao datamining mineração de dados*, Editora Ciência Moderna.
- Wettayaprasit, W., Laosen, N., Chevakidagarn, S. (2007). Data filtering technique for neural network forecasting, 7th WSEAS International Conference on Simulation, Modeling and Optimization, Beijing, China, September, 15-17.
- Wu, Y. K., Hong, J. S. (2007). A literature review of wind forecasting technology in the world, *Power Tech*, Lausanne, Switzerland, 1-5, July.