

A GA-Based Approach for Building Regularized Sparse Polynomial Models for Wind Turbine Power Curves

Haroldo C. Maya, Guilherme A. Barreto

¹Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI)
Universidade Federal do Ceará (UFC), Campus do Pici, Bloco 725, Fortaleza-CE

mayaharoldo@gmail.com, gbarreto@ufc.br

Abstract. *In this paper, the classical polynomial model for wind turbines power curve estimation is revisited aiming at an automatic and parsimonious design. In this regard, using genetic algorithms we introduce a methodology for estimating a suitable order for the polynomial as well its relevant terms. The proposed methodology is compared with the state of the art in estimating the power curve of wind turbines, such as logistic models (with 4 and 5 parameters), artificial neural networks and weighted polynomial regression. We also show that the proposed approach performs better than the standard LASSO approach for building regularized sparse models. The results indicate that the proposed methodology consistently outperforms all the evaluated alternative methods.*

1. Introdução

Com o crescimento da indústria eólica, aerogeradores estão sendo instalados em diversos tipos de ambientes que se diferem, por exemplo, em relação à umidade relativa do ar, densidade do ar, altitude, condições climáticas sazonais e complexidade do terreno. Estas características, além de outras não mencionadas e outras nem conhecidas, influenciam significativamente a produção energética local. Em face de tais dificuldades, faz-se necessário uma modelagem adequada da relação velocidade do vento versus potência gerada (conhecida genericamente como *curva de potência*) pelos aerogeradores em certo parque eólico, com o intuito de prever corretamente sua produção energética, bem como o monitorar o seu desempenho.

Dada sua importância, a modelagem (ou estimação) de curvas de potência é um tema que sempre despertou (e ainda desperta) o interesse da comunidade de energias renováveis haja visto o grande número de artigos disponíveis na literatura especializada, principalmente na forma de surveys [Marciukaitis et al. 2017, Wang et al. 2018, Sohoni et al. 2016, Lee et al. 2015, Shokrzadeh et al. 2014, Lydia et al. 2014, Clifton et al. 2013, Lydia et al. 2013, Li et al. 2001]. Dentre os métodos encontrados na literatura, destacam-se o modelo polinomial como um dos mais utilizados, o polinomial ponderado, o modelo logístico de 4 e 5 parâmetros e redes neurais artificiais. Em particular, o modelo polinomial tem seu uso disseminado entre os praticantes da área devido não somente à sua formulação mais simples quando comparado a modelos mais complexos, tais como redes neurais, mas principalmente porque encontra-se disponível em planilhas de cálculo populares (e.g. Excel e LibreOffice Calc).

Mesmo sendo de uso mais simples, o modelo polinomial requer a especificação *a priori* da ordem do polinômio. A especificação correta da ordem do polinômio é essencial para um bom desempenho preditivo do modelo, mas também para evitar problemas

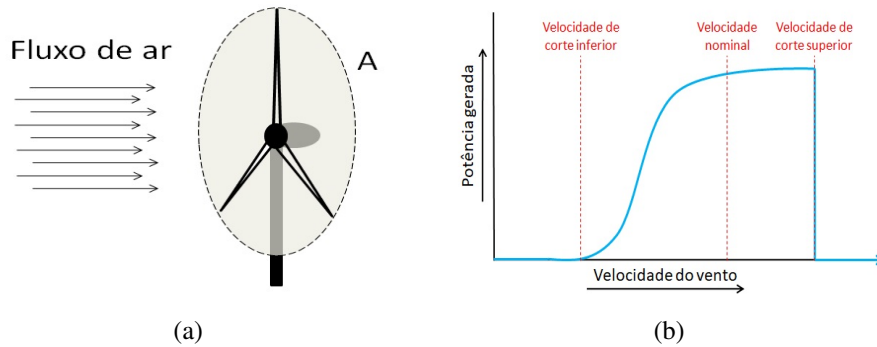


Figure 1. (a) Fluxo de ar pela área circular de varredura das pás do aerogerador. (b) Curva de potência típica de um aerogerador com controle do ângulo de passo.

numéricos. Uma vez escolhida a ordem, um método de estimação de parâmetros (geralmente, mínimos quadrados ordinários) é usado para estimar os coeficientes associados a cada termo do polinômio. No modelo polinomial clássico, o número de coeficientes a estimar é igual à ordem do polinômio mais 1.

Tendo em mente as dificuldades supracitadas, revisitaremos o modelo polinomial convencional a fim de propor uma metodologia de projeto baseada em computação evolucionária que retorna a ordem ótima do modelo polinomial, bem como seleciona os termos relevantes do polinômio. A metodologia proposta é avaliada em um amplo estudo usando dados reais e comparada com métodos que compõem o estado da arte em estimação da curva de potência de aerogeradores. Uma comparação com uma abordagem clássica na estimação de modelos esparsos regularizados também é realizada. Os resultados indicam um desempenho consistentemente superior da metodologia proposta em relação a todas as alternativas avaliadas.

O restante do artigo está organizado como segue. Na Seção 2, os princípios físicos que regem o comportamento da curva de potência de um aerogerador são brevemente apresentados. Todos os modelos usados neste artigo para estimação da curva de potência de aerogeradores são discutidos na Seção 3. A metodologia proposta é introduzida na Seção 4 em detalhes. Os resultados são apresentados e discutidos na Seção 5. O artigo é concluído na Seção 6.

2. Fundamentos da Curva de Potência

Seja um aerogerador de eixo horizontal qualquer. Denotaremos por A a área circular (em m^2) varrida por suas pás, tal como esboçado na Figura 1(a). Considere que o ar do local onde o aerogerador está instalado tenha densidade ρ constante (em kg/m^3) e que a componente de velocidade do vento (m/s) que atravessar perpendicularmente a área A é constante e igual a v . Assim, sabe-se que a potência P (em Watts) é calculada como

$$P = \frac{\Delta E_c}{\Delta t}, \quad (1)$$

em que t é a medida do tempo (em segundos) e $E_c = \frac{1}{2}Mv^2$ é a energia cinética do movimento (em Joules), com M denotando a massa (em kg) e v a velocidade. Para v

constante, pode-se escrever a variação da energia cinética como

$$\Delta E_c = \frac{1}{2} \Delta M v^2, \quad (2)$$

tal que a potência eólica disponível associada a v em uma área A , denotada doravante por P_w , equivale a energia cinética por unidade de tempo, ou seja

$$P_w = \frac{\Delta E_c}{\Delta t} = \frac{1}{2} \frac{\Delta M}{\Delta t} v^2 = \frac{1}{2} (\rho A v) v^2 = \frac{1}{2} \rho A v^3. \quad (3)$$

Porém, apenas uma parte da potência eólica disponível será aproveitada para o movimento do rotor, sendo chamada de potência absorvida (P_a). O coeficiente de potência $C_P = P_a/P_w$ é uma grandeza adimensional que caracteriza a capacidade de um aerogerador absorver a potência eólica disponível [Shokrzadeh et al. 2014, Clifton et al. 2013]. O valor do coeficiente de potência, que pode ser obtido a partir dos dados do fabricante, é influenciado por aspectos morfológicos e aerodinâmicos do aerogerador, como por exemplo o desenho de suas pás. O seu valor máximo teórico, conhecido como limite de Betz [Betz 1920], é de $16/27 \approx 0,5926$. Ademais, ainda deve-se contar com a eficiência do rotor, η , para se obter a potência elétrica gerada a partir do fluxo de vento, P , chegando assim à seguinte equação teórica:

$$P = \frac{1}{2} \eta C_p \rho A v^3. \quad (4)$$

Turbinas reais são projetadas para fornecer, no máximo, a sua potência nominal, podendo sofrer avarias caso a exceda. Sendo assim, a potência elétrica fornecida tem comportamento não linear proporcional ao cubo da velocidade do vento, semelhante ao desenvolvimento teórico previamente apresentado, para baixas velocidades. Porém, a medida que se aproxima da potência nominal, os aerogeradores comerciais dispõem de meios para limitar a energia eólica absorvida e, conseqüentemente, controlar a potência fornecida não exceda a potência nominal do rotor [Manwell et al. 2009].

Há vários métodos que visam controlar a potência absorvida pelo rotor, dentre os quais destacamos dois como os principais utilizados atualmente: (i) *controle do ângulo de passo*, que consiste no controle ativo do ângulo de ataque das pás do aerogerador. Dessa forma pode-se alterar o valor do seu C_p , que influencia na potência eólica absorvida; e (ii) *controle por estol*, que é um tipo de controle passivo, que consiste em projetar as pás de forma que o efeito de estol¹ ocorra em velocidades superiores à velocidade eólica nominal.

Além disso, há uma velocidade de vento mínima para o funcionamento do aerogerador, conhecida como velocidade de corte inferior (*cut-in speed*), na qual a potência gerada é nula para velocidades iguais ou inferiores. Há também um limite superior, conhecido como velocidade de corte superior (*cut-out speed*), em que o funcionamento do aerogerador é interrompido com a finalidade de proteger a integridade de seus componentes por conta da alta intensidade do vento [Sohoni et al. 2016]. Um gráfico típico da curva de potência de um aerogerador comercial com controle de *pitch* está mostrado na Figura 1(b), em que fica evidenciado o comportamento não linear da curva de potência.

¹ Este efeito ocorre quando o fluxo de ar na superfície da pá deixa, abruptamente, de fixar-se à superfície, passando a girar de forma turbulenta, diminuindo assim a força de sustentação e aumentando a força de arrasto.

3. Estimação da Curva de Potência

Nesta seção são apresentados os modelos a serem avaliados neste artigo, começando pelo modelo polinomial clássico e o respectivo estimador de mínimos quadrados ordinários (*ordinary least squares*, OLS) e uma versão ponderada deste estimador, passando pelos modelos logísticos de 4 e 5 parâmetros, chegando ao modelo baseado em redes neurais artificiais.

3.1. Modelo de regressão polinomial

Regressão polinomial muito provavelmente é a técnica mais utilizada na prática para estimar a curva de potência de aerogeradores [Lydia et al. 2014, Shokrzadeh et al. 2014]. Este modelo tenta ajustar um polinômio de ordem k a um conjunto de n observações do par entrada-saída $\{(p_i, v_i)\}_{i=1}^n$, sendo a p_i a i -ésima medição da potência elétrica gerada e v_i é a velocidade do vento correspondente.

Do ponto de vista da modelagem estocástica, o ajuste do polinômio aos dados assume que o processo gerador dos dados pode ser descrito pela seguinte expressão:

$$p_i = \beta_0 + \beta_1 v_i + \beta_2 v_i^2 + \dots + \beta_k v_i^k + \epsilon_i, \quad (5)$$

em que β_j , $j = 0 \dots l$, são os coeficientes do polinômio e ϵ_i é uma variável aleatória normal, com média nula e variância σ_ϵ^2 . Pode-se escrever a Eq. 5 na forma matricial como

$$\mathbf{p} = \mathbf{V}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (6)$$

em que $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_n]^T$, $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \dots \ \beta_k]^T$, $\boldsymbol{\epsilon} = [\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_n]^T$, com a matriz \mathbf{V} , de dimensão $n \times (k + 1)$, sendo definida como

$$\mathbf{V} = \begin{bmatrix} 1 & v_1 & \dots & v_1^k \\ 1 & v_2 & \dots & v_2^k \\ \vdots & \vdots & & \vdots \\ 1 & v_n & \dots & v_n^k \end{bmatrix}.$$

A estimativa de $\boldsymbol{\beta}$ é obtida via método OLS, a partir da minimização do funcional $J(\boldsymbol{\beta}) = \|\boldsymbol{\epsilon}\|^2 = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \sum_{i=1}^n \epsilon_i^2$, cuja derivada em relação a $\boldsymbol{\beta}$ é dada por

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial \left[(\mathbf{p} - \mathbf{V}\boldsymbol{\beta})^T (\mathbf{p} - \mathbf{V}\boldsymbol{\beta}) \right]}{\partial \boldsymbol{\beta}} = -2\mathbf{V}^T (\mathbf{p} - \mathbf{V}\boldsymbol{\beta}). \quad (7)$$

Igualando o resultado ao vetor nulo de dimensão $k + 1$ chega-se à formula do estimador OLS para os parâmetros do modelo polinomial:

$$\hat{\boldsymbol{\beta}} = (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T \mathbf{p}. \quad (8)$$

A matriz \mathbf{V} é uma matriz de Vandermonde. Isto faz com que uma escolha inadequada da ordem do polinômio (e.g. $k > 8$) produza valores elevados para as algumas de suas colunas. Por exemplo, com valores típicos de velocidade do vento na faixa de 4 a 12 m/s, este fato conduz em geral a erros numéricos durante a inversão existente na Eq. (8). A metodologia proposta tem a propriedade inerente de evitar a ocorrência de tais problemas numéricos causados por valores elevados da ordem do polinômio.

3.2. Modelo polinomial ponderado

Os dados para estimação de curva de potência costumam ser obtidos via sistemas SCADA², estando assim susceptíveis a erros de medição, falha em sensores e erros no sistemas de comunicação [Sohoni et al. 2016]. Os dados também são afetados por aerogeradores que não estão em produção pois foram desligadas pelo sistema de controle por outra razão que não a operação anômala. Assim, sistemas SCADA podem ter entradas nulas ou dados errôneos, denominados genericamente de *outliers*, resultando em modelos inexatos da curva de potência. Pode-se optar por uma inspeção visual dos dados e a remoção manual de observações supostamente anômalas antes de realizar a estimação de parâmetros do modelo polinomial.

Uma alternativa é usar alguma técnica de estimação de parâmetros que seja robusta a *outliers*, tal como *estimação-M* [Huber 1964]. Grosso modo, esta técnica consiste em reescrever o funcional do método OLS em uma forma mais geral dada por $\sum_{i=1}^n w_i^2(\epsilon_i)\epsilon_i^2$, em que $w_i(\epsilon)$ é o peso associado ao erro ϵ_i . Note, porém, que os pesos dependem dos erros (ou resíduos), os resíduos dependem dos coeficientes estimados, e os coeficientes estimados dependem dos pesos. Esta dependência faz com que um método iterativo seja necessário para estimar os parâmetros, sendo o *iteratively reweighted least-squares*³ (IRLS) o mais usado para este fim.

Há várias funções que podem ser usadas para especificar os pesos, tal como a função de Huber:

$$w(\epsilon_i) = \begin{cases} 1, & |\epsilon_i| \leq k \\ \frac{k}{|\epsilon_i|}, & |\epsilon_i| > k \end{cases} \quad (9)$$

em que k é um limiar acima do qual o valor absoluto do erro é considerado como sendo causado por um *outlier*. Enquanto o valor absoluto do erro estiver abaixo do limiar k , o seu peso tem valor igual a 1 e a função-custo comporta-se como a função do método OLS. Os pesos decrescem de valor à medida que a amplitude dos erros aumentam. Quanto maior o erro, menor o peso e, conseqüentemente, menor sua influência na função-custo.

3.3. Rede Perceptron Multicamadas (MLP)

A rede MLP(p, q, c) é uma rede neural com p variáveis de entrada, uma camada oculta com q neurônios e uma camada de saída com c neurônios. As funções de ativação dos neurônios ocultos devem ser não-lineares (e.g. sigmoidais). Neurônios de saída pode usar tanto funções sigmoidais quanto lineares, a depender da aplicação de interesse. Em classificação de padrões geralmente funções sigmoidais são usadas, enquanto funções lineares são usadas em problemas de regressão ou ajuste de curvas.

A saída da rede MLP para um único neurônio linear (i.e. $c = 1$) é dada por

$$y_n = \mathbf{w}^T \mathbf{h}_n + \theta, \quad (10)$$

onde $\mathbf{w} = [w_1 \cdots w_q]^T$ é o vetor⁴ de pesos (incluindo o limiar γ) que conecta os neurônios ocultos ao neurônio de saída e θ é o limiar deste neurônio. O vetor $\mathbf{h}_n \in \mathbb{R}^q$, que contém

²Supervisory control and data acquisition.

³Este método é usado na função `robustfit` implementada no Matlab®.

⁴Assumimos que todos os vetores são vetores-coluna, salvo indicação contrária.

as saídas atuais dos neurônios ocultos, é calculado como

$$\mathbf{h}_n = [\phi(\mathbf{m}_1^T \mathbf{x}_n + b_1), \dots, \phi(\mathbf{m}_q^T \mathbf{x}_n + b_q)]^T, \quad (11)$$

tal que $\mathbf{x}_n = [x_1 \cdots x_p]^T$ é o vetor de entrada atual, $\mathbf{m}_j = [m_{j1} \cdots m_{jq}]^T$ é o vetor de pesos do j -ésimo neurônio oculto, b_j denota seu limiar, e $\phi(\cdot)$ é a função de ativação não-linear associada.

Visto que estamos lidando com um problema de uma única entrada (velocidade do vento, v_n) e uma única saída (potência gerada, p_n), é preciso treinar apenas uma rede MLP(1, q , 1), i.e. uma rede com uma única entrada (além do limiar) e um neurônio de saída. Neste caso, $y_n = \hat{p}_n$ e $\mathbf{x}_n = [-1 \ v_n]^T$, com \hat{p}_n denotando a potência predita para a velocidade do vento atual v_n . Neste artigo, os pesos e limiares da rede MLP aplicada ao problema de estimação da curva de potência de aerogeradores são computados via algoritmo de Levenberg-Marquardt [Hagan and Menhaj 1994].

3.4. Modelos logísticos de 4 e 5 parâmetros

Modelos logísticos são funções sigmoidais (i.e. em forma de S) que possuem parâmetros para ajustar o seu formato, sendo comumente usadas na aproximação da curva de potência de aerogeradores [Sohoni et al. 2016, Lydia et al. 2014]. Neste trabalho iremos comparar o desempenho de dois modelos logísticos, cujos parâmetros serão ajustados por meio da técnica de otimização evolucionária conhecida como *evolução diferencial* [Storn and Price 1997].

Para o modelo logístico de 4 parâmetros (4PL), a potência predita para um dado valor da velocidade do vento (v_i), é definido pela seguinte expressão:

$$p_i = a \left(\frac{1 + m e^{-\frac{v_i}{\tau}}}{1 + n e^{-\frac{v_i}{\tau}}} \right), \quad (12)$$

em que definimos $\beta = [a \ m \ n \ \tau]^T$ como o seu vetor de parâmetros. Não há nenhuma restrição para os valores de β . No caso do modelo logístico de 5 parâmetros (5PL), temos a seguinte expressão:

$$p_i = d + \frac{a - d}{\left[1 + \left(\frac{v_i}{c} \right)^b \right]^g}, \quad (13)$$

em que $\beta = [a \ b \ c \ d \ g]^T$ é o vetor de parâmetros correspondentes. Tem-se agora as seguintes restrições: $c, g > 0$.

Ajustaremos os parâmetros de cada modelo logístico utilizando um algoritmo de evolução diferencial (DE) [Storn and Price 1997] com as seguintes características: (i) DE/best/1; (ii) critério de parada ocorre quando a aptidão for a mesma ao longo de 200 gerações seguidas; (iii) população de tamanho 150; (iv) fator de mutação $F = 0,9$; (v) taxa de cruzamento $CR = 0,8$; (vi) número de execuções = 1.

3.5. O Método LASSO

Considere a função de penalização para valores elevados do vetor de parâmetros β na forma $[\ell_d(\beta)]^d$, $d \in \mathbb{N}^+$. Assim, uma regressão por mínimos quadrados regularizado pela com penalidade ℓ_d tem como objetivo minimizar o seguinte funcional

$$J(\beta) = (\mathbf{p} - \mathbf{V}\beta)^T (\mathbf{p} - \mathbf{V}\beta) + \lambda [\ell_d(\beta)]^d. \quad (14)$$

em que $\lambda \geq 0$ é uma constante usada para controlar a intensidade da penalização. Para $d = 1$, tem-se a função de penalização dada por

$$\lambda [\ell_1(\boldsymbol{\beta})] = \lambda \sum_{i=1}^n |\beta_i|, \quad (15)$$

que não é diferenciável quando seu argumento é zero. Como a derivada da função é constante em valor absoluto (com exceção, como supracitado, quando seu argumento for zero), tal função de penalização tem a tendência de reduzir o valor dos coeficientes com a mesma intensidade até zerá-los. Isso resulta em uma solução esparsa para o vetor de coeficientes $\boldsymbol{\beta}$; ou seja, alguns coeficientes menos relevantes para a regressão serão zerados. A estimativa de parâmetros por mínimos quadrados com penalização ℓ_1 tem sido popularizada sob o nome *Least Absolute Selection and Shrinkage Operator* (LASSO) por [Tibshirani 1996].

Os parâmetros da regressão LASSO não podem ser estimados através de método tradicionais baseados na diferenciação da função objetivo. Esta desvantagem levou ao desenvolvimento de uma infinidade de técnicas para determinar os parâmetros ótimos. [Schmidt 2005] desenvolveu um estudo que elenca 8 métodos distintos de otimização. Vários desses algoritmos usam diretamente o problema de otimização sem restrições

$$\underset{\boldsymbol{\beta}}{\text{minimizar}} \quad (\mathbf{p} - \mathbf{V}\boldsymbol{\beta})^T (\mathbf{p} - \mathbf{V}\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1, \quad (16)$$

enquanto outras técnicas usam formulações restritas equivalentes, como por exemplo

$$\begin{aligned} \underset{\boldsymbol{\beta}}{\text{minimizar}} \quad & (\mathbf{p} - \mathbf{V}\boldsymbol{\beta})^T (\mathbf{p} - \mathbf{V}\boldsymbol{\beta}) \\ \text{sujeito a} \quad & \|\boldsymbol{\beta}\|_1 \leq \frac{1}{\lambda}. \end{aligned} \quad (17)$$

3.6. Seleção de Modelos

Para fins de seleção do melhor modelo usaremos neste artigo o *critério de informação bayesiano* (BIC, sigla em Inglês) [Schwarz 1978]:

$$BIC(k_\beta) = N \ln(SSE) + (k_\beta + 1) \ln(N), \quad (18)$$

em que N é o número total de observações usadas na construção do modelo e $SSE(k_\beta)$ é a soma dos erros (i.e. resíduos) quadráticos para um dado modelo com k_β parâmetros:

$$SSE(k_\beta) = \sum_{i=1}^n (p_i - \hat{p}_i)^2, \quad (19)$$

com \hat{p}_i sendo o valor predito da potência gerada.

O valor do BIC será usado também para a determinação do número adequado de parâmetros para os modelos polinomiais (usual e ponderado) e da rede MLP. Para os modelos polinomiais, o número de parâmetros é igual a $k_\beta = k + 1$, em que k é a ordem do maior expoente. No caso do modelo neural, o número de parâmetros depende apenas do número de neurônios ocultos (q), pois o número de entradas e o número de saídas é fixado em $p = c = 1$. Assim, o número total de parâmetros da rede MLP é dado por $k_\beta = (p + 1) \cdot q + (q + 1) \cdot m = (1 + 1) \cdot q + (q + 1) \cdot 1 = 3q + 1$.

4. Abordagem Proposta

Suponha uma regressão polinomial conforme apresentada na Seção 3.1. Considere agora a introdução de coeficiente extra α_j associado a cada termo da regressão, segundo a seguinte definição:

$$\hat{p}_i = \alpha_0\beta_0 + \alpha_1\beta_1v_i + \cdots + \alpha_{k_*}\beta_{k_*}v_i^{k_*} = \sum_{j=0}^{k_*} \{\alpha_j\beta_jv_i^j\} = (\boldsymbol{\alpha} \circ \boldsymbol{\beta})^T \mathbf{v}_i \quad (20)$$

em que k_* é o maior grau permitido ao polinômio, $\boldsymbol{\beta}$ é o vetor de coeficientes do polinômio, $\boldsymbol{\alpha}$ é um vetor de elementos binários (i.e. $\alpha_j \in \{0, 1\}$, $j = 0 \dots k_*$) chamados de *coeficientes de ativação*, \mathbf{v}_i é o vetor de potências associadas à i -ésima medida de velocidade v_i . O símbolo \circ denota um produto de Hadamard.

É importante enfatizar o papel dos coeficientes α_j no modelo proposto. Um coeficiente α_j com valor 0 elimina completamente o termo correspondente do polinômio. Segundo a formulação mostrada na Eq. (20), o modelo polinomial original torna-se um caso particular do metodologia proposta, em que $\alpha_j = 1, \forall j$. Do exposto, buscamos uma solução que simultaneamente minimize a soma dos erros quadráticos (SSE) do modelo e o número de termos do modelo polinomial. Isto é equivalente a procurar uma solução ótima *esparsa* para o vetor $\boldsymbol{\alpha}$ e, conseqüentemente, para o vetor de coeficientes $\boldsymbol{\beta}$. Um exemplo hipotético de uma solução esparsa para o problema de interesse em que $k_* = 4$ é dada por $\boldsymbol{\alpha} = [1 \ 1 \ 0 \ 1 \ 0]^T$. Neste caso, usamos apenas os coeficientes β_0, β_1 e β_3 , o que leva à seguinte expressão para o preditor polinomial:

$$\hat{p}_i = \hat{\beta}_0 + \hat{\beta}_1v_i + \hat{\beta}_3v_i^3, \quad (21)$$

com os coeficientes $\hat{\beta}_0, \hat{\beta}_1$ e $\hat{\beta}_3$ sendo estimados pela mesma expressão usada na Eq. (8), mas com as colunas para os quais $\alpha_j = 0$ retiradas. No presente exemplo, as colunas $j = 3$ e $j = 5$ seriam eliminadas.

Para encontrar o vetor-solução $\boldsymbol{\alpha}$ ótimo lançaremos mão de um método de otimização metaheurística, mais especificamente, algoritmos genéticos (AG) [Engelbrecht 2007]. Como os elementos de $\boldsymbol{\alpha}$ são binários, a codificação binária para os cromossomos será usada. Deste modo, o l -ésimo indivíduo de uma população de M cromossomos é definido como

$$\text{Indivíduo}_l = \boldsymbol{\alpha}_l = [\alpha_{l,0} \ \alpha_{l,1} \ \cdots \ \alpha_{l,k_*}]^T, \quad (22)$$

para $l = 1, \dots, M$.

Para fins de uma justa comparação com as abordagens clássicas apresentadas na Seção 3, adotamos como função de aptidão do AG o índice BIC, conforme definido na Eq. (18). Assim, a aptidão de cada indivíduo é calculada como

$$\text{Aptidão}(\boldsymbol{\alpha}_l) = BIC(\boldsymbol{\alpha}_l), \quad (23)$$

para $l = 1, \dots, M$. Limitamos o valor máximo da ordem do polinômio em $k_* = 10$, com o AG possuindo as seguintes especificações: (i) $N_g = 20$ gerações; (ii) população de tamanho $M = 100$; (iii) cromossomo binário de tamanho $k_* + 1 = 11$; (iv) seleção por

Table 1. Comparação de desempenho dos modelos quanto ao BIC.

Modelo	Num. de Parâmetros	BIC	Diferença acumulada
PRISMA	4	234526,3	-
5PL	5	234599,2	+72,9
Polinomial	8	234602,7	+76,4
MLP	10	234628,7	+102,4
4PL	4	234670,1	+152,2
LASSO	4	234769,4	+243,1
Ponderado	6	234837,2	+310,9

torneio; (v) cruzamento de um ponto; (vi) probabilidade de cruzamento $p_c = 80\%$; (vii) probabilidade de mutação $p_m = 0,1\%$; (viii) sem elitismo; (ix) número de execuções = 1.

A abordagem proposta, por produzir um modelo polinomial esparso, pode ser entendida como uma alternativa ao tradicional método LASSO e será denotada doravante pelo acrônimo PRISMA (Polynomial Reduction Induced by Sparse Method Algorithm).

5. Resultados e Discussão

O conjunto de dados para validação da metodologia proposta foi utilizado primeiramente em [Lee et al. 2015] e está disponível para uso público⁵. São disponibilizados os dados de velocidade e potência de 6 aerogeradores ao todo (4 inshore e 2 offshore), denotadas WT1 a WT6, respectivamente. No presente artigo, usamos apenas os dados referentes à turbina WT5, mas a tecnologia proposta e as conclusões são válidas para todo aerogerador com controle do ângulo de passo. Os dados ambientais do aerogerador WT5, dentre eles a velocidade do vento, são coletados por sensores em uma torre meteorológica, enquanto os dados de potência são medidos na própria turbina. Cada torre meteorológica está associada a 2 aerogeradores, o que significa que cada medida de velocidade feita naquela torre é pareado com as de potência das 2 torres associadas.

Os dados relativos à turbina WT5 (no total de 16.444 medidas) foram coletados de 01/01/2009 a 31/12/2009. É prática corrente arranjar os dados coletados em fazendas eólicas em blocos de 10min, período em que velocidades do vento são consideradas estacionárias (i.e. aproximadamente constantes). consequentemente, tomam-se as médias das leituras da potência de saída e da velocidade do vento a cada intervalo de 10min. Todos os modelos implementados e experimentos foram realizados no ambiente de programação do software Octave⁶.

Na Figura 2 são mostradas as curvas de potência obtidas pelo uso da metodologia PRISMA em comparação com modelos alternativos, sendo 3 deles modelos não-esparsos - 5PL, POLI e MLP(1,3,1) - e um esparso (LASSO). Como esperado, todos os modelos geram curvas de potência com a forma sigmoideal, semelhantes àquela da Figura 1(b), que é a forma típica da curva de potência de aerogeradores com controle de *pitch*. Contudo, numericamente falando, os modelos tiveram desempenho geral muito diferentes, conforme podemos observar na Tabela 1. O modelo PRISMA obteve o melhor desempenho

⁵<https://aml.engr.tamu.edu/2001/09/01/publications/>

⁶<http://octave.org>

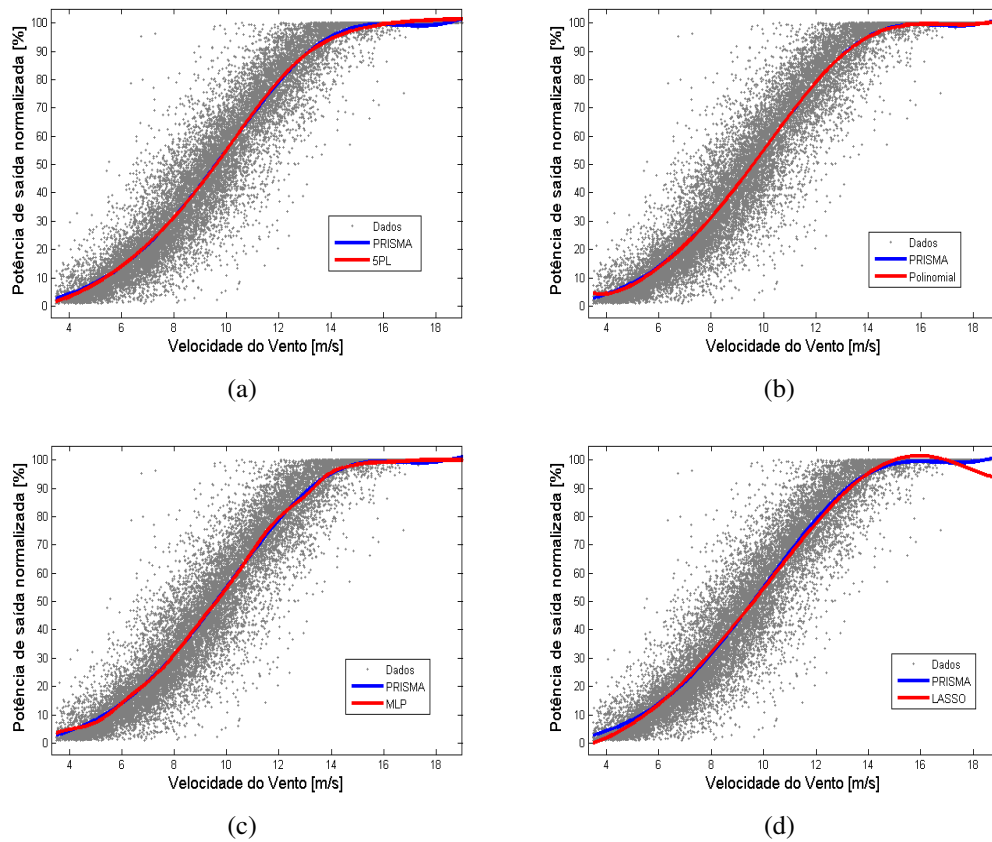


Figure 2. Curva de potência do modelo proposto. Os pontos cinza são os dados do aerogerador WT5 e a potência de saída foi normalizada pela potência nominal.

combinado (i.e. menor BIC + menor número de parâmetros), seguido pelos modelos não-esparcos 5PL, polinomial e MLP, nesta ordem. Ao compararmos os modelos esparcos PRISMA e LASSO, ambos chegaram a uma solução que utiliza 4 termos para a curva de potência, porém podemos facilmente perceber que, para o conjunto de dados utilizado, que o modelo PRISMA se aproxima mais da potência nominal para velocidade de vento superior a 15 m/s, apresentando-se, portanto, um melhor ajuste aos dados se comparado ao método LASSO.

Para o modelo PRISMA, as ativações não-nulas resultantes foram $\{\alpha_3, \alpha_8, \alpha_9, \alpha_{10}\}$, de modo que os coeficientes estimados foram $\{\hat{\beta}_3, \hat{\beta}_8, \hat{\beta}_9, \hat{\beta}_{10}\}$. Vale ressaltar que mesmo que o modelo LASSO tenha chegado a uma solução com apenas 4 parâmetros, nem por isso este modelo produziu o menor BIC. Pelo contrário, foi um dos piores, juntamente com o modelo polinomial ponderado e o 4PL. A rede MLP com $q = 3$ neurônios ocultos foi o modelo que necessitou de mais parâmetros (10, no total) para chegar a uma solução razoavelmente boa.

A título de completude, mostramos na Figura 3 os boxplots dos resíduos para todos os modelos avaliados. Como era de se esperar para o tipo de aplicação que estamos abordando, há muitos *outliers* presentes nos dados [Sohoni et al. 2016]. A presença de *outliers* acaba interferindo negativamente no processo de estimação de parâmetros de todos os modelos e gerando assim uma distribuição de resíduos que também apresenta outliers,

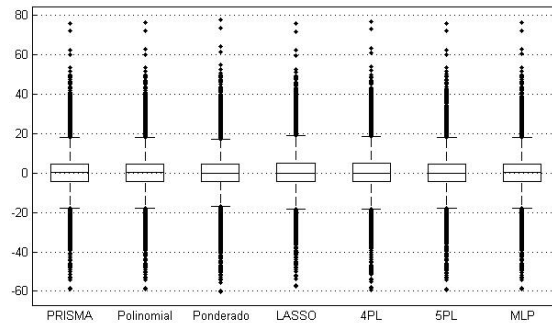


Figure 3. Boxplots dos resíduos para cada um dos modelos utilizados.

Table 2. Resultado do teste de Kolmogorov-Smirnov para os modelos avaliados neste artigo.

Teste KS	Polin.	Ponder.	LASSO	4PL	5PL	MLP
PRISMA	0	1	1	1	1	0

conforme pode ser verificado na referida figura. O mais importante aqui é que o modelo proposto obteve o melhor desempenho, mesmo em um cenário com vários *outliers*. E também que o modelo polinomial ponderado, que supostamente deveria lidar bem com *outliers*, teve o pior desempenho.

Para finalizar, usamos o teste de Kolmogorov-Smirnov (KS) a fim de avaliar a similaridade estatística entre os resíduos produzidos pelos vários modelos em comparação aos do modelo proposto. A hipótese nula do teste KS é a de que os resíduos dos dois modelos comparados provêm da mesma distribuição de probabilidade. Dois modelos não são considerados *equivalentes estatisticamente* se a hipótese nula for rejeitada. A Tabela 2 traz os resultados dos testes, onde se percebe o modelo polinomial clássico e o MLP produziram resultados similar, estatisticamente falando, ao do modelo proposto. Isto posto, combinando os resultados das Tabelas 1 e 2 é possível concluir que a abordagem proposta não descaracteriza o modelo polinomial clássico, mas sim o refina, tornando-o mais eficiente e parcimonioso.

6. Conclusões e Trabalhos Futuros

Neste artigo uma nova metodologia foi proposta visando o projeto automático e parcimonioso do modelo polinomial aplicado à estimação da curva de potência de aerogeradores. Usando dados reais, o desempenho do método proposto foi comparado ao de vários outros modelos que compõem o estado da arte na estimação da curva de potência de aerogeradores, de tal forma que os resultados obtidos indicaram claramente um melhor desempenho para o modelo proposto.

Ainda que, tanto na indústria de aerogeradores quanto na academia, a prática no desenvolvimento de modelos da curva de potência leve em consideração apenas a velocidade do vento (e, às vezes, a direção do vento), estamos atualmente refinando o modelo proposto a fim de introduzir outras variáveis de entrada, além da velocidade do vento, que influenciam no valor da potência gerada pelo aerogerador, tais como direção, densidade do ar, umidade, intensidade da turbulência, etc.

Agradecimentos: Os autores agradecem ao CNPq (processo no. 309451/2015-9) e à CAPES (processo no. 1707928/DS) pelo suporte financeiro.

References

- Betz, A. (1920). Das maximum der theoretisch möglichen ausnützung des windes durch windmotoren. *Zeitschrift für das gesamte Turbinenwesen*, 26:307–309.
- Clifton, A., Kilcher, L., Lundquist, J. K., and Fleming, P. (2013). Using machine learning to predict wind turbine power output. *Environmental Research Letters*, (8):1–8.
- Engelbrecht, A. P. (2007). *Computational Intelligence: An Introduction*. Wiley, 2nd edition.
- Hagan, M. T. and Menhaj, M. (1994). Training feed-forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6):989–993.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101.
- Lee, G., Ding, Y., Genton, M. G., and Xie, L. (2015). Power curve estimation with multivariate environmental factors for inland and offshore wind farms. *Journal of the American Statistical Association*, 110:56–67.
- Li, S., Wunsch, D. C., O’Hair, E., and Giesselmann, M. G. (2001). Comparative analysis of regression and artificial neural network models for wind turbine power curve estimation. *Journal of Solar Energy Engineering*, 123(4):327–332.
- Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P. (2014). A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30:452–460.
- Lydia, M., Suresh Kumar, S., Immanuel Selvakumar, A., and Edwin Prem Kumar, G. (2013). Advanced algorithms for wind turbine power curve modeling. *IEEE Transactions on Sustainable Energy*, 4(3):827.
- Manwell, J., McGowan, J., and Rogers, A. (2009). *Wind energy explained : theory, design, and application*. John Wiley & Sons, Ltd., 2nd edition.
- Marciukaitis, M., Zutautaitė, I., Martišauskas, L., Jokšas, B., Gecevičius, G., and Sfetsos, A. (2017). Non-linear regression model for wind turbine power curve. *Renewable Energy*, 113:732–741.
- Schmidt, M. (2005). Least squares optimization with L_1 -norm regularization. Technical report, The University of British Columbia. CS542B Project Report.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.
- Shokrzadeh, S., Jozani, M. J., and Bibeau, E. (2014). Wind turbine power curve modeling using advanced parametric and nonparametric methods. *IEEE Transactions on Sustainable Energy*, 5(4):1262–1269.
- Sohoni, V., Gupta, S. C., and Nema, R. K. (2016). A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems. *Journal of Energy*, 2016:1–18.
- Storn, R. and Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society - Series B*, 58(1):267–288.
- Wang, J., Zhao, X., Guo, X., and Li, B. (2018). Analyzing the research subjects and hot topics of power system reliability through the Web of Science from 1991 to 2015. *Renewable and Sustainable Energy Reviews*, 82(1):700–713.