

# Previsão de potência eólica – treino de sistemas com critérios de entropia

V. Miranda, *Fellow IEEE*, C. Cerqueira e C. Monteiro

**Resumo** — Este artigo aplica à previsão de potência eólica um novo método de treino de sistemas de inferência difusa ou redes neuronais que conduz a uma distribuição dos erros de previsão com Entropia mínima, implicando maior frequência de erros diminutos que os sistemas treinados com o clássico critério do Erro Quadrático Médio. O método combina a definição de Renyi de Entropia com a técnica das janelas de Parzen coo medida do conteúdo de informação da distribuição dos erros, num paradigma designado ITL – Information Theoretic Learning.

Apresenta-se um caso prático de aplicação a um parque eólico europeu, com treino de um sistema difuso de Takagi-Sugeno com os parâmetros ajustados por um método de Enxame Evolucionário de Partículas (EPSO).

**Palavras chave** — Sistemas de Inferência, Sistemaas difusos, energia eólica, entropia.

## I. OBJECTIVO

Este artigo discute o problema de produzir uma previsão de potência gerada num parque eólico, a partir de estimativa de velocidade e direcção do vento. O método consiste na aplicação de um Sistema de Inferência Difusa de Takagi-Sugeno, treinado com um algoritmo do tipo Enxame Evolucionário de Partículas (EPSO – Evolutionary Particle Swarm Optimization), com a particularidade de o critério de treino não ser o clássico Erro Quadrático Médio mas sim uma medida da Entropia da Informação, seguindo o paradigma designado ITL – Information Theoretic Learning[1][2].

O problema existe e é importante porque

- um parque é uma estrutura geograficamente dispersa e que sofre influência das características locais do terreno
- a velocidade do vento (e direcção) pode variar de gerador para gerador
- a jusante de um gerador aparecem efeitos de esteira e turbulência que reduzem a energia do vento e podem afectar outros geradores em função da disposição do parque e da direcção geral do vento
- um terreno de orografia complicada pode introduzir efeitos inesperados
- complexidade adicional resulta da não linearidade da curva de potência vs. velocidade do vento.
- em regra existe uma medida de velocidade e direcção do

vento num local próximo de um parque, mas não medidas por cada gerador.

A previsão da potência gerada por um parque eólico é um procedimento necessário em todos os métodos de previsão que exigem uma previsão de vento intermédia. Isto ocorre em sistemas como o *Prediktor*[3] ou *eWind*[4], mas também em métodos de base estatística [5] ou baseados em técnicas de inteligência computacional [6].

Esta previsão é actualmente muito importante na Europa, onde a crescente penetração de energia eólica atingirá percentagens ainda há pouco tempo não imagináveis, em particular na Alemanha, Espanha ou Portugal, em função do esforço europeu colectivo de cumprimento das obrigações do protocolo de Quioto.

Alguns métodos convertem previsões de vento em potência por meio de uma curva de potência empírica que tenta modelizar o comportamento não linear dos grupos eólicos. Em [7], por exemplo, Lange derivou um modelo para as incertezas de previsão em função da derivada de uma curva desse tipo. Todavia, a experiência mostra que a relação entrada-saída é demasiado complexa para ser bem representada por uma simples função e, portanto, sistemas mapeadores do tipo das redes neuronais ou dos sistemas de inferência difusa são mais apropriados para emular essa relação.

O treino deste tipo de sistemas é feito classicamente minimizando um critério de Erro Quadrático Médio (MSE – Mean Square Error). O motivo para se investigar a aplicação alternativa de um critério de Entropia fundamenta-se no facto de que a distribuição dos erros de previsão de potência eólica não é Gaussiana (normal). Mesmo que as previsões de vento exibissem erros Gaussianos, a não linearidade da curva característica dos grupos eólicos produziria erros não normalmente distribuídos.

Tipicamente, as distribuições dos erros de previsão de potência eólica são assimétricas positivas (para a direita) e exibem curtose platicúrtica (mais achatadas que as Gaussianas). As distribuições normais são as únicas que acumulam toda a informação nos dois primeiros momentos (média e variância). Quando se trata de distribuições de erros como se fossem Gaussianas, desperdiça-se informação contida nos momentos de ordem superior. É o caso quando se usa um critério de mínimos quadrados ou de variância (como o MSE) – o processo de treino apenas passa uma parte da informação dos dados para os parâmetros do sistema, deixando na distribuição dos erros uma fracção da informação.

A definição de Entropia de Shannon[8] tem sido muito usada como medida de conteúdo de informação, mas

---

V. Miranda colabora com o INESC Porto e também com a FEUP, Faculdade de Engenharia da Universidade do Porto, Portugal (e-mail: vmiranda@inescporto.pt).

Cristina Cerqueira colabora com o INESC Porto and também com a FCUP, Faculdade de Ciências da Universidade do Porto, Portugal (email: caac@inescporto.pt).

Claudio Monteiro colabora com o INESC Porto e também com a FEUP (email: cmonteiro@inescporto.pt).

apresenta dificuldades práticas de aplicação ao treino de sistemas. Pelo contrário, o paradigma ITL usa a definição de Entropia de Renyi e combina-a com o método das janelas de Parzen (que estima a função densidade de probabilidade dos erros) num procedimento tratável computacionalmente.

No trabalho descrito neste artigo, as previsões de potência eólica são geradas por um Sistema de Inferência Difusa de Takagi-Sugeno de 1ª ordem (TS-FIS) [9]. Em vez de se adoptar o método clássico de retropropagação, optámos por um algoritmo do tipo Enxame Evolucionário de Partículas – EPSO [10] para encontrar os pesos ou valores dos parâmetros que minimizassem a Entropia da distribuição dos erros.

Os nossos resultados mostram que este procedimento se consegue distribuições dos erros mais esbeltas do que com o critério MSE, evidenciando frequências maiores de erros próximos de zero [11]. Como bónus, mostramos também que um algoritmo do tipo EPSO pode descobrir os pesos de um sistema do tipo TS-FIS, quer com critérios de MSE quer de Entropia.

## II. TREINANDO MAPEADORES

Designaremos por *mapeador* uma rede neuronal ou um Sistema de Inferência Difusa ou, em geral, um sistema que emula uma função de transferência entrada-saída e cujo desempenho depende da afinação de pesos ou parâmetros internos.

Podemos dividir um mapeador em três módulos básicos: a sua estrutura interna, o critério de desempenho e o mecanismo de treino. Uma vez definido o tipo de mapeador (classe, estrutura), podemos intervir em cada módulo do seguinte modo (Figura 1):

- Na estrutura interna, modificando os pesos
- No critério de desempenho, seleccionando um processo de avaliação adequado
- No mecanismo de treino, seleccionando um algoritmo que actualiza os pesos em função do critério de desempenho.

No trabalho deste artigo tomámos as seguintes opções:

- estrutura:
- desempenho: minimização da entropia de Renyi, por alternativa ao MSE
- treino: uso de um algoritmo EPSO, por oposição à retropropagação clássica.

O aspecto técnico mais importante abordado neste artigo é o critério de treino. No paradigma ITL, o uso de Entropia como medida de desempenho num contexto de treino supervisionado (onde se define um alvo  $T$  para a saída do mapeador) tem o seguinte intuito: se puderem ser descobertos pesos  $W$  modelando um mapeador cuja saída apresente uma distribuição de erros do tipo função (ou impulso) de Dirac (significando que todos os erros seriam iguais, ver Figura 2), teremos obtido um modelo cuja resposta reproduzirá exactamente a resposta real – bastando para isso subtrair à saída o desvio (bias) para zero correspondendo à efectiva média da função densidade de probabilidade (fdp) dos erros.

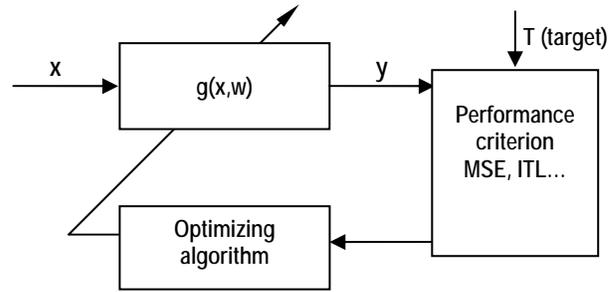


Figura 1 – Organização básica de um mapeador e seus três módulos

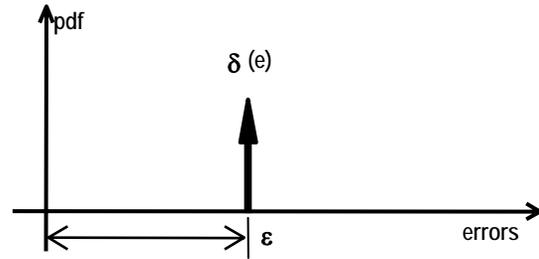


Figura 2 – Um mapeador produzindo um erro sistemático para todas as saídas apresentará uma função densidade de probabilidade dos erros do tipo função de Dirac

Portanto, o objectivo orientando o desenvolvimento do mapeador deverá ser descobrir os valores  $W$  dos pesos por forma a que a fdp dos erros se aproxime o mais possível da função de Dirac. Isto pode conseguir-se minimizando a Entropia da distribuição dos erros. O sucesso do paradigma ITL é ter descoberto uma função custo representando este objectivo e ter organizado um procedimento para computar a solução e em tempo útil.

## III. ENTROPIA

A Entropia é um conceito desenvolvido em teoria da Informação que formaliza a noção de conteúdo de informação. Quanto menos previsível for uma informação, maior o seu conteúdo de informação; uma mensagem perfeitamente conhecida *a priori* tem um conteúdo de informação nulo.

Shannon [8] definiu Entropia como uma distribuição de probabilidade  $P = (p_1, p_2, \dots, p_n)$  tal que

$$H_S(P) = \sum_{k=1}^N p_k \log \frac{1}{p_k} \quad \text{com} \quad \sum_{k=1}^N p_k = 1 \quad \text{e} \quad p_k \geq 0$$

Embora esta definição tenha tido aplicação vasta, nomeadamente em comunicações, outras definições de entropia foram distribuídas. A Entropia de Renyi [13] define-se como

$$H_{R\alpha} = \frac{1}{1-\alpha} \log \sum_{k=1}^N p_k^\alpha \quad \text{com} \quad \alpha > 0, \alpha \neq 1$$

Na verdade, a Entropia de Renyi's é uma família de funções  $H_{R\alpha}$  dependendo de um parâmetro  $\alpha$ . Há uma relação entre as Entropias de Shannon e Renyi:

$$H_{R\alpha} \geq H_S \geq H_{R\beta} \quad \text{se} \quad \beta > 1 > \alpha > 0$$

$$\lim_{\alpha \rightarrow 1} H_{R\alpha} = H_S$$

Quando  $\alpha = 2$ , temos a Entropia Quadrática de Renyi:

$$H_{R2} = -\log \sum_{k=1}^N p_k^2$$

Esta definição pode ser generalizada para uma variável aleatória contínua  $Y$  com fdp  $f_Y(z)$ :

$$H_{R2} = -\log \int_{-\infty}^{+\infty} f_Y^2(z) dz$$

Podemos ver que a Entropia de Renyi, com a sua soma de probabilidades, é muito mais propícia a uma implementação algorítmica expedita que a Entropia de Shannon, com a sua soma pesada de logaritmos de probabilidades.

#### IV. JANELAS DE PARZEN

A estimativa da fdp dos dados a partir de uma amostra constituída por pontos discretos  $\mathbf{y}_i \in \mathbb{R}^M$ ,  $i = 1, \dots, N$  num espaço  $M$ -dimensional pode ser feita pelo método das Janelas de Parzen [13]. Esta técnica usa uma função de *kernel* centrada em cada ponto, que é interpretado como sendo descrito localmente por uma função de densidade de Dirac, e é substituído ou aproximado por um conjunto contínuo cuja densidade é representada pelo *kernel*. Se for usado um *kernel* Gaussiano, as expressão para a estimativa  $\hat{f}_Y$  da fdp real  $f_Y$  de um conjunto de  $N$  pontos é a soma das contribuições individuais

$$\hat{f}_Y(\mathbf{z}) = \frac{1}{N} \sum_{i=1}^N G(\mathbf{z} - \mathbf{y}_i, \sigma^2 \mathbf{I})$$

onde  $G(\dots)$  é o *kernel* Gaussiano e  $\sigma^2 \mathbf{I}$  é a matriz de covariância (assumindo-se que corresponde a variâncias independentes e iguais em todas as dimensões). Em cada dimensão temos

$$G(z_k - y_{ik}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(z_k - y_{ik})^2}$$

Compreende-se que o “tamanho” da janela, aqui definido pelo valor de  $\sigma$ , é importante na obtenção de uma estimativa mais “suave” (para  $\sigma$  grande) ou mais “irregular” de  $f_Y$ .

#### V. O CRITÉRIO DA ENTROPIA SEGUNDO ITL

Combinando a Definição de Renyi da Entropia de uma fdp com a estimativa de uma fdp por janelas de Parzen, obtém-se um estimador da Entropia para um conjunto discreto de pontos  $\{\mathbf{y}\}$  como sendo

$$H_{R2}(\mathbf{y}) = -\log \int_{-\infty}^{+\infty} \hat{f}_Y^2(\mathbf{z}) d\mathbf{z} = -\log V(\mathbf{y}), \text{ onde}$$

$$V(\mathbf{y}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int_{-\infty}^{+\infty} G(\mathbf{z} - \mathbf{y}_i, \sigma^2 \mathbf{I}) G(\mathbf{z} - \mathbf{y}_j, \sigma^2 \mathbf{I}) d\mathbf{z}$$

Nesta expressão, reconhecemos a convolução de Gaussianas, o que permite escrever uma interessante expressão equivalente:

$$V(\mathbf{y}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(\mathbf{y}_i - \mathbf{y}_j, 2\sigma^2 \mathbf{I})$$

Isto significa que, para se calcular a Entropia, não é preciso calcular nenhum integral mas apenas os valores das funções Gaussianas das distâncias entre pares de pontos. NO vocabulário ITL,  $V(\mathbf{y})$  designa-se *potencial de informação* (IP) do conjunto de dados. Como o objectivo é minimizar  $H$ , pode-se para isso maximizar  $V$ . Assim,  $\text{Max } V$  transforma-se na função custo para otimizar um mapeador com treino supervisionado e obter entropia mínima nos erros da saída [14].

A descoberta dos pesos num mapeador pode ser feita empregando-se qualquer algoritmo de optimização que encontre os valores de  $\mathbf{w}$  que minimizem a função objectivo

$$\min H_{R2}(\mathbf{w})$$

Como um algoritmo evolucionário do tipo EPSO.

#### VI. APLICAÇÃO A UM TS-FIS

Os Sistemas de Inferência Difusa de Takagi-Sugeno (TS FIS) são entendidos como optimizáveis por treino supervisionado. Significa isto que se dispõe previamente de conjuntos de treino e de teste com valores alvo  $\mathbf{T}$  e que o objectivo do treino é fazer com que a saída do sistema se assemelhe tanto quanto possível ao alvo. Neste caso, não estamos a lidar com o conteúdo de informação da saída do sistema  $\mathbf{y}$  mas com o conteúdo em informação dos erros  $\boldsymbol{\varepsilon} = \mathbf{T} - \mathbf{y}$ .

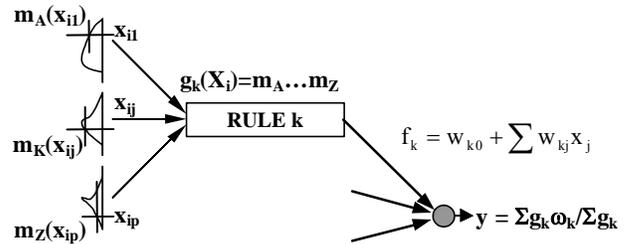


Figura 3 – Esquema de um TS FIS. Cada padrão de entrada  $X$  activa algumas funções de pertinência; a combinação destas dispara a regra  $k$  com força  $g_k$ . A combinação pesada das forças de disparo das regras produz a saída do sistema.

Num TS FIS, existem regras que são difusas no antecedente mas determinísticas no consequente. A forma geral de da regra de ordem  $k$  com saída  $y_k$  é

$$\text{SE } (x_1 \text{ is } A \text{ and } \dots \text{ and } x_p \text{ is } Z) \text{ ENTÃO } y_k = y(\mathbf{x}, \mathbf{w})$$

O antecedente da regra  $k$  é um conjunto difuso cuja função de pertinência  $g_k$  é a intersecção dos conjuntos difusos descrevendo as condições  $A, \dots, Z$ . Normalmente, para operar a intersecção usa-se a T-norma produto (dos valores de pertinência de cada variável de entrada).

O consequente da regra  $k$  é uma função  $f_k$  das entradas. Num TS FIS de ordem 0,  $f_k$  é constante e, portanto,  $f_k = w_k$ . Num TS FIS de 1ª ordem,  $f_k$  é uma combinação linear das entradas

$$f_k = w_k + w_{k1}x_1 + \dots + w_{kp}x_p$$

A saída de um TS FIS é uma soma pesada das respostas de todas as regras (ver Figura 3):

$$y_i = \frac{\sum_{k=1}^R g_k w_k}{\sum_{k=1}^R g_k} = \sum_{k=1}^R \bar{g}_k w_k, \quad \text{with } \bar{g}_k = \frac{g_k}{\sum_{k=1}^R g_k}$$

Em modo de treino supervisionado, um padrão de entrada  $\mathbf{X}_i$  com resposta  $y_i$  gerará um erro relativo a um alvo  $T_i$  como sendo  $\varepsilon_i = T_i - y_i$ . Adoptando-se o paradigma ITL, o objectivo do treino será maximizar o potencial de informação associado à distribuição dos erros.

O principal conjunto de parâmetros é o dos pesos  $w$  das regras. Outros parâmetros podem ser ajustados – por exemplo, se as funções de pertença das entradas são Gaussianas, podem-se afinar os seus valores centrais e dispersões. Todavia, isto não é conveniente em muitos casos porque as entradas estão associadas a valores linguísticos e a mudança na sua forma e localização cria uma dissociação delas com as etiquetas linguísticas que é suposto representarem.

Para cálculo dos pesos segundo o paradigma ITL, foram propostas adaptações ao método de retropropagação clássico [15] mas neste trabalho optámos por uma abordagem distinta.

## VII. EPSO COMO OPTIMIZADOR

O método do Enxame Evolucionário de Partículas (EPSO – Evolutionary Particle Swarm Optimization) é um híbrido de conceitos dos Algoritmos Evolucionários e da Optimização por Enxame de Partículas (PSO – Particle Swarm Optimization) proposto inicialmente em [10] com aplicações em Sistemas de Potência [16]. É um Algoritmo Evolucionário em que o operador mutação apenas se aplica sobre parâmetros estratégicos e o operador recombinação é não-convencional: trata-se, de facto, da regra do movimento a que uma partícula obedece. Por limitações de espaço, o leitor é remetido para a consulta das publicações em referência.

Numa diversidade de problemas, o EPSO já foi confirmado como tendo melhor desempenho que outras meta-heurísticas como Algoritmos Genéticos ou Enxames de Partículas clássicos [17][18].

## VIII. ESTIMAÇÃO DA POTÊNCIA PRODUZIDA POR UM PARQUE EÓLICO

Esta secção apresenta o resultado da aplicação conjunta das técnicas referidas anteriormente, para previsão da potência de um parque eólico a partir da velocidade e direcção de vento.

Os dados respeitam a um parque no norte de Portugal, com geradores de potência média de cerca de 1 MW totalizando cerca de 40 MW, dispersos sobre a crista de uma serra. Não se publicam detalhes específicos por razões de confidencialidade. Os dados foram recolhidos de 1 de Janeiro de 2004 a 20 de Fevereiro de 200 e as potências geradas são representadas em p.u. da capacidade instalada.

O objectivo é mostrar que a aplicação de um critério de Entropia segundo o paradigma ITL produz um mapeador melhor que a aplicação do clássico critério de Erro Quadrático

Médio. Para o demonstrar, treinámos dois TS FIS de ordem 0 usando o algoritmo EPSO; o modelo MSE determinou valores de parâmetros de modo a minimizar o Erro Médio Quadrático; o modelo Entropia calculou os mesmos parâmetros de modo a minimizar a Entropia de Renyi na distribuição dos erros. Seleccionámos 5000 pontos dos dados, usando 1000 pontos para treino e dedicando os restantes a teste.

Na Figura 4 apresenta-se uma representação de 9993 medidas de potência vs. velocidade do vento. Este conjunto apresenta alguns valores estranhos e foi sujeito a posterior tratamento. Por exemplo, notam-se pontos com velocidade elevada de vento e potência gerada nula, o que se deve à medidas corresponderem a momentos de desligamento do parque.

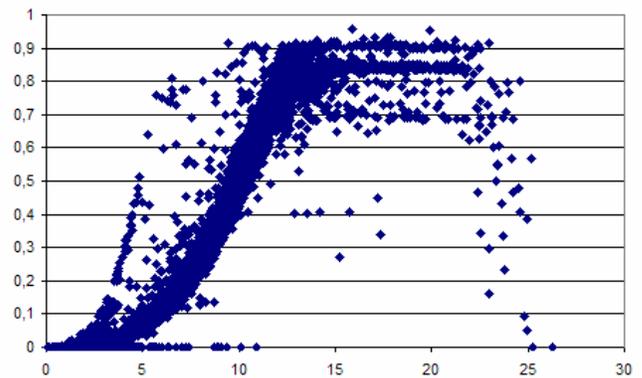


Figura 4 – Gráfico da potência gerada no parque (eixo y) em função da velocidade do vento (eixo x). Potência em p.u. da capacidade do parque. Dados não tratados.

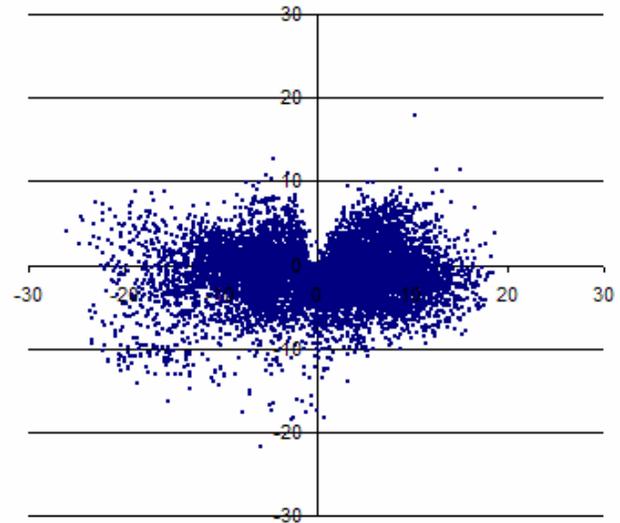


Figura 5 – Distribuição da velocidade do vento (em m/s) e direcção, conforme medidas num local nas vizinhanças do parque. Cada ponto é a extremidade de um vector cuja dimensão é proporcional à velocidade do vento e cujo ângulo com a origem e está relacionado com a direcção do vento..

Na Figura 5 apresentam-se os mesmos pontos mas relacionando velocidade de vento e sua direcção.

Para testar o critério de Entropia, definimos previamente um TS FIS de ordem 0 com as seguintes características:

- o Duas variáveis de entrada: velocidade de vento  $S$ , em m/s, e direcção do vento  $D$ , em graus
- o Um intervalo de viabilidade  $[0, 30]$  para  $S$ , e  $[0, 360]$  para  $D$
- o Uma gama de variação de  $P$  como  $[0, 1]$
- o Uma partição do universo de discurso de  $S$  em cinco conjuntos difusos com funções de pertença Gaussianas
- o Uma partição do universo de discurso de  $D$  em dois conjuntos difusos com funções de pertença Gaussianas

Mantivemos essas funções fixas e apenas optimizámos os pesos  $w$  das 10 regras difusas do sistema. Para determinar os seus valores, usámos 20 partículas (indivíduos) e um factor de replicação  $r=2$  (cada indivíduo dá origem a dois descendentes). Usámos mutações Gaussianas e uma taxa de aprendizagem  $\tau = 0.5$ .

Como já referido, treinámos 2 modelos: MSE e Entropia. Usámos o mesmo critério de paragem e os modelos só diferiram, de facto, na função de adaptação adoptada. Para o modelo Entropia, usámos ainda janelas de Parzen de tamanho fixo ( $\sigma = 0.01$ ).

A Figura 6 e a Figura 7 representam as fdp dos erros obtidos nos conjuntos de treino e de teste, após o treino, e mostram como o sistema treinado com MSE tem menos erros próximo de zero que o sistema treinado com o critério de Entropia.

O cálculo da Entropia de Renyi para a distribuição dos erros no conjunto de teste deu  $-0.6648$  para os erros resultantes da aplicação do MSE e  $-1.1782$  para os resultantes da aplicação do critério da Entropia. Não sendo surpreendente, este resultado confirma que a distribuição dos erros resultante da aplicação do novo critério contém menos informação que a outra – portanto, mais informação passou dos dados para os pesos do modelo.

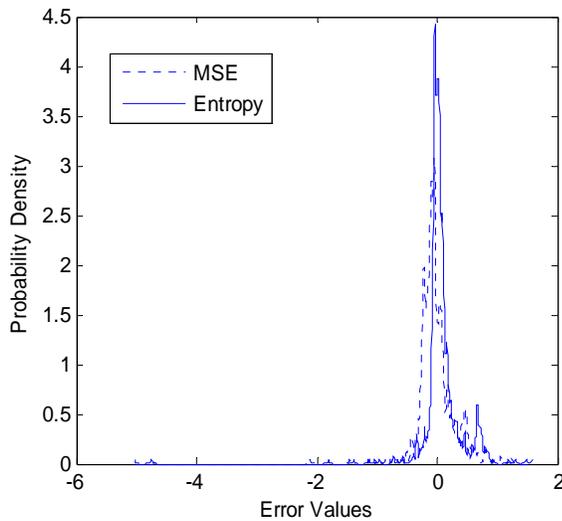


Figura 6 – Fdp dos erros de previsão no conjunto de treino, comparando os modelos com critérios MSE e Entropia (representação conseguida com janelas de Parzen aplicadas no conjunto discreto dos erros). Os erros estão em décimas de p.u. A forma mais “aguçada” da distribuição associada ao modelo Entropia mostra que este critério conduz a erros mais próximos de zero.

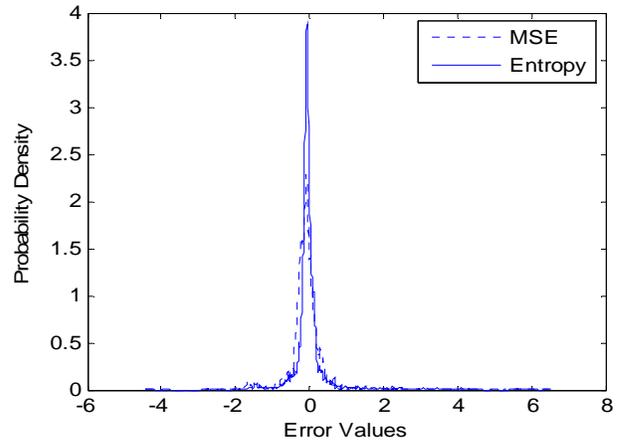


Figura 7 – Fdp dos erros de previsão no conjunto de teste, comparando os modelos com critérios MSE e Entropia (representação conseguida com janelas de Parzen aplicadas no conjunto discreto dos erros). Os erros estão em décimas de p.u.

O impacto destes resultados pode ser apreciado nas Figura 8 e Figura 9, onde se vê no domínio dos tempos as previsões dos dois modelos e os valores horários efectivamente medidos. Num caso encontra-se um período de boa concordância e noutro um período de menos boa concordância com os valores medidos e registados no SCADA do parque. O desempenho do modelo treinado com critério de Entropia é superior.

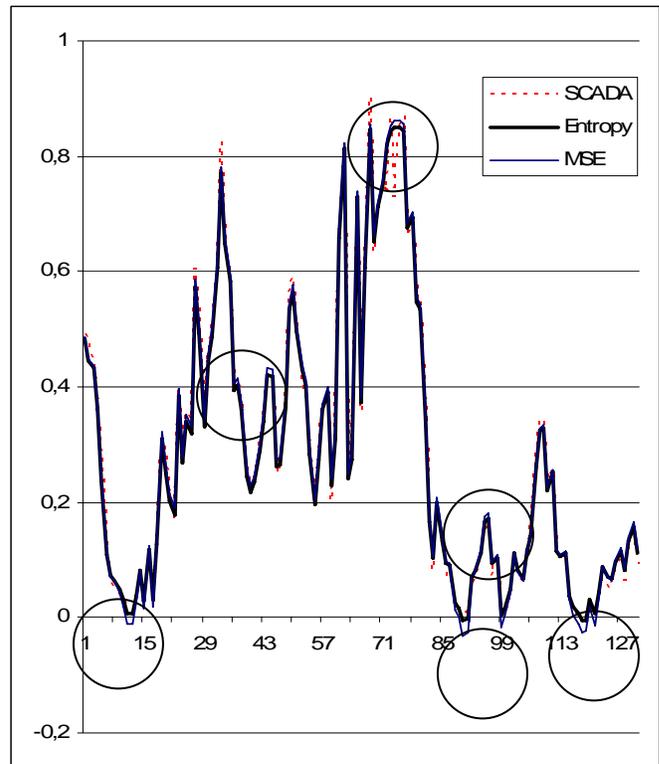


Figura 8 – Comparação do desempenho dos dois modelos, numa sequência de 19 semanas do conjunto de teste onde há boa concordância. Os valores do gráfico são potências horárias e o eixo dos x está graduado em dias. O eixo dos y está em p.u. da potência nominal do parque. Os círculos indicam algumas horas onde é possível ver que modelo adoptando o critério MSE não teve comportamento tão bom como o que adoptou o critério da Entropia.

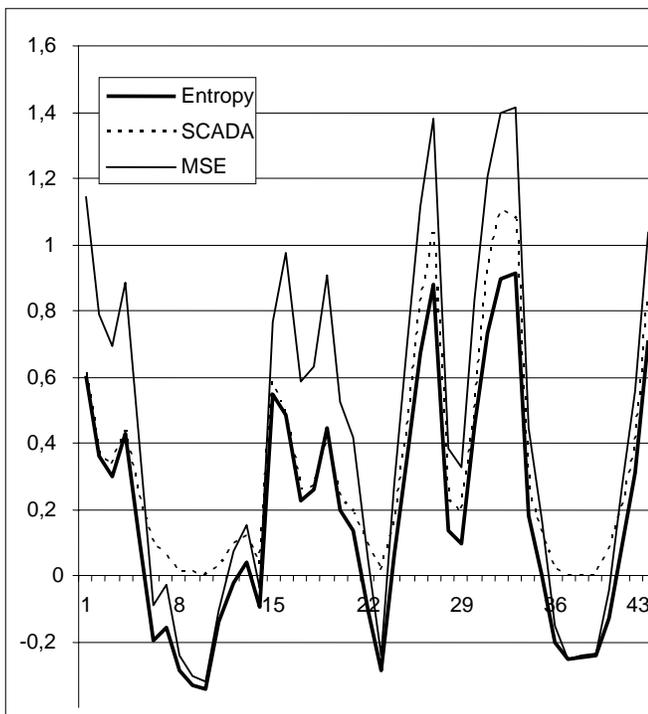


Figura 9 – Comparação do desempenho dos dois modelos, numa sequência de 7 semanas do conjunto de teste onde a concordância com os valores medidos não é tão boa.. Os valores do gráfico são potências horárias e o eixo dos x está graduado em dias. O eixo dos y está em p.u. da potência nominal do parque. Os círculos indicam algumas horas onde é possível ver que modelo adoptando o critério MSE não teve comportamento tão bom como o que adoptou o critério da Entropia.

### IX. CONCLUSÕES

Se o modelo de previsão de potência eólica de um parque ainda pode ser objecto de afinação, o certo é que os resultados apresentados mostram que a aplicação de um critério de Entropia permitiu obter melhores resultados do que a aplicação do clássico critério do Erro Quadrático Médio.

Este resultado deriva do facto de a distribuição dos erros da previsão de potência eólica de um parque não ser Gaussiana e, portanto, um critério que se baseie na medida de informação da fdp dos erros naturalmente há-de produzir melhores resultados que um critério que apenas se baseie na variância.

Este resultado constitui uma contribuição importante para os modelos de maior fôlego de previsão de geração eólica que necessitam de uma conversão da previsão de vento em previsão de potência eólica, pois podem permitir melhores modelos num problema de grande importância operacional em cada vez mais países do mundo, à medida que a penetração da produção eólica assume percentagens significativas e deixa de ser marginal.

### X. AGRADECIMENTO

Trabalho parcialmente desenvolvido durante uma estadia de visita à Universidade da Florida, Gainesville, FL, USA, em Abril 2005. V. Miranda agradece o apoio científico do Prof. J. Principe e financeiro da FCT – Fundação para a Ciência e Tecnologia, Portugal.

- [1] J. C. Principe and Dongxin Xu, "Introduction to information theoretic learning", *Proc. International Joint Conference on Neural Networks (IJCNN'99)*, Washington DC, USA, 10-16 July 1999, pp. 1783-1787
- [2] J. C. Principe and D. Xu "Information-theoretic learning using Renyi's quadratic Entropy", in J.-F. Cardoso, C. Jutten, and P. Loubaton, editors, *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation*, Aussois, France, pages 407-412, 1999.
- [3] L. Landberg, "Short term prediction of the power production of wind farms", *Journal of Wind Eng. and Ind. Aerodynamics*, no.80, pp. 207-220, 1999
- [4] B. Bailey, M. C. Brower and J. C. Stack "Short term wind forecasting – development and application of a mesoscale model!", *Proceedings of the 1999 European Wind Energy Conference EWEC'99*, Nice, France, pag. 1062-1065, March 1999
- [5] G. Giebel, L. Landberg and T. S. Nielsen, "The ZEPHYR project: the next generation prediction tool", *Proceedings of the 2001 European Wind Energy Conference EWEC'01*, pp. 777-781, Copenhagen, Denmark, June 2001
- [6] S. Li, D. C. Wunsch E. A. O'Hair, "Using neural networks to estimate wind power turbine generation", *IEEE Transactions on Energy Conversion*, vol. 13, no.3, pp. 276-282, September 2001
- [7] M. Lange, "Analysis of the uncertainty of wind power predictions", PhD Thesis, Carl von Ossietzky University, Oldenburg, Germany, 2003
- [8] C.E. Shannon, "A Mathematical Theory of Communications", *Bell Systems Technical Journal*, vol. 27, pp. 379-423, pp. 623-656, 1948.
- [9] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control", *IEEE Transactions on Systems, Man and Cybernetics*, v. 15, pp. 116-132, 1985
- [10] V. Miranda and N. Fonseca, "EPSO – Best-of-Two-Worlds Meta-Heuristic Applied To Power System Problems ", *Proceedings of WCCI 2002 - World Congress on Computational Intelligence - CEC - Conference on Evolutionary Computing*, Honolulu, Hawaii, U.S.A., May, 2002
- [11] D. Erdogmus and J. C. Principe, "Comparison of Entropy and mean square error criteria in adaptive system training using higher order statistics", in P. Pajunen and J. Karhunen, editors, *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation*, Helsinki, Finland, Otamedia, Espoo, Finland, 2000, , pages75-80
- [12] A. Renyi, "Some Fundamental Questions of Information Theory", *Selected Papers of Alfred Renyi*, vol 2, pp. 526-552, Akademia Kiado, Budapest, 1976.
- [13] E. Parzen, "On the estimation of a probability density function and the mode", *Annals Math. Statistics*, v. 33, 1962, p. 1065
- [14] D. Erdogmus and J. C. Principe, "Generalized Information Potential Criterion for Adaptive System Training", *IEEE Transactions on Neural Networks*, vol. 13, no. 5, September 2002, pp. 1035-1044
- [15] R. A. Morejon and J. C. Principe, "Advanced search algorithms for information-theoretic learning with kernel-based estimators", *IEEE Transactions on Neural Networks*, vol. 15, no. 4, July 2004, pp. 874-84
- [16] V. Miranda and N. Fonseca, "EPSO - Evolutionary Particle Swarm Optimization, a New Algorithm with Applications in Power Systems", *Proceedings of the IEEE Transmission and Distribution Asia-Pacific Conference 2002*, Yokohama, Japan, Oct 2002
- [17] H. Mori and Y. Komatsu, "A Hybrid Method of Optimal Data Mining And Artificial Neural Network for Voltage Stability Assessment", *Proceedings of IEEE St. Petersburg PowerTech Conference*, Russia, June 2005
- [18] N. W. Oo and V. Miranda, "Multi-energy Retail Market Simulation with Intelligent Agents", *Proceedings of IEEE St. Petersburg PowerTech Conference*, Russia, June 2005