FACULDADE DE ECONOMIA DA UNIVERSIDADE DO PORTO

TESE DE MESTRADO EM ANÁLISE DE DADOS E SISTEMAS DE APOIO À DECISÃO

Treino *On Line* de Redes Neuronais com Critérios de Informação Aplicado à Previsão Eólica

Autor

Ricardo Jorge Gomes de Sousa Bento Bessa

Orientador

Professor Doutor João Gama

Co-Orientador

Professor Doutor Vladimiro Miranda

Junho de 2008

Dedicado aos meus

pais e avós

Resumo

Nesta tese são estudadas redes neuronais MLP (*Multilayer Perceptron*) para previsão de produção eólica com treino *on line* para um horizonte temporal de 72 horas. O principal objectivo consiste em aplicar à previsão de produção eólica critérios de treino *on line* da rede neuronal, baseados na teoria da informação (*ITL - Information Theoretic Learning*).

O principal motivo para a aplicação dos critérios ITL fundamenta-se no facto de a distribuição dos erros de previsão eólica não ser Gaussiana. Mesmo que as previsões de vento exibissem erros Gaussianos, a não linearidade da curva característica dos grupos eólicos produz erros não-Gaussianos, sendo as distribuições dos erros de previsão de potência eólica assimétricas positivas e mais achatadas que as Gaussianas. Quando se tratam distribuições de erros como se fossem Gaussianas desperdiça-se informação contida nos momentos de ordem superior. Quando se utiliza um critério de mínimos quadrados ou de variância (MSE) o processo de treino apenas passa uma parte da informação dos dados para os parâmetros do sistema, deixando na distribuição dos erros uma parte da informação.

Com base no critério ITL foram estudados três critérios de treino da rede neuronal: minimização da entropia do erro (MEE), maximização da correntropia (MCC) e minimização da entropia do erro com pontos de referência (MEEF). Estes critérios de treino procuram conduzir a uma distribuição dos erros de previsão com entropia (informação) mínima, implicando maior frequência de erros próximos de zero. Os critérios combinam a definição de *Renyi* de Entropia com a técnica das janelas de *Parzen*, construindo uma medida do conteúdo de informação da distribuição dos

erros.

Tratando-se de um problema com um fluxo contínuo de dados não-estacionários foram desenvolvidas metodologias de treino *on line* da rede neuronal. O Treino *on line* permite incorporar nova informação proveniente do sistema de aquisição de dados do parque eólico, permitindo também processar um fluxo contínuo de dados sem necessidade de armazenar informação numa base de dados. Duas estratégias de treino *on line* foram desenvolvidas utilizando critérios de treino ITL: 1) retropropagação da correntropia do erro de um novo valor medido; 2) estimação recursiva da entropia do erro sobre uma janela temporal de tamanho pré-definido.

Os testes dos diferentes critérios e modos de treino foram realizados sobre três parques eólicos situados em locais distintos de Portugal Continental e com diferente complexidade de terreno. Dos resultados obtidos foi possível apurar a superioridade da qualidade das previsões obtidas com os critérios ITL, quando comparadas com as obtidas com o MSE. O treino *on line* da rede neuronal quando comparado com o treino *offline* permitiu obter melhores previsões nos três parques e num caso especial em que ocorre uma mudança de conceito.

Para o treino *offline* foi elaborada uma metodologia de estandardização dos parâmetros da rede de forma a tornar o treino da rede neuronal independente da complexidade do terreno e localização do parque eólico. No caso do treino *on line* é avaliado o impacto dos diferentes parâmetros no erro e exploradas algumas técnicas adaptativas dos parâmetros. Verificou-se que valores fixos dos parâmetros levavam a melhores resultados, no entanto, a utilização de abordagens adaptativas não deverá ser abandonada.

As redes neuronais desenvolvidas nesta tese são uma ferramenta que se revela bastante importante e útil quando integrada em sistemas de previsão mais sofisticados.

Palavras Chave: previsão eólica, redes neuronais, critério ITL, entropia, correntropia, treino *on line*.

Abstract

This thesis reports new results in adopting entropy concepts to the training of neural networks to perform wind power prediction as a function of wind characteristics (speed and direction) in wind parks connected to a power grid. Renyi's Entropy is combined with a Parzen Windows estimation of the error pdf to form the basis of two criteria (Minimum Error Entropy and Maximum Correntropy) under which neural networks are trained. Also, the merits of on line training against off line training are evaluated, as a consequence of concept drift in wind pattern behaviour and wind farm operation. The results are favourably compared with the traditional minimum square error (MSE) criterion.

When observing the literature, one realizes that models that are built relying one way or another in a training process, adopt usually the Minimum Square Error as a quality criterion to conduct the training process. Also, most proposals are based on an off line training mode, building a model that is then used to produce predictions. These options, however, rely in assumptions of Gaussianity and stationarity of the wind-electric power model, which must be strongly questioned.

First of all, using the Minimum Square Error to define the quality of a model (a neural network) assumes that the pdf (probability distribution function) of the prediction errors is Gaussian. In fact, it is well known that the power curve is highly non linear: the power curve has a linear relation with to the cube of wind speed for small values, a rather steep slope for medium wind speeds and saturation for large wind speed. Small deviations of the prediction to the real value in the steep part of the curve are transferred to large differences between power prediction and

real power. In contrast, a small deviation in the almost-linear and flat parts of the curve originates a small error in power prediction. The error distributions from wind power prediction models are usually right skewed and had positive excess of kurtosis, meaning that they were asymmetrical, presents a higher frequency of errors to the left of the mean and were flatter than the Gaussian distribution. These reasons have motivated the use of techniques that would train adaptive systems (i.e. neural networks) based on minimizing the information content of the error distribution instead of minimizing its variance. A measure of information content is entropy and incorporating entropy as a corner-stone concept in training mappers has been the object of ITL (Information Theoretic Learning).

An off line training assumes stationarity, i.e., the model once trained remains a stable representation of the real system. However, it has been observed that wind speed behaviour and wind farm operation exhibits a trait called concept drift in the vocabulary of specialists of data streams, meaning that it changes characteristics over time. Therefore, a model trained off line will display, after some time, a pattern of growing error in prediction values.

This thesis brings further contributions to build better wind power forecasting models. It describes the newly proposed criteria of Minimum Entropy and of Maximum Correntropy and presents, for three real cases of wind parks in Portugal, the results from neural networks, trained to produce 72-hour ahead forecasts, comparing the performance of models trained off line and on line under three paradigms: MSE (Minimum Square Error), MEE (Minimum Entropy Error) and MCC (Maximum Correntropy Criterion). The results reported in this thesis confirm that on line training may lead to better prediction models, allowing to adequately deal with data streams in the presence of concept drift or concept changes. This characteristic is present not only in wind behaviour but also in the practical operation of wind parks, namely due to the variation in the available generating capacity, either because of maintenance, failure or simply because of capacity additions.

Keywords: Wind power forecasting, neural networks, ITL criterion, correntropy, entropy, on line train

Prefácio

Esta tese foi elaborada no âmbito do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão. O assunto desta tese é o treino *on line* de redes neuronais aplicado a um problema real de previsão eólica baseado em critérios de treino da teoria da informação (*ITL - Information Theoretic Learning*). As principais razões para o desenvolvimento desta tese prendem-se com os elevados níveis de produção eólica no sistema produtor de muitos países europeus, o que motiva o desenvolvimento de ferramentas de apoio à decisão para produtores e operadores da rede eléctrica. A previsão eólica possuiu diversas aplicações, que vão desde a operação segura da rede eléctrica por parte do operador até à participação no mercado de electricidade por parte do produtor.

Gostaria neste espaço, de agradecer a todas as pessoas e entidades que de uma forma ou de outra contribuíram para o desenvolvimento desta tese e cuja contribuição foi valiosa. Assim, agradeço:

Ao Professor Doutor João Gama, por me ter apresentado os conceitos de treino *on line*, pela sua disponibilidade, conselhos e orientação.

Ao Professor Doutor Vladimiro Miranda, por me ter apresentado os conceitos ITL e a sua aplicação à previsão eólica, pela confiança depositada em mim, pela sua disponibilidade, conselhos e motivação.

Aos meus colegas do INESC Porto e ao corpo docente deste Mestrado, que contribuíram para a minha evolução intelectual e a abertura a novas áreas cientificas.

Ao Professor Doutor Cláudio Monteiro que me apresentou ao problema de previsão eólica, os seus conselhos e ideias foram determinantes para iniciar esta tese.

Ao Engenheiro João Sousa, que em conjunto há dois anos trabalhamos com o problema de previsão eólica. A sua amizade e disponibilidade na revisão desta tese foram determinantes para a sua conclusão.

Ao INESC Porto pela disponibilização dos dados e ao Professor Doutor Pedro Miranda pelas previsões do modelo MM5.

Aos amigos António Pina, Ana Vanessa, Joaquim Matos, Lourenço Moura, Bruno Ferreira, André Moreira, pela amizade e pelos incansáveis incentivos.

Abreviaturas e Notação Matemática

Abreviaturas

CIM	*Correntropy Induced Metric*
DPN	Desvio Padrão Normalizado
ECMWF	*European Centre for Medium-Range Weather Forecasts*
EMAN	Erro Médio Absoluto Normalizado
EPSO	*Evolutionary Particle Swarm Optimization*
EWEA	*Wind Energy Association*
fdp	Função Densidade de Probabilidade
HIRLAM	*High Resolution Limited Area Model*
IF	Força de Informação
IP	Potêncial de Informação
iRPROP	*Improved Resilient Backpropagation*
ITL	*Information Theoretic Learning*
LMS	*Least Mean Square*
MCC	*Maximization Correntropy Criterion*/Maximização da Correntropia
MEE	*Minimization Error Entropy*/Minimização da Entropia dos Erros
MEEF	*Minimization Error Entropy with Fiducial Points*

MLP	*MultiLayer Perceptron*
MM5	Modelo Meso-escala
MSE	*Mean Square Error*/Minimização do Erro Quadrático
NBIAS	*Bias* Normalizado
NWP	*Numerical Weather Prediction*
RBF	*Radial Basis Function*
REMQN	Raiz do Erro Médio Quadrático Normalizado
RIU	*Remote Interface Unit*
RPROP	*Resilient Backpropagation*
SCADA	*Supervisory Control and Data Acquisition*
SEN	Sistema Eléctrico Nacional
UTM	*Universal Transverse Mercator*
v.a.	Variável Aleatória

Notação Matemática

P_n	Potência nominal do aerogerador
$h_(v)$	Distribuição de Weibull
p	Potência eólica produzida
$\hat{p}$	Potência eólica prevista
t	Tempo
$\hat{p}_{t+k\|k}$	Previsão da produção para o instante $t+k$ dado que se está no instante t
pz	neurónio z da camada p
O_{pz}	saída do neurónio z da camada p
$w_{kx,pz}$	Peso da ligação entre o neurónio pz e neurónio kx
net_{kx}	saída do somatório do neurónio x da camada k
f_{kx}	função de activação do neurónio x da camada k
b_{kx}	*bias* do neurónio x da camada k
$\Delta w_{kx,pz}$	Termo de actualização dos pesos da ligação entre o neurónio pz e neurónio kx
η	Taxa de aprendizagem
E	Erro Quadrático
ω	Termo de momento
e_i	Erro do exemplo i
H_s	Entropia de *Shannon*
p_k	Probabilidade de $x = x_k$
$H_{R\alpha}$	Entropia de *Renyi* de ordem α
H_{R2}	Entropia quadrática de *Renyi*
$f_Y(z)$	função de densidade de probabilidade da v.a. Y
K	função *Kernel*
σ	tamanho da janela de *Parzen* e desvio padrão

G	*Kernel* Gaussiano
N	Número de exemplos de treino
λ	Factor de memória
M	Tamanho da janela da estimação recursiva da entropia
$V_{\sigma)}$	Correntropia com janela de *Parzen* de tamanho σ
$E\{\}$	Valor esperado
D_i	Valor desejado à saída da rede
y_i	Saída da rede
γ	Número de pontos de referência ou peso do critério MCC no critério MEEF
$g(x, w)$	Função genérica que transforma as entradas x em saídas y, através dos pesos w
$\bar{e}$	Média do erro
$card$	Cardinal
κ	Parâmetro da taxa de aprendizagem adaptativa
τ	Parâmetro da taxa de aprendizagem adaptativa
η_+	Factor de incremento
η_-	Factor de decremento
Δ_{min}	Limite mínimo de actualização dos pesos do iRPROP
Δ_{max}	Limite máximo de actualização dos pesos do iRPROP
Δ_0	Valor inicial de actualização dos pesos do iRPROP
v	Velocidade do vento
d_{sen}	Componente seno da direcção do vento
d_{cos}	Componente cosseno da direcção do vento
m_{sen}	Componente seno da antiguidade das previsões do MM5
m_{cos}	Componente cosseno da antiguidade das previsões do MM5

Conteúdo

Lista de Figuras **xv**

Lista de Tabelas **xxii**

1 Introdução **1**

1.1 Contexto Geral 1

1.2 Previsão Eólica 2

1.3 Objectivos da Tese 4

1.3.1 Aprendizagem com Critérios de Informação 4

1.3.2 Treino *On Line* 5

1.4 Principais Contribuições da Tese 6

1.5 Estrutura da Tese 7

2 Descrição do Problema de Produção de Energia Eólica **9**

2.1 Relação entre Potência e Velocidade do Vento 9

2.2 Natureza do Vento 11

2.2.1 Variação Anual e Sazonal 13

2.3 Parâmetros que Condicionam o Potencial Eólico 14

2.4 Variabilidade da Produção Eólica em Horizontes Curtos 15

2.5 Fluxo de Dados de um Parque Eólico 17

2.5.1 SCADA 17

2.5.2 Estação Meteorológica 19

2.5.3 Modelo de Previsão Meteorológica Numérica 20

3 Estado da Arte **27**

3.1 Formulação da Previsão Eólica 28

3.2 Métodos de Previsão de Referência 29

3.3 Horizonte de Previsão 30

3.4 Distribuição dos Erros de Previsão 31

3.4.1 Sub-amostragem 33

3.5 Abordagem Estatística 34

3.5.1 Função de Custo 35

3.5.2 Modelos com Treino *On Line* 39

4 Treino de Sistemas Adaptativos **42**

4.1 Conceitos Básicos de Redes Neuronais 43

4.1.1 Termo de Momento 45

4.1.2 Modos de Treino 45

4.2 Funções de Custo 47

4.3 Aprendizagem Baseada em Teoria da Informação 49

4.3.1 Entropia 49

4.3.2 Janelas de Parzen e Estimação da Entropia 50

4.3.3 Correntropia 55

4.3.4 Critérios ITL para Treino de Sistemas Adaptativos 58

4.3.5 Algoritmos de Retropropagação dos Critérios ITL 62

4.4 Optimização do Algoritmo de Treino 64

4.4.1 Treino *Off Line* 64

4.4.2 Treino *On Line* 68

5 Teste e Avaliação dos Modelos de Previsão Eólica **71**

5.1 Descrição dos Parques Eólicos 72

5.1.1 Pré-processamento dos Dados 73

5.2 Teste *Off Line* para um Horizonte Temporal de 24 Horas 75

5.2.1 Conjunto de Dados 75

5.2.2 Metodologia de Teste e Avaliação 76

5.2.3 Características da Rede Neuronal 77

5.2.4 Análise dos Resultados 78
5.3 Previsão Eólica para um Horizonte Temporal de 72 Horas 82
5.3.1 Conjunto de Dados 82
5.3.2 Metodologia de Teste e Avaliação 84
5.3.3 Comparação *Off Line* de Critérios de Treino 85
5.3.4 Avaliação do Impacto dos Parâmetros da Rede Neuronal no Treino *On Line* 94
5.3.5 Comparação entre o Treino *On Line* e *Off Line* da Rede Neuronal 100

6 Conclusões **108**
6.1 Resumo das Principais Contribuições e Resultados 108
6.2 Trabalho Futuro 110

Bibliografia **112**

A Algoritmos de Treino Batch e Sequencial **124**

B Forças de Informação **127**

C Algoritmos de Retropropagação dos Critérios ITL **129**

D Testes Preliminares da Estrutura da Rede Neuronal **132**

E Algoritmos de Teste de Previsões para um Horizonte de 72 Horas **135**

F Métricas para Avaliação das Previsões **137**

G Resultados *Off Line* para 72 Horas **139**

H Variação dos Parâmetros do Treino *On Line* **142**

I Resultados do Treino *On Line* para 72 horas **150**

J Artigos Publicados **152**

Lista de Figuras

2.1 *Curva de Potência de um aerogerador com potência nominal de 1800 kW.* 10

2.2 *Espectro da velocidade do vento de Brookhaven, baseado no estudo de Van der Hoven.* 13

2.3 *Sistema SCADA de comunicação de um parque eólico* 17

2.4 *Produção em função velocidade do vento registada pelo SCADA de um aerogerador.* 19

2.5 *Malha de um modelo global.* 23

2.6 *Aninhamento de domínios na zona de Portugal.* 25

4.1 *Esquema de um neurónio biológico.* 43

4.2 *Operações executadas num nó genérico, designado como de ordem x na camada k, recebendo sinais da camada anterior p, depois de afectados pelos pesos w das ligações.* 44

4.3 *Representação de um neurónio x na camada k produzindo uma resposta* O_{kx} *com m entradas oriundas da camada anterior p e ainda um limiar* b_{kx} *independente do processamento das camadas anteriores.* 44

4.4 *Ilustração gráfica do impacto de* σ *na estimação da fdp com o método das janelas de Parzen.* 52

4.5 *Curvas de nível de CIM(X,0) num espaço bi-dimensional com* $\sigma = 1$. . 57

4.6 *Erro quadrático* $MSE(X,Y) = (Y - X)^2$. 60

4.7 *Correntropia* $V(X,Y) = k(Y - X)$ *com* $\sigma = 1$. 60

4.8 *Instabilidade no processo de treino provocada por um valor de h em função do desvio padrão.* 67

5.1 *Erro de previsão do MM5 no parque A e aproximação pela distribuição normal (média=-0.107 desv. padrão=2.90).* 73

5.2 *Erro de previsão do MM5 no parque B e aproximação pela distribuição normal (média=-0.66 desv. padrão=3.50).* 73

5.3 *fdp do erro dos critérios MEE e MEEF para o Parque A.* 79

5.4 *fdp do erro dos critérios MEEF,MCC e MSE no Parque A.* 79

5.5 *fdp do erro dos critérios MEE e MEEF para o Parque B.* 79

5.6 *fdp do erro dos critérios MEEF,MCC e MSE no Parque B.* 79

5.7 *fdp do erro dos critérios MEE e MEEF para o Parque C.* 80

5.8 *fdp do erro dos critérios MEEF,MCC e MSE no Parque C.* 80

5.9 *Esquema temporal das previsões meteorológicas do modelo MM5.* 84

5.10 *EMAN para o Parque A no modo de treino off line.* 86

5.11 *REMQN para o Parque A no modo de treino off line.* 86

5.12 *NBIAS para o Parque A no modo de treino off line.* 87

5.13 *NBIAS para o Parque B no modo de treino off line.* 88

5.14 *NBIAS para o Parque C no modo de treino off line.* 90

5.15 *Frequência dos valores medidos de produção do Parque A e dos valores previstos com os diferentes critérios.* 91

5.16 *NBIAS do erro de previsão obtido com os diferentes critérios em função da classe de produção do parque eólico A.* 91

5.17 *Desvio padrão normalizado da distribuição dos erros de previsão em função das classes de previsão para o Parque A.* 93

5.18 *Previsão e valor medido para os dias 26-28 de Outubro de 2006 para o Parque A.* 94

5.19 *Previsão e valor medido para os dias 10-12 de Novembro de 2006 para o Parque C.* 94

5.20 *EMAN obtido com o critério MSE para o parque A em função do horizonte temporal para diferentes valores da taxa de aprendizagem.* 96

5.21 *EMAN obtido com o critério MCC para o Parque A em função do horizonte temporal para diferentes valores da taxa de aprendizagem.* . . . 96
5.22 *EMAN obtido com o critério MSE para o Parque A em função do horizonte temporal para diferentes valores do termo de momento.* 97
5.23 *EMAN obtido com o critério MCC para o Parque A em função do horizonte temporal para diferentes valores do termo de momento.* 97
5.24 *REMQN obtido com o critério MSE para o parque A em função do horizonte temporal para diferentes valores do termo de momento* . . . 98
5.25 *REMQN em função do horizonte temporal para diferentes valores de* λ *e com* $M = 4000$. 99
5.26 *REMQN em função do horizonte temporal para diferentes valores de* M *e com* $\lambda = 0.08$. 99
5.27 *EMAN obtido com o critério MCC para o Parque A em função do horizonte temporal para diferentes valores da janela de Parzen.* 100
5.28 *EMAN obtido com o critério MEEF para o Parque A em função do horizonte temporal para diferentes valores da janela de Parzen.* 100
5.29 *EMAN em função do horizonte temporal obtido com o treino on line para o Parque A.* 101
5.30 *REMQN em função do horizonte temporal obtido com o treino on line para o Parque A.* 101
5.31 *NBIAS em função do horizonte temporal obtido com o treino on line para o Parque A.* 102
5.32 *NBIAS em função do horizonte temporal obtido com o treino on line para o Parque B.* 103
5.33 *EMAN em função do horizonte temporal obtido com o treino on line para o parque C.1.* 104
5.34 *REMQN em função do horizonte temporal obtido com o treino on line para o parque C.1.* 104
5.35 *NBIAS em função do horizonte temporal obtido com o treino on line para o parque C.1.* 105

5.36 *Comparação do EMAN obtido num novo treino da rede usando todo o ano de 2005 e com uma rede treinada em modo on line.* 106
5.37 *Comparação do REMQN obtido num novo treino da rede usando todo o ano de 2005 e com uma rede treinada em modo on line.* 106
5.38 *Previsão e valor medido para os dias 09-11 de Junho de 2006 para o Parque A.* 107
5.39 *Previsão e valor medido para os dias 12-14 de Novembro de 2005 para o parque C.1.* 107

D.1 *Evolução do erro no modo de treino off line para diferentes configurações da rede para o Parque A.* 134
D.2 *Evolução do erro no modo de treino off line para diferentes configurações da rede para o Parque B.* 134
D.3 *Evolução do erro no modo de treino on line para diferentes configurações da rede para o Parque A.* 134
D.4 *Evolução do erro no modo de treino on line para diferentes configurações da rede para o Parque B.* 134

G.1 *EMAN para o Parque B no modo de treino off line.* 139
G.2 *REMQN para o Parque B no modo de treino off line.* 139
G.3 *EMAN para o Parque C no modo de treino off line.* 140
G.4 *REMQN para o Parque C no modo de treino off line.* 140
G.5 *Frequência dos valores medidos de produção do Parque B e dos valores previstos com os diferentes critérios.* 140
G.6 *Frequência dos valores medidos de produção do Parque C e dos valores previstos com os diferentes critérios.* 140
G.7 *NBIAS do erro de previsão dos diferentes critérios em função da classe de produção do parque eólico B.* 141
G.8 *NBIAS do erro de previsão dos diferentes critérios em função da classe de produção do parque eólico C.* 141
G.9 *Desvio padrão da distribuição dos erros de previsão em função das classes de previsão para o Parque B.* 141

G.10 *Desvio padrão da distribuição dos erros de previsão em função das classes de previsão para o Parque C.* 141

H.1 *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque A com o critério MSE.* 142

H.2 *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque A com o critério MCC.* 142

H.3 *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque B com o critério MSE.* 143

H.4 *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque B com o critério MCC.* 143

H.5 *EMAN obtido com o critério MSE para o Parque B em função do horizonte temporal para diferentes valores da taxa de aprendizagem.* . . . 143

H.6 *EMAN obtido com o critério MCC para o Parque B em função do horizonte temporal para diferentes valores da taxa de aprendizagem.* . . . 143

H.7 *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque C com o critério MSE.* 144

H.8 *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque C com o critério MCC.* 144

H.9 *EMAN obtido com o critério MSE para o Parque C em função do horizonte temporal para diferentes valores da taxa de aprendizagem.* . . . 144

H.10 *EMAN obtido com o critério MCC para o Parque C em função do horizonte temporal para diferentes valores da taxa de aprendizagem.* . . . 144

H.11 *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque A com o critério MCC.* 145

H.12 *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MSE.* 145

H.13 *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MCC.* 145

H.14 *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MSE.* 146

H.15 *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MCC.* 146
H.16 *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MSE.* 146
H.17 *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MCC.* 146
H.18 *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MSE.* 147
H.19 *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MCC.* 147
H.20 *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque A com o critério MCC.* 147
H.21 *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque A com o critério MEEF.* 147
H.22 *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MCC.* 148
H.23 *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MEEF.* 148
H.24 *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MCC.* 148
H.25 *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MEEF.* 148
H.26 *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MCC.* 149
H.27 *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MEEF.* 149
H.28 *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MCC.* 149
H.29 *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MEEF.* 149

I.1 *EMAN em função do horizonte temporal obtido com o treino on line para o parque B* 150
I.2 *REMQN em função do horizonte temporal obtido com o treino on line para o Parque B.* 150
I.3 *EMAN em função do horizonte temporal obtido com o treino on line para o Parque C.* 151
I.4 *REMQN em função do horizonte temporal obtido com o treino on line para o Parque C.* 151
I.5 *NBIAS em função do horizonte temporal obtido com o treino on line para o Parque C.* 151

Lista de Tabelas

2.1 *Escalas espaço temporais para modelos atmosféricos.* 23

5.1 *Conjunto de dados dos três parques eólicos.* 74

5.2 *Conjunto de dados do treino off line do parque A.* 75

5.3 *Conjunto de dados do treino off line do parque B.* 76

5.4 *Conjunto de dados do treino off line do parque C.* 76

5.5 *Bias relativo à potência instalada do erro nos três parques eólicos.* . . 80

5.6 *Avaliação dos erros de previsão sob diferentes critérios:* H_{R2} *entropia do erro; EMQ – erro médio quadrático (% potência instalada); variância das simulações.* 82

5.7 *Meses dos conjuntos de validação, teste e treino dos três parques eólicos.* 83

D.1 *Características dos dados do Parque A no ano de 2005.* 132

D.2 *Características dos dados do Parque B no ano de 2005.* 132

Capítulo 1

Introdução

1.1 Contexto Geral

A energia eólica com uma potência instalada em toda a Europa que excede os 40 GW é actualmente a fonte de energia com maior crescimento tecnológico. Os objectivos traçados pela indústria e pela Comissão Europeia para o desenvolvimento da energia eólica durante a última década foram largamente excedidos, tendo a European Wind Energy Association (EWEA) definido novas metas a atingir. Os países membros da Comissão Europeia devem ter uma potência instalada total de 75 GW em 2010 e 180 GW em 2020 (Zervos, 2003), o que representa um aumento de 25% das metas definidas em 2000 e quase o dobro das metas definidas em 1997. Devido a uma grande evolução tecnológica na produção de energia eólica e ao rápido desenvolvimento do mercado a energia eólica é essencial para a União Europeia cumprir as metas definidas em (Commission, 2001). A Meta consiste em aumentar a percentagem de produção com recurso a energia renovável de 14% em 1997 para 22.1% em 2020.

Para Portugal foi estabelecido que em 2010 39% da electricidade consumida deverá ter proveniência "renovável". Considerando o crescimento previsível do consumo e considerando as hipóteses de outras tecnologias estima-se que para atinigir este objectivo, seria necessário instalar mais de 3750 MW de potência eólica. Ao longo do 1º semestre de 2007 entraram em actividade nove parques eólicos, representando um crescimento de cerca de 15% na potência ligada à rede que no final

do semestre era de 1731 MW (correspondentes a 148 parques). A potência eólica instalada actualmente representa cerca de 17% do total da capacidade instalada no SEN(Sistema Eléctrico Nacional), ou seja, o dobro da contribuição verificada em fins de 2005. Do conjunto dos parques em funcionamento, o de maior dimensão tem 121 MW e o menor 0.5 MW (REN, 2007).

Em Portugal o crescente interesse pelo desenvolvimento de ferramentas de previsão de produção com base eólica deve-se ao concurso público criado em 2006 com o objectivo de promover um "Cluster Industrial Eólico", que entre vários pontos, engloba a atribuição de direitos de ligação para 1700 MW de potência com base eólica. Está prevista a construção de três centros de despacho para gestão da produção eólica, para os quais será necessário desenvolver ferramentas de previsão eólica.

1.2 Previsão Eólica

As estratégias para a operação em segurança da rede eléctrica são baseadas em previsões das condições futuras de várias variáveis, nomeadamente em previsões de consumo e produção. Em sistemas eléctricos com elevada potência eólica instalada (como é o caso de Portugal e Espanha) a previsão eólica torna-se uma importante informação para a definição de estratégias de operação de forma a minimizar os custos e assegurar o desempenho e segurança do sistema (Doherty & O'Malley, 2005). O operador da rede eléctrica monitoriza a área de operação em tempo real e efectua ajustes quando as condições actuais se desviam das que foram previstas, estes ajustes e desvios dos planos óptimos de operação e actividades programadas incrementam os custos de operação (Agency, 2005).

Não existem dúvidas sobre os benefícios da expansão da produção de energia com recurso a fontes eólicas. No entanto, quanto mais energia eólica for ligada ao sistema eléctrico mais importante será perceber e gerir o impacto da produção eólica nas operação do sistema.

Segundo (Ahlstrom et al., 2005), a previsão de produção eólica é uma ferramenta importante quando é necessário integrar na rede grandes quantidades de potência eólica, também a forma como se usa a previsão é muito importante. A previsão é um

desperdício se não for integrada de modo a que a informação seja útil na operação do sistema e nos mercados eléctricos. Se os operadores do sistema não têm confiança nas previsões, ou não é disponibilizada de uma forma que lhe seja útil utilizar, são tentados a operar o sistema num modo mais conservador aumentando assim os custos da integração da energia eólica.

Actualmente, o desenvolvimento e expansão da energia eólica na Europa ocorre paralelamente ao processo de desrregularização dos mercados eléctricos de energia. Isto significa que mesmo a energia eólica, beneficiando de incentivos, em poucos anos terá que ser considerada igual às outras fontes energéticas que participam nos mercados eléctricos. Isto implica que os produtores serão penalizados no caso de incumprimento das propostas de venda de energia apresentadas no mercado, sendo a variabilidade do recurso eólico uma fonte de penalizações num ambiente de mercado (Milborrow, 2001). Esta situação já se verifica nos países nórdicos com a participação dos produtores de energia eólica no chamado *Nord Pool*.

Dependendo da falta de energia ou excesso (devido a avarias no sistema produtor ou à variabilidade da produção com base eólica), o operador de mercado determina penalizações que serão pagas pelos produtores eólicos que falham as suas obrigações. Em (Pinson et al., 2004) e (Pinson et al., 2007a) é demonstrado que o uso de ferramentas avançadas de previsão aumentam substancialmente o lucro de um produtor eólico, reduzindo a diferença entre o valor proposto no mercado e o valor produzido.

Uma forma de compensar o impacto da variabilidade do recurso eólico (e o erro de previsão) consiste em recorrer a sistemas de armazenamento de energia. A associação entre sistemas de armazenamento e parques eólicos é de todo o interesse para o promotores eólicos, isto porque permitem diminuir as penalizações no mercado de ajustes. É possível com esta associação controlar o fornecimento de energia, podendo-se armazenar energia eólica em períodos com baixos preços (ou períodos com elevada velocidade do vento) e mais tarde vender energia quando o preço de mercado é elevado (Korpaas et al., 2003) (Bathurst & Strbac, 2003)(Castronuovo & Peças Lopes, 2004). É reconhecido na literatura que a utilização de métodos de armazenamento de energia em conjunto com sistemas de previsão tornará possível

optimizar os benefícios da venda de energia com base eólica.

A previsão de produção eólica de curto prazo para um horizonte temporal até 72 horas é constituída por duas fases: primeiro, a ”fase meteorológica”, que consiste em previsões do vento para as próximas horas ou dias ao nível do parque eólico; a segunda fase, a ”conversão de energia”, envolve a transformação da velocidade do vento em potência produzida. Normalmente, a primeira fase é baseada num modelo de previsão numérica das condição atmosféricas (NWP - *Numerical Weather Prediction*) que fornece uma malha em torno do parque eólico com previsões das variáveis que caracterizam o vento. A segunda fase corresponde a modelizar o sistema de conversão da velocidade do vento em potência (chamado de curva de potência) do parque eólico, tendo em consideração a curva de potência individual de cada aerogerador, as características do terreno, a densidade do ar, a turbulência e outros parâmetros que influenciam a produção.

Das duas fases apresentadas acima, a que provoca maior impacto no erro de previsão é a meteorológica. É fácil imaginar que quanto melhor for a previsão do vento, melhor será a previsão da produção do parque eólico. Esta afirmação é sustentada por vários estudos que procuraram quantificar a importância das previsões NWP no erro de previsão da produção (Holttinen & Hirvonen, 2005) (Lange, 2003).

1.3 Objectivos da Tese

1.3.1 Aprendizagem com Critérios de Informação

Uma metodologia de previsão ideal para a previsão eólica teria um erro sistemático nulo e parte aleatória do erro com uma forma de ruído branco (Pinson, 2006). O ruído branco é visto como uma sequência de erros independentemente distribuídos que podem ser aproximados por uma distribuição normal com média nula (de acordo com o Teorema do Limite Central). No entanto os erros de previsão de produção eólica na prática não são Gaussianos e são geralmente correlacionados como demonstrado em (Nielsen, 2002). O problema da correlação entre erros não será abordado nesta tese, apenas a distribuição não-Gaussiana dos erros será alvo de estudo.

A grande maioria dos métodos de previsão eólica assume que a distribuição dos erros é Gaussiana e que os dados apresentam um carácter estacionário, sendo que estes dois pressupostos deverão ser fortemente questionados.

A quase generalidade dos métodos de previsão utiliza como critério de treino (função de custo) a minimização do erro médio quadrático. O uso deste critério para treino de um sistema adaptativo (uma rede neuronal) assume que a função densidade de probabilidade (fdp) dos erros de previsão é Gaussiana, sendo esta distribuição a única que é totalmente caracterizada pelos seus dois primeiros momentos (média e variância). A assimetria e o achatamento são propriedades que não são capturadas por um método que apenas se baseia na minimização da variância, existindo momentos de ordem superior que contêm informação relevante que deveria ser "transferida" para os parâmetros (pesos) da rede neuronal em vez de permanecerem na distribuição do erro.

A conversão de velocidade do vento em potência (curva de potência) num aerogerador é um processo não-linear, a potência tem uma relação linear com o cubo da velocidade para valores pequenos da velocidade, uma parte não linear para valores intermédios do cubo da velocidade e uma parte saturada para valores elevados do cubo da velocidade. A transformação não-linear altera as propriedades estatísticas da distribuição do erro de previsão do modelo NWP. O facto de se estar na presença de um problema de previsão com características não-Gaussianas motiva a utilização de um critério de treino baseado no conteúdo de informação da distribuição dos erros de previsão. Uma medida do conteúdo de informação é a entropia, a sua incorporação como critério de treino de sistemas adaptativos foi alvo de estudo no âmbito do critério ITL (*Information Theoretic Learning*) (Principe et al., 1999b) (Principe & Erdogmus, 2006).

O objectivo é a aplicação de critérios de treino de redes neuronais baseados no critério ITL ao problema de previsão de produção eólica.

1.3.2 Treino *On Line*

O uso de um treino *off line* assume que os dados são estacionários e apenas é necessário treinar o modelo uma vez, permanecendo um modelo estático e sem necessi-

dade de integrar nova informação no treino. Existem no entanto ao longo da exploração de parque eólico inúmeros problemas e situações que provocam uma operação não-estacionária, alguns exemplos são: alteração da potência instalada do parque eólico com a instalação de novas máquinas ou com o incremento da potência das máquinas; construção de um novo parque eólico nas proximidades do parque em estudo; avaria/manutenção prolongada de uma ou várias máquinas; limitação da potência produzida pelo parque eólico[1]; degradação do desempenho das máquinas do parque; construção de um parque eólica nas proximidades; mudança do modelo de previsão meteorológica (ou dos parâmetros do modelo) que fornece as previsões de velocidade e direcção; alteração das condições atmosféricas e regime de ventos do parque.

Todos estes acontecimentos são muito comuns em parques eólicos, sendo absolutamente necessário possuir modelos que sejam adaptativos e que incorporem informação proveniente de um novo conceito. Para além de tornar possível lidar com alterações do conceito, também permite processar um fluxo contínuo de dados sem necessidade de armazenar a informação numa base de dados.

O objectivo consiste em aplicar treino *on line*[2] de redes neuronais ao problema de previsão eólica, onde nova informação disponível sobre a produção do parque eólico é utilizada para actualizar o modelo de previsão.

1.4 Principais Contribuições da Tese

As contribuições desta Tese para o estado da arte de previsão eólica são:

- Treino de redes neuronais com os três critérios (MEE, MCC e MEEF) que derivam do critério ITL aplicado ao problema de previsão da produção eólica. Aplicação do critério de maximização da correntropia em dados reais;

- Avaliação da utilidade económica das previsões com os critérios ITL para a participação do produtor eólico no mercado de electricidade com propostas de

[1] Por vezes é útil para o promotor limitar a um máximo a potência produzida nas máquinas, podendo ser feito com recurso a electrónica de potência ou software.

[2] ao longo desta tese a palavra *on line* será utilizada em vez de ”tempo real”.

venda de energia. Avaliação do erro de previsão obtido para um horizonte temporal de 72 horas com a aplicação dos critérios ITL em três parques eólicos reais;

- Estandardização dos parâmetros da rede neuronal de forma a tornar o treino da rede neuronal independente da complexidade do terreno e da localização do parque eólico;

- Aplicação e avaliação da utilidade do treino *on line* de redes neuronais nos três parques eólicos e numa situação real com alteração da operação do parque. Aplicação do treino *on line* com cálculo recursivo da entropia do erro em dados reais;

- Estudo sobre o impacto dos diferentes parâmetros da rede neuronal no treino *on line*.

1.5 Estrutura da Tese

A tese encontra-se dividida em seis capítulos, seguindo uma sequência lógica para o desenvolvimento do tema abordado. No primeiro capítulo é apresentada a utilidade da previsão eólica e o fim a que se destina, também é abordado o facto de se estar na presença de um problema não-Gaussiano com dados não-estacionários.

O segundo capítulo consiste numa descrição dos conceitos básicos e essenciais sobre a energia eólica, bem como das diferentes variáveis e dados do parque eólico. Neste capítulo é igualmente descrito o modelo de previsão das condições meteorológicas que é a principal fonte de dados dos sistemas de previsão.

O terceiro capítulo consiste no estado da arte da previsão eólica, sendo dada especial ênfase às diferentes funções de custo e modos de treino utilizados. Também neste capítulo encontra-se uma formulação matemática do problema de previsão eólica. O quarto capítulo aborda os conceitos básicos sobre redes neuronais e os critérios de treino de sistemas adaptativos que derivam do critério ITL.

No quinto capítulo são analisados os três casos de estudo com o objectivo de avaliar os diferentes critérios e modos de treino das redes neuronais. São realizadas

previsões para horizontes temporais completos de 24 e 72 horas em modo *off line* e *on line*.

No sexto e último capítulo são apresentadas conclusões sobre o trabalho desenvolvido e sugestões para um trabalho futuro.

Capítulo 2

Descrição do Problema de Produção de Energia Eólica

Neste capítulo serão expostos os conceitos e especificidades da produção de energia eólica. Serão explicados conceitos relevantes sobre a tecnologia eólica para a previsão de produção, em especial o sistema de conversão de velocidade do vento em potência e a elevada variabilidade da produção eólica.

Serão também descritas as várias fontes de dados de um parque eólico, bem como o modelo previsão das condições meteorológicas NWP que fornece previsões das condições meteorológicas para um horizonte que poderá ir até 72 horas.

2.1 Relação entre Potência e Velocidade do Vento

Uma turbina eólica capta a energia cinética do vento com um rotor constituído por duas ou mais pás acoplado a um gerador eléctrico. O sistema de conversão de energia eólica divide-se em três partes: rotor, cabine e torre.

O rotor em relação à superfície de ataque do vento incidente nas pás pode ser colocado a montante ou a jusante da torre. A opção *upwind*, em que o vento ataca as pás pelo lado da frente, generalizou-se devido ao facto de o vento incidente não ser perturbado pela torre. Na cabina estão alojados, entre outros equipamentos, o veio principal, o travão de disco, a caixa de velocidades (quando existe), o gerador e o mecanismo de orientação direccional. A torre suporta a *nacelle* e eleva o rotor até

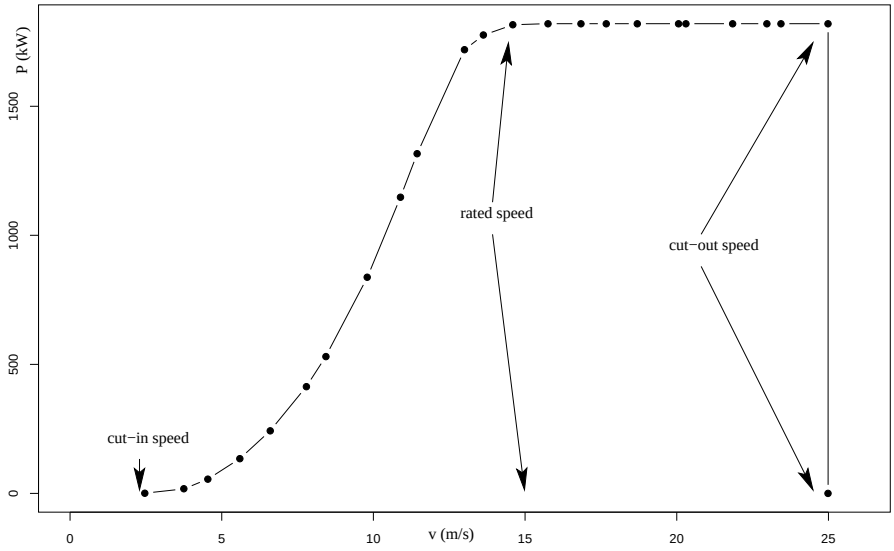

Figura 2.1: *Curva de Potência de um aerogerador com potência nominal de 1800 kW.*

uma cota em que a velocidade do vento é maior e menos perturbada do que junto ao solo. O processo de conversão de energia de um aerogerador é descrito pela sua curva característica, conhecida como curva de potência. As curvas de potência independentemente do fabricante não apresentam grandes diferenças na sua forma, a figura 2.1 dá o exemplo da curva de potência de um aerogerador com potência nominal de 1800 kW. Não existe produção para velocidades do vento inferiores à *cut-in speed* (aprox. 2-3 m/s), a produção aumenta abruptamente entre a velocidade *cut-in* e *rated speed* (aprox. 12-16 m/s), sendo que na *rated speed* atinge uma produção muito próxima da produção nominal P_n. A produção é quase constante entre a *rated speed* e a *cut-out speed* (aprox. 25-30 m/s), parando a turbina na *cut-out speed* por razões de segurança. O valor nominal da turbina P_n é uma referência da capacidade de produção do aerogerador. A rápida subida da curva característica para baixas velocidade do vento é explicada pela equação que determina a relação entre as duas variáveis. A potência, $P(W)$, é proporcional ao cubo da velocidade do vento, e dada por:

$$P = \frac{1}{2}\rho A v^3 \tag{2.1}$$

onde $A(m^2)$ é a secção plana transversal do rotor da turbina e ρ a massa específica do ar ($1.225\,kg/m^3$ em condições de pressão e temperatura normais.

Na maior parte do tempo está-se na zona não linear da curva de potência porque menores velocidades do vento são mais comuns do que velocidades elevadas. É visível que a forma da curva irá influenciar significativamente estudos de avaliação da performance de um parque eólico e o erro de um modelo de previsão. Se a previsão original da velocidade do vento apresenta um ligeiro desvio do valor medido, na parte não linear da curva este ligeiro desvio é transferido para uma diferença relativamente grande entre a potência prevista e a potência medida. Em contraste, na parte saturada da curva este pequeno desvio também resulta numa ligeira diferença entre a potência prevista e potência medida.

2.2 Natureza do Vento

A energia eólica tem a sua origem na energia solar, como a maior parte das fontes energéticas do planeta. O aquecimento de massas de ar como consequência da radiação solar contribuiu para a aparição de correntes ascendentes, sendo que o espaço que libertam é ocupado por outras massas adjacentes de ar mais frio. O movimento destas massas de ar provoca a aparição de diferenças de pressão, que por sua vez, contribuem para o próprio movimento de ar. A variação cúbica da energia produzida por uma turbina eólica com a velocidade do vento torna crucial perceber as características do recurso para identificar os locais passíveis de construção de parques, avaliação económica dos investimentos, projecto de aerogeradores e previsão de produção eólica. Do ponto de vista da produção de energia a característica mais crucial do recurso é a sua variabilidade. O vento tem uma variabilidade geográfica e temporal, sendo esta amplificada pela relação cúbica com a produção.

Numa larga escala a variação espacial descreve o facto de existirem diferentes climas nas várias regiões do mundo, uns mais ventosos que outros, estando este aspecto dependente da quantidade de irradiação solar. Estas regiões são ditadas pela latitude, por exemplo, em latitudes temperadas o vento tende a ser mais forte durante o Inverno que no Verão. Numa pequena escala existe uma grande variação largamente ditada pelo impacto da orografia, proximidade dos oceanos, do tipo de vegetação, da presença de montanhas, etc.. O tipo de vegetação tem influência

significativa devido ao efeito de absorção ou reflexão da irradiação solar, afectando a temperatura da superfície e humidade. A topografia mais localmente tem um papel importante na obtenção de velocidades elevadas, por exemplo, o vento é mais forte no topo de montes e montanhas do que em vales ou planícies. Em termos ainda mais locais, o vento é influenciado por árvores, edifícios e outros detalhes do local em análise.

Numa dada localização, a variabilidade temporal numa larga escala significa que a quantidade de vento pode variar de um ano para o outro. Estas variações de longo termo não estão bem estudadas e pode por isso ser complicado efectuar estudos exactos sobre a viabilidade económica de um parque eólico. Esta variação poderá estar relacionada com variações de temperatura a longo termo – o aquecimento global – causada pela actividade humana. Para além desta variação a longo termo poderão existir variações de um ano para o outro numa dada região devido a fenómenos tais como: o *el niño*; variações na quantidade de partículas na atmosfera (vulcões); actividade solar; etc.

Numa escala de tempo mais curta que um ano as variações sazonais são muito mais previsíveis, apesar de existirem grandes variações em escalas mais curtas, que, apesar de bem estudadas, não são muito previsíveis para horizontes superiores a alguns dias. Estas são as chamadas variações *synoptic* e estão relacionadas com o movimento de grandes massas de ar do tipo depressões ou anti-ciclones. Dependendo da localização poderão existir também variações com a altura do dia (variações diurnas) que também são previsíveis.

Em variações de escala ainda mais curtas, na ordem dos minutos ou menos, são conhecidas como turbulência. A turbulência é caracterizada por um movimento de curta duração, inferior a 10 minutos. A turbulência tem origem em fenómenos mecânicos — gradientes de velocidade — e em fenómenos térmicos — estratificação instável da atmosfera. A turbulência tem grande influência no projecto e performance individual dos aerogeradores e também na qualidade da potência entregue à rede.

Van der Hoven construiu um espectro da velocidade do vento para registos longos e curtos. A figura 2.2 mostra claramente picos correspondentes à variabilidade

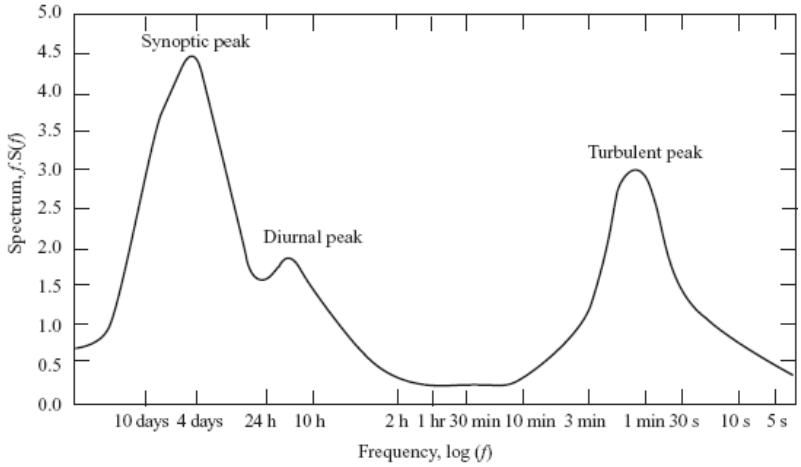

Figura 2.2: *Espectro da velocidade do vento de Brookhaven, baseado no estudo de Van der Hoven.*

synoptic, diurna e efeitos de turbulência. Particularmente interessante é o *gap* espectral entre o *diurnal peak* e o *turbulent peak*, há uma energia muito pequena no espectro no intervalo entre 2h e 10 minutos, mostrando que as variações *synoptic peak* e *diurnal peak* podem ser tratadas de forma independente das variações de alta frequência da turbulência. Existe pouca energia no espectro na região entre 2 horas e 10 minutos. Segundo (Simões, 2004), o cálculo da média da velocidade entre medições deve ser realizado na zona do espectro entre os 10 minutos e as 8 horas, que é uma zona de vazio espectral, onde não existem grandes variações. O período de tempo utilizado para o cálculo do valor médio da velocidade horizontal é, normalmente, de uma hora ou dez minutos por forma a minimizar o desvio padrão do vento com o tempo.

2.2.1 Variação Anual e Sazonal

Enquanto que a variação média anual de ano para ano continua muito difícil de prever, a variação da velocidade do vento pode ser bem caracterizada na forma de uma distribuição de probabilidade, em períodos superiores a uma semana. A variação da velocidade do vento descreve-se melhor com uma função densidade de probabilidade de Weibull – $h(v)$ – com dois parâmetros, o parâmetro de forma k que descreve a variabilidade em torno da média, e o parâmetro de escala c que está está relacio-

nado com a velocidade média anual (Justus et al., 1978). A expressão matemática da função densidade de probabilidade é:

$$h(v) = \left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{(k-1)} e^{-\left(\frac{v}{c}\right)^k} \qquad \text{para } 0 < v < \infty \tag{2.2}$$

Os parâmetros c e k estão relacionados com a velocidade média $\overline{v}$ anual e variância.

A distribuição de probabilidade do vento tem grande aplicação em problemas de estudo e avaliação do potencial de locais para construção de um parque eólico. Permite aos engenheiros optimizar a potência das turbinas, minimizar custos de produção e avaliação do retorno financeiro dos investimentos. A descrição das condições do vento num grande número de locais da Europa estão reunidos no Atlas Europeu do Vento (Troen & Petersen, 1989), no qual as condições do vento são resumidas com os dois parâmetros dados como função dos vários sectores da direcção a uma altura aproximadamente de 10 metros acima do nível do solo.

2.3 Parâmetros que Condicionam o Potencial Eólico

Existem diversos parâmetros e factores que condicionam o escoamento atmosférico de um dado local.

A medição da velocidade e direcção do vento num determinado local são essenciais para a caracterização do escoamento atmosférico. Em locais onde se pretende instalar um parque eólico são realizadas medições pelo menos durante um ano antes da sua construção, em alguns casos é aconselhado um período mais longo.

As medições normalmente obtidas podem ir desde a altura meteorológica de referência (10 m) a 80 m, sendo que o ideal seria obter as medidas o mais perto possível do rotor da turbina eólica a instalar. Por vezes é necessário transpor verticalmente os valores da velocidade do vento. É importante salientar que a velocidade do vento não aumenta infinitamente com a altura, segundo (Patel, 1999) aumenta com a altura até aos 450 metros e depois diminuiu. A velocidade do vento a 450 metros pode ser quatro ou cinco vezes maior do que na superfície.

A direcção do vento também influência a produção de potência, mas quando

comparada com a direcção a sua influência é muito menor, isto porque as turbinas são dispostas de forma a ficarem de frente para o seu melhor sector de direcção. Segundo (Li et al., 2001), geralmente para a mesma velocidade de vento não existe grande diferença na produção de potência para diferentes direcções do vento. No entanto, para pequenas produções a variação na direcção do vento é muito maior do que nas potências mais elevadas.

Outros parâmetros são a complexidade do terreno em estudo (terreno plano, elevações de declive suave ou terreno montanhoso) e a rugosidade (tipo de ocupação do solo característica de uma dada região ou local), no entanto, estes dois factores não são quantificáveis em variáveis nem modelizáveis do ponto de vista estatístico.

O escoamento atmosférico apresenta uma elevada dependência da complexidade do terreno em estudo (terreno plano, elevações de declive suave ou terreno montanhoso), por isso a orografia representa um papel muito importante na sua caracterização. As montanhas podem alterar as características do escoamento atmosférico de diferentes formas. O aquecimento das encostas das montanhas durante o dia e arrefecimento durante a noite fazem com que o ar adjacente à montanha aqueça ou arrefeça por condução e mistura. A diferença de temperatura que se faz sentir entre o ar perto da montanha e o ambiente circundante vai dar origem a fenómenos de brisa.

2.4 Variabilidade da Produção Eólica em Horizontes Curtos

Muitas vezes a energia eólica é incorrectamente classificada como um fonte de energia intermitente. Ao nível de um sistema de energia a potência eólica não ”aparece” e ”desaparece” em intervalos irregulares, de um momento para o outro não deixa de estar a ser injectada na rede energia eólica, mesmo em situações limite são necessárias horas para que todos os aerogeradores de uma área sejam desligados. Períodos onde não há qualquer produção de energia são facilmente previstos e o decréscimo de potência até zero é gradual. Como já foi explicado, a variabilidade da potência eólica num horizonte curto (de minutos a horas) é importante para a integração

da potência eólica na rede, pois afecta o escalonamento das unidades de produção de energia, gestão das reservas e operação de dispositivos de armazenamento de energia. Esta variabilidade é determinada pela variação na velocidade do vento e a localização geográfica dos parques eólicos.

A análise dos dados disponíveis da operação dos parques eólicos e medidas meteorológicas permitem quantificar as variações da potência injectada na rede que podem ser esperadas para um determinado período de tempo (minutos, horas, várias horas). Da observação de dados de operação de parques eólicos, segundo (Association, 2006) foi observado o seguinte:

- Em grandes turbinas individuais a variação da potência produzida numa escala de segundos é muito pequena, em grande parte devido à atenuação do campo do vento à altura do rotor e principalmente ao efeito de filtro passa-baixo[1] da inércia da turbina.

- Num parque eólico individual a variação da potência produzida agregada de todas as máquinas é pequena em escalas de 10 segundos, devido à agregação da potência produzida por cada turbina individual.

- Para um número de parque eólicos espalhados por uma grande área, por exemplo todo o sistema eléctrico, a variação da potência total produzida por todos os parques é pequena para escalas de tempo de inferiores ou iguais a 10 minutos.

As variações relativas a várias horas afectam o planeamento de operação do sistema eléctrico, no entanto, estas variações podem ser previstas usando ferramentas de previsão de potência eólica. Em (Association, 2006) encontra-se um estudo destas variações como uma percentagem relativa à potência instalada, por exemplo, num parque eólico da Dinamarca a maior variação verificada foi 50 % da potência instalada; enquanto que num parque da Alemanha verifica-se 20 %.
Em (Wan & Bucaneg, 2002) são analisadas as variações de dois parques localizados nos Estados Unidos, tendo-se verificado que 94.5% de variações minuto a minuto

[1] filtro que apenas deixa passar as baixas frequências

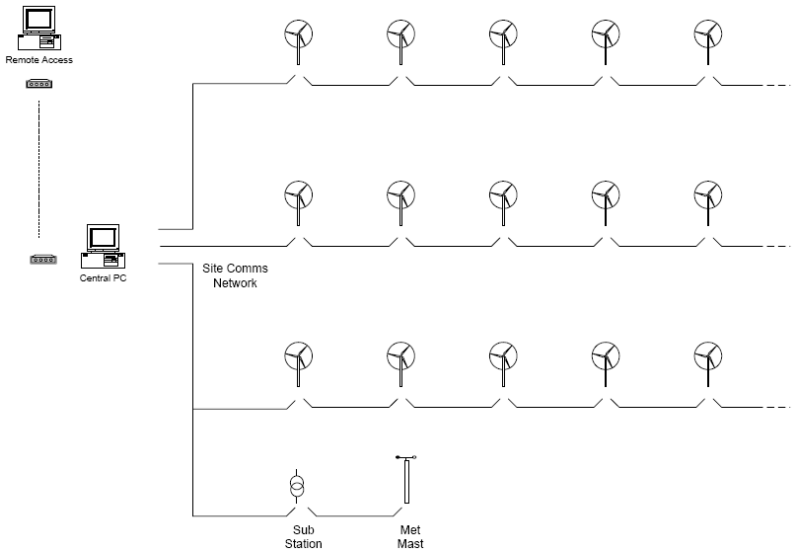

Figura 2.3: *Sistema SCADA de comunicação de um parque eólico*

estão contidas em ±2.8% da potência instalada dos parques, as variações segundo a segundo são ainda mais pequenas (98% das variações entre ±0.93%). A produção eólica demonstra distintos padrões de produção para cada estação do ano. Igualmente demonstrado foi que a variabilidade da produção diminuía com o aumento do número de turbinas no parque eólico e da distância entre elas.

2.5 Fluxo de Dados de um Parque Eólico

2.5.1 SCADA

Cada aerogerador possuiu o seu próprio controlador RIU (*Remote Interface Unit*) na base da torre e opera de forma independente. O controlo geral do parque eólico é realizado através de um sistema SCADA (*Supervisory Control and Data Acquisition*) (figura 2.3). Cada aerogerador está ligado ao SCADA através de linhas de comunicação. O sistema SCADA permite o controlo remoto e monitorização individual de cada aerogerador e do parque eólico, quer a partir do computador central ou de um computador pessoal remoto.

O RIU tem a função de fazer o interface com cada aerogerador, permitindo comunicar com diferentes tipos de aerogeradores. A tecnologia moderna de comunicação

deste tipo de sistemas resume-se ao seguinte: Ethernet; (TCP/IP) Wireless connectivity IEEE 802.11.

O sistema SCADA permite registar medições de diversas variáveis existentes num parque eólico, sendo que o número de variáveis varia de sistema para sistema. Algumas das variáveis que podem ser registados por estes tipos de sistemas são: produção de potência activa e reactiva à saída do parque e de cada aerogerador; velocidade do vento em cada aerogerador; direcção da nacelle (cabine); temperatura; tensão nas fases; controlo da turbina.

Devido a questões técnicas de protocolo de comunicação por vezes existe um problema ao nível do histórico de informação. É então casual existirem falhas no histórico de informação relativo aos aerogeradores e estação de medida. O mesmo não acontece com dispositivos que estejam ligados directamente ao *master*, como é o caso dos contadores de energia da subestação do parque. Será portanto de esperar algumas falhas no histórico das leituras por aerogerador e poucas falhas no histórico das variáveis do parque eólico.

Um anemômetro é um instrumento de medição da velocidade do vento que se encontra instalado no cimo da *nacelle*. As suas medições são perturbadas pelo movimento das pás do rotor (efeito de esteira) e constantemente sujeitos às condições atmosféricas. Sendo que no inverno apresentam condições críticas de funcionamento, podendo em muitos casos congelar. Portanto, as medições do anemômetro não apresentam a qualidade necessária para serem utilizadas em qualquer estudo.

Relativamente à direcção do vento, o SCADA regista a direcção da *nacelle*. Direcção essa que nem sempre coincide com a direcção do vento, devido ao sistema de controlo do movimento da *nacelle*, que depende do sistema de controlo de forma a diminuir o número de rotações. A alteração do movimento só se verifica após o valor da direcção de vento estar consolidado, e convém também realçar que por vezes a *nacelle* procede ao movimento de destorcimento, que não é mais do que um movimento de rotação sobre a torre de forma a garantir que não existe uma torção dos cabos no interior da torre.

Os dados registados pelo SCADA apresentam alguns erros, tal acontece em situações em que as máquinas por razões de avaria ou manutenção, estão indisponí-

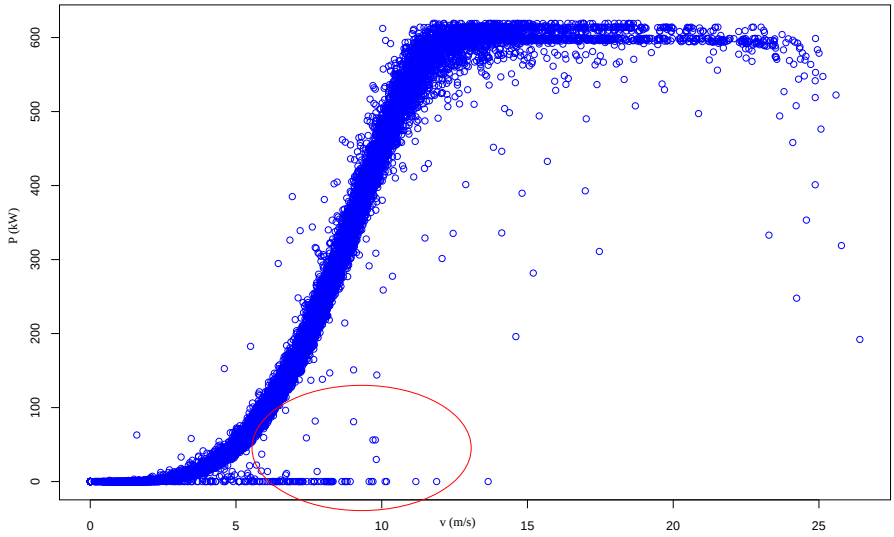

Figura 2.4: *Produção em função velocidade do vento registada pelo SCADA de um aerogerador.*

veis. Por vezes, a velocidade do vento é superior à *cut-in speed*, mas a produção dos aerogeradores não corresponde à produção que seria esperada. Na figura 2.4 são visíveis as situação de indisponibilidade das máquinas, onde por vezes a sua produção é nula ou muito afastada do que deveria produzir segundo a curva de potência. Também é visível nas situações em que a velocidade do vento atinge o ponto *cut-out speed* o decréscimo da potência produzida até zero é gradual.

No caso da análise da potência total produzida pelo parque podem ocorrer situações em que apesar de a velocidade medida apresentar uma valor elevado, a produção total encontra-se muito afastada da potência nominal, isto porque poderão existir máquinas indisponíveis ou situações de velocidades extremas em que por razões de segurança é necessários desligar as máquinas.

2.5.2 Estação Meteorológica

Com o objectivo de avaliar o potencial energético de um local, determinar a melhor colocação dos aerogeradores e escolha da potência das turbinas a instalar no local são colocadas estações de medição das condições meteorológicas, também conhecidas como mastros anemométricos. Em locais onde a orografia é muito complexa, a avaliação do potencial eólico é, normalmente, válida em pequenas áreas em redor

dos mastros instalados para medição do vento (tipicamente 5x5km). Esta limitação leva a que os potenciais promotores no início dos seu projectos necessitem da realização de campanhas experimentais de medição do vento.

Nos últimos anos, com o aparecimento dos grandes parques, quer em área quer em número de aerogeradores, foi necessária a instalação de mais do que um mastro anemométrico dentro da área que engloba o parque eólico. Na fase de exploração do parque eólico pelo menos uma das estações de medição continua em funcionamento. As estações na fase de exploração têm como objectivo permitir a avaliação da performance do parque, comparando as curvas de potência inicialmente estimadas para o parque com as curvas actuais.

A recolha de dados na fase de avaliação deve ter a duração de pelo menos 1 ano, sendo que a altura de recolha dos dados deve ser o mais próxima possível da altura a que se vai instalar a *nacelle*. De acordo com a norma IEC 61400-12 devem ser usadas para estudo velocidades médias em períodos de 10 minutos, que, como já foi explicado, permitem conduzir a uma estimativa razoável da energia produzida anualmente. Em muitos parques, os dados das estações não se encontram acessíveis remotamente através do SCADA, nestes casos a sua utilidade reside na avaliação das previsões meteorológicas.

As medidas do vento realizadas pelo mastro são as seguintes: velocidade, direcção do Vento, pressão, temperatura e densidade do ar.

2.5.3 Modelo de Previsão Meteorológica Numérica

Um modelo de previsão numérica das condições meteorológicas (NWP - *Numerical Weather Prediction*) é um modelo computacional desenvolvido para simular o comportamento da atmosfera. Este resolve um sistema complexo de equações matemáticas baseadas em leis físicas (lei da conservação de energia; massa e momento; equação de estado dos gases) de modo a prever o estado futuro da atmosfera, partindo de condições iniciais específicas. Estas equações matemáticas têm a forma de equações diferenciais que relacionam as derivadas totais de certas variáveis físicas com a variável tempo. Processos físicos, como a radiação, nuvens, precipitação e trocas à superfície, são representadas no modelo usando esquemas de parametrização.

Estas equações diferenciais que correspondem à modelização matemática das leis físicas, que regem o comportamento dinâmico da atmosfera, recebem a denominação de equações primitivas.

O objectivo da previsão numérica é determinar o estado da atmosfera num horizonte futuro a partir da condição inicial actual. É necessário um modelo matemático que resolva as equações primitivas, sendo que as equações primitivas não são lineares. Para a resolução das equações diferenciais recorre-se a técnicas numéricas. As técnicas mais utilizadas são o método das diferenças finitas, elementos finitos e esquemas de interpolação.

Com as soluções numéricas obtém-se o estado futuro da atmosfera, o que não é mais do que a previsão do tempo. Como as equações são não lineares, pequenos erros nas condições iniciais amplificam erros à saída com o avanço da simulação no tempo, de modo que a diferença entre as condições atmosféricas previstas e as condições reais da atmosfera aumenta ao longo da previsão. Devido a estes erros, as previsões realistas usando modelos de previsão numérica só podem ser efectuadas para um período máximo de 5-10 dias. Este comportamento da atmosfera é um exemplo de um sistema caótico não linear.

Para a obtenção das soluções numéricas os modelos discretizam o espaço e o tempo. O planeta é dividido em várias camadas verticais que representam níveis da atmosfera, dividindo a superfície do planeta numa malha de caixas horizontais separadas por linhas semelhantes às linhas de longitude e latitude. O planeta é coberto por uma malha com quatro dimensões (três dimensões espaciais e tempo), é sobre esta malha que as equações diferenciais são integradas no tempo para prever o estado da atmosfera. A resolução espacial da malha influencia directamente a precisão do modelo atmosférico; uma diminuição do tamanho da malha significa um aumento da precisão, mas também um aumento do tempo computacional para a resolução das equações. Uma área de investigação neste tipo de modelos centra-se em reduzir o tamanho da malha sem aumentar o tempo de computação.

O significado da coordenada vertical (altura) na malha depende do modelo utilizado. Existem modelos que trabalham com coordenadas verticais z definidas sobre um nível de referência (por exemplo, nível médio do mar num dado ponto). Outros

modelos utilizam uma coordenada conhecida como σ , que é uma coordenada vertical que ”segue” o terreno; um modelo que use esta coordenada necessita de uma caracterização precisa da orografia do terreno.

O processo de previsão de um modelo NWP consiste em três fases: pré-processamento, processamento e pós-processamento (Kalnay, 2003). Na fase de pré-processamento realizam-se as tarefas da definição do domínio em análise (zona de estudo), resolução da malha, características do solo, e assimilação dos dados meteorológicos disponíveis. Esta assimilação contempla o estabelecimento do estado inicial da atmosfera e consiste numa interpolação sobre os pontos da malha das variáveis meteorológicas disponíveis. O resultado do pré-processamento é uma malha com os valores da pressão, velocidade, direcção do vento, temperatura e humidade, para cada um dos pontos da malha.

A fase de processo consiste na resolução das equações diferenciais para cada um dos elementos da malha. A resolução das equações é de forma iterativa com um intervalo de tempo definido da ordem dos segundos.

Na fase de pós-processamento são determinadas a partir de variáveis básicas (pressão, temperatura, velocidade, ...) outras variáveis de interesse meteorológico.

Modelos Atmosféricos

Nem todos os modelos atmosféricos trabalham com a mesma escala espaço-temporal. No caso da previsão eólica é crucial a previsão da velocidade do vento; neste tipo de problemas a escala do modelo deve ser pequena no espaço devido aos efeitos que a orografia tem sobre o vento e pequena no tempo devido à variabilidade do vento.

Existem três classes de modelos segundo a escala: macro-escala, meso-escala e micro-escala. Em (Orlanski, 1975) foram propostas 8 subdivisões para as três escalas. Os modelos atmosféricos com grande resolução espacial (escala espacial pequena) têm um alcance temporal baixo. Na tabela 2.1 encontram-se as divisões bem como a sua escala temporal e espacial. Os modelos de previsão numérica do tempo podem ser classificados segundo a sua escala espacial dentro da qual a previsão é feita:

- **Global:** São modelos macro-escala que simulam o comportamento da atmos-

Tabela 2.1: *Escalas espaço temporais para modelos atmosféricos.*

Nome	Escala Espacial	Escala Temporal
$Macro-\alpha$	10000 km	1 mês
$Macro-\beta$	2000 km	1 semana
$Macro-\gamma$	200 km	1 dia
$Meso-\beta$	20 km	6 horas
$Meso-\gamma$	2 km	1 hora
$Meso-\delta$	200 m	30 minutos
$Micro-\beta$	20 m	1 minuto
$Micro-\gamma$	2 m	1 segundo

fera em todo o planeta, e portanto realizam a integração das equações primitivas sobre uma malha 4-D que cobre toda a Terra. O tamanho típico de uma malha de um modelo NWP global é aproximadamente de 150km na direcção este-oeste e 100km na direcção norte-sul, com 20 a 40 níveis verticais (Figura 2.5). Isto significa que a atmosfera é representada por cerca de um milhão ou mais de cubos individuais. Em cada um destes cubos, as componentes do vento, temperatura, pressão e humidade são calculadas.

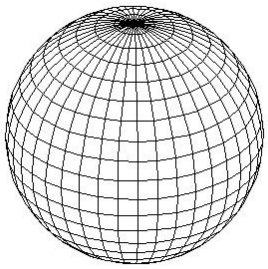

Figura 2.5: *Malha de um modelo global.*

- **Regional ou local:** com resolução espacial de alguns metros a cerca de 50km, visa identificar com maior detalhe o comportamento da atmosfera sobre uma região específica, sendo capazes de identificar fenómenos meteorológicos de

pequena escala. Têm a desvantagem relativamente ao modelos globais de para além de precisarem de condições iniciais, também necessitarem de umas condições de contorno nas fronteiras do domínio. Estas condições de controlo são conseguidas interpolando os valores de saída de um modelo atmosférico global.

Como é possível verificar pela tabela 2.1, tanto as micro-escalas como as macro-escalas não permitem estudar os fenómenos relacionados com a previsão de produção eólica. Pela resolução da escala parcial os únicos modelos que permitem um estudo da velocidade do vento são os modelos meso-escala, no entanto, em termos temporais são muito limitados. Uma forma de aumentar a escala temporal de um modelo meso-escala é as suas condições iniciais serem estabelecidas a partir de um modelo de maior escala. A qualidade das previsões dos modelos meso-escala estão fortemente dependentes da qualidade das condições iniciais e de fronteira fornecidas.

Uma característica deste tipo de modelos é a possibilidade de aninhamento. O aninhamento consiste na definição de domínios menores, com maior resolução espacial, dentro de outros domínios. Por exemplo, um domínio de 50x50 km pode corresponder a um domínio aninhado a outro maior 150x150 km, que por sua vez está aninhado a outro maior 450x450 km. O domínio maior toma as condições de contorno de um modelo global.

MM5

O modelo de meso-escala da PSU/NCAR[2] é um modelo de área limitada (áreas limitadas a centenas de quilómetros) com resoluções médias de 20 – 30 km, e com um sistema de coordenadas σ. O modelo foi desenvolvido para simular e prever circulações de meso-escala ou escala regional. O MM5 é composto por um conjunto de programas independentes de pré e pós-processamento. O conjunto de informação processada por cada programa constitui a base de dados para as simulações do programa principal, MM5, que fornece os resultados climáticos. O sistema de

[2] Pennsylvania State University/National Central for Atmospheric Research

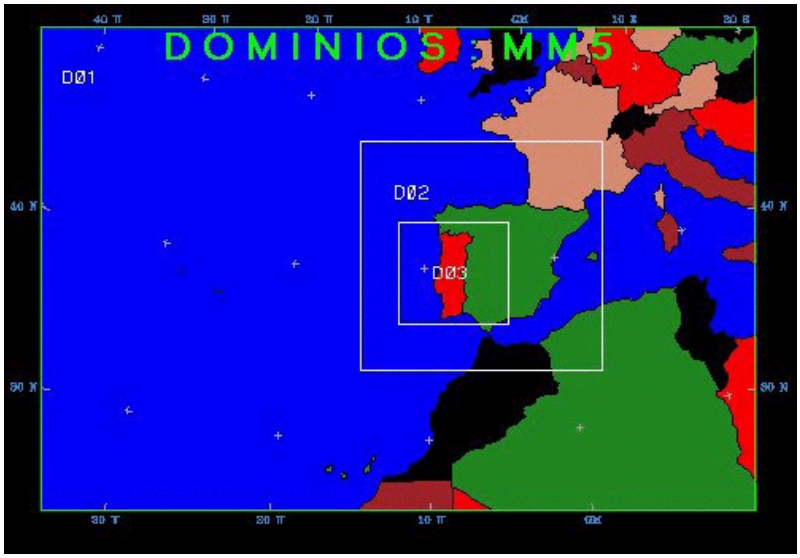

Figura 2.6: *Aninhamento de domínios na zona de Portugal.*

modelação MM5 é distribuído livremente e principalmente escrito em Fortran, e foi desenvolvido na Penn State e na NCAR como um modelo de meso-escala comunitário. O MM5 tem a capacidade de múltiplo aninhamento até um limite de nove domínios a correr ao mesmo tempo e interagindo entre eles. Um exemplo do aninhamento de domínios para Portugal encontra-se na figura 2.6, onde são visíveis três domínios.

O MM5 guarda os resultados para todos os pontos da malha de cada um dos domínios e para cada um dos intervalos temporais até alcançar o horizonte final de previsão. O programa MM5 oferece a possibilidade de gerar uma série temporal das variáveis meteorológicas para cada um dos passos de integração até alcançar o horizonte final, a altura é 10 metros.

MM5 - Previsão das Condições Meteorológicas de um Parque Eólico

Uma metodologia de previsão numérica das variáveis relacionados com o vento, por razões já explicadas nos capítulos anteriores, deve ser capaz de integrar três fases de modelos de previsão: um modelo global disponível no *National Weather Services*; um modelo meso-escala cuja operação possa ser realizadas de diversas formas; um modelo de previsão à escala de um parque eólico. As previsões da velocidade de

vento são afectadas pela resolução espacial, especialmente nos terrenos complexos como é caso do Norte de Portugal. Os erros típicos dos modelos de meso-escala baseiam-se em erros de amplitude e erros de fase. Os erros de amplitude podem ser reduzidos recorrendo a uma maior resolução da grelha, enquanto que os erros de fase não são resolvidos com esta técnica, mas sim recorrendo a mais do que um modelo global, procedendo à actualização das condições de fronteira e aumentando o domínio de actuação.

O modelo MM5 utilizado nas previsões desta tese apresenta uma configuração com 3 malhas aninhadas com resoluções de 54x54 km, 18x18 km, 6x6 km. A fronteira do domínio exterior foi forçada por análises, de 6 em 6 horas, do ECMWF (*European Centre for Medium-Range Weather Forecasts*). Partindo destes dados um mapa tri-dimensional com o estado da atmosfera é construído cada 24 horas. São obtidas previsões meteorológicas diariamente com um horizonte temporal de 72 horas, devido ao processamento destas previsões ser demoroso e depender do resultado de centros de meteorologia internacionais, esta informação só se encontra disponível depois das 6 horas da manhã de cada dia. Isto implica que durante este período seja necessário recorrer às previsões do dia anterior. O período de amostragem das previsões é 30 minutos.

As previsões podem ser obtidas para um ponto de referência dentro do parque eólico ou por ponto de localização de cada aerogerador, com base nas suas coordenadas geográficas UTM[3]. As previsões por aerogerador são construídas com base numa interpolação bilinear dos quatros pontos do quadrado da grelha do modelo MM5 mais próximos do ponto que se pretende obter a previsão. Sendo assim, no caso do ponto de referência dentro do parque que foi definido como um ponto central, a interpolação é efectuada com base nos quatro vértices do quadrado que engloba o ponto de previsão. Neste caso, a resolução do modelo meso-escala não necessita de ser tão elevada como no caso da previsão meteorológica por aerogerador.

As grandezas previstas pelo modelo são: velocidade e direcção do vento, pressão atmosférica, temperatura e densidade do ar.

[3] Universal Transverse Mercator

Capítulo 3

Estado da Arte

Vários centros de investigação e empresas investiram na investigação e desenvolvimento de métodos/modelos de previsão, o que levou a um grande número de sistemas de previsão, alguns dos quais já se encontram em fase de operacionalização e comercialização. Os sistemas de previsão são essencialmente caracterizados pelo horizonte de previsão (minutos, horas, dias), complexidade computacional (um computador, *cluster* de computadores) e valor do erro de previsão.

Este capítulo não tem como objectivo apresentar um estudo detalhado de todas as metodologias de previsão existentes na literatura; uma visão geral das metodologias desenvolvidas para este problema pode ser encontrada em (Yuan-Kang & Jing-Shan, 2007)(Giebel, 2002) e (Costa et al., 2008). Um objectivo é demonstrar que grande parte das metodologias de previsão usam como critério de treino a minimização do erro médio quadrático, sendo que a mais recente investigação na área passa por desenvolver novas funções de custo. Serão apresentados trabalhos onde a distribuição dos erros de previsão é estudada.

Um outro objectivo consiste em apresentar trabalhos que demonstram o interesse crescente no desenvolvimento de modelos com capacidade de se adaptarem a alterações na operação do parque eólico. Actualmente já existe alguma investigação onde a estimação dos parâmetros dos modelos é realizada de forma recursiva.

As conclusões deste capítulo irão apontar no sentido de desenvolver novos critérios de treino de modelos de previsão e a necessidade de estes serem capazes de se adaptar a alterações no regime de exploração do parque eólico.

3.1 Formulação da Previsão Eólica

Previsão é a arte de prever o que vai acontecer no futuro partindo do conhecimento que se tem sobre situações actuais e passadas. É importante referir que se trata de uma forma de extrapolação: um modelo de previsão é construído com base num conjunto de dados e usado fora desse conjunto de dados. De certa forma, as previsões são condicionais e devem ser formuladas do seguinte modo: "dado um conjunto de dados e assumindo que o comportamento modelado continua no futuro, pode-se prever que ...". Isto significa que se assume que o futuro é igual ao passado, esta é a primeira fonte de incerteza nas previsões. Logo, alimentar os modelos de previsão com a mais recente informação tem importância fundamental.

A previsão concentra-se na evolução no tempo da variável a prever p. A variável a prever apresenta um intervalo de amostragem fixo e a evolução de p é representada por uma série temporal discreta $(p_t, t \in \mathbb{T})$. Os valores de $\{p_t\}_{t\in\mathbb{T}}$ podem ser instantâneos ou médias dos valores no intervalo de amostragem considerado.

A variável p a prever no problema de previsão eólica é a potência produzida por uma turbina eólica, parque eólico ou uma área de rede com vários parques eólicos. Os períodos de amostragem da potência produzida são normalmente de 30 minutos ou 1 hora. Estes são os intervalos mais utilizados nas previsões por razões de ordem técnica quer dos mercados (neste caso é 1 hora), quer dos operadores do sistema eléctrico. A previsão do valor de p realizada no instante t para o instante $t+k$ tem a designação de $\hat{p}_{t+k/t}$, não é mais do que o valor previsto da potência para o instante $t+k$ dado que estamos no instante t,onde k tem várias designações tais como horizonte de previsão ou instante de previsão.

Em estatística existe a distinção entre dois tipos de modelos: univariados e multivariados. Um modelo univariado apenas considera os valores passados da variável p, este modelo indica que p_t pode ser expressa como a soma da função g_u composta pelos l valores passados de p e uma perturbação aleatória e_t:

$$p_t = g_u(p_{t-1}, p_{t-2}, \ldots, p_{t-l}) + e_t \tag{3.1}$$

onde e_t erro aleatório de observação (ou ruído branco), uma distribuição normal com média zero e variância comum σ_e^2.

Alternativamente pode ser considerada a formulação de um modelo multivariado, onde, para além dos valores passados da variável que se pretende prever, são também utilizadas variáveis explicativas, mais propriamente previsões das variáveis explicativas. No problema de previsão da produção eólica as variáveis explicativas são normalmente: velocidade do vento, direcção, temperatura, pressão e densidade do ar. Num modelo multivariado p_t é expressa em função dos valores passados de p, valores passados das variáveis explicativas x, previsões das variáveis explicativas e também uma perturbação aleatória:

$$p_t = g_m(p_{t-1}, p_{t-2}, \ldots, p_{t-l}, x_{t-1}, x_{t-2}, \ldots, x_{t-m}, \hat{x}_{t/t-1}) + e_t \tag{3.2}$$

Ambos os modelos apresentados podem ser utilizados para descrever a relação entre as variáveis consideradas e p_{t+k}.

O problema de previsão de produção eólica pode então ser formulado da seguinte forma:

$$p_{t+k/t} = f(p_t, p_{t-1}, \ldots, p_{t-l}, x_t, x_{t-1}, \ldots, x_{t-m}, \hat{x}_{t+k/t}) + e_t \tag{3.3}$$

O trabalho da previsão passa por determinar uma função f que melhor descreve o processo considerado. Em termos físicos, o conceito que um sistema de aprendizagem pretende adquirir é a curva de potência do aerogerador/parque, ou seja, a função f não é mais do que a equação que caracteriza a curva de potência do parque.

3.2 Métodos de Previsão de Referência

Um sistema de previsão apenas tem interesse se conseguir ser melhor do que os métodos de previsão de referência. Estes métodos não são resultados da criação de modelos, mas apenas de simples considerações.

O método mais conhecido é a persistência. É um método de previsão *naive* onde o valor previsto de produção é o mesmo do último valor medido:

$$\hat{p}^{p}_{t+k/t} = p_t \tag{3.4}$$

Apesar da sua simplicidade, este método pode ser difícil de superar até às primeiras 6 horas devido à escala de mudanças na atmosfera. O método da persistência

também possuiu uma justificação estatística para o seu excelente desempenho nas primeiras horas, isto porque não é mais do que uma versão simplificada de um modelo autorregressivo. A simples interpretação de um diagrama de autocorrelação e autocorrelação parcial da série de produção do parque permite determinar que estamos na presença de um modelo autoregressivo de segunda ordem (Rodrigues et al., 2007).

Uma generalização do método da persistência é substituir o ultimo valor medido pela média dos últimos n valores medidos.

$$\hat{p}^{MA,n}_{t+k/t} = \frac{1}{n}\sum_{i=0}^{n-1} p_{t-i} \tag{3.5}$$

Este método é muitas vezes conhecido como média móvel. Quando n tende para infinito, este método tende para a média global:

$$\hat{p}^{0}_{t+k/t} = \overline{p_t} \tag{3.6}$$

onde $\overline{p_t}$ é a média de todos os valores medidos da potência produzida até ao instante t.

3.3 Horizonte de Previsão

Uma previsão de produção eólica é para um horizonte temporal com um determinado alcance, sendo que este alcance poderá ter vários limites dependendo do fim a que se destina a previsão. No caso da previsão de produção eólica não existe uma unanimidade nas fronteiras que separam uma classificação da outra. Muitos autores não classificam a sua previsão quanto ao horizonte, e os que o fazem apresentam classificações diferentes. Da consulta da literatura existente foram extraídas as seguintes conclusões sobre a classificação do horizonte temporal:

- **Previsões de muito curto prazo:** o horizonte de previsão é apenas de algumas horas, não existindo unanimidade para esse número de horas. Segundo (Mohrlen, 2004), esse horizonte tem um limite de 9 horas, segundo (Giebel, 2002) pode variar entre 3 e 10 horas;

- **Previsões de curto prazo:** estas previsões vão desde o limite da previsão de ”muito curto prazo” até 48 ou 72 horas. Na maioria dos trabalhos este horizonte encontra-se limitado até 48 horas, no entanto a tendência actual é efectuar previsões até 72 horas (Yuan-Kang & Jing-Shan, 2007)(Giebel et al., 2007). Esta previsão é essencial para a participação nos mercados eléctricos diários de energia. No caso do mercado Espanhol, o prazo de apresentação das ofertas de venda no mercado diário termina às 10 horas de cada dia, pelo que um horizonte de previsão de 38 horas consegue cobrir todo o dia seguinte. No entanto, em outros países a hora de encerramento do mercado pode ser diferente, pelo que o horizonte necessário para cobrir o dia seguinte pode variar;

- **Previsões de médio prazo:** correspondem a horizontes acima das 72 horas e até 5-7 dias. Quanto maior o horizonte, maior os erros de previsão (Giebel et al., 2007). Actualmente é muito difícil realizar previsões atmosféricas para horizontes acima dos 7 dias.

3.4 Distribuição dos Erros de Previsão

A forma das distribuições dos erros de previsão eólica tem sido alvo de estudo, principalmente em trabalhos em que o objectivo é estimar a incerteza da previsão eólica. Grande parte dos actuais sistemas e algoritmos de previsão eólica têm como entrada previsões de variáveis meteorológicas provenientes de um outro modelo de previsão numérica. Depois é uma tarefa do modelo estatístico converter estas previsões meteorológicas em produção de energia eléctrica.

No entanto, o sistema de conversão de velocidade do vento em energia eléctrica tem uma forma não linear. Para algumas gamas de valores de velocidade do vento pequenos erros na previsão da velocidade reflectem-se em grandes erros na produção. Em contraste na zona linear (”achatada”) da curva pequenos desvios de velocidade reflectem-se em insignificantes desvios na produção. No fundo o que a curva de potência faz é amplificar ou atenuar erros na previsão da velocidade do vento em função do valor da sua derivada local.

Lange em (Lange, 2004) da análise dos erros de previsão de produção e velo-

cidade do vento em seis parques eólicos na Alemanha conclui que o desvio padrão (relativo à potência instalada) do erro da previsão da produção é cerca de 1.8 a 2.6 vezes maior do que o erro de previsão da velocidade do vento. Uma inspecção visual do histograma dos erros de previsão dos seis parques permite também verificar que os erros do modelo de previsão da velocidade do vento apresentam características de uma distribuição Gaussiana. A distribuição do erro de previsão da velocidade do vento apresenta uma forma quase simétrica centrada em zero, em outros estudos também este resultado é verificado (Giebel, 2000) (Landberg, 1994).
As distribuições do erro de previsão da velocidade do vento foram testadas para os diferentes locais com o teste estatístico do qui-quadrado tendo-se verificado em quase todos que a distribuição pode ser aproximada por uma distribuição Gaussiana.

Em contraste, a distribuição dos erros de previsão da produção eólica já não apresenta as mesmas características, estando bastante longe de ter uma forma Gaussiana. Em (Lange, 2005) o teste paramétrico do qui-quadrado e o não-paramétrico de *Lilliefors* foram aplicados às distribuições do erro de 6 parques eólicos situados na Alemanha, não tendo sido válida para qualquer parque a hipótese de distribuição Gaussiana do erro.

A análise da distribuição do erro de previsão é também alvo de estudo em (Bludszuweit et al., 2008). O autor utilizou dados medidos de dois parques eólicos para simular três cenários de previsão baseados na persistência. Verificou-se que a distribuição dos erros de previsão apresenta um achatamento variável, possuindo valores para o achatamento desde 3 até 10. A distribuição dos erros de previsão pode então ser caracterizada como uma distribuição com ”caudas pesadas”.

Igualmente nos diversos sistemas de previsão, distribuições com uma forma não-Gaussianas dos erros são encontradas (Sideratos & Hatziargyriou, 2007) (Pinson, 2006) (Tsikalakis et al., 2006) (Lange, 2003). As distribuições são geralmente assimétricas, com um elevado achatamento em ambas as ”caudas” e um elevado ”pico” perto do erro nulo. Da análise dos resultados dos diversos sistemas de previsão verifica-se que a característica não linear da curva de potência altera as propriedades estatísticas da distribuição do erro.

3.4.1 Sub-amostragem

Um aspecto que tem sido ignorado em todos os estudos sobre a previsão eólica é o problema de se estar na presença de ”sinais” sub-amostrados (Bessa et al., 2008). O teorema da amostragem do processamento de sinal diz que:

> ... a quantidade de amostras por unidade de tempo de um sinal, chamada taxa ou frequência de amostragem, deve ser maior que o dobro da maior frequência contida no sinal a ser amostrado, para que possa ser reproduzido integralmente sem erro de *aliasing*...

As medida da velocidade do vento são usualmente realizadas com médias de 10 minutos, no entanto, o ”sinal” da velocidade do vento incluiu componentes com uma frequência mais elevada. No espaço de 10 minutos existem grandes variações do vento, sendo que este número de variações ainda é maior em intervalos de 30 minutos. No entanto, este problema não é totalmente tratável de forma estatística, teria que ser realizada uma ”fusão” de conceitos estatísticos e físicos para abordar o problema (tratado na física como turbulência).

Relativamente às medidas de produção elas poderão ser medidas instantâneas de produção a cada 10 minutos ou valores médios (energia gerada dividida pela duração do intervalo de tempo). A combinação de um problema de sub-amostragem com a não linearidade da curva de potência leva à introdução de uma incerteza de natureza não-gaussiana à previsão da produção.

O impacto da sub-amostragem na distribuição dos erros foi alvo de estudo em (Ahmadzadeh & Petrou, 2001) num outro tipo de problema. Verificou-se que quanto mais informação os dados contêm nas frequências elevadas (é o caso do vento) maior será o erro e mais não-gaussiana será a distribuição dos erros para um período de amostragem maior.

Um resultado muito interessante e que vai de encontro ao que foi exposto nesta secção é que a previsão da energia produzida[1] ao invés de prever a potência produzida apresenta melhores resultados (Makarov et al., 2002). Como se trata de um

---

[1] Relação entre energia e potência: $E(t) = \int_0^T P(t)\,dt$

problema de sub-amostragem a previsão do valor em MWh em vez de MW permite incorporar mais informação e tornar a curva menos variável e mais fácil de prever.

3.5 Abordagem Estatística

A abordagem estatística consiste em um ou vários modelos que procuram estabelecer a relação entre o histórico de potência produzida e previsões das variáveis explicativas (velocidade, direcção, temperatura, etc.). Os modelos da abordagem estatística são criados através de um ajuste a um conjunto de treino (definindo a função f da equação 3.3) e são usados sobre um conjunto de teste, ou em situações reais de previsão. O ajuste aos dados consiste num problema de optimização onde a função de custo (ou função objectivo) a ser minimizada é em função do erro de previsão.

Os modelos estatísticos modernos são constituídos por uma parte autorregressiva que modela a persistência do comportamento do vento, e uma parte que consiste numa transformação não linear das variáveis meteorológicas explicativas em potência produzida. As variáveis de entrada dos modelos são os valores previstos das condições meteorológicas de um modelo NWP meso-escala e os valores passados de produção.

Sem as previsões do modelo NWP seria impossível efectuar previsões com qualidade para um horizonte superior a 6 horas. Os modelos puramente baseados em séries temporais apresentam um excelente desempenho quando comparados com o modelo de referência até 6-10 horas, enquanto que os modelos baseados em previsões NWP apresentam um melhor desempenho a partir das 6 horas (Kariniotakis, 2006) (Giebel, 2002).

Uma outra abordagem é a chamada abordagem física (Landberg, 1994) (Bailey et al., 1999). A ideia principal da abordagem física é refinar os resultados das previsões NWP para determinar o vento na área do parque e à altura da *nacelle*. São usadas considerações físicas sobre o terreno tais como: rugosidade, orografia e obstáculos. Um método conhecido é o uso de modelos CFD(*Computer Fluid Dynamics*) (Castro, 1997) que permitem captar com grande exactidão o campo do vento

no parque eólico considerando uma descrição completa do terreno. Quando o vento ao nível do parque eólico e à altura da torre é conhecido o segundo passo consiste em converter a velocidade do vento em potência. Usualmente esta tarefa é realizada recorrendo a curvas de potência teóricas (curvas do dadas pelo fabricante), embora existam estudos ((Cabezon et al., 2004)) que revelam o interesse do uso de curvas de potência empíricas em vez das teóricas.

3.5.1 Função de Custo

O critério de treino do erro médio quadrático assume que o erro do sistema adaptativo apresenta uma distribuição Gaussiana, e mesmo que tal não se verifique esta aproximação continua a apresentar bons resultados tal como é demonstrado em (Bishop, 1995). No entanto, e apesar da qualidade dos resultados usando este critério ser bastante aceitável, o desenvolvimento de funções de custo que não assumam a distribuição Gaussiana do erro poderá ser uma alternativa benéfica para o problema. Nas secções seguintes serão analisadas as funções de custo utilizadas no actual estado da arte de previsão eólica.

Abordagens com o Erro Médio Quadrático

Não é objectivo desta secção uma explicação detalhada das técnicas utilizadas nos sistemas de previsão, o principal objectivo é salientar que existem na bibliografia diferentes tipos de modelos estatísticos de previsão. O ponto comum em todos eles é a função de custo (ou função objectivo) utilizada no problema de optimização que permite estimar os parâmetros dos seus modelos. Grande parte dos modelos utiliza a minimização dos quadrados dos desvios entre previsão e valor real para estimar os seus parâmetros.

Os modelos baseados em sistemas de inferência difusa utilizam o erro médio quadrático no seu processo de treino supervisionado. O treino do sistema de inferência difusa é também realizado através de um algoritmo de retropropagação do erro onde o objectivo é determinar o peso associado a cada uma das regras difusas. Por exemplo, em (Damousis et al., 2004) é descrito um modelo difuso que é treinado com recurso a um algoritmo genético cuja função de adaptação é o erro

médio quadrático. Outros exemplos de modelos baseados em lógica difusa poderão ser encontrados em (Sideratos & Hatziargyriou, 2007) (Pinson & Kariniotakis, 2003) (W. Potter & Negnevitsky, 2006). Todos estes sistemas têm como variáveis de entrada previsões das condições meteorológicas.

Os modelos que utilizam redes neuronais têm como função de custo no seu treino o erro médio quadrático. Na previsão eólica são utilizadas redes neuronais de diferentes tipos (MLP, RBF, recursivas, modulares, etc.), no entanto, todas elas têm como elemento comum a sua função de custo. Em (Ramírez-Rosado et al., 2006) é descrito um modelo de previsão da produção baseado em quatro famílias distintas de modelos de redes neuronais. É construído um modelo de cada uma das quatro famílias (MLP, modulares, RBF e com análise de componentes principais) para cada um dos instantes do horizonte de previsão (1 a 48 horas). Os 96 modelos de redes neuronais são optimizados usando um algoritmo genético que permite automaticamente seleccionar as entradas, o número de neurónios e o valor dos parâmetros do algoritmo de treino. Outros trabalhos na área das redes neuronais aplicadas à previsão eólica com a função de custo erro médio quadrático são: (Ernst & Rohrig, 2002) (Barbounis et al., 2006) (Singh et al., 2007) (Watanabe & Wakao, 2007). Todos estes modelos têm como variáveis de entrada previsões das condições meteorológicas.

O uso de metodologias baseadas em séries temporais para previsões de muito curto-prazo é a opção mais comum e que leva a melhores resultados. Também os modelos baseados na teoria das séries temporais (modelos AR, ARIMA, ...) utilizam o erro médio quadrático para estimar os seus parâmetros. Em (Milligan et al., 2004) são descritos vários modelos estatísticos que podem ser úteis para os chamados *hour-ahead markets*. O objectivo do artigo não é desenvolver modelos de previsão que possam competir comercialmente com os já existentes, em vez disso, pretende-se investigar até que horizonte os modelos baseados em séries temporais conseguem ser melhores que a persistência. Os modelos estudados são os ARMA (apresentam uma parte autorregressiva e média móvel), sendo apenas usados os valores passados da potência para prever a potência nos instantes futuros. Os modelos são também aplicados para prever a velocidade do vento a partir dos seus

valores passados. Os autores concluíram que a capacidade de previsão do modelo ARMA dependia do período de tempo ao qual estava a ser aplicado. É sugerida a possibilidade de usar vários modelos para prever para o mesmo horizonte, em vez de apenas um modelo. Também neste caso o erro médio quadrático foi usado como função de custo para estimar os parâmetros do modelo.

Outros trabalhos relacionados com séries temporais e que não usam como entradas previsões de um modelo meso-escala poderão ser encontrados em (Makarov et al., 2002) (Larson, 2004)(El-Fouly et al., 2006).

Abordagens Não-paramétricas

Recentemente têm sido desenvolvidos métodos probabilísticos não-paramétricos onde não é assumida qualquer fdp para o erro do modelo. Em (Nielsen et al., 2006) e (Bremnes, 2004) o objectivo consiste em estimar os quantis da distribuição da potência produzida utilizando para isso a regressão de quantis. A utilidade das previsões probabilísticas consiste no facto de tornar possível produzir informação sobre a incerteza associada às previsões para posteriormente ser utilizada em ferramentas de apoio à decisão.

Foram desenvolvidas as duas técnicas: regressão local de quantis (Bremnes, 2004) e a regressão de quantis adaptativa (Møller et al., 2006). Ambas permitem prever a distribuição da produção para cada um dos instantes do horizonte de previsão, no entanto, a regressão local de quantis apresenta a desvantagem de apenas produzir um quantil de cada vez que é resolvida, o que obrigaria a correr 10 vezes o algoritmo para produzir 10 diferentes quantis. O conceito do regressão de quantis adaptativa foi aplicado com o objectivo de prever a distribuição do erro e não a distribuição da potência produzida.

Em (Juban et al., 2007) é proposta uma metodologia para prever a função densidade de probabilidade para cada instante de previsão do horizonte baseada numa técnica de *Kernel density estimation*. Tal como nos métodos anteriores também esta técnica é não-paramétrica.

Uma abordagem diferente encontra-se descrita em (Miranda et al., 2006), onde pela primeira vez é testada a aplicação de uma função de custo baseada no critério

ITL (Principe et al., 1999b) em alternativa ao clássico mínimos quadrados num problema de previsão de produção eólica. No entanto, os dados utilizados no referido artigo consistem apenas em valores medidos do parque eólico, não sendo utilizadas previsões da velocidade do vento e direcção. Em alternativa ao uso de uma rede neuronal e ao clássico algoritmo de retropropagação do erro foi usado um sistema de inferência difusa (*Takagi-Sugeno*) treinado com recurso ao algoritmo de optimização evolucionário EPSO (*Evolutionary Particle Swarm Optimization*). Os resultados apresentados apenas se resumem à distribuição de probabilidade do erro usando as duas funções de custo analisadas. Não são apresentados resultados para um horizonte de previsão completo nem comparados com um modelo de referência. Apesar de não terem sido usadas previsões do vento e dos poucos resultados disponíveis, foram verificados indícios para uma exploração da aplicação das técnicas baseadas no critério ITL.

Estimação Robusta

A elevada presença de ruído e *outliers* nas variáveis do problema de previsão eólica levou também ao desenvolvimento de técnicas que apresentam elevada robustez. Os desenvolvimentos em estimação robusta aplicada à previsão eólica são recentes e encontram-se descritos em (Aa. Nielsen et al., 2007) e (Pinson et al., 2007b). Estes modelos são também adaptativos com o intuito de lidar com o facto de a produção eólica apresentar um carácter não-estacionário.

A ideia base destes modelos é que a estimação dos parâmetros dos modelos (neste caso são funções polinomiais) com recurso ao clássico mínimos quadrados é afectada por ”outliers” e ruído dos dados. O método proposto é uma estimação robusta M (Huber, 1981) assimétrica, mas convexa, de forma a lidar com a forma não gaussiana dos erros. O objectivo principal é retirar ”peso” aos pontos considerados suspeitos, para isso a função de custo é tornada linear a partir de um valor pré-definido de erro (c). Como se trata de um problema não-estacionário o valor de c não poderá ser constante ao longo do tempo, em vez disso é definida uma proporção α dos erros que deverão ser considerados ”suspeitos”. O ponto c é estimado a partir da distribuição empírica dos erros mais recentes do modelo para que a mesma

proporção de erros positivos e negativos seja considerada, c torna-se dependente da altura do ano. Esta alteração transforma a função de custo numa função assimétrica, o que permite lidar também com a assimetria da distribuição dos erros.

3.5.2 Modelos com Treino *On Line*

Num trabalho recente sobre a história da previsão eólica foram identificadas as próximas linhas de investigação (Costa et al., 2008). Entre vários aspectos destaca-se a necessidade de desenvolver modelos em que a estimação dos parâmetros seja em tempo real, sendo até apontada como uma característica essencial a necessidade dos modelos se adaptarem a mudanças no parque eólico e no meio que os rodeia.

O problema da mudança das condições do parque eólico (mudança do conceito) já é referido em algumas publicações. Por exemplo, em (Sanchez, 2006) é detectada a mudança da curva de potência de funcionamento em dados reais de um parque eólico situado em Espanha. Também a capacidade de controlo que as actuais máquinas possuem é identificada como possível razão para estas alterações de funcionamento.

O sistema de previsão descrito em (Sanchez, 2006) é para um horizonte de previsão de 36 horas e encontra-se actualmente em funcionamento em Espanha com o nome Sipreolico. O modelo faz uso das previsões das variáveis meteorológicas obtidas com o modelo HIRLAM(Machenhauer, 1988). O sistema baseia-se na combinação adaptativa linear das previsões obtidas com nove modelos dinâmicos lineares e não lineares. Estes modelos adaptativos incluem um modelo autorregressivo puro, modelo autorregressivo com variável endógena e modelos autoregressivos com componentes de regressão linear não paramétrica. Os parâmetros dos novos modelos são estimados recursivamente com os mínimos quadrados e usando um factor de esquecimento que também é ajustado ao longo do tempo. A previsão final é obtida combinando linearmente as previsões dos modelos que melhores resultados têm oferecido nas últimas previsões. O autor apenas usa funções polinomiais e outros modelos simples de previsão porque o sistema de previsão destina-se a ser utilizado pelo operador da rede eléctrica espanhola e tem como principal requisito a capacidade de ser aplicado a diferentes parques sem a necessidade de ajustar parâmetros.

Por esta razão, o autor não recomenda o uso de redes neuronais em sistemas de previsão por possuírem um treino estático e uma necessidade de ajustar os seus parâmetros cada vez que se muda de parque.

Em (Aa. Nielsen et al., 2007) e (Pinson et al., 2007b) é justificada a necessidade de os modelos serem adaptativos com a possibilidade de rapidamente um modelo ter que passar a fazer previsões para outro parque eólico. O objectivo é tornar desnecessário ajustar os coeficientes do modelo de forma diferente para cada parque, o modelo terá de ser capaz de o fazer automaticamente. Também é referido o facto de a curva de potência apresentar uma natureza não-estacionária. Os coeficientes do modelo são estimados de forma recursiva com um factor de esquecimento que evoluiu ao longo do tempo.

Outros sistemas com estimação recursiva de parâmetros com recurso aos mínimos quadrados poderão ser encontrados em (Makarov et al., 2002) e (Nielsen & H.Madsen, 2001). Um outro exemplo de um sistema adaptativo é a regressão de quantis adaptativa descrita em (Møller et al., 2006).

Uma abordagem adaptativa com redes neuronais é descrita em (Barbounis & Theocharis, 2006) e (Barbounis et al., 2006), o objectivo é realizar previsão de produção eólica de um parque eólico usando redes neuronais MLP localmente recorrentes. Uma classe de algoritmos de aprendizagem em tempo real foi desenvolvida para treino das redes neuronais localmente recorrentes (LF-MLN:*Local Feedback Multilayer Networks*) onde a recorrência é localmente introduzida em cada neurónio. Os algoritmos são baseados no erro recursivo (RPE - *Recursive Prediction Error*). O RPE algoritmo é um método *Gauss-Newton* de pesquisa estocástico com grandes semelhanças aos mínimos quadrados recursivos.

Um algoritmo RPE global é desenvolvido onde todos os pesos das ligações são tratados de forma simultânea. Devido ao elevado peso computacional e necessidade de armazenar muitos dados três algoritmos de aprendizagem local são sugeridos para dividir o algoritmo global RPE num conjuntos de subproblemas ao nível da rede neuronal. Os algoritmos locais são o RPE desacoplado (DRPE:*Decoupled RPE*), o modelo linear do neurónio (NLM-RPE:*Neuron LinearModel RPE*) e o modelo não-linear do neurónio (NNM-RPR: *Neuron Non-linear-Model RPE*).

É também comum nestes algoritmos adaptativos a utilização de métricas que se baseiam no pressuposto de que a distribuição dos erros é Gaussiana, um exemplo desta aproximação é encontrado na distância de *Cook* que é utilizada em (Sanchez, 2006) para alteração do factor de esquecimento dos mínimos quadrados recursivos.

Grande parte dos sistemas e algoritmos de previsão presentes na bibliografia (por exemplo, todos os modelos apresentados na secção 3.5.1) apesar dos novos desenvolvimentos ainda não são treinados em tempo real, sendo essencialmente modelos de aprendizagem estáticos.

Capítulo 4

Treino de Sistemas Adaptativos

Este capítulo inicia-se com uma breve explicação da teoria das redes neuronais artificiais. Serão apresentados os conceitos básicos relacionados com as redes neuronais. Apenas é abordada a teoria das redes neuronais MLP (*multilayer perceptrons*) porque são o tipo mais comum de redes neuronais e também as que melhores resultados apresentam na previsão eólica (Rosado et al., 2006).

De uma forma mais detalhada, serão abordados conceitos que permitem extrair informação directamente dos dados, conceito este que é o objectivo principal de sistemas de aprendizagem. A ideia de treinar algoritmos de aprendizagem baseada nos conceitos da teoria da informação (Critério ITL–Information Theoretic Learning) (Principe et al., 1999b) (Principe & Erdogmus, 2006) surgiu na Universidade da Flórida. A abordagem ITL usa uma forma menos habitual de calcular a quantidade de informação, através da entropia de *Renyi* e combina-a com as janelas de *Parzen*. No fundo o critério ITL substitui os momentos de segunda ordem: a variância é substituída pela **Entropia**; a correlação substituída pela **Correntropia**; a Minimização do Erro Quadrático (MSE - *Mean Square Error*) pela **Minimização da Entropia dos Erros (MEE - *Minimization Error Entropy*)** ou **Maximização da Correntropia (MCC - *Maximization Correntropy Criterion*)**; as distâncias no espaço dos dados (i.e. Euclidiana) por distâncias num **espaço probabilístico**. Nas secções seguintes será apresentada a abordagem ITL, bem como técnicas de optimização do treino de redes neuronais com estes critérios.

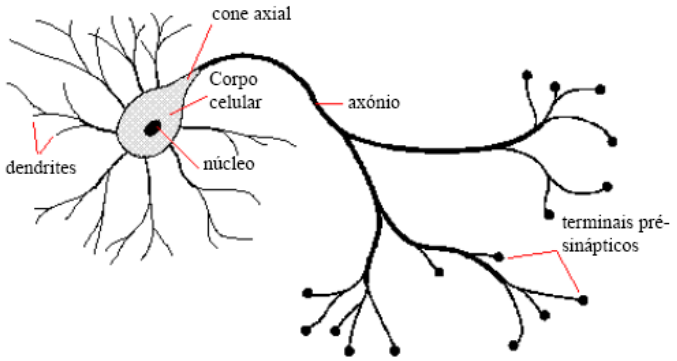

Figura 4.1: *Esquema de um neurónio biológico.*

4.1 Conceitos Básicos de Redes Neuronais

Como elemento base de uma rede neuronal, tem-se o neurónio, que é uma unidade de processamento elementar. A ideia do neurónio como unidade de computação inspira-se na biologia dos neurónios do sistema nervoso, nomeadamente humano. Os sinais electroquímicos captados pelas dendrites, cada uma com a sua intensidade própria, adicionam-se no corpo celular e, se a resultante ultrapassar um certo limiar, criam uma onda solitária de despolarização eléctrica que tem início no cone axial e avança pelo axónio, até atingir os terminais pré-sinápticos onde, por intermediação química, activará dendrites de outros neurónios.

Nas redes neuronais artificiais também se têm neurónios com entradas e uma saída e um processamento interno não linear que transforma as entradas numa saída. Os neurónios encontram-se organizados em camadas, com ligações unidireccionais sempre num sentido progressivo, da entrada para a saída da rede. As ligações entre neurónios estão associadas a pesos $w_{kx,pz}$, por forma a que um sinal emitido por um neurónio é multiplicado pelo peso da ligação antes de dar entrada num neurónio seguinte. Um neurónio índice x da camada k recebe sinais O_{pz} dos neurónios da camada anterior p e efectua as seguintes transformações representadas na figura 4.2, antes de produzir um sinal de saída O_{kx} que, por sua vez, irá activar neurónios na camada seguinte:

$$net_{kx} = \sum_{z=1}^{m} w_{kx,pz} \cdot O_{pz} \quad ; \quad O_{kx} = f_{kx}(net_{kx}) \tag{4.1}$$

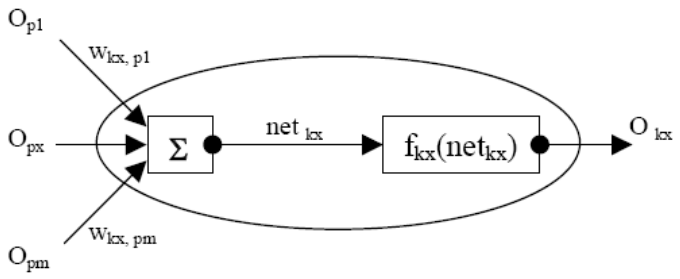

Figura 4.2: *Operações executadas num nó genérico, designado como de ordem x na camada k, recebendo sinais da camada anterior p, depois de afectados pelos pesos w das ligações.*

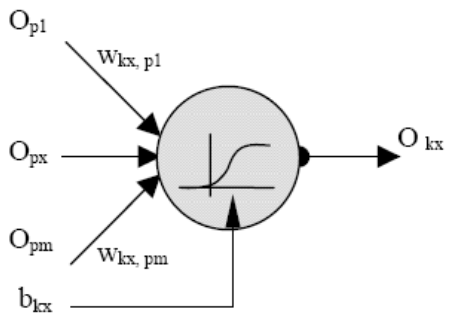

Figura 4.3: *Representação de um neurónio x na camada k produzindo uma resposta O_{kx} com m entradas oriundas da camada anterior p e ainda um limiar b_{kx} independente do processamento das camadas anteriores.*

A função f é chamada função de activação do neurónio. As funções mais correntemente adoptadas são a linear, em degrau e a sigmóide e suas variantes.

Um dos sinais de entrada num neurónio pode ser o correspondente a um limiar de activação (também chamado de *bias* do neurónio) da resposta do neurónio, como se pode ver na figura 4.3. Considerando o *bias* fica-se com

$$O_{kx} = f_{kx}\left(\sum_{z=1}^{m} w_{kx,pz} O_{pz} - b_{kx}\right) \tag{4.2}$$

O *bias* $b_{kx} > 0$ representa um deslocamento para a direita da função de activação, exigindo que a acumulação de sinais de entrada ultrapasse um certo limiar para originar uma resposta significativa do neurónio. A existência de um valor para o *bias* é facilmente incluído no modelo, considerando que apenas se trata de mais um peso cujo o valor também deverá ser determinado no decurso do treino da rede

neuronal. A equação 4.2 passaria a ter a seguinte forma

$$O_{kx} = f_{kx}\left(\sum_{z=1}^{m} w_{kx,pz} O_{pz} - b_{kx} \cdot (-1)\right) \tag{4.3}$$

No fundo é assumido que (-1) é o sinal de uma entrada suplementar com um peso b_{kx} a determinar.

4.1.1 Termo de Momento

O método de treino usualmente utilizado é o algoritmo de retropropagação do erro. Este algoritmo tem como base matemática o método da descida mais íngreme, trata-se de um método de pesquisa local, logo o treino das redes neuronais sofre dos mesmos problemas dos métodos gradientes locais. Estes métodos são muito susceptíveis a ficarem aprisionados em óptimos locais, precisamente porque, nas vizinhanças de um óptimo local, o gradiente aponta sempre na "direcção errada" e não permite que a solução corrente evolua em outro sentido. Da mesma forma "vales" alongados na função objectivo constituem armadilhas para os algoritmos de gradiente, levando à estagnação do progresso na direcção do óptimo com dispêndio do esforço de inúmeras iterações com reduzida eficiência.

Este problema foi alvo de estudo em diversas publicações com o objectivo de aperfeiçoar o algoritmo de treino e acelerar o treino da rede neuronal. Alguns destes trabalhos poderão ser encontrados em (Silva & Almeida, 1991) (Fahlman, 1988) (Riedmiller, 1994) (Bishop, 1995).

Um método heurístico e com bons resultados é a introdução do chamado "termo de momento", que mais não é do que uma tentativa de estabilizar o progresso das iterações numa direcção descoberta como produtiva, como que "saltando por cima" de "rugosidades" locais da função objectivo. O termo de momento é introduzido no algoritmo de actualização dos pesos da seguinte forma:

$$\Delta w_t = -\eta \frac{\partial E}{\partial w} + \omega \Delta w_{t-1} \tag{4.4}$$

4.1.2 Modos de Treino

O treino da rede neuronal com o algoritmo de retropropagação do erro é realizado de forma supervisionada, onde são apresentados um a um exemplos de um conjunto

de treino. Durante a fase de treino, uma passagem completa pelo conjunto de treino é chamada de "época". Uma prática comum é realizar o ajuste dos pesos e *bias* da rede durante várias "épocas" até que um determinado critério de convergência seja atingido. Este procedimento deve ser executado com extremo cuidado para evitar um sobre-ajustamento da rede neuronal aos dados de treino, sendo necessário garantir a capacidade de generalização da rede neuronal.

A maioria das técnicas existentes para evitar o sobre-ajustamento consistem na paragem antecipada do treino da rede neuronal. Um método muito comum e que se encontra implementado no MATLAB[1] consiste em: dividir os dados em treino e validação, este último conjunto é utilizado para aferir a qualidade do treino e a capacidade de generalização da rede (estimação do erro de generalização); durante a fase de treino, em cada época é calculado o erro no conjunto de validação parando-se quando ao fim de um número definido de iterações o erro aumentar sucessivamente; a melhor solução de pesos e bias para a rede é a solução com menor erro obtida durante o processo de treino. Outros métodos poderão ser encontrados em (Kohavi, 1995).

O treino de retroprogação do erro pode ser realizados de três formas distintas:

- **Modo *Batch* ou *Off Line*:** neste modo de treino a actualização dos pesos apenas é realizada ao fim de uma época. A função de custo no caso dos mínimos quadrados é o erro médio quadrático. No Apêndice A o algoritmo 2 descreve o procedimento deste método de treino para o critério MSE.

- **Modo Sequencial ou estocástico:** neste modo de treino a actualização dos pesos é realizada cada vez que um exemplo do conjunto de treino da rede neuronal é apresentado ao algoritmo de treino da rede. Para cada exemplo de treino x_i o erro entre o saída y_i e o valor desejado d_i é calculado, sendo depois esse erro retropropagado em todos os pesos da rede neuronal. Em cada "época" tem-se um número de actualizações dos pesos igual ao número de exemplos do conjunto de treino. O algoritmo 1 do Apêndice A descreve o procedimento deste método de treino para o critério MSE. Este modo de treino deriva do

[1] www.mathworks.com/

método LMS (*Least Mean Square*) proposto em (Widrow & Stearns, 1985), onde o gradiente não é calculado como um valor médio sobre um conjunto de dados, é sim um gradiente "instantâneo" sobre apenas um exemplo de treino. A estimativa do gradiente desta forma apresenta muito ruído, mas como o processo de treino não encontra o mínimo em apenas um passo, então durante o processo iterativo o ruído associado ao gradiente será "filtrado".

- **Modo *On Line*:** O funcionamento do método sequencial em tempo real será apelidado de treino *on line* ao longo da tese. No treino *on line* cada novo exemplo de treino disponível (i.e. chega um novo valor medido pelo SCADA) é propagado desde a entrada da rede até à saída, sendo depois o erro à saída retropropagado apenas uma vez, actualizando de imediato os pesos da rede. Esta estratégia de treino em tempo real de redes neuronais foi aplicada em (Rodrigues & Gama, 2007) com bons resultados na previsão de carga de um sistema eléctrico. Poderá ser conveniente previamente realizar um treino *off line* com os dados disponíveis (se existirem), para determinar os pesos iniciais do treino *on line*.

Os modos de treino *Batch* e *sequencial* têm as suas vantagens e desvantagens. O treino *Batch* permite estimar com maior rigor o gradiente, o que garante pelo menos uma convergência para um mínimo local. Por outro lado, o treino sequencial não tem garantias de convergência para um óptimo local devido ao seu carácter estocástico, no entanto, permite um menor esforço computacional e evita ficar "preso" em óptimos locais.

4.2 Funções de Custo

Como se sabe, os pesos da redes neuronal são ajustados de forma a minimizar (ou maximizar) uma função de custo. Uma função de custo é utilizada independentemente de qual o algoritmo de treino da rede neuronal, tanto as meta-heurísticas (i.e. enxame de partículas) (Miranda et al., 2006) como o algoritmo de retropropagação usam uma função de custo (nas meta-heurísticas é por vezes chamada de função de adaptação). Todas as funções de custo são em função do erro ($e_i = y_i - d_i$) cometido

pela rede neuronal nos exemplos de treino. São várias as funções que são usadas em redes neuronais, em (Bishop, 1995) são apresentadas as funções mais comuns em problemas de regressão e classificação:

- soma do quadrados dos erro: $\frac{1}{2}\sum_{i=1}^{N} e_i^2$;
- erro de *Minkowski*: $\sum_{i=1}^{N} |e_i|^R$;
- erro médio quadrático: $\frac{1}{N}\sum_{i=1}^{N} e_i^2$;
- *cross-entropy*: $-\sum_{i=1}^{N} [t_i \cdot \log(y_i) + (1-t_i)\log(y_i)]$, onde t_i é a classe do exemplo i;

Outras funções de custo utilizadas em redes neuronais poderão ser encontradas em (Gangal et al., 2007).

De todas as funções apresentadas, o erro médio quadrático é a função de custo mais utilizada em diversos problemas de regressão, no quais se engloba a previsão de produção eólica. Esta função de custo tem uma aplicação óptima quando o erro é caracterizado por uma função Gaussiana, como é demonstrado em (Bishop, 1995). Em problemas não lineares, com elevado ruído, *outliers* e não-Gaussianos esta função de custo apresenta várias limitações. As limitações advêm do facto de não apresentar qualquer robustez quanto ao ruído e basear-se apenas na minimização da variância do erro, desprezando informação contida em momentos de ordem superior.

A função de custo MSE também apresenta algumas vantagens: a sua superfície de erro é suave e continuamente derivável; é uma função quadrática convexa, logo, com apenas um mínimo global; não implica o uso de um elevado esforço computacional, quando aplicada no treino de sistemas adaptativos.

4.3 Aprendizagem Baseada em Teoria da Informação

4.3.1 Entropia

Num sistema de comunicação a informação contida num certo acontecimento é maior quanto menor for a sua probabilidade de ocorrência. *Claude Shannon* definiu Entropia da seguinte forma: se o conteúdo de uma mensagem é totalmente conhecido à priori o seu conteúdo de informação é nulo, no entanto quanto menos previsível uma mensagem é, maior é o seu conteúdo de informação. *Shannon* em (Shannon, 1948) introduziu os conceitos de Entropia e informação mútua, e lançou os conceitos base para o campo da teoria da informação. Definiu entropia H_S de uma distribuição discreta de probabilidade $P = (p_1, p_2, \ldots, p_N)$ como:

$$H_s(P) = \sum_{k=1}^{N} p_k \cdot \log\left(\frac{1}{p_k}\right), \; com \quad \sum_{k=1}^{N} p_k = 1 \quad e \quad p_k \geq 0 \quad e \quad 0 \cdot \log_2(0) = 0 \tag{4.5}$$

é portanto a quantidade média de informação contida numa única observação de uma variável aleatória (v.a.) X que toma valores $x_1, x_2, \ldots, x_N$ com probabilidade $p_k = P(x = x_k)$ com $k = 1, 2, \ldots, N$. A entropia mede a quantidade média de informação "transportada" pela mensagem x. Esta definição encontrou muita aplicação, nomeadamente em sistemas de comunicação. No entanto existem várias definições possíveis, um exemplo é a Entropia de *Renyi* (Renyi, 1976), que se define como:

$$H_{R\alpha}(P) = \frac{1}{1-\alpha} \log \sum_{k=1}^{N} p_k^{\alpha}, \; com \quad \alpha > 0, \alpha \neq 1 \tag{4.6}$$

A entropia de *Renyi* é na verdade uma família de funções $H_{R\alpha}$ dependentes de um parâmetro real α.

Existe uma relação bem conhecida entre as definições de *Shannon* e *Renyi*:

$$H_{R\alpha} \geq H_S \geq H_{R\beta} \quad se \quad \beta > 1 > \alpha > 0 \tag{4.7}$$

$$\lim_{\alpha \to 1} H_{R\alpha} = H_S \tag{4.8}$$

Quando $\alpha = 2$, $H_{R2} = -\log \sum_{k=1}^{N} p_k^2$, temos a chamada Entropia quadrática, devido ao quadrado da probabilidade.

Para v.a. contínuas com função de densidade de probabilidade $f_Y(z)$, pode-se obter a versão diferenciável da Entropia de *Renyi* (Renyi, 1976):

$$H_{R\alpha}(Y) = \frac{1}{1-\alpha}\log\left(\int_{-\infty}^{+\infty} f_Y(z)^{\alpha}\, dz\right) \tag{4.9}$$

$$H_{R2} = -\log\left(\int_{-\infty}^{+\infty} f_Y(z)^{2}\, dz\right) \tag{4.10}$$

No uso do critério ITL aplicado à aprendizagem baseada em exemplos é desejado estimar directamente a Entropia a partir dos dados sem assumir qualquer fdp.

A definição de entropia de *Shannon*, que consiste numa soma pesada de logaritmos de probabilidades, não é atractiva para o objectivo de estimar a entropia com algoritmos simples. Estimadores computacionalmente mais simples da entropia e que sejam contínuos e diferenciáveis são mais desejados neste tipo de problemas. A definição de entropia de *Renyi*, logaritmo de uma soma de probabilidades, permite uma estimação mais simples da entropia.

Na secção seguinte será apresentado um modo eficiente de estimar o valor da Entropia a partir dos dados.

4.3.2 Janelas de Parzen e Estimação da Entropia

Vários autores apresentaram expressões para o cálculo da Entropia de v.a. contínuas, algumas dessas formas de estimação podem ser encontradas em (Gokhale, 1975) e (Lazo & Rathie, 1978). Existem várias formas de estimação não-paramétrica de fdp, em (Marques de Sá, 2006) são revistos vários métodos.

A estimação de densidade de *Kernel* ou janelas de *Parzen* (Parzen, 1962) é um método de estimação de densidade não-paramétrico muito utilizado. A técnica das janelas de *Parzen* consiste em estimar a fdp de uma variável aleatória a partir de uma amostra discreta de N pontos , i = 1,2,. . . , N num espaço M-dimensional. Esta técnica usa uma função nuclear (*Kernel*) centrada em cada ponto da amostra – interpreta cada ponto, como sendo localmente descrito por uma fdp de Dirac; o ponto é substituído (aproximado) por um conjunto contínuo cuja densidade de probabilidade é descrita pelo *Kernel*. A expressão da estimativa da fdp real $f_Y(z)$ da variável aleatória $Y \in \Re^M$, com base numa amostra formada por um conjunto de N pontos,

consegue-se a partir da soma das contribuições individuais do *Kernel* aplicado em cada ponto, na forma:

$$f_Y(z) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(z - y_i) \tag{4.11}$$

onde o vector $y_i \in \Re^M$ são amostras da v.a., e $K(.)$ é a função Kernel e σ o tamanho da janela de *Parzen*. A função *Kernel* e o tamanho da janela pode ser optimizado com o principio da máxima verosimilhança (Duin, 1976) ou regras empíricas para obter parâmetros próximos dos óptimos (Devroye & Lugosi, 2001) (Silverman, 1986).

Ao longo desta tese a função *Kernel* utilizada é o *Kernel* Gaussiano simétrico. O *Kernel* Gaussiano foi a escolha preferida porque é derivável e permite simplificar computacionalmente o algoritmo (Principe et al., 1999b). A sua propriedade mais interessante e importante para o problema é: sendo $y_i \in \Re^M$ e $y_j \in \Re^M$ duas amostras de v.a. independentes, Σ_1 e Σ_2 duas matrizes de covariância, tem-se a relação

$$\int_{-\infty}^{+\infty} G(z - y_i, \Sigma_1) G\left(z - y_j, \Sigma_2\right) dz = G\left(\left(z_i - z_j\right), (\Sigma_1 + \Sigma_2)\right) \tag{4.12}$$

onde G é o *Kernel* Gaussiano.

A equação 4.12 representa a convolução de duas funções Gaussianas, isto é não só um resultado conhecido como extremamente simplificador, porque a convolução de duas Gaussianas é ainda uma Gaussiana cuja variância é a soma das variâncias. O *Kernel* Gaussiano simétrico tem a seguinte forma:

$$K\left(z_k - y_{ik}, \sigma^2\right) = G\left(z_k - y_{ik}, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(z_k - y_{ik})^2} \tag{4.13}$$

$\sigma^2 I$ é a matriz de covariâncias (que se assume corresponder a variâncias independentes e iguais em todas as dimensões). Por isso, a matriz de covariâncias é diagonal com todos os elementos da diagonal principal iguais e os restantes nulos e pode representar-se a partir de uma matriz identidade I multiplicada por uma constante $\sigma^2 I$. No fundo tem-se uma fdp centrada no valor do ponto i na sua dimensão k e com variância σ^2.

O tamanho da janela de *Parzen* na estimativa da fdp tem grande importância, valores elevados de σ levam a estimativas mais "suaves" e valores pequenos levam a estimativas com mais picos, tal facto pode ser verificado na figura 4.4.

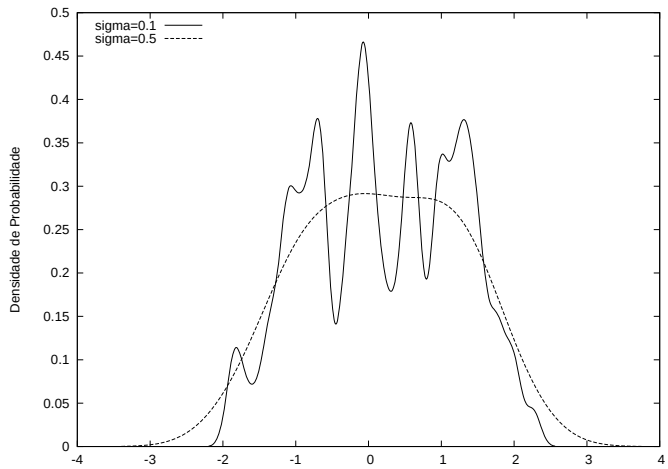

Figura 4.4: *Ilustração gráfica do impacto de σ na estimação da fdp com o método das janelas de Parzen.*

A estimativa da entropia H_{R2} será feita aplicando o conceito de janelas de *Parzen*. Assim, substituindo na equação 4.10 da entropia de Renyi a fdp $f_Y(y)$ pela sua aproximação conseguida pela aplicação de janela de *Parzen* a cada um dos pontos y:

$$H_{R2} = -\log \int_{-\infty}^{+\infty} f_Y^2(z)\, dz \quad com \quad \hat{f}_Y(z) = \frac{1}{N}\sum_{i=1}^{N} G\left(z - y_i, \sigma^2 I\right) \tag{4.14}$$

fica então:

$$H_{R2} = -\log \int_{-\infty}^{+\infty} \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G\left(z - y_i, \sigma^2 I\right) G\left(z - y_j, \sigma^2 I\right) dz \tag{4.15}$$

o que permite escrever:

$$H_{R2} = -\log \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \int_{-\infty}^{+\infty} G\left(z - y_i, \sigma^2 I\right) G\left(z - y_j, \sigma^2 I\right) dz \tag{4.16}$$

ou ainda:

$$H_{R2}(y) = -\log V(y), \quad sendo \tag{4.17}$$

$$IP(y) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \int_{-\infty}^{+\infty} G\left(z - y_i, \sigma^2 I\right) G\left(z - y_j, \sigma^2 I\right) dz \tag{4.18}$$

Na terminologia ITL, $IP(y)$ recebe a designação de potencial de informação do conjunto y. Como o logaritmo é uma função monótona e crescente, minimizar a entropia equivale a maximizar o potencial de informação (Erdogmus & Principe, 2002b).

Aplicando na definição de potencial de informação à propriedade da equação 4.12, fica-se então com:

$$IP(y) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G\left(y_i - y_j, 2\sigma^2 I\right) \quad ou \quad IP(y) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} V_{ij} \tag{4.19}$$

A avaliação das propriedades estatísticas de um sinal em modo *on line* necessita de uma fórmula recursiva. Em (Erdogmus, 2002) são propostas duas formas de estimação recursiva da entropia quadrática. A derivação da fórmula recursiva concentrou-se no potencial de informação devido à relação existente com a entropia. A primeira forma de estimação da entropia é a chamada *estimação exacta da entropia*, supondo que no instante k, já temos k amostras, a actualização do potencial quando no instante $k+1$ chega um novo exemplo consiste em:

$$IP_{k+1} = (1-\lambda)\,IP_k + \frac{2\lambda}{2k+1}\sum_{i=1}^{k} K_\sigma(y_{k+1} - y_i) + \frac{\lambda}{2k+1}K_\sigma(0) \tag{4.20}$$

onde λ é o factor de memória variável com o tempo, definido como $1 - k^2/(k+1)^2$. Uma vez actualizado o potencial de informação, o novo valor da entropia pode ser calculado aplicando $H_{k+1} = -\log(IP_{k+1})$. O termo *exacto* provém do facto de esta fórmula recursiva levar ao mesmo valor da entropia estimada com a equação 4.19. Esta fórmula recursiva um vez que necessita de todos os exemplos para o cálculo apenas tem utilidade em ambientes estacionários e sem um fluxo contínuo de dados. A complexidade computacional da estimação da entropia foi reduzida de $O\left(N^2\right)$ para $O(k)$.

Para a segunda fórmula recursiva alterou-se a estratégia de estimação, em vez de se estimar recursivamente o potencial de informação, a estratégia consiste em recursivamente actualizar a fdp e usar um factor de memória fixo. Considerando um factor de memória λ e uma média amostral sobre uma janela de tamanho M, tem-se o seguinte estimador recursivo:

$$IP_{k+1} = (1-\lambda)\,IP_k + \frac{\lambda}{M}\sum_{i=k-M+1}^{k} K_\sigma(x_i - x_{k+1}) \tag{4.21}$$

O estimador recursivo da equação 4.21 é chamado de *estimador recursivo com esquecimento*. Ainda é necessário armazenar alguns exemplos para a janela de tamanho fixo M, mas muito menos do que os N exemplos necessários na equação 4.19.
A complexidade do algoritmo foi reduzida de $O\left(N^2\right)$ para $O(M)$.

Em (Erdogmus, 2002) e (Erdogmus & Principe, 2002b) a estimação da entropia é generalizada para qualquer ordem α da entropia de *Renyi*, no entanto, em testes realizados em dados artificiais os melhores resultados obtidos foram com $\alpha = 2$. Uma outra razão que levou à utilização da entropia de *Renyi* quadrática está relacionada com o facto de a equação 4.20 apenas ser válida para a entropia quadrática (Erdogmus, 2002).

Propriedades da Entropia

Todas as propriedades encontram-se descritas e demonstradas em (Erdogmus, 2002) e (Erdogmus & Principe, 2002b), sendo apenas apresentadas algumas propriedades essenciais para compreender a sua aplicação no treino de sistemas adaptativos.

Propriedade 1: O estimador da entropia é insensível à média da fdp. No treino supervisionado de sistemas esta propriedade indica que o algoritmo vai convergir para um conjunto de pesos que originam erros com média não nula (Erdogmus & Principe, 2000). Este problema é corrigido somando um *bias* à resposta do algoritmo de aprendizagem, depois do processo de treino, de forma a forçar o erro no conjunto de treino a ser nulo.

Propriedade 2: Se o máximo valor do *Kernel* $K(\xi)$ é conseguido quando $\xi = 0$, então o valor mínimo da entropia é obtido quando todas as amostras são iguais umas às outras, i.e. $x_1 = x_2 = x_N = c$. No treino supervisionado é imperativo que a função de custo atinja o óptimo global, quando todas as amostras do erro sejam zero. Esta propriedade, torna o critério de minimização da entropia do erro (que será apresentado mais tarde neste capítulo) válido para o treino de algoritmos de aprendizagem.

Propriedade 3: Se a função *Kernel* $K_\sigma()$ é contínua, derivável e simétrica, então o mínimo global descrito na propriedade anterior tem gradiente nulo e uma matriz Hessiana com valores próprios semi-finitos. Sobre estas condições o valor máximo do *Kernel* apenas é atingido quando o seu argumento é zero.

Propriedade 4: Minimizando a entropia de *Renyi* do erro está-se a minimizar a

distância *Csiszar* (Csiszár & Körner, 1981), definida como:

$$D_g(q;p) = \int_{-\infty}^{+\infty} q(x)\, g\left(\frac{p(x)}{q(x)}\right) dx \tag{4.22}$$

onde $g()$ tem de ser convexa com $g(1) = 0$. No treino de sistemas adaptativos consiste na minimização entre a fdp dos valores desejados (”reais”) e a fdp da resposta do algoritmo de aprendizagem.

4.3.3 Correntropia

A abordagem ITL inspirou o aparecimento de novas medidas estatísticas, como é o caso de: *symmetric information potential* (Liu et al., 2007b), autocorrentropia (Santamaría et al., 2006) mais tarde aplicada a duas v.a. (Liu et al., 2006)e (Liu et al., 2007a) sob o nome de correntropia. Em (Santamaría et al., 2006) é demonstrado que tal como a autocorrelação, também a autocorrentropia mede a similaridade entre instantes temporais, mas o valor médio ao longo dos instantes converge assintoticamente para o potencial de informação estimado com recurso às janelas de *Parzen* (logo para a entropia da v.a.), daí o seu nome. Esta propriedade demonstra que a correntropia possui informação sobre a fdp, ou seja, contém momentos de ordem superior da fdp.

A correntropia está directamente relacionada com a probabilidade do quanto similares duas v.a. são na vizinhança do espaço de junção controlado pela largura da janela de *Parzen*. No fundo, o tamanho da janela funciona como um *zoom* que controla a janela de observação na qual a similaridade é avaliada. A possibilidade de ajustar o tamanho da janela *Parzen* fornece um mecanismo eficaz que permite eliminar o efeito de *outliers*.

A correntropia de um processo estocástico $\{x_t\}$ é definida como:

$$V_x(t,s) = E(\delta(x_t - x_s)) = \int \delta(x_t - x_s)\, p(x)\, dx \tag{4.23}$$

podendo perfeitamente ser estimada usando a teoria dos Kernels:

$$\hat{V}_x(t,s) = E(k(x_t - x_s)) \tag{4.24}$$

A correntropia entre duas v.a. escalares X e Y é definida por:

$$V_\sigma(X,Y) = E\left[K_\sigma(X-Y)\right] \tag{4.25}$$

Na prática, a correntropia é estimada a partir de uma amostra finita de dados $\{(x_i, y_i)\}_{i=1}^N$, levando ao seguinte estimador para a correntropia:

$$\hat{V}_{N,\sigma}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(x_i - y_i) \tag{4.26}$$

Propriedades da Correntropia

Todas as propriedades da correntropia encontram-se descritas em (Liu et al., 2007a) e (Santamaría et al., 2006). Apenas serão descritas três propriedades essenciais para compreender a sua aplicação no treino de sistemas adaptativos.

Propriedade 1: A correntropia é positiva e limitada: $0 < V(X,Y) \leq 1/\sqrt{2\pi}\sigma$. Atinge o seu valor máximo se e só se $X = Y$.

Propriedade 2: Assumindo que os pontos $\{(x_i, y_i)\}_{i=1}^N$ provêm de uma fdp conjunta $f_{X,Y}(X,Y)$, e $\hat{f}_{X,Y;\sigma}(x,y)$ a sua estimativa com recurso a janelas de *Parzen* com tamanho σ. A correntropia estimada com um $\sigma' = \sqrt{2}\sigma$ é o integral de $\hat{f}_{X,Y;\sigma}(x,y)$ ao longo da linha $x = y$ (Liu et al., 2006):

$$\hat{V}_{\sqrt{2}\sigma}(X,Y) = \int_{-\infty}^{+\infty} \hat{f}_{X,Y;\sigma}(x,y)|_{x=y=u}\, du \tag{4.27}$$

Esta propriedade indica que a correntropia é mais do que uma extensão da correlação cruzada, ela também permite quantificar a probabilidade de dois eventos serem iguais. É de salientar que a correntropia não avalia a similaridade em todo o espaço conjunto, mas para $X = Y$ fornece uma estimativa da densidade de probabilidade.

Em aplicações práticas, a fdp conjunta é desconhecida e apenas é conhecido um conjunto finito de dados para estimar a correntropia. O facto de existir um número finito de dados também limita inferiormente o tamanho da janela, janelas com um tamanho muito pequeno poderão levar a estimativas erradas. Assumindo que tamanho da janela usado é σ que é relativamente pequeno quando comparado com variância da distribuição dos dados, é possível usar um rectângulo com largura $\sqrt{\pi/2}\sigma$ como aproximação em vez do *Kernel* Gaussiano e obtém-se uma estimação

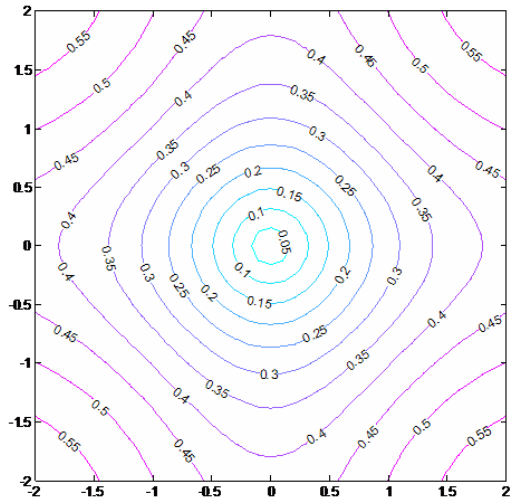

Figura 4.5: *Curvas de nível de CIM(X,0) num espaço bi-dimensional com* $\sigma = 1$.

mais precisa da entropia na forma:

$$V(X,Y) = E[K_\sigma(X-Y)] \approx \frac{P\left(|Y-X| < \sqrt{\pi/2}\sigma\right)}{\sqrt{\pi/2}\sigma} \tag{4.28}$$

Visto que a correntropia avalia directamente dos dados a probabilidade ($|Y-X| < 1.25\sigma$) para um Kernel com tamanho σ, ela poderá ser usada como medida local de similaridade em aplicações de treino supervisionado onde o critério MSE é tradicionalmente usado. É uma medida resistente a *outliers* porque não se preocupa com o quanto diferentes são os dois eventos fora do intervalo definido por 1.25σ.

Propriedade 3: Correntropia, como estimador amostral, induz uma métrica no espaço amostral. Dado dois vectores $X = (x_1, x_2, \ldots, x_N)^T$ e $Y = (y_1, y_2, \ldots, y_N)^T$ no espaço amostral, a função $CIM(X,Y) = (K(0) - V(X,Y))^{1/2}$ define a métrica no espaço amostral – *Correntropy Induced Metric (CIM)*. Esta propriedade revela o significado geométrico da correntropia no espaço amostral. A métrica CIM respeita as propriedades necessárias para ser considerada uma métrica (Liu et al., 2007a).

A figura 4.5 mostra curvas de nível da distância $CIM(X,O)$ de X à origem num espaço bi-dimensional. Na figura 4.5 observa-se que CIM comporta-se como uma norma L2 quando dois pontos estão próximos, e esta área é chamada de zona Euclidiana; fora da zona Euclidiana CIM comporta-se como como uma norma L1 na chamada zona de transição; quando os dois pontos se encontram muito afastados a

métrica satura e torna-se insensível à distância (norma L0), é a chamada zona de rectificação. Esta propriedade mostra a robustez inerente à métrica CIM.

Uma outra observação é que o tamanho do Kernel σ controla o tamanho das zonas. Um pequeno valor de σ leva a uma zona Euclidiana apertada e uma grande zona de rectificação, enquanto que um valor elevado implica uma zona Euclidiana grande.

4.3.4 Critérios ITL para Treino de Sistemas Adaptativos

Minimização da Entropia do Erro

Um dos critérios que resulta da abordagem ITL, nomeadamente da estimação de entropia, foi a minimização da entropia do erro (MEE) segundo (Erdogmus & Principe, 2002b) e (Erdogmus & Principe, 2002a). No caso de uma rede neuronal, o critério MEE propõe que o ajuste dos pesos seja conseguido pela aplicação de um método de optimização apropriado que encontre os pesos w que originam um valor mínimo da equação 4.14. Num problema de treino supervisionado tem-se um erro dado por $\epsilon = Di - y_i$ (T_i o valor desejado e y_i a saída da rede), logo o objectivo será minimizar a entropia da distribuição dos erros, levando a que toda a informação fique representada na rede neuronal pelos seus pesos w e seja o mínimo possível na distribuição dos erros.

O critério MEE é formulado da seguinte forma:

$$\min H_{R2}(\varepsilon) \tag{4.29}$$

Este critério ao contrário do clássico MSE, onde a minimização da variância repousa apenas na informação sobre cada ponto individualmente, exige a consideração da relação mútua entre pares de pontos. Sendo por isso possível com este critério extrair mais informação dos dados. No entanto, quando o tamanho do *Kernel* tende para infinito equivale a um problema de minimização da variância (Principe et al., 2002).

O critério da minimização da entropia é assintoticamente robusto ao ruído com média nula, seja qual for a fdp do ruído (Erdogmus, 2002).

É importante salientar que o principal objectivo não é ter uma estimação o melhor possível da fdp e consequentemente da entropia, isto porque ter janelas de *Parzen* com um tamanho muito reduzido coloca problemas de óptimos locais no treino de sistemas adaptativos. No caso da rede neuronal e outros métodos que realizem o seu treino com base em gradientes, o mais importante é o sinal do gradiente e não o seu valor. A estimativa com janelas de *Parzen* é tendenciosa (enviesada) e implica um *trade-off* entre conseguir estimativas menos tendenciosas e limitar a variância da estimativa. A estimativa com enviesamento nulo seria com janela de *Parzen* de largura $\sigma = 0$ (equivale a colocar um impulso de *Dirac*) em cada amostra e faria "explodir" a variância. Por outro lado, uma janela demasiado larga impede a localização de um óptimo.

Maximização da Correntropia

No treino supervisionado quando se está a comparar a saída de um algoritmo com o valor real (desejado) está-se no fundo a fazer a seguinte pergunta: "qual é a probabilidade das duas variáveis (medidas) serem iguais?"

O objectivo do critério MCC (*Maximization Correntropy Criterion*) consiste em maximizar a correntropia, ou seja, maximizar a probabilidade de a reposta do sistema adaptativo e o valor real serem iguais.

Em (Liu et al., 2007a) é demonstrada a estreita relação do critério MCC com a estimação-M (Huber, 1981).

Quando comparado com o critério MEE, o critério MCC tem a vantagem de possuir uma complexidade computacional de $O(N)$ ao contrário do critério MEE que apresenta uma complexidade de $O\left(N^2\right)$.

Na figura 4.6 encontra-se a comparação entre as funções de custo dos critérios MSE e MCC.

Sendo X e Y duas v.a. e $E = Y - X$, o erro quadrático $MSE(X,Y)$ é definido da seguinte forma:

$$MSE(X,Y) = E\left[(X-Y)^2\right] = \int_{x,y}\int (x-y)^2 f_{X,Y}(x,y)\,dx\,dy = \int_e e^2 f_E(e)\,de \qquad (4.30)$$

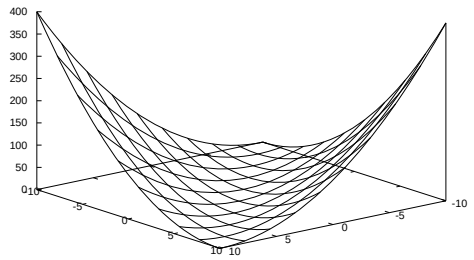

Figura 4.6: *Erro quadrático* $MSE(X,Y) = (Y-X)^2$.

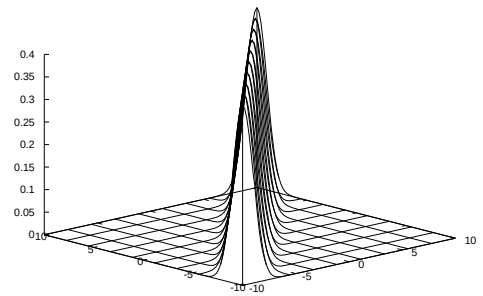

Figura 4.7: *Correntropia* $V(X,Y) = k(Y-X)$ *com* $\sigma = 1$.

enquanto que a correntropia é definida da seguinte forma:

$$V(X,Y) = E[k(X-Y)] = \int_{x,y}\int k(x-y)\, f_{X,Y}(x,y)\, dx\, dy = \int_{e} k(e)\, f_E(e)\, de \qquad (4.31)$$

Na figura 4.6 verifica-se que existe um aumento quadrático para valores afastados da linha $x = y$, que resulta na amplificação da contribuição de pontos que estão afastados do valor médio da distribuição do erro. É por esta razão que resíduos com distribuição gaussiana garantem optimalidade para o critério MSE. É também esta a razão pela qual resíduos com distribuições não-Gaussianas tornam o critério MSE não óptimo, em particular se as distribuições dos erros apresentarem *outliers*, forem assimétrica ou de média não nula. Ao contrário do critério MSE que é um critério global, o critério MCC é um critério local como se pode verificar na figura 4.7. Dizer que o critério é global significa que todas as amostras no espaço conjunto vão contribuir com bastante peso para a medida de similaridade, enquanto que num critério local o valor da contribuição de cada amostra é regulado pelo tamanho da função Kernel ao longo da linha $x = y$.

Minimização da Entropia do Erro com Pontos de Referência

O MCC é um critério local, logo, apenas tem em consideração a parte da distribuição do erro que se encontra ”dentro” do tamanho do *Kernel*, o que não permite minimizar a informação contida na parte da distribuição não abrangida pelo *Kernel*.

Pelo contrário, o critério MEE tem em consideração toda a fdp do erro o que torna o treino muito mais eficiente do que o MCC, pois permite extrair informação que seria desprezada pelo MCC. Como se sabe, também o critério MEE apresenta uma elevada robustez a *outliers*, o único defeito deste critério para o treino é a sua insensibilidade à média do erro. Um forma de resolver o problema consiste em adicionar um *bias* à resposta do algoritmo, de forma a forçar a média do erro no conjunto de treino a ter valor nulo. No entanto, em problemas com uma fdp assimétrica ou com caudas muito "pesadas" a estimação da média do erro é um procedimento problemático. Um método muito mais atractivo do que efectuar um ajuste na distribuição é fixar o pico da fdp do erro na origem, para garantir erro com média nula.

Como o objectivo no treino supervisionado é ter o máximo de erros igual a zero, pode ser construído um critério que minimiza a entropia do erro com respeito a 0. Em (Weifeng et al., 2006) é demonstrado que este método não é mais do que uma combinação pesada dos dois critérios MCC e MEE, onde o termo que corresponde ao MEE minimiza a entropia do erro, enquanto que o termo correspondente ao MCC coloca o pico da fdp na origem. A função custo do novo critério é:

$$\max \quad (1-\gamma)\, IP(e) + \gamma V(e) \tag{4.32}$$

onde γ representa o número de pontos de referência devem ser colocados na origem, quanto maior o valor de γ maior o número de pontos de referência na origem. Quando $\gamma = 0$ o critério reduz-se ao MEE, quando $\gamma = 1$ ao MCC, $0 < \gamma < 1$, é a chamada minimização da entropia do erro com pontos de referência (MEEF - Minimization Error Entropy with Fiducial Points).

Este critério permite automaticamente localizar e fixar o pico da fdp do erro na origem mesmo nos casos em que a distribuição do erro é assimétrica e caudas "pesadas". Isto porque, o critério MCC influencia a localização da fdp, uma vez que tenta colocar todos os erros próximos da origem.

Em (Weifeng et al., 2006) é demonstrado que esta função de custos também induz uma métrica, que apresenta a característica de ser sensível ao sinal do erro. Favorece erros com o mesmo sinal e penaliza erros com sinais diferentes.

4.3.5 Algoritmos de Retropropagação dos Critérios ITL

Critério MEE

Usando uma analogia com a física, as amostras poderão ser vistas como partículas de informação num campo de informação. Essas partículas interagem umas com as outras de acordo com leis de interacção definidas, sendo os parâmetros do sistema adaptativo modificados de acordo com a interacção entre as partículas. No Apêndice B encontra-se uma adaptação da metodologia descritiva da relação mútua entre as amostras presente em (Erdogmus, 2002) e (Principe et al., 1999b).

No caso do critério MEE tem-se o potencial de informação e força de informação (IF) dados por:

$$IP(y) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} IP_{ij} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} K_{2\sigma^2}\left(y_i - y_j\right) \tag{4.33}$$

$$IF(y) = \frac{\partial V(y)}{\partial y_i} = -\frac{1}{N^2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_{2\sigma^2}\left(y_i - y_j\right)\left(y_i - y_j\right) \tag{4.34}$$

Sendo que, a força de informação é a derivada do potencial de informação (em analogia com a física).

Se o objectivo é minimizar a entropia das amostras, as forças são atractivas (com sinal positivo devido à maximização do potencial) e procuram mover as partículas em direcção ao estado de potencial máximo. Por outro lado, se o objectivo é maximizar a entropia, as forças são repulsivas (com sinal negativo), o movimento das partículas é espalharem-se para preencher o espaço.

O mecanismo que faz movimentar as partículas e corresponde à manipulação dos pesos w do sistema adaptativo $g(x,w)$ que transforma as entradas x nos erros e. Por isso, é necessário propagar o efeito das forças de informação, transformando o movimento das partículas no espaço da saída em movimento dos pesos no seu próprio espaço. Isso consegue-se com o cálculo do gradiente do potencial de informação relativamente aos pesos do sistema adaptativo, que é dado por (Erdogmus, 2002):

$$\frac{\partial IP(e)}{\partial w} = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_{2\sigma^2}\left(e_i - e_j\right)\cdot\left(e_i - e_j\right)\left(\frac{\partial y_i}{\partial w} - \frac{\partial y_j}{\partial w}\right) \tag{4.35}$$

O gradiente da equação 4.35 é válido para o caso unidimensional. O gradiente para situações com multiplas saídas encontra-se descrito em (Erdogmus, 2002).

O algoritmo 4 de retropropagação, que se encontra no Apêndice C, descreve uma situação com apenas uma saída da rede neuronal e com a nomenclatura adaptada da secção 4.1.

A equação 4.21 da secção 4.3.2 é utilizada para a estimação *on line* da entropia à medida que novos exemplos de treino estão disponíveis. Calculando a derivada de 4.21 relativamente ao vector dos pesos de um sistema adaptativo obtém-se o seguinte gradiente:

$$\frac{\partial H_{k+1}}{\partial w} = -\frac{\partial IP_{k+1}/\partial w}{IP_{k+1}} = \frac{(1-\lambda)\frac{\partial IP_k}{\partial w} + \frac{\lambda}{M}\sum_{i=k-M+1}^{k} K'_\sigma(e_i - e_{k+1})\left(\frac{\partial y_i}{\partial w} - \frac{\partial y_{k+1}}{\partial w}\right)}{(1-\lambda)IP_k + \frac{\lambda}{M}\sum_{i=k-M+1}^{k} K_\sigma(e_i - e_{k+1})} \tag{4.36}$$

A única alteração para aplicar o modo recursivo ao algoritmo 4 consiste em:

$$\left(w_{ij}\right)_{k+1} = \left(w_{ij}\right)_k + \eta \cdot \left[(1-\lambda)\left(\Delta w_{ij}\right)_k + \lambda\left(\Delta w_{ij}\right)_{k+1}\right] \tag{4.37}$$

Este método pode ser visto como uma alternativa ao clássico RLS (*Recursive Least Square*), descrito em (Haykin, 1995).

Em (Erdogmus, 2002) são efectuadas duas recomendações relativamente ao treino *on line*: 1) a estimação recursiva do gradiente apenas é uma aproximação válida, se a taxa de aprendizagem η for suficientemente pequena para que $\partial IP_k/\partial w_{k+1}$ e $\partial IP_k/\partial w_k$ tenham um valor próximo; 2) o gradiente do instante k deve ser o último gradiente calculado e não o gradiente correspondente ao último vector de pesos.

Critério MCC

A retropropagação da correntropia é em tudo semelhante à utilizada no critério MSE. As diferenças fundamentais encontram-se apenas na derivada da função de custo e no sinal positivo da actualização dos pesos, visto que se pretende maximizar a correntropia. O algoritmo de retropropagação para o caso do critério MCC encontra-se descrito no algoritmo 5 do Apêndice C. O treino *on line* da rede neuronal é realizado com recurso à retropropagação estocástica, que consiste em: à medida que chegam novos exemplos é calculada a correntropia entre a resposta do sistema adaptativo e o valor desejado, sendo depois retropropagada de forma a actualizar os pesos e *bias* da rede neuronal. Este método é uma alternativa ao clássico LMS (*Least-Mean-Square*), descrito em (Haykin, 1995).

Critério MEEF

O critério MEEF consiste na soma pesada dos critérios MEE e MCC, logo também o seu algoritmo de retropropagação consiste na combinação pesada dos dois algoritmos já apresentados. Basicamente, são aplicados paralelamente os algoritmos 4 e 5 do Apêndice C e a actualização dos pesos é realizada com a soma pesada das contribuições dos dois algoritmos, como demonstrado no algoritmo 6 do Apêndice C. O treino *on line* é realizado seguindo a mesma lógica do treino *off line*, a entropia do erro é calculada de forma recursiva e paralelamente é calculada a correntropia do novo exemplo. A alteração dos peso e *bias* é novamente uma soma ponderada das contribuições dos dois critérios.

4.4 Optimização do Algoritmo de Treino

Nesta secção serão estudadas e abordadas diferentes técnicas com o objectivo de atingir uma rápida convergência e um melhor desempenho no treino *off line* e *on line* das rede neuronais. É essencial diminuir o tempo de convergência dos algoritmos de treino devido ao elevado esforço computacional da função de custo MEE. Em problemas com conjuntos de dados de elevada dimensão, o treino de uma rede neuronal com a função de custo MEE poderá demorar um elevado número de horas/dias. O estudo irá incidir sobre a taxa de aprendizagem, tamanho do *Kernel* e a combinação do treino *batch* e sequencial da rede neuronal. As técnicas abordadas nesta secção serão utilizadas e analisadas no capítulo 5.

4.4.1 Treino *Off Line*

Taxa de Aprendizagem Adaptativa

Na extensa bibliografia sobre redes neuronais existem vários trabalhos onde é explorado o uso de uma taxa de aprendizagem variável ao longo do processo de treino. São vários o métodos utilizados para variar a taxa de aprendizagem com o objectivo de atingir uma rápida convergência e um maior desempenho da rede neuronal (Silva & Almeida, 1991) (Fahlman, 1988), sendo que todos eles foram desenvol-

vidos usando como função de custo o MSE.

Em (Morejon & Principe, 2004) são analisados os diferentes algoritmos de treino de uma rede neuronal (*gradient-descent*, *conjugate gradient*, *Levenberg–Marquardt*, etc.), aplicados ao critério MEE. Dos diferentes algoritmos analisados, os únicos que não necessitam de nenhuma alteração na sua metodologia são o método do gradiente sem o termo momento e o RPROP (*Resilient Backpropagation* (Riedmiller, 1994). Nos resultados dos casos de teste apresentados em (Morejon & Principe, 2004) verificou-se um excelente desempenho em termos de convergência do algoritmo RPROP. Quando comparado com os restantes algoritmos é o mais rápido a convergir, perdendo no entanto eficiência na parte final do treino. O algoritmo RPROP consiste numa variante mais sofisticada do algoritmo de retropropagação com o método gradiente. A diferença está na utilização de uma taxa de aprendizagem adaptativa que apenas depende do sinal do gradiente e não do módulo do gradiente. O fraco desempenho do algoritmo na fase final já foi alvo de estudo e foram propostas versões aperfeiçoadas do algoritmo, em (Igel & Husken, 2000) são propostas diferentes versões aperfeiçoadas (iRPROP).

Tamanho das Janelas de Parzen Adaptativo

Em (Erdogmus & Principe, 2002a) concluiu-se experimentalmente que a dimensão da janela de *Parzen* desde que contida no intervalo $[0.01 : 0.1]$ não é crítica para o desempenho do algoritmo. Uma regra empírica utilizada em (Erdogmus, 2002) indica que a largura da janela deve cobrir em média cerca de 10 amostras. Um método alternativo e analisado em (Erdogmus & Principe, 2002b) e (Erdogmus, 2002) consiste em iniciar o processo de treino com uma janela de *Parzen* muito mais larga e progressivamente reduzi-la até convergir para o tamanho definido. Com uma janela larga, muitos óptimos locais são eliminados e o algoritmo progride com facilidade na direcção geral do óptimo a partir de uma inicialização aleatória, podendo depois ser guiado de forma mais fina à medida que se comprime a janela. No fundo, a utilização de uma estratégia deste género permite transformar um problema de optimização local num problema de optimização global.

Em (Erdogmus & Jeong, 2005) é utilizada uma regra simples para a redução

do valor de h[2] ao longo do processo de treino. A regra consiste em iniciar o treino com um valor elevado para h (por exemplo 0.5) sendo ao longo do processo de treino reduzido de acordo com a seguinte regra:

$$h^{(k)} = \begin{cases} h^{(k-1)} \cdot 0.95 & se \quad \left| IP_{norm}^{k-1} - IP_{norm}^{k-2} \right| < 0.002 \wedge IP_{norm}^{k-1} > IP_{norm}^{k-2} \\ h^{(k-1)} & \end{cases} \tag{4.38}$$

onde h possui um limite inferior de 10% do valor inicial. Um aspecto importante na aplicação desta regra é que se deve aplicar antes de se iniciar o treino da rede na época k e como tal devem ser apenas comparados potenciais calculados com o mesmo valor de h.

Uma alternativa seria h ser adaptado em função da variância do erro. A estratégia baseia-se na ideia de que à medida que o processo de treino se aproxima do fim os erros tendem para zero, e então faria sentido diminuir h visto que os erros se encontram mais próximos. Este algoritmo tem o problema de o potencial (e a entropia) e suas derivadas dependerem do valor de h, pequenos valores de h originam baixos valores de potencial (altos valores de entropia). Se o valor de h for reduzido de uma iteração para a outra está-se a criar um conflito com o critério de maximização do potencial. Na proximidade do erro de treino mínimo o algoritmo pode tornar-se muito instável e perder a capacidade de convergência.

Na figura 4.8 encontra-se um resultado para um pequeno conjunto de dados de um parque eólico (Parque A), sobre a instabilidade na convergência provocada por um valor de h em função do desvio padrão do erro. Não se pretende nesta secção estudar qual a melhor forma de usar uma janela de *Parzen* com tamanho adaptativo, nem provar que a abordagem adaptativa é a ideal. Em (Erdogmus & Principe, 2002b) são claras as vantagens do uso de uma janela adaptativa quando comparados os resultados obtidos com uma janela de tamanho fixo.

A adaptação de h para o critério MCC pode ser realizada de uma forma simples. O processo de treino começa inicialmente com o critério MSE e em certo momento do treino o critério é alterado para correntropia com um h de tamanho reduzido (Liu et al., 2007a).

[2] usou-se h em vez de σ para evitar confusão com o desvio padrão

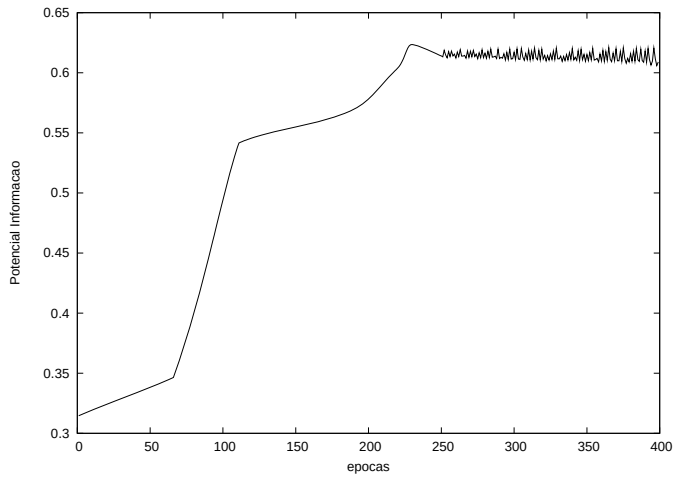

Figura 4.8: *Instabilidade no processo de treino provocada por um valor de h em função do desvio padrão.*

Treino Batch-Sequencial

O treino com os critérios MEE e MEEF é muito exigente computacionalmente para grandes conjuntos de dados, como é caso da previsão eólica. Devido à estimação da fdp do erro ser realizada com recurso a janelas de *Parzen* são necessários todas as amostras de erro para estimar o valor da entropia. Esta condição obriga ao uso do treino *batch*, impedindo o uso do treino sequencial. Em (Bishop, 1995) era aconselhado e em (Santos et al., 2005) é proposto o algoritmo Batch-Sequencial para ultrapassar as limitações do treino *batch* no treino de redes neuronais com o critério MEE. O algoritmo consiste numa combinação dos dois métodos.

Este método possuiu as vantagens de ambos, continua a ser possível estimar a entropia a partir dos dados e torna menos provável o algoritmo de treino ficar preso num óptimo local. O algoritmo consiste em dividir o conjunto de treino em vários subconjuntos, que são apresentados de forma sequencial ao algoritmo de treino da rede neuronal, a cada subconjunto é aplicado o treino *batch*. Sendo T o conjunto de treino e Ts os subconjuntos de igual tamanho obtidos através de uma amostragem aleatória sem reposição, tem-se então:

$$card(T) = n + \sum_{j=1}^{L} card\left(Ts_j\right) \tag{4.39}$$

onde L é o número de subconjuntos, n o número de exemplos não incluídos e $card$

a cardinalidade do conjunto. A divisão em subconjuntos é realizada em cada época, sendo deixados de fora n exemplos, o que não é crítico visto que na próxima época poderão ser novamente incluídos. O método encontra-se descrito no algoritmo 3 do Apêndice A.

A complexidade original do treino com o MEE era $O\left(card\left(Ts^2\right)\right)$, com o batch-sequencial passa-se a ter uma complexidade de:

$$O\left(L\left(\frac{card(T_s)}{L}\right)^2\right) \tag{4.40}$$

A redução é proporcional ao tamanho de L. O número de exemplos nos subconjuntos é determinado pelas características dos dados, mas nunca deverá ser muito reduzido.

4.4.2 Treino *On Line*

Taxa de Aprendizagem Adaptativa

Na secção 4.1.2 foi apresentada a definição de treino *on line* da rede neuronal. No entanto, por vezes existe na literatura uma incorrecta utilização do termo *on line* quando se trata de treino de redes neuronais. Não é correctamente distinguido o treino incremental do treino *on line*, e nem sempre uma taxa adaptativa para o treino incremental pode ser utilizada da mesma forma no treino *on line*. Um exemplo é o método apresentado em (Almeida, 1997), que segundo o autor pode também ser utilizado no modo *on line*. Não é no entanto possível adaptar o método para um algoritmo que funcione sem épocas de treino, sendo que no modo *on line* não existe o conceito ”época”, cada exemplo é tratado individualmente à medida que ”chega”.

Um método desenvolvido especificamente para este modo de treino é o algoritmo descrito em (Almeida et al., 1998). O método baseia-se no clássico algoritmo estocástico de retroprogapação, diferindo apenas no facto de cada sinapse possuir a sua própria taxa de aprendizagem, e essa mesma taxa evoluir ao longo do processo de treino. A taxa de aprendizagem é actualizada segundo uma expressão determinada

em (Almeida et al., 1998) e que consiste em:

$$\eta_i^n = \eta_i^{n-1}\left[1+\kappa\frac{d_i^{n-1}\cdot d_i^n}{v_i^n}\right] \tag{4.41}$$

onde $d_i^n = \left(\frac{\partial E}{\partial w}\right)_i^n$, v_i^n é a exponencial média do quadrado de d_i^n obtido com:

$$v_i^n = \tau v_i^{n-1} + (1-\tau)\left[d_i^n\right]^2 \tag{4.42}$$

Este método possuiu dois parâmetros, os valores recomendados para cada parâmetro são os seguintes: $\kappa = 0.01$; $\tau = 0.9$. O limite inferior para η é 10^{-10}.

Termo de Momento

No *on line* é importante avaliar a utilidade de possuir um termo momento também ele adaptativo. Em (Almeida et al., 1998) experiências realizadas sugerem a não utilização de um termo adaptativo. É recomendado um termo de momento fixo de 0.9 para ser utilizado no algoritmo com taxa de aprendizagem adaptativa, conforme descrito na secção anterior.

Neste contexto surge a necessidade de avaliar qual o impacto de um termo momento, sem descurar o fim a que se destina e qual a sua influência na actualização dos pesos. Em (Silva, 1998) é feita uma analogia do termo momento com um filtro passa baixo (para ω<1), contribuindo para uma atenuação de componentes oscilatórias do vector gradiente. Esta analogia indica que a inclusão de um termo momento provoca uma amplificação da taxa de aprendizagem que pode atingir $1/(1-\omega)$. Este efeito de aumento da taxa de aprendizagem não é desejado quando se está a realizar um treino *on line*, onde se poderá pretender ter uma taxa de aprendizagem de valor reduzido.

Uma outra razão para a necessidade de verificar experimentalmente a utilidade do termo momento no treino *on line* é o facto de a superfície de erro apenas depender do valor das variáveis de entrada do modelo. O valor do mínimo da superfície de erro e a sua localização dependem do valor das variáveis de entrada do modelo e do valor desejado da variável a prever (Principe et al., 1999a). Será então expectável que estas três características da superfície de erro num problema com fluxo contínuo de dados e não-estacionário estejam em constante alteração. Tal como também já

foi analisado em (Almeida et al., 1998), um valor elevado do termo momento resulta num termo de memória excessivamente longo, o que pode dificultar a adaptação do processo devido às variações da superfície de erro.

Tamanho das Janelas de Parzen Adaptativo

O uso de um tamanho das janelas de *Parzen* adaptativo aplicado ao critério MEE, ainda não tem sido alvo de qualquer estudo na bibliografia existente sobre o assunto. Recentemente em (Heinz, 2007) é proposto um algoritmo para estimação de densidade com recurso a *Kernels* Gaussianos em tempo real. Também no treino *on line* o tamanho do *Kernel* poderá ser adaptativo (sendo alterado à medida que chegam novos exemplos). Na secção 4.3.2 a fórmula recursiva apresentada consiste igualmente em recursivamente actualizar a fdp, ou seja, é numa estimação recursiva de densidade de probabilidade. Portanto, uma técnica empírica semelhante à utilizada em (Heinz, 2007) poderá também ser testada neste tipo de problemas.

De acordo com a regra empírica de (Silverman, 1986) tem-se:

$$h^k = 1.06 \cdot \sigma^k \cdot k^{1/5} \tag{4.43}$$

Sendo que o valor de k deverá ser fixo e de valor igual ao M da equação 4.21. Como se trata de uma estimativa recursiva do tamanho do *Kernel*, também o desvio padrão tem de ser calculado de forma recursiva. Uma equação muito conhecida de estimação recursiva da variância consiste em:

$$Var^k = \frac{k-1}{k} Var^{k-1} + \frac{1}{k-1}\left(x_k - \mu_k\right)^2 \tag{4.44}$$

com μ igual à média. Uma aproximação válida, para o problema em estudo, consiste em considerar que a média do erro é nula e o termo de memória igual ao utilizado na estimação recursiva do potencial. A variância passa a ser estimada com:

$$Var^k = (1-\lambda) Var^{k-1} + \lambda \cdot x_k^2 \tag{4.45}$$

Cada vez que chega um novo exemplo, a variância é recursivamente estimada, sendo depois o tamanho do *Kernel* actualizado com equação 4.43.

Capítulo 5

Teste e Avaliação dos Modelos de Previsão Eólica

Neste capítulo serão comparadas as previsões obtidas com a aplicação dos critérios ITL e erro médio quadrático. A vantagem do uso de uma metodologia de treino *on line* de redes neuronais será também avaliada. Os dados para treino e teste dos modelos são referentes a três parques eólicos com diferentes complexidades de terreno situados em Portugal Continental. Os erros que resultam dos diferentes critérios e modos de treino serão analisados de diferentes perspectivas, procurando incorporar a utilidade da previsão para efeitos de mercado de electricidade.

Os testes realizados dividem-se em duas secções:

- Comparação dos diferentes critérios ITL e o critério MSE no modo de treino *off line* para previsões meso-escala com um horizonte temporal de 24 horas. O principal objectivo é comparar os critérios de forma similar à existente na bibliografia sobre o critério ITL.

- Previsão de produção eólica para o horizonte completo de 72 horas com recurso a previsões MM5 para 72 horas. O objectivo consiste em avaliar as vantagens da utilização do treino *on line* sobre o *off line* e dos critérios ITL sobre o MSE.

5.1 Descrição dos Parques Eólicos

O Parque Eólico A situa-se no interior Centro-Norte e tem uma potência instalada de 21.6 MW repartida por 12 máquinas. O terreno tem características complexas, com zonas de declive acentuado a norte e a sul do parque e vários vales encaixados de pequena dimensão.

O Parque Eólico B situado na Região Centro Sul do país é constituído por 20 máquinas, num total de 16.2 MW de potência instalada. O terreno apresenta uma elevada complexidade topográfica.

O Parque Eólico C situa-se na Região Centro Oeste a cerca de 25 km da linha da costa, e é constituído por 15 máquinas, num total de 34.5 MW de potência instalada. Apresenta uma topografia muito menos elevada e mais suave que os outros dois parques considerados.

Os critérios para a selecção destes três parques foram vários, desde a disponibilidade de informação de exploração à desejável diversidade das soluções tecnológicas. Pesou ainda o facto de se situarem em áreas climatologicamente distintas, com relevo e cobertura do solo também diferenciados, como forma de conferir aos exercícios alguma representatividade relativamente a outras zonas e parques eólicos em Portugal.

O conjunto de dados disponível relativo aos três parques encontra-se descrito na tabela 5.1. Na tabela encontram-se os valores máximo, mínimos e médios de potência produzida medida e velocidade prevista no parque eólico para os anos em estudo.

Nas figuras 5.1 e 5.2 encontra-se o histograma e a aproximação normal do erro de previsão para 24 horas da velocidade do vento do modelo MM5 relativamente aos valores de velocidade registados no mastro de medição para o ano de 2005. As figuras indicam que o erro de previsão do modelo MM5 apresenta uma distribuição muito próxima da Gaussiana.

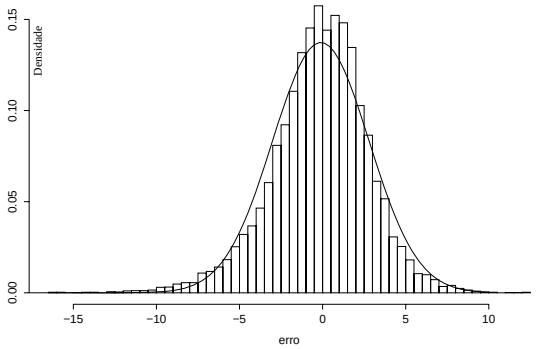

Figura 5.1: *Erro de previsão do MM5 no parque A e aproximação pela distribuição normal (média=-0.107 desv. padrão=2.90).*

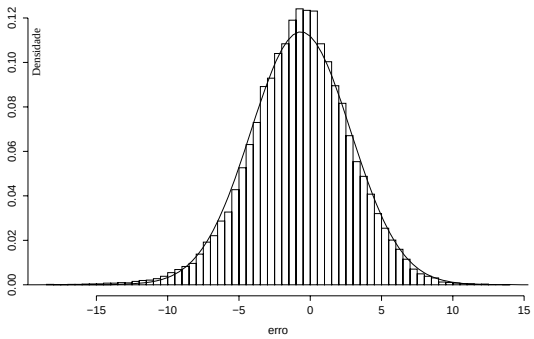

Figura 5.2: *Erro de previsão do MM5 no parque B e aproximação pela distribuição normal (média=-0.66 desv. padrão=3.50).*

5.1.1 Pré-processamento dos Dados

As entradas da rede neuronal são a velocidade e direcção do vento previstas pelo modelo MM5. Não é considerado relevante para os testes realizados no âmbito desta tese uma análise de significância das diferentes variáveis do problema, uma análise de significância das variáveis poderá ser encontrada em (Rosado et al., 2007). Optou-se por utilizar as variáveis mais comuns em problemas de previsão eólica e também as que mais importância física têm no problema de previsão (Joensen et al., 1997). A necessidade de introduzir a velocidade do vento é completamente justificada pela relação cúbica com a potência, e analisando a disposição dos aerogeradores num parque eólico verifica-se também a dependência da direcção do vento. A velocidade de vento nos aerogeradores depende da disposição dos outros aerogeradores, dos obstáculos do terreno, da vegetação e da topografia.

Os dados utilizados foram devidamente pré-processados. Tendo sido retirados erros de medição e preenchidos alguns valores em falta com o recurso à persistência. As previsões do modelo MM5 e os valores medidos pelo SCADA foram alvo de uma sincronização. O processo de sincronismo é necessário para garantir que os conjuntos de entrada estão na mesma base temporal. Os dados do SCADA são registados com uma base temporal de 10 minutos enquanto que os dados do MM5 são numa base temporal de 30 minutos. Para tal, recorre-se a uma interpolação dos dados do SCADA para a base temporal do MM5.

Tabela 5.1: *Conjunto de dados dos três parques eólicos.*

Ano	Pmax	Pmin	Pmed	Vmax	Vmed
			Parque A		
2005	21680.17	-77.67	6113.98	28.7	7.65
2006	21677	-90.33	5676.57	26.9	9.62
			Parque B		
2005	16142.67	-80.17	5206.87	20.7	6.71
			Parque C		
2005	34520	0	7956.68	25.4	7.34
2006	34850	0	8142.99	24.7	7.24

O SCADA utiliza a hora legal e o MM5 a hora universal, foi portanto necessário proceder à devida correcção de data e hora dos registos do SCADA de forma a transformar os dados na mesma referência temporal, a hora universal. Este processo seria simples se os sistemas de SCADA alterassem a data e hora automaticamente, segundo as mudanças de horário Verão/Inverno e Inverno/Verão estipuladas. Esta alteração advém de intervenção humana, pelo que foi necessário proceder a uma análise de correlação dos dados, ao longo do ano, para identificar os instantes de mudança de hora de forma a possibilitar a devida correcção.

Não foram filtrados pontos em que para uma dada velocidade a potência efectivamente medida é muito diferente da potência esperada. Para servirem de entradas para uma rede neuronal, os dados foram normalizados para um intervalo entre -1 e 1 usando o método min-max.

5.2 Teste *Off Line* para um Horizonte Temporal de 24 Horas

5.2.1 Conjunto de Dados

Nesta secção os diferentes critérios ITL e o critério MSE serão comparados num modo de treino *Batch* ou *off line*. Os conjuntos de dados utilizados apenas se referem a previsões MM5 para um horizonte de 24 horas. Estes conjuntos de dados não são constituídos por todos os dados disponíveis, tendo o estudo procurado conjuntos de dados relativamente pequenos para tornar possível realizar diferentes simulações e aplicar todos os critérios. O conjunto de treino foi escolhido de forma a ser suficientemente representativo (vários meses de dados, cerca de 70%) e o conjunto de teste com um tamanho de 30% do conjunto de dados. Nas tabelas 5.2, 5.3 e 5.4 encontram-se estatísticas básicas dos conjuntos de treino, teste e validação referentes aos três parques analisados.

Tabela 5.2: *Conjunto de dados do treino off line do parque A.*

Variável	Treino	Validação	Teste
Média (kW)	7947.06	2567.81	3986.83
Variância (kW)	53959535	12806109	23704424
Assimetria	0.576	2.154	1.527
Mínimo (kW)	-74.33	-68.17	-73.50
Máximo (kW)	21662.70	19177.50	21439.00
Tamanho	5184	1267	2204

Devido ao carácter cíclico da direcção, esta é decomposta em duas componentes usando as seguintes equações

$$\begin{aligned} d_{sen} &= sen\left(\frac{d\cdot\pi}{180}\right) \\ d_{\cos} &= \cos\left(\frac{d\cdot\pi}{180}\right) \end{aligned} \tag{5.1}$$

Tabela 5.3: *Conjunto de dados do treino off line do parque B.*

Variável	Treino	Validação	Teste
Média (kW)	6184.00	4791.77	4266.19
Variância (kW)	27276937	29329712	18641596
Assimetria	0.459	0.77	1.14
Mínimo (kW)	-60.50	-80.17	-67.50
Máximo (kW)	16142.67	16105.67	16105.17
Tamanho	4893	815	2446

Tabela 5.4: *Conjunto de dados do treino off line do parque C.*

Variável	Treino	Validação	Teste
Média (kW)	7632.08	7418.37	8994.55
Variância (kW)	44600767	42377876	73471458
Assimetria	0.88	1.05	0.97
Mínimo (kW)	0.0	0.0	0.0
Máximo (kW) 31920	30920	33950	
Tamanho	5195	1192	2450

5.2.2 Metodologia de Teste e Avaliação

A avaliação do desempenho das funções de custo do critério ITL deve ser realizada de acordo com a bibliografia sobre o tema. Na bibliografia apresentada ao longo do capitulo 4, a comparação entre as função de custo ITL e o clássico MSE é realizada com recurso a funções de distribuição de probabilidade do erro estimadas com recurso a janelas de *Parzen*. Este método de avaliação baseia-se numa inspecção visual das distribuições de probabilidade do erro no conjunto teste.

É relevante avaliar o erro segundo os diferentes critérios considerados no treino da rede. O erro pode ser avaliado segundos dois critérios: entropia e erro médio quadrático. Não foi considerado o uso da correntropia para avaliação dos resultados porque apenas consegue aferir a semelhança entre duas variáveis dentro do espaço

controlado pelo tamanho da janela de *Parzen*.

Os critérios de avaliação da qualidade das previsões e do treino de modelos adaptativos apresentam um carácter subjectivo. Um critério que minimiza o quadrado dos desvios atribuiu uma maior importância aos erros de maior amplitude. Por outro lado, a entropia procura tornar os pequenos erros todos iguais e próximos de zero, mas não trata os erros de maior amplitude da forma semelhante ao erro quadrático médio. Se o utilizador final da previsão estiver mais interessado em não ter erros de grande amplitude, então provavelmente o critério mais útil corresponde à função de custo que penaliza os erros de grande amplitude (como é o caso dos mínimos quadrados). A escolha da melhor previsão é no fundo um problema multicritério.

O algoritmo de retropropagação é um método de pesquisa local, portanto para evitar que o treino da rede fique "preso" num mínimo local, deverão ser realizadas várias simulações com diferentes pesos iniciais. Em cada teste serão realizadas 25 simulações com pesos iniciais diferentes, sendo este número suficiente de acordo com (Iyer & Rhinehart, 1999). A inicialização dos pesos é realizada com recurso a uma distribuição uniforme entre -1 e 1. O resultado final consiste na média de todas as simulações.

5.2.3 Características da Rede Neuronal

A rede neuronal utilizada é composta por apenas uma camada interna com 7 neurónios. Não se encontra no âmbito desta tese procurar a estrutura óptima da rede neuronal. No entanto, numa fase de estudos preliminares verificou-se que a variação da estrutura da rede não apresenta um impacto significativo no erro de previsão, quer em treino *off line*, quer em treino *on line*; os resultados preliminares encontram-se no Apêndice D.

Cada neurónio tem como função de activação a tangente hiperbólica. O algoritmo de retropropagação da rede neuronal é o iRPROP, com os parâmetros iguais aos recomendados na literatura existente sobre o método, e que de forma empírica permitem obter os melhores resultados independentemente do problema em estudo (Riedmiller, 1994) e (Igel & Husken, 2000). Experimentalmente foram tes-

tados diferentes valores para os parâmetros, sem ter sido encontrada uma melhor solução do que a recomendada. Os parâmetros utilizados foram:

- factor de incremento e decremento: $\eta_+ = 1.2$ e $\eta_- = 0.5$;
- limites máximo e mínimo de actualização dos pesos: $\Delta_{min} = 0$ e $\Delta_{max} = 50$;
- valor inicial de actualização dos pesos: $\Delta_0 = 0.0125$.

Relativamente ao tamanho do *Kernel* foi utilizada a estratégia adaptativa para o critério MEE, descrita pela relação 4.38 do capítulo 4. A estratégia para o critério MCC consiste no principio de nas primeiras épocas treinar a rede neuronal com o critério MSE (10 iterações) e nas restantes épocas treinar com o MCC com um valor do *Kernel* reduzido. O valor inicial do *Kernel* para o critério MEE é $\sigma^2 = 0.3$ e o utilizado no critério MCC foi $\sigma^2 = 0.02$. Os mesmos valores e estratégias adaptativas foram utilizadas no critério MEEF. O valor da constante que pesa os dois critérios no MEEF é 0.3.

O modo de treino utilizado é o *Batch*-sequencial com subconjuntos de tamanho igual a 800 pontos. O critério de paragem consiste na paragem antecipada, o treino termina ao fim de 15 épocas em que o erro no conjunto de validação aumenta sucessivamente ou ao fim de um número máximo de 150 épocas.

5.2.4 Análise dos Resultados

As figuras 5.3 – 5.8 comparam as fdp dos erros de previsão obtidos com os critérios ITL (MEE, MCC e MEEF) e o critério MSE para os três parques eólicos.

Da análise visual das figuras 5.3 e 5.7 verifica-se que a fdp encontra-se mais centrada em zero, quando a rede neuronal é treinada com pontos de referência (critério MEEF). O uso de pontos de referência em vez do simples cálculo da média do erro sobre o conjunto de treino é para este problema uma solução mais robusta, visto estar-se na presença de distribuições do erro com elevada assimetria e *outliers*. Na tabela 5.5 encontra-se a comparação do *bias* do erro (em % da potência instalada) cálculado para os dois critérios. A tabela 5.5 permite confirmar que os resultados do critério MEEF encontram-se mais centrados em zero, facto esse que já era visível

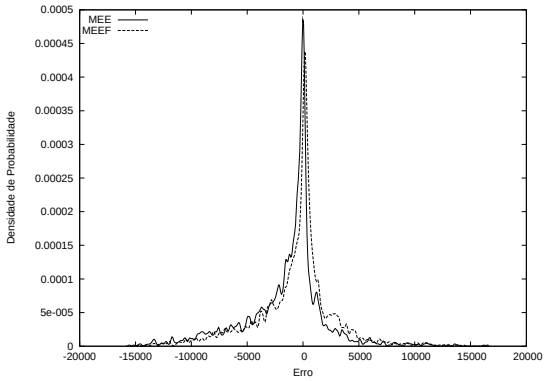

Figura 5.3: *fdp do erro dos critérios MEE e MEEF para o Parque A.*

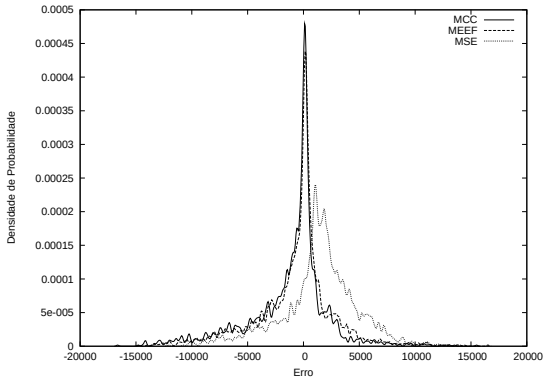

Figura 5.4: *fdp do erro dos critérios MEEF,MCC e MSE no Parque A.*

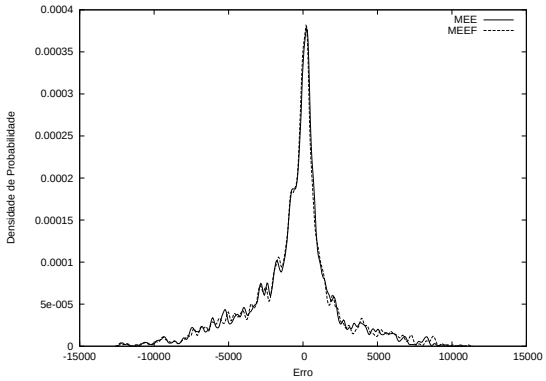

Figura 5.5: *fdp do erro dos critérios MEE e MEEF para o Parque B.*

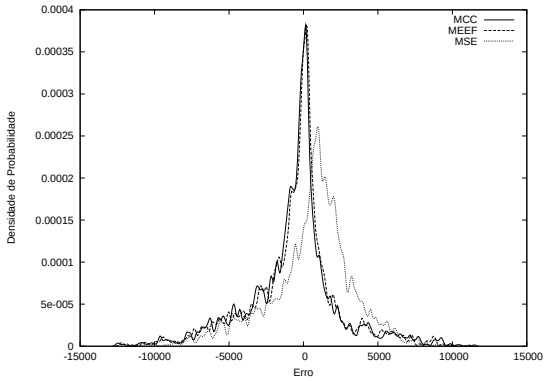

Figura 5.6: *fdp do erro dos critérios MEEF,MCC e MSE no Parque B.*

na inspecção visual da fdp. O valor positivo do *bias* do Parque C indica uma maior frequência de previsões de valor acima do valor medido pelo SCADA.

As figuras 5.4, 5.6 e 5.8, comparam as fdp obtidas com os critérios MEEF, MCC e MSE para os três parques eólicos. A análise visual da forma da fdp dos erros obtidos com o critério MSE permite verificar que o erro de previsão apresenta características não-Gaussianas. No caso dos critérios ITL é possível obter uma fdp com um pico de elevada frequência de erros centrado na origem (o que origina entropia mínima), ao contrário do que é obtido com a minimização da variância do erro. Se o erro fosse de facto Gaussiano seria esperado um desempenho igual entre os critérios que derivam do ITL e o MSE, o que não é o caso.

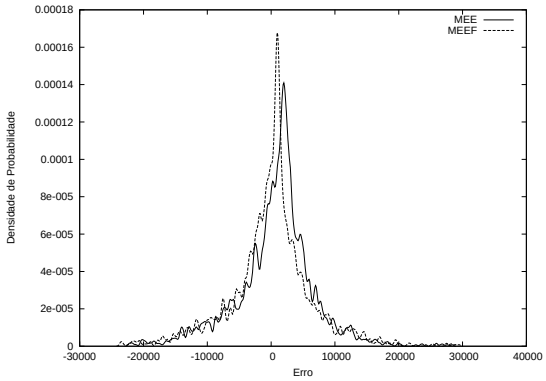

Figura 5.7: *fdp do erro dos critérios MEE e MEEF para o Parque C.*

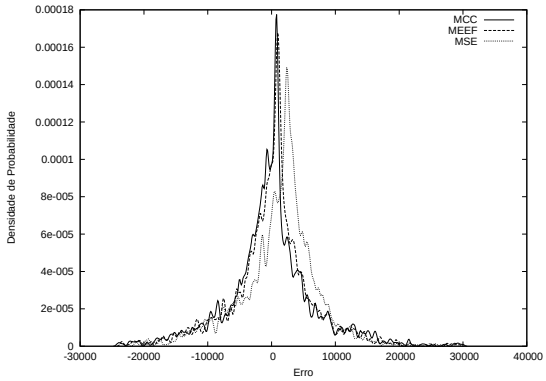

Figura 5.8: *fdp do erro dos critérios MEEF,MCC e MSE no Parque C.*

Tabela 5.5: *Bias relativo à potência instalada do erro nos três parques eólicos.*

Parque	MEEF	MEE
A	-0.030	-0.066
B	-0.036	-0.038
C	0.0017	0.018

Nos três parques verificou-se a preferência dos critérios de entropia por um elevado e mais concentrado número de erros próximos da origem e um pequeno número de picos em zonas de erro elevado. A forma da distribuição do erro de previsão que resulta dos critérios ITL advém do objectivo de obter uma distribuição de erros do tipo função (ou impulso) de *Dirac* para que todos os erros sejam iguais. O critério MSE pelo contrário prefere erros largamente distribuídos num pequeno intervalo, isto porque assume que a distribuição dos erros é Gaussiana.

Uma comparação visual dos resultados com os critérios MCC e MEEF, para os três parques, indica que o critério MCC apresenta uma maior probabilidade junto a zero, no entanto, tem erros mais elevados na parte negativa da distribuição. A existência de um maior número de erros de amplitude elevada no critério MCC justifica-se pela insensibilidade a *outliers* controlada pelo tamanho do *Kernel* que o critério apresenta.

O elevado esforço computacional requerido na estimação da entropia representa um grande obstáculo na aplicação dos critérios MEE e MEEF a este problema. As técnicas de optimização do treino da rede neuronal, descritas na secção 4.4 do capítulo 4, permitiram reduzir consideravelmente o esforço computacional com o aumento da capacidade de convergência da rede neuronal (redução do número de épocas para convergência) e redução do conjunto de dados para estimação da entropia. O tempo de treino do conjunto de dados da tabela 5.2 com o clássico algoritmo de retropropagação era de aproximadamente 3 horas[1] para apenas uma simulação. Com a optimização do treino, o tempo de processamento foi reduzido para 10 minutos.

Na tabela 5.6 encontra-se a comparação dos resultados obtidos com os diferentes critérios de treino sob diferentes critérios de avaliação.

Como era expectável o critério MEE produz a solução com menor entropia do erro, mas é interessante constatar que também os critérios MCC e MEEF minimizam a entropia do erro com um excelente desempenho.

Ao contrário do esperado, nem sempre o critério MSE obtém uma solução com um erro quadrático mínimo no conjunto de teste, seria esperado que uma rede treinada com um critério que pretende minimizar o quadrado dos desvios tivesse um desempenho superior quando avaliada com este critério. Tanto o MEE como o MEEF apresentam resultados da mesma ordem de grandeza dos obtidos com o MSE para os três parques para o critério EMQ, dando uma indicação sobre a qualidade dos resultados obtidos com os critérios ITL. O critério MCC apresenta o desempenho menos satisfatório no EMQ devido à sua insensibilidade a *outliers*, não sendo os erros de grande amplitude tratados com a mesma importância que os restantes erros.

Da análise da variância total[2] do erro de 25 simulações verifica-se que o critério MSE apresenta o menor valor em dois parques, sendo que no terceiro parque a variância também tem um valor reduzido. O critério que maior variância apresenta é o MCC.

---

[1] com um Pentium IV 3.0 GHz e 1 GB de RAM

[2] admitindo que os erros são independentes é possível somar as variâncias

Tabela 5.6: *Avaliação dos erros de previsão sob diferentes critérios: H_{R2} entropia do erro; EMQ – erro médio quadrático (% potência instalada); variância das simulações.*

Critério de Treino	H_{R2}	EMQ	Variância
	Parque A		
MSE	3.931	639.18	**35255.46**
MCC	3.751	661.02	210053.94
MEEF	3.767	**612.52**	94209.57
MEE	**3.747**	714.82	41580.74
	Parque B		
MSE	3.873	**511.84**	8045.45
MCC	3.783	611.86	65026.13
MEEF	3.789	579.33	**7782.63**
MEE	**3.779**	561.36	8119.16
	Parque C		
MSE	4.141	1097.20	**33229.75**
MCC	4.121	1340.68	64709.18
MEEF	4.123	1188.52	36013.77
MEE	**4.118**	**1087.54**	92694.61

5.3 Previsão Eólica para um Horizonte Temporal de 72 Horas

5.3.1 Conjunto de Dados

As principais características do conjunto de dados utilizados na previsão para um horizonte temporal completo de 72 horas já foram descritas na tabela 5.1 da secção 5.1. Na tabela 5.7 encontram-se os meses que compõem os conjuntos de validação, teste e treino dos três parques em estudo.

O segundo conjunto de dados do Parque C (chamado C.1) tem como principal finalidade representar um conjunto de dados com uma mudança de conceito. O

Tabela 5.7: *Meses dos conjuntos de validação, teste e treino dos três parques eólicos.*

Parque	Validação	Treino	Teste	nº previsões
A	Jan (2005)	Fev-Dez (2005)	Jan-Dez (2006)	363
B	Jan (2005)	Fev-Ago (2005)	Set-Dez (2005)	111
C	Fev (2005)	Mar-Dez (2005)	Mar-Dez (2006)	225
C.1	Fev (2005)	Mar-Jun (2005)	Jul-Dez (2005)	184

parque eólico C no período de tempo do conjunto de treino C.1 possui uma potência instalada de 29.9 MW, sendo mais tarde adicionados mais dois aerogeradores, perfazendo uma potência instalada de 34.5 MW.

As entradas da rede neuronal são a velocidade e direcção do vento previstas pelo modelo MM5 para um horizonte temporal de 72 horas, com uma frequência de actualização de 24 horas (entre as 6:00 e 7:00 da manhã de cada dia). A actualização das previsões do MM5 ocorre em cada 24 horas, estão portanto disponiveis três previsões para cada dia:

- $v_{D|t}$: previsão da velocidade (ou direcção) para cada 30 minutos do dia D (D é um dia genérico) realizadas no instante t, sendo t o instante em que as previsões do MM5 são actualizadas. Estas previsões são as mais actuais para o dia D;

- $v_{D|t-24h}$: previsão da velocidade (ou direcção) para cada 30 minutos do dia D realizadas no instante $t-24h$. Previsões realizadas 24 horas antes para o dia D;

- $v_{D|t-48h}$: previsão da velocidade (ou direcção) para cada 30 minutos do dia D realizadas no instante $t-48h$. Previsões realizadas 48 horas antes para o dia D.

A existência de três previsões para o mesmo dia obriga à inclusão de uma variável auxiliar que represente a "antiguidade" das previsão do modelo MM5, variável essa que será chamada de m. Também esta variável deverá ser decomposta em duas

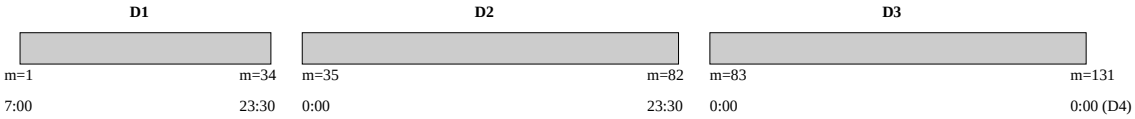

Figura 5.9: *Esquema temporal das previsões meteorológicas do modelo MM5.*

componentes tal como a direcção do vento, as componentes são:

$$\begin{aligned} m_{sen} &= sen\left(\frac{m \cdot 2\pi}{131}\right) \\ m_{cos} &= cos\left(\frac{m \cdot 2\pi}{131}\right) \end{aligned} \tag{5.2}$$

Note-se que o valor de m para as previsão mais recentes apenas varia entre 1 e 34, isto porque a primeira previsão realizada pelo modelo MM5 é referente às 7:00 do dia D. Na prática a previsão não é realizada para 72 horas completas, isto porque não é feita nenhuma previsão para o horizonte entre as 0h e as 7h da manhã do primeiro dia. No entanto, usa-se o termo 72 horas porque o modelo NWP Global começa efectivamente a prever às 0h do primeiro dia, estando depois as previsões meso-escala apenas disponíveis entre as 6h e 7h.

Na figura 5.9 encontra-se um esquema temporal das previsões do MM5 realizadas às 6:00 da manhã de cada dia.

5.3.2 Metodologia de Teste e Avaliação

Nesta secção serão comparados dois critérios ITL (MCC e MEEF) com o MSE num modo de treino *off line*. A principal diferença em relação à secção anterior consiste em terem sido realizadas previsões para um horizonte completo de previsão de 72 horas, tornando possível a análise dos erros de previsões em função do horizonte de previsão. Esta metodologia também tem como objectivo simular o funcionamento real de um sistema de previsão eólica.

A estrutura da rede neuronal utilizada neste teste é constituída por 5 entradas e uma camada interior com 9 neurónios. Os respectivos parâmetros para o treino *off line* são idênticos aos utilizados e descritos na secção 5.2.3. O tamanho das janelas de *Parzen* para o teste *off line* é também o mesmo utilizado anteriormente para os critérios MEEF e MCC. Utilizou-se o mesmo tamanho das janelas de *Parzen* nos três

parques, indo de encontro ao objectivo de possuir uma rede neuronal de parâmetros normalizados para qualquer parque eólico. O tamanho dos subconjuntos do treino *Batch*-sequencial é 2000 exemplos.

O treino das redes neuronais utilizadas no treino *on line* e *off line* é exactamente igual para os dois casos. Os dois métodos diferem apenas no teste, enquanto que para a rede treinada *off line* assume-se que o modelo permanece estático, na rede *on line* nova informação proveniente das leituras do SCADA é retropropagada pela rede. É portanto realizado um treino *off line* com o algoritmo *Batch*-sequencial para as duas redes, sendo que no treino *on line* os pesos e "bias" são apenas os pesos iniciais da rede e são actualizados com nova informação ao longo do processo de teste.

A metodologia de avaliação e teste das previsões para o horizonte completo de 72 horas é a que se encontra descrita no protocolo de avaliação e teste estabelecido no âmbito do projecto europeu ANEMOS (Madsen et al., 2004). As métricas consideradas para a avaliação dos diferentes resultados encontram-se descritas no Apêndice F.

5.3.3 Comparação *Off Line* de Critérios de Treino

Análise Baseada em Medidas de Erro

Nas figuras 5.10 e 5.11 são apresentados os resultados relativos ao Parque A dos critérios de avaliação $EMAN_k$ (erro médio absoluto normalizado) e $REMQN_k$ (raiz do erro médio quadrático normalizado) respectivamente em função do horizonte de previsão.

Na figuras 5.10 é clara a vantagem do uso dos critérios ITL em vez do MSE, ao longo de todo o horizonte temporal o EMAN obtido com o critério ITL é sempre inferior ao obtido com os critério MSE. Não se verifica uma diferença significativa de erro entre os critérios MCC e MEEF. A comparação com o erro da persistência permite perceber que a persistência apresenta um excelente desempenho nas primeiras 6 horas, isto porque a rede neuronal não tem como entradas os valores passados de produção do parque. No entanto, para horizontes superiores a 6 horas,

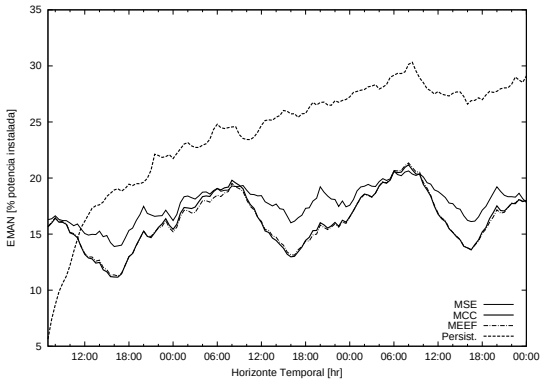

Figura 5.10: *EMAN para o Parque A no modo de treino off line.*

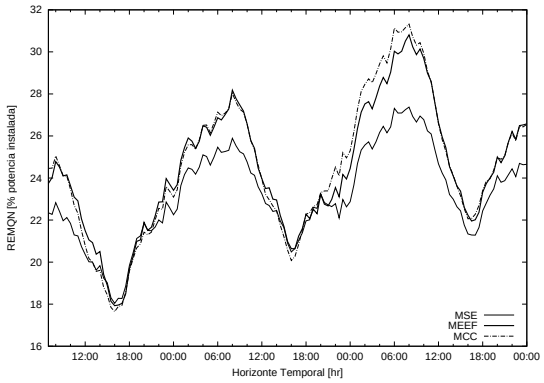

Figura 5.11: *REMQN para o Parque A no modo de treino off line.*

o erro da persistência cresce constantemente, permitindo perceber que se trata de um parque com uma uma complexidade considerável na previsão da produção.

Em relação ao critério de avaliação REMQN, dado que o critério MSE minimiza o quadrado dos desvios então é esperado que uma rede neuronal treinada com o MSE apresente numa medida de erro relacionada directamente com esse critério um desempenho superior aos critérios ITL. O critério MCC é o que apresenta pior desempenho no REMQN devido à insensibilidade a erros de grande amplitude.

A grandeza do erro do REMQN permite avaliar se a qualidade dos dados é elevada ou não. Não sendo aconselhável utilizar o REMQN para comparação de modelos se a confiança não for total na qualidade dos dados utilizados para teste. Nos dados utilizados nos testes não foram retiradas situações de manutenção/avaria das máquinas que são uma das principais fontes de erro, apesar de ter sido efectuado um pré-processamento dos dados. Em sistemas de previsão mais elaborados, como por exemplo em (Rodrigues et al., 2007), previsões por aerogerador (em vez de uma abordagem por parque) são realizadas com o auxilio de um plano de disponibilidades da máquina com o intuito de corrigir este tipo de erros.

Na figura 5.12 encontra-se o resultado do critério de avaliação $NBIAS_k$ (*bias* normalizado), relativo ao Parque A e em função do horizonte de previsão.

Os resultados do NBIAS dão indicações sobre se o método tende a sobre ou subestimar a potência produzida pelo parque. O seu valor calculado sobre todo um

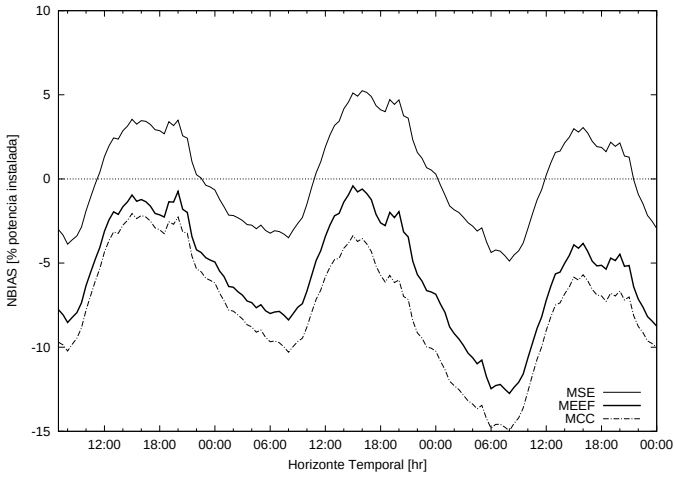

Figura 5.12: *NBIAS para o Parque A no modo de treino off line.*

conjunto de teste (como foi realizado na secção 5.2.3) apenas dá uma indicação geral sobre a tendência da previsão. O BIAS calculado para vários subconjuntos de dados, com diferentes condições meteorológicas ou níveis de produção, permite detectar condições em que o método produz subestimativas significativas da potência produzida.

Nas medidas EMAN, REMQN e NBIAS é visível um padrão ao longo do horizonte de previsão. Às 6h da manhã o erro decresce durante doze horas, atingindo o seu mínimo às 18h da tarde, começando novamente a aumentar até às 6h do dia seguinte. A variação do NBIAS apresenta uma tendência contrária, como se pode constatar pela figura 5.12, o menor valor do NBIAS ocorre durante a noite, existindo uma variação de cerca de 8% entre o dia e a noite. O padrão corresponde a um ciclo diurno, com um mínimo próximo das 18h e um máximo próximo das 6h, sendo que durante o dia o erro diminuiu, aumentado depois durante a noite.

Uma explicação física das variações diurnas não se encontra no âmbito desta tese, uma explicação detalhada do fenómeno atmosférico poderá ser encontrada em (Lange, 2004). É no entanto importante reter que em Portugal Continental o vento apresenta velocidades mais elevadas durante a noite, sendo que nesta situação a amplitude dos erros é maior. Durante o dia as velocidades do vento são menores, e portanto a amplitude do erro será mais reduzida.

Ainda relativamente ao NBIAS, é de notar que enquanto que o MSE varia entre

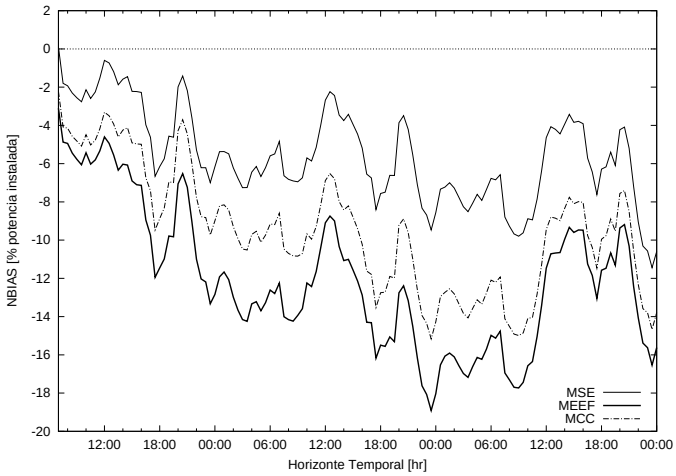

Figura 5.13: *NBIAS para o Parque B no modo de treino off line.*

valores positivos e negativos, os critérios ITL apenas apresentam valores negativos traduzindo-se numa subestimação da potência produzida. No critério MSE durante a noite a potência é subestimada e durante o dia a potência é sobrestimada, nos critérios ITL tal variação já não se verifica.

Os resultados dos critérios de avaliação $EMAN_k$ e $REMQN_k$ relativos ao Parque B encontram-se nas figuras G.1 e G.2 do Apêndice G. A diferença entre os critérios ITL e MSE não é tão evidente no EMAN como no caso do Parque A. No entanto, para o critério REMQN as conclusões retiradas para o Parque B são iguais às já verificadas para o Parque A.

Na figura 5.13 é apresentado o resultado do critério de avaliação $NBIAS_k$ para o parque B.

Ao contrário do observado no Parque A, no Parque B as variações diurnas não estão tão bem definidas no NBIAS. Este fenómeno deve-se ao facto de em latitudes como Portugal continental a variação diurna ser mais significativa durante o Verão. Durante o Inverno é usual não existir (ou existir apenas uma) variação diurna durante a semana. Os meses para os quais foram realizadas as previsões no parque B são apenas meses de Outono, pelo que não são muito frequentes ciclos diurnos de vento.

Um outro ponto de dissemelhança com o Parque A são os valores negativos do NBIAS para o critério MSE, sendo que os valores negativos dos critérios ITL já ocor-

riam no Parque A. Da literatura sobre previsão eólica sabe-se que o NBIAS varia de tipo de terreno para tipo de terreno (Lange, 2004) (Pinson, 2006), em terrenos muito complexos o NBIAS apresenta geralmente valores negativos , e em terrenos de menor complexidade apresenta valores positivos. A variação do NBIAS com a complexidade do terreno está directamente relacionada com o facto de o MM5 ser um modelo meso-escala com uma determinada malha que apenas modela a orografia e rugosidade que se encontra no interior da malha, sem ter em conta as características locais de todo o parque.

Devido à elevada complexidade do terreno do Parque B são esperados valores negativos do NBIAS. O mesmo não se deveria verificar nos resultados do NBIAS para o Parque A com os critérios ITL. O valor negativo do NBIAS no parque A trata-se de um novo resultado tendo em conta a literatura sobre previsão eólica, principalmente porque as analises dos erros de previsão realizadas provêm de sistemas de previsão treinados com o critério MSE.

Os resultados dos critérios de avaliação $EMAN_k$ e $REMQN_k$ relativos ao Parque C encontram-se nas figuras G.3 e G.4 do Apêndice G. Os resultados obtidos e conclusões para o Parque C são em tudo semelhantes aos do Parque A. Salientando-se apenas que o Parque C, aparentemente, é o que apresenta menor ”dificuldade” de previsão da produção, porque o erro da persistência é o menor dos três parques. O valor do REMQN poderá também indicar que a qualidade dos dados utilizados deste parque será superior à dos restantes parques.

Na figura 5.14 é apresentado o resultado do critério de avaliação $NBIAS_k$ para o Parque C.

Tal como o Parque A, também este parque não está localizado num terreno de complexidade muito elevada, sendo então possível detectar uma variação do NBIAS do critério MSE, entre valores positivos e negativos.

É conhecido de (Palmer, 2000) e (Lange, 2004) que a amplitude do erro de previsão da velocidade do vento aumenta com o horizonte temporal. Nos resultados obtidos para os três parques, em relação à previsão da produção, o horizonte de previsão não apresenta um elevado impacto na amplitude do erro das métricas EMAN e REMQN (apenas um ligeiro aumento do erro).

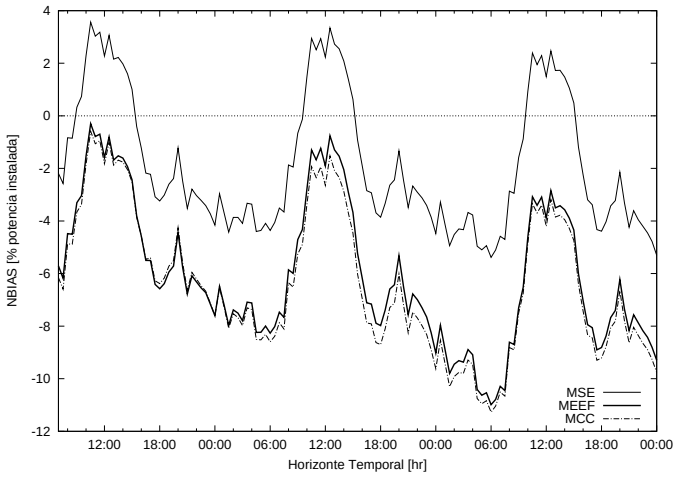

Figura 5.14: *NBIAS para o Parque C no modo de treino off line.*

Verificou-se que o sinal do BIAS para além da dependência da complexidade e do tipo de terreno do parque eólico também depende do critério (ou função de custo) utilizado no treino do sistema de previsão.

Efeito da Curva de Potência

Em 5.15 a potência é dividida em intervalos que correspondem a 10% da potência instalada do Parque A, sendo apresentado um histograma com a frequência da produção medida no parque, e a freqûência dos valores de produção previstos com os diferentes critérios. Não foi efectuada qualquer separação por horizonte de previsão porque, como já foi verificado, não existe uma forte relação na variação do erro com o horizonte de previsão. A figura 5.16 apresenta o valor normalizado do BIAS do erro de previsão em função das classes de potência produzida.

A função densidade de probabilidade do vento num parque eólico indica que é mais frequente a ocorrência de situações de velocidade reduzida, logo, períodos de produção reduzida serão também mais frequentes. Cada parque eólico possuiu o seu próprio regime de vento (e portanto a sua própria fdp do vento) e diferentes tipos de máquinas, o que leva a um histograma da produção único para cada parque. Quando se está a prever a produção para um determinado parque eólico com um método de previsão para um período de tempo longo, é esperado que as previsões si-

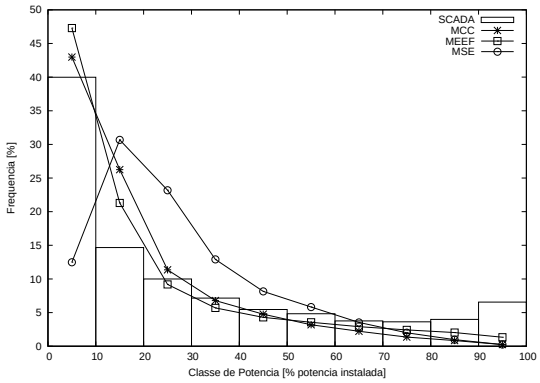

Figura 5.15: *Frequência dos valores medidos de produção do Parque A e dos valores previstos com os diferentes critérios.*

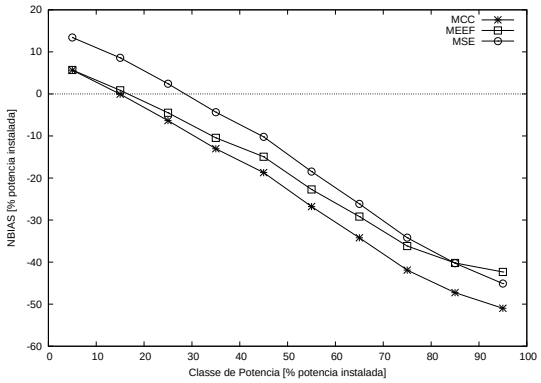

Figura 5.16: *NBIAS do erro de previsão obtido com os diferentes critérios em função da classe de produção do parque eólico A.*

gam o histograma da produção medida no local. Por exemplo, se os valores medidos indicam que a produção eólica é 30% do tempo, entre 0 e 10% da potência instalada, então, o método de previsão deverá realizar previsões naquela classe de potência com uma frequência aproximada.

A barra de produção inferior a 10% da potência instalada da figura 5.15 apresenta uma maior frequência que as restantes: 40 % da produção corresponde a valores menores que 10% da nominal (2160 kW), ou até valores nulos, enquanto que as frequência das restantes encontra-se entre 15% e 5%. As frequências das menores produções variam de parque para parque, mas em todos eles a barra com maior frequência corresponde a baixos valores de produção. Este comentário pode ser generalizado, como se pode verificar nos resultados dos dois parques que se encontram nas figuras G.5 e G.6 do Apêndice G.

Na figura 5.15 é visível que o critério MSE não prevê correctamente as potências de valor mais reduzido (entre 0% e 10% da potência instalada). Pelo contrário, os critérios ITL conseguem prever bem melhor nesta gama de potência. O critério MSE falha nas previsões de potência inferior a 10% da instalada, em quase 30% das vezes. Visto que as potências de menor valor são mais frequentes, também desta análise torna-se claro que o critério MSE tem tendência a sobreestimar a potência produzida quando a produção é reduzida. Nestas situações a potência prevista

encaixa-se numa das outras barras, com produção superior a 10% da potência instalada, o que explica o aumento da frequência das previsões entre 10% e 20% com o critério MSE.

Para potências de valor elevado, entre 70% e 100%, todos os critérios tendem a subestimar a potência produzida, daí a frequência das previsões serem sempre inferiores às frequências dos valores medidos.

As mesmas conclusões são verificadas nos restantes dois parques, cujos resultados se encontram nas figuras G.5 e G.6 do Apêndice G.

Na figura 5.15 verifica-se que as frequências das previsões com os critérios ITL são as que mais se aproximam das frequências dos valores produzidos pelo parque. Este resultado advém da propriedade 4 descrita na secção 4.3.2 do capítulo 4, onde é afirmado que minimizando a entropia de *Renyi* do erro está-se a minimizar a distância entre a fdp dos valores medidos e a fdp dos valores previstos.

Na figura 5.16 o objectivo é estudar a distribuição condicional das previsões dado o valor medido $q(\hat{p}|p)$. Esta distribuição condicional dá informação sobre o que é chamado em (Murphy, 1993) de discriminação. Um exemplo de uma fraca discriminação seria um método de previsão que fornece previsões semelhantes (por exemplo entre 80-85%) quando a potência produzida varia entre 80% e 100%. O BIAS em função da potência medida é uma medida de discriminação.

A primeira conclusão a retirar da figura 5.16 é a variação linear do NBIAS com a potência, o módulo aumenta à medida que a potência aumenta. O valor do NBIAS é positivo para pequenos valores de produção, indicando uma sobrestimação da potência, que é mais acentuada no critério MSE. Para elevados valores de potência o valor do BIAS é negativo atingindo valores muito elevados (cerca de 50%), indicando uma subestimação da potência.

É interessante verificar que todos os critérios, em média, subestimam a potência produzida, uns mais do que outros. Esta subestimação deve-se essencialmente ao modelo de previsão NWP (neste caso o MM5) que tem tendência a subestimar a velocidade do vento, como é demonstrado em (Lange, 2004).

O comportamento para os restantes parques é semelhante, como se pode verificar nas figuras G.7 e G.8 do Apêndice G.

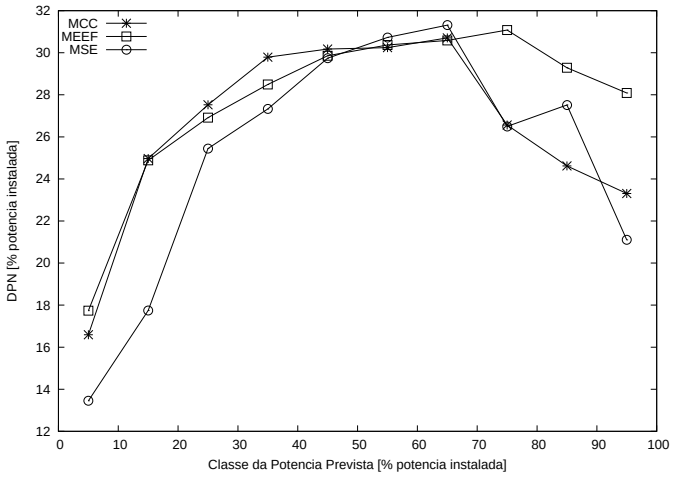

Figura 5.17: *Desvio padrão normalizado da distribuição dos erros de previsão em função das classes de previsão para o Parque A.*

Na figura 5.17 encontra-se o desvio padrão normalizado em função das diferentes classes de potência prevista. A informação proveniente do desvio padrão permite ter informação relativa ao impacto da curva de potência sobre a incerteza da previsão[3], e qual dos critérios de treino provoca uma maior incerteza na previsão. Verificou-se que o nível da potência prevista provoca impacto no desvio padrão do erro. É notória uma ligeira simetria no comportamento das curvas do DPN, em relação a 50% da potência nominal, sendo que nos Parque A e B (figura G.9 do Apêndice G) a simetria encontra-se bem definida. O desvio padrão é elevado na zona intermédia da curva, decrescendo quando se aproxima das zonas "achatadas" da curva de potência, sendo o desvio padrão nas zonas lineares da curva de valor muito próximo. O critério MSE quando comparado com os critério ITL apresenta uma menor "incerteza".

Previsões

Nas figuras 5.18 e 5.19 encontram-se previsões para o Parque A e C realizadas com os diferentes critérios.

[3] a incerteza de um determinado processo é muitas vezes vista como a variabilidade das observações.

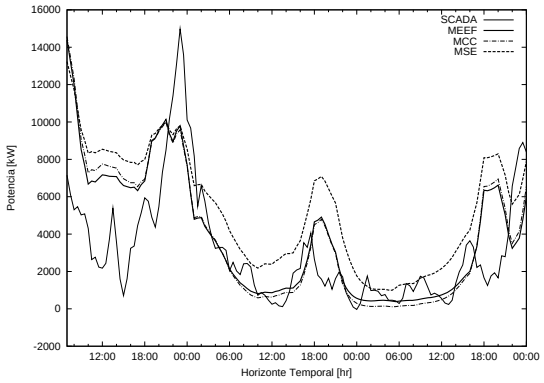

Figura 5.18: *Previsão e valor medido para os dias 26-28 de Outubro de 2006 para o Parque A.*

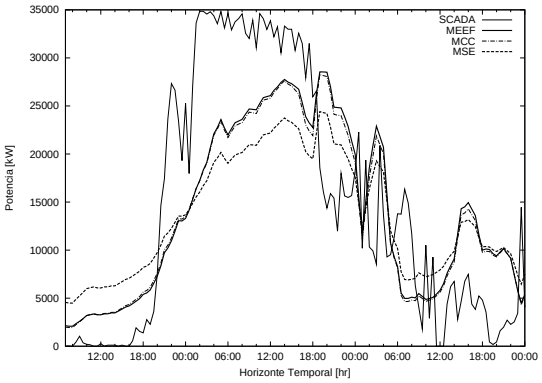

Figura 5.19: *Previsão e valor medido para os dias 10-12 de Novembro de 2006 para o Parque C.*

Na figura 5.18 são visíveis alguns erros de fase introduzidos pelo modelo MM5, os picos que são previstos ocorrer por volta das 18h do segundo e terceiro dia já ocorreram às 17h desse mesmo dia. O erro de fase dos modelos MM5 é um assunto bem conhecido da previsão eólica e tem sido alvo de estudo em diversas publicações (Lange, 2004) e (Hou et al., 2001). Na previsão do Parque A é visível a sobrestimação da produção por parte do critério MSE.

Na figura 5.19 no início da previsão (até às 18h do primeiro dia) é perfeitamente visível um erro de previsão devido à indisponibilidade das máquinas do parque eólico. Apesar de a produção do parque ser nula, o valor da velocidade prevista pelo MM5 é suficientemente elevado para existir uma previsão de produção no parque diferente de zero.

5.3.4 Avaliação do Impacto dos Parâmetros da Rede Neuronal no Treino *On Line*

Nesta secção será avaliado o impacto dos vários parâmetros da rede neuronal no modo treino *on line*. Os parâmetros que mais impacto podem apresentar num modo de treino *on line* são: taxa de aprendizagem, termo de momento e o tamanho do *Kernel* (ou das janelas de *Parzen*). A avaliação dos diferentes parâmetros consiste em testar diferentes valores para os parâmetros e analisar da evolução do erro de

previsão ao longo do horizonte temporal.

O treino *off line* que antecede o treino *on line* e que determina os pesos iniciais do treino *on line* foi realizado com os mesmos parâmetros e rede neuronal da secção anterior.

No critério MEEF apenas será alvo de estudo o impacto da variação do tamanho do *Kernel*, isto porque, como foi exposto na secção 4.3.5 do capítulo 4 o treino com este critério requer uma taxa de aprendizagem de valor muito reduzido. Em (Morejon & Principe, 2004) não é aconselhado o uso de um termo de momento no treino de redes neuronais com o critério MEE, e portanto também para o MEEF não deverá ser utilizado o termo de momento.

Optou-se por variar apenas um parâmetro de cada vez, evitando ajustar mais do que um parâmetro em simultâneo e assim possuir um número infindável de combinações.

O primeiro teste consiste em avaliar diferentes taxas de aprendizagem fixas e o método com taxa de aprendizagem adaptativa descrito na secção 4.4.2 do capítulo 4. Neste teste o termo de momento é nulo e o tamanho do *Kernel* é fixo de valor $\sigma^2 = 0.5$. No método com taxa de aprendizagem variável não foi utilizado termo de momento, isto porque para qualquer valor do termo de momento o seu desempenho era inferior ao termo de momento nulo.

O parque analisado com mais detalhe foi o Parque A, sendo que as conclusões retiradas para os restantes parques são idênticas, como se pode verificar nos resultados do Apêndice H.

Nas figuras 5.20 e 5.21 encontram-se os resultados para o erro EMAN referentes ao Parque A obtidos com o critério MSE e MCC, variando a taxa de aprendizagem no modo de treino *on line*.

Os resultados com uma taxa de aprendizagem fixa e de valor reduzido são significativamente melhores do que com uma taxa de aprendizagem elevada ou variável. O valor máximo da taxa de aprendizagem da metodologia adaptativa foi definido como sendo 0.1, este valor máximo poderia ser menor, mas o resultado final acabaria for ser uma aproximação da solução com uma taxa de aprendizagem fixa com esse valor. Isto porque rapidamente a taxa de aprendizagem atinge um valor máximo e

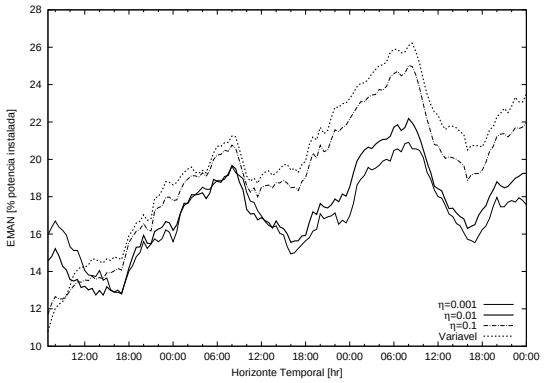

Figura 5.20: *EMAN obtido com o critério MSE para o parque A em função do horizonte temporal para diferentes valores da taxa de aprendizagem.*

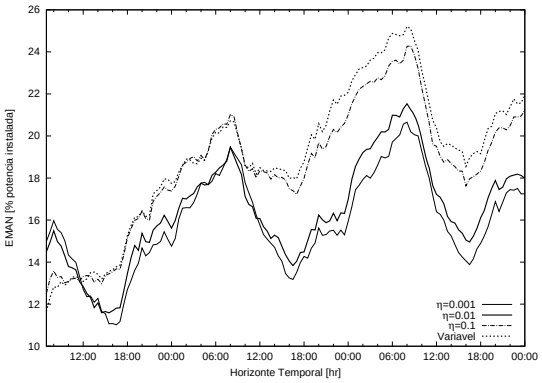

Figura 5.21: *EMAN obtido com o critério MCC para o Parque A em função do horizonte temporal para diferentes valores da taxa de aprendizagem.*

raramente se ”desvia” significativamente desse valor, o que explica o desempenho próxima da taxa de aprendizagem fixa de 0.1.

O fraco desempenho da taxa de aprendizagem adaptativa indica que num processo *on line*, a simples monitorização do sinal do gradiente, entre duas iterações consecutivas, não é por si só uma solução muito robusta e com um desempenho superior. O mesmo resultado foi obtido quando se pretendeu testar o algoritmo iR-PROP em modo *on line*, neste caso, os resultados obtidos também apresentavam baixo desempenho. Sabe-se que a utilização de um η, com valor elevado, permite obter normalmente um processo mais rápido de convergência em zonas de menor curvatura da superfície de erro, mas pode gerar um processo de convergência lento e instável em regiões de maior curvatura. Uma taxa de aprendizagem de valor reduzido garante geralmente um processo de convergência mais regular, no entanto, torna o processo de convergência um pouco lento. Esta lentidão no processo de convergência não constitui um problema para o treino *on line*, como se pode constatar nas figuras 5.20 e 5.21. Com uma taxa de aprendizagem de valor muito reduzido, o treino *on line* apresenta um excelente desempenho em ambos os critérios.

Um resultado relevante acontece na zona do horizonte temporal em que a persistência no treino *off line* apresentava um desempenho superior ao obtido com a

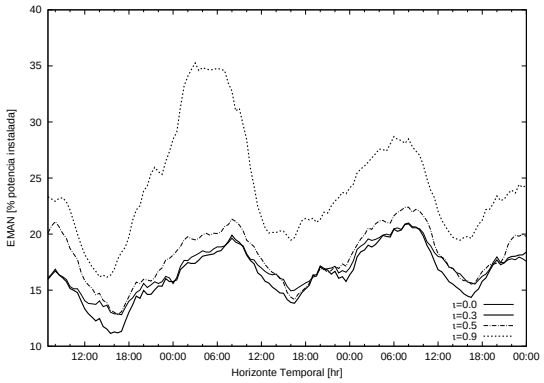

Figura 5.22: *EMAN obtido com o critério MSE para o Parque A em função do horizonte temporal para diferentes valores do termo de momento.*

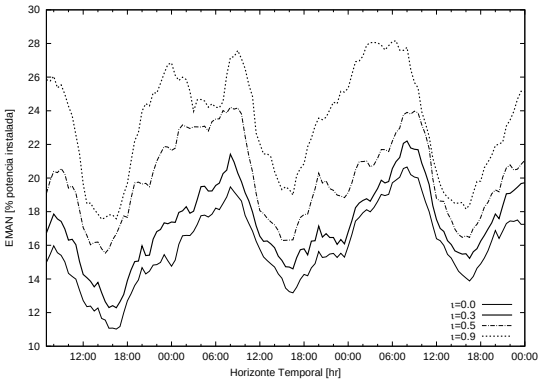

Figura 5.23: *EMAN obtido com o critério MCC para o Parque A em função do horizonte temporal para diferentes valores do termo de momento.*

rede neuronal. No treino *on line* uma taxa de aprendizagem de valor elevado permite obter resultados muito melhores relativamente à persistência do que com uma taxa de aprendizagem de valor reduzido. O uso de uma taxa de aprendizagem de valor elevado nas primeiras 6 horas do horizonte pode ser uma alternativa ao uso de um modelo univariado (i.e. autorregressivo).

Nas figuras 5.22 e 5.23 encontram-se os resultados para o erro EMAN referentes ao Parque A, obtidos com o critério MSE e MCC. Variou-se o termo de momento da rede neuronal e manteve-se a taxa de aprendizagem fixa em 0.001 e o tamanho do *Kernel* em $\sigma^2 = 0.5$. Na figura 5.24 encontra-se o resultado obtido com o critério MSE para a métrica de avaliação REMQN.

Da análise dos resultados do critério MCC com o termo de momento (figura 5.23) concluiu-se que não deve ser utilizado um termo de momento no treino *on line* com este critério. A diferença de desempenho ente a situação com e sem termo de momento é muito elevada. Em (Morejon & Principe, 2004) já era referida a incompatibilidade entre o termo de momento e janelas de *Parzen* devido ao impacto que o valor do *Kernel* tem na superfície de erro.

Na figura 5.22 os resultados com valores do termo de momento muito elevados apresentam um fraco desempenho, sendo que ao contrário do critério MCC, neste

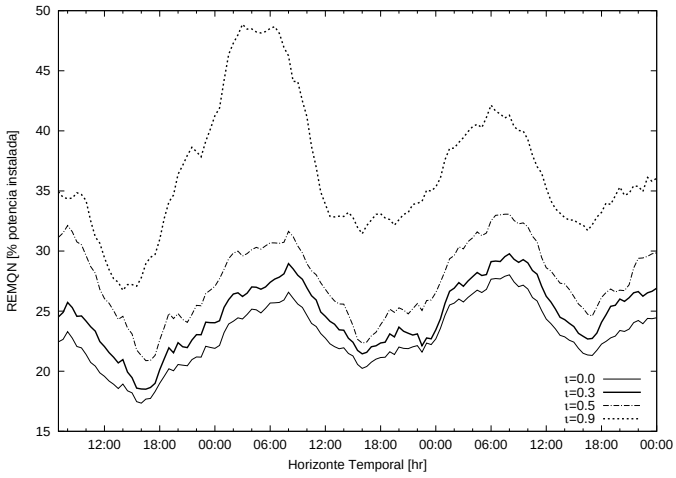

Figura 5.24: *REMQN obtido com o critério MSE para o parque A em função do horizonte temporal para diferentes valores do termo de momento*

caso, o momento de valor 0.3 permite obter melhores resultados do que a solução sem momento. No entanto, analisando segundo o critério REMQN que possuiu uma forte relação com o MSE, a melhor solução é sem o termo de momento.

Também a literatura sobre redes neuronais (secção 4.4.2 do capítulo 4) sustenta a indicação sobre a não utilização de um termo de momento no treino *on line*. Na próxima secção são apresentadas evidências de que se está na presença de um problema de previsão com dados não estacionários, pelo que a superfície de erro muda à medida que nova informação é incorporada na rede neuronal. Um valor elevado do termo momento pode dificultar a adaptação do processo devido às variações da superfície de erro, o que torna muito instável a convergência do treino.

O treino *on line* com base no critério MEEF necessita do ajuste de mais dois parâmetros adicionais, um que representa o tamanho da janela de exemplos sobre a qual se estima a entropia (denotado de M), e um termo de memória (denotado de λ). Na escolha do valor dos dois parâmetros deve ser realizado um *trade-off* entre velocidade e variância da convergência. Nas figuras 5.25 e 5.26 encontram-se os resultados obtidos variando os parâmetros M e λ para o conjunto de dados C.1 (onde existe uma notória alteração do conceito). A métrica de avaliação utilizada foi o REMQN, tendo-se utilizado o valor $\gamma = 0.3$ como peso entre os critérios MEE e MCC. Da análise das figuras concluiu-se que a escolha destes dois parâmetros não

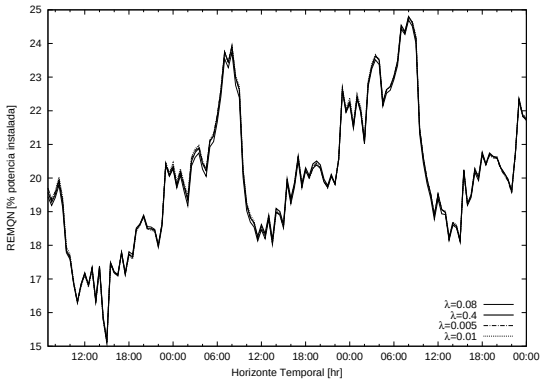

Figura 5.25: *REMQN em função do horizonte temporal para diferentes valores de* λ *e com* $M = 4000$.

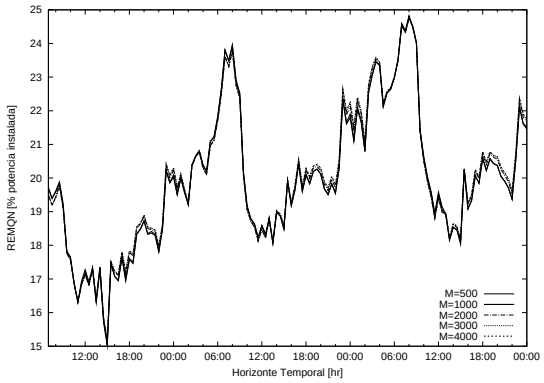

Figura 5.26: *REMQN em função do horizonte temporal para diferentes valores de* M *e com* $\lambda = 0.08$.

é crítica, sendo o desempenho de cada solução muito próximo.

Nas figuras 5.27 e 5.28 encontra-se o erro EMAN referente ao Parque A obtido com a variação do tamanho das janelas de *Parzen* para os critérios MCC e MEEF. A taxa de aprendizagem foi fixada em 0.001 e não foi utilizado termo de momento. Note-se que no caso do MEEF o tamanho da janela foi aumentado de forma homotética nos dois critérios que o constituem. Testou-se a mesma estratégia adaptativa do tamanho da janela de *Parzen* utilizada na estimação da entropia, técnica essa que foi descrita na secção 4.4.2 do capítulo 4. Os parâmetros do MEEF utilizados no teste foram: $\gamma = 0.4$; $M = 4000$ e $\lambda = 0.08$.

Em situações onde o conceito é alterado (ambiente não-estacionário) não é recomendável possuir no critério MCC uma janela de *Parzen* de tamanho reduzido. O uso de uma janela de *Parzen* com tamanho reduzido torna a rede neuronal resistente a *outliers*, mas também a torna resistente a qualquer mudança de conceito, não conseguindo portanto ”aprender” um novo conceito. Este comportamento explica os fracos resultados obtidos por $\sigma^2 = 0.01$ na figura 5.27.

O critério MEEF em parte é influenciado pelo critério MCC, portanto sofre dos mesmo problemas quando o tamanho do *Kernel* é muito reduzido. Na componente de estimação da entropia do MEEF o tamanho da janela de *Parzen* tem uma influência diferente da que é exercida no critério MCC, neste caso, no treino *on line*,

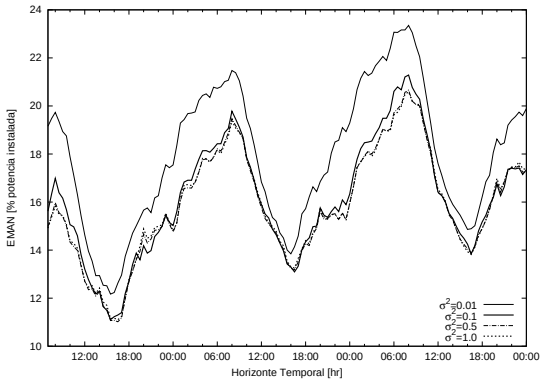

Figura 5.27: *EMAN obtido com o critério MCC para o Parque A em função do horizonte temporal para diferentes valores da janela de Parzen.*

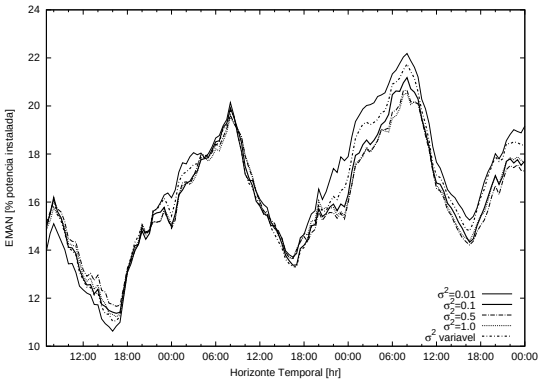

Figura 5.28: *EMAN obtido com o critério MEEF para o Parque A em função do horizonte temporal para diferentes valores da janela de Parzen.*

pequenos valores de σ originam uma superfície de erro com muitos óptimos locais. Com uma superfície de erro muito irregular (com muitos óptimos locais) o treino *on line* tem muitas dificuldades em convergir, resultando numa degradação do desempenho. Tanto para o MEEF como para o MCC um valor de σ reduzido, num modo de treino *on line*, provoca um mau desempenho da rede neuronal, sendo recomendável utilizar um valor entre 0.2 e 1.0.

Os resultados obtidos com um σ variável, no critério MEEF, têm um desempenho inferior ao obtido com os valores fixos (0.1;0.5;1.0). No entanto, esta abordagem possuiu a vantagem de não ser necessário definir um valor para σ, sendo o seu valor calculado directamente dos dados, e sem uma degradação elevada de desempenho.

5.3.5 Comparação entre o Treino *On Line* e *Off Line* da Rede Neuronal

Análise Baseada em Medidas de Erro

Nesta secção são comparados os dois modos de treino da rede neuronal para os três parques, incluindo um quarto exemplo com uma alteração do conceito. Também os dois critérios ITL (MCC e MEEF) são comparados com o MSE no modo de treino

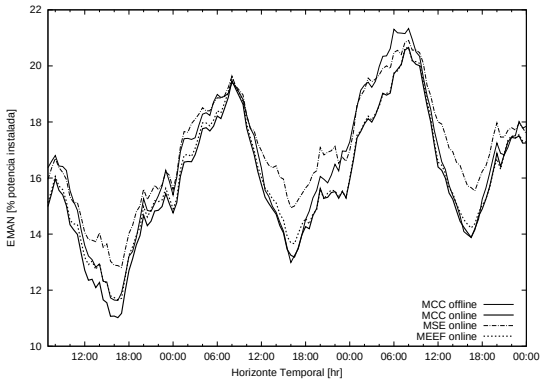

Figura 5.29: *EMAN em função do horizonte temporal obtido com o treino on line para o Parque A.*

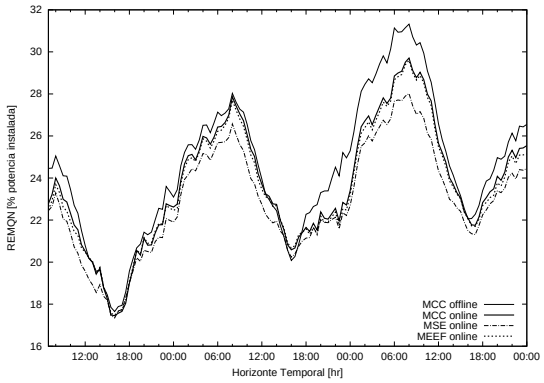

Figura 5.30: *REMQN em função do horizonte temporal obtido com o treino on line para o Parque A.*

on line. Os critérios de avaliação e comparação utilizados apenas se resumem ao EMAN e REMQN. O resultado de referência para comparação dos dois modos de treino é o erro do treino *off line* com o critério MCC.

Foram utilizados os mesmos parâmetros para o treino *on line* em todos os parques eólicos. Os parâmetros da rede neuronal para o treino *on line* são: taxa de aprendizagem de valor 0.001, sem termo de momento, e com um tamanho das janelas de *Parzen* de $\sigma^2 = 0.5$.

Nas figuras 5.29 e 5.30 são comparados os resultados dos diferentes critérios de treino, em modo de treino *on line* para o Parque A, com recurso às métricas de avaliação $EMAN_k$ e $REMQN_k$. O treino *on line* é comparado com o resultado do treino *off line* obtido com o critério MCC.

Os resultados do EMAN obtidos com os critérios MCC e MEEF em modo de treino *on line* são muito semelhantes, não sendo visíveis grandes diferenças. Tal como no treino *off line*, também no treino *on line* existe uma diferença significativa de desempenho entre os critérios ITL e MSE. O treino *on line* permitiu reduzir ao longo de todo o horizonte o erro de previsão o erro EMAN.

Na figura 5.30 o desempenho dos critérios MCC e MEEF é também semelhante, sendo que o critério MSE, tal como no treino *off line*, apresenta melhores resultados. Um resultado relevante é o inferior desempenho do treino *off line* neste critério

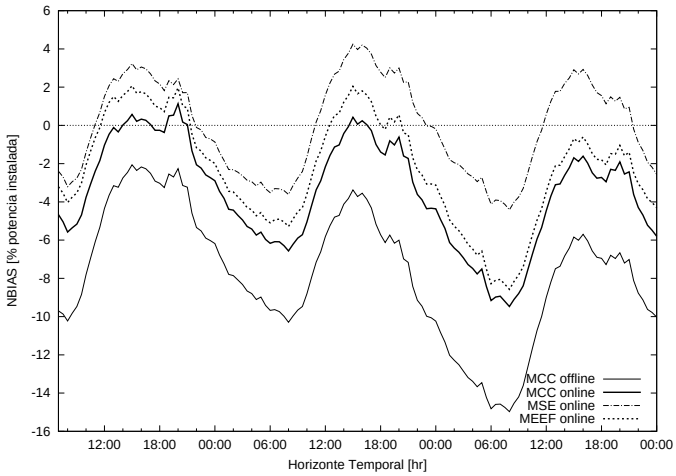

Figura 5.31: *NBIAS em função do horizonte temporal obtido com o treino on line para o Parque A.*

de avaliação, o que indica que o treino *on line* reduz os erros de grande amplitude. Estes erros de grande amplitude podem estar relacionados com diferentes factores, sendo os mais prováveis a má qualidade dos dados ou variações no regime de operação do parque eólico.

Na figura 5.31 é apresentado o resultado do critério de avaliação $NBIAS_k$, relativo ao Parque A e em função do horizonte de previsão.

Da análise da figura 5.31 verifica-se que o treino *on line* tem uma forte influência no erro sistemático da rede neuronal. Enquanto que no treino *off line* os critério ITL apresentam um BIAS sempre negativo, no treino *on line* os valores do BIAS são alvo de um ajuste que os leva a valores positivos. Este ajuste para valores positivos indica um aumento da sobrestimação da potência produzida. Este efeito de ajuste não é verificado no critério MSE. A amplitude em valor absoluto entre o valor máximo e mínimo do BIAS foi reduzida do treino *off line* para o *on line*, passando de 12.91% para 8.33%.

Nas figuras I.1 e I.2 do Apêndice I são apresentados os resultados dos critérios de avaliação $EMAN_k$ e $REMQN_k$ respectivamente, relativos ao Parque B e em função do horizonte de previsão. As conclusões são semelhantes às do Parque A. Neste parque é notória a diferença de desempenho entre os critérios ITL e MSE no treino *on line*, ao contrário do que se verificava no treino *off line*.

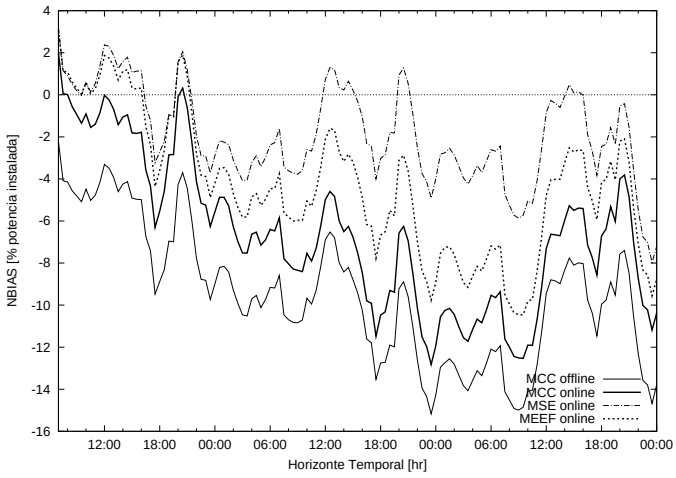

Figura 5.32: *NBIAS em função do horizonte temporal obtido com o treino on line para o Parque B.*

Na figura 5.32 é apresentado o resultado do critério de avaliação $NBIAS_k$ respectivamente, relativo ao Parque B e em função do horizonte de previsão.

Na figura 5.32 verifica-se que também o BIAS do critério MSE sofre alterações com o treino *on line*, sendo que no Parque B esta alteração é mais nítida do que no Parque A. Enquanto que no treino *off line* o BIAS do critério MSE apenas apresentava valores negativos (figura 5.13), no treino *on line* sofre um ajuste para valores positivos. Também neste parque a diferença em valor absoluto entre o máximo e mínimo do BIAS foi reduzida do treino *on line* para o *off line*, passando de 12.9% para 10.8%.

Os resultados para o Parque C não acrescentam nada de novo às análises apresentadas nesta secção. Os resultados encontram-se nas figuras I.3, I.4 e I.5 do Apêndice I.

Uma situação mais interessante do Parque C é a representada pelo conjunto de dados C.1 da tabela 5.7. Neste conjunto de dados encontra-se um exemplo real de alteração de conceito, exemplificado pela entrada em funcionamento de dois aerogeradores. Este exemplo não é uma situação muito frequente em parques eólicos, no entanto no futuro poderão ser realizados incrementos na potência das máquinas de parques eólicos. Este conjunto de dados também permite simular uma situação de limitação da potência produzida pelas máquinas. A limitação de potência das má-

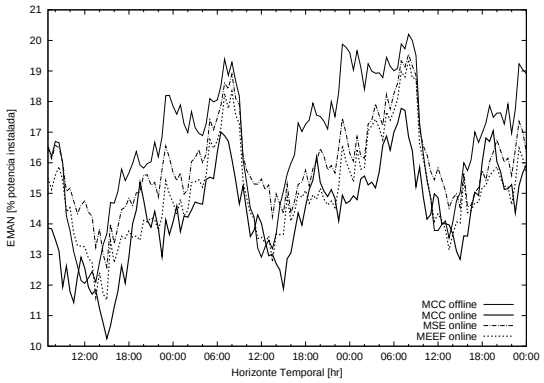

Figura 5.33: *EMAN em função do horizonte temporal obtido com o treino on line para o parque C.1.*

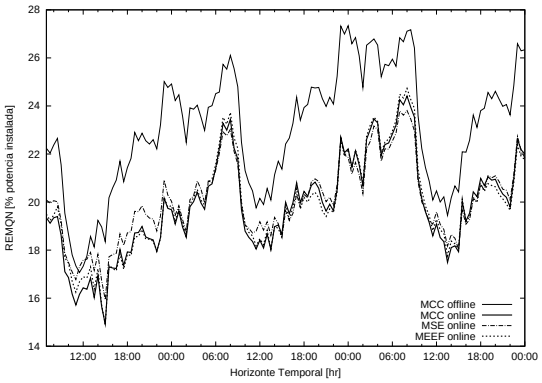

Figura 5.34: *REMQN em função do horizonte temporal obtido com o treino on line para o parque C.1.*

quinas é uma operação frequente num parque eólico, e na participação no mercado de electricidade será uma ferramenta essencial para evitar penalizações da venda de energia ao mercado. A consequência da limitação da potência produzida é o parque eólico possuir duas ou mais curvas de potência, sendo que cada uma das curvas representa um conceito diferente.

Nas figuras 5.33 e 5.34 são apresentados os resultados relativos ao caso particular do parque C (C.1) dos critérios de avaliação $EMAN_k$ e $REMQN_k$ respectivamente.

Os resultados da figura 5.33 demonstram a superioridade do treino *on line* sobre o *off line* quando ocorre uma mudança de conceito. Também numa situação de mudança de conceito os critérios ITL apresentam melhores resultados do que o MSE.

Na figura 5.34 a diferença entre os modos de treino *on line* e *off line* é ainda mais acentuada, a principal contribuição provém dos erros de grande amplitude provocados pela alteração do conceito. O ”domínio” do MSE sobre os critérios ITL no REMQN não se verifica, sendo que em grande parte do horizonte temporal os critérios ITL apresentam um melhor desempenho.

Apenas com base no resultado obtido, para esta situação, é possível concluir que o critério MSE quando comparado com os critérios ITL apresenta uma maior difi-

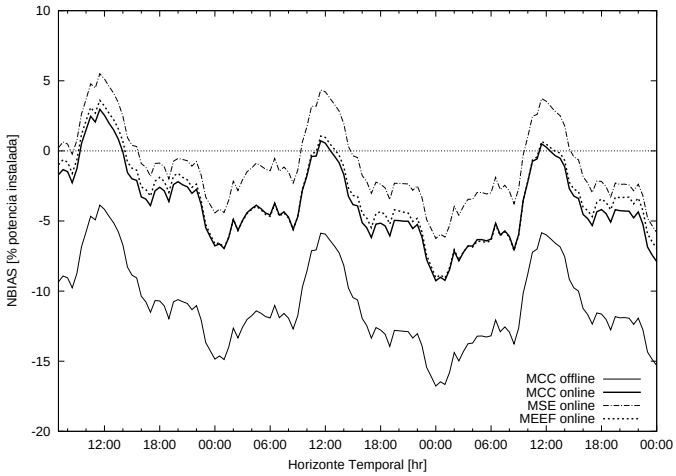

Figura 5.35: *NBIAS em função do horizonte temporal obtido com o treino on line para o parque C.1.*

culdade em incorporar informação sobre um novo conceito, o que consequentemente leva a um pior desempenho no critério REMQN.

Na figura 5.35 é apresentado o resultado do critério de avaliação $NBIAS_k$ relativo ao parque C.1.

Considere-se agora que a rede neuronal treinada com o conjunto de treino C.1 é mantida em funcionamento e usando o treino *on line* os novos valores medidos pelo SCADA foram retropropagados ao longo do ano de 2005. Considere-se outra situação em que a rede neuronal treinada em modo *off line* com o conjunto C.1 foi novamente treinada no fim do ano de 2005, mas com todos os dados disponíveis desse ano. É interessante comparar os resultados obtidos com estas duas redes para o conjunto de teste do Parque C (ano de 2006), e utilizado o critério de treino MCC. Os resultados da comparação com as métricas de avaliação EMAN e REMQN encontram-se nas figuras 5.36 e 5.37 respectivamente.

O erro EMAN da figura 5.36 permite perceber que uma rede treinada em modo *on line* leva a um desempenho da mesma ordem de grandeza do obtido com uma rede treinada com todo o ano de 2005 em modo *off line*, conseguindo até em algumas zonas do horizonte temporal um desempenho superior.

Em relação ao REMQN, a rede *on line* tem os melhores resultados, sendo a diferença de erro bastante considerável. A principal vantagem do treino *on line* neste

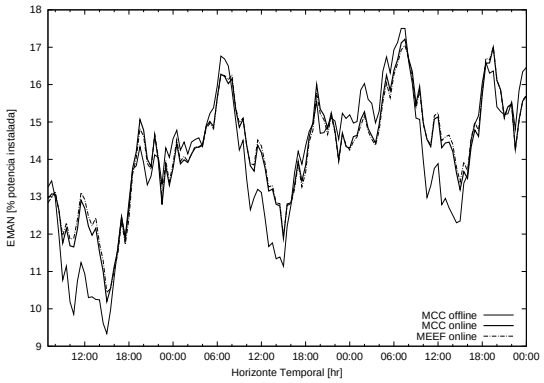

Figura 5.36: *Comparação do EMAN obtido num novo treino da rede usando todo o ano de 2005 e com uma rede treinada em modo on line.*

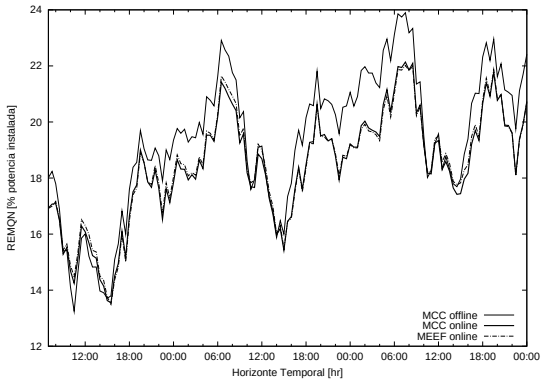

Figura 5.37: *Comparação do REMQN obtido num novo treino da rede usando todo o ano de 2005 e com uma rede treinada em modo on line.*

cenário consiste em tornar desnecessário o armazenamento de dados e manutenção do modelo de previsão.

O resultado da figura 5.35 demonstra que o BIAS do modo de treino *off line* é sempre negativo, indicando uma sistemática subestimação da potência. Enquanto que o treino *on line* ajusta a escala das previsões reduzindo significativamente o erro sistemático.

Previsões

Na figura 5.38 encontra-se o resultado da previsão para o Parque A com os diferentes critérios e modos de treino. O ciclo diurno de Verão encontra-se perfeitamente definido na figura. O treino *on line* permite obter uma melhor aproximação do valor medido. Também a melhor previsão dos critérios ITL é visível na figura, os resultados do treino *on line* com o MSE apenas são próximos dos obtidos com o MCC em modo *off line*.

Nas primeiras 6 horas a previsão não consegue "seguir" o valor medido pelas razões explicadas anteriormente, no entanto, também nesta zona temporal os critérios apresentam previsões muito semelhantes.

Na figura 5.39 encontra-se uma previsão para três dias do conjunto de dados

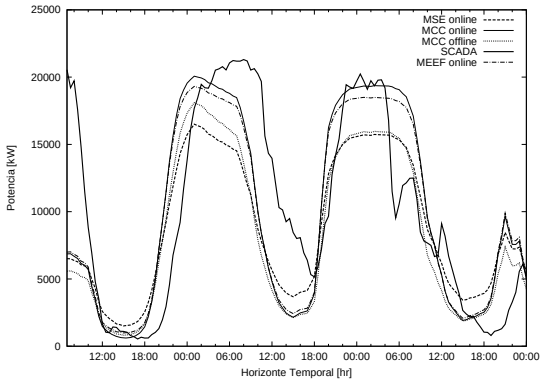

Figura 5.38: *Previsão e valor medido para os dias 09-11 de Junho de 2006 para o Parque A.*

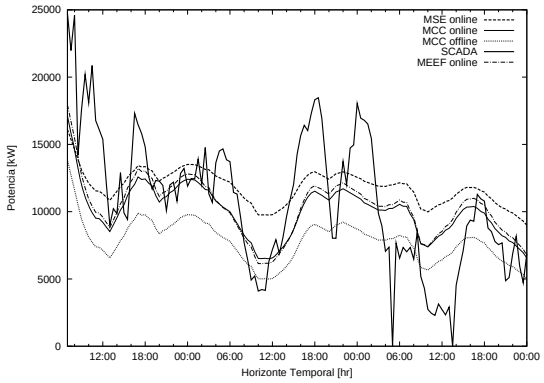

Figura 5.39: *Previsão e valor medido para os dias 12-14 de Novembro de 2005 para o parque C.1.*

C.1. Tal como na figura anterior, também nesta figura percebe-se que o treino *on line* introduz um ajuste de escala ao valor da previsão. Pelo contrário, a previsão *off line* fornece um valor subestimado da potência produzida.

Capítulo 6

Conclusões

Actualmente encontram-se operacionais vários sistemas de previsão de produção eólica para horizontes temporais desde 36 horas até 72 horas. Estes sistemas de previsão são extremamente completos na sua estrutura, englobando diferentes algoritmos e estruturas de previsão. Nesta tese foi proposta a afinação e melhoramento em alguns dos seus módulos de previsão.

As contribuições dos critério ITL e do treino *on line* em conjunto com uma robustez dos parâmetros da rede neuronal, quando integradas nos sistema de previsão mais sofisticados, representam um significativo acréscimo de qualidade. É, no entanto, importante promover um debate sobre o que é de facto uma boa previsão, visto que não é uma definição absoluta. A qualidade das previsões deverá sempre ser analisada em função do fim a que se destinam. Só a operacionalização dos sistemas de previsão eólica e a sua posterior avaliação em ambiente real de funcionamento irá permitir tirar conclusões sobre o que realmente são previsões "úteis".

6.1 Resumo das Principais Contribuições e Resultados

- A distribuição dos erros de previsão da potência produzida num parque eólico não é Gaussiana, portanto um critério que se baseie na medida de informação da fdp dos erros produz melhores resultados que um critério que apenas se baseie na minimização da variância. Os critérios ITL permitem produzir

melhores previsões para um horizonte temporal de 72 horas a serem utilizadas na gestão operacional da rede eléctrica, e reduzir consideravelmente o risco financeiro da participação do produtor de energia eólica no mercado de electricidade.

- Os resultados obtidos com o treino *on line* da rede neuronal apresentam fortes evidências de os dados serem não-estacionários. Uma estratégia adaptativa, de treino dos modelos de previsão, permite obter erros de menor amplitude e fornece ao sistema de previsão uma elevada robustez a mudanças na operação do parque eólico. O caso particular em que a potência instalada do parque eólico é alterada em certo momento da operação, como demonstrado no Parque C, revela a utilidade do uso de uma estratégia de integração de novos valores medidos pelo SCADA na rede neuronal. Sabe-se do contacto com as entidades que exploram os parques, que é muito usual alterar o regime de exploração do parque eólico. Os peritos em previsão meteorológica referem também a frequente alteração dos modelos meso-escala e globais, o que influencia significativamente o erro dos modelos de previsão.

- Não se identifica uma grande diferença de desempenho entre os critérios MCC e MEEF, pelo que a utilização do critério MCC neste tipo de problemas é mais aconselhável. A combinação dos critérios MEE e MCC é uma alternativa ao critério MCC, no entanto sofre da elevada complexidade computacional necessária para estimar a entropia do erro.

- O sinal do erro sistemático (BIAS) não só depende da complexidade e do tipo de terreno do parque eólico, mas também depende do critério (ou função de custo) utilizado no treino dos modelos estatísticos. Os resultados obtidos com os critérios ITL têm tendência a subestimar a potência produzida pelo parque, enquanto que o critério MSE produz com maior frequência sobrestimativas da potência produzida.

- As técnicas de optimização do treino de redes neuronais aplicadas nesta tese revelaram-se muito úteis, tendo sido possível reduzir consideravelmente o

tempo de treino do critério MEE em conjuntos de treino constituídos em média por 30000 exemplos.

- Os parâmetros da rede neuronal no treino *on line* não terão que necessariamente ser dinâmicos, sendo aconselhada uma taxa de aprendizagem de valor reduzido e a não existência de um termo de momento. No caso dos critérios ITL, o tamanho das janelas de *Parzen* apresenta uma importância capital na performance do treino *on line*, valores pequenos originam muitos óptimos locais e insensibilidade à alteração do conceito dos dados. O tamanho das janelas (σ^2) deverá variar entre 0.2 e 1.0 para não existir uma degradação elevada de performance.

- A aplicação do algoritmo iRPROP e do *Kernel* adaptativo ao treino *off line* permitiu desenvolver uma estratégia que torna desnecessário ajustar constantemente os parâmetros da rede cada vez que se muda de parque eólico. Ultrapassou-se no treino *off line* um dos problemas que tornava o uso de redes neuronais na previsão eólica pouco atractivo.

6.2 Trabalho Futuro

- O treino *on line* de redes neuronais é um método capaz de realizar previsões para parques sem histórico ou que apenas se encontram em funcionamento à pouco tempo. Neste caso, não existe um histórico para realizar um treino *off line* da rede neuronal. O método consiste em durante o funcionamento do parque usar todos os novos valores medidos pelo SCADA para treinar a rede neuronal em ”tempo real”. A metodologia proposta consistiria em usar como ”histórico” as curvas dos fabricantes dos aerogeradores para determinar os pesos iniciais da rede neuronal, mais tarde novos valores medidos pelo SCADA são retropropagados pela rede neuronal à medida que vão estando disponíveis. Num trabalho futuro, esta metodologia de treino para parques eólicos sem histórico será avaliada.

- Um aspecto que não foi abordado e que poderá ser estudado num trabalho

futuro é a aplicação da metodologia de treino *on line* numa abordagem por aerogerador. Sabe-se que a variabilidade na produção de um parque com 5 aerogeradores é superior à de um com 20, portanto, em cada aerogerador existe uma variabilidade de produção muito elevada. O uso do treino *on line* em cada aerogerador permite detectar situações de indisponibilidade das máquinas e também lidar com a maior variabilidade da produção dos aerogeradores. A estratégia de treino *on line* também permite ultrapassar o elevado esforço computacional necessário no treino *Batch* de vários aerogeradores em simultâneo.

- Um trabalho futuro deverá também consistir no desenvolvimento de estratégias adaptativas para o tamanho das janelas de *Parzen*, de forma a lidar com alterações do conceito e fluxo contínuo de dados. Um outro parâmetro que não deverá ser tido como constante é o termo de memória da estimação recursiva da entropia, este termo deverá ser alterado em função de um algoritmo de detecção de mudança de conceito. Não deverá também ser abandonada a hipótese de possuir uma taxa de aprendizagem adaptativa, sendo interessante explorar a sua integração com um mecanismo de mudança de conceito.

- Uma outra linha de investigação é o desenvolvimento de funções de custo que permitam incorporar preferências dos promotores dos parques e dos operadores da rede eléctrica, o utilizador final das previsões deverá ser visto como um Agente de Decisão num processo de ajuda à decisão.

Bibliografia

Aa. Nielsen, H., Pinson, P., Christiansen, L. E., Nielsen, T. S., Madsen, H., Badger, J., Giebel, G., & Ravn, H. F. (2007). Improvement and automation of tools for short term wind power forecasting. In *European Wind Energy Conference, Scientific Track*. Milan, Italy.

Agency, I. E. (2005). Variability of wind power and other renewables. Tech. rep., Paris.

Ahlstrom, M., Jones, L., Zavadil, R., & Grant, W. (2005). The future of wind forescasting and utility operation. *IEEE Power and Energy Magazine, vol. 3(6)*, pp. 57–64.

Ahmadzadeh, M. R., & Petrou, M. (2001). Error statistics for slope and aspect when derived from interpolated data. *IEEE Transactions on Geoscience and Remote Sensing, vol. 39(9)*, pp. 1823–1833.

Almeida, L. B. (1997). *Handbook of Neural Computation*, chap. Multilayer perceptrons. Oxford University Press.

Almeida, L. B., Langlois, T., Amaral, J. D., & Plakhov, A. (1998). Parameter adaptation in stochastic optimization. *On-Line Learning in Neural Networks, Cambridge University Press*.

Association, E. W. E. (2006). Large scale integration of wind energy in the european power supply: analysis, issues and recommendations. Tech. rep.

Bailey, B., Brower.M.C., & Zack, J. (1999). Short-term wind forecasting - development and application of a mesoscale model. In *Proceedings of the European Wind Energy Conference EWEC*. Nice.

Barbounis, T., & Theocharis, J. (2006). Locally recurrent neural networks for long-term wind speed and power prediction. *Neurocomputing, vol. 69*, pp. 466–496.

Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Transactions on Energy Conversion, vol. 21*, pp. 273–284.

Bathurst, G., & Strbac, G. (2003). Value of combining energy storage and wind in short-term energy and balancing markets. *Electric Power Systems Research, vol. 67*, pp. 1–8.

Bessa, R., Miranda, V., & Gama, J. (2008). Improvement in wind power forecasting based on information entropy-related concepts. In *2008 IEEE Power Engineering Society General Meeting*.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Bludszuweit, H., Domínguez-Navarro, J., & Llombart, A. (2008). Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems, (Aceite)*.

Bremnes, J. B. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy, vol. 7*, pp. 47–54.

Cabezon, D., Marti, I., San Isidro, M., & Perez, I. (2004). Comparison of methods for power curve modelling. In *CD-Proc. of the 2004 Global Windpower Conference*. Chicago, Illinois (USA).

Castro, F. (1997). *Métodos Numéricos para a Simulação de Escoamentos Atmosféricos sobre Topografia Complexa*. Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto.

Castronuovo, E., & Peças Lopes, J. (2004). On the optimization of the daily operation of a wind-hydro power plant. *IEEE Transactions on Power Systems, vol. 19(3)*, pp. 1599–1606.

Commission, E. (2001). *Directive 2001/77/EC of the European Parliament and of the Council, on the promotion of electricity produced from renewable energy sources in the internal electricity market*. Official Journal of the European Commission.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., & Feitosa, F. (2008). A review on the young history of the wind power short-term prediction. *Renewable & Sustainable Energy Reviews*. (Aceite).

Csiszár, I., & Körner, J. (1981). *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic.

Damousis, I. G., Alexiadis, M. C., Theocharis, J. B., & Dokopoulos, P. S. (2004). A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion, vol. 19(2)*, pp. 352–361.

Devroye, L., & Lugosi, G. (2001). *Combinatorial Methods ind density estimation*. New York: Springer.

Doherty, R., & O'Malley, M. (2005). A new approach to quantify reserve demand in systems with significant installed wind capacity. *IEEE Transactions on Power Systems, vol. 20(2)*, pp. 587–595.

Duin, R. P. W. (1976). On the choice of smoothing parameters for parzen estimators of probability density functions. *IEEE Transactions on Computers, vol. 25*(11), pp. 1175–1179.

El-Fouly, T. H. M., El-Saadany, E. F., & Salama, M. M. A. (2006). Grey predictor for wind energy conversion systems output power prediction. *IEEE Transaction on Power Systems, vol. 21*, pp. 1450–1452.

Erdogmus, D. (2002). *Information Theoretic Learning: Renyi's Entropy and Its Applications to Adaptive System Training*. Ph.D. thesis, University of Florida.

Erdogmus, D., & Jeong, K.-H. (2005). Entropy minimization algorithm for multi-layer perceptron training. MATLAB code.
URL `http://www.cnel.ufl.edu/`

Erdogmus, D., & Principe, J. (2002a). An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems. *IEEE Transactions on Signal Processing, vol. 50*(7), pp. 1780–1786.

Erdogmus, D., & Principe, J. (2002b). Generalized information potential criterion for adaptive system training. *IEEE Transactions on Neural Networks, vol. 13*(5), pp. 1035–1044.

Erdogmus, D., & Principe, J. C. (2000). Comparison of entropy and mean square error criteria in adaptive system training using higher order statistics. In *Proc. ICA*. Helsinki, Finland.

Ernst, B., & Rohrig, K. (2002). Online-monitoring and prediction of wind power in german transmission system operation centres. In *Proceedings of IEA Joint Action Symposium on Wind Forecasting Techniques*.

Fahlman, S. E. (1988). An empirical study of learning speed in back-propagation networks. Tech. rep.

Gangal, A. S., Kalra, P. K., & Chauhan, D. S. (2007). Performance evaluation of complex valued neural networks using various error functions. In *Proceedings of World Academy of Science, Engineering and Technology*, vol. 23.

Giebel, G. (2000). *On the benefits of distributed generation of wind energy in Europe*. Ph.D. thesis, University Carl von Ossietzky, Alemanha.

Giebel, G. (2002). The state-of-the-art in short-term prediction of wind power - a literature overview. Anemos deliverable report d1.1, Risø National Laboratory.

Giebel, G., Brownsword, R., & Kariniotakis, G. (2007). Best practice in short-term forecasting. a users guide. In *Proceedings of the European Wind Energy Conference & Exhibition (EWEC 2007)*. Milan.

Gokhale, D. V. (1975). Maximum entropy characterization of some distributions. *Statistical Distributions in Scientific Work, vol. 3*, pp. 299–304.

Haykin, S. (1995). *Adaptive Filter Theory*. Prentice Hall.

Heinz, C. (2007). *Density Estimation over Data Streams*. Ph.D. thesis, Philipps-Universität, Marburg.

Holttinen, H., & Hirvonen, R. (2005). Power system requirements for wind power. *Capítulo 8 em Wind Power in Power Systems, John Wiley&Sons*, (pp. pp. 144–167).

Hou, D., Kalnay, E., & Droegemeier, K. (2001). Objective verification of the samex '98 ensemble forecast. *Monthly Weather Review, vol. 129*, pp. 73–91.

Huber, P. (1981). *Robust Statistics*. John Wiley & Sons.

Igel, C., & Husken, M. (2000). Improving the rprop learning algorithm. In I. A. Press (Ed.) *Proceedings of the Second International ICSC Symposium on Neural Computation (NC 2000)*, (p. 115.121).

Iyer, M. S., & Rhinehart, R. R. (1999). A method to determine the required number of neural-network training repetitions. *IEEE Transactions on Neural Networks, vol. 10(2)*, pp. 427–432.

Joensen, A., Madsen, H., & Nielsen, T. S. (1997). Statistical methods for predicting wind power. In *Proc. of the 1997 European Wind Energy Conference, EWEC 1997*.

Juban, J., Fugon, L., & Kariniotakis, G. (2007). Probabilistic short-term wind power forecasting based on kernel density estimators. In *European Wind Energy Conference EWEC 2007*. Milan, Italy.

Justus, C., Hargraves, W., Mikhail, A., & Graber, D. (1978). Methods for estimating wind speed frequency distribution. *J. App.Met., vol. 17*, pp. 350–353.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

Kariniotakis, G. (2006). Short-term forecasting of wind power - state-of-the-art, requirements & challenges. In *EES-UETP Course*. Madrid.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. Montreal, Canada.

Korpaas, M., Holen, A., & Hildrum, R. (2003). Operation and sizing of energy storage for wind power plants in a market system. *International Journal of Electrical Power and Energy Systems*, *vol. 25*(8), pp. 599–606.

Landberg, L. (1994). Short-term prediction of local wind conditions. Tech. rep., Riso National Laboratory.

Lange, M. (2003). *Analysis of the Uncertainty of Wind Power Predictions*. Ph.D. thesis, University of Oldenburg.

Lange, M. (2004). On the uncertainty of wind power predictions. In *Proceedings of the EWEA Special Topic Conference: "The Science of Making Torque from Wind"*.

Lange, M. (2005). On the uncertainty of wind power predictions - analysis of the forecast accuracy and statistical distribution of errors. *Transactions of the ASME, Journal of Solar Energy Engineering*, *vol. 2*(no. 127), pp. 177–194.

Larson, K. A. (2004). Advanced short-range wind energy forecasting technologies – challenges, solutions, and validation. In *Global Windpower*.

Lazo, A. V., & Rathie, P. N. (1978). On the entropy of continuous probability distributions. *IEEE Trans. on Information Theory*, *vol. 24*, pp. 120–122.

Li, S., Wunsch, D. C., O'Hair, E. A., & Giesselmann, M. G. (2001). Using neural networks to estimate wind turbine power generation. *IEEE Transactions On Energy Conversion*, *vol. 16*, pp. 276–282.

Liu, W., Pokharel, P., & Principe, J. (2006). Correntropy: A localized similarity measure. In *IJCNN '06. International Joint Conference on Neural Networks*, (pp. 4919– 4924).

Liu, W., Pokharel, P., & Principe, J. (2007a). Correntropy: properties and applications in non-gaussian signal processing. *Accepted in IEEE Transactions on Signal Processin*.

Liu, W., Pokharel, P., & Principe, J. (2007b). A new information theoretic measure of pdf symmetry. In *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (pp. 529–532). Honolulu.

Machenhauer, B. (1988). Hirlam final report. Tech. rep., Copenhagen.

Madsen, H., Kariniotakis, G., Nielsen, T., H.Aa.and Nielsen, & Pinson, P. (2004). A protocol for standardizing the performance evaluation of short-term wind power prediction models. Tech. rep., Project ANEMOS.

Makarov, Y. V., Hawkins, D. L., Leuze, E., & Vidov, J. (2002). California iso wind generation forecasting service design and experience. In *Proc. of the 2002 AWEA Windpower Conference*. Portland, Oregon.

Marques de Sá, J. (2006). Nonparametric density and entropy estimation(with a focus on the parzen window method). Tutorial text, NNIG Neural Network Interest Group.

Milborrow, D. (2001). Penalties for intermittent energy sources. In *PIU Energy Review*.

Milligan, M., Schwartz, M., & Wan, Y.-H. (2004). Statistical wind power forecasting for u.s. wind farms. In *17th Conference on Probability and Statistics in the Atmospheric Sciences/2004 American Meteorological Society Annual Meeting*. Seattle,Washington.

Miranda, V., Cerqueira, C., & Monteiro, C. (2006). Training a fis with epso under an entropy criterion for wind power prediction. In *International Conference*

on Probabilistic Methods Applied to Power Systems (PMAPS 2006), (pp. 1–8). Stockholm.

Møller, J. K., Nielsen, H. A., & Madsen, H. (2006). Algorithms for adaptive quantile regression - and a matlab implementation. Tech. rep., IMM-Technical report-2006-08.

Mohrlen, C. (2004). *Uncertainty in Wind Energy Forecasting*. Ph.D. thesis, National University of Ireland.

Morejon, R., & Principe, J. (2004). Advanced search algorithms for information-theoretic learning with kernel-based estimators. *IEEE Transactions on Neural Networks, vol. 15(4)*, pp. 874–884.

Murphy, A. H. (1993). What is a good forecast? an essay on the nature of goodness in weather forecasting. *Weather and Forecasting, vol. 8(2)*, pp. 281–293.

Nielsen, H. A. (2002). Analysis and simulation of prediction errors for wind power productions reported to nord pool. Informatics and mathematical modeling - eltra project report, Technical University of Denmark, Denmark.

Nielsen, H. A., Madsen, H., & Nielsen, T. S. (2006). Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy, vol. 9*, pp. 95–108.

Nielsen, T. S., & H.Madsen (2001). Zephyr - the prediction models. In *Proc. of the 2001 European Wind Energy Conference, EWEC 2001*. Copenhagen, Denmark.

Orlanski, I. (1975). A rational subdivision of scales of atmospheric processes. *Bull. Amer. Meteor. Soc., vol. 102*, pp. 373–530.

Palmer, T. N. (2000). Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys., vol. 63(2)*, pp. 71–116.

Parzen, E. (1962). On the estimation of a probability density function an the mode. *Annals Math Statistics, vol. 33*.

Patel, M. R. (1999). *Wind and Solar Power Systems*. CRC Press.

Pinson, P. (2006). *Estimation of the Uncertainty in Wind Power Forecasting*. Ph.D. thesis, Ecole des Mines de Paris.

Pinson, P., Chevallier, C., & Kariniotakis, G. (2004). Optimizing benefits from wind power participation in electricity markets using advanced tools for wind power forecasting and uncertainty assessment. In *Proc. of the 2004 EWEC Conference*.

Pinson, P., Chevallier, C., & Kariniotakis, G. (2007a). Trading wind generation with short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems, vol. 22(3)*, pp. 1148–1156.

Pinson, P., & Kariniotakis, G. (2003). Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. In *IEEE Bologna PowerTech Conference*. Bologna, Italy.

Pinson, P., Nielsen, H. A., & Madsen, H. (2007b). Robust estimation of time-varying coefficient functions - application to the modeling of wind power production. Tech. rep., Informatics and Mathematical Modelling, Technical University of Denmark, DTU.

Principe, J., & Erdogmus, D. (2006). From linear adaptive filtering to nonlinear information processing - the design and analysis of information processing systems. *IEEE Signal Processing Magazine, vol. 23*(6), pp. 14–33.

Principe, J., Erdogmus, D., & Kenneth, E. H. I. (2002). Beyond second-order statistics for learning: A pairwise interaction model for entropy estimation. *Natural Computing, vol. 1*(Number 1), pp. 85–108(24).

Principe, J., Euliano, N. R., & Lefebvre, W. C. (1999a). *Neural and Adaptive Systems - Fundamentals Through Simulations*. John Wiley and Sons.

Principe, J., Xu, D., & Fisher, J. (1999b). Information theoretic learning. *Unsupervised Adaptive Filtering, vol. 1*, pp. 265–319. Simon Haykin Editor,Wiley.

Ramírez-Rosado, I., Fernández-Jiménez, L., & Monteiro, C. (2006). Artificial neural network models for wind power short-term forecasting using weather predictions. In *Proceedings International Conference on Modelling, Identification and Control*.

REN (2007). *A Energia Eólica em Portugal – 1º Semestre de 2007*.

Renyi, A. (1976). Some fundamental questions of information theory. *Selected Papers of Alfred Renyi, vol. 2*, pp. 526–552. Akademia Kiado, Budapest.

Riedmiller, M. (1994). Rprop - description and implementation details. Tech. rep., University of Karlsruhe.

Rodrigues, A., Lopes, J., Miranda, P., Palma, L., Monteiro, C., Sousa, J., Bessa, R., Rodrigues, C., & Matos, J. (2007). Eprev - wind power forecasting tool. In *European Wind Energy Conference EWEC 2007*. Milan.

Rodrigues, P., & Gama, J. (2007). Electricity load forecast using data streams techniques. In *Proceedings of the European Workshop on Data Stream Analysis*, (pp. pp. 129–135).

Rosado, I. J. R., Jiménez, L. A. F., & Monteiro, C. (2006). Artificial neural network models for wind power short-term forecasting using weather predictions. In *Proceedings of the 25th IASTED international conference on Modeling, Indentification, and Control*. Lanzarote, Spain: ACTA Press.

Rosado, I. J. R., Jiménez, L. A. F., Monteiro, C., Bessa, R., & Sousa, J. (2007). Comparison of two new short-term wind power forecasting systems. *IEEE Transactions on Power Systems*. (submetido).

Sanchez, I. (2006). Short-term prediction of wind energy production. *Int. J. Forecasting, vol. 22*, pp. 43–56.

Santamaría, I., Pokharel, P., & Principe, J. C. (2006). Generalized correlation function: Definition, properties, and application to blind equalization. *IEEE Transactions on Signal Processing, vol. 54*(6), pp. 2187–2197.

Santos, J. M., Marques de Sá, J., & Alexandre, L. A. (2005). Batch-sequential algorithm for neural networks trained with entropic criteria. In *Proceedings of the 15th Int. Conf. on Artificial Neural Networks*, (p. 91–96).

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal, vol. 27*, pp. 379–423, pp. 623–656.

Sideratos, G., & Hatziargyriou, N. D. (2007). An advanced statistical method for wind power forecasting. *IEEE Transactions on Power Systems, vol. 22*, pp. 258–265.

Silva, F., & Almeida, L. (1991). Speeding-up backpropagation by data orthonormalization. In *Proceedings of International Conference on Artificial Neural Networks*. Espoo, Finland.

Silva, F. M. (1998). *Descorrelação de Dados e Aceleração da Aprendizagem em Redes Neuronais*. Ph.D. thesis, Instituto Superior Técnico (IST), Universidade de Lisboa.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simões, T. (2004). *Base de Dados do Potêncial Energético do Vento em Portugal - Metodologia e Desenvolvimento*. Master's thesis, Faculdade de Ciências da Universidade de Lisboa.

Singh, S., Bhatti, T. S., & Kothari, D. P. (2007). Wind power estimation using artificial neural network. *Journal of Energy Engineering, vol. 133(1)*, pp. 46–52.

Troen, I., & Petersen, E. (1989). Europeanwind atlas. Tech. rep., Risø National Laboratory,.

Tsikalakis, A. G., Katsigiannis, Y. A., Georgilakis, P. S., & Hatziargyriou, N. D. (2006). Determining and exploiting the distribution function of wind power forecasting error for the economic operation of autonomous power systems. In *Proceedings of the IEEE Power Engineering Society General Meeting*.

W. Potter, C., & Negnevitsky, M. (2006). Very short-term wind forecasting for tasmanian power generation. *IEEE Transactions on Power Systems, vol. 21(2)*, pp. 965–972.

Wan, Y.-h., & Bucaneg, D. (2002). Short-term power fluctuations of large wind power plants. In *21st ASME Wind Energy Symposium*. Reno, Nevada.

Watanabe, Y., & Wakao, S. (2007). Application of non-linear modeling methods to wind power estimation. In *European Wind Energy Conference EWEC 2007*. Milan.

Weifeng, L., Pokharel, P., & Principe, J. (2006). Error entropy, correntropy and m-estimation. In *Proceedings of the 2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, (pp. 179–184). Maynooth, Ireland.

Widrow, B., & Stearns, S. (1985). *Adaptive Signal Processing*. New Jersey.

Yuan-Kang, W., & Jing-Shan, H. (2007). A literature review of wind forecasting technology in the world. In *Proceedings of the IEEE Power Tech 2007*. Lausanne, Switzerland.

Zervos, A. (2003). Developing wind energy to meet the kyoto targets in the european union. *Wind Energy, vol. 6*, pp. 309 – 319.

Apêndice A

Algoritmos de Treino Batch e Sequencial

Notação Matemática:

$epocas$**:** Número de épocas de treino da rede neuronal;

N**:** Número de exemplos do conjunto de treino;

$e(s)$**:** erro do exemplo de treino s;

f'**:** derivada da função de activação do neurónio;

net_x^k**:** resultado do somatório no neurónio x da camada k;

δ_x^k**:** ”sensibilidade” do neurónio x da camada k;

w_{yx}^k**:** peso da ligação entre o neurónio y e x que sai da camada k;

O_x^k**:** saída do neurónio x da camada k;

η**:** taxa de aprendizagem;

Algorithm 1: Algoritmo de retropropagação sequential para o critério MSE

Input: Saída da rede y e valor desejado d, número de camadas l, N exemplos do conjunto de treino

Output: Pesos W e bias B da rede neuronal

begin

for $z \in [1 : epocas]$ **do**

for $s \in [1 : N]$ **do**

1. Avaliação dos gradientes para a camada de saída l

$\delta^l(e(s)) = (y - d) \cdot f'\left(net^l(s)\right)$

2. Avaliação dos gradientes locais para a camada com índice o de $l-1$ até 1 e para o neurónio de índice j

$\delta_j^0(e(s)) = f'\left(net_j^0(s)\right) \sum_{k=1}^{m_o+1} \delta_k^{0+1}(e(s))\, w_{kj}^{0+1}$

3. Para cada camada de índice o que vai desde 1 até l actualizar os pesos $\left(w_{ji}^o\right)_{k+1} = \left(w_{ji}^o\right)_k - \eta \cdot \left(\delta_j^0(e(s)) \cdot O_i^{0-1}(s)\right)$

end

Algorithm 2: Algoritmo de retropropagação batch para o critério MSE

Input: Saída da rede y e valor desejado d, número de camadas l, N exemplos do conjunto de treino

Output: Pesos W e bias B da rede neuronal

begin

- **for** $z \in [1 : epocas]$ **do**
  - **for** $s \in [1 : N]$ **do**
    - 1. Avaliação dos gradientes para a camada de saída l
      $\delta^l(e(s)) = (y - d) \cdot f'\left(net^l(s)\right)$
    - 2. Avaliação dos gradientes locais para a camada com índice o de $l-1$ até 1 e para o neurónio de índice j
      $\delta_j^0(e(s)) = f'\left(net_j^0(s)\right) \sum_{k=1}^{m_o+1} \delta_k^{0+1}(e(s))\, w_{kj}^{0+1}$
    - 3. Para cada camada de índice o que vai desde 1 até l avaliar a actualização dos pesos $\Delta w_{ji}^o = \eta\left(\delta_j^o(e(s)) \cdot O_i^{o-1}(s)\right)$
  - Para cada camada de índice o que vai desde 1 até l os pesos são actualizados $\left(w_{ji}^o\right)_{k+1} = \left(w_{ji}^o\right)_k - \Delta w_{ji}^o$

end

Algorithm 3: Treino Batch-Sequencial

Input: Número de subconjuntos L

begin

- **for** $k \in [1 : epocas]$ **do**
  - Criar L subconjuntos Ts
  - **for** $j \in [1 : L]$ **do**
    - Calcular o gradiente da entropia usando treino *Batch* do subconjunto Ts_j
    - Actualização dos pesos

end

Apêndice B

Forças de Informação

Supondo que o resultado do sistema adaptativo são as amostras $\{y_1, \ldots, y_N\}$, podem ser as amostras do erro cometido pelo sistema (supervisionado) ou a resposta do sistema (não-supervisionado). Por simplicidade, assumindo que se está a lidar com uma variável unidimensional, no entanto a extensão do problema para várias dimensões é trivial. A cada partícula está associado um campo de potencial, o valor da partícula representa a posição da partícula no campo. No caso de vectores com mais de uma dimensão, a posição é um vector de coordenadas. Supondo que cada partícula y_i gera um campo de energia potencial, dado por $v(\xi)$[1], torna-se necessário que esta função seja contínua e diferenciável, também é necessário que satisfaça a condição simétrica $v(\xi) = v(-\xi)$. Devido à simetria e diferenciabilidade o gradiente do potencial na origem é zero, o que significa que a força (gradiente do potencial) que uma partícula exerce sobre ela própria é zero. Da física, também se sabe que a força aplicada a uma partícula por outra é em função da posição relativa da partícula afectada, com respeito à fonte da força. Este conceito permite verificar que o potencial energético da partícula y_j devido à relação com a partícula e_j, denotado de $v\left(y_j|y_i\right)$, é dado por $V\left(y_j|y_i\right) = v\left(y_j - y_i\right)$, onde o vector das diferenças $\left(y_j - y_i\right)$ representa a posição relativa de y_j com respeito a y_i.

A energia potencial total de y_j, devido à relação desta partícula com as restantes, é

[1] a nomenclatura usada na física para o campo potencial é V

dada por:

$$V\left(y_j\right) = \sum_{i=1}^{N} V\left(y_j|y_i\right) = \sum_{i=1}^{N} v\left(y_j - y_i\right) \tag{B.1}$$

Em analogia com a física a força de interacção entre as partículas é dada por:

$$F\left(y_j|y_i\right) = \frac{\partial V\left(y_j|y_i\right)}{\partial y_j} = \frac{\partial v\left(y_j - y_i\right)}{\partial\left(y_j - y_i\right)} \tag{B.2}$$

Da propriedade de sobreposição das forças, obtém-se a força total que actua sobre a partícula y_j:

$$F\left(y_j\right) = \sum_{i=1}^{N} F\left(y_j|y_i\right) = \sum_{i=1}^{N} v'\left(y_j - y_i\right) \tag{B.3}$$

Como esperado, a força aplicada à partícula, por ela própria, apresenta valor nulo. Finalmente, o campo potencial de todo o conjunto de amostras é a soma do campo potencial de cada partícula, o que origina:

$$V(y) = \sum_{i=1}^{N}\sum_{j=1}^{N} v\left(y_j - y_i\right) \tag{B.4}$$

A mesma analogia também é válida para o critério MSE. Neste caso, a interacção acontece entre cada partícula e a média amostral(Principe et al., 2002). Considerando o problema de maximização do potencial energético tem-se:

$$V_{MSE} = -\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{B.5}$$

onde $\bar{y}$ é a média amostral. Este problema de maximização não é mais do que a minimização da variância. A contribuição de uma única partícula para o campo potencial é dada por:

$$V_{MSE}(y_i;\bar{y}) = -\frac{1}{N}(y_i - \bar{y})^2 \tag{B.6}$$

onde a fonte do campo potencial é a média amostral e o potencial da partícula, depende da distância Euclidiana à fonte. A força que actua na partícula y é naturalmente definida como a derivada do potencial relativamente à posição da partícula, tem-se então:

$$F_{MSE}(y) = \frac{\partial V_{MSE}(y)}{\partial y} = -\frac{2}{N}(y - \bar{y}) \tag{B.7}$$

A diferença essencial, entre o critério MEE e MSE, encontra-se na fonte do potencial (e consequentemente na força), no caso do MSE a fonte é a média amostral, enquanto que no caso do MEE cada partícula individualmente é um fonte de potencial.

Apêndice C

Algoritmos de Retropropagação dos Critérios ITL

Notação Matemática:

$epocas$**:** Número de épocas de treino da rede neuronal;

N**:** Número de exemplos do conjunto de treino;

$e(s)$**:** erro do exemplo de treino s;

f'**:** derivada da função de activação do neurónio;

net_x^k**:** resultado do somatório no neurónio x da camada k;

δ_x^k**:** ”sensibilidade” do neurónio x da camada k;

$\delta(e(s)|e(t))$**:** ”sensibilidade” da diferença entre os erros $e(s)$ e $e(t)$;

IF**:** força da informação da diferença mútua entre os erros;

w_{yx}^k**:** peso da ligação entre o neurónio y e x que sai da camada k;

O_x^k**:** saída do neurónio x da camada k;

η**:** taxa de aprendizagem;

Algorithm 4: Algoritmo de retropropagação para o critério MEE

Input: Vector dos erros do sistema adaptativo($E = D - Y$), tamanho do *kernel* σ, número de camadas l, N exemplos do conjunto de treino.

Output: Pesos W e bias B da rede neuronal.

begin

for $s \in [1 : N]$ **do**

for $t \in [1 : N]$ **do**

1. Avaliação dos gradientes locais para a camada de saída l

$$\delta^l(e(s)|e(t)) = -IF(e(s)|e(t)) \cdot f'\left(net^l(s)\right)$$

$$\delta^l(e(t)|e(s)) = -IF(e(t)|e(s)) \cdot f'\left(net^l(t)\right)$$

2. Avaliação dos gradientes locais para a camada com índice o de $l-1$ até 1 e para o neurónio de índice j

$$\delta_j^0(e(s)|e(t)) = f'\left(net_j^0(s)\right) \cdot \sum_{k=1}^{m_0+1} \delta_k^{o+1}(e(s)|e(t))\, w_{kj}^{o+1}$$

$$\delta_j^0(e(t)|e(s)) = f'\left(net_j^0(t)\right) \cdot \sum_{k=1}^{m_0+1} \delta_k^{o+1}(e(t)|e(s))\, w_{kj}^{o+1}$$

3. Para cada camada de índice o que vai desde 1 até l avaliar a actualização dos pesos

$$\Delta w_{ji}^o = \eta\left(\delta_j^0(e(s)|e(t)) \cdot O_i^{o-1}(s) + \delta_j^0(e(t)|e(s)) \cdot O_i^{o-1}(t)\right)$$

end

Algorithm 5: Algoritmo de retropropagação para o critério MCC

Input: Saída da rede y e valor desejado d, tamanho do *kernel* σ, número de camadas l, N exemplos do conjunto de treino.

Output: Pesos W e bias B da rede neuronal.

begin

for $s \in [1:N]$ **do**

1. Avaliação dos gradientes para a camada de saída l

$\delta^l(e(s)) = K'_\sigma(y-d) \cdot f'\left(net^l(s)\right)$

2. Avaliação dos gradientes para a camada com índice o de $l-1$ até 1 e para o neurónio de índice j

$\delta_j^0(e(s)) = f'\left(net_j^0(s)\right) \sum_{k=1}^{m_o+1} \delta_k^{0+1}(e(s))\, w_{kj}^{0+1}$

3. Para cada camada de índice o que vai desde 1 até l avaliar a actualização dos pesos $\Delta w_{ji}^o = \eta\left(\delta_j^o(e(s)) \cdot O_i^{o-1}(s)\right)$

end

Algorithm 6: Algoritmo de retropropagação para o critério MEEF

Input: Saída da rede y e valor desejado d, vector dos erros do sistema adaptativo ($E = D - Y$), tamanho do *kernel* σ_{MCC} e σ_{MEE}, γ,número de camadas l, N exemplos do conjunto de treino

Output: Pesos W e bias B da rede neuronal

begin

for $s \in [1:N]$ **do**

Algoritmo de retropropagação MCC

for $t \in [1:N]$ **do**

Algoritmo de retropropagação MEE

Depois de determinado o δ de cada critério os pesos são actualizados da seguinte forma: $\left(w_{ji}\right)_{k+1} = \left(w_{ji}\right)_k + \eta \cdot \left[(1-\gamma)\left(\Delta w_{ji}\right)_{k+1}^{MEE} + \gamma\left(\Delta w_{ji}\right)_{k+1}^{MCC}\right]$

end

Apêndice D

Testes Preliminares da Estrutura da Rede Neuronal

Os resultados apresentados neste anexo, são testes preliminares realizados sobre conjuntos de dados dos parques A e B, para um horizonte temporal de 24 horas. As características dos dados encontram-se nas tabelas D.1 e D.2.

Tabela D.1: *Características dos dados do Parque A no ano de 2005.*

Medida	Treino	Validação	Teste
Média (kW)	5499.11	8061.21	7561.12
Mínimo (kW)	-74.33	-71.83	-75.83
Máximo (kW)	21651.00	21662.66	21680.16
Número	8445	1689	5067

Tabela D.2: *Características dos dados do Parque B no ano de 2005.*

Medida	Treino	Validação	Teste
Média (kW)	6184.00	4791.77	4266.19
Mínimo (kW)	-60.50	-80.17	-767.50
Máximo (kW)	16142.67	16105.67	16105.17
Número	4893	815	2446

Em cada teste foram realizadas 25 simulações, com diferentes pesos iniciais. Os pesos iniciais foram gerados utilizando uma distribuição uniforme, num intervalo entre -1 e 1. Os resultados finais são uma média dos resultados das 25 simulações. O critério usado nos testes foi o MSE.

A métrica utilizada para avaliar a performance das redes neuronais foi uma janela deslizante de tamanho igual a duas semanas (672 pontos), sobre a qual é calculado o erro absoluto médio.

Neste estudo é analisado o impacto da topologia da rede neuronal, nos modos de treino *off line* e *on line*. Foram para isso testados diferentes números de neurónios e topologias com uma e duas camadas escondidas. Para a escolha das topologias a analisar foi utilizada a funcionalidade *Intelligent Problem Solver* do software *Statistica*[1], para determinar quais as melhores topologias da rede no problema em questão. Em todas as experiências os parâmetros encontram-se fixos, com os seguintes valores: taxa de aprendizagem de 0.1 e termo de momento de 0.5. A nomenclatura '3-5-1' utilizada para caracterizar a rede significa que a rede possui 3 entradas, 5 neurónios na camada interior e 1 neurónio para a saída da rede.

Para os dados de ambos os parques não se verificou um impacto significativo da topologia da rede no modo *off line*, a performance da rede é praticamente semelhante. A comparação da evolução do erro na janela deslizante, em modo de treino *off line* para ambos os parques, encontra-se nas figuras D.1 e D.2. No caso do treino *on line* existe uma diferença de erro entre as topologias, no entanto as redes com um menor número de neurónios e apenas com uma camada, apresentam melhor desempenho. Nas figuras D.3 e D.4 encontram-se os resultados relativos aos parques A e B, respectivamente.

[1] http://www.statsoft.com/

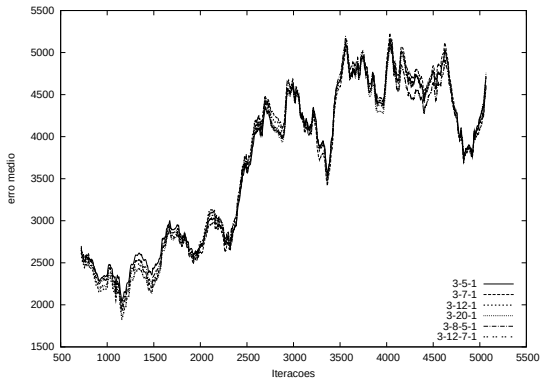

Figura D.1: *Evolução do erro no modo de treino off line para diferentes configurações da rede para o Parque A.*

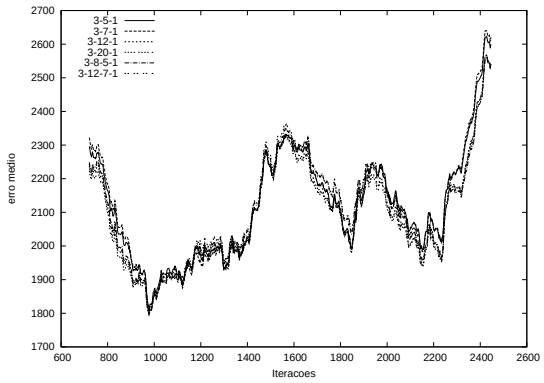

Figura D.2: *Evolução do erro no modo de treino off line para diferentes configurações da rede para o Parque B.*

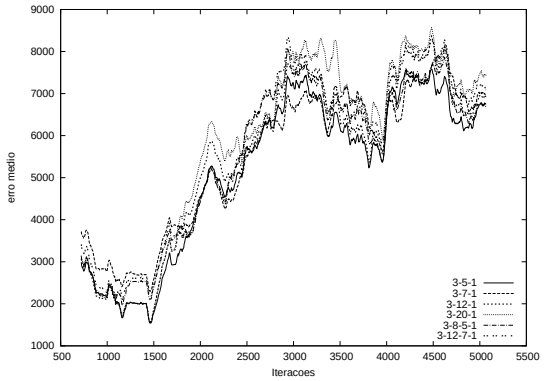

Figura D.3: *Evolução do erro no modo de treino on line para diferentes configurações da rede para o Parque A.*

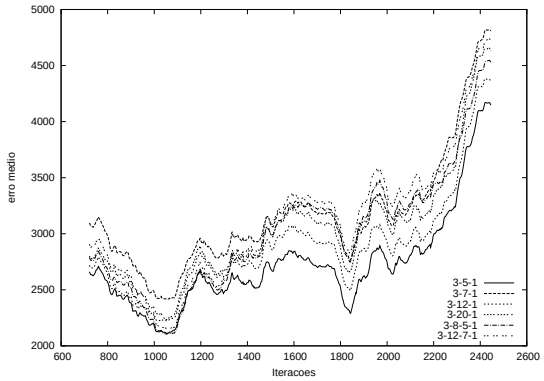

Figura D.4: *Evolução do erro no modo de treino on line para diferentes configurações da rede para o Parque B.*

Apêndice E

Algoritmos de Teste de Previsões para um Horizonte de 72 Horas

Algorithm 7: Teste off line

Input: Velocidade $\widehat{v}_{t+k|t}$ e direcção $\widehat{d}_{t+k|t}$ do vento; número de previsões do conjunto de teste $ndias$; rede neuronal treinada com um treino batch-sequencial RN.

Output: Previsão do parque $\widehat{p}_{t+k|t}$.

begin

  for $D \in [1 : ndias]$ **do**

    Previsão realizada às 6:30 para o resto do dia e mais os dois próximos dias completos:

    for $k \in [1 : 131]$ **do**

      $prev \longleftarrow \widehat{p}_{t+k|t+D} = RN\left\{\widehat{v}_{t+k|t+D}, \widehat{d}_{t+k|t+D}\right\}$

end

Algorithm 8: Teste on line

Input: Velocidade $\widehat{v}_{t+k|t}$ e direcção $\widehat{d}_{t+k|t}$ do vento; novo valor de produção p_s medido no instante s pelo SCADA; número de previsões do conjunto de teste $ndias$; rede neuronal treinada com um treino batch-sequencial RN.

Output: Previsão do parque $\widehat{p}_{t+k|t}$.

begin

  for $D \in [1 : ndias]$ **do**

    Previsão realizada às 6:30 para o resto do dia e mais os dois próximos dias completos:

    for $k \in [1 : 131]$ **do**

      $prev \longleftarrow \widehat{p}_{t+k|t+D} = RN\left\{\widehat{v}_{t+k|t+D}, \widehat{d}_{t+k|t+D}\right\}$

    A cada 30 minutos chegam novos valores medidos pelo SCADA, são retropropagados os novos valores medidos desde as 6:30 até às 6:30 do dia seguinte.

    Detecção de um novo máximo ou mínimo das variáveis de entrada:

    if $v_s > v_{max}$ **then**

      $v_{max} = \hat{v}_s$

    if $v_s < v_{min}$ **then**

      $v_{min} = \hat{v}_s$

    if $p_s > p_{max}$ **then**

      $p_{max} = p_s$

    if $p_s < p_{min}$ **then**

      $p_{min} = p_s$

    $prev \longleftarrow \widehat{p}_{s|t+D} = RN\left\{\widehat{v}_s, \widehat{d}_s\right\}$

    $E \longleftarrow prev - p_s$

    retropropagar E **pela rede neuronal;**

end

Apêndice F

Métricas para Avaliação das Previsões

O erro de previsão é definido como a diferença entre o valor medido e o valor previsto. Assim sendo, e porque se separou cada previsão para cada instante do horizonte temporal, tem-se o erro de previsão para o instante $t+k$ dado que se está em t definido da seguinte forma:

$$e(t+k|t) = \hat{p}(t+k|t) - p(t+k) \tag{F.1}$$

O erro sistemático do modelo, chamado ”bias”, é estimado para cada instante do horizonte de previsão, como o erro médio em todo o período de teste:

$$BIAS(k) = \hat{\mu}_e(k) = \frac{1}{N}\sum_{t=1}^{N} e(t+k|t) \tag{F.2}$$

Existem dois critérios para ilustrarem a performance de previsão: o erro médio absoluto (EMA) e a raiz do erro médio quadrático (REMQ) . O EMA é dado por

$$EMA(k) = \frac{1}{N}\sum_{t=1}^{N} |e(t+k|t)| \tag{F.3}$$

O erro REMQ é dado por:

$$REMQ(k) = \sqrt{\frac{\sum_{t=1}^{N} (e(t+k|t))^2}{N}} \tag{F.4}$$

Tanto os erros aleatórios como sistemático contribuem para o valor do EMA e REMQ. Os valores de $BIAS(k)$ e $EMA(k)$ estão associados com o momento de primeira ordem do erro de previsão, e portanto, são medidas directamente relacionadas com a

energia produzida. Os valores de $REMQ(k)$ estão associados com o momento de segunda ordem, e portanto, estão associados à variância dos erros de previsão. Neste ultima métrica os elevados erros de previsão apresentam uma grande contribuição.

Todos os erros de previsão apresentados são em relação à potência instalada do parque, para permitir a comparação de erros, entre diferentes parques eólicos. Neste problema não é possível utilizar a definição de MAPE (*Mean Absolute Percentage Error*) porque muitas vezes a produção do parque é nula. A comparação com a persistência tem grande relevância, porque o valor do erro da persistência dá uma indicação sobre a dificuldade de previsão num determinado parque eólico. Existem situações em que os erros de previsão são muito baixos, mas também o erro da persistência é muito baixo, podendo até apresentar um valor inferior.

Apêndice G

Resultados *Off Line* para 72 Horas

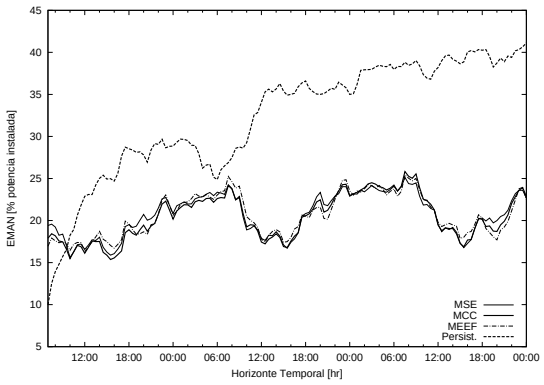

Figura G.1: *EMAN para o Parque B no modo de treino off line.*

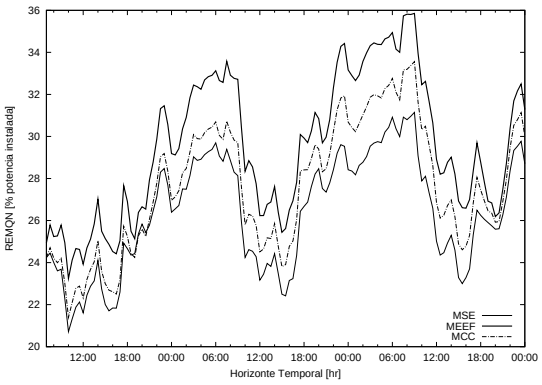

Figura G.2: *REMQN para o Parque B no modo de treino off line.*

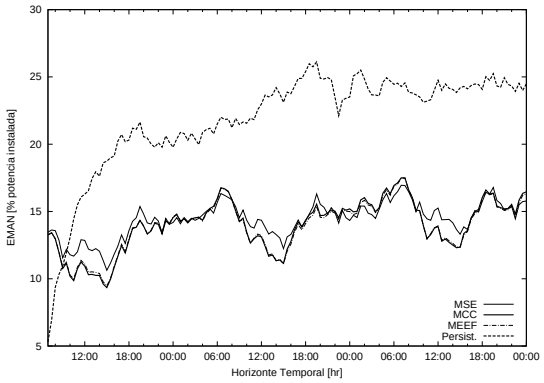

Figura G.3: *EMAN para o Parque C no modo de treino off line.*

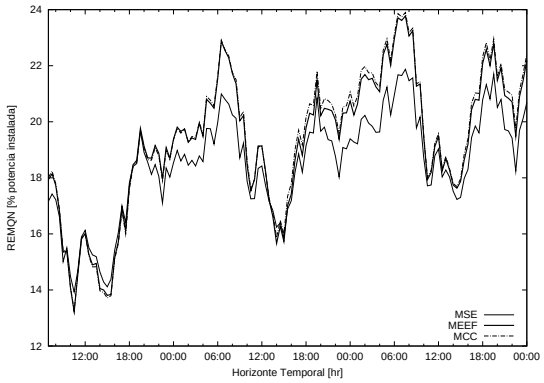

Figura G.4: *REMQN para o Parque C no modo de treino off line.*

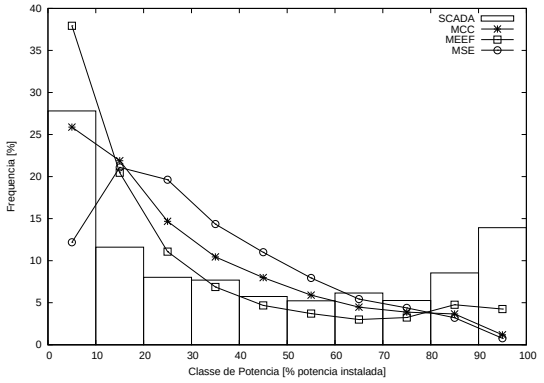

Figura G.5: *Frequência dos valores medidos de produção do Parque B e dos valores previstos com os diferentes critérios.*

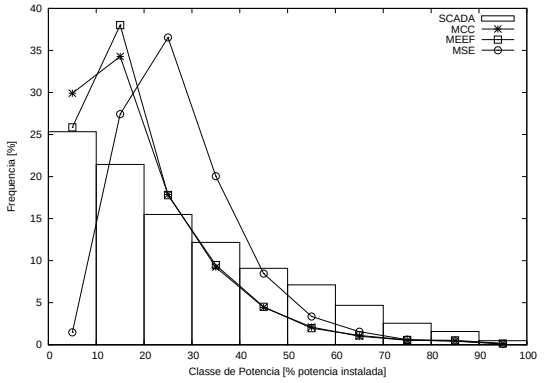

Figura G.6: *Frequência dos valores medidos de produção do Parque C e dos valores previstos com os diferentes critérios.*

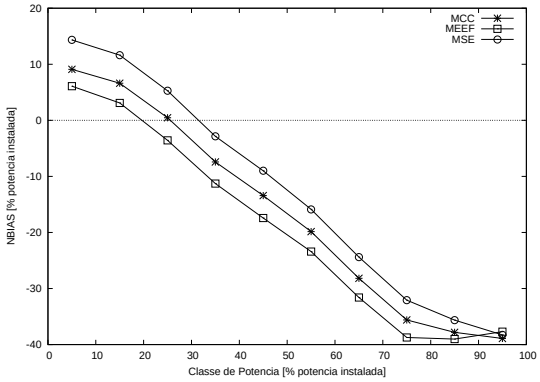

Figura G.7: *NBIAS do erro de previsão dos diferentes critérios em função da classe de produção do parque eólico B.*

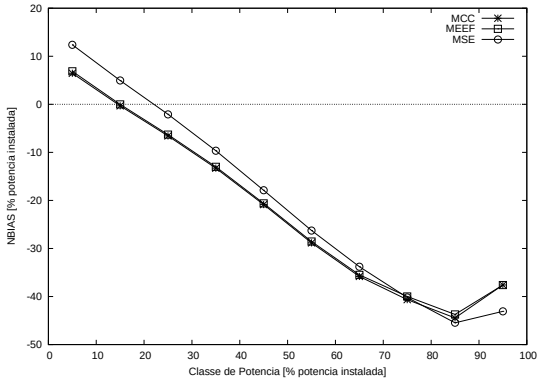

Figura G.8: *NBIAS do erro de previsão dos diferentes critérios em função da classe de produção do parque eólico C.*

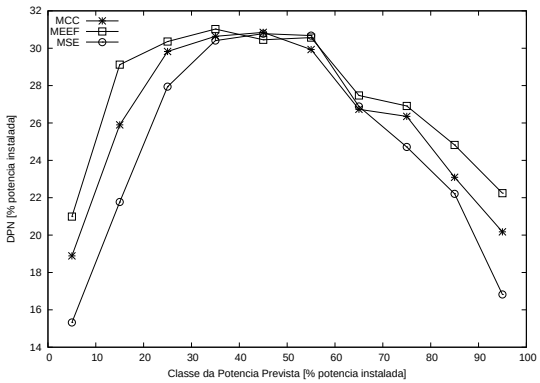

Figura G.9: *Desvio padrão da distribuição dos erros de previsão em função das classes de previsão para o Parque B.*

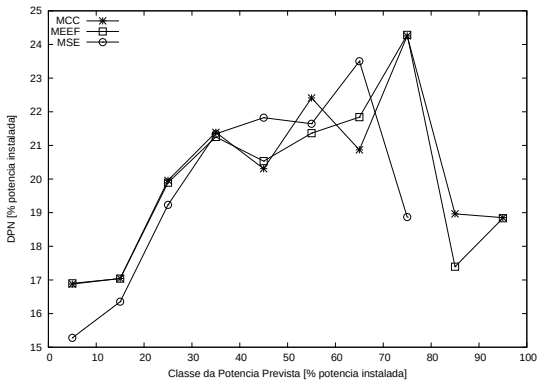

Figura G.10: *Desvio padrão da distribuição dos erros de previsão em função das classes de previsão para o Parque C.*

Apêndice H

Variação dos Parâmetros do Treino *On Line*

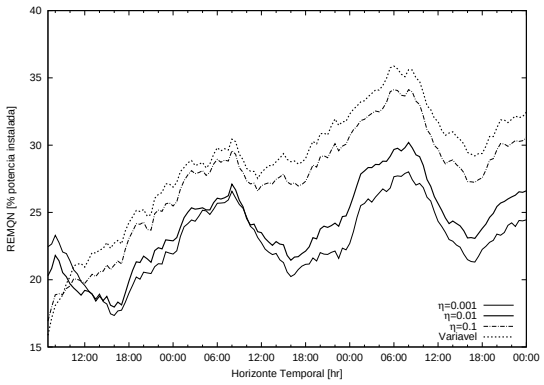

Figura H.1: *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque A com o critério MSE.*

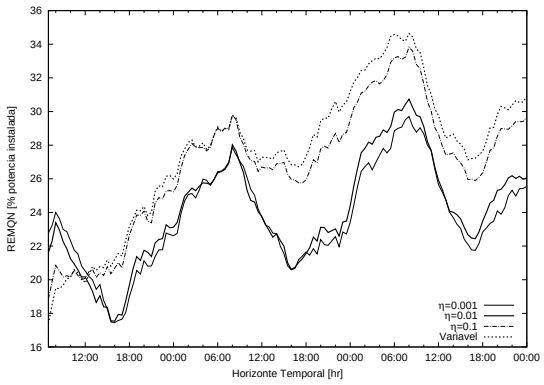

Figura H.2: *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque A com o critério MCC.*

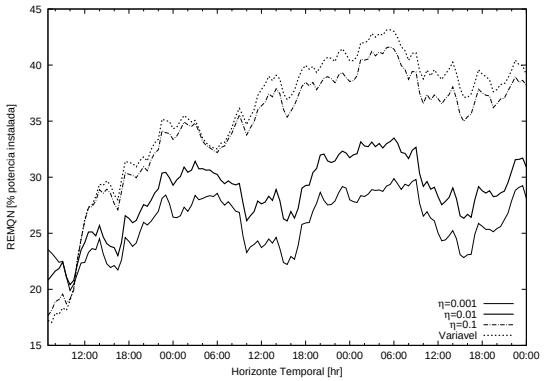

Figura H.3: *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque B com o critério MSE.*

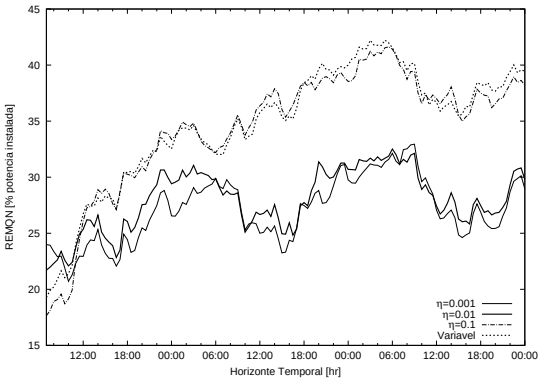

Figura H.4: *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque B com o critério MCC.*

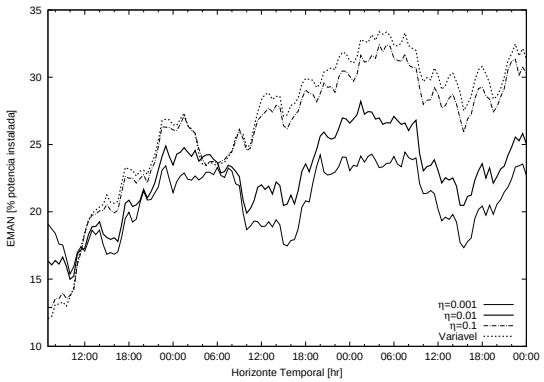

Figura H.5: *EMAN obtido com o critério MSE para o Parque B em função do horizonte temporal para diferentes valores da taxa de aprendizagem.*

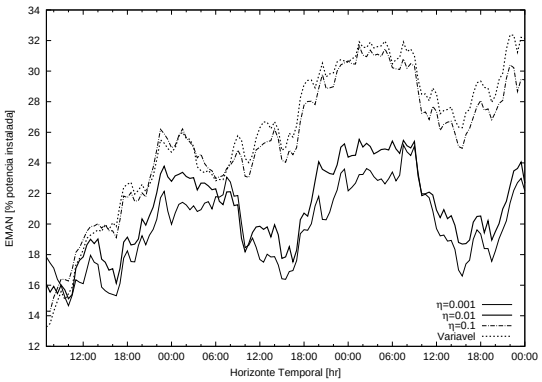

Figura H.6: *EMAN obtido com o critério MCC para o Parque B em função do horizonte temporal para diferentes valores da taxa de aprendizagem.*

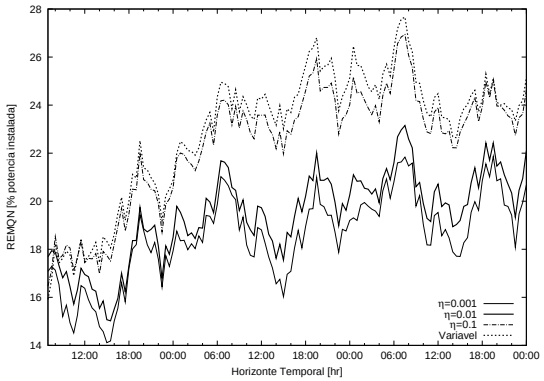

Figura H.7: *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque C com o critério MSE.*

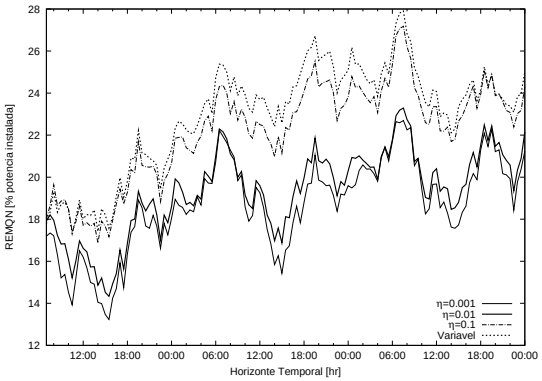

Figura H.8: *REMQN em função do horizonte temporal para diferentes valores da taxa de aprendizagem para o Parque C com o critério MCC.*

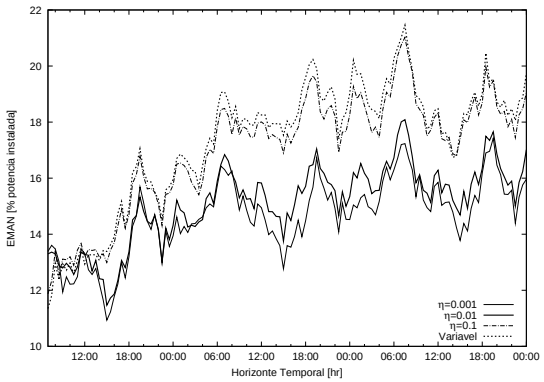

Figura H.9: *EMAN obtido com o critério MSE para o Parque C em função do horizonte temporal para diferentes valores da taxa de aprendizagem.*

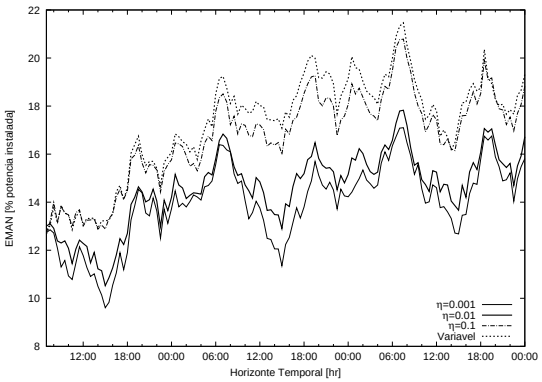

Figura H.10: *EMAN obtido com o critério MCC para o Parque C em função do horizonte temporal para diferentes valores da taxa de aprendizagem.*

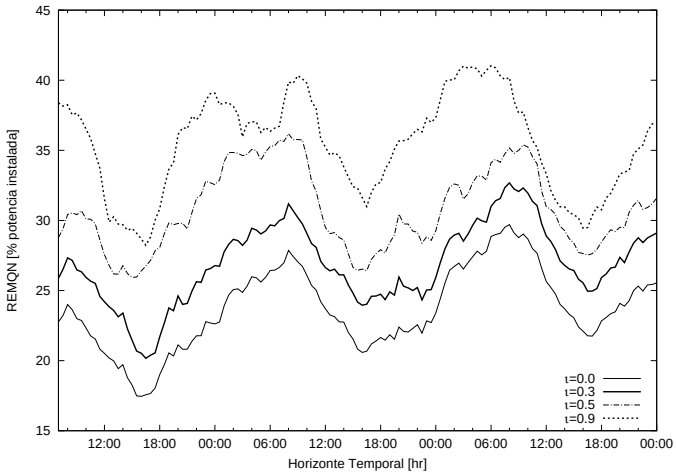

Figura H.11: *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque A com o critério MCC.*

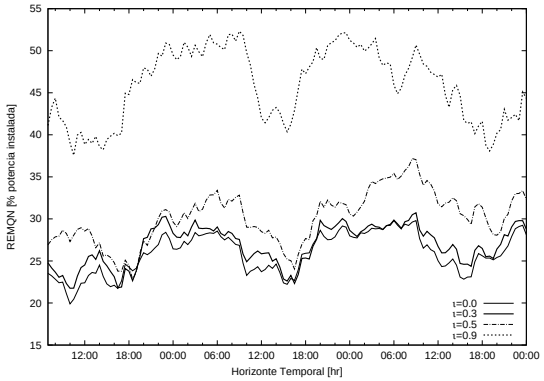

Figura H.12: *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MSE.*

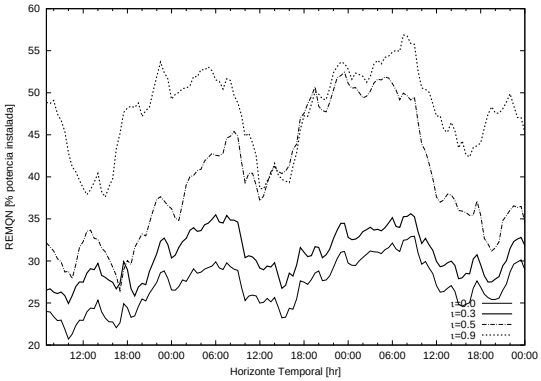

Figura H.13: *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MCC.*

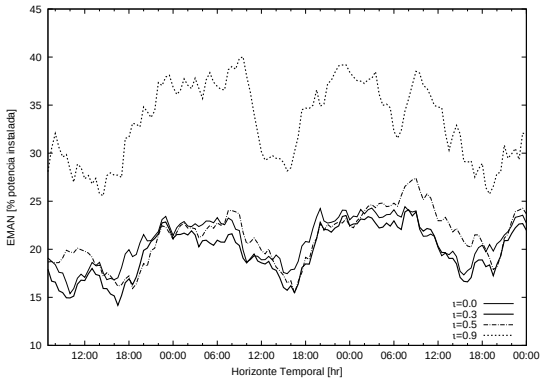

Figura H.14: *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MSE.*

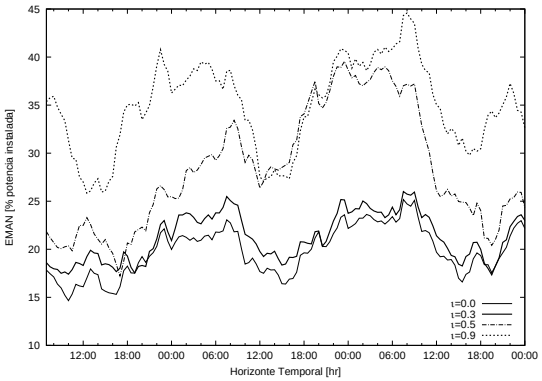

Figura H.15: *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque B com o critério MCC.*

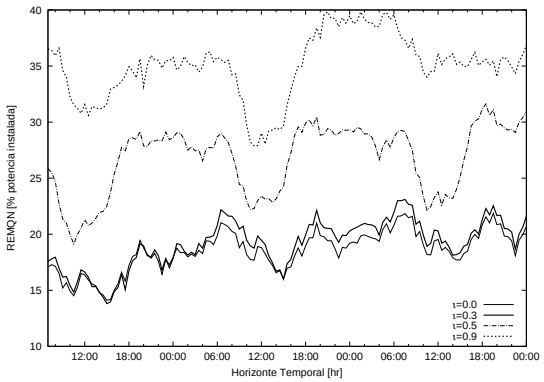

Figura H.16: *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MSE.*

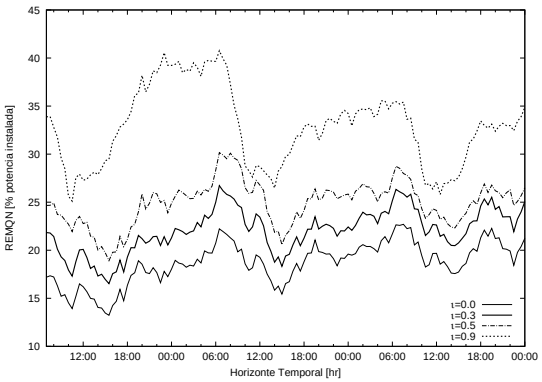

Figura H.17: *REMQN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MCC.*

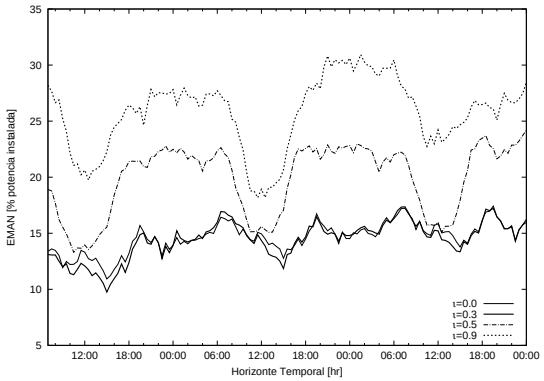

Figura H.18: *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MSE.*

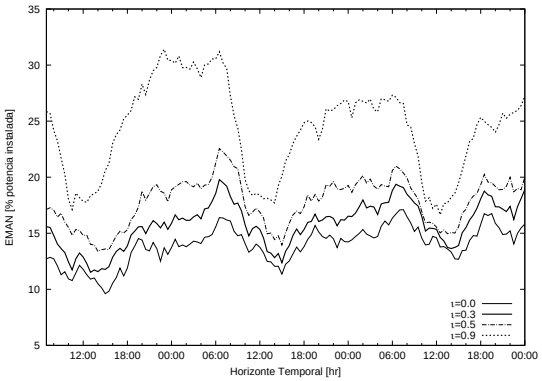

Figura H.19: *EMAN em função do horizonte temporal para diferentes valores do termo de momento para o Parque C com o critério MCC.*

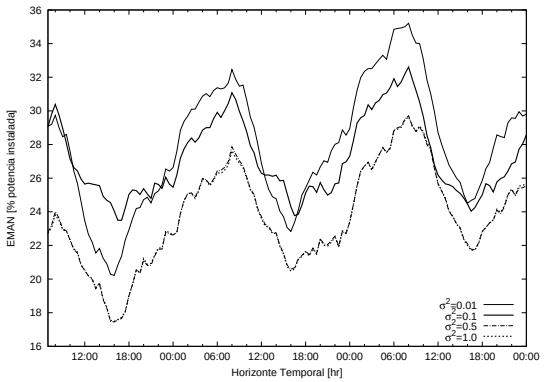

Figura H.20: *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque A com o critério MCC.*

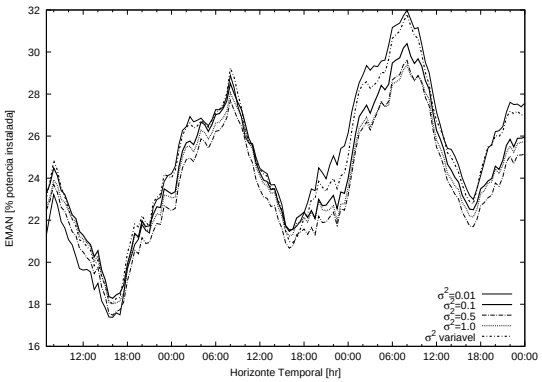

Figura H.21: *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque A com o critério MEEF.*

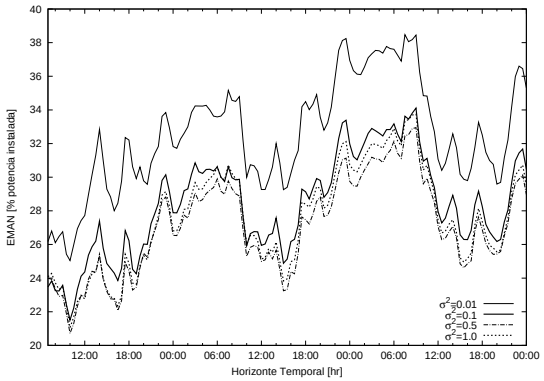

Figura H.22: *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MCC.*

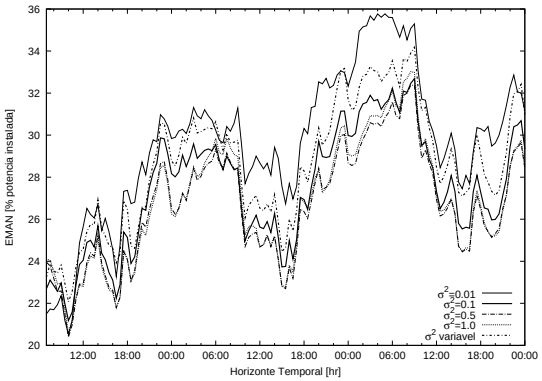

Figura H.23: *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MEEF.*

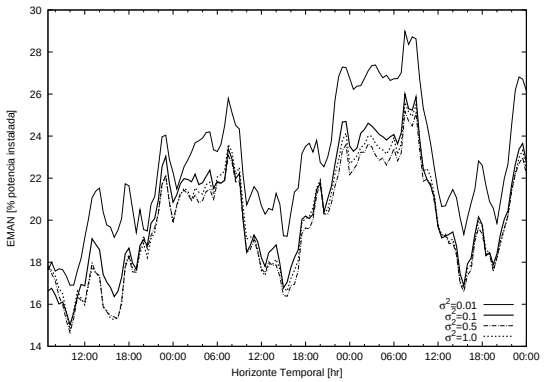

Figura H.24: *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MCC.*

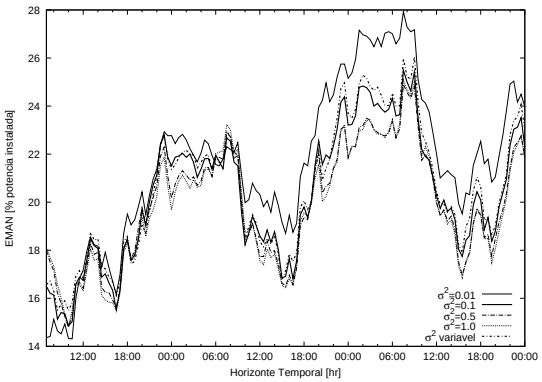

Figura H.25: *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque B com o critério MEEF.*

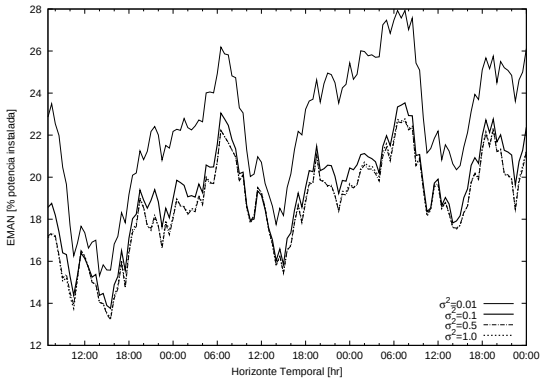

Figura H.26: *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MCC.*

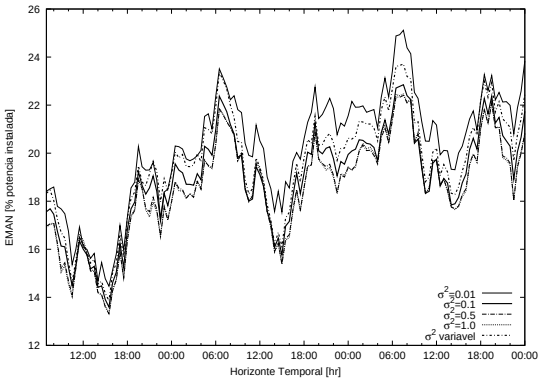

Figura H.27: *REMQN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MEEF.*

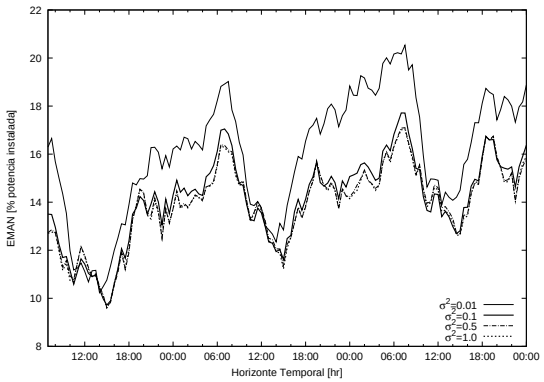

Figura H.28: *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MCC.*

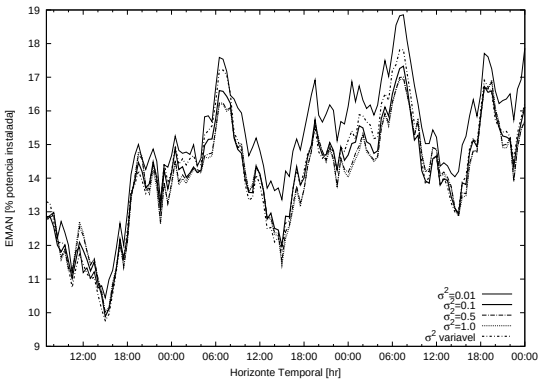

Figura H.29: *EMAN em função do horizonte temporal para diferentes valores da janela de Parzen para o Parque C com o critério MEEF.*

Apêndice I

Resultados do Treino *On Line* para 72 horas

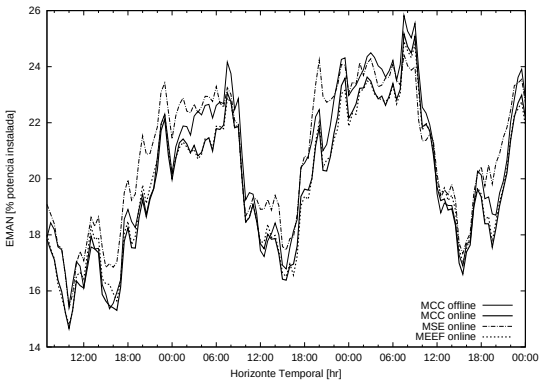

Figura I.1: *EMAN em função do horizonte temporal obtido com o treino on line para o parque B*

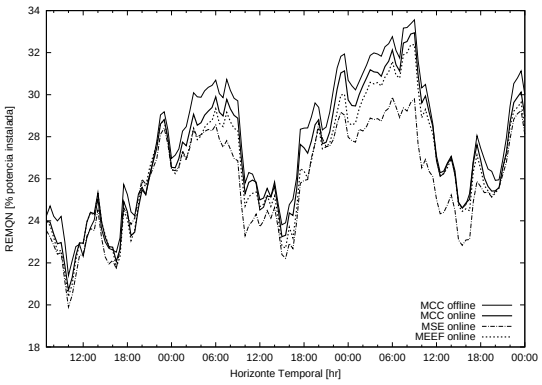

Figura I.2: *REMQN em função do horizonte temporal obtido com o treino on line para o Parque B.*

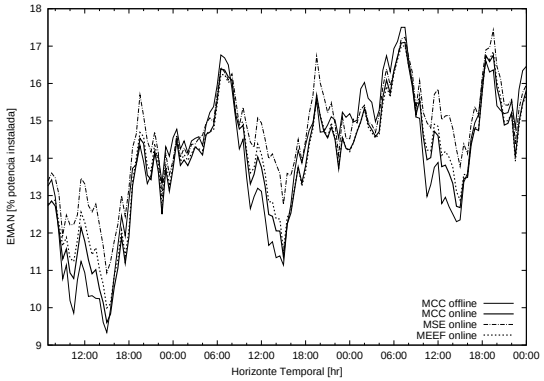

Figura I.3: *EMAN em função do horizonte temporal obtido com o treino on line para o Parque C.*

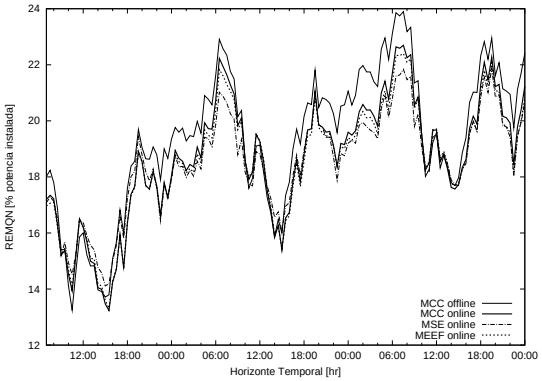

Figura I.4: *REMQN em função do horizonte temporal obtido com o treino on line para o Parque C.*

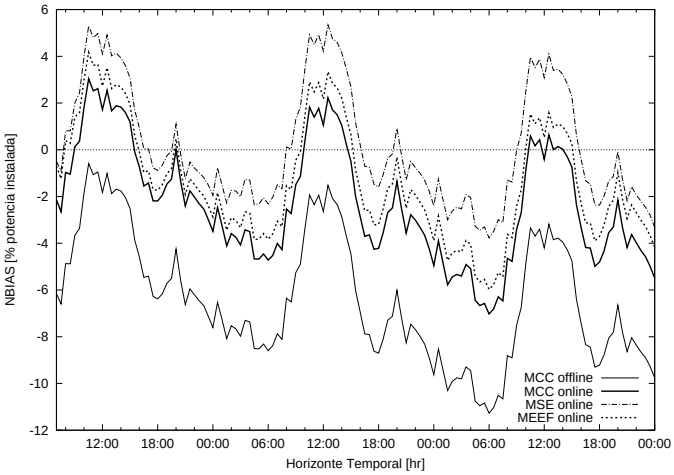

Figura I.5: *NBIAS em função do horizonte temporal obtido com o treino on line para o Parque C.*

Apêndice J

Artigos Publicados

- **Bessa, Ricardo, Miranda, Vladimiro e Gama, João (2008). Entropy and Correntropy against Minimum Square Error in Off-Line and On-Line 3-day ahead Wind Power Forecasting. IEEE Transactions on Power Systems (submetido)**

  Este artigo não se encontra em anexo por na altura da elaboração desta teste ainda se encontrar em fase de revisão. O resumo do artigo é o seguinte:
  This paper reports new results in adopting entropy concepts to the training of neural networks to perform wind power prediction as a function of wind characteristics (speed and direction) in wind parks connected to a power grid. Renyi's Entropy is combined with a Parzen Windows estimation of the error pdf to form the basis of two criteria (minimum Entropy and maximum Correntropy) under which neural networks are trained. Also, the merits of on-line training against off-line training are evaluated, as a consequence of concept drift in wind pattern behavior. The results are favourably compared with the traditional minimum square error (MSE) criterion. Real case examples for two distinct wind parks are presented.

- **Bessa, Ricardo, Miranda, Vladimiro e Gama, João (2008). Wind Power Forecasting With Entropy-Based Criteria Algorithms. Proceedings of PMAPS 2008 - 10th International Conference on Probabilistic Methods Applied to Power Systems, Puerto Rico, Maio**

Este artigo foi aceite e apresentado numa conferência internacional.

- **Bessa, Ricardo, Miranda, Vladimiro e Gama, João (2008). Improvement in Wind Power Forecasting Based on Information Entropy-Related Concepts. Proceedings of IEEE Power Engineering Society General Meeting, Pittsburgh, Julho**

Este artigo foi aceite e apresentado numa conferência internacional.

- **Bessa, Ricardo, Miranda, Vladimiro e Gama, João (2008). Wind Power Forecasting using streaming techniques. Proceedings of ISF 2008 - The 28th Annual International Symposium on Forecasting, Nice, Junho**

Este artigo foi aceite e apresentado numa conferência internacional. O resumo submetido é o seguinte:

This work is devoted to reporting further advancement in adopting entropy related concepts in training mappers for wind power prediction for a time horizon of 72 hours, rejecting the traditional approach of tuning systems based on the variance of the error distribution, i.e., based on the Mean Square Error (MSE). The rationale behind this approach lies in the fact that an entropy evaluation may allow one to extract more information from data that just using a variance-based method, because entropy takes in account all the moments of a probability distribution. Therefore, if errors are not Gaussian (as it seems to be the case in wind power prediction), one may hope to obtain error distributions with higher frequencies close to zero and approximating a Dirac function, which has minimum entropy. Furthermore, this approach will in principle be less sensitive to outliers and noise. Renyi's Entropy is combined with a Parzen Windows estimation of the error pdf to form the basis of the minimum error minimization (MEE) criterion under which neural networks are trained. Another criterion is the maximum correntropy criterion (MCC), which is a local crite-

rion of similarity and very useful for cases with non-zero mean, non-Gaussian, with large outliers. For handling concept drift we use incremental (online) train, when new instances arrive we process then and update the model. Two approaches were used for the online train: 1) a stochastic train of the neural network, when a new example arrives the error is backpropagated throw the neural network and the weights are updated; 2) a recursive estimation of the error entropy with a forgetting factor to deal with concept change. Real case examples for three distinct wind parks located in different regions of Portugal are presented. The input variables used on the cases study were: forecasted meteorological variable values obtained from a numerical weather prediction model (MM5); and electric power generation registers from the SCADA system of the wind farm.

Wind Power Forecasting With Entropy-Based Criteria Algorithms

Ricardo Bessa, Vladimiro Miranda, *Fellow*, *IEEE* and João Gama

Abstract-- This paper reports new results in adopting entropy concepts to the training of mappers such as neural networks to perform wind power prediction as a function of wind characteristics (mainly speed and direction) in wind parks connected to a power grid. Renyi's Entropy is combined with a Parzen Windows estimation of the error pdf to form the basis of three criteria (MEE, MCC and MEEF) under which neural networks are trained. The results are favourably compared with the traditional minimum square error (MSE) criterion. Real case examples for two distinct wind parks are presented.

***Index Terms*—Wind power forecasting, neural networks, correntropy, entropy, Parzen windows.**

I. Introduction

THE benefits of accurate wind power forecasting in electrical power systems are being increasingly recognised. Nowadays power systems with renewable energy generation must live with uncertainty both on supply and on demand, but a high uncertainty in wind generation for systems with large penetration of this source (the case in several European Union countries) may cause difficulties in system operation and in operation planning or conventional generation scheduling. The potential for reduced operation costs (via better scheduling decisions) of accepting high penetrations of wind energy [1] has become an attractive factor for the development of new models and tools.

Wind park owners also benefit from better wind power prediction. Short-term wind power forecasting tools are necessary to support a competitive participation of wind power in electricity markets against more predictable energy sources [2].

Short-term wind power forecasting employs mathematical models to map a function whose analytic form is unknown. The search for the best model parameters is usually made in a supervised training mode and we may witness a generalized adoption of a criterion for measuring the performance quality in terms of the prediction error: the MSE (Minimum Square Error).

Wind power prediction at a wind park level is contaminated with uncertainty from two origins: one inherited from the process of wind prediction and the other related with the complex wind park structure and terrain characteristics. Some issues must be underlined:

- The wind turbine conceptually has a non-linear transfer function. The power curve has a linear relation with to the cube of wind speed for small values, a rather steep slope for medium wind speeds and saturation for large wind speed.
  The wind speed must be above a cut-in value to get the turbine into operation. Small deviations of the prediction to the real value in the steep part of the curve are transferred to large differences between power prediction and real power. In contrast, a small deviation in the almost-linear and flat parts of the curve originates a small error in power prediction.
  The transformation of wind speed to wind power changes the statistical properties of the errors, the non-linearity of the power curve amplifies (increases by a factor 2 to 2.6) initial errors in the wind speed prediction according to its local derivative at a certain wind speed. Gaussian errors in wind speed prediction are no longer gaussian when going through this "filter".
- Measurements of wind speed and are usually taken in 10 minute intervals in many facilities. However, the wind speed signal includes components with much shorter wavelengths. Wind speed may fluctuate up or down within the interval of 10 minutes and therefore, one faces a problem of undersampling.
- There is usually no instant power measurement but average power in 10 minute periods (energy generated divided by the duration of the interval). Wind speed is measured every 10 minutes or calculated in terms of average speed within the 10 minute period. This combined with the non-linearity of the generator model is a cause for the introduction of additional uncertainty of non-gaussian nature.

We believe that mapping methods such as neural networks or fuzzy inference systems are well suited to emulate a speed/power relation. Accuracy is of the utmost importance and because average errors are significant (2 digits in percentage) any improvement will be economically valuable. This paper is devoted to reporting further advancement in adopting entropy related concepts in training mappers for wind power prediction, rejecting the traditional approach of tuning systems based on the variance of the error distribution, i.e., based on the Mean Square Error (MSE). The rationale behind this approach lies in the fact that an entropy evaluation may allow one to extract more information from data that just

R. J. Bessa is with INESC Porto, Instituto de Engenharia de Sistemas e Computadores do Porto, Portugal, and also with FEP, Faculty of Economy of the University of Porto, Portugal (email: rbessa@inescporto.pt).

V. Miranda is with INESC Porto and also with FEUP, Faculty of Engineering of the University of Porto, Portugal (email: vmiranda@inescporto.pt).

J. Gama is with INESC Porto LA/LIAAD, Laboratory of Artificial Intelligence and Decision Aid, and with FEP, Porto, Portugal (email: jgama@liaad.up.pt).

using a variance-based method, because entropy takes in account all the moments of a probability distribution. Therefore, if errors are not Gaussian (as it seems to be the case in wind power prediction), one may hope to obtain error distributions with higher frequencies close to zero and approximating a Dirac function, which has minimum entropy.

Furthermore, this approach will in principle be less sensitive to outliers and noise.

II. Entropy And Parzen Window Pdf Estimation

Renyi's entropy [3] of a discrete probability distribution P = ($p_1, p_2, \ldots, p_n$) is defined as

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

Renyi's entropy is a family of functions $H_{R\alpha}$ depending on a real parameter α. When α = 2, we have what is called quadratic entropy

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

This definition can be generalized for a continuous random variable Y with pdf $f_Y(z)$:

$$H_{R2} = -\log\int_{-\infty}^{+\infty} f_Y(z)^2 dz \tag{3}$$

The estimation of the pdf of data from a sample constituted by discrete points $y_i \in R^M$, i=1,…,N in a M-dimensional space, may be done by the Parzen Window method[4]. This technique uses a kernel function centered on each point; it looks at a point as being locally described by a probability density Dirac function, which is replaced or approximated by a continuous set whose density is represented by the kernel. If a Gaussian kernel is used, the expression of the estimation $\hat{f}_Y$ for the real pdf f_Y of a set of N points is a summation of individual contributions:

$$\hat{f}_Y(z) = \frac{1}{N}\sum_{i=1}^{N} G\left(z - y_i, \sigma^2 I\right) \tag{4}$$

Where G(.,.) is the Gaussian kernel and $\sigma^2 I$ is the covariance matrix (here assumed with independent and equal variances in all dimensions). In each dimension we have

$$G\left(z_k - y_{ik}, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(z_k - y_{ik})^2} \tag{5}$$

It is easy to understand that the "size" of the window, here defined by the value of σ, is important in obtaining a smoother or more "spiky" estimate for f_Y.

III. Information Theoretic Learning

A breakthrough has been achieved by signal processing researchers when they proposed combining Renyi's entropy definition with an estimate of a pdf by the Parzen window method [5][6] – this has been called Information Theoretic Learning. An entropy estimator for a discrete set of data points {y} is

$$H_{R2}(y) = -\log\int_{-\infty}^{+\infty} \hat{f}_Y(z)^2 dz = -\log V(y) \tag{6}$$

Where, using (6)

$$V(y) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{-\infty}^{+\infty} G\left(z - y_i, \sigma^2 I\right) G\left(z - y_j, \sigma^2 I\right) dz \tag{7}$$

In this expression we recognize the convolution of Gaussian functions, and the integral of two Gaussians with equal standard deviations is a Gaussian with twice the standard deviation. Then we have the following result:

$$V(y) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G\left(yi - yj, 2\sigma^2 I\right) \tag{8}$$

which allows the practical evaluation of entropy by simply calculating the Gaussian function values of the vector distances between pairs of samples. V(y) is called the information potential (IP) of the data set.

As the objective in supervised training is to minimize H of the errors *e*, one can instead maximize the information potential V. So, Max V(e) becomes the cost function for optimizing a mapper with minimum output entropy [7] – the MEE criterion, for Minimum Entropy Error.

The discovery of weights in a mapper may be done by applying a suitable optimization method that will discover the weights *w* that minimize the objective function

$$\min_{w} H_{R2}(e) \tag{9}$$

This can be achieved by the classical back-propagation algorithm [8] but in [9] we have applied an evolutionary algorithm to minimize entropy: EPSO, Evolutionary Particle Swarm Optimization. The entropy concept is independent of the training method but several avenues must be explored (backpropagation with a gradient search or other methods) in order to build efficient algorithms to optimize mappers because different objective functions may lead to distinct performance in different algorithms.

IV. Correntropy

Correntropy is a generalized similarity measure between two arbitrary scalar random variables X and Y defined by:

$$V_\sigma(X,Y) = E\left[k_\sigma(X - Y)\right] \tag{10}$$

where k_σ is the kernel function (usually gaussian).

In real problems, the joint pdf is unknown and we only have available a finite number of points. The estimator is:

$$\hat{V}_\sigma(X,Y) = \frac{1}{N}\sum_{i=1}^{N} k_\sigma(x_i - y_i) \tag{11}$$

Correntropy is directly related to the probability of how similar two random variables are in a neighbourhood of the joint space defined by the kernel bandwidth, and provides the probability density of the event p(X=Y). Using Parzen windows, the bandwidth controls the observation window in which the similarity is assessed but makes one unable to assess similarity in the whole joint space.

In [11] one may find a discussion on the properties of

correntropy. It is proved that a measure CIM(X,Y) = (k(0)-V(X,Y) related with correntropy satisfies all the properties of a metric. CIM may be divided in three different regions: when the error is close to zero CIM is equivalent to L2 norm; when the error grows CIM becomes a L1 norm; when the error is very large CIM becomes a L0 norm, the metric saturates and becomes very insensitive to large errors. This property shows the robustness of CIM and the importance of kernel bandwidth. A small kernel size leads to a small Euclidean zone while a large kernel size will increase the Euclidean region where the metric behaves like the MSE criterion.

When we use correntropy to train adaptive systems, we actually make the system output close to the desired response in the CIM sense if we maximize the correntropy of the error distribution – MCC criterion. We can use the MCC as a new performance function, with the advantage over MSE of being a local criterion of similarity and very useful for cases with non-zero mean, non-Gaussian, with large outliers. It does not require the computing effort of MEE but tends to minimize entropy because it tends to maximize the pdf value at the origin.

The MCC criterion becomes

$$\max_{w} J(e) = \frac{1}{N}\sum_{i=1}^{N} K_{\sigma}\left(g(w,x_i) - T\right) \qquad (12)$$

where g(w,x) represents the mapper producing y = g(w,x) responses from input x as a function of weights w, and T represents the target values. This criterion may be differentiated and a backpropagation algorithm used to train a neural network (NN).

V. Hybrid Approach MEEF

MEE is an exact criterion in terms of entropy concept while MCC is only an approximation. However, MEE is much more consuming in computing effort. Also, MEE has degenerate minima because it is insensitive to the mean of the error. There are methods to deal with the problem. The first method is to correct the result by properly modifying the output bias of the neural network to yield zero mean error over the training data set just after training ends. The other way is to add a so-called MCC term as a function of the errors to the MEE cost function (a function of the difference of errors) as

$$J(e) = \gamma\sum_{i=1}^{N} K_{\sigma\sqrt{2}}(e_i) + (1-\gamma)\sum_{j=1}^{N}\sum_{i=1}^{N} K_{\sigma\sqrt{2}}\left(e_j - e_i\right) \qquad (13)$$

where γ is a weighting constant between 0 and 1 and K is the kernel function. This cost function is called Minimum Error Entropy with Fiducial points (MEEF) [10]. The MEE term minimizes the error entropy and the MCC term anchors the mean of error at zero.

VI. Training Procedure Optimization

In this paper we are training neural networks in order to produce wind power predictions. For this purpose, the sample entropy estimation is very heavy computationally. We are usually working with very large data sets and using an algorithm with complexity $O(N^2)$ by adopting MEE or MEEF we could easily have processing times of hours in a modern fast PC. Therefore, following the analysis performed in [8] we used instead of the classic backpropagation, the iRPROP[12] algorithm, which with a variable learning rate can achieve a faster convergence and maintaining a good performance. Other strategy that can lead to improved results is using an adaptive kernel size during training [13]. In each test case we start with a large kernel size and during the adaptation gradually and slowly decrease it toward 10% of the initial size with the following empirical rule:

$$\begin{aligned} &\text{IF } \left|Vnorm_t - Vnorm_{t-1}\right| < 0.002 \text{ and } Vnorm_t > Vnorm_{t-1} \\ &\text{THEN } \sigma_t = 0.95\cdot\sigma_{t-1} \end{aligned} \qquad (14)$$

where V_{norm} is the normalized information potential relative to $V_{\sigma}(0)$.

For the MCC criteria the same logic was followed but instead of annealing the kernel size, we first train with MSE criterion during the first epochs, and switch the criterion to MCC during the remaining epochs. Finally, to reduce the large processing time of batch training we use the batch-sequential approach tested in [14]: the algorithm combines the two methods and takes all the advantages of each. The batch-sequential algorithm consists in randomly dividing the training set in several groups with an equal number of samples (size L) that are presented to the algorithm in a sequential way. In each group we apply the batch training mode and a new subset division is performed in each epoch of the training process. The results of using this method is the MEE complexity reduction from $O(N^2)$ to $O([N/L]/^2)$, the fast convergence and reduction of the probability of the algorithm getting trapped in local minima.

VII. Comparing ITL Criteria With MSE

In this section we present results for training of neural networks in real wind power forecasting problems, comparing the performance of MSE and three ITL inspired criteria (MCC, MEE and MEEF).

The real test cases refer to two wind parks. Wind park A has a rated power of 21600 kW and comprises 12 equal wind turbines of 1.8 MW. The wind park is situated in the north of Portugal, in a complex mountainous region 65 km far from the Atlantic Ocean. Wind park B has a rated power of 16200 kW and comprises 17 equal wind turbines of 0.6 MW and 3 turbines of 2 MW. The wind park is situated in the center of Portugal, in a complex mountainous region 68 km far from the Atlantic Ocean and 200 km far from the wind park A.

The data collected from the wind park included SCADA registers with mean electric power delivered by the wind park into the substation connecting it to the electric power network.

Other model's input variables include forecasts generated from a NWP (numerical weather prediction) model including the well known MM5 mesoscale model (see http://www.mmm.ucar.edu/mm5/) for mean wind speed and wind direction, for a reference point in the wind park with forecasting horizons ranging from 0 to 24 hours.

Fig. 1 displays one year of the NWP/MM5 prediction errors against real wind speed values measured by the metering station at wind park A.

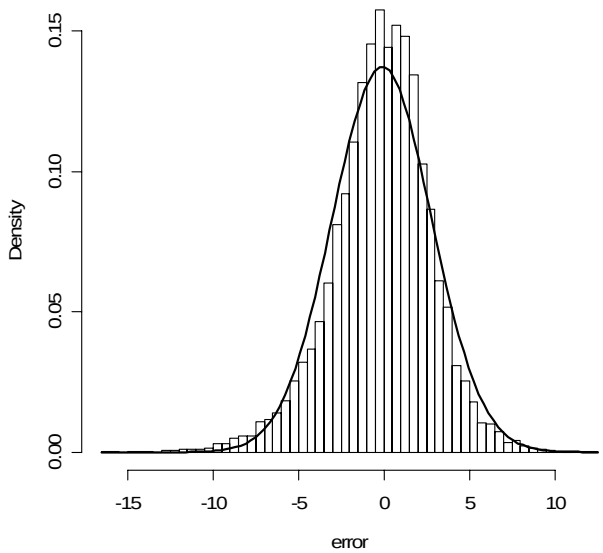

Fig. 1. MM5 wind speed prediction error in wind park A and the normal distribution (curve) with mean of -0.10 and standard deviation of 2.90

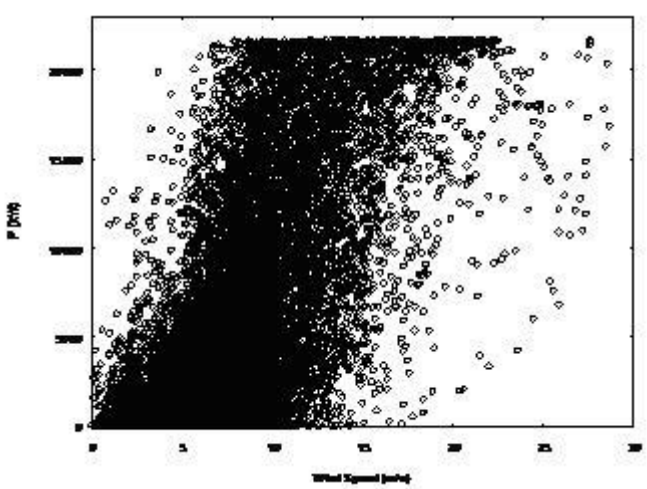

Fig. 2. Wind speed prediction(x axis) against power output measured (y axis) in 30 minute intervals.

It might seem that the NWP error could be approximated by a Gaussian distribution, but if we apply the *Kolmogorov-Smirnov* test to the error data then the null hypothesis (error with normal distribution) is rejected. Even if the NWP error is Gaussian, the non-linearity of the power curve changes the statistical properties of the distribution of deviations between forecasts and measurement.

Fig. 2 displays one year of measurements of power output against wind speed predictions by the MM5 model with an interleave period of 30 minutes. From Fig. 2 it's clear the utility of a local criterion like MCC and the noisy robustness of MEE. We have variables with noisy and a very large number of outliers, some duo to NWP forecast errors and others duo to wind park maintenance.

For example, we have points where the wind speed is above wind speed cut-in but the wind form production is zero or a low value, the cause is some wind park turbine maintenance or malfunction.

These elements show that the application of the ITL criteria have the potential to produce a better mapper than the application of the classical MSE. To demonstrate this, we have trained a feed forward MLP neural network with only one hidden layer comprising 7 neurons, using a hyperbolic tangent activation function.

For the iRPROP training algorithm we chose the standard parameter values for the increase and decrease factor $\eta^+=1.2$, $\eta^-=0.5$ because they actually produce the best results independent of the learning problem [12]. The remaining parameters of iRPROP are the range of the weight update values $\Delta_{min}=0$, $\Delta_{max}=50$; and the initial weight update value $\Delta_0=0.0125$. The tuning of these parameters is not critical.

The initial kernel size for MEE is 0.3, and the kernel window size for MCC is 0.02. Also the value of these parameters is not critical. In the case of the MEEF criterion the same values for kernel size and same strategy were used.

This training methodology is very useful for wind forecasting as we can train a neural network for any wind park without concerning about tuning parameters. The stopping criterion for each cost function is the early stopping and the maximal number of iterations (we set this value to 150). We stopped the training when the validation error started to increase. The batch-sequential subsets length was 300 points.

To validate the results we adopted a Monte Carlo procedure running 25 simulations, generating randomly weights in each case by a uniform distribution in the interval [-1,1]. The final conclusions derive from the average of all simulations. When comparing criteria, the same weights were used in all cases.

The inputs of the MLP are NWP meteorological forecasted values: mean wind speed values, mean wind direction values. Duo to the cyclic character of the wind direction, this variable comprised two components, i.e. the sine and cosine components. The model totalled 3 input variables, and they were standardized using the min-max method.

The available data of the two wind parks were divided into three data sets. Table I and II shows the statistical characteristics of the mean electric power series data of the two wind parks. The training set is constituted of data from January to March on 2005; the validation set of data from April; the test set of data from May to June on 2005.

TABLE I
CHARACTERISTICS OF THE WIND PARK A DATA

	Training	**Validation**	**Test**
Mean (kW)	7947.06	2567.81	3986.83
Variance (kW^2)	53959535	12806109	23704424
Skewness	0.576	2.154	1.527
Minimum (kW)	-74.33	-68.17	-73.50
Maximum (kW)	21662.70	19177.50	21439.00
Number	5184	1267	2204

TABLE I
CHARACTERISTICS OF THE WIND PARK B DATA

	Training	Validation	Test
Mean (kW)	6184.00	2567.81	3986.83
Variance (kW^2)	27276937	29329712	18641596
Skewness	0.459	2.154	1.527
Minimum (kW)	-60.50	-68.17	-73.50
Maximum (kW)	16142.67	19177.50	21439.00
Number	4893	1267	2204

The training set is composed of data from July to September 2005; the validation set, from October 2005; the test set, from November to December 2005.The negative values indicate electric power consumption of the wind park.

Fig. 3 and 4 compare the pdf of errors obtained with MEE and MEEF in batch-sequential training for wind parks A and B. We see that for case A the pdf of errors is more centred in zero error when the NN is trained with MEEF. It's clear that for wind power forecasting the use of fiducial points as reference for the pfd peak is a better method than the zero mean error. One of the drawbacks in the MEE criteria is the insensitivity to the mean and in this case we have error distributions with large skewness and kurtosis. For case B the results for each criterion are similar.

One drawback of these models was the long processing time of the training of the algorithm but with the training process explained in section VI we have reduced the complexity of the algorithm and the number of training epochs until convergence.

With the classical batch backpropagation training, the processing time for the MEE model in the case of wind park A was about 3 hours for only one simulation, with a Pentium IV 3.0 GHz and 1 GB of RAM. With the optimization of the training, the processing time of one simulation became 10 minutes, with less computational complexity and number of epochs.

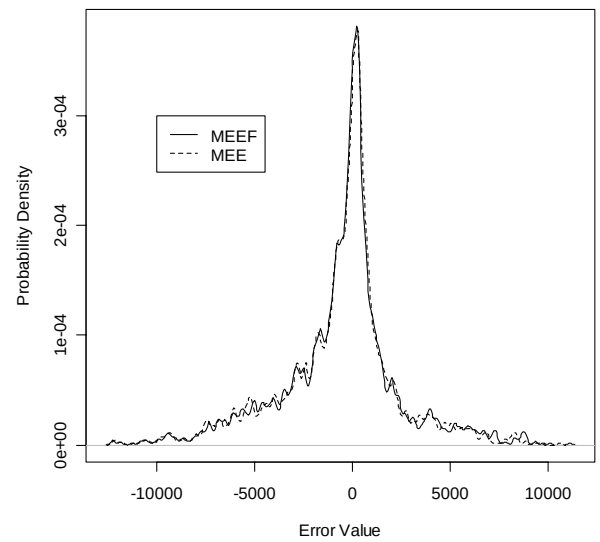

Fig. 4. Comparison of the error pdf generated by NN trained with MEEF (dotted) and MEE (solid) criteria for wind park B

Fig. 5 and 6 compare the pdf of errors obtained with MEEF, MCC and MSE in batch-sequential training for wind park A and B respectively.

From these figures we confirm that the prediction errors are not Gaussian: it was possible to obtain narrower pdf with entropy and correntropy criteria than with a variance-based criterion. In fact, if the error pdf were gaussian, the MSE criterion would perform as well as an entropy-based criterion, but this was not the case.

Therefore, in agreement with the theory, it was possible to design a mapper that would produce a predictor with a higher frequency of errors close to zero, a characteristic associated with smaller entropy of the pdf. In general, this is desirable. For wind park A the MCC presents higher probability density near zero than MEEF, but we have larger errors in the negative part of the PDF.

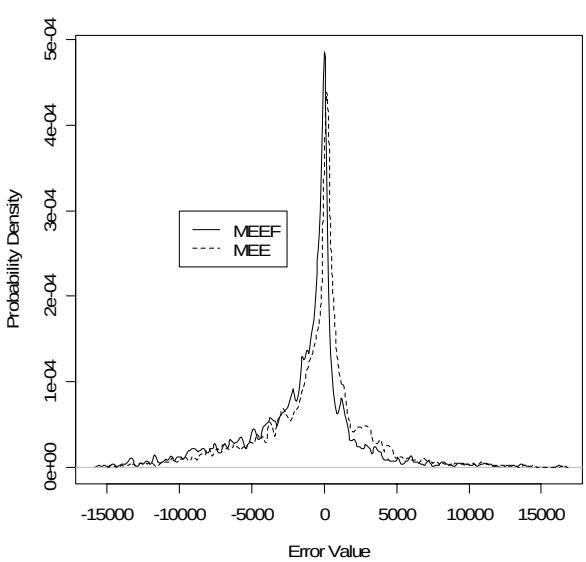

Fig. 3. Comparison of the error pdf generated by NN trained with MEEF (dotted) and MEE (solid) criteria for wind park A

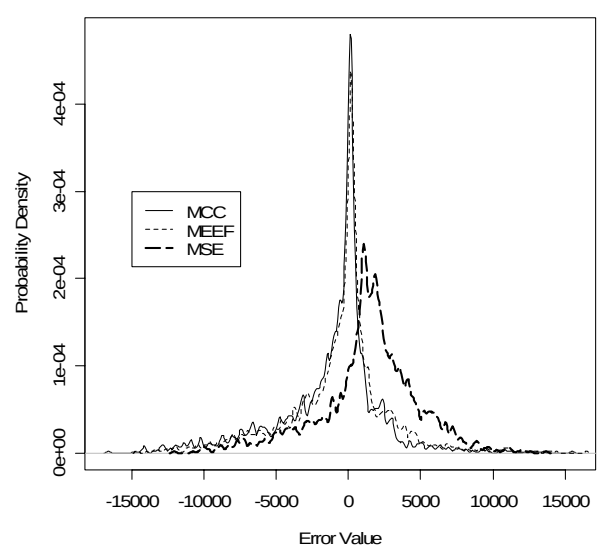

Fig. 5. Comparison of the error pdf generated by NN trained with MEE (dotted), MCC (solid) and MSE (dotted line width) criteria for wind park A

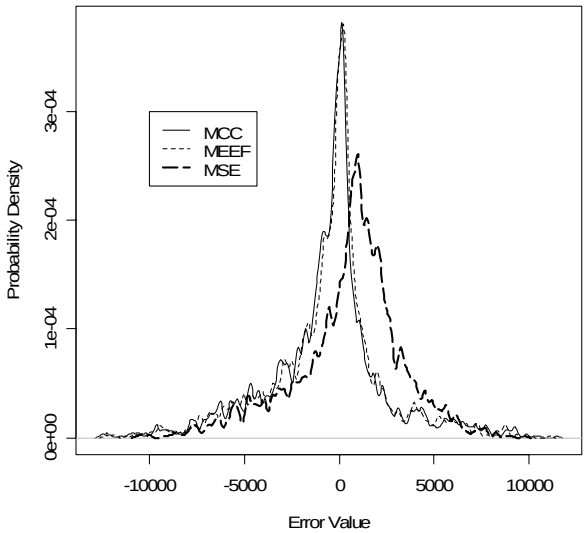

Fig. 6. Comparison of the error pdf generated by NN trained with MEE (dotted), MCC (solid) and MSE (dotted line width) criteria for wind park B

Table III shows the Renyi's Entropy estimated with kernel size of 100, for the error distributions of each criteria and wind park.

TABLE III
RENYI'S ENTROPY OF THE ERROR PDF PER CRITERION

	Wind Park A	Wind Park B
MSE	3.931	3.873
MCC	3.751	3.783
MEE	3.747	3.779
MEEF	3.767	3.789

Table III shows, as expected, that the MEE criterion produces a solution with minimal error entropy, but it's interesting to note that the MCC also minimizes the error entropy with excellent performance.

Fig. 7 shows the forecasted values for the mean 30-minute electric power from the wind park obtained with three criteria (2 ITL criteria and MSE), when compared to the real value registered in the park SCADA. It's clear that the ITL criteria made a better fit and followed better the SCADA measured. It's logical that in some points the MSE produced small errors, but as we see in Fig. 7 the ITL produces a more number of errors close to zero.

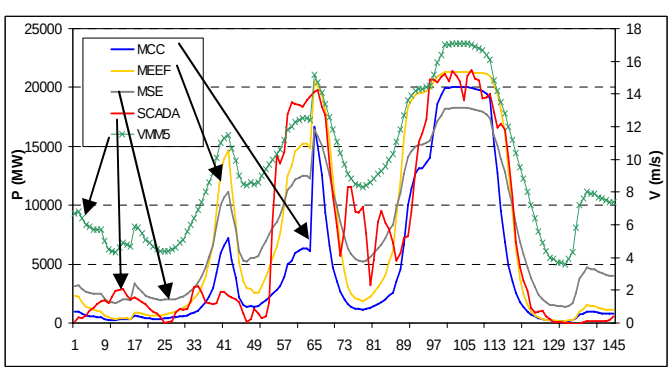

Fig. 7. Comparison among predictions by three criteria on 2005 May 26, 27 and 28 for wind park A. The graph also displays the wind speed prediction provided by the MM5 model (wind speed scale on the right)

It is important to emphasize the fact that we used explanatory variables that are predictions from a NWP model. So, the divergence between the predictions and the real value isn't upsetting. In some cases we can see that the wind speed has a peak which the neural network follows, but the wind park is producing much less; this may due to some turbines being in maintenance, and in these cases the MCC criterion lead to a smaller prediction value because of its robustness to outliers.

VIII. CONCLUSIONS

Wind power prediction for wind parks displays obvious non-Gaussian characteristics in the error probability distributions and therefore training input-output systems (mappers) to perform wind power prediction gains from moving away from criteria based on Variance (such as MSE) and instead adopting criteria based on Entropy concepts. For this purpose, a direct use of an entropy measure (the MEE criterion) is possible, but an approximation such as the MCC criterion measuring correntropy has clear computing advantages. The combination of these two criteria is a robust alternative; we can get all the advantages of the two criteria. Also optimization procedure techniques are needed to deal with the large computing effort of calculating entropy for large data sets.

This paper also contributes with evidence from real cases supporting that Entropy-based criteria perform better than a variance-based criterion, based on studies performed with real data from wind parks in Europe and comparisons between NN trained under the three competing criteria – MSE, MEEF and MCC.

VI. REFERENCES

[1] Mark L. Ahlstrom, Robert M. Zavadil, "The Role of Wind Forecasting in Grid Operations & Reliability" *IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific*, China, pp. 1-5, 2005.

[2] P. Pinson, C. Chevallier, G. Kariniotakis, "Optimizing Benefits from Wind Power Participation in Electricity Markets using Wind Power Forecasting Embedded with Uncertainty Management Tools", *European Wind Energy Conference EWEC 2004*, London, UK, 22-25 Nov. 2004.

[3] A. Renyi, "Some Fundamental Questions of Information Theory", *Selected Papers of Alfred Renyi*, vol 2, pp. 526-552, Akademia Kiado, Budapest, 1976.

[4] [10] E. Parzen, "On the estimation of a probability density function an the mode", Annals Math Statistics, vol 33, 1962

[5] 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.

[6] 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.

[7] 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.

[8] R.A. Morejon, J.C. Principe, "Advanced search algorithms for information-theoretic learning with kernel-based estimators", *IEEE Transactions on Neural Networks*, vol. 15(4), pp. 874 - 884, July 2004

[9] V. Miranda, C. Cerqueira, C. Monteiro, "Training a FIS with EPSO under an Entropy Criterion for Wind Power prediction", *International*

Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2006, Stockholm, 11-15 June 2006.

[10] Liu Weifeng, P.P. Pokharel, J.C. Principe, "Error Entropy, Correntropy and M-Estimation", *16th IEEE Signal Processing Society Workshop on Machine learning for Signal Processing*, pp. 179–184, 2006

[11] W. Liu, P. Pokharel, J. Principe, "Correntropy: properties and applications in non-Gaussian signal processing", *Accepted in IEEE Transactions on Signal Processing*, 2007.

[12] C. Igel, M. Husken, "Improving the Rprop learning algorithm", *Proceedings of the Second International ICSC Symposium on Neural Computation (NC 2000)*, ICSC Academic Press, pp.115-121, 2000

[13] D. Erdogmus and J. C. Principe, "Generalized information potential criterion for adaptive system training," *IEEE Transactions on Neural Networks*, vol. 13, pp. 1035–1044, Sept. 2002.

[14] Jorge M. Santos, J. Marques de Sá, Luís A. Alexandre, "Batch-sequential algorithm for neural networks trained with entropic criteria", *Proceedings of the 15th Int. Conf. on Artificial Neural Networks*, pp. 91–96, 2005

VII. Biographies

Ricardo Bessa received his Licenciado degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 2005 in Electrical and Computer Engineering. He then joined INESC Porto and is a researcher in its Power Systems Unit while completing a M.Sc. degree in the Faculty of Economy of the University of Porto (FEP).

Vladimiro Miranda (M'90, SM'04, F'05) received his Licenciado, Ph.D. and Agregado degrees from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 1977, 1982 and 1991, all in Electrical Engineering. In 1981 he joined FEUP and currently holds the position of Professor Catedrático. He is also currently a Director of INESC Porto, an advanced research institute in Portugal. He has authored many papers and been responsible for many projects in areas related with the application of Computational Intelligence to Power Systems.

João Gama received his Licenciado degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in Electrical and Computer Engineering.. In 2000 he received his Ph.D. degree in Computer Science from the Faculty of Sciences of the same University. He joined the Faculty of Economy (FEP) where he holds the position of Assistant Professor, He is also a senior researcher at LIAAD (Laboratory for Artificial Intelligence and Decision Aid), a group belonging to INESC Porto LA. He has worked in projects and authored papers in areas related to machine learning, data streams and adaptive learning systems and is a member of the editorial board of international journals in his area of expertise.

Improvement in Wind Power Forecasting Based on Information Entropy-Related Concepts

Ricardo Bessa, Vladimiro Miranda, *Fellow, IEEE* and João Gama

***Abstract* — This paper reports new results in adopting entropy concepts to the training of mappers such as neural networks to perform wind power prediction as a function of wind characteristics (mainly speed and direction) in wind parks connected to a power grid. It also addresses the differences relevant to power system operation between off-line and on-line training of neural networks. Real case examples are presented.**

I. Introduction

Wind generation has become a relevant component of many power systems in Europe and will most certainly assume such role in the US and other countries in the near future. For instance, in Iberia (Portugal + Spain) the global peak is currently just under 52 GW and the installed capacity of wind generation will likely be of some 25 GW in the end of 2010. When the wind penetration becomes important, system operators must rely on forecasts for wind power generation in order to adequately plan the power dispatch in short and medium term time horizons, with the objective of obtaining economic gains (and security) from improved accuracy in predictions [1].

Wind power is generated usually in wind parks connected to the grid where a number of generators exhibit a geographical dispersion that may be sensitive to wind direction and where each generator may inject a different power value in the grid due to irregularities in wind speed. Nowadays, generators of above 2.5 MW each unit are currently installed in on-shore facilities and larger units are considered for off-shore generation.

One must not mix concepts: system operators require wind power predictions and not wind speed predictions. However, a wind prediction phase is usually mandatory prior to forecasting the wind power output from a wind park.

In *Prediktor*[2], for instance, from a model of fluid dynamics equations one obtains wind speed as seen by a wind park and then derive power from theoretical power curves. Another model (*eWind* [3]) also forecasts first wind speed and only then wind power. Pure statistical methods [4] and methods based on computational intelligence techniques [5] also use this two-step approach.

Wind prediction carries uncertainty, which is combined with the uncertainty associated with the power generation by the complex wind park structure to generate wind power prediction uncertainty. Two facts must be underlined:

- Measurements of wind speed and are usually taken in intervals of 10 minutes in many facilities. However, wind speed presents variations that indicate a presence of frequencies with much shorter wavelengths (high frequencies). There is no power measurement but, in fact, average power in 10 minute periods (energy generated divided by the duration of the measurement interval) – but there is presently no integration of wind speed, so wind may fluctuate up or down within the interval of 10 minutes. Therefore, one faces a problem of undersampling when dealing with wind time series and no one can guarantee that errors in power prediction will display a Gaussian distribution.
- The theoretical power curve of wind generators is a squashing function. Therefore, even if the wind prediction models have Gaussian error distributions, the error distribution associated with wind power will not be Gaussian.

We believe that mapping methods such as neural networks or fuzzy inference systems are well suited to emulate a speed/power relation. Accuracy is of the utmost importance and because average errors are significant (2 digits in percentage) any improvement will be economically valuable.

This paper is devoted to reporting further advancement in adopting entropy related concepts in training mappers for wind power prediction, rejecting the traditional approach of tuning systems based on the variance of the error distribution, i.e., based on the Mean Square Error (MSE). The rationale behind this approach lies in the fact that an entropy evaluation may allow one to extract more information from data that just using a variance-based method, because entropy takes in account all the moments of a probability distribution. Therefore, if errors are not Gaussian (as it seems to be the case in wind power prediction), one may hope to obtain error distributions with higher frequencies close to zero and approximating a Dirac function (which has minimum entropy. Furthermore, this approach will in principle be less sensitive to outliers.

In a previous paper [6] some preliminary results were

This work was in part developed under Project POSC/EEA-ESE/60980/2004, FCT – Foundation for Science and Technology, Portugal.

Ricardo Bessa is with INESC Porto, Instituto de Engenharia de Sistemas e Computadores do Porto, Portugal (email: rbessa@inescporto.pt).

V. Miranda is with INESC Porto and also with FEUP, Faculty of Engineering of the University of Porto, Portugal (email: vmiranda@inescporto.pt).

J. Gama is with INESC Porto LA/LIAAD, Laboratory of Artificial Intelligence and Decision Aid, and with FEP, Faculty of Economy of the University of Porto, Portugal (email: jgama@liaad.up.pt).

presented in this line. A 0-order Takagi-Sugeno FIS (Fuzzy Inference System, see the Appendix) was trained to predict wind power output from a real wind park located in northern Portugal, having a total installed capacity of about 40 MW with generators having each a capacity of about 1 MW located on the top of a mountain range. Training and test sets were organized from measurements of wind speed, wind direction and power output of the wind park, collected every ten minutes from January 1, 2004 to February 20, 2005.

The novelty in this approach is the use of a performance criterion based on minimizing Renyi's Quadratic Entropy of the error distribution. This requires the estimation of the probability distribution function (pdf) of the errors which is done using the concept of Parzen windows.

II. Entropy and Parzen Window pdf Estimation

Renyi's entropy [6] of a discrete probability distribution P = ($p_1, p_2, \ldots, p_n$) is defined as

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

Renyi's entropy is a family of functions $H_{R\alpha}$ depending on a real parameter α.When $\alpha = 2$, we have what is called quadratic entropy

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

This definition can be generalized for a continuous random variable Y with pdf $f_Y(z)$:

$$H_{R2} = -\log\int_{-\infty}^{+\infty} f_Y{}^2(z)dz \tag{3}$$

The estimation of the pdf of data from a sample constituted by discrete points $\mathbf{y}_i \in R^M$, i = 1,…,N in a M-dimensional space, may be done by the Parzen window method [7]. This technique uses a kernel function centered on each point; it looks at a point as being locally described by a probability density Dirac function, which is replaced or approximated by a continuous set whose density is represented by the kernel. If a Gaussian kernel is used, the expression of the estimation $\hat{f}_Y$ for the real pdf f_Y of a set of N points is a summation of individual contributions

$$\hat{f}_Y(\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} G(\mathbf{z}-\mathbf{y}_i, \sigma^2\mathbf{I}) \tag{4}$$

where G(.,.) is the Gaussian kernel and $\sigma^2\mathbf{I}$ is the covariance matrix (here assumed with independent and equal variances in all dimensions). In each dimension, we have

$$G(z_k - y_{ik}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(z_k - y_{ik})^2} \tag{5}$$

It is easy to understand that the "size" of the window, here defined by the value of σ, is important in obtaining a smoother or more "spiky" estimate for f_Y.

III. Information Theoretic Learning

A breakthrough has been achieved by signal processing researchers when they proposed combining Renyi's entropy definition with an estimate of a pdf by the Parzen window method [8][9] – this has been called Information Theoretic Learning. An entropy estimator for a discrete set of data points {**y**} is

$$H_{R2}(\mathbf{y}) = -\log\int_{-\infty}^{+\infty} \hat{f}_Y{}^2(\mathbf{z})d\mathbf{z} = -\log V(\mathbf{y}) \tag{6}$$

where, using (6),

$$V(\mathbf{y}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{-\infty}^{+\infty} G(\mathbf{z}-\mathbf{y}_i, \sigma^2\mathbf{I})G(\mathbf{z}-\mathbf{y}_i, \sigma^2\mathbf{I})d\mathbf{z} \tag{7}$$

In this expression we recognize the convolution of Gaussian functions, which has the following consequence:

$$V(\mathbf{y}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G(\mathbf{y}_i - \mathbf{y}_j, 2\sigma^2\mathbf{I}) \tag{8}$$

which allows the practical evaluation of entropy by simply calculating the Gaussian function values of the vector distances between pairs of samples.

V(**y**) is called the *information potential (IP)* of the data set. As the objective in supervised training is to minimize H of the errors $\boldsymbol{\varepsilon}$, one can instead maximize the information potential V. So, Max V($\boldsymbol{\varepsilon}$) becomes the cost function for optimizing a mapper with minimum output entropy [10] – the MEE criterion, for Minimum Entropy Error.

The discovery of weights in a mapper may be done by applying a suitable optimization method that will discover the weights **w** that minimize the objective function

$$\min \quad H_{R2}(\mathbf{w}) \tag{9}$$

This can be achieved by the classical back-propagation algorithm [11] but in [6] we have applied an evolutionary algorithm to minimize entropy: EPSO, Evolutionary Particle Swarm Optimization. The entropy concept is independent of the training method but several avenues must be explored (backpropagation with a gradient search or other methods) in order to build efficient algorithms to optimize mappers because different objective functions may lead to distinct performance in different algorithms.

IV. Discussion of Previous Results

In [6] we have trained a 0-order Takagi-Sugeno FIS based on data organized in two input variables (wind speed, direction angle) and one output (10-minute average wind park power output). Fig. 1 illustrates the anisotropic characteristic of wind direction.

A 10 rule system was formed and trained optimizing both a classical MSE criterion and a Minimum Entropy criterion. The result has been promising, because we could verify that the distribution of errors both in the training and the test sets was sharper for the system trained under the Minimum Entropy criterion, as seen in Fig. 2.

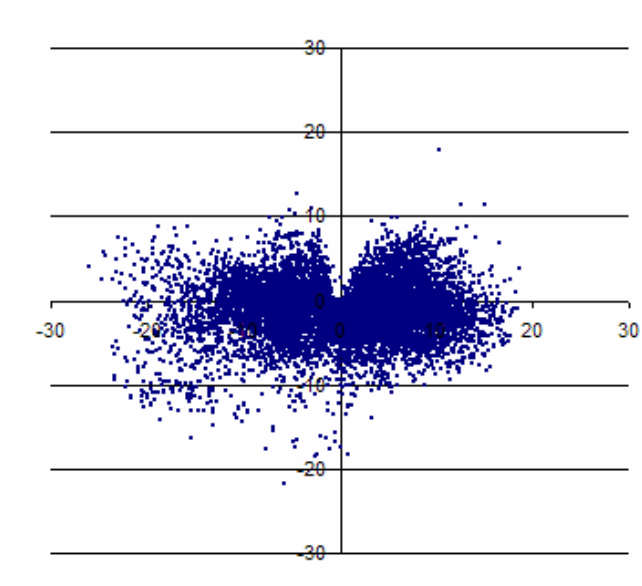

Fig. 1. Distribution of wind speed (in m/s) and direction at the measuring location near the wind park [6]. Each point is the tip of a vector whose size is proportional to wind speed and angle is related to wind direction. The distribution of winds may be explained not only from global dominant winds but also from the local characteristics of the terrain.

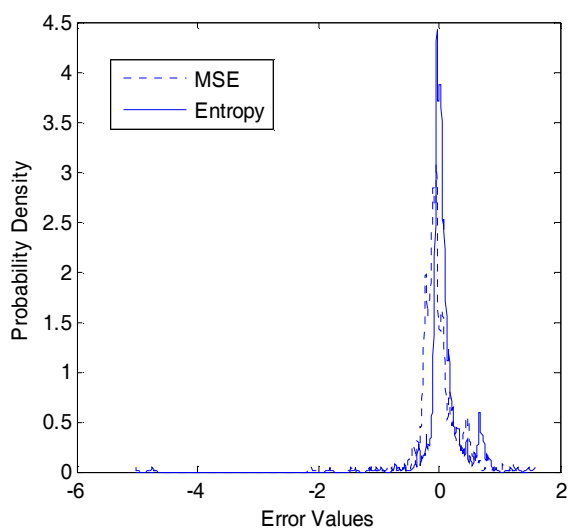

Fig. 2. Probability density functions of TS-FIS prediction errors, for both models (MSE and MEE), estimated with Parzen windows, for the training set [6]. Plots generated with Parzen windows with $\sigma = 0.01$.

We have concluded therefore that, in fact, the prediction errors were not Gaussian and therefore, in agreement with the theory, it was possible to design a mapper that would produce a predictor with a higher frequency of errors close to zero, a characteristic associated with smaller entropy of the pdf. In general, this is desirable.

An inconvenience of adopting the ITL paradigm and using (9) as the performance criterion is however the following: the entropy estimation given by equation (6) and (7) requires a much larger computing effort that the simple calculation of MSE. In fact, MSE depends on determining the square of simple errors from a set while Entropy evaluation depends on calculating the Gaussian of the difference between all pairs of errors, so the effort grows quadratically with the dimension of the sets instead of linearly as with MSE.

This could be seen as not too much troubling for a procedure involving off-line training. However, wind has another complex characteristic: it displays what is called "concept drift", meaning that the probabilistic model that could be used to explain its behavior changes over time – so a model trained with a batch of data may become outdated and require re-training at some point. Therefore, when predictions must serve the operation of a power system, it is much better to adopt a strategy of on-line training – a non-parametric model such as one based on entropy is a plus, but the time consuming trait of the entropy calculation becomes a drawback. We have therefore experimented with a new concept.

V. Correntropy

Correntropy V_σ is a generalized similarity measure between two arbitrary scalar random variables X and Y defined by:

$$V_\sigma(X,Y) = E\left[k_\sigma(X-Y)\right] \tag{10}$$

where k_σ is the Gaussian kernel function. In real problems, the joint PDF is unknown and we only have available a finite number of points, the sample estimator is:

$$\hat{V}_\sigma(X,Y) = \frac{1}{N}\sum_{i=1}^{N} k_\sigma(x_i - y_i) \tag{11}$$

Correntropy is directly related to the probability of how similar two random variables are in a neighbourhood of the joint space defined by the kernel bandwidth, and provides the probability density of the event p(X=Y). Using Parzen windows, the bandwidth controls the observation window in which similarity is assessed but makes one unable to assess similarity in the whole joint space.

In [12] one may find a discussion on the properties of correntropy. It is proved that a measure CIM(X,Y) = (k(0)-V(X,Y) related with correntropy satisfies all the properties of a metric. CIM may be divided in three different regions: when the error is close to zero CIM is equivalent to L2 norm; when the error grows CIM becomes a L1 norm; when the error is very large CIM becomes a L0 norm, the metric saturates and becomes very insensitive to large errors. This property shows the robustness of CIM and the importance of kernel bandwidth. A small kernel size leads to a small Euclidean zone while a large kernel size will increase the Euclidean region where the metric behaves like the MSE criterion.

When we use correntropy to train adaptive systems, we actually make the system output close to the desired response in the CIM sense if we maximize the correntropy of the error distribution – MCC criterion. We can use the MCC as a new performance function, with the advantage over MSE of being a local criterion of similarity and very useful for cases with non-zero mean, non-Gaussian, with large outliers. It does not require the computing effort of MEE but tends to minimize entropy because it tends to maximize the pdf value at the

origin. It is therefore an attractive criterion to adopt in on-line training of mappers for adaptive wind prediction.

The MCC criterion becomes

$$\min_{w} J(w) = \frac{1}{N}\sum_{i=1}^{N} k_{\sigma}\left(g\left(w, x_i\right) - T\right) \qquad (12)$$

where g(w,x) represents the mapper producing y = g(w,x) responses from input x as a function of weights w, and T represents the target values. This criterion may be differentiated and a backpropagation algorithm used to train a neural network (NN).

VI. On-line training with concept drift

By on-line training we refer to a successive backpropagation procedure that keeps updating the weights of a neural network with new data items arriving. This happens with wind and power data where we have a case of data arriving in a continuous flow to a wind power operation control center.

To show the importance of dealing with concept drift in data streams, we present the results obtained with online and offline NN training modes under the presence of concept drift. The artificial data set consists on 12667 examples generated using the function from Friedman 2 problem and 4223 generated using the function from Friedman 3 problem [13]. The white Gaussian error parameters were the same as used in Friedman problems.

To independently evaluate the NN performance over time we used, after training completed, the cumulative mean absolute error (CMAE) of the normalized values. Fig. 2 shows the evolution of the CMAE with online and offline over the test process for the artificial data with a concept change in iteration 2581 on the test data – the NN trained off-line clearly mis-adjusts while the on-line training actually improves the NN adjustment to a concept drift.

Although in wind forecasting problems it is difficult to detect the point where there is a concept change there is an agreement that the concept changes with high frequency.

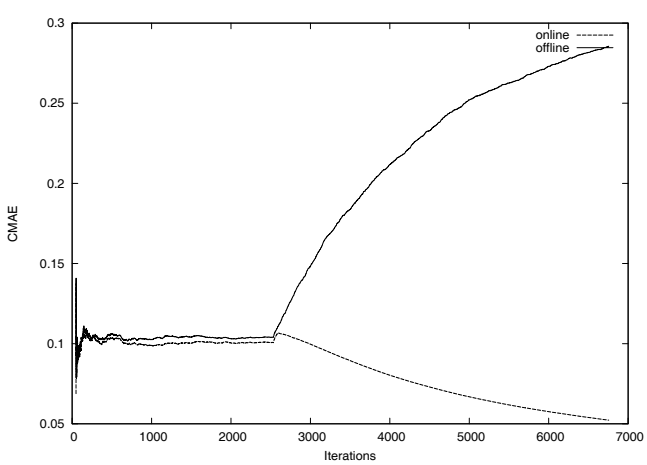

Fig 3. Comparison of errors between online (dashed) and offline (solid) training for artificial data with concept change

VII. Comparing MSE with MCC

In this section we present results for on-line training of neural networks in a wind park comparing the performance of MSE against MCC. This comparison is made in two ways: by comparing the pdf of the errors generated by two NN trained with different criteria and by showing the evolution of CMAE for both networks when subject to on-line training under data streaming. Using CMAE for quality evaluation actually makes sense whenever any prediction error will imply a penalty proportional to the error.

The real cases refer to wind park with a rated power of 21.6 MW and comprising 12 equal wind turbines of 1.8 MW. The facility is situated in the north of Portugal, in a complex mountainous region 65 km away from the Atlantic Ocean. Power output data were collected by the SCADA systems in each park and wind forecasts were generated form the known MM5 mesoscale model (http://www.mmm.ucar.edu/mm5/) in a range from 0 to 24 hours in the forecasting horizon.

Fig. 4 displays one year of measurements of power output against wind speed predictions by the MM5 model with an interleave period of 30 minutes.

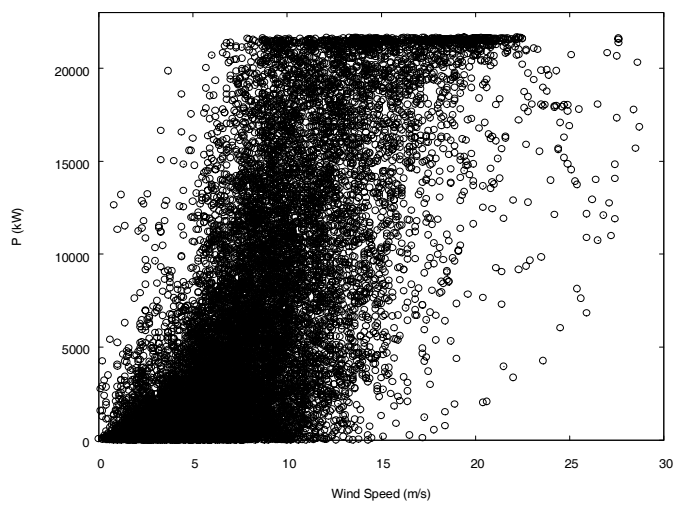

Fig 4. Wind speed prediction(x axis) against power output measured (y axis) in 30 minute intervals.

The NN were designed with 3 inputs (wind speed, and the sin and cos of the direction angle) and only one hidden layer comprising 5 neurons, using an hyperbolic tangent activation function.

In the off-line training we used a batch back-propagation approach with a defined training data set. In the online mode we predict for t+48 (24 hours) on the time stamp t, $\hat{P}_{t+48|t} = NN\left\{\hat{v}_{t+48|t}, \hat{d}_{t+48|t}\right\}$. The real value is known past 24 hours and then we compute the forest error committed in t-48, then we forecast $\hat{P}_{t|t} = NN\left\{\hat{v}_{t|t}, \hat{d}_{t|t}\right\}$ and backpropagate the error.

For MCC we first trained the NN with MSE criterion during the first 200 epochs, which is equivalent to kernel size annealing [14] and switched the criterion to MCC for the following 700 epochs. The learning rate for both criteria was

set to 0.1 and we used a fixed size of 0.002 for the Parzen windows during the batch training. The momentum rate was set to 0.5.

To validate the results we adopted a Monte Carlo procedure running 50 simulations, generating randomly weights in each case by a uniform distribution in the interval [-1,1]. The final conclusions derive from the average of all Monte Carlo simulations. When comparing MSE with MCC, the same weights were used in both cases.

Fig. 4 compares the pdf of errors obtained with MSE and MCC in batch off-line training. Again we see that the pdf of errors has a smaller entropy when the NN is trained with MCC and displays a higher frequency of errors close to zero in this case.

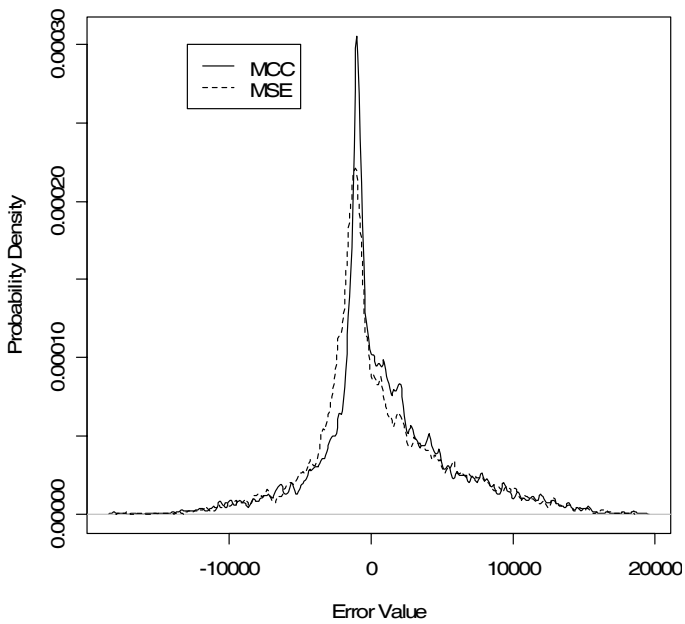

Fig. 5. Comparison of the error pdf generated by NN trained with MSE (dotted) and MCC (solid) criteria.

Fig. 6 compares the performance of NN trained with MSE and MCC criteria in online mode, during more than 3300 hours of simulation.

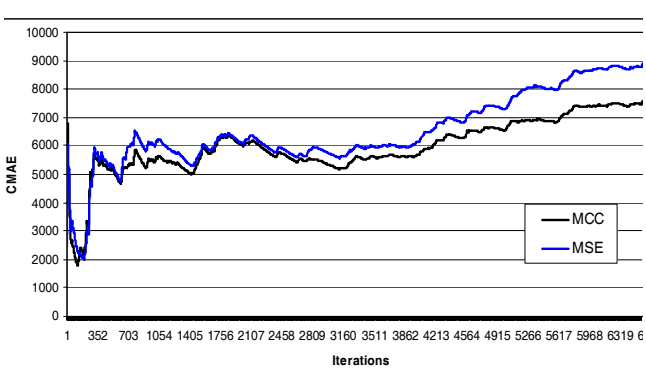

Fig. 6. CMAE comparison of errors produced by NN in online training mode under MSE (line above) and MCC (line below) showing that the MCC criterion produces a mapper with better global performance.

VIII. Conclusions

Wind power prediction for wind parks displays obvious non-Gaussian characteristics in the error probability distributions and therefore training input-output systems (mappers) to perform wind power prediction gains from moving away from criteria based on Variance (such as MSE) and instead adopting criteria based on Entropy concepts. For this purpose, a direct use of an entropy measure (the MEE criterion) is possible, but an approximation such as the MCC criterion measuring correntropy has clear computing advantages.

Wind prediction errors suffer from concept drift, meaning that the underlying probabilistic distributions that may explain them change characteristics along time. Therefore, an online training mode for NN designed to produce power forecasts for operation purposes is preferable to an off-line training mode.

This paper contributes with evidence supporting both assertions, based on studies performed with real data from wind parks in Europe and comparisons between NN trained under the two competing criteria – MSE and MCC.

IX. Appendix – Takagi-Sugeno FIS

A Takagi-Sugeno Fuzzy Inference System (TS-FIS) may be viewed as neuro-fuzzy mapper and is commonly trained with the back-propagation algorithm under supervised training. The most widespread tool is the ANFIS [15]. One has training and test sets with target values **T** and the training task deals with the information content of the errors $\boldsymbol{\varepsilon} = \mathbf{T} - \mathbf{y}$ between target and the output **y**.

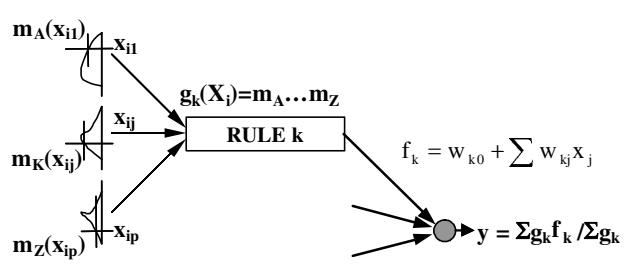

Figure 7 – TS-FIS scheme. Each input pattern X activates some membership functions; the combination of these fires a rule k with strength g_k. The weighted combination of rule firing strengths gives the output of the system.

In a TS FIS, one has rules that are fuzzy in their antecedent and crisp in their consequent (Fig. 7). A general form of a rule k with output y_k is

IF (x_1 is A and … and x_p is Z) THEN $y_k = y(\boldsymbol{x}, \boldsymbol{w})$

The antecedent of rule k is a fuzzy set whose membership function g_k is the intersection of fuzzy sets describing conditions A,…,Z. Usually, the T-norm used to represent intersection is the product (of the membership values of each input variable).

The consequent of a rule k is a function f_k of inputs. In 0-order TS-FIS, f_k is constant and, therefore, $f_k = w_k$. In 1st-order TS-FIS, f_k is a linear combination of inputs such as in

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

The output of a TS-FIS is a weighted sum of the responses of all the rules

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

To optimize a performance criterion, one has to discover the adequate weight **w** values for the rules. Other parameters may be adjusted by training in TS-FIS.

X. References

[1] Mark L. Ahlstrom, Robert M. Zavadil, "The Role of Wind Forecasting in Grid Operations & Reliability", Proceedings of IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific, China, pp. 1-5, 2005.

[2] 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

[3] 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

[4] 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

[5] 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

[6] A. Renyi, "Some Fundamental Questions of Information Theory", *Selected Papers of Alfred Renyi*, vol 2, pp. 526-552, Akademia Kiado, Budapest, 1976.

[7] E. Parzen, "On the estimation of a probability density function and the mode", *Annals Math. Statistics*, v. 33, 1962, p. 1065

[8] 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

[9] 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.

[10] 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

[11] 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

[12] W. Liu, P. Pokharel, J. Principe, "Correntropy: properties and applications in non-Gaussian signal processing", Accepted in IEEE Transactions on Signal Processing, 2007.

[13] J. H. Friedman, "Multivariate adaptive regression splines", *The Annals of Statistics, vol. 19*(1), pp. 1-67, 1991.

[14] D. Erdogmus, J. Principe, An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems, *IEEE Transactions on Signal Processing, vol. 50*, pp. 1780-1786, July,2002.

[15] J.-S. Roger Jang, "ANFIS – Adaptive Network Based Fuzzy Inference Systems", IEEE Transactions on Systems, Man and Cybernetics, vol. 23, no. 3, May 1993

XI. Biographies

Ricardo Bessa received his Licenciado degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 2005 in Electrical and Computer Engineering. He then joined INESC Porto and is a researcher in its Power Systems Unit while completing a M.Sc. degree in the Faculty of Economy of the University of Porto (FEP).

Vladimiro Miranda (M'90, SM'04, F'05) received his Licenciado, Ph.D. and Agregado degrees from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 1977, 1982 and 1991, all in Electrical Engineering. In 1981 he joined FEUP and currently holds the position of Professor Catedrático. He is also currently a Director of INESC Porto, an advanced research institute in Portugal. He has authored many papers and been responsible for many projects in areas related with the application of Computational Intelligence to Power Systems.

João Gama received his Licenciado degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in Electrical and Computer Engineering.. In 2000 he received his Ph.D. degree in Computer Science from the Facuky of Sciences of the same University. He joined the Faculty of Economy (FEP) where he holds the position of Assistant Professor, He is also a senior researcher at LIAAD (Laboratory for Artificial Intelligence and Decision Aid), a group belonging to INESC Porto LA. He has worked in projects and authored papers in areas related to machine learning, data streams and adaptive learning systems and is a member of the editorial board of international journals in his area of expertise.