

APLICAÇÃO DE MÉTODOS ESTATÍSTICOS E REDES NEURASIS NO PÓS-  
PROCESSAMENTO DE PRODUTOS DE PREVISÃO NUMÉRICA DE TEMPO.

Nilza Barros da Silva

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM  
ENGENHARIA CIVIL.

Aprovada por:

---

Prof. José Luis Drummond Alves, D.Sc.

---

Dr. Ricardo Carvalho de Almeida, D.Sc.

---

Prof. Gutemberg Borges França, Ph.D.

---

Dr. Reinaldo Bomfim da Silveira, Ph.D.

RIO DE JANEIRO, RJ - BRASIL  
SETEMBRO DE 2007

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

SILVA, NILZA BARROS DA

Aplicação de Métodos Estatísticos  
e Redes Neurais no Pós-Processamento  
de Produtos de Previsão Numérica de  
Tempo [Rio de Janeiro] 2007

IX, 207 p. 29,7 cm (COPPE/UFRJ,  
M.Sc., Engenharia Civil, 2007)

Dissertação - Universidade Federal do Rio de  
Janeiro, COPPE

1. Redes Neurais Artificiais
2. Model Output Statistics
3. Componentes Principais
4. Pós-processamento de produtos numéricos
5. Previsão de eventos meteorológicos severos

I. COPPE/UFRJ II. Título (série)

*“Entrega o teu caminho ao Senhor, confia nele e o mais Ele fará.”*

*(Salmos 37:5)*

Dedico este trabalho, aos meus pais, Fátima e Nilson  
e ao meu marido André.

# *Agradecimentos*

Gostaria de agradecer primeiramente a Deus, por tudo que Ele me proporciona nesta vida.

Minha gratidão muito especial ao meu co-orientador Ricardo Carvalho de Almeida pela confiança e inestimável ajuda desde o início até o último momento do meu mestrado.

Ao meu orientador José Alves pelo apoio e ajuda durante os momentos cruciais desta dissertação.

Aos meus ex-chefes, CMG Garcez, CMG Claudio e CMG Costa Neves que permitiram desde o início a realização deste curso e confiaram que eu poderia chegar até aqui.

Aos meus chefes, CF Obino, CF Chen e CC Luiz Claudio, que me apoiaram compreenderam as ausências que se fizeram necessárias para realização deste curso.

Aos meus amigos e colegas, Ilma, Angélica, Fábio David, CC Giovana e Dandan que sempre confiaram em mim e me ouviram nos momentos em que me sentia cansada ou um pouco desanimada.

Aos meus familiares que estiveram ao meu lado em todos os momentos.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## APLICAÇÃO DE MÉTODOS ESTATÍSTICOS E REDES NEURASIS NO PÓS-PROCESSAMENTO DE PRODUTOS DE PREVISÃO NUMÉRICA DE TEMPO

Nilza Barros da Silva

Setembro/2007

Orientadores: José Luiz Drummond Alves

Ricardo Carvalho de Almeida

Programa: Engenharia Civil

As saídas do modelo numérico HRM (*High resolution Reginal Model*) e as observações de três estações meteorológicas de superfície foram utilizadas no desenvolvimento de redes neurais e regressões para previsões de ocorrência de vento forte e visibilidade restrita. As redes e as equações foram desenvolvidas de tal forma que fossem capazes de prever os eventos de forma independente. As amostras para as três estações costeiras ficaram limitadas em função do período de configuração do modelo numérico e pelo fato de apresentarem registros das observações de vento e visibilidade somente em quatro horários (00Z, 06Z, 12Z e 18Z). Foram empregadas várias medidas de verificação para comparar os resultados da rede neural com o MOS (*Model Output Statistics*) e com as previsões de vento do modelo numérico. Foram desenvolvidas redes neurais e regressões para diferentes configurações das amostras de dados nas três estações estudadas. As variáveis preditoras foram obtidas com uso de componentes principais ou selecionadas por *screening regression*. Verificou-se que os dois métodos de pós-processamento apresentam melhor desempenho que o modelo numérico apenas na previsão dos eventos de interesse, e que a rede neural apresenta maior acurácia e confiabilidade do que o MOS. Observou-se também que a melhora obtida pelo pós-processamento depende da localização da estação.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

STATISTICAL METHODS AND NEURAL NETWORK APLLIED IN THE POST-  
PROCESSING OF NUMERICAL WEATHER PREDICTION

Nilza Barros da Silva

September/2007

Advisors: José Luiz Drummond Alves  
Ricardo Carvalho de Almeida

Department: Civil Engineering

Output from HRM (High resolution Regional Model) numerical prediction model and surfaces observations from three stations were used to develop neural networks and statistical regressions for predicting high speed wind and low visibility events. The neural networks and regression equations were developed to predict visibility or high speed wind independently. The data sample was limited by the numerical model configuration and because the surface observations were available only four times a day (00Z, 06Z, 12Z e 18Z). Some performance measures were applied to compare the neural network, MOS (Model Output Statistics) and HRM predictions with observational data. Neural networks and regressions were developed with different data set configurations for each of the three stations. Principal Components or Screening Regression was employed to select the predictors variables. It was found that both post-processing methodologies produce forecasts that outperform the original numerical prediction forecasts and that the neural networks predictions present more accuracy and reliability than MOS. The extent of the improvement depends on the station location.

# Sumário

<b>SUMÁRIO</b> .....	<b>VII</b>
<b>CAPÍTULO 1</b> .....	<b>1</b>
<b>INTRODUÇÃO</b> .....	<b>1</b>
<b>CAPÍTULO 2</b> .....	<b>6</b>
<b>REVISÃO BIBLIOGRÁFICA E FUNDAMENTOS TEÓRICOS</b> .....	<b>6</b>
<b>2.1</b> Métodos de pós –processamento .....	6
<b>2.1.1</b> Model Output Statistics .....	8
2.1.1.1 Introdução .....	8
2.1.1.2 Breve Histórico .....	8
2.1.1.3 Desenvolvimento das equações de previsão do MOS .....	9
2.1.1.4 Regressão Linear .....	10
2.1.1.4.1 Regressão Linear Simples .....	10
2.1.1.4.1.1 Significado dos parâmetros da regressão .....	12
2.1.1.4.1.2 Estimação da função de regressão .....	13
2.1.1.4.1.3 Distribuição dos Resíduos .....	14
2.1.1.4.2 Regressão Linear Múltipla .....	18
2.1.1.4.2.1 Inferências sobre os parâmetros da regressão .....	19
2.1.1.4.3 Screening Regression .....	21
2.1.1.4.4 Componentes Principais como variáveis preditoras .....	23
2.1.1.4.5 Regressão para estimação de eventos probabilísticos .....	27
<b>2.1.2</b> Fundamentos de Redes Neurais .....	29
2.1.2.1 Introdução .....	29
2.1.2.2 Breve Histórico .....	30
2.1.2.3 A Inspiração Biológica .....	33
2.1.2.4 Neurônio Artificial .....	34
2.1.2.4.1 Funções de Ativação .....	37
2.1.2.5 Rede Perceptron .....	39
2.1.2.6 Rede Perceptron de Múltiplas Camadas .....	40
2.1.2.6.1 Definindo a arquitetura da rede neural .....	41
2.1.2.7 Processo de aprendizagem .....	44
<b>CAPÍTULO 3</b> .....	<b>48</b>
<b>ÁREAS DE ESTUDO</b> .....	<b>48</b>
<b>3.1</b> Descrição das áreas .....	48
<b>CAPÍTULO 4</b> .....	<b>55</b>
<b>DADOS E METODOLOGIA</b> .....	<b>55</b>
<b>4.1</b> Dados .....	55



<b>4.2</b>	Metodologia .....	58
<b>4.2.1</b>	Controle de qualidade dos dados .....	58
<b>4.2.2</b>	Escolha das variáveis preditoras .....	58
<b>4.2.3</b>	Características das amostras .....	60
<b>4.2.4</b>	Implementação da Rede Neural .....	63
<b>4.2.5</b>	Implementação da MOS .....	66
<b>4.2.6</b>	Avaliação do desempenho do pós-processamento .....	67
<b>CAPÍTULO 5</b>	.....	<b>71</b>
<b>RESULTADOS E DISCUSSÃO</b>	.....	<b>71</b>
<b>5.1</b>	Redes Neurais e MOS como classificadores – Ocorrência de Vento Forte e de Visibilidade Restrita .....	71
<b>5.1.1</b>	Ocorrência de Vento Forte .....	74
5.1.1.1	Salinópolis .....	74
5.1.1.1.1	Rodadas de 00Z e 12Z .....	75
5.1.1.1.1.a	Dados Replicados .....	75
5.1.1.1.1.b	Dados Não Replicados .....	84
5.1.1.2	Santa Marta .....	89
5.1.1.2.1	Rodadas de 00Z e 12Z .....	90
5.1.1.2.1.a	Dados Replicados .....	90
5.1.1.2.1.b	Dados Não Replicados .....	93
5.1.1.2.2	Rodada de 00Z .....	95
5.1.1.2.2.a	Dados Replicados .....	95
5.1.1.2.2.b	Dados Não Replicados .....	98
5.1.1.2.3	Rodada de 12Z .....	101
5.1.1.2.3.a	Dados Replicados .....	101
5.1.1.2.3.b	Dados Não Replicados .....	103
5.1.1.3	Mostardas .....	106
5.1.1.3.1	Rodadas de 00Z e 12Z .....	106
5.1.1.3.1.a	Dados Replicados .....	106
5.1.1.3.1.b	Dados Não Replicados .....	110
<b>5.1.2</b>	Ocorrência de Visibilidade Restrita .....	114
5.1.2.1	Salinópolis .....	115
5.1.2.1.1	Rodadas de 00Z e 12Z .....	116
5.1.2.1.1.a	Dados Replicados .....	116
5.1.2.1.1.b	Dados Não Replicados .....	118
5.1.2.2	Santa Marta .....	120
5.1.2.2.1	Rodadas de 00Z e 12Z .....	120
5.1.2.2.1.a	Dados Replicados .....	120
5.1.2.2.1.b	Dados Não Replicados .....	124
5.1.2.2.2	Rodada de 00Z .....	128
5.1.2.2.2.a	Dados Replicados .....	128
5.1.2.2.2.b	Dados Não Replicados .....	130
5.1.2.2.3	Rodada de 12Z .....	131
5.1.2.2.3.a	Dados Replicados .....	131
5.1.2.2.3.b	Dados Não Replicados .....	134
5.1.2.3	Mostardas .....	136
5.1.2.3.1	Rodadas de 00Z e 12Z .....	136
5.1.2.3.1.a	Dados Replicados .....	136
5.1.2.3.1.b	Dados Não Replicados .....	140
<b>CAPÍTULO 6</b>	.....	<b>144</b>
<b>CONCLUSÃO</b>	.....	<b>144</b>

<b>CAPÍTULO 7</b> .....	<b>151</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>151</b>
<b>Apêndice A</b> Sumário estatístico do pós-processamento (MOS e RN) para ocorrência de vento forte e visibilidade restrita.....	<b>156</b>
<i>Sumário Estatístico para previsão de ocorrência de vento forte -Salinópolis</i> .....	<b>156</b>
<i>Sumário Estatístico para previsão de ocorrência de vento forte -Mostardas</i> .....	<b>159</b>
<i>Sumário Estatístico para previsão de ocorrência de visibilidade Restrita -Salinópolis</i> .....	<b>163</b>
<i>Sumário Estatístico para previsão de ocorrência de visibilidade Restrita –Mostardas</i> .....	<b>165</b>
<b>Apêndice B</b> – Regressões Lineares Múltiplas do MOS.....	<b>168</b>
<i>Equações do MOS para previsão de ocorrência de vento forte -Salinópolis</i> .....	<b>168</b>
<i>Equações do MOS para previsão de ocorrência vento forte - Santa Marta</i> .....	<b>175</b>
<i>Equações do MOS para previsão de ocorrência vento forte - Mostardas</i> .....	<b>180</b>
<i>Equações do MOS para previsão de ocorrência de visibilidade restrita - Salinópolis</i> .....	<b>187</b>
<i>Equações do MOS para previsão de ocorrência de visibilidade restrita –Santa Marta</i> .....	<b>194</b>
<i>Equações do MOS para previsão de ocorrência de visibilidade restrita –Mostardas</i> .....	<b>201</b>

# **Capítulo 1**

## **Introdução**

Condições de tempo adversas podem ameaçar a segurança da navegação marítima de forma que a previsão adequada e antecipada de tais fenômenos pode evitar situações de riscos.

As condições de vento forte e de visibilidade restrita podem ser enquadradas em situações de risco, pois a ocorrência de tais eventos implica, normalmente, em interrupção e conseqüentes prejuízos das atividades marítimas e, em alguns casos, de riscos à vida humana. Contudo, apesar dos avanços na previsão do tempo, estes fenômenos ainda são de difícil previsão, e contam com alto grau de subjetividade do previsor do tempo. Portanto, a aplicação de metodologias que produzam resultados que possam servir de subsídios ao previsor quando da previsão destes eventos podem ser grande valia.

O nevoeiro é uma nuvem com base em contato com o solo. Pode se formar quando o ar se torna saturado através do resfriamento radiativo, resfriamento advectivo, resfriamento por expansão ou por adição de vapor d'água. No Manual do Observador Meteorológico (DHN, 1992) o nevoeiro é definido como a condensação de vapor d'água nas baixas camadas da atmosfera, reduzindo a visibilidade horizontal a menos de 1km.

A restrição da visibilidade é um motivo de preocupação para órgãos gerenciadores de meios de transportes aéreos, marítimos e rodoviários, pois, normalmente, a ocorrência de nevoeiros pode ser uma situação precursora de acidentes

graves e, em alguns casos, fatais. O mesmo se pode afirmar da ocorrência de ventos fortes.

O Serviço Meteorológico Marinho (SMM) nacional, operado pelo Centro de Hidrografia da Marinha (CHM), é responsável pela emissão de boletins de previsão meteorológica para a região marítima entre os paralelos 7° N e 35.8° S, a oeste do meridiano de 20° W. Inclui-se em suas atribuições a emissão de avisos de mau tempo, pelos quais são reportadas aos navegantes condições de tempo que possam representar ameaça à segurança da navegação. A Organização Meteorológica Mundial (OMM) define, entre outras, duas condições para a emissão de avisos de mau tempo: baixa visibilidade horizontal, quando esta fica reduzida a menos que 1 km, e ocorrência de ventos fortes, quando a velocidade do vento atinge valores iguais ou superiores a 13.9 m/s (WMO, 1990). Caso esses fenômenos sejam previstos, deve-se emitir um boletim especial aos navegantes, denominado “Aviso de Mau Tempo”.

A atividade de previsão meteorológica envolve a integração e análise de dados e informações produzidos por diversos métodos. Entre esses pode-se citar como um dos mais importantes a previsão numérica de tempo. Este método consiste na integração numérica de sistemas de equações que prognosticam os processos físicos, dinâmicos e termodinâmicos da atmosfera. Essas equações são resolvidas em uma grade de pontos discreta, gerando valores previstos de diversas variáveis relevantes para a caracterização das condições meteorológicas, tais como, magnitude das componentes meridional e latitudinal do vento, temperatura do ar, umidade relativa e pressão atmosférica.

Apesar da sofisticação e confiabilidade dos modelos numéricos, eles ainda apresentam limitações na representação de fenômenos locais e de pequena escala que não podem ser resolvidos explicitamente. KALNAY (2003) argumenta que apesar do contínuo aumento da resolução dos modelos numéricos existem alguns processos que

ocorrem em escala molecular como condensação, evaporação, etc., que não podem ser resolvidos explicitamente pelos atuais e nem por modelos futuros. Assim, torna-se necessária a parametrização de tais fenômenos. Além disso, o modelo numérico, necessariamente, simplifica condições de superfície, uma vez que sua representação do “mundo” baseia-se em pontos de grade.

Dentre as restrições do modelo numérico, inclui-se a previsão dos eventos de interesse neste estudo: a visibilidade, que não é prevista diretamente pelo modelo e conta com a subjetividade e experiência do previsor para sua previsão. O vento previsto pelo modelo, por sua vez, apresenta baixa resolução espacial e não representa adequadamente as situações de vento forte.

Desta forma, constata-se a necessidade de aplicação de técnicas objetivas de previsão para os casos de visibilidade restrita e de vento forte que possam servir de auxílio ao previsor do tempo na tomada de decisão.

Neste trabalho serão utilizadas redes neurais artificiais e regressões do *Model Output Statistics* (MOS) para serem aplicadas na previsão de ocorrência de visibilidade restrita e de vento forte em três estações meteorológicas a partir do mapeamento entre as previsões numéricas de tempo realizadas pelo modelo *High Resolution Regional Model* (HRM) e os valores de visibilidade e vento registrados nas observações meteorológicas de superfície daquelas estações. Serão desenvolvidas redes e regressões para diferentes configurações das amostras de dados nas três estações estudadas. Para variáveis preditoras adotou-se o uso de componentes principais, ou estas foram escolhidas por *screening regression*.

Os resultados das previsões das redes neurais e do MOS serão comparados a fim de avaliar qual deles apresenta melhor desempenho na previsão destes fenômenos. Além disso, serão avaliados os métodos utilizados na escolha das variáveis preditoras.

O SMM utiliza o modelo numérico HRM, para apoiar suas atividades de previsão. Atualmente, o modelo apresenta uma configuração diferente da que apresentava no período em estudo, pois se considerou o período em que não houve alteração na configuração do modelo. Desta forma, não será possível a implementação imediata das técnicas apresentadas no presente estudo.

O estudo foi feito para três estações de observação meteorológica de superfície, Salinópolis (PA), Santa Marta (SC) e Mostardas (RS), que foram escolhidas em função de apresentarem maior volume de informações no período de estudo. Contudo, as diferentes posições geográficas possibilitarão a avaliação do desempenho dos métodos de pós-processamento, rede neural e MOS, em diferentes condições climáticas.

Neste estudo a análise da visibilidade observada nas estações mostrou a baixa ocorrência de casos em que ela era menor que 1 km. Adotou-se, portanto, uma margem que caracterizasse uma situação de alerta, mas que fornecesse informações que possibilitassem o adequado uso do pós-processamento. Assim os casos de visibilidade restrita foram classificados quando a visibilidade horizontal fosse menor que 2 km. Desta forma, estão abrangidas as ocorrências de visibilidade restrita causadas por nevoeiro ou névoa úmida, que ocorrem quando da condensação do vapor d'água.

É importante ressaltar a complexidade do pós-processamento e da previsão das ocorrências de visibilidade restrita e de vento forte, pois são classificados como eventos raros. Esses eventos apresentam baixa frequência de ocorrência no conjunto de dados observados nas estações, o que torna a tarefa de prever estes eventos ainda mais difícil para a rede neural e para MOS. A visibilidade apresenta uma dificuldade ainda maior, que começa no momento do registro nas estações meteorológicas de superfície. As observações de visibilidade no mar não contam, normalmente, com auxílio de

aparelhos. Utilizam-se pontos de referências cujas distâncias são conhecidas, ou seja, toda avaliação da visibilidade será sempre subjetiva.

Podemos resumir os objetivos deste trabalho em três perguntas que pretendemos responder ao final:

- 1- Utilizar redes neurais para o pós-processamento das previsões do modelo numérico melhora as previsões locais?
- 2- A utilização de métodos estatísticos (MOS) melhora a previsão local?
- 3- Qual dos dois métodos é melhor?

O presente trabalho está organizado em sete capítulos. Neste, são apresentadas as motivações e os objetivos do trabalho bem como as perguntas que esta dissertação pretende responder ao final. No capítulo 2, são apresentadas as revisões bibliográficas e o embasamento teórico dos métodos de rede neurais e MOS utilizados neste trabalho. No capítulo 3 são descritas as principais características da região de estudo quanto à localização e frequência dos eventos estudados. No capítulo 4 apresentam-se os dados e as metodologias aplicadas para implementação dos métodos de pós-processamento. No capítulo 5 apresentam-se os resultados obtidos nos vários experimentos realizados nas três estações estudadas para diferentes configurações das amostras. No capítulo 6 são apresentadas as conclusões e recomendações e no capítulo 7 as referências bibliográficas utilizadas na presente dissertação. No apêndice A são apresentadas as tabelas com o sumário estatístico dos experimentos realizados e os gráficos com as medidas de desempenho para as três estações estudadas nos horários de 00Z e 12Z. No apêndice B são apresentadas as tabelas com os coeficientes estimados para as regressões do MOS para as três estações bem como algumas estatísticas utilizadas nas inferências destas regressões.

# **Capítulo 2**

## **Revisão Bibliográfica e Fundamentos Teóricos**

Neste capítulo serão apresentadas as bases teóricas para as técnicas de pós-processamento que serão utilizadas nos capítulos seguintes. Além disso, serão apresentados os dois métodos usados na definição da forma como as variáveis preditoras são utilizadas no treinamento da rede neural e no desenvolvimento das equações de regressão do *Model Output Statistics* (MOS). São dois métodos distintos, denominados *Screening Regression* e Análise de Componentes Principais.

### **2.1 Métodos de pós –processamento**

Vários estudos têm mostrado que as previsões oriundas de modelos numéricos podem ser aperfeiçoadas por meio do pós-processamento destas informações (GLAHN, LOWRY, 1972, KULIGOWSKI, BARROS, 1998, MARZBAN, 2003).

Com base na estatística clássica houve o desenvolvimento de previsões do tempo sem a utilização de dados provenientes dos modelos numéricos. Contudo, com o avanço nas previsões resultantes dos modelos numéricos, novos tipos de previsões estatísticas com uso de produtos oriundos destes modelos, surgiram a fim de aperfeiçoar estas previsões.

Há várias maneiras de se realizar este pós-processamento. Os dois métodos mais conhecidos são chamados *perfect-prog* e *Model Output Statistics*. Ambos têm suas vantagens e desvantagens, mas o último tem sido preferido pela sua capacidade incluir diretamente nas equações de regressão as influências de características de diferentes modelos numéricos em diferentes projeções no futuro (WILKS, 2006).



Mais recentemente, métodos não lineares têm sido usados para o pós-processamento. Entre eles, o interesse pelo uso de redes neurais tem aumentado entre os meteorologistas (HSIEH, TANG, 1998).

WILKS (2006) argumenta que há três razões para o uso do pós-processamento das saídas dos modelos numéricos:

1- Os modelos por sua natureza intrínseca, de representar o mundo em pontos de grades, simplificam e homogeneízam condições de superfície. Efeitos de pequenas escalas, importantes para previsões locais, dependendo da resolução, podem não ser incluídos nos modelos numéricos. Além disso, algumas variáveis importantes na previsão não são explicitamente representadas ou fornecidas pelos modelos numéricos. Contudo, relações estatísticas entre as informações do modelo e o valor desejado podem ser desenvolvidas a fim de resolver este problema;

2- Os modelos numéricos estão sujeitos a erros, os quais, por serem sistemáticos, podem ser minimizados por meio de previsões estatísticas baseadas nas informações dos modelos numéricos a fim de compensar e corrigir as previsões; e

3- Os modelos numéricos são determinísticos, isto é, dado um conjunto de condições iniciais, produzem um mesmo resultado para determinado elemento meteorológico. O uso desta informação, em conjunção com métodos estatísticos, permite expressar e quantificar as incertezas associadas com diferentes situações de previsão por meio, por exemplo, de equações cujas previsões sejam de probabilidades de ocorrência do evento de interesse.

Neste trabalho serão utilizados como métodos de pós-processamento o MOS e as redes neurais artificiais. No desenvolvimento destas técnicas serão utilizadas, como variáveis preditoras, as previsões do modelo numérico em conjunção com variáveis observadas.

## **2.1.1 Model Output Statistics**

### **2.1.1.1 Introdução**

A técnica de MOS consiste em determinar uma relação estatística entre os campos de variáveis meteorológicas oriundos de modelos numéricos de previsão de tempo (variáveis preditoras ou independentes) e observações locais (variáveis preditadas ou dependentes). As equações de previsão são desenvolvidas por meio de técnicas estatísticas que serão detalhadas adiante.

### **2.1.1.2 Breve Histórico**

Uma das primeiras técnicas estatísticas a ser desenvolvida com uso das saídas de modelos numéricos recebeu o nome de *perfect prog* (KLEIN *et al.*, 1959). Neste método, climatologias históricas são usadas na obtenção dos coeficientes das equações de previsão e, somente durante a implementação as previsões do modelo numérico são utilizadas como variáveis preditoras para obtenção das previsões.

O MOS proposto por GLAHN e LOWRY (1972) foi o segundo método a incorporar as informações de modelos numéricos. Contudo, enquanto o *perfect prog* usa as previsões do modelo numérico apenas para fazer as previsões, o MOS usa esses preditores no desenvolvimento e na implementação das equações. No trabalho desses autores foram usadas as saídas de dois modelos numéricos operados pelo NWS (*National Weather Service*) dos EUA, para derivar as equações de regressão para previsões objetivas de probabilidade de precipitação, vento a superfície, probabilidade de ocorrência de granizo, temperatura máxima e cobertura de nuvem. O MOS mostrou-se útil na previsão objetiva do tempo, e seus resultados foram disseminadas para o público pelo NWS.

Carter *et al.* (1989) relatam o uso do MOS e do *perfect prog* nos Estados Unidos da América pelo NWS nas duas décadas anteriores. Informam que mais de 100.000 equações estavam sendo processadas e produziam em torno de 820.000 previsões diárias. Relatam também a superioridade das previsões do MOS em relação ao *perfect prog*.

Experimentos com o uso MOS são reportadas em vários países, entre eles, a Austrália (WOODCOCK, 1984, TAPP *et al.*, 1986), e o Canadá (WILSON, YACOWAR, 1980).

### **2.1.1.3 Desenvolvimento das equações de previsão do MOS**

O *perfect prog* utiliza dados históricos de observações para o desenvolvimento de suas equações. As saídas do modelo numérico são aplicadas apenas para obtenção das previsões.

$$\hat{y}_0 = f_{PP}(x_0) \quad \text{no desenvolvimento da equação} \quad (2-1)$$

$$\hat{y}_t = f_{PP}(x_t) \quad \text{na implementação da equação} \quad (2-2)$$

A função de regressão,  $f_{PP}$ , é a mesma nos dois casos, mas o desenvolvimento das equações, que implica em estimar os parâmetros da regressão, se faz com dados observados no mesmo tempo  $0$  (zero) para os preditores e o preditando. Na fase de implementação, os preditores no tempo  $t$  ( $x_t$ ) são obtidos a partir das previsões do modelo numérico.

Finalmente, o MOS usa a mesma equação tanto no desenvolvimento quanto na implementação.

$$\hat{y}_t = f_{MOS}(x_t) \quad (2-3)$$

A equação acima é desenvolvida usando como preditores ( $x_t$ ) que se referem à previsão para o futuro tempo  $t$ , e é implementada da mesma forma.

O MOS é baseado em regressão múltipla. A seguir serão apresentadas informações teóricas para sobre esta ferramenta estatística.

#### **2.1.1.4 Regressão Linear**

Nesta seção, os fundamentos de regressão linear serão apresentados. Informações mais completas podem ser obtidas em KUTNER *et al.* (2005) e em NETER e WASSERMAN (1974).

##### **2.1.1.4.1 Regressão Linear Simples**

Considere o caso mais simples de regressão onde apenas uma variável independente é tratada e a função de regressão é linear. O modelo pode ser estabelecido da seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2-4)$$

onde:

$Y_i$  é o valor da variável resposta no  $i$ -ésimo ensaio ou observação

$\beta_0$  e  $\beta_1$  são parâmetros

$X_i$  é uma constante, ou seja, o valor da variável independente no  $i$ -ésimo ensaio

$\varepsilon_i$  é um erro aleatório com média igual a zero e variância constante ( $E(\varepsilon_i) = 0$  e  $\sigma^2(\varepsilon_i) = \sigma^2$ );  $\varepsilon_i$  e  $\varepsilon_j$  não são correlacionados de modo que a covariância entre eles é zero ( $\sigma(\varepsilon_i, \varepsilon_j) = 0$  para todo  $i, j$ ;  $i \neq j$ )

$i = 1, 2, \dots, n.$

O modelo (2-4) é denominado *simples, linear nos parâmetros e linear na variável independente*. Ele é “simples” porque tem apenas uma variável independente, “linear nos parâmetros” porque nenhum parâmetro aparece em expoente ou é multiplicado ou dividido por outro parâmetro, é “linear na variável independente” porque esta aparece apenas elevada à primeira potência. O modelo que é linear nos parâmetros e na variável independente é também chamado modelo de primeira ordem.

Seguem abaixo algumas características importantes dos modelos de regressão:

a) a variável resposta  $Y_i$  no  $i$ -ésimo ensaio é a soma de dois componentes: (1) o termo constante  $\beta_0 + \beta_1 X_i$ , e (2) o termo aleatório  $\varepsilon_i$ . Logo  $Y_i$  é uma variável aleatória.

b) Uma vez que  $E(\varepsilon_i) = 0$ , segue que

$$E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_0 + \beta_1 X_i + E(\varepsilon_i) = \beta_0 + \beta_1 X_i$$

Assim, a variável resposta  $Y_i$ , quando o nível de  $X$  no  $i$ -ésimo ensaio é  $X_i$ , vem de uma distribuição de probabilidade cuja média é:

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad (2-5)$$

Portanto, a função de regressão do modelo (2-4) é:

$$E(Y) = \beta_0 + \beta_1 X \quad (2-6)$$

Assim, diz-se que a função de regressão para o modelo (2-4) é a expressão (2-6), já que a equação de regressão relaciona as médias das distribuições de probabilidade de  $Y$  para dado  $X$ , com os níveis de  $X$ .

c) A resposta  $Y_i$  na  $i$ -ésima prova, excede ou resulta em valor abaixo do valor de função de regressão, de uma quantidade  $\varepsilon_i$  chamada termo do erro.

d) Assume-se que os termos de erro  $\varepsilon_i$  têm variância constante  $\sigma^2$ . Daí segue-se que a variância da variável resposta  $Y_i$  tem também variância constante:

$$\sigma^2(Y_i) = \sigma^2 \quad (2-7)$$

Assim, o modelo (2-4) assume que as distribuições de probabilidade de  $Y$  têm a mesma variância  $\sigma^2$ , qualquer que seja o nível da variável independente  $X$ .

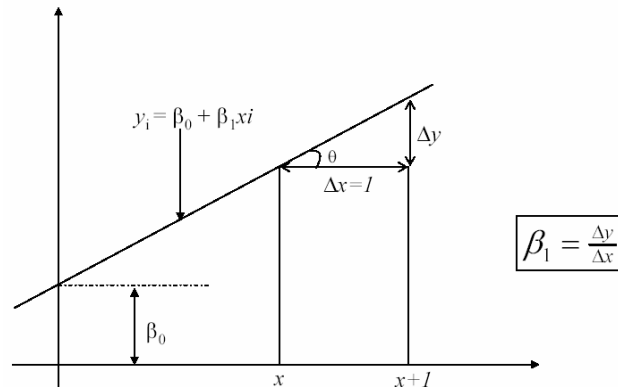
e) Assume-se que os termos de erro não são correlacionados. Portanto, o resultado de qualquer ensaio não tem nenhum efeito sobre o termo do erro para qualquer outro ensaio, ou seja, ser positivo ou negativo, grande ou pequeno, etc. Sendo não correlacionados os termos do erro, também não serão os  $Y$ 's.

f) Em resumo, o modelo (2-4) considera que as observações da variável resposta  $Y_i$  vêm de distribuições de probabilidade cujas médias são  $E(Y) = \beta_0 + \beta_1 X$  e cujas variâncias são  $\sigma^2$ , a mesma para todos os níveis de  $X$ . Adicionalmente, quaisquer duas observações  $Y_i$  e  $Y_j$  são não correlacionadas.

#### **2.1.1.4.1.1 Significado dos parâmetros da regressão**

Os parâmetros  $\beta_0$  e  $\beta_1$  do modelo (2-4), Figura 2-1, são chamados coeficientes de regressão.  $\beta_1$  é a inclinação da linha de regressão. Ele indica as variações na média da distribuição de probabilidade de  $Y$  por acréscimo unitário em  $X$ . O parâmetro  $\beta_0$  é o intercepto da linha de regressão com o eixo da ordenada. Se a amplitude do modelo inclui  $X = 0$ ,  $\beta_0$  fornece a média da distribuição de probabilidade de  $Y$  no ponto  $X=0$ .

Quando a amplitude do modelo não inclui  $X=0$ ,  $\beta_0$  não tem um significado particular como termo separado do modelo de regressão.



**Figura 2-1** Significado dos parâmetros do modelo de regressão linear simples

### 2.1.1.4.1.2 *Estimação da função de regressão*

Os parâmetros do modelo  $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$  a serem estimados a partir de dados observados são:  $\beta_0$ ,  $\beta_1$  e  $\sigma^2$ . Os dados consistem de n pares de observações  $(X_i, Y_i)$ . Para encontrar os “melhores” estimadores dos parâmetros da regressão será empregado o método dos mínimos quadrados (MMQ).

Para cada observação  $(X_i, Y_i)$  o MMQ considera o desvio de  $Y_i$  em relação ao seu valor esperado:

$$Y_i - (\beta_0 + \beta_1 X_i) \tag{2-8}$$

O MMQ exige que consideremos a soma quadrada dos n desvios. Este critério é denominado Q:

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{2-9}$$

Usando aproximações analíticas (KUTNER *et al.*, 2005), mostra-se que para o modelo de regressão (2-4) os valores  $b_0$  e  $b_1$  que minimizam  $Q$  para qualquer amostra de dados são as equações simultâneas abaixo:

$$\sum Y_i = nb_0 + b_1 \sum X_i \quad (2-10)$$

$$\sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2 \quad (2-11)$$

As equações (2-10) e (2-11) são chamadas **equações normais**;  $b_0$  e  $b_1$  são chamados **estimadores pontuais** de  $\beta_0$  e  $\beta_1$ , respectivamente.

As equações normais podem ser resolvidas simultaneamente para  $b_0$  e  $b_1$ :

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (2-12)$$

$$b_0 = \frac{1}{n} \left( \sum Y_i - b_1 \sum X_i \right) = \bar{Y} - b_1 \bar{X} \quad (2-13)$$

### 2.1.1.4.1.3 Distribuição dos Resíduos

Conforme visto acima, o ajuste de uma reta não envolve conceitos estatísticos. Basicamente, define-se o erro mínimo para a média do erro quadrático mínimo e manipulação matemática dos pares  $(x, y)$  de dados observados. Os conceitos estatísticos são apresentados a partir de suposições feitas sobre quantidades  $e_i$  (resíduos ou erros). De tal forma que, uma das convenções é considerar que os erros são variáveis aleatórias independentes com média zero e variância constante. Uma consideração adicional é a de que os resíduos têm uma distribuição normal. Conforme apresentado em KUTNER *et*



al. (2005), há dois motivos que justificam esta suposição: primeiro, os termos de erro usualmente representam os efeitos dos fatores omitidos do modelo. Ou seja, a distribuição do termo de erro tende a estar em conformidade com a hipótese de normalidade quando todas as variáveis (independentes) consideradas importantes para a regressão estão consideradas no modelo. O segundo é o fato de que procedimentos de estimação e testes baseados na distribuição t são mais sensíveis quando os dados fogem bastante da não normalidade. De tal forma que, a menos que a não normalidade seja séria, principalmente quando se considera a assimetria, os valores dos intervalos de confiança dos coeficientes serão próximos dos níveis de normalidade.

Uma das conveniências do ajustamento pelo MMQ é sua garantia de que a soma dos resíduos é zero:

$$\sum_{i=1}^n e_i = 0 \quad (2-14)$$

Observe na equação 2-14 que a média dos resíduos também é zero.

Considerar que os resíduos podem ser caracterizados em termos de sua variância é ponto principal em que a estatística se baseia na análise da regressão. A idéia implícita é que os resíduos se distribuem aleatoriamente em torno de uma média. A equação 2-14 mostra que o valor médio em torno do qual o erro se distribui é zero.

Um ponto central na inferência estatística para construção das regressões é a estimação da variância dos resíduos a partir dos resíduos amostrais.

Uma vez que a equação 2-14 garante que a média dos resíduos é zero e considerando que os desvios são os resíduos, temos que

$$Y_i - \hat{Y}_i = e_i \quad (2-15)$$

Logo,

$$SQE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 \quad (2-16)$$

A soma dos quadrados SQE tem (n-2) graus de liberdade. Dois graus foram perdidos, pois  $\beta_0$  e  $\beta_1$  são estimados para obtenção de  $\hat{Y}_i$ . Logo,

$$s^2 = QME = \frac{SQE}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum e_i^2}{n-2} \quad (2-17)$$

Onde o QME é o Erro Quadrático Médio. Mostra-se que QME é o estimador não tendencioso do desvio-padrão ( $\sigma^2$ ) (NETER, WASSERMAN, 1974) do modelo de regressão (2-4), isto é,  $E(QME) = \sigma^2$ .

A equação 2-15 pode ser decomposta da seguinte forma:

$$Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + Y_i - \hat{Y}_i \quad (2-18)$$

Ou seja, o desvio total  $Y_i - \bar{Y}$  pode ser visto como a soma de duas componentes:

- 1)  $\hat{Y}_i - \bar{Y}$ : desvio do valor ajustado em relação à média;
- 2)  $Y_i - \hat{Y}_i$ : desvio de  $Y_i$  em torno da linha de regressão.

Uma vez que a equação (2-18) elevada ao quadrado mantém a mesma relação (KUTNER *et al.*, 2005), temos:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2-19)$$

$$SQR = SQT - SQE \quad (2-20)$$

Onde SQR é a Soma de Quadrados da Regressão, SQT é a Soma de Quadrados Total e SQE é a Soma de Quadrados dos Erros.

A soma de quadrados dividida pelos graus de liberdade é chamada de quadrado médio. Logo,

$$QMR = \frac{SQR}{1} \quad (2-21)$$

$$QME = \frac{SQE}{n-2} \quad (2-22)$$

$$QMT = \frac{SQT}{n-1} \quad (2-23)$$

A qualidade do ajuste da reta pode ser medida em termos de algumas medidas estatísticas. Uma delas é o coeficiente de determinação ou  $R^2$ .

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT} \quad (2-24)$$

O  $R^2$  pode ser interpretado como a proporção de variação do preditando (proporcional a SQT) que é descrita ou considerada pela regressão (SQR).

Um ajuste perfeito resultaria em  $R^2 = 1$ . Um ajuste muito bom acarretaria um valor próximo de 1. Um ajuste fraco ocasiona um  $R^2$  próximo de zero.

Outra medida da qualidade do ajustamento da reta de regressão é a taxa F ou estatística F:

$$F = \frac{QMR}{QME} \quad (2-25)$$

O valor de F cresce de acordo com a qualidade da regressão, desde que uma forte relação entre a variável independente (X) e a variável dependente (Y) implica em alto valor para QMR e baixo QME. Assumindo que os resíduos são independentes e que seguem a distribuição normal, e sob a hipótese nula de que não existe relação linear, a distribuição de F tem uma distribuição paramétrica conhecida. Esta distribuição forma a

base do teste estatístico que é aplicado no caso de regressão linear simples, e de forma genérica, no caso de regressão múltipla (mais de uma variável independente).

#### **2.1.1.4.2 Regressão Linear Múltipla**

A regressão linear múltipla é a forma mais geral da regressão linear simples. Como no caso da simples, também apresenta uma variável dependente (Y), mas utiliza mais de uma variável independente (ou preditora).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (2-26)$$

onde,

$\beta_0, \beta_1, \dots, \beta_{p-1}$  são parâmetros;

$X_{i1}, X_{i2}, \dots, X_{i,p-1}$  são constantes conhecidas;

$\varepsilon_i$  são independentes com  $N(0, \sigma^2)$ ;

$i=1, \dots, n$ .

Cada uma das (p-1) variáveis preditoras tem seu próprio coeficiente, de forma análoga à inclinação da regressão simples, Os p preditores são chamados coeficientes da regressão. Em particular, o coeficiente  $\beta_0$  é denominado constante da regressão.

A equação (2-15) para os resíduos ainda é válida, se considerarmos que  $\hat{Y}_i$  é uma função de preditores,  $X_i$ ,  $i = 1, \dots, p-1$ . Quando temos duas variáveis preditoras, os resíduos podem ser visualizados como uma distância vertical. Neste caso, a função de regressão (2-26) é uma superfície ao invés de uma reta.

Os  $p$  parâmetros apresentados na equação (2-26) são obtidos, como visto para regressão simples, minimizando a soma dos erros ao quadrado. Isto é feito com o uso de matrizes, os detalhes não serão mostrados aqui. Na prática, estes cálculos são feitos por programas ou pacotes estatísticos.

Os valores de SQR, SQT e SQE continuam sendo calculados conforme apresentado na equação (2-19); e o valor de QMT conforme equação (2-23).

$$QMR = \frac{SQR}{p-1} \quad (2-27)$$

$$QME = \frac{SQE}{n-p} \quad (2-28)$$

A coeficiente de determinação é computado de acordo com a equação 2-24.

#### **2.1.1.4.2.1 Inferências sobre os parâmetros da regressão**

Neste trabalho serão utilizados para avaliação da qualidade das regressões, os coeficientes da regressão e as estatísticas mais relevantes: coeficiente de determinação múltiplo ( $R^2$ ), p-valor do teste t para coeficientes da regressão, erro padrão estimado dos coeficientes ( $s(b_k)$ ), estatística F da regressão e respectivo p-valor.

O teste de hipótese (estatística t), foi utilizado para testar se o parâmetro ajustado é estatisticamente significativo. As hipóteses do teste são:

$$H_0: \beta_k = 0$$

$$H_1: \text{Nem todos os } \beta_k (k = 1, \dots, p-1) \text{ igual a zero}$$

Nas tabelas do apêndice B, os p-valor associados com os coeficientes  $b_0, b_1, \dots, b_p$  foram calculados sob a hipótese nula  $H_0: \beta_k = 0$ , usando a estatística do teste ( $t^*$ ):

$$t^* = \frac{b_k}{s(b_k)} \sim t_{(n-p)} \quad k = 0, 1, \dots, p-1 \quad (2-29)$$

O teste de hipótese (estatística F), foi utilizado como teste auxiliar para avaliar a qualidade do ajuste. As hipóteses do teste são:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

$$H_1: \beta_k \neq 0 \quad (k = 1, 2, \dots, p-1)$$

Os p-valor associados à regressão foram calculados sob a hipótese nula  $H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ , usando a estatística do teste:

$$F^* = \frac{QMR}{QME} \sim F_{(1-\alpha; p-1; n-p)} \quad (2-30)$$

Os p-valor dão o grau de confiança ao rejeitarmos uma hipótese nula, a qual poderá ser rejeitada se o p-valor for menor ou igual a um dado nível de significância  $\alpha$ . Neste trabalho usaremos um nível de significância de 5% para avaliarmos os resultados dos experimentos. Ou seja, neste trabalho, valores de p-valor menores que 5% não sustentam a hipótese nula e esta poderá ser rejeitada.

O p-valor é considerado uma medida de confiança na hipótese nula, isto é, quanto maior seu valor maior a probabilidade da hipótese nula ser verdadeira. O seu cálculo baseia-se na estatística do teste considerando que a hipótese nula é verdadeira. Seu valor é obtido por meio de consulta feita à tabela da respectiva estatística do teste.

O  $R^2$  mostra o quanto da variação total da variável dependente é reduzida quando se utiliza a reta de regressão. Outra definição é que o  $R^2$  dá a proporção da variação total de Y que é considerada por todo o conjunto de regressores em uma regressão múltipla. O  $R^2$  assume o valor 0 quando todos os  $b_k = 0$  ( $k=1, \dots, p-1$ ) e

assume o valor 1 quando todas as observações se ajustam perfeitamente na superfície prevista, isto é, quando  $Y_i = \hat{Y}_i$  para todo  $i$ .

No desenvolvimento desta dissertação estas estatísticas serão utilizadas a fim de se avaliar da qualidade das regressões e, conseqüentemente, de suas previsões.

### **2.1.1.4.3 Screening Regression**

*Screening regression* é o termo usado em estatística que combina regressão linear múltipla com um método objetivo de selecionar um conjunto de preditores considerados “ótimos” para ser usado na equação, a partir de um conjunto inicial maior de potenciais preditores.

A necessidade de técnicas deste tipo ocorre, pois variáveis preditoras são quase sempre correlacionadas entre si, ou seja, apresentam informações redundantes. A inclusão de preditores altamente correlacionados entre si, tende a fornecer estimativas dos coeficientes com alta variância amostral. Esta situação não deve ser ignorada, pois os indicadores de qualidade (como o  $R^2$ ) e os testes são afetados, podendo levar a tomar decisões erradas sobre variáveis importantes para o modelo.

O procedimento adotado no presente estudo para a seleção dos preditores é o chamado Seleção para Frente (*forward selection*). O procedimento pode ser resumido como segue:

1) É calculada a regressão linear simples para todas (p-1) variáveis independentes. Para cada regressão simples é calcula a estatística  $F^*$  (equação 2-31) parcial<sup>1</sup> para testar se a inclinação da reta é zero. A variável independente com o maior valor de  $F^*$  é candidata a entrar. Se o valor de  $F^*$  excede um determinado valor, a

---

<sup>1</sup> A estatística  $F$  do teste apresentada em (2-30) testa se todos os coeficientes  $\beta_k = 0$ , isto é, se há ou não relação entre  $Y$  e o conjunto de variáveis independentes  $X$ . A estatística apresentada em (2-31) é chamada estatística  $F$  parcial e mede se um coeficiente em particular é igual a zero.

variável independente será incluída no modelo. Caso contrário a seleção termina considerando que não há variável suficientemente significativa para entrar no modelo de regressão.

$$F_K^* = \frac{QMR(X_k)}{QME(X_k)} \quad (2-31)$$

2) Considere-se, por exemplo, que a variável independente  $X_7$  foi incluída no modelo no passo 1. Calcula-se agora todas as regressões com duas variáveis independentes, e  $X_7$  é uma delas. Para cada uma das regressões, a estatística  $F^*$ , dada abaixo, é obtida.

$$F_K^* = \frac{QMR(X_k / X_7)}{QME(X_k / X_7)} = \left[ \frac{b_k}{s(b_k)} \right]^2 \quad (2-32)$$

Esta estatística testa se  $\beta_k = 0$  quando  $X_7$  e  $X_k$  são as variáveis independentes no modelo.

3) Suponha agora que a variável  $X_3$  foi incluída no passo 2. Agora o modelo tem duas variáveis independentes,  $X_3$  e  $X_7$ . Calculam-se então todas as regressões com três variáveis independentes e avalia-se, conforme passo 2, qual a candidata a entrar e daí por diante. O cálculo só finaliza quando nenhuma variável pode ser incluída no modelo.

Observe na fórmula (2-32) a equivalência da estatística  $F$  com o teste  $t$ . Neste trabalho foi utilizada a rotina do *Fortran 90* chamada RSTEP (IMSL 1987b), para a construção do modelo de regressão usando a seleção para frente. Nesta rotina utiliza-se a estatística  $t^*$  para seleção das variáveis. A demonstração da equivalência das estatísticas pode ser encontrada em KUTNER *et al.* (2005).



#### **2.1.1.4.4 Componentes Principais como variáveis preditoras**

KUTNER *et al.* (2005) sugerem o uso de componentes principais como variáveis preditoras para resolver o problema de multicolinearidade<sup>2</sup>. Esta técnica tornou-se popular na análise de dados atmosféricos após estudo feito por LORENZ (1956, *apud* WILKS, 2006).

A análise de componentes principais (CP) fornece um ou vários índices, chamados componentes. Estas componentes, que são combinações lineares das variáveis originais, não são correlacionadas entre si. Normalmente, apenas algumas destas componentes capturaram a maior parte da variabilidade das variáveis originais. A partir daí é possível reduzir um conjunto de dados contendo um grande número de variáveis para um conjunto contendo um número bem menor de novas variáveis.

Segue abaixo uma breve descrição da Análise de Componentes Principais.

Dado um vetor  $\mathbf{x}$  (com  $K \times 1$  observações), espera-se encontrar ( $M \times 1$ ) vetores  $\mathbf{u}$ , cujos elementos são combinações lineares dos elementos de  $\mathbf{x}$ , que contêm a maior parte da informação da coleção original de valores de  $\mathbf{x}$ , e cuja dimensionalidade é  $M \ll K$ . Este objetivo pode ser atingido se existe substancial correlação entre as variáveis contidas em  $\mathbf{x}$ , o que significa que  $\mathbf{x}$  contém informação redundante. Os elementos destes novos vetores  $\mathbf{u}$  são chamados de componentes principais.

A análise de CP é baseada na matriz de covariância amostral  $S$  (ou matriz de correlação). Considere a matriz para  $k$  variáveis:

---

<sup>2</sup> Multicolinearidade ou intercorrelação ocorre quando as variáveis preditoras são correlacionadas entre si. Normalmente, o primeiro termo é usado quando esta correlação é muito alta.

$$S = \begin{bmatrix} s_1^2 & s_{12} & \dots & s_{1k} \\ s_{21} & s_2^2 & \dots & s_{2k} \\ & & \dots & \\ s_{k1} & s_{k2} & \dots & s_k^2 \end{bmatrix} \quad (2-33)$$

Onde,  $s_i^2$  é a variância da  $i$ -ésima variável  $x_i$ , e  $s_{ij}$  é a covariância entre a  $i$ -ésima e  $j$ -ésima variáveis. Se as covariâncias não são iguais a zero, isso indica que a relação linear existe entre essas duas variáveis sendo a ‘intensidade’ dessa relação representada pelo coeficiente de correlação  $r_{ij} = \frac{s_{ij}}{(s_i s_j)}$ .

O uso da matriz de correlação é indicado quando as variáveis observadas possuem dimensão ou escalas muito distintas. Neste estudo foi utilizada a matriz de correlação, pois as variáveis usadas apresentavam escalas muito diferentes.

Observe que [S] possui grande parte das informações da variabilidade apresentadas em  $x$ . Para o cálculo das CP utilizam-se as anomalias ( $x'$ ) das variáveis. A anomalia é obtida pela subtração da média amostral dos valores  $x_k$ .

$$x' = x - \bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_k \end{bmatrix} \quad (2-34)$$

As anomalias apresentam a mesmas variâncias e correlações mútuas das variáveis originais. Mostra-se que a matriz de covariância de  $x$  e  $x'$  são idênticas (WILKS, 2006).

A nova variável  $u_m$ , que leva em conta a máxima variabilidade de  $x'$  (e portanto de  $x$ ), é obtida a partir do cálculo dos auto-vetores de [S]. A  $m$ -ésima componente

principal  $\mathbf{u}_m$  é encontrado a partir da projeção do vetor de dados  $\mathbf{x}'$  sobre o m-ésimo autovetor  $\mathbf{e}_m$ ,

$$\mathbf{u}_m = \mathbf{e}_m^T \mathbf{x}' = \sum_{k=1}^K e_{km} x'_k, \quad m=1, \dots, M \quad (2-35)$$

Observe que cada uma das componentes principais é um conjunto de médias ponderadas dos valores  $\mathbf{x}'$ . Se o conjunto de dados consiste de  $n$  observações, haverá  $n$  valores em cada componente principal, ou nova variável,  $\mathbf{u}_m$ .

Geometricamente, o primeiro autovetor  $\mathbf{e}_1$ , aponta na direção (no espaço dimensional  $K$  de  $\mathbf{x}'$ ) em que o vetor de dados juntamente exhibe a maior variabilidade. O primeiro autovetor  $\mathbf{e}_1$  está associado ao maior autovalor  $\lambda_1$ . O segundo autovetor  $\mathbf{e}_2$ , associado com o segundo maior autovalor  $\lambda_2$ , é definido tal que seja perpendicular a  $\mathbf{e}_1$ , é sob esta condição que ele se alinha na direção em que os vetores  $\mathbf{x}'$  exibem suas maiores variações. Os autovetores subsequentes  $\mathbf{e}_m$ ,  $m=3, 4, \dots, M$ , são também numerados em ordem decrescente de magnitude, sendo cada um perpendicular ao autovetor anterior. Sob estas condições, esses autovetores buscam localizar as direções em que os dados originais, conjuntamente, apresentam a maior variabilidade.

Em resumo, os autovetores definem um novo sistema de coordenadas para o dados. Este sistema de coordenadas é orientado de forma que cada novo eixo é alinhado ao longo da direção de maior variabilidade conjunta dos dados. Observe que este sistema será diferente para cada conjunto de dados utilizados, pois o cálculo utiliza a matriz de covariância amostral  $[\mathbf{S}_x]$  a partir de um conjunto de dados particular.

Uma das propriedades mais importantes das componentes principais é que elas não são correlacionadas. Ou seja, a matriz de correlação de  $\mathbf{u}_m$  é a matriz identidade,  $[\mathbf{I}]$ . A partir desta informação temos que a covariância para todos pares  $\mathbf{u}_m$  é zero, ou seja,

sua matriz de covariância é diagonal. Esta informação permite-nos obter a matriz de covariância dos componentes principais a partir da diagonalização de  $[S_x]$ :

$$[S_u] = Var([E]^T x) = [E]^T [S_x] [E] = [E]^{-1} [S_x] [E] = [\Lambda] \quad (2-36)$$

Ou seja, a variância da  $m$ -ésima componente principal  $u_m$  é o  $m$ -ésimo autovalor  $\lambda_m$ . Mostra-se que cada componente principal representa uma parte da variância total de  $x$ .

$$R_m^2 = \frac{\lambda_m}{\sum_{k=1}^K \lambda_k} \times 100\% = \frac{\lambda_m}{\sum_{k=1}^K s_{k,k}} \times 100\% \quad (2-37)$$

A fórmula acima mostra que a variação total dos dados originais é completamente considerada no conjunto total dos  $K$   $u_m$ .

Considere a matriz ortogonal  $[E]$  cujas colunas são autovetores, temos:

$$u = [E]^t x' \quad (2-38)$$

A equação acima define a transformação do vetor de dados  $x'$  vetor de componentes principais  $u$ . Para obtenção dos dados originais a partir das componentes principais, sob certas considerações, temos que:

$$\begin{matrix} \mathbf{x}' \\ (\mathbf{K} \times 1) \end{matrix} = \begin{matrix} [E] \\ (\mathbf{K} \times \mathbf{K}) \end{matrix} \begin{matrix} \mathbf{u} \\ (\mathbf{K} \times 1) \end{matrix} \quad (2-39)$$

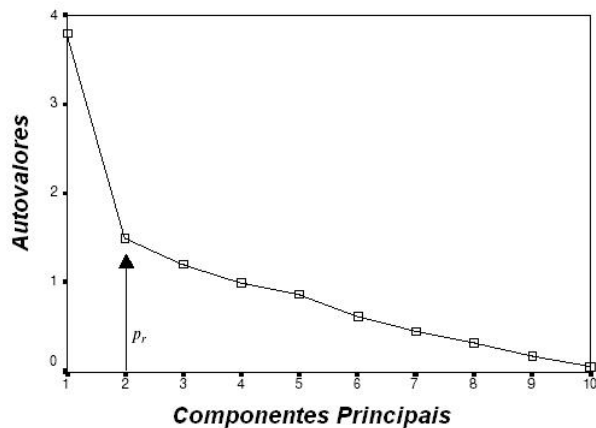
Considerando que na prática são utilizados  $M < K$  componentes principais (normalmente, os  $M$  maiores autovalores), a reconstrução dos dados é aproximada:

$$\begin{matrix} \mathbf{x}' \\ (\mathbf{K} \times 1) \end{matrix} \sim \begin{matrix} [E] \\ (\mathbf{K} \times \mathbf{M}) \end{matrix} \begin{matrix} \mathbf{u} \\ (\mathbf{M} \times 1) \end{matrix} \quad (2-40)$$

Uma das principais qualidades do uso de CP, além do fato, considerado nesta dissertação, da não correlação entre elas, é a sua capacidade de compressão (PREISENDORFER, 1988, *apud* HSIEH, TANG, 1998). Ou seja, o uso de  $M \ll K$  componentes principais. Não há um critério único que forneça a quantidade ideal de

componentes que devem ser considerados em determinado estudo. Nesta dissertação será utilizado, inicialmente, o método subjetivo apresentado em WILKS (2006) que se baseia na análise do gráfico *scree plot*

O número de componentes a reter é identificado quando a curva evidenciar uma estabilização. Pode-se truncar na componente anterior à estabilização da curva. Vide exemplo na Figura 2-2. Nos caso em que a curva não apresente uma inclinação tão evidente, será utilizado o critério sugerido por JOLLIFFE (2002, *apud* WILKS, 2006) em que serão retidos componentes principais que considerem entre 70% e 90% da variação total dos dados.



**Figura 2-2** *Scree plot* – distribuição dos autovalores em função dos respectivas componentes principais.

#### **2.1.1.4.5 Regressão para estimação de eventos probabilísticos**

A capacidade da estatística de produzir previsões probabilísticas é uma de suas vantagens sobre os métodos de previsão determinísticos (fornecidos por modelos numéricos). WILKS (2006) observa duas capacidades importantes:

- a) fornece uma expressão explícita da incerteza inerente sobre a previsão do tempo futuro;

b) permite ao previsor extrair mais informações das probabilidades no momento da tomada de decisão.

Uma das formas mais usadas de obter previsões probabilísticas é o desenvolvimento de regressões onde a variável dependente é binária. Em Meteorologia é chamada Regressão para Estimação de Eventos Probabilísticos (REEP), mas em Estatística pode ser demonstrado formalmente como abaixo:

Considere a regressão linear simples abaixo:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad Y_i = 0,1 \quad (2-41)$$

Onde  $Y$  só assume os valores 0 ou 1. O valor esperado  $E(Y_i)$  tem um significado especial. Uma vez que  $E(\varepsilon_i) = 0$ . Temos,

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad (2-42)$$

Considere  $Y_i$  como uma variável aleatória de Bernoulli para a qual estabelecemos a distribuição de probabilidade conforme abaixo:

$Y_i$	Probabilidade
0	$P(Y_i = 0) = 1 - p_i = q_i$
1	$(Y_i = 1) = p_i$

Portanto,  $p_i$  é a probabilidade de que  $Y_i = 1$  e  $q_i$  é a probabilidade de que  $Y_i = 0$ .

Pela definição de valor esperado de uma variável aleatória, nós temos que:

$$E(Y_i) = 0(q_i) + 1(p_i) \quad (2-43)$$

Das equações (2-42) e (2-43), encontramos:

$$\beta_0 + \beta_1 X_i = p \quad (2-44)$$

Ou seja,  $E(Y_i) = \beta_0 + \beta_1 X_i$  é a probabilidade de que  $Y_i = 1$ . Esta interpretação se aplica também para as regressões lineares múltiplas.

Um dos problemas deste método é que as previsões não são limitadas por zero e um, isto é, podem ocorrer probabilidades negativas ou maiores que um, contrariando uma das propriedades básicas da probabilidade. No entanto, isto causa pouca dificuldade operacionalmente, pois equações de probabilidades com vários preditores raramente produzem este tipo de resultado. Quando o problema ocorre, a probabilidade costuma ser próxima de zero ou de um, e os resultados operacionais podem ser considerados como tal.

Outro fator a ser considerado é que o fato de a variável dependente ser binária implica na violação das suposições necessárias para inferências sobre os coeficientes e sobre a reta de regressão. Contudo, a experiência tem mostrado que, geralmente, podemos obter resultados estáveis quando utilizamos amostras grandes (MURPHY, KATZ, 1985). Assim as inferências relacionadas com a regressão e seus coeficientes podem ser feitas (NETER, WASSERMAN, 1974).

## **2.1.2 Fundamentos de Redes Neurais**

### **2.1.2.1 Introdução**

A Inteligência Artificial busca, através de técnicas inspiradas no comportamento do cérebro, o desenvolvimento de sistemas inteligentes que imitem aspectos do comportamento humano, tais como: aprendizado, percepção, raciocínio, evolução e adaptação.

### **2.1.2.2 Breve Histórico**

Em 1943, Warren McCulloch e o estatístico Walter Pitts publicaram um artigo com o título “A Logical Calculus of the Ideas Immanent in Nervous Activity” que hoje é referência para a teoria das redes neurais artificiais. Descontadas todas as imprecisões, foi o primeiro artigo de redes neurais.

HEBB (1949) publicou o livro “The Organization of Behavior”, apresentando pela primeira vez, uma regra de aprendizagem fisiológica para a modificação sináptica, conhecida como postulado de aprendizagem.

Em 1951 Minski constrói o Snark, primeiro neurocomputador com capacidade de aprendizado, ou seja, ajustava automaticamente os pesos entre as sinapses.

NEUMANN (1958) publicou o livro chamado “The Computer and the Brain”, cujo principal tema foi a questão da memória no sistema nervoso biológico, comentando que o significado exato ou o modo de armazenagem da informação ainda não era conhecido.

Frank Rosenblatt, posteriormente, inventou uma classe de redes simples, baseadas em neurônios, por ele denominadas PERCEPTRON, que foram apresentadas em seu livro, “Principles of Neurodynamics” (1962).

Na mesma época em que Rosenblatt trabalhava no PERCEPTRON, WIDROW (1962) desenvolveu um modelo neural linear muito simples, que ele chamou de ADALINE (ADAPtative LInear NETwork), e a sua generalização multidimensional MADALINE (Many ADALINE). A contribuição realmente importante deste trabalho foi a invenção de um princípio de treinamento extremamente poderoso para as redes de adalines conhecido como a Regra Delta, que foi mais tarde generalizado para redes neurais mais elaboradas.



MINSKY e PAPPERT (1969) no livro “Perceptrons,” provam que, o *perceptron* de uma única camada descrito por Rosenblatt era incapaz de resolver problemas simples de classificação de padrão. Após a publicação deste livro deu-se início a uma era de obscurantismo em pesquisas com redes neurais que só foi encerrada no início da década de 1980 com o trabalho de Hopfield em memórias associativas.

Na década de 80, HOPFIELD (1982) utilizou a idéia de uma função de energia para formular um novo modo de entender a computação executada por redes recorrentes com conexões sinápticas simétricas, estabelecendo assim, o isomorfismo entre uma rede recorrente e modelos utilizados na Física Estatística. Este caso particular de redes neurais com realimentação passou a ser chamada de redes de Hopfield.

Por vários anos não houve um algoritmo para treinar redes de neurônios de múltiplas camadas o que implicava em uma limitação fundamental para o treinamento de redes complexas. A invenção do algoritmo *backpropagation* ou da retropropagação do erro tem grande influência na retomada dos estudos das redes neurais artificiais, contribuíram nesse desenvolvimento Rumelhart *et al.* (1986), PARKER(1985), LeCUN (1985, *apud* HAYKIN, 1994) e WERBOS (1974).

BROOMHEAD e LOWE (1988) descreveram o procedimento para o projeto de redes neurais alimentadas adiante (*feedforward*) em camadas, utilizando funções de base radial, as quais fornecem alternativas aos “perceptrons” de múltiplas camadas.

Atualmente, o interesse na aplicação de redes neurais nas áreas de Meteorologia e Oceanografia tem aumentado. Principalmente no pós-processamento das previsões de modelos atmosféricos, esperando assim que estas possam ser beneficiadas.

MARZBAN (1998) desenvolveu uma rede neural para diagnosticar quais circulações detectadas por um algoritmo de detecção de mesociclones podem ser

classificadas como ventos severos. O desempenho da rede foi avaliado por meio de várias medidas escalares.

HSIEH e TANG (1998) afirmam que são três os obstáculos para aplicação de redes neurais em Meteorologia e Oceanografia: (a) instabilidade não linear quando o conjunto de dados é pequeno, (b) campos espacialmente grandes com um número elevado de dados a serem processados, e (c) dificuldade para interpretar os resultados não lineares das redes neurais. Estudos recentes, contudo, avançaram na solução dos problemas acima. No caso do item (a) emprega-se a técnica de média dos conjuntos (*ensemble average*). Para o item (b) utiliza-se método dos componentes principais. Para (c) emprega-se a análise espectral para ajudar no entendimento das relações não lineares da rede neural.

KULIGOWSKI e BARROS (1998) descrevem um modelo de previsão de precipitação baseado em rede neurais artificiais. Neste estudo só foram utilizados dados observados como preditores da rede neural.

MARZBAN (2003) utilizou rede neural para a previsão de temperatura para o Sistema de Previsão Regional Avançado (ARPS). Utilizou 31 (trinta e uma) estações e desenvolveu uma rede para cada uma delas. A temperatura prevista pela rede apresenta bom desempenho quando avaliada em termos de algumas medidas escalares.

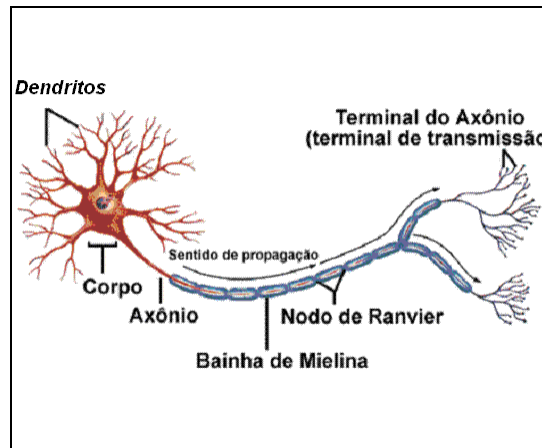
KRETZSCHMAR *et al.* (2004) avaliaram o desempenho da rede neural para classificação de velocidades de vento e rajadas. Eles utilizaram séries temporais locais e dados de modelo numérico como variáveis preditoras. A rede neural mostrou bom desempenho em relação à persistência para as medidas de alarme falso e probabilidade de detecção.

### **2.1.2.3 A Inspiração Biológica**

As Redes Neurais Artificiais surgiram em consequência da busca por conhecimento a respeito da mente humana. Com isso, houve um grande interesse em pesquisar o papel do funcionamento de estruturas do sistema nervoso, o qual motivou a construção de modelos matemático-computacionais que pudessem auxiliar na elucidação de aspectos neurobiológicos envolvidos em diversas atividades cognitivas. As Redes Neurais Artificiais têm como base as funções mais elementares de um neurônio biológico.

Os neurônios são células presentes no sistema nervoso e estão envolvidos na recepção, transmissão e processamento dos estímulos que chegam ou partem do organismo. As operações individuais destes neurônios são bastante simples, mas são as conexões entre as células que permitem a realização de várias tarefas pelo sistema nervoso.

A figura abaixo apresenta as diferentes partes dos neurônios: corpo (soma), dendritos e axônio. Os dendritos têm por função receber os estímulos transmitidos pelos outros neurônios. O corpo de neurônio, também chamado de soma, é responsável por coletar e combinar informações vindas de outros neurônios, e finalmente o axônio, que é constituído de uma fibra tubular que pode alcançar até alguns metros é responsável por transmitir os estímulos para outras células.



**Figura 2-3 Estrutura de um neurônio.**

#### **2.1.2.4 Neurônio Artificial**

O cérebro humano tem bilhões de células nervosas (neurônios), cada uma conectada a outras centenas de células. Um estímulo qualquer (ouvir música, por exemplo) inicia um processo que vai do ouvido ao cérebro. O primeiro neurônio recebe a informação (entrada) então decide se esta será disparada ou não. Se ele decide que não será disparada, o sinal é finalizado; mas caso seja disparado, este envia o sinal do seu axônio para um outro neurônio. Cada conexão na junção das sinapses tem químicas que inibem ou estimulam o sinal. O próximo neurônio então soma as informações (ou entradas) do neurônio anterior e de todos os outros aos quais ele esta conectado. Então a decisão de disparar ou não dependerá desta soma. O processo continua até que o cérebro reconheça o estímulo. Decidir a resposta (resultado ao estímulo) é parte deste processo. O cérebro, no futuro, ao receber o mesmo estímulo terá uma resposta previsível. Isto é o que chamamos aprendizagem.

No neurônio artificial cada informação (entrada) é multiplicada pelos pesos de suas conexões com os neurônios. Os pesos das conexões podem ser positivos (excitam) ou negativos (inibem). Estes pesos ajudam a determinar quais entradas são importantes e quais não são. Todas as entradas são somadas. Então uma função de transferência (ou

ativação) determina a saída. As saídas de um neurônio podem ser conectadas a vários outros e se tornar as entradas destes.

O neurônio artificial é uma estrutura lógico-matemática que procura simular a forma, o comportamento e as funções de um neurônio biológico. Assim sendo, os dendritos foram substituídos por entradas, cujas ligações com o corpo celular artificial são realizadas através de elementos chamados de peso (simulando as sinapses). Os estímulos captados pelas entradas são processados pela função de soma, e o limiar de disparo do neurônio biológico foi substituído pela função de ativação.

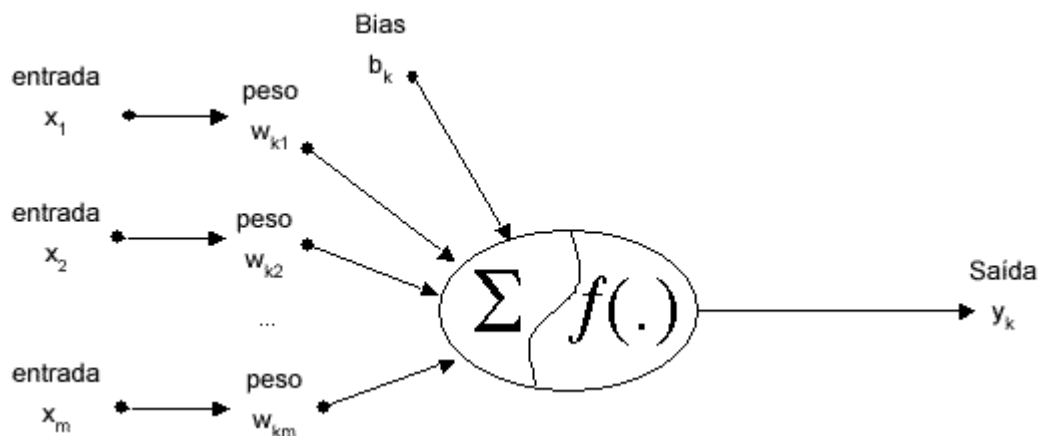
Os neurônios são as unidades de processamento das redes neurais e, portanto, são fundamentais para o seu funcionamento. A Figura 2-4 exibe um modelo simplificado de neurônio artificial, que é composto de :

Estímulos de entrada (X): onde as componentes são representadas pelo vetor X;

Ligações sinápticas (W): onde as componentes são representadas pelo vetor W;

Saída (Y): onde as componentes são representadas pelo vetor Y que produz segundo o valor de ativação alcançado pelo neurônio.

Neste modelo, os pesos sinápticos são fixos e a saída obtida pela aplicação de uma função de limiar (ou degrau).



**Figura 2-4 Neurônio artificial**

O modelo acima foi apresentado por McCulloch e Pitts em 1940. Depois disso, surgiram outras abordagens, dando flexibilidade aos pesos e maior capacidade à rede neural; através dos neurônios com funções de ativação não-lineares, das arquiteturas com mais de uma camada e algoritmos apropriados para alterar os pesos sinápticos.

Nos neurônios artificiais os seguintes elementos estão envolvidos:

- Conjunto de sinapses ( $W$ ): Ligações entre neurônios. Cada ligação possui um valor (peso), que representa a sua força: os estímulos de entrada são multiplicados pelos respectivos pesos de cada ligação, podendo gerar um sinal tanto positivo (excitatório) quanto negativo (inibitório).
- Combinador Linear ( $\Sigma$ ): Executa o somatório dos sinais produzidos pelo produto entre os pesos sinápticos e as entradas fornecidas ao neurônio. Em outras palavras, é o integrador dos sinais que chegam ao neurônio.

A saída do neurônio é definida pelo seu valor de ativação calculado da seguinte forma:

$$v_j = \sum_{i=1}^m w_{ji} x_i + b \quad (2-45)$$

Onde:

$v$  é o valor de ativação do neurônio  $k$ ;

$w$  são os pesos das conexões do neurônio  $k$ ;

$x$  é o valor de cada um dos  $m$  estímulos que chegam ao neurônio  $k$ ;

$b$  é o valor do *bias* que será somado ao valor do combinador linear para compor o valor de ativação. A função de ativação fornece o valor da saída de um neurônio.

### 2.1.2.4.1 Funções de Ativação

#### Função Limiar (Degrau)

Utilizada no modelo de McCulloch e Pits (HAYKIN, 1994), a função limiar modela a característica "tudo-ou-nada" deste neurônio.

É expressa da seguinte forma:

$$f(v) = \begin{cases} 1, & \text{se } v \geq 0; \\ 0, & \text{se } v < 0; \end{cases} \quad (2-46)$$

Nos neurônios construídos com essa função, a saída  $y$  será igual a zero, caso o valor de ativação  $v$  seja negativo e 1 nos casos em que o valor de ativação seja positivo.

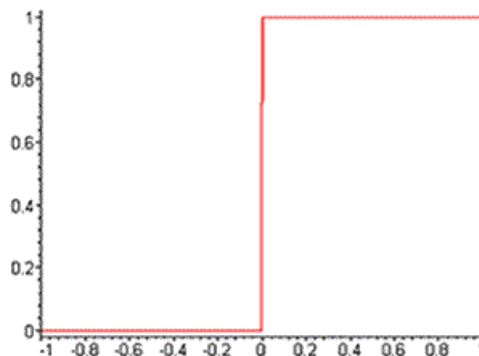


Figura 2-5 Gráfico da função de Limiar

Devemos lembrar que o valor de ativação ( $v$ ) é composto pelo combinador linear e pelo *bias*:

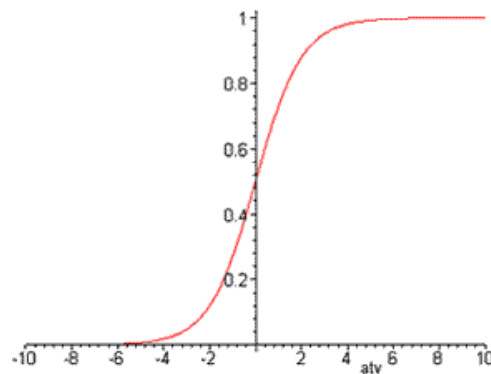
$$v_j = \sum_{i=1}^m w_{ji} x_i + b \quad (2-47)$$

### Função Sigmóide:

Esta função, ao contrário da função limiar, pode assumir todos os valores entre 0 e 1. A representação mais utilizada para esta função é a função logística, definida por:

$$f(v) = \frac{1}{1 + e^{(-av)}} \quad (2-48)$$

Onde  $a$  é o parâmetro de inclinação da função sigmóide e  $v$  é o valor de ativação do neurônio.



**Figura 2-6** Gráfico da função sigmóide

Pode ser necessário, porém, que a função de ativação assuma valores entre 1 e -1. Esta característica traz benefícios analíticos. Para obtermos tal intervalo de valores utilizamos as funções Signum, no caso da função de limiar, e a função tangente hiperbólica, no caso da Função Sigmóide. Segue abaixo a definição da tangente hiperbólica.

Como a Função Logística, também possui forma de "s", assumindo valores entre 1 e -1, sendo representada por:



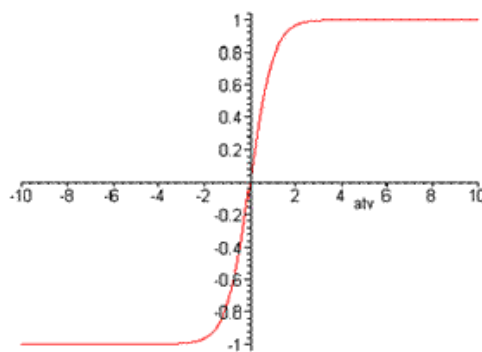
$$f(v) = a \frac{e^{(bv)} - e^{(-bv)}}{e^{(bv)} + e^{(bv)}} \quad (2-49)$$

Onde:

a é o parâmetro de inclinação da curva;

b são os limites inferiores e superiores ( $b = |1|$  no gráfico);

v é o valor de ativação da unidade.



**Figura 2-7** Gráfico da função tangente hiperbólica

### 2.1.2.5 Rede Perceptron

No final da década de 1950, Rosenblatt deu prosseguimento às idéias de McCulloch & Pitts. Ele criou uma rede de múltiplos neurônios do tipo discriminadores lineares e chamou esta rede de *Perceptron*. Esse modelo consiste de um único neurônio com pesos sinápticos ajustáveis e *bias*, treinado para classificar um conjunto de padrões em duas classes, devido a sua saída binária. O treinamento é feito com o algoritmo do *perceptron* que converge na superfície limite entre as duas classes. A convergência desse algoritmo ficou conhecida como Teorema de Convergência do *Perceptron*. Por meio deste algoritmo a rede busca uma solução através de pesos corretos, iniciando com alguma estimativa do processo e modificando-os com base na sua própria experiência.

### 2.1.2.6 Rede Perceptron de Múltiplas Camadas

Uma rede neural típica consiste de uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (Figura 2-8). Os sinais da entrada são propagados para frente através das camadas da rede. Estas redes são chamadas *Perceptrons* de Múltiplas Camadas e são uma generalização dos *perceptrons* de uma única camada.

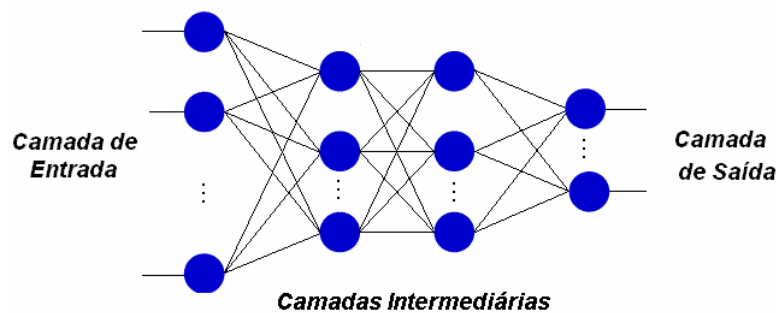


Figura 2-8 – *Perceptron* de múltiplas camadas

As principais características dessas redes são:

- Utilização de função de ativação não-linear com forma suave, diferenciável em qualquer ponto. A presença da não linearidade é importante, pois do contrário as relações entre as entradas e as saídas da rede estariam reduzidas a uma rede *perceptron* de só uma camada;
- a existência de uma ou mais camadas de neurônios intermediárias permite que a rede execute tarefas complexas extraíndo as características mais importantes dos padrões de entrada;
- exibe alto grau de conectividade determinado pelas sinapses da rede (HAYKIN, 1994).

A capacidade computacional destas redes surge a partir das características listadas acima e de sua capacidade da rede aprender a partir da experiência. Essas redes se destacam pela capacidade de abstração em representar as características do fenômeno através das informações encontradas na camada de entrada. Possuem alto grau de generalização, mesmo quando os padrões não pertencem ao conjunto de dados fornecidos à rede. Outra característica é a robustez, ou seja, a retirada de algum neurônio não compromete, normalmente, a funcionalidade da rede, sendo capaz de reconhecer padrões mesmo na presença de distorções.

A característica mais significativa de redes neurais está em sua habilidade de aproximar qualquer função contínua exatamente. Esta habilidade das redes neurais as tem tornado útil para modelar sistemas não lineares.

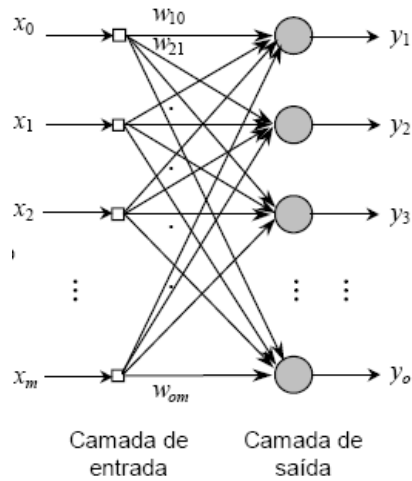
#### **2.1.2.6.1 Definindo a arquitetura da rede neural**

A modelagem da arquitetura de uma Rede Perceptron Múltiplas Camadas envolve a escolha da quantidade de camadas e o número de neurônios em cada camada.

A topologia da rede se caracteriza pela definição dos seguintes parâmetros: número de camadas, número de neurônios ou nós em cada camada e a conexão entre os neurônios.

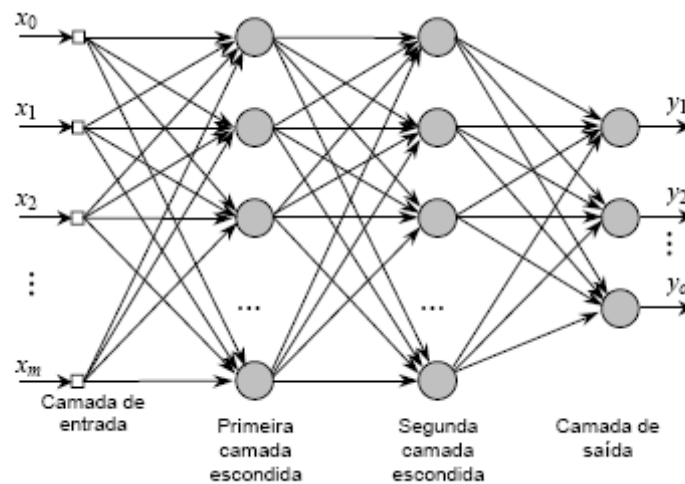
Quanto às camadas as redes neurais podem ser classificadas em:

- 1) Rede de camada única, onde existe apenas uma unidade de processamento (Figura 2-9). Os neurônios da camada de entrada correspondem aos neurônios sensoriais que possibilitam a entrada de sinais na rede (não fazem processamento). Os neurônios da camada de saída fazem processamento.



**Figura 2-9 Rede de camada única**

2) Redes de múltiplas camadas têm uma ou mais camadas intermediárias ou escondidas.



**Figura 2-10 Rede de múltiplas camadas**

O processamento de cada unidade é influenciado pelo processamento efetuado pelas unidades das camadas anteriores. Cada camada desempenha um papel específico, por exemplo:

- Camada de Entrada: receptora de estímulos, ou seja, são unidades de entrada que fornecem para a rede os padrões que serão distribuídos para as outras camadas;

- Camadas escondidas ou intermediárias: onde são realizados os processamentos por meio de funções de ativação e atualização dos pesos sinápticos. Esta camada define a não linearidade da rede neural;

- Camada de Saída: combina as regiões formadas pela camada anterior, definindo o espaço de saída da rede. Constitui o resultado final da rede o qual é comparado com o valor desejado.

As camadas intermediárias da rede são como detectores de características, as quais serão representadas, internamente, através dos pesos sinápticos.

Para a escolha do número de neurônios em cada camada, deve-se considerar:

- O número de épocas de treinamento;
- A quantidade de ruído presente nas épocas;
- A complexidade da função a ser aprendida pela rede;

Ao determinar o número de neurônios em cada camada, os seguintes cuidados devem ser tomados:

- Não utilizar um número de neurônios maior que o necessário: número alto de unidades pode fazer com que a rede memorize os dados do treinamento; com isto ela torna-se incapaz de generalizar e, portanto, reconhecer padrões não vistos durante o treinamento. Isto é chamado de *overfitting*.

- Não utilizar um número de neurônios inferior ao número necessário, isto pode fazer com que a rede gaste muito tempo para aprender, podendo não alcançar os pesos adequados, ou seja, a rede pode não convergir, ou generalizar demais os padrões de entrada.

Quanto ao tipo de conexão dos neurônios a redes classificam-se em:

1) Redes neurais diretas ou *feedforward*: nesta rede as conexões entre as unidades não formam ciclos. O sinal passa em todos os pontos de conectividade da rede em uma única direção.

2) Redes recorrentes: nesta rede as conexões formam ciclos. O sinal passa através de caminhos fechados onde ocorrem laços de realimentação.

### **2.1.2.7 Processo de aprendizagem**

Para que uma rede neural artificial possa fornecer resultados convenientes, é necessário que passe por uma fase de treinamento, onde seus pesos são ajustados de forma que ela se adapte aos diferentes estímulos de entrada. Durante esta fase de treinamento ocorre o aprendizado.

Há vários processos de aprendizado, os quais, de forma geral, podem ser classificados em:

- **Aprendizado supervisionado:** é fornecida uma referência do objetivo a ser alcançado. Os valores esperados ou desejados são conhecidos e são apresentados à rede durante o treinamento. O ajuste dos pesos é feito a partir da diferença entre o valor desejado e a saída prevista pela rede.
- **Aprendizado não supervisionado:** neste caso não é fornecida nenhuma referência externa.

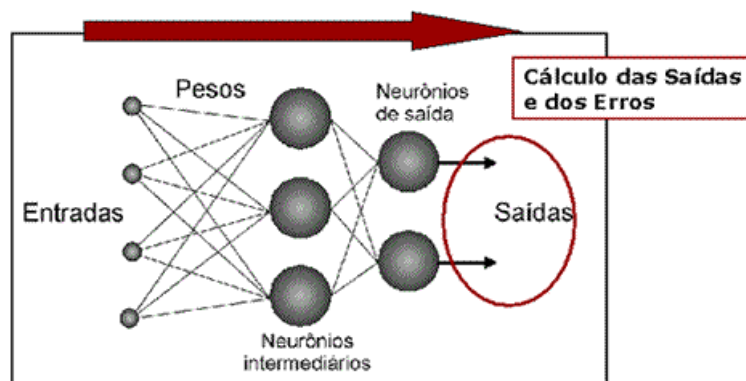
O algoritmo de treinamento mais usado é o algoritmo de retropropagação do erro, conhecido com *backpropagation*, onde a minimização de uma função custo, no caso, a soma dos erros médios quadráticos sobre um conjunto de treinamento, utiliza a técnica de busca do gradiente descendente.

O método do gradiente descendente é utilizado para correção dos pesos sinápticos e seu objetivo é encontrar o mínimo global da função erro a ser minimizada.

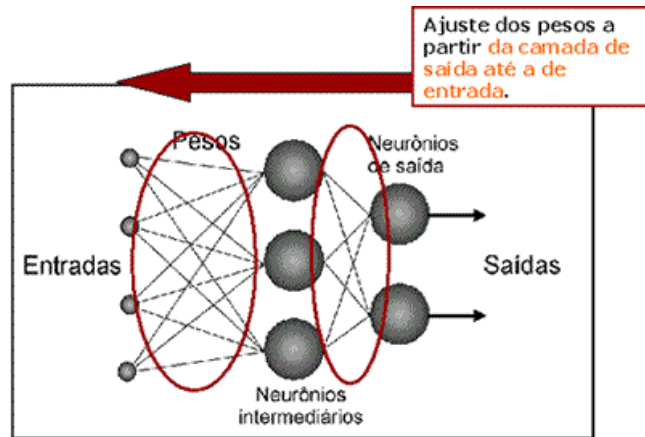
A localização do mínimo global significa a solução ótima do treinamento. Em alguns casos, porém, podem surgir mínimos locais, são nestas situações que a correta definição da taxa de aprendizado é importante para controlar o tempo de aprendizagem da rede.

O treinamento *backpropagation* baseia-se no aprendizado supervisionado por correção de erros. Segue abaixo a descrição sucinta das etapas para aplicação do algoritmo:

- Selecionar os pares do conjunto treinamento e apresentar à camada de entrada. Depois de apresentado o padrão de entrada, a resposta de uma unidade é propagada como entrada para as unidades na camada seguinte, até a camada de saída, onde é obtida a resposta da rede. Esta etapa, denominada propagação, pode ser visualizada na Figura 2-11;
- O erro é calculado, isto é, a diferença entre o valor desejado e a saída da rede;
- O ajuste dos pesos (ou sinapses) ocorre a partir da retropropagação (Figura 2-12). Desde a camada de saída até a camada de entrada, são feitas alterações nos pesos sinápticos. O objetivo ao ajustar as sinapses é que a diferença entre a saída da rede e o valor alvo seja minimizado.



**Figura 2-11 - Direção de Propagação do Sinal de entrada**



**Figura 2-12 - Direção da retropropagação (*backpropagation*) do erro**

O erro é definido conforme a equação abaixo:

$$e_j(n) = d_j(n) - y_j(n) \quad (2-50)$$

onde, o neurônio  $j$  é um nó de saída, e  $d$  é o valor desejado e  $y$  é o resultados da rede.

A soma dos erros ao quadrada é dada por:

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (2-51)$$

Onde  $C$  inclui todos os neurônios da camada da saída da rede. O objetivo do processo de aprendizagem é ajustar os parâmetros livres da rede a fim de minimizar a função de custo  $\varepsilon_{MED}$  (HAYKIN, 1994):

$$\varepsilon_{MED} = \frac{1}{N} \sum_{i=1}^N \varepsilon(n) \quad (2-52)$$

onde,  $N$  é o número total de pares no conjunto treinamento.

A correção  $\Delta w_{ji}(n)$  aplicada à  $w_{ji}(n)$  é definido pela Regra Delta:

$$\Delta w_{ji}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} \quad (2-53)$$

onde:



$\eta$  é o parâmetro da taxa de aprendizagem do algoritmo *backpropagation*. Na prática esta taxa define a velocidade do aprendizado. O sinal negativo na equação acima indica a descida do gradiente no espaço dos pesos.

Mostra-se então que a atualização do peso da conexão da unidade  $i$  para a unidade  $j$  é dada pela equação abaixo:

$$\omega_{ji}(n+1) = \omega_{ji}(n) + \Delta\omega_{ji}(n+1) \quad (2-54)$$

A atualização dos pesos pode ser de duas maneiras básicas: local e por lote.

Na atualização local a atualização é feita imediatamente após a apresentação de cada amostra de treinamento, é também chamado de método de atualização *on-line* ou padrão a padrão. Requer um menor armazenamento para cada conexão, e apresenta menos possibilidade de convergência para um mínimo local.

Na atualização por lote ou batelada a atualização dos pesos só é feita após a apresentação de todas as amostras de treinamento que constituem uma época. Esse procedimento é conhecido como método de atualização *off-line* ou *batch*, e o ajuste relativo a cada apresentação de uma amostra é acumulado, possibilitando uma melhor estimativa do vetor gradiente.

Neste trabalho, serão desenvolvidas redes neurais *feedforward* com uso do algoritmo *backpropagation* com os pesos sendo atualizados por *batching*.

# **Capítulo 3**

## **Áreas de Estudo**

Entre o conjunto de estações disponíveis foram escolhidas aquelas onde houvesse o maior número de observações para o período analisado.

Foram utilizados os dados observados em três estações de superfície, duas localizadas na Costa Sul e uma na Costa Norte do Brasil. A localização destas estações permitirá uma comparação entre o desempenho das previsões baseadas no pós-processamento, por meio de redes neurais e MOS, em duas regiões, Sul e Norte, com características climáticas diferentes.

### **3.1 Descrição das áreas**

As estações escolhidas para o estudo são mantidas pela Marinha do Brasil representada pelo Centro de Hidrografia da Marinha (CHM) em apoio ao Serviço Meteorológico Marinho (SMM).

Todas as estações meteorológicas da Marinha do Brasil são equipadas com sistemas convencionais de coleta de dados e algumas possuem também sistemas automáticos de coleta e transmissão de dados. As estações meteorológicas automatizadas participam diretamente do intercâmbio internacional de dados meteorológicos através do Sistema de Telecomunicação Global (*Global Telecommunication System* - GTS). O GTS é uma grande rede internacional de dados onde todos os usuários deste sistema podem acessar informações meteorológicas de seu interesse.

Os dados das estações meteorológicas convencionais são enviados periodicamente ao CHM, onde são analisados e qualificados, e posteriormente incluídos no Sistema de Informações Meteorológicas (SIM) para serem disponibilizados no GTS. Dentre as estações escolhidas para este estudo, apenas Santa Marta possui, além do sistema convencional de coleta, o sistema automático.

As figuras abaixo foram obtidas no *site* da MARINHA DO BRASIL (MB) e exibem fotos das estações estudadas.



**Figura 3-1 Farol Salinópolis –PA.**

Posição: 00° 37'S - 047° 21'W

Altitude: 26m

Estabelecimento: SET/1978



**Figura 3-2 RF Sta. Marta – SC.**

Posição: 28° 36'S - 048° 49'W

Altitude: 52m

Estabelecimento: JAN/1963



**Figura 3-3 Farol Mostardas -RS.**

Posição: 31° 15'S - 050° 54'W

Altitude: 1m

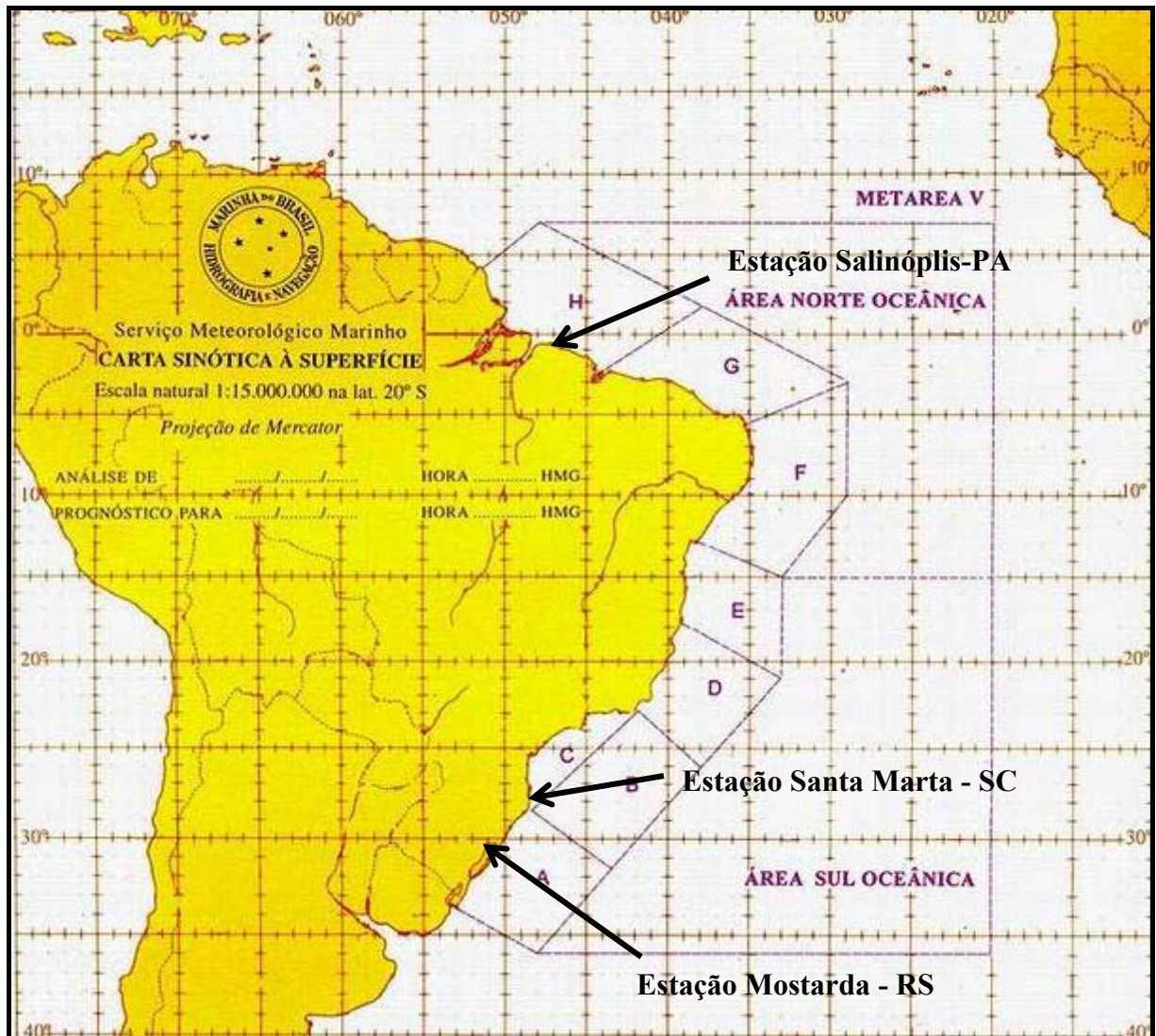
Estabelecimento: JAN/1957

A Figura 3-4 exibe a localização geográfica das três estações estudadas. Observe que as estações de Santa Marta e Mostardas localizam-se em região de clima temperado enquanto Salinópolis está sob clima tropical.

A estação Mostardas, localizada no Estado do Rio Grande do Sul, é caracterizada por clima subtropical úmido, constituído por invernos e verões definidos, separados por estações intermediárias com aproximadamente dois meses de duração, e chuvas bem distribuídas ao longo do ano.

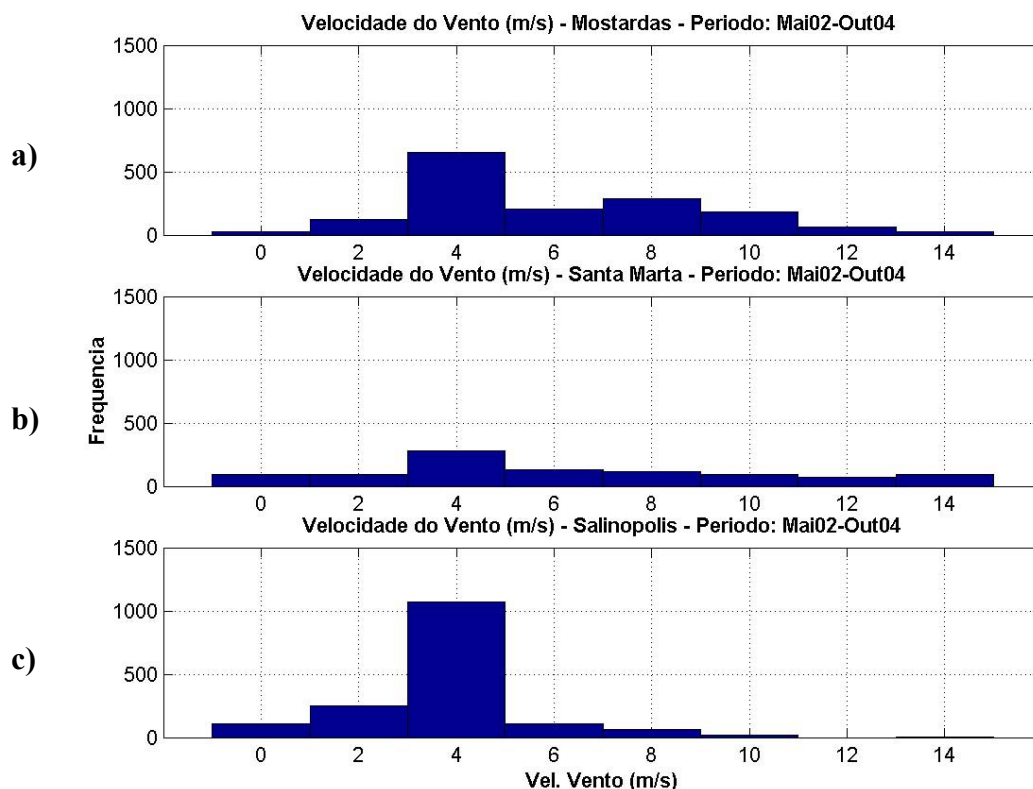
Na Figura 3-5a e na Figura 3-6a são apresentadas as distribuições de vento e visibilidade em Mostardas no período de estudo. Observa-se a baixa frequência dos casos de vento forte (velocidade > 13.9 m/s) e de visibilidade restrita (< 2km)

registrados na estação. O percentual de ocorrência de vento forte é de 1.27% e de visibilidade restrita é de 3.5%. Os eventos de interesse são considerados eventos raros, portanto este comportamento é esperado.



**Figura 3-4 Áreas de Estudo**

A estação Santa Marta, localizada no Estado de Santa Catarina, possui um clima definido como subtropical úmido mesotérmico, onde as quatro estações do ano são bem definidas. Sua localização, em uma área sujeita à passagem de sistemas frontais intensos, favorece um maior número de ocorrência dos eventos de interesse que serão estudados.

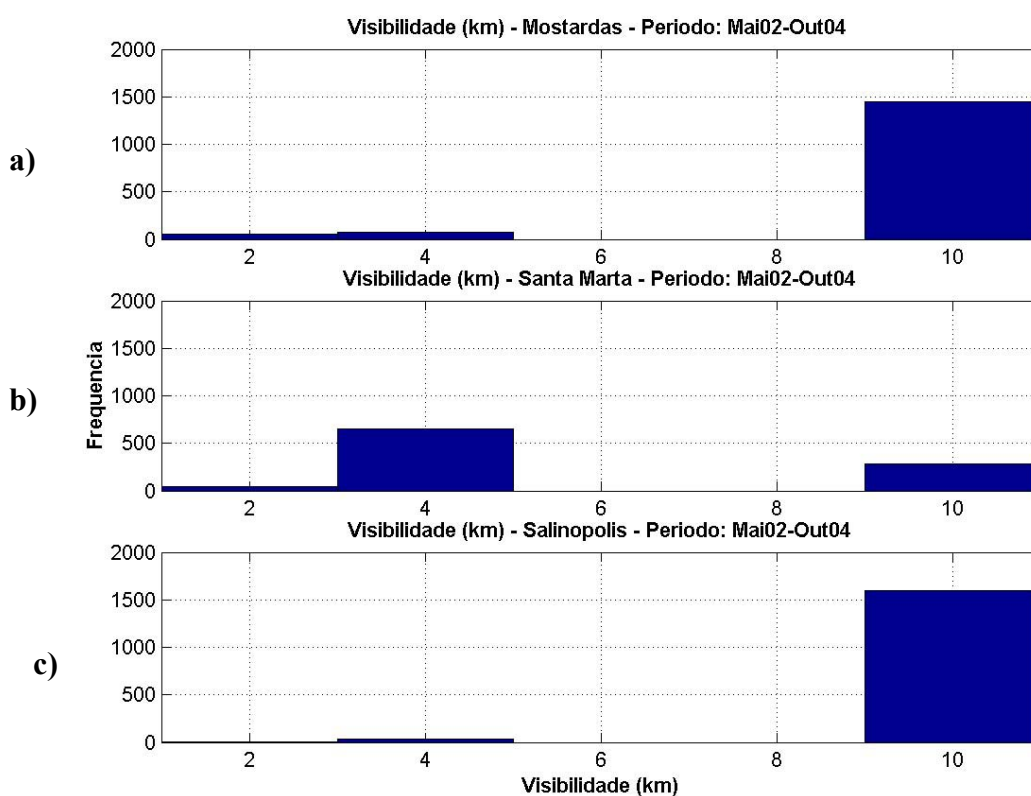


**Figura 3-5** Histograma de freqüência da velocidade do vento (m/s) observada em (a) Mostardas (RS), (b) Santa Marta (SC) e (c) Salinópolis (PA) no período de 01mai02 a 31out04 nos horários de 00Z e 12Z.

Na Figura 3-5b e na Figura 3-6b são apresentadas as distribuições do vento e da visibilidade na região para o período estudado. Observe, no histograma que os casos de vento forte (velocidade  $> 13.9$  m/s) e de visibilidade restrita ( $< 2$ km) registradas na estação apresentam baixa freqüência em relação às outras faixas. O percentual de ocorrência de vento forte é de 6.6% e de visibilidade restrita é de 4.1%. Este comportamento é esperado, tendo em vista que os eventos estudados são considerados raros. Contudo, as ocorrências são percentualmente maiores do que em Mostardas. Veremos durante o estudo que esta característica de Santa Marta permitirá que o pós-processamento dos dados tenha melhor desempenho, pois as redes neurais e as regressões do MOS terão maiores subsídios para a correta previsão dos casos.



Por outro lado, a análise do comportamento dos campos de visibilidade e de vento observados em Salinópolis mostra que as frequências observadas dos eventos são ainda menores do que nas outras duas estações. O percentual de ocorrência de vento forte registrado na região foi de 0.19% e de visibilidade restrita é 0.5%. Salinópolis esta sob um regime tropical onde as amplitudes térmicas e as variações climáticas são pequenas. Este comportamento, que pode ser confirmado por meio de análise da Figura 3-5c e da Figura 3-6c que exibem as baixas variações nos campos em relação ao comportamento das outras estações.



**Figura 3-6** Histograma de frequência da visibilidade (km) observada em : (a) Mostardas (RS), (b) Santa Marta (SC) e (c) Salinópolis (PA), no período de 01mai02 a 31out04, nos horários de 00Z e 12Z.

Veremos durante os estudos que esta característica de Salinópolis implicará em maior dificuldade para se obter um bom desempenho dos métodos de pós-processamento, pois a baixa representatividade dos eventos estudados impedirá o

treinamento da rede neural e o ajustamento das regressões do MOS de forma adequada, muito embora as variabilidades sejam relativamente pequenas.

É importante ressaltar que, embora sejam considerados eventos raros, as ocorrências de ventos fortes e de visibilidade restrita são consideradas críticas para a segurança da navegação, pois ameaçam a segurança da vida humana no mar, e podem causar prejuízos econômicos decorrentes de acidentes de navegação. Além disso, as ocorrências de vento forte não são adequadamente previstas por modelos numéricos e a visibilidade não é diretamente prevista por estes. Por esta razão esses casos costumam ser diagnosticados de forma subjetiva pelo previsor do tempo. Destarte, a utilização de formas objetivas de previsão destes eventos que possam servir de guia para a tomada de decisão dos previsores poderá ser de grande utilidade.



# **Capítulo 4**

## **Dados e Metodologia**

Neste capítulo serão detalhados os procedimentos realizados para a implementação dos métodos de pós-processamento, rede neural e MOS, para previsão de ocorrência de vento forte e visibilidade restrita em três estações localizadas na costa brasileira.

### **4.1 Dados**

No presente estudo foram utilizados os seguintes conjuntos de dados:

- Dados de vento, de visibilidade e de umidade relativa observados nos horários de 00 UTC (*Universal Time Zone*) e 12 UTC em três estações meteorológicas de superfície: Salinópolis (PA), Santa Marta (SC) e Mostardas (RS);

- Análise e previsões de 12 e 24 horas do modelo numérico HRM (*High Resolution Regional Model*), obtidas a partir de processamentos inicializados nos horários de 00Z e 12Z. Foram selecionados os pontos de grade do modelo mais próximos da estação, posicionados sobre o continente e sobre o mar, respectivamente. As variáveis escolhidas, apresentada na Tabela 4-1, foram aquelas que, do ponto de vista físico, estão relacionadas com os eventos a serem previstos (ocorrência de vento forte e visibilidade restrita).

Tendo em vista a necessidade de se obter uma amostra homogênea, o período de estudo ficou restrito ao período em não houve alteração na configuração do modelo HRM, ou seja, 01 de maio de 2002 a 31 de outubro de 2004.

No período estudado o modelo HRM apresentava a seguinte configuração:

- Limites: 50°S, 15°N, 018°W, 072°W
- Grade horizontal: 0,3° x 0,3° (cerca de 30 km x 30 km).
- Níveis verticais: 26 (com melhor resolução na camada limite)
- Validade das previsões: 78 horas
- Rodadas: 00Z e 12Z

O modelo HRM é um modelo numérico hidrostático de escala meso- $\alpha$  e meso- $\beta$  e utiliza condições iniciais e de contorno provenientes do modelo global alemão (GME). Com relação as parametrizações físicas do modelo, podemos destacar:

- esquema de precipitação, incluindo microfísica de nuvens parametrizada;
- esquema de radiação (RITTER, GELEYN, 1992) incluindo *feedback* total de radiação de nuvens, derivação parcial de cobertura de nuvens (umidade relativa e convecção), entre outros;
- esquema de convecção de fluxo de massa (TIEDTKE, 1989) que diferencia entre convecção profunda, rasa ou de níveis médios.

Atualmente, o modelo apresenta outra configuração. Este é um dos motivos que impossibilita a implementação imediata dos métodos de pós-processamento que serão estudados nesta dissertação. Outras informações sobre o modelo HRM podem ser obtidas em MAJEWSKI (2006).

Para obtenção dos dados do modelo numérico foram desenvolvidos programas utilizando o aplicativo *GrADS v1.9b4* (GRADS/COLA), uma vez que os dados do modelo estavam no formato binário GRIB, e em linguagem *Fortran 90*. Além disso, algumas rotinas em *Shell script* foram criadas a fim de permitir a consolidação dos dados oriundos do modelo numérico com os dados observados nas estações.

**Tabela 4-1 Descrição das variáveis de entrada (ou preditoras)**

no.	Variáveis	Descrição
1	hh	Hora (00Z ou 12Z)
2	tp	Intervalo da previsão (00,12,24)
3	an	Hora da inicialização dos dados do modelo (00Z ou 12Z)
4	tsfclf	Temperatura da superfície do solo
5	tsfcsf	Temperatura da superfície do mar
6	t2ml	Temperatura do ar ao nível de 2m, posição em terra
7	t2ms	Temperatura do ar ao nível de 2m, posição no mar
8	dpt2ml	Temperatura do ponto de orvalho no nível de 2m, posição em terra
9	dpt2ms	Temperatura do ponto de orvalho no nível de 2m, posição no mar
10	u10ml	Valor da componente latitudinal do vento, posição em terra
11	u10ms	Valor da componente latitudinal do vento, posição no mar
12	v10ml	Valor da componente meridional do vento, posição em terra
13	v10ms	Valor da componente meridional do vento, posição no mar
14	clct	Cobertura total de nuvens, posição em terra
15	clcts	Cobertura total de nuvens, posição no mar
16	psfcf	Pressão Atmosférica no nível do mar
17	vort1000	Vorticidade na superfície geopotencial de 1000 hPa
18	div850	Divergência na superfície geopotencial de 850 hPa
19	urel850	Umidade relativa na superfície geopotencial de 850 hPa
20	hgt500	Altura da superfície geopotencial em 500 hPa
21	vort500	Vorticidade na superfície geopotencial de 500 hPa
22	div250	Divergência na superfície geopotencial de 250 hPa
23	magv250	Magnitude do Vento na superfície geopotencial de 500 hPa
24	urfct	Umidade relativa
25	urel12hr <sub>obs</sub>	Umidade relativa observada na estação
26	magv12hr <sub>obs</sub> ou visib12hr <sub>obs</sub>	Velocidade do Vento observada na estação ou Visibilidade observada na estação

## **4.2 Metodologia**

### **4.2.1 Controle de qualidade dos dados**

Foi realizada a análise exploratória dos dados (TRIOLA, 1998) a fim de se avaliar a natureza da distribuição das variáveis estudadas bem como suas variações. Foram utilizados histogramas e diagramas *boxplot* para esta análise.

O *boxplot* é útil para revelar tendências centrais, dispersão, a distribuição dos dados e a presença de *outliers* (valores extremos). Os valores extremos foram analisados também subjetivamente a fim de avaliar se poderiam ser considerados valores espúrios ou se eram apenas valores com pouca representatividade na amostra, mas que poderiam ocorrer do ponto de vista climatológico da região estudada. No primeiro caso, o valor considerado espúrio foi substituído pela média e no segundo caso o valor foi mantido. Observe que a substituição pela média não afeta os resultados da regressão e o treinamento da rede.

Nesta etapa foram utilizadas algumas funções do aplicativo *Matlab v.6.0.0.88r12* para a geração dos gráficos e produzidos alguns programas em *Fortran 90* para identificação dos *outliers*.

### **4.2.2 Escolha das variáveis preditoras**

Uma vez que a utilização de variáveis preditoras que apresentem uma alta correlação entre si implica em informações redundantes (WILKS, 2006). Foram utilizados dois métodos a fim de minimizar os problemas inerentes ao uso de preditoras que fossem correlacionadas entre si.

No primeiro método utilizou-se a análise de componentes principais (WILKS, 2006), detalhado no Capítulo 2. As 26 variáveis apresentadas na Tabela 4-1 foram utilizadas para obtenção das componentes que são combinações lineares das variáveis originais e que não são correlacionados. Aplicando-se também a capacidade de compressão da análise de componentes principais, utilizou-se um número de preditores menor que o conjunto total de variáveis. Em todos os experimentos foram utilizadas 16 componentes que respondem, em média, por mais de 95% da variação total dos dados originais. A escolha destas componentes foi feita de forma subjetiva com uso do gráfico *scree plot* ( Figura 2-2) e o uso de 16 componentes mostrou-se satisfatório quando da realização dos experimentos. O método de componentes principais foi implementado por meio de programa em *Fortran 90* com a utilização da subrotina EVCSF da biblioteca matemática IMSL (IMSL, 1987a) para obtenção dos autovalores e autovetores associados com estas componentes.

As componentes foram utilizadas durante o treinamento das redes neurais e para o ajustamento das regressões.

O segundo método baseou-se no *Screening Regression* (WILKS, 2006, MURPHY, KATZ, 1985), detalhado no Capítulo 2. Neste método as variáveis preditoras foram escolhidas pelo método de Seleção para Frente (WILKS, 2006, NETER, WASSERMAN, 1974) onde as variáveis que apresentam maior significância com a variável dependente são incluídas na regressão. O método é feito de forma iterativa e a inclusão de novas variáveis pára quando não existem variáveis consideradas estatisticamente significativas. Neste estudo foi utilizado um nível de significância de 5%. O método foi implementado por meio da rotina do *Fortran 90*, chamada RSTEP da biblioteca estatística IMSL (IMSL, 1987b), que ajusta o modelo de regressão usando a seleção para frente.

Observe que os dois métodos acima permitem que seja escolhido um conjunto de variáveis preditoras que seja considerado “ótimo” do ponto de vista estatístico. Além disso, permitem que seja resolvido o problema da correlação entre as variáveis preditoras.

Durante este estudo será avaliado se os dois métodos apresentam diferenças significativas quando da aplicação do pós-processamento. Ou seja, se as previsões feitas pelo MOS e pela rede neural apresentam diferenças significativas em termos de identificação do evento de interesse quando se usa *screening* ou componentes principais.

### **4.2.3 Características das amostras**

Foram selecionadas 26 variáveis (Tabela 4-1) que do ponto de vista físico estão relacionadas com a ocorrência de visibilidade restrita e de vento forte. Observe a inclusão de campos de umidade relativa e das próprias variáveis a serem previstas disponíveis 12 horas antes da observação. O procedimento de inclusão dos dados observados como variáveis preditoras em conjunção com as previsões do modelo numérico, já foi utilizado em VISLOCKY e FRITSCH (1997) e em HALL *et al.* (1998).

Quando da união dos dados oriundos do modelo numérico com os dados observados para organização das amostras, houve uma diminuição no número total de registros utilizados nos experimentos das três estações estudadas. Isto ocorreu em virtude de faltarem dados do modelo HRM e/ou das observações das estações em vários horários.

Outro fator limitador do tamanho da amostra foi a disponibilidade de poucos dados de observações, que são coletados em apenas quatro horários (00Z, 06Z, 12Z e 18Z), de acordo com o padrão adotado pela Organização Mundial de Meteorologia para

estações meteorológicas costeiras, diferente do que ocorre com dados oriundos de aeródromos, onde as observações são feitas a cada hora. A opção de se utilizar observações costeiras criou mais uma dificuldade para o estudo em virtude da pequena amostra. Contudo, a utilização destes dados era necessária, uma vez que este estudo pretende avaliar a possível utilização dos métodos aqui analisados no Serviço Meteorológico Marinho. Assim será possível, ao final deste estudo, ter-se uma idéia das dificuldades para a implementação do método escolhido, de forma que medidas poderão ser adotadas quando da implementação dos mesmos.

Em todos os experimentos cada amostra foi dividida aleatoriamente em dois conjuntos: 80% para o conjunto de treinamento e 20% para o conjunto de teste. Este procedimento é útil para controle do *overfitting* além de permitir a validação dos resultados. O *overfitting* da rede neural ocorre quando esta ajusta os dados de forma tão perfeita que inclui os ruídos das informações, com isso perde a capacidade de generalização produzindo previsões de baixo desempenho durante a validação. No caso das regressões, a utilização do conjunto teste servirá para validação de suas previsões já que neste método o controle do *overfitting* se dá com a escolha de um número adequado de preditores. Uma das regras é utilizar uma amostra que seja entre 6 ou 10 vezes o número de preditores utilizados para obtenção dos coeficientes (KUTNER *et al.*, 2005). Em virtude do pequeno tamanho das amostras disponíveis, não foi possível utilizar-se um terceiro conjunto de dados independente, para a validação dos resultados, sendo tal validação realizada com o próprio conjunto de teste.

O conjunto de treinamento foi utilizado na fase de aprendizado da rede e para a obtenção dos coeficientes das regressões do MOS. Utilizou-se o segundo conjunto para validação dos resultados, por meio da obtenção das medidas de desempenho que serão detalhadas abaixo.

Outro importante procedimento adotado foi a replicação dos dados. Este procedimento consiste em uniformizar o número total de eventos de interesse, de tal forma que, durante o treinamento da rede e no ajuste das regressões para obtenção dos coeficientes, o número total de casos extremos (ocorrência de vento forte ou visibilidade restrita) seja o mesmo que o número total de não evento, por meio da repetição dos eventos extremos dentro da amostra. Este procedimento aumenta artificialmente a representatividade dos casos de interesse, uma vez que neste estudo o objetivo é a previsão de dois eventos considerados raros e reconhecidamente de difícil previsão.

Os experimentos foram classificados de acordo com a organização das amostras em local da estação, horários de inicialização das variáveis do modelo, de acordo com a replicação ou não dos dados e conforme o método de escolha das variáveis preditoras, a saber:

- local: Salinópolis (PA), Santa Marta (SC) ou Mostardas (RS);
- rodadas: de 00Z e 12Z integradas ou 00Z e 12Z separadas;
- eventos de interesse: replicação ou não replicado; e
- método para escolha das variáveis preditoras: *screening* ou componentes principais.

O termo usado “rodadas de 00Z” (ou “de 12Z”) refere-se ao horário de inicialização do modelo numérico e para o qual foram obtidas a análise, previsão de 12 e 24 horas das variáveis preditoras do modelo e listadas na Tabela 4-1. O horário Zulu ou Hora do Meridiano Zero é o mesmo que o horário UTC.

A Tabela 4-2 ilustra a forma de organização das amostras utilizadas na implementação dos métodos de pós-processamento para cada uma das três estações. Observe que para cada estação foram utilizadas 12 amostras com configurações



diferentes. Ou seja, foram implementadas 12 redes neurais e 12 regressões para cada estação.

Os experimentos serão realizados de forma que reflitam a organização das amostras.

<b>Rodada</b>	<b>Replicação</b>	<b>Variáveis preditoras</b>	<b>Amostra</b>
00Z e 12Z	Sim	<i>Screening</i>	1
	Não	<i>Screening</i>	2
	Sim	CP	3
	Não	CP	4
00Z	Sim	<i>Screening</i>	5
	Não	<i>Screening</i>	6
	Sim	CP	7
	Não	CP	8
12Z	Sim	<i>Screening</i>	9
	Não	<i>Screening</i>	10
	Sim	CP	11
	Não	CP	12

**Tabela 4-2** Esquema de organização das amostras utilizadas nos experimentos com redes neurais e MOS para as três estações estudadas.

#### **4.2.4 Implementação da Rede Neural**

Após o pré-processamento dos dados, as redes neurais foram implementadas de acordo com as configurações das amostras exibidas na Tabela 4-2. Desta forma, foram desenvolvidas quatro redes neurais para cada um dos horários de inicialização totalizando 12 redes para cada uma das três estações estudadas. Com relação ao método para escolha das variáveis preditores, *screening* ou componentes principais, temos tratamentos diferentes na implementação das redes neurais.

Para o cálculo das componentes principais ou para aplicação do *screening* utilizou-se as 26 variáveis apresentadas na Tabela 4-1 nos horários de 00Z e 12Z. Nas redes com utilização de apenas um horário de inicialização utilizou-se 25 variáveis, pois a variável referente à rodada do modelo (AN) foi utilizada para separar as amostras. Ou seja, foram utilizadas as mesmas variáveis predictoras tanto na rede utilizada para previsão de vento forte quanto para previsão de visibilidade restrita.

As redes desenvolvidas com as componentes principais foram testadas com diferentes configurações, onde se variou o número de variáveis de entrada, o número de neurônios na camada intermediária e o número de épocas a fim de se obter o desempenho desejado em termos de detecção dos eventos de interesse.

Nas redes desenvolvidas com as *screening* as diferentes arquiteturas foram alteradas de acordo com o número de neurônios na camada intermediária e o número de épocas de treinamento. Observe que nestes experimentos não foram alteradas as variáveis de entrada uma vez que se considera que o conjunto “ótimo” de variáveis predictoras foi selecionado pelo *screening*.

Foram desenvolvidas redes neurais distintas para classificação de ocorrências de vento forte e de visibilidade restrita em cada uma das estações. As redes foram configuradas com uma camada oculta, com neurônios com função de ativação tangente hiperbólica nas camadas intermediárias e de saída.

O conjunto treinamento foi usado para adaptar os pesos sinápticos das redes *feedforward* de múltiplas camadas. Utilizou-se o algoritmo *backpropagation* com o método do gradiente descendente para obtenção do mínimo da função erro (HAYKIN, 1994). O passo de busca  $\alpha$  (alfa) adotado<sup>3</sup>, utilizado no método do gradiente descendente, foi, a princípio, de 0.1. Em alguns experimentos, entretanto, foram

---

<sup>3</sup> Refere-se ao  $\eta$  apresentado na equação (2-53).

utilizados valores menores para se obter a convergência nas soluções. As sinapses foram inicializadas aleatoriamente, com valores variando entre -0.2 e +0.2. Em cada treinamento foi adotado o critério de se salvar o conjunto de sinapses que produzia o menor erro, antes de ter atingido a condição de *overtraining*. A situação de *overtraining* pode ser controlada pela análise dos gráficos de erros onde é exibida a evolução do erro do conjunto teste.

Foram testadas redes para diferentes épocas de treinamento com as sinapses sendo corrigidas por batelada (*batch*).

As variáveis de entrada das redes foram normalizadas de forma que a maioria dos valores ficassem entre -1 e +1. Para isso foi subtraída a média de cada variável e o resultado foi dividido por duas vezes o desvio padrão de cada variável. Este procedimento foi aplicado somente quando se utilizou o método *screening*. As componentes principais já se apresentam normalizadas. A variável dependente (a ser prevista) foi codificada de forma que recebeu valor 1 quando da ocorrência do evento e -1 quando da não ocorrência.

Outra característica aplicada neste trabalho foi sugerida por HSIEH e TANG (1998), que propõem o emprego da técnica do *ensemble average*. Neste método são realizados conjuntos de treinamento, cada um iniciado de uma determinada posição de busca aleatória no método do gradiente descendente. Posteriormente, é tomada a média das diversas soluções de cada membro do conjunto, que produzirá uma melhor estimativa da solução verdadeira. Os autores atestam que esta técnica é útil para resolver o problema de mau condicionamento que ocorrem nas redes neurais ao tentar resolver vários parâmetros a partir de uma amostra pequena. Tendo em vista, que em estudos meteorológicos e oceanográficos as amostras costumam ser pequenas, a utilização deste

método é indicada. Desta forma, optou-se por adotar a técnica de média dos conjuntos em todos os experimentos deste trabalho.

Para a implementação das redes foi desenvolvido um programa em *Fortran 90*.

#### **4.2.5 Implementação da MOS**

Para o desenvolvimento das regressões, as variáveis preditores e a organização das amostras foram processadas da mesma forma que na implementação das redes neurais. Ou seja, foram desenvolvidas quatro regressões para cada um dos horários de inicialização e para cada uma das três estações, e regressões diferentes para previsão de ocorrência de vento forte ou visibilidade restrita.

Contudo, a variável dependente foi transformada em variável binária, isto é, assume o valor 1 quando da ocorrência do evento (velocidade do vento  $> 13.9$  m/s ou visibilidade  $< 2$ km). Com este procedimento, os resultados das regressões são probabilidades de ocorrência do evento a ser previsto. Assim, as regressões do MOS fornecerão resultados entre 0 e 1 (WILKS, 2006, NETER, WASSERMAN, 1974). A fim de que fosse possível comparar os resultados probabilísticos do MOS com os resultados categóricos da rede neural, utilizou-se um limiar (*threshold*) para transformar os resultados do MOS em valores não probabilísticos. O limiar é o valor da probabilidade acima do qual considera-se que a previsão considera a ocorrência do evento, valores menores que este limiar são classificados como não ocorrência do evento.

O fato de a variável dependente ser binária implica na violação das suposições necessárias para inferências sobre os coeficientes e sobre a reta de regressão. Contudo, a experiência tem mostrado que, geralmente, podemos obter resultados estáveis quando

utilizamos amostras grandes (MURPHY, KATZ, 1985). Dessa forma, inferências relacionadas com a regressão e seus coeficientes podem ser feitas (NETER, WASSERMAN, 1974).

As inferências dos coeficientes das regressões serão feitas, conforme descrito no capítulo 2, por meio de estatísticas que medem a qualidade do ajuste das regressões, coeficiente de determinação múltiplo (equação 2-24) a estatística  $F^*$  (equação 2-30). Contudo, a existência de uma relação entre a variável dependente e suas preditoras não pode assegurar a utilidade ou relevâncias das previsões baseadas neste modelo (NETER, WASSERMAN, 1974). Em função disso, os coeficientes encontrados foram apresentados a uma amostra independente.

A amostra independente consiste de um conjunto de dados escolhidos aleatoriamente a partir do conjunto total de dados, que corresponde ao conjunto de teste utilizado no treinamento das redes neurais. As componentes principais para esta amostra foram obtidas a partir da aplicação dos *scores* calculados com a amostra de treinamento.

A utilização das componentes principais e do *screening* se deu da mesma forma que na rede neural.

Assim, a padronização na utilização das amostras e dos métodos de seleção das variáveis preditoras permitiu normalizar os dados de entrada das metodologias, visando atender a um dos objetivos do trabalho que é comparar os resultados das previsões do MOS com os resultados da rede neural.

#### **4.2.6 Avaliação do desempenho do pós-processamento**

A validação será baseada em algumas medidas escalares as quais são calculadas através de uma tabela de contingência 2 x 2.

		<b>Observação</b>		
		<b>sim</b>	<b>não</b>	
<b>Previsão</b>	<b>sim</b>	a	b	a+b
	<b>não</b>	c	d	c+d
		a+c	b+d	n=a+b+c+d

A tabela de contingência acima tem dimensionalidade três, ou seja, as informações sobre o desempenho das previsões contidas nesta tabela não podem ser expressas completamente com menos de três parâmetros. Tendo em vista, o grande número de nomenclaturas apresentadas na literatura para estas medidas, neste trabalho usaremos as definições apresentadas em WILKS (2006).

- a) **Proporção correta (PC):** representa a proporção de previsões corretas incluindo os eventos e não eventos. Quando se analisam eventos raros esta medida nem sempre é útil, uma vez que penaliza igualmente os dois tipos de erros (alarmes falso e falhas na previsão do evento). O melhor resultado para esta medida é 1 e o pior resultado é 0.

$$PC = \frac{a + d}{n} \quad (4-1)$$

- b) **Threat Score (TS):** representa o número de previsões corretas divididas pelo número total de vezes em que o evento foi previsto e/ou observado. Pode ser interpretada como a proporção correta do evento que está sendo previsto desconsiderando as previsões corretas do não evento. Esta medida é importante, principalmente quando a ocorrência do evento a ser previsto (no nosso estudo,

ocorrência de vento forte ou de baixa visibilidade) é muito menor que a não ocorrência do evento (vento leve ou moderado e visibilidade boa ou moderada). O melhor resultado para esta medida é 1 e o pior resultado é 0.

$$TS = \frac{a}{a+b+c} \quad (4-2)$$

c) **Bias (B)**: é uma comparação entre a média das previsões com a média das observações. O *bias* é a taxa do número de previsões do evento com o número de vezes que o evento foi observado. Previsões sem tendência apresentam  $B=1$ , indicando que o evento foi previsto o mesmo número de vezes em que foi observado. *Bias* maior que um indica que evento foi previsto mais vezes do que foi observado, ou seja, foi superestimado. O *bias* menor que um indica que o evento foi subestimado, isto é, o evento foi previsto menos vezes do que observado.

$$B = \frac{a+b}{a+c} \quad (4-3)$$

d) **Taxa de Alarme Falso (FAR, *false alarm ratio*)**: é o número de vezes que foi o evento foi previsto, mas não ocorreu. Pode ser interpretado como a proporção previsões do evento. O melhor resultado para esta medida é 0 e o pior é 1.

$$FAR = \frac{b}{a+b} \quad (4-4)$$

e) **Taxa de acerto (H, *hit rate*)**: é a proporção das vezes em que o evento de interesse ocorreu e também foi previsto. Ou seja, mostra o desempenho na detecção do evento. Esta medida também é conhecida com probabilidade de detecção (POD, *probability of detection*). A melhor previsão ocorre quando POD é igual a 1 e a pior quando for 0.

$$H = POD = \frac{a}{a + c} \quad (4-5)$$

f) **Taxa de alarme (F, *alarm rate*):** é a taxa de alarmes falsos sobre o total de vezes que o evento não ocorreu. Esta medida pode ser interpretada também como a frequência condicional relativa da previsão incorreta dado que o evento não ocorreu. Esta medida também é conhecida como probabilidade de falsa detecção (POFD, *probability of false detection*).

$$POFD = F = \frac{b}{b + d} \quad (4-6)$$

Estas medidas de desempenho serão utilizadas em todos os experimentos e servirão de base para avaliação do pós-processamento.



# **Capítulo 5**

## **Resultados e Discussão**

Neste capítulo são apresentados os resultados obtidos no pós-processamento do modelo numérico HRM utilizando redes neurais (RN) e *Model Output Statistics* (MOS). Estes resultados serão comparados entre si a fim de avaliar o desempenho dos métodos de escolha das variáveis independentes, isto é, qual a melhor técnica, se a utilização de componentes principais (CP) ou uso do *screening* (SCR) e também quais das técnicas de pós-processamento, MOS ou RN, apresenta o melhor resultado quando comparadas com os resultados do modelo numérico (HRM).

Serão utilizadas medidas escalares para avaliar o desempenho de cada experimento de acordo com suas configurações.

### **5.1 Redes Neurais e MOS como classificadores – Ocorrência de Vento Forte e de Visibilidade Restrita**

Neste tópico a rede neural e o MOS foram utilizados para identificar situações de vento forte ( $>13.9$  m/s) ou de visibilidade restrita (visibilidade  $<2$ km), consideradas críticas para o navegante. Importante ressaltar que o modelo numérico não produz previsões de visibilidade e que os ventos previstos pelo modelo possuem baixa resolução espacial, não representando adequadamente situações de vento forte. Essas situações são diagnosticadas de forma subjetiva pelos previsores do tempo.

Os resultados destes experimentos serão apresentados por meio de tabelas onde medidas escalares são tabuladas e exibem os resultados dos experimentos referentes às

amostras de acordo com os horários de inicialização do modelo HRM e, conseqüentemente, de acordo com a forma em que as amostras foram selecionadas.

As duas primeiras colunas exibem os resultados dos experimentos para redes neurais separadas de acordo com a forma de escolha das variáveis independentes, quais sejam, utilizando componentes principais (coluna 1) ou utilizando *screening* (coluna 2). As colunas 3 e 4 referem-se aos resultados do MOS e da mesma forma, separadas de acordo com o método usado na escolha das variáveis independentes. A quinta coluna, não apresentada no experimento para visibilidade, exibe os resultados das estatísticas para as rodadas do modelo HRM. Ou seja, os experimentos são apresentados de forma a evidenciar os resultados relacionados com a forma em que as amostras foram selecionadas.

As linhas referentes à topologia exibem a configuração das redes neurais utilizadas nos experimentos, da seguinte forma: **número de neurônios na camada de entrada - número de neurônios na camada oculta - número de neurônio na camada de saída - número de épocas de treinamento - número de iterações externas** (aplicada na técnica dos conjuntos). Nesta mesma linha, quando os resultados se referem aos experimentos do MOS, é apresentado o número de coeficientes utilizados nas equações de regressão do MOS.

A linha referente ao valor de  $\alpha$  (alfa) refere-se ao passo de busca adotado no método do gradiente descendente para obtenção do erro mínimo durante o treinamento da rede.

A linha referente ao valor do limiar exibe o valor utilizado para transformar as previsões do MOS em valores não probabilísticos.

Para o desenvolvimento das retas de regressão utilizando a técnica de MOS o preditando foi codificado com valores 0 (não ocorrência do evento) ou 1 (ocorrência do

evento). Por exemplo, a variável dependente assumirá o valor 1 se a velocidade do vento (visibilidade) observada for maior que 13.9 m/s (menor que 2 km).

A regressão estimada com uma variável dependente binária sempre representa a probabilidade desta assumir o valor 1 (WILKS, 2006, NETER, WASSERMAN, 1974). A técnica de regressão pelo método dos mínimos quadrados produzirá uma equação que, quando apresentada a um novo dado, produzirá previsões entre 0 e 1 a qual poderá ser interpretada como a probabilidade de ocorrência do evento (neste caso, a ocorrência de vento forte ou de visibilidade restrita). A fim de transformar as saídas do MOS em valores não probabilístico, utilizou-se um limiar a partir do qual considera-se que a previsão do MOS previu a ocorrência do evento. As previsões menores que o limiar escolhido implicam em previsão de não ocorrência do evento.

Para a avaliação da consistência das regressões e, conseqüentemente, de suas previsões, serão feitas inferências sobre estas. Desta forma, neste estudo, as inferências serão feitas em termos do coeficiente de determinação ( $R^2$ ), equação (2-24), e da estatística F, equação (2-30). Um ajuste perfeito da equação de regressão resulta em  $R^2=1$  e para um ajuste bom teremos  $R^2$  próximo de 1. A estatística F avalia a hipótese de relação entre a variável dependente e o conjunto de variáveis independentes, de forma que um p-valor, a ela associado, maior que um dado nível de significância (adotaremos 5% ) implica que está relação não existe, ou seja, não há um relação linear entre as variáveis preditoras e o preditando.

Ao final, os resultados serão comparados do ponto de vista geográfico a fim de se avaliar se o pós-processamento do modelo HRM apresenta diferenças de desempenho para estações de clima temperado (Santa Marta e Mostardas) e para a estação de clima tropical (Salinópolis).

### **5.1.1 Ocorrência de Vento Forte**

Para o estudo da ocorrência de vento forte as amostras foram organizadas conforme esquema exibido na Tabela 4-2. Contudo, a fim de tornar a apresentação dos resultados mais concisa, e ao mesmo tempo informativa, serão apresentados apenas os resultados referentes aos horários de 00Z e 12Z, pois os experimentos mostraram que a utilização de dados referentes a dois horários de inicialização permite o adequado treinamento da rede neural e um melhor ajustamento das regressões do MOS, de tal forma que seus desempenhos superam os resultados quando se utiliza apenas um horário de inicialização. A única exceção será para a Estação Santa, cujos métodos de pós-processamento apresentam bom desempenho mesmo quando se utiliza apenas um horário de inicialização.

Os resultados dos experimentos para os horários de 00Z e 12Z, separadamente, podem ser consultados no Apêndice A.

#### **5.1.1.1 Salinópolis**

Nas próximas seções serão apresentados os resultados dos experimentos para a estação Salinópolis utilizando RN e MOS. Estes resultados serão comparados entre si de forma a avaliar qual dos dois métodos apresenta o melhor desempenho na previsão da ocorrência de ventos fortes. Os resultados serão divididos de acordo com os horários das rodadas 00Z e 12Z e de acordo com a replicação ou não dos eventos de interesse.

Será avaliado o impacto da replicação dos casos de interesse nos resultados, de forma que em todos os métodos de pós-processamento será utilizado este recurso.

### **5.1.1.1.1 Rodadas de 00Z e 12Z**

#### **5.1.1.1.1.a Dados Replicados**

A fim de se avaliar a significância dos resultados, serão exibidos os gráficos de erros de treinamento e teste, no caso da RN; os coeficientes obtidos na análise de regressão, e algumas estatísticas importantes em inferências, no caso do MOS.

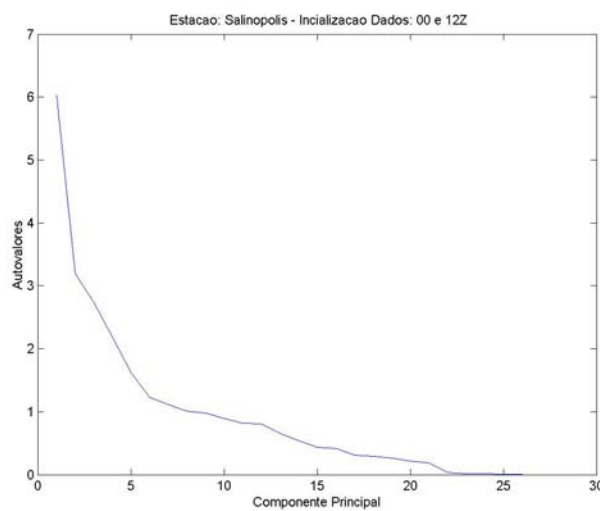
Para o treinamento da rede e para o ajuste das regressões do MOS as variáveis de entrada foram obtidas de duas formas: na primeira empregou-se o método de componentes principais; na segunda, o método *screening*.

Para este experimento, que utiliza as previsões do modelo numérico nos dois horários de inicialização, 00Z e 12Z, as variáveis utilizadas para obtenção das componentes principais e para uso pelo *screening* foram as 26 apresentadas na Tabela 4-1 do capítulo 4.

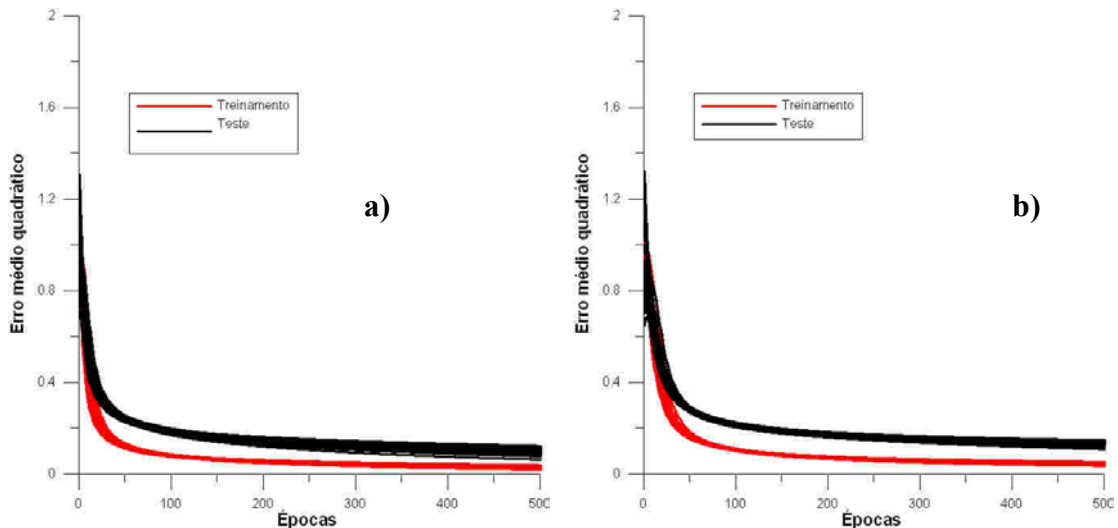
Como variáveis preditoras, foram selecionadas 16 componentes principais as quais acumulam, aproximadamente, 95% da variação total das variáveis originais utilizadas.

O critério de escolha destas componentes baseou-se na análise subjetiva do gráfico denominado *scree plot* conforme apresentado em WILKS (2006). Normalmente, a diferença entre os primeiros autovalores é grande e diminui em direção aos últimos. Sugere-se o corte quando a estabilização dos autovalores passa a ser pequena. O *scree plot* é apresentado na Figura 5-1 e a escolha pelo critério dado seria entre 15 e 20 componentes principais. Dentre os vários experimentos realizados os melhores resultados foram obtidos com 16 componentes principais. O aumento do número de componentes não mostrou grandes diferenças de desempenho tanto do MOS quanto da RN.

A Figura 5-2 exibe a evolução dos erros de treinamento e de teste do conjunto de 31 treinamentos com 500 épocas cada. Observe que o erro converge rapidamente para o mínimo. A Figura 5-2a exibe os treinamentos para a rede neural onde os preditores são Figura 5-2b os preditores foram selecionados pelo método *screening*. Não se observam grandes diferenças de comportamento entre os dois gráficos.



**Figura 5-1** *Scree plot* exibindo os autovalores e as componentes principais calculados a partir das 26 variáveis originais utilizadas na análise de RN e MOS para previsão de ocorrência de vento forte.



**Figura 5-2** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* – Rodadas 00Z e 12Z, com dados replicados, referentes à estação Salinópolis.

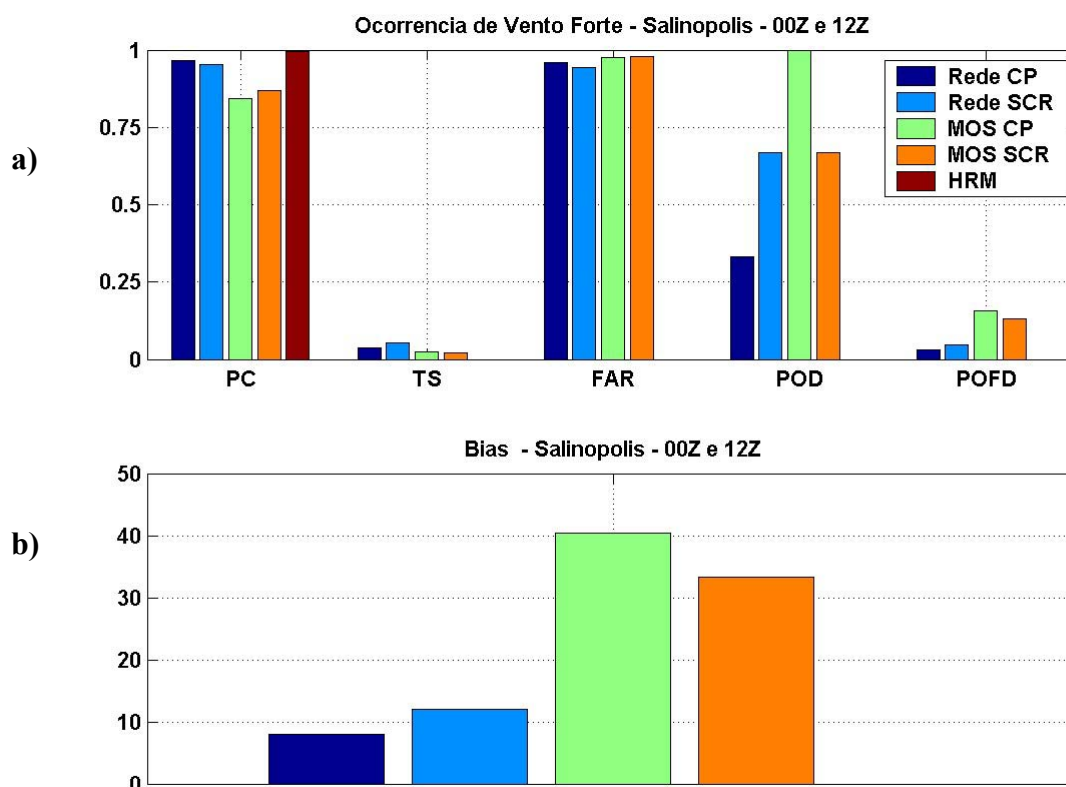
A Tabela B- 1 apresenta os coeficientes da regressão obtidos com a técnica do MOS, e a Tabela B- 2 exibe os coeficientes das regressões onde a variáveis preditores foram escolhidas pela técnica de *screening*. Nestas tabelas podemos observar:  $R^2=72.23\%$  e  $R^2=74.36\%$ , isto é, próximo do valor ideal (100%); p-valor, associado à estatística F, com valor igual a 0 (menor que o nível de significância), ou seja, rejeita-se a hipótese nula de que não há relação entre as variáveis preditoras e o preditando. Além disso, observe na quarta coluna de ambas as tabelas a significância de cada um dos coeficientes, representada pelo p-valor. A partir desta análise, podemos concluir que as equações são significantes a um nível de 5%. As inferências para as regressões e seus coeficientes podem ser consultadas no capítulo 4.

Finalmente, na Tabela 5-1 na **Figura 5-3** são apresentados os resultados do MOS e RN para previsão de ocorrência de vento forte (maior que 13.9 m/s) em Salinópolis.

**Tabela 5-1 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Salinópolis.**

Estatísticas	Rodadas 00+12Z				
	Média = 3.80 m/s Desvio Padrão= 1.77 m/s				
	RN PC	RN SCRN	MOS PC	MOS SCRN	HRM
PC	0.9667	0.9534	0.8429	0.8682	0.9960
TS	0.0385	0.0541	0.0248	0.0198	0
POD	0.3333	0.6667	1	0.6667	0
FAR	0.9583	0.9444	0.9752	0.9800	-
POFD	0.0301	0.0455	0.1578	0.1310	0
B	8	12	40.3333	33.3333	0
Topologia (RN)	16-8-1 500-31	24-12-1 500-31	17 coef	25 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.5	0.5	

Nota: Dados replicados. Eventos no conjunto teste = 3; Não Eventos = 748.  
Eventos no conjunto treinamento = 6; Não Eventos = 2999.



**Figura 5-3 Estatísticas do pós-processamento (MOS e RN) e do modelo HRM para previsão de vento forte – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Salinópolis.**



Na análise da Tabela 5-1 e da **Figura 5-3**, levando-se em conta o interesse principal, que é a previsão de vento forte, observamos que o MOS, utilizando componentes principais, consegue identificar todos os eventos. Contudo, apresenta um *bias* (medida de tendência) de 40.33. Ou seja, para obter este alto valor de detecção (POD=1), o método superestimou o evento em torno de 40 vezes. O MOS com preditores escolhidos pelo método *screening* também apresenta também um alto *bias*, igual a 33.33. A rede neural também superestima, (**Figura 5-3b**) a previsão de vento forte, mas em magnitudes bem menores quando comparadas com os resultados do MOS.

Em termos de acurácia para a previsão de eventos extremos, medida pela TS, não se observam diferenças significativas entre os métodos.

Nas medidas de alarmes falsos, observam-se altos valores para o MOS e para a rede neural. A probabilidade de falsa detecção (POFD) é alta para ambas configurações do MOS, e o melhor desempenho nesta medida, isto é, valor próximo de zero, é obtido pelas previsões da rede neural. Observe que os altos valores de POFD vão implicar em queda no total de previsões corretas (PC) realizadas pelo MOS.

A Figura 5-4a e a Figura 5-4b exibem um comparação entre os eventos observados em Salinópolis e os previstos pela rede neural. Conforme já observado pelas medidas escalares, podemos verificar que a maior parte dos acertos se dá na identificação dos eventos de maior frequência que são as ocorrências de ventos fracos.

A Figura 5-5 e a Figura 5-6 exibem os resultados do MOS em forma de probabilidade. Observe na Figura 5-5b que para velocidades de vento maiores que 13.9 m/s todas as probabilidades atingem o valor 1. Ou seja, todas as ocorrências de vento forte foram identificadas corretamente, contudo alguns eventos com velocidades

menores também receberam, para o limiar de 0.5, probabilidade 1, o que caracteriza os alarmes falsos apresentados na Tabela 5-1.

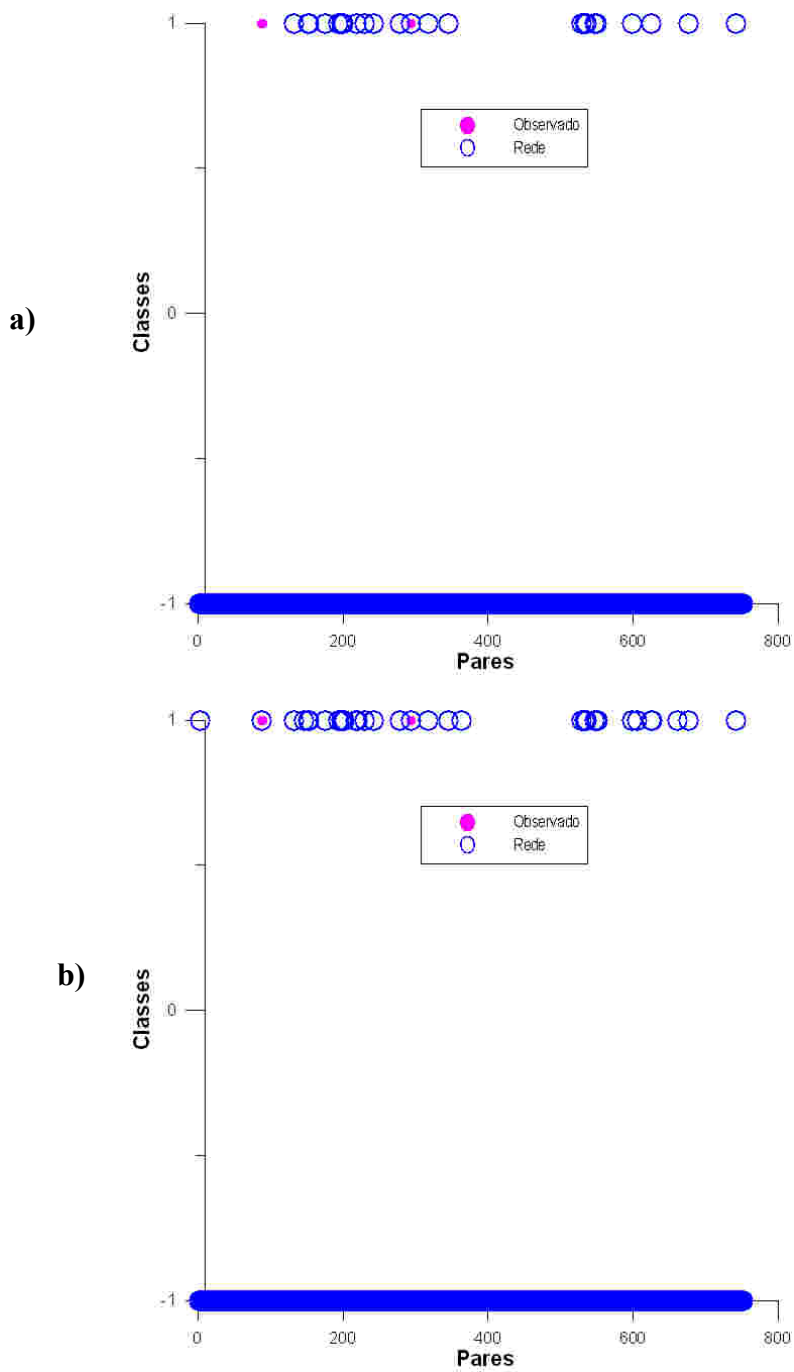
Na Figura 5-6b pode-se verificar que um dos casos de vento forte obteve probabilidade de ocorrência igual a 0, isto caracteriza o erro na previsão mostrada na tabela.

Ao se analisar os resultados em relação ao uso de componentes principais ou screening, observa-se melhor acurácia e confiabilidade no MOS com uso de screening e, no caso da rede neural, não se observam diferenças significativas.

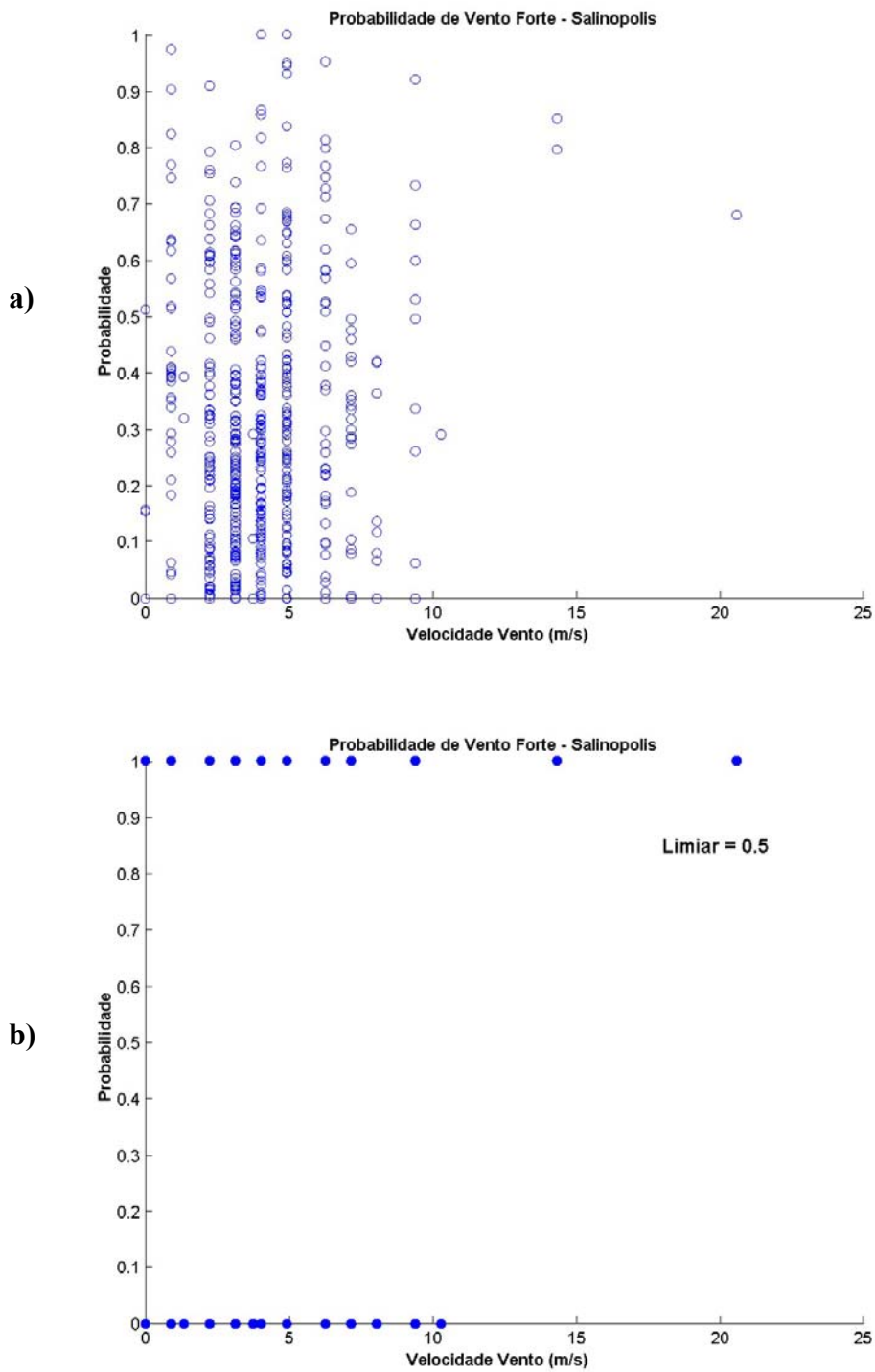
Ao se analisar o desempenho do modelo numérico observamos uma proporção de acerto de 99.6%. Ocorre que o modelo numérico não consegue identificar nenhum das ocorrências de vento forte. Ou seja, o alto grau de acerto não reflete a identificação dos eventos de interesse. Desta forma, podemos verificar que o pós-processamento pode ser uma ferramenta importante quando precisamos identificar eventos raros como é o caso da ocorrência de vento forte.

Em resumo, as configurações das redes neurais, embora não consigam identificar todos os eventos, mantêm a acurácia das previsões e, contrário ao que ocorre com as previsões do MOS, não produzem um número de alarmes falsos tão elevado que ponha em dúvida a confiabilidade de suas previsões. Desta forma, a escolha da melhor configuração entre as quatro apresentadas deverá, além de capacidade discriminatória de evento, unir acurácia e confiabilidade nas previsões.

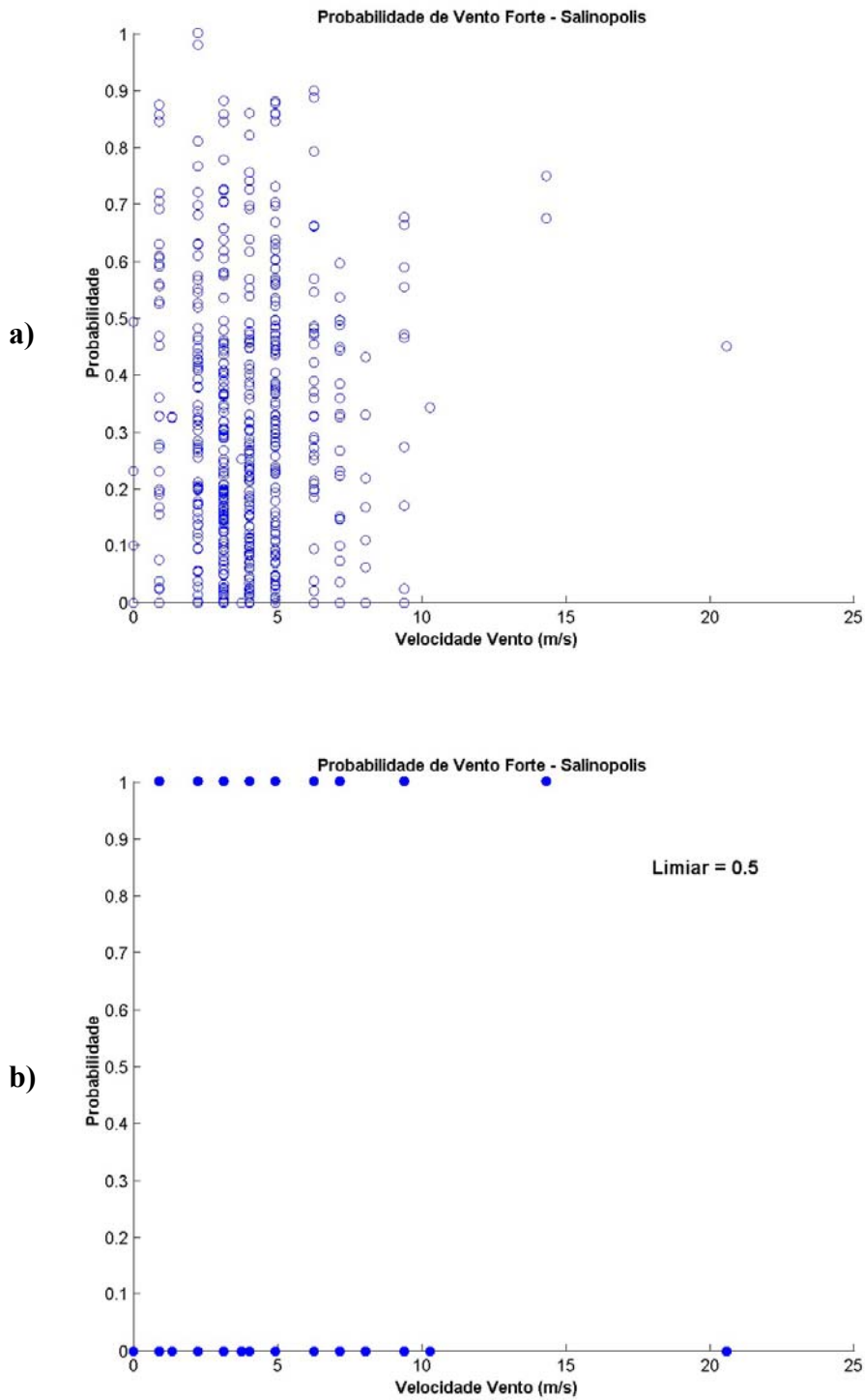
O que se observa, na realização do pós-processamento, é que a tentativa de identificar ou prever os eventos raros implicará, normalmente, em uma diminuição na proporção total de acertos e aumento dos alarmes falsos. Contudo, este será um dilema quando se tenta prever eventos raros e a escolha do melhor método dependerá dos objetivos da previsão.



**Figura 5-4** Comparação da classificação de ocorrência de vento forte a partir dos dados observados em Salinópolis (círculos rosas) e calculadas pela rede neural (círculos azuis) utilizando (a) componentes principais e (b) *screening*. Rodadas de 00Z e 12Z, com dados replicados. A classe 1 corresponde ao evento de velocidade do vento > 13.9 m/s.



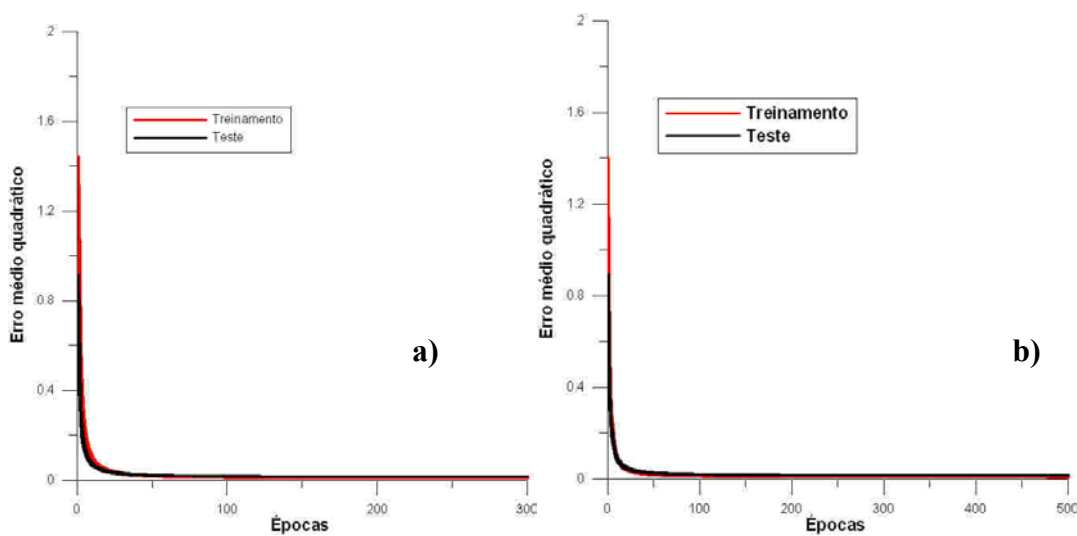
**Figura 5-5** Probabilidades de ocorrência de vento forte obtidas com a equação de regressão utilizando componentes principais (a) e as mesmas probabilidade após definição do limiar de 0.5 (b).



**Figura 5-6** Probabilidades de ocorrência de vento forte obtidas com a equação de regressão utilizando *screening* (a) e probabilidade após definição do limiar de 0.5 (b).

### 5.1.1.1.1.b Dados Não Replicados

Nesta seção os resultados serão apresentados sem que os eventos de interesse tenham sido replicados. Poderemos confirmar a importância da replicação dos casos de menor frequência quando a intenção é a previsão de eventos raros.



**Figura 5-7** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* – Rodadas 00Z e 12Z, com dados não replicados, referentes à estação Salinópolis.

A Figura 5-7 exibe a evolução dos erros de treinamento e teste. Observe a rápida convergência para o mínimo. Veremos a seguir que isso faz com que os métodos de pós-processamento alcancem um alto percentual de acerto, mas sem identificarem nenhuma ocorrência de vento forte.

Na Tabela B- 3 e Tabela B- 4 na são apresentados os resultados das regressões. Observa-se que os resultados não são tão significativos quanto os apresentados quando os dados de interesse foram replicados.

Os coeficientes de determinação que, medem o quanto da variação total do valor a ser previsto (probabilidade de ocorrência de vento forte), é “explicada” pela equação de regressão. Neste caso, apresentam valores baixos, 0.53% e 1.15% .

**Tabela 5-2 Sumário estatístico do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodada de 00Z e 12Z, com dados não replicados, referentes à estação Salinópolis.**

Estatísticas	Rodadas 00+12Z – Não replicados				
	Média = 3.80 m/s Desvio Padrão= 1.77 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
<b>PC</b>	0.9960	0.9960	0.9960	0.9960	0.9960
<b>TS</b>	0	0	0	0	0
<b>FAR</b>	-	-	-	-	-
<b>POD</b>	0	0	0	0	0
<b>POFD</b>	0	0	0	0	0
<b>B</b>	0	0	0	0	0
<b>Topologia (RN)</b>	16-8-1 300-31	3-3-1 500-31	17 coef	4 coef	-
<b>Alfa (RN)</b>	<b>0.1</b>	<b>0.1</b>	-	-	-
<b>Limiar (MOS)</b>	-	-	<b>0.5</b>	<b>0.5</b>	

Nota: Dados replicados. Eventos no conjunto teste = 3; Não Eventos = 748. Eventos no conjunto treinamento = 6; Não Eventos = 2999.

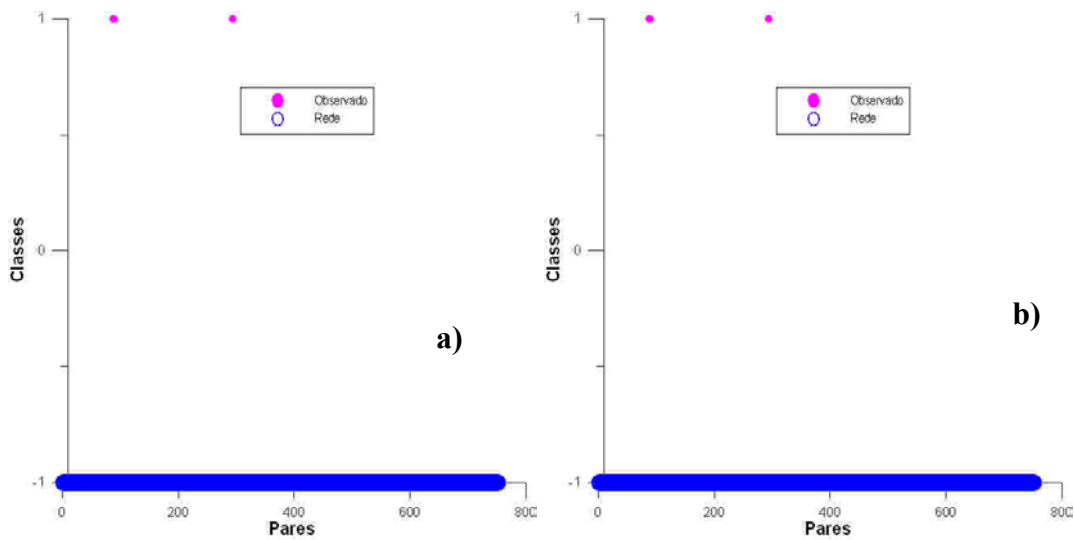
A Tabela 5-2 mostra que a não replicação dos dados no conjunto treinamento implica em resultados iguais para todos os métodos de pós-processamento. O alto percentual de acerto refere-se aos eventos onde a velocidade do vento era menor que 13.9 m/s.

Ao compararmos os resultados dos experimentos com rede neural e MOS verifica-se que não houve melhora na previsão ao se fazer o pós-processamento.

O fraco desempenho dos métodos pode ser confirmado pela análise gráfica. A Figura 5-8 mostra que a rede neural não consegue identificar o eventos de maior interesse. Com relação ao MOS isto pode ser verificado na Figura 5-9a e na Figura

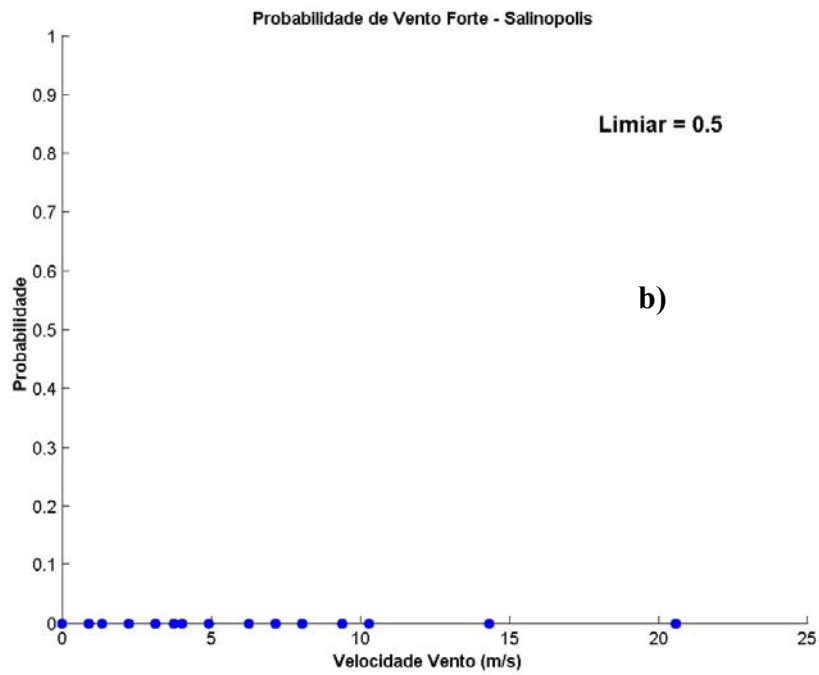
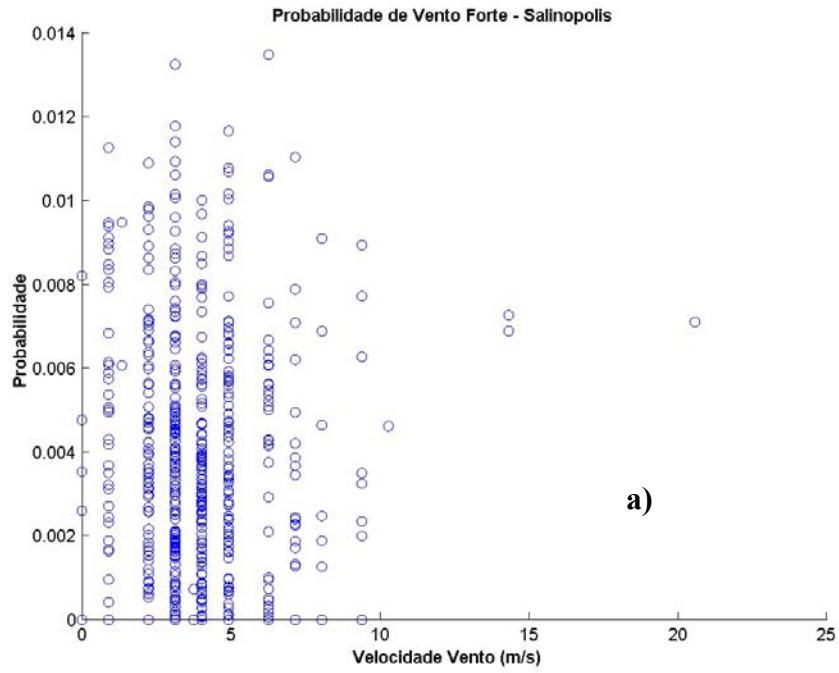
5-10a onde as probabilidades são, devido à baixa significância das regressões, pouco representativas.

Na Figura 5-9b e na Figura 5-10b são exibidas as probabilidades após a definição do limiar. Verifica-se que todos os casos de vento forte receberam probabilidade de ocorrência igual a zero. Conforme mostrado na tabela, esta situação caracteriza a incapacidade do MOS em prever os eventos de interesse.



**Figura 5-8** Comparação da classificação de ocorrência de vento forte a partir dos dados observados em Salinópolis (círculos rosas) e calculadas pela rede neural (círculos azuis) utilizando (a) componentes principais e (b) *screening*. Rodadas de 00Z e 12Z, com dados não replicados. A classe 1 corresponde ao evento de velocidade do vento > 13.9 m/s.





**Figura 5-9** Probabilidades para ocorrência de vento forte obtidas com a equação de regressão utilizando componentes principais (a) e probabilidade após definição do limiar de 0.5 (b).

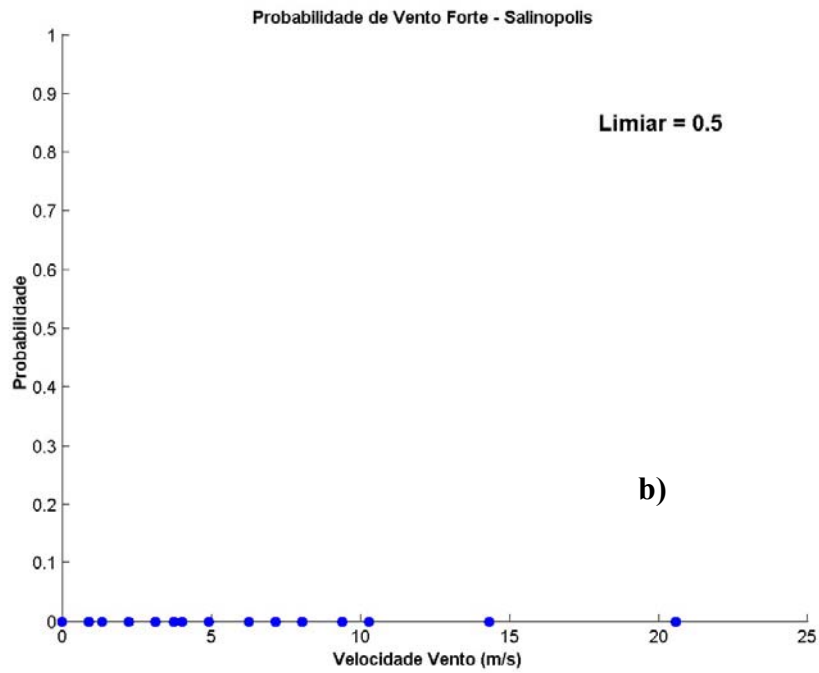
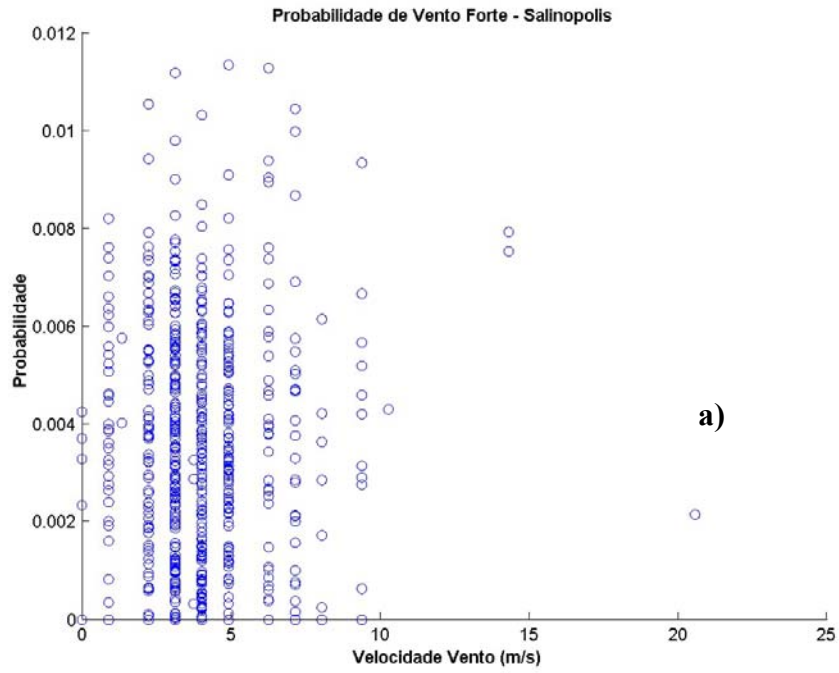


Figura 5-10 Probabilidades para ocorrência de vento forte obtidas com a equação de regressão utilizando *screening* (a) e as mesmas probabilidade após definição do limiar de 0.5 (b).

Em resumo, verificamos que a replicação dos eventos de menor frequência tem grande influência nos resultados do pós-processamento. A replicação realmente faz com que os métodos de rede neural e MOS consigam identificar os eventos de interesse.

Contudo, o MOS na tentativa de identificar os eventos perde em confiabilidade e acurácia nas suas previsões. Isso ocorre, principalmente, ao verificarmos o grande número de alarmes falsos. A rede neural, por outro lado, não consegue identificar todos os eventos, mas mantém controle do número de alarmes falsos de forma a não tornar suas previsões excessivamente suspeitas.

Importante notar que a comparação dos resultados do MOS e da rede neural com os resultados dos modelos mostram a relevância do pós-processamento, uma vez que consegue identificar os casos de vento forte, o que não é possível com o uso exclusivo das previsões do modelo numérico. Mesmo com o número reduzido de eventos de interesse nos conjuntos de treinamento e teste, é possível verificar o potencial dos pós-processamento na previsão desses.

### **5.1.1.2 Santa Marta**

Para esta estação, apesar da amostra relativamente pequena, são observados resultados significativos na previsão da ocorrência de vento forte.

A localização da estação em uma região montanhosa e próxima ao mar, em uma latitude sujeita à passagem de sistemas frontais intensos, favorece a maior ocorrência de ventos fortes, o que permite que os métodos de pós-processamento tenham mais subsídios para previsão.

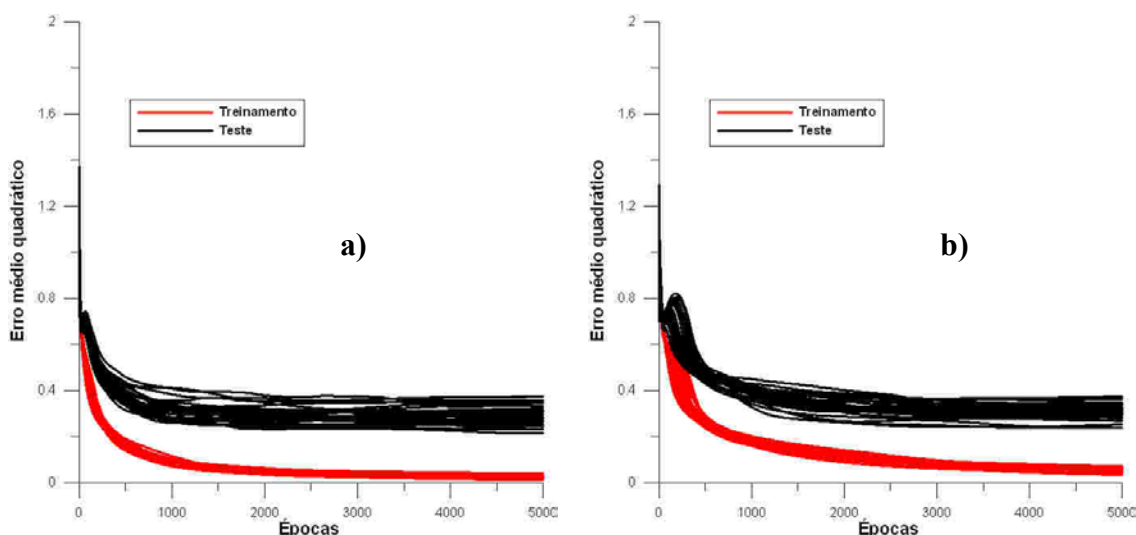
A fim de manter a clareza do texto não serão exibidos, em virtude da grande semelhança entre eles, os gráficos de distribuição das probabilidades. Os coeficientes e as estatísticas das regressões estão listados no apêndice B.

As 16 componentes principais usadas no treinamento da rede e no ajustes das regressões do MOS, respondem por, aproximadamente, 96% da variação total das variáveis. O número de componentes foi escolhido, conforme detalhado na seção anterior.

### 5.1.1.2.1 Rodadas de 00Z e 12Z

#### 5.1.1.2.1.a Dados Replicados

As evoluções dos erros, Figura 5-11a e Figura 5-11b, apresentam comportamentos diferentes de acordo com o tipo de variáveis independentes usadas. A rede que utilizou componentes principais converge para o mínimo mais rapidamente em comparação com a rede em que as variáveis utilizadas foram escolhidas pelo método *screening*. Os gráficos de erro são úteis na avaliação da consistência dos resultados que veremos a seguir.



**Figura 5-11** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* – Rodadas de 00Z e 12Z, com dados replicados, referentes à estação Santa Marta.

As regressões apresentam coeficiente de determinação de 36.32% e 40.22% de acordo com uso de componentes principais e *screening*, respectivamente.

Os coeficientes obtidos pelo método da regressão podem ser encontrados no apêndice B. As equações são estatisticamente significantes ao nível de 5%.

A Tabela 5-3 e a **Figura 5-12** exibem os resultados dos experimentos para o conjunto teste.

O MOS não apresenta diferenças significativas para as previsões utilizando componentes principais ou *screening*.

A rede neural utilizando *screening* apresenta maior detecção dos eventos, mas não apresenta diferenças significativas para as outras medidas quando comparada com os resultados utilizando componentes principais.

O MOS apresenta maior detecção dos eventos em suas duas configurações, mas perde a confiabilidade e em acurácia quando analisamos os resultados em termos de alarmes falsos e tendência.

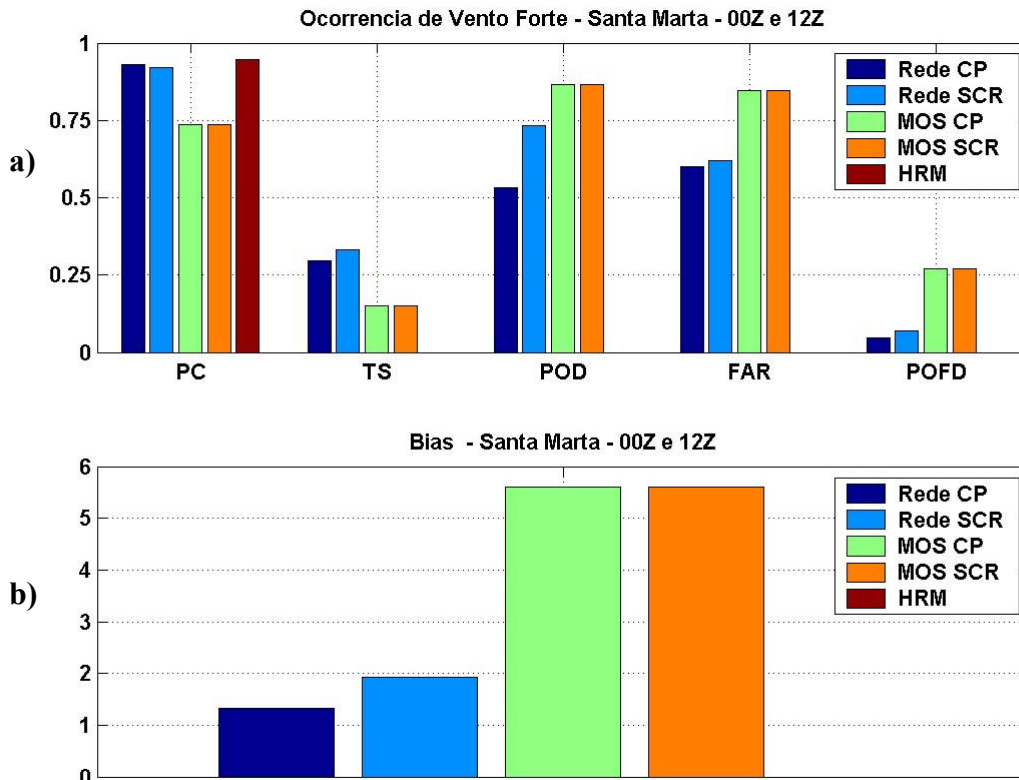
De forma geral, a rede neural tem melhor desempenho do que o MOS em termos de acurácia dos resultados, representada pelas medidas TS e PC, e de emissão de alarmes falsos.

A comparação com o modelo numérico HRM mostra que o pós-processamento é uma ferramenta útil para a previsão de ocorrência de vento forte, uma vez que o modelo não consegue prever nenhum evento de interesse. O alto percentual de acerto não reflete a previsão dos casos de vento maior que 13.9 m/s .

**Tabela 5-3 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte – Rodadas de 00Z e 12Z, com dados replicados, referentes à estação Santa Marta.**

Estatísticas	Rodadas 00+12Z				
	Média = 6.61 m/s Desvio Padrão= 4.38 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9314	0.9206	0.7365	0.7365	0.9458
TS	0.2963	0.3333	0.1512	0.1512	0
POD	0.5333	0.7333	0.8667	0.8667	0
FAR	0.6000	0.6207	0.8452	0.8452	-
POFD	0.0458	0.0687	0.2710	0.2710	0
B	1.3333	1.9333	5.6	5.6	0
Topologia (RN)	16-8-1 5000-31	16-8-1 5000-31	17 coef	17 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.5	0.5	-

Nota: Dados replicados. Eventos no conjunto teste =15; Não Eventos = 262 Eventos no conjunto treinamento = 86; Não Eventos = 1025.

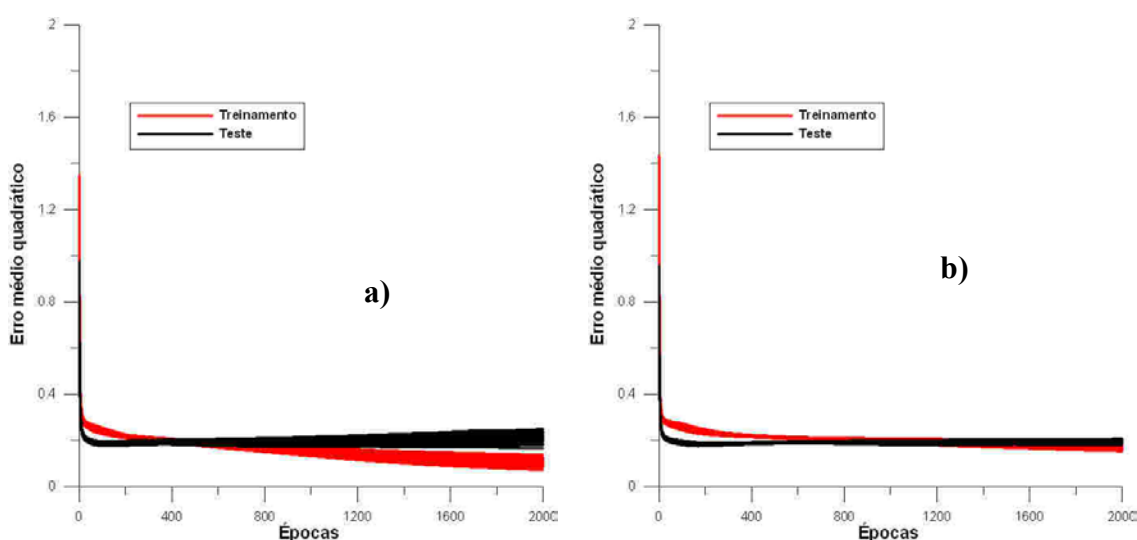


**Figura 5-12 Estatísticas do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 00Z e 12Z, com dados replicados, referentes à estação Santa Marta.**

### 5.1.1.2.1.b Dados Não Replicados

Observe, nos gráficos abaixo, a rápida convergência dos erros em virtude da menor ocorrência de eventos de interesse, pois o conjunto treinamento não foi replicado.

Os coeficientes de determinação foram menores do que os obtidos com a replicação dos dados, 13.11% e 14.69%, para as retas obtidas com componentes principais e *screening*, respectivamente. As regressões podem ser consideradas estatisticamente significantes ao nível de 5%.



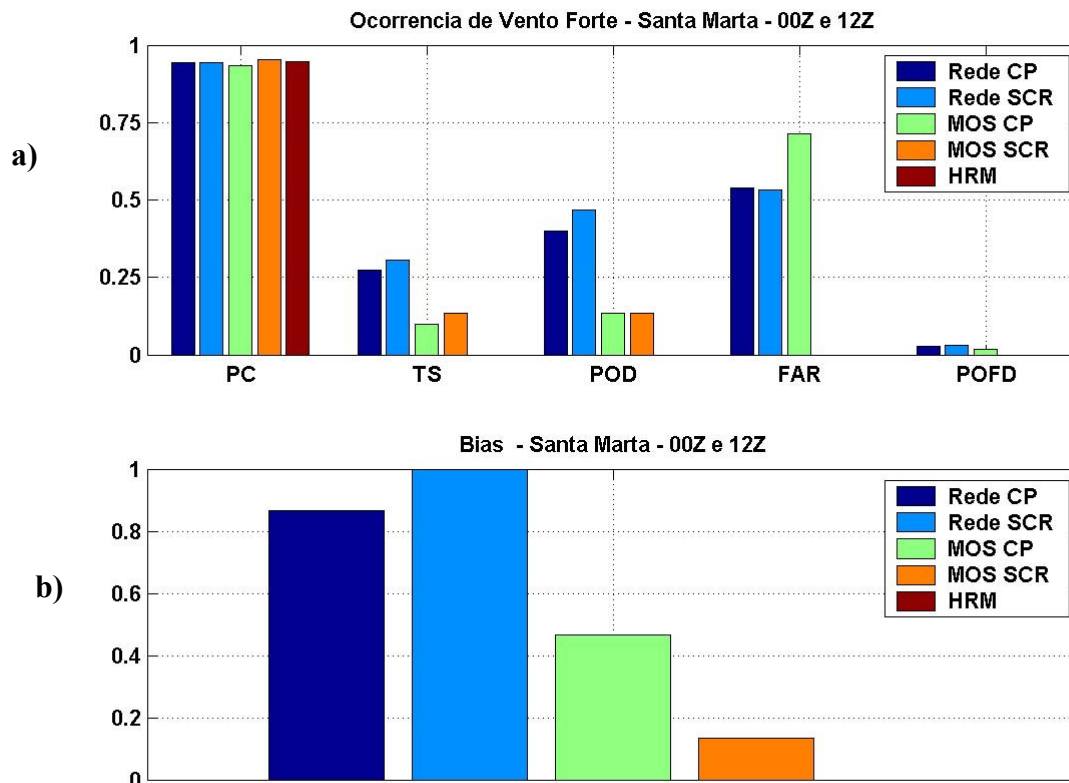
**Figura 5-13** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* - Rodadas de 00Z e 12Z com os dados não replicados, referentes à estação Santa Marta.

Neste experimento, observa-se uma característica importante, que evidencia a relevância de amostras com maior frequência dos eventos de interesse. Os resultados, apesar de não serem melhores do que os obtidos com os dados replicados, apresentaram-se mais expressivos do que os obtidos em Salinópolis sem a replicação, pois o clima tropical acarreta menor ocorrência de ventos fortes na região.

**Tabela 5-4 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte – Rodadas de 00Z e 12Z com os dados não replicados, referentes à estação Santa Marta.**

Estatísticas	Rodadas 00+12Z				
	Média = 6.61 m/s Desvio Padrão= 4.38 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9422	0.9422	0.9350	0.9531	0.9458
TS	0.2727	0.3043	0.1000	0.1333	0
POD	0.4000	0.4667	0.1333	0.1333	0
FAR	0.5385	0.5333	0.7143	0	-
POFD	0.0267	0.0305	0.0191	0	0
B	0.8667	1	0.4667	0.1333	0
Topologia (RN)	16-8-1 2000-31	11-5-1 2000-31	17 coef	12 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.35	0.35	-

Nota: Dados não replicados. Eventos no conjunto teste =15; Não Eventos = 262  
Eventos no conjunto treinamento = 86; Não Eventos = 1025.



**Figura 5-14 Estatísticas do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 00Z e 12Z com os dados não replicados, referentes à estação Santa Marta.**



Observe na Figura 5-14b que todos os métodos, com exceção da rede neural com *screening*, que não apresentou tendência ( $bias=1$ ), subestimaram a previsão do evento.

As previsões das duas configurações da rede neural apresentaram maior valor do TS (Figura 5-14a), que é uma medida importante para avaliar o desempenho quando o interesse é a previsão de eventos raros.

Neste experimento, a rede neural utilizando *screening* apresenta melhor desempenho em comparação com todas as outras configurações do MOS e da própria rede neural utilizando componentes principais.

O pós-processamento mostrou-se uma ferramenta útil, pois consegue detectar os eventos de vento forte. Observe que o modelo numérico acerta apenas os casos onde a velocidade do vento não é extrema, em função do maior número de ocorrência destes eventos.

#### **5.1.1.2.2 Rodada de 00Z**

##### **5.1.1.2.2.a Dados Replicados**

Abaixo, na evolução dos erros (Figura 5-15a e Figura 5-15b) podemos observar uma variação maior no conjunto teste para as diferentes realizações do que nos experimentos anteriores.

As regressões apresentaram coeficientes de determinação de 38.71% e 41.97%, para as regressões obtidas com componentes principais e pelo método *screening*, respectivamente.

As regressões serão consideradas estatisticamente significantes a um nível de 5%.

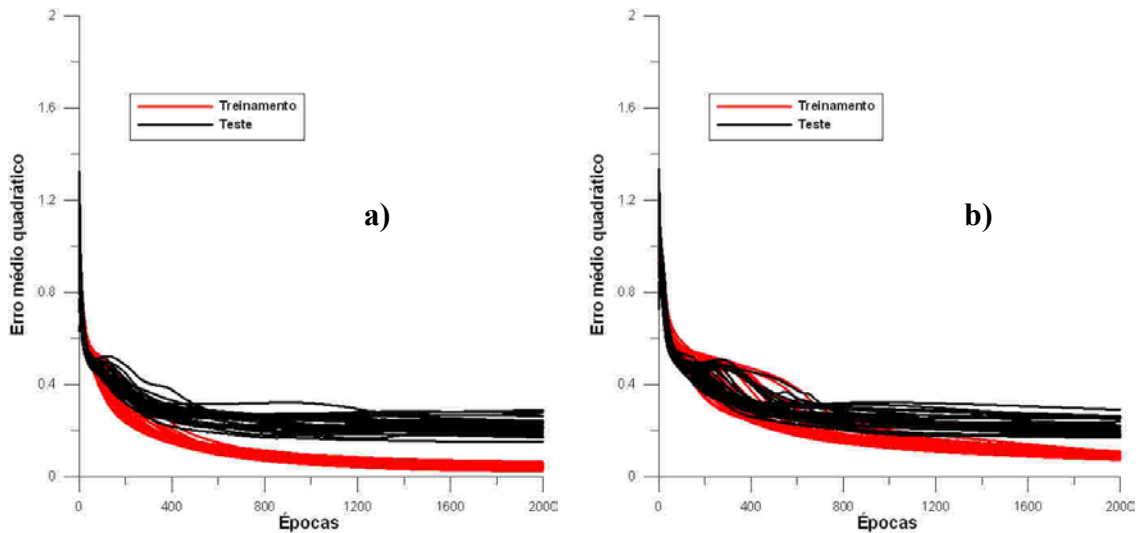
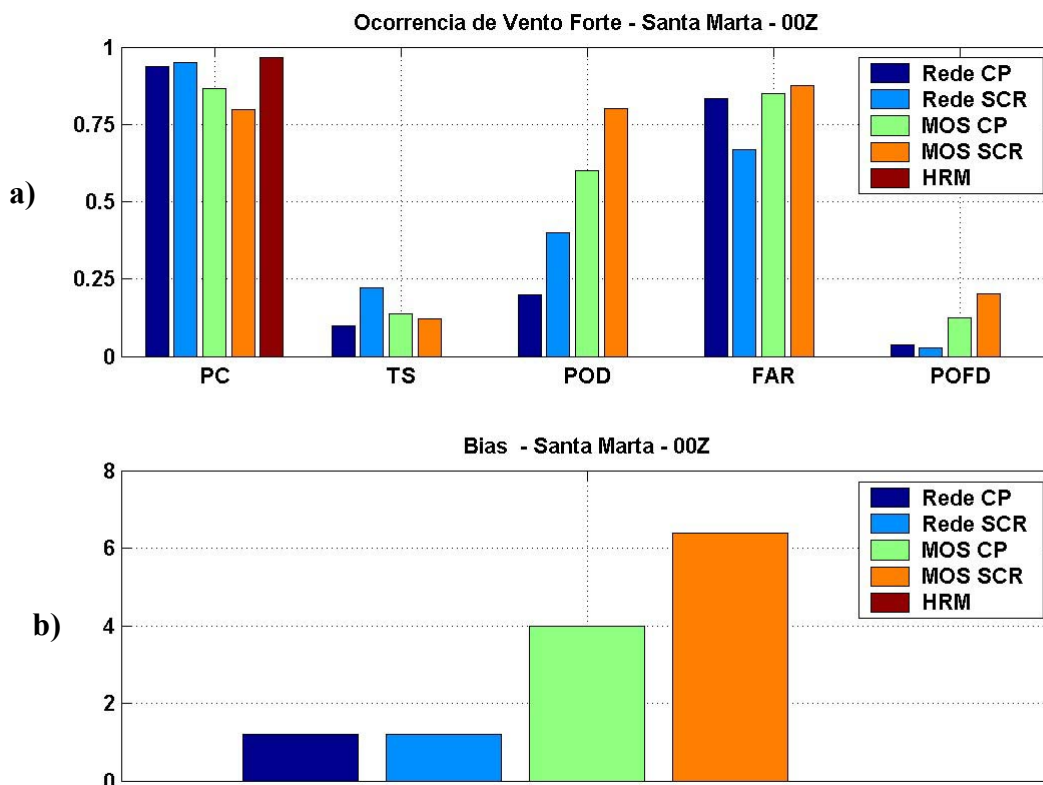


Figura 5-15 Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* – Rodada 00Z, com dados replicados, referentes à estação Santa Marta.

Tabela 5-5 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodada 00Z, com dados replicados, referentes à estação Santa Marta.

Estatísticas	Rodadas 00Z				
	Média = 6.8 m/s Desvio Padrão= 4.5 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9371	0.9510	0.8671	0.7972	0.9650
TS	0.1000	0.2222	0.1364	0.1212	0
POD	0.2000	0.4000	0.6000	0.8000	0
FAR	0.8333	0.6667	0.8500	0.8750	-
POFD	0.0362	0.0290	0.1232	0.2029	0
B	1.2	1.2	4	6.4	0
Topologia (RN)	16-8-1 2000-31	14-7-1 2000-31	17 coef	15 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.5	0.5	-

Nota: Dados replicados. Eventos no conjunto teste =5; Não Eventos = 138  
Eventos no conjunto treinamento = 51; Não Eventos = 522.



**Figura 5-16 Estatísticas do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte – Rodada 00Z, com dados replicados, referentes à estação Santa Marta.**

Os resultados do MOS apresentam maior detecção dos casos de vento forte, contudo seus resultados perdem em acurácia. A emissão de alarmes falsos é alta, tanto para os emitidos em relação ao total de eventos previstos (FAR) quanto para a medida de detecção falsa (POFD), esta última relacionada com o total de não eventos. Como consequência, o MOS diminui a percentual de acertos.

O MOS, utilizando componentes principais, embora com menor desempenho na detecção dos eventos, apresenta melhor acurácia e menor tendência.

A rede neural com utilização de *screening* apresenta melhores resultados em termos de PC, TS e FAR quando comparada com a rede que utiliza componentes principais.

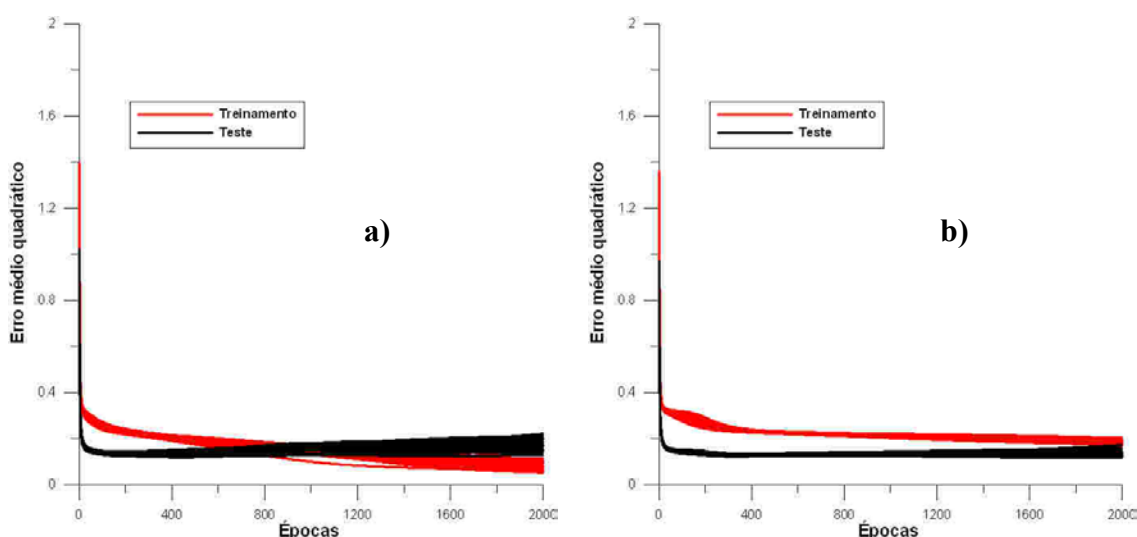
De forma geral, a rede neural apresenta melhores resultados em termos de acurácia e tendência das previsões do que o MOS.

Ao compararmos os métodos de pós-processamento com os resultados do modelo numérico HRM observamos, novamente, que o modelo não consegue identificar nenhum dos casos de vento forte.

### 5.1.1.2.2.b Dados Não Replicados

Nas evoluções dos erros (Figura 5-17a Figura 5-17b) podemos observar a rápida convergência para o mínimo.

A regressão apresentou coeficiente de determinação de 17.2% e 16.39%, para as regressões obtidas com componentes principais e pelo método *screening*, respectivamente. Observe que os valores são baixos em relação ao experimento anterior onde a replicação dos dados foi aplicada.



**Figura 5-17** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* – Rodada de 00Z, com dados não replicados, referentes à estação Santa Marta.

As regressões são consideradas estatisticamente significantes a um nível de 5%.

Na Tabela 5-6 e na Figura 5-18 são apresentados os resultados dos experimentos para o conjunto teste. Apesar dos valores baixos obtidos pelos coeficientes de determinação e da pequena amostra, observamos que as medidas de desempenho apresentam bons resultados para o MOS e para a rede neural quando comparação com os experimentos sem a replicação dos dados.

Podemos citar duas características da amostra utilizando informações da estação Santa Marta: a primeira é o número maior de casos extremos de vento no grupo treinamento e, a segunda é que, climatologicamente, a maior ocorrência destes eventos em Santa Marta permite a melhor classificação dos eventos pela rede e pelo MOS.

Observamos melhores resultados para as previsões da rede neural com utilização de *screening* em termos das medidas de acurácia, representadas pelo TS, e pelo POD. Além disso, a Figura 5-18b mostra que, contrária às outras previsões onde se nota a subestimação dos eventos, não há tendências em suas previsões, isto é, obteve *bias*=1.

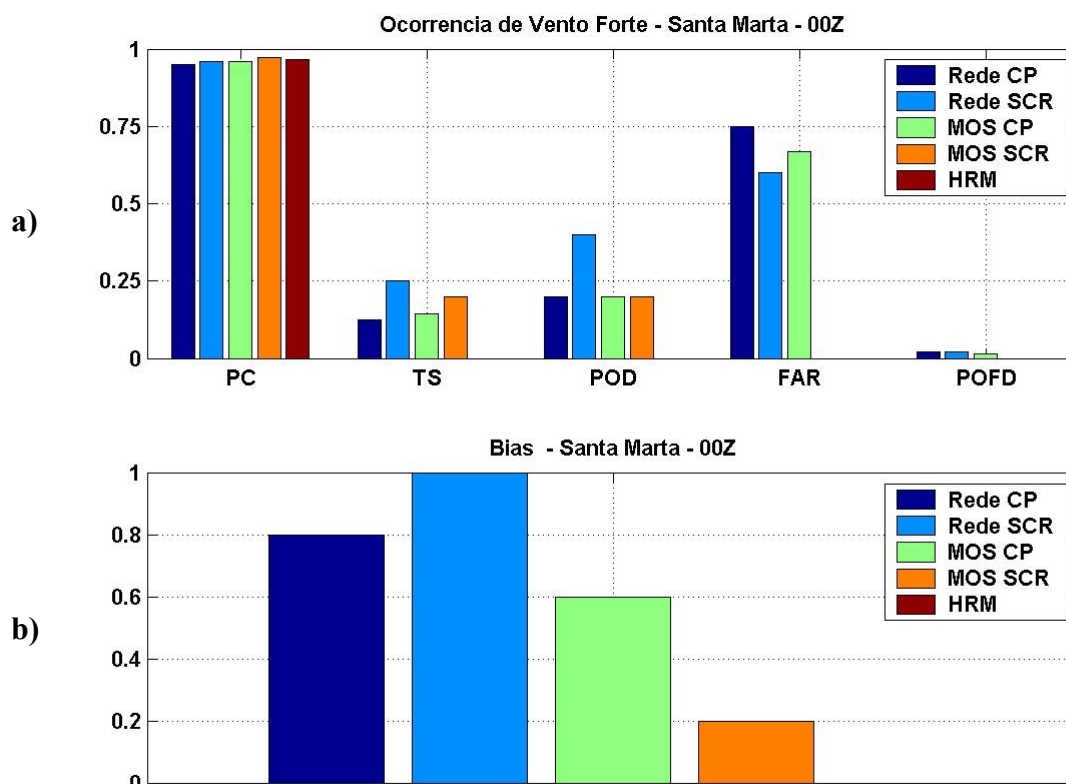
Observamos ainda que os métodos, contrário ao que ocorre nas previsões do modelo numérico, conseguem identificar alguns casos de vento forte.

De forma geral, a previsão da rede neural com uso de *screening* apresenta melhor desempenho. Observe que consegue manter alta proporção de acerto, valor próximo ao obtido pelo HRM, e consegue identificar 2 dos 5 casos de vento forte, ao passo que o modelo numérico não prevê nenhum dos casos.

**Tabela 5-6 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodada de 00Z, com dados não replicados, referentes à estação Santa Marta.**

Estatísticas	Rodadas 00Z				
	Média = 6.8 m/s Desvio Padrão= 4.5 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9510	0.9580	0.9580	0.9720	0.9650
TS	0.1250	0.2500	0.1429	0.2000	0
POD	0.2000	0.4000	0.2000	0.2000	0
FAR	0.7500	0.6000	0.6667	0	-
POFD	0.0217	0.0217	0.0145	0	0
B	0.8000	1	0.6000	0.2000	0
Topologia (RN)	16-8-1 2000-31	8-4-1 2000-31	17 coef	9 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.35	0.4	-

Nota: Dados não replicados. Eventos no conjunto teste =5; Não Eventos = 138  
Eventos no conjunto treinamento = 51; Não Eventos = 522.



**Figura 5-18 Estatísticas do pós-processamento (MOS e RN) e do modelo numérico para previsão de ocorrência de vento forte – Rodadas de 00Z, com dados não replicados, referentes à estação Santa Marta.**

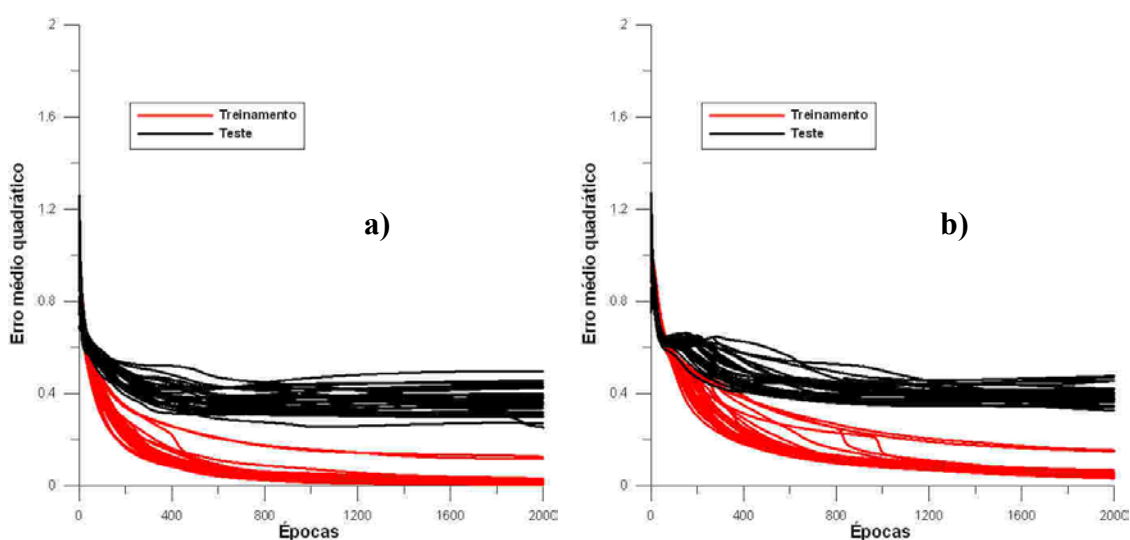
### 5.1.1.2.3 Rodada de 12Z

#### 5.1.1.2.3.a Dados Replicados

Na evolução dos erros (Figura 5-19a e Figura 5-19b) podemos observar variação maior no conjunto do teste para as diferentes realizações.

As regressões apresentaram coeficientes de determinação de 38.06% e 46.82%, para as regressões obtidas com componentes principais e pelo método *screening*, respectivamente.

As regressões podem ser consideradas estatisticamente significantes a um nível de 5%.



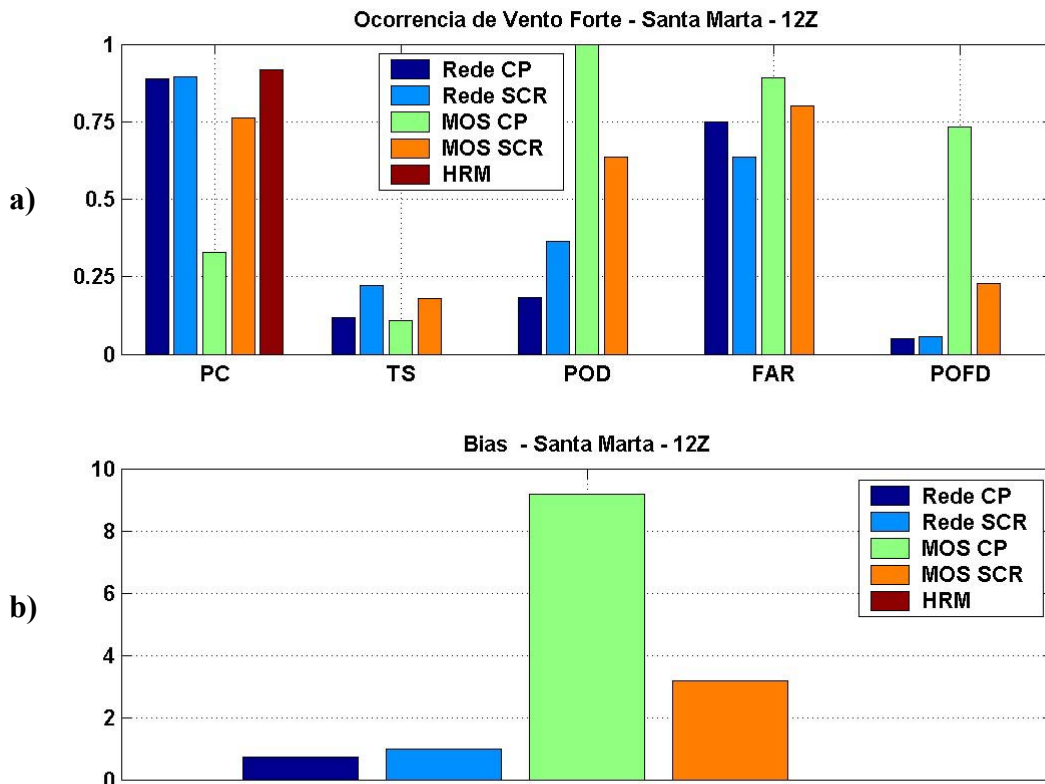
**Figura 5-19** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* – Rodada 12Z, com dados replicados, referentes estação Santa Marta.

Na Tabela 5-7 e na Figura 5-20 temos os resultados dos experimentos. Observe que o MOS, com utilização de componentes principais, embora tenha previsto todos os eventos,  $POD=1$ , perde em acurácia, tendo menor PC dentre os experimentos, e emite alarmes falsos que colocam em dúvida e confiabilidade de suas previsões.

**Tabela 5-7 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodada 12Z, com dados replicados, referentes estação Santa Marta.**

Estatísticas	Rodadas 12Z				
	Média = 6.43 m/s Desvio Padrão= 4.21 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.8881	0.8955	0.3284	0.7612	0.9179
TS	0.1176	0.2222	0.1089	0.1795	0
POD	0.1818	0.3636	1	0.6364	0
FAR	0.7500	0.6364	0.8911	0.8000	-
POFD	0.0488	0.0569	0.7317	0.2276	0
B	0.7213	1	9.1818	3.1818	0
Topologia (RN)	16-8-1 2000-31	16-8-1 2000-31	17 coef	17 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.1	0.5	-

Nota: Dados replicados. Eventos no conjunto teste =11; Não Eventos = 123  
Eventos no conjunto treinamento = 34; Não Eventos = 504.



**Figura 5-20 Estatísticas do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodada 12Z, com dados replicados, referentes estação Santa Marta.**



Os resultados das redes neurais apresentam baixo desempenho na detecção dos eventos e subestima a previsão do evento.

Novamente, na comparação com o modelo numérica verifica-se que o mesmo não consegue identificar os casos de vento forte.

Neste experimento, o pós-processamento não se mostrou significativo quando se avalia a detecção do evento e o número de alarmes falsos.

#### **5.1.1.2.3.b      *Dados Não Replicados***

Abaixo, na evolução dos erros (Figura 5-21a e Figura 5-21b) podemos observar rápida convergência para o mínimo.

As regressões apresentaram coeficientes de determinação de 11.79% e 8.07% para as regressões obtidas com componentes principais e pelo método *screening*, respectivamente.

As regressões podem ser consideradas estatisticamente significantes a um nível de 5%, mas apresentam baixa significância quando analisamos os coeficientes das regressões individualmente.

Os resultados das regressões não se mostraram estatisticamente significantes quando se analisam os coeficientes obtidos com utilização de componentes principais. O MOS com utilização de *screening* apresenta baixas probabilidades para os vários limiares testados de forma que seu poder de discriminação entre evento e não evento é baixa. Os resultados apresentados na Tabela 5-8 mostram que a alta detecção do evento é acompanhada da baixa acurácia e de confiabilidade de suas previsões.

As previsões das duas configurações da rede neural podem ser consideradas poucos significantes ao se analisar os resultados de detecção do evento e o TS.

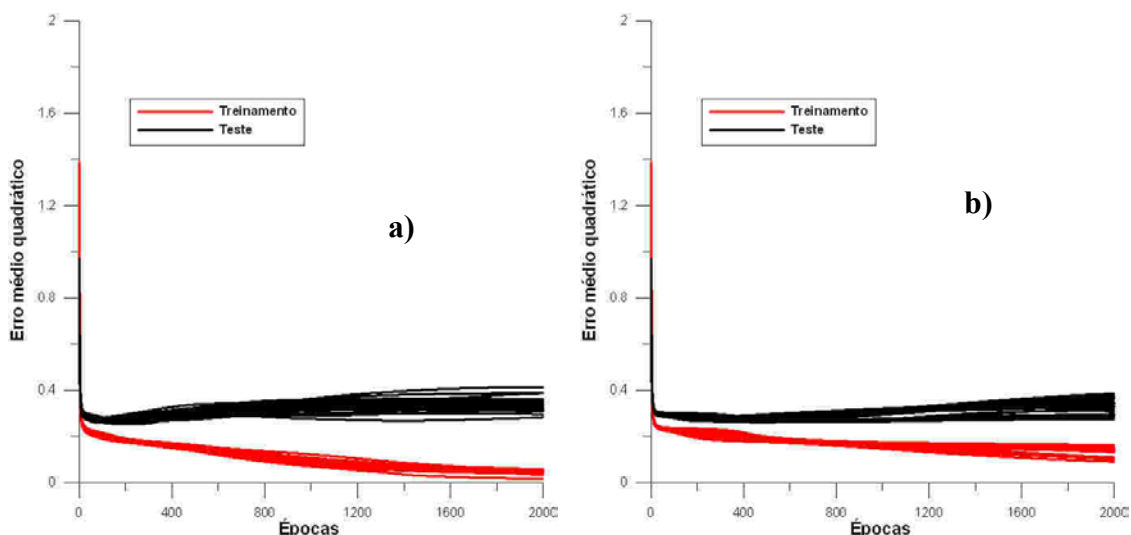


Figura 5-21 Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) CP e (b) *screening* – Rodada 12Z, com dados não replicados, referentes estação Santa Marta.

Tabela 5-8 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodada 12Z, com dados não replicados, referentes estação Santa Marta.

Estatísticas	Rodadas 12Z				
	Média = 6.43 m/s Desvio Padrão= 4.21 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9030	0.9179	0.7388	0.7388	0.9179
TS	0.0714	0	0.16667	0.1667	0
POD	0.0909	0	0.6364	0.6364	0
FAR	0.7500	-	0.8158	0.8158	-
POFD	0.0244	0	0.2520	0.2520	0
B	0.3636	0	3.4545	3.4545	0
Topologia (RN)	16-8-1 2000-31	4-4-1 2000-31	17 coef	5 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.1	0.1	-

Nota: Dados não replicados. Eventos no conjunto teste = 11; Não Eventos = 123  
Eventos no conjunto treinamento = 34; Não Eventos = 504.

Em resumo, verifica-se a relevância do pós-processamento quando o objetivo é a previsão de eventos de vento forte, pois o modelo HRM embora identifique todos os

casos de vento fraco, cuja ocorrência é predominante, falha na identificação dos casos de vento forte. Os métodos de pós-processamento, por outro lado, conseguem identificar alguns casos de vento forte, embora isto acarrete a emissão de alarmes falsos e queda no percentual total de acerto, contudo, verificamos que a rede neural consegue manter estes índices de desempenho em patamares aceitáveis.

Não se identificam diferenças significativas quando se compara o uso de *screening* ou componentes principais.

Observou-se também a importância da replicação dos dados para o aumento da detecção dos casos de vento forte. Contudo, uma consequência, ao se utilizar este recurso, é o aumento na emissão de alarmes falsos e na tendência a superestimar os casos de vento forte, mas verificamos também que as configurações das redes neurais conseguem manter estas medidas de desempenho em patamares aceitáveis.

Observa-se que os experimentos utilizando os horários de 00Z apresentam um bom desempenho em termos de detecção e alarmes falsos, este comportamento se mantém mesmo quando os dados não são replicados. Podemos verificar que este comportamento é diferente do que ocorre nos experimentos das outras duas estações, onde o pós-processamento não apresenta uma boa previsão quando os eventos não são replicados. Uma das explicações para este comportamento é o fato de a amostra, apesar de pequena, apresentar um número razoável de eventos extremos. Além disso, a localização da estação, em uma região sujeita a passagem de sistemas frontais intensos, favorece as ocorrências de vento forte, de tal forma que os métodos de pós-processamento conseguem os subsídios necessários para uma correta previsão do evento.

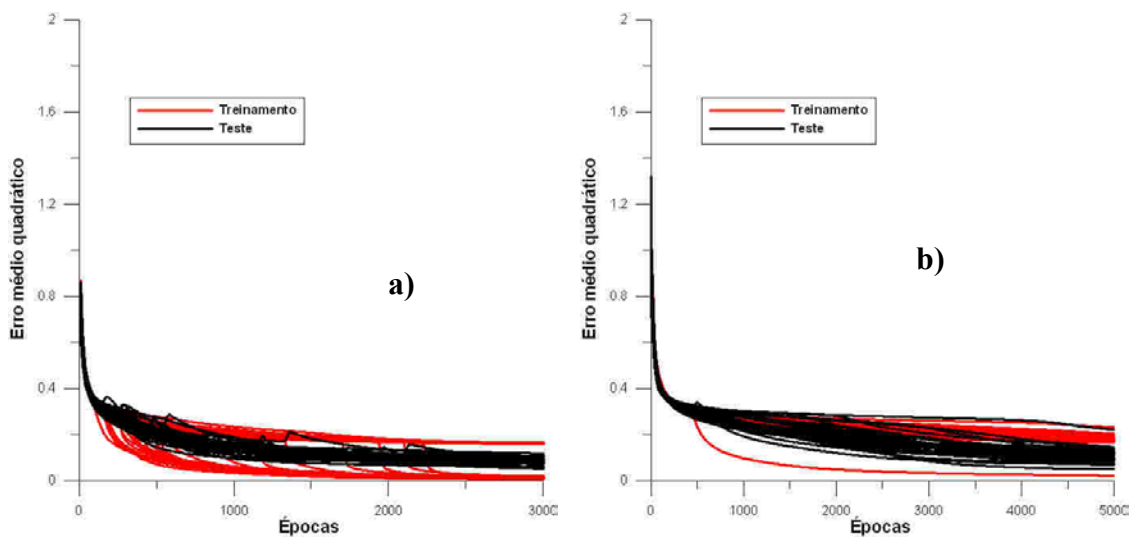
As previsões que utilizam os dois horários e, em particular, com uso de redes neurais, apresentaram melhor desempenho na detecção do evento e no equilíbrio entre estes acertos e emissão de alarmes falsos.

### 5.1.1.3 Mostardas

#### 5.1.1.3.1 Rodadas de 00Z e 12Z

##### 5.1.1.3.1.a Dados Replicados

Os gráficos abaixo exibem o comportamento dos erros durante o treinamento da rede neural. Observa-se a necessidade de mais épocas no experimento onde as variáveis foram escolhidas pelo método *screening*, Figura 5-22b, demonstrando comportamento diferente na distribuição dos erros.



**Figura 5-22** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) componentes principais e (b) *screening* – Rodadas de 00Z e 12Z, com dados replicados, referentes à estação Mostardas.

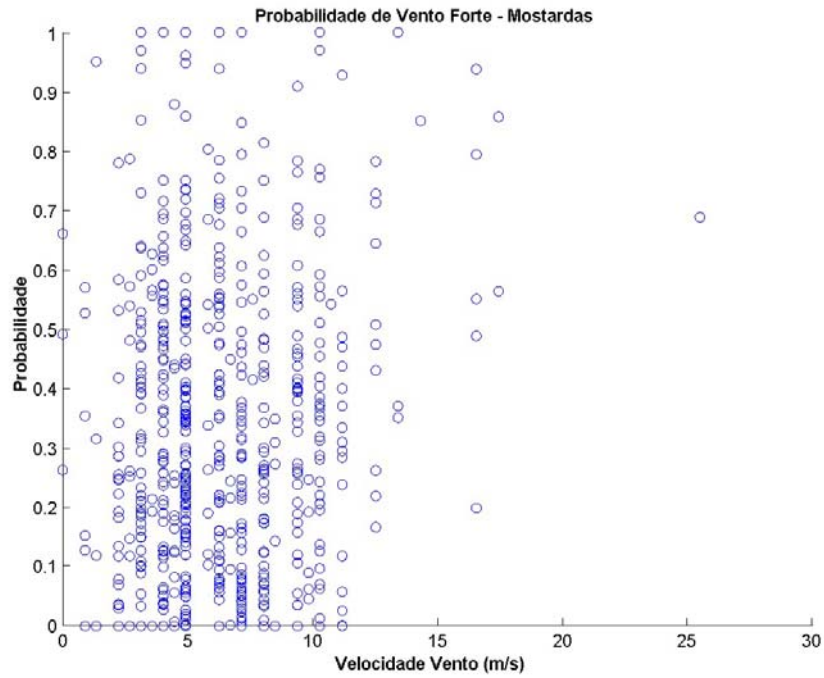
As regressões, obtidas com componentes principais e pelo método *screening*, apresentaram coeficientes de determinação de 43.45% e 47.65%, respectivamente. A Figura 5-23 ilustra os resultados obtidos com as regressões.

Na Tabela 5-9 e na Figura 5-24 são apresentados os resultados dos experimentos.

A rede neural não detecta todos os eventos ( $POD = 0.22$ ), mas consegue manter a acurácia, medida pelo PC, TS e *bias*. O MOS, por outro lado, identifica um número maior de eventos ( $POD=0.67$  e  $POD =0.78$ ), contudo, o número de alarmes falsos, ( $FAR=0.96$  e  $FAR=0.95$ ) e a superestimação do evento ( $BIAS=18$  e  $BIAS=17$ ) ocorrem dentro de patamares que colocam em dúvida a confiabilidade de suas previsões. Além disso, o MOS apresenta alta probabilidade de detecção falsa ( $POFD=0.23$  e  $POFD=0.21$ ) e, em consequência, menor proporção de acerto (PC).

Quando se consideram os resultados do modelo numérico, observa-se que a RN com utilização de componentes principais, embora subestime a previsão do evento ( $Bias=0.56$ ), apresenta melhores resultados em termos de detecção do evento, alarmes falsos e acurácia. O modelo numérico não detecta nenhum dos casos de vento forte.

a)



b)

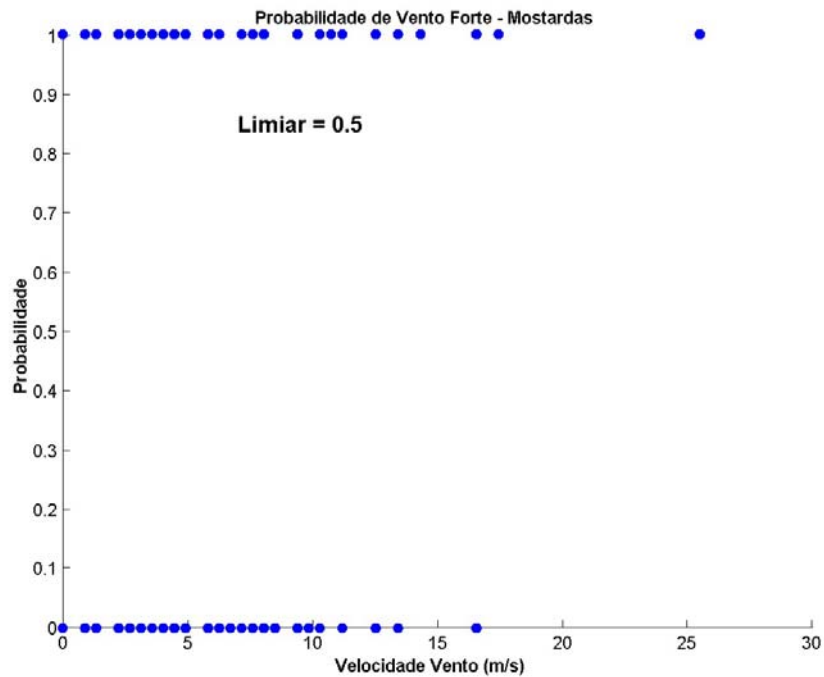
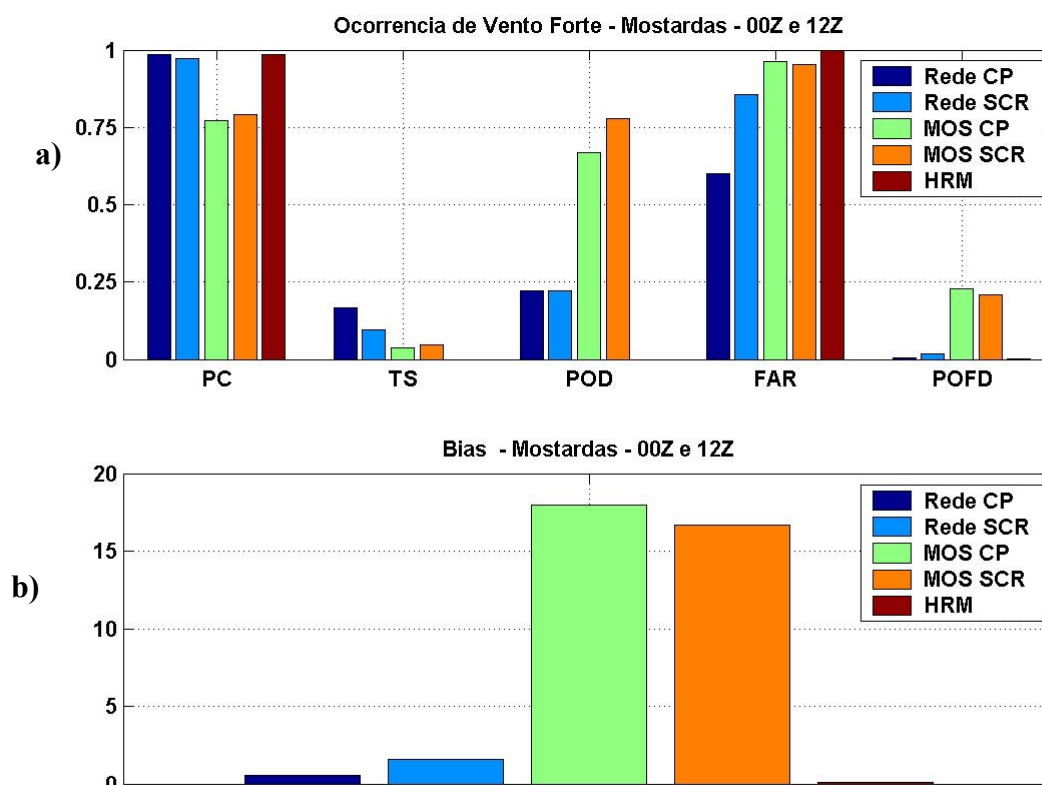


Figura 5-23 Probabilidades para previsão de ocorrência de vento forte obtidas com a equação de regressão utilizando (a) *screening* e as mesmas probabilidade (b) após definição do limiar de 0.5.

**Tabela 5-9** Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodadas de 00Z e 12Z , com dados replicados, referentes à estação Mostardas.

Estatísticas	Rodadas 00Z e 12Z				
	Média = 6.09 m/s Desvio Padrão= 2.88 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9856	0.9727	0.7712	0.7914	0.9856
TS	0.1667	0.0952	0.0364	0.0461	0
POD	0.2222	0.2222	0.6667	0.7778	0
FAR	0.6000	0.8571	0.9630	0.9533	1
POFD	0.0044	0.0175	0.2274	0.2085	0.0015
B	0.5556	1.5556	18	16.6667	0.1111
Topologia (RN)	16-8-1 3000-31	18-9-1 5000-31	17 coef	19 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.1	0.1	-

Nota: Dados replicados. Eventos no conjunto teste =9; Não Eventos = 686  
Eventos no conjunto treinamento = 27; Não Eventos = 2756.

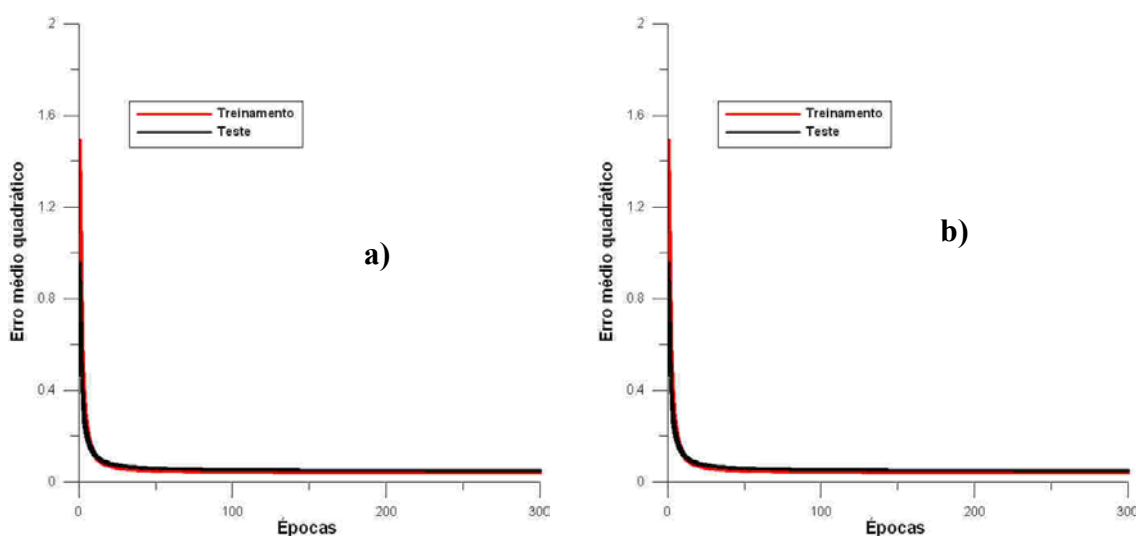


**Figura 5-24** Estatísticas do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 00Z e 12Z , com dados replicados, referentes à estação Mostardas.

### 5.1.1.3.1.b Dados Não Replicados

Os gráficos abaixo (Figura 5-25a e Figura 5-25b) exibem a rápida convergência dos erros, pois, veremos adiante, que a rede classifica corretamente todos os eventos não críticos, mas falha na detecção dos eventos de interesse. O mesmo ocorre para as previsões do MOS, cujas equações, não apresentadas, apresentam pouca significância estatísticas nos dois métodos usados.

As regressões, obtidas com componentes principais e pelo método *screening*, apresentaram coeficientes de determinação de 2.46% e 2.63%, respectivamente.



**Figura 5-25** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de ocorrência de vento forte realizado pela RN utilizando (a) CP e (b) *screening* – Rodadas de 00Z e 12Z, dados não replicados, referentes à estação Mostardas.

Sem a replicação dos dados, os métodos de pós-processamento não conseguem identificar os eventos de interesse.

Embora o desempenho do MOS e da RN superem o do modelo numérico HRM, isto ocorre apenas para a proporção de correção (98.71%), pois esta medida considera



também os eventos de maior ocorrência, não sendo um bom parâmetro de desempenho quando analisamos eventos raros.

**Tabela 5-10 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodadas de 00Z e 12Z , dados não replicados, referentes à estação Mostardas.**

Estatísticas	Rodadas 00Z e 12Z				
	Média = 6.09 m/s Desvio Padrão= 2.88 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
<b>PC</b>	0.9871	0.9871	0.9871	0.9871	0.9856
<b>TS</b>	0	0	0	0	0
<b>POD</b>	0	0	0	0	0
<b>FAR</b>	-	-	-	-	1
<b>POFD</b>	0	0	0	0	0.0015
<b>B</b>	0	0	0	0	0.1111
<b>Topologia (RN)</b>	16-8-1 300-31	4-4-1 300-31	17 coef	5 coef	-
<b>Alfa (RN)</b>	0.1	0.1	-	-	-
<b>Limiar (MOS)</b>	-	-	0.1	0.1	-

Nota: Dados não replicados. Eventos no conjunto teste =9; Não Eventos = 686  
Eventos no conjunto treinamento = 27; Não Eventos = 2756.

Observe que o modelo numérico apresenta menor percentual de acerto, pois previu incorretamente a ocorrência de vento forte. Confirmando uma característica observada em vários experimentos deste trabalho, qual seja, queda na acurácia e aumento de alarmes falsos quando o método tenta prever um evento raro.

Novamente, a importância de se aumentar artificialmente a frequência dos eventos de interesse é demonstrada com esses resultados.

Segue abaixo a conclusão para os experimentos realizados para as três estações:

- a comparação das previsões do modelo numérico com as previsões do MOS e rede neural confirmou a baixa acurácia dos modelos numéricos para a previsão de eventos extremos. O modelo apresenta alto percentual de acerto, considerando-se eventos e não eventos, mas erra na maioria dos casos de vento forte;
- com relação ao uso de componentes principais ou *screening*, não se verificam diferenças significativas;
- a tentativa de identificar ou prever os eventos raros implicará, normalmente, em uma diminuição na proporção total de acertos e aumento dos alarmes falsos;
- a rede neural apresenta melhores resultados em suas previsões que o MOS, embora não consiga identificar todos os casos extremos;
- o MOS não apresenta confiabilidade em suas previsões em vista de produzir um número elevado de falsos alarmes, apesar de identificar, na maior parte dos experimentos, todos os eventos extremos;
- a utilização da amostra com os dados previstos pelo modelo nos dois horários de inicialização apresentou melhores resultados do que o pós-processamento especializado por horários;
- a comparação dos resultados do MOS e da rede neural com os resultados dos modelos mostram a relevância do pós-processamento, uma vez que conseguem identificar um número razoável de casos de vento forte, o que não é possível com o uso exclusivo dos resultados do modelo numérico;
- mesmo nos casos em que havia um número reduzido de eventos de interesse o MOS e a rede neural conseguiram, na maior parte dos casos, prever um número razoável desses eventos;

- observa-se a importância do tamanho da amostra e, em especial, de uma maior ocorrência de eventos para que ocorra um treinamento da rede e um ajustamento da regressão do MOS adequados;

- os resultados obtidos na estação Santa Marta, quando utilizando todos os dados integrados, mostraram-se melhores que nas demais estações, pois está localizada em uma região sujeita a maior ocorrência de ventos fortes, permitindo que os métodos de pós-processamento tenham mais subsídios para uma adequada implementação;

- Observa-se que os resultados dos experimentos apresentam diferenças de acordo com a localização da estação:

a) Salinópolis, por encontrar-se em uma região tropical, ou seja, região com pouca variação climática, apresenta poucas ocorrências de vento forte, o que ocasiona dificuldade para a adequada implementação dos métodos de pós-processamento. Esta característica da estação implica em um desempenho, em termos de detecção dos eventos e emissão de alarmes falsos inferiores aos resultados obtidos nas estações de Santa Marta e Mostardas.

b) O fato de Santa Marta localizar-se em uma região com mais ocorrências de ventos fortes permite que o treinamento da rede e o ajustamento das regressões sejam mais efetivos na previsão dos casos de interesse. Santa Marta apresenta, com exceção do horário de 12Z, uma boa detecção do evento mesmo sem a replicação dos dados. Mostardas embora apresente melhor desempenho que Salinópolis não consegue suplantar os resultados de Santa Marta. Observe que Mostardas necessita da replicação em todos os experimentos. Contudo, Mostardas, também sujeita a um clima temperado e, portanto, com um número maior de casos de ventos forte, apresenta melhores resultados na detecção dos eventos do que Salinópolis.

- Podemos observar a relevância da aplicação de métodos de pós-processamento, em especial uso de redes neurais, quando se comparam suas previsões com os resultados do modelo numérico HRM, que não consegue prever a maioria dos casos de vento forte.

### **5.1.2 Ocorrência de Visibilidade Restrita**

O objetivo desta seção é avaliar os resultados do MOS e da rede neural para a previsão de ocorrência de visibilidade horizontal restrita, definida como inferior a 2 km.

A visibilidade horizontal na superfície não é prevista diretamente pelo modelo. A previsão desta variável é feita subjetivamente pelo previsor do tempo a partir da análise de campos gerados por modelos numéricos, dados meteorológicos de diversas fontes e, principalmente, modelos conceituais locais.

A comparação entre as previsões subjetivas emitidas pelos previsores do tempo e os resultados do pós-processamento não foi possível em virtude da falta da uniformidade na elaboração e arquivamento das informações. Contudo, sabe-se que a previsão de visibilidade é uma tarefa difícil para o previsor, em função da alta subjetividade inerente ao tipo de previsão.

O objetivo é obter resultados, por meio de uma técnica objetiva de previsão, que possam servir de subsídios para a tomada de decisão dos previsores de tempo quando da ocorrência de visibilidade restrita.

O estudo foi feito para três estações costeiras: Salinópolis, Santa Marta e Mostardas.

Serão apresentados apenas os resultados dos experimentos com utilização dos dois horários de inicialização do modelo (00Z e 12Z), pois apresentaram melhor desempenho quando comparados com os resultados de 00Z e 12Z separadamente. A exceção será os resultados obtidos na Estação Santa Marta, cujo desempenho do pós-

processamento foi satisfatório mesmo com a utilização de apenas um horário de previsão. Os resultados dos experimentos para os horários de 00Z e 12Z, separadamente, podem ser consultados no Apêndice A.

### 5.1.2.1 Salinópolis

Neste experimento será analisada a aplicação dos métodos de pós-processamento para a estação Salinópolis.

Salinópolis localiza-se em uma região de clima tropical onde as condições meteorológicas não apresentam grandes variações. Conforme detalhado no capítulo 3, a frequência de visibilidade restrita nesta estação é muito pequena, ocasionando poucos eventos. Esta característica da estação dificulta a aprendizagem da rede neural e o ajustamento das regressões do MOS, em virtude da baixa representatividade dos eventos de interesse.

Para o treinamento da rede e o ajustamento das regressões foram utilizadas como variáveis de entrada (ou preditoras) as componentes principais que respondem por determinado percentual da variação total dos dados e as variáveis escolhidas pelo método de *screening*.

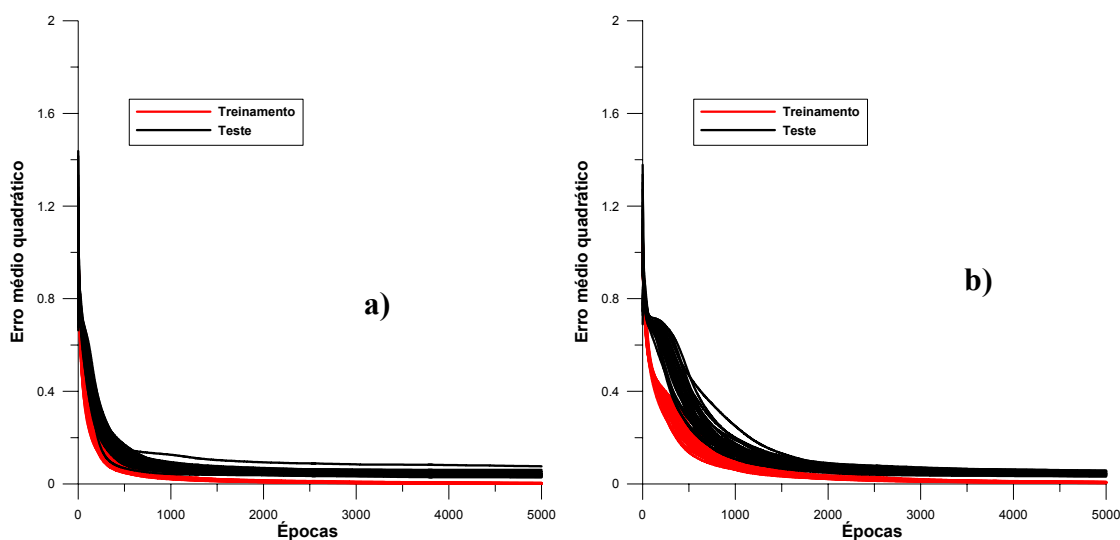
As inferências para as regressões e seus coeficientes serão feitas da mesma forma apresentada no experimento para previsão de vento forte, ou seja, uso do coeficiente de determinação ( $R^2$ ); o p-valor para o estudo da regressão e o p-valor referente aos coeficientes da regressão. Todas as regressões, bem como suas estatísticas, podem ser consultadas no apêndice B. Como ocorreu no estudo sobre previsão de situações de vento forte, serão apresentados resultados considerando três situações: utilização de dados referentes aos horários de 00Z e 12Z integrados, e utilizando-se os dados de 00Z e 12Z separadamente, especializando as implementações por horários.

## 5.1.2.1.1 Rodadas de 00Z e 12Z

### 5.1.2.1.1.a Dados Replicados

A Figura 5-26a e a Figura 5-26b apresentam as evoluções dos erros para os conjuntos de teste e treinamento. Observa-se a rápida convergência para o erro mínimo da rede utilizando componentes principais, enquanto a rede utilizando screening converge mais lentamente.

A análise da evolução do erro durante o treinamento da rede é importante, pois permite uma avaliação da consistência das previsões bem como impedir que estas memorizem as informações e forneçam previsões de baixa qualidade quando se utiliza o conjunto teste. Este ajuste perfeito da rede é denominado *overfitting*.



**Figura 5-26** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de visibilidade restrita realizado pela RN utilizando (a) Componentes principais e (b) *screening* – Rodadas de 00Z e 12Z, dados replicados, referentes à estação Salinópolis.

As 16 componentes principais utilizadas em todos os experimentos respondem por 95% da variação total das 26 variáveis originais utilizadas neste estudo e são

apresentadas na Tabela 4-1, do capítulo 4. A Figura 5-27 ilustra o comportamento dos autovalores em relação às componentes principais.

As regressões, para as retas obtidas com componentes principais e pelo método *screening*, apresentaram coeficientes de determinação de 40.51% e 46.93%, respectivamente. As regressões podem ser consideradas estatisticamente significantes a um nível de 5%. Os coeficientes e as estatísticas das equações do MOS podem ser consultadas no apêndice B.

**Tabela 5-11 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z e 12Z, dados replicados, referentes à estação Salinópolis**

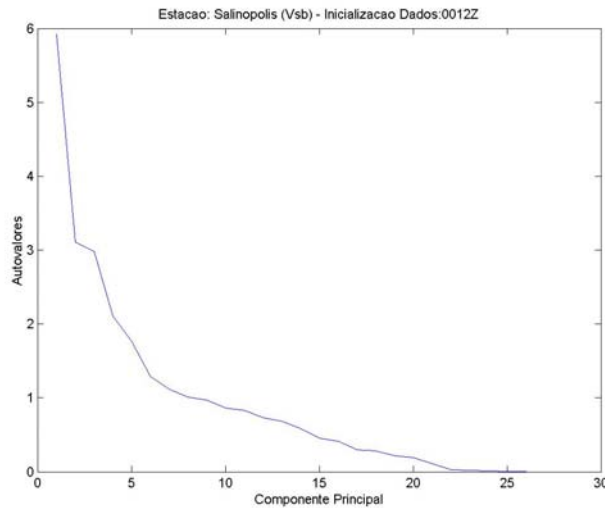
Estatísticas	Rodadas 00Z e 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9934	0.9921	0.7325	0.7351
<b>TS</b>	0	0	0	0
<b>POD</b>	0	0	0	0
<b>FAR</b>	1	1	1	1
<b>POFD</b>	0.0027	0.0040	0.2646	0.2620
<b>B</b>	0.6667	1	66.3333	65.6667
<b>Topologia (RN)</b>	16-8-1 5000-31	24-12-1 5000-31	17 coef.	25 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.5	0.5

Nota: Dados replicados. Eventos no conjunto teste =3; Não Eventos = 752  
Eventos no conjunto treinamento = 11; Não Eventos = 3012.

É importante ressaltar a necessidade de escolha de um limiar para as probabilidades a fim de transformar as previsões dos MOS em valores não probabilísticos.

A Tabela 5-11 exhibe os resultados do pós-processamento. Podemos observar que nenhum dos métodos consegue identificar os eventos de interesse (POD=0).

Conforme já observado no experimento para previsão de vento forte, a baixa representatividade dos casos de visibilidade restrita não permite o treinamento adequado da rede bem como dificulta o ajuste das regressões do MOS.



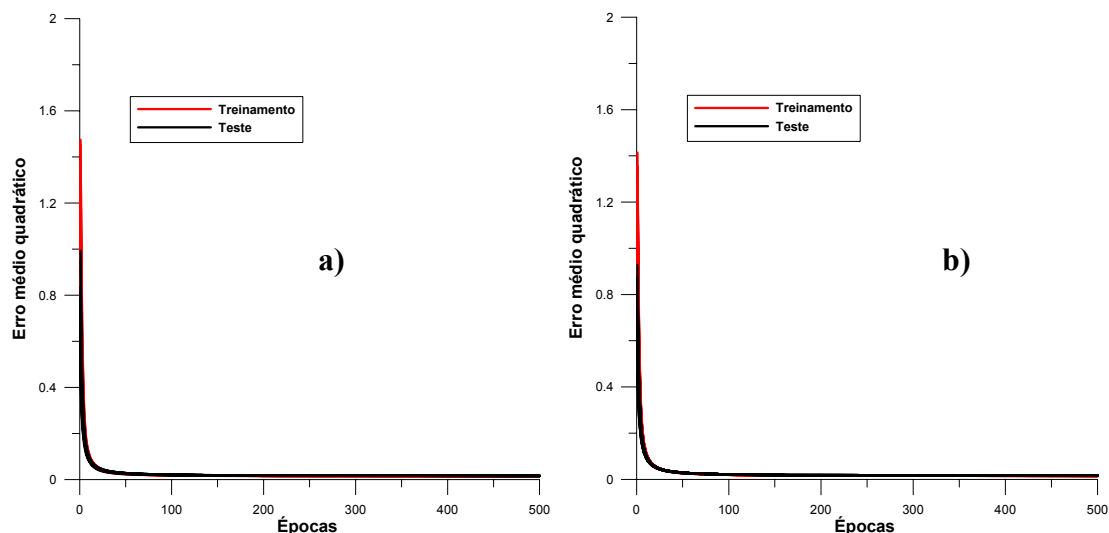
**Figura 5-27** *Scree plot* exibindo os autovalores e as componentes principais calculadas a partir das 26 variáveis originais utilizadas na análise de RN e MOS para previsão de ocorrência de visibilidade restrita para estação Salinópolis.

#### **5.1.2.1.1.b** *Dados Não Replicados*

A Figura 5-28a e a Figura 5-28b apresentam-se as evoluções dos erros para o conjunto teste e treinamento. Observa-se a rápida convergência para o mínimo para as duas topologias da rede. Isto ocorre, pois as duas configurações da rede obtêm alto percentual de acerto ao identificarem todos os casos de boa visibilidade, mas não identificam nenhum dos 3 casos de visibilidade restrita observados na estação.

As regressões, obtidas com componentes principais e pelo método *screening*, apresentaram coeficientes de determinação de 0.63% e 0.25%, respectivamente.





**Figura 5-28** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de visibilidade restrita realizado pela RN utilizando (a) Componentes principais e (b) *screening* - Rodadas de 00Z e 12Z, com dados não replicados, referentes à estação Salinópolis.

**Tabela 5-12** Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z e 12Z, com dados não replicados, referentes à estação Salinópolis.

Estatísticas	Rodadas 00Z e 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9960	0.9960	0.9960	0.9960
<b>TS</b>	0	0	0	0
<b>POD</b>	0	0	0	0
<b>FAR</b>	-	-	-	-
<b>POFD</b>	0	0	0	0
<b>B</b>	0	0	0	0
<b>Topologia (RN)</b>	16-8-1 500-31	1-1-1 500-31	17 coef.	2 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.5	0.5

Nota: Dados replicados. Eventos no conjunto teste = 3; Não Eventos = 752  
Eventos no conjunto treinamento = 11; Não Eventos = 3012.

A regressão com uso de componentes principais não foi considerada significativa. O p-valor de 0.26, obtido pela regressão, não sustenta a hipótese de que

existe uma relação de regressão entre as 16 componentes principais e a variável a ser prevista (ocorrência de visibilidade menor que 2 km).

O modelo de regressão com uso de *screening*, embora seja considerado significativo a um nível de 5%, selecionou apenas a variável umidade relativa em 850hPa, dentre as 25 variáveis apresentadas, e obteve coeficiente de determinação muito baixo.

A Tabela 5-12 exhibe os resultados do pós-processamento. Podemos observar que nenhum dos métodos consegue identificar os eventos de interesse.

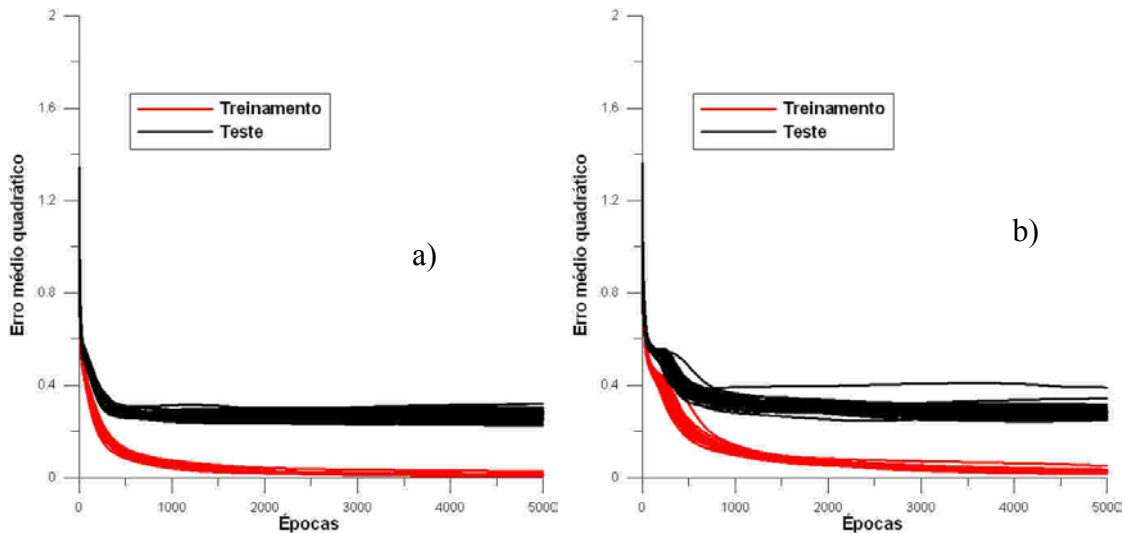
## **5.1.2.2 Santa Marta**

### **5.1.2.2.1 Rodadas de 00Z e 12Z**

#### **5.1.2.2.1.a Dados Replicados**

A Figura 5-29a e a Figura 5-29b apresentam as evoluções dos erros para os conjuntos de teste e treinamento. Observa-se a rápida convergência do erro para o mínimo para na rede que utiliza componentes principais e uma convergência mais lenta para a rede que selecionou a variáveis pelo *screening*.

As regressões com uso de componentes principais e *screening* apresentaram coeficientes de determinação iguais a 43.53% e 48.51% , respectivamente e são consideradas significantes a um nível de 5%. A Figura 5-30a e a Figura 5-30b ilustram os resultados das regressões antes e depois de definido o valor do limiar. Observe que um dos casos de visibilidade menor que 2 km recebeu probabilidade de ocorrência igual a zero, ou seja, para qualquer limiar escolhido não será possível a previsão deste caso.



**Figura 5-29** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de visibilidade restrita realizado pela RN utilizando (a) Componentes principais e (b) *screening* – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Santa Marta.

Na análise da Tabela 5-13 e da Figura 5-31a, observa-se que as duas configurações do MOS apresentam alta probabilidade de detecção do evento. Conseguem identificar 17 dos 19 casos. Contudo, perdem em acurácia, como pode ser verificado pelos baixos valores de PC e TS, e pelo alto número de alarmes falsos. Emitem 67 alarmes falsos em um total de 84 tentativas ( $FAR=0.8$ ) e para um total de 258 casos de não eventos apresentam  $POFD=0.26$ .

As duas configurações da rede neural não conseguem detectar todos os eventos. A rede com uso de componentes principais identifica 9 dos 19 casos ( $POD=0.47$ ) e a que usa *screening* detecta 10 ( $POD= 0.53$ ). Entretanto, a emissão de alarmes falsos fica na ordem de 8 em 18 tentativas ( $FAR=0.44$ ) e em relação ao total de não eventos temos  $POFD = 0.03$ . A rede apresenta melhor desempenho em termos de outra medida de acurácia, importante para previsão de eventos raros, o TS.

Na análise da tendência ( Figura 5-31b) temos que as duas configurações da rede neural subestimam as previsões, mas apresentam *bias* próximo do valor ideal. O MOS superestima a previsão do evento em torno 4 vezes.

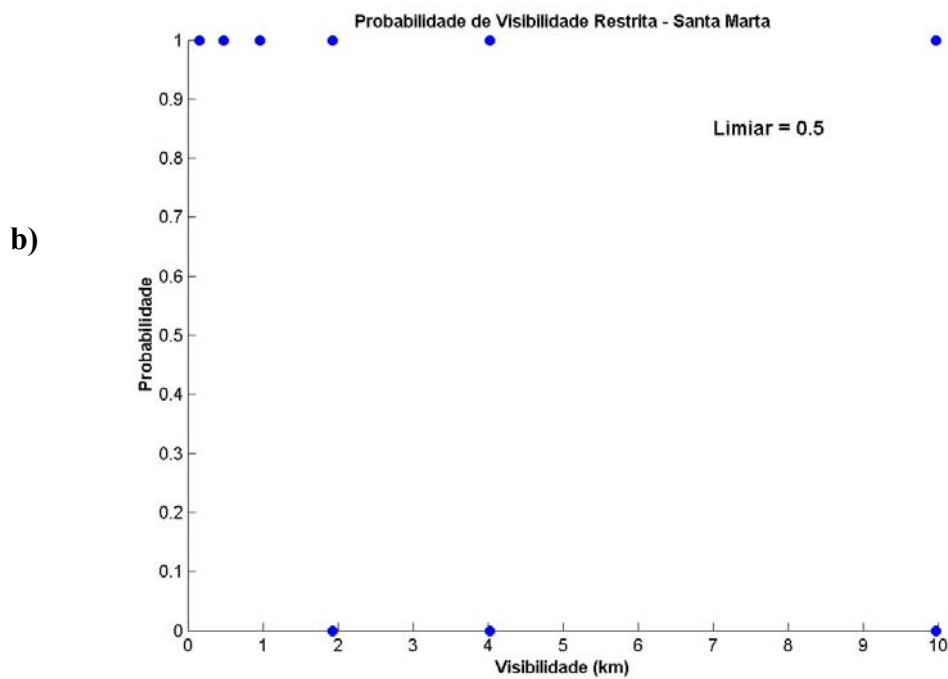
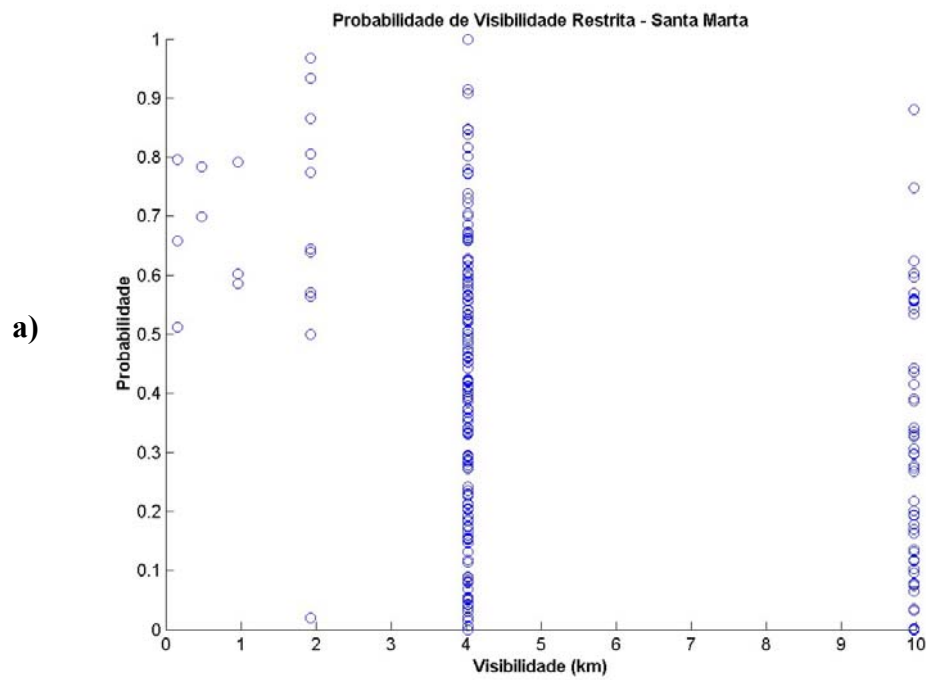
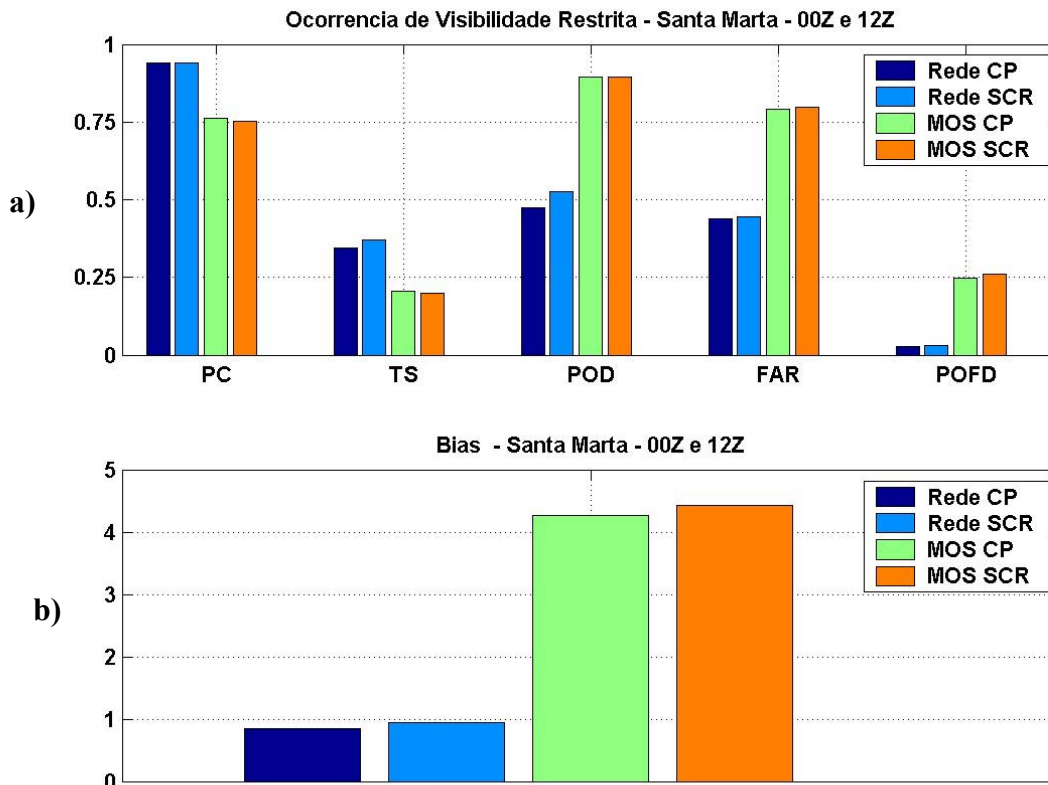


Figura 5-30 Probabilidades para previsão de visibilidade restrita obtidas com a equação de regressão utilizando *screening* (a) antes e (b) depois de definido limiar de 0.5

**Tabela 5-13 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Santa Marta.**

Estatísticas	Rodadas 00Z e 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
PC	0.9386	0.9386	0.7617	0.7509
TS	0.3462	0.3704	0.2048	0.1977
POD	0.4737	0.5263	0.8947	0.8947
FAR	0.4375	0.4444	0.7901	0.7976
POFD	0.0271	0.0310	0.2481	0.2597
B	0.8421	0.9474	4.2632	4.4211
Topologia (RN)	16-8-1 5000-31	16-8-1 5000-31	17 coef.	17 coef.
Alfa (RN)	0.1	0.1	-	-
Limiar (MOS)	-	-	0.5	0.5

Nota: Dados replicados. Eventos no conjunto teste =19; Não Eventos = 258. Eventos no conjunto treinamento = 46; Não Eventos = 1065.



**Figura 5-31 Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Santa Marta.**

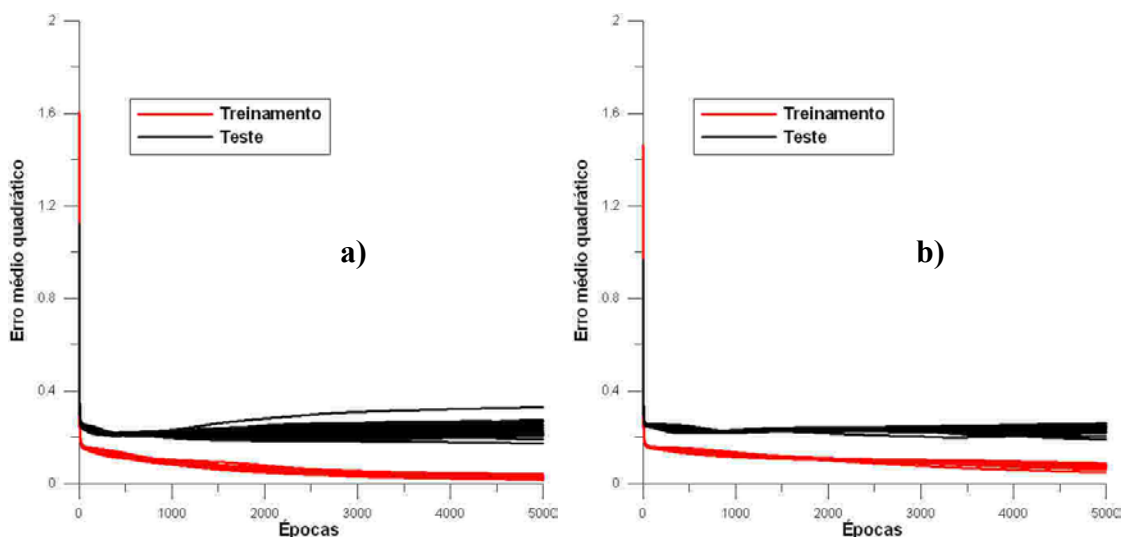
A análise dos resultados, em relação ao método para uso das variáveis preditores, componentes principais ou *screening*, não revelam diferenças significativas.

Pode-se observar que as configurações da rede neural apresentam maior confiabilidade e conseguem manter a confiabilidade e acurácia nas previsões.

### 5.1.2.2.1.b Dados Não Replicados

A Figura 5-32a e a Figura 5-32b apresentam a evolução dos erros para os conjuntos de teste e treinamento. Observa-se a rápida convergência para o mínimo para as duas configurações da rede.

As regressões com uso de componentes principais e *screening* apresentaram o mesmo coeficiente de determinação com valor de 7.3% e são consideradas significantes a um nível de 5%. Observa-se a queda no coeficiente em relação a experimento com replicação dos dados.

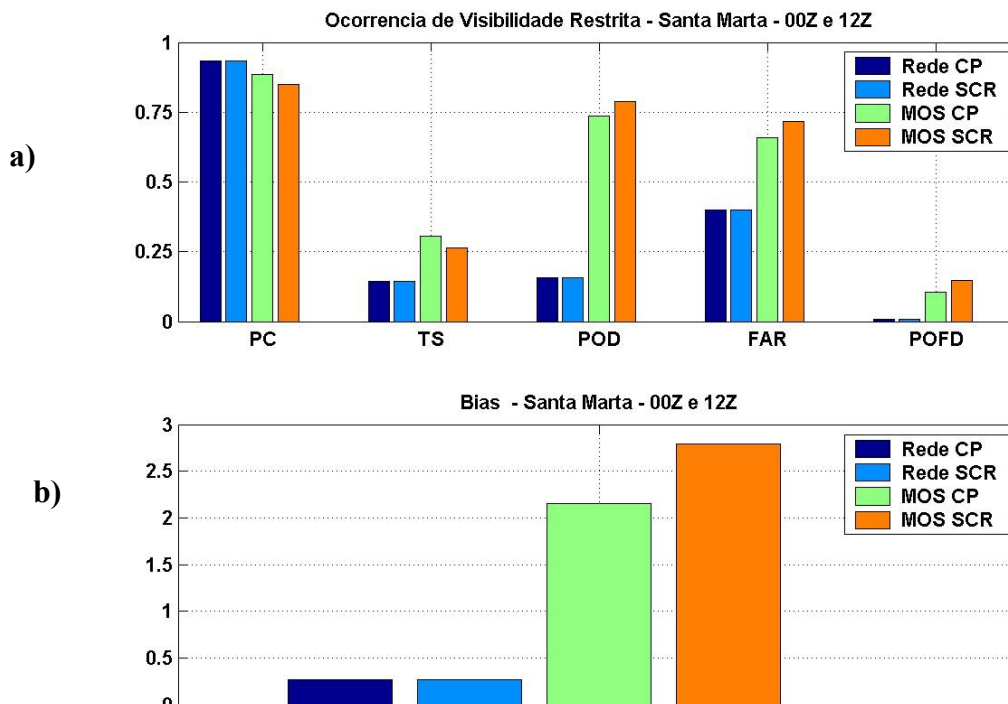


**Figura 5-32** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de visibilidade restrita realizado pela RN utilizando (a) Componentes principais e (b) *screening* – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Santa Marta.

**Tabela 5-14 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Santa Marta.**

Estatísticas	Rodadas 00Z e 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9350	0.9350	0.8845	0.8484
<b>TS</b>	0.1429	0.1429	0.3043	0.2632
<b>POD</b>	0.1579	0.1579	0.7368	0.7895
<b>FAR</b>	0.4000	0.4000	0.6585	0.7170
<b>POFD</b>	0.0078	0.0078	0.1047	0.1473
<b>B</b>	0.2632	0.2632	2.1579	2.7895
<b>Topologia (RN)</b>	16-8-1 5000-31	7-7-1 5000-31	17 coef.	8 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.1	0.1

Nota: Dados não replicados. Eventos no conjunto teste =19; Não Eventos = 258. Eventos no conjunto treinamento = 46; Não Eventos = 1065.



**Figura 5-33 Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Santa Marta.**

Apesar da baixa correlação, as regressões conseguem manter um alto valor de detecção em ambas configurações. O MOS com uso de componentes principais detecta 14 dos 19 casos de visibilidade restrita,  $POD=0.74$ . O MOS com uso de *screening* prevê 15 casos,  $POD=0.79$ . Os alarmes falsos são altos em comparação com os resultados da rede neural. Contudo, as duas configurações da rede subestimam a previsão do evento. Além disso, a análise do TS mostra o melhor desempenho das previsões do MOS.

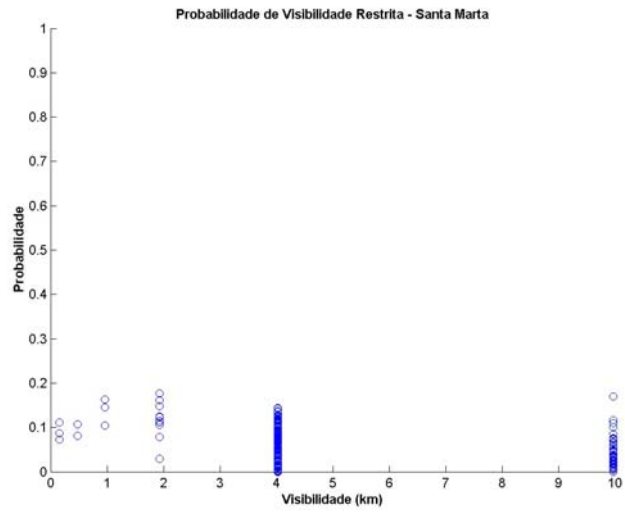
Não se verificam diferenças significativas quanto ao uso de *screening* ou componentes principais.

Apesar do bom resultado, as regressões do MOS apresentam baixa significância de suas probabilidades, conforme demonstrado na Figura 5-34a e Figura 5-34b, de tal forma que o método apresenta baixa discriminação dos eventos. Ou seja, em termos de significância estatística seria indicado o uso das regressões apresentadas na seção anterior, onde os dados foram replicados.

Em resumo, observa-se o baixo desempenho do pós-processamento sem o aumento superficial da amostra de eventos de interesse quando se compara com os resultados com replicação dos dados. Observa-se que a rede neural apresenta melhor aproveitamento da aplicação do método de replicação, pois aumenta sua taxa de detecção, mantendo as taxas de alarmes falsos em patamares razoáveis, além de não diminuir excessivamente o percentual total de acertos. Enquanto o MOS, ao aumentar sua probabilidade de detecção, perde a acurácia e a confiabilidade de suas previsões, em virtude do alto número de alarmes falsos emitidos.



a)



b)

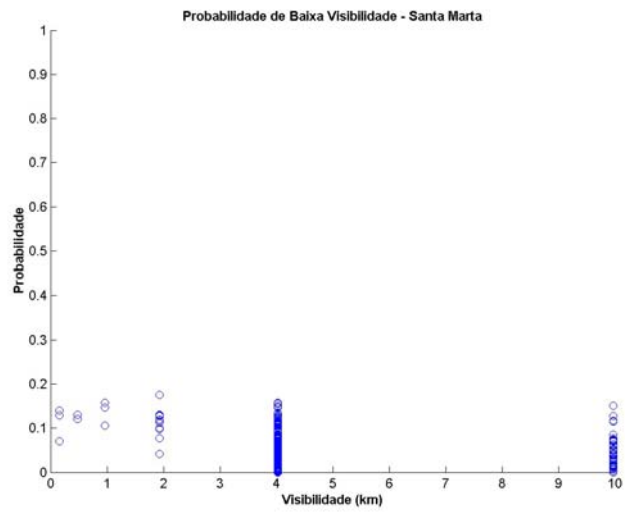


Figura 5-34 Probabilidades obtidas para previsão de visibilidade restrita com a equação de regressão utilizando (a) componentes principais e (b) *screening*.

## 5.1.2.2.2 Rodada de 00Z

### 5.1.2.2.2.a Dados Replicados

Para o conjunto de dados utilizados neste experimento, as regressões com uso de componentes principais e *screening* apresentaram coeficientes de determinação iguais a 51.48% e 52.78% , respectivamente. As regressões são consideradas significantes a um nível de 5%.

Os gráficos das evoluções dos erros dos conjuntos de teste e treinamento para as 5000 épocas e 31 conjuntos (não apresentados) mostram rápida convergência para o mínimo nas duas configurações da rede.

Na análise da Figura 5-35 e da Tabela 5-15 observa-se que as duas configurações do MOS e a rede neural utilizando *screening* apresentam maior detecção do evento, identificando 2 dos 3 casos de visibilidade restrita. A rede neural com uso de componentes principais identificou apenas 1 dos 3 casos.

Em termos de alarmes falsos as duas configurações da rede apresentam menores taxas, FAR e POFD.

O MOS perde em acurácia ao diminuir o percentual total de acertos e no valor do TS.

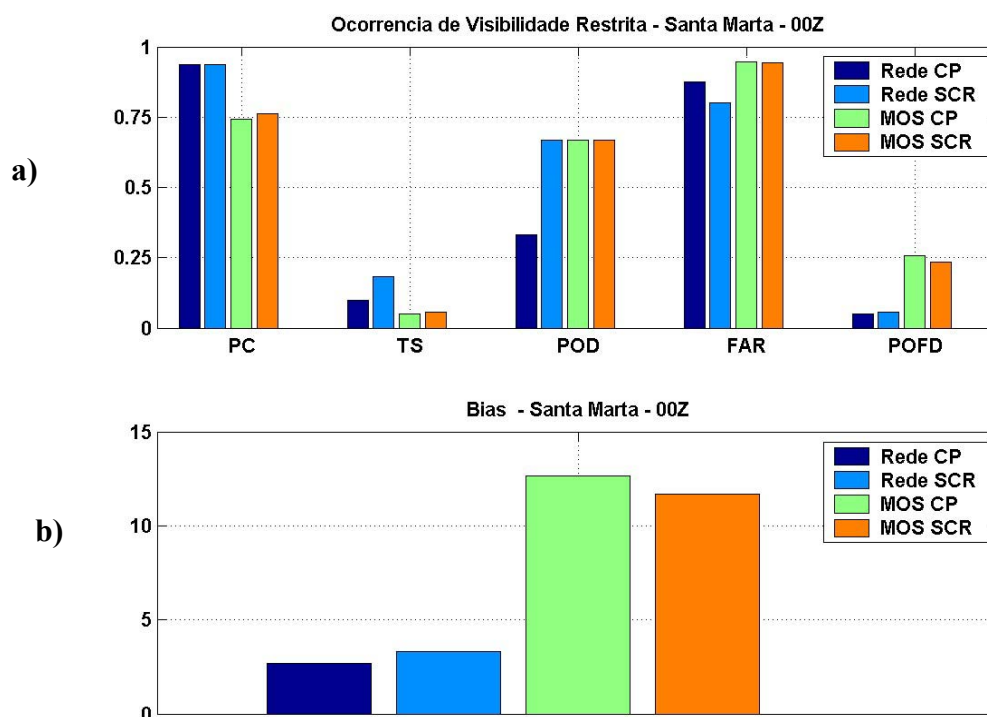
Ambos métodos de pós-processamento superestimam a previsão do evento, mas as redes apresentam *bias* inferiores (Figura 5-35b).

As previsões das redes neurais apresentam maior confiabilidade, embora não consigam prever todos os casos. Especificamente, a rede onde as variáveis foram escolhidas pelo método de *screening* apresentou maior acurácia em suas previsões.

**Tabela 5-15 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 00Z, com os dados replicados, referentes à estação Santa Marta.**

Estatísticas	Rodada 00Z			
	RN CP	RN SCR	MOS CP	MOS SCR
PC	0.9371	0.9371	0.7413	0.7622
TS	0.1000	0.1818	0.0513	0.0556
POD	0.3333	0.6667	0.6667	0.6667
FAR	0.8750	0.8000	0.9474	0.9429
POFD	0.0500	0.0571	0.2571	0.2357
B	2.6667	3.3333	12.6667	11.6667
Topologia (RN)	16-8-1 5000-31	14-7-1 5000-31	17 coef.	15 coef.
Alfa (RN)	0.1	0.1	-	-
Limiar (MOS)	-	-	0.5	0.5

Nota: Dados replicados. Eventos no conjunto teste =3; Não Eventos = 140. Eventos no conjunto treinamento = 24; Não Eventos = 549.



**Figura 5-35 Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodada de 00Z, com os dados replicados, referentes à estação Santa Marta.**

### 5.1.2.2.2.b Dados Não Replicados

Na Tabela 5-16 observa-se que os resultados do MOS apresentam maior detecção dos eventos, contudo mantêm comportamento similar ao encontrado no experimento com utilização dos dois horários de previsão e sem replicação dos dados, onde as probabilidades obtidas pelas regressões não se mostraram significativas. Além disso, apresentam baixos coeficientes de determinação, 9.98% e 8.84% para a regressão com utilização de componentes principais e *screening*, respectivamente.

**Tabela 5-16 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 00Z, com os dados não replicados, referentes à estação Santa Marta.**

Estatísticas	Rodada 00Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9790	0.9790	0.8322	0.8112
<b>TS</b>	0.2500	0	0.0769	0.0690
<b>POD</b>	0.3333	0	0.6667	0.6667
<b>FAR</b>	0.5000	-	0.9200	0.9286
<b>POFD</b>	0.0071	0	0.1643	0.1857
<b>B</b>	0.6667	0	8.3333	9.3333
<b>Topologia (RN)</b>	16-8-1 500-31	5-5-1 500-31	17 coef.	6 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.1	0.1

Nota: Dados não replicados. Eventos no conjunto teste =3; Não Eventos = 140. Eventos no conjunto treinamento = 24; Não Eventos = 549.

A rede neural com *screening* é a mais afetada pela não replicação dos dados, não identificando nenhum dos 3 casos de visibilidade restrita, enquanto a rede com utilização de componentes principais detecta 1 dos casos.

De forma geral, as duas configurações do MOS apresentam melhores resultados do que os obtidos com a replicação dos dados, e o mesmo ocorre para a rede neural com

uso de componentes principais. Para estas configurações observa-se queda nos índices de alarmes falsos e de tendência enquanto a proporção de detecção do evento é mantida.

Observa-se o comportamento diferente das previsões obtidas em Santa Marta em comparação com os resultados de Salinópolis. O maior número de casos e a climatologia da região fornecem os subsídios necessários para o treinamento da rede e para a modelagem do MOS. De tal forma que os métodos conseguem, mesmo sem que os dados sejam replicados, identificar os casos de interesse.

Em resumo, embora os métodos tenham conseguido identificar alguns casos de visibilidade restrita, a baixa significância das regressões indica que a replicação dos dados é necessária a fim se obter resultados estatisticamente consistentes.

### **5.1.2.2.3 Rodada de 12Z**

#### **5.1.2.2.3.a Dados Replicados**

Para o conjunto de dados utilizado neste experimento, as regressões com uso de componentes principais e *screening* apresentaram coeficientes de determinação iguais a 45.66% e 50.82% , respectivamente. As regressões são consideradas significantes a um nível de 5%.

A Tabela 5-17 e a Figura 5-36 apresentam os resultados dos experimentos. Observa-se que o MOS com uso de *screening* apresenta maior detecção do evento, identificando 10 dos 11 casos de visibilidade restrita, enquanto o MOS com uso de componentes principais identifica 9 dos casos de interesse. Contudo, repete o comportamento dos outros experimentos no que se refere ao alto número de alarmes falsos, mas em proporções menores que nos experimentos anteriores. Além disso, apresenta um bom desempenho em relação ao TS, medida relacionado com a previsão de eventos raros.

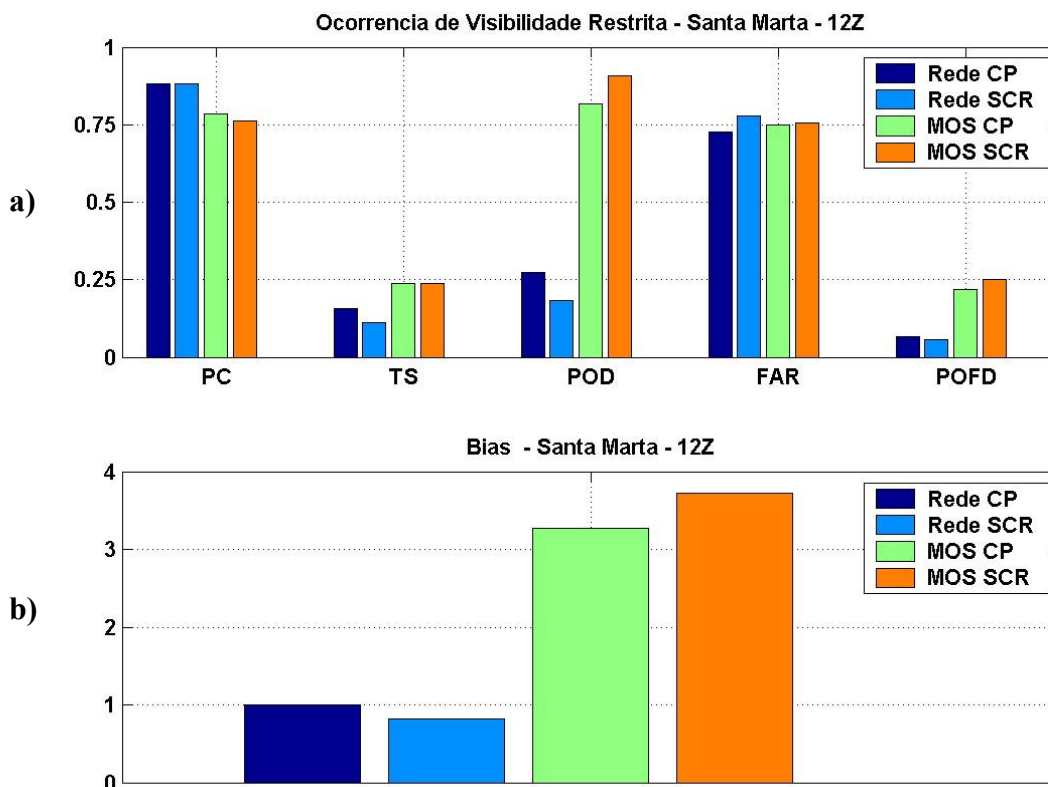
Para obtenção das altas proporções de detecção, o MOS emite altos números de alarmes falsos, principalmente a proporção de detecção falsa (POFD) 27 (componentes principais) e 31 (*screening*) alarmes em um total de 123 não eventos. Observe que esta medida está diretamente relacionada com a redução no percentual de acertos, medido pela PC, de tal forma que altos valores do POFD implicam em menor proporção de acertos, isto é, o método tende a errar os eventos de visibilidade boa.

As duas configurações da rede neural, embora mantenham alta a proporção de acertos, identifica apenas 3 dos 11 casos de visibilidade restrita na rede com componentes principais e 2 dos casos na rede com uso de *screening*.

**Tabela 5-17 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 12Z, com os dados replicados, referentes à estação Santa Marta.**

Estatísticas	Rodada 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.8806	0.8806	0.7836	0.7612
<b>TS</b>	0.1579	0.1111	0.2368	0.2381
<b>POD</b>	0.2727	0.1818	0.8182	0.9091
<b>FAR</b>	0.7273	0.7778	0.7500	0.7561
<b>POFD</b>	0.065	0.0569	0.2195	0.2520
<b>B</b>	1	0.8182	3.2727	3.7273
<b>Topologia (RN)</b>	16-8-1 5000-31	19-10-1 5000-31	17 coef.	20 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.5	0.5

Nota: Dados replicados. Eventos no conjunto teste =11; Não Eventos = 123. Eventos no conjunto treinamento = 27; Não Eventos = 511.



**Figura 5-36 Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodada de 12Z, com os dados replicados, referentes à estação Santa Marta.**

Apesar do menor número de eventos de interesse observados (apenas 3) o experimento anterior (rodada 00Z e replicado) obteve maior detecção dos eventos na configuração de sua rede em comparação com os resultados da rede do presente experimento. Contudo, o MOS apresentado aqui tem melhores resultados. Esta situação pode estar relacionada com o fato de casos de visibilidade restrita ocorrem mais no período da manhã. Como as previsões com dados inicializados em 00Z estão mais próximas dos eventos, seus acertos podem ser atribuídos ao fato de haver menores erros cumulativos, ou degradação da previsão, num período mais curto que as previsões que utilizam os dados de 12Z.

Neste experimento, a rede neural não consegue bom desempenho na detecção dos eventos. As previsões do MOS, apesar de não poderem ser consideradas melhores, em função da emissão de alarmes falsos (POFD alto e queda no PC), apresenta desempenho melhor que nos experimentos anteriores.

#### **5.1.2.2.3.b      *Dados Não Replicados***

Para o conjunto de dados utilizado neste experimento, as regressões com uso de componentes principais e *screening* apresentaram coeficientes de determinação iguais a 9.7% e 6.94% , respectivamente. As regressões são consideradas significantes a um nível de 5%.

A Tabela 5-18 e a Figura 5-37 apresentam os resultados. O MOS apresenta resultado similar ao obtido no experimento anterior. Contudo, as baixas probabilidades obtidas a partir suas regressões (figura não apresentada) bem como os baixos valores dos coeficientes de determinação tornam estas previsões pouco significantes estatisticamente.

As configurações da rede neural não apresentam boa detecção do evento e subestimam o mesmo quando o experimento usa componentes principais. A rede neural com *screening* não identifica nenhum dos casos de visibilidade restrita.

Em resumo, para os experimentos feitos para a estação Santa Marta verificamos que os melhores resultados ocorrem quando os dados são replicados e quando a amostra utiliza as previsões do modelo numérico dos dois horários de análise. O uso dos dois horários, além de aumentar o tamanho da amostra, aumenta o número de casos de visibilidade restrita, facilitando o aprendizado da rede e o ajustamento das regressões para os casos mais extremos.



Contudo, a estação Santa Marta apresenta bons resultados, mesmo quando os dados não são replicados, por se localizar em uma região onde a ocorrência de casos de visibilidade restrita é maior.

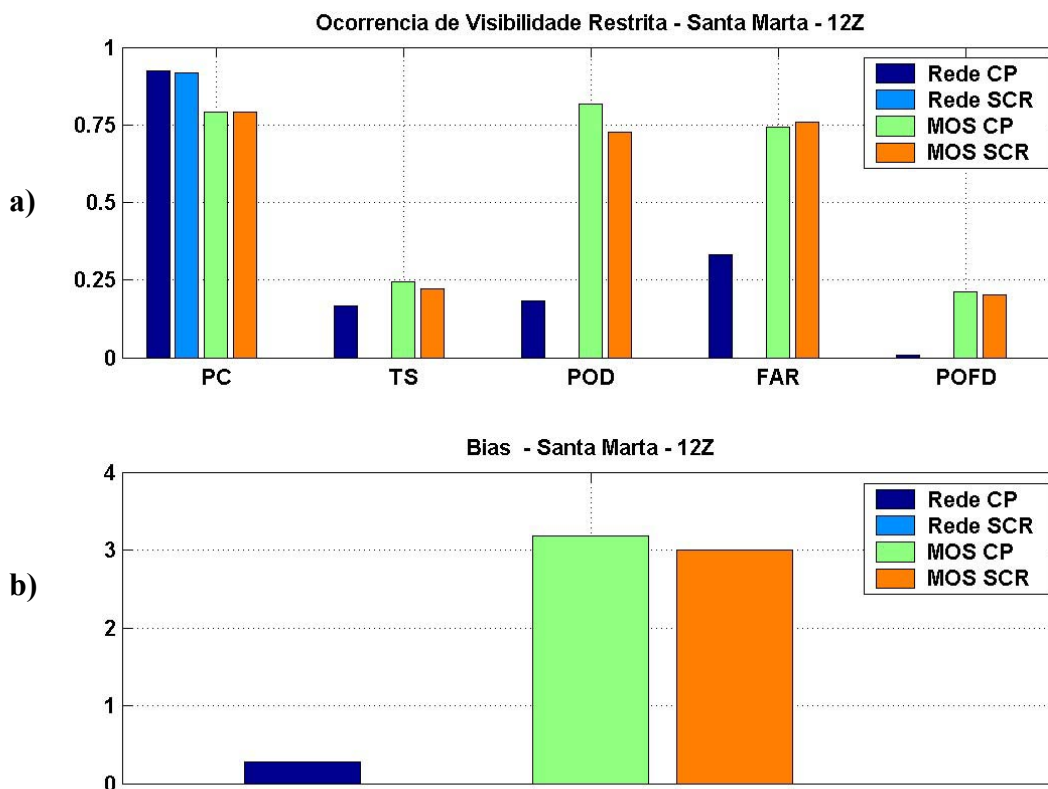
**Tabela 5-18 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita Rodada de 12Z, com os dados não replicados, referentes à estação Santa Marta.**

Estatísticas	Rodada 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9254	0.9179	0.7910	0.7910
<b>TS</b>	0.1667	0	0.2432	0.2222
<b>POD</b>	0.1818	0	0.8182	0.7273
<b>FAR</b>	0.3333	-	0.7429	0.7576
<b>POFD</b>	0.0081	0	0.2114	0.2033
<b>B</b>	0.2727	0	3.1818	3
<b>Topologia (RN)</b>	16-8-1 5000-31	4-4-1 5000-31	17 coef.	5 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.1	0.1

Nota: Dados não replicados. Eventos no conjunto teste =11; Não Eventos = 123. Eventos no conjunto treinamento = 27; Não Eventos = 511.

O MOS apresenta como característica uma alta detecção dos casos de visibilidade restrita, mas um alto número de alarmes falsos, de tal forma que suas previsões perdem a confiabilidade. Observe que esta característica também foi apresentada nos resultados do MOS para previsão de ocorrência de ventos fortes.

A rede neural, apesar de não identificar todos os eventos, apresenta melhor confiabilidade em suas previsões, pois embora não identifique todos os eventos, o número de alarmes falsos e o percentual total de acertos se mantêm em patamares razoáveis. Além disso, apesar de superestimar os eventos, apresentou *bias* menores do que o MOS em todos os experimentos.



**Figura 5-37 Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodada de 120Z, com os dados não replicados, referentes à estação Santa Marta.**

Observe o melhor desempenho do pós-processamento e a relevância do mesmo quando se compara os resultados obtidos em Santa Marta com os obtidos em Salinópolis. Indicando que a eficiência do pós-processamento também se relaciona com a localização e com as características climáticas do local de interesse.

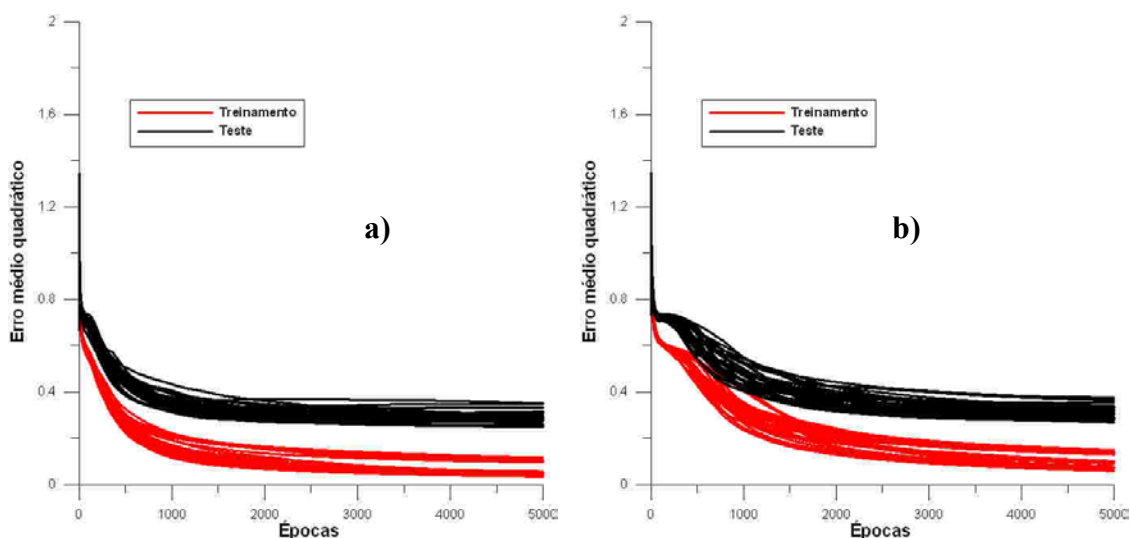
### 5.1.2.3 Mostardas

#### 5.1.2.3.1 Rodadas de 00Z e 12Z

##### 5.1.2.3.1.a Dados Replicados

A Figura 5-38 exibe a evolução dos erros do treinamento e do teste para as duas configurações da rede neural. Observa-se a rápida convergência para o mínimo na rede

utilizando componentes principais, enquanto a rede com *screening* apresenta uma convergência mais lenta. O objetivo, ao apresentar a evolução dos erros, é demonstrar a consistência dos resultados obtidos a partir de suas configurações.



**Figura 5-38** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de visibilidade restrita realizado pela RN utilizando (a) Componentes principais e (b) *screening* – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Mostardas.

Para o conjunto de dados utilizados neste experimento as regressões com uso de componentes principais e *screening* apresentaram coeficientes de determinação iguais a 33.21% e 36.17% , respectivamente. As regressões são consideradas significantes a um nível de 5%.

Na **Tabela 5-19** e na Figura 5-39 são apresentados os resultados do pós-processamento. Não se observam grandes diferenças nos experimentos quando os resultados são analisados em termos do uso de componentes principais e *screening*.

O resultado do MOS com uso de *screening* mostra maior detecção do evento, identificando 26 entre os 29 casos de visibilidade restrita. Contudo, o número de alarmes falsos emitidos foi de 213 alarmes em 239 tentativas (FAR=0.89) e em relação ao total de não eventos obtém POFD=0.32. Conforme já observado, esta última medida

de alarmes falsos implica em diminuição no percentual total de acertos medido pelo PC. Além disso, apresenta superestimação do evento.

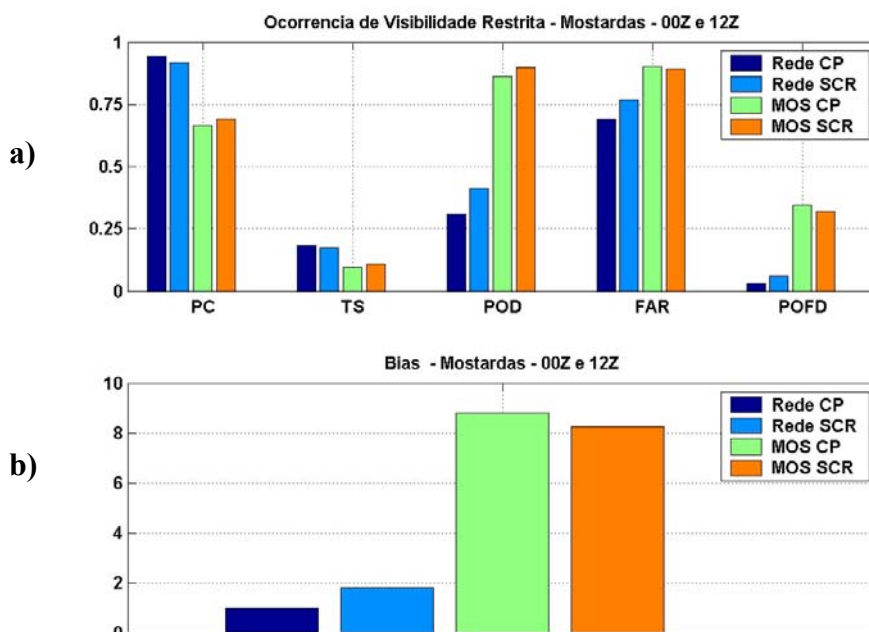
A rede neural não consegue prever todos os casos de visibilidade restrita, identificando 9 e 12 dos 29 casos, para as redes com componentes principais e *screening*, respectivamente. Entretanto, observamos que o número de alarmes falsos e as tendências de suas previsões apresentam-se de tal forma que mantêm a confiabilidade de suas previsões. A rede com componentes principais emite 20 alarmes falsos em 29 tentativas de prever o evento ( $FAR=0.6897$ ) em relação ao total de não eventos temos  $POFD=0.03$ , o qual não implica em queda brusca no percentual de acerto ( $PC=0.94$ ). Enquanto, a rede com *screening* emite 40 alarmes falsos em 57 tentativas de prever o evento ( $FAR=0.77$ ) e apresenta  $POFD =0.06$ , de tal forma que o percentual total de acerto não diminui tanto quanto no MOS. Com relação ao *bias* (Figura 5-39b) não apresenta tendência.

Em resumo, verificamos que a rede neural consegue manter a confiabilidade de suas previsões.

**Tabela 5-19 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Mostardas.**

Estatísticas	Rodadas 00Z e 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
PC	0.9425	0.9181	0.6638	0.6897
TS	0.1837	0.1739	0.0965	0.1074
POD	0.3103	0.4138	0.8621	0.8966
FAR	0.6897	0.7692	0.9020	0.8912
POFD	0.0300	0.0600	0.3448	0.3193
B	1	1.7931	8.7931	8.2414
Topologia (RN)	16-8-1 5000-31	16-8-1 5000-31	17 coef.	17 coef.
Alfa (RN)	0.1	0.1	-	-
Limiar (MOS)	-	-	0.5	0.5

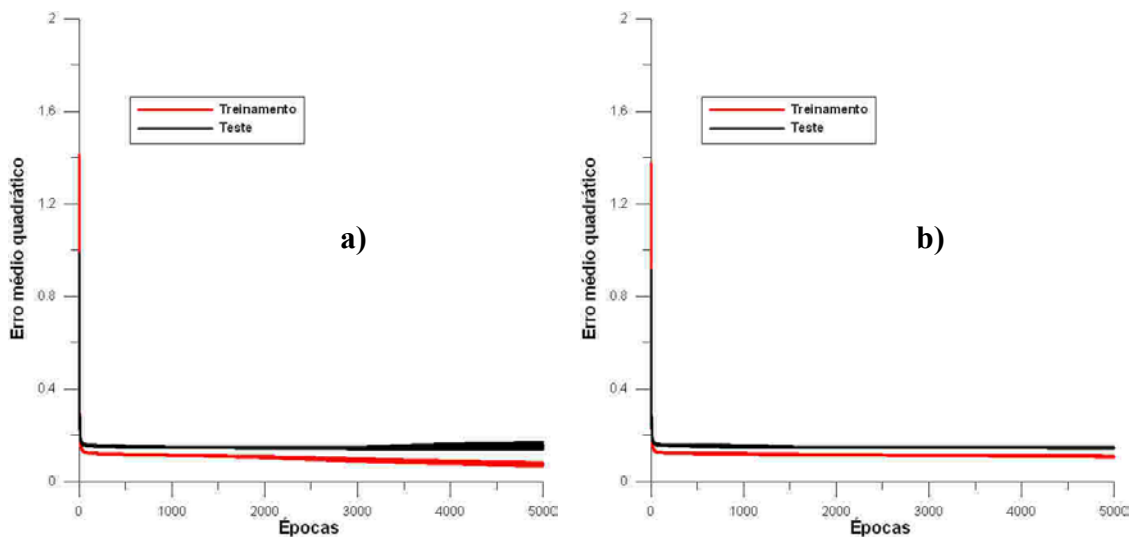
Nota: Dados replicados. Eventos no conjunto teste =29; Não Eventos = 667. Eventos no conjunto treinamento = 90; Não Eventos = 2698.



**Figura 5-39 Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Mostardas.**

### 5.1.2.3.1.b Dados Não Replicados

A Figura 5-40 apresenta a evolução dos erros do teste e treinamento para as duas configurações da rede. Observe a rápida convergência do erro. Em ambas topologias obtém-se um alto percentual de acerto, mas apenas 1 dos 29 casos de visibilidade restrita é identificado.



**Figura 5-40** Evolução dos erros de treinamento (linhas vermelhas) e do teste (linhas pretas) no experimento para previsão de visibilidade restrita realizado pela RN utilizando (a) Componentes principais e (b) *screening* – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Mostardas.

Para o conjunto de dados utilizados neste experimento as regressões com o uso de componentes principais e *screening* apresentaram coeficientes de determinação iguais a 4.16% e 4.07% , respectivamente. As regressões são consideradas significantes a um nível de 5%.

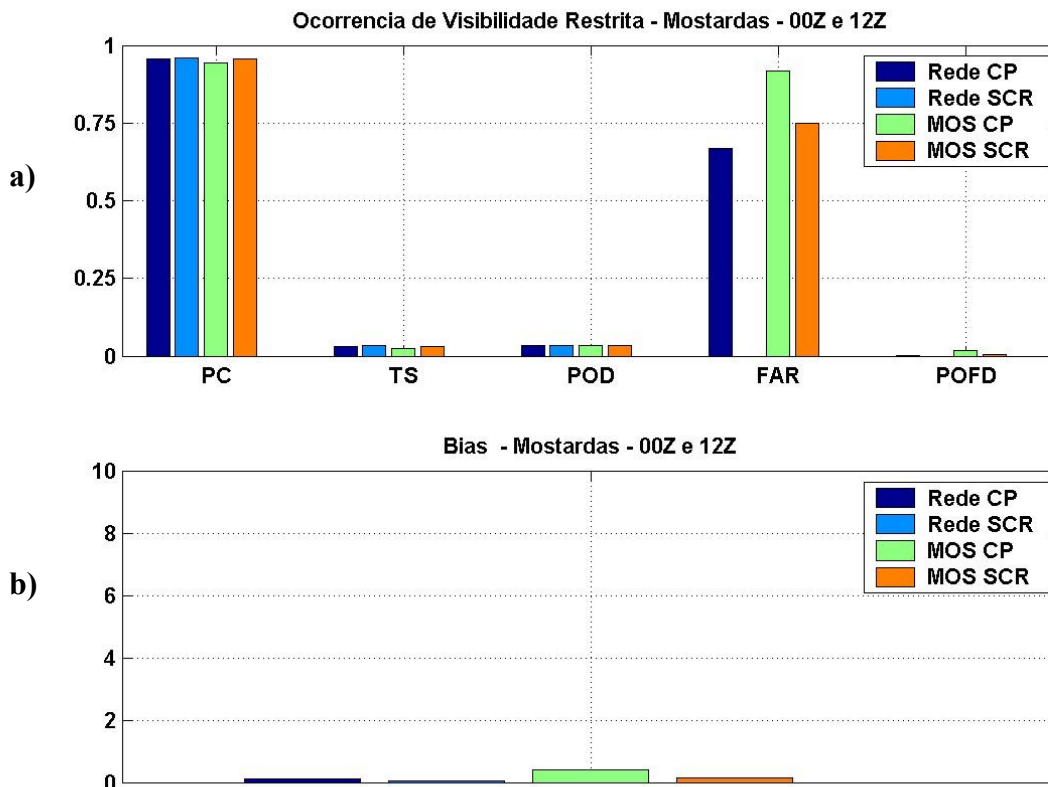
A Tabela 5-20 e a Figura 5-41 exibem os resultados dos experimentos. Observa-se a baixa detecção dos eventos para todos os métodos e que o alto percentual de acerto, em torno de 95%, refere-se aos casos de visibilidade boa.

Pode-se verificar a importância da replicação dos dados quando o interesse é identificar eventos raros.

**Tabela 5-20 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Mostardas.**

Estatísticas	Rodadas 00Z e 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
PC	0.9569	0.9598	0.9440	0.9555
TS	0.0323	0.0345	0.0250	0.0313
POD	0.0345	0.0345	0.0345	0.0345
FAR	0.6667	0	0.9167	0.7500
POFD	0.0030	0	0.0165	0.0045
B	0.1034	0.0345	0.4138	0.1379
Topologia (RN)	16-8-1 5000-31	5-5-1 5000-31	17 coef.	6 coef.
Alfa (RN)	0.1	0.1	-	-
Limiar (MOS)	-	-	01	0.1

Nota: Dados não replicados. Eventos no conjunto teste =29; Não Eventos = 667. Eventos no conjunto treinamento = 90; Não Eventos = 2698



**Figura 5-41 Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita. Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Mostardas.**

Segue abaixo a conclusão para os experimentos para detecção de visibilidade restrita para as três estações:

- Nenhum dos métodos de pós-processamento mostrou-se eficiente na detecção dos casos de visibilidade restrita observados em Salinópolis. A baixa frequência destes eventos na região impossibilitou o treinamento da rede neural e o ajustamento da regressão de forma eficaz;
- Observa-se melhor desempenho do pós-processamento e a relevância do mesmo quando se comparam os resultados obtidos em Santa Marta com os obtidos em Salinópolis. Os resultados indicam que a eficácia do pós-processamento também se relaciona com a localização e com as condições de clima local;
- Diferente do que ocorre em Mostardas, onde verificamos a relevância da replicação, com exceção de um experimento (rodada de 12Z), as previsões para Santa Marta apresentam bons resultados, mesmo quando os dados não são replicados. Possivelmente, deve-se ao fato de se localizar em uma região onde a ocorrência de casos de visibilidade restrita é maior, o que facilita o aprendizado da rede e o ajustamento das equações de regressão do MOS;
- A rede neural apresenta maior confiabilidade em suas previsões, pois embora não identifique todos os eventos, o número de alarmes falsos e o percentual total de acertos se mantêm em patamares razoáveis;
- Observaram-se melhores resultados quando se utiliza os dois horários de inicialização. Ocorre que este procedimento, além de aumentar o tamanho da amostra, aumenta o número de casos de visibilidade restrita, facilitando o aprendizado da rede e o ajustamento das regressões para os casos mais extremos;



- O MOS apresenta como característica uma alta detecção dos casos de visibilidade restrita, mas emite um alto número de alarmes falsos, de tal forma que suas previsões perdem a confiabilidade;
- De forma geral, a utilização dos dois horários de inicialização apresentam melhores previsões;
- Não foram observadas diferenças significativas quando se comparam os resultados em relação ao uso de *screening* ou componentes principais;
- O pós-processamento mostrou-se mais eficiente nas previsões feitas para Mostardas e Santa Marta, ambas sujeitas ao clima temperado do Sul do país. Por outro lado, em Salinópolis não foram obtidos resultados satisfatórios;
- Podemos considerar a razoável eficácia do uso de redes neurais na detecção dos eventos de visibilidade restrita, embora não tenhamos condições de comparar os resultados do modelo já que esta previsão é feita subjetivamente pelo previsor a partir da consulta a dados e informações meteorológicas de diferentes fontes.

# **Capítulo 6**

## **Conclusão**

Este trabalho apresentou o resultado do pós-processamento das previsões do modelo numérico HRM por meio de redes neurais e MOS.

A motivação fundamenta-se na necessidade de desenvolver métodos objetivos para apoiar previsão local de vento forte e visibilidade restrita, eventos cuja ocorrência é precursora de acidentes graves e até fatais. No que tange ao Serviço Meteorológico Marinho (SMM), responsável pela emissão de boletins de previsão meteorológica para área marítima, a ocorrência destes eventos justificam a emissão de avisos de mau tempo, pois são condições de tempo que colocam em risco a segurança da navegação. Contudo, em vista das limitações dos modelos numéricos, as previsões destes eventos contam em grande parte com a experiência e com a subjetividade do previsor do tempo.

Assim, o objetivo desta dissertação é testar e avaliar metodologias de previsão objetiva que possam subsidiar o previsor do tempo na tomada de decisão para emissão de avisos de mau tempo relacionados com os eventos estudados.

Foram feitos vários experimentos para as três estações estudadas considerando diferentes configurações das amostras. Para cada estação, foram desenvolvidas diferentes regressões e redes neurais. Nas regressões as previsões foram convertidas em resultados não probabilísticos a fim de permitir a comparação com os resultados categóricos da rede neural.

Os experimentos do pós-processamento para previsão de ocorrência de vento forte foram comparados com os resultados do modelo numérico. No entanto, nos experimentos para a previsão de visibilidade, não foi possível a comparação com outro

método de previsão, por dois motivos. Primeiro, por se tratar de uma variável que não é prevista diretamente pelo modelo numérico e, segundo, a forma de registro da visibilidade nas estações é diferente da forma em que o evento é previsto no boletim emitido pelo SMM, onde as previsões são em intervalos de ocorrência do evento.

Medidas de verificação foram utilizadas a fim de avaliar o desempenho do MOS e das redes neurais, bem como comparar os resultados entre ambos. O uso destas medidas torna-se relevante, pois por meio de medidas simples podemos obter informações de aspectos diferentes dos métodos de pós-processamento tais como: acurácia, emissão de alarmes falsos e capacidade de discriminação dos eventos e não eventos.

A replicação dos dados, que consiste em aumentar artificialmente a frequência dos eventos de interesse na amostra, mostrou-se um importante recurso para aumentar a detecção dos eventos de interesse por parte dos métodos de pós-processamento.

Nos experimentos para previsão de ocorrência de vento forte, podemos destacar o seguinte:

- na comparação com os resultados do modelo numérico, os métodos de pós-processamento tiveram melhor desempenho na detecção dos eventos;
- o pós-processamento com uso dos dois horários de inicialização apresenta melhor desempenho do que o pós-processamento especializado por horário;
- as previsões realizadas nas duas estações localizadas em região de clima temperado, Santa Marta e Mostardas, apresentaram melhores desempenhos quando se compara com Salinópolis, que está sujeita ao clima tropical;

- os resultados com o uso de componentes principais ou *screening* não mostraram diferenças significantes no resultado do pós-processamento, de forma que em trabalhos futuros pode ser escolhido o uso de qualquer um dos métodos;
- as previsões da rede neural apresentaram melhor acurácia e confiabilidade na previsão dos casos de vento forte, quando se compara com as previsões do MOS e com os resultados do modelo numérico.

Nos experimentos para previsão de ocorrência de visibilidade restrita, podemos destacar o seguinte:

- as poucas ocorrências de visibilidade restrita em Salinópolis, impossibilitaram o adequado treinamento e ajuste das redes neurais e do MOS, respectivamente, de tal forma que nenhum evento foi previsto no pós-processamento;
- o MOS apresenta comportamento semelhante ao encontrado quando da previsão de vento forte, ou seja, nos experimentos em que detecta a maioria dos eventos de interesse, a emissão de alarmes falsos é tão alta que torna suas previsões pouco confiáveis;
- não houve diferença significativa no desempenho dos pós-processamento quando se usa *screening* ou componentes principais.

Considerando a aplicação dos métodos de pós-processamento de uma forma geral, isto é, tanto na previsão de visibilidade quanto na previsão de vento forte, podemos concluir que:

- o pós-processamento com uso das regressões do MOS apresenta pouca confiabilidade em suas previsões, de forma que a emissão excessiva de alarmes falsos as tornam poucos confiáveis e com baixa acurácia. As previsões da rede neural, embora não consigam acertar todos os eventos de interesse, apresentam melhor desempenho, pois conseguem controlar e manter o número de emissões de alarmes falsos em patamares que não coloca em risco sua confiabilidade;

- a rede neural apresenta melhor aproveitamento quando os dados são replicados, de forma que o uso deste recurso proporciona aumento na detecção dos eventos de interesse, mas sem a emissão excessiva de alarmes e falsos. No MOS, embora o uso deste recurso implique na detecção de, praticamente, todos os eventos de interesse, ocasiona a queda no desempenho quando se avalia a previsão conjunta dos eventos e não eventos, além de tornar suas previsões pouco confiáveis em função do número de alarmes falsos emitidos;

- verifica-se um melhor desempenho nas previsões para a estação Santa Marta, indicando que a maior frequência dos eventos de interesse permitem o melhor aproveitamento dos recursos do pós-processamento. Este fato também é confirmado quando se analisam os resultados dos pós-processamento em Salinópolis, cuja previsão do vento forte apresentou pior desempenho em relação às outras estações, em virtude de baixa ocorrência dos eventos na amostra;

- a partir dos resultados em Salinópolis, onde as previsões dos eventos não se mostraram satisfatórias, podemos inferir que, apesar da baixa variabilidade climática da região, os métodos de pós-processamento apresentaram maior dificuldade de prever os casos de interesse, pois a rede neural e o MOS

necessitam de um maior número de eventos de interesse na amostra para que possam ser treinados e ajustados de forma adequada;

- a tentativa de identificar eventos raros implicará normalmente em queda na proporção total de acertos e em aumento na emissão de alarmes falsos. Para lidar com este dilema, caberá ao usuário definir um critério que caracterize um limite na emissão de alarmes que não represente uma queda excessiva na confiabilidade e no poder discriminatório das previsões;
- observaram-se melhores resultados quando se utilizam os dois horários de inicialização. Ocorre que este procedimento, além de aumentar o tamanho da amostra, aumenta o número de casos de interesse, facilitando o aprendizado da rede e o ajustamento das regressões para os casos mais extremos.

Em resumo, a rede neural e o MOS melhoram a previsão local dos eventos estudados. Contudo, a rede neural, além de detectar os eventos de interesse, consegue manter a confiabilidade de suas previsões, o que não ocorre com o MOS, pois a emissão excessiva de alarmes falsos torna suas previsões poucos confiáveis.

Desta forma, verifica-se que o emprego do pós-processamento melhora a previsão local dos eventos de interesse e que, em particular, as redes neurais apresentam melhor desempenho em termos de confiabilidade, detecção e discriminação destes eventos.

Assim, as perguntas formuladas no capítulo 1 podem ser respondidas, de acordo com os limites estabelecidos nesta dissertação, como segue:

- 1- Utilizar redes neurais para o pós-processamento das previsões do modelo numérico melhora as previsões locais?

Resposta: Sim.

2- A utilização de métodos estatísticos (MOS) melhora a previsão local?

Resposta: Não, pois apresenta baixa confiabilidade e baixa discriminação na previsão dos eventos.

3- Qual dos dois métodos é melhor?

Resposta: A rede neural, apresentou maior confiabilidade e discriminação nos resultados de tal forma que, para as configurações dos experimentos aplicadas nesta dissertação, apresentou melhor resultado quando comparada com o MOS.

É importante destacar que as amostras utilizadas nos experimentos podem ser consideradas pequenas. Isto ocorre em função da forma de registro das observações nas estações, onde os dados são registrados apenas em quatro horários sinóticos (00Z, 06Z, 12Z e 18Z), assim como a forma de arquivamento dos dados do modelo numérico, que estavam disponíveis em intervalos de 12 horas. Estes foram fatores que limitaram os tamanhos das amostras utilizadas no estudo e mostrou a necessidade de se aprimorar a forma de arquivamento destes dados, a fim de possibilitar a criação de amostras mais representativas para serem utilizadas em futuros estudos.

Seguem abaixo algumas recomendações a partir dos estudos desta dissertação:

- os modelos numéricos são considerados produtos perecíveis em virtude das modificações e alterações nas configurações a que estão sujeitos. Desta forma, sugere-se que haja planejamento e organização na forma de armazenamento e coleta dos dados observados a serem usados no pós-processamento;

- sugere-se que seja organizado um banco de dados de forma que as observações de visibilidade e as suas previsões sejam armazenados de forma homogênea a fim de permitir a comparação entre elas;
- a fim de permitir uma maior representatividade dos eventos na amostra, desenvolver redes neurais a partir dos dados observados em várias estações localizadas em regiões próximas e com situações climáticas semelhantes;
- desenvolver redes neurais especializadas para as diferentes estações do ano, por exemplo, primavera-verão e outono-inverno.



# **Capítulo 7**

## **Referências Bibliográficas**

BROOMHEAD, D.S., LOWE, D., “Multivariable functional interpolation and adaptive networks” **In: Complex Systems**, v. 2, pp. 321-355,1988.

CARTER, G. M., DALLAVALLE, J. P., GLAHN, H.R, “Statistical Forecasts Based on the National Meteorological Center’s Numerical Weather Prediction System”, **Weather and Forecasting**, 4: 401-412, 1989.

DHN- Diretoria de Hidrografia e Navegação, **Manual do Observador Meteorológico**, Rio de Janeiro, A Diretoria, pp 150, 1992.

GLAHN, H. R.; LOWRY, D. A., “The use of Model Output Statistics (MOS) in Objective Weather Forecasting.” **Journal of Applied Meteorology**, v.11, pp. 1203-1211, Dec. 1972.

GRADS/COLA, disponível em <<http://www.iges.org/grads/>> acessado em 10/09/07.

HALL, T., H, BROOKS, E.B., DOSWELL, C.A., “Precipitation Forecasting Using a Neural Networks”, **Weather and Forecasting**, v. 14, pp. 338-345, Jun. 1999.

HAYKIN, S., **Neural Networks: A comprehensive Foundation**, Macmillan College Publishing Company, Inc., 1<sup>st</sup> ed., New York,1994.

HEBB, D. O., **The Organization of Behavior: A Neuropsychological Theory.**, pp. New York, Wiley, 1949.

HOPFIELD, J.J., “Neural Networks and Physical Systems with Emergent Collective Computational Abilities”, **Proceedings of the National Academy of Sciences of the U.S.A.**, v. 79, pp. 2554-2558, 1982.

HSIEH, W. W., TANG, B.. “Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography”, **Bulletin of the American Meteorological Society**, Vancouver, n.79 , p.1855-19870, May 1998.

IMSL Library, **Fortran Subroutines for Mathematical Applications (on line)**, vol. 1, Houston, Visual Numerics Inc.,1987.

\_\_\_\_\_, **Fortran Subroutines for Statistical Applications (on line)**, vol. 1, Houston, Visual Numerics Inc.,1987.

JOLLIFFE, I.T., **Principal Components Analysis**, 2<sup>nd</sup> ed. Springer, 2002. 487pp

KALNAY, E., **Atmospheric Modeling, Data Assimilation and Predictability**, 1<sup>st</sup> ed., Cambridge, Cambridge, pp. 341, 2003.

KLEIN, W. H., LEWIS, B. M., ENGER, I., “Objective Prediction of Five-Day Mean Temperatures During Winter”, **Journal of Meteorology**, v.16, pp. 672-682. Dec. 1959.

KRETZSCHMAR, R., ECKERT, P., CATTANI, D., *et al.*, “Neural Network for Local Wind Prediction“, **Journal of Applied Meteorology**, vol. 43, pp. 727-738, May 2004.

KULIGOWSKI, R.J., BARROS, A.P., “Experiments in Short-Term Precipitation Forecasting using Artificial Neural Networks”, **Monthly Weather Review**, vol. 126, pp. 470-482, Feb, 1998.

KUTNER, MICHAEL H., NACHTSHEIM, C.J., NETER, J. *et al.*, **Applied Linear Statistical Models**. 5<sup>th</sup> ed. New York: Mcgraw-hill Irwin, 2005. 1396 p.

LORENZ, E.N., “Empirical Orthogonal Functions and Statistical Weather Prediction.” **Science Report 1, Statistical Forecasting Project, Department of Meteorology, MIT** (NTIS AD 110268),49 pp. 1956.

LECUN, Y., “Une Procedure D’apprentissage Pour Reseau A Seuil Assymetrique”, **Cognitva**, vol. 85, pp. 599-604, 1985.

MAJEWSKI, D., HRM – User’s Guide, 2006.

MB - MARINHA DO BRASIL - Diretoria de Hidrografia e Navegação, disponível em <[http://www.mar.mil.br/dhn/chm/oce\\_dados\\_coleta.html](http://www.mar.mil.br/dhn/chm/oce_dados_coleta.html)>, Acessado em 10/09/07.

MARZBAN, C. “Neural Networks for Postprocessing Model Output: ARPS”, **Monthly Weather Review**, Seattle, vol. 131, p.1103-1111, Jun. 2003.

MARZBAN, C., STUMPF, G.J., “A Neural Networks for Damaging Wind Prediction”, **Weather and Forecasting**, Norman, vol. 13, p.151-163, Mar. 1998.

MARZBAN, C., LEYTON, S., COLMAN, B.. “Ceiling and Visibility via Neural Networks”, **Weather and Forecasting**, Seattle, v. 22, p.466-479, Jun. 2007.

MINSKY, M.L., PAPPERT, S.A., **Perceptrons**, Cambridge, MA: MIT Press.

MURPHY, A. H., KATZ, R. W., **Probability, Statistics, and Decision Making in the Atmospheric Sciences**. 1st ed. Boulder: Westview Press, Inc, 545p., 1985.

NETER, J., WASSERMAN, W., **Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs**. 1<sup>st</sup> ed. Homewood: Richard D. Irwin, Inc., 1974. 842 p.

PARKER, D.B., “Learning-logic: Casting the Cortex of the Human Brain in Silicon”, **Technical Report TR-47, Center for Computational Research in Economics and Management Science**, MIT, Cambridge, MA, 1985.

PREISENDORFER, R.W., **Principal Component Analysis in Meteorology and Oceanography**, New York, Elsevier, 1988.

RITTER, B, GELEYN, J.F., “A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations.”, **Monthly Weather Review**, 109, p. 758-766, 1992

ROSENBLATT, F., 1962, **Principles of Neurodynamics**, Washington, DC, Spartan Books.

RUMELHART , D.E., HINTON, G.E., WILLIAMS, R.J., “Learning Representations by Backpropagation Errors”, *Nature (London)*, vol. 323, pp. 533-539, 1986.

TAPP, R.G., WOODCOCK, F., MILLS, G., “The Application of Model Output Statistics to Precipitation Prediction in Australia”, **Monthly Weather Review**, 114: 50-61, 1986.

TIEDTKE, M., “A Comprehensive Mass Flux Scheme for Cumulus Parametrization in Large Scale Models”, **Monthly Weather Review**, 117, p. 1779-1800, 1989.

TRIOLA, M. F., **Introdução à Estatística**, 7 ed., Rio de Janeiro, Livros Técnicos e Científicos Editora S.A., 1998.

VISLOCKY, R. L., FRITSCHI, M., “An Automated, Observations-Based System for Short-Term Prediction of Ceiling and Visibility”, **Weather and Forecasting**, v.12, pp. 31-43, Mar. 1997.

von NEUMANN, J., **The computer and Brain**, New Haven, CT: Yale University Press, 1958.

WIDROW, B., “Generalization and Information Storage in Networks of Adaline Neurons”, in **Self Organization Systems**, pp. 435-461, 1962.

WILKS, D.S., **Statistical Methods in the Atmospheric Sciences**. 2nd ed. San Diego: Academic Press, 2006. 627 p.

WILSON, L. J., ACOWAR, N. “Statistical Weather Element Forecasting in the Canadian Weather Service”, **Proc. WMO Symp. on Probabilistic and Statistical Methods in Weather Forecasting**, Nice, pp. 40-406, 1980.

WOODCOCK, F., “Australian Experimental Model Output Statistics Forecasts of Daily Maximum and Minimum Temperature”, **Monthly Weather Review**, v. 112: pp. 2112-2121, May 1984.

WMO, World Meteorological Organization, Manual on Marine Meteorological Services, vol. I, pp 85, 1990.

# Apêndices

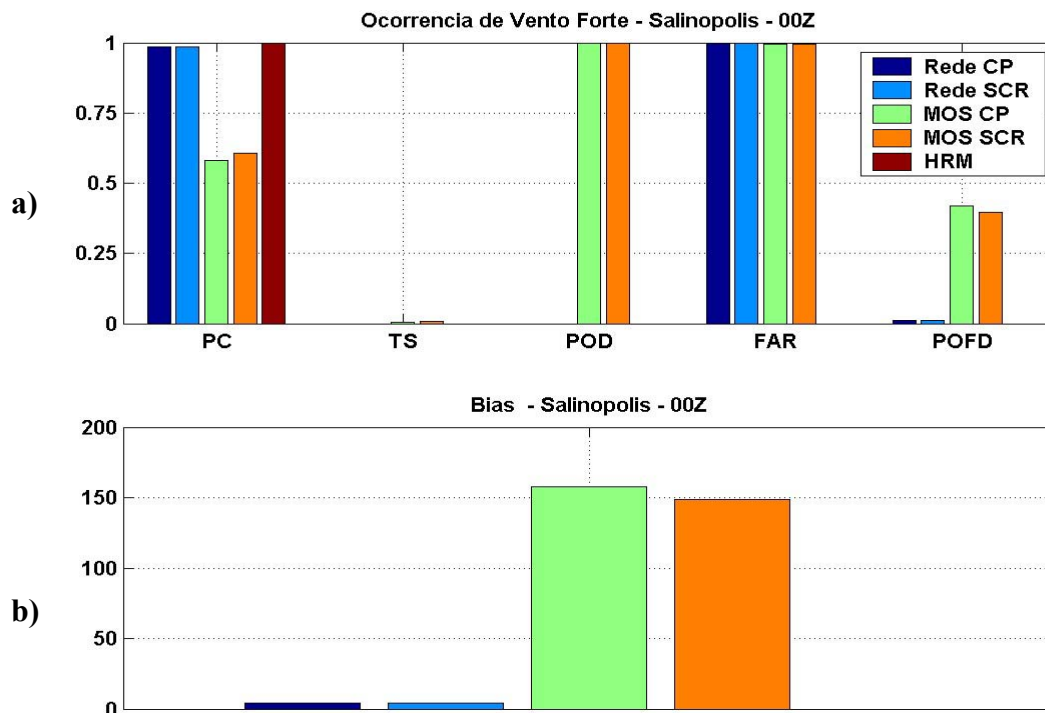
## **Apêndice A** Sumário estatístico do pós-processamento (MOS e RN) para ocorrência de vento forte e visibilidade restrita

### *Sumário Estatístico para previsão de ocorrência de vento forte -Salinópolis*

Tabela A-1 Sumário estatístico do pós-processamento (MOS e RN) para ocorrência de vento forte Rodadas de 00Z, com os dados replicados, referentes à Salinópolis.

Estatísticas	Rodadas 00Z				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9867	0.9867	0.5813	0.6053	0.9973
TS	0	0	0.0063	0.0067	0
POD	0	0	1	1	0
FAR	1	1	0.9937	0.9933	-
POFD	0.0107	0.0107	0.4198	0.3957	0
B	4	4	158	149	0
Topologia (RN) Coef (MOS)	16-8-1 500-31	20-10-1 500-31	17 coef	21 coef	-
Alfa (RN)	0.1		-	-	-
Limiar (MOS)	-	-	0.1	0.1	-

Nota: Dados replicados. Eventos no conjunto teste = 1; Não Eventos = 374. Eventos no conjunto treinamento = 2; Não Eventos = 1500.



**Figura A-1** Sumário estatístico do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 00Z, com os dados replicados, referentes à Salinópolis.

**Tabela A- 2** Resultado estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte – Rodadas de 00Z, com os dados não replicados , referentes à Salinópolis

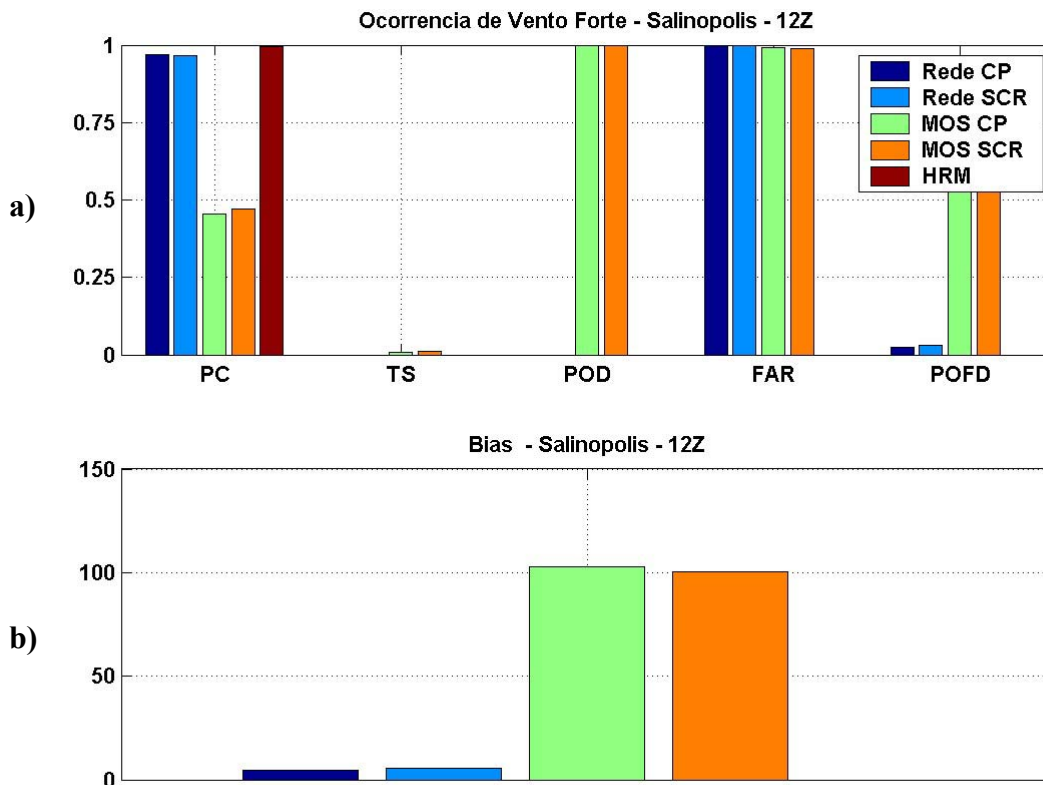
Estatísticas	Rodadas 00Z				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
<b>PC</b>	0.9973	0.9973	0.9973	0.9973	0.9973
<b>TS</b>	0	0	0	0	0
<b>POD</b>	0	0	0	0	0
<b>FAR</b>	-	-	-	-	-
<b>POFD</b>	0	0	0	0	0
<b>B</b>	0	0	0	0	0
<b>Topologia (RN)</b>	16-8-1	2-2-1			
<b>Coef (MOS)</b>	300-31	300-31	17 coef	3 coef	-
<b>Alfa (RN)</b>	0.1		-	-	-
<b>Limiar (MOS)</b>	-	-	0.1	0.1	-

Nota: Dados replicados. Eventos no conjunto teste = 1; Não Eventos = 374. Eventos no conjunto treinamento = 2; Não Eventos = 1500.

**Tabela A- 3 Sumário estatístico do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 12Z, com os dados replicados, referentes à Salinópolis.**

Estatísticas	Rodadas 12Z				
	Média = m/s Desvio Padrão= m/s				
	RN CP	RN SCRN	MOS CP	MOS SCRN	HRM
PC	0.9707	0.9653	0.4560	0.4693	0.9947
TS	0	0	0.0097	0.0100	0
POD	0	0	1	1	0
FAR	1	1	0.9903	0.9900	-
POFD	0.0241	0.0295	0.5500	0.5300	0
B	4.5	5.5	103	100.5	0
Topologia (RN) Coef (MOS)	16-8-1 300-31	19-9-1 300-31	17 coef	20 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.1	0.1	-

Nota: Dados replicados. Eventos no conjunto teste = 2; Não Eventos = 373  
Eventos no conjunto treinamento = 4; Não Eventos = 1500.



**Figura A- 2 Estatísticas do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 12Z, com os dados replicados, referentes à Salinópolis.**



**Tabela A- 4 Sumário estatístico do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 12Z, com os dados não replicados, referentes à Salinópolis.**

Estatísticas	Rodadas 12Z				
	RN CP	RN SCR N	MOS CP	MOS SCR N	HRM
PC	0.9947	0.9947	0.9947	0.9947	0.9947
TS	0	0	0	0	0
POD	0	0	0	0	0
FAR	-	-	-	-	-
POFD	0	0	0	0	0
B	0	0	0	0	0
Topologia (RN) Coef (MOS)	16-8-1 300-31	2-2-1 300-31	17 coef	3 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.1	0.1	-

Nota: Dados replicados. Eventos no conjunto teste = 2; Não Eventos = 373  
Eventos no conjunto treinamento = 4; Não Eventos = 1500.

### *Sumário Estatístico para previsão de ocorrência de vento forte -Mostardas*

**Tabela A- 5 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodadas de 00Z, com dados replicado, referentes à estação Mostardas.**

Estatísticas	Rodadas 00Z				
	Média =6.16 m/s Desvio Padrão=2.92 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9595	0.9566	0.8237	0.8150	0.9855
TS	0.0667	0.0625	0.0615	0.0303	0
POD	0.2000	0.2000	0.8000	0.4000	0
FAR	0.9091	0.9167	0.9375	0.9683	-
POFD	0.0293	0.0323	0.1760	0.1789	0
B	2.2	2.4	12.8	12.6	0
Topologia (RN) Coef (MOS)	16-10-1 5000-31	20-10-1 5000-31	17 coef	21 coef	-
Alfa (RN)	0.025	0.1	-	-	-
Limiar (MOS)	-	-	0.5	0.1	-

Nota: Dados não replicados. Eventos no conjunto teste =5; Não Eventos = 341  
Eventos no conjunto treinamento = 15; Não Eventos = 1373.

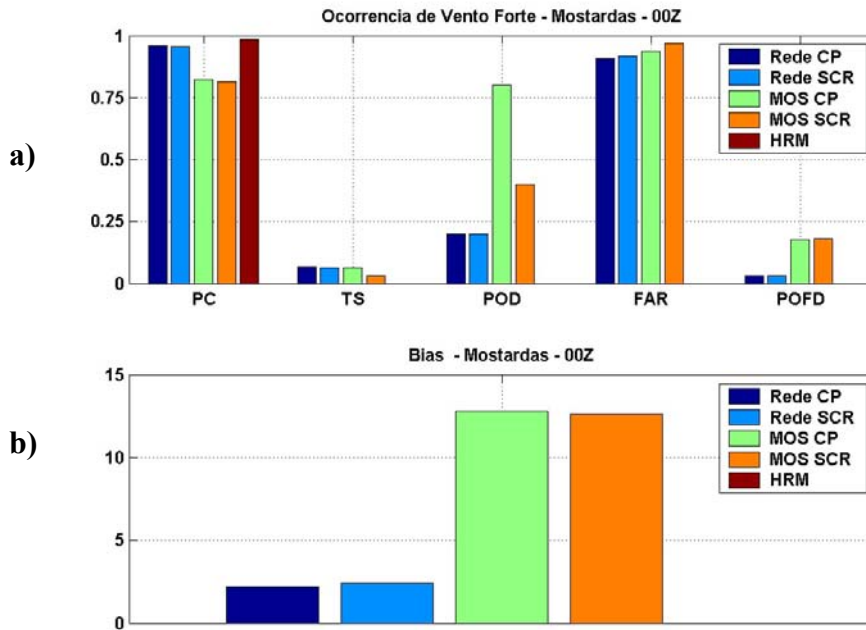


Figura A- 3 Estatísticas do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 00Z, com dados replicado, referentes à estação Mostardas.

Tabela A- 6 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodadas de 00Z, com dados não replicados, referentes à estação Mostardas.

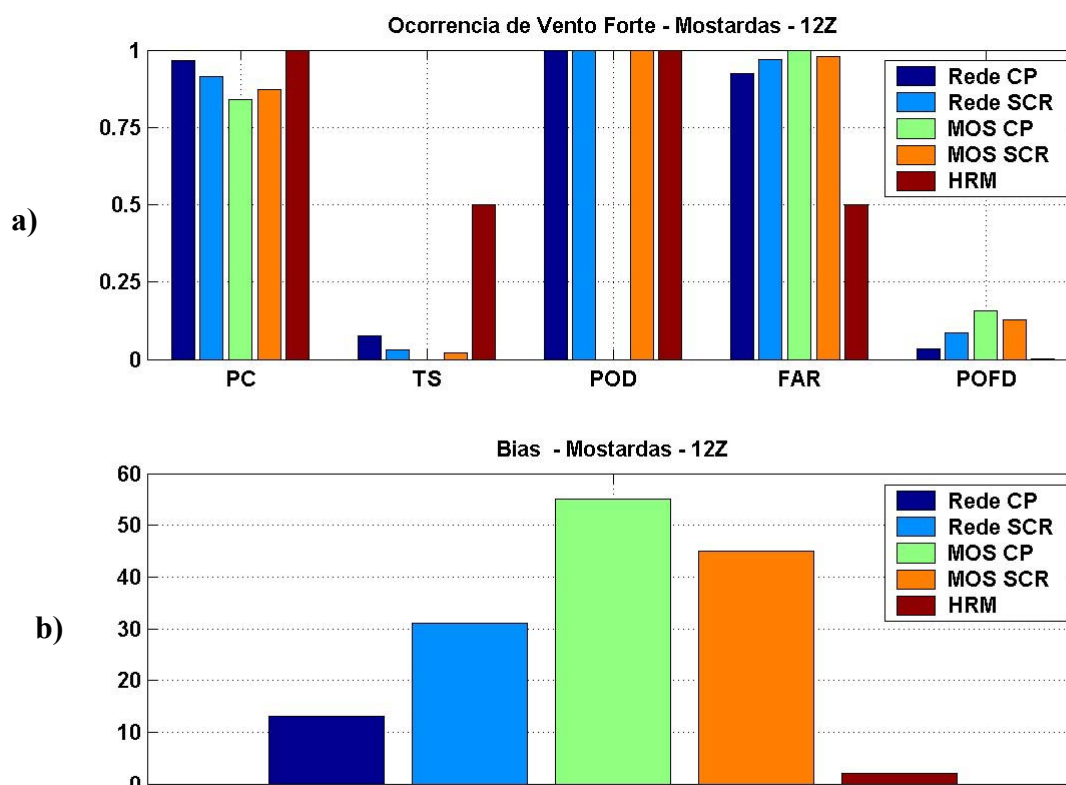
Estatísticas	Rodadas 00Z				
	Média =6.16 m/s Desvio Padrão=2.92 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
<b>PC</b>	0.9855	0.9855	0.9855	0.9855	0.9855
<b>TS</b>	0	0	0	0	0
<b>POD</b>	0	0	0	0	0
<b>FAR</b>	-	-	-	-	-
<b>POFD</b>	0	0	0	0	0
<b>B</b>	0	0	0	0	0
<b>Topologia (RN)</b>	16-8-1 1000-31	4-4-1 1000-31	17 coef	5 coef	-
<b>Alfa (RN)</b>	0.1	0.1	-	-	-
<b>Limiar (MOS)</b>	-	-	-	-	-

Nota: Dados não replicados. Eventos no conjunto teste =5; Não Eventos = 341 Eventos no conjunto treinamento = 15; Não Eventos = 1373.

**Tabela A- 7 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodadas de 12Z, com dados replicados, referentes à estação Mostardas.**

Estatísticas	Rodadas 12Z				
	Média =6.06 m/s Desvio Padrão=2.90 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
PC	0.9655	0.9138	0.8391	0.8736	0.9971
TS	0.0769	0.0323	0	0.0222	0.5000
POD	1	1	0	1	1
FAR	0.9231	0.9677	1	0.9778	0.5000
POFD	0.0346	0.0865	0.1585	0.1268	0.0029
B	13	31	55	45	2
Topologia (RN)	16-8-1 1000-31	20-10-1 1000-31	17 coef	5 coef	-
Alfa (RN)	0.1	0.1	-	-	-
Limiar (MOS)	-	-	0.5	0.5	-

Nota: Dados replicados. Eventos no conjunto teste =1; Não Eventos = 347 Eventos no conjunto treinamento = 15; Não Eventos = 1381



**Figura A- 4 Estatísticas do pós-processamento (MOS e RN) para ocorrência de vento forte – Rodadas de 12Z, com dados replicados, referentes à estação Mostardas.**

**Tabela A- 8 Sumário estatístico do pós-processamento (MOS e RN) e do modelo HRM para ocorrência de vento forte - Rodadas de 12Z, com dados não replicados, referentes à estação Mostardas.**

Estatísticas	Rodadas 12Z				
	Média =6.06 m/s Desvio Padrão=2.90 m/s				
	RN CP	RN SCR	MOS CP	MOS SCR	HRM
<b>PC</b>	0.9971	0.9971	0.9943	0.9971	0.9971
<b>TS</b>	0	0	0	0	0.5000
<b>POD</b>	0	0	0	0	1
<b>FAR</b>	-	-	1	-	0.5000
<b>POFD</b>	0	0	0	0	0.0029
<b>B</b>	0	0	1	0	2
<b>Topologia (RN)</b>	16-8-1 2000-31	6-6-1 2000-31	17 coef	7 coef	-
<b>Alfa (RN)</b>	0.1	0.1	-	-	-
<b>Limiar (MOS)</b>	-	-	0.1	0.1	-

Nota: Dados não replicados. Eventos no conjunto teste =1; Não Eventos = 347 Eventos no conjunto treinamento = 15; Não Eventos = 1381.

***Sumário Estatístico para previsão de ocorrência de  
visibilidade Restrita -Salinópolis***

**Tabela A- 9 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z, com dados replicados, referentes à estação Salinópolis.**

Estatísticas	Rodadas 00Z			
	RN CP	RN SCR	MOS CP	MOS SCR
PC	0.9894	0.9894	0.7487	0.8519
TS	0	0	0	0
POD	0	0	0	0
FAR	1	1	1	1
POFD	0.0053	0.0053	0.2473	0.1436
B	1	1	46.5	27
Topologia (RN)	16-8-1 5000-31	19-9-1 5000-31	17 coef.	20 coef.
Alfa (RN)	0.1	0.1	-	-
Limiar (MOS)	-	-	0.5	0.5

Nota: Dados replicados. Eventos no conjunto teste = 2; Não Eventos = 376. Eventos no conjunto treinamento = 4; Não Eventos = 1508.

**Tabela A- 10 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodadas de 00Z, com dados não replicados, referentes à estação Salinópolis.**

Estatísticas	Rodadas 00Z			
	RN CP	RN SCR	MOS CP	MOS SCR
PC	0.9947	0.9947	0.9947	0.9947
TS	0	0	0	0
POD	0	0	0	0
FAR	-	-	-	-
POFD	0	0	0	0
B	0	0	0	0
Topologia (RN)	16-8-1 500-31	1-1-1 500-31	17 coef.	2 coef.
Alfa (RN)	0.1	0.1	-	-
Limiar (MOS)	-	-	0.1	0.1

Nota: Dados replicados. Eventos no conjunto teste = 2; Não Eventos = 376. Eventos no conjunto treinamento = 4; Não Eventos = 1508.

**Tabela A- 11 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 12Z, com dados replicados, referentes à estação Salinópolis.**

Estatísticas	Rodadas 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9920	0.9920	0.3581	0.7454
<b>TS</b>	0	0	0.0122	0.0204
<b>POD</b>	0	0	1	0.6667
<b>FAR</b>	-	-	0.9878	0.9794
<b>POFD</b>	0	0	0.6471	0.2540
<b>B</b>	0	0	81.6667	32.3333
<b>Topologia (RN)</b>	16-8-1 5000-31	22-11-1 5000-31	17 coef.	23 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.5	0.4

Nota: Dados replicados. Eventos no conjunto teste =3, Não Eventos = 374. Eventos no conjunto treinamento = 5; Não Eventos = 1506.

**Tabela A- 12 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 12Z, com dados não replicados, referentes à estação Salinópolis.**

Estatísticas	Rodadas 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9920	-	0.9920	-
<b>TS</b>	0	-	0	-
<b>POD</b>	0	-	0	-
<b>FAR</b>	-	-	-	-
<b>POFD</b>	0	-	0	-
<b>B</b>	0	-	0	-
<b>Topologia (RN)</b>	16-8-1 500-31	-	17 coef.	-
<b>Alfa (RN)</b>	0.1	-	-	-
<b>Limiar (MOS)</b>	-	-	0.5	-

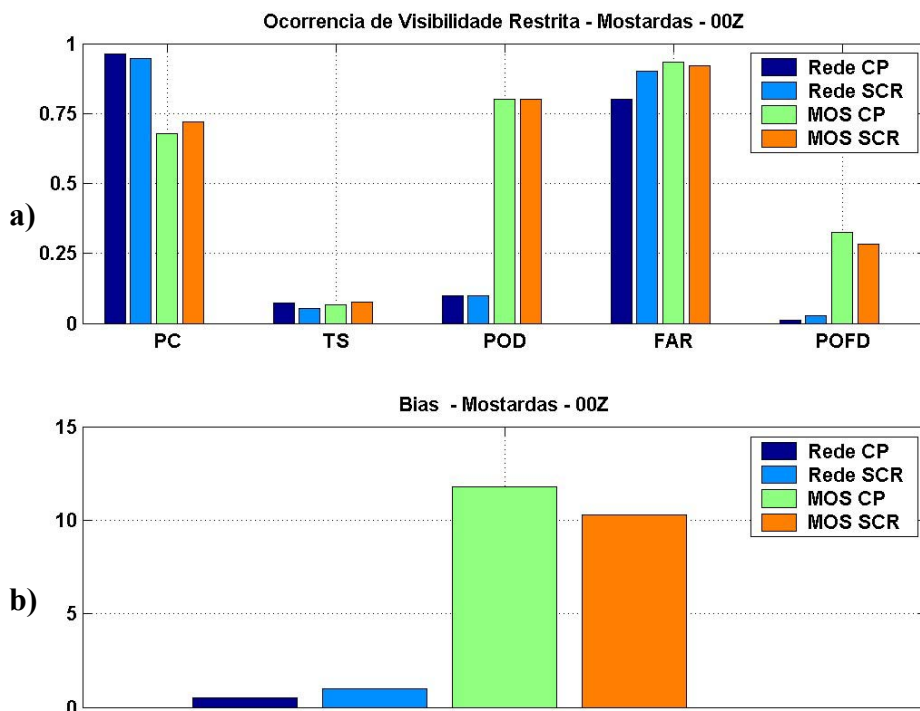
Nota: Dados não replicados. Eventos no conjunto teste =3 Não Eventos = 374. Eventos no conjunto treinamento = 5; Não Eventos = 1506.

## *Sumário Estatístico para previsão de ocorrência de visibilidade Restrita – Mostardas*

**Tabela A- 13** Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 00Z, com os dados replicados, referentes à estação Mostardas.

Estatísticas	Rodada 00Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9625	0.9481	0.6772	0.7205
<b>TS</b>	0.0714	0.0526	0.0667	0.0762
<b>POD</b>	0.1000	0.1000	0.8000	0.8000
<b>FAR</b>	0.8000	0.9000	0.9322	0.9223
<b>POFD</b>	0.0119	0.0267	0.3264	0.2819
<b>B</b>	0.5000	1	11.8	10.3
<b>Topologia (RN)</b>	16-8-1 5000-31	16-8-1 5000-31	17 coef.	17 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.5	0.5

Nota: Dados não replicados. Eventos no conjunto teste =10; Não Eventos = 337. Eventos no conjunto treinamento = 41; Não Eventos = 1351



**Figura A- 5** Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodada de 00Z, com os dados replicados, referentes à estação Mostardas.

**Tabela A- 14 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 00Z, com os dados não replicados, referentes à estação Mostardas.**

Estatísticas	Rodada 00Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9712	0.9712	0.9712	0.9712
<b>TS</b>	0	0	0	0
<b>POD</b>	0	0	0	0
<b>FAR</b>	-	-	-	-
<b>POFD</b>	0	0	0	0
<b>B</b>	0	0	0	0
<b>Topologia (RN)</b>	16-8-1 500-31	4-4-1 500-31	17 coef.	5 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.5	0.5

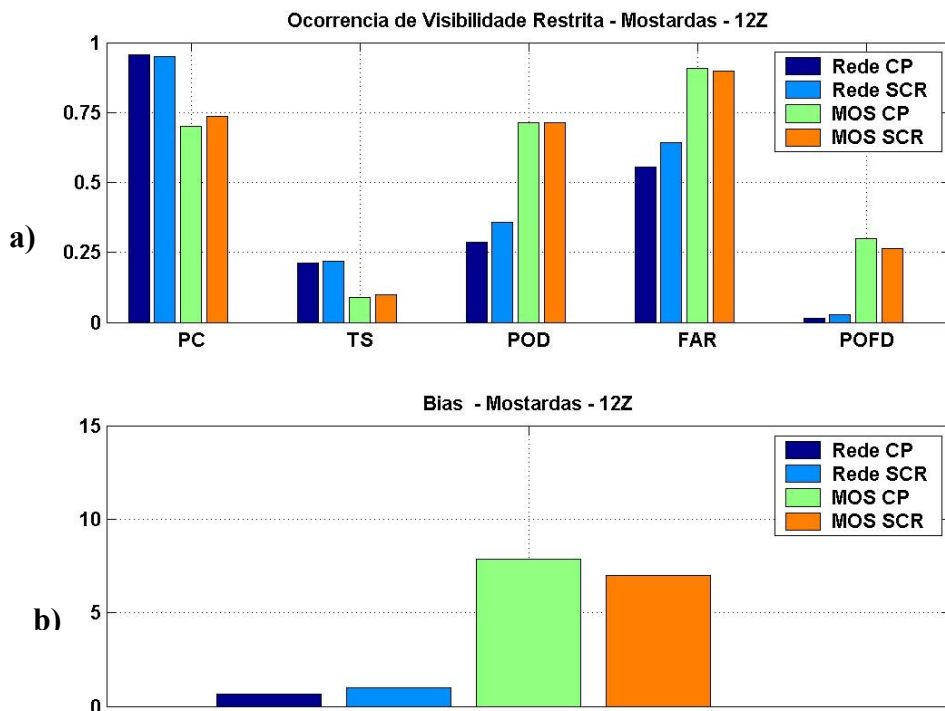
Nota: Dados não replicados. Eventos no conjunto teste =10; Não Eventos = 337. Eventos no conjunto treinamento = 41; Não Eventos = 1351.

**Tabela A- 15 Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 12Z, com os dados replicados, referentes à estação Mostardas.**

Estatísticas	Rodada 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9570	0.9484	0.7020	0.7364
<b>TS</b>	0.2105	0.2174	0.0877	0.0980
<b>POD</b>	0.2857	0.3571	0.7143	0.7143
<b>FAR</b>	0.5556	0.6429	0.9091	0.8980
<b>POFD</b>	0.0149	0.0269	0.2985	0.2627
<b>B</b>	0.6429	1	7.8571	7
<b>Topologia (RN)</b>	16-8-1 5000-31	17-9-1 5000-31	17 coef	18 coef
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.5	0.5

Nota: Dados não replicados. Eventos no conjunto teste =14; Não Eventos = 335. Eventos no conjunto treinamento = 54; Não Eventos = 1342





**Figura A- 6** Estatísticas do pós-processamento (MOS e RN) para previsão de ocorrência de visibilidade restrita – Rodada de 12Z, com os dados replicados, referentes à estação Mostardas.

**Tabela A- 16** Sumário estatístico do pós-processamento (MOS e RN) para previsão de ocorrência visibilidade restrita - Rodada de 12Z, com os dados não replicados, referentes à estação Mostardas.

Estatísticas	Rodada 12Z			
	RN CP	RN SCR	MOS CP	MOS SCR
<b>PC</b>	0.9628	0.9542	0.9026	0.9226
<b>TS</b>	0.0714	0	0.1282	0.1563
<b>POD</b>	0.0714	0	0.3571	0.3571
<b>FAR</b>	0	1	0.8333	0.7826
<b>POFD</b>	0	0.0030	0.0746	0.0537
<b>B</b>	0.0714	0.1429	2.1429	1.6429
<b>Topologia (RN)</b>	16-8-1 5000-31	4-4-1 5000-31	17 coef.	5 coef.
<b>Alfa (RN)</b>	0.1	0.1	-	-
<b>Limiar (MOS)</b>	-	-	0.1	0.1

Nota: Dados não replicados. Eventos no conjunto teste =14; Não Eventos = 335. Eventos no conjunto treinamento = 54; Não Eventos = 1342

## **Apêndice B – Regressões Lineares Múltiplas do MOS**

Serão apresentadas tabelas com os coeficientes das regressões lineares múltiplas obtidos com o método MOS para os vários experimentos realizados.

### *Equações do MOS para previsão de ocorrência de vento forte -Salinópolis*

**Tabela B- 1** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Salinópolis.

$R^2 = 72.23\%$	$F(16,5981) = 972.23$		$P_{\text{valor}} = 0$
Preditores	Coeficientes ( $b_k$ )	Desvio-Padrões ( $s(b_k)$ )	Pvalor
Constante ( $b_0$ )	0.4998	0.0034	0.0000
CP 1	0.0274	0.0011	0.0000
CP 2	-0.0203	0.0017	0.0000
CP 3	0.1863	0.0020	0.0000
CP 4	-0.0504	0.0024	0.0000
CP 5	-0.1883	0.0028	0.0000
CP 6	0.0146	0.0034	0.0000
CP 7	0.0108	0.0036	0.0025
CP 8	0.0014	0.0041	0.7317
CP 9	0.0125	0.0043	0.0035
CP 10	0.0702	0.0045	0.0000
CP 11	-0.0499	0.0048	0.0000
CP 12	-0.0746	0.0052	0.0000
CP 13	-0.0580	0.0054	0.0000
CP 14	-0.1096	0.0056	0.0000
CP 15	0.0929	0.0068	0.0000
CP 16	0.0623	0.0075	0.0000

**Tabela B- 2** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário - Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Salinópolis.

$R^2 =$ 74.36%	$F(24,5974) = 753.27$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$
Constante ( $b_0$ )	0.4995	0.0033	0.0000
hora	0.2878	0.0086	0.0000
tp	-0.0284	0.0117	0.0154
an	0.1114	0.0083	0.0000
tsfcl	0.5285	0.0996	0.0000
tsfcs	-0.1514	0.0994	0.1267
t2ml	0.4166	0.0946	0.0000
t2ms	-0.3346	0.0289	0.0000
dpt2ml	-0.1416	0.1671	0.3928
dpt2ms	-0.3409	0.0294	0.0000
u10ml	-0.4069	0.0647	0.0000
u10ms	0.5806	0.0690	0.0000
v10ml	-0.0812	0.0350	0.0202
v10ms	0.0058	0.0368	0.8726
clctl	0.0643	0.0117	0.0000
psfc	-0.1566	0.0120	0.0000
vort1000	0.4088	0.0169	0.0000
urel850	0.1384	0.0142	0.0000
hgt500	0.0641	0.0136	0.0000
vort500	-0.0440	0.0076	0.0000
magv250	-0.1472	0.0073	0.0000
urfct	0.6546	0.2265	0.0038
urel12hr <sub>obs</sub>	0.1789	0.0112	0.0000
magv12hr <sub>obs</sub>	0.3722	0.0125	0.0000

Nota: variáveis descartadas no processo *screening* cobertura de nuvens (clcts), divergência em 850 Hpa e divergência em 250Hpa

**Tabela B- 3** Equação para previsão de ocorrência de vento forte onde preditores são as componentes principais (CP) e o preditando é binário - Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Salinópolis.

$R^2 = 1.152\%$			
$F(16,2988) = 2.177$		$p_{\text{valor}} = 0.0044$	
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$
Constante ( $b_0$ )	0.0020	0.0008	0.9999
CP 1	0.0003	0.00033	0.3076
CP 2	0.0007	0.00046	0.1283
CP 3	-0.0001	0.00049	0.7652
CP 4	-0.0004	0.00055	0.4842
CP 5	0.0007	0.00067	0.2736
CP 6	-0.0014	0.00074	0.0503
CP 7	-0.0005	0.00078	0.5028
CP 8	0.0011	0.00081	0.1607
CP 9	0.0016	0.00082	0.0530
CP 10	-0.0023	0.00085	0.0062
CP 11	-0.0003	0.00089	0.6964
CP 12	0.0007	0.00092	0.4590
CP 13	0.0032	0.00100	0.0016
CP 14	0.0012	0.00101	0.2328
CP 15	0.0003	0.00118	0.8128
CP 16	0.0001	0.00122	0.9171

**Tabela B- 4** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário - Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Salinópolis.

$R^2 = 0.532$			
$F(3,3001) = 5.352$		$p_{\text{valor}} = 0.0011$	
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$
Constante ( $b_0$ )	0.0020	0.0008	0.0141
hora	0.0050	0.0017	0.0027
vort1000	0.0047	0.0017	0.0059
urfct	0.0034	0.0017	0.0389

**Tabela B- 5** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodada de 00Z , com os dados replicados, referentes à estação Salinópolis.

<b>R<sup>2</sup> = 86.549</b>		<b>F(20,2979) = 1199.628</b>		<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>	
Constante (b <sub>0</sub> )	0.5001	0.0034	0.0000	
CP 1	-0.0346	0.0010	0.0000	
CP 2	-0.2002	0.0018	0.0000	
CP 3	0.0104	0.0024	0.0000	
CP 4	0.1369	0.0026	0.0000	
CP 5	-0.1413	0.0032	0.0000	
CP 6	-0.0221	0.0034	0.0000	
CP 7	0.0369	0.0037	0.0025	
CP 8	0.0417	0.0040	0.7317	
CP 9	0.0070	0.0042	0.0035	
CP 10	0.0307	0.0043	0.0000	
CP 11	-0.0685	0.0051	0.0000	
CP 12	-0.0233	0.0052	0.0000	
CP 13	0.0351	0.0059	0.0000	
CP 14	-0.1147	0.0064	0.0000	
CP 15	0.0247	0.0071	0.0000	
CP 16	-0.0356	0.0081	0.0000	

**Tabela B- 6** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário Rodada de 00Z , com os dados replicados, referentes à estação Salinópolis.

$R^2 = 87.272$	$F(20,2979) = 1021.31$	$p_{\text{valor}} = 0$	
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	$P_{\text{valor}}$
Constante ( $b_0$ )	0.5001	0.0033	0.0000
hora	0.2065	0.0096	0.0000
tp	-0.1058	0.0099	0.0000
tsfcl	-0.2544	0.0822	0.0020
tsfcs	0.3465	0.0845	0.0000
t2ml	0.1337	0.0208	0.0000
t2ms	-0.2797	0.0179	0.0000
u10ml	-0.5945	0.0934	0.0000
u10ms	0.4019	0.0987	0.0000
v10ml	-0.1004	0.0098	0.0000
clctl	0.1540	0.0140	0.0000
clcts	-0.1425	0.0139	0.0000
vort1000	0.3167	0.0185	0.0000
div850	0.0846	0.0100	0.0000
urel850	0.0689	0.0128	0.0000
hgt500	-0.2126	0.0107	0.0000
vort500	-0.0391	0.0067	0.0000
div250	-0.0164	0.0082	0.0446
magv250	-0.0884	0.0074	0.0000
urel12hr <sub>obs</sub>	0.1768	0.0119	0.0000
magv12hr <sub>obs</sub>	0.2779	0.0172	0.0000

Nota: variáveis descartadas no processo *screening*:  
dpt2ml, dpt2ms, v10ms, psfc e urfct.

**Tabela B- 7 Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditor é binário – Rodada de 12Z , com os dados replicados, referentes à estação Salinópolis.**

<b><math>R^2 = 75.729</math>   <math>F(16,2983) = 581.715</math>   <math>p_{\text{valor}} = 0.000</math></b>			
<b>Preditores</b>	<b>Coefficientes (<math>b_k</math>)</b>	<b>Desvio-Padrão s(<math>b_k</math>)</b>	<b>Pvalor</b>
Constante ( $b_0$ )	0.4998	0.0045	0.0000
CP 1	0.0602	0.0016	0.0000
CP 2	-0.0732	0.0021	0.0000
CP 3	0.1100	0.0026	0.0000
CP 4	0.1051	0.0033	0.0000
CP 5	-0.1974	0.0040	0.0000
CP 6	0.0689	0.0047	0.0000
CP 7	-0.1190	0.0049	0.0000
CP 8	-0.0346	0.0054	0.0000
CP 9	0.0025	0.0054	0.6444
CP 10	0.0716	0.0061	0.0000
CP 11	-0.0771	0.0067	0.0000
CP 12	0.0750	0.0071	0.0000
CP 13	-0.0346	0.0075	0.0000
CP 14	-0.0754	0.0087	0.0000
CP 15	0.0258	0.0092	0.0052
CP 16	0.1136	0.0106	0.0000

**Tabela B- 8** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditor é binário - Rodada de 12Z , com os dados replicados, referentes à estação Salinópolis.

<b>R<sup>2</sup> = 77.965</b>		<b>F(19,2980) = 554.956</b>		<b>p<sub>valor</sub> = 0</b>	
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>		
Constante (b <sub>0</sub> )	0.4999	0.0043	0.0000		
hora	0.1668	0.0118	0.0000		
tp	-0.2166	0.0134	0.0000		
an	0.3451	0.0160	0.0000		
tsfcs	0.1102	0.0267	0.0000		
t2ml	-0.2941	0.0271	0.0000		
t2ms	-0.2169	0.0215	0.0000		
dpt2ms	0.6668	0.0465	0.0000		
v10ml	-0.8561	0.0515	0.0000		
v10ms	0.0672	0.0127	0.0000		
clcts	-0.1278	0.0157	0.0000		
psfc	0.2460	0.0157	0.0000		
vort1000	-0.1133	0.0120	0.0000		
div850	0.0871	0.0149	0.0000		
urel850	-0.0561	0.0159	0.0004		
hgt500	0.0336	0.0106	0.0015		
vort500	0.0658	0.0124	0.0000		
magv250	-0.1744	0.0096	0.0000		
urel12hr <sub>obs</sub>	0.1575	0.0149	0.0000		
magv12hr <sub>obs</sub>	0.4047	0.0179	0.0000		

Nota: No processo de *screening* saíram as variáveis: tsfcl, dpt2ml, u10ml, u10ms, clctl, urfet.



***Equações do MOS para previsão de ocorrência vento forte - Santa Marta***

**Tabela B- 9** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Santa Marta.

$R^2 = 36.32\%$	$F(16,2023) = 72.463$		$p_{\text{valor}} = 0$
Preditores	Coeficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	$p_{\text{valor}}$
Constante ( $b_0$ )	0.500	0.0089	0.000
CP 1	-0.004	0.0031	0.2577
CP 2	-0.079	0.0043	0.0000
CP 3	0.000	0.0057	0.9445
CP 4	-0.010	0.0069	0.1467
CP 5	0.168	0.0075	0.0000
CP 6	-0.029	0.0084	0.0006
CP 7	0.010	0.0090	0.2851
CP 8	0.038	0.0096	0.0001
CP 9	0.013	0.0098	0.1970
CP 10	0.069	0.0099	0.0000
CP 11	-0.071	0.0104	0.0000
CP 12	0.130	0.0113	0.0000
CP 13	0.011	0.0126	0.3824
CP 14	-0.015	0.0131	0.2497
CP 15	0.085	0.0140	0.0000
CP 16	-0.069	0.0168	0.0000

**Tabela B- 10** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Santa Marta.

$R^2 = 40.22\%$		$F(16,2033) = 85.498$		$p_{\text{valor}} = 0$	
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	P <sub>valor</sub>		
Constante ( $b_0$ )	0.5	0.0086	0.0000		
hh	-0.059	0.0245	0.0158		
tp	-0.089	0.0191	0.0000		
tsfclf	0.799	0.1702	0.0000		
tsfcsf	-0.076	0.0262	0.0040		
t2ml	-1.362	0.1776	0.0000		
dpt2ml	0.311	0.0573	0.0000		
dpt2ms	-0.213	0.0456	0.0000		
v10ms	-0.265	0.0669	0.0001		
clcts	-0.113	0.0208	0.0000		
psfcf	-0.567	0.0347	0.0000		
vort1000	0.207	0.0663	0.0018		
div850	-0.111	0.0265	0.0000		
urel850	-0.082	0.0232	0.0004		
hgt500	0.171	0.0359	0.0000		
urel12hr <sub>obs</sub>	-0.221	0.0203	0.0000		
magv12hr <sub>obs</sub>	0.228	0.0190	0.0000		

**Tabela B- 11** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Santa Marta.

$R^2 = 13.11\%$		$F(16,2023) = 10.318$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$	
Constante ( $b_0$ )	0.0774	0.0075	0.000	
CP 1	0.0014	0.0028	0.6227	
CP 2	-0.0256	0.0038	0.0000	
CP 3	-0.0005	0.0046	0.9094	
CP 4	0.0115	0.0055	0.0368	
CP 5	-0.0316	0.0066	0.0000	
CP 6	-0.0463	0.0069	0.0000	
CP 7	-0.0112	0.0074	0.1335	
CP 8	0.0178	0.0079	0.0243	
CP 9	0.0184	0.0082	0.0252	
CP 10	-0.0195	0.0084	0.0204	
CP 11	-0.0292	0.0087	0.0008	
CP 12	0.0234	0.0091	0.0103	
CP 13	-0.0092	0.0103	0.3727	
CP 14	-0.0326	0.0110	0.0030	
CP 15	0.0048	0.0116	0.6803	
CP 16	-0.0198	0.0140	0.1560	

**Tabela B- 12** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Santa Marta.

$R^2 = 14.69$		$F(11,1099) = 17.21$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$	
Constante ( $b_0$ )	0.0774	0.0075	0.000	
hh	-0.0432	0.01688	0.0107	
t2ms	-0.0224	0.02897	0.4395	
dpt2ms	-0.1170	0.03368	0.0005	
v10ms	-0.0503	0.02128	0.0183	
clcts	-0.0410	0.01855	0.0271	
psfcf	-0.1364	0.02108	0.0000	
div850	-0.0374	0.02074	0.0718	
urel850	-0.0477	0.01932	0.0138	
urfct	0.0889	0.02087	0.0000	
urel12hr <sub>obs</sub>	-0.0602	0.01677	0.0003	
magv12hr <sub>obs</sub>	0.0904	0.01559	0.0000	

**Tabela B- 13** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodada de 00Z, com os dados replicados, referentes à estação Santa Marta.

$R^2 = 38.71$		$F(16,1027) = 40.54$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	Pvalor	
Constante ( $b_0$ )	0.5000	0.0122	0.0000	
CP 1	-0.0162	0.0042	0.0001	
CP 2	-0.0778	0.0059	0.0000	
CP 3	-0.0372	0.0082	0.0000	
CP 4	-0.0843	0.0100	0.0000	
CP 5	-0.1630	0.0107	0.0000	
CP 6	-0.0257	0.0115	0.0259	
CP 7	-0.0235	0.0124	0.0588	
CP 8	0.0605	0.0134	0.0000	
CP 9	-0.0735	0.0140	0.0000	
CP 10	-0.0075	0.0152	0.6194	
CP 11	0.0088	0.0157	0.5751	
CP 12	0.0118	0.0173	0.4950	
CP 13	-0.0766	0.0182	0.0000	
CP 14	0.1321	0.0199	0.0000	
CP 15	0.0031	0.0230	0.8936	
CP 16	0.0971	0.0256	0.0002	

**Tabela B- 14** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodada de 00Z, com os dados replicados, referentes à estação Santa Marta.

$R^2 = 41.97\%$		$F(14,1029) = 53.14$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	Pvalor	
Constante ( $b_0$ )	0.5000	0.0119	0.0000	
tp	-0.098	0.0291	0.0008	
tsfclf	0.886	0.2155	0.0000	
t2ml	-0.060	0.5682	0.9158	
t2ms	-0.171	0.0420	0.0000	
dpt2ml	-1.468	0.5349	0.0062	
clcts	-0.063	0.0291	0.0305	
psfcf	-0.653	0.0482	0.0000	
urel850	-0.178	0.0335	0.0000	
hgt500	0.326	0.0560	0.0000	
vort500	-0.126	0.0323	0.0001	
div250	-0.116	0.0257	0.0000	
urfct	0.834	0.2608	0.0014	
urel12hr <sub>obs</sub>	-0.210	0.0267	0.0000	
magv12hr <sub>obs</sub>	0.264	0.0260	0.0000	

**Tabela B- 15** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodada de 12Z, com os dados replicados, referentes à estação Santa Marta.

$R^2 = 38.06\%$	$F(16,991) = 38.00$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrões ( $s(b_k)$ )	$P_{\text{valor}}$
Constante ( $b_0$ )	0.5000	0.0125	0.000
CP 1	0.0053	0.0043	0.2213
CP 2	-0.0914	0.0065	0.0000
CP 3	-0.0053	0.0076	0.4891
CP 4	-0.0212	0.0096	0.0278
CP 5	0.1433	0.0105	0.0000
CP 6	0.0148	0.0120	0.2171
CP 7	0.0270	0.0125	0.0306
CP 8	0.0621	0.0132	0.0000
CP 9	-0.0994	0.0140	0.0000
CP 10	-0.1139	0.0152	0.0000
CP 11	0.0687	0.0160	0.0000
CP 12	0.0961	0.0181	0.0000
CP 13	-0.0509	0.0204	0.0127
CP 14	-0.0525	0.0214	0.0144
CP 15	0.0795	0.0244	0.0011
CP 16	-0.1198	0.0305	0.0001

**Tabela B- 16** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodada de 12Z, com os dados replicados, referentes à estação Santa Marta.

$R^2 = 46.82\%$	$F(16,991) = 54.53$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrões ( $s(b_k)$ )	$P_{\text{valor}}$
Constante ( $b_0$ )	0.5000	0.0125	0.0000
tsfelf	0.825	0.2259	0.0003
tsfcsf	-0.157	0.0375	0.0000
t2ml	-1.032	0.2707	0.0001
dpt2ms	-0.213	0.0658	0.0012
v10ms	-1.020	0.1045	0.0000
clcts	-0.157	0.0294	0.0000
psfcf	-0.551	0.0456	0.0000
vort1000	0.884	0.1017	0.0000
div850	-0.064	0.0381	0.0956
hgt500	0.194	0.0518	0.0002
vort500	0.100	0.0267	0.0002
div250	0.062	0.0253	0.0139
magv250	0.104	0.0333	0.0018
urfct	0.222	0.0439	0.0000
urel12hr <sub>obs</sub>	-0.187	0.0301	0.0000
magv12hr <sub>obs</sub>	0.234	0.0257	0.0000

***Equações do MOS para previsão de ocorrência vento forte  
- Mostardas***

**Tabela B- 17** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Mostardas.

<b>R<sup>2</sup> = 43.45%</b>	<b>F(16,5495) = 263.91</b>		<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>
Constante (b <sub>0</sub> )	0.5000	0.0051	0.0000
CP 1	0.0055	0.0020	0.0068
CP 2	0.0187	0.0026	0.0000
CP 3	-0.1042	0.0029	0.0000
CP 4	-0.0226	0.0032	0.0000
CP 5	0.1350	0.0039	0.0000
CP 6	-0.0585	0.0045	0.0000
CP 7	0.0892	0.0049	0.0000
CP 8	0.0612	0.0052	0.0000
CP 9	-0.0843	0.0053	0.0000
CP 10	-0.0574	0.0059	0.0000
CP 11	0.1028	0.0064	0.0000
CP 12	0.0630	0.0067	0.0000
CP 13	0.0422	0.0072	0.0000
CP 14	0.1048	0.0081	0.0000
CP 15	-0.0895	0.0085	0.0000
CP 16	0.0129	0.0091	0.1553

**Tabela B- 18** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Mostardas.

<b>R<sup>2</sup>=47.65%</b>		<b>F(18, 5493) =277.73</b>		<b>p<sub>valor</sub> = 0</b>	
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>		
Constante (b <sub>0</sub> )	0.5000	0.0049	0.000		
tp	0.0395	0.01057	0.0002		
an	-0.0785	0.01071	0.0000		
tsfclf	-0.6722	0.06899	0.0000		
t2ml	0.4295	0.08030	0.0000		
dpt2ms	-0.0853	0.02182	0.0001		
u10ms	-0.1043	0.01544	0.0000		
v10ml	-0.1030	0.01379	0.0000		
clct	-0.2070	0.02304	0.0000		
clcts	0.0568	0.02223	0.0106		
psfcf	-0.6009	0.02020	0.0000		
vort1000	-0.0617	0.01481	0.0000		
div850	-0.0328	0.01171	0.0050		
urel850	0.1943	0.01478	0.0000		
hgt500	0.1205	0.02490	0.0000		
vort500	-0.0665	0.01463	0.0000		
div250	-0.0893	0.01131	0.0000		
magv250	0.3170	0.01301	0.0000		
magv12hr <sub>obs</sub>	0.1532	0.01241	0.0000		

**Tabela B- 19** Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Mostardas.

$R^2=2.46\%$		$F(16,5495) = 4.35$		$p_{\text{valor}} = 0$
Preditores	Coeficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	Pvalor	
Constante ( $b_0$ )	0.0097	0.0018	0.0000	
CP 1	0.0001	0.00069	0.8453	
CP 2	0.0039	0.00105	0.0002	
CP 3	0.0000	0.00108	0.9699	
CP 4	0.0011	0.00147	0.4595	
CP 5	-0.0088	0.00156	0.0000	
CP 6	-0.0032	0.00161	0.0470	
CP 7	-0.0003	0.00174	0.8641	
CP 8	0.0032	0.00187	0.0892	
CP 9	0.0019	0.00195	0.3320	
CP 10	-0.0008	0.00200	0.6952	
CP 11	0.0009	0.00214	0.6705	
CP 12	-0.0030	0.00216	0.1722	
CP 13	0.0075	0.00226	0.0009	
CP 14	0.0021	0.00248	0.3926	
CP 15	-0.0010	0.00254	0.6937	
CP 16	-0.0026	0.00272	0.3302	

**Tabela B- 20** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Mostardas.

$R^2 = 2.63\%$		$F(4,2778) = 18.72$		$p_{\text{valor}} = 0$
Preditores	Coeficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	Pvalor	
Constante ( $b_0$ )	0.0097	0.0018	0.0000	
dpt2ms	-0.01672	0.005406	0.0020	
psfcf	-0.03432	0.004988	0.0000	
magv250	0.00896	0.004104	0.0290	
magv12hr <sub>obs</sub>	0.00951	0.003718	0.0106	



**Tabela B- 21 Equação para previsão de ocorrência de Vento Forte onde preditores são as componentes principais (CP) e o preditando é binário - Rodadas de 00Z, com os dados replicados, referentes à estação Mostardas.**

$R^2 = 45.25\%$		$F(16, 2729) = 140.99$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$p_{\text{valor}}$	
Constante ( $b_0$ )	0.500	0.0071	0.0000	
CP 1	0.025	0.0027	0.0000	
CP 2	0.034	0.0037	0.0000	
CP 3	-0.125	0.0039	0.0000	
CP 4	-0.027	0.0045	0.0000	
CP 5	0.099	0.0058	0.0000	
CP 6	0.059	0.0063	0.0000	
CP 7	-0.046	0.0068	0.0000	
CP 8	-0.057	0.0074	0.0000	
CP 9	0.038	0.0086	0.0000	
CP 10	0.095	0.0092	0.0000	
CP 11	-0.029	0.0096	0.0030	
CP 12	-0.055	0.0103	0.0000	
CP 13	0.171	0.0110	0.0000	
CP 14	-0.120	0.0120	0.0000	
CP 15	-0.068	0.0131	0.0000	
CP 16	0.049	0.0150	0.0012	

**Tabela B- 22** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z, com os dados replicados, referentes à estação Mostardas.

$R^2 = 55.76$		$F(20, 2725) = 171.74$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	Pvalor	
Constante ( $b_0$ )	0.5000	0.0064	0.0000	
hh	0.103	0.0173	0.0000	
tsfcsf	-0.322	0.0575	0.0000	
t2ml	-5.623	0.4525	0.0000	
t2ms	0.678	0.0694	0.0000	
dpt2ml	5.944	0.5053	0.0000	
dpt2ms	-0.609	0.0418	0.0000	
u10ml	-0.604	0.0896	0.0000	
u10ms	0.695	0.0961	0.0000	
clcts	-0.109	0.0200	0.0000	
psfcf	-0.757	0.0305	0.0000	
vort1000	-0.169	0.0267	0.0000	
div850	-0.165	0.0164	0.0000	
urel850	0.167	0.0191	0.0000	
hgt500	0.374	0.0390	0.0000	
vort500	-0.285	0.0234	0.0000	
div250	-0.078	0.0153	0.0000	
magv250	0.200	0.0182	0.0000	
urfct	-2.142	0.1907	0.0000	
urel12hr <sub>obs</sub>	0.120	0.0169	0.0000	
magv12hr <sub>obs</sub>	0.144	0.0176	0.0000	

**Tabela B- 23** Equação para previsão de ocorrência de Visibilidade restrita onde preditores são as componentes principais (CP) e o preditando é binário – Rodadas de 12Z, com os dados replicados, referentes à estação Mostardas.

$R^2=63.51\%$	$F(16,2745)=298.70$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$
Constante ( $b_0$ )	0.5001	0.0058	0.0000
CP 1	-0.0027	0.0022	0.2239
CP 2	-0.0718	0.0028	0.0000
CP 3	0.1760	0.0034	0.0000
CP 4	0.0739	0.0036	0.0000
CP 5	-0.0232	0.0051	0.0000
CP 6	-0.0042	0.0052	0.4200
CP 7	-0.1199	0.0056	0.0000
CP 8	0.0950	0.0060	0.0000
CP 9	0.0150	0.0069	0.0312
CP 10	0.0416	0.0074	0.0000
CP 11	-0.0562	0.0081	0.0000
CP 12	0.0558	0.0083	0.0000
CP 13	0.0020	0.0090	0.8244
CP 14	-0.0375	0.0096	0.0001
CP 15	-0.0790	0.0100	0.0000
CP 16	-0.0458	0.0110	0.0000

**Tabela B- 24** Equação para previsão de ocorrência de Vento Forte onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 12Z, com os dados replicados, referentes à estação Mostardas.

$R^2 = 70.42\%$		$F(20,2741) = 326.21$		$p_{\text{valor}} = 0$	
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	Pvalor		
Constante ( $b_0$ )	0.5001	0.0052	0.0000		
hh	0.1146	0.0147	0.0000		
tp	0.0374	0.0119	0.0017		
tsfclf	-0.2629	0.0815	0.0013		
tsfesf	0.1853	0.0525	0.0004		
t2ml	0.0994	0.1027	0.3332		
t2ms	-0.3992	0.0688	0.0000		
dpt2ms	0.0921	0.0278	0.0010		
u10ml	0.1941	0.0673	0.0040		
u10ms	-0.3571	0.0687	0.0000		
clcts	-0.1650	0.0164	0.0000		
psfcf	-0.8268	0.0265	0.0000		
vort1000	-0.0438	0.0181	0.0159		
div850	0.1171	0.0126	0.0000		
urel850	0.1766	0.0159	0.0000		
hgt500	0.1433	0.0330	0.0000		
vort500	-0.1011	0.0169	0.0000		
div250	0.0877	0.0134	0.0000		
magv250	0.3104	0.0145	0.0000		
urel12hr <sub>obs</sub>	-0.1223	0.0150	0.0000		
magv12hr <sub>obs</sub>	0.0608	0.0123	0.0000		

***Equações do MOS para previsão de ocorrência de visibilidade restrita - Salinópolis***

**Tabela B- 25** Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Salinópolis.

<b>R<sup>2</sup> = 40.51%</b>		<b>F(16,6007) = 255.612</b>		<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrões (b<sub>k</sub>)</b>	<b>Pvalor</b>	
Constante (b <sub>0</sub> )	0.5001	0.0050	0.0000	
CP 1	0.0517	0.0018	0.0000	
CP 2	0.0227	0.0029	0.0000	
CP 3	0.0558	0.0030	0.0000	
CP 4	0.0328	0.0035	0.0000	
CP 5	-0.0518	0.0036	0.0000	
CP 6	0.0500	0.0041	0.0000	
CP 7	-0.0875	0.0044	0.0000	
CP 8	-0.1055	0.0052	0.0000	
CP 9	0.0252	0.0054	0.0000	
CP 10	-0.0490	0.0059	0.0000	
CP 11	-0.0812	0.0061	0.0000	
CP 12	0.0097	0.0070	0.1630	
CP 13	-0.1616	0.0074	0.0000	
CP 14	0.0496	0.0076	0.0000	
CP 15	0.2321	0.0080	0.0000	
CP 16	0.0022	0.0087	0.8025	

**Tabela B- 26** Equação para previsão de ocorrência de Visibilidade Restrita onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com os dados replicados, referentes à estação Salinópolis.

$R^2 = 46.93\%$		$F(24,5999) = 221.030$		$p_{\text{valor}} = 0$	
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	Pvalor		
Constante ( $b_0$ )	0.4994	0.0047	0.0000		
hh	0.1226	0.0267	0.0000		
tp	-0.0274	0.0172	0.1100		
an	0.1398	0.0114	0.0000		
tsfclf	1.8549	0.1038	0.0000		
tsfcsf	-1.7368	0.1029	0.0000		
t2ml	-0.8218	0.1399	0.0000		
t2ms	-0.2200	0.0395	0.0000		
dpt2ml	1.6853	0.1932	0.0000		
dpt2ms	-0.0872	0.0329	0.0083		
u10ml	0.2575	0.0635	0.0001		
u10ms	-0.1367	0.0656	0.0376		
v10ml	0.0712	0.0124	0.0000		
clct	0.0666	0.0190	0.0005		
clcts	-0.1053	0.0186	0.0000		
psfcf	-0.2694	0.0166	0.0000		
vort1000	0.3040	0.0222	0.0000		
div850	0.2151	0.0106	0.0000		
urel850	0.5430	0.0162	0.0000		
hgt500	0.0529	0.0129	0.0000		
vort500	-0.0974	0.0109	0.0000		
div250	-0.0944	0.0142	0.0000		
urfct	-2.1319	0.2825	0.0000		
urel12hr <sub>obs</sub>	-0.0859	0.0134	0.0000		
visib12hr <sub>obs</sub>	-0.1904	0.0280	0.0000		

**Tabela B- 27 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Salinópolis.**

$R^2 = 0.64\%$		$F(16,3006) = 1.204$		$p_{\text{valor}} = 0.25570$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	Pvalor	
Constante ( $b_0$ )	0.0036	0.0011	0.0009	
CP 1	0.0008	0.0005	0.0665	
CP 2	0.0006	0.0006	0.3475	
CP 3	0.0003	0.0006	0.5914	
CP 4	-0.0006	0.0008	0.3915	
CP 5	-0.0011	0.0008	0.1891	
CP 6	-0.0012	0.0010	0.2258	
CP 7	0.0006	0.0010	0.5625	
CP 8	-0.0010	0.0011	0.3459	
CP 9	0.0008	0.0011	0.4962	
CP 10	-0.0009	0.0012	0.4366	
CP 11	0.0006	0.0012	0.5962	
CP 12	-0.0001	0.0013	0.9251	
CP 13	0.0003	0.0013	0.8086	
CP 14	-0.0020	0.0014	0.1662	
CP 15	0.0010	0.0016	0.5385	
CP 16	0.0041	0.0017	0.0155	

**Tabela B- 28 Equação para previsão de ocorrência de Visibilidade Restrita onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com os dados não replicados, referentes à estação Salinópolis.**

$R^2 = 0.25\%$		$F(1,3021) = 7.533$		$p_{\text{valor}} = 0.0061$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	Pvalor	
Constante ( $b_0$ )	0.0036	0.0011	0.0009	
urel850	0.0060	0.0022	0.0061	

**Tabela B- 29 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 00Z, com os dados replicados, referentes à estação Salinópolis.**

$R^2 = 59.66\%$		$F(16,2999) = 277.159$		$P_{valor} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	Pvalor	
Constante ( $b_0$ )	0.5000	0.0058	0.0000	
CP 1	0.0452	0.0022	0.0000	
CP 2	0.0723	0.0028	0.0000	
CP 3	0.0215	0.0037	0.0000	
CP 4	0.2113	0.0041	0.0000	
CP 5	-0.0533	0.0046	0.0000	
CP 6	-0.0799	0.0051	0.0000	
CP 7	-0.0027	0.0055	0.6290	
CP 8	-0.0221	0.0058	0.0001	
CP 9	0.0113	0.0066	0.0841	
CP 10	0.0109	0.0071	0.1236	
CP 11	-0.0206	0.0079	0.0089	
CP 12	-0.0287	0.0084	0.0006	
CP 13	-0.0516	0.0088	0.0000	
CP 14	0.1328	0.0098	0.0000	
CP 15	0.0156	0.0107	0.1432	
CP 16	0.1015	0.0121	0.0000	



**Tabela B- 30 Equação para previsão de ocorrência de Visibilidade Restrita onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z, com os dados replicados, referentes à estação Salinópolis**

$R^2 = 67.38\%$		$F(19,2996) = 325.76$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	$p_{\text{valor}}$	
Constante ( $b_0$ )	0.5006	0.0052	0.0000	
hh	0.4295	0.0269	0.0000	
tp	0.1376	0.0166	0.0000	
tsfclf	-1.3905	0.1084	0.0000	
tsfcsf	1.2824	0.1049	0.0000	
t2ms	-0.0567	0.0173	0.0011	
u10ml	0.9485	0.0642	0.0000	
u10ms	-1.0875	0.0639	0.0000	
v10ml	0.3299	0.0589	0.0000	
v10ms	-0.2696	0.0533	0.0000	
clct	-0.3844	0.0214	0.0000	
clcts	0.2008	0.0158	0.0000	
psfcf	-0.0587	0.0185	0.0016	
vort1000	-0.1979	0.0198	0.0000	
urel850	0.1837	0.0208	0.0000	
hgt500	-0.0530	0.0133	0.0001	
vort500	0.0948	0.0128	0.0000	
div250	-0.0417	0.0159	0.0088	
urel12hr <sub>obs</sub>	-0.0655	0.0155	0.0000	
visib12hr <sub>obs</sub>	-0.1040	0.0293	0.0004	

**Tabela B- 31 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 12Z, com os dados replicados, referentes à estação Salinópolis**

$R^2 = 52.19\%$		$F(16, 2995) = 204.327$		$p_{\text{valor}} = 0$
<b>Preditores</b>	<b>Coefficientes (<math>b_k</math>)</b>	<b>Desvio-Padrão s(<math>b_k</math>)</b>	<b>p<sub>valor</sub></b>	
Constante ( $b_0$ )	0.5000	0.0063	0.0000	
CP 1	0.0447	0.0022	0.0000	
CP 2	0.0530	0.0034	0.0000	
CP 3	0.0876	0.0041	0.0000	
CP 4	0.0206	0.0043	0.0000	
CP 5	0.0475	0.0050	0.0000	
CP 6	-0.1569	0.0058	0.0000	
CP 7	0.1824	0.0066	0.0000	
CP 8	-0.1306	0.0068	0.0000	
CP 9	-0.0092	0.0073	0.2089	
CP 10	0.0306	0.0077	0.0001	
CP 11	0.0426	0.0083	0.0000	
CP 12	-0.0925	0.0098	0.0000	
CP 13	0.0099	0.0102	0.3311	
CP 14	0.0450	0.0103	0.0000	
CP 15	0.0692	0.0122	0.0000	
CP 16	-0.0379	0.0139	0.0064	

**Tabela B- 32 Equação para previsão de ocorrência de Visibilidade Restrita onde preditores foram selecionados pelo método screening e o preditando é binário – Rodadas de 12Z, com os dados replicados, referentes à estação Salinópolis**

<b>R<sup>2</sup> = 64.33%</b>			
		<b>F(22, 2989) = 245.05</b>	<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>
Constante (b <sub>0</sub> )	0.5000	0.0055	0.0000
hh	0.2140	0.0343	0.0000
tsfclf	0.5544	0.1233	0.0000
tsfcsf	-0.6041	0.1195	0.0000
t2ml	-0.3697	0.1569	0.0188
t2ms	-0.6157	0.0490	0.0000
dpt2ml	2.3782	0.2291	0.0000
dpt2ms	-0.4515	0.0444	0.0000
u10ml	-0.3356	0.0739	0.0000
u10ms	0.7303	0.0753	0.0000
v10ml	1.2465	0.0532	0.0000
v10ms	-0.9665	0.0531	0.0000
psfcf	-0.1772	0.0200	0.0000
vort1000	0.1209	0.0215	0.0000
div850	0.2102	0.0136	0.0000
urel850	0.1685	0.0196	0.0000
hgt500	0.0463	0.0155	0.0028
vort500	-0.0910	0.0137	0.0000
div250	-0.1222	0.0164	0.0000
magv250	-0.1468	0.0127	0.0000
urfct	-2.6265	0.3364	0.0000
urel12hr <sub>obs</sub>	-0.1213	0.0139	0.0000
visib12hr <sub>obs</sub>	0.0975	0.0332	0.0034

***Equações do MOS para previsão de ocorrência de  
visibilidade restrita –Santa Marta***

**Tabela B- 33 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 0012Z, com dados replicados, referentes à estação Santa Marta.**

<b>R<sup>2</sup> = 43.560</b>		<b>F(16, 2113) = 101.926</b>		<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio- Padrão s(b<sub>k</sub>)</b>	<b>p<sub>valor</sub></b>	
Constante (b <sub>0</sub> )	0.5000	0.0082	0.0000	
CP 1	0.0090	0.0029	0.0021	
CP 2	0.1092	0.0042	0.0000	
CP 3	0.0820	0.0048	0.0000	
CP 4	0.0835	0.0062	0.0000	
CP 5	0.0919	0.0071	0.0000	
CP 6	-0.0191	0.0076	0.0119	
CP 7	0.0129	0.0079	0.1012	
CP 8	-0.0534	0.0084	0.0000	
CP 9	0.0696	0.0089	0.0000	
CP 10	-0.0127	0.0095	0.1799	
CP 11	-0.0189	0.0100	0.0575	
CP 12	-0.0901	0.0108	0.0000	
CP 13	0.0716	0.0113	0.0000	
CP 14	-0.0709	0.0135	0.0000	
CP 15	-0.0178	0.0147	0.2255	
CP 16	-0.1157	0.0156	0.0000	

**Tabela B- 34 Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método screening e o preditando é binário – Rodadas de 0012Z, com dados replicados, referentes à estação Santa Marta.**

<b>R<sup>2</sup> = 48.509</b>		<b>F(16,2113) = 124.41</b>		<b>p<sub>valor</sub> = 0</b>	
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>		
Constante (b <sub>0</sub> )	0.5000	0.0078	0.0000		
hora	0.2786	0.0184	0.0000		
tp	-0.0837	0.0169	0.0000		
tsfcs	0.2831	0.0390	0.0000		
t2ms	-0.2194	0.0424	0.0000		
u10ml	0.1351	0.0347	0.0001		
u10ms	0.1810	0.0382	0.0000		
clctl	-0.1425	0.0247	0.0000		
clcts	0.2076	0.0254	0.0000		
vort1000	-0.5192	0.0348	0.0000		
div850	0.0963	0.0210	0.0000		
urel850	0.2345	0.0248	0.0000		
vort500	0.1068	0.0165	0.0000		
magv250	0.0992	0.0209	0.0000		
urfct	0.1504	0.0206	0.0000		
urel12hr <sub>obs</sub>	0.0885	0.0204	0.0000		
visib12hr <sub>obs</sub>	-0.1714	0.0190	0.0000		

**Tabela B- 35** equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 0012Z, com dados não replicados, referentes à estação Santa Marta.

$R^2 = 7.3$		$F(16,1094) = 5.384$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio- Padrão $s(b_k)$	Pvalor	
Constante ( $b_0$ )	0.0414	0.0058	0.0000	
CP 1	0.0034	0.0021	0.1096	
CP 2	0.0122	0.0029	0.0000	
CP 3	0.0123	0.0035	0.0005	
CP 4	0.0184	0.0044	0.0000	
CP 5	0.0098	0.0050	0.0478	
CP 6	-0.0045	0.0054	0.3988	
CP 7	-0.0079	0.0057	0.1633	
CP 8	-0.0007	0.0058	0.9073	
CP 9	-0.0041	0.0063	0.5097	
CP 10	-0.0087	0.0064	0.1767	
CP 11	0.0013	0.0069	0.8457	
CP 12	0.0053	0.0075	0.4764	
CP 13	0.0400	0.0081	0.0000	
CP 14	0.0060	0.0085	0.4794	
CP 15	-0.0110	0.0088	0.2125	
CP 16	-0.0094	0.0109	0.3887	

**Tabela B- 36** Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método screening e o preditando é binário – Rodadas de 0012Z, com dados não replicados, referentes à estação Santa Marta.

$R^2 = 7.352$		$F(7,1103) = 12.504$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio- Padrão $s(b_k)$	Pvalor	
Constante ( $b_0$ )	0.041404	0.0058	0.0000	
hora	0.05793	0.0121	0.0000	
clctl	-0.04319	0.0186	0.0204	
clcts	0.05641	0.0181	0.0019	
urel850	0.04414	0.0139	0.0015	
vort500	0.02382	0.0118	0.0427	
urfct	0.04001	0.0134	0.0030	
visib12hr <sub>obs</sub>	-0.04292	0.0119	0.0003	

**Tabela B- 37 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodada de 00Z, com dados os replicados, referentes à estação Santa Marta.**

<b>R<sup>2</sup> = 51.484</b>		<b>F(16,1081) = 71.697</b>		<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>p<sub>valor</sub></b>	
Constante (b <sub>0</sub> )	0.5000	0.0106	0.0000	
CP 1	0.0122	0.0038	0.0015	
CP 2	0.1189	0.0053	0.0000	
CP 3	0.0894	0.0063	0.0000	
CP 4	0.0111	0.0084	0.1871	
CP 5	0.1279	0.0087	0.0000	
CP 6	-0.0508	0.0103	0.0000	
CP 7	0.0637	0.0106	0.0000	
CP 8	-0.0204	0.0111	0.0673	
CP 9	-0.0086	0.0121	0.4778	
CP 10	0.0803	0.0130	0.0000	
CP 11	-0.0707	0.0144	0.0000	
CP 12	-0.0858	0.0157	0.0000	
CP 13	0.0255	0.0170	0.1334	
CP 14	0.1217	0.0191	0.0000	
CP 15	0.0297	0.0200	0.1378	
CP 16	0.0423	0.0220	0.0547	

**Tabela B- 38** Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método screening e o preditando é binário – Rodadas de 00Z, com dados replicados, referentes à estação Santa Marta.

<b>R<sup>2</sup> = 52.781</b>		<b>F(14,1083) = 86.470</b>		<b>p<sub>valor</sub> = 0</b>	
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrões(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>		
Constante (b <sub>0</sub> )	0.5000	0.0105	0.0000		
hh	0.2647	0.0237	0.0000		
tsfcsf	0.1518	0.0560	0.0067		
t2ms	-0.2943	0.0673	0.0000		
dpt2ms	0.2023	0.0458	0.0000		
u10ml	0.0915	0.0458	0.0456		
u10ms	0.1161	0.0539	0.0312		
v10ms	-0.3347	0.0577	0.0000		
clcts	0.1047	0.0285	0.0002		
vort1000	0.1860	0.0698	0.0078		
urel850	0.3380	0.0314	0.0000		
hgt500	0.1545	0.0387	0.0001		
vort500	0.1404	0.0269	0.0000		
urel12hr <sub>obs</sub>	0.0805	0.0266	0.0025		
visib12hr <sub>obs</sub>	-0.2531	0.0251	0.0000		



**Tabela B- 39 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodada de 12Z, com dados os replicados, referentes à estação Santa Marta**

<b>R<sup>2</sup> = 45.665</b>		<b>F(16, 1005) = 52.791</b>		<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>p<sub>valor</sub></b>	
Constante (b <sub>0</sub> )	0.5000	0.0116	0.0000	
CP 1	-0.0067	0.0041	0.1069	
CP 2	0.1034	0.0059	0.0000	
CP 3	0.0705	0.0070	0.0000	
CP 4	-0.0533	0.0090	0.0000	
CP 5	0.1537	0.0100	0.0000	
CP 6	-0.0520	0.0107	0.0000	
CP 7	0.0128	0.0115	0.2659	
CP 8	0.0267	0.0127	0.0350	
CP 9	-0.0010	0.0135	0.9425	
CP 10	0.0542	0.0142	0.0001	
CP 11	-0.0996	0.0151	0.0000	
CP 12	0.0653	0.0180	0.0003	
CP 13	0.0553	0.0190	0.0036	
CP 14	-0.0310	0.0207	0.1339	
CP 15	0.1196	0.0226	0.0000	
CP 16	0.1104	0.0249	0.0000	

**Tabela B- 40** Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método screening e o preditando é binário – Rodadas de 12Z, com dados replicados, referentes à estação Santa Marta.

<b><math>R^2 = 50.815</math></b>			
<b><math>F(19,1002) = 54.485</math></b>		<b><math>p_{valor} = 0</math></b>	
<b>Preditores</b>	<b>Coefficientes (<math>b_k</math>)</b>	<b>Desvio- Padrão s(<math>b_k</math>)</b>	<b>Pvalor</b>
Constante ( $b_0$ )	0.5000	0.0111	0.0000
hh	0.3196	0.0282	0.0000
tsfclf	-1.1638	0.2219	0.0000
tsfcsf	0.2499	0.0619	0.0001
t2ml	2.1640	0.6019	0.0003
t2ms	-0.2226	0.0885	0.0120
dpt2ml	-1.0665	0.5072	0.0355
u10ms	0.2570	0.0412	0.0000
v10ml	0.1370	0.0496	0.0057
clct	-0.1776	0.0355	0.0000
clcts	0.1571	0.0385	0.0000
vort1000	-0.5226	0.0600	0.0000
div850	0.0652	0.0323	0.0437
urel850	0.2537	0.0369	0.0000
hgt500	0.1634	0.0414	0.0001
vort500	0.0501	0.0287	0.0806
magv250	0.2190	0.0336	0.0000
urfct	0.6759	0.2522	0.0074
urel12hr <sub>obs</sub>	0.1265	0.0291	0.0000
visib12hr <sub>obs</sub>	-0.1678	0.0275	0.0000

***Equações do MOS para previsão de ocorrência de  
visibilidade restrita –Mostardas***

**Tabela B- 41 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 0012Z, com dados replicados, referentes à estação Mostardas.**

<b>R<sup>2</sup> = 33.21%</b>		<b>F(16,5379) = 167.174</b>		<b>Pvalor = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>Pvalor</b>	
Constante (b <sub>0</sub> )	0.5000	0.0056	0.0000	
CP 1	0.0395	0.0022	0.0000	
CP 2	-0.0995	0.0030	0.0000	
CP 3	0.0674	0.0032	0.0000	
CP 4	0.0453	0.0039	0.0000	
CP 5	0.0437	0.0047	0.0000	
CP 6	-0.0362	0.0051	0.0000	
CP 7	0.0478	0.0054	0.0000	
CP 8	-0.0530	0.0056	0.0000	
CP 9	-0.0166	0.0060	0.0056	
CP 10	-0.0636	0.0063	0.0000	
CP 11	-0.0283	0.0065	0.0000	
CP 12	0.0553	0.0069	0.0000	
CP 13	-0.0227	0.0073	0.0020	
CP 14	0.0424	0.0078	0.0000	
CP 15	0.0178	0.0079	0.0246	
CP 16	-0.0845	0.0083	0.0000	

**Tabela B- 42** Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com dados replicados, referentes à estação Mostardas.

<b>R<sup>2</sup> = 36.17%</b>			
<b>F(16,5379) = 190.538</b>		<b>p<sub>valor</sub> = 0</b>	
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio-Padrão s(b<sub>k</sub>)</b>	<b>P<sub>valor</sub></b>
Constante (b <sub>0</sub> )	0.5000	0.0054	0.0000
tsfcsf	-0.3432	0.0224	0.0000
t2ml	0.9828	0.3259	0.0025
dpt2ml	-1.3402	0.3689	0.0003
dpt2ms	0.3741	0.0302	0.0000
u10ml	-0.1057	0.0138	0.0000
v10ml	0.0829	0.0138	0.0000
clct	-0.1199	0.0203	0.0000
clcts	0.1730	0.0204	0.0000
psfcf	-0.1081	0.0162	0.0000
vort1000	0.0704	0.0142	0.0000
vort500	0.0474	0.0113	0.0000
div250	-0.0598	0.0115	0.0000
magv250	-0.0458	0.0128	0.0003
urfct	0.6333	0.1478	0.0000
urel12hr <sub>obs</sub>	0.1403	0.0139	0.0000
visib12hr <sub>obs</sub>	-0.2367	0.0127	0.0000

**Tabela B- 43** Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodadas de 00Z e 12Z, com não dados replicados, referentes à estação Mostardas.

$R^2 = 4.16\%$		$F(16,2771) = 7.522$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$	
Constante ( $b_0$ )	0.0323	0.0033	0.0000	
CP 1	0.0037	0.0012	0.0031	
CP 2	0.0083	0.0018	0.0000	
CP 3	-0.0129	0.0019	0.0000	
CP 4	0.0078	0.0023	0.0008	
CP 5	0.0033	0.0028	0.2328	
CP 6	0.0093	0.0030	0.0019	
CP 7	-0.0038	0.0031	0.2258	
CP 8	-0.0103	0.0033	0.0021	
CP 9	-0.0051	0.0036	0.1505	
CP 10	0.0056	0.0037	0.1224	
CP 11	0.0010	0.0038	0.7922	
CP 12	0.0009	0.0040	0.8224	
CP 13	0.0020	0.0043	0.6352	
CP 14	-0.0028	0.0045	0.5361	
CP 15	0.0077	0.0047	0.1010	
CP 16	-0.0119	0.0054	0.0268	

**Tabela B- 44** Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodadas de 00Z e 12Z, com não dados replicados, referentes à estação Mostardas.

$R^2 = 4.07\%$		$F(5,2782) = 23.619$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$P_{\text{valor}}$	
Constante ( $b_0$ )	0.0323	0.0033	0.0000	
tsfcsf	-0.0514	0.0107	0.0000	
dpt2ms	0.0401	0.0117	0.0006	
urfct	0.0194	0.0085	0.0217	
urel12hr <sub>obs</sub>	0.0153	0.0077	0.0460	
visib12hr <sub>obs</sub>	-0.0285	0.0068	0.0000	

**Tabela B- 45 Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodada de 00Z , com dados replicados, referentes à estação Mostardas**

<b>R<sup>2</sup> = 36.61%</b>		<b>F(16,2685) = 96.913</b>		<b>p<sub>valor</sub> = 0</b>
<b>Preditores</b>	<b>Coefficientes (b<sub>k</sub>)</b>	<b>Desvio- Padrão s(b<sub>k</sub>)</b>	<b>p<sub>valor</sub></b>	
Constante (b <sub>0</sub> )	0.5000	0.0077	0.0000	
CP 1	0.0453	0.0029	0.0000	
CP 2	0.1131	0.0042	0.0000	
CP 3	-0.0470	0.0045	0.0000	
CP 4	0.0226	0.0061	0.0002	
CP 5	0.0185	0.0062	0.0030	
CP 6	-0.0518	0.0069	0.0000	
CP 7	-0.0094	0.0074	0.2027	
CP 8	0.0103	0.0081	0.2017	
CP 9	-0.0453	0.0086	0.0000	
CP 10	0.0468	0.0091	0.0000	
CP 11	0.1044	0.0098	0.0000	
CP 12	0.0678	0.0104	0.0000	
CP 13	-0.0056	0.0106	0.6003	
CP 14	0.0604	0.0114	0.0000	
CP 15	-0.0592	0.0120	0.0000	
CP 16	-0.1392	0.0127	0.0000	

**Tabela B- 46** Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodada de 00Z, com dados replicados, referentes à estação Mostardas.

$R^2 = 39.67\%$		$F(17,2684) = 103.814$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	Pvalor	
Constante ( $b_0$ )	0.5000	0.0075	0.0000	
hh	-0.0807	0.0210	0.0001	
tp	0.0572	0.0186	0.0021	
tsfcf	-0.7234	0.1930	0.0002	
tsfcsf	0.2733	0.0757	0.0003	
t2ml	2.1748	0.5270	0.0000	
t2ms	-0.7230	0.0932	0.0000	
dpt2ml	-1.8215	0.5637	0.0012	
dpt2ms	0.4269	0.0516	0.0000	
u10ml	-0.1982	0.0194	0.0000	
v10ml	0.1143	0.0204	0.0000	
psfcf	-0.1037	0.0228	0.0000	
vort1000	-0.0585	0.0196	0.0029	
vort500	0.0575	0.0159	0.0003	
magv250	0.0878	0.0175	0.0000	
urfct	0.7922	0.2176	0.0003	
urel12hr <sub>obs</sub>	0.1399	0.0192	0.0000	
visib12hr <sub>obs</sub>	-0.3330	0.0192	0.0000	

**Tabela B- 47** Equações para previsão de ocorrência de visibilidade restrita onde preditores são PC e o preditor é binário – Rodada de 12Z , com dados replicados, referentes à estação Mostardas

$R^2 = 38.50\%$		$F(16,2667) = 104.358$		$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão $s(b_k)$	$p_{\text{valor}}$	
Constante ( $b_0$ )	0.5000	0.0076	0.0000	
CP 1	0.0413	0.0029	0.0000	
CP 2	-0.0895	0.0041	0.0000	
CP 3	0.0828	0.0044	0.0000	
CP 4	0.0781	0.0057	0.0000	
CP 5	0.0581	0.0064	0.0000	
CP 6	-0.0461	0.0069	0.0000	
CP 7	-0.0377	0.0073	0.0000	
CP 8	-0.0422	0.0078	0.0000	
CP 9	-0.0433	0.0085	0.0000	
CP 10	-0.0895	0.0088	0.0000	
CP 11	0.0390	0.0090	0.0000	
CP 12	0.0007	0.0095	0.9446	
CP 13	0.0432	0.0107	0.0001	
CP 14	-0.0612	0.0113	0.0000	
CP 15	0.0887	0.0115	0.0000	
CP 16	-0.0298	0.0128	0.0200	



**Tabela B- 48** Equação para previsão de ocorrência de Visibilidade restrita onde preditores foram selecionados pelo método *screening* e o preditando é binário – Rodada de 12Z , com dados replicados, referentes à estação Mostardas

$R^2 = 43.92\%$			
		$F(17,2666) = 122.803$	$p_{\text{valor}} = 0$
Preditores	Coefficientes ( $b_k$ )	Desvio-Padrão s( $b_k$ )	Pvalor
Constante ( $b_0$ )	0.5000	0.0073	0.0000
tsfcsf	-0.2394	0.0628	0.0001
t2ml	2.4096	0.4097	0.0000
t2ms	-0.2043	0.0795	0.0101
dpt2ml	-2.7793	0.4562	0.0000
dpt2ms	0.4938	0.0475	0.0000
u10ml	0.3813	0.0658	0.0000
u10ms	-0.4362	0.0682	0.0000
v10ms	0.0675	0.0196	0.0006
clct	-0.2867	0.0268	0.0000
clcts	0.2672	0.0269	0.0000
vort1000	0.1150	0.0210	0.0000
div850	-0.1094	0.0166	0.0000
div250	-0.1041	0.0152	0.0000
magv250	-0.0610	0.0165	0.0002
urfct	1.2747	0.1857	0.0000
urel12hr <sub>obs</sub>	0.1155	0.0192	0.0000
visib12hr <sub>obs</sub>	-0.2142	0.0169	0.0000

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)