



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-14172-TDI/1089

**EMPREGO DE REDES NEURAS ARTIFICIAIS E REGRESSÃO
LINEAR MÚLTIPLA NO REFINAMENTO DAS PREVISÕES DE
RADIÇÃO SOLAR DO MODELO ETA**

Ricardo André Guarnieri

Dissertação de Mestrado do Curso de Pós-Graduação em Meteorologia, orientada pelo
Dr. Enio Bueno Pereira, aprovada em 12 de abril de 2006.

551.521.37:621.3.05

Guarnieri, R. A.

Emprego de redes neurais artificiais e regressão linear múltipla no refinamento das previsões de radiação solar do modelo ETA / Ricardo André Guarnieri. – São José dos Campos: INPE, 2006.

171p. ; (INPE-14172-TDI/1089)

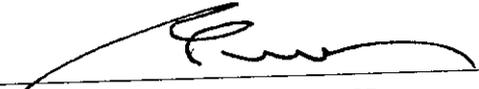
1.Previsão de tempo. 2. Radiação solar. 3.Energia solar.
4.redes neurais artificiais. 5.Análise estatística
multivariada. I.Título.

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Mestre** em
Meteorologia

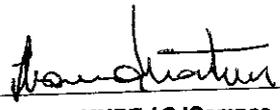
Dr. Clóvis Angeli Sansigolo


Presidente / INPE / SJCampos - SP

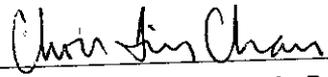
Dr. Enio Bueno Pereira


Orientador(a) / INPE / SJCampos - SP

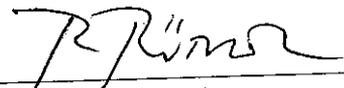
Dr. Fernando Ramos Martins


Membro da Banca / INPE / SJCampos - SP

Dra. Chou Sin Chan


Membro da Banca / INPE / Cachoeira Paulista - SP

Dr. Ricardo Rüther


Convidado(a) / UFSC / Florianópolis - SC

Aluno (a): **Ricardo André Guarnieri**

São José dos Campos, 12 de abril de 2006

“A terra fornece o suficiente para satisfazer todas as necessidades, mas não o suficiente para a cobiça de cada um”.

MAHATMA GANDHI

*A meus pais,
Dioni Antônio e
Dirce Teresinha Guarnieri*

AGRADECIMENTOS

A meus pais, Dioni e Dirce, pelo amor, carinho e apoio constante – emocional e financeiro – e por sempre acreditarem em mim e meus irmãos, estando ao nosso lado em todos os momentos de nossas vidas, independentemente da situação e das dificuldades.

Aos meus irmãos, em particular meu irmão mais velho Fernando, que além de ser a minha família em São José dos Campos, sempre foi grande amigo, confidente e professor. Agradecimento não menos importante aos meus irmãos Jean e Ângela, pela amizade e carinho, mesmo apesar de estarmos tão afastados fisicamente.

Aos meus amigos e colegas, inclusive os mais distantes, pelo companheirismo e por terem contribuído para minha formação profissional e humana. Agradecimento especial à minha namorada, Caren, pelo apoio, amor e carinho, mesmo apesar da distância.

De forma alguma poderia deixar de fazer um agradecimento especial ao meu orientador, Dr. Enio Bueno Pereira, por sempre ter disponibilizado tempo para me ensinar, discutir os trabalhos e para conversar até mesmo questões de ordem pessoal.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, pelo auxílio financeiro de dois anos de bolsa de mestrado (Processo: 132148/2004-8).

Ao Instituto Nacional de Pesquisas Espaciais – INPE e ao seu Centro de Previsão de Tempo e Estudos Climáticos – CPTEC/INPE, pela oportunidade de estudos e utilização de dados e instalações.

Ao Projeto SONDA, financiado pela FINEP (22.01.0569.00) e Petrobrás, pela disponibilização dos dados empregados nesta dissertação de mestrado.

A todos os professores do Curso de Pós-graduação em Meteorologia do CPTEC e a todos os professores do Curso de Engenharia Química da Universidade Federal de Santa Maria, por todos os conhecimentos e experiências transmitidos, que muito contribuíram para a realização de meus trabalhos e que levarei comigo por toda a minha vida.

Agradecimentos especiais aos professores e pesquisadores do CPTEC Dra. Chou Sin Chan, Dr. Clóvis Sansigolo e Dr. Fernando Ramos Martins, pelos inúmeros ensinamentos e por terem sempre se empenhado em sanar minhas dúvidas.

Agradecimento ao Dr. Nelson Jorge Schuch, meu primeiro orientador e grande incentivador, pessoa que me colocou no caminho da ciência e que sempre procurou inculcar em mim princípios humanos que vão muito além da área acadêmica.

Aos colegas do grupo de pesquisa, especialmente Hugo e Sheila, pela companhia diária, pela amizade e pelos conhecimentos transmitidos. Aos colegas de apartamento, Alisson e Luciano, pelo companheirismo, amizade e paciência ao longo desses dois anos de estudos. Aos colegas de mestrado, pela amizade e pelo compartilhamento de conhecimentos ao longo do período de disciplinas, e mesmo no ano de dissertação, em especial a Virginia. E à secretaria da PGMET, em especial a Lilian.

E a todas as demais pessoas que participaram de mais essa fase de minha vida.

RESUMO

A pesquisa abordada nesta dissertação teve como objetivo o desenvolvimento e avaliação de uma metodologia de previsão de irradiação solar incidente, voltada ao setor de geração e distribuição de energia, constituindo assim, uma pesquisa de meteorologia aplicada. O trabalho desenvolvido consiste no refinamento da previsão numérica de tempo produzida pelo modelo Eta/CPTEC, visando reduzir as incertezas associadas à previsão de radiação solar deste modelo. As variáveis de saída do modelo Eta, representando as condições atmosféricas previstas, foram empregadas como preditores em modelos de Redes Neurais Artificiais (RNAs) e Regressões Lineares Múltiplas (RLMs), ajustados para calcular a radiação solar incidente para duas localidades onde se situam estações do projeto SONDA: Florianópolis-SC (FLN) e São Martinho da Serra-RS (SMS). As medidas de radiação solar global realizadas nestes locais foram utilizadas na seleção de preditores, no treinamento de RNAs, no ajuste de RLMs e para avaliação das previsões. Foram utilizados dados de janeiro/2002 a outubro/2005 para FLN e de julho/2004 a outubro/2005 para SMS. Diversos conjuntos de preditores, constituídos com base em considerações físicas e estatísticas, foram testados no ajuste e simulação de modelos RNA e RLM. Parâmetros de avaliação de erros, determinados frente aos dados observacionais de cada estação de medida, foram calculados para cada previsão, a cada teste, permitindo a comparação de RNAs e RLMs entre si, e com a previsão de radiação solar diretamente determinada pelo modelo Eta. Visando maximizar o ganho de desempenho sobre o modelo Eta e minimizar o número de variáveis, encontrou-se um grupo de oito preditores, com o qual análises mais aprofundadas foram realizadas, incluindo avaliação de desempenho em cada estação do ano e avaliação de previsões com maiores antecedências. Para todos os conjuntos de preditores testados, erros de previsão muito semelhantes para RNAs e RLMs foram observados, indicando, portanto, não haver diferenças significativas de desempenho entre os dois métodos. Ambos os métodos, ao empregar o grupo de oito preditores, promoveram redução do viés e do RMSE em comparação com a previsão de radiação solar do modelo Eta. Confrontando previsões de irradiação solar integradas de 12:00UT a 18:00UT de cada dia, verificou-se que, enquanto o modelo Eta apresenta valores de viés superiores a 25%, os métodos RNA e RLM conduzem a valores de viés inferiores a 2%. Em termos de RMSE relativo, as previsões do modelo Eta produzem erros maiores que 40%, enquanto os métodos RNA e RLM apresentam erros entre 26 e 28%. A avaliação do *Skill* em termos do RMSE, aponta para ganhos superiores a 30% das novas previsões sobre as previsões do modelo Eta. Esta elevação de desempenho através do refinamento também é verificada nos coeficientes de correlação: de 0,72 (Eta) para 0,80 (RNA e RLM) em FLN, e de 0,78 (Eta) para 0,85 (RNA e RLM) em SMS.

ARTIFICIAL NEURAL NETWORKS AND MULTIPLE LINEAR REGRESSIONS APPLIED TO REFINING THE SOLAR RADIATION FORECAST PROVIDED BY THE ETA MODEL

ABSTRACT

This study aims to develop and evaluate a methodology to increase the accuracy of solar radiation forecasts generated by meteorological mesoscale models. The work is therefore an application of meteorological concepts and tools to meet the needs of the electricity production and distribution sector. The process is based on refining the numerical weather predictions yielded by the Eta/CPTEC model. Eta model output data, representing the forecasted atmospheric conditions, were used as predictors by Artificial Neural Networks (ANNs) and Multiple Linear Regressions (MLRs). These refining models were adjusted for calculation of the incident solar radiation at two sites where ground data from the SONDA project were available: Florianópolis (FLN) and São Martinho da Serra (SMS). The measurements of global solar radiation acquired in FLN and SMS were used for predictor's selection, for training/fitting of ANNs and MLRs and for forecasts evaluation. This study included data for FLN from January/2002 to October/2005, and for SMS from July/2004 to October/2005. Several sets of predictors were tested in the adjustment and simulation of ANN and MLR models. Error evaluation parameters were calculated for each forecast and test. These parameters allowed the comparison among solar radiation forecasts provided by ANNs, MLRs, and Eta model. In order to maximize the improvement over the Eta model forecasts, and to minimize the number of variables to be processed, a set of eight predictors were chosen. This predictors' set was employed in the performance evaluation for each season and evaluation of forecasts generated with more time in advance. Forecast errors for ANNs and MLRs were very close, for all sets of tested predictors, indicating high similarities between the two methods. The ANN and MLR, using the selected set of eight predictors, provided solar radiation estimates with lower bias and RMSE than the ones provided by the unadjusted Eta model. While Eta model forecasts presented bias higher than 25%, ANNs and MLRs provided forecasts with bias lower than 2%, for the solar radiation integrals between 12:00UT and 18:00UT. Concerning to the RMSE, the Eta model forecasts presented errors higher than 40%, while the ANNs and MLRs led to errors between 26% and 28%. The *skill*, based on RMSE, indicates improvements higher than 30% in ANN and MLR forecasts over Eta model forecasts. This increase in performance is also demonstrated by the correlation coefficients rising: from 0.72 (Eta) to 0.80 (ANN and MLR) for FLN and from 0.78 (Eta) to 0.85 (ANN and MLR) for SMS.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS

LISTA DE SIGLAS E ABREVIATURAS

| | |
|---|-----------|
| CAPÍTULO 1 – INTRODUÇÃO E OBJETIVOS..... | 25 |
| 1.1. Justificativa para os Estudos de Energia Solar | 26 |
| 1.2. Objetivos..... | 30 |
| 1.3. Organização da Dissertação | 32 |
| | |
| CAPÍTULO 2 – RADIAÇÃO SOLAR E A ATMOSFERA TERRESTRE | 33 |
| 2.1. Descrição Quantitativa da Radiação | 33 |
| 2.2. Leis da Radiação..... | 35 |
| 2.3. Radiação Solar e Radiação Terrestre | 37 |
| 2.4. Interações da Radiação com a Atmosfera Terrestre..... | 38 |
| 2.5. Radiação Solar no Topo da Atmosfera | 42 |
| 2.6. Equação de Transferência Radiativa..... | 46 |
| 2.7. Radiação Solar e Nuvens..... | 47 |
| 2.8. Modelagem da Radiação Solar Incidente..... | 50 |
| | |
| CAPÍTULO 3 – REDES NEURAIS ARTIFICIAIS..... | 53 |
| 3.1. Descrição Básica | 53 |
| 3.2. Histórico | 54 |
| 3.3. Perceptrons de Múltiplas Camadas | 55 |
| 3.4. Modelo de Neurônio Artificial..... | 56 |
| 3.5. Aprendizado de RNAs..... | 58 |
| 3.5.1. Backpropagation (Retropropagação)..... | 60 |
| 3.5.2. Algoritmo Rprop | 62 |
| 3.6. Generalização..... | 63 |
| 3.7. Modelagem de Radiação Solar utilizando RNAs..... | 64 |

| | |
|--|------------|
| CAPÍTULO 4 – DADOS: MODELO ETA E PROJETO SONDA..... | 67 |
| 4.1. O Modelo Eta/CPTEC | 67 |
| 4.1.1. Descrição do Modelo Eta | 68 |
| 4.1.2. Parametrização de Radiação | 70 |
| 4.1.3. Variáveis de Saída do Modelo Eta/CPTEC | 71 |
| 4.2. O Projeto SONDA: Estações SMS e FLN | 74 |
| | |
| CAPÍTULO 5 – METODOLOGIA | 79 |
| 5.1. Extração de Dados de Saída do Modelo Eta | 79 |
| 5.2. Dados Derivados de Saídas do Modelo Eta | 83 |
| 5.3. Qualificação e Integração dos Dados de Radiação Solar | 83 |
| 5.4. Conjuntos de Dados Disponíveis | 85 |
| 5.5. Treinamento das RNAs e Ajustes de RLMs..... | 92 |
| 5.6. Estudos Planejados | 94 |
| 5.7. Definição dos Grupos de Preditores | 95 |
| 5.8. Avaliação das Previsões | 97 |
| | |
| CAPÍTULO 6 – RESULTADOS | 101 |
| 6.1. Avaliação da Previsão de Radiação Solar do Modelo Eta | 101 |
| 6.2. Experimentos: Testes de Grupos de Preditores..... | 108 |
| 6.2.1. Previsões Eta e por Persistência para Subconjuntos de Testes..... | 109 |
| 6.2.2. RNAs e RLMs Baseadas em Grupos Extensos de Preditores | 114 |
| 6.2.3. Otimização dos Grupos de Preditores | 118 |
| 6.2.4. Comparações entre as Previsões..... | 130 |
| 6.2.5. Teste de Generalização de RLMs e RNAs | 135 |
| 6.3. Avaliação das Previsões para cada Estação do Ano..... | 136 |
| 6.4. Previsões de Integral Diária de Radiação | 141 |
| 6.5. Análise de Previsões empregando Subdivisão Não-Homogênea | 144 |
| 6.6. Avaliação da Antecedência da Previsão | 149 |
| | |
| CAPÍTULO 7 – CONCLUSÕES | 155 |
| | |
| CAPÍTULO 8 – SUGESTÕES DE TRABALHOS FUTUROS | 161 |
| | |
| REFERÊNCIAS BIBLIOGRÁFICAS | 165 |

LISTA DE FIGURAS

| | |
|---|-----|
| 1.1 – Matrizes de energia elétrica mundial e brasileira. | 28 |
| 2.1 – Espectro solar fora da atmosfera, comparado com o espectro de um corpo negro a 6000 K e com o espectro solar no nível do mar. | 38 |
| 2.2 – Absorção de radiação na atmosfera. (a) Curvas de corpo negro para a radiação solar (6000K) e radiação terrestre (255K). (b) Espectro de absorção para toda a extensão vertical da atmosfera. (c) Espectro de absorção para a atmosfera acima de 11 km. (d) Espectros de absorção para vários gases atmosféricos entre o topo da atmosfera e a superfície da Terra. | 40 |
| 2.3 – A órbita da Terra em torno do Sol, a obliquidade do eixo de rotação da Terra e as estações do ano. | 44 |
| 2.4 – Massa de ar para uma atmosfera esférica homogênea. | 46 |
| 2.5 – Elemento de volume diferencial contendo material que altera um feixe de radiação que o atravessa. | 47 |
| 2.6 – Distribuições de tamanho de partículas para alguns tipos de nuvem. | 49 |
| 2.7 – Comparação de incertezas na estimativa de radiação solar, entre o emprego de modelo baseado em satélite e redes de radiômetros com diferentes distâncias entre os equipamentos. | 50 |
| 3.1 – Neurônio de McCulloch e Pitts. | 54 |
| 3.2 – Exemplo de uma RNA-MLP com duas camadas intermediárias. | 56 |
| 3.3 – Modelo de um neurônio de uma rede MLP. | 57 |
| 3.4 – Comportamento dos sinais de saída de diferentes funções de ativação. | 58 |
| 4.1 – Grade horizontal tipo E de Arakawa. | 68 |
| 4.2 – Seção vertical idealizada da topografia do modelo Eta e distribuição das variáveis na vertical. | 69 |
| 4.3 – Adicionamento de duas camadas de espessuras óticas δ_1 e δ_2 | 71 |
| 4.4 – Estações do Projeto SONDA. | 75 |
| 4.5 – Plataforma radiométrica do Observatório Espacial do Sul, suportando os equipamentos do projeto SONDA e os sensores de radiação ultravioleta e ozônio do PMOA. | 77 |
| 5.1 – Área do modelo Eta e área de extração dos dados. | 80 |
| 5.2 – Imagens do satélite Landsat mostrando a localização do sítio da estação SMS. Na imagem maior são mostrados os pontos de grade do Eta e suas respectivas “caixas” | 81 |
| 5.3 – Imagens do satélite Landsat mostrando a localização do sítio da estação FLN. Na imagem maior são mostrados os pontos de grade do Eta e suas respectivas “caixas” | 82 |
| 5.4 – Transformações aplicadas aos dados do SONDA e do modelo Eta. | 87 |
| 5.5 – Agrupamento dos dados de radiação com previsões de diferentes antecedências. | 88 |
| 5.6 – Integrais de energia radiante para dias ensolarados. | 91 |
| 6.1 – Diagramas de dispersão para as previsões P00UT de radiação solar incidente do modelo Eta para três períodos diários e integral diária, para FLN. | 104 |

| | | |
|-------------|---|-----|
| 6.2 | – Diagramas de Dispersão para as previsões de radiação solar incidente do modelo Eta P00UT para três períodos diários e integral diária, para SMS..... | 107 |
| 6.3 | – Diagramas de dispersão para as previsões de radiação solar por persistência e calculadas pelo modelo Eta (<i>ocis</i>), para FLN e SMS..... | 111 |
| 6.4 | – Diagramas de dispersão para as previsões derivadas da previsão de radiação solar do modelo Eta (EtaA e EtaB), para FLN e SMS..... | 113 |
| 6.5 | – Diagramas de dispersão para previsões por RLM e RNA do Experimento A. | 115 |
| 6.6 | – Diagramas de dispersão para previsões por RLM e RNA do Experimento 7. | 128 |
| 6.7 | – Gráficos comparativos de ME%, RMSE%, R e R ² , para as previsões de radiação solar avaliadas, para FLN e SMS. | 131 |
| 6.8 | – Diagramas de caixa (<i>box plots</i>) para observações e previsões de radiação solar P00UT-Rad18UT para FLN e SMS. | 132 |
| 6.9 | – Histogramas indicando o número de ocorrências de desvios entre previsões e observações em cada categoria de 2MJ/m ² de intervalo, para FLN e SMS. | 133 |
| 6.10 | – Séries temporais de observações e previsões de radiação solar Rad18UT compreendendo os dias dos subconjuntos de testes de FLN e SMS pertencentes aos períodos de Verão 2004/2005 e Inverno 2005. | 134 |
| 6.11 | – Coeficientes de Correlação calculados para várias previsões (P00UT-Rad18UT – subconjunto de testes) confrontadas com observações, em cada estação do ano, para FLN..... | 137 |
| 6.12 | – ME% para várias previsões (P00UT- Rad18UT – subconjunto de testes) confrontadas com observações, em cada estação do ano, para FLN..... | 139 |
| 6.13 | – RMSE% para várias previsões (P00UT-Rad18UT – subconjunto de testes) confrontadas com observações, em cada estação do ano, para FLN..... | 139 |
| 6.14 | – Skill(RMSE,Eta) para várias previsões (P00UT-Rad18UT – subconjunto de testes) confrontadas com observações, em cada estação do ano, para FLN..... | 140 |
| 6.15 | – Diagramas de dispersão para previsões (P00UT) de integrais diárias de energia solar (Eta, RLM e RNA), para FLN e SMS. | 143 |
| 6.16 | – Histogramas ilustrando as subdivisões homogênea e não-homogênea dos dados P00UT-Rad18UT de FLN e SMS em subconjuntos de treinamento, validação e testes. | 145 |
| 6.17 | – Diagramas de dispersão para previsões Eta, RLM e RNA, geradas com base na subdivisão não-homogênea, para FLN e SMS. | 148 |
| 6.18 | – Variação dos coeficientes de correlação entre previsões e observações para previsões (<i>Phh</i> UT-Rad18UT) com diferentes antecedências, para FLN e SMS.. | 151 |
| 6.19 | – Variação dos coeficientes de determinação entre previsões e observações para previsões com diferentes antecedências, para FLN e SMS. | 152 |
| 6.20 | - Erros médios percentuais (ME% e RMSE%) para previsões (<i>Phh</i> UT-Rad18UT) com diferentes antecedências, para FLN e SMS. | 153 |

LISTA DE TABELAS

| | |
|--|-----|
| 2.1 – Altitudes aproximadas de nuvens (km)..... | 48 |
| 4.1 – Variáveis de saída do modelo Eta/CPTEC candidatas a preditores de radiação solar..... | 73 |
| 5.1 – Preditando e preditores a serem utilizados no ajuste dos modelos..... | 90 |
| 6.1 – Parâmetros de avaliação relativos aos diagramas de dispersão da Figura 6.1 – Florianópolis (FLN)..... | 104 |
| 6.2 – Parâmetros de avaliação relativos aos diagramas de dispersão da Figura 6.2 – São Martinho da Serra (SMS)..... | 107 |
| 6.3 – Quantidade de dados e média das observações nos subconjuntos de testes para FLN e SMS (P00UT-Rad18UT)..... | 109 |
| 6.4 – Parâmetros de avaliação das previsões (P00UT-Rad18UT – subconj. de testes) de radiação por persistência e pelo modelo Eta (<i>ocis</i>), para FLN e SMS..... | 111 |
| 6.5 – Parâmetros de avaliação das previsões Eta (<i>ocis</i>), EtaA e EtaB, frente radiação medida, para FLN e SMS (P00UT-Rad18UT – subconj. de testes)..... | 113 |
| 6.6 – Parâmetros de avaliação das previsões por RLM e RNA frente radiação medida, utilizando todos os 36 preditores da Tabela 5.1 (Experimento A)..... | 115 |
| 6.7 – Skill em termos de RMSE, calculados para os modelos RLM e RNA do Experimento A, com relação a outras previsões, para as estações FLN e SMS... | 116 |
| 6.8 – Parâmetros de avaliação calculados para os Experimentos A a F..... | 117 |
| 6.9 – Regressões Múltiplas Passo-a-passo para FLN (P00UT-Rad18UT)..... | 120 |
| 6.10 – Regressões Múltiplas Passo-a-passo para SMS (P00UT-Rad18UT)..... | 120 |
| 6.11 – Parâmetros de avaliação calculados para os Experimentos 1 a 5..... | 124 |
| 6.12 – Parâmetros de avaliação calculados para os Experimentos 6 a 9..... | 126 |
| 6.13 – Parâmetros de avaliação das previsões por RLM e RNA frente radiação medida, para o Experimento 7..... | 128 |
| 6.14 – Skill em termos de RMSE, calculados para os modelos RLM e RNA do Experimento 7, com relação a outras previsões, para as estações FLN e SMS.... | 129 |
| 6.15 – Resultados de teste de previsão de radiação solar para cada estação empregando RLMs e RNAs treinadas com dados da outra estação..... | 135 |
| 6.16 – Parâmetros de avaliação para as previsões Eta, RLM e RNA, para três períodos diários e integral diária, para FLN e SMS..... | 142 |
| 6.17 – Parâmetros de avaliação das previsões Eta, RLM e RNA, frente radiação medida, utilizando conjunto de testes da subdivisão não-homogênea de P00UT-Rad18UT..... | 146 |

LISTA DE SÍMBOLOS

Radiação Solar (Capítulo 2)

- λ - Comprimento de Onda. Unidade: μm . Como subscrito de outro símbolo: denomina grandeza monocromática ou espectral.
- θ - Ângulo zenital. Unidade: graus ou radianos.
- ϕ - Ângulo azimutal. Unidade: graus ou radianos.
- Ω - Ângulo Sólido. Unidade: sr .
- E - Densidade de Fluxo Radiante: emitância ou irradiância. Unidade: $W \cdot m^{-2}$.
- E_λ - Irradiância monocromática ou espectral. Unidade: $W \cdot m^{-2} \cdot \mu m^{-1}$.
- L - Radiância. Unidade: $W \cdot m^{-2} \cdot sr^{-1}$.
- L_λ - Radiância monocromática ou espectral. Unidade: $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$.
- B - Radiância teoricamente emitida por um corpo negro, segundo a função de Planck.
- $*$ - Como sobrescrito de outro símbolo: denomina grandeza relativa a corpo negro.
- T - Temperatura. Unidades: $^{\circ}C$ ou K .
- S - Constante Solar ($S = 1368 W \cdot m^{-2}$).
- χ - Parâmetro de tamanho para espalhamento.
- d - Distância Terra-Sol.
- d_m - Distância média Terra-Sol.
- E_0 - Irradiância solar incidente em plano perpendicular à direção do feixe solar no TDA.
- Γ - Ângulo Diário
- θ_s - Ângulo zenital solar.
- E_h - Irradiância solar incidente em superfície horizontal no TDA.
- δ - Ângulo de declinação solar.
- h - Ângulo horário.

Redes Neurais Artificiais (Capítulo 3)

- θ - Limiar de um neurônio com saída binária.
- x_i - Entrada de um neurônio.
- $w_{j,i}$ - Peso sináptico de um neurônio j , que pondera uma entrada x_i .
- y_j - Saída de um neurônio j .
- b_j - Limiar, ou *bias*, de um neurônio j .
- v_j - Nível de atividade interna de um neurônio j .
- φ - Função de ativação.
- t - Iteração da fase de treinamento.
- e_j - Erro entre a saída de um neurônio j e sua saída desejada.

- d_j - Valor-alvo ou saída desejada para o neurônio j .
- Δw_{ji} - Ajuste a ser aplicado ao peso $w_{j,i}$ numa dada iteração do treinamento.
- E - Função de custo.
- η - Taxa de aprendizado.
- \vec{w} - Vetor contendo todos os pesos e *bias* de um neurônio ou de uma RNA.
- δ_j - Gradiente local para um neurônio j .
- λ - Parâmetro de regularização.

Outros

- η - Coordenada vertical “eta”, empregada pelo modelo Eta.
- p - Pressão.
- z - Altitude.
- u/v - Vento ou as componentes zonal (u) e meridional (v) do vento.
- T - Temperatura.
- δ - Espessura ótica.
- e_s - Pressão de vapor de saturação do ar em relação à água.

Tabela 4.1 (pág. 73): símbolos das variáveis de saída do modelo Eta/CPTEC.

Tabela 5.1 (pág. 90): símbolos de todos os preditores empregados nas análises.

Unidades

| | | |
|-----------------------------|---|--|
| m | - metro (unidade de comprimento no sistema SI) | |
| kg | - quilograma (unidade de massa no sistema SI) | |
| s | - segundos (unidade de tempo no sistema SI) | |
| K | - Kelvin (unidade de temperatura no sistema SI) | |
| N | - Newton (unidade de força no SI) | $N = kg \cdot m \cdot s^{-2}$ |
| Pa | - Pascal (unidade de pressão no SI) | $Pa = N \cdot m^{-2} = kg \cdot m^{-1} \cdot s^{-2}$ |
| J | - Joules (unidade de energia no sistema SI) | $J = N \cdot m = kg \cdot m^2 \cdot s^{-2}$ |
| W | - Watt (unidade de potência no sistema SI) | $W = J \cdot s^{-1} = kg \cdot m^2 \cdot s^{-3}$ |
| sr | - esterradianos (unidade de ângulo sólido) | |
| $^{\circ}C$ | - graus Celsius (unidade de temperatura) | $^{\circ}C = K - 273,15$ |
| h | - horas (unidade de tempo) | $h = 3600 s$ |
| $W \cdot h$ | - Watt-hora (unidade de energia) | $W \cdot h = 3600 J$ |
| mb | - milibar (unidade de pressão) | $mb = 10^2 Pa$ |
| <i>Prefixos empregados:</i> | | |
| | c (<i>centi</i> – 10^{-2}) | m (<i>mili</i> – 10^{-3}) |
| | h (<i>hecto</i> – 10^2) | k (<i>quilo</i> – 10^3) |
| | | μ (<i>micro</i> – 10^{-6}) |
| | | M (<i>mega</i> – 10^6) |

LISTA DE SIGLAS E ABREVIATURAS

| | |
|----------|--|
| ADALINE | - <i>ADaptive LInear NEuron</i> (Neurônio Linear Adaptativo). |
| BSRN | - <i>Baseline Surface Radiation Network</i> . |
| CPTEC | - Centro de Previsão de Tempo e Estudos Climáticos (CPTEC/INPE). |
| FINEP | - Financiadora de Estudos e Projetos. |
| FLN | - Estação SONDA de Florianópolis (FLN). |
| GRADS | - <i>Grid Analysis and Display System (Software)</i> . |
| IR | - <i>Infrared Radiation</i> (Radiação Infravermelha). |
| LABSOLAR | - Laboratório de Energia Solar (LABSOLAR/UFSC). |
| LACESM | - Laboratório de Ciências Espaciais de Santa Maria (LACESM/UFSM). |
| MAE | - <i>Mean Absolute Error</i> (Erro Absoluto Médio). |
| ME | - <i>Mean Error</i> (Erro Sistemático Médio ou Viés). |
| MLP | - <i>Multilayer Perceptron</i> (RNA do tipo Perceptron de Múltiplas Camadas). |
| NCAR | - <i>National Center for Atmospheric Research</i> . |
| NCEP | - <i>National Centers for Environmental Prediction</i> . |
| NIR | - <i>Near Infrared Radiation</i> (Infravermelho Próximo). |
| OES | - Observatório Espacial do Sul (OES/RSU/INPE). |
| PhhUT | - Conjunto de preditores provenientes de previsões do modelo Eta geradas <i>hh</i> horas antes das 00UT de cada dia. |
| PMOA | - Programa de Monitoramento do Ozônio Atmosférico (UFSM). |
| PNT | - Previsão Numérica de Tempo. |
| R | - Coeficiente de Correlação. |

| | |
|-------------|---|
| R^2 | - Coeficiente de Determinação. |
| Rad hh UT | - Radiação solar integrada no período de 6 horas que antecede hh UT. |
| RLM(s) | - Regressão(ões) Linear(es) Múltipla(s). |
| RMSE | - <i>Root Mean Squared Error</i> (Raiz do Erro Quadrático Médio). |
| RNA(s) | - Rede(s) Neural(is) Artificial(is). |
| Rprop | - Algoritmo de treinamento <i>Resilient Backpropagation</i> . |
| RSU | - Unidade Regional Sul do INPE (RSU/INPE). |
| SI | - Sistema Internacional de Unidades. |
| SMS | - Estação SONDA de São Martinho da Serra (RS). |
| SONDA | - Sistema de Organização Nacional de Dados Ambientais para o setor de energia. |
| TDA | - Topo da Atmosfera. |
| UA | - Unidade Astronômica. |
| UFSC | - Universidade Federal de Santa Catarina. |
| UFMS | - Universidade Federal de Santa Maria. |
| UT | - <i>Universal Time</i> (Hora Universal). |
| UV | - <i>Ultraviolet Radiation</i> (Radiação Ultravioleta). |
| VIS | - <i>Visible Radiation</i> (Radiação Visível). |
| WMO | - <i>World Meteorological Organization</i> (Organização Mundial de Meteorologia). |

CAPÍTULO 1

INTRODUÇÃO E OBJETIVOS

O estudo da radiação solar incidente na superfície terrestre tem implicações diretas na meteorologia, especialmente nos estudos sobre o clima e suas mudanças. Informações oriundas do estudo da radiação solar são também importantes para atividades econômicas como a agropecuária, a arquitetura – mais especificamente as áreas de conforto térmico e de iluminação de ambientes – e diversos outros setores do conhecimento. O enfoque que irá ser dado nesse trabalho, contudo, contemplará a radiação solar como fonte de energia natural e renovável.

A energia solar vem ganhando importância nos últimos anos, acompanhando a crescente demanda energética mundial, as perspectivas de escassez dos combustíveis fósseis e as motivações ambientais ligadas ao aquecimento global. Embora seja uma fonte de energia abundante, tem a desvantagem de apresentar-se na forma disseminada (não-concentrada e de difícil captação, em comparação com a energia hidroelétrica) e possuir oferta inconstante na medida em que está sujeita à variabilidade climática, principalmente devido às nuvens e aos ciclos diurno e sazonal. Portanto, para ser aproveitada, são demandados estudos que possibilitem conhecer melhor sua disponibilidade regional, variabilidade temporal e previsibilidade.

Informações provenientes da *previsão* de radiação solar incidente são de vital importância na operação de *sistemas híbridos*¹ de geração de eletricidade, permitindo um gerenciamento mais eficiente das fontes de energia e o emprego otimizado de energia solar-fotovoltaica, em prol da economia de outros recursos. Além de permitir o conhecimento antecipado da disponibilidade de energia solar para geração fotovoltaica em locais isolados, as previsões de radiação solar permitem estimar a demanda por aquecimento resistivo complementar em sistemas solar-térmicos. Não obstante, estas

¹ Sistemas híbridos de energia (*hybrid power systems*) são sistemas autônomos de geração de energia elétrica a partir de duas ou mais fontes de energia. Em geral, combinam fontes de energia renováveis (solar ou eólica) e geradores convencionais (movidos a combustíveis fósseis). Nesses sistemas, objetiva-se produzir o máximo de eletricidade a partir das fontes renováveis, sendo o gerador convencional empregado na manutenção da disponibilidade e qualidade da energia elétrica.

previsões também encontram aplicações no setor de distribuição de energia, visto que o despacho de carga através das linhas de transmissão é ligado à temperatura dos cabos elétricos, em grande parte dependente da irradiação solar (além de outros fatores ambientais como vento e umidade).

O presente trabalho é um estudo sobre previsão de radiação solar incidente, tendo como objetivo principal produzir informações que atendam principalmente as atividades ligadas à geração e distribuição de energia. Assim, além de contribuir para um melhor aproveitamento de sua conversão em energia elétrica e térmica, uma previsão confiável da incidência de radiação solar vem a atender uma demanda do Operador Nacional do Sistema Elétrico, no que diz respeito à otimização do despacho de carga nas linhas de transmissão.

Com o intuito de verificar o balanço de energia atmosférico, modelos de previsão numérica de tempo calculam campos de radiação solar incidente, bem como outras componentes radiativas. O modelo Eta operacional no CPTEC/INPE, produz previsões para os horários sinóticos (00:00, 06:00, 12:00 e 18:00, no horário universal) de até 7 dias futuros, em que a radiação solar incidente é apresentada na forma de valores médios de seis horas, numa grade de 40×40 quilômetros. Contudo, estas previsões apresentam um conhecido viés, superestimando a intensidade da radiação incidente. Neste trabalho, objetiva-se utilizar as demais variáveis previstas pelo modelo Eta para obter previsões de radiação solar com menor erro, testando o emprego de Redes Neurais Artificiais (RNAs) comparativamente a modelos de Regressão Linear Múltipla (RLM).

1.1. Justificativa para os Estudos de Energia Solar

A demanda energética mundial primária (energia elétrica, automotora, industrial, etc.) depende quase totalmente dos combustíveis fósseis. O emprego dessas fontes de energia lança na atmosfera uma grande quantidade de poluentes, especialmente óxidos de carbono, relacionados à intensificação do efeito estufa. A preocupação mundial quanto às mudanças climáticas, devido à elevação dos níveis de carbono atmosférico e o conseqüente aquecimento global, tem sido objeto de acordos internacionais, como o

recentemente criado *Protocolo de Kyoto*². Contudo, ao lado da preocupação ambiental, existe uma preocupação estratégica decorrente da percepção de que os reservatórios destes combustíveis fósseis são limitados, isto é, trata-se de recursos de energia esgotáveis (não-renováveis).

A maior e mais importante fonte de energia convencional a partir do século passado tem sido, sem dúvida, o petróleo. Segundo Geller (2003), num cenário otimista, a depleção desse recurso energético deverá iniciar a partir de um pico de produção/demanda por volta de 2030 quando, então, terá início o seu esgotamento gradual a nível mundial. Contudo, estudos mais pessimistas antecipam esse pico de produção para 2010 e, portanto, antecipam o processo de esgotamento dos recursos (Bentley, 2002). Ainda segundo estudos compilados por Bentley, o declínio na produção do petróleo deverá ser superior a 3-5% ao ano após o seu apogeu de produção. Raciocínio análogo pode ser feito para as outras fontes não-renováveis de energia, como o carvão e o gás natural, mas obviamente com escalas bem diferentes de apogeu e declínio.

Assim, uma nova consciência mundial, quanto à utilização dos recursos naturais e energéticos, têm apontado para a inserção de fontes de energia alternativas, que não apenas apresentem reduzido impacto ambiental, mas que também venham a ser economicamente viáveis³ e que sejam renováveis. Entende-se por “energias renováveis” todas as formas de energia cuja taxa de utilização é inferior a sua taxa de renovação, podendo-se citar a energia solar, eólica, da biomassa, geotérmica e hidroelétrica. No caso do Brasil, sua matriz energética já contempla de forma exemplar o emprego das energias renováveis, principalmente através da hidroeletricidade e da biomassa (emprego do álcool no setor de transporte automotivo). A Figura 1.1 apresenta a matriz de energia elétrica do Brasil, comparativamente à matriz mundial.

² O Protocolo de Kyoto é um aditivo à Convenção do Tratado das Nações Unidas para Mudanças Climáticas (*United Nations Framework Convention on Climate Change - UNFCCC*), criado em 1997, com entrada em vigor em 2005. Países que ratificam este protocolo comprometem-se a reduzir suas emissões de dióxido de carbono e outros cinco gases do efeito estufa, ou assumem o compromisso de comercializar as emissões se houver manutenção ou aumento das taxas de emissão (Wikipedia, 2006).

³ Os cenários energéticos atuais levam em conta tanto os aspectos ambientais como econômicos, e apontam para uma crescente inserção, nas matrizes energéticas (principalmente dos países desenvolvidos), de fontes de energia alternativas e que apresentem maiores chances de se tornarem cada vez mais viáveis economicamente.

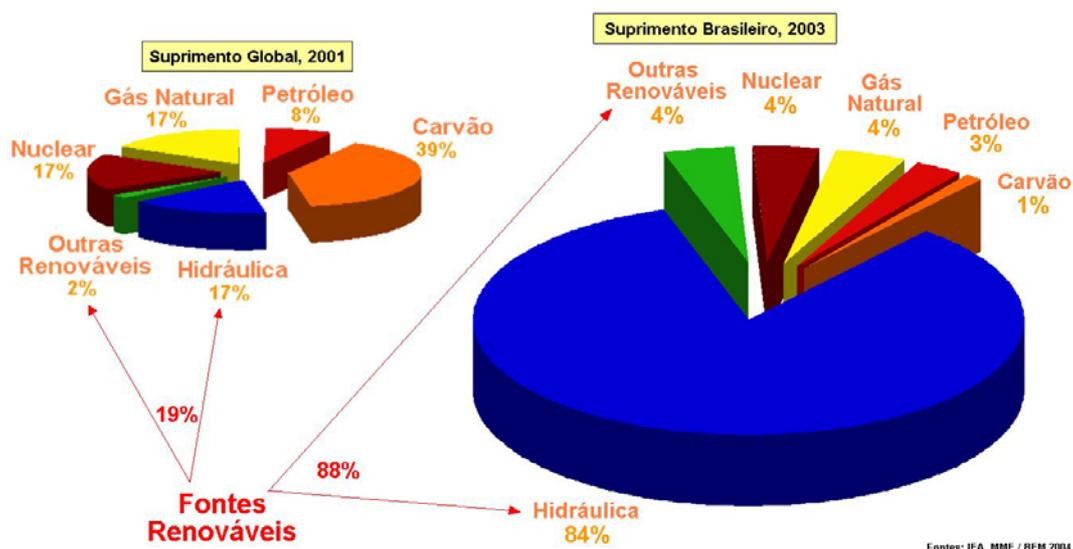


Figura 1.1: Matrizes de energia elétrica mundial e brasileira.
 Fonte dos dados: IEA (2001) e MME (2004).

Contudo, com o esperado crescimento da economia brasileira, haverá uma demanda proporcionalmente crescente por mais energia (Goldemberg e Villanueva, 2003). As energias renováveis solar e eólica representam importante opção, não só pelas vantagens inerentes à diversificação e complementaridade das fontes de energia, como também pelo fato de que haverá grande pressão mundial para estabilizar as taxas de emissão de carbono para a atmosfera, levando a reprimir o crescimento dos combustíveis fósseis.

A crise na geração de energia elétrica enfrentada pelo Brasil recentemente expôs fragilidades do sistema energético atual. Como apontado na Figura 1.1, a energia hidroelétrica possui papel central na matriz energética brasileira, no que diz respeito à geração de energia elétrica. Trata-se de uma forma de energia limpa quanto à emissão de poluentes para a atmosfera, embora a construção dos reservatórios ocasione alagamento de grandes áreas, afetando a fauna e a vegetação locais, e modificando o micro-clima regional. Porém, além desses impactos ambientais ainda não adequadamente avaliados, a dependência dessa forma de energia ao regime de chuvas cria vulnerabilidades em épocas de estiagem, como ocorreu no período de 1999 a 2001. Visando sustentar a oferta de energia a longo prazo e aumentar a confiabilidade do sistema, a reestruturação energética nacional atualmente em curso prevê a inclusão de recursos alternativos aos atualmente empregados.

A conversão de energia solar em energia elétrica ou térmica possui impactos reduzidos e emissão de poluentes nula, atendendo às preocupações de cunho ambiental – servindo ao Protocolo de Kyoto – e a demanda por fontes de energia alternativas e renováveis que complementem as fontes atuais. Contudo, além das barreiras tecnológicas ainda não solucionadas, quanto à sua variabilidade de oferta e dependência às condições de tempo e clima, outra barreira que se opõe à exploração e uso pleno desta abundante fonte energética é o custo relativamente alto (cerca de 5-10 vezes superior ao da energia hidroelétrica). Desta forma, no cenário atual, a energia solar e outras energias renováveis não serão empregadas em substituição aos combustíveis fósseis ou à energia hidráulica, mas serão gradativamente inseridas de forma complementar às atuais fontes, em sistemas energéticos híbridos.

As limitações ao uso da energia solar tendem a ser reduzidas ou até mesmo eliminadas com o tempo, principalmente em resposta ao crescimento da demanda de energia mundial, devido ao crescimento das economias de países em desenvolvimento (Brasil, Índia e China, principalmente), o aparecimento de novas tecnologias de armazenamento de energia e o crescimento da escala de mercado dessas tecnologias. Por ser um país localizado em sua maior parte na região intertropical, o Brasil possui grande potencial de energia solar ao longo de todo ano (Tiba, 2000; Colle e Pereira, 1998). Devido à possibilidade de geração elétrica descentralizada (não conectada à rede elétrica das distribuidoras de energia), através de painéis fotovoltaicos, o uso deste recurso viabilizaria o desenvolvimento de regiões remotas, onde o custo de eletrificação pela rede convencional é economicamente inviável.

Nesse contexto, o Governo Federal iniciou em 2004 o desafio de acabar com a exclusão elétrica no país e lançou o programa *Luz Para Todos*, que tem o objetivo de levar energia elétrica para 10 milhões de pessoas do meio rural até 2008, incluindo a energia solar-fotovoltaica nos locais mais remotos e distantes das linhas de transmissão (MME, 2006). Mas para que se possa otimizar os vultosos investimentos de implantação de estruturas coletoras de energia solar e sua exploração eficiente, faz-se necessário conhecer os locais com maior disponibilidade, os potenciais de utilização e as incertezas da variabilidade climática que influenciam este recurso de energia.

Contudo, o Brasil possui insuficiência de dados confiáveis que suportem esses estudos de viabilidade. Na tentativa de suprir tal carência, duas iniciativas do INPE, através do CPTEC merecem destaque: o emprego de modelos de radiação baseados em imagens de satélite, como os modelos BRASIL-SR (Pereira *et al.*, 1996; Martins, 2001) e GL (Ceballos *et al.*, 2004), e a ampliação da rede de estações medidoras de radiação solar, através de projetos como o SONDA (<http://www.cptec.inpe.br/sonda/>). Com a disponibilidade desses novos dados, estudos são demandados para que sejam produzidas informações e ferramentas úteis ao setor energético, incluindo modelos de previsão de curto prazo.

1.2. Objetivos

O objetivo deste trabalho é melhorar a capacidade de previsão da radiação solar incidente atualmente disponibilizada pelo modelo de previsão de tempo Eta/CPTEC operacional, visando atender a uma grande demanda do setor de produção e distribuição de energia elétrica no Brasil e no mundo. Tem como finalidades possibilitar o melhor aproveitamento da energia solar-elétrica em futuros projetos de exploração e comercialização dessa fonte renovável de energia, bem como prover meios de otimizar os processos de despacho de carga elétrica em linhas de transmissão. Este estudo visa contribuir para a implantação de uma ferramenta operacional destinada a fornecer essas previsões às concessionárias, geradoras e distribuidoras de energia elétrica e ao Operador Nacional do Sistema Elétrico. Trata-se, pois, de um trabalho de aplicação da meteorologia.

Para tal finalidade, são empregados modelos estatísticos simples alimentados com dados previstos diversos, gerados como saídas do modelo Eta, e que expressam as condições atmosféricas e da superfície para instantes futuros. No desenvolvimento e avaliação de tais modelos, além dos dados de saída do modelo Eta, são empregados dados de radiação medida em duas estações solarimétricas do projeto SONDA. O problema consiste, pois, em estabelecer relações entre os preditores (dados provenientes do modelo Eta) e o preditando (radiação solar medida nas estações SONDA), e expressá-las nos modelos.

Também se almeja verificar se, na solução específica do problema deste trabalho, o emprego da moderna técnica de Redes Neurais Artificiais (RNAs) oferece algum ganho frente aos modelos mais simples, ajustados através de Regressão Linear Múltipla (RLM). O emprego de RNAs é justificável pela sua conhecida capacidade de reconhecimento de padrões e ajuste funcional, e por proporcionar bons resultados em problemas que apresentem elevado grau de não-linearidades, muito embora não seja verificada a mesma eficiência para todos os problemas.

As metas principais para atingir os objetivos deste trabalho são:

- Tratamento dos dados de saída do modelo Eta e de radiação medida pelas estações SONDA de São Martinho da Serra (SMS) e Florianópolis (FLN), dispondo-os com igual resolução temporal, e constituindo conjuntos preditores/preditando.
- Verificação do viés e do erro quadrático médio da previsão pontual de radiação dada pelo modelo Eta, para as duas estações em estudo.
- Construção de modelos baseados em RNAs e RLMs, utilizando os dados gerados pelo modelo Eta, em sua previsão mais recente para cada dia, para as localidades das duas estações SONDA empregadas.
- Realização de experimentos com várias combinações de preditores, verificando qual o melhor e mais compacto conjunto de variáveis de saída do modelo Eta que permite obter estimativas de radiação solar pelos modelos de RNA e RLM.
- Comparação de desempenho das RNAs frente a RLMs, verificando se é justificável o uso de tal ferramenta estatística na tarefa a que este trabalho se propõe.
- Quantificação dos erros das previsões frente à radiação medida nas duas estações SONDA para os períodos de dados empregados.

- Estudo do crescimento do erro das previsões obtidas pelo modelo Eta, por RNAs e modelos ajustados por RLM com o aumento da antecedência da previsão.
- Avaliação do desempenho das previsões obtidas com os modelos ajustados em comparação com a previsão de radiação disponibilizada diretamente pelo modelo Eta, e quantificação do ganho obtido.

1.3. Organização da Dissertação

Esta dissertação é dividida em oito capítulos. Após esta introdução, alguns aspectos teóricos sobre radiação solar e interações com a atmosfera terrestre são apresentados no Capítulo 2. No Capítulo 3 são apresentadas noções gerais sobre redes neurais artificiais. O Capítulo 4 compreende informações sobre o modelo Eta e sobre o projeto SONDA, sendo descritos os dados empregados neste trabalho.

A seguir, no Capítulo 5, é apresentada a metodologia empregada, bem como os tratamentos realizados nos dados, e são indicadas as análises objetivadas. Uma avaliação do viés da previsão de radiação do modelo Eta, os resultados obtidos com os diferentes conjuntos de preditores utilizando RNAs e RLMs, estudos de previsões diversas utilizando o melhor conjunto de preditores e discussões dos resultados, são apresentados no Capítulo 6. O Capítulo 7 reúne as conclusões obtidas neste trabalho, e sugestões de trabalhos futuros são apresentadas no Capítulo 8.

CAPÍTULO 2

RADIAÇÃO SOLAR E A ATMOSFERA TERRESTRE

O Sol é a principal fonte da energia incidente no planeta Terra. A radiação solar constitui a força motriz para os movimentos atmosféricos e para outros processos térmicos, dinâmicos e químicos que ocorrem na atmosfera e na superfície terrestre (Brasseur e Solomon, 1986; Kondratyev, 1969).

A energia solar, na forma de luz, exhibe uma faixa contínua de comprimentos de onda (λ) no espectro eletromagnético, segundo uma lei conhecida como função de Planck, descrita a seguir, na seção 2.2. Por razões históricas e didáticas, o espectro luminoso é subdividido em regiões ou faixas espectrais que recebem nomes específicos, embora esses limites não sejam rigidamente definidos. Segundo Kondratyev (1969), a região da radiação visível (VIS) corresponde ao intervalo de comprimentos de onda entre 0,4 e 0,75 micrômetros (μm). A radiação ultravioleta (UV) é geralmente definida no intervalo de 0,01 a 0,4 μm (Coulson, 1975). Hudson (1969) define como radiação infravermelha (*infrared* – IR) os comprimentos de onda entre 0,75 a 1000 μm , e classifica o subintervalo entre 0,75 e 3 μm como *infravermelho próximo* (*near infrared* – NIR).

2.1. Descrição Quantitativa da Radiação

A taxa de transferência de energia pela radiação eletromagnética é chamada de *fluxo radiante*, expresso em unidades de energia por unidade de tempo: joules por segundo ($J \cdot s^{-1}$) ou watts (W). A *densidade de fluxo radiante* é o fluxo radiante que atravessa uma unidade de área, expressa em watts por metro quadrado ($W \cdot m^{-2}$), sendo também denominada *emitância* ou *irradiância* (E), quando representa o fluxo emergindo ou incidindo em uma área (Kidder e Vonder Haar, 1995).

A radiação é função da direção e essa dependência direcional é levada em conta empregando-se o *ângulo sólido* (Ω), expresso em *esterradianos* (sr). Definições formais e explicações didáticas acerca do ângulo sólido são encontradas em inúmeras

referências, podendo-se citar Kidder e Vonder Haar (1995). Salvo no caso de fontes luminosas pontuais, a irradiância incidente sobre um ponto representa contribuições oriundas de diferentes fontes e direções. Muitas vezes é necessário identificar a parte da irradiância que provém de direções dentro de um determinado arco infinitesimal de ângulo sólido $d\Omega$. Define-se, assim, a *radiância* (L), que é a irradiância por unidade de ângulo sólido, expressada em watts por metro quadrado por esterradiano ($W \cdot m^{-2} \cdot sr^{-1}$) (Wallace e Hobbs, 1977). Essa grandeza permite detalhar as porções de radiação provenientes de cada direção em torno de um ponto de observação.

A radiação pode atingir uma superfície plana com diferentes inclinações. O *ângulo zenital* (θ) é o ângulo entre a normal (ou, *zênite*) de uma superfície de incidência e a direção da radiação. Como a radiância representa a energia radiante que atravessa uma área perpendicular ao feixe de radiação, a componente da radiância na direção da normal da superfície de incidência é dada por $L \cdot \cos \theta$.

A irradiância incidente sobre uma superfície, como a superfície terrestre, representa o efeito combinado das componentes normais dos feixes de radiação provenientes de todas as direções do hemisfério acima da superfície. Um hemisfério equivale a um ângulo sólido de 2π sr. Assim, a relação entre a irradiância e a radiância é dada pela seguinte expressão:

$$E = \int_0^{2\pi} L(\theta, \phi) \cdot \cos \theta \cdot d\Omega \quad (2.1)$$

em que $L(\theta, \phi)$ é a radiância proveniente de direções representadas pelo ângulo zenital (θ) e pelo ângulo azimutal (ϕ).

Para expressar a intensidade da radiação em cada comprimento de onda do espectro eletromagnético, define-se a *radiância monocromática* (L_λ) e a *irradiância monocromática* (E_λ) – também denominadas *radiância espectral* e *irradiância espectral* – como sendo a radiância e a irradiância por intervalo unitário de comprimento de onda (Liou, 1980; Kidder e Vonder Haar, 1995). Tomando como unidade de comprimento de onda o micrômetro, as unidades para E_λ e L_λ são $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$ e $W \cdot m^{-2} \cdot \mu m^{-1}$,

respectivamente. A irradiância total, levando em conta todos os comprimentos de onda ($W \cdot m^{-2}$), é obtida pela integração de E_λ sobre todo o espectro eletromagnético.

A quantidade de energia radiante que atravessa uma superfície durante certo intervalo de tempo, por unidade de área ($J \cdot m^{-2}$), é a *irradiação* (Plana-Fattori e Ceballos, 2005), correspondendo à integração da irradiância ao longo do intervalo de tempo em questão. Muito embora o sistema internacional de unidades (SI) seja recomendado, comumente emprega-se como unidade de irradiação o watt-hora por metro quadrado ($W \cdot h \cdot m^{-2}$).

2.2. Leis da Radiação

Um corpo negro representa um corpo hipotético que absorve ou emite com máxima eficiência toda radiação eletromagnética em todas as direções e em todos os comprimentos de onda do espectro eletromagnético (Wallace e Hobbs, 1977). Planck derivou, de um ponto de vista teórico, a chamada *função de Planck* ou *lei de Planck*:

$$B_\lambda(T)^* = \frac{2hc^2}{\lambda^5 \left[\exp\left(\frac{hc}{\lambda kT}\right) - 1 \right]} \quad (2.2)$$

em que h é a constante de Planck, c é a velocidade da luz no vácuo, k é a constante de Boltzmann, T é a temperatura absoluta e λ é o comprimento de onda (Liou, 1980). A função de Planck quantifica a radiância monocromática, emitida em cada comprimento de onda, por um corpo negro a uma dada temperatura. Quanto maior a temperatura, maiores são os valores da radiância monocromática para todos os comprimentos de onda e menor é o comprimento de onda em que ocorre a máxima emissão. O comprimento de onda em que a emissão pelo corpo é máxima é dado por uma relação conhecida como *lei de deslocamento de Wien* (Kidder e Vonder Haar, 1995):

$$\lambda_m = \frac{2897,9}{T} \quad (2.3)$$

em que λ_m é dado em micrômetros (μm) e T em kelvins (K). A radiância total de um corpo negro é obtida pela integração de $B_\lambda(T)$ em todos os comprimentos de onda:

$$B(T)^* = \int_0^{\infty} B_{\lambda}(T) \cdot d\lambda = \left(\frac{2\pi^4 k^4}{15c^2 h^3} \right) T^4 \quad (2.4)$$

Como a radiância de um corpo negro é isotrópica, a sua emitância é dada pela seguinte expressão, conhecida como *lei de Stefan-Boltzmann*, em que σ é a constante de Stefan-Boltzmann (Liou, 1980):

$$E(T)^* = \pi \cdot B(T)^* = \sigma T^4 \quad (2.5)$$

Sendo a radiação de corpo negro a máxima emissão teoricamente possível para um corpo real a uma dada temperatura em um dado comprimento de onda, define-se a *emissividade* (ε_{λ}) de um corpo pela seguinte expressão (Wallace e Hobbs, 1977):

$$\varepsilon_{\lambda} = \frac{E_{\lambda}}{E_{\lambda}^*} \quad (2.6)$$

A emissividade de um corpo negro é unitária em todos os comprimentos de onda, sendo a emissividade de qualquer substância real compreendida entre 0 e 1, podendo variar com o comprimento de onda. Considerando, ao invés das emitâncias monocromáticas, as emitâncias totais dos corpos, define-se a *emissividade de corpo cinza* (ε):

$$\varepsilon = \frac{E}{E^*} = \frac{E}{\sigma T^4} \quad (2.7)$$

A *absortividade* (a_{λ}) e *absortividade de corpo cinza* (a) são definidas pela razão entre irradiância absorvida por um corpo e aquela incidente sobre ele, sendo a absortividade de um corpo negro igual à unidade para todos os comprimentos de onda. Define-se ainda a *refletividade* ρ_{λ} (ou *albedo*), e a *transmissividade* τ_{λ} (ou *transmitância*), como a fração da irradiância incidente que é refletida e transmitida, respectivamente.

A *lei de Kirchhoff* afirma que os materiais em equilíbrio termodinâmico local são tão bons absorvedores quanto são emissores de radiação em um comprimento de onda particular: $a_{\lambda} = \varepsilon_{\lambda}$. Com boa aproximação, a lei de Kirchhoff pode ser aplicada à atmosfera terrestre (Liou, 1980; Kidder e Vonder Haar, 1995).

2.3. Radiação Solar e Radiação Terrestre

O Sol emite um espectro contínuo de radiação (Figura 2.1), com uma distribuição espectral equivalente a de corpo negro a uma temperatura de aproximadamente 6000 K, de acordo com a lei de Plank para a emissão de radiação (Coulson, 1975; Brasseur and Solomon, 1986; Lenoble, 1993). A radiação solar cobre todo o espectro eletromagnético, mas a porção mais significativa do espectro, associada com transferência de energia radiativa no sistema climático, é compreendida entre o ultravioleta e o infravermelho próximo (Peixoto e Oort, 1992).

Acima de 99% da energia solar ocorre em comprimentos de onda inferiores a 4 μm , com o máximo de intensidade na região do visível, em aproximadamente 0,5 μm (Hoyt e Schatten, 1997), que corresponde à cor amarela. Segundo Lutgens e Tarbuck (1995), do total emitido pelo Sol, a radiação visível ($0,4 < \lambda < 0,7 \mu m$) representa 43%, o infravermelho próximo ($\lambda > 0,7 \mu m$) responde por 49%, a radiação ultravioleta ($\lambda < 0,4 \mu m$) por 7%, e menos de 1% é emitido como raios X, raios gama ou ondas de rádio.

A média anual da irradiância solar total (todos os comprimentos de onda) incidente no topo da atmosfera (TDA), em um plano perpendicular à direção da radiação, é conhecida historicamente como *constante solar* (S). Um valor de consenso é $S = 1368 \text{ W}\cdot\text{m}^{-2}$. Variações de $\pm 3,4\%$ em torno desse valor ocorrem ao longo do ano, devido à excentricidade da órbita da Terra em torno do Sol, sendo mais apropriado, portanto, referir-se a S como a *irradiância solar à distância média Terra-Sol*. Adicionalmente, S varia em $\pm 0,6 \text{ W}\cdot\text{m}^{-2}$ ao longo do ciclo solar de 11 anos (Kidder e Vonder Haar, 1995).

A emissão de radiação pela Terra ocorre em comprimentos de onda longos, entre 4 e 100 μm (Coulson, 1975), uma vez que as temperaturas da atmosfera e da superfície terrestre ($T < 300 \text{ K}$) são muito inferiores quando comparadas à temperatura da superfície do Sol ($T \cong 6000 \text{ K}$). O pico das emissões da superfície terrestre para a atmosfera ou da Terra como um todo para o espaço, ocorre entre 10 e 12 μm .

Costuma-se dividir o espectro eletromagnético em aproximadamente 4 μm (Coulson, 1975). A radiação menor que 4 μm é chamada *radiação de onda curta* ou *radiação*

solar. A radiação maior que $4 \mu\text{m}$ é chamada *radiação de onda longa*. A radiação de onda longa emitida pela Terra e pela atmosfera é chamada *radiação terrestre* (Iqbal, 1983), embora se utilize também o termo *radiação atmosférica*. Vale ressaltar que a emissão de radiação por cada porção de gás na atmosfera ocorre em todas as direções, sendo comum utilizar-se os termos *ascendente* e *descendente*, para expressar a radiação no sentido da superfície para o espaço e do espaço para a superfície, respectivamente.

2.4. Interações da Radiação com a Atmosfera Terrestre

A radiação solar sofre diversas interações com os constituintes atmosféricos, as nuvens e, finalmente, com a superfície do planeta. Basicamente, os processos mais importantes que influenciam a intensidade de radiação solar que atinge a superfície da Terra são a *absorção* e o *espalhamento* atmosféricos (Liou, 1980). A Figura 2.1 mostra o espectro de radiação solar que atinge o TDA e o espectro de radiação solar que atinge a superfície da Terra (linhas contínuas), em termos de irradiância espectral, permitindo verificar a atenuação sofrida pela radiação ao atravessar atmosfera (Robinson, 1966). Apenas para comparação, uma curva de emissão de um corpo negro a 6000 K também é apresentada (linha tracejada).

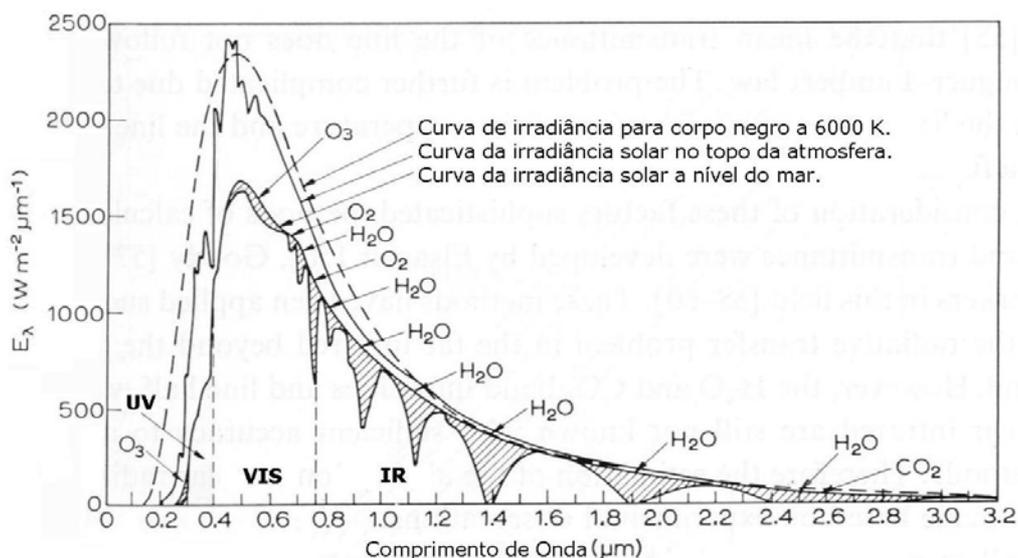


Figura 2.1: Espectro solar fora da atmosfera, comparado com o espectro de um corpo negro a 6000 K e com o espectro solar no nível do mar.
Fonte: Modificado de Robinson (1966).

A área entre as curvas (linhas contínuas) da Figura 2.1 representa a redução sofrida pela radiação solar incidente, durante sua passagem através da atmosfera. Essa redução é dividida em duas partes: a área não sombreada representa o espalhamento sofrido pela radiação e a área sombreada a absorção pelas moléculas do ar, primariamente por água (H₂O), dióxido de carbono (CO₂), ozônio (O₃) e oxigênio (O₂).

A radiação pode ser absorvida pelos os gases atmosféricos por meio de *interações de ionização-dissociação, transições eletrônicas, transições vibracionais e transições rotacionais* (Kidder e Vonder Haar, 1995). Dependendo do tipo de transição ou interação, a absorção da radiação pode ocorrer num *continuum* de comprimentos de onda, ou em comprimentos de onda discretos. Os gases atmosféricos absorvem diferentemente os comprimentos de onda da radiação solar e da radiação terrestre como se observa na Figura 2.2.

Devido à baixa eficiência de absorção dos comprimentos de onda da radiação visível (VIS), em caso de ausência de nuvens, a atmosfera é considerada transparente à radiação solar. As regiões do espectro onde a atmosfera é transparente à radiação são denominadas de “janelas espectrais”. Por outro lado, a absorção pelos gases atmosféricos é forte para a radiação de onda longa emitida pela superfície (radiação terrestre), de maneira que, na maioria dos comprimentos de onda do infravermelho (IR), a atmosfera é praticamente opaca. Nas Figuras 2.1 e 2.2 podem ser observadas faixas de radiação ultravioleta (UV) e infravermelha em que a atmosfera é completamente opaca.

Através do espalhamento de fótons de radiação pelas não-homogeneidades presentes em um meio propagante, a energia radiante transmitida numa dada direção pode ser reduzida ou aumentada. Estas não-homogeneidades são primeiramente as próprias moléculas, mas também as partículas maiores compostas desde poucas a muitas moléculas. Na atmosfera terrestre essas não-homogeneidades são basicamente as moléculas do ar, os aerossóis e as nuvens, sendo que o espalhamento da radiação ocorre devido à diferença entre o índice de refração das partículas e do meio propagante (Houghton, 1985). Segundo Liou (1980), o espalhamento ocorre em todos os comprimentos de onda do espectro eletromagnético.

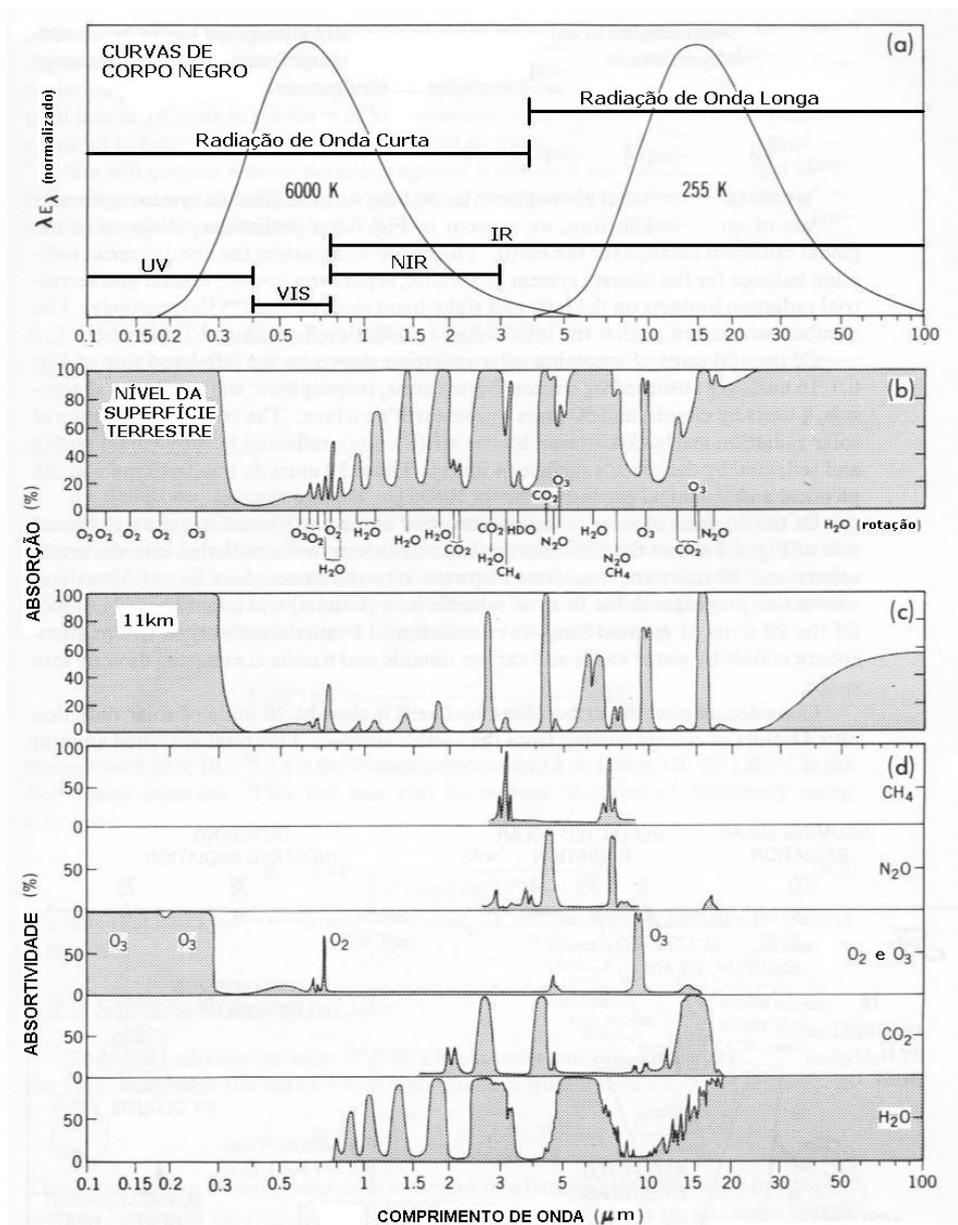


Figura 2.2: Absorção de radiação na atmosfera. (a) Curvas de corpo negro para a radiação solar (6000K) e radiação terrestre (255K). (b) Espectro de absorção para toda a extensão vertical da atmosfera. (c) Espectro de absorção para a atmosfera acima de 11 km. (d) Espectros de absorção para vários gases atmosféricos.
 Fonte: Modificado de Peixoto e Oort (1992).

Ao contrário do que ocorre na absorção, a energia radiante espalhada permanece na forma de radiação, sendo apenas removida da direção de incidência da radiação e dispersada em todas as direções. Uma vez que parte da energia solar é espalhada para trás e para os lados, a quantidade de energia que atinge a superfície da Terra, após atravessar a atmosfera, é atenuada (Peixoto e Oort, 1992). Na atmosfera, as partículas

responsáveis por espalhamento possuem tamanhos variáveis desde as moléculas de gás ($\sim 10^8 \text{ cm}$) até as grandes gotas de chuva e granizo ($\sim 1 \text{ cm}$) (Liou, 1980).

O regime de espalhamento da radiação é determinado pela razão entre o tamanho da partícula espalhadora, e o comprimento de onda da radiação incidente, expressa pelo *parâmetro de tamanho* (χ), $\chi = (2\pi \cdot r) / \lambda$, em que r é o raio da partícula, considerada esférica (Kidder e Vonder Haar, 1995). Quando as partículas são muito menores que o comprimento de onda ($\chi \ll 1$), o espalhamento é chamado *espalhamento Rayleigh*. Para partículas com tamanhos aproximadamente iguais ou maiores que o comprimento de onda ($0,1 < \chi < 50$), o espalhamento é referido como *espalhamento Mie*. Se o tamanho da partícula for muito maior que o comprimento de onda ($\chi > 50$), o desvio da radiação incidente é explicado pelos princípios da ótica geométrica (Liou, 1980).

O espalhamento da radiação solar visível e ultravioleta pelas moléculas do ar ocorre segundo o regime Rayleigh. Neste regime, o espalhamento é inversamente proporcional à quarta potência do comprimento de onda (*Espalhamento* $\sim \lambda^{-4}$), de maneira que o espalhamento é mais intenso para comprimentos de onda menores. A radiação espalhada pelo regime Rayleigh é uniformemente dividida entre pró-espalhamento (espalhamento “para frente” em relação ao plano perpendicular à direção de incidência) e retroespalhamento (espalhamento "para trás").

No regime Mie, a dispersão é menos seletiva ao comprimento de onda e ocorre predominância do pró-espalhamento sobre o retroespalhamento (van de Hulst, 1957). Ocorre, neste regime, o espalhamento da luz do Sol por partículas maiores, como aerossóis, neblina, fumaça, *smog* e poeira. O espalhamento da radiação visível pelas gotículas de nuvens, gotas de chuva e partículas de gelo, é descrito pelos princípios da ótica geométrica, havendo reflexão e refração (Kidder e Vonder Haar, 1995).

Após atravessar uma camada com agentes espalhadores, como é a atmosfera, a radiação de ondas curtas que atinge uma superfície pode ser subdividida entre uma componente proveniente diretamente da fonte e uma componente que engloba a radiação vinda de todas as demais direções devido ao espalhamento. Segundo Plana-Fattori e Ceballos (2005) podem-se definir as seguintes irradiâncias:

- **Irradiância difusa:** Irradiância descendente numa superfície horizontal, decorrente do espalhamento do feixe solar direto pelos constituintes atmosféricos (moléculas, material particulado, nuvens, etc.).
- **Irradiância normal:** Irradiância proveniente diretamente do Sol numa superfície normal à incidência do feixe direto. É a irradiância solar incidente no TDA (constante solar) que ainda resta no nível de observação, somados os efeitos de pró-espalhamento atmosférico na direção do feixe solar.
- **Irradiância direta:** Irradiância do feixe solar direto numa superfície horizontal. É o produto entre a irradiância normal e o co-seno do ângulo zenital solar.
- **Irradiância global:** Irradiância descendente numa superfície horizontal, constituída pelo somatório das irradiâncias direta e difusa.

2.5. Radiação Solar no Topo da Atmosfera

A Terra realiza órbitas elípticas em torno do Sol, estando o Sol situado no foco da elipse. A excentricidade da órbita da Terra é definida pela razão entre o máximo desvio a partir de uma órbita circular e o próprio raio médio, assumindo um valor atual de 0,0167 (Peixoto e Oort, 1992). A distância média Terra-Sol (d_m) é $1,496 \times 10^{11} \text{ m}$, e é conhecida como Unidade Astronômica (UA). A distância máxima Terra-Sol (*afélio*) é $1,521 \times 10^{11} \text{ m}$ (1,017 UA), e a distância mínima (*periélio*) é $1,471 \times 10^{11} \text{ m}$ (0,983 UA). Segundo Iqbal (1983), o periélio ocorre aproximadamente em 3 de Janeiro, o afélio em 4 de julho, e a distância média Terra-Sol ocorre em 4 de Abril e 5 de Outubro.

Devido às ligeiras variações da distância Terra-Sol ao longo do ano, decorrentes da excentricidade da órbita terrestre, a irradiância solar que atinge o TDA da Terra sofre pequenas alterações. Num dado instante, a irradiância incidente sobre um **plano perpendicular à direção do feixe de radiação solar**, situado no TDA, pode ser relacionada com a constante solar ($S = 1368 \text{ W} \cdot \text{m}^{-2}$) e com d_m pela seguinte expressão, em que d é a distância Terra-Sol no instante em questão (Liou, 1980; Hartmann, 1994):

$$E_0 = S \left(\frac{d_m}{d} \right)^2 \quad (2.8)$$

A relação $(d_m/d)^2$ é chamada *fator de correção de excentricidade*. Este fator pode ser calculado para cada dia do ano através de uma equação desenvolvida em termos de uma expansão em série de Fourier por Spencer (1971):

$$\left(\frac{d_m}{d}\right)^2 = 1,000110 + 0,034221\cos\Gamma + 0,001280\text{sen}\Gamma + \quad (2.9)$$

$$+ 0,000719\cos 2\Gamma + 0,000077\text{sen } 2\Gamma$$

em que Γ , em radianos, é chamado de *ângulo diário*, e é dado por:

$$\Gamma = \frac{2\pi(\text{dia} - 1)}{365} \quad (2.10)$$

em que *dia* é o dia do ano, variando de 1 (1 de janeiro) a 365 (31 de dezembro).

Contudo, devido ao formato aproximadamente esférico da Terra, as superfícies horizontais de incidência de radiação solar assumem diferentes orientações em relação à direção do fluxo radiante, em cada ponto do globo. O ângulo entre a direção de incidência da radiação solar sobre um ponto da esfera terrestre e o zênite (normal da superfície) deste local é denominado *ângulo zenital solar* (θ_s).

Enquanto, num dado instante, a radiação incide perpendicularmente sobre um ponto do globo ($\theta_s = 0^\circ$), para todos os demais pontos: $\theta_s > 0^\circ$. Ao incidir perpendicularmente, um feixe de radiação com área de seção transversal, A_0 , intercepta uma área $A = A_0$ da superfície de incidência. Contudo, ao incidir com $0^\circ < \theta_s < 90^\circ$, a energia contida no feixe de radiação solar se dispersa em uma área de incidência maior, equivalente a $A = A_0 \cdot (\cos(\theta_s))^{-1}$. Desta forma, quanto maior for θ_s , menor a quantidade de energia solar recebida por unidade de área da superfície de incidência.

Assim, a irradiância solar incidente sobre um **plano horizontal no TDA** (um plano paralelo à superfície terrestre abaixo), num dado instante, pode ser expressa por:

$$E_h = E_0 \cos(\theta_s) = S \left(\frac{d_m}{d}\right)^2 \cos(\theta_s) \quad (2.11)$$

em que θ_s é calculado para o instante e coordenadas geográficas do plano em questão.

O plano de revolução da Terra em torno do Sol é chamado *plano de eclíptica*. Além deste movimento de translação, a Terra realiza um movimento de rotação em torno de seu *eixo polar*, responsável pelo ciclo diurno da incidência de radiação em cada ponto do globo (*ciclo dia-noite*). O eixo de rotação apresenta uma inclinação em relação ao plano de eclíptica de $23^{\circ}27'$ que faz com que a orientação da Terra em relação ao Sol mude continuamente ao longo do ano, como se pode observar na Figura 2.3.

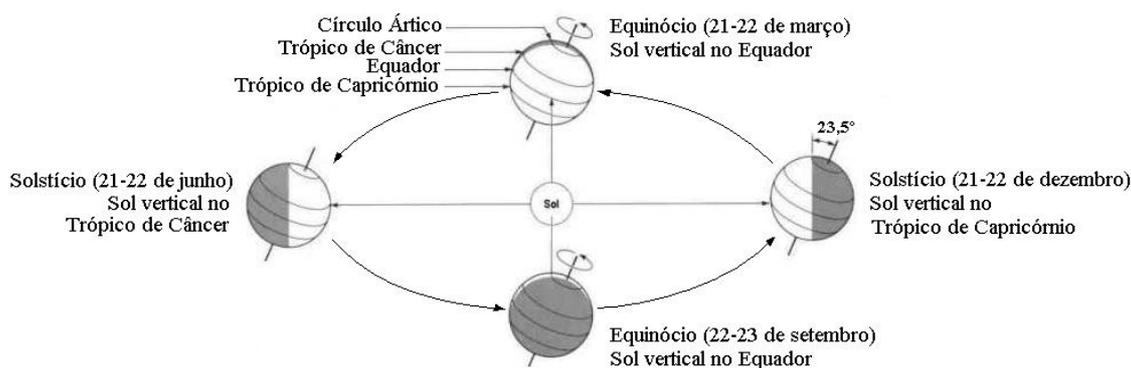


Figura 2.3: A órbita da Terra em torno do Sol, a obliquidade do eixo de rotação da Terra e as estações do ano.

Fonte: Modificado de Lutgens e Tarbuck (1995).

Esta mudança da orientação da Terra em relação ao Sol ao longo do ano ocasiona uma variação sazonal na incidência de radiação sobre cada latitude do globo terrestre, definindo as estações do ano. No dia 21 ou 22 de Junho, a Terra está orientada de maneira que a radiação solar incide perpendicularmente sobre o Trópico de Câncer (latitude $23^{\circ}27'$ N), ocorrendo o *solstício* de inverno para o Hemisfério Sul (de verão para o Hemisfério Norte). Em 21 ou 22 de Dezembro, a radiação solar incide verticalmente nos pontos sobre o Trópico de Capricórnio (latitude $23^{\circ}27'$ S), ocorrendo o solstício de verão para o Hemisfério Sul. Os *equinócios* ocorrem entre os solstícios, e a radiação solar atinge perpendicularmente o Equador. Para o Hemisfério Sul, os equinócios de outono e primavera ocorrem respectivamente nos dias 21 ou 22 de março e nos dias 22 ou 23 de setembro (Lutgens e Tarbuck, 1995).

O *ângulo de declinação do Sol* (δ) corresponde à latitude em que a radiação solar incide perpendicularmente, isto é, a latitude dos pontos da superfície terrestre nos quais a radiação solar incide com $\theta_s = 0^{\circ}$ ao meio-dia local. A declinação varia de $-23^{\circ}27'$

(23°27'S) a +23°27' (23°27'N) ao longo do ano, podendo ser calculada para cada dia do ano através da seguinte equação (Spencer, 1971):

$$\delta = (0,006918 - 0,399912 \cos \Gamma + 0,070257 \sin \Gamma - 0,006758 \cos 2\Gamma + 0,000907 \sin 2\Gamma - 0,002697 \cos 3\Gamma + 0,00148 \sin 3\Gamma) \cdot \left(\frac{180}{\pi} \right) \quad (2.12)$$

em que Γ , em radianos, é o *ângulo diário*, definido pela Equação 2.10.

Quanto mais distante, em termos de latitude, estiver um ponto em relação ao ângulo de declinação solar, maiores serão os valores de θ_s e, portanto, menor a incidência de radiação. O ciclo sazonal do ângulo de declinação solar é responsável pelas alterações na distribuição de irradiação sobre a atmosfera da Terra, ao longo do ano, bem como influi nas trajetórias aparentes do Sol no céu e na duração do período de dia e noite, de cada ponto da superfície terrestre e cada dia do ano.

Assim, o ângulo zenital solar é função da latitude do local (φ), da época do ano e do instante do dia, sendo a época do ano expressa pelo ângulo de declinação solar (δ). O instante do dia é expresso pelo *ângulo horário* (h), que assume valor zero no meio-dia local e aumenta 15° para cada hora antes ou depois do meio-dia local. Assim, o ângulo zenital solar pode ser expresso pela seguinte equação:

$$\cos(\theta_s) = \sin(\varphi)\sin(\delta) + \cos(\varphi)\cos(\delta)\cos(h) \quad (2.13)$$

Além de o ângulo zenital solar influir na intensidade da irradiação solar incidente sobre uma superfície horizontal no TDA, de acordo como descrito pela Equação 2.11, o θ_s também é relacionado com a atenuação proporcionada à radiação solar, pela atmosfera. Quanto maior o θ_s , maior é o caminho percorrido pela radiação solar ao longo da atmosfera, desde o TDA até atingir um dado ponto na superfície terrestre. Com um maior comprimento de caminho ótico, maior também é a frequência de interações da radiação com os elementos atmosféricos, e maior a atenuação por absorção e espalhamento. O efeito da atmosfera contribui, então, para que ainda menos radiação solar atinja a superfície para altos valores de θ_s , influenciando na sazonalidade da distribuição de irradiação que atinge a superfície terrestre.

O comprimento do caminho ótico é quantificado em termos da *massa de ar* (*air mass*), m_{ar} , comumente definida como a secante do ângulo zenital solar ao considerar-se a atmosfera plana (sendo esta uma boa aproximação para o caso de ângulos zenitais inferiores a 70°). Para o caso de uma atmosfera esférica e homogênea (Figura 2.4), utiliza-se uma equação apresentada por Robinson (1966), considerando uma espessura atmosférica de $H = 7991 \text{ km}$ e raio da Terra de $R = 6370 \text{ km}$:

$$m_{ar} = \frac{(BA)}{(CA)} = \left[\left(\frac{R}{H} \cdot \cos \theta_s \right)^2 + 2 \cdot \frac{R}{H} + 1 \right]^{\frac{1}{2}} - \frac{R}{H} \cdot \cos \theta_s \quad (2.14)$$

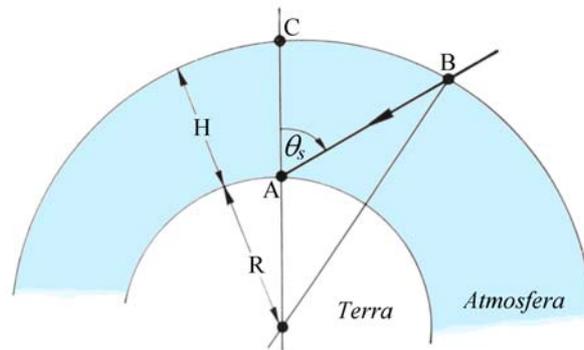


Figura 2.4: Massa de ar para uma atmosfera esférica homogênea.
Fonte: Modificado de Robinson (1966).

A massa de ar assume valor unitário, para o caso de $\theta_s = 0^\circ$, e valores tanto maiores quanto maior for θ_s , representando quantas vezes o caminho da radiação na atmosfera é maior do que o caminho que seria percorrido com $\theta_s = 0^\circ$.

2.6. Equação de Transferência Radiativa

Devido à diversidade de detalhes envolvidos e visando não prolongar este capítulo, o desenvolvimento de todos os termos da equação de transferência radiativa não será apresentado, recomendando-se a consulta dos seguintes trabalhos para deduções completas: Kidder e Vonder Haar (1995), Kondratyev (1969), Liou (1980), Peixoto e Oort (1992), Lenoble (1993).

Considerando a radiação incidindo num volume diferencial (Figura 2.5) preenchido por algum material, como o gás atmosférico, a equação de transferência radiativa trata das mudanças que ocorrem no feixe de luz, ao atravessar o volume. Uma vez que os efeitos

de absorção, emissão e espalhamento ocorrem diferentemente em cada comprimento de onda, utiliza-se a radiância espectral. Desconsiderando os efeitos de polarização, quatro processos podem alterar a luz ao longo da passagem pelo volume:

- A. Radiação do feixe pode ser absorvida pelo material.
- B. Radiação pode ser emitida pelo material, na direção do feixe.
- C. Radiação pode ser espalhada da direção do feixe para outras direções.
- D. Radiação de outras direções pode ser espalhada na direção do feixe.

A taxa de variação da radiância espectral com a distância, dL_λ / ds , consiste, assim, de quatro termos:

$$\frac{dL_\lambda}{ds} = A + B + C + D \quad (2.15)$$

Os termos A e C removem radiação da direção da radiação incidente (ou, do feixe de radiação) e são conhecidos como *termos de atenuação*. Os termos B e D adicionam radiação ao feixe, sendo conhecidos como *termos fonte*. A modelagem da radiação solar que incide na superfície terrestre pode ser realizada através da quantificação dos termos fonte e de atenuação, resolvendo assim, a equação de transferência radiativa. A resolução da equação de transferência radiativa através de diversos métodos computacionais é discutida na seção 2.8.

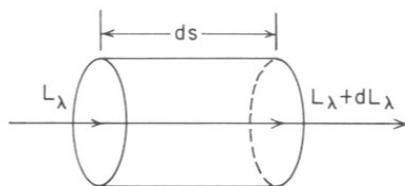


Figura 2.5: Elemento de volume diferencial contendo material que altera um feixe de radiação que o atravessa.
Fonte: Kidder e Vonder Haar (1995).

2.7. Radiação Solar e Nuvens

A todo instante, cerca de 50% da superfície da Terra é coberta por nuvens que ocorrem em altitudes variáveis desde a superfície até aproximadamente 20 km (Wallace e Hobbs, 1977). As nuvens possuem um papel importante para o balanço de radiação, assim como para o clima como um todo, afetando o albedo, a absortividade e a

transmissividade da radiação incidente. Os efeitos radiativos das nuvens no balanço energético do planeta vêm sendo extensivamente estudados (Iribarne e Cho, 1980; Liou, 1992; Harrison *et al.*, 1993; Hartmann, 1993). A utilização de satélites tem aumentado a qualidade e o detalhamento das informações sobre a distribuição das nuvens.

As nuvens formam-se quando o ar torna-se supersaturado com respeito à água líquida, ou ao gelo, geralmente através de ascensão de parcelas de ar, acompanhadas por expansão adiabática e resfriamento. O vapor supersaturado condensa-se na superfície de núcleos de condensação formando gotículas de nuvem (Wallace e Hobbs, 1977). Os aerossóis atuam como núcleos de condensação, e suas características influem na supersaturação necessária para que ocorra a condensação e formação da gotícula.

As nuvens são normalmente classificadas de acordo com características físicas como altitude e forma. Atualmente são dez os nomes básicos de nuvens empregados: *cirrus* (Ci), *cirrocumulus* (Cc), *cirrostratus* (Cs), *altocumulus* (Ac), *altostratus* (As), *nimbostratus* (Ns), *stratus* (St), *stratocumulus* (Sc), *cumulus* (Cu) e *cumulonimbus* (Cb). De acordo com Lenoble (1993) estes tipos de nuvens são classificados de acordo com a altitude, como apresentado na Tabela 2.1.

Tabela 2.1: Altitudes aproximadas de nuvens (km).

| Classificação | Tipo | Regiões Polares | Latitudes médias | Regiões Tropicais |
|----------------------|-------------|------------------------|-------------------------|--------------------------|
| Nuvens altas | Ci, Cc, Cs | 3-8 | 5-13 | 6-18 |
| Nuvens médias | Ac, As, Ns | 0,5-5 | 0,5-9 | 0,5-12 |
| Nuvens baixas | Sc, St | 0-2 | 0-2 | 0-2 |
| Nuvens convectivas | Cb,Cu | 0,5-8 | 0,5-13 | 0,5-18 |

Fonte: Lenoble (1993).

Propriedades radiativas detalhadas de nuvens não são bem conhecidas e mudam substancialmente com o tipo e a forma da nuvem. O albedo de uma nuvem *stratus* delgada é de aproximadamente 30%, varia entre 60 e 70% para *stratus* espessas, atinge 70% para *nimbostratus* e 20% para *cirrus* (Houghton, 1985). A espessura da nuvem e o ângulo zenital solar são os fatores que mais influenciam o seu albedo.

Dentro da nuvem, o espalhamento múltiplo é a forma de atenuação predominante da radiação solar, e uma nuvem de poucas dezenas de metros de espessura é suficiente para

espalhar toda a radiação visível incidente. A absorção de radiação visível é desprezível, pois a água líquida não absorve eficientemente estes comprimentos de onda. A maior parte da radiação emerge da nuvem em pontos diversos após ter sido espalhada várias vezes. Welch *et al.* (1980) efetuaram cálculos para nuvens *stratus* com 2 km de espessura, com radiação de $\lambda = 0,55 \mu m$ incidindo na vertical, e encontraram que 0,2% da radiação é absorvida, 79,8% é espalhada para fora da nuvem pelo topo e somente 20% é espalhada pela base em direção à superfície. Para comprimentos de onda maiores a absorção aumenta. Considerando uma média para todo espectro solar, verificou-se absorção de 10%, e espalhamento de 73,8% e 16,2% pelo topo e base, respectivamente.

Considerando suas propriedades óticas – albedo elevado e espalhamento eficiente – **as nuvens constituem o principal fator modulador da radiação solar que chega à superfície**. Segundo Paltridge e Platt (1976), as características radiativas das nuvens dependem da distribuição de tamanhos das partículas de água, do número de partículas por volume unitário, do conteúdo e estado físico da água, características estas que variam de acordo com o tipo de nuvem. Stephens (1978) apresenta valores típicos para o raio médio, o número de gotículas e o conteúdo de água líquida, bem como distribuições de tamanhos de partículas (Figura 2.6), para alguns tipos de nuvens, que podem ser empregados na modelagem da radiação solar.

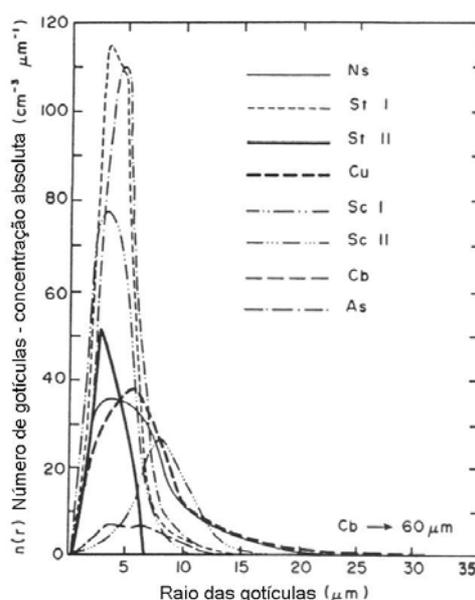


Figura 2.6: Distribuições de tamanho de partículas para alguns tipos de nuvem.
Fonte: Stephens (1978).

2.8. Modelagem da Radiação Solar Incidente

O estudo da radiação solar incidente sobre a superfície da Terra pode ser feito com a utilização de uma rede de radiômetros e de técnicas de interpolação das medidas de radiação. Contudo, as informações obtidas têm sua confiabilidade dependente de um grande número de radiômetros e de pequenas distâncias entre eles, inviabilizando tal metodologia em algumas situações. A adoção de modelos computacionais para obter estimativas da radiação solar constitui-se uma alternativa à metodologia anterior.

Perez *et al.* (1997) e Perez (2003) realizaram estudos sobre as incertezas de estimativas de radiação solar baseadas em interpolações de dados de estações radiométricas situadas a distâncias crescentes e estimativas baseadas em modelos computacionais empregando dados de satélites. A Figura 2.7 apresenta os valores dos erros quadráticos médios relativos (*relative root mean squared errors – RMSE%*) para estimativas por interpolação de medidas com diferentes distâncias e os níveis de erros encontrados para estimativas de modelos baseados em dados de satélites.

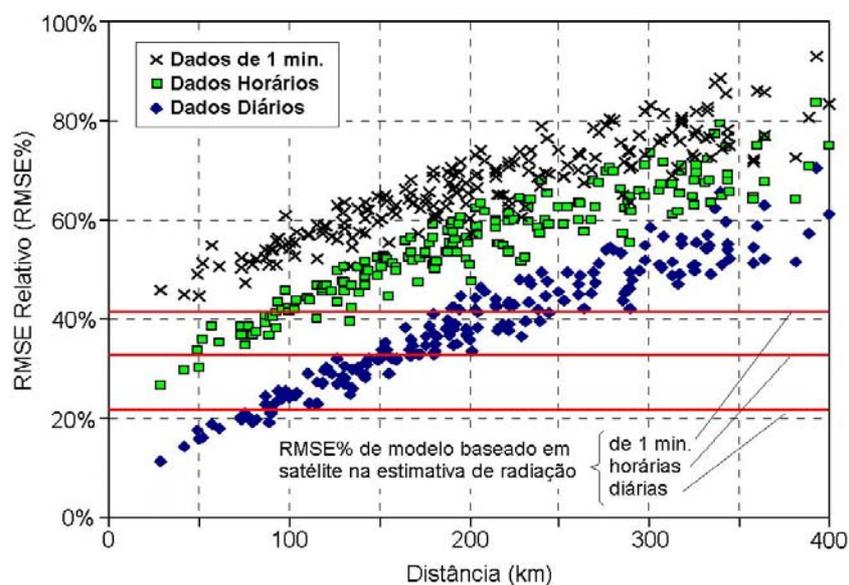


Figura 2.7: Comparação de incertezas na estimativa de radiação solar, entre o emprego de modelo baseado em satélite e redes de radiômetros com diferentes distâncias entre os equipamentos.
Fonte: Perez (2003).

Foi demonstrado que para estimativas do total diário de radiação por interpolação, faz-se necessária uma rede de radiômetros com distâncias inferiores a aproximadamente 75

km para que estas estimativas sejam melhores do que aquelas obtidas por modelos. Para estimativas horárias, radiômetros deveriam ser dispostos a distâncias inferiores a 50 *km*. Para distâncias superiores, as incertezas das interpolações serão sempre maiores do que aquelas dos modelos, demonstrando a vantagem de um método sobre o outro. Para estimativas com resolução temporal de 1 minuto, os modelos são superiores às interpolações, para qualquer distância entre sensores maior que 25 *km*.

Os modelos computacionais para estimativa de radiação solar incidente podem ser classificados em *modelos físicos* e *modelos estatísticos*. Os modelos físicos caracterizam-se por simular os processos físicos atmosféricos, resolvendo a equação de transferência radiativa através de métodos matemáticos (Gautier *et al.*, 1980; Stuhlmann *et al.*, 1990; Whitlock *et al.*, 1995; Pereira *et al.*, 1996; Ceballos *et al.*, 2004). Os modelos estatísticos caracterizam-se por utilizar expressões empíricas para estimativa da radiação incidente em função de propriedades do sistema Terra-atmosfera, sendo os modelos ajustados com o uso de medidas de radiação em uma ou mais estações solarimétricas (Tarpley, 1979; Hay e Hanson, 1978; Diabaté *et al.*, 1988).

Em geral, modelos físicos e estatísticos apresentam resultados comparáveis no que concerne aos níveis de confiança dos resultados para um mesmo local. Contudo, os modelos físicos têm a vantagem de poderem ser generalizados para áreas extensas, mantendo mais ou menos os mesmos níveis de confiança nos resultados desde que sejam disponíveis como entradas os dados atmosféricos e ambientais corretos. Os modelos estatísticos apresentam validade restrita à região para a qual as expressões empíricas foram desenvolvidas, possuindo menor capacidade de regionalização. Ainda, os modelos estatísticos dependem de estações solarimétricas de referência para avaliar os coeficientes de ajuste, e os coeficientes podem variar com o tempo, sendo necessária validação sistemática.

Para emprego nos modelos físicos, vários métodos foram desenvolvidos para solucionar a equação de transferência radiativa de forma precisa, podendo-se citar os métodos que utilizam harmônicos esféricos (Dave e Canosa, 1974), ordenadas discretas (Liou, 1976), ordens sucessivas de espalhamento (Raschke, 1972), Monte Carlo e diferenças finitas

(Lenoble, 1985). Estes métodos demandam elevado tempo computacional, inviabilizando, sob o ponto de vista prático, seu uso operacional. Como alternativa a tais métodos, foram desenvolvidos métodos aproximados que demandam menor tempo computacional. O *método de dois fluxos*, por exemplo, é empregado em diversos modelos (Kerschgens *et al.*, 1978; Gautier *et al.*, 1980; Dedieu *et al.*, 1987), incluindo o modelo BRASIL-SR (Pereira *et al.*, 1996; Martins, 2001), que é um dos modelos empregados no CPTEC/INPE para estimativa de radiação solar incidente a partir de imagens de satélite.

Os modelos de Redes Neurais Artificiais e Regressões Lineares Múltiplas empregados neste trabalho podem ser considerados modelos estatísticos, utilizando informações de condições atmosféricas geradas por previsão numérica de tempo para estimar a radiação solar incidente para instantes futuros.

CAPÍTULO 3

REDES NEURAIIS ARTIFICIAIS

Redes Neurais Artificiais (RNAs) são uma forma de computação não algorítmica caracterizada por sistemas que, em algum nível, lembram a estrutura de raciocínio do cérebro humano (Braga *et al.*, 1998). São ferramentas estatísticas capazes de armazenar conhecimento a partir de exemplos e ser empregadas na solução de problemas de ajuste funcional, reconhecimento de padrões e diversas outras aplicações. Este capítulo tem por intuito descrever as RNAs, suas arquiteturas, características gerais e aplicações. Também é apresentado um breve histórico do desenvolvimento desta ferramenta e aborda-se o processo de aprendizado. Diante dos diversos tipos de RNAs e da diversidade de detalhes envolvidos, procurar-se-á focar apenas os aspectos mais importantes e enfatizar as informações necessárias à compreensão deste trabalho. Para maiores detalhes, recomenda-se a leitura de Haykin (2001) ou de Braga *et al.* (1998).

3.1. Descrição Básica

Redes Neurais Artificiais são compostas por unidades de processamento simples, denominadas *neurônios*, que computam determinadas funções matemáticas, geralmente não-lineares, a partir de entradas recebidas. Estas unidades são dispostas paralelamente constituindo camadas e são interligadas com as unidades de camadas vizinhas por conexões geralmente associadas a pesos. Por analogia às interligações dos neurônios no cérebro humano, as conexões entre unidades de uma rede neural são chamadas *sinapses*, sendo os pesos denominados *pesos sinápticos*.

Os pesos sinápticos ponderam as entradas recebidas por cada neurônio e são os responsáveis pelo armazenamento do conhecimento representado pela RNA. A solução de um problema com RNAs passa inicialmente por uma fase de aprendizagem, onde os pesos das conexões são ajustados até proporcionarem à rede a capacidade de representar o problema. Uma vez treinada, os pesos são fixados e a rede pode ser empregada como um modelo, estimando saídas a partir de um conjunto de dados de entrada.

3.2. Histórico

O primeiro modelo artificial de um neurônio biológico foi apresentado por McCulloch e Pitts (1943), mas sem a introdução de técnicas de aprendizado. Em tal modelo, o neurônio possui entradas (x_1, x_2, \dots, x_n), provenientes dos dados de entrada da rede ou das saídas de outros neurônios, as quais são ponderadas pelos pesos (w_1, w_2, \dots, w_n) das ligações (ou *sinapses*), sendo estes pesos fixos, não ajustáveis. O neurônio realiza o somatório (Σ) dos produtos $x_i w_i$, e caso este valor seja igual ou superior a um limiar θ , o sinal de saída (y) do neurônio será igual a 1, e diz-se que o neurônio tem sua saída ativa ou que está excitado. Se o somatório resultar número menor que o limiar, a saída será zero. Este modelo de neurônio de saídas binárias é ilustrado pela Figura 3.1.

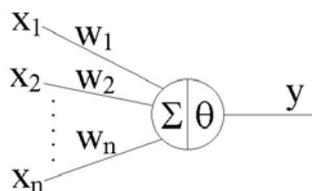


Figura 3.1: Neurônio de McCulloch e Pitts.

O aprendizado foi alvo de estudo de Hebb (1949), que mostrou que a plasticidade da aprendizagem de redes neurais é conseguida através da variação dos pesos de entrada dos neurônios. Ele propôs uma teoria para explicar o aprendizado em neurônios biológicos baseada no reforço das ligações sinápticas entre neurônios excitados.

Rosenblatt (1958) introduziu o *perceptron*, composto por neurônios de McCulloch e Pitts, porém dotados de sinapses ajustáveis. Rosenblatt propôs um algoritmo para treinar a rede para executar determinados tipos de funções, onde os pesos sinápticos são ajustados de acordo com a saída desejada. Por possuir pesos ajustáveis apenas na camada de saída, o modelo de Rosenblatt é tratado como *perceptron de uma camada*.

Quase simultaneamente à apresentação do perceptron por Rosenblatt, o ADALINE (*ADaptive Linear NEuron*) foi desenvolvido por Widrow e Hoff (1960). Embora ambos os modelos sejam baseados em elementos de processamento executando operações de soma ponderada e comparação com um limiar, o ADALINE possui saídas binárias bipolares (-1 ou 1) ao invés das saídas unipolares (0 ou 1) do perceptron. Contudo, a

maior diferença entre os modelos encontra-se em seus algoritmos de aprendizado. Enquanto o perceptron ajusta seus pesos comparando a saída desejada com a saída binária do neurônio, o ADALINE adapta seus pesos em função do erro de sua saída analógica ($y = \sum w_i x_i$), produzida antes da comparação com o limiar θ . A regra de aprendizado do ADALINE é baseada no método do gradiente para minimização do erro, conhecida como *regra delta* ou *regra de Widrow-Hoff* (Haykin, 2001).

Minsky e Papert (1969) destacaram a inépcia do uso do perceptron de uma camada na resolução de certos problemas, independentemente do algoritmo de treinamento. Também foi argumentado que para perceptrons com mais de uma camada, os algoritmos de aprendizado não garantiam convergência como para modelos de camada única. Estas observações levaram as RNAs a serem relegadas a um plano secundário durante toda a década de 1970 e início da década de 1980. Segundo Braga *et al.* (1998), as pesquisas na área foram retomadas, em parte, devido a um trabalho de Hopfield (1982) destacando as propriedades associativas das RNAs. A descrição do algoritmo de treinamento *backpropagation* alguns anos depois (Rumelhart *et al.*, 1986) mostrou que a visão de Minsky e Papert sobre o perceptron era bastante pessimista e que é possível treinar RNAs de múltiplas camadas para resolver problemas complexos.

3.3. Perceptrons de Múltiplas Camadas

Dentre todas as possíveis topologias (arquiteturas), as RNAs podem ser classificadas em *não-recorrentes* (ou *feedforward*) e *recorrentes* (ou *feedback*). As RNAs não-recorrentes são as que não possuem realimentação de suas saídas nas entradas, sendo deste tipo as RNAs empregadas neste trabalho. Estas RNAs têm sua estrutura organizada em camadas, podendo ser constituídas por uma ou mais camadas.

Redes neurais multicamadas contêm um conjunto de neurônios de entrada, uma *camada de saída* e uma ou mais *camadas intermediárias*, também chamadas *camadas escondidas* ou *ocultas*. Segundo Wasserman (1989) a entrada não é considerada uma camada da rede, pois apenas recebe os dados e os distribui, não realizando qualquer ponderação, somatório ou comparação com limiares.

RNAs não-recorrentes com mais de uma camada são denominadas *perceptrons de múltiplas camadas (multilayer perceptron - MLP)*. Como exemplo, a Figura 3.2 mostra uma MLP com três entradas, duas camadas intermediárias com quatro neurônios cada e uma camada de saída com um neurônio, produzindo uma única informação de saída.

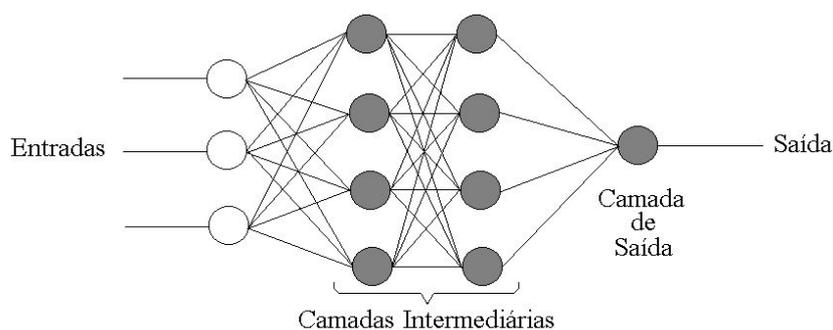


Figura 3.2: Exemplo de uma RNA-MLP com duas camadas intermediárias.

Dentre os vários tipos e topologias de RNAs, o modelo MLP de treinamento supervisionado é uma das arquiteturas mais empregadas na atualidade, em numerosas áreas de ciência e tecnologia (Haykin, 1994). Uma de suas propriedades principais é seu potencial de uso para ajustes funcionais, devido suas universais capacidades de aproximação (Hornik *et al.*, 1989). Segundo Cybenko (1989), uma rede MLP com uma camada intermediária pode implementar qualquer função contínua. A utilização de duas camadas intermediárias permite a aproximação de qualquer função (Cybenko, 1988).

Não há método que predetermine a melhor combinação de camadas e neurônios em uma MLP, pois isso depende do modelo específico, dos processos físicos e do conjunto de dados de treinamento que a rede simulará. Segundo Tymvios *et al.* (2005) existem na bibliografia algumas relações empíricas para resolver este problema, mas o melhor método até o momento é a construção de vários modelos pelo pesquisador, escolhendo aquele que atinge um treinamento satisfatório e melhor se ajusta para a aplicação desejada.

3.4. Modelo de Neurônio Artificial

Os neurônios utilizados nas redes MLP diferem daquele definido por McCulloch e Pitts, em que o somatório das entradas ponderadas é simplesmente comparado com um limiar,

gerando um sinal binário. A partir desse neurônio primordial foram derivados outros modelos que permitem a produção de uma saída qualquer, não necessariamente zero ou um, e com diferentes funções aplicadas ao somatório de entradas ponderadas (Braga *et al.*, 1998). Assim, cada neurônio representado por um círculo cinza na rede MLP da Figura 3.2 deve ser visto como o modelo para um neurônio genérico j da Figura 3.3.

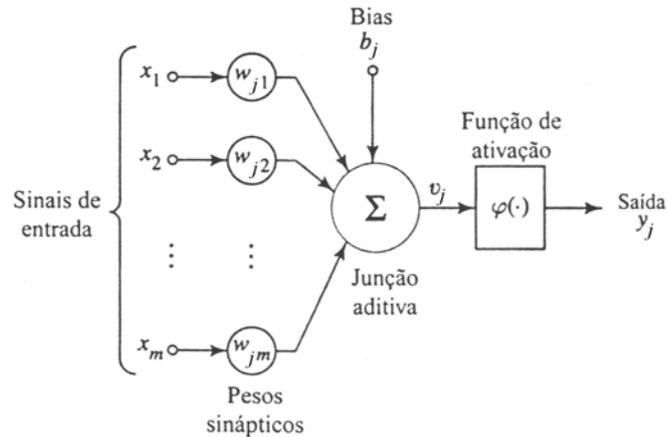


Figura 3.3: Modelo de um neurônio de uma rede MLP.
 Fonte: Adaptado de Haykin (2001).

O chamado *nível de atividade interna* do neurônio (v_j) é obtido pela combinação do somatório das entradas ponderadas com outro valor ajustável, denominado *bias* (b_j). A saída do neurônio, y_j , é produzida pela aplicação do nível de atividade interna a uma *função de ativação* (φ). O *bias* tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo, respectivamente, possuindo papel semelhante ao limiar θ do neurônio de McCulloch e Pitts (Haykin, 1994). Assim, o neurônio j pode ser descrito pelas seguintes equações:

$$v_j = \sum_{i=1}^m w_{ji} x_i + b_j \quad (3.1)$$

$$y_j = \varphi(v_j) \quad (3.2)$$

A função de ativação define a saída de um neurônio de acordo com o nível de atividade produzida pelas suas entradas. Funções de ativação comumente empregadas, de acordo Braga *et al.* (1998) e Haykin (2001), são:

a) Função Limiar ou Degrau:
$$\varphi(v_j) = \begin{cases} +\gamma & \text{se } v_j \geq 0 \\ -\gamma & \text{se } v_j < 0 \end{cases} \quad (3.3)$$

b) Função Linear:
$$\varphi(v_j) = \alpha v_j \quad (3.4)$$

c) Funções Sigmóides:

c1) Função Logística:
$$\varphi(v_j) = \frac{1}{1 + \exp(-\alpha v_j)} \quad (3.5)$$

c2) Função Tangente Hiperbólica:
$$\varphi(v_j) = \operatorname{tgh}\left(\frac{v_j}{2}\right) = \frac{1 - \exp(-v_j)}{1 + \exp(-v_j)} \quad (3.6)$$

O comportamento das funções de ativação definidas acima é esboçado na Figura 3.4. A função degrau é geralmente empregada com $\gamma=1$, e pode ser redefinida para que as saídas situem-se dentro do intervalo de 0 a 1 (neurônio de McCulloch e Pitts). A função linear definida com $\alpha =1$ permite que a saída do neurônio seja igual ao nível de atividade interna. As funções logística e tangente hiperbólica são dois tipos de funções sigmóides, produzindo como saídas valores entre 0 e 1, e entre -1 e 1, respectivamente. Por serem contínuas, crescentes, diferenciáveis e não-lineares, as funções sigmóides são as mais comumente utilizadas nas camadas ocultas de RNAs tipo MLP.

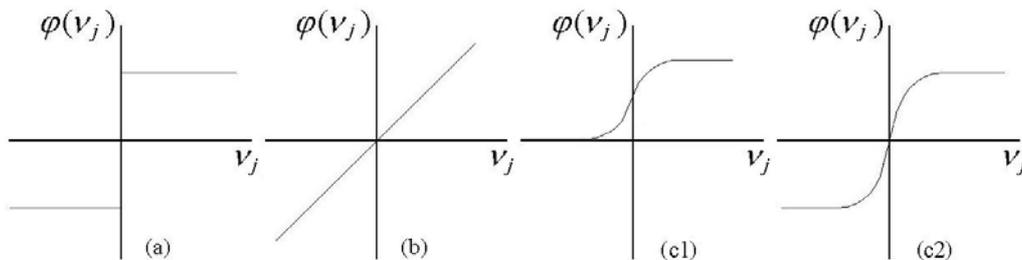


Figura 3.4: Comportamento dos sinais de saída de diferentes funções de ativação: (a) função degrau; (b) função linear; (c1) função logística; (c2) função tangente hiperbólica.

3.5. Aprendizado de RNAs

A solução de um problema com RNAs passa inicialmente por uma fase de aprendizado ou treinamento. Nesta fase, a RNA ajusta os pesos das sinapses seguindo um procedimento chamado *algoritmo de aprendizado*. Durante o treinamento, os pesos da

rede neural gradualmente convergem para determinados valores, de maneira que a aplicação dos vetores de dados de entrada produza as saídas necessárias. Os treinamentos podem ser classificados em *supervisionados* e *não-supervisionados*. Neste trabalho, é empregado um treinamento supervisionado.

No treinamento supervisionado, a cada iteração um vetor de entradas e um vetor alvo (saídas desejadas) são apresentados à rede. A partir das entradas, a rede calcula as saídas e compara-as com os alvos. Os pesos são ajustados de maneira que os valores das saídas aproximem-se dos alvos. Assim, o algoritmo de treinamento gradativamente extrai as características necessárias para representar a informação fornecida. O ajuste dos pesos pode ser realizado a cada iteração (*treinamento seqüencial*) ou após a apresentação à rede de todos os exemplos do conjunto de dados de treinamento (*treinamento por lote*).

Considerando-se um neurônio de saída j em treinamento supervisionado seqüencial, a cada iteração t é produzido um sinal de erro $e_j(t)$, obtido pela diferença entre o valor de saída calculado pelo neurônio e o valor alvo (ou, saída desejada, $d_j(t)$):

$$e_j(t) = d_j(t) - y_j(t) \quad (3.7)$$

Com base em $e_j(t)$, e objetivando obter um erro menor na próxima iteração ($t+1$), ajustes corretivos são calculados e aplicados aos pesos de j :

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) \quad (3.8)$$

em que $w_{ji}(t)$ é o peso da conexão do neurônio j com o neurônio i da camada anterior e $\Delta w_{ji}(t)$ é o ajuste a ser aplicado ao peso, fazendo-o assumir novo valor na iteração seguinte: $w_{ji}(t+1)$. O ajuste é obtido pela minimização da chamada *função de custo*:

$$E(t) = \frac{1}{2} e_j^2(t) \quad (3.9)$$

A minimização desta função para o caso particular de um neurônio de saída com função de ativação linear, resulta na já mencionada *regra delta* (Widrow e Hoff, 1960) que

afirma que o ajuste de um peso é proporcional ao produto entre o erro e a entrada da sinapse (x_i):

$$\Delta w_{ji}(t) = \eta e_j(t) x_i(t) \quad (3.10)$$

em que η é a *taxa de aprendizado*, que determina a velocidade com que os pesos são ajustados na direção de produzirem o menor erro (Braga *et al.*,1998). A regra delta pressupõe que o erro seja diretamente mensurável, de maneira que a sua aplicabilidade restringe-se apenas à última camada da RNA (Haykin, 2001).

A função de custo constitui uma medida de desempenho do sistema em fase de aprendizado e é uma função dos parâmetros livres do sistema (pesos e bias). Assim, a função de custo pode ser visualizada como uma *superfície de erro*, com os parâmetros livres como coordenadas. A superfície de erro pode apresentar um único ponto de mínimo, no caso de a rede fazer uso de funções de ativação lineares, ou pode apresentar vários mínimos locais além do mínimo global, se houverem neurônios não-lineares (Braga *et al.*,1998). Ao longo do aprendizado, sucessivas combinações de parâmetros livres conduzem a pontos sucessivamente mais baixos, em direção a um ponto de mínimo da superfície de erro, sendo esta direção definida através do *gradiente* da superfície de erro em cada ponto. Assim, a minimização da função de custo que leva à regra delta é chamada de *método do gradiente para minimização do erro*, em que o ajuste dos pesos deve ser feito sempre no sentido inverso do gradiente:

$$\vec{\Delta w} = -\eta \cdot \nabla E(\vec{w}) \quad (3.11)$$

em que \vec{w} é um vetor contendo todos os pesos e o bias do neurônio, $E(\vec{w})$ é a função de custo (função de todos os parâmetros livres do neurônio contidos no vetor \vec{w}) e $\nabla E(\vec{w})$ é o gradiente, que é um vetor de igual dimensão a \vec{w} , e onde cada componente é a derivada parcial de $E(\vec{w})$ com relação a cada peso em \vec{w} (Haykin, 2001).

3.5.1. Backpropagation (Retropropagação)

O mais difundido e utilizado algoritmo de treinamento supervisionado para redes MLP é o *backpropagation* (Rumelhart *et al.*, 1986). Neste algoritmo, ocorre numa primeira fase

a propagação do sinal funcional (*feedforward*) e numa segunda fase a retropropagação do erro (*backpropagation*). Na primeira fase, a partir das entradas, o sinal funcional propaga-se pela rede até a geração de uma saída, mantendo-se fixos os pesos das sinapses. Na segunda fase, a saída é comparada com um alvo produzindo um sinal de erro. O sinal de erro propaga-se da saída para entrada, e os pesos são ajustados de maneira a minimizar o erro. O backpropagation é baseado na regra delta, mas oferece uma maneira de ajustar os pesos dos neurônios de todas as camadas da rede neural e não apenas da camada de saída. A dedução completa do backpropagation foge do objetivo desta Dissertação e maiores detalhes podem ser obtidos em Rumelhart *et al.* (1986), Haykin (2001) ou Braga *et al.* (1998).

A cada passo t do treinamento backpropagation, o sinal de erro pode ser calculado para cada neurônio j de saída de uma rede MLP. Para uma RNA com J neurônios de saída, a função de custo é redefinida como o somatório dos J sinais de erro:

$$E(t) = \frac{1}{2} \sum_{j=1}^J e_j^2(t) \quad (3.12)$$

Pelo método do gradiente para minimização do erro, a correção a ser aplicada a cada peso de um neurônio j de saída é dada pela derivada parcial de $E(t)$ em relação ao peso em questão, calculada utilizando-se a regra da cadeia (Haykin, 2001):

$$\Delta w_{ji}(t) = -\eta \frac{\partial E(t)}{\partial w_{ji}(t)} = \eta \cdot e_j(t) \cdot \varphi'_j(v_j(t)) \cdot x_i(t) \quad (3.13)$$

Em que $\varphi'_j(v_j(t))$ é a derivada da função de ativação aplicada ao nível de atividade interna do neurônio. O erro $e_j(t)$ e a derivada da função de ativação são comumente agrupados no chamado *gradiente local*, $\delta_j(t)$:

$$\delta_j(t) = e_j(t) \cdot \varphi'_j(v_j(t)) \quad (3.14)$$

$$\Delta w_{ji}(t) = \eta \delta_j(t) x_i(t) \quad (3.15)$$

Apesar de não ser possível calcular seus erros diretamente, os neurônios ocultos compartilham a responsabilidade pelo erro cometido na saída da rede. Assim, o sinal de erro deve ser retropropagado para que se possa ajustar os pesos das camadas ocultas. Considerando agora que o neurônio j é um neurônio oculto, conectado a neurônios k da camada de saída, o backpropagation calcula o gradiente local por:

$$\delta_j(t) = \varphi'_j(v_j(t)) \cdot \sum_k \delta_k(t) w_{kj}(t) \quad (3.16)$$

Desta forma, o gradiente local para um neurônio j oculto é calculado através do somatório dos gradientes locais de cada neurônio k da camada seguinte, ponderados pelas ligações sinápticas que ligam cada neurônio k ao neurônio j . O sinal de erro para um neurônio oculto é, portanto, uma estimativa determinada recursivamente em termos dos sinais de erro de todos os neurônios da camada posterior. O ajuste dos pesos da camada oculta é realizado pela equação 3.15, utilizando-se a equação 3.16 para $\delta_j(t)$.

O treinamento backpropagation seqüencial é normalmente mais rápido, requer menos memória, mas é mais instável. O treinamento por lote oferece uma melhor estimativa do gradiente e é mais estável, ao custo de maior tempo computacional e memória (Braga *et al.*, 1998). Algumas variantes do backpropagation, como o algoritmo Rprop apresentado a seguir, são definidos apenas para treinamento por lote.

No treinamento por lote, a apresentação de todos os exemplos de treinamento ($n = 1, \dots, N$) constitui uma *época*. Durante cada época, os pesos mantêm-se constantes e uma nova função de custo é definida como sendo a média das funções de custo definidas pela equação 3.12 para cada exemplo n . Em decorrência do cálculo do gradiente dessa nova função de custo, os ajustes dos pesos ao final da época são dados pelas médias de todos os ajustes calculados com cada exemplo (Haykin, 2001).

3.5.2. Algoritmo Rprop

Segundo Demuth *et al.* (2005), devido ao comportamento assintótico das funções de ativação sigmóides, suas derivadas assumem valores muito pequenos para argumentos

afastados de zero, conduzindo a gradientes locais de pequena magnitude, de maneira que o ajuste de pesos é pequeno mesmo estando afastados de seus valores ótimos. Como redes MLP empregam funções sigmóides nas camadas ocultas, o backpropagation é computacionalmente lento e o tempo de convergência é grande. Variantes desse algoritmo de treinamento foram desenvolvidas para aumentar o desempenho e a velocidade de treinamento. Uma dessas variantes é o algoritmo Rprop - *Resilient Backpropagation* (Riedmiller e Braun, 1993; Riedmiller, 1994), que é o tipo de treinamento empregado neste trabalho.

O propósito do algoritmo Rprop é eliminar a influência da magnitude dos gradientes no ajuste de pesos. O algoritmo utiliza apenas o sinal do gradiente no ajuste dos pesos, de maneira a conhecer a direção em que o ajuste deve ser realizado. A magnitude do ajuste é variável, e aumenta ou diminui dependendo se a direção do ajuste mantém-se ou altera-se, respectivamente, ao longo das épocas de treinamento (Demuth *et al.*, 2005).

3.6. Generalização

O encerramento da fase de treinamento ocorre quando o erro atingir um valor mínimo pré-estabelecido ou quando a taxa de variação desse erro atingir um valor definido como critério de parada (Haykin, 1994). Contudo, treinar a rede até o mínimo da superfície de erro não garante que a rede possua capacidade de generalizar o problema aprendido, isto é, não é garantida a eficiência da RNA para um conjunto de dados diferente daquele utilizado para o treinamento, de maneira que o erro para esse novo conjunto de dados pode ser elevado.

Quando a rede não adquire capacidade de generalização do problema, diz-se que ocorreu um conhecido problema no treinamento de RNAs, chamado *excesso de ajuste* ou *overfitting*. O *overfitting* geralmente ocorre quando é utilizado um número muito elevado de neurônios ocultos (Demuth *et al.*, 2005), permitindo à RNA “decorar” os dados, ao invés de extrair seus padrões. Existem alguns métodos para melhorar a capacidade de generalização de uma RNA, sendo os mais conhecidos a *regularização* e o *treinamento com parada antecipada*.

A regularização implica na modificação da função de custo, de maneira a minimizar não apenas o erro, mas também o tamanho dos pesos, ao longo do treinamento (Hinton e Nowlan, 1987; Braga *et al.*, 1998), uma vez que pesos elevados estão relacionados à ocorrência de *overfitting* no treinamento:

$$E = \frac{1}{2} \sum_{j=1}^J (e_j)^2 + \frac{1}{2} \lambda \|\vec{w}\|^2 \quad (3.17)$$

em que J é o número de neurônios da camada de saída, $\|\vec{w}\|$ é a norma do vetor de pesos que contém todos os pesos da RNA e λ é o *parâmetro de regularização*.

O treinamento com parada antecipada consiste em dividir os dados disponíveis em um *conjunto de treinamento* e um *conjunto de validação*. O ajuste dos pesos é realizado utilizando-se apenas o conjunto de treinamento, sendo o conjunto de validação utilizado para avaliação da capacidade de a RNA generalizar para um conjunto de dados diferente daquele utilizado no aprendizado. Assim, ao longo do treinamento, o erro de ambos os conjuntos é medido, e o treinamento é interrompido quando o erro para o conjunto de validação começar a subir (Demuth *et al.*, 2005; Braga *et al.*, 1998).

3.7. Modelagem de Radiação Solar utilizando RNAs

O emprego de RNAs é por muitos considerado um método deslegante, por não levar em conta aspectos físicos, e por não empregar o equacionamento das conhecidas (e muitas vezes, complexas) relações físicas entre as variáveis. Mas como destacado no final do capítulo anterior, tal ferramenta pode ser classificada como um modelo estatístico para estimativa de radiação solar, estabelecendo relações empíricas entre a radiação incidente e variáveis meteorológicas. A utilização de RNAs para obter estimativas aproximadas ou previsões de radiação solar incidente segue uma tendência, observada em vários setores de pesquisa, para a utilização de métodos alternativos aos convencionais e que apresentem boa destreza frente a problemas não-lineares ou de natureza caótica. Diante da relativa simplicidade do emprego de RNAs, a sua utilização permite atender de forma mais rápida a demanda por um determinado setor.

O emprego de RNAs para a modelagem da radiação incidente é recente. Al-Alawi e Al-Hinai (1998) empregaram RNAs para estimativa de radiação solar, utilizando como entradas dados diários de pressão, temperatura, umidade relativa, velocidade do vento, duração de brilho do Sol, mês do ano e latitude da estação. Mohandes *et al.* (1998) realizaram estimativas de radiação com RNAs alimentadas por duração de brilho do Sol, latitude, longitude e altitude das estações. Soares *et al.* (2004) utilizaram RNAs para modelar a radiação solar difusa horária na cidade de São Paulo. Dentre outros estudos, todos recentes, pode-se ainda citar os trabalhos de Hontoria *et al.* (2002, 2005), Tymvios *et al.* (2005) e Elminir *et al.* (2005), que, fazendo uso de RNAs do tipo MLP, também realizaram estimativas de radiação solar para estações específicas, com base em dados meteorológicos. Segundo Tymvios *et al.* (2005), RNAs do tipo MLP são as mais amplamente empregadas em tópicos meteorológicos.

A grande maioria dos estudos de radiação solar com RNAs, tratam de modelar radiação solar a partir de outras variáveis medidas por estações meteorológicas, estimando a radiação para os mesmos instantes das medidas. Estudos de previsão de radiação solar com RNAs são menos frequentes. Sfetsos e Coonick (2000) realizaram previsões futuras de radiação solar baseadas em dados passados e combinando uma série de outras ferramentas com diversas topologias de RNAs recorrentes. Mihalakakou *et al.* (2000) também realizaram previsões horárias de radiação utilizando valores antecedentes de radiação. Mas como essas previsões levam em conta apenas as periodicidades observadas nas séries temporais de radiação antecedentes e como a radiação solar incidente é dependente de fatores atmosféricos não conhecidos de antemão para o instante futuro (por exemplo, a cobertura de nuvens), tal incerteza pode levar a previsões completamente diferentes da realidade a ser observada no instante futuro, dependendo da localidade de estudo.

Neste trabalho, RNAs são treinadas para apenas modelar a radiação solar medida com base em variáveis atmosféricas para o mesmo instante, cabendo, pois, o aspecto de “previsão”, ao fato destas variáveis serem geradas por previsões de tempo em instantes anteriores. Após treinamento, essas RNAs podem atuar como um pós-processamento das saídas de previsão de tempo, produzindo estimativas de radiação solar.

CAPÍTULO 4

DADOS: MODELO ETA E PROJETO SONDA

Este capítulo tem por objetivo descrever os dados utilizados neste trabalho, bem como suas fontes: o modelo Eta operacional do CPTEC e as estações solarimétricas do projeto SONDA. Como mencionado anteriormente, as variáveis de saída do modelo Eta serão empregadas como preditores nos modelos de RNAs e de RLM, enquanto os dados de radiação das estações SONDA serão utilizados para treinamento/ajuste dos modelos e para verificação das previsões de radiação do Eta e dos modelos ajustados.

4.1. O Modelo Eta/CPTEC

O modelo Eta é um modelo de previsão de tempo de mesoescala, em ponto de grade, de equações primitivas (INPE.CPTEC, 2005; MetEd, 2005). Trata-se de um modelo internacional, empregado em previsão meteorológica operacional e em pesquisa, em diversos países. Foi criado na Iugoslávia, por Zaviša Janjić e Fedor Mesinger, mas seu desenvolvimento contou com a colaboração de pesquisadores de diversos países, e grande parte das parametrizações e o desenvolvimento posterior do modelo foram realizados no *National Center for Environmental Prediction* (NCEP).

O modelo Eta é utilizado operacionalmente no Centro de Previsão de Tempo e Estudos Climáticos (CPTEC/INPE-MCT) desde 1996, sendo sua configuração descrita por Chou (1996). O modelo Eta/CPTEC é originário do NCEP, e cobre a maior parte da América do Sul e oceanos adjacentes: latitudes entre 50,2°S e 12,2°N, e longitudes entre 83°W e 25,8°W. A resolução horizontal é de 40 km e a vertical de 38 níveis. Uma versão do modelo com resolução horizontal de 20 km vem sendo testada no CPTEC, embora neste trabalho tenham sido utilizados dados do modelo de 40 km. Uma breve descrição do modelo é apresentada no item a seguir, e maiores detalhes podem ser obtidos em Mesinger *et al.* (1988), Janjić (1994), Black (1994) e Ničković *et al.* (1998).

4.1.1. Descrição do Modelo Eta

As variáveis prognósticas do modelo Eta são: temperatura do ar, componentes zonal e meridional do vento, umidade específica, água líquida da nuvem, pressão à superfície e energia cinética turbulenta. A grade horizontal utilizada é a grade E de Arakawa (Arakawa e Lamb, 1977), e um esquema desta grade é apresentado na Figura 4.1. Nos pontos denominados por h (pontos de massa) são definidas as variáveis de pressão em superfície, temperatura, umidade específica, velocidade vertical, água líquida da nuvem, energia cinética turbulenta e outras variáveis termodinâmicas. Nos pontos v (pontos de vento) são definidas as componentes horizontais do vento. O comprimento d indica o espaçamento entre dois pontos de h , ou de v , adjacentes, sendo que a magnitude desse comprimento é comumente usada para definir a resolução horizontal do modelo. Cada ponto de grade consiste em um ponto de massa localizado no centro de quatro pontos de velocidade do vento.

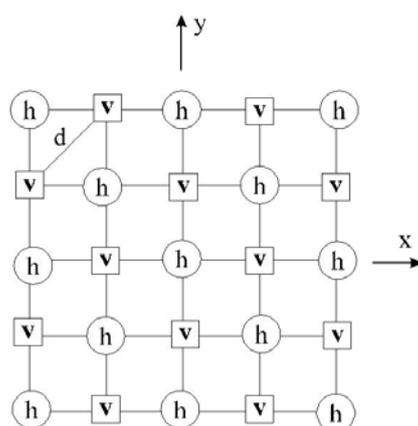


Figura 4.1: Grade horizontal tipo E de Arakawa.
Fonte: Modificado de Ničković *et al.* (1998).

A coordenada vertical *eta* (η), que é utilizada no modelo, e a partir da qual provém o seu nome, foi desenvolvida por Mesinger (1984) para reduzir problemas relacionados com a coordenada *sigma*, bastante empregada em modelos meteorológicos. Por acompanhar a topografia, a coordenada *sigma* produz erros em gradientes horizontais na vizinhança de regiões de terreno montanhoso. As superfícies de *eta* constante estão dispostas quase na horizontal tanto em áreas planas quanto nas montanhosas, reduzindo o problema da falsa contribuição para o gradiente de pressão observado na coordenada *sigma* (Ničković *et al.*, 1998). A coordenada η é definida através da seguinte relação:

$$\eta = \left(\frac{p - p_T}{p_{sfc} - p_T} \right) \eta_{srf} \quad (4.1)$$

em que

$$\eta_{srf} = \frac{p_{ref}(z_{sfc}) - p_T}{p_{ref}(0) - p_T} \quad (4.2)$$

Nestas equações, p_T é a pressão no topo do domínio do modelo (isto é, a pressão no mais elevado nível atmosférico modelado), p_{sfc} e z_{sfc} são respectivamente a pressão e a altitude à superfície e p_{ref} é uma pressão de referência, função da distância acima do nível do mar obtida a partir de uma atmosfera padrão. Das Equações 4.1 e 4.2 decorre que: $\eta = 0$ em $p = p_T$; $\eta = 1$ em $z = z_{sfc}=0$; e $\eta = \eta_{srf}$ em $z = z_{sfc}$.

A topografia do modelo é representada por degraus discretos, cujos topos coincidem com a interface das camadas do modelo, de acordo com o diagrama esquemático da Figura 4.2. As alturas das interfaces são previamente determinadas de acordo com a atmosfera padrão. As variáveis de temperatura e umidade específica (representadas por T, na Figura 4.2), e as componentes horizontais do vento (representadas por u) são dispostas no meio das camadas, enquanto a velocidade vertical e o geopotencial são dispostos nas interfaces das camadas. Os pontos de vento são situados nos lados verticais dos contornos dos degraus. Na figura, as linhas horizontais tracejadas representam as alturas reais do relevo, que são ajustadas pela topografia do modelo aos níveis de pressão (p_s) das interfaces das camadas. Descrições mais detalhadas podem ser obtidas em Ničković *et al.* (1998).

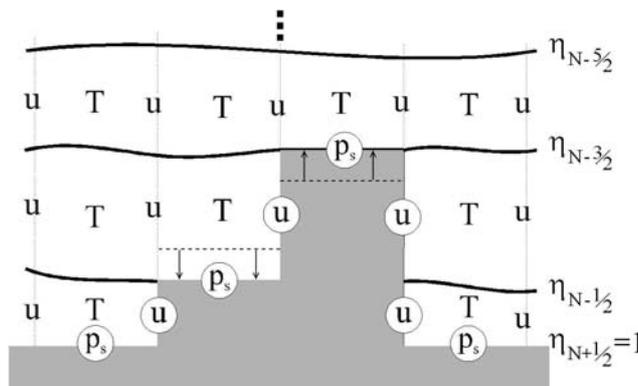


Figura 4.2: Seção vertical idealizada da topografia do modelo Eta e distribuição das variáveis na vertical.

Fonte: Modificado de Ničković *et al.* (1998).

O modelo Eta/CPTEC operacional é rodado duas vezes ao dia, com condições iniciais nos horários 00:00 UT e 12:00 UT (UT = *Universal Time*), que correspondem às 21:00 e às 9:00 no horário oficial brasileiro. Uma vez que os horários utilizados ao longo desta dissertação correspondem sempre a horas inteiras (*hh:00*), se utilizará daqui por diante apenas a notação *hhUT*, para representar os instantes em horário universal.

Como condições iniciais o modelo utiliza as análises de modelos do NCEP/NCAR. A temperatura da superfície do mar é obtida diretamente de um valor médio das temperaturas observadas na semana anterior, sendo esta variável mantida constante durante o período de integração. A umidade do solo e o albedo são obtidos de climatologias globais mensal e sazonal, respectivamente. As condições de contorno lateral são obtidas a partir das previsões do Modelo Global do CPTEC e atualizadas a cada 6 horas. Assim, o modelo Eta realiza previsões para cada 6 horas após o horário da condição inicial, coincidindo com os horários sinóticos de cada dia (00UT, 06UT, 12UT e 18UT).

4.1.2. Parametrização de Radiação

A parametrização de radiação do modelo foi desenvolvida no *Geophysical Fluid Dynamics Laboratory* (GFDL). Para o cálculo da radiação de onda curta o esquema utilizado é o de Lacis e Hansen (1974), e para radiação de onda longa o esquema de Fels e Schwarzkopf (1975). A nebulosidade é baseada na umidade relativa (nuvens estratiformes) e na taxa de chuva convectiva do modelo (nuvens convectivas). As frações de nuvens calculadas pelo modelo são indiretamente usadas na parametrização de processos radiativos.

O esquema de Lacis e Hansen (1974) utilizado para o cálculo de radiação de onda curta é baseado num método computacional conhecido como *método de adicionamento* (*adding method*). Segundo Lenoble (1993), a essência do método é simples e consiste na subdivisão da atmosfera em camadas delgadas de maneira que suas espessuras óticas δ_i sejam suficientemente pequenas para que se possa conhecer suas funções de reflexão e transmissão de radiação com a aproximação de espalhamento simples. Através do cômputo das sucessivas reflexões entre duas camadas (Figura 4.3), pode-se conhecer as

funções de reflexão e transmissão para a camada combinada. Assim, uma camada inteira com espessura ótica δ^* é constituída pelas sucessivas adições de camadas delgadas de espessura ótica δ_i . O equacionamento detalhado do método de adicionamento, utilizado no esquema de radiação do modelo Eta, pode ser encontrado em Lacis e Hansen (1974) ou em Lenoble (1985).

A comparação da radiação solar e outros fluxos de superfície simulados pelo modelo com dados observacionais mostrou que o modelo Eta/CPTEC sistematicamente superestima a radiação solar incidente, bem como os fluxos de calor sensível e latente na superfície (Chou *et al.*, 2002). Hinkelman *et al.* (1999) também compararam dados de fluxos radioativos calculados pelo Eta/NCEP com medidas de radiação solar e verificaram que apesar de os fluxos de energia na superfície estarem aproximadamente balanceados, um excesso sistemático de radiação solar incidente é observado, sendo tais erros atribuídos a conhecidas deficiências da parametrização de Lacis e Hansen, como sugerido por Ramaswamy e Freidenreich (1992). Problemas na parametrização de nuvens e a deficiente atenuação da radiação de onda curta por vapor d'água e aerossóis são apontados como causadores do viés observado (Hinkelman *et al.*, 1999).

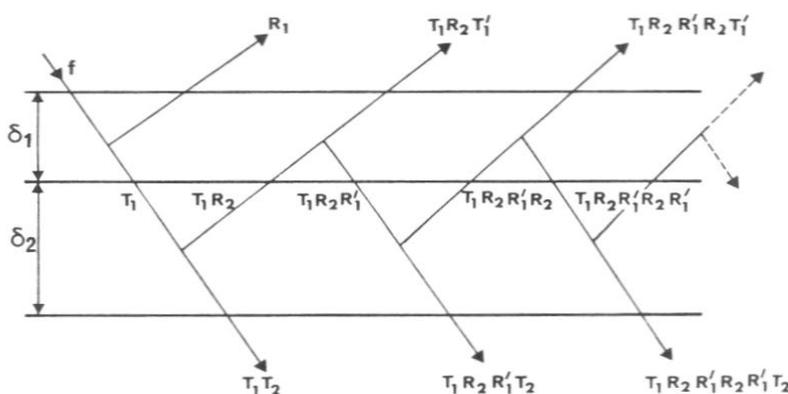


Figura 4.3: Adicionamento de duas camadas de espessuras óticas δ_1 e δ_2 . T refere-se à transmissão total (direta e difusa) e R à reflexão.
Fonte: Lenoble (1993).

4.1.3. Variáveis de Saída do Modelo Eta/CPTEC

Como mencionado anteriormente, a cada dia o modelo Eta/CPTEC gera duas saídas, correspondentes às rodadas de 00UT e 12UT. Em cada rodada são gerados arquivos de saída contendo dados das condições atmosféricas para diversos instantes futuros,

denominados *tempos*. O *tempo 1* (ou análise) corresponde ao horário da condição inicial (00UT ou 12UT). O *tempo 2* corresponde a um instante futuro, 6 horas após a condição inicial. Os demais *tempos* (3, 4, 5, 6, ...) correspondem a instantes futuros, múltiplos de 6 horas (12 horas, 18 horas, 24 horas, 30 horas, ...), após a hora de rodada do modelo.

Até meados de setembro de 2003, o modelo Eta era rodado no supercomputador SX4 do CPTEC, e os seus dados de saída eram armazenados em fitas magnéticas. Os arquivos de saída possuíam informação para 13 *tempos* (prazo de integração de 72 horas, ou 3 dias futuros). Com a migração do modelo para o supercomputador SX6 em 2003, os arquivos de saída passaram a conter 29 *tempos* (prazo de integração de 168 horas, ou 7 dias futuros), e são, até o momento, disponibilizados em um servidor de dados.

Os arquivos de saída do Eta possuem dados para 58 variáveis, sendo 49 variáveis de superfície, ou variáveis de valor único para toda a coluna atmosférica, e 9 variáveis contendo valores em 19 níveis de pressão. A fim de reduzir o volume de dados a serem processados e de maneira a simplificar o conjunto de preditores a serem empregados neste estudo, optou-se por não utilizar as variáveis dispostas em níveis. Ainda com o intuito de reduzir o número de variáveis, para que se pudesse armazenar os dados para emprego futuro em outras estações solarimétricas, foram desprezadas variáveis que se apresentam constantes ao longo do tempo para cada ponto de grade, ou que são pouco representativas das condições do tempo, como por exemplo: topografia, máscara mar-terra e variáveis de temperatura e umidade dentro do solo. Foram preservadas 33 variáveis como possíveis preditores a serem empregados neste trabalho.

Estas 33 variáveis são apresentadas na Tabela 4.1, representadas por suas siglas-padrão de 4 letras, juntamente com suas descrições e unidades. As temperaturas são dadas na unidade Kelvin (K), as pressões em hecto-Pascals (hPa), velocidades em metros por segundo ($m \cdot s^{-1}$), energia em metro quadrado por segundo quadrado ($m^2 \cdot s^{-2}$ – equivalente a joule por quilograma, $J \cdot kg^{-1}$). As taxas de precipitação são expressas em quilograma por metro quadrado por dia ($kg \cdot m^{-2} \cdot dia^{-1}$). Os fluxos de energia (calor e radiação) são expressos em Watts por metro quadrado ($W \cdot m^{-2}$). Água precipitável é dada em quilogramas por metro quadrado ($kg \cdot m^{-2}$) e umidade específica em quilogramas de água

por quilogramas de ar ($kg(H_2O) \cdot kg(ar)^{-1}$). Albedo e frações de cobertura de nuvens são adimensionais variando entre 0 e 1. A fração de cobertura de nuvens é discretizada em fração de cobertura de nuvens baixas (*lwnv*), médias (*mdnv*) e altas (*hinv*), sendo também apresentada a fração total de cobertura de nuvens (*cbnt*).

Tabela 4.1: Variáveis de saída do modelo Eta/CPTEC candidatas a preditores de radiação solar.

| Variável | Descrição (unidades) |
|-------------|--|
| <i>pslm</i> | Pressão ao nível médio do mar (hPa) |
| <i>pslc</i> | Pressão à superfície (hPa) |
| <i>tp2m</i> | Temperatura a 2 metros acima da superfície (K) |
| <i>dp2m</i> | Temperatura do Ponto de Orvalho a 2 metros acima da superfície (K) |
| <i>u10m</i> | Componente zonal do vento a 10 metros acima da superfície ($m \cdot s^{-1}$) |
| <i>v10m</i> | Componente meridional do vento a 10 metros acima da superfície ($m \cdot s^{-1}$) |
| <i>prec</i> | Precipitação total ($kg \cdot m^{-2} \cdot dia^{-1}$) |
| <i>prcv</i> | Precipitação convectiva ($kg \cdot m^{-2} \cdot dia^{-1}$) |
| <i>prge</i> | Precipitação de larga escala ($kg \cdot m^{-2} \cdot dia^{-1}$) |
| <i>neve</i> | Precipitação de neve ($kg \cdot m^{-2} \cdot dia^{-1}$) |
| <i>clsf</i> | Fluxo de calor latente à superfície ($W \cdot m^{-2}$) |
| <i>cssf</i> | Fluxo de calor sensível à superfície ($W \cdot m^{-2}$) |
| <i>ghfl</i> | Fluxo de calor para o solo ($W \cdot m^{-2}$) |
| <i>tsfc</i> | Temperatura da superfície (K) |
| <i>qsfc</i> | Umidade específica à superfície ($kg(H_2O) \cdot kg(ar)^{-1}$) |
| <i>tgsc</i> | Temperatura da camada superficial do solo – 10 cm (K) |
| <i>ussl</i> | Umidade do solo à superfície (0 a 1 - adimensional) |
| <i>lwnv</i> | Fração de cobertura de nuvens baixas (0 a 1 - adimensional) |
| <i>mdnv</i> | Fração de cobertura de nuvens médias (0 a 1 - adimensional) |
| <i>hinv</i> | Fração de cobertura de nuvens altas (0 a 1 - adimensional) |
| <i>cbnt</i> | Fração média de cobertura de nuvens (0 a 1 - adimensional) |
| <i>ocis</i> | Fluxo de radiação de onda curta descendente na superfície ($W \cdot m^{-2}$) |
| <i>olis</i> | Fluxo de radiação de onda longa descendente na superfície ($W \cdot m^{-2}$) |
| <i>oces</i> | Fluxo de radiação de onda curta ascendente na superfície ($W \cdot m^{-2}$) |
| <i>oles</i> | Fluxo de radiação de onda longa ascendente na superfície ($W \cdot m^{-2}$) |
| <i>roce</i> | Fluxo de radiação de onda curta ascendente no topo da atmosfera ($W \cdot m^{-2}$) |
| <i>role</i> | Fluxo de radiação de onda longa ascendente no topo da atmosfera ($W \cdot m^{-2}$) |
| <i>albe</i> | Albedo (0 a 1 - adimensional) |
| <i>cape</i> | Energia potencial convectiva disponível ($m^2 \cdot s^{-2}$) |
| <i>cine</i> | Energia para inibição da convecção ($m^2 \cdot s^{-2}$) |
| <i>agpl</i> | Água precipitável instantânea ($kg \cdot m^{-2}$) |
| <i>pcbs</i> | Pressão da base da nuvem (hPa) |
| <i>pctp</i> | Pressão no topo da nuvem (hPa) |

Com exceção das variáveis de fluxos (expressos em $W \cdot m^{-2}$) e das variáveis de precipitação (expressas em $kg \cdot m^{-2} \cdot dia^{-1}$), todas as demais variáveis são valores

instantâneos para cada *tempo*, isto é, são valores válidos apenas para o horário representado pelo *tempo*. Por sua vez, as variáveis de fluxo para um dado *tempo* são, na realidade, médias para o período de 6 horas entre o *tempo* anterior e o *tempo* em questão, e as variáveis de precipitação são valores acumulados de precipitação no período de 6 horas que antecede o *tempo* considerado.

Foram tomados dados de saída do Eta a partir de janeiro de 2002 até outubro de 2005. Este é também o período de dados de radiação medidos pela estação de Florianópolis que será empregado neste trabalho. Para os cálculos para a estação de São Martinho da Serra, são empregados dados do Eta de julho de 2004 a outubro de 2005, pois este é o período de dados de radiação disponíveis para esta estação.

4.2. O Projeto SONDA: Estações SMS e FLN

O SONDA - Sistema de Organização Nacional de Dados Ambientais para o setor de energia - é um projeto do Instituto Nacional de Pesquisas Espaciais (INPE), coordenado pelo CPTEC, para implementação de uma infra-estrutura física e de recursos humanos destinada a levantar e melhorar a base de dados de superfície necessária ao levantamento e planejamento do uso dos recursos de energia solar e eólica no Brasil (SONDA, 2005). O projeto é financiado pela FINEP - Financiadora de Estudos e Projetos, através do Fundo Setorial de Energia - CT-ENERG, e, a partir de 2006, pela Petrobrás. O objetivo central do projeto é a instalação de uma rede de estações para medição da radiação solar incidente, da velocidade e direção dos ventos, e de outras variáveis meteorológicas, no território brasileiro. A Figura 4.4 indica a localização das estações do projeto.

As *estações solares básicas* são dotadas de sensores para medida de radiação solar global, radiação solar difusa, radiação PAR (radiação fotossinteticamente ativa) e iluminância. Além dos sensores presentes nas estações solares básicas, as *estações solares avançadas* possuem medida de radiação solar direta e de radiação de onda longa descendente. As *estações eólicas* possuem sensores para medida de vento em diferentes alturas. Por fim, as *estações de referência* são as estações mais completas, compostas por todos os sensores presentes nas estações solares avançadas e nas estações eólicas, e

A estação de São Martinho da Serra (SMS) está situada dentro das instalações do Observatório Espacial do Sul, pertencente à Unidade Regional Sul do INPE (OES/RSU/INPE), e também conta com suporte do Laboratório de Ciências Espaciais de Santa Maria, vinculado ao Centro de Tecnologia, da Universidade Federal de Santa Maria (LACESM/CT/UFSM). A estação de Florianópolis (FLN) encontra-se dentro do campus da Universidade Federal de Santa Catarina (UFSC), situado na porção insular da cidade, e sua manutenção encontra-se a cargo do Laboratório de Energia Solar (LABSOLAR/UFSC).

As estações entraram em operação sob os padrões SONDA em julho de 2004. Contudo, algumas estações solarimétricas mais antigas, e que apresentavam dados de boa qualidade, foram incorporadas ao projeto SONDA, como é o caso da estação de Florianópolis. Neste trabalho, conforme mencionado em um item anterior, foram empregados dados de Florianópolis de janeiro de 2002 a outubro de 2005, e dados de São Martinho da Serra de julho de 2004 a outubro de 2005.

Embora seja planejada a aplicação da metodologia aqui apresentada para as demais estações do projeto, as estações SMS e FLN foram escolhidas para utilização nesse trabalho por diferentes razões: SMS é atualmente a estação de referência com melhor qualidade e continuidade dos dados desde o início de operação em julho de 2004; e FLN apresenta uma série de dados bastante longa, muito anterior à sua integração ao projeto SONDA, tendo dados disponíveis desde a década de 1990 quando se tornou uma estação pertencente à BSRN.

As estações solarimétricas básicas, avançadas e de referência possuem em comum um piranômetro para medida da radiação solar global incidente. O equipamento utilizado é um piranômetro *Kipp & Zonen CM21* (Figura 4.5). Este sensor mede radiação solar global na faixa espectral de 0,3 a 2,8 μm e é considerado um medidor de radiação *padrão secundário* pela *WMO* e pela *ISO 9060* (Kipp & Zonen, 2006).

A Figura 4.5 mostra a plataforma onde estão instalados os equipamentos radiométricos do projeto SONDA da estação SMS, no Observatório Espacial do Sul, juntamente com os sensores de medida de radiação ultravioleta e ozônio do Programa de Monitoramento

do Ozônio Atmosférico – PMOA (projeto do LACESM-UFSM em parceria com o INPE). Instalações similares são utilizadas na estação de Florianópolis. No detalhe da Figura 4.5, é apresentada uma fotografia do piranômetro *CM21*, instalado dentro de seu acessório de ventilação, que promove a estabilidade térmica e mantém o domo limpo.



Figura 4.5: Plataforma radiométrica do Observatório Espacial do Sul, suportando os equipamentos do projeto SONDA e os sensores de radiação ultravioleta e ozônio do PMOA. No detalhe, o piranômetro *CM21* montado em seu acessório de ventilação.

Os sensores de radiação das estações SONDA medem radiação com resolução temporal de 1 segundo, em Watts por metro quadrado ($W \cdot m^{-2}$), mas são armazenados valores médios de cada minuto, juntamente com seu desvio padrão, valor mínimo e valor máximo no minuto. Neste trabalho, foram utilizados dados de radiação solar global, medidos por piranômetros *CM21*, com resolução de 1 minuto, mas que foram posteriormente integrados em intervalos de 6 horas, para estar em conformidade com a resolução temporal dos valores disponíveis do modelo Eta.

CAPÍTULO 5

METODOLOGIA

A previsão de radiação solar incidente em um instante futuro pode ser realizada tendo-se uma estimativa das condições atmosféricas em tal instante, uma vez que são essas condições que modulam a incidência de radiação na superfície. Outras variáveis de superfície, correlacionáveis com a radiação solar incidente, como temperatura e fluxos de energia, também podem ser empregadas na estimativa. Estas informações podem ser supridas com o uso de dados de saída de modelos de previsão de tempo, como os dados gerados pelo modelo Eta/CPTEC operacional, utilizados neste trabalho. Este capítulo tem por objetivo descrever as etapas de preparação dos dados, os preditores utilizados, o ajuste dos modelos de RNAs e de RLM, e os parâmetros de avaliação das previsões.

5.1. Extração de Dados de Saída do Modelo Eta

Os dados de saída das rodadas do modelo Eta são disponibilizados em pontos de grade, a cada 0,4° de latitude e longitude, em uma área entre as latitudes 12,2° Norte e 50,2° Sul e longitudes entre 83,0° Oeste e 25,8° Oeste, como mostrado na Figura 5.1.

Como descrito no capítulo anterior, em cada ponto de grade, encontram-se dados para 58 variáveis, 9 delas dispostas em 19 níveis de pressão (totalizando $49 + (9 \times 19) = 220$ dados para cada ponto), e disponíveis para 13 *tempos* (janeiro de 2002 a setembro de 2003) ou para 29 *tempos* (setembro de 2003 a outubro de 2005). Assim, os arquivos de saída de cada rodada do Eta ocupam uma memória bastante elevada.

Como é planejada a aplicação da metodologia deste trabalho para outras localidades, realizando testes para as demais estações do projeto SONDA e até mesmo estações de outros projetos, procurou-se armazenar os dados do Eta de todo território brasileiro. Para que se pudesse armazenar uma quantidade de dados compatível com a disponibilidade de memória, os dados de saída do Eta, das rodadas de 00UT e 12UT, foram “recortados” empregando uma ferramenta do software GRADS, chamada LATS4D. Assim, os arquivos binários do Eta foram reduzidos, passando a abranger

apenas a área do território brasileiro (foi adotado um quadrado entre as latitudes 7°N e 35°S e entre as longitudes 75°W e 33°W), como indicado na Figura 5.1. Também foram tomadas apenas as 33 variáveis indicadas na Tabela 4.1, e dos arquivos com 29 *tempos* foram tomados apenas 21 *tempos* (equivalente a um prazo de integração de 120 horas ou 5 dias). Conseguiu-se assim fazer com que os dados passassem a ocupar um espaço equivalente a aproximadamente 5% do volume de dados iniciais de 29 *tempos*.



Figura 5.1: Área do modelo Eta (linhas azuis) e área de extração dos dados (linhas pretas).

A partir dos dados para todo o Brasil foram separados apenas os dados dos pontos de grade do Eta mais próximos às estações SMS e FLN, como representativos destas. Na Figura 5.2 pode-se observar uma imagem do satélite *Landsat*, para a região de São Martinho da Serra, obtida através do software *NASA World Wind* (NASA, 2005). Na imagem são indicadas as linhas de latitude e longitude com suas respectivas coordenadas. Os pontos azuis correspondem aos pontos de grade do modelo Eta, sendo que cada ponto representa toda a área do retângulo azul concêntrico. Assim, cada “caixa” do Eta possui 0,4° de lado, tanto em latitude como em longitude, sendo representada pelo ponto central. O ponto vermelho representa a localização do

Observatório Espacial do Sul, onde se situa a estação SMS do projeto SONDA. Desta forma, foram tomados para a estação SMS os dados do Eta para o ponto de grade 29,4°S e 53,8°O. A imagem menor, no canto direito superior da Figura 5.2, indica a localização da imagem maior, dentro do Estado do Rio Grande do Sul. A imagem do canto direito inferior é uma ampliação da região marcada pelo retângulo vermelho tracejado, onde é indicada a posição exata do sítio.

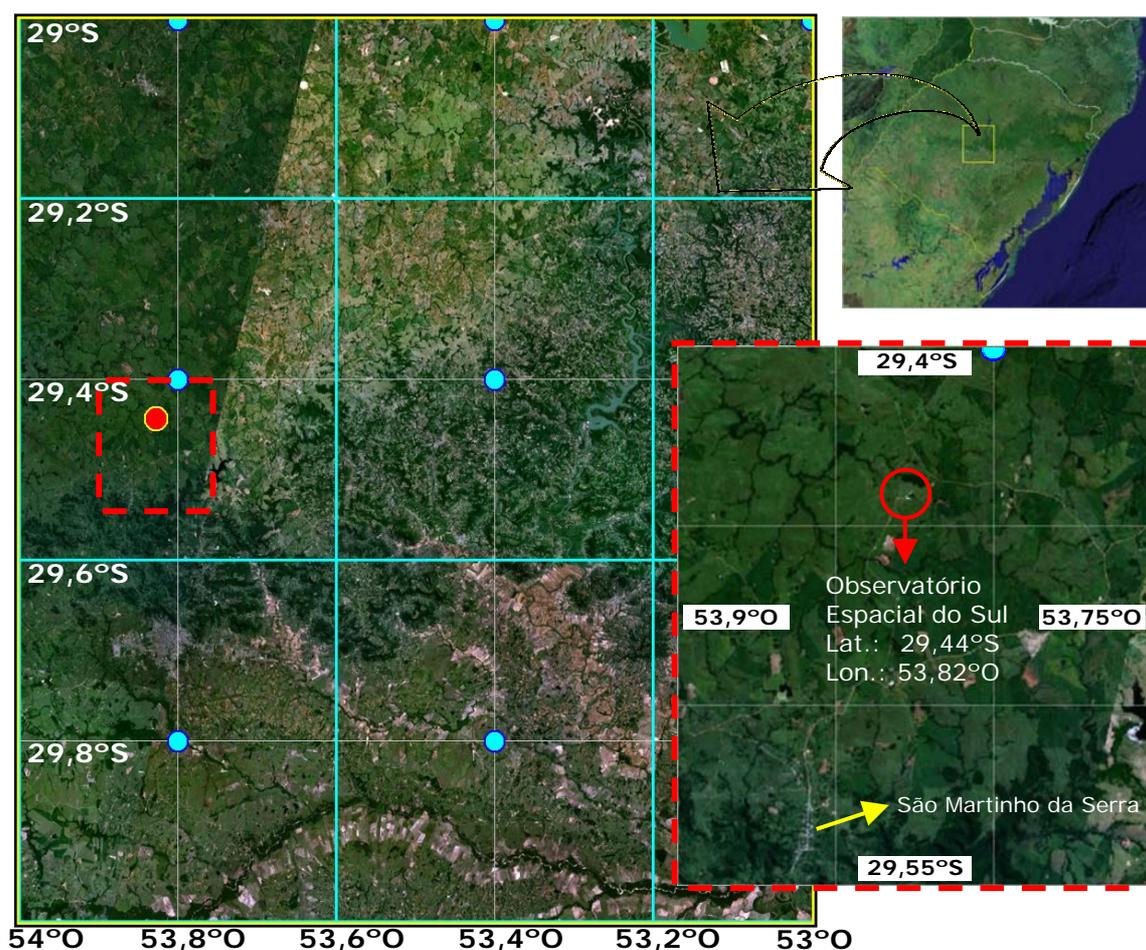


Figura 5.2: Imagens do satélite Landsat mostrando a localização do sítio da estação SMS. Na imagem maior são mostrados os pontos de grade do Eta e suas respectivas “caixas”.
Fonte: Modificado de NASA, 2005.

A Figura 5.3 apresenta um conjunto de imagens equivalente à Figura 5.2, porém mostrando a região de Florianópolis, bem como a localização exata da estação FLN e os pontos de grade do modelo Eta. Enquanto a estação SMS encontra-se bastante próxima do ponto central de uma “caixa” do modelo Eta, a estação FLN encontra-se exatamente

no limiar entre duas “caixas”. Assim, pressupôs-se que as condições observadas na estação sejam influenciadas pela “caixa” a Norte e pela “caixa” a Sul da estação. Os dados para ambos os pontos de grade – 27,4°S; 48,6°O e 27,8°S; 48,6°O – foram separados, e foi tomada a média de cada uma das 33 variáveis nos dois pontos como previsões do modelo Eta representativas da estação.

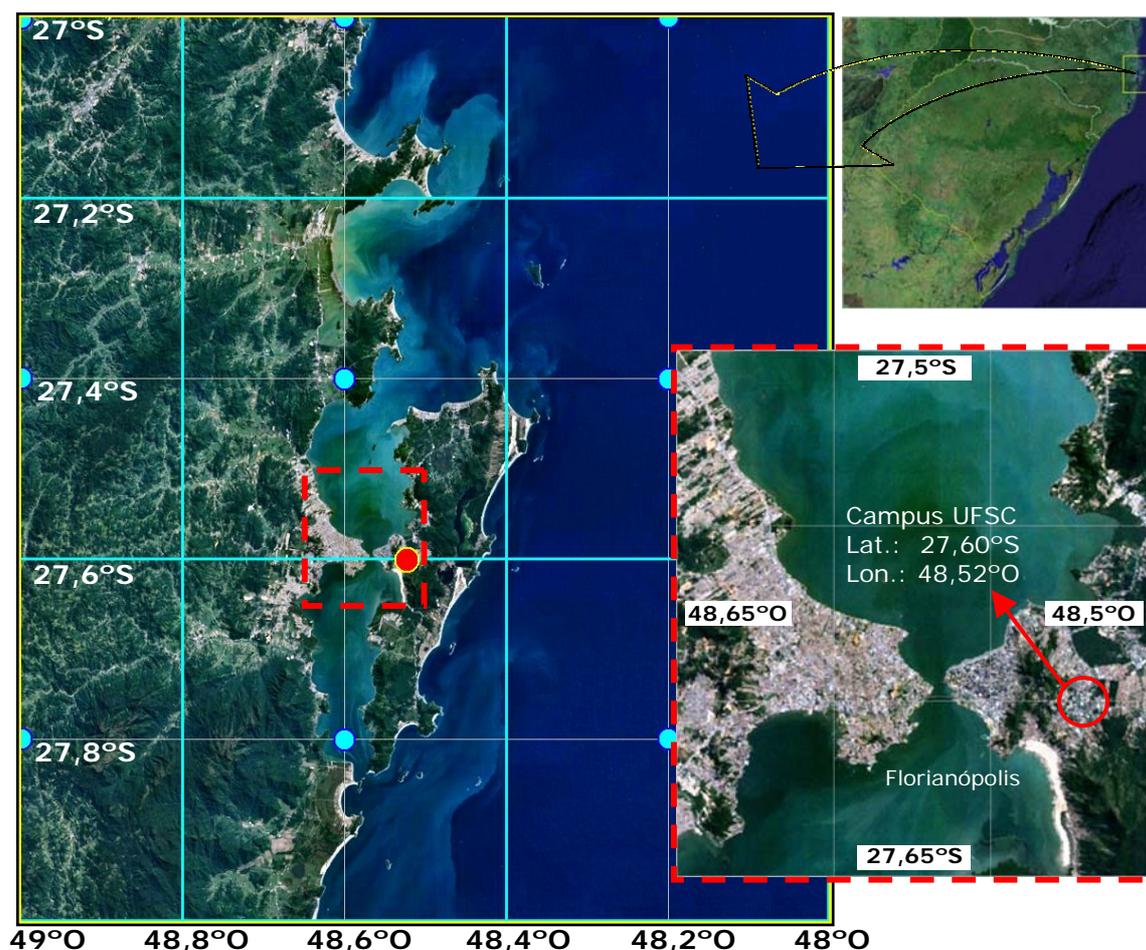


Figura 5.3: Imagens do satélite Landsat mostrando a localização do sítio da estação FLN. Na imagem maior são mostrados os pontos de grade do Eta e suas respectivas “caixas”.
Fonte: Modificado de NASA, 2005.

Para cada dia e para cada estação, foram gerados arquivos em formato texto para cada rodada do Eta (00UT e 12UT), contendo as 33 variáveis da Tabela 4.1, em cada um dos 13 ou 21 *tempos*. Todo o procedimento de “recorte” dos dados do Eta, até a geração dos arquivos para cada estação, em cada dia e rodada, foi automatizado pela criação de rotinas computacionais.

5.2. Dados Derivados de Saídas do Modelo Eta

A partir dos dados de temperatura (*tp2m*) e temperatura de ponto de orvalho (*dp2m*), fornecidos pelo modelo Eta para uma altura de 2 metros acima da superfície, foi calculada a umidade relativa. O cálculo foi realizado de acordo com Wallace e Hobbs (1977), através da razão entre as pressões de vapor de saturação do ar em relação à água (e_s) na temperatura de ponto de orvalho (T_d) e na temperatura ambiente (T_a):

$$RH = \frac{e_s(T_d)}{e_s(T_a)} = \frac{e_s(dp2m)}{e_s(tp2m)} \quad (5.1)$$

A pressão de vapor de saturação (e_s), em milibares (*mb*), para cada temperatura em graus Celsius ($^{\circ}C$), foi calculada de acordo com a seguinte equação (Bolton, 1980):

$$e_s(T) = 6,112 \cdot e^{\left(\frac{17,67 \cdot T}{T+243,5}\right)} \quad (5.2)$$

Outra nova variável foi criada utilizando os dados das componentes zonal (*u10m*) e meridional (*v10m*) do vento a 10 metros de altura: a velocidade resultante do vento a 10 metros (*wnds* – abreviatura para *wind speed*):

$$wnds = \sqrt{(u10m)^2 + (v10m)^2} \quad (5.3)$$

Estas duas variáveis novas, umidade relativa e velocidade do vento, foram calculadas para cada rodada e cada *tempo*, de todos os dias que compõem o conjunto de dados do Eta. Assim como as variáveis de temperatura e das componentes de vento, essas novas variáveis são instantâneas, previstas para o horário do *tempo* considerado. A umidade relativa (representada, a partir daqui, por *rh2m*) é adimensional, com valores entre 0 e 1, e a velocidade resultante do vento é dada em metros por segundo.

5.3. Qualificação e Integração dos Dados de Radiação Solar

Devido à recente implantação das estações SONDA e a fase de qualificação e disponibilização dos dados ainda estar em curso, alguns defeitos foram encontrados em

seus dados. Embora esta etapa tenha demandado bastante tempo e tenha tido como resultado um conjunto de dados qualificado e coerente, foge do escopo deste trabalho detalhar todos os tipos de erros encontrados e suas correções. Como primeira etapa de redução dos dados de radiação global, os dados de ambas as estações foram visualizados graficamente, juntamente com dados teóricos de radiação solar incidente no topo da atmosfera, de maneira a verificar inconsistências, valores espúrios ou mesmo ausência de dados. Em seguida, correções pontuais foram aplicadas aos dados, quando tal procedimento era possível, e alguns dados foram excluídos. Vale mencionar que a estação SMS possui uma série de dados praticamente ininterrupta desde o início de operação, enquanto a estação FLN possui alguns intervalos sem dados ou com dados que tiveram de ser excluídos.

Com base no dia do ano, no minuto do dia, e na latitude da estação, o ângulo zenital solar foi calculado para cada minuto do dia, para todos os dias dos conjuntos de dados. Calculados o ângulo zenital solar (equação 2.13) e o fator de correção de excentricidade (equação 2.9), a irradiância solar incidente no TDA (em $W \cdot m^{-2}$) foi calculada para cada minuto (equação 2.11), para ambas as estações. A massa de ar também foi calculada para cada minuto, utilizando-se a equação 2.14.

Como a resolução temporal dos dados SONDA de radiação global é de 1 minuto, e os dados do Eta são fornecidos para intervalos de 6 horas, as irradiâncias disponíveis nas estações (médias de cada minuto) foram integradas. Assim, para cada período de 6 horas, obteve-se a energia radiante integrada, expressa em Joules por metro quadrado ($J \cdot m^{-2}$). Da mesma forma, as irradiâncias solares no TDA, calculadas com resolução de 1 minuto, foram integradas nos mesmos períodos de 6 horas. Também foi tomado o valor médio do ângulo zenital solar e da massa de ar nesses períodos (não tendo sido incluídos nas médias os valores equivalentes a instantes anteriores ao nascer e posteriores ao pôr-do-sol). Estas integrais e valores médios foram armazenados com a hora final dos períodos de 6 horas que representam. Por exemplo, a integral entre 12UT e 18UT foi armazenada como 18UT, similarmente ao que ocorre com os valores médios ou acumulados nas saídas do modelo Eta.

Os cálculos teóricos da radiação solar incidente no TDA, do ângulo zenital solar e da massa de ar, foram realizados com o objetivo de, inserindo tais dados no conjunto de preditores de radiação solar global incidente, dar aos modelos de RNA e RLM uma estrutura mais assemelhada a um modelo de cálculo de radiação. A inclusão da radiação solar no TDA justifica-se pelo fato de esta grandeza ser o valor limite de energia solar radiante que entra na atmosfera, e por ser a interação desta radiação com as condições atmosféricas que determina a incidência de radiação na superfície. O uso da radiação no TDA permite também embutir a informação de sazonalidade nos modelos ajustados. A inclusão do ângulo zenital solar médio e da massa de ar média deve-se ao fato dessas variáveis quantificarem o comprimento do caminho percorrido pela radiação solar ao longo da atmosfera, expressando a intensidade da influência das condições atmosféricas sobre a radiação.

5.4. Conjuntos de Dados Disponíveis

Como mencionado em itens anteriores, algumas variáveis de saída do modelo Eta são médias ou quantidades acumuladas em um intervalo de 6 horas que antecede cada *tempo* nos arquivos de previsões. Por exemplo, a variável *ocis*, que representa a radiação solar incidente prevista pelo Eta, é apresentada como média das irradiâncias observadas desde o *tempo* anterior até o *tempo* considerado. Outras variáveis são instantâneas e correspondem especificamente aos horários de cada *tempo*, como é o caso dos valores de temperatura, vento e pressão. Tanto os dados instantâneos como os valores médios foram transformados antes da sua utilização nos ajustes dos modelos de RNA e RLM, sendo que apenas os dados acumulados de precipitação permaneceram inalterados.

Os dados de radiação e de fluxos de calor, que são originalmente disponibilizados como médias de 6 horas em Watts por metro quadrado ($\text{W}\cdot\text{m}^{-2}$) foram apenas multiplicados por 21600 segundos (6 horas), sendo então convertidos em estimativas de integrais de energia acumulada em Joules por metro quadrado ($\text{J}\cdot\text{m}^{-2}$), no período considerado. Assim, os dados de radiação do Eta passaram a ter equivalência de unidades com os

dados de radiação medida pelas estações SONDA que, como descrito na seção anterior, foram integrados em intervalos equivalentes de 6 horas.

Os dados instantâneos para um *tempo* específico permitem que apenas se tenha informações do instante final do período de 6 horas precedente, em que as medidas de radiação foram integradas. Com o intuito de utilizar, nos modelos a serem ajustados, dados que representem melhor as condições atmosféricas e de superfície ao longo dos períodos em que a radiação foi integrada, procurou-se tomar também informações do instante inicial desses períodos, que estão disponíveis se forem considerados os dados do *tempo* anterior. Assim, as variáveis instantâneas foram transformadas, de maneira que seus valores em um dado *tempo* foram substituídos pela média dos valores do *tempo* em questão e do *tempo* anterior.

A Figura 5.4 ilustra as integrações e transformações aplicadas aos dados de radiação das estações SONDA e aos dados do modelo Eta, para um dado dia do conjunto de dados. Embora sejam apresentadas as modificações realizadas em um arquivo diário da rodada de 00UT do Eta, o mesmo foi realizado com os dados da rodada 12UT de cada dia. Como anteriormente explanado, os dados de radiação medida na superfície e calculada para o TDA são integrados e armazenados com a hora final dos 4 períodos diários. A figura também apresenta a substituição dos dados instantâneos, pelas médias dos dados entre o *tempo* corrente e o anterior. É importante salientar que, após as transformações aplicadas, todo o conjunto de dados passou a ter resolução temporal de 6 horas, sendo que cada dado é armazenado com a hora final do período que representa.

Previsões com resolução temporal de 6 horas, sejam valores instantâneos ou médios nesses períodos, não permitem o detalhamento de toda a variabilidade existente ao longo de um dia. Além disso, o uso de dados médios ou integrados em períodos de 6 horas acarreta algumas limitações aos modelos de RLM e RNA. Com a utilização de valores médios em um período relativamente longo, acaba-se por remover parte das não-linearidades inerentes à transferência de radiação na atmosfera. Isto é, enquanto na realidade ocorre grande variabilidade das condições atmosféricas, da radiação solar e da própria inter-relação entre as variáveis ao longo de um período de 6 horas, ao tomar-se

valores médios acaba-se por simplificar as influências dos preditores no preeditando, por assumir-se um único peso de contribuição dos preditores para com o preeditando, ao longo do período.

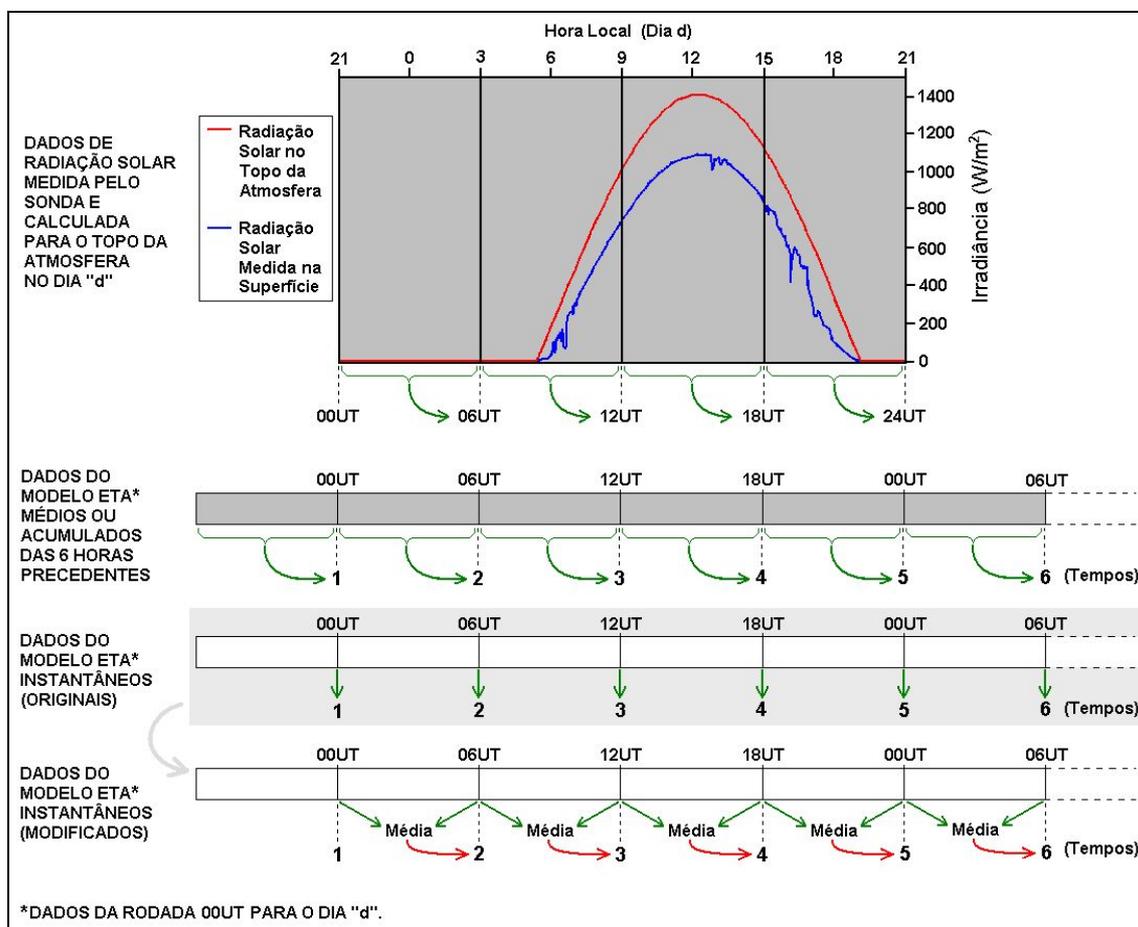


Figura 5.4: Transformações aplicadas aos dados de radiação SONDA e do modelo Eta.

Por outro lado, a adoção de previsões de 6 horas, evita a ocorrência de grandes erros que ocorreriam se fosse utilizada previsão horária ou em intervalos de tempo ainda mais curtos. A imprevisibilidade, com resolução temporal maior, de certos processos atmosféricos (como cobertura de nuvens e precipitação), por um modelo meteorológico (Eta), poderia resultar em previsões que, na realidade, não corresponderiam às condições atmosféricas que viriam a ser observadas no instante de medida do preeditando. Uma vez que a radiação incidente é fortemente dependente das condições atmosféricas, especialmente nuvens, os erros de previsão poderiam resultar na constituição de um conjunto incoerente de preditores e preeditando. Essa incoerência

levaria à dificuldade de aprendizado de RNAs, ou simplesmente impossibilitaria a construção de relações entre as variáveis em qualquer modelo empírico.

Como os arquivos de saída do modelo Eta guardam previsões para vários *tempos* futuros a partir do horário da rodada, é possível obter-se informações para um dado dia com diferentes intervalos de tempo de antecedência. Por exemplo, as informações do Eta para um dado dia, podem ser obtidas a partir da rodada da 00UT do dia em questão, ou das rodadas 12UT e 00UT do dia anterior. Obviamente, quanto mais recente a previsão dada pelo Eta, mais precisa ela será. Contudo, às vezes faz-se necessária uma previsão com maior antecedência, mesmo que esta apresente uma margem de erro maior. Visando um breve estudo sobre o crescimento do erro dos modelos ajustados e do próprio Eta, à medida que se aumenta a antecedência da previsão, os dados de radiação medida (variável alvo a ser ajustada pelas RNAs ou por RLM, isto é, o preditando) foram agrupados com dados previstos pelo Eta (preditores dos modelos de RNA e RLM) a partir de diferentes rodadas, como ilustrado pela Figura 5.5.

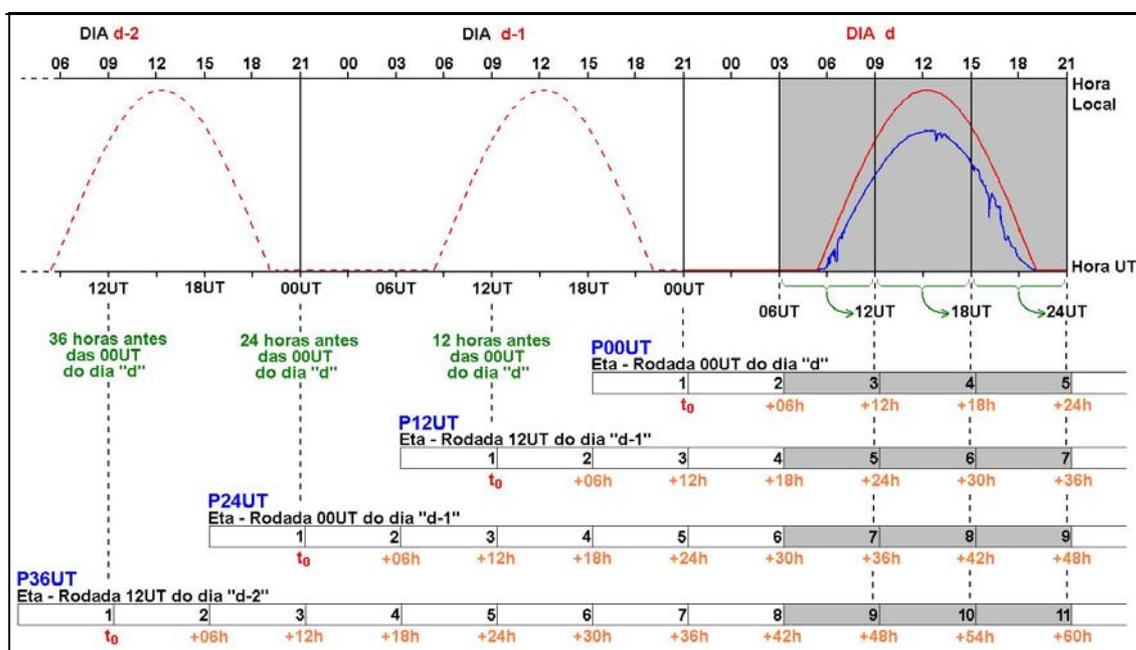


Figura 5.5: Agrupamento dos dados de radiação com previsões de diferentes antecedências.

Assim, foram gerados arquivos contendo preditando e preditores, com os preditores provenientes de diferentes rodadas anteriores. Na Figura 5.5 são mostrados os *tempos*

das rodadas do Eta que equivalem aos horários das integrais não-nulas de radiação para um dado dia. Os arquivos com preditores provenientes da rodada 00UT do mesmo dia dos dados de radiação medida serão denominados **P00UT**. Quando os preditores são tomados das rodadas de dias anteriores aos dados de radiação, os conjuntos de dados serão denominados **P12UT**, **P24UT**, **P36UT**, **P48UT**, e assim sucessivamente, indicando que os preditores para cada dia são provenientes de previsões geradas, respectivamente, 12, 24, 36 e 48 horas antes do início dos dias.

O conjunto total de preditores utilizados é apresentado na Tabela 5.1, onde cada variável é apresentada por uma sigla de 4 letras. As variáveis provenientes do Eta utilizam a mesma sigla padrão empregada nos dados de saída do modelo (Tabela 4.1). As novas variáveis derivadas do Eta, umidade relativa e velocidade do vento, são representadas por *rh2m* e *wnds*, respectivamente. O ângulo zenital médio, a massa de ar média e a radiação solar no topo da atmosfera receberam respectivamente as siglas *szam*, *airm* e *rtoa*. O preditendo, ou dado alvo a ser modelado, isto é, a radiação solar global medida, é representada por *rsfc*.

As variáveis do modelo Eta para radiação, fluxos de calor, precipitação e nuvens foram agrupadas e poderá se fazer referência aos grupos através dos nomes indicados na Tabela 5.1. O grupo *Sol* representa os valores calculados para a radiação solar no TDA, o ângulo zenital solar e a massa de ar. A variável *neve* (Tabela 4.1) foi excluída do conjunto de dados, por apresentar valores nulos para todo o período de dados. A variável *pslm* também foi excluída, pois apresentava perfeita correlação com a variável *pslc*, que é a variável preferida para a pressão das estações.

Após verificação da qualidade dos dados de radiação medida e das previsões do modelo Eta, e após exclusão dos dias com dados nulos ou defeituosos, os arquivos P00UT, P12UT, P24UT, P36UT e P48UT criados passaram a conter informações, respectivamente, para 472, 470, 472, 470 e 472 dias para a estação SMS (entre julho/2004 e outubro/2005) e 1150, 1149, 1149, 1148 e 1148 dias para a estação FLN (entre janeiro/2002 e outubro/2005). Para cada dia, os arquivos P00UT, P12UT, ..., P48UT possuem valores de preditendo e preditores para os 3 períodos diários (de 6

horas cada), representados pela hora final dos períodos, em que a integral de radiação é não-nula: 12UT, 18UT, 24UT (Figura 5.5).

Tabela 5.1: Preditando e preditores a serem utilizados no ajuste dos modelos.

| Preditando | Descrição (unidades) | |
|-------------------|---|----------|
| <i>rsfc</i> | Radiação solar global medida - integral de 6 horas ($J \cdot m^{-2}$) | |
| Preditores | Descrição (unidades) | |
| <i>szam</i> | Ângulo zenital solar médio de 6 horas (graus) | Sol |
| <i>airm</i> | Massa de ar média de 6 horas (adimensional) | |
| <i>rtoa</i> | Radiação solar no topo da atmosfera - integral de 6 horas ($J \cdot m^{-2}$) | |
| <i>pslc</i> | Pressão à superfície (hPa) | |
| <i>tp2m</i> | Temperatura a 2 metros (K) | |
| <i>dp2m</i> | Temperatura do ponto de orvalho a 2 metros (K) | |
| <i>rh2m</i> | Umidade relativa a 2 metros (0 a 1 - adimensional) | |
| <i>u10m</i> | Componente zonal do vento a 10 metros ($m \cdot s^{-1}$) | |
| <i>v10m</i> | Componente meridional do vento a 10 metros ($m \cdot s^{-1}$) | |
| <i>wnds</i> | Velocidade do vento a 10 metros ($m \cdot s^{-1}$) | |
| <i>tsfc</i> | Temperatura da superfície (K) | |
| <i>qsfc</i> | Umidade específica à superfície ($kg(H_2O) \cdot kg(ar)^{-1}$) | |
| <i>tgsc</i> | Temperatura da camada superficial do solo - 10 cm (K) | |
| <i>ussl</i> | Umidade do solo à superfície (0 a 1 - adimensional) | |
| <i>albe</i> | Albedo (0 a 1 - adimensional) | |
| <i>cape</i> | Energia potencial convectiva disponível ($m^2 \cdot s^{-2}$) | |
| <i>cine</i> | Energia para inibição da convecção ($m^2 \cdot s^{-2}$) | |
| <i>agpl</i> | Água precipitável instantânea ($kg \cdot m^{-2}$) | |
| <i>pcbs</i> | Pressão da base da nuvem (hPa) | |
| <i>pctp</i> | Pressão no topo da nuvem (hPa) | |
| <i>lwnv</i> | Fração de cobertura de nuvens baixas (0 a 1 - adimensional) | Nuvem |
| <i>mdnv</i> | Fração de cobertura de nuvens médias (0 a 1 - adimensional) | |
| <i>hinv</i> | Fração de cobertura de nuvens altas (0 a 1 - adimensional) | |
| <i>cbnt</i> | Fração média de cobertura de nuvens (0 a 1 - adimensional) | |
| <i>prec</i> | Precipitação total ($kg \cdot m^{-2} \cdot dia^{-1}$) | Chuva |
| <i>prcv</i> | Precipitação convectiva ($kg \cdot m^{-2} \cdot dia^{-1}$) | |
| <i>prge</i> | Precipitação de larga escala ($kg \cdot m^{-2} \cdot dia^{-1}$) | |
| <i>clsf</i> | Fluxo de calor latente à superfície (integral de 6 horas - $J \cdot m^{-2}$) | Fluxos |
| <i>cssf</i> | Fluxo de calor sensível à superfície (integral de 6 horas - $J \cdot m^{-2}$) | |
| <i>ghfl</i> | Fluxo de calor para o solo (integral de 6 horas - $J \cdot m^{-2}$) | |
| <i>ocis</i> | Radiação de onda curta descend. na superf. (int. 6h - $J \cdot m^{-2}$) | Radiação |
| <i>olis</i> | Radiação de onda longa descend. na superf. (int. 6h - $J \cdot m^{-2}$) | |
| <i>oces</i> | Radiação de onda curta ascend. na superf. (int. 6h - $J \cdot m^{-2}$) | |
| <i>oles</i> | Radiação de onda longa ascend. na superf. (int. 6h - $J \cdot m^{-2}$) | |
| <i>roce</i> | Radiação de onda curta ascend. no topo da atmosf. (int. 6h - $J \cdot m^{-2}$) | |
| <i>role</i> | Radiação de onda longa ascend. no topo da atmosf. (int. 6h - $J \cdot m^{-2}$) | |

Os modelos de RNA e RLM foram ajustados com os dados de cada um desses 3 períodos diários, individualmente. Isto foi feito devido ao fato de que a consideração dos 3 períodos conjuntamente levaria a constituir um grupo de dados estatisticamente não-homogêneo, uma vez que seus intervalos de valores diferem consideravelmente. Um conjunto de dados não-homogêneo pode levar a elevadas falsas-correlações entre variáveis, por apresentar grupos de valores com distintas magnitudes.

Para se verificar as diferenças de magnitude entre os períodos de radiação integrada, a Figura 5.6 apresenta curvas de radiação e integrais de energia radiante para as estações FLN e SMS em dias de céu limpo próximos aos solstícios de inverno e verão. Verifica-se que, enquanto os dados para 18UT oscilam entre aproximadamente 63 e 80% do total diário de energia radiante, os dados para 12UT e 24UT oscilam entre aproximadamente 4 e 16% e entre 13 e 23%, respectivamente.

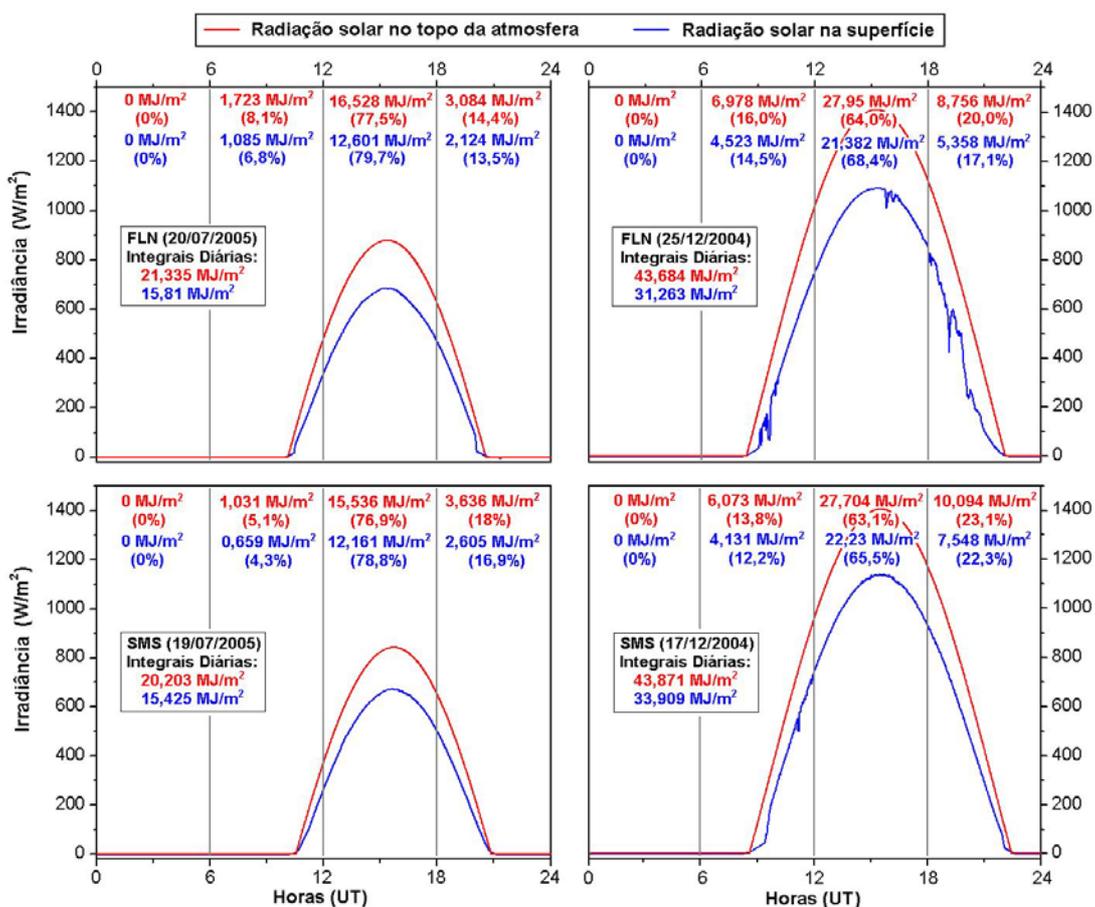


Figura 5.6: Integrais de energia radiante para dias ensolarados para as estações FLN e SMS.

Assim, os arquivos P00UT, P12UT, ..., P48UT de cada estação foram subdivididos passando a conter apenas dados para determinados períodos de cada dia. Por exemplo, os arquivos P00UT, para ambas as estações, foram subdivididos em **P00UT-Rad12UT**, **P00UT-Rad18UT** e **P00UT-Rad24UT**. O ajuste de modelos utilizando integrais diárias de radiação foi evitado, uma vez que se faria necessária a obtenção de valores diários para os preditores, utilizando média. Isso atribuiria o mesmo peso aos valores de um preditor de cada período de 6 horas, enquanto na realidade, os preditores do período 18UT (12UT a 18UT, ou 9 às 15 horas locais) são os de maior importância para a quantidade total diária de radiação incidente, pois modulam o período de maior intensidade de radiação que atravessa a atmosfera.

5.5. Treinamento das RNAs e Ajustes de RLMs

Neste trabalho, são empregadas RNAs do tipo MLP (perceptrons de múltipla camada). A implementação computacional das RNAs foi realizada utilizando um conjunto de ferramentas do software MATLAB[®], chamado *Neural Networks Toolbox* (Demuth *et al.*, 2005), que permite a construção e treinamento de RNAs com recursos recentes da área de neurocomputação. Com comandos básicos de operação matricial foram realizadas regressões lineares múltiplas (RLMs) com os mesmos conjuntos de preditores utilizados como entradas nas RNAs.

Numa etapa preliminar, foram testadas várias estruturas de RNAs, com diferentes números de camadas e neurônios e com diferentes tipos de treinamento. Em concordância com a literatura, verificou-se que o treinamento *backpropagation* tradicional é bastante lento, levando a um tempo de convergência muito grande. Empregou-se então, o algoritmo de treinamento *R-prop*, que pareceu levar a melhores resultados e demandando menor tempo de aprendizado. Também foi verificado que RNAs com duas camadas escondidas levaram a resultados ligeiramente melhores do que o emprego de uma única camada escondida.

Para denominar a arquitetura de RNA utilizada, será utilizada a seguinte notação: *RNA-ne-n1-n2-ns*; em que *ne* é o número de entradas utilizadas pela rede, *n1* é o número de neurônios da primeira camada escondida, *n2* é o número de neurônios da segunda

camada escondida, e ns é o número de saídas. O número de saídas será sempre 1, pois almeja-se produzir apenas estimativas de radiação solar global incidente. Os neurônios das camadas escondidas utilizam funções de ativação do tipo tangente hiperbólica, e a camada de saída, funções de ativação lineares. Como bastante conhecido da literatura, a utilização de funções sigmoidais nas camadas escondidas é o que proporciona às RNAs tipo MLP a capacidade de aprender não-linearidades (Haykin, 2001).

Nos testes preliminares, observou-se que, para pequeno número de entradas, a estrutura $RNA-(n)-(2n)-(n)-1$ produz melhores resultados. Para grande número de entradas, tal estrutura leva a instabilidades, pois o número de parâmetros livres torna-se elevado em comparação ao número de exemplos de treinamento, produzindo resultados bastante incoerentes e até mesmo valores negativos de radiação. Para conjuntos de preditores maiores, a estrutura $RNA-(n)-(n)-(0,5n)-1$ foi empregada.

Como mencionado no Capítulo 3, um procedimento comum e que evita o *overfitting* da RNA é a subdivisão dos dados em conjunto de treinamento e validação (*treinamento com parada antecipada*). Durante o treinamento, a RNA extrai informações do conjunto de treinamento e acompanha a redução do erro no conjunto de validação. O treinamento é interrompido quando a curva do erro de validação apresentar uma inflexão, com o erro passando a aumentar. Um terceiro conjunto, totalmente independente do treinamento, é reservado para a avaliação dos modelos e comparação entre eles. É importante mencionar que, para o ajuste de qualquer modelo estatístico, é procedimento comum realizar o ajuste e o teste do modelo com conjuntos de dados independentes.

Assim, quando é apresentado à rede um conjunto de dados, com preditores oriundos de um arquivo P00UT-Rad18UT, por exemplo, contendo um número N de exemplos, ocorre a seguinte subdivisão: $0,5 \cdot N$ dos exemplos passam a pertencer a um *conjunto de treinamento*, $0,25 \cdot N$ a um *conjunto de validação* e os $0,25 \cdot N$ restantes compõe um *conjunto de testes*.

Os mesmos conjuntos de treinamento e validação utilizados para treinar uma RNA, contendo os mesmos preditores, são conjuntamente empregados para ajustar um modelo de RLM. O conjunto de testes é empregado por ambos os modelos ajustados, RNA e RLM, para simulação e cálculo dos erros de previsão. Essas simulações são comparadas

entre si e com dados da variável *ocis* (previsão de radiação de onda curta pelo Eta) para os mesmos dias do conjunto de testes.

Outras informações referentes ao treinamento das RNAs neste trabalho:

- Antes de ser iniciado o treinamento das RNAs, preditores e preditando são pré-processados através de uma normalização simples, com relação aos valores máximos e mínimos de cada variável no conjunto de treinamento, passando a assumir valores entre 0 e 1, de acordo com a expressão:

$$valor\ normalizado = \frac{valor - mínimo}{máximo - mínimo}$$

- No início do treinamento, os pesos da RNA são inicializados assumindo valores aleatórios.
- Ao longo do treinamento, para avaliar a performance do aprendizado, adotou-se uma função de custo similar à equação 3.17 com parâmetro de regularização igual a 0,5, buscando a minimização simultânea do erro quadrático médio e dos valores dos pesos, e assim contribuindo para evitar *overfitting*.

5.6. Estudos Planejados

Como mencionado anteriormente, o objetivo deste trabalho é ajustar modelos de RNAs e RLMs capazes de estimar a radiação solar incidente, tendo como entradas dados meteorológicos previstos pelo modelo Eta e valores calculados para radiação solar no TDA, para um instante futuro. Além de se tentar obter uma melhor previsão de radiação frente àquela disponibilizada diretamente pelo modelo Eta, pretende-se verificar se o uso de RNAs oferece ganhos em relação a um modelo linear (RLM), mais simples. Trata-se, pois, do desenvolvimento e avaliação de um refinamento estatístico da previsão de radiação solar do modelo Eta, através de um procedimento que faz o papel de um MOS (*Model Output Statistics*).

No ajuste de ambos os tipos de modelos, RNA e RLM, foi utilizada como preditando, ou variável-alvo a ser modelada, a variável *rsfc* – *radiação solar medida na superfície*. Os modelos foram ajustados separadamente para as estações SMS e FLN do projeto

SONDA. Os dados da variável *ocis* – radiação solar prevista pelo modelo Eta – constituem a previsão que se planeja aperfeiçoar com os novos modelos ajustados, e tais dados foram empregados para comparação com as novas previsões.

Como inúmeras combinações de preditores podem ser estudadas, um número limitado de experimentos e análises foi planejado. Para o teste de combinações de preditores com diferentes complexidades, buscando encontrar conjuntos ótimos de preditores, foram utilizados somente os arquivos P00UT-Rad18UT de cada estação. Essa escolha deve-se ao fato de que as previsões P00UT são as que produzem melhores previsões das condições atmosféricas e de superfície, por serem baseadas em condições iniciais mais recentes, e pelo fato de que a radiação solar entre 12:00UT e 18:00UT compreende a maior fração de energia diária, como observado na Figura 5.6. Assim, foram realizados diversos “experimentos”, empregando diferentes combinações de preditores em previsões P00UT-Rad18UT. Esses experimentos serão discutidos no item seguinte.

Utilizando o conjunto mais compacto de preditores que proporcionou previsões com menor erro, foram realizadas análises de energia solar total diária, pelo cálculo adicional de previsões P00UT-Rad12UT e P00UT-Rad24UT. Ainda, utilizando previsões de várias antecedências, foi também estudado o crescimento do erro com o uso de previsões P12UT, P24UT, P36UT e P48UT.

5.7. Definição dos Grupos de Preditores

Como se pode verificar na Tabela 5.1, existem 36 preditores que podem ser agrupados em conjuntos com complexidade variada, e empregados para ajuste de modelos de cálculo de radiação solar incidente. Testar e analisar todas as combinações possíveis sem utilizar um embasamento físico ou estatístico, demandaria um longo tempo. Para amparar a escolha de grupos de preditores testados em cada *experimento*, além dos argumentos físicos, foi utilizada uma ferramenta estatística denominada *regressão linear múltipla passo-a-passo (stepwise linear regression)*.

A regressão passo-a-passo é uma ferramenta bastante conhecida, utilizada para verificar quais os melhores preditores para um determinado preditando, levando em conta apenas

a dependência linear entre as variáveis. Neste tipo de análise, um modelo é desenvolvido pela inserção de um preditor a cada passo, a partir do conjunto total de preditores. O primeiro preditor inserido é aquele que apresentar maior correlação com o preditando, isto é, aquele que melhor explica a variância do preditando. A partir daí, a cada passo, é inserido um novo preditor que assegure a representação da maior fração de variância que não foi explicada pela inserção dos preditores anteriores.

Cada novo preditor possui a maior independência em relação ao conjunto de preditores que já compõe o modelo, e leva a explicar uma maior fração de variância dentre os preditores ainda não inseridos. Ao longo do processo, preditores podem também vir a ser excluídos do modelo, caso a inserção de outros preditores, consiga explicar melhor a variância, tornando-o desnecessário.

O critério adotado para entrada e saída no modelo foi o nível de significância (p) de cada variável independente (preditores) com relação à variável dependente (preditando) em cada passo. Foi fixado como critério de entrada $p < 0,05$ (intervalo de confiança de 95%) e como critério de saída $p > 0,1$ (intervalo de confiança de 90%). Assim, a variável que explicar uma fração de variância que levar a um menor valor de p em cada passo, será inserida. Se ao longo do processo uma variável inserida anteriormente apresentar $p > 0,1$ (deixar de pertencer ao intervalo de confiança de 90%), ela é excluída do modelo. A inserção de variáveis é encerrada quando, num dado passo, o menor valor de p , dentre as variáveis não inseridas, não satisfizer $p < 0,05$ (nenhum novo preditor é significativo a um nível de confiança de 95%).

Contudo, a não inserção de um preditor não significa que ele não possui qualquer informação que leve ao preditando, mas indica que outras variáveis permitem explicar melhor a variância. Ainda, a análise de regressão passo-a-passo leva em conta apenas a dependência linear entre as variáveis, e para o conjunto de dados específico empregado na análise, não sendo levados em conta aspectos físicos. Portanto, cabe uma análise criteriosa de seus resultados. Esta análise foi empregada durante os experimentos apenas como um critério objetivo para **auxiliar** a escolha de preditores. Sempre que possível, buscou-se empregar preditores fisicamente relacionados com o preditando.

No Capítulo 6, os resultados das regressões passo-a-passo, para os dados P00UT-Rad18UT de ambas as estações, são apresentados, juntamente com os resultados dos experimentos com diversos grupos de preditores.

5.8. Avaliação das Previsões

Para avaliação das previsões de radiação solar dadas pelas RNAs, pelos modelos baseados em RLM e pelo modelo Eta, os valores previstos foram comparados com os valores de radiação medidos na localidade de interesse. Os valores observados (O) e previstos (P) foram comparados graficamente em diagramas de dispersão, apresentados no Capítulo 6. Em cada diagrama de dispersão é apresentada a reta $O = P$, representando a previsão ideal. Também é apresentada uma reta determinada por um ajuste linear entre O e P , juntamente com o intercepto e o coeficiente angular, com único objetivo de observar o comportamento geral das previsões e os seus desvios em relação à previsão ótima ($O = P$).

A performance dos modelos foi verificada com alguns índices comumente empregados na literatura e descritos por vários autores (Jolliffe e Stephenson, 2003; WWRP/WGNE Joint Working Group on Verification, 2005):

– **Coeficiente de Correlação de Pearson entre Observações e Previsões (R):**

Medida de associação linear entre observações e previsões e é definido por:

$$R = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{P_i - \bar{P}}{S_P} \right) \left(\frac{O_i - \bar{O}}{S_O} \right) \quad (5.4)$$

em que S_P e S_O são, respectivamente, os desvios padrões das previsões e observações, \bar{P} e \bar{O} são as médias das previsões e observações, e N é o número de observações e suas respectivas previsões. Substituindo S_P e S_O pelas suas equações, o coeficiente de correlação entre P e O pode ser calculado por:

$$R = \frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2} \cdot \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (5.5)$$

Pode variar de -1 a 1, e seu valor para previsão perfeita é 1. Por ser independente da escala e por não levar em conta o viés da previsão, é possível encontrar uma correlação elevada para uma previsão com erros grandes (viés). É adimensional.

- **Coefficiente de Determinação (R^2):** Medida da fração de variância das observações explicada pelas previsões e é definido por:

$$R^2 = \frac{\sum_{i=1}^N (P_i - \bar{O})^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (5.6)$$

Pode ser calculado elevando o coeficiente de correlação (R) ao quadrado. Varia de 0 a 1, indicando que as previsões explicam, respectivamente, 0 a 100% da variância total das observações. É adimensional.

- **Viés ou Erro Sistemático Médio ($ME - Mean Error$):** Medida da média dos erros (Erro ou desvio = $P_i - O_i$) e permite verificar se a previsão é sistematicamente subestimada ou superestimada. Varia de infinito negativo a infinito positivo e seu valor para previsão perfeita é 0, mas pode atingir esse valor mesmo para uma previsão ruim, caso hajam erros compensatórios. Possui a mesma unidade das previsões e observações.

$$ME = \frac{1}{N} \sum_{i=1}^N (P_i - O_i) \quad (5.7)$$

- **Erro Absoluto Médio ($MAE - Mean Absolute Error$):** Medida da magnitude média dos erros de previsão. Varia de 0 a infinito e seu valor para previsão perfeita é 0. Possui a mesma unidade das previsões ou observações.

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (5.8)$$

- **Raiz do Erro Quadrático Médio ($RMSE - Root Mean Squared Error$):** Medida da magnitude média dos erros de previsão. Varia de 0 a infinito e seu

valor para previsão perfeita é 0. Como é uma média quadrática, erros maiores possuem maior influência sobre o valor final do que os erros menores. Possui a mesma unidade das previsões ou observações.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad (5.9)$$

Destes índices, dois deles são mais comumente empregados na literatura: o *ME* e o *RMSE*. Essas medidas de erro são comumente expressas como frações percentuais em relação à média das observações, isto é, como *erros relativos*. Assim, *ME%* e *RMSE%* representam as razões entre os valores de erro e a média das observações (\bar{O}), multiplicadas por 100.

Para comparar os modelos entre si, utilizando suas medidas de erro, utiliza-se um índice chamado *Skill* (destreza), que quantifica o ganho (ou aperfeiçoamento) obtido pelo uso de uma previsão, em relação ao uso de outra previsão disponível e tomada como referência. O *Skill* pode ser calculado com a seguinte expressão:

$$Skill(\acute{I}NDICE, ref) = \frac{\acute{I}NDICE - \acute{I}NDICE_{ref}}{\acute{I}NDICE_{perf} - \acute{I}NDICE_{ref}} \quad (5.10)$$

em que *ÍNDICE* pode ser qualquer dos índices de avaliação apresentados anteriormente, calculado para uma previsão a ser avaliada, *ÍNDICE_{perf}* é o valor para previsão perfeita desse índice (0, para ME, MAE ou RMSE; 1 para R ou R²), *ÍNDICE_{ref}* é o valor deste índice calculado para uma outra previsão tomada como referência.

Skill(ÍNDICE,ref) simboliza o valor de *Skill* obtido utilizando um índice de nome *ÍNDICE* e utilizando um previsão de referência de nome *ref*. No caso do viés (*ME*), que pode apresentar valores negativos e positivos, foi utilizado seu módulo. Um *Skill* igual a 1 significa que a previsão perfeita foi atingida e *Skill* = 0 significa que não houve ganho em relação à previsão de referência. Valores negativos significam que a previsão avaliada é pior do que a previsão de referência.

É comum, na avaliação de previsões, utilizar como previsão de referência a climatologia ou a *previsão por persistência*. A previsão por persistência consiste em tomar como previsão de radiação para o dia futuro, o valor conhecido (neste caso, a radiação incidente medida) no presente.

CAPÍTULO 6

RESULTADOS

Neste capítulo, os resultados obtidos pela comparação entre previsões e medidas de radiação solar são apresentados e discutidos. Em uma primeira parte, uma avaliação da previsão de radiação solar incidente dada pelo modelo Eta é apresentada, juntamente com uma descrição dos diagramas de dispersão e uma discussão sobre os erros médios. No item seguinte, previsões obtidas a partir de modelos RLM e RNA são avaliadas em experimentos utilizando diversos conjuntos de preditores, sendo apontado um conjunto mínimo de preditores que leva a uma previsão satisfatória. Avaliações de desempenho do modelo Eta e dos demais modelos ajustados, em cada estação do ano, são discutidas. Na última seção, uma avaliação de previsões realizadas com diferentes antecedências é apresentada, sendo disponibilizados gráficos de erros. Em cada seção são discutidos os resultados obtidos pelos novos modelos frente ao modelo Eta, e pelas RNAs frente aos modelos de RLM.

6.1. Avaliação da Previsão de Radiação Solar do Modelo Eta

Como mencionado nos capítulos anteriores, o modelo Eta possui um conhecido viés na previsão de radiação solar incidente na superfície. Este foi um dos fatores que motivaram a realização desse trabalho, uma vez que refinamentos e a obtenção de previsões com menores erros são necessários para diversas aplicações. Nesta seção, uma avaliação da previsão de radiação solar dada pelo modelo Eta (variável *ocis*) para as estações de Florianópolis (FLN) e São Martinho da Serra (SMS) é apresentada, sendo quantificados o viés e o RMSE.

A avaliação foi realizada apenas sobre as previsões P00UT (Figura 5.5) do modelo Eta, isto é, as previsões mais recentes disponíveis (antecedência mínima). Assim, as estimativas (previsões) de radiação para cada dia, geradas no início dos dias (00UT), foram comparadas com as medidas de radiação (observações), nos três períodos diários em que a incidência de radiação é não-nula: 06–12UT, 12–18UT e 18–24UT. Também

foram comparadas as integrais diárias das estimativas com as integrais diárias de energia solar medida, obtidas pelos somatórios dos três períodos diários.

Na Figura 6.1 são apresentados diagramas de dispersão para a estação FLN, em que as previsões (ordenadas) são comparadas com as observações (abscissas) de radiação solar. Foram empregados dados para 1150 dias entre janeiro/2002 e outubro/2005. Cada ponto nos diagramas representa a previsão e a observação referente a um dia do conjunto de dados. São apresentados diagramas de dispersão para as previsões P00UT-Rad12UT, P00UT-Rad18UT, P00UT-Rad24UT e para integrais diárias baseadas em P00UT.

Em todos os diagramas de dispersão apresentados neste capítulo, o comportamento geral das previsões (P) dos modelos, em função das observações (O), é descrito por ajustes lineares ($P = a + b \cdot O$) calculados com base em todos os pontos e representados por linhas vermelhas. Os coeficientes lineares (a) e angulares (b), bem como os coeficientes de correlação (R), calculados para os ajustes lineares, são apresentados nos cantos inferiores direitos de cada diagrama de dispersão. As linhas azuis diagonais, que passam pelas origens dos gráficos e possuem coeficientes angulares unitários, representam os casos ideais, isto é, elas representam os casos em que as previsões seriam perfeitas, assumindo o mesmo valor das observações ($P = O$).

As distâncias verticais dos pontos às linhas azuis, correspondem à diferença, ou desvios, da previsão em relação à observação ($P_i - O_i$). Observa-se, em todos os diagramas de dispersão da Figura 6.1, que, em geral, a maior parte dos pontos situa-se acima das linhas azuis, indicando que o modelo Eta estima intensidades de radiação maiores do que as efetivamente observadas, em todos os períodos diários considerados. O fato de as retas que representam o comportamento geral das previsões (vermelhas) situarem-se sempre acima das retas que representam as previsões ideais (azuis), enfatiza a tendência de o modelo sistematicamente superestimar a radiação solar, devido possivelmente, a uma deficiência na modelagem dos processos que atenuam a radiação solar, como por exemplo, o espalhamento por aerossóis. A ocorrência de maior superestimativa para os dias com menores valores de observações, também poderia estar ligada a uma geração deficiente da cobertura de nuvens em dias nublados, pelo modelo Eta.

Vale mencionar que os modelos meteorológicos, como o modelo Eta, são empregados para geração de campos de variáveis, não sendo objetivada a previsão pontual. As previsões para um dado ponto de grade representam uma condição média prevista para uma área extensa. Assim, ao compararmos as previsões com observações, os erros avaliados não representam apenas o erro do procedimento de cálculo do modelo, mas também englobam os erros inerentes a comparar um campo com um ponto específico dentro do campo, bem como os erros decorrentes de imprecisões nas condições iniciais e de contorno empregadas pelo modelo Eta. O processo de refinamento busca, justamente, “corrigir” o valor do campo para o ponto de interesse e remover erros sistemáticos decorrentes de imprecisões existentes no modelo e nos dados por ele empregados.

A partir dos desvios, calculados entre previsões e observações em cada dia, foram calculados o viés, ou erro sistemático (ME), o erro absoluto médio (MAE) e a raiz do erro quadrático médio (RMSE), de acordo com as equações 5.7, 5.8 e 5.9, respectivamente, e os resultados são apresentados na Tabela 6.1. Ao longo deste capítulo, estes erros médios serão referidos apenas por suas siglas: ME, MAE e RMSE.

Como mencionado no Capítulo 5, o ME permite avaliar o erro sistemático do modelo em superestimar ou subestimar suas previsões, fornecendo uma estimativa média de quanto o modelo erra para mais ou para menos suas previsões. Um ME positivo indica que o modelo superestima as previsões de radiação.

Por sua vez, os erros médios MAE e RMSE são estimativas médias dos módulos dos desvios entre previsões e observações, isto é, são medidas do tamanho médio dos desvios, sem considerar seus sinais. O MAE e o RMSE medem, portanto, a distância vertical média dos pontos à linha azul, independentemente de os pontos estarem situados acima ou abaixo dela nos diagramas de dispersão, podendo ser considerados medidas do “espalhamento” da nuvem de pontos em torno da linha de previsão perfeita. A diferença entre os dois é que o MAE é a média aritmética dos módulos dos desvios, enquanto o RMSE é uma média quadrática, sendo mais influenciado pelos maiores desvios entre previsão e observação.

FLN – P00UT – Modelo Eta (*ocis*)

N = 1150 dias (jan/2002-out/2005)

— P = O
— P = a + b*O

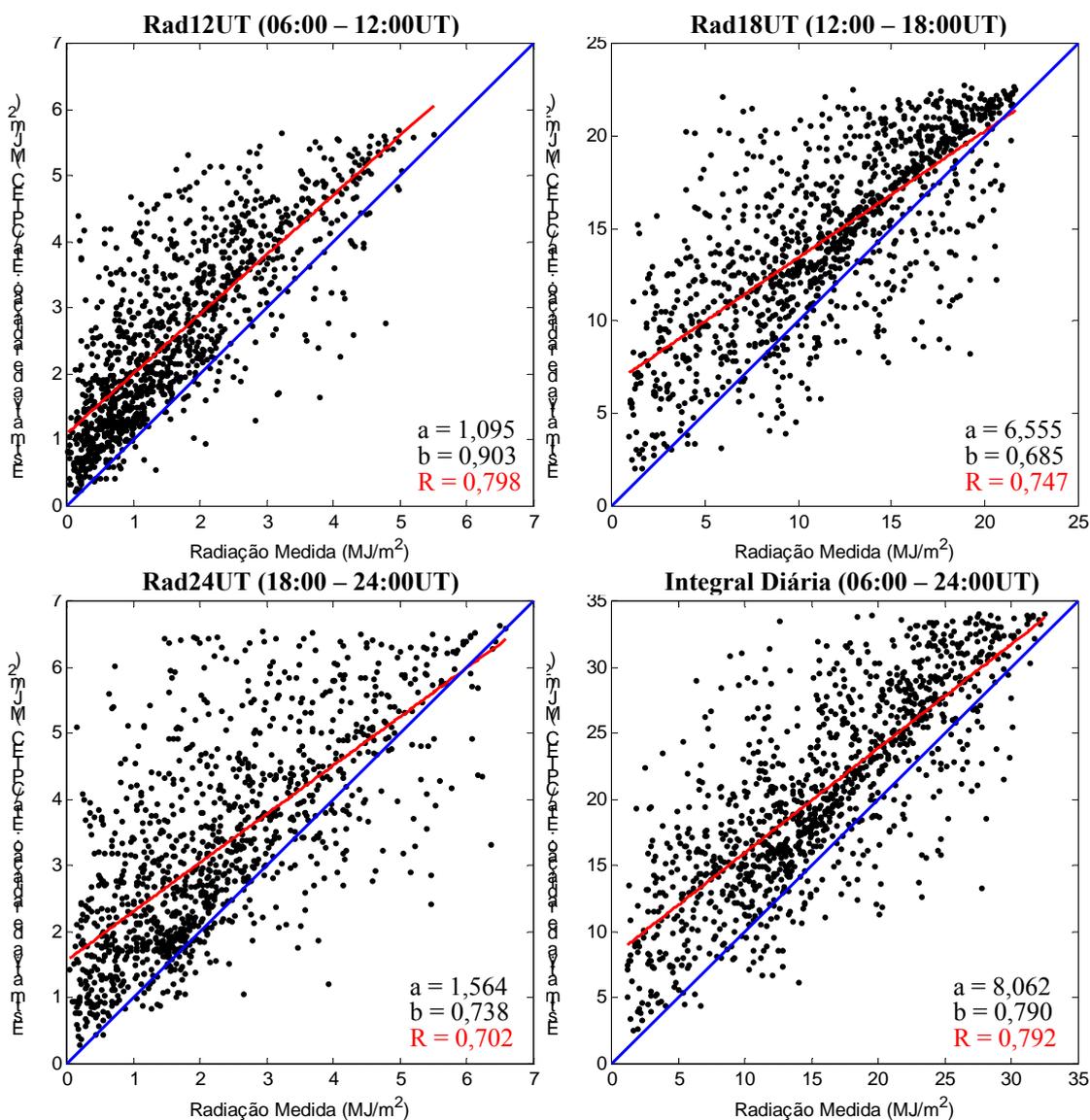


Figura 6.1: Diagramas de dispersão para as previsões P00UT de radiação solar incidente do modelo Eta (variável *ocis*) para três períodos diários e integral diária, para FLN.

Tabela 6.1: Parâmetros de avaliação relativos aos diagramas de dispersão da Figura 6.1 (FLN).

| Índices | Rad12UT | Rad18UT | Rad24UT | Integral Diária |
|----------------|--------------|---------------|--------------|-----------------|
| R | 0,798** | 0,747** | 0,702** | 0,792** |
| R ² | 0,636 | 0,558 | 0,493 | 0,627 |
| ME* | 0,934 | 2,887 | 0,979 | 4,800 |
| MAE* | 0,997 | 3,688 | 1,142 | 5,498 |
| RMSE* | 1,241 | 4,635 | 1,509 | 6,757 |
| \bar{O} * | 1,663 | 11,667 | 2,232 | 15,562 |
| ME% | 56,2 % | 24,7 % | 43,8 % | 30,8 % |
| RMSE% | 74,6 % | 39,7 % | 67,6 % | 43,4 % |

* Unidade: MJ/m².

** Nível de significância p<0,001.

É importante mencionar, nesta breve discussão sobre os erros médios, que o erro sistemático (ME) influencia as medidas de erro aleatório MAE e RMSE. Ao haver um viés elevado (ME diferente de zero), positivo ou negativo, as previsões encontram-se sistematicamente afastadas das observações, contribuindo para um maior valor nos módulos dos desvios e conduzindo a maiores valores de MAE e RMSE. Assim, um outro parâmetro de avaliação empregado para medir o “espalhamento” dos pontos, é o coeficiente de correlação (R). O coeficiente de correlação mede o grau de linearidade observado entre previsões e observações, e grosseiramente pode-se dizer que mede o espalhamento dos pontos em torno da linha vermelha nos diagramas de dispersão, não sendo influenciado pelo viés.

Além dos erros médios, a Tabela 6.1 apresenta os valores dos coeficientes de correlação (R) e determinação (R^2), para cada diagrama de dispersão da Figura 6.1, e os valores médios das 1150 observações (\bar{O}) em cada período diário. A razão dos erros médios pelas médias das observações, permite a determinação dos erros médios percentuais, ME% e RMSE%, também apresentados. Em concordância com os diagramas de dispersão, verifica-se que o ME é sempre positivo para todos os períodos diários, e o ME% revela um viés relativamente elevado nas previsões do modelo Eta.

Verifica-se, na tabela 6.1, que os maiores valores para ME, MAE e RMSE ocorrem para o período Rad18UT. Porém, como geralmente neste período ocorre a maior incidência de radiação, os erros médios relativos, ME% e RMSE%, são menores dos que os observados nos demais períodos diários.

Observa-se ainda que, apesar do modelo Eta possuir um viés elevado e um RMSE também expressivo, os coeficientes de correlação indicam bons níveis de correlações entre previsões e observações, sempre superiores a 0,7. Os coeficientes de determinação (R^2) expressam que entre aproximadamente 50 e 65% da variância das observações é comum às previsões, considerando os quatro períodos avaliados.

O exame conjunto dos diagramas de dispersão e dos resultados da tabela 6.1, revela que, como esperado, o período Rad18UT é o que mais influencia a Integral Diária de radiação. Tanto o comportamento dos diagramas de dispersão como os valores

percentuais dos erros médios, ME% e RMSE%, são bastante parecidos para Rad18UT e para Integral Diária. Portanto, a realização dos experimentos empregando apenas os preditores de previsões P00UT-Rad18UT, na seção seguinte, é plausível, pois além de o período Rad18UT abranger a maior fração diária de energia solar (Figura 5.6), este período é bastante representativo da Integral Diária, no que diz respeito aos erros.

Uma observação importante decorrente dos diagramas de dispersão para Rad18UT e para integral diária, é que apesar de haver um viés deslocando os pontos, observam-se conjuntos densos de pontos dispostos em linha, bastante paralelos às linhas azuis. Isto mostra que apesar de todos os erros inerentes a um modelo de previsão e todas as dificuldades ligadas à imprevisibilidade e não-linearidades atmosféricas, o modelo Eta possui uma considerável habilidade em simular a radiação solar incidente, especialmente para os dias com maior incidência de radiação.

A avaliação das previsões de radiação solar do modelo Eta para São Martinho da Serra, foi realizada da mesma maneira que para Florianópolis e resultou nos diagramas de dispersão da Figura 6.2 e nos parâmetros de avaliação apresentados na Tabela 6.2. Foram utilizados, neste caso, dados para 472 dias entre julho/2004 e outubro/2005.

Apesar de ser um conjunto de dados menor do que o disponível para Florianópolis, os erros médios observados, bem como seus valores percentuais (ME% e RMSE%), são bastante parecidos para as duas estações em cada período diário. Contudo, foram observados coeficientes de correlação ligeiramente maiores para SMS em comparação com FLN, e os coeficientes de determinação indicam que as frações das variâncias das observações comuns às previsões, variam entre aproximadamente 60 e 75%.

Assim como observado para FLN, os diagramas de dispersão para Rad18UT e para Integrais Diárias para SMS apresentam conjuntos densos de pontos em que as previsões são aproximadamente iguais às observações, especialmente em dias com maior incidência de radiação. Porém, novamente verifica-se um erro sistemático, e as previsões para SMS parecem estar deslocadas para valores maiores do que as observações, destacando a tendência do modelo Eta em superestimar a radiação solar.

SMS – P00UT – Modelo Eta (*ocis*)

N = 472 dias (jul/2004-out/2005)

— P = O
— P = a + b*O

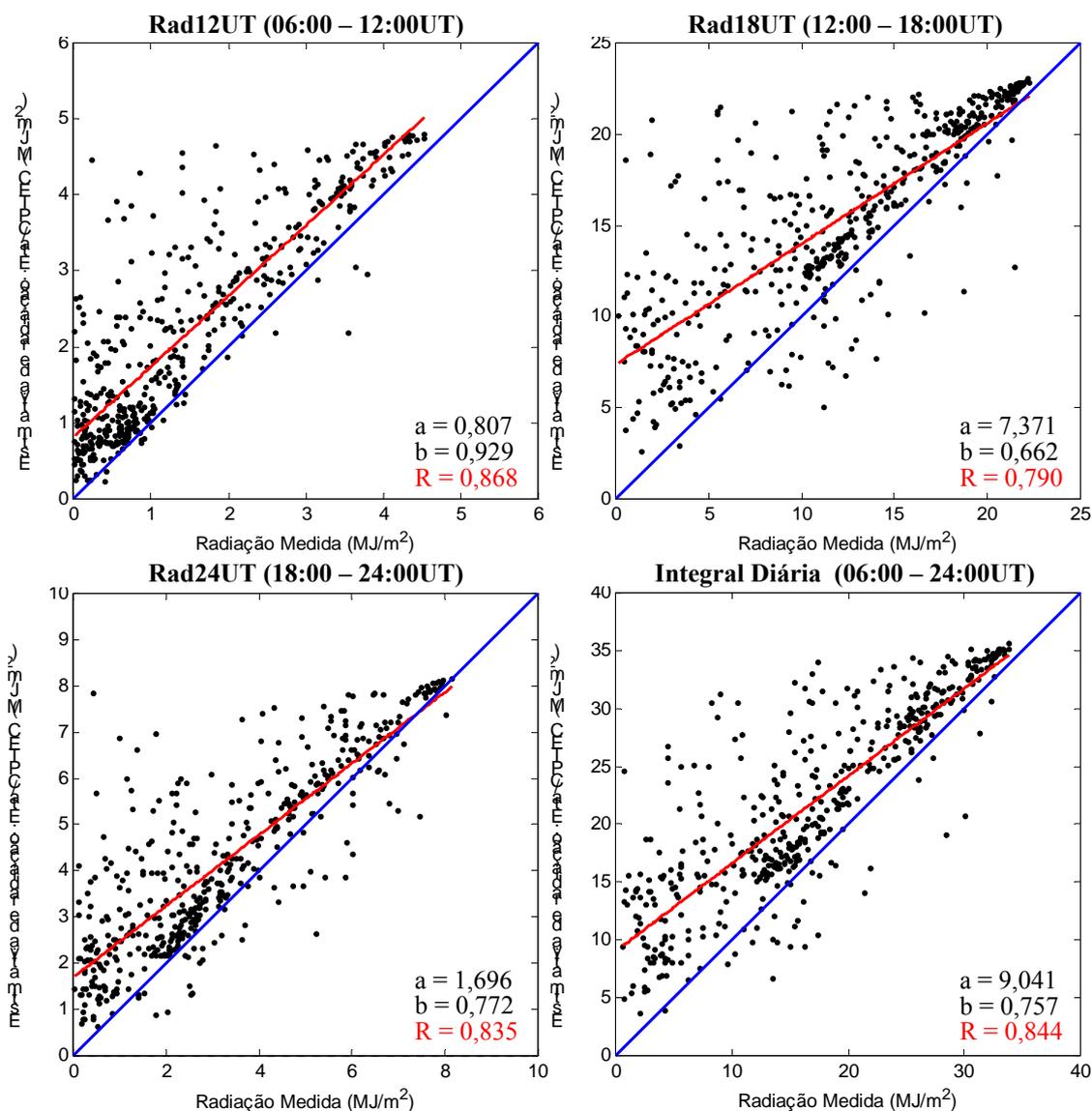


Figura 6.2: Diagramas de Dispersão para as previsões de radiação solar incidente do modelo Eta P00UT (variável *ocis*) para três períodos diários e integral diária, para SMS.

Tabela 6.2: Parâmetros de avaliação relativos aos diagramas de dispersão da Figura 6.2 (SMS).

| Índices | Rad12UT | Rad18UT | Rad24UT | Integral Diária |
|----------------|---------|---------|---------|-----------------|
| R | 0,868** | 0,790** | 0,835** | 0,844** |
| R ² | 0,754 | 0,624 | 0,697 | 0,712 |
| ME* | 0,704 | 3,324 | 0,970 | 4,998 |
| MAE* | 0,724 | 3,766 | 1,125 | 5,434 |
| RMSE* | 0,973 | 5,023 | 1,532 | 6,985 |
| \bar{O} * | 1,468 | 11,974 | 3,178 | 16,621 |
| ME% | 47,9 % | 27,8 % | 30,5 % | 30,1 % |
| RMSE% | 66,3 % | 41,9 % | 48,2 % | 42,0 % |

* Unidade: MJ/m².

** Nível de significância p<0,001.

Como as previsões para P00UT-Rad18UT são abordadas com maior ênfase nas seções subsequentes, é importante destacar que os valores absolutos para ME observados nas previsões P00UT de radiação solar do modelo Eta, para o período diário Rad18UT, foram de aproximadamente 2,9 e 3,3 MJ·m⁻², para FLN e SMS, respectivamente. Esses valores de ME equivalem a valores de ME% de aproximadamente 25 e 28%, para FLN e SMS, respectivamente. Por sua vez, os valores de RMSE observados para as previsões P00UT-Rad18UT do modelo Eta foram de aproximadamente 4,6 e 5,0 MJ·m⁻², correspondendo a 40 e 42% em RMSE%, para FLN e SMS, respectivamente. A redução desses erros de previsão de radiação é o objetivo almejado pelo ajuste dos modelos de RNA e RLM.

6.2. Experimentos: Testes de Grupos de Preditores

Como descrito no Capítulo 5, podem ser utilizados como entradas de modelos RNA e RLM um total de 36 preditores (Tabela 5.1). Contudo, alguns destes preditores contêm pouca ou nenhuma informação que leve a uma melhoria na previsão de radiação solar incidente. Como é almejado encontrar um grupo reduzido de preditores que permita obter previsões satisfatórias, testes com várias combinações de entradas foram realizados. Para evitar testar todas as combinações possíveis, critérios estatísticos e considerações físicas foram utilizados para eleger grupos de preditores a serem testados.

Os experimentos com grupos de preditores foram realizados considerando apenas os conjuntos de dados P00UT-Rad18UT, como enfatizado na seção anterior e no Capítulo 5. Estes conjuntos de dados, contendo 1150 dias para Florianópolis e 472 dias para São Martinho da Serra, foram subdivididos de maneira a serem compostos os subconjuntos de treinamento, validação e testes. A subdivisão realizada para esta etapa de experimentos consistiu em, a cada 4 dias de dados ao longo das seqüências de dados completas, fazer com que 2 dias pertençam ao conjunto de treinamento, 1 dia ao conjunto de validação e 1 dia ao conjunto de testes. Desta forma, esta subdivisão consiste numa *subdivisão homogênea*, em que cada subconjunto possui amostras igualmente distribuídas de todo o conjunto de dados, sendo cada um bastante representativo de todo o período de dados.

Os subconjuntos de treinamento (N=575 para FLN e N=236 para SMS) e validação (N=288 para FLN e N=118 para SMS) foram empregados para o treinamento de RNAs e ajuste de modelos de RLM. As avaliações dos modelos ajustados, apresentadas ao longo desta seção, foram realizadas com os dados dos subconjuntos de testes, e as informações de quantidade de dados e média das observações (\bar{O}), nesses subconjuntos, são destacadas na Tabela 6.3, para ambas as estações.

Tabela 6.3: Quantidade de dados e média das observações nos subconjuntos de testes para FLN e SMS (P00UT-Rad18UT).

| | FLN | SMS |
|-----------|--------------------------|--------------------------|
| N | 287 dias | 118 dias |
| \bar{O} | 11,810 MJ/m ² | 11,831 MJ/m ² |

6.2.1. Previsões Eta e por Persistência para Subconjuntos de Testes

Como é almejada a comparação das previsões novas com as atualmente disponíveis, e como os modelos de RNA e RLM são avaliados apenas com os dados do subconjunto de testes, avaliações e cálculos dos erros das previsões de radiação do modelo Eta foram realizados empregando apenas os dados deste subconjunto. Somente assim, tendo os erros de cada modelo sido calculados com base em dados para os mesmo dias, é possível compará-los através de estimativas de *Skill*.

Apenas para comparação, previsões por persistência para o subconjunto de testes são avaliadas juntamente com as previsões de radiação solar do modelo Eta. A previsão por persistência consiste em, na ausência de outra previsão confiável, supor que as condições observadas no dia atual irão persistir até o dia futuro. Desta forma, foram tomadas como previsões por persistência para cada dia, o valor de radiação solar medida no dia anterior. Uma previsão é considerada útil se possuir uma habilidade superior à previsão por persistência.

Assim, antes de apresentar as avaliações dos modelos de RNA e RLM com diferentes conjuntos de preditores, as avaliações das previsões por persistência e do modelo Eta (variável *ocis*), para os subconjuntos de testes das estações FLN e SMS, são apresentadas. Diagramas de dispersão, comparando as previsões com as observações, são apresentados na Figura 6.3. Parâmetros de avaliação, incluindo os erros médios, os

coeficientes de correlação e determinação, e valores de *Skill*, são apresentados na Tabela 6.4. Os valores de *Skill* (destreza) permitem verificar o ganho em termos de R e RMSE proporcionado pelo modelo Eta, frente às previsões por persistência. Os erros percentuais, ME% e RMSE% foram calculados através da razão dos valores de ME e RMSE com as médias das observações apresentadas na Tabela 6.3.

Verifica-se, para ambas as estações, um considerável ganho pelo uso das previsões do modelo Eta, frente à persistência. Enquanto as previsões por persistência possuem uma fração de variância comum às observações (R^2) de aproximadamente 20%, essa fração chega a 50 e 60%, aproximadamente, para FLN e SMS, respectivamente. Os valores de *Skill* para o RMSE, do modelo Eta em relação à persistência, revelam ganhos em torno de 13% para FLN e 20% para SMS. O menor RMSE e maior R^2 para o modelo Eta podem ser visualizados nos diagramas de dispersão, onde se verifica um menor espalhamento dos pontos, em comparação aos diagramas para persistência.

Os diagramas de dispersão para o modelo Eta permitem verificar novamente o viés positivo de suas previsões. Os valores encontrados para ME nos subconjuntos de testes estão em concordância com os valores encontrados na seção anterior deste trabalho, em que o conjunto total de dados foi empregado. Embora a previsão por persistência possua valores pequenos para ME, pode-se verificar, pelas linhas vermelhas dos diagramas de dispersão, que esta previsão está longe da realidade, sendo as estimativas de radiação superestimadas em uma região e subestimadas em outra. Assim, quando ocorre compensação de erros, a qualidade da previsão não pode ser atestada apenas pelo cálculo do ME, sendo imprescindível a observação do diagrama de dispersão.

Os parâmetros de avaliação da Tabela 6.4, para o modelo Eta, constituem os valores de referência que se pretende melhorar com as novas previsões, e serão empregados ao longo desta seção para comparação com os novos modelos ajustados. Os valores para o modelo Eta na Tabela 6.4, podem ser comparados com os valores para Rad18UT das Tabelas 6.1 e 6.2, e as semelhanças entre os erros evidenciam que os subconjuntos de testes são representativos dos conjuntos totais de dados.

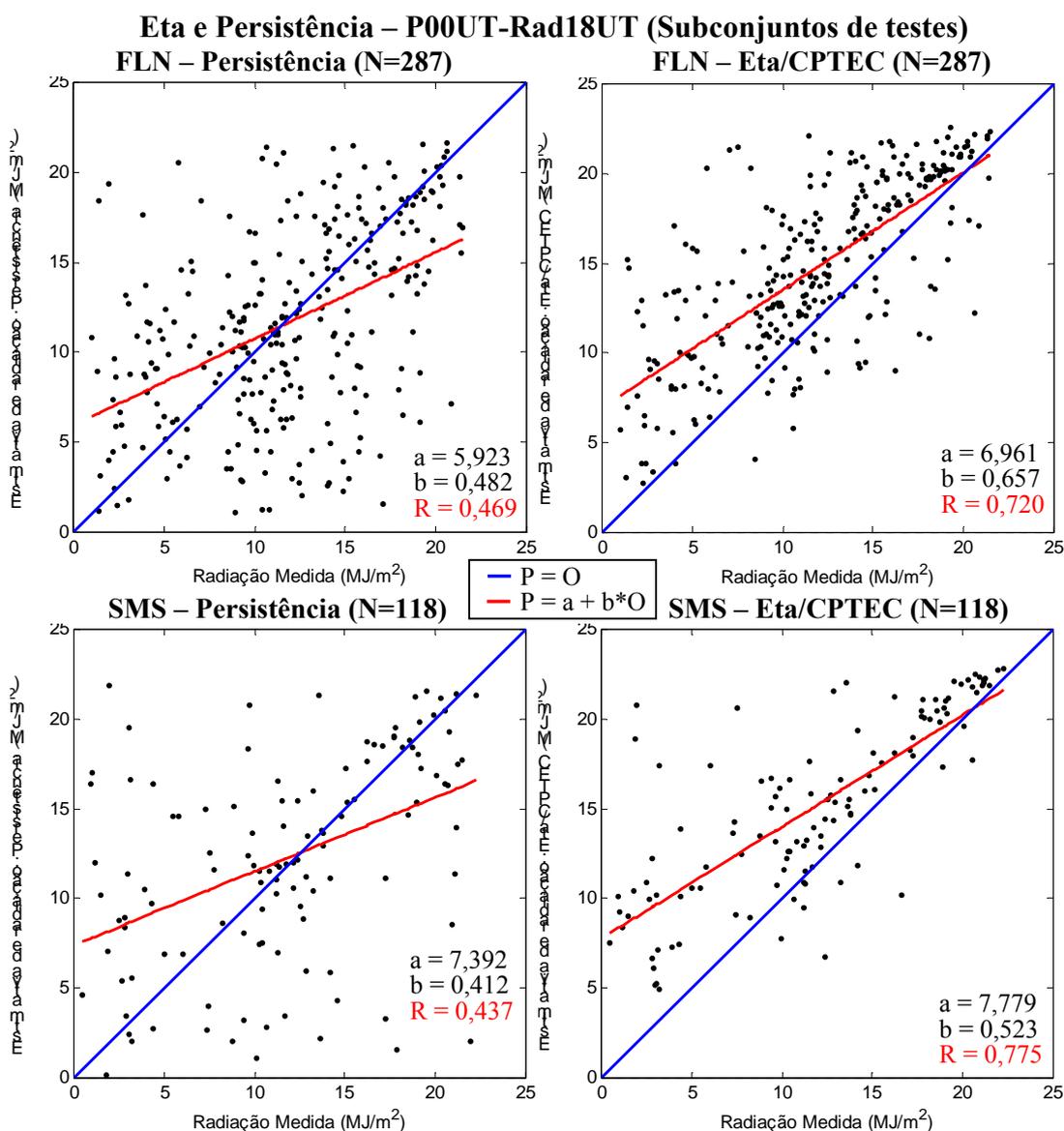


Figura 6.3: Diagramas de dispersão para as previsões (P00UT-Rad18UT – subconj. de testes) de radiação solar por persistência e calculadas pelo modelo Eta, para FLN e SMS.

Tabela 6.4: Parâmetros de avaliação das previsões (P00UT-Rad18UT – subconj. de testes) de radiação por persistência e pelo modelo Eta (*ocis*), para FLN e SMS.

| ÍNDICES | FLN | | SMS | |
|---------------------------|--------------|------------|--------------|------------|
| | Persistência | Modelo Eta | Persistência | Modelo Eta |
| R | 0,469* | 0,720* | 0,437* | 0,775* |
| R ² | 0,220 | 0,519 | 0,191 | 0,600 |
| ME (MJ/m ²) | -0,194 | 2,906 | 0,438 | 3,314 |
| MAE (MJ/m ²) | 4,055 | 3,759 | 4,467 | 3,776 |
| RMSE (MJ/m ²) | 5,427 | 4,729 | 6,364 | 5,115 |
| ME% | 1,6% | 24,6% | 3,7% | 28,0% |
| RMSE% | 45,9% | 40,0% | 53,8% | 43,2% |
| Skill(R,Persist.) | | 0,474 | | 0,599 |
| Skill(RMSE,Persist.) | | 0,129 | | 0,196 |

* Nível de significância $p < 0,001$.

Como o modelo Eta apresenta elevados valores para ME, que contribuem, em parte, para os elevados valores de RMSE, o refinamento mais simples aplicável às previsões de radiação é a remoção do viés, através de ajustes utilizando apenas a variável *ocis*. A fim de verificar qual o RMSE das previsões do modelo Eta, sem a influência do viés, foram criadas, experimentalmente, duas novas previsões derivadas da previsão de radiação do modelo Eta. A primeira, denominada *EtaA*, consistiu em, utilizando o conjunto de treinamento e validação, realizar o ajuste de uma regressão linear entre a variável *ocis* e a radiação medida (*rsfc*), obtendo-se então uma equação linear com uma única entrada (*ocis*) que produz uma nova estimativa de radiação. A segunda, *EtaB*, consistiu em simplesmente subtrair da variável *ocis* um valor constante, equivalente ao viés do modelo Eta, estimado para as duas estações com os conjuntos completos de dados, na seção 6.1 deste capítulo ($ME_{FLN} = 2,887 \text{ MJ}\cdot\text{m}^{-2}$; $ME_{SMS} = 3,324 \text{ MJ}\cdot\text{m}^{-2}$).

Diagramas de dispersão para as previsões *EtaA* e *EtaB* são apresentados na Figura 6.4. Acima dos diagramas, são apresentadas as equações de ajuste aplicadas a *ocis* na obtenção das novas previsões, para cada estação. Os parâmetros de avaliação para *EtaA* e *EtaB*, para FLN e SMS, são apresentados na Tabela 6.5, juntamente com os coeficientes de correlação e erros médios para o modelo Eta, anteriormente apresentados na Tabela 6.4. Considerando as duas estações, as previsões *EtaA* e *EtaB* apresentam valores de ME% menores que 0,2%, em módulo, e valores de RMSE% entre 30 e 33%. Embora estas previsões sejam métodos simples de refinar a previsão do modelo Eta, conduzindo a erros menores, verifica-se que os coeficientes de correlação e determinação permanecem inalterados, uma vez que ambas são transformações lineares da previsão dada diretamente pelo modelo Eta.

Com os ajustes de RNAs e RLMs, envolvendo outras variáveis previstas pelo modelo Eta, objetivou-se obter modelos que não apenas apresentassem erros ainda menores que *EtaA* e *EtaB*, mas que levassem a um ganho em R e R^2 (previsões que representem maiores frações da variância das observações). Como os modelos de RNAs e RLMs são alimentados por previsões do modelo Eta, também são refinamentos do modelo Eta, envolvendo, porém, maior complexidade de entradas do que num refinamento baseado em apenas ajustar linearmente a previsão de radiação do Eta.

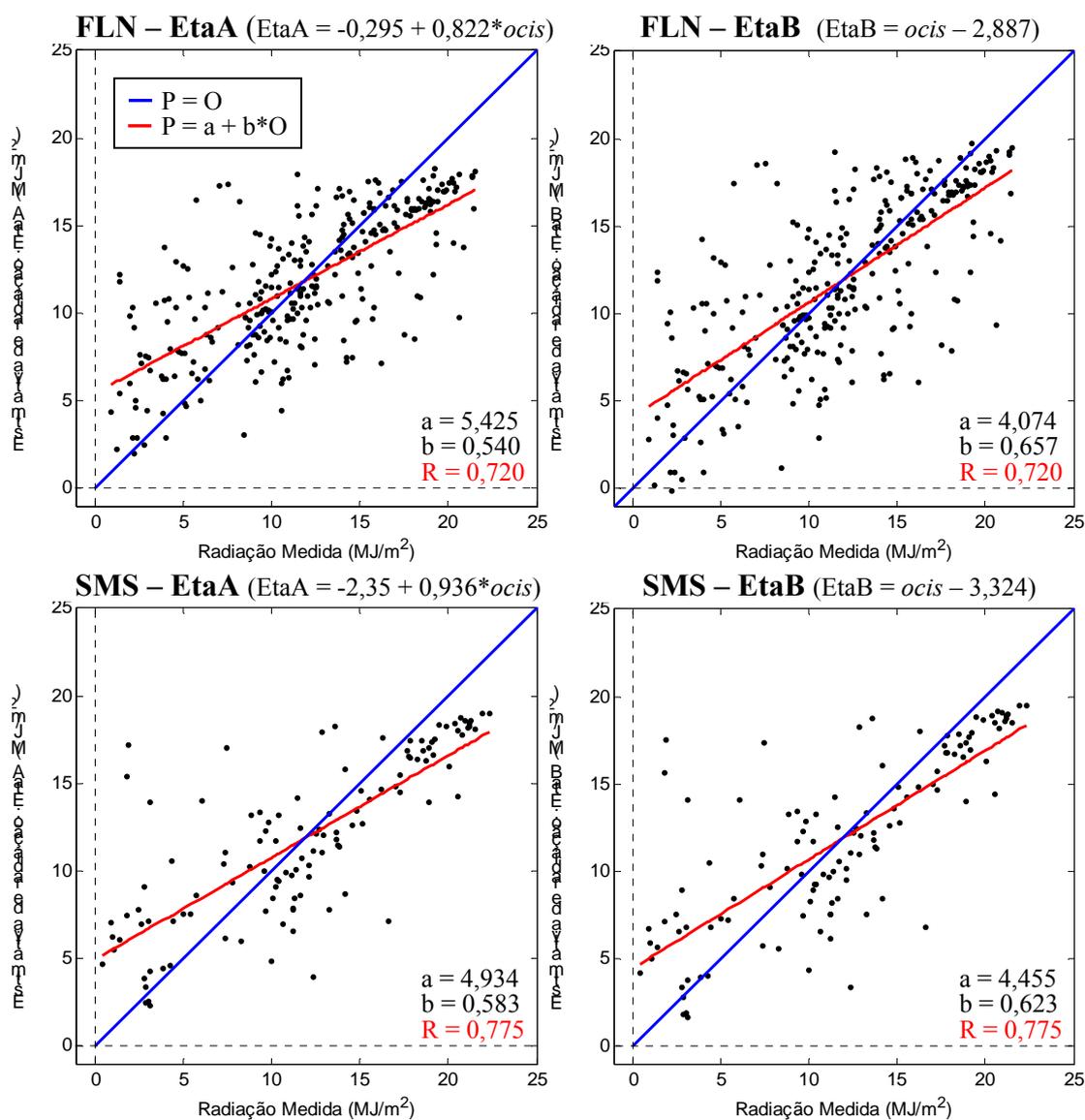


Figura 6.4: Diagramas de dispersão para as previsões derivadas da previsão de radiação solar do modelo Eta (EtaA e EtaB), para FLN e SMS.

Tabela 6.5: Parâmetros de avaliação das previsões Eta (*ocis*), EtaA e EtaB, frente radiação medida, para FLN e SMS (P00UT-Rad18UT – subconj. de testes).

| Índices | FLN | | | SMS | | |
|-----------------|---------|---------|---------|---------|---------|---------|
| | Eta | EtaA | EtaB | Eta | EtaA | EtaB |
| R | 0,720** | 0,720** | 0,720** | 0,775** | 0,775** | 0,775** |
| R ² | 0,519 | 0,519 | 0,519 | 0,600 | 0,600 | 0,600 |
| ME* | 2,906 | -0,013 | 0,019 | 3,314 | -0,001 | -0,010 |
| MAE* | 3,759 | 2,707 | 2,733 | 3,776 | 2,955 | 2,888 |
| RMSE* | 4,729 | 3,600 | 3,731 | 5,115 | 3,895 | 3,896 |
| ME% | 24,6% | -0,1% | -0,2% | 28,0% | -0,01% | -0,1% |
| RMSE% | 40,0% | 30,5% | 31,6% | 43,2% | 32,9% | 32,9% |
| Skill(RMSE,Eta) | | 0,239 | 0,211 | | 0,239 | 0,238 |

* Unidade: MJ/m².

** Nível de significância $p < 0,001$.

6.2.2. RNAs e RLMs Baseadas em Grupos Extensos de Preditores

O primeiro treinamento de RNAs, e ajuste de um modelo de RLM, foi realizado empregando o conjunto total de 36 preditores, isto é, todas as variáveis disponíveis no conjunto de dados P00UT-Rad18UT referentes às condições atmosféricas e de superfície do instante a ser estimada a radiação solar incidente. Os diagramas de dispersão para as previsões geradas por estes modelos de RNA e RLM, comparadas com as medidas de radiação, para FLN e SMS, são apresentados na Figura 6.5. A Tabela 6.6 apresenta um sumário dos parâmetros de avaliação dessas previsões frente às observações.

Este experimento foi realizado para ser utilizado como parâmetro de comparação para os experimentos que empregam menor número de preditores. Assim, fez-se possível verificar se a informação contida nesses conjuntos reduzidos de preditores é suficiente para ser atingida uma previsão satisfatória, com nível de erro comparável aos verificados na Tabela 6.6.

Em todos os experimentos, com qualquer conjunto de preditores, ao ser realizado o ajuste de um modelo RLM, foram treinadas 25 RNAs, e tomadas as previsões obtidas apenas pela RNA que apresentou menor erro. Esse procedimento foi adotado, pois, em cada treinamento, os pesos são inicializados aleatoriamente e sofrem seqüências de ajustes diferentes, levando a pesos finais diferentes. Eventualmente o treinamento pode ser interrompido em um mínimo local da superfície do erro de aprendizagem, com erro ainda elevado e distante do mínimo global, acarretando em RNAs ineficientes.

Neste experimento com todos os preditores, denominado *Experimento A*, verifica-se que o desempenho da RNA de menor erro, dentre as 25 treinadas, não apresentou um desempenho superior ao dos modelos RLM. Os coeficientes de correlação e os erros médios para RLM e RNA não apresentaram diferenças significativas, em ambas as estações. Para FLN, a RNA apresentou um desempenho apenas ligeiramente maior do que o modelo de RLM, enquanto para SMS, o modelo RLM foi ligeiramente melhor.

Experimento A – Todos os preditores (36) – P00UT-Rad18UT

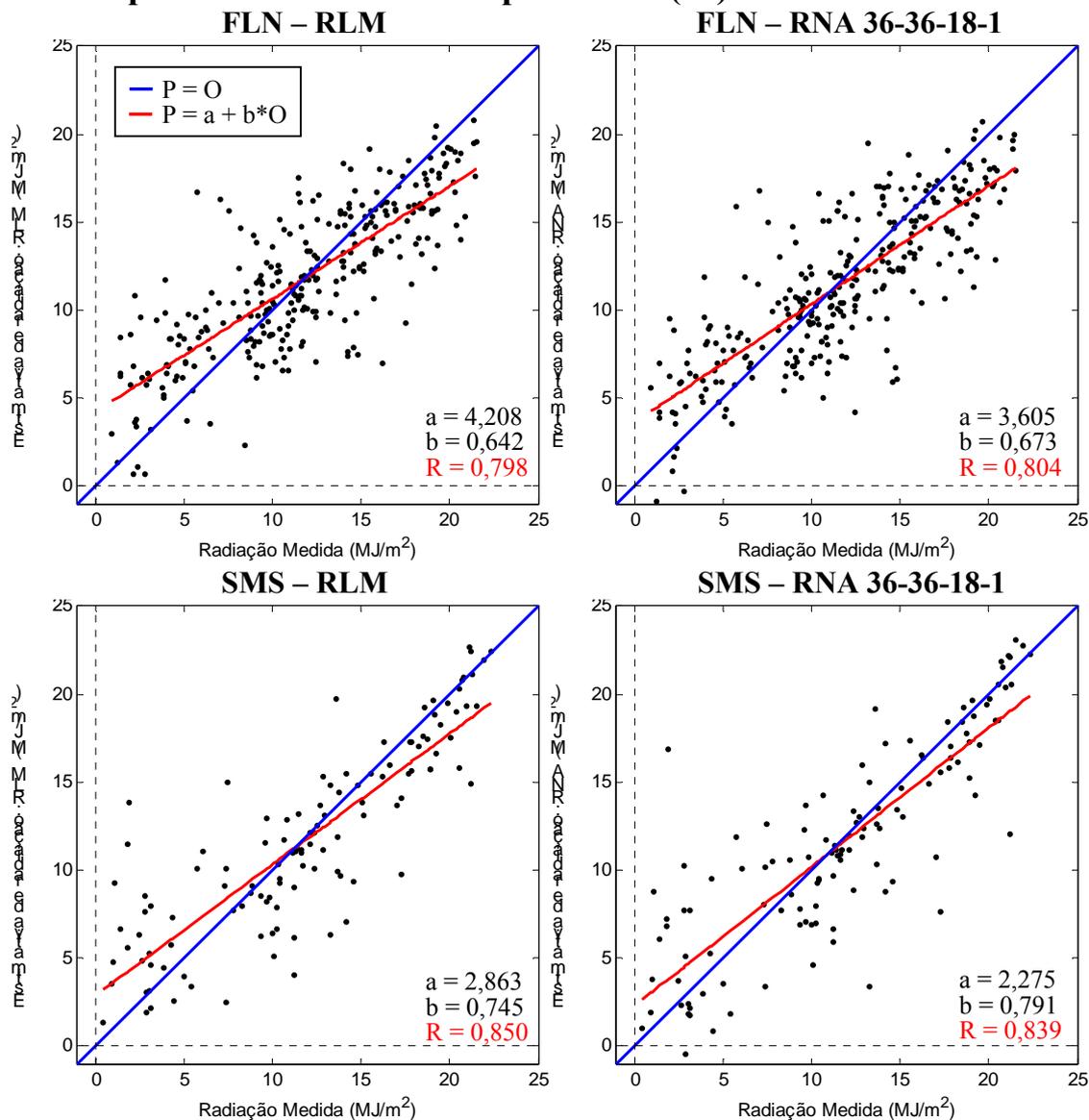


Figura 6.5: Diagramas de dispersão para previsões por RLM e RNA do Experimento A.

Tabela 6.6: Parâmetros de avaliação das previsões por RLM e RNA frente radiação medida, utilizando todos os 36 preditores da Tabela 5.1 (Experimento A).

| ÍNDICES | FLN | | SMS | |
|---------------------------|---------|---------|---------|---------|
| | RLM | RNA* | RLM | RNA* |
| R | 0,798** | 0,804** | 0,850** | 0,839** |
| R ² | 0,637 | 0,646 | 0,723 | 0,704 |
| ME (MJ/m ²) | -0,021 | -0,252 | -0,160 | -0,202 |
| MAE (MJ/m ²) | 2,414 | 2,419 | 2,306 | 2,392 |
| RMSE (MJ/m ²) | 3,122 | 3,100 | 3,245 | 3,411 |
| ME% | -0,2% | -2,1% | -1,4% | -1,7% |
| RMSE% | 26,4% | 26,2% | 27,4% | 28,8% |

* RNA-36-36-18-1

** Coeficientes de correlação com nível de significância $p < 0,001$.

Apesar de o uso de RNAs parecer não proporcionar um ganho sobre os modelos RLM, ambos os modelos apresentaram coeficientes de correlação na ordem de 0,8 para FLN e 0,85 para SMS. Estes valores representam um ganho em coeficiente de correlação frente às previsões de radiação Eta, EtaA e EtaB, que apresentaram valores para R de 0,72 e 0,75, para FLN e SMS, respectivamente.

Os modelos de RNA e RLM levaram a previsões com viés bastante reduzido ($|ME\%$ menores que 2,1%), embora maiores que os encontrados para EtaA e EtaB ($|ME\%$ menores que 0,2%). Contudo, verificaram-se valores de RMSE menores. Enquanto o RMSE% para o modelo Eta é de aproximadamente 40% para ambas as estações, e os modelos EtaA e EtaB possuem erros superiores a 30%, os novos modelos apresentaram RMSE% entre 26 e 28%. Essa diferença em RMSE reflete-se nos diagramas de dispersão, em que se verifica um menor espalhamento dos pontos em torno da linha de previsão perfeita para RLM e RNA, em comparação com as previsões Eta, EtaA e EtaB.

Os ganhos obtidos em termos RMSE foram quantificados em termos de *Skill*, avaliado em relação às demais previsões, sendo os valores apresentados na Tabela 6.7. De acordo com os valores de *Skill* apresentados, verificam-se ganhos superiores a 40% dos novos modelos sobre a previsão por persistência, ganhos superiores a 30% sobre o modelo Eta, e ganhos entre 12 e 17% sobre as previsões EtaA e EtaB.

Tabela 6.7: *Skill* em termos de RMSE, calculados para os modelos RLM e RNA do Experimento A, com relação a outras previsões, para as estações FLN e SMS.

| ÍNDICES | FLN | | SMS | |
|-----------------------|-------|-------|-------|--------|
| | RLM | RNA* | RLM | RNA* |
| Skill (RMSE,Persist.) | 0,425 | 0,429 | 0,490 | 0,464 |
| Skill (RMSE, Eta) | 0,340 | 0,344 | 0,366 | 0,333 |
| Skill (RMSE,EtaA) | 0,133 | 0,139 | 0,167 | 0,124 |
| Skill (RMSE, EtaB) | 0,163 | 0,169 | 0,167 | 0,124 |
| Skill (RMSE, RLM) | | 0,007 | | -0,051 |

* RNA-36-36-18-1

Após o ajuste de modelos empregando todos os preditores, foram realizados novos ajustes, restringindo determinados grupos de preditores, com a finalidade de observar se determinado grupo pode ser excluído e qual o impacto dessas exclusões sobre a qualidade da previsão de radiação solar incidente. Após o experimento anterior,

utilizando todos os 36 preditores (Experimento A), foram realizados experimentos excluindo a variável *ocis* e excluindo todo o grupo Radiação. Em seguida, foram adicionalmente excluídos os grupos Fluxos, Chuva e Nuvens. A Tabela 6.8 apresenta os parâmetros de avaliação (R, ME e RMSE) para RNAs e RLMs em cada um desses experimentos, identificados pelas letras B a F, nas estações FLN e SMS.

Tabela 6.8: Parâmetros de avaliação calculados para os Experimentos A a F.

| Experimento | ÍNDICES | FLN $\bar{O} = 11,810 \text{ MJ/m}^2$ N=287 | | SMS $\bar{O} = 11,831 \text{ MJ/m}^2$ N=118 | |
|-----------------------|---|---|--------|---|--------|
| | | RLM | RNA | RLM | RNA |
| A (RNA-36-36-18-1) | Preditores: todos (36) | | | | |
| | R | 0,798 | 0,804 | 0,850 | 0,839 |
| | ME* | -0,021 | -0,252 | -0,160 | -0,202 |
| | RMSE* | 3,122 | 3,100 | 3,245 | 3,411 |
| B (RNA-35-36-18-1) | Preditores: todos, menos ocis (35) | | | | |
| | R | 0,798 | 0,790 | 0,853 | 0,837 |
| | ME* | -0,011 | -0,089 | -0,120 | -0,392 |
| | RMSE* | 3,122 | 3,197 | 3,215 | 3,452 |
| C (RNA-30-30-15-1) | Preditores: todos, menos grupo radiação (30) | | | | |
| | R | 0,792 | 0,793 | 0,856 | 0,843 |
| | ME* | 0,002 | -0,203 | -0,060 | -0,368 |
| | RMSE* | 3,165 | 3,166 | 3,179 | 3,395 |
| D (RNA-27-28-14-1) | Preditores: todos, menos grupos radiação e fluxos (27) | | | | |
| | R | 0,786 | 0,801 | 0,853 | 0,833 |
| | ME* | -0,034 | -0,063 | -0,063 | -0,151 |
| | RMSE* | 3,207 | 3,107 | 3,214 | 3,444 |
| E (RNA-24-24-12-1) | Preditores: todos, menos grupos radiação, fluxos e chuva (24) | | | | |
| | R | 0,785 | 0,789 | 0,854 | 0,853 |
| | ME* | -0,037 | -0,138 | -0,074 | -0,460 |
| | RMSE* | 3,213 | 3,198 | 3,203 | 3,330 |
| F (RNA-20-20-10-1) | Preditores: todos, menos grupos rad., flux., chuva e nuvens (20) | | | | |
| | R | 0,765 | 0,782 | 0,851 | 0,839 |
| | ME* | 0,018 | -0,148 | -0,135 | -0,277 |
| | RMSE* | 3,339 | 3,235 | 3,239 | 3,407 |

* Unidade: MJ/m²

Considerando os valores de R e RMSE para FLN, na Tabela 6.8, verifica-se que, para RLM, as exclusões dos grupos Radiação, Fluxos e Chuva, não provocam diferenças significativas nos parâmetros de avaliação, sendo que a exclusão do grupo Nuvens, parece promover um aumento mais significativo em RMSE. Os modelos RNA para

FLN apresentam resultados com diferenças insignificantes em relação aos modelos RLM, e resultados apenas ligeiramente melhores para RNA são observados nos experimentos D e F. Para SMS, o modelo RLM não apresentou diferenças significativas em seu desempenho com a redução do número de preditores, e os valores de R e RMSE no experimento F são muito próximos dos encontrados para o experimento A. As RNAs para SMS apresentaram desempenhos inferiores a RLM em todos os experimentos.

Considerando as duas estações e todos os experimentos, verifica-se que ambos os modelos apresentam viés negativo (ME na Tabela 6.8), embora o viés dos modelos RLM é sempre mais próximo de zero do que o viés para RNAs. Verifica-se ainda que, frente aos modelos de RLM, as RNAs não levam a um ganho considerável, apresentando desempenhos sempre aproximadamente iguais, em termos de R e RMSE.

Com esse conjunto de experimentos (A-F), constatou-se que a redução do número de preditores de 36 para 20, não leva a diferenças expressivas no desempenho dos modelos RLM e RNA, confirmando a existência de variáveis desnecessárias ou redundantes no conjunto de preditores. As variáveis do grupo de Nuvens, que se espera apresentarem maior relação com a radiação solar incidente, parecem não influir significativamente no desempenho dos modelos para SMS, mas levam a uma redução expressiva de desempenho para o modelo RLM para FLN.

Além de permitir verificar que um menor número de preditores não reduz significativamente o desempenho dos modelos, o Experimento F (Tabela 6.8) revela que a informação contida em variáveis mais simples, não pertencentes a qualquer dos grupos eliminados, é suficiente para obter previsões com desempenho ainda superior a Eta, EtaA e EtaB (Tabela 6.5), em termos de R e RMSE.

6.2.3. Otimização dos Grupos de Preditores

Com a finalidade de verificar estatisticamente quais os preditores com maior potencialidade de contribuir para o cálculo da radiação solar incidente (*rsfc*) em cada estação, empregou-se a técnica de *regressão linear múltipla passo-a-passo*, descrita no capítulo 5. Nestas análises, não se empregou apenas os subconjuntos de testes, mas os

conjuntos completos de dados P00UT-Rad18UT de ambas as estações. Os resultados são apresentados nas Tabelas 6.9 e 6.10, para FLN e SMS, respectivamente.

Similarmente aos experimentos A-F, a análise foi realizada primeiramente sobre todos os preditores disponíveis, e repetida em mais quatro análises, reduzindo-se o número de preditores em cada análise, pela exclusão de um grupo (Radiação, Fluxos, Chuva, e Nuvens). Para cada análise, são apresentados, nas Tabelas 6.9 e 6.10, quais os preditores significativos identificados, dispostos na ordem com que foram inseridos, a cada passo, nos modelos de regressão que procuram simular o preditando (*rsfc*). Também são apresentados os coeficientes de correlação e determinação para cada modelo. Alguns preditores encontram-se realçados pelas cores verde e azul, sendo o significado disto apresentado na discussão que se segue.

Verifica-se que, assim como observado nos Experimentos A-F, a disponibilização de apenas 20 preditores leva a modelos que não apresentam diferenças significativas em termos dos coeficientes de correlação e determinação, em relação aos modelos gerados com base em todos os 36 preditores. Para todas as análises, em ambas as estações, o método de regressão passo-a-passo (com os critérios estabelecidos na seção 5.7) adotou um número de preditores sempre bastante menor do que o número de preditores disponibilizados, nunca tendo sido necessários mais do que 12 preditores.

Os preditores inseridos nos primeiros passos das regressões passo-a-passo são os que mais contribuem para a fração da variância do preditando explicada pelo modelo linear criado, e à medida que os passos avançam, os preditores inseridos representam contribuições cada vez menores, pouco modificando o coeficiente de determinação. Assim, nas seqüências de variáveis inseridas apresentadas nas Tabelas 6.9 e 6.10, as primeiras são as de maior importância na modelagem da radiação solar incidente.

Tabela 6.9: Regressões Múltiplas Passo-a-passo para FLN (P00UT-Rad18UT).

| Var. Dependente: <i>rsfc</i> | Var. Independentes: 36 preditores da Tabela 5.1 | |
|---|--|--------------------------------------|
| Análise 1 – Preditores: todos (36) | | |
| Variáveis inseridas: 11 <i>ocis, hinv, roce, u10m, rh2m, airm, role, oces, ghfl, mdny, v10m</i> | R = 0,803 | R ² = 0,645 |
| Análise 2 – Preditores: todos, menos grupo radiação (30) | | |
| Variáveis inseridas: 10 <i>cssf, hinv, ghfl, szam, u10m, agpl, qsfc, cbnt, v10m, prcv</i> | R = 0,806 | R ² = 0,650 |
| Análise 3 – Preditores: todos, menos grupos radiação e fluxos (27) | | |
| Variáveis inseridas: 10 <i>rtoa, cbnt, hinv, u10m, agpl, tsfc, v10m, cape, prge, ussl</i> | R = 0,803 | R ² = 0,644 |
| Análise 4 – Preditores: todos, menos grupos radiação, fluxos e chuva (23) | | |
| Variáveis inseridas: 9 <i>rtoa, cbnt, hinv, u10m, mdny, agpl, tsfc, v10m, cape</i> | R = 0,802 | R ² = 0,642 |
| Análise 5 – Preditores: todos, menos grupos radiação, fluxos, chuva e nuvens (20) | | |
| Variáveis inseridas: 6 <i>rh2m, rtoa, agpl, u10m, cape, tsfc</i> | R = 0,777 | R ² = 0,603 |
| Regressão Múltipla com os 36 preditores inseridos: | | R = 0,819 R²=0,671 |

Tabela 6.10: Regressões Múltiplas Passo-a-passo para SMS (P00UT-Rad18UT).

| Var. Dependente: <i>rsfc</i> | Var. Independentes: 36 preditores da Tabela 5.1 | |
|---|--|--------------------------------------|
| Análise 1 – Preditores: todos (36) | | |
| Variáveis inseridas: 10 <i>rh2m, u10m, wnds, oces, agpl, v10m, tsfc, szam, cbnt, pslc</i> | R = 0,887 | R ² = 0,787 |
| Análise 2 – Preditores: todos, menos grupo radiação (30) | | |
| Variáveis inseridas: 12 <i>rh2m, u10m, cbnt, szam, agpl, tsfc, tgsc, airm, rtoa, qsfc, pslc, ghfl</i> | R = 0,891 | R ² = 0,794 |
| Análise 3 – Preditores: todos, menos grupos radiação e fluxos (27) | | |
| Variáveis inseridas: 11 <i>rh2m, u10m, cbnt, szam, agpl, tsfc, tgsc, airm, rtoa, qsfc, pslc</i> | R = 0,890 | R ² = 0,792 |
| Análise 4 – Preditores: todos, menos grupos radiação, fluxos e chuva (23) | | |
| Variáveis inseridas: 11 <i>rh2m, u10m, cbnt, szam, agpl, tsfc, tgsc, airm, rtoa, qsfc, pslc</i> | R = 0,890 | R ² = 0,792 |
| Análise 5 – Preditores: todos, menos grupos radiação, fluxos, chuva e nuvens (20) | | |
| Variáveis inseridas: 10 <i>rh2m, u10m, agpl, tsfc, tgsc, szam, airm, rtoa, qsfc, pslc</i> | R = 0,887 | R ² = 0,786 |
| Regressão Múltipla com os 36 preditores inseridos: | | R = 0,897 R²=0,804 |

Para a estação FLN (Tabela 6.9) verifica-se que a não-disponibilização dos grupos Radiação e Fluxos, não leva a alterações significativas no coeficiente de determinação do modelo de regressão passo-a-passo gerado, apesar de os grupos de preditores escolhidos diferirem bastante entre as Análises 1, 2 e 3. A remoção do grupo Chuva não provocou alterações consideráveis no grupo de preditores escolhidos, nem em sua ordem. A remoção do grupo Nuvens provocou a maior modificação no coeficiente de determinação, embora o modelo de regressão tenha sido composto por apenas 6 variáveis tomadas como significativas. Em quase todas as análises verificou-se a inclusão de pelo menos dois preditores do grupo Nuvens, sendo que na Análise 4 foram incluídos 3 destes preditores (*cbnt*, *hinv* e *mdnv*) em passos iniciais da Análise.

Para a estação SMS (Tabela 6.10) verifica-se que o coeficiente de correlação é aproximadamente igual para as Análises 1 a 5. Em todas as análises, mesmo naquelas em que há disponibilidade de variáveis dos grupos Radiação e Fluxos, as variáveis inseridas, em geral, são variáveis mais simples, não pertencentes a qualquer dos grupos de preditores. Dentre os preditores do grupo Nuvens, apenas a variável *cbnt* foi inserida nos modelos. A disponibilidade ou não das variáveis do grupo Chuva, em nada alterou a escolha dos preditores, sendo obtidos os mesmos modelos nas Análises 3 e 4.

Dentre as 20 variáveis de saída do modelo Eta que não pertencem aos grupos Radiação, Fluxos, Chuva e Nuvens, as seguintes variáveis parecem ter importância comum às duas estações nas análises de regressão passo-a-passo: *rh2m*, *u10m*, *rtoa*, *agpl* e *tsfc*. Esses preditores estão destacados com a cor verde em todas as análises das Tabelas 6.9 e 6.10, onde também estão destacadas as variáveis do grupo Nuvens, com a cor azul.

A variável *rh2m* foi inserida em todas as análises para SMS e nas Análises 1 e 5 para FLN. Sua importância pode ser explicada pela elevada correlação negativa com o preditor *rsfc* observada, possivelmente devido ao fato de que em dias nublados e chuvosos a umidade relativa é elevada e a incidência de radiação é baixa, enquanto em dias de céu claro ocorre o contrário.

As demais variáveis, apesar de não terem apresentado correlações diretas com *rsfc* ao terem sido considerados todos os dias de dados, podem possuir alguma relação com *rsfc*

para dias com condições de tempo particulares (dias de céu claro ou dias nublados). Assim, essas variáveis podem conter informações que, combinadas, consigam explicar grande parte da variância de *rsfc*. Pode-se explicar a importância da variável *rtoa* pelo fato de que a radiação solar no TDA possui alta correlação com a radiação incidente em dias de céu claro. A variável *agpl*, por guardar a informação do conteúdo de água precipitável, deve apresentar valores elevados em dias chuvosos ou com nuvens de chuva, quando a radiação solar incidente é reduzida. A temperatura na superfície, *tsfc*, correlaciona-se com a radiação solar incidente em dias de céu limpo.

Embora presente em todos os modelos de regressão passo-a-passo, para as duas estações, é mais difícil supor uma ligação entre *u10m* e a radiação solar incidente. Possivelmente algum fator climático que influencia as condições do tempo e a incidência de radiação, também está vinculado a maiores ou menores intensidades de vento zonal em FLN e SMS. Contudo, as duas estações possuem diferenças em suas características meteorológicas: SMS possui um regime fortemente frontal e FLN apresenta características litorâneas.

É importante mencionar que o fato de um preditor ou grupo de preditores não ter mostrado grande importância na tentativa de modelar a radiação solar incidente, nas regressões passo-a-passo ou nos Experimentos A-F, não significa que não existe ligação física entre as variáveis e não prova a força da ligação física entre elas. A única conclusão a que se chega é que, estatisticamente, a variável de saída do Eta não leva a uma melhora na representação da radiação medida, o que pode ser devido à existência de outra variável ou combinação de variáveis que leve a resultados melhores, ou a uma imperfeição do modelo Eta em simular tal variável. Também não se pode esquecer que o modelo Eta prevê campos, isto é, estimativas suavizadas de valores pontuais.

Diante da verificação de que variáveis mais simples possuem informação suficiente para estimar a radiação solar incidente, procurou-se encontrar um grupo de preditores reduzido e que não incluía preditores do grupo Radiação. Dessa forma, pode-se verificar se um grupo de preditores que não inclui as previsões de radiação do modelo Eta permite que seja obtida uma previsão de radiação com desempenho superior ao

apresentado pelo código radiativo do modelo Eta. Uma vez que as variáveis de fluxos de calor na superfície são diretamente controladas pelas quantidades de radiação previstas pelo modelo Eta, procurou-se também não empregar os preditores do grupo Fluxos.

Assim, adotou-se como ponto de partida, um conjunto básico e arbitrário de preditores. O primeiro experimento de treinamento de RNAs e ajustes de RLMs com grupos reduzidos de preditores foi realizado empregando, além da radiação incidente no TDA (*rtoa*), apenas as variáveis meteorológicas básicas: umidade (*rh2m*), temperatura (*tsfc*), pressão (*pslc*) e vento (*u10m*, *v10m* e *wnds*). Desta forma, esse conjunto de preditores engloba quatro das cinco variáveis identificadas anteriormente como tendo grande importância para ambas as estações (*rh2m*, *u10m*, *rtoa*, *agpl* e *tsfc*). Os coeficientes de correlação e os erros médios (ME e RMSE) dos modelos de RNA e RLM, para ambas as estações, são apresentados na Tabela 6.11, identificados como *Experimento 1*.

Os modelos RLM e RNA para o Experimento 1 apresentam desempenho parecido ao observado para as previsões EtaA e EtaB (Tabela 6.5), embora para SMS, os novos modelos levem a resultados ligeiramente melhores em termos de R e RMSE. Contudo, o desempenho observado no Experimento 1 é bastante inferior ao obtido no Experimento A (Tabela 6.6). Objetivando um desempenho comparável ao encontrado no Experimento A, novos experimentos foram realizados com inclusão e exclusão de preditores, levando em conta aspectos físicos e os resultados da regressão múltipla passo-a-passo.

O Experimento 2 consistiu em tomar o conjunto de preditores do Experimento 1 e incluir adicionalmente o preditor *agpl* – identificado como bom preditor para ambas as estações, nas regressões passo-a-passo. Verifica-se, na Tabela 6.11, um considerável ganho do Experimento 2 sobre o Experimento 1, com aumento dos coeficientes de correlação, e redução do RMSE de 3,6 para 3,3 MJ·m⁻², em média.

Uma vez que alguns dos preditores utilizados no Experimento 2, não haviam sido identificados como fortes preditores nas regressões passo-a-passo, novos experimentos foram realizados, excluindo-os. No Experimento 3, foram removidos *v10m* e *wnds*, e

não foram observadas mudanças significativas no desempenho de RLMs e RNAs. Como foi atingido um desempenho equivalente, mas com um grupo de preditores menor, o Experimento 3 representa um ganho nos modelos de RLM e RNA.

Tabela 6.11: Parâmetros de avaliação calculados para os Experimentos 1 a 5.

| Experimento | ÍNDICES | FLN $\bar{O} = 11,810 \text{ MJ/m}^2$ N=287 | | SMS $\bar{O} = 11,831 \text{ MJ/m}^2$ N=118 | |
|---|---------|---|--------|---|--------|
| | | RLM | RNA | RLM | RNA |
| Preditores: rtoa, rh2m, tsfc, pslc, u10m, v10m, wnds (7) | | RNA-7-14-7-1 | | | |
| 1 | R | 0,717 | 0,731 | 0,802 | 0,805 |
| | ME* | -0,108 | -0,185 | 0,005 | 0,041 |
| | RMSE* | 3,614 | 3,541 | 3,685 | 3,670 |
| Preditores: rtoa, rh2m, tsfc, pslc, u10m, v10m, wnds, agpl (8) | | RNA-8-16-8-1 | | | |
| 2 | R | 0,764 | 0,775 | 0,845 | 0,841 |
| | ME* | 0,009 | -0,013 | -0,096 | 0,019 |
| | RMSE* | 3,344 | 3,275 | 3,294 | 3,338 |
| Preditores: rtoa, rh2m, tsfc, pslc, u10m, agpl (6) | | RNA-6-12-6-1 | | | |
| 3 | R | 0,766 | 0,770 | 0,842 | 0,843 |
| | ME* | 0,014 | -0,060 | -0,197 | -0,175 |
| | RMSE* | 3,331 | 3,308 | 3,330 | 3,324 |
| Preditores: rtoa, rh2m, tsfc, u10m, agpl (5) | | RNA-5-10-5-1 | | | |
| 4 | R | 0,767 | 0,764 | 0,837 | 0,836 |
| | ME* | 0,014 | -0,005 | -0,219 | -0,017 |
| | RMSE* | 3,329 | 3,350 | 3,385 | 3,384 |
| Preditores: rtoa, rh2m, tsfc, agpl (4) | | RNA-4-8-4-1 | | | |
| 5 | R | 0,756 | 0,755 | 0,834 | 0,831 |
| | ME* | 0,019 | 0,003 | -0,234 | -0,147 |
| | RMSE* | 3,391 | 3,408 | 3,404 | 3,433 |

* Unidade: MJ/m^2

O experimento 4 consistiu na remoção do preditor *pslc*, de maneira a restarem apenas 5 preditores. Estes preditores são os 5 preditores destacados nas Tabelas 6.9 e 6.10 e que foram identificados como importantes para ambas as estações. Esta nova redução no grupo de preditores não ocasionou diminuição significativa de desempenho, apesar de ser observado um ligeiro acréscimo em RMSE para SMS, do Experimento 3 para o Experimento 4.

No Experimento 5, testou-se a exclusão da variável *u10m*, uma vez que, como mencionado anteriormente, não há uma boa explicação física para a permanência desta

variável nos modelos. Contudo, verificou-se uma redução de desempenho com os valores de RMSE atingindo valores superiores a $3,4 \text{ MJ}\cdot\text{m}^{-2}$. Optou-se por manter o preditor *u10m* nos experimentos subsequentes.

Assim, com apenas 5 preditores (Experimento 4), os modelos RLM e RNA apresentam desempenho superior à previsão de radiação do modelo Eta, e previsões EtaA e EtaB. Enquanto os novos modelos apresentam valores para o coeficiente de correlação de aproximadamente 0,77 e 0,84, para FLN e SMS, respectivamente, as previsões do modelo Eta levam a coeficientes 0,72 e 0,78. A redução de RMSE, observada a partir das previsões EtaA e EtaB (Tabela 6.5) para os modelos RLM e RNA (Experimento 4), é de 3,6 para $3,3 \text{ MJ}\cdot\text{m}^{-2}$, para FLN, e de 3,9 para $3,4 \text{ MJ}\cdot\text{m}^{-2}$, para SMS. As diferenças de desempenho entre RLM e RNA foram insignificantes nos experimentos realizados.

Objetivando atingir valores de RMSE mais próximos daqueles obtidos para o Experimento A, novos experimentos foram empregados, adicionando aos 5 preditores do Experimento 4, preditores do grupo Nuvens. Além de as variáveis de nuvens terem sido apontadas como bons preditores nas regressões passo-a-passo, especialmente para FLN, essas variáveis têm conhecida relação física com a radiação solar incidente, como mencionado na seção 2.7 deste trabalho.

O Experimento 6 consistiu em inserir as 4 variáveis do grupo Nuvens ao grupo de preditores do Experimento 4. Pode-se verificar na Tabela 6.12, que este grupo de 9 preditores elevou os coeficientes de correlação para FLN para aproximadamente 0,79, sendo observados para SMS coeficientes de 0,85 e 0,84 para RLM e RNA, respectivamente. Os valores de RMSE para FLN, considerando ambos os modelos, é de aproximadamente $3,2 \text{ MJ}\cdot\text{m}^{-2}$, e o modelo RLM de SMS apresentou RMSE menor que $3,3 \text{ MJ}\cdot\text{m}^{-2}$.

Uma vez que a variável *lwnv*, que representa nuvens em baixos níveis atmosféricos, não foi identificada como bom preditor para nenhuma das estações, nas regressões passo-a-passo, o Experimento 7 (Tabela 6.12) consistiu em excluí-la. Esta redução para 8 preditores não afetou significativamente o desempenho dos modelos.

Os experimentos 8 e 9 consistiram em remover as variáveis *mdnv* (nuvens em níveis médios) e *hinv* (nuvens em altos níveis), mantendo apenas a variável *cbnt* (fração de cobertura de nuvens considerando todos os níveis) no Experimento 9. Embora não tenha havido redução de desempenho significativa com relação ao Experimento 7, os valores de RMSE para FLN aumentaram ligeiramente nos Experimentos 8 e 9, e os coeficientes de correlação diminuíram de 0,79 para 0,78. Para SMS os coeficientes de correlação permaneceram em aproximadamente 0,85, e apenas uma pequena redução em RMSE é observada. Optou-se por adotar o Experimento 7 nos estudos subseqüentes.

Tabela 6.12: Parâmetros de avaliação calculados para os Experimentos 6 a 9.

| Experimento | ÍNDICES | FLN $\bar{O} = 11,810 \text{ MJ/m}^2$ N=287 | | SMS $\bar{O} = 11,831 \text{ MJ/m}^2$ N=118 | |
|---|--------------|---|--------|---|--------|
| | | RLM | RNA | RLM | RNA |
| Preditores: rtoa, rh2m, tsfc, u10m, agpl, cbnt, hinv, mdnv, lwnv | | RNA-9-18-9-1 | | | |
| 6 | R | 0,786 | 0,793 | 0,848 | 0,839 |
| | ME* | -0,025 | 0,012 | -0,105 | -0,134 |
| | RMSE* | 3,205 | 3,161 | 3,261 | 3,366 |
| Preditores: rtoa, rh2m, tsfc, u10m, agpl, cbnt, hinv, mdnv | | RNA-8-16-8-1 | | | |
| 7 | R | 0,789 | 0,790 | 0,848 | 0,848 |
| | ME* | -0,026 | -0,088 | -0,124 | -0,080 |
| | RMSE* | 3,187 | 3,178 | 3,262 | 3,267 |
| Preditores: rtoa, rh2m, tsfc, u10m, agpl, cbnt, hinv | | RNA-7-14-7-1 | | | |
| 8 | R | 0,787 | 0,784 | 0,849 | 0,851 |
| | ME* | -0,038 | -0,052 | -0,127 | -0,040 |
| | RMSE* | 3,198 | 3,219 | 3,257 | 3,238 |
| Preditores: rtoa, rh2m, tsfc, u10m, agpl, cbnt | | RNA-6-12-6-1 | | | |
| 9 | R | 0,781 | 0,778 | 0,849 | 0,850 |
| | ME* | -0,008 | -0,013 | -0,134 | -0,152 |
| | RMSE* | 3,236 | 3,256 | 3,254 | 3,246 |

* Unidade: MJ/m^2

Ao longo dos nove experimentos com conjuntos reduzidos de preditores (Experimentos 1-9), e mesmo ao longo dos seis experimentos com conjuntos mais completos (Experimentos A-F), não foi verificado um desempenho destacadamente maior pelo uso de RNAs. O desempenho das RNAs parece sempre oscilar em torno do desempenho dos modelos RLM, apresentando sempre uma diferença pouco significativa em RMSE, ME e R, ora ligeiramente melhor e ora ligeiramente pior.

Uma das possíveis razões por não ser encontrada uma melhora em desempenho com o uso de RNAs, que constitui-se uma ferramenta mais complexa e poderosa, é a possibilidade de a inter-relação entre as variáveis ser mais linear do que se supunha no início deste trabalho. O fato de se estar trabalhando com médias das variáveis em longos intervalos de tempo pode ter tornado o sistema ainda mais linear. Assim, como o diferencial das RNAs é justamente sua habilidade de ajuste de não-linearidades, a sua utilização não promove ganhos sobre modelos lineares gerados por RLM.

É importante mencionar que ambas as metodologias, RNA e RLM, consistem em criar modelos que minimizam o erro quadrático médio, de maneira que a ausência de não-linearidades nos dados leva o treinamento de RNAs a modelos que produzam saídas semelhantes aos modelos RLM.

Outro aspecto importante é o fato de que as RNAs dependem da qualidade e quantidade dos exemplos apresentados durante o treinamento, para que possa extrair padrões corretos e representativos do comportamento e inter-relação entre as variáveis. Por empregar dados de previsão de tempo, que naturalmente possuem erros, o algoritmo de aprendizado pode ter encontrado dificuldades de, com o conjunto de exemplos disponibilizado, promover associações entre as entradas que pudessem levar a valores mais aproximados dos alvos (medidas reais) empregados na fase de treinamento.

Uma vez que os oito preditores do Experimento 7 foram adotados como grupo reduzido de preditores a ser empregado nas próximas análises, cabe uma análise mais detalhada das previsões obtidas neste experimento e uma confrontação com os parâmetros de desempenho das demais previsões disponíveis.

Diagramas de dispersão para as previsões por RLM e RNA obtidas no Experimento 7, para as estações FLN e SMS, são apresentados na Figura 6.6. Verificam-se grandes semelhanças com relação aos diagramas para o Experimento A (Figura 6.5), em termos da dispersão dos pontos e comportamento geral das previsões. Os parâmetros de avaliação do Experimento 7 são apresentados na Tabela 6.13, alguns deles tomados diretamente da Tabela 6.12.

Experimento 7 – P00UT-Rad18UT

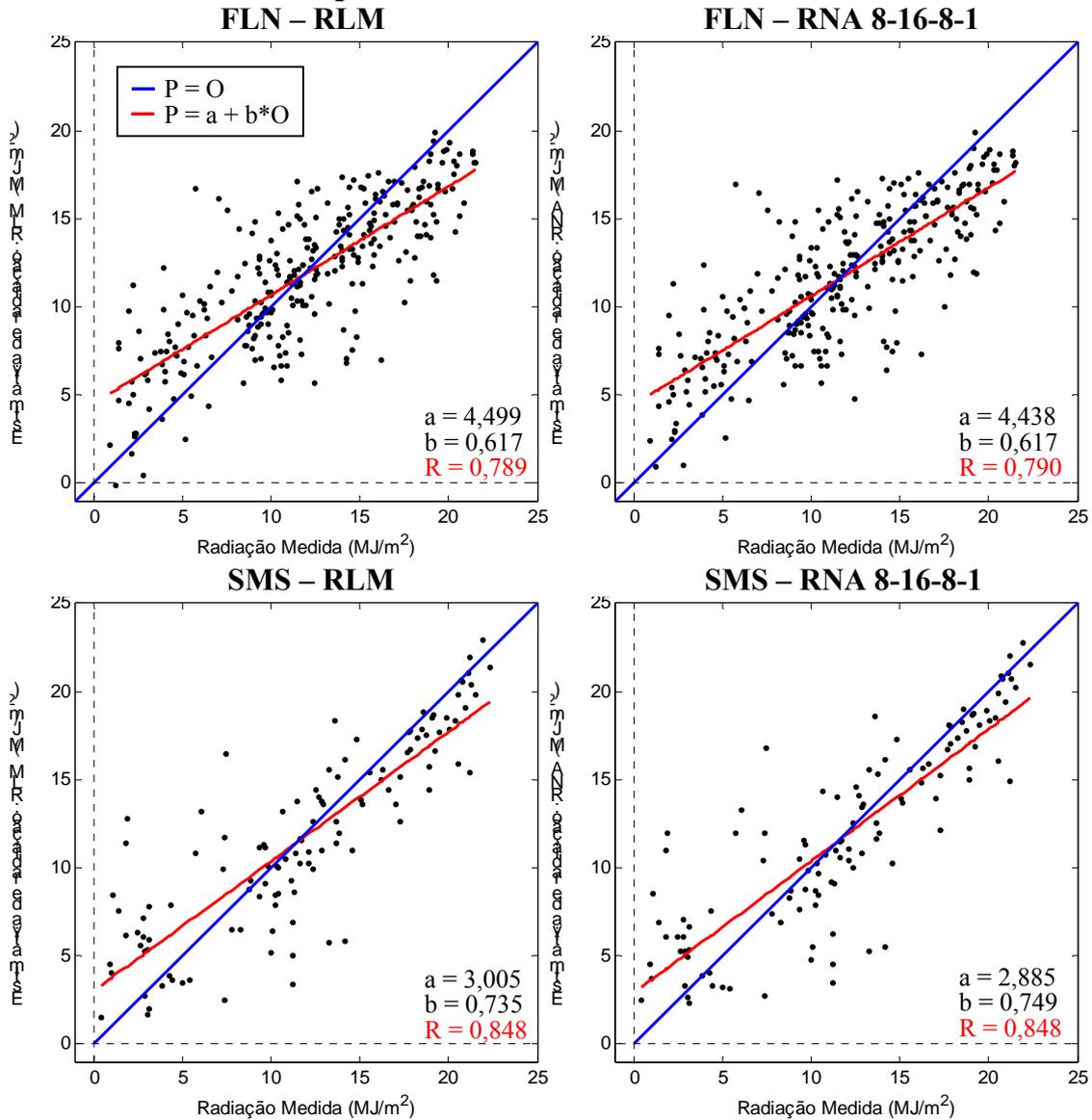


Figura 6.6: Diagramas de dispersão para previsões por RLM e RNA do Experimento 7.

Tabela 6.13: Parâmetros de avaliação das previsões por RLM e RNA frente radiação medida, para o Experimento 7.

| ÍNDICES | FLN | | SMS | |
|---------------------------|---------|---------|---------|---------|
| | RLM | RNA* | RLM | RNA* |
| R | 0,789** | 0,790** | 0,848** | 0,848** |
| R ² | 0,622 | 0,625 | 0,720 | 0,720 |
| ME (MJ/m ²) | -0,026 | -0,088 | -0,124 | -0,080 |
| MAE (MJ/m ²) | 2,454 | 2,438 | 2,398 | 2,349 |
| RMSE (MJ/m ²) | 3,187 | 3,178 | 3,262 | 3,267 |
| ME% | -0,2% | -0,8% | -1,0% | -0,7% |
| RMSE% | 27,0% | 26,9% | 27,6% | 27,6% |

* RNA-8-16-8-1

** Coeficientes de correlação com nível de significância p<0,001.

Na Tabela 6.14 são apresentados os valores de *Skill* para o Experimento 7, em termos de RMSE, avaliados em relação à outras previsões. Pela comparação dos valores da Tabela 6.13 com a Tabela 6.6, e dos valores de *Skill* da Tabela 6.14 com a Tabela 6.7, verifica-se que o desempenho encontrado no Experimento 7 (8 preditores) é muito próximo ao encontrado no Experimento A (36 preditores).

Em termos de ME%, apesar de ambos os experimentos (A e 7) apresentarem erros muito baixos em comparação com a previsão do modelo Eta, valores para ME% menores do que no Experimento A foram encontrados para o Experimento 7 ($|ME\%|$ menores que 1%). Em ambos os experimentos, os coeficientes de correlação para SMS foram de aproximadamente 0,85, enquanto para FLN os coeficientes foram aproximadamente 0,8 e 0,79, para os experimentos A e 7, respectivamente. Os valores de RMSE no Experimento 7 são apenas ligeiramente maiores do que no Experimento A, exceto para a RNA de SMS que se mostrou melhor no Experimento 7.

Tabela 6.14: *Skill* em termos de RMSE, calculados para os modelos RLM e RNA do Experimento 7, com relação a outras previsões, para as estações FLN e SMS.

| ÍNDICES | FLN | | SMS | |
|------------------------|-------|-------|-------|--------|
| | RLM | RNA* | RLM | RNA* |
| Skill (RMSE, Persist.) | 0,413 | 0,414 | 0,487 | 0,487 |
| Skill (RMSE, Eta) | 0,326 | 0,328 | 0,362 | 0,361 |
| Skill (RMSE, EtaA) | 0,115 | 0,117 | 0,163 | 0,161 |
| Skill (RMSE, EtaB) | 0,146 | 0,148 | 0,163 | 0,161 |
| Skill (RMSE, RLM) | | 0,003 | | -0,002 |

* RNA-8-16-8-1

De acordo com a Tabela 6.14, os modelos do Experimento 7 apresentam valores de *Skill* semelhantes aos do Experimento A (Tabela 6.7), com os ganhos em termos de RMSE situados entre 32 e 36% sobre as previsões do modelo Eta, e entre 11 e 16% sobre as previsões EtaA e EtaB. Em ambos os experimentos, os valores de *Skill* revelaram ganhos superiores a 40% sobre a previsão por persistência. Assim, os novos modelos levam a previsões com erros reduzidos, superando as previsões já existentes. A diferença no número de preditores, de 36 para 8, e as pequenas diferenças observadas em desempenho, justificam o emprego das variáveis do grupo de preditores do Experimento 7, ao invés da utilização de todos os preditores disponíveis.

Outros experimentos foram realizados após o Experimento 9, consistindo na adição e remoção de preditores, e no emprego de grupos de preditores escolhidos pelas regressões passo-a-passo. Contudo, não foram observados desempenhos significativamente superiores que os observados nos Experimentos A e 7. Nestes experimentos, apesar de as modificações levarem a ligeiras melhoras de desempenho para uma estação, ocasionam perda de desempenho para a outra estação, não sendo encontrado um modelo que levasse a ganhos simultâneos para FLN e SMS. Porém, verificou-se que a exclusão do preditor *rh2m* do Experimento 7 não altera o desempenho de RNAs e RLMs para FLN, mas ocasiona elevação em RMSE para SMS. A adição da variável *pslc* leva a melhorias nas previsões para SMS, mas ocasiona perdas para FLN.

6.2.4. Comparações entre as Previsões

Com a finalidade de comparar visualmente os resultados encontrados nas sub-sessões anteriores, os parâmetros de avaliação das previsões analisadas são apresentados graficamente na Figura 6.7. São mostrados os valores de ME%, RMSE%, R e R^2 para as previsões por persistência, Eta, EtaA, EtaB, e os modelos RLM e RNA dos Experimentos 7 e Experimento A, para FLN e SMS. Nesta seção e nas seguintes, os modelos dos Experimentos A e 7, serão denominados por: RLM-Ex.7 e RNA-Ex.7, para o Experimento 7; e RLM-Ex.A e RNA-Ex.A, para o Experimento A.

Verifica-se, nos gráficos da Figura 6.7, que qualquer das previsões derivadas do modelo Eta (EtaA, EtaB, RLMs e RNAs) promovem a remoção do viés, redução do RMSE e aumento dos coeficientes de correlação e determinação, para ambas as estações solarimétricas. Como observado anteriormente, embora os modelos RLM e RNA apresentem desempenhos semelhantes, suas previsões possuem desempenhos superiores aos observados para EtaA e EtaB. Além disso, verificam-se diferenças insignificantes entre a utilização de 8 ou 36 preditores.

Na Figura 6.8 são apresentados diagramas de caixa (*box plots*) para as diversas previsões, comparadas às observações, para FLN e SMS. Os diagramas de caixa mostram as similaridades e diferenças entre as distribuições de valores de previsões e

observações, em termos da localização e espalhamento das distribuições. Os limites inferior e superior da caixa correspondem ao 25° e ao 75° percentil dos dados, sendo a mediana representada pela linha horizontal dentro da caixa. As extremidades das linhas verticais correspondem ao 5° e ao 95° percentil, os símbolos “x” representam o 1° e o 99° percentil, os símbolos “-” indicam os valores extremos das séries de dados, e o quadrado no interior das caixas corresponde às médias das observações ou previsões.

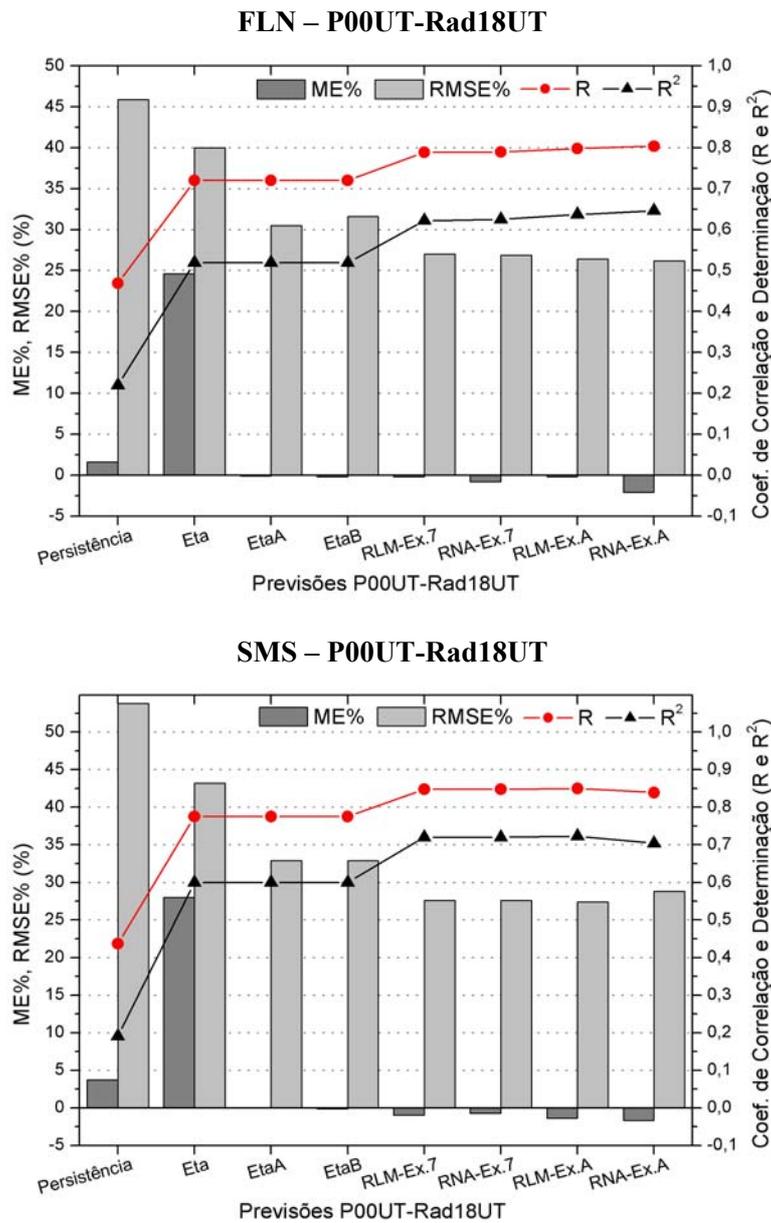


Figura 6.7: Gráficos comparativos de ME%, RMSE%, R e R², para as previsões de radiação solar avaliadas, para FLN e SMS.

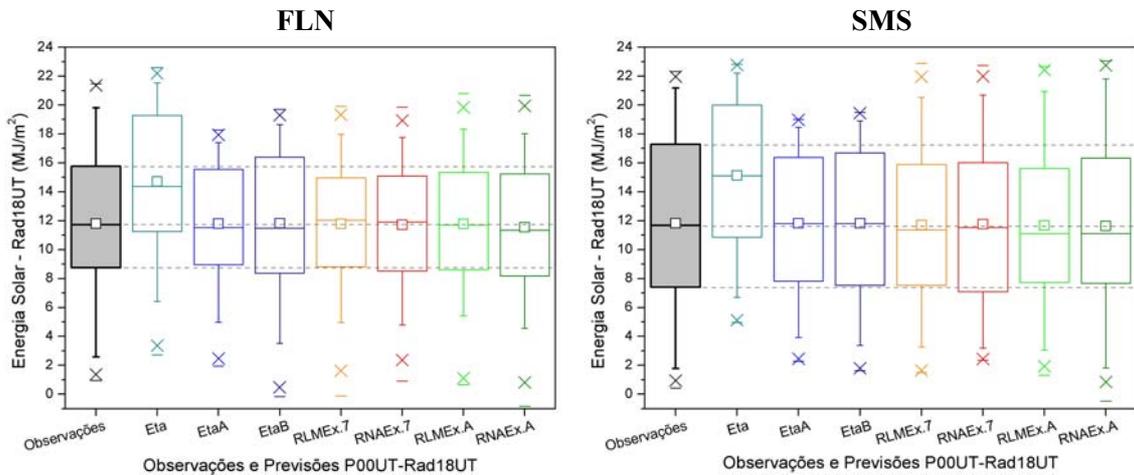


Figura 6.8: Diagramas de caixa (*box plots*) para observações e previsões de radiação solar P00UT-Rad18UT para FLN e SMS.

A caixa da esquerda, em cada gráfico da Figura 6.8, corresponde às observações de radiação, e linhas pontilhadas são traçadas a partir da mediana, do 25º percentil e do 75º percentil, para comparação com as caixas das previsões. Verifica-se, em ambas as estações, que as distribuições de valores das previsões Eta encontram-se deslocadas para valores maiores do que as observações, comportamento característico do viés do modelo. A previsão EtaA apresenta os limites das caixas próximos aos limites das caixas das observações, embora seja verificada uma pobre representação dos valores extremos. A previsão EtaB, possui distribuição idêntica à previsão Eta, porém deslocada em direção à média das observações, pela remoção do viés. As previsões RLM e RNA possuem distribuições semelhantes entre si, com medianas e médias próximas das encontradas para as observações, porém apresentando uma representação precária dos maiores e menores valores de radiação.

A fim de verificar a magnitude e ocorrência dos desvios entre previsões e observações, para cada previsão avaliada e para ambas as estações, foram construídos os histogramas apresentados na Figura 6.9. Cada barra vertical dos histogramas simboliza o número de ocorrências de desvios em cada categoria, onde cada categoria abrange desvios em intervalos de 2 MJ/m^2 . As barras na posição “0” simbolizam o número de ocorrências dos menores desvios, entre -1 e $+1 \text{ MJ/m}^2$.

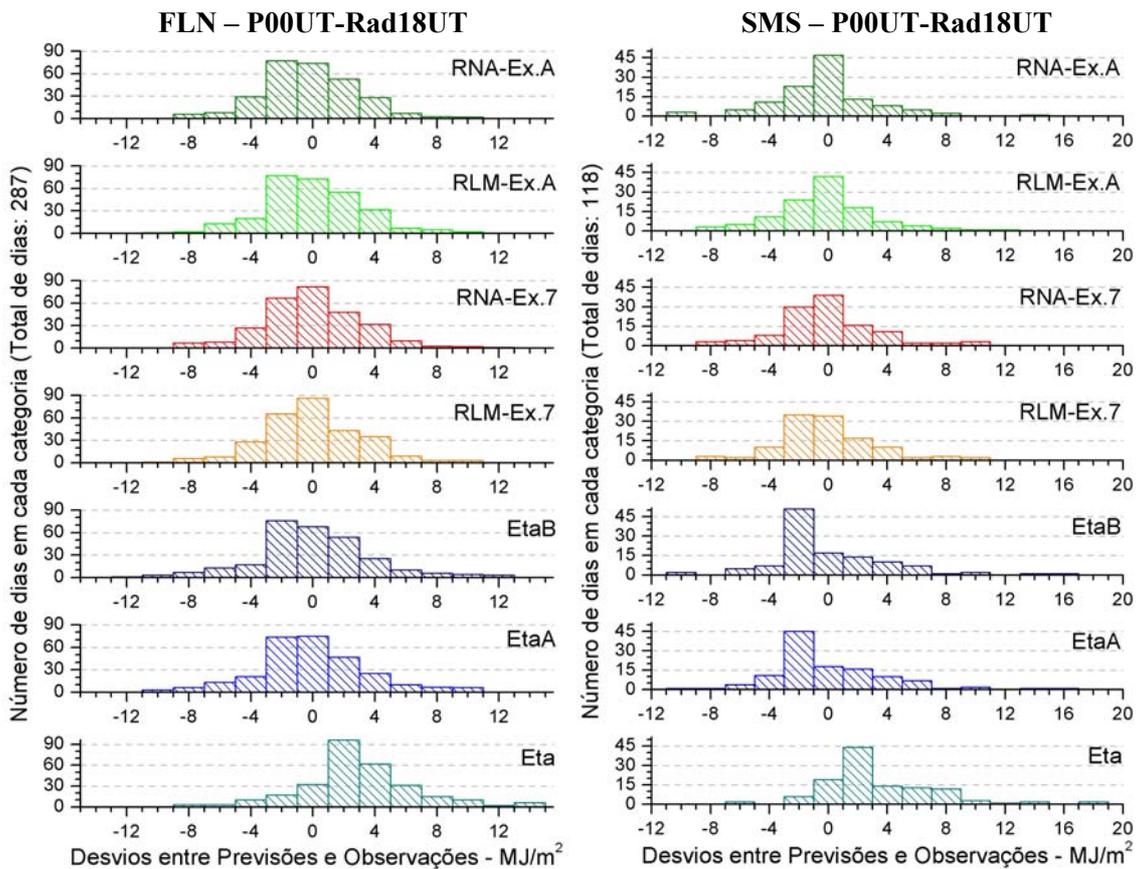


Figura 6.9: Histogramas indicando o número de ocorrências de desvios entre previsões e observações em cada categoria de $2MJ/m^2$ de intervalo, para FLN e SMS.

Verifica-se que, para o modelo Eta, a maior ocorrência é verificada para os desvios entre 1 e $3 MJ/m^2$, o que é relacionado ao viés positivo do modelo. As previsões EtaA e EtaB, apresentam maiores ocorrências de desvios menores, embora para SMS seja verificada grande ocorrência de desvios negativos entre previsões e observações. Por fim, os modelos RLM e RNA parecem apresentar as melhores distribuições de desvios, com maiores ocorrências para os menores desvios e baixas ocorrências para os grandes desvios. RLMs e RNAs com 8 preditores (Ex. 7) parecem apresentar melhor distribuição de desvios para FLN, enquanto os modelos com 36 preditores (Ex. A) parecem levar à maior ocorrência de desvios menores para SMS.

Na Figura 6.10, são apresentados fragmentos das séries temporais dos subconjuntos de testes, relativos ao Verão 2004/2005 e ao Inverno 2005, para as estações FLN e SMS. As observações estão representadas pela cor vermelha. São apresentados os valores das

previsões EtaA (previsão obtida a partir da variável *ocis* do modelo Eta, representando a previsão do modelo Eta livre de viés) na cor verde e previsões dos modelos RLM e RNA com 8 (Ex.7) e 36 (Ex. A) preditores em tons de azul. Pode-se verificar que, na maioria dos dias apresentados, as previsões RLM e RNA encontram-se mais próximas das observações do que as previsões EtaA, confirmando o que foi anteriormente atestado pelos parâmetros de avaliação. Apenas eventualmente, as previsões RLM, RNA e EtaA apresentam valores bastante próximos entre si, e em poucos dias a previsão EtaA apresenta valores mais próximos às observações.

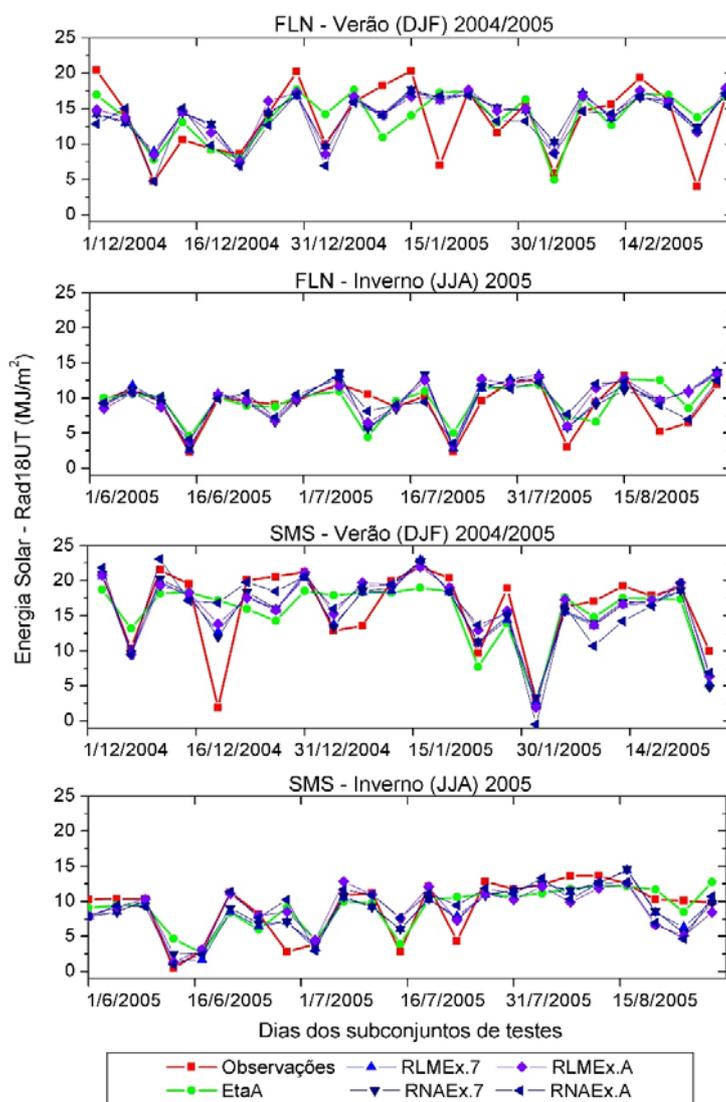


Figura 6.10: Séries temporais de observações e previsões de radiação solar Rad18UT compreendendo os dias dos subconjuntos de testes de FLN e SMS pertencentes aos períodos de Verão 2004/2005 e Inverno 2005.

6.2.5. Teste de Generalização de RLMs e RNAs

Com o objetivo de testar a capacidade de generalização dos modelos RLM e RNA para simular outras localidades, um teste adicional foi realizado. Pretende-se, pois, verificar se os modelos ajustados com dados de uma estação proporcionam bons resultados ao serem aplicados na simulação de previsões de radiação solar para a outra estação, já que, via de regra, modelos estatísticos possuem baixa generalidade. Os modelos RLM e RNA ajustados/treinados para FLN, nos experimentos A e 7, foram empregados na realização de previsões com o subconjunto de testes de SMS, e vice-versa. Os resultados da avaliação destas simulações são apresentados na Tabela 6.15.

Tabela 6.15: Resultados de teste de previsão de radiação solar para cada estação empregando RLMs e RNAs treinadas com dados da outra estação.

| | Preditores | Modelos | ME% | RMSE% |
|--|------------|-------------------------------|---------|---------|
| Previsões para SMS empregando modelos ajustados com dados de FLN | 8 | RLM _{FLN-Ex.7} → SMS | 3,6 % | 38,2 % |
| | 8 | RNA _{FLN-Ex.7} → SMS | 4,9 % | 35,6 % |
| | 36 | RLM _{FLN-Ex.A} → SMS | 216,9 % | 224,0 % |
| | 36 | RNA _{FLN-Ex.A} → SMS | 122,4 % | 133,1 % |
| | | Previsão Eta para SMS | 28,0 % | 43,2 % |
| Previsões para FLN empregando modelos ajustados com dados de SMS | 8 | RLM _{SMS-Ex.7} → FLN | -24,6 % | 38,9 % |
| | 8 | RNA _{SMS-Ex.7} → FLN | -28,0 % | 42,1 % |
| | 36 | RLM _{SMS-Ex.A} → FLN | 175,0 % | 196,7 % |
| | 36 | RNA _{SMS-Ex.A} → FLN | 169,9 % | 176,1 % |
| | | Previsão Eta para FLN | 24,6 % | 40,0 % |

Verifica-se com este teste, que os valores de erros foram maiores do que aqueles encontrados ao empregar os modelos na simulação da própria estação para a qual foram ajustados. Em relação aos erros de previsão de radiação do modelo Eta, os modelos com 8 preditores (Experimento 7) levaram a erros menores para SMS e de mesma ordem para FLN. Para FLN, verificou-se um forte viés negativo, especialmente na previsão dos valores elevados de radiação, possivelmente devido ao fato de que os modelos foram ajustados com base no menor número de dados disponíveis para SMS e que podem não

proporcionar aos modelos a capacidade de generalizar a relação entre as variáveis. Para os modelos com 36 preditores, resultados espúrios foram obtidos em ambas as estações, possivelmente devido ao fato de um ou mais preditores possuírem valores e variâncias específicos para cada estação (fatores locais). O melhor desempenho dos modelos com menor número de preditores, neste teste de generalização, reforça a escolha do conjunto de 8 preditores empregado.

6.3. Avaliação das Previsões para cada Estação do Ano

Uma vez que se observam condições de tempo e clima bastante distintas ao longo das estações do ano, e como os sistemas e processos meteorológicos atuantes em cada estação podem ser bastante distintos para uma dada localidade, os modelos de previsão de tempo nem sempre apresentam o mesmo desempenho na previsão de certas variáveis, para todas as épocas do ano. Ao ser analisado o desempenho de um modelo considerando todos os dados de um ou mais anos, suas deficiências e habilidades para épocas específicas do ano podem ser ocultadas.

A fim de analisar as previsões de radiação solar P00UT-Rad18UT em cada estação do ano separadamente, as previsões geradas para cada dia do subconjunto de testes, foram separadas de acordo com o dia do ano. Foram considerados como dias da estação *Verão* aqueles pertencentes aos meses de Dezembro, Janeiro e Fevereiro. Da mesma forma, os dias dos demais meses foram classificados em: *Outono* – Março, Abril e Maio; *Inverno* – Junho, Julho e Agosto; e *Primavera* – Setembro, Outubro e Novembro.

Dados para pouco mais de um ano encontram-se disponíveis para a estação SMS, de maneira que as estações do ano estariam representadas apenas pelo comportamento observado em poucos dias de um único ano, nos subconjuntos de testes. Assim, a análise das previsões por estação do ano foi realizada apenas para a estação FLN. Os dias com dados válidos entre 2002 e 2005, para FLN, e que constituem o subconjunto de testes (N = 287 dias), foram então divididos por estações: Verão – N = 75 dias; Outono – N = 64 dias; Inverno – N = 73 dias; e Primavera – N = 75 dias. Esses conjuntos de dados para FLN também são um tanto reduzidos para se obter resultados conclusivos, mas foram empregados em caráter preliminar de avaliação.

As previsões Eta, EtaA, EtaB, e dos modelos RLM e RNA (Experimentos A e 7), foram avaliadas separadamente para os conjuntos de dias de cada estação do ano, sendo as previsões confrontadas com as observações (medidas). Foram calculados os coeficientes de correlação, o viés (ME) e o RMSE. Os valores de ME e RMSE foram convertidos em ME% e RMSE%, através da divisão pelas médias das observações estimadas com base nos dias pertencentes a cada estação do ano no subconjunto de testes.

Os coeficientes de correlação (R) calculados para as previsões do modelo Eta e para os modelos dos Experimentos A e 7, em cada estação do ano, são apresentados no gráfico de barras da Figura 6.11. Vale lembrar que os valores de R para EtaA e EtaB são iguais aos encontrados para a previsão de radiação diretamente proveniente do modelo Eta. Além dos valores de R de cada estação do ano, são apresentados os valores estimados para todo o subconjunto de testes, na parte direita do gráfico, com fundo cinza.

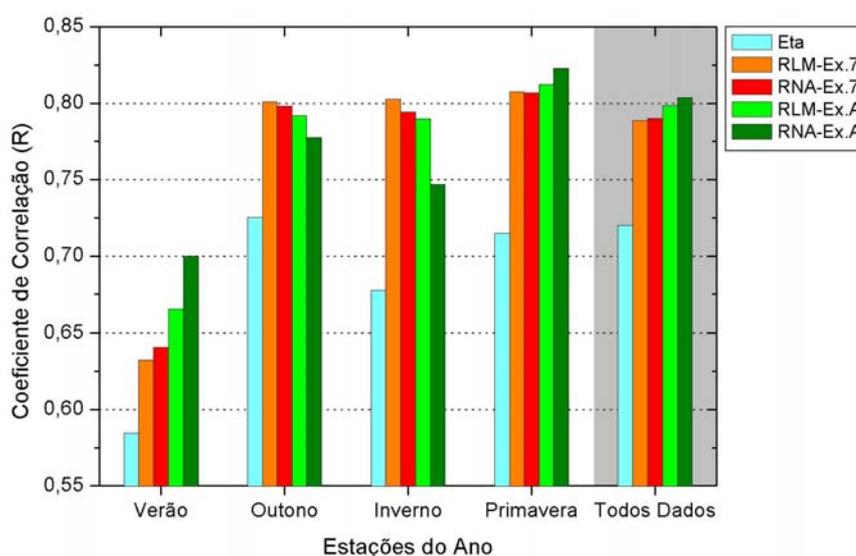


Figura 6.11: Coeficientes de Correlação calculados para várias previsões (P00UT-Rad18UT – subconj. de testes) confrontadas com observações, em cada estação do ano (FLN).

Verifica-se na, Figura 6.11, que o pior desempenho para o modelo Eta ocorre no Verão. Isso pode ser explicado pelo fato de que as maiores temperaturas nesta estação, favorecem a ocorrência de nuvens e precipitação convectivas, sendo que as maiores instabilidades desta época são mais difíceis de serem simuladas pelos modelos de previsão de tempo. O desempenho para o modelo Eta no inverno é melhor do que no verão, mas ainda menor do que o observado considerando todos os dias do ano. Apesar

de o modelo Eta simular melhor os sistemas frontais, predominantes no inverno, pequenos erros na simulação do deslocamento dos sistemas podem ocasionar condições de nuvens completamente diferentes num dado instante, em uma dada estação. Soma-se a isso o fato de Florianópolis estar situada numa região costeira, combinando um regime frontal com influências litorâneas. Os melhores desempenhos para o modelo Eta são observados no Outono e na Primavera.

Uma vez que os modelos RLM e RNA são dependentes das condições de tempo previstas, e estas mesmas condições de tempo estão ligadas à própria previsão de radiação do modelo Eta, é esperado que os desempenhos dos modelos, ao longo das estações, tenham um comportamento semelhante. Contudo, com exceção do verão, os modelos RLM e RNA apresentam um desempenho mais constante para as demais estações, ainda que a RNA com 36 preditores (RNA-Ex.A) apresente maiores variações.

Em todas as estações, os modelos RLM e RNA apresentam coeficientes de correlação mais elevados do que os encontrados para o modelo Eta. O modelo RNA-Ex.A apresentou os maiores valores para R no verão e na primavera. No outono e no inverno, o melhor desempenho coube ao modelo RLM-Ex.7. Os modelos RLM e RNA com 8 preditores (Ex.7) não apresentam diferenças significativas comparados aos modelos de 36 preditores (Ex.A), exceto no verão, quando a diferença é maior. É bastante provável que exista uma diferença de influência de diferentes preditores em cada estação do ano, em cada localidade.

Na Figura 6.12, os valores de ME% para várias previsões, calculados para cada estação do ano, em FLN, são apresentados. Verifica-se que, apesar de as modelos RLM e RNA e as previsões EtaA e EtaB apresentarem viés muito próximo de zero, ao avaliar o ano todo, o viés é maior ao serem consideradas as estações do ano individualmente (-5 a +5%). Contudo, o viés do modelo Eta é muito maior e positivo em todas as estações (+18 a +29%), apresentando seu menor valor no verão. Os valores de ME% para as previsões EtaA, EtaB, RLM e RNA, são, em geral, negativos no verão, positivos no outono e negativos no inverno. Na primavera, as previsões EtaA e EtaB possuem viés positivo, e os modelos RLM e RNA, viés negativo.

Os valores de RMSE% para diversas previsões, em cada estação do ano, são mostrados no gráfico de barras da Figura 6.13. Verifica-se que, em todas as estações, o RMSE% para modelos RLM e RNA é menor do que para EtaA e EtaB, e muito menor do que para a previsão do modelo Eta. O menor valor de RMSE% para o modelo Eta é verificado no verão.

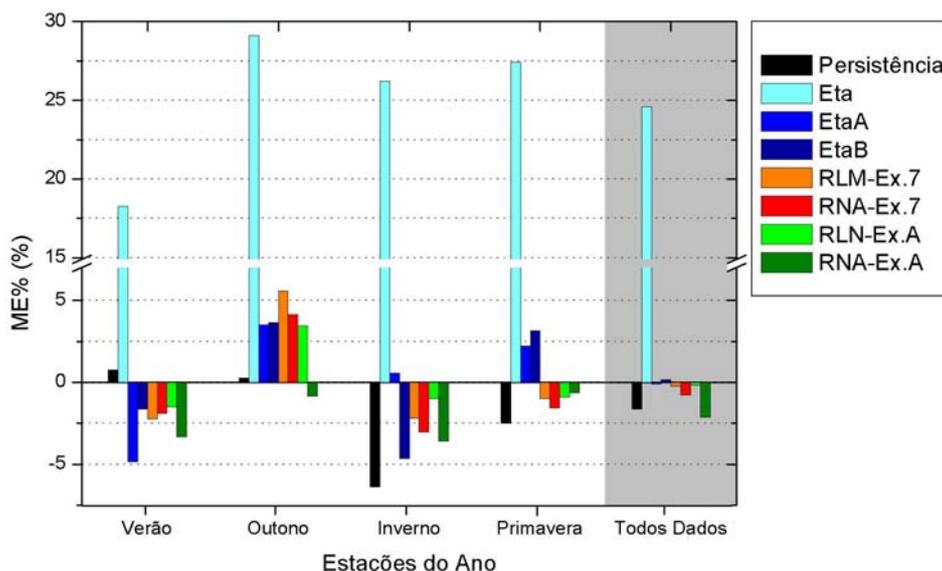


Figura 6.12: ME% para várias previsões (P00UT- Rad18UT – subconjunto de testes) confrontadas com observações, em cada estação do ano, para FLN.

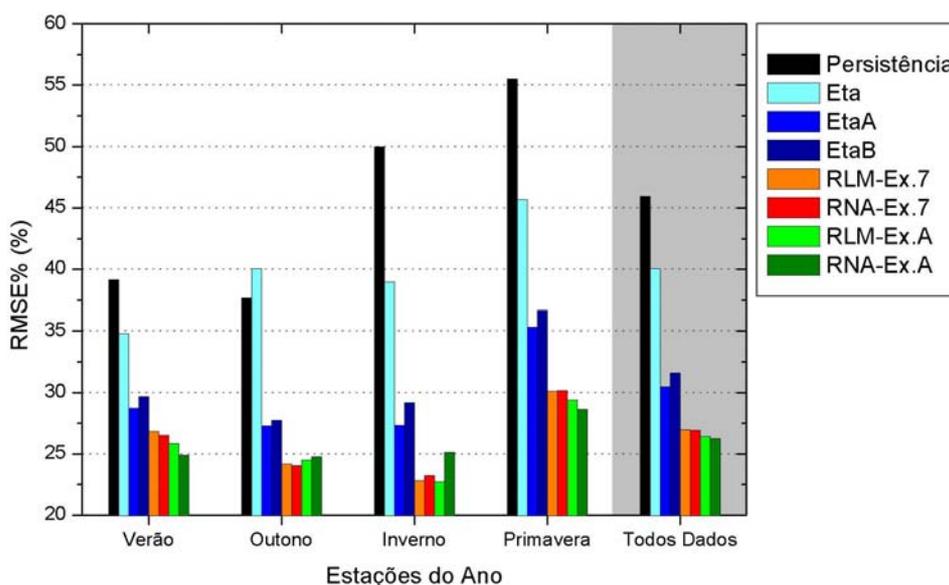


Figura 6.13: RMSE% para várias previsões (P00UT-Rad18UT – subconjunto de testes) confrontadas com observações, em cada estação do ano, para FLN.

Os valores para RMSE% para os modelos RLM e RNA, são bastante próximos em todas as estações, e assumem seus maiores valores na primavera, quando também ocorrem os maiores valores de RMSE para todas as previsões. O fato de os maiores valores de RMSE% ocorrerem na estação com coeficientes de correlação mais elevados (primavera) parece algo contraditório. Parece haver um comportamento linear entre observações e previsões (R elevado), mas afastado da reta de previsão perfeita ($P = O$), de maneira ao RMSE ser alto. Contudo, nesta estação, o viés (ME%) é menor do que nas outras estações do ano. A explicação deste impasse está no fato de que, na primavera, as previsões de radiação são superestimadas nos dias de menor incidência e superestimadas nos dias de maior incidência, de maneira a haver uma compensação no viés, resultando em valores de ME pequenos e de RMSE grandes.

Com o objetivo de verificar o ganho em RMSE dos novos modelos frente à previsão de radiação solar do modelo Eta, em cada estação do ano, o $Skill(RMSE, Eta)$ é apresentado na Figura 6.14 para as previsões EtaA, EtaB, RLM e RNA. Os modelos de RLM e RNA apresentam ganhos sempre superiores aos observados para EtaA e EtaB. O maior ganho é verificado no Inverno, quando os modelos RLM-Ex.7, RNA-Ex.7 e RLM-Ex.A, apresentam $Skill$ superior a 0,4 sobre o Eta (40% de ganho sobre o Eta) e os menores valores de RMSE%.

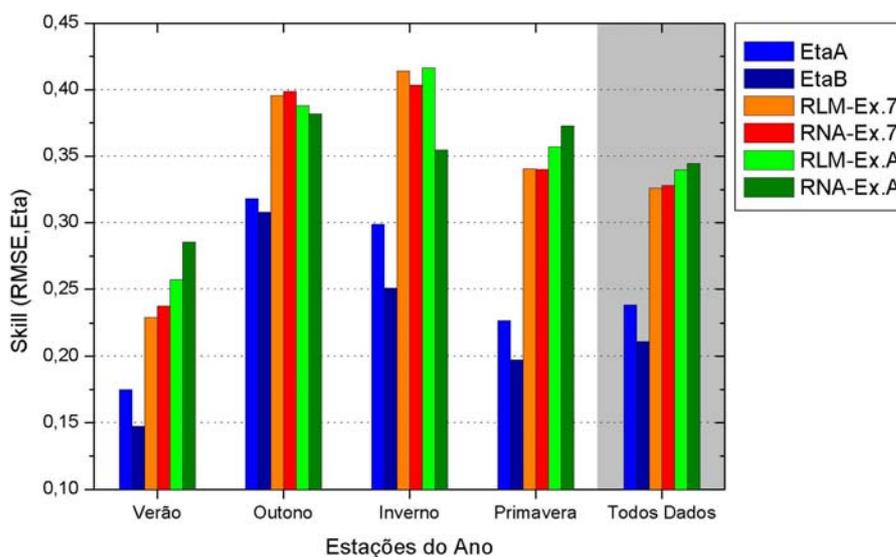


Figura 6.14: $Skill(RMSE, Eta)$ para várias previsões (P00UT-Rad18UT – subconjunto de testes) confrontadas com observações, em cada estação do ano, para FLN.

Os menores valores de *Skill*, para todos os modelos apresentados na Figura 6.14, ocorrem no verão, período em que o modelo Eta apresenta o menor valor de RMSE% (Figura 6.13). No outono, inverno e primavera, os valores de *Skill* dos modelos RLM e RNA são maiores do que aqueles encontrados na análise dos dados de todo o ano.

6.4. Previsões de Integral Diária de Radiação

Objetivando verificar o desempenho de RLMs e RNAs na estimativa de integrais diárias de radiação, o conjunto de preditores do Experimento 7 (8 preditores) foi empregado para ajustes de RLMs e RNAs para os períodos diários Rad12UT e Rad24UT. O procedimento foi o mesmo realizado nos experimentos anteriores, tendo o ajuste/treinamento sido realizado com os subconjuntos de treinamento e validação, e a avaliação realizada com o subconjunto de testes. Os dados P00UT-Rad12UT e P00UT-Rad24UT foram subdivididos empregando a mesma subdivisão homogênea aplicada anteriormente para P00UT-Rad18UT, de maneira que os subconjuntos são compostos de dados para os mesmos dias.

Após o ajuste dos modelos, e a simulação, empregando os dados do conjunto de testes, os valores de radiação previstos para Rad12UT e Rad24UT foram somados aos valores previstos para Rad18UT (Experimento 7), obtendo-se, para cada dia, a integral diária de energia solar incidente. Para comparação, as medidas de radiação e previsões do modelo Eta para os mesmos períodos foram tomadas dos subconjuntos de testes, e as integrais diárias foram calculadas.

Os parâmetros de avaliação para as previsões Eta, RLM e RNA foram calculados para cada período diário e para a integral diária, através da comparação com os respectivos valores das observações. Na Tabela 6.16 são apresentados os coeficientes de correlação e os erros médios ME e RMSE para cada previsão, em cada período diário e para integral diária. São apresentados valores para RMSE%, calculados empregando as médias das observações, apresentadas para cada período diário. Também são apresentados valores de *Skill* em termos de RMSE, em relação ao modelo Eta. Diagramas de dispersão para integrais diárias, comparando as previsões Eta, RLM e RNA com as observações, para FLN e SMS, são apresentados na Figura 6.15.

Tabela 6.16: Parâmetros de avaliação para as previsões Eta, RLM e RNA, para três períodos diários e integral diária, para FLN e SMS.

| | Índices | FLN | | | SMS | | |
|--|------------------|-------|--------|--------|-------|--------|--------|
| | | Eta | RLM | RNA** | Eta | RLM | RNA** |
| Rad12UT $\bar{O}_{FLN}=1,69^*$ $\bar{O}_{SMS}=1,42^*$ | R | 0,789 | 0,794 | 0,804 | 0,867 | 0,894 | 0,907 |
| | ME | 0,889 | -0,041 | -0,050 | 0,730 | 0,028 | 0,014 |
| | RMSE | 1,208 | 0,728 | 0,713 | 0,992 | 0,568 | 0,532 |
| | RMSE% | 71,3% | 42,9% | 35,5% | 69,7% | 39,9% | 37,4% |
| | Skill (RMSE,Eta) | | 0,397 | 0,395 | | 0,427 | 0,464 |
| Rad18UT $\bar{O}_{FLN}=11,81^*$ $\bar{O}_{SMS}=11,83^*$ | R | 0,720 | 0,789 | 0,790 | 0,775 | 0,848 | 0,848 |
| | ME* | 2,906 | -0,026 | -0,088 | 3,314 | -0,124 | -0,080 |
| | RMSE* | 4,729 | 3,187 | 3,178 | 5,115 | 3,262 | 3,267 |
| | RMSE% | 40% | 27% | 26,9% | 43,2% | 27,6% | 27,6% |
| | Skill (RMSE,Eta) | | 0,326 | 0,328 | | 0,362 | 0,361 |
| Rad24UT $\bar{O}_{FLN}=2,25^*$ $\bar{O}_{SMS}=3,09^*$ | R | 0,663 | 0,783 | 0,799 | 0,789 | 0,837 | 0,854 |
| | ME* | 1,041 | 0,054 | 0,036 | 1,021 | 0,042 | 0,011 |
| | RMSE* | 1,601 | 0,896 | 0,867 | 1,644 | 1,140 | 1,079 |
| | RMSE% | 71,3% | 39,9% | 38,6% | 53,2% | 36,9% | 34,9% |
| | Skill (RMSE,Eta) | | 0,440 | 0,458 | | 0,307 | 0,344 |
| Integral Diária $\bar{O}_{FLN}=15,75^*$ $\bar{O}_{SMS}=16,34^*$ | R | 0,766 | 0,829 | 0,832 | 0,833 | 0,883 | 0,885 |
| | ME* | 4,835 | 0,001 | -0,102 | 5,066 | -0,038 | -0,054 |
| | RMSE* | 6,922 | 4,064 | 4,034 | 7,124 | 4,244 | 4,212 |
| | RMSE% | 43,9% | 25,8% | 25,6% | 43,6% | 26% | 25,8% |
| | Skill (RMSE,Eta) | | 0,413 | 0,417 | | 0,404 | 0,409 |

* Unidade: MJ/m². ** RNA-8-16-8-1 (Ex.7)

Além de os valores para ME serem mais próximos de zero para os modelos RLM e RNA, estes modelos apresentam os menores valores de RMSE, em ambas as estações (FLN e SMS) e para todos os períodos diários. Os coeficientes de correlação para RLMs e RNAs são maiores dos que os observados para o modelo Eta em todos os períodos diários e ambas as estações.

O ganho em RMSE sobre o modelo Eta, estimado para RLM e RNA pelo *Skill*, é de 40%, ou superior, para Rad12UT e Rad24UT, em FLN. O menor *Skill* é observado para Rad24UT em SMS, ainda assim indicando um ganho de 30% sobre o modelo Eta. Os parâmetros de avaliação para Rad18UT são os mesmos apresentados anteriormente nas Tabelas 6.13 e 6.14 (Experimento 7).

Previsões de Integrais Diárias de Energia Solar

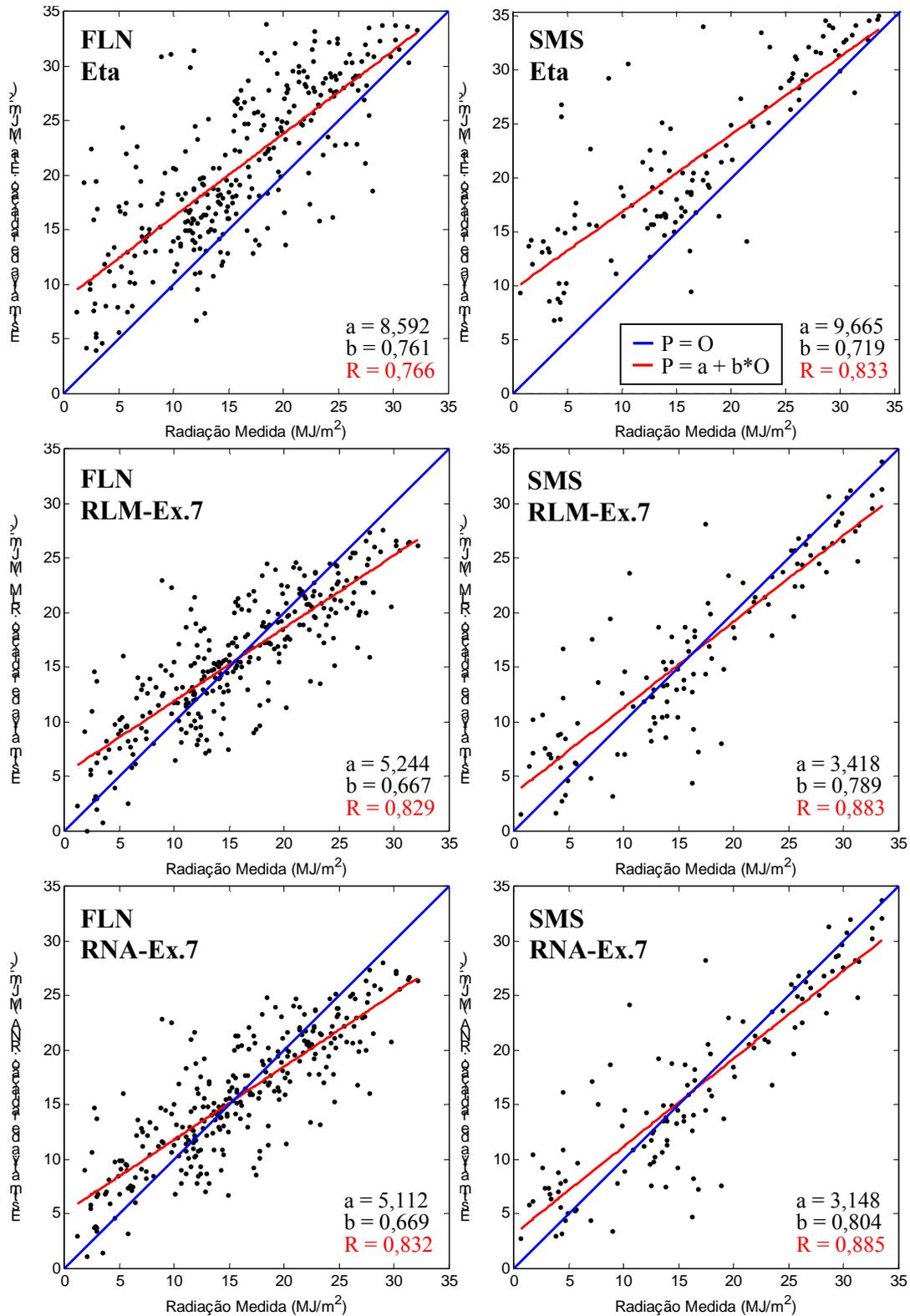


Figura 6.15: Diagramas de dispersão para previsões (P00UT) de integrais diárias de energia solar (Eta, RLM e RNA), para FLN e SMS.

Para as previsões de integrais diárias de energia solar, foram verificados RMSE% em torno de 26% para ambos os modelos, RNA e RLM, e ambas as estações. As previsões do modelo Eta, em ambas as estações, apresentaram RMSE% de aproximadamente 44%. Assim, os ganhos em RMSE conseguidos pelos novos modelos sobre o modelo Eta, foram superiores a 40%. Esta redução em RMSE conseguida pelos novos modelos para as integrais diárias pode ser verificada graficamente na Figura 6.15, onde são apresentados diagramas de dispersão.

Pelos diagramas de dispersão da Figura 6.15, pode-se verificar que, para os modelos RLM e RNA ocorre uma menor dispersão dos pontos, situando-se mais próximos da linha de previsão perfeita, em comparação ao modelo Eta. A maior linearidade entre previsões e observações, nos novos modelos, está de acordo com a elevação dos coeficientes de correlação, de 0,77 (Eta) para 0,83 (RLM e RNA), aproximadamente, para FLN, e de 0,83 (Eta) para 0,88 (RLM e RNA), para SMS. Novamente, não é observado um desempenho significativamente maior para os modelos de RNA, em comparação aos modelos RLM, havendo grande semelhança de comportamento entre os diagramas de dispersão dos dois modelos.

6.5. Análise de Previsões empregando Subdivisão Não-Homogênea

Nas análises apresentadas até a seção anterior, os ajustes e avaliações de modelos RLM e RNA, foram realizados com base em subconjuntos de treinamento, validação e teste, gerados através de uma subdivisão homogênea dos dados, como descrito no início da seção 6.2. A fim de verificar a habilidade dos modelos em estabelecer relações entre os preditores e simular a radiação solar utilizando outras porções do conjunto de dados, um novo teste foi realizado, empregando uma nova subdivisão dos dados P00UT-Rad18UT das estações FLN e SMS.

A nova subdivisão consistiu numa subdivisão não-homogênea, em que uma seqüência contínua de dias foi tomada como subconjunto de testes, sendo o restante dos dados separados em subconjuntos de treinamento e validação. Assim, foram tomados para os subconjuntos de testes os dias finais dos conjuntos de dados das duas estações, totalizando a quarta parte dos dias válidos. Os dias restantes foram divididos

homogeneamente de maneira que, a cada três dias, dois dias compõem o subconjunto de treinamento e um dia compõe o subconjunto de validação. Para uma melhor verificação das diferenças entre as subdivisões, histogramas são apresentados na Figura 6.16. Os histogramas apresentam o número de dias que compõe cada subconjunto, a cada 50 dias das seqüências de dias válidos para as estações FLN e SMS.

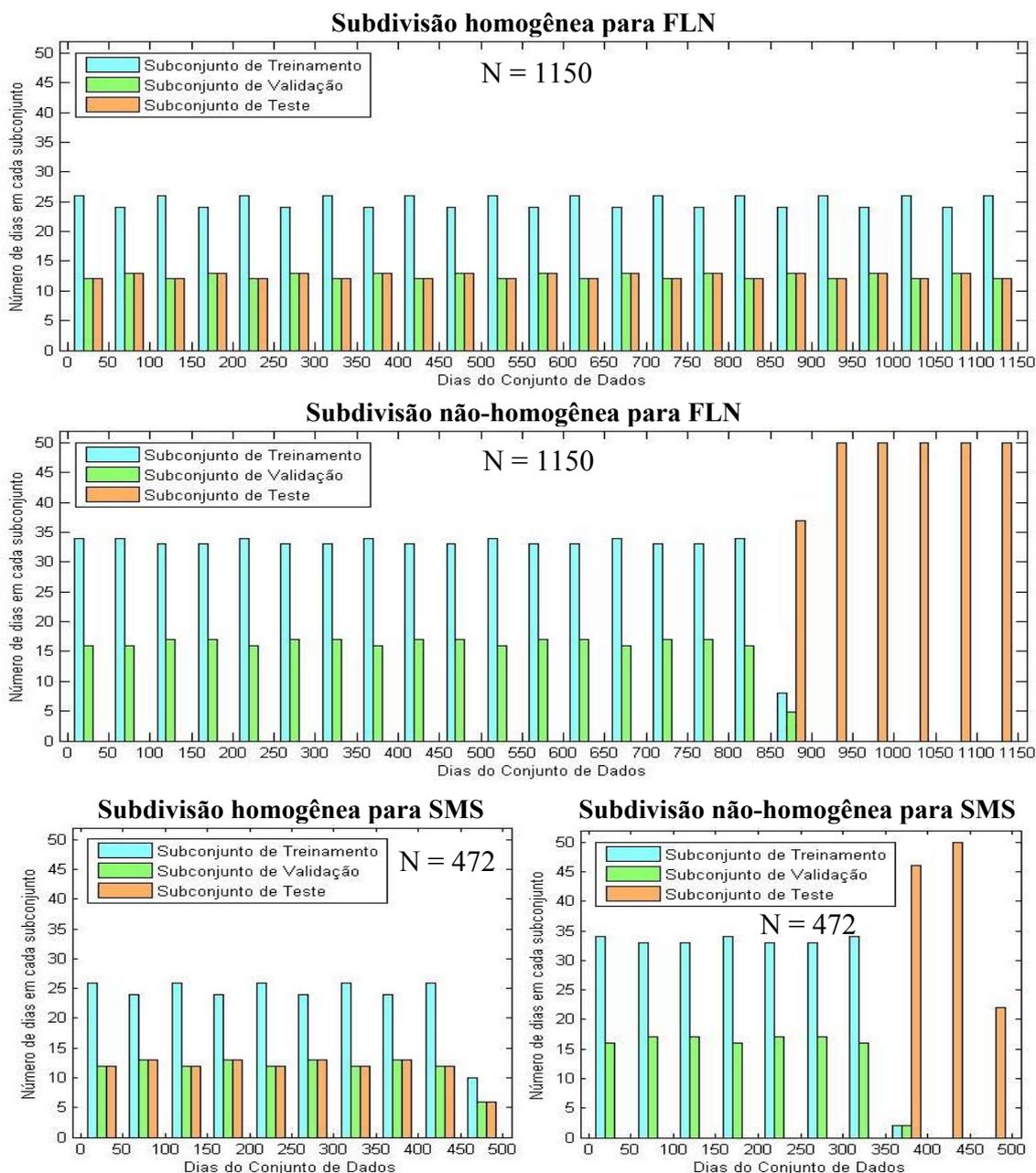


Figura 6.16: Histogramas ilustrando as subdivisões homogênea e não-homogênea dos dados P00UT-Rad18UT de FLN e SMS em subconjuntos de treinamento (FLN: N=575; SMS: N=236), validação (FLN: N=288; SMS: N=118) e testes (FLN: N=287; SMS: N=118).

Em ambas as subdivisões, homogênea e não-homogênea, 50% dos dados compõem os subconjuntos de treinamento (576 dias para FLN e 236 dias para SMS), 25% os subconjuntos de validação (287 dias para FLN e 118 dias para SMS) e 25% pertencem aos subconjuntos de testes (287 dias para FLN e 118 dias para SMS).

Empregando os subconjuntos gerados pela subdivisão não-homogênea dos dados P00UT-Rad18UT e o grupo de preditores utilizado no Experimento 7, novos ajustes de RLMs e treinamento de RNAs foram realizados para ambas as estações. As simulações foram realizadas sobre os novos subconjuntos de testes, de onde também foram tomados os dados da variável *ocis* (previsão de radiação do modelo Eta). As previsões Eta, RLM e RNA foram avaliadas frente às observações, e os parâmetros de avaliação foram calculados. Os valores de R, R², ME, MAE, RMSE, RMSE% para as previsões são apresentados na Tabela 6.17, juntamente com os valores de *Skill* das previsões RLM e RNA frente ao modelo Eta, em termos de RMSE. Na Figura 6.17, diagramas de dispersão para as previsões Eta, RLM e RNA comparadas às observações são apresentados, para FLN e SMS.

Tabela 6.17: Parâmetros de avaliação das previsões Eta, RLM e RNA, frente radiação medida, utilizando conjunto de testes da subdivisão não-homogênea de P00UT-Rad18UT.

| Índices | FLN | | | SMS | | |
|---------------------------|--|---------|---------|---|---------|---------|
| | $\bar{O} = 11,04 \text{ MJ/m}^2$ N = 287 | | | $\bar{O} = 9,98 \text{ MJ/m}^2$ N = 118 | | |
| | Eta | RLM | RNA* | Eta | RLM | RNA* |
| R | 0,728** | 0,798** | 0,802** | 0,687** | 0,781** | 0,755** |
| R ² | 0,530 | 0,636 | 0,643 | 0,472 | 0,609 | 0,571 |
| ME (MJ/m ²) | 2,559 | 0,014 | 0,186 | 4,463 | 1,158 | 0,772 |
| MAE (MJ/m ²) | 3,583 | 2,362 | 2,346 | 4,543 | 2,612 | 2,710 |
| RMSE (MJ/m ²) | 4,381 | 3,024 | 2,995 | 6,001 | 3,539 | 3,591 |
| ME% | 23,2% | 0,1% | 1,7% | 44,7% | 11,6% | 7,7% |
| RMSE% | 39,7% | 27,4% | 27,1% | 60,1% | 35,5% | 36% |
| Skill (RMSE,Eta) | | 0,310 | 0,316 | | 0,410 | 0,402 |

* RNA-8-16-8-1

** Nível de significância p<0,001

Os valores da Tabela 6.17 podem ser confrontados com os resultados encontrados para as avaliações realizadas com a subdivisão homogênea, para o modelo Eta (Tabela 6.4 ou 6.5) e para os modelos RLM e RNA do Experimento 7 (Tabelas 6.13 e 6.14). Verifica-se que as previsões do modelo Eta, neste novo subconjunto de testes, apresentam ME e

RMSE mais elevados do que no subconjunto de testes utilizado anteriormente, apesar de as diferenças serem pequenas para FLN. O coeficiente de correlação para o modelo Eta no novo subconjunto de testes é aproximadamente igual ao anterior para FLN, e menor para SMS.

Para FLN, os valores de coeficiente de correlação, ME, MAE, RMSE, RMSE% e Skill(RMSE,Eta), para os modelos RLM e RNA, são aproximadamente iguais aos encontrados no Experimento 7, sendo verificados valores para R em torno de 0,8, RMSE% em torno de 27% e ganho em RMSE sobre o Eta de aproximadamente 31%. Para SMS, as diferenças são grandes entre os testes realizados com os subconjuntos de testes das duas subdivisões, e os erros são maiores na subdivisão não-homogênea. Porém, como o modelo Eta apresenta RMSE também elevado no novo subconjunto de testes para SMS, os modelos RLM e RNA proporcionam ganhos superiores a 40% em RMSE para esta estação.

Para os modelos RLM e RNA, valores de ME e ME% menores do que para o modelo Eta são observados na Tabela 6.17, para ambas as estações. Assim como no Experimento 7, os valores de ME para FLN foram próximos de zero neste novo subconjunto de testes. Porém, para SMS, a subdivisão não-homogênea dos dados levou a valores mais elevados de ME para RLM e RNA, em relação ao Experimento 7.

As maiores diferenças observadas para SMS entre as duas subdivisões de dados, nos parâmetros de avaliação para as previsões Eta, RLM e RNA, podem ser explicadas pelo fato de que a quantidade de dados para esta estação ser pouco maior do que um ano. Assim, na subdivisão não-homogênea, as avaliações dos modelos são realizadas com base em um período não representativo do comportamento anual, que abrange dias de julho a outubro de 2005. Na subdivisão homogênea, as avaliações são realizadas sobre um subconjunto de testes composto por amostras de todo o período de dados, entre julho de 2004 e outubro de 2005. Para FLN, as diferenças entre as subdivisões foram menores, pois o conjunto de dias válidos corresponde a um período de dados de pouco menos do que quatro anos, de maneira que o subconjunto de testes da subdivisão não-homogênea abrange quase todos os dias disponíveis do ano de 2005.

Análise de previsões para subdivisão não-homogênea dos dados

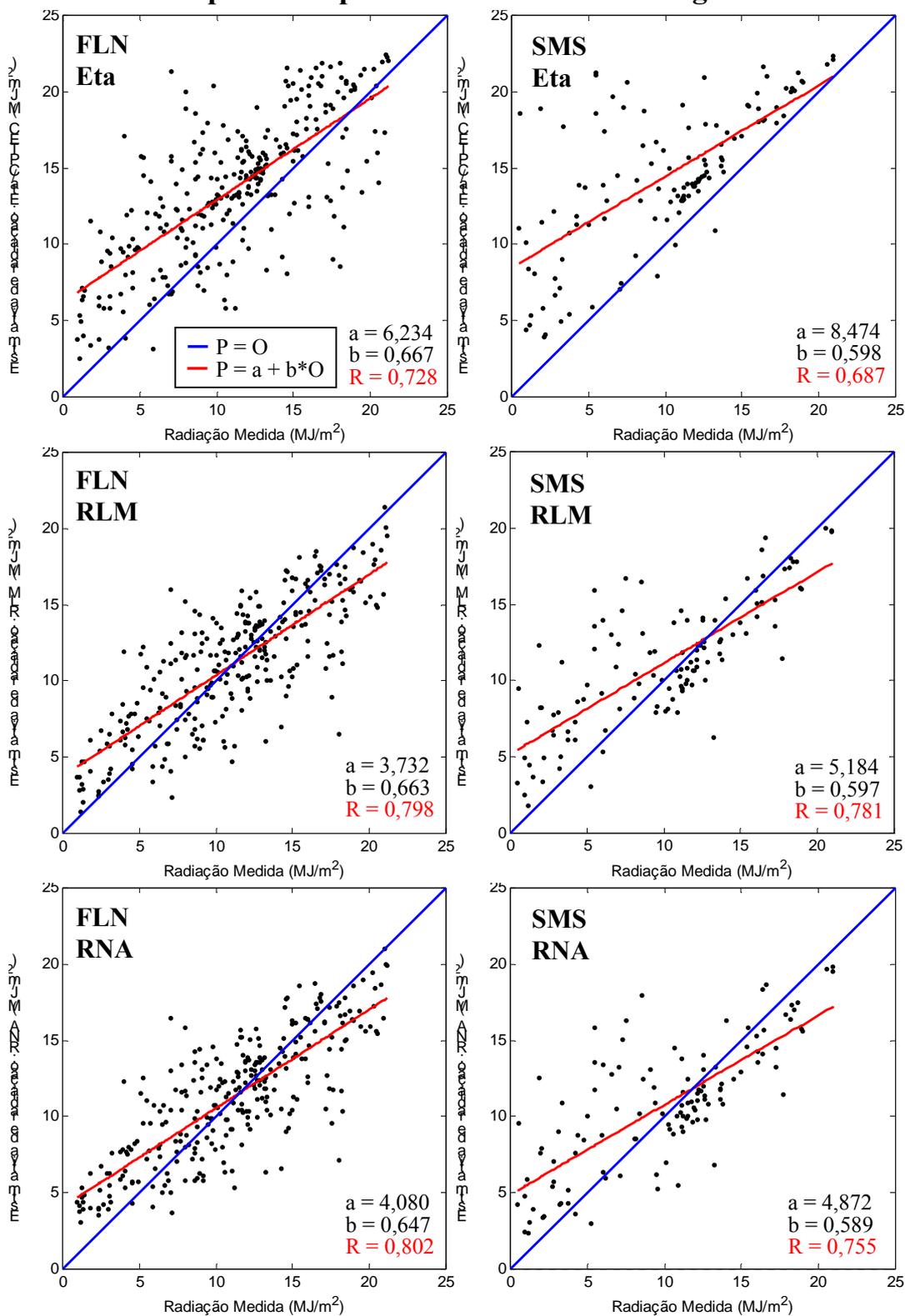


Figura 6.17: Diagramas de dispersão para previsões Eta, RLM e RNA, geradas com base na subdivisão não-homogênea dos dados P00UT-Rad18UT, para FLN e SMS.

Os modelos RLM e RNA apresentaram parâmetros de avaliação muito parecidos entre si, para FLN. Essas semelhanças são também observadas nos diagramas de dispersão da Figura 6.17. Para SMS, o modelo RNA apresentou um desempenho ligeiramente menor em termos de R e RMSE, mas o valor de ME foi menor que o observado para RLM. A comparação dos diagramas de dispersão para RLM e RNA para SMS, revelam diferenças pequenas de comportamento entre as duas previsões.

6.6. Avaliação da Antecedência da Previsão

Uma análise de desempenho de previsões com diferentes antecedências foi realizada utilizando os dados P00UT, P12UT, P24UT, P36UT e P48UT, para o período diário Rad18UT. Cada período de dados *PhhUT*-Rad18UT foi dividido em subconjuntos de treinamento, validação e teste, a fim de serem ajustados modelos de RLM e RNA.

A subdivisão empregada foi a subdivisão não-homogênea introduzida na seção anterior (Figura 6.16). A subdivisão homogênea não foi empregada, pois, nesta subdivisão, ao haver um único dia ausente em um conjunto *PhhUT* e que não esteja ausente em outro conjunto, leva os subconjuntos de teste a serem constituídos por dias totalmente diferentes. Assim, ao realizar as avaliações sobre os subconjuntos de teste, não seria coerente comparar parâmetros gerados com base em dias diferentes. Na subdivisão não-homogênea esse problema é evitado, uma vez que os 287 últimos dias para FLN e 118 últimos dias para SMS são simplesmente tomados para os subconjuntos de testes.

Os modelos de RLM e RNA foram ajustados empregando os 8 preditores do Experimento 7 e foram utilizados para simular a radiação solar incidente nos dias que constituem os subconjuntos de teste. A partir dos subconjuntos de testes, também foram tomadas as previsões de radiação solar pelo Eta (*ocis*) e as previsões por persistência. As previsões por persistência para P00UT-Rad18UT para cada dia são os valores de radiação medida do período Rad18UT do dia anterior. Para as previsões P12UT e P24UT, foram tomados como persistência os valores das medidas para dois dias antes. Por fim, para as previsões por persistência P36UT e P48UT foram utilizados os dados medidos três dias antes.

Para cada previsão (Persistência, Eta, RLM e RNA) foram calculados os parâmetros de avaliação para cada antecedência hh , através da comparação das previsões $P_{hh}UT$ -Rad18UT com as observações, nos dias que compõe o subconjunto de testes. Os valores encontrados para R , R^2 , $ME\%$ e $RMSE\%$ são apresentados graficamente, permitindo ser mais bem observada a variação desses parâmetros com o aumento da antecedência das previsões.

A Figura 6.18 apresenta a variação do coeficiente de correlação (R) em função da antecedência das previsões. Para FLN, observa-se o já esperado decréscimo do coeficiente de correlação com o aumento da antecedência, uma vez que previsões geradas com maior antecedência naturalmente possuem erros maiores. Para SMS, verifica-se um ligeiro aumento em R para as previsões $P_{12}UT$ dos modelos Eta, RLM e RNA, seguindo-se o decréscimo dos valores para as maiores antecedências. Em ambas as estações, os coeficientes de correlação para os modelos RLM e RNA são maiores que os observados para o modelo Eta.

A aparente “melhora” da previsão $P_{12}UT$ sobre $P_{00}UT$, tanto para o modelo Eta, como para os modelos RLM e RNA, para SMS, não é um comportamento incoerente. Normalmente, quanto maior a antecedência das previsões, maiores são os erros e pior a qualidade de previsão. Contudo, as previsões $P_{12}UT$, geradas às 12UT das vésperas das observações, podem vir a apresentar um melhor desempenho que a previsão mais recente, $P_{00}UT$, devido ao fato de que o modelo leva certo tempo para se estabilizar e para “construir” sistemas e processos atmosféricos. Assim, ao dispor de mais tempo entre o instante inicial e o instante de interesse, as previsões $P_{12}UT$ podem vir a prover informações de maior qualidade acerca da cobertura de nuvens e outras propriedades atmosféricas.

Para FLN, não se observa grandes diferenças entre os coeficientes de correlação para as previsões por com antecedência de 0, 12 e 24 horas, para os modelos RLM e RNA. Para o modelo Eta parece também não haver uma queda tão abrupta em R , com o aumento da antecedência. Essa queda é mais destacada ao se considerar o coeficiente de determinação (R^2), que corresponde ao quadrado do valor de R e representa a fração da

variância das observações que é comum às previsões. O comportamento de R^2 com a antecedência das previsões é apresentado na Figura 6.19.

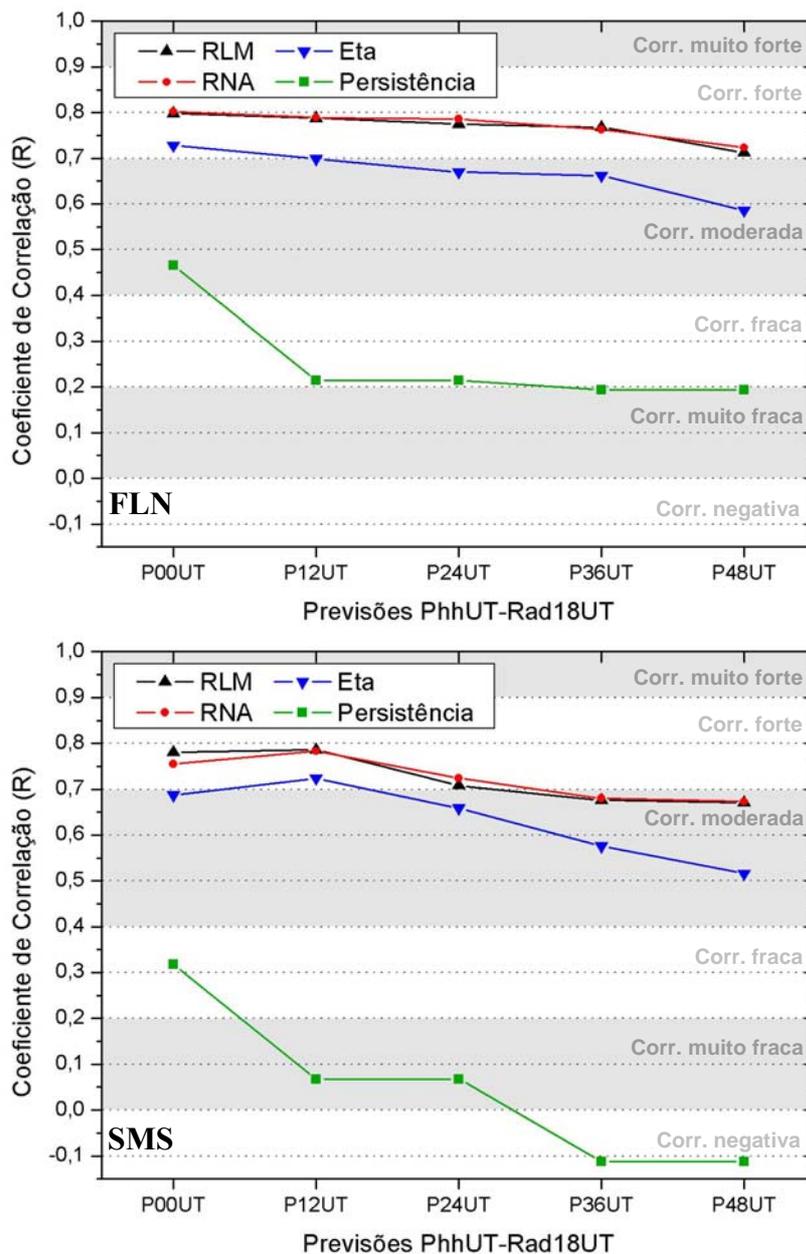


Figura 6.18: Variação dos coeficientes de correlação entre previsões e observações para previsões (*PhhUT-Rad18UT*) com diferentes antecedências, para FLN e SMS.

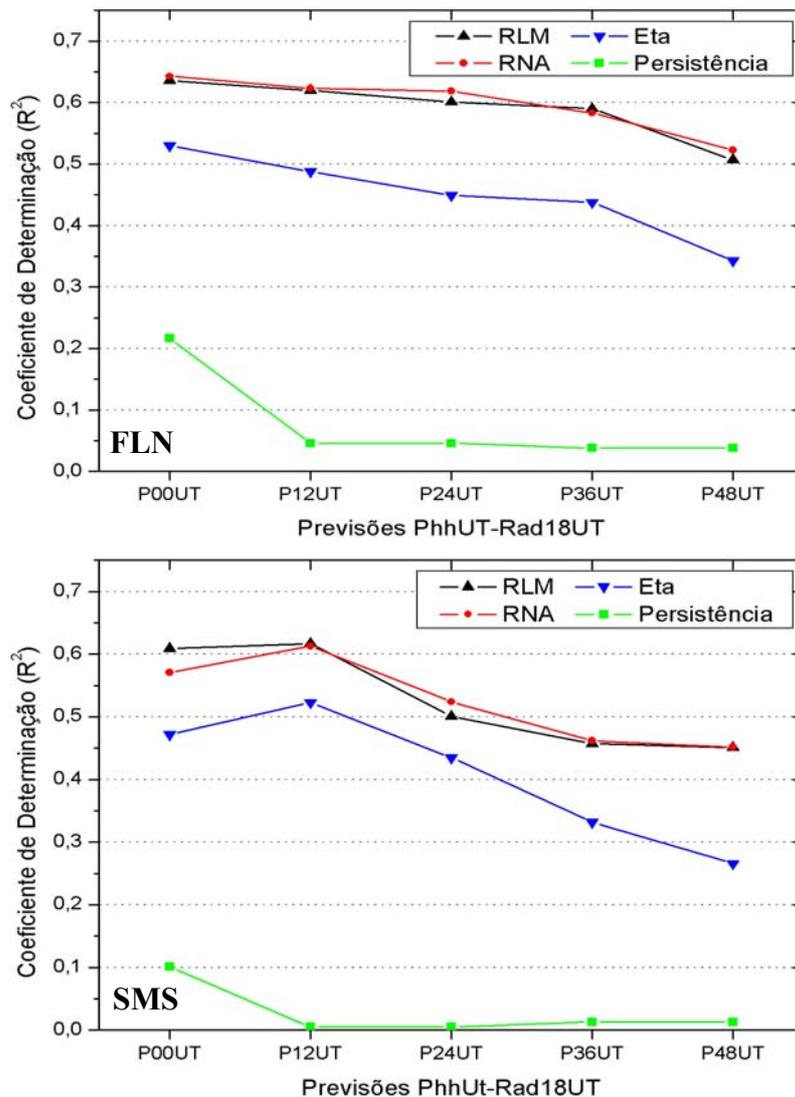


Figura 6.19: Variação dos coeficientes de determinação entre previsões e observações para previsões (*Phh*UT-Rad18UT) com diferentes antecedências, para FLN e SMS.

O comportamento dos erros médios, ME% e RMSE%, com o aumento da antecedência das previsões, é apresentado na Figura 6.20. Para FLN, os valores de ME% são próximos de zero para os modelos RLM e RNA, enquanto para o modelo Eta observam-se valores em torno de 20%, para todas as previsões. Para SMS, os valores de ME% parecem não variar muito com a antecedência das previsões, assumindo valores em torno de 40% para o modelo Eta e em torno de 10% para os modelos RLM e RNA.

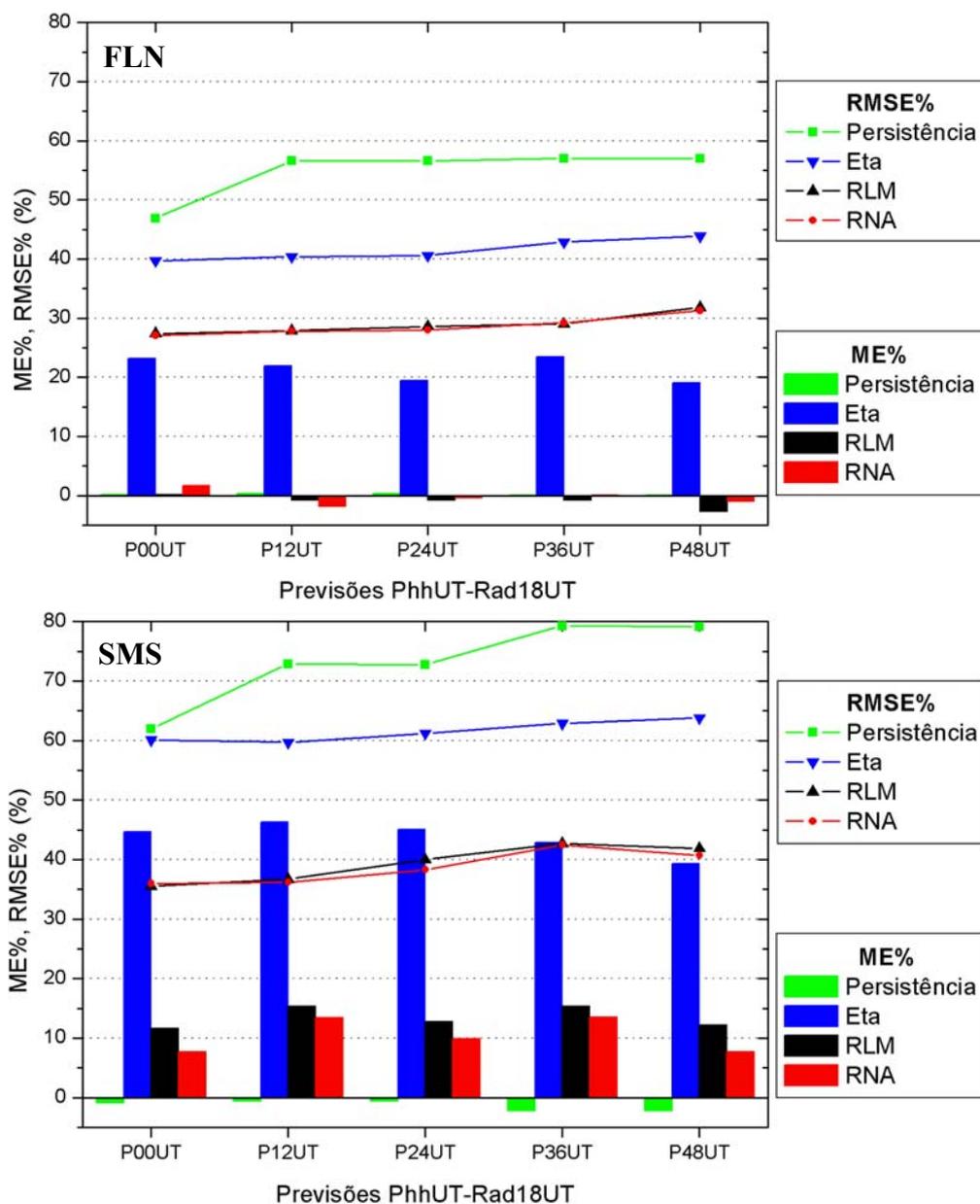


Figura 6.20: Erros médios percentuais (ME% e RMSE%) para previsões (*PhhUT-Rad18UT*) com diferentes antecedências, para FLN e SMS.

Na Figura 6.20, não parece haver grande aumento dos valores de RMSE% com o aumento da antecedência das previsões, ao menos para aquelas analisadas. Em todas as antecedências, as previsões RLM e RNA levam a valores menores de RMSE% em ambas as estações. Considerando as duas estações e todas as antecedências analisadas, verifica-se um ganho em RMSE dos modelos RLM e RNA sobre o Eta, entre 28 e 40%.

Novamente, nesta análise, não foi verificada diferença significativa de desempenho entre RNAs e RLMS, embora ambos os modelos promovam melhorias na previsão de radiação solar. Apesar de não verificado o esperado ganho das RNAs sobre as RLMS, foi verificado para ambas as estações (FLN e SMS) que o uso de outras variáveis de saída do modelo Eta, como preditores de modelos de refinamento, leva a previsões com maiores coeficientes de correlação e menores erros.

CAPÍTULO 7

CONCLUSÕES

Os objetivos propostos neste trabalho foram atingidos, sendo obtido um ganho nas previsões de radiação solar de curto prazo, para as estações SONDA de Florianópolis (FLN) e São Martinho da Serra (SMS), em comparação com aquelas disponibilizadas diretamente pelo modelo Eta operacional do CPTEC. Contudo foi verificado não ocorrer um ganho significativo em desempenho com o uso de Redes Neurais Artificiais (RNAs), em relação aos modelos ajustados por Regressão Linear Múltipla (RLM).

A comparação das previsões P00UT de radiação solar do modelo Eta (previsões mais recentes para cada dia) com as observações, para as estações FLN e SMS, revelou um viés elevado, de aproximadamente 30% (ME%), e um RMSE% superior a 40%, para as integrais diárias de energia solar de ambas as estações. Valores de erros médios percentuais semelhantes foram encontrados para o período diário Rad18UT (ME% entre 25 e 28% e RMSE% entre 40 e 42%), que engloba o intervalo de maior incidência de radiação, entre 12:00UT e 18:00UT. Os valores elevados de RMSE são em grande parte devido ao viés observado nas previsões.

Os modelos de previsão numérica de tempo comumente apresentam erros devido às não-linearidades dos processos atmosféricos, às aproximações empregadas, aos erros inerentes à análise e assimilação dos dados meteorológicos globais e à representação horizontal e vertical discretizada da atmosfera (pontos de grade representando grandes áreas ou volumes). Contudo, os diagramas de dispersão para as previsões de radiação solar do modelo Eta revelam que, em grande parte dos dias analisados, as previsões são bastante próximas às observações, porém deslocadas por um viés positivo. Assim, um desempenho razoável em termos de linearidade entre previsões e observações é verificado, com coeficientes de correlação superiores a 0,7 para FLN, e superiores a 0,8 para SMS, muito embora tenha sido mostrado que refinamentos levam a previsões com desempenho superior.

A aplicação de refinamentos simples na previsão de radiação solar do modelo Eta, empregando um simples ajuste linear entre previsões e observações (EtaA) ou simplesmente subtraindo o viés das previsões de radiação solar (EtaB), além de produzir previsões (P00UT-Rad18UT) com viés próximo a zero ($|ME\%|$ inferiores a 0,2%), promovem ganhos entre 21 e 24% em RMSE, verificado através do cálculo do *Skill* frente à previsão de radiação do modelo Eta (*ocis*). Os valores de RMSE% para estes refinamentos foram de aproximadamente 31 e 33%, para FLN e SMS, respectivamente. Por serem baseados unicamente nas previsões de radiação do modelo Eta, estes ajustes (EtaA e EtaB) não promovem uma elevação nos coeficientes de correlação, apresentando valores idênticos aos encontrados para a previsão Eta.

O uso, em modelos estatísticos (RNAs e RLMs), de outras variáveis previstas pelo modelo Eta, representando as condições da atmosfera e da superfície terrestre, juntamente com variáveis teoricamente calculadas relacionadas à radiação incidente no TDA, permitiram a obtenção de previsões de radiação com maiores coeficientes de correlação e menores valores de RMSE em relação a Eta, EtaA e EtaB, e valores de viés (ME) pequenos comparados às previsões do modelo Eta.

O ajuste de modelos RLM e RNA baseados não apenas nas previsões de radiação solar do modelo Eta, mas em todos os 36 preditores disponíveis (variáveis previstas pelo modelo Eta e variáveis calculadas), permitiu serem obtidas previsões (P00UT-Rad18UT) com viés próximo a zero ($|ME\%|$ inferiores a 2%) e valores de RMSE% de aproximadamente 26 e 28%, para FLN e SMS, respectivamente. Esses valores de RMSE correspondem a ganhos entre 34 e 36% sobre o modelo Eta, e entre 13 e 17% sobre EtaA e EtaB. Apesar de serem empregados muitos preditores, os modelos RLM e RNA revelaram ser possível promover melhorias na previsão de radiação, tomando informações diversas sobre as condições atmosféricas e de superfície fornecidas pelo modelo Eta para o instante a ser estimada a radiação solar incidente.

Testes com treinamentos de RNAs e ajustes de modelos RLM realizados com base em menor número de preditores, através da exclusão dos grupos Radiação, Fluxos e Chuva, revelaram ser possível construir modelos com menos entradas e baseados em preditores

mais simples, não pertencentes a estes grupos. Esses modelos não apresentaram um desempenho menor, apesar de não empregarem as previsões de radiação solar do modelo Eta. Conclusões semelhantes foram obtidas pelas análises de regressões múltiplas passo-a-passo. As regressões passo-a-passo indicaram ainda que o emprego de grupos com menos de 10 preditores parece ser suficiente para constituir modelos lineares com coeficientes de correlação aproximadamente iguais aos verificados para modelos baseados nos 36 preditores.

Experimentos realizados com exclusão e inserção de preditores, a partir de um conjunto arbitrário inicial de preditores constituído apenas por variáveis meteorológicas básicas (temperatura, umidade relativa, pressão e vento) e radiação solar no TDA, permitiram a obtenção de um conjunto mínimo de 8 preditores, composto apenas por algumas das variáveis iniciais e outras pertencentes ao grupo Nuvens (não foram utilizados os grupos Radiação, Fluxos e Chuva). Os modelos RLM e RNA ajustados com este conjunto reduzido de preditores levaram a previsões com desempenhos comparáveis aos observados para os modelos com 36 preditores. Embora desempenhos ligeiramente superiores tenham sido encontrados para cada estação individualmente, através da inserção/exclusão de determinados preditores, os 8 preditores mencionados, parecem constituir o grupo que leva ao melhor desempenho conjunto das duas estações.

Os modelos RNA e RLM ajustados no experimento que empregou os 8 preditores mencionados no parágrafo anterior, apresentaram previsões (P00UT-Rad18UT) com viés próximo a zero ($|ME\%$ inferiores a 1%) e valores de RMSE% de aproximadamente 27 e 28%, para FLN e SMS, respectivamente. Esses valores de RMSE correspondem a ganhos entre 32 e 36% sobre o modelo Eta, e entre 11 e 16% sobre EtaA e EtaB. Estes valores de erros e *Skill*, comparáveis aos encontrados para o experimento com 36 preditores, enfatizam a desnecessidade de empregar um conjunto extenso de preditores e de utilizar como entrada a própria previsão de radiação solar do modelo Eta.

Em termos de coeficientes de correlação, tanto os modelos RNA e RLM com 36 preditores, como os modelos com 8 preditores, apresentaram valores em torno de 0,80 para FLN e 0,85 para SMS. As previsões Eta, EtaA e EtaB, apresentaram coeficientes

de correlação de 0,72 e 0,78, para FLN e SMS, respectivamente. Assim, os novos modelos ajustados, empregando outras saídas previstas pelo modelo Eta, além de produzir ganhos em termos da redução dos erros médios, promovem um ganho em termos de fração da variância das previsões comum às observações.

Em todos os experimentos realizados, com grupos extensos e reduzidos de preditores, as RNAs não apresentaram ganhos significativos com relação aos modelos de RLM, sendo os desempenhos sempre muito aproximados. Como uma das habilidades mais importantes das RNAs é a capacidade de ajuste de não-linearidades, pode-se concluir que o problema em estudo mostrou-se bastante linear, de maneira a não ter sido observado um ganho das RNAs sobre modelos puramente lineares (RLMs).

Apesar de não observado um ganho pelo uso de RNAs neste trabalho, não se pode afirmar a inaptidão de seu uso na resolução de problemas de modelagem de radiação solar. O ajuste de RNAs capazes de simular não-linearidades e reconhecer padrões específicos é ligada à quantidade e qualidade dos exemplos apresentados. Assim, uma das possíveis razões pelo desempenho similar das RNAs em comparação com as RLMs, pode estar relacionado ao conjunto de dados relativamente curto, não abrangendo todas as possibilidades de situações que influenciam a incidência de radiação solar. Além disso, apesar de RNAs serem robustas frente a entradas com ruídos, a alimentação de exemplos totalmente errôneos, como por exemplo, previsões de condições de tempo ligadas a céu aberto quando foi observado de fato um dia de chuva, pode levar a RNA a um treinamento inadequado.

É importante enfatizar que os modelos RNA e RLM não são independentes e concorrentes do Eta, mas apenas refinamentos, ou *downscaling*, para previsões pontuais, alimentados pelas saídas do modelo Eta. As RNAs e RLMs simplesmente promovem ajustes entre a radiação solar no TDA, as variáveis atmosféricas e superficiais previstas (relacionadas à transmitância de radiação na atmosfera e incidência na superfície) e a radiação solar observada para o mesmo instante de tempo, cabendo o aspecto de previsão propriamente dita, ao modelo Eta. Apesar da redução dos erros médios observada com o emprego de RLMs e RNAs, grandes diferenças entre as previsões

desses modelos e as observações de radiação solar foram verificadas em determinados dias. Obviamente, estes modelos não levam a previsões de radiação corretas, se as condições de tempo previstas alimentadas diferirem das condições observadas.

A avaliação das previsões de radiação por estação do ano, para FLN, mostrou que os modelos apresentam desempenhos diferentes em cada estação, com coeficientes de correlação e erros médios assumindo valores maiores e menores do que aqueles encontrados ao se considerar dados do ano todo. Contudo, em todas as estações, os modelos RNA e RLM apresentaram desempenhos melhores do que a previsão de radiação solar do modelo Eta, e mesmo de EtaA e EtaB, com menores valores de RMSE e maiores valores de coeficientes de correlação. Para todos os modelos, os piores desempenhos em termos de coeficientes de correlação são observados no verão, o que pode ser atribuído ao fato de os modelos de previsão numérica de tempo apresentarem maiores dificuldades de simular as condições do tempo nesta estação, devido à maior ocorrência de processos convectivos.

Apesar do desempenho ligeiramente pior dos modelos RLM e RNA de 8 preditores ao considerar o ano todo, o desempenho desses modelos é ligeiramente superior àquele observado para os modelos com 36 preditores em algumas estações do ano. As diferenças variáveis em desempenho entre os modelos, ao longo das estações do ano, indicam a possibilidade de diferentes grupos de preditores e tipos de modelos levarem a melhores previsões de radiação solar em cada época do ano.

Apesar de os experimentos com grupos de preditores, as regressões múltiplas passo-a-passo e a seleção dos preditores terem sido baseados apenas nas previsões para o período diário Rad18UT, modelos RLM e RNA com 8 preditores promoveram redução de ME e RMSE e elevação de coeficiente de correlação para as previsões P00T-Rad12UT e P00UT-Rad24UT, em comparação às previsões de radiação solar do modelo Eta. As integrais diárias calculadas pelo somatório dos três períodos diários previstos com RLMs e RNAs promoveram um ganho de aproximadamente 40% em termos de RMSE, sobre as integrais diárias de radiação solar previstas pelo modelo Eta, bem como valores de ME próximos a zero e maiores coeficientes de correlação.

Possivelmente a utilização de outros grupos de preditores possa levar a desempenhos ainda melhores para os períodos diários Rad12UT e Rad24UT, bem como para a integral diária.

Como esperado, o aumento da antecedência das previsões de radiação solar leva a um crescimento dos erros e a uma redução dos coeficientes de correlação. Porém, para todas as antecedências analisadas (P00UT, P12UT, P24UT, P36UT, P48UT), os modelos RLM e RNA com 8 preditores apresentaram valores menores de ME% e RMSE%, e coeficientes de correlação mais elevados do que modelo Eta.

Estudos mais detalhados, empregando ajustes de RLMs e RNAs para outras estações solarimétricas que permitam uma maior representação do território brasileiro, bem como a identificação dos conjuntos de preditores mais importantes e característicos para cada região e para cada época do ano, poderão permitir a obtenção de uma metodologia de pós-processamento das previsões do modelo Eta que leve a previsões de radiação solar com erros menores que os atuais. Estudos envolvendo estações solarimétricas de latitudes menores e maior insolação, bem como o uso de maiores séries de dados, viriam a assegurar a qualidade dos métodos.

De modo geral, pode-se concluir que os modelos descritos nesse trabalho podem ser estendidos a locais ou regiões de interesse de concessionárias e distribuidoras de energia elétrica ou do Operador Nacional do Sistema Elétrico, de forma a oferecer previsões de radiação com bom grau de confiabilidade e de forma operacional. Diversas aplicações são possíveis para o setor de energia a partir dessa fonte de dados:

- 1) Previsão de despacho de carga em linhas de transmissão;
- 2) Especificação e estudo de durabilidade de transformadores abaixadores de tensão;
- 3) Estudos e desenvolvimento de projetos de geração elétrica fotovoltaica e de aquecimento solar de água;
- 4) Aplicação em projetos de geração híbrida fóssil-fotoelétrica de eletricidade;
- 5) Estudos de complementaridade entre regimes hídrico-eólico-solares para geração de energia elétrica renovável, entre outros.

CAPÍTULO 8

SUGESTÕES DE TRABALHOS FUTUROS

Ao longo da realização deste trabalho e após a análise dos resultados obtidos, observou-se a necessidade de estudos complementares que poderão levar a maiores aperfeiçoamentos das previsões de radiação solar e sua avaliação em outras localidades, além das estações FLN e SMS. Pode-se sugerir como trabalhos e estudos futuros:

- **Verificar a importância de outros preditores na obtenção de previsões para os períodos diários Rad12UT e Rad24UT.** Neste trabalho, os preditores foram selecionados com base nos dados para os períodos diários Rad18UT (previsões P00UT-Rad18UT), de maneira que podem existir outros preditores que estejam mais bem relacionados com a incidência de radiação nos demais períodos diários, quando a radiação solar atravessa a atmosfera com maior inclinação. Pode-se verificar se o uso de preditores diferentes em cada período diário proporciona ganhos na previsão de integrais diárias de energia solar.

- **Selecionar conjuntos de preditores e ajustar modelos RLM e RNA individuais para cada estação do ano.** Neste trabalho, os modelos foram ajustados empregando dados de todas as épocas do ano e com um conjunto fixo de preditores (escolhidos a partir de experimentos e análises empregando dados de todo ano), mas desempenhos variados foram observados para as previsões em cada estação do ano. Pode-se testar se resultados melhores podem ser obtidos através do ajuste de modelos e uso de grupos de preditores variados para cada estação do ano, ou para época seca e época chuvosa.

- **Realizar estudos de seleção de preditores e ajustes de modelos estatísticos baseados em previsões de tempo para outras localidades.** Neste trabalho, foram estudadas previsões de radiação pontuais para duas estações solarimétricas, situadas na região sul do Brasil. Planeja-se a aplicação da metodologia deste trabalho, para outras estações do projeto SONDA e de outros projetos, e utilizando séries de dados mais extensas. Assim, poderá ser estudado o desempenho das previsões de radiação do

modelo Eta e de modelos estatísticos em todas as regiões. Pode-se verificar a possibilidade de criação de uma metodologia de pós-processamento das previsões do modelo Eta, que possa levar a uma melhoria na previsão de radiação em todo o território. Pode-se testar treinamentos/ajustes de RNAs e RLMs generalizadas, utilizando conjuntamente os dados de várias estações.

- **Testar se RLMs e RNAs promovem ganho sobre as previsões horárias de radiação do modelo Eta.** Neste trabalho, foram empregados dados de radiação e de previsão de tempo representando períodos de 6 horas, mas dados de previsão de tempo do modelo Eta, com resolução temporal de 1 hora, são disponibilizados para determinados locais. Apesar de valores horários permitirem uma melhor descrição do comportamento da incidência de radiação ao longo do dia, quanto menores os intervalos de tempo considerados, maior é a possibilidade de a previsão diferir da situação real observada (maiores erros de previsão).

- **Treinar RNAs apenas com “exemplos corretos” e utilizando outras arquiteturas.** Neste trabalho foram empregadas RNAs do tipo perceptron de múltiplas camadas, treinadas a partir de um período aproximadamente contínuo de dados de previsões e observações. Além do emprego de outras arquiteturas de RNAs e tipos de treinamentos, critérios de acerto das previsões de tempo do modelo Eta poderiam ser definidos (empregando dados meteorológicos disponíveis para a estação), sendo utilizados no treinamento apenas dados de dias em que os erros das previsões das condições de tempo sejam menores. A eliminação dos dias em que as previsões pontuais estejam distantes das condições reais observadas, poderia permitir às RNAs representar melhor as relações entre preditores e preditando. O emprego de séries mais longas de dados, também pode vir a promover uma melhoria de aprendizado das RNAs, uma vez que o aprendizado depende da qualidade e quantidade de exemplos.

- **Avaliar as previsões de radiação solar de outros modelos de PNT do CPTEC e o uso de outras variáveis previstas em modelos estatísticos.** Neste trabalho, foi empregado o modelo Eta operacional com resolução horizontal de 40 km. Outros modelos, como o B-RAMS, o modelo Global, e o modelo Eta com resolução de 20 km,

poderiam ser empregados. Como os modelos apresentam habilidades variadas, poderiam ser utilizados preditores de diferentes modelos em modelos estatísticos de previsão de radiação ou combinar previsões em previsões por conjunto.

- **Promover modificações no código radiativo do modelo Eta.** Neste trabalho, são apresentados modelos estatísticos que, a partir de variáveis previstas diversas geradas pelo modelo Eta operacional, levam a uma previsão de radiação solar alternativa àquela fornecida por este modelo e com erros reduzidos. Contudo, a realização de estudos diretamente sobre as parametrizações e o código de cálculo de radiação do modelo Eta, podem conduzir a aperfeiçoamentos que resultem em uma melhoria na previsão de radiação fornecida diretamente pelo modelo.

- **Emprego de outros modelos de transferência radiativa, alimentados com dados de previsão de tempo.** O modelo BRASIL-SR, desenvolvido na Divisão de Clima e Meio Ambiente (DMA/CPTEC), possui como principal dado de entrada a fração de cobertura efetiva de nuvens, determinada a partir de imagens de satélite e que guarda a informação de espessura óptica atmosférica. Através da determinação de uma função empírica para aproximação da cobertura efetiva de nuvens a partir de variáveis de previsão de tempo, o modelo BRASIL-SR poderia ser empregado para realizar estimativas da incidência futura de radiação. Adicionalmente, dados de algumas variáveis de previsão de tempo poderiam ser empregados no modelo BRASIL-SR, substitutivamente aos dados climatológicos atualmente utilizados pelo modelo. Testes poderiam ser também realizados com o modelo GL desenvolvido na Divisão de Satélites e Sistemas Ambientais (DSA/CPTEC).

REFERÊNCIAS BIBLIOGRÁFICAS

- Al-Alawi, S. M.; Al-Hinai, H. A. An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. **Renewable Energy**, v. 14, n. 1-4, p. 199-204, 1998.
- Arakawa, A.; Lamb, V. R. Computational design of the basic dynamical processes of the UCLA general circulation model. **Methods in Computational Physics**, v. 17, p. 173-265, 1977.
- Bentley, R.W. Global oil and gas depletion: an overview. **Energy Policy**, v. 30, n. 3, p. 189-205, 2002.
- Black, T. L. The new NMC mesoscale Eta model: description and forecast examples. **Weather and Forecasting**, v. 9, n. 2, p. 265-278, Feb. 1994.
- Bolton, D. The computation of equivalent potential temperature. **Monthly Weather Review**, v. 108, n. 7, p. 1046-1053, 1980.
- Braga, A. P.; Carvalho A. C. P. F.; Ludermir, T. B. **Fundamentos de redes neurais artificiais**. Rio de Janeiro: 11^a Escola de Computação/Universidade Federal do Rio de Janeiro, 1998. 246p.
- Brasseur, G.; Solomon, S. **Aeronomy of the middle atmosphere**. Dordrecht, Holland: D. Reidel Publishing Company, 1986. 452p.
- Ceballos, J. C.; Bottino, M. J.; de Souza J. M. A simplified physical model for assessing solar radiation over Brazil using GOES 8 visible imagery. **J. Geophys. Res.**, v. 109, D02211, doi:10.1029/2003JD003531, 2004.
- Chou, S. C. Modelo Regional Eta. **Climanálise**, Edição comemorativa de 10 anos, 1996. Disponível em: <<http://www.cptec.inpe.br/products/climanalise/>>. Acesso em: 12 abr. 2005.
- Chou, S. C.; Tanajura C. A. S.; Xue, Y.; Nobre, C. A. Validation of the coupled Eta/SsiB model over South America. **J. Geophys. Res.**, v.107, n. D20, LBA 56, p.1-20, 2002.
- Cybenko, G. **Continuous valued neural networks with two hidden layers are sufficient**. Medford, MA: Department of Computer Science, Tufts University, 1988. Technical Report.
- Cybenko, G. Approximation by superpositions of a sigmoid function. **Mathematics of Control, Signals and Systems**, v. 2, p. 303-314, 1989.
- Colle, S.; Pereira, E. B. **Atlas de irradiação solar do Brasil** (primeira versão para irradiação global derivada de satélite e validada na superfície). Brasília: INMET, Outubro 1998.

- Coulson, K. L.. **Solar and terrestrial radiation**: methods and measurements, New York: Academic Press, 1975. 322p.
- Dave, J. V.; Canosa, Z. A direct solution of the radiative transfer equation: application to atmospheric models with arbitrary vertical non-homogeneities. **Journal of the Atmospheric Sciences**, v. 31, n. 4, p. 1089-1101, 1974.
- Dedieu, G.; Deschamps, P. Y.; Kerr, Y. H. Satellite estimation of solar irradiance at the surface of the earth and of surface albedo using a physical model applied to Meteosat data. **Journal of Climate and Applied Meteorology**, v. 26, n. 1, p. 79-87, 1987.
- Demuth, H.; Beale, M.; Hagan, M. **Neural network toolbox user's guide** – Version 4 – for use with MATLAB®. The MathWorks Inc. Disponível em: <http://www.mathworks.com/access/helpdesk/help/pdf_doc/nnet/nnet.pdf>. Acesso em: 15 ago. 2005.
- Diabaté, L.; Moussu G.; Wald. L. Description of an operational tool for determining global solar radiation at ground using geostationary satellite images. **Solar Energy**, v. 42, n. 3, p. 201-207, 1989.
- Elminir, H. K.; Areed, F. F.; Elsayed, T. S. Estimation of solar radiation components incident on Helwan site using neural networks. **Solar Energy**, v. 79, n.3, p. 270-279, 2005.
- Fels, S. B.; Schwarzkopf, M. D. The simplified exchange approximation: a new method for radiative transfer calculations. **Journal of the Atmospheric Sciences**, v. 32, n.7, p. 1475-1488, 1975.
- Gautier, C.; Diak, G.; Masse, S. A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. **Journal of Applied Meteorology**, v. 19, n. 8, p. 1005-1012, 1980.
- Geller, H. S. **Revolução energética**: políticas para um futuro sustentável. Rio de Janeiro: Relume Dumará, 2003. 299p.
- Goldenberg, J.; Villanueva, L. D. **Energia, meio ambiente & desenvolvimento**, 2. ed. São Paulo: Edusp – Editora da Universidade de São Paulo, 2003. 232p.
- Harrison, E. F.; Minnis, P.; Barkstrom, B. R.; Gibson, G. G. Radiation budget at the top of atmosphere. In: Gurney, R. J.; Foster, J. L.; Parkinson, C. L. (eds.). **Atlas of satellite observations related to global change**. San Diego: Cambridge University Press. Cap. 2, p.19-38, 1993.
- Hartmann, D. L. Radiative effects of clouds on Earth's climate. In: Hobbs, P. V. (ed.). **Aerosol-cloud-climate interactions**. San Diego: Academic Press. Cap. 6, p. 151-173, 1993.
- Hartmann, D.L. **Global physical climatology**. San Diego: Academic Press, 1994. 411p.

Hay, J. E.; Hanson, K. J. A satellite-based methodology for determining solar irradiance at the ocean surface during GATE. **Bulletin of the American Meteorological Society**, v. 59, p. 1549, 1978.

Haykin, S. **Neural networks: a comprehensive foundation**. New York: Macmillan College Publishing Company, 1994.

Haykin, S. **Redes neurais: princípios e prática**, 2. ed. Porto Alegre: Bookman Companhia Editora, 2001. 900p.

Hebb, D. O. **The organization of behavior**. New York: Wiley, 1949.

Hinkelman, L. M.; Ackerman, T. P.; Marchand, R. T. An evaluation of NCEP Eta model predictions of surface energy budget and cloud properties by comparison with measured ARM data. **J. Geophys. Res.**, v.104, n. D16, p. 19535-19549, 1999.

Hinton, G. E; Nowland, S. J. How learning can guide evolution. **Complex Systems**, v. 1, n. 1, p. 495-502, 1987.

Hontoria, L; Aguilera, J.; Zufiria P. Generation of hourly irradiation synthetic series using the neural network multilayer perceptron. **Solar Energy**, v. 72, n. 5, p. 441-446, 2002.

Hontoria, L.; Aguilera, J.; Zufiria P. An application of the multilayer perceptron: solar radiation maps in Spain. **Solar Energy**, v. 79, n. 5, p. 523-530, 2005.

Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences of the United States of America**, v. 79, n. 8, p. 2554-2558, 1982.

Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward network are universal approximators. **Neural Networks**, v. 2, n. 5, p. 359-366, 1989.

Houghton, H. G. **Physical meteorology**. Cambridge, MA: MIT, 1985. 442p.

Hoyt, D. V.; Schatten, K. H. **The role of the sun in climate change**. New York: Oxford University Press, 1997.

Hudson Jr., R. D. **Infrared system engineering**. New York: John Wiley & Sons Inc., 1969. 642p.

Instituto Nacional de Pesquisas Espaciais. Centro de Previsão de Tempo e Estudos Climáticos (INPE.CPTEC). **Previsões Numéricas – Modelo Eta**. Disponível em: <http://www.cptec.inpe.br/prevnum/exp_eta.shtml>. Acesso em: 10 abr. 2005.

International Energy Agency (IEA). **IEA Energy Statistics**. Disponível em: <<http://www.iea.org/>>. Acesso em: 15 jun. 2001.

Iqbal, M. **An introduction to solar radiation**. Toronto: Academic Press Canada, 1983. 390p.

- Iribarne, J. V.; Cho, H. R. **Atmospheric physics**. Dordrecht, Holland: Reidel Publishing, 1980. 212p.
- Janjić, Z. L. The step-mountain Eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. **Monthly Weather Review**, v. 122, n. 5, p. 927-945, 1994.
- Jolliffe, I. T.; Stephenson, D. B. (eds.). **Forecast Verification: a practitioner's guide in atmospheric science**. Chichester, England: John Wiley & Sons Inc., 2003. 240p.
- Kerschgens, M.; Pilz, U.; Raschke, E. A modified two-stream approximation for computations of the solar radiation budget in a cloudy atmosphere. **Tellus**, v. 30, p. 429-435, 1978.
- Kidder, S. Q.; Vonder Haar, T. H. **Satellite meteorology: an introduction**. San Diego: Academic Press, 1995. 466p.
- Kipp & Zonen, **CM-21 precision pyranometer: description, specification and manuals**. Disponível em: <<http://www.kippzonen.com/>>. Acesso em: 5 jan. 2006.
- Kondratyev, K. Ya. **Radiation in the atmosphere**. New York: Academic Press, 1969. 912p.
- Lacis, A. A.; Hansen, J. E. A parameterization of the absorption of solar radiation in earth's atmosphere. **Journal of the Atmospheric Sciences**, v. 31, n. 1, p. 118-133, 1974.
- Lenoble J. **Radiative transfer in scattering and absorbing atmospheres: standard computational procedures**. Hampton, Virginia, USA: A. DEEPAK, 1985. 300p.
- Lenoble, J. **Atmospheric radiative transfer**. Hampton, Virginia, USA: A. DEEPAK 1993. 532p.
- Liou, K. On the absorption, reflection and transmission of solar radiation in cloudy atmospheres. **Journal of the Atmospheric Sciences**, v. 33, n. 5, p. 798-805, 1976.
- Liou, K. N. **An introduction to atmospheric radiation**. New York: Academic Press Inc., 1980. 392p.
- Liou, K. N. **Radiation and cloud processes in the atmosphere: theory, observation and modeling**. New York: Oxford University Press, 1992.
- Lutgens, F. K.; Tarbuck E.J. **The atmosphere: an introduction to meteorology**, 6. ed. New Jersey: Prentice Hall Inc., 1995. 462p.
- Martins, F. R. **Influência do processo de determinação da cobertura de nuvens e dos aerossóis de queimada no modelo físico de radiação BRASIL-SR**. 2001. 330p. Tese (Doutorado em Geofísica Espacial) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos.

- McCulloch, W. S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115-133, 1943.
- Mesinger, F. A blocking technique for representation of mountains in atmospheric models. **Rivista di Meteorologia Aeronautica**, v. 44, n.1-4, p. 195-202, 1984.
- Mesinger, F.; Janjić, Z. L.; Ničković, S.; Deaven, D. G. The step-mountain coordinate: model description and performance for cases of alpine lee cyclogenesis and for a case of Appalachian redevelopment. **Mon. Weather Rev.**, v. 116, n. 7, p. 1493-1518, 1988.
- MetEd, Meteorology Education & Training – MetEd/UCAR/NCEP. **Eta introduction**. Disponível em: <<http://meted.ucar.edu/nwp/pcu2/etintro.htm>>. Acesso em: 10 abr. 2005.
- Mihalakakou, G.; Santamouris, M.; Asimakopoulos, D. N. The total solar radiation time series simulation in Athens, using neural networks. **Theor. and Appl. Climatol.**, v.66, n. 3-4, p. 185-197, 2000.
- Ministério de Minas e Energia (MME). **Balanco energético nacional 2004** (BEN-2004). Brasília. 2004. Disponível em: <http://www.mme.gov.br/site/menu/select_main_menu_item.do?channelId=1432>.
- Ministério de Minas e Energia (MME). **Programa luz para todos**. Brasília. Disponível em: <http://www.mme.gov.br/programs_display.do?prg=8>. Acesso em: 5 fev. 2006.
- Minsky, M.; Papert, S. **Perceptrons: an introduction to computational geometry**. Massachusetts: MIT Press, 1969.
- Mohandes, M.; Rehman, S.; Halawani, T. O. Estimation of global solar radiation using artificial neural networks. **Renewable Energy**, v. 14, n. 1-4, p. 179-184, 1998.
- National Aeronautics and Space Administration (NASA), **NASA World Wind software download**. Disponível em: <<http://worldwind.arc.nasa.gov/download.html>>. Acesso em: 18 nov. 2005.
- Ničković, S.; Mihailović, D.; Rajković, B. **Scientific documentation of the Eta model**. Unofficial documentation for WMO publication. Geneva, Switzerland: 1998. Disponível em: <<http://www.cptec.inpe.br/~eta/NewsLetter/docs/SciDoc/>>.
- Paltridge, G. W.; Platt, C. M. R. **Radiative processes in meteorology and climatology**. Amsterdam: Elsevier Scientific Publishing Company, 1976.
- Peixoto, J. P.; Oort, A. H. **Physics of climate**. New York: American Institute of Physics (AIP) Press, 1992. 520p.
- Pereira, E. B.; Abreu, S. L.; Stuhlmann, R.; Rieland, M.; Colle, S. Survey of the incident solar radiation in Brazil by the use of the METEOSAT satellite data. **Solar Energy**, v. 57, n. 2, p. 125-132, 1996.

Perez, R.; Seals, R.; Zelenka, A. Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data. **Solar Energy**, v. 60, n. 2, p. 89-96, 1997.

Perez, R. **Comunicação pessoal**. 2003.

Plana-Fattori, A.; Ceballos, J. C. **Glossário de termos técnicos em radiação atmosférica - versão 2.0**. IAG/USP – CPTEC/INPE. Disponível em: <<http://www.cptec.inpe.br/satelite/metsat/pesquisa/gloss1/glossar.htm>>. Acesso em: 17 mar. 2005.

Ramaswamy, V.; Freidenreich, S. M. A study of broadband parametrizations of the solar radiative interactions with water vapor and water drops. **Journal of Geophysical Research**, v.97, n. D11, p. 11487-11512, 1992.

Raschke, E. Multiple scattering calculations of the transfer of solar radiation in an atmosphere-ocean system. **Beitr. Phys. Atm.**, v. 45, p. 1-19, 1972.

Riedmiller, M.; Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: IEEE International Conference on Neural Networks (ICNN'93), 1993, San Francisco, CA, EUA. **Proceedings...** San Francisco: IEEE, p. 586-591, 1993.

Riedmiller, M. **Rprop** – description and implementation details. Karlsruhe, Germany: Universität of Karlsruhe, 1994. Technical Report.

Robinson, N. (ed.) **Solar radiation**. Amsterdam: Elsevier Publishing Company, 1966. 347p.

Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychol. Rev.**, v. 65, p. 386-408, 1958.

Rumelhart, D. E.; Hinton G. E.; Williams, R. J. Learning representations by back-propagation errors. **Nature**, v. 323, p. 533-536, doi:10.1038/323533a0, Out.1986.

Sfetsos, A.; Coonick, A. H. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. **Solar Energy**, v. 68, n. 2, p. 169-178, 2000.

Sistema Nacional de Organização Nacional de Dados Ambientais (SONDA). **Estações do projeto SONDA**, São José dos Campos, SP: INPE, 2005. Disponível em: <<http://www.cptec.inpe.br/sonda/>>. Acesso em: 19 dez. 2005.

Soares, J.; Oliveira, A. P.; Boznar, M. Z.; Mlakar, P.; Escobedo, J. F.; Machado, A. J. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique. **Applied Energy**, v. 79, n. 2, p.201-214, 2004.

Spencer, J. W. Fourier series representation of the position of the Sun. **Search**, v. 2, n. 5, p. 172, 1971.

- Stephens, G. L. Radiation profiles in extended water clouds. I: theory. **Journal of the Atmospheric Sciences**, v. 35, n. 11, p. 2111-2122, 1978.
- Stuhlmann, R.; Rieland, M.; Raschke, E. An improvement of the IGMK model to derive total and diffuse solar radiation at the surface from satellite data. **Journal of Applied Meteorology**, v. 29, n. 7, p. 586-603, 1990.
- Tarpley, J. D. Estimating incident solar radiation at the Earth's surface from geostationary satellite data. **Journal of Applied Meteorology**, v. 18, n. 9, p. 1172-1181, 1979.
- Tiba, C. (ed.) **Atlas Solarimétrico do Brasil: banco de dados terrestres**. Recife: Editora Universitária da UFPE, 2000. 111p.
- Tymvios F. S.; Jacovides, C. P.; Michaelides, S. C.; Scouteli, C. Comparative study of Angström's and artificial neural networks' methodologies in estimating global solar radiation. **Solar Energy**, v. 78, n. 6, p. 752-762, 2005.
- van de Hulst, H. C. **Light scattering by small particles**. New York: Wiley, 1957. 470p.
- Wallace, J. M.; Hobbs, P. V. **Atmospheric science: an introductory survey**. San Diego: Academic Press, 1977. 467p.
- Wasserman, P. D. **Advanced methods in neural computing**. New York: Van Nostrand Reinhold, 1989.
- Welch, R. M.; Cox, S. K.; Davis, J. M. **Solar radiation and clouds**, v. 17, n. 39, Boston, MA: American Meteorological Society, 1980. 96p. Meteorological Monograph Series.
- Whitlock, C. H.; Charlock, T. P.; Staylor, W. F.; Pinker, R. T.; Laszlo, I.; Ohmura, A.; Gilgen, H.; Konzelman, T.; Di Pasquale, R. C.; Moats, C. D.; LeCroy, S.R.; Ritchey, N. A. First global WCRP shortwave surface radiation budget dataset. **Bull. Amer. Meteor. Soc.**, v.76, n. 6, p. 905-922,1995.
- Widrow, B.; Hoff, M. E. Adaptative switching circuits. In: Western Electronic Show and Convention (WESCON), 1960, Los Angeles, CA, EUA. **Proceedings of IRE WESCON Convention Record**. Los Angeles: Institute of Radio Engineers - IRE, v. 4, p. 96-104, 1960.
- Wikipedia: the free Encyclopedia. **Kyoto protocol**. Disponível em: <http://en.wikipedia.org/wiki/Kyoto_Protocol>. Acesso em: fev. 2006.
- WWRP/WGNE Joint Working Group on Verification. **Forecast verification - issues, methods and FAQ**. Disponível em: <http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html>. Acesso em: 23 nov. 2005.