



Ministério da
**Ciência, Tecnologia
e Inovação**



sid.inpe.br/mtc-m19/2013/04.17.13.55-TDI

ESTIMATIVA DO PERFIL DA CONCENTRAÇÃO DE CLOROFILA EM ÁGUAS NATURAIS COM O USO DE REDES NEURAIS ARTIFICIAIS

Fábio Dall Cortivo

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelo Dr. Ezzat Selim Chalhoub, aprovada em 26 de março de 2013.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3DTPM2H>>

INPE
São José dos Campos
2013

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):

Presidente:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

Luciana Manacero - Serviço de Informação e Documentação (SID)



Ministério da
**Ciência, Tecnologia
e Inovação**



sid.inpe.br/mtc-m19/2013/04.17.13.55-TDI

ESTIMATIVA DO PERFIL DA CONCENTRAÇÃO DE CLOROFILA EM ÁGUAS NATURAIS COM O USO DE REDES NEURAIS ARTIFICIAIS

Fábio Dall Cortivo

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelo Dr. Ezzat Selim Chalhoub, aprovada em 26 de março de 2013.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3DTPM2H>>

INPE
São José dos Campos
2013

Dados Internacionais de Catalogação na Publicação (CIP)

Dall Cortivo, Fabio.
C818e Estimativa do perfil da concentração de clorofila em águas naturais com o uso de redes neurais artificiais / Fábio Dall Cortivo. – São José dos Campos : INPE, 2013.
 xxviii + 188 p. ; (sid.inpe.br/mtc-m19/2013/04.17.13.55-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013.

Orientador : Dr. Ezzat Selim Chalhoub.

1. ótica hidrológica. 2. cor do oceano. 3. perfil da concentração de clorofila. 4. equação de transferência radiativa. 5. perceptron de múltiplas camadas. 6. método quasi-Newton. I.Título.

CDU 004.8:551.46.09

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Doutor(a)** em
Computação Aplicada

Dr. Haroldo Fraga de Campos Velho



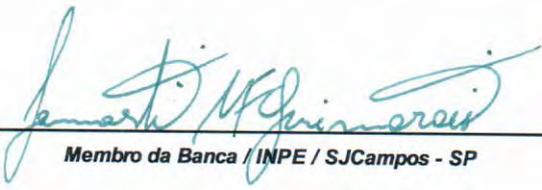
Presidente / INPE / São José dos Campos - SP

Dr. Ezzat Selim Chalhoub



Orientador(a) / INPE / SJCampos - SP

Dr. Lamartine Nogueira Frutuoso
Guimarães



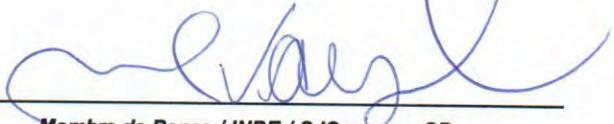
Membro da Banca / INPE / SJCampos - SP

Dr. Stephan Stephany



Membro da Banca / INPE / SJCampos - SP

Dr. Milton Kampel



Membro da Banca / INPE / SJCampos - SP

Dr. Carlos Alberto Eiras Garcia



Convidado(a) / FURG / Rio Grande do Sul - RS

Dr. José Alfredo Ferreira Costa



Convidado(a) / UFRN / Rio Grande do Norte - RN

Este trabalho foi aprovado por:

maioria simples

unanimidade

Aluno (a): **Fábio Dall Cortivo**

“Os problemas significativos que enfrentamos não podem ser resolvidos no mesmo nível de pensamento em que estávamos quando os criamos.”

ALBERT EINSTEIN

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.”

MARTIN LUTHER KING

A minha família e à minha querida Livia

AGRADECIMENTOS

Primeiramente à Deus, por permitir que eu exista.

Aos meus pais, Algemiro e Nilce, meu eterno muito obrigado por toda a vida de sacrifícios que tiveram, a fim de me proporcionar todas as condições para que eu pudesse chegar até aqui, e à minha irmã Cristiane, pelo incentivo, apoio e correções ortográficas sugeridas.

Ao Dr. Mario Rocha Retamoso, pela confiança depositada durante a graduação e pelo incentivo para que eu seguisse os estudos após a graduação, meu muito obrigado.

Ao meu orientador Dr. Ezzat Selim Chalhoub, pelas conversas, pela orientação, pelo conhecimento, pelo(s) susto(s) e, principalmente, pela confiança depositada em mim, além de tudo, um exemplo de dedicação e sabedoria a ser seguido, meu muito obrigado.

A minha querida Lívia, por estar ao meu lado e, principalmente, pela paciência nos momentos finais da elaboração deste trabalho.

Ao grande amigo que fiz logo que cheguei, Jonas Bianchini Fulindi, sendo companhia para muitas conversas. Que nossa amizade seja duradoura!

Aos professores e pesquisadores dos Cursos de Pos-Graduação em Computação Aplicada (CAP) pelo conhecimento transmitido.

Aos colegas do curso e aos demais amigos que fiz.

Não posso deixar de agradecer às oportunidades que tive em participar do Conselho de Curso da CAP (04/2010 à 05/2011) e do Conselho dos Cursos de Pos-Graduação (09/2010 à 03/2012), e pelo convívio e aprendizagem que tive durante esses períodos.

À CAPES pela bolsa de estudos, pois sem ela não teria tido condições de realizar este trabalho.

E, por fim, não poderia deixar de agradecer a todos aqueles que não acreditaram na minha vitória, pois isso me ajudou a buscar força nos momentos difíceis.

RESUMO

A estimativa da concentração de clorofila, em águas naturais, a partir da radiação emergente do oceano, medida pelos sensores a bordo dos satélites de sensoriamento remoto da cor da água, possibilita a quantificação da biomassa fitoplanctônica. O fitoplâncton é responsável pela produção primária, *i.e.*, a fixação do carbono pelo processo da fotossíntese, e é a base da cadeia alimentar dos oceanos, regulando a produtividade biológica marinha, e é uma das principais fontes geradoras de oxigênio para o planeta. Os algoritmos que convertem a radiação emergente dos oceanos em medidas de concentração de clorofila fornecem uma estimativa para essa concentração para a camada superficial. Apesar de os valores estimados por estes algoritmos estarem muito próximos dos valores reais, principalmente em águas do Caso 1 (oceano aberto), sabe-se que estes algoritmos são incapazes de inferir sobre a concentração da clorofila abaixo da superfície da água, ou então, sobre o perfil dessa concentração ao longo da profundidade. Em virtude das deficiências apresentadas por esses algoritmos, esta tese propõe uma metodologia capaz de fazer a estimativa desse perfil da concentração de clorofila. A nossa metodologia é baseada em Redes Neurais Artificiais do tipo Perceptron de Múltiplas Camadas, treinadas com o algoritmo quasi-Newton e implementadas em linguagem FORTRAN 90. Os padrões para o treinamento, teste e validação da rede são obtidos a partir da resolução da Equação de Transferência Radiativa (dados sintéticos), em que, para a modelagem dos coeficientes de absorção e de espalhamento, foram utilizados os modelos bio-ópticos, com os perfis de concentração de clorofila seguindo um comportamento gaussiano. A solução numérica da Equação de Transferência Radiativa é obtida através do código PEENSA, o qual implementa o método de Ordenadas Discretas Analítico (AS_N). Na análise efetuada, foram considerados 14720 perfis de concentração de clorofila, que foram gerados a partir de dois valores atribuídos à quantidade de biomassa no sistema, e a partir de dois conjuntos com diferentes comprimentos de onda da radiação e de três conjuntos contendo as direções nas quais a radiação emerge da superfície da água. A utilização de dados sintéticos faz com que seja necessária a adição de ruído aos padrões de treinamento, teste e validação. Desse modo, foram utilizados três níveis de ruído no treinamento e teste, e um nível de ruído (não presente no treinamento e no teste) no processo de validação da rede. O critério de parada do treinamento se deu pelo processo de validação cruzada utilizando o conjunto de teste. Os resultados obtidos mostraram que a metodologia sugerida é promissora, apresentando resultados entre 60% e 70% (taxa de acerto da rede) para padrões de entrada corrompidos com 3% de ruído gaussiano aditivo. Além disso, a demanda de tempo computacional necessário para o treinamento das redes é relativamente pequeno.

ESTIMATION OF THE PROFILE OF THE CHLOROPHYLL CONCENTRATION IN NATURAL WATERS USING AN ARTIFICIAL NEURAL NETWORK

ABSTRACT

The estimation of the chlorophyll concentration, in natural waters, from upwelled radiation at the surface of the ocean, measured by orbital sensors are used for the quantification of the phytoplanktonic biomass. Phytoplankton that is responsible for the primary production, *i.e.*, the carbon fixation by the photosynthesis process, is the basis of the chain food of the oceans, that controls the marine biologic productivity, and is one of the most important source of oxygen for the planet. The algorithms that convert the radiation from the oceans into chlorophyll concentrations produces an estimation of this concentrations only at the surface of the water. Although the estimated values by those algorithms are very close to the real ones, mainly for Case 1 waters (open ocean), they cannot infer about the chlorophyll concentration below the surface or even its profile along the depth of the water. To circumvent these difficulties, this thesis proposes a methodology that is capable of making the estimation of the profile of the chlorophyll concentration. Our methodology is based on Artificial Neural Networks of Multilayer Perceptron type, trained with the quasi-Newton algorithm, and implemented in FORTRAN 90. The patterns for the training, testing and validation are obtained from the resolution of the Radiative Transfer Equation (synthetic data), where the absorption and scattering coefficients are represented by bio-optical models, with the profile of the chlorophyll concentrations following a gaussian model. The numerical solution of the Radiative Transfer Equation is performed by PEESNA code, which implements the Analytical Discrete-Ordinates (AS_N) method. In the performed analysis, we used 14720 profiles of chlorophyll concentration, that were generated by attributing two values to the biomass quantity in the system, and by considering two sets of different radiation wavelength values and three sets containing the directions, with which the radiation is emitted at the surface of the water. When using synthetic data, it becomes necessary to apply noise for the corruption of the patterns for the training, testing and validation. So, we used three noise levels for the training and the testing, and one more noise level (not present in the training and the testing) to perform the validation of the network. The stopping criteria to end the training process, using the test set, was given by the cross-validation method. The results obtained show that the methodology proposed is quite promising, presenting results between 60% and 70% (hits rate) for the input patterns corrupted with 3% of gaussian noise level. Besides that, the computational time spent to train the network is relatively small.

LISTA DE FIGURAS

	<u>Pág.</u>
2.1	Processo de interação da radiação eletromagnética com águas naturais. 16
2.2	Representação gráfica de um problema de multiregiões. 17
3.1	Estrutura de um neurônio presente no encéfalo humano. 23
3.2	Modelo do neurônio artificial de McCulloch e Pitts. 24
3.3	Representação gráfica de possíveis funções de ativação que podem ser utilizadas no neurônio artificial. 26
3.4	Conjunto de padrões linearmente separáveis e não separáveis. 28
3.5	Representação gráfica de um perceptron de múltiplas camadas. 28
3.6	Diagrama de blocos a aprendizagem supervisionada. 31
3.7	Ilustração da generalização de uma rede e o comportamento do erro quadrático médio no procedimento de validação cruzada. 32
3.8	Diagrama representando problema direto e problema inverso. 41
3.9	Organograma contendo as possíveis alternativas para buscar a solução de problemas inversos (blocos pretos), adicionada a possibilidade da utilização das RNAs como metodologia de resolução e as alternativas para o treinamento (blocos vermelhos). 43
4.1	Coefficientes de absorção e de espalhamento da água pura e coeficiente de absorção específico do fitoplâncton. 48
4.2	Perfil da concentração de clorofila para o Mar Céltico no mês de maio, segundo Platt e Sathyendranath (1988). 49
4.3	Representação dos valores médios adotados em cada região espacial. 50
4.4	Concentração de clorofila de superfície considerando a composição anual de 2009. Sensor SeaWiFS. 52
4.5	Esquema gráfico da combinação dos valores de τ_m com os valores de σ 52
4.6	Perfis de concentração de clorofila gerados a partir da combinação de três valores de pico com os valores de desvio padrão adotados. 53
4.7	Penetração da radiação solar em águas de oceano aberto para os comprimentos de onda da faixa do visível. 54
4.8	Curvas do albedo de espalhamento simples obtidas considerando o perfil mostrando na Figura 4.2 para diferentes comprimentos de onda compreendidos no intervalo $\lambda \in [412, 678]$ nm. 55

4.9	Curvas do albedo de espalhamento simples obtidas considerando o perfil mostrando na Figura 4.2 para diferentes comprimentos de onda compreendidos no intervalo $\lambda \in [500, 610]$ nm.	55
4.10	Curvas da radiação emergente na superfície obtidas pela simulação da ETR, levando em conta o perfil mostrado na Figura 4.2, para diferentes comprimentos de onda compreendidos no intervalo $\lambda \in [500, 610]$ nm. . .	56
4.11	Variação da radiação emergente na superfície, em $\theta = 0^\circ$ e $\theta = 60^\circ$, para diferentes λ para o perfil de concentração de clorofila mostrado na Figura 4.2.	57
4.12	Radiação emergente na superfície, calculada pelo modelo adotado, para os comprimentos de onda definidos na Eq. (4.8) para dois perfis de concentração de clorofila.	58
4.13	Diagrama ilustrando os problemas considerados em cada um dos grupos.	58
4.14	Diagrama simplificado dos passos a serem seguidos para a resolução do problema direto.	61
4.15	Representação esquemática do problema direto e do problema inverso. . .	62
4.16	Representação dos padrões de entrada e de saída para cada uma das redes.	63
4.17	Perfis de concentração de clorofila em $\tau_m = 5$ m e $\tau_m = 35$ m para $\sigma = 5$ m, 9 m e 17 m.	64
4.18	Curvas do albedo de espalhamento simples geradas a partir dos perfis mostrados nas Figuras 4.17(a) e 4.17(b).	65
4.19	Radiação emergente medida nas direções definidas na Eq. (4.10) levando em conta os perfis mostrados nas Figuras 4.17(a) e 4.17(b).	65
4.20	Radiação emergente exata gerada a partir dos perfis mostrados nas Figuras 4.17(a) e 4.17(b), e a radiância emergente corrompida com 5% de ruído.	66
4.21	Problema de região única com perfil constante.	67
4.22	Diagrama ilustrando a combinação dos problemas multiregiões e região única.	67
4.23	Diagrama da atuação conjunta das redes de classificação e das redes que determinam os valores do desvio padrão e da profundidade do pico da concentração de clorofila.	68
4.24	Interface gráfica para a configuração da RNA implementada.	72
4.25	Diagrama em blocos do funcionamento da RNA.	73
5.1	Estimativas obtidas pelas redes de filtragem quando apresentados os padrões de entrada obtidos a partir dos problemas $P_{1h_{144}}$ e $P_{1h_{30}}$ com uma única região espacial e corrompidos com o nível de ruído ε_3	80

5.2	Combinações de valores de profundidade do pico com valores de desvio padrão que foram classificados para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.	82
5.3	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{Aqua}_{h_{144}}$, e os respectivos erros dessas estimativas.	90
5.4	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{P}_{1h_{144}}$, e os respectivos erros dessas estimativas.	90
5.5	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{P}_{2h_{144}}$, e os respectivos erros dessas estimativas.	91
5.6	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{P}_{7Dh_{144}}$, e os respectivos erros dessas estimativas.	91
5.7	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{Aqua}_{h_{30}}$, e os respectivos erros dessas estimativas.	91
5.8	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{P}_{1h_{30}}$, e os respectivos erros dessas estimativas.	92
5.9	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{P}_{2h_{30}}$, e os respectivos erros dessas estimativas.	92
5.10	Estimativa do desvio padrão e da profundidade do pico, para o problema $\text{P}_{7Dh_{30}}$, e os respectivos erros dessas estimativas.	92
5.11	Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas Aqua_{h_*} .	93
5.12	Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas P_{1h_*} .	93
5.13	Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas P_{2h_*} .	93
5.14	Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas P_{7Dh_*} .	94
5.15	Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas Aqua_{h_*} .	94
5.16	Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas P_{1h_*} .	94
5.17	Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas P_{2h_*} .	95
5.18	Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas P_{7Dh_*} .	95

5.19	Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema Aqua _{h*} .	98
5.20	Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P _{1h*} .	99
5.21	Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P _{2h*} .	100
5.22	Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P _{7Dh*} .	101
5.23	Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema Aqua _{h*} .	102
5.24	Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P _{1h*} .	103
5.25	Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P _{2h*} .	104
5.26	Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problemas P _{7Dh*} .	105
5.27	Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para o problemas Aqua _{h*} .	107
5.28	Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para o problemas P _{1h*} .	108
5.29	Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para o problemas P _{2h*} .	109
5.30	Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para os problemas P _{7Dh*} .	110
5.31	Perfis da concentração de clorofila estimados a partir das radiâncias, obtidas levando em conta o perfil mostrado na Figura 4.2, corrompidas com 1% e 2% de ruído.	112
5.32	Perfis obtidos com dados (radiâncias) corrompidas com 1% de ruído.	113
5.33	Perfis obtidos com dados (radiâncias) corrompidas com 2% de ruído.	113
5.34	Perfis recuperados por Souto et al. (2007).	114
5.35	Perfis da concentração de clorofila estimados a partir das radiâncias, obtidas levando em conta o perfil mostrado na Figura 4.2, corrompidas com 3% e 5% de ruído.	114
A.1	Problema com uma única região.	146

C.1	Anatomia fisiológica da sinapse.	156
C.2	Três estados de um neurônio.	158
C.3	Potenciais pós-sinápticos excitatórios.	159
C.4	Respostas características de diferentes tipos de neurônios a diferentes níveis do estado excitatório.	160
D.1	Ilustração do cálculo dos gradientes para variáveis da camada oculta. . .	174
E.1	Redes utilizadas na Metodologia A: Rede Única.	179
E.2	Treinamento do comitê utilizado na Metodologia B.	180
E.3	Estrutura do comitê utilizado na Metodologia B: Comitê de Redes. . . .	180
E.4	Radiação emergente calculada pelo modelo adotado para os problemas $P_{7Dh_{144}}$ e $P_{1h_{144}}$ para $\tau_{m_1} = 1$ m e $\tau_{m_2} = 2$ m.	181
E.5	Representação das redes utilizadas na Metodologia C: Redes Alternadas. . .	182
E.6	Representação do procedimento de validação das redes utilizadas na Metodologia C: Redes Alternadas.	182

LISTA DE TABELAS

	<u>Pág.</u>
1.1 Propriedades óticas estimadas e metodologias inversas adotadas.	5
4.1 Valores adotados para os parâmetros I_0 , h , chl_0 , θ_0 , φ_0 , ζ , R e ρ	51
4.2 Tempo médio gasto pelo código PEESNA, em segundos, para a resolução de cada problema direto.	60
5.1 Número de redes utilizadas para cada problema de cada classe, total de perfis para multiregiões e região única, e o valor de concentração de clorofila de corte adotado para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$	75
5.2 Valores mínimos e máximos utilizados para a normalização da radiação emergente para os problemas multiregiões.	77
5.3 Níveis de ruído adotados para os conjuntos de treinamento, teste e validação.	78
5.4 Percentagens de padrões de cada conjunto em cada nível de ruído para o treinamento das redes responsáveis pela filtragem dos problemas.	79
5.5 Número de entradas e de neurônios ocultos para cada rede de filtragem.	79
5.6 Quantidade de perfis selecionados e não selecionados pelas redes de filtragem, quando aplicados os padrões pertencentes aos problemas do grupo $\mathbf{G}_{h_{144}}$, que consideram 20 regiões espaciais e corrompidos com o ruído ε_3	81
5.7 Quantidade de perfis selecionados e não selecionados pelas redes de filtragem, quando aplicados os padrões pertencentes aos problemas do grupo $\mathbf{G}_{h_{30}}$, que consideram 20 regiões espaciais e corrompidos com o ruído ε_3	81
5.8 Total de padrões para os conjuntos de treinamento, teste e validação para as redes RNA_σ e RNA_{τ_m}	83
5.9 Total de padrões nos conjuntos de treinamento, teste e validação, após as réplicas, para as redes RNA_σ e RNA_{τ_m}	83
5.10 Resultados obtidos pelas redes sobre o conjunto de validação para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$	85
5.11 Melhores estimativas para os valores do desvio padrão e os respectivos valores da profundidade do pico da concentração de clorofila para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$	96
5.12 Melhores estimativas para os valores da profundidade do pico da concentração de clorofila e os respectivos valores do desvio padrão para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$	96

5.13	Piores estimativas para os valores do desvio padrão e os respectivos valores da profundidade do pico da concentração de clorofila para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$	97
5.14	Piores estimativas para os valores da profundidade do pico da concentração de clorofila e os respectivos valores do desvio padrão para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$	97
5.15	Piores estimativas gerais, sobre o conjunto de validação, para os valores do desvio padrão e os respectivos valores da profundidade do pico da concentração de clorofila, para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$. . .	106
5.16	Piores estimativas gerais, sobre o conjunto de validação, para os valores da profundidade do pico da concentração de clorofila e os respectivos valores do desvio padrão, para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$. . .	106
5.17	Estimativas do desvio padrão a partir da radiância emergente, obtida ao se considerar o perfil mostrado na Figura 4.2, corrompida com quatro níveis de ruído.	115
5.18	Estimativas da profundidade do pico a partir da radiância emergente, obtida ao se considerar o perfil mostrado na Figura 4.2, corrompida com quatro níveis de ruído.	115

LISTA DE SÍMBOLOS

a	– coeficiente de absorção
a_λ^c	– coeficiente de absorção específico da clorofila
a_λ^w	– coeficiente de absorção da água pura
\mathbb{B}	– matriz/vetor que contém todos os níveis de <i>bias</i>
b	– coeficiente de espalhamento ou então nível de <i>bias</i>
b_λ^w	– coeficiente de espalhamento da água pura
c	– coeficiente de atenuação do feixe
chl	– concentração de clorofila
chl ₀	– concentração de clorofila de fundo
chl ₀ ^S	– concentração de clorofila de superfície
chl _r	– concentração de clorofila média para cada região
d_{kp}	– padrão de saída do conjunto de treinamento, teste ou validação
\mathbb{E}_t	– matriz que contém todos os padrões de entrada para o treinamento
e_σ	– erro cometido nas estimativas dos valores do desvio padrão
e_τ	– erro cometido nas estimativas da profundidade do pico
G^\pm	– condições de contorno
\mathbb{H}	– matriz Hessiana
h	– quantidade total de biomassa
I	– intensidade de radiação (radiância)
I_0	– intensidade do feixe incidente
I_s	– intensidade de radiação espalhada
I_u	– intensidade de radiação não-espalhada
$\mathcal{J}(\cdot)$	– funcional de diferenças quadráticas
$\nabla \mathcal{J}$	– gradiente do funcional de diferenças quadráticas
N_P	– número total de padrões no conjunto de teste
N_p	– número total de padrões de treinamento
N_r	– número total de níveis de ruído
N_s	– número total de neurônios na camada de saída
P_l	– polinômios de Legendre
P_l^m	– funções associadas de Legendre
$p(\cos \Theta)$	– função de fase
R	– número de regiões
\mathcal{R}	– conjunto dos números reais
\mathcal{R}^*	– conjunto dos números reais excluindo o zero
\mathcal{R}_+^*	– conjunto dos números reais positivos excluindo o zero
S_0	– fonte internas de radiação
v	– campo local induzido
\mathbb{W}	– matriz que contém todos os pesos sinápticos
w	– pesos sinápticos
\vec{X}	– vetor que contém todas as variáveis livres da rede

y_k	–	sinal de saída de um neurônio k
z	–	profundidade geométrica
α	–	constante de momento
$\beta_{l,r}$	–	coeficientes dos polinômios de Legendre para cada região
Γ	–	matriz contendo todos os parâmetros das funções de ativação
γ	–	parâmetro das funções de ativação
δ_j^n	–	gradiente local de um neurônio j com relação a w (ou b) na n -ésima iteração
ε	–	níveis de ruído adotados
ζ	–	profundidade ótica
η	–	taxa de aprendizagem ou pesos de quadratura de Gauss-Legendre
Θ	–	ângulo formado entre a direção incidente e a direção espalhada
θ	–	ângulo polar
θ_0	–	ângulo polar de incidência do feixe de fótons
λ	–	comprimento de onda
λ_g	–	comprimento de onda médio da banda selecionada
μ	–	co-seno do ângulo polar $\mu = \cos \theta$ ou nós de quadratura de Gauss-Legendre
ρ	–	coeficiente de assimetria da função de fase
ϱ_j^n	–	gradiente local de um neurônio j com relação a γ na n -ésima iteração
ϖ	–	albedo de espalhamento simples
ϖ_r	–	albedo de espalhamento simples para cada região
σ	–	desvio padrão da concentração de clorofila
σ^*	–	desvio padrão da concentração de clorofila estimado pela rede
τ	–	variável ótica
τ_m	–	profundidade do pico da concentração de clorofila
τ_m^*	–	profundidade do pico da concentração de clorofila estimado pela rede
$\bar{\tau}_{m,r}$	–	valor médio de τ para cada região
Υ	–	limiar de classificação
ϕ	–	função de ativação
φ	–	ângulo azimutal
φ_0	–	ângulo azimutal de incidência do feixe de fótons

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
1.1 Objetivos e Metodologia	9
1.2 Organização do Trabalho	11
2 EQUAÇÃO DE TRANSPORTE DE PARTÍCULAS	13
2.1 O Problema da Ótica e o Modelo Matemático	15
3 REDES NEURAIS ARTIFICIAIS	21
3.1 Neurônio Biológico	22
3.2 Neurônio Artificial	23
3.2.1 Funções de Ativação	25
3.3 Características das RNAs	26
3.4 Perceptron de Múltiplas Camadas	27
3.5 Aprendizagem das Redes Neurais Artificiais	29
3.5.1 Aprendizagem Supervisionada	30
3.5.2 Validação Cruzada	31
3.6 Algoritmos de Aprendizagem	32
3.6.1 Algoritmo <i>Backpropagation</i> Tradicional	34
3.6.2 Algoritmo <i>Backpropagation</i> Generalizado	36
3.6.3 Algoritmo quasi-Newton	37
3.7 Redes Neurais e Problemas Inversos	40
4 PROBLEMAS ABORDADOS	47
4.1 Detalhes Gerais dos Problemas	47
4.1.1 Detalhes Específicos para os Problemas Abordados	50
4.1.2 Padrões de Entrada e de Saída	61
4.2 Redes Classificadoras	64
4.3 Metodologia de Treinamento	68
4.4 Rede Neural Implementada	71
5 RESULTADOS E DISCUSSÕES	75
5.1 Treinamento das Redes	78
5.1.1 Treinamento das Redes de Filtragem	78

5.1.2	Treinamento das Redes RNA_σ e RNA_{τ_m}	81
5.2	Resultados Obtidos	84
5.3	Discussão dos Resultados	87
6	CONCLUSÕES	117
	REFERÊNCIAS BIBLIOGRÁFICAS	121
	APÊNDICE A - APROXIMAÇÃO DA EQUAÇÃO DE TRANSFERÊNCIA RADIATIVA E O MÉTODO AS_N	141
A.1	Aproximação da Equação de Transferência Radiativa	141
A.2	Método das Ordenadas Discretas Analítico – AS_N	143
	APÊNDICE B - TEOREMA DA ADIÇÃO DOS HARMÔNICOS ESFÉRICOS E A APROXIMAÇÃO DA FUNÇÃO DE FASE	151
B.1	Teorema da Adição dos Harmônicos Esféricos	151
B.2	Aproximação da Função de Fase de Henyey-Greenstein	151
	APÊNDICE C - O NEURÔNIO BIOLÓGICO	155
	APÊNDICE D - FORMALISMOS MATEMÁTICOS DOS ALGORITMOS <i>BACKPROPAGATION</i>, <i>BACKPROPAGATION</i> GENERALIZADO E QUASI-NEWTON	161
D.1	Algoritmo <i>Backpropagation</i> Tradicional	161
D.1.1	O Neurônio está na Camada de Saída	161
D.1.2	O Neurônio está em uma Camada Oculta	163
D.2	Algoritmo <i>Backpropagation</i> Generalizado	166
D.2.1	O Neurônio está na Camada de Saída	166
D.2.2	O Neurônio está em uma Camada Oculta	168
D.3	Cálculo dos Gradientes para o Método quasi-Newton	170
D.3.1	O Neurônio está na Camada de Saída	172
D.3.2	O Neurônio está na Camada Oculta	173
D.4	Derivadas das Funções de Ativação Seleccionadas	176
D.4.1	Função Sigmoides	176
D.4.2	Função Tangente Hiperbólica	177
D.4.3	Função Logaritmo Natural	177
	APÊNDICE E - METODOLOGIAS QUE NÃO PRODUZIRAM RESULTADOS SATISFATÓRIOS	179
E.1	Metodologia A: Rede Única	179

E.2	Metodologia B: Comitê de Redes	179
E.3	Metodologia C: Redes Alternadas	181
APÊNDICE F - TRABALHOS PUBLICADOS.		183
F.1	Estimativa do Albedo de Espalhamento Simples Usando uma Rede Neural de Múltiplas Camadas	183
F.2	Comparison of Two Learning Strategies for a Supervised Neural Network	184
F.3	A Committee of MLP with Adaptive Slope Parameter Trained by the quasi-Newton Method to Solve Problems in Hydrologic Optics	185
F.4	Comparação entre um Algoritmo Genético e uma Rede Neural Artificial na Estimativa de Parâmetros em Oscilações Harmônicas	186
F.5	Análise Temporal da Concentração de Clorofila- <i>a</i> da Superfície do Mar na Região Oceânica Sudeste Brasileira	187
F.6	Estimativa do Perfil da Concentração de Clorofila em Águas Naturais Através de um Perceptron de Múltiplas Camadas	188

1 INTRODUÇÃO

A possibilidade de reproduzir fenômenos da natureza por meio de equações (ou modelos) matemáticas(os) é um sonho parcialmente realizado pelo homem. A representação desses fenômenos por meio de modelos, mesmo que de forma aproximada, permite entender seu comportamento e, assim, corroborar para o desenvolvimento de novas tecnologias, ou, até mesmo, fazer previsões de “estados” futuros desses fenômenos, como a previsão numérica do tempo. A previsão numérica do tempo é de extrema importância na agricultura, e, por consequência, na economia, podendo ainda ser usada para prever desastres naturais.

No século I a.C, Lucrécio (poeta e filósofo) escreveu que a luz solar e seu calor eram compostos por pequenas partículas. Durante o século XVII, Newton e seus seguidores desenvolveram amplamente a teoria corpuscular da luz, capaz de explicar vários fenômenos óticos. No entanto, nesse mesmo século, Christian Huygens, René Descartes, entre outros, defendiam a teoria de que a luz deveria ser interpretada como onda. A razão pela qual Newton rejeitava a teoria ondulatória residia na propagação retilínea da luz. Porém, com a explicação dos fenômenos de interferência e difração por Thomas Young e Augustin Fresnel, no século XIX, a teoria ondulatória seguiu seu desenvolvimento. Vale lembrar que ambas as teorias consideravam a velocidade de propagação da luz finita, o que foi comprovado pelos experimentos de Ole Christensen Romer e Armand Hippolyte Louis Fizeau. O apogeu da teoria ondulatória aconteceu quando o físico James Clerk Maxwell unificou a ótica com o eletromagnetismo em seu Tratado de Eletricidade e Magnetismo em 1873.

Em 1905, o físico alemão Albert Einstein, usando o conceito de Max Planck sobre *quanta* de luz, explicou por que metais emitem elétrons quando expostos à radiação luminosa (efeito fotoelétrico). Esse fenômeno, não explicado pela teoria ondulatória, fez ressurgir a teoria corpuscular da luz. No século XX, foi atribuído à luz o comportamento dual, *i.e.*, um comportamento tanto de onda como de partícula, tornando as duas teorias aceitas. A teoria (ondulatória ou corpuscular) a ser considerada ao representar um fenômeno físico por meio de um modelo matemático está associada ao fenômeno físico observado.

Em ótica hidrológica, a luz é tratada como um feixe de fótons (partículas, portanto, teoria corpuscular) e a representação da interação desses fótons com o meio aquático é feita pela Equação de Transferência Radiativa (ETR). A partir da ETR e do conhecimento das propriedades óticas do meio aquático, bem como das condições de contorno e possíveis fontes internas de radiação, é possível simular a interação da

radiação eletromagnética com o meio e determinar a quantidade de fótons que estão emergindo¹ na superfície do meio aquático após a interação com este (problema direto). Por outro lado, a partir da ETR e da radiação emergente na superfície, é possível determinar as propriedades óticas do meio e/ou condições de contorno, e/ou ainda, possíveis fontes internas de radiação (problema inverso).

A modelagem de problemas diretos e inversos em ótica hidrológica, bem como metodologias gerais para solucionar esses problemas, têm sido tema de estudo de alguns pesquisadores do Laboratório Associado de Computação e Matemática Aplicada (LAC) do Instituto Nacional de Pesquisas Espaciais (INPE), desde 1997. Dentre os principais resultados obtidos na resolução de problemas inversos, destacam-se: a estimativa de fontes internas de bioluminescência e/ou a estimativa de propriedades óticas inerentes (STEPHANY, 1998; STEPHANY et al., 2000b; STEPHANY et al., 2000a); a estimativa da função de fase em Chalhoub et al. (2000); e a identificação de condições de contorno por Retamoso et al. (2002) e Campos Velho et al. (2002).

Chalhoub e Campos Velho (2001) fizeram a estimativa conjunta da função de fase e do albedo de espalhamento simples usando radiâncias medidas *in situ* e metodologia implícita. A minimização do funcional de diferenças quadráticas se deu por meio do método de Levenberg-Marquardt (LEVENBERG, 1944; MARQUARDT, 1963). Essa mesma estimativa conjunta, utilizando dados sintéticos, foi feita por Cortivo et al. (2012a) por meio de um comitê de Redes Neurais Artificiais (RNAs) especialistas. A estimativa apenas do albedo de espalhamento simples, também com RNAs, foi feita por Cortivo et al. (2010) e Cortivo et al. (2012b). No entanto, em ambos os problemas foram utilizados dados sintéticos. Em Chalhoub e Campos Velho (2002), é feita a estimativa simultânea da função de fase, do albedo de espalhamento simples e da espessura ótica do meio, utilizando apenas a radiação emergente na superfície. Em Chalhoub e Campos Velho (2003), é feita a estimativa de fontes internas de bioluminescência para dados de sensoriamento remoto (multiespectrais). Outros detalhes acerca dos resultados obtidos ao longo dos anos são encontrados em Campos Velho et al. (2000) e Campos Velho et al. (2003).

Os primeiros resultados obtidos na estimativa do perfil vertical da concentração de clorofila para águas naturais do Caso 1², a partir da radiação emergente na superfície

¹As partículas que emergem do meio são chamadas de “radiação emergente” e sua unidade de medida é a candela (potência por unidade de ângulo sólido).

²Em geral, águas do Caso 1 são aquelas em que a concentração de clorofila é predominante quando comparada aos demais constituintes. Por outro lado, águas do Caso 2 são aquelas em que as partículas inorgânicas ou a matéria orgânica dissolvida são predominantes. Na página 48 será apresentada uma definição mais detalhada para as águas do Caso 1 e do Caso 2.

da água, se devem a Souto (2006), Souto et al. (2007) e Souto et al. (2008). Nesses trabalhos foi adotada uma formulação implícita, considerando apenas a radiância emergente na superfície, medida em diferentes direções polares, como “fonte” de informação sobre a distribuição vertical da clorofila. Já a busca do mínimo para a funcional de diferenças quadráticas se deu por meio do algoritmo ACO (*Ant Colony Optimization*) (DORIGO, 1992). Em Cortivo et al. (2012c) é feita, por meio de RNAs, a estimativa do perfil médio da concentração de clorofila, \overline{chl} , para águas naturais do Caso 1, utilizando apenas a radiância emergente da superfície da água em diferentes direções polares. A Tabela 1.1 resume os trabalhos citados.

Como comentado anteriormente, a possibilidade de representar fenômenos físicos por meio de modelos matemáticos torna possível a previsão de seus estados futuros e um melhor entendimento desses fenômenos. Diante disso, a representação das propriedades óticas do meio aquático, por meio de modelos (bio-óticos) matemáticos, permite entender a variação dos índices de absorção e/ou espalhamento da luz, ao longo do espectro eletromagnético, e qual (ou quais) constituinte(s) colabora(m) para o aumento ou redução desses índices. Dessa forma, é possível expressar a cor do oceano aberto em função desses modelos bio-óticos (propriedades óticas inerentes³) (SATHYENDRANATH et al., 2001), permitindo entender as variações de tonalidade do oceano (regiões do oceano que apresentam cor esverdeada e regiões que apresentam cor azul escuro).

As variações da cor do oceano estão ligadas diretamente à absorção seletiva da radiação solar, nos comprimentos de onda do azul e do verde, pela clorofila. Dessa forma é possível quantificar a biomassa fitoplanctônica oceânica por meio de medidas (da cor do oceano) obtidas pelos sensores a bordo de satélites (MOREL; PRIEUR, 1977; GORDON; MOREL, 1983). A cor do oceano é uma fonte útil de informações sobre algumas de suas características e seu estudo possibilita determinar propriedades químicas, físicas e biológicas, ou ainda, auxiliar pescadores na escolha dos melhores locais para a pesca.

O fato de a clorofila estar presente na célula do fitoplâncton torna possível inferir a quantidade dessas células por meio da estimativa da concentração de clorofila. Dessa forma, a estimativa da concentração de clorofila nos oceanos recebe

³As propriedades óticas inerentes (*Inherent Optical Properties – IOPs*) são as propriedades do meio aquático que não dependem do campo de luz, *i.e.*, dependem exclusivamente das características da água. As duas principais IOPs são os coeficientes de absorção e de espalhamento. Por outro lado, existem as propriedades óticas aparentes (*Apparent Optical Properties – AOPs*) que são influenciadas pela distribuição angular do campo de luz, assim como pela natureza e quantidade de substâncias presentes no meio (PREISENDORFER, 1976). As AOPs são influenciadas pelas IOPs.

muita atenção da comunidade oceanográfica devido à possibilidade de estimar a quantidade de biomassa fitoplanctônica.

O fitoplâncton é o principal produtor primário dos oceanos e é representado por um conjunto de microalgas unicelulares que se desenvolvem na região eufótica. A atividade fotossintética do fitoplâncton é o primeiro passo na fixação do carbono inorgânico em carbono orgânico particulado no ambiente pelágico marinho (LALLI; PARSONS, 1993). Visto que os organismos fitoplanctônicos necessitam do dióxido de carbono (CO_2) para realizar a fotossíntese, populações de fitoplâncton, sustentadas por longos períodos, podem reduzir o efeito estufa causado por esse gás. Ainda, durante o processo de fotossíntese, átomos de oxigênio (O_2) são liberados para a atmosfera. Dessa forma, o fitoplâncton faz ligação entre a energia solar e a produção de recursos biológicos marinhos, dos quais dependem outros níveis tróficos (CLARK; SHERMAN, 1986). Além disso, o fitoplâncton é a base da cadeia trófica dos oceanos, servindo de alimento para pequenos crustáceos até grandes predadores. Assim, é possível dizer que a quantidade de clorofila nos oceanos é um parâmetro que pode ser utilizado para avaliar a saúde do planeta. Recentes estudos têm mostrado que mudanças na estrutura das comunidades de fitoplâncton podem alterar completamente a função de um ecossistema e levar à diferentes processos de causa-efeito no sistema oceano-atmosfera (FINKEL et al., 2010).

Em virtude da importância do fitoplâncton, a comunidade científica, apoiada pelos governos, vem se esforçando para monitorar a cor do oceano. Esse monitoramento pode ser feito em escala “local”, por meio de cruzeiros científicos, ou então em escala “global” por meio de sensores passivos a bordo de satélites. Nesse último caso é possível monitorar de grandes áreas do oceano.

O primeiro sensor orbital responsável pelo sensoriamento remoto da cor do oceano foi o CZCS (*Coastal Zone Color Scanner*), colocado em órbita pela agência espacial americana NASA (*National Aeronautics and Space Administration*) em 1978. Esse sensor deu origem a uma série de instrumentos cada vez mais sofisticados para o monitoramento da cor dos oceanos. As imagens obtidas pelo CZCS possibilitaram o entendimento da distribuição global do fitoplâncton, permitindo obter mapas com a variabilidade sinóptica, espacial e temporal da biomassa fitoplanctônica, parcialmente entendida, até então, pela amostragem feita por barcos/cruzeiros (YODER et al., 1988; FELDMAN et al., 1989; AIKEN et al., 1992; MCCLAIN, 1993; YODER et al., 1993; MITCHELL, 1994).

Tabela 1.1 - Propriedades óticas estimadas e metodologias inversas adotadas.

Propriedade	Mod. Direto	Tipo	Otimizador	Regularização	Dados	Ruído	Referência
$S_B(z)$, $a(z)$ e $b(z)$	Ins. Invar.	par/fun	QN	MaxEnt-0	E_λ	2% e 5%	Stephany (1998), Stephany et al. (2000b) e Stephany et al. (2000a)
$p(\cos \Theta)$	S_N - Dif. Fin.	par.	QN	MaxEnt-0	E_λ/I_λ	2% - 20%	Chalhoub et al. (2000)
C-C	LTS _N	par/fun	QN	Tikhonov-1	E_λ/I_λ	2% - 10%	Retamoso et al. (2002) e Campos Velho et al. (2002)
$p(\cos \Theta)$, ϖ_0	AS _N	par.	LM	-	I_λ	2%, 5%, 10%	Chalhoub e Campos Velho (2001)
$p(\cos \Theta)$, ϖ_0 , ζ	AS _N	par.	LM	-	E_λ/I_λ	5% - 20%	Chalhoub e Campos Velho (2002)
$S_0(z)$	AS _N	par.	LM	-	I_λ	0%, 1%, 2%, 5%	Chalhoub e Campos Velho (2003)
$\text{chl}(z)$ e $\overline{\text{chl}}$	LTS _N	fun.	ACO	Intrínseca	I_λ	2%, 5%, 10% e 20%	Souto (2006), Souto et al. (2007), Souto et al. (2008) e Souto et al. (2012)
ϖ_0	AS _N	par.	RNA+QN	-	I_λ	0%, 2%, 5%, 7%, 10%	Cortivo et al. (2010) e Cortivo et al. (2012b)
$p(\cos \Theta)$, ϖ_0	AS _N	par.	RNA+QN	-	I_λ	2.5%, 5%, 10%, 20%, 30%	Cortivo et al. (2012a)
$\overline{\text{chl}}$	AS _N	par.	RNA+QN	-	I_λ	5%, 10%, 15%, 20%	Cortivo et al. (2012c)

Fonte: Adaptada de Souto (2006).

O sucessor do CZCS foi o sensor OCTS (*Ocean Color and Temperature Scanner*), lançado pela agência espacial japonesa NASDA (*National Space Development Agency*) em 1996. Esse sensor deixou de operar sete meses depois do lançamento devido a problemas no painel solar do satélite ADEOS (*Advanced Earth Observing Satellite*) (KAWAMURA; THE OCTS TEAM, 1998). Em 1997, a NASA lançou o sensor SeaWiFS (*Sea-viewing Wide Field-of-view Sensor*), com várias melhorias em relação ao CZCS, e que operou até dezembro de 2010 (OCEAN COLOR, 2013a). A partir de 1996/1997 outros sensores foram lançados e os principais avanços estão ligados às características radiométricas e espectrais. A qualidade radiométrica foi aperfeiçoada em termos da amplitude dinâmica e da elevação da razão sinal/ruído, enquanto que o aperfeiçoamento espectral ocorreu em função do aumento do número de canais espectrais e da redução das larguras de banda (KAMPEL; NOVO, 2005).

A estimativa da concentração de clorofila de superfície a partir das radiâncias e/ou reflectâncias medidas pelos sensores a bordo de satélites é feita por meio de algoritmos específicos, obtidos empiricamente a partir de conjuntos de dados que contenham medidas *in situ*. O primeiro algoritmo proposto (GORDON et al., 1983; FELDMAN et al., 1989; EVANS; GORDON, 1994) para processar os dados do sensor CZCS foi obtido a partir do conjunto de dados *Nimbus Experiment Team* (ACKER, 1994), o qual continha menos de 60 estações.

Com o lançamento do sensor SeaWiFS e com o aumento do número de medidas *in situ*, foram propostos algoritmos mais robustos para converter a radiação medida pelos sensores em mapas de concentração de clorofila de superfície. Em O'Reilly et al. (1998), é utilizado o conjunto de dados SeaBAM (*SeaWiFS Bio-optical Algorithm Mini-workshop*) (FIRESTONE; HOOKER, 1998), contendo 919 estações, a fim de testar o desempenho de dois algoritmos semianalíticos e 15 empíricos. Conforme O'Reilly et al. (1998), os algoritmos que apresentaram melhor desempenho foram os algoritmos OC2 (*Ocean Chlorophyll 2*) e OC4 (*Ocean Chlorophyll 4*). Ambos algoritmos são empíricos, consideram um polinômio cúbico modificado, e são baseados na razão de bandas. O algoritmo OC2 considera a razão entre as reflectâncias de sensoriamento remoto medidas em $R_{sr} = 490 \text{ nm}$ e $R_{sr} = 555 \text{ nm}$, enquanto o algoritmo OC4 considera a razão entre as reflectâncias medidas em $R_{sr} = 443 \text{ nm}$, $R_{sr} = 490 \text{ nm}$, $R_{sr} = 510 \text{ nm}$ e $R_{sr} = 555 \text{ nm}$.

Melhorias no algoritmo OC2 foram propostas por Maritorena e O'Reilly (2000) a partir da inclusão de novas medidas *in situ* no conjunto SeaBAM, resultando no algoritmo OC2v2. O'Reilly et al. (2000) sugeriam a versão quatro para ambos os

algoritmos, OC2v2 e OC4, a partir da inclusão de outras novas medidas *in situ* no conjunto SeaBAM, resultando nas versões OC2v4 e OC4v4. Versões mais atualizadas desses algoritmos, ou então, outros algoritmos para outros sensores orbitais, são encontradas em [Ocean Color \(2013b\)](#). Os algoritmos citados, em geral, apresentam ótimas estimativas para a concentração de clorofila de superfície para águas naturais do Caso 1, no entanto, acabam sobrestimando os valores da concentração de clorofila de superfície em águas naturais do Caso 2 devido à presença de material inorgânico em suspensão e matéria orgânica dissolvida ([IOCCG, 2000](#)).

Metodologias como RNAs são utilizadas na estimativa da concentração de clorofila de superfície e se mostram eficientes. [Gross et al. \(2000\)](#) utilizaram RNAs do tipo Perceptron de Múltiplas Camadas (PMC) treinadas com o algoritmo *Backpropagation*, para estimar a concentração da clorofila próxima à superfície, para águas naturais do Caso 1, a partir da reflectância medida pelo sensor SeaWiFS. Segundo [Gross et al. \(2000\)](#), os resultados obtidos pelas RNAs mostraram uma redução significativa do erro relativo e absoluto dessa estimativa, quando comparados aos resultados obtidos pelo algoritmo OC2. Uma aplicação de RNAs do tipo PMC, também treinada com o algoritmo de retropropagação do erro, no entanto, em águas do Caso 2, é encontrada em [Schiller e Doerffer \(1999\)](#). Nesse trabalho, os autores determinam a concentração de clorofila, material inorgânico em suspensão e matéria orgânica dissolvida, a partir das reflectâncias de 16 bandas espectrais. [Richardson et al. \(2002\)](#) investigaram características dos perfis verticais da concentração de clorofila em águas costeiras, na África do Sul, por meio de RNAs do tipo mapas auto-organizáveis (*Self Organization Maps – SOM*) ([KOHONEN, 1982](#); [KOHONEN, 1989](#); [HAYKIN, 2001](#)).

O emprego de sensores a bordo de satélites para o monitoramento dos oceanos, juntamente com algoritmos capazes de processar os dados coletados por esses sensores, em mapas de concentração de clorofila, por exemplo, foi e continua sendo um grande passo em ótica marinha. Além disso, o monitoramento da concentração de clorofila na superfície, em grandes áreas de oceano, permite observar sinoticamente feições biológicas de sistemas dinâmicos como os grandes giros subtropicais, frentes oceânicas, ressurgências e vórtices de mesoescala ([MONGER et al., 1997](#); [RYAN et al., 2001](#); [KAMPEL, 2003](#); [GARCIA et al., 2004](#); [GONZÁLEZ-SILVERA et al., 2004](#)).

Apesar de os algoritmos propostos produzirem mapas de concentração de clorofila acurados, principalmente para águas do Caso 1, permitindo, assim, o monitoramento de grandes áreas do oceano, esses mapas fornecem informações apenas da camada superficial da água. Em situações em que a concentração de clorofila é uniforme,

ao longo da profundidade, esses mapas são válidos para extrair informações sobre a estrutura vertical da concentração de clorofila. No entanto, na maioria das vezes, os perfis verticais da concentração de clorofila apresentam um pico máximo de concentração e se assemelham às curvas gaussianas (LEWIS et al., 1983). Platt et al. (1994) sugerem que seja adicionado um valor constante de concentração de clorofila de fundo (*background*) ao modelo sugerido por Lewis et al. (1983), estando esse valor associado a uma quantidade mínima de clorofila na coluna d'água. Já os trabalhos de Siegel et al. (2002) e Henson et al. (2009) fazem uma análise da concentração de clorofila de superfície, ao longo do tempo, para o Atlântico Norte. Dandonneau et al. (2004) fazem um estudo da concentração de superfície, também ao longo do tempo, para o Pacífico. Valério et al. (2012) utilizaram uma série temporal de 13 anos de medidas de concentração de clorofila de superfície em águas naturais do Caso 1, do Atlântico Sul, adjacente à plataforma sudeste brasileira, e é possível observar que a variação dessa concentração na superfície também segue um comportamento gaussiano ao longo de cada ano.

Em virtude da variação da concentração de clorofila ao longo da profundidade, e da inviabilidade para estimar esses perfis pelos algoritmos propostos (comentados anteriormente), Souto (2006) apresentou uma metodologia inversa para solucionar esse problema. Nesse trabalho, o autor utiliza a ETR para representar a interação da radiação eletromagnética com o meio aquático, e modelos bio-óticos que dependem da distribuição vertical da concentração de clorofila para representar os coeficientes de absorção (MOREL, 1991) e espalhamento (GORDON; MOREL, 1983). O modelo inverso, adotado pelo autor, consiste em resolver o problema direto definido pela ETR, em que os valores da concentração de clorofila, ao longo da profundidade, são determinados durante o processo de minimização do funcional de diferenças quadráticas. Esse funcional foi definido pela diferença entre a radiação experimental⁴ emergente na superfície e a calculada pelo modelo inverso, para cada conjunto de valores de concentração de clorofila. Em seu trabalho, Souto (2006) obteve ótimas estimativas para concentrações até aproximadamente 20 metros e resultados não satisfatórios para profundidades entre 20 e 40 metros. Esse resultados, principalmente para os valores de concentração de clorofila após os 20 metros, foram melhorados por meio do uso do critério da derivada segunda durante o processo de otimização (SOUTO et al., 2007; SOUTO et al., 2008; SOUTO et al., 2012).

Apesar dos ótimos resultados obtidos com o critério da derivada segunda, a desvan-

⁴Nesse caso, entende-se por radiação experimental emergente na superfície a radiação calculada pelo modelo direto e corrompida com algum ruído gaussiano.

tagem da metodologia proposta por Souto (2006) reside no grande tempo computacional necessário para a resolução do problema inverso, mesmo com códigos paralelizados sob o paradigma MPI (*Message Passing Interface*). A grande demanda de tempo se deve a dois fatores: a) o uso do algoritmo estocástico ACO para minimização do funcional de diferenças quadráticas; b) o elevado número de chamadas do modelo direto para a resolução do problema inverso. Melhorias sob o ponto de vista do desempenho computacional foram propostas em Souto et al. (2010). No entanto, nesse problema, o autor considerou algumas simplificações no modelo matemático adotado (problema isotrópico e com simetria azimutal), o que pode distanciá-lo um pouco da realidade em determinadas situações. As simplificações feitas no modelo e a grande demanda de tempo computacional (para modelos mais próximos à realidade) prejudicam o processamento de dados para grandes áreas oceânicas.

1.1 Objetivos e Metodologia

Em função da inviabilidade de se determinar o perfil da concentração de clorofila pelos algoritmos propostos para processar os dados de cor do oceano, e da grande demanda de tempo computacional necessário pela metodologia proposta por Souto (2006) e Souto et al. (2007), esta tese propõe uma metodologia alternativa para solucionar o problema. Portanto, o objetivo é desenvolver uma metodologia alternativa que demande um menor tempo computacional e que continue a produzir resultados satisfatórios, permitindo, no entanto, o processamento de dados de grandes áreas do oceano.

A metodologia proposta é baseada em RNAs do tipo PMC treinadas com o algoritmo quasi-Newton e são utilizadas três redes: a RNA_{Cte} ; a RNA_{σ} ; e a RNA_{τ_m} . A RNA_{Cte} utiliza a radiação emergente para classificar os problemas em duas classes: a primeira classe é composta por aqueles problemas em que há a possibilidade de determinar o perfil da concentração de clorofila (ao longo da profundidade), e a segunda classe é composta pelos problemas em que apenas é possível determinar o perfil médio de concentração de clorofila. A inviabilidade de se determinar o perfil da concentração de clorofila está associada a limitações físicas do problema e/ou da formulação adotada. Uma descrição detalhada das dificuldades em se determinar o perfil da concentração de clorofila será apresentada no Capítulo 4. As outras duas redes resolvem o problema de modo cascata, *i.e.*, a estimativa feita pela primeira rede, dessas duas, será considerada como um dado adicional aos padrões de entrada da última rede.

O modelo adotado para representar o perfil vertical da concentração de clorofila segue um comportamento gaussiano com um valor de concentração de clorofila de

fundo e que foi proposto por [Platt et al. \(1994\)](#). Os parâmetros a serem estimados pelas redes que operam de modo cascata são o desvio padrão e a profundidade do pico da concentração de clorofila, sendo que, cada uma das redes, determina um dos parâmetros. O parâmetro estimado pela primeira dessas duas redes passa a compor uma das entradas para a segunda rede. O valor da concentração de clorofila de fundo e a quantidade total de biomassa, presentes no modelo sugerido por [Platt et al. \(1994\)](#), serão considerados constantes.

A rede RNA_{Cte} , que faz a classificação, foi treinada para determinar um valor (médio) de concentração de clorofila a partir da radiação emergente na superfície. Dessa forma, após a saída dessa rede, há uma condição baseada em um limiar, que irá definir se é possível determinar o valor dos dois parâmetros a serem estimados, ou então, se será considerado um perfil médio. Quando o valor da concentração estimada por essa rede for superior ao limiar, o processo passa para o passo seguinte (determinar o desvio padrão e a profundidade do pico da concentração).

A rede seguinte, denominada RNA_{σ} , determina o valor do desvio padrão do modelo matemático adotado para representar a distribuição de clorofila ao longo da profundidade, a partir da radiação emergente na superfície da água e da concentração de clorofila de superfície. Essa estimativa, feita pela rede RNA_{σ} , é adicionada às entradas dos padrões de entrada da rede seguinte, denominada RNA_{τ_m} , que faz a estimativa da profundidade do pico da concentração. A radiação emergente na superfície, que é entrada para a rede RNA_{Cte} , é a mesma para as redes RNA_{σ} e RNA_{τ_m} . Em todos os problemas considerados, a radiação emergente é a radiação calculada pelo modelo matemático adotado e descrito no [Capítulo 2](#) com os valores para os parâmetros descritos no [Capítulo 4](#).

Como este trabalho tem o intuito de apresentar uma metodologia alternativa à proposta por [Souto \(2006\)](#), foi escolhido o modelo gaussiano, citado anteriormente, para representar a distribuição da concentração de clorofila ao longo da profundidade. Portanto, os resultados obtidos são válidos para perfis que “respeitam” o modelo sugerido por [Platt et al. \(1994\)](#). Para regiões oceânicas nas quais esse perfil não é válido, como por exemplo em regiões próximas ao Japão ([KAMEDA; MATSUMURA, 1998](#)), as estimativas obtidas por estas redes poderão não ser boas. O perfil adotado por [Kameda e Matsumura \(1998\)](#) foi obtido por [Matsumura e Shiimoto \(1993\)](#). Esse perfil considera um termo adicional, dado pelo produto da profundidade pelo gradiente vertical da concentração de clorofila, sendo formado por três componentes: concentração de clorofila de fundo; produto do gradiente vertical da concentração de

clorofila pela profundidade; curva gaussiana. Outros perfis da distribuição da concentração de clorofila ao longo da profundidade podem ser encontrados no trabalho de Uitz et al. (2006). Nesse trabalho, os autores também examinam a possibilidade de inferir a quantidade de biomassa fitoplanctônica a partir da concentração de clorofila de superfície, em diversas regiões oceânicas do planeta, para águas do Caso 1.

As principais contribuições deste trabalho são:

- desenvolvimento de uma metodologia promissora, por meio de RNAs, para a estimativa dos perfis verticais da concentração de clorofila para águas oceânicas do Caso 1;
- utilização de uma RNA a fim de determinar a possibilidade de resolução de um problema inverso;
- estruturação de RNAs para atuarem em cascata, *i.e.*, a estimativa feita por uma rede precedente passa a compor uma das entradas para a rede subsequente;
- aplicação de RNAs na estimativa de parâmetros em um problema inverso de ótica hidrológica;
- aplicação do método quasi-Newton, para o treinamento das RNAs, com o cálculo dos gradientes do funcional de diferenças quadráticas feito de modo analítico;
- adição de cada parâmetro das funções de ativação ao conjunto de variáveis da rede (pesos sinápticos níveis de *bias*) e busca dos valores ótimos durante o treinamento por meio de fórmulas analíticas;
- economia de tempo computacional com o uso de RNAs e do cálculo dos gradientes, do funcional de diferenças quadráticas, de forma analítica.

1.2 Organização do Trabalho

O Capítulo 2 apresenta uma breve descrição de alguns métodos para se obter a solução da ETR, bem como a descrição do problema da ótica (interação da radiação solar com o meio aquático), e do modelo matemático adotado considerado nesta tese.

O Capítulo 3 apresenta uma descrição histórica do surgimento das RNAs, uma descrição do neurônio biológico, do neurônio artificial e suas possíveis funções de

ativação. Esse capítulo versa também sobre as características de uma RNA, do PMC, da aprendizagem das RNAs e dos algoritmos de aprendizagem, dando ênfase para os algoritmos *Backpropagation* “tradicional”, *Backpropagation* “generalizado” e quasi-Newton. Apresenta, ainda, os detalhes da formulação do problema de treinamento de uma RNA, considerando o parâmetro de cada função de ativação como variável a ser otimizada durante o treinamento, e como as RNAs podem atuar na resolução de problemas inversos.

O Capítulo 4 descreve os detalhes gerais e específicos do problema descrito no Capítulo 2, bem como a justificativa dos valores dos parâmetros utilizados. Descreve, ainda, os padrões de entrada e saída da rede, metodologia de treinamento adotada, classificação dos problemas, além de dar uma breve descrição do algoritmo desenvolvido, que implementa as redes neurais utilizadas.

O Capítulo 5 apresenta os resultados obtidos pela técnica adotada e descrita no Capítulo 4. Apresenta, também, a discussão desses resultados e uma comparação com os resultados obtidos por Souto (2006) e Souto et al. (2007).

Por fim, o Capítulo 6 apresenta as conclusões obtidas com a elaboração desta tese e sugestões para trabalhos futuros.

No Apêndice A, são encontrados os formalismos matemáticos sobre a aproximação da ETR em um sistema de equações diferenciais ordinárias de primeira ordem a coeficientes constantes, e detalhes do método AS_N . No Apêndice B, é descrito brevemente o teorema da adição dos harmônicos esféricos e a aproximação da função de fase. No Apêndice C, são apresentados detalhes sobre o funcionamento do neurônio biológico, e, no Apêndice D, são encontrados os formalismos matemáticos sobre os algoritmos de aprendizagem comentados no Capítulo 3. No Apêndice E, são apresentadas, brevemente, metodologias que foram “testadas” durante a elaboração deste trabalho e que, no entanto, não produziram resultados satisfatórios. Finalmente, no Apêndice F, são apresentadas as primeiras páginas de cada trabalho resultante das atividades de pesquisa desenvolvidas durante a realização desta tese.

2 EQUAÇÃO DE TRANSPORTE DE PARTÍCULAS

A equação de transporte de partículas, proposta por Ludwig Boltzmann, é um modelo matemático eficiente, porém complexo, na maioria das vezes, para modelar fenômenos de transporte de partículas como nêutrons e fótons. Suas aplicações estão em diversas áreas da ciência como: medicina, processos de transferência de calor (condução e convecção), plasmas, reatores nucleares, ótica hidrológica e na área espacial. No caso específico da ótica hidrológica, a equação de transporte recebe o nome de Equação de Transferência Radiativa (ETR), pois, nesse caso, as partículas consideradas são fótons.

Aplicações da equação de transporte na medicina, em específico em tomografia, podem ser encontradas em [Klose et al. \(2002\)](#). Nesse trabalho, os autores apresentam um modelo direto, utilizando a ETR, para a propagação de fótons em meios (tecidos) com alto índice de espalhamento, mas que apresentam, no entanto, regiões inertes (*void-like regions*) nas quais os índices de absorção e de espalhamento são muito baixos. Segundo [Klose et al. \(2002\)](#), o uso da equação de difusão não permite a modelagem desse tipo de problema. O modelo inverso, para o problema citado, é encontrado em [Klose e Hielscher \(2002\)](#). Já em [Klose et al. \(2005\)](#) são apresentados os resultados obtidos na reconstrução da distribuição espacial de fontes fluorescentes em tecidos biológicos com um alto índice de espalhamento.

Aplicações da equação de transporte em transferência de calor podem ser encontradas em [Modest \(2003\)](#). Na geração de energia, a equação de transporte de partículas é amplamente aplicada na modelagem de reatores nucleares ([DUDERSTADT; MARTIN, 1975](#); [BELL; GLASSTONE, 1970](#); [LEWIS; MILLER, 1993](#)). Em Astrofísica, a ETR pode ser utilizada no estudo de transferência radiativa de fótons em atmosferas planetárias e estelares ([AMBARZUMIAN, 1944](#); [CHANDRASEKHAR, 1950](#)).

A ETR também pode ser aplicada em ótica hidrológica ([MOBLEY, 1994](#)). Nesse caso, a luz é tratada como um feixe de fótons (partículas) e a ETR é utilizada a fim de modelar os fenômenos de absorção e de espalhamento, fontes internas de bioluminescência, entre outras propriedades do meio aquático. Deduções da equação de transporte sob o ponto de vista da transferência radiativa, neutrônica e ótica hidrológica podem ser encontradas em [Chandrasekhar \(1950\)](#), [Duderstadt e Martin \(1975\)](#) e [Mobley \(1994\)](#), respectivamente.

A grande aplicação da equação de transporte em várias áreas da ciência, principalmente em Astrofísica e posteriormente em neutrônica, fez com que a comunidade

científica buscasse soluções desde as versões simples até as versões complexas da equação de transporte, resultando no surgimento de várias técnicas para a obtenção dessas soluções. Uma versão clássica da equação de transporte é a chamada equação de Case (CASE; ZWEIFEL, 1967). Nessa versão, os autores consideram-na como unidimensional, apresentando apenas um grupo de energia, espalhamento isotrópico e ausência de fontes internas de radiação. A solução para essa versão da equação de transporte é baseada na expansão em autovalores e autofunções singulares, cujos detalhes estão em Case e Zweifel (1967). Nessa referência, também são encontrados outros métodos e soluções para versões idealizadas da equação de transporte, como por exemplo, soluções baseadas em funções de Green em um meio infinito, solução para o problema de Milne, etc.

O método mais popular para se obter a solução da equação de transporte é o método S_N , que foi introduzido por Wick (WICK, 1943) e Chandrasekhar (CHANDRASEKHAR, 1950). Esse método consiste em aproximar a integral angular por um esquema de quadratura numérica e, então, resolver o conjunto de equações diferenciais ordinárias resultante por diferenças finitas. Uma implementação numericamente estável pode ser encontrada em Stamnes e Swanson (1981) e Stamnes et al. (1988).

Outros métodos populares utilizados para se obter soluções para a equação de transporte, também populares, são: o método P_N (JEANS, 1917), baseado na expansão do fluxo em um conjunto de harmônicos esféricos; e o método da inserção invariante (*invariant imbedding*) (AMBARZUMIAN, 1944), no qual seus operadores são calculados pela integração das equações de Riccati. O método da inserção invariante foi adaptado para problemas em ótica hidrológica por Preisendorfer (1961) e Mobley (1994). Ainda, tem-se o método F_N , que reduz a equação íntegro-diferencial e condições de contorno associadas em um sistema de equações integrais singulares para as intensidades de radiação (SIEWERT, 1978; SIEWERT; BENOIST, 1979; GRANDJEAN; SIEWERT, 1979; DEVAUX; SIEWERT, 1980; GARCIA, 1985), e o método LTS_N , que segue a linha do S_N , no entanto, a solução do sistema de equações diferenciais ordinárias resultante é obtida por meio da aplicação da transformada de Laplace na variável espacial (VILHENA; BARICHELLO, 1991; BARICHELLO; VILHENA, 1993; SEGATTO; VILHENA, 1994). Há também as variantes do método LTS_N , como o LTP_N (VILHENA; STRECK, 1992), o $LTC h_N$ (CARDONA; VILHENA, 1994; CARDONA, 1996), entre outros. O método LTP_N é uma extensão do método P_N , e o método $LTC h_N$ aplica polinômios de Chebyshev em vez de polinômios de Legendre, para aproximar o fluxo angular, e em seguida aplica a transformada de Laplace para resolver o sistema de equações diferenciais.

Há também o método das ordenadas discretas analítico AS_N (CHALHOUB, 1997; CHALHOUB; GARCIA, 2000), que também segue a linha do método S_N , no entanto, a solução para o conjunto de equações diferenciais ordinárias resultante é obtida por meio da decomposição espectral da matriz de espalhamento (matriz S_N). Desse modo, toda solução é calculada analiticamente e a única aproximação numérica é o cálculo dos autovalores e autovetores da matriz S_N .

Os métodos e referências citadas cobrem os principais caminhos para a busca de soluções para a equação de transporte, na linha dos métodos determinísticos. A outra linha de busca de soluções para a equação de transporte segue uma abordagem probabilística. Nessa linha, é utilizado o método de Monte Carlo (BELL; GLASSTONE, 1970; SPANIER; GELBARD, 2008), que se baseia especificamente na essência física do problema, descrevendo os históricos dos fótons, que são tomados individualmente e aleatoriamente, gerados pelas sucessivas colisões durante a sua trajetória. O histórico é construído de forma estocástica a partir do coeficiente de atenuação. O método de Monte Carlo é utilizado em situações nas quais a geometria do problema é complexa.

A próxima seção apresenta os detalhes do processo radiativo em águas naturais. Posteriormente, são apresentados alguns detalhes sobre o formalismo matemático envolvido e sobre o método de resolução adotado para a ETR. Com relação ao método de resolução, são apresentados alguns detalhes na seção seguinte e no Apêndice A. O método de resolução segue o trabalho de Chalhoub (2005) e Cortivo (2008).

2.1 O Problema da Ótica e o Modelo Matemático

O processo de interação de um feixe de fótons com o meio aquático pode ser influenciado por muitos fatores: estado da superfície da água, devido à ação dos ventos; condições de fundo; propriedades óticas inerentes, como absorção e espalhamento; existência ou não de fontes de luz submersas, como bioluminescência e fluorescência; micro-organismos e a presença de constituintes opticamente ativos, como matéria orgânica dissolvida e material particulado. Todos esses fatores influenciam no fluxo de radiação emergente na superfície que, em geral, é a única medida disponível a ser utilizada na tentativa de estimar alguma(s) propriedade(s) do corpo d'água. Ainda, acrescenta-se o fato de que todas essas condições citadas podem variar na escala temporal, espacial e espectral, tornando o problema ainda mais complexo. A Figura 2.1 ilustra de forma simplificada esse processo de interação dos fótons com o meio aquático.

Na Figura 2.1, θ_0 e φ_0 representam, respectivamente, o ângulo polar, em que

$\mu = \cos \theta$, e o ângulo azimutal de incidência de um feixe de fótons provindo do Sol, em relação à normal \vec{n} à superfície da água. No momento em que o feixe de fótons atinge a superfície da água pode ocorrer tanto a reflexão quanto a refração. Aqueles fótons que são refletidos voltam para o ar. Já aqueles que penetram no meio aquático mudam de direção devido à refração e, então, começam a se propagar na direção θ' até o momento em que colidam com alguma partícula/molécula (água, matéria orgânica, fitoplâncton, etc.). Nessa colisão, pode ocorrer absorção, espalhamento ou ambos. Os fótons que são absorvidos desaparecem do sistema e aqueles que são espalhados podem ser continuamente espalhados até que sejam “emitidos” para o ar, ou então, até o momento em que são absorvidos no corpo d’água. Para aqueles fótons que são “emitidos” para o ar, no momento da troca de meio (água-ar), ocorre novamente a refração¹. Nesse momento, o(s) fóton(s) passa(m) a se propagar com uma direção polar θ e em uma direção azimutal φ .

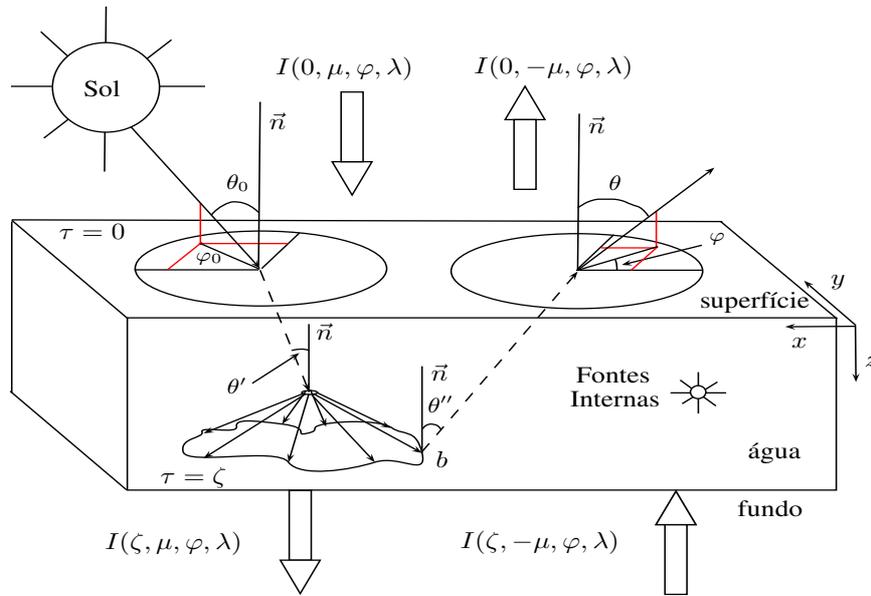


Figura 2.1 - Processo de interação da radiação eletromagnética com águas naturais.
Fonte: Adaptada de Stephany (1998).

Além disso, na Figura 2.1, a representa a absorção, b representa o espalhamento, (x, y, z) representam o sistema de coordenadas cartesianas, τ é a variável ótica que está associada à variável geométrica z , ζ é a profundidade ótica. Há também

¹Na primeira situação, *i.e.*, ar-água, o fóton passa de um meio menos refringente (ar) para um meio mais refringente (água). Nesse caso, o ângulo de refração θ' é menor que o ângulo de incidência θ_0 em relação à normal, *i.e.*, se aproximam da normal. Já na outra situação, quando o fóton é lançada para o ar, ocorre o contrário, *i.e.*, o ângulo de refração θ é maior que o ângulo de incidência θ'' em relação à normal, *i.e.*, se afastam da normal.

a representação das possíveis fontes internas. Por fim, $I(0, \mu, \varphi, \lambda)$, $I(\zeta, \mu, \varphi, \lambda)$, $I(0, -\mu, \varphi, \lambda)$ e $I(\zeta, -\mu, \varphi, \lambda)$ representam, respectivamente, as radiâncias descendentes (incidentes) na superfície e fundo, e as radiâncias ascendentes (emergentes) na superfície e fundo.

Nesta tese, o processo de interação da radiação eletromagnética provinda do Sol com o meio aquático é modelado pela ETR. No problema abordado neste trabalho, é considerada geometria plano-paralela², com multiregiões espaciais³ relacionadas à profundidade, com dependência nos ângulos zenital e azimutal, bem como no comprimento de onda, e ainda, o espalhamento é anisotrópico (veja Figura 2.2).

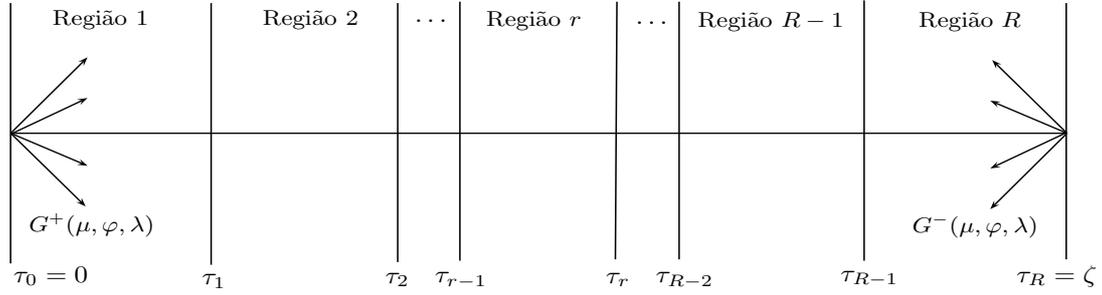


Figura 2.2 - Representação gráfica de um problema de multiregiões.
Fonte: Chalhoub (2005).

Para um problema com essas características, a ETR é escrita como

$$\mu \frac{\partial}{\partial \tau} I_r(\tau, \mu, \varphi, \lambda) + I_r(\tau, \mu, \varphi, \lambda) = \frac{b_r(\tau, \lambda)}{c_r(\tau, \lambda)} \int_{-1}^1 \int_0^{2\pi} \int_{\lambda} p_r(\cos \Theta, \lambda) I_r(\tau, \mu', \varphi', \lambda') d\lambda' d\varphi' d\mu' + S_0(\tau, \lambda), \quad (2.1)$$

sujeita às condições de contorno,

$$I_1(\tau_0, \mu, \varphi, \lambda) = G^+(\mu, \varphi, \lambda) \quad \text{e} \quad I_R(\tau_R, -\mu, \varphi, \lambda) = G^-(\mu, \varphi, \lambda), \quad (2.2)$$

com condições de interface, para $r = 1, 2, \dots, R - 1$, dadas por

$$I_r(\tau_r, \pm\mu, \varphi, \lambda) = I_{r+1}(\tau_r, \pm\mu, \varphi, \lambda), \quad (2.3)$$

²Geometria plano-paralela é adotada quando é possível considerar que, para uma determinada região (do oceano), não há variações “significativas” com relação as direções x e y . Sendo assim, as variações mais significativas estão na direção z , *i.e.*, com relação a profundidade. Ver Figura 2.1.

³Um problema multiregiões é resultado da extratificação em planos (paralelos) da profundidade, *i.e.*, é feita discretização na variável espacial (profundidade).

em que $I_r(\tau, \mu, \varphi, \lambda)$ denota a intensidade de radiação (radiância) na região r , $\tau \in (0, \zeta)$ é a variável ótica, com ζ representando a profundidade ótica do meio, $\mu \in [-1, 1]$ e $\varphi \in [0, 2\pi]$ são, respectivamente, o cosseno do ângulo polar medido a partir do eixo positivo τ e o ângulo azimutal, os quais especificam a direção de propagação Θ da radiação no meio, e λ representa o comprimento de onda do fóton. Além disso, $c_r(\tau, \lambda) = a_r(\tau, \lambda) + b_r(\tau, \lambda)$ é o coeficiente de atenuação do feixe, em que $a_r(\tau, \lambda)$ e $b_r(\tau, \lambda)$ são, respectivamente, o coeficiente de absorção e de espalhamento, $p_r(\cos \Theta, \lambda)$ é a função de fase de espalhamento de $\{\mu', \varphi', \lambda'\}$ para $\{\mu, \varphi, \lambda\}$, $S_0(\tau, \lambda)$ é uma fonte interna de radiação, e, por fim, $G^+(\mu, \varphi, \lambda)$ e $G^-(\mu, \varphi, \lambda)$ são as radiâncias incidentes em $\tau_0 = 0$ e $\tau_R = \zeta$, respectivamente.

Discretizando as Eqs. (2.1), (2.2) e (2.3), para cada região, na variável λ e considerando todas as variáveis dependentes do comprimento de onda como valores médios no intervalo (banda) $\Delta\lambda_g$, tem-se para uma variável $F(\lambda)$,

$$F_g = F(\lambda_g) = \frac{1}{\Delta\lambda_g} \int_{\Delta\lambda_g} F(\lambda) d\lambda,$$

em que λ_g é um comprimento de onda médio no intervalo de banda g . Além disso, considerando que não há espalhamento para baixo (*downscattering*), e nem espalhamento para cima (*upscattering*), *i.e.*, a partícula só pode ser espalhada dentro da mesma banda, e que os coeficientes de espalhamento e atenuação do feixe também assumem valores médios em cada região, com relação à variável ótica τ , a Eq. (2.1) passa a ser escrita como

$$\mu \frac{\partial}{\partial \tau} I_{r,g}(\tau, \mu, \varphi) + I_{r,g}(\tau, \mu, \varphi) = \frac{b_{r,g}}{c_{r,g}} \int_{-1}^1 \int_0^{2\pi} p_r(\cos \Theta) I_{r,g}(\tau, \mu', \varphi') d\varphi' d\mu' + S_{0,g}(\tau), \quad (2.4)$$

sujeita às condições de contorno,

$$I_{1,g}(\tau_0, \mu, \varphi) = I_{0,g} \delta(\mu - \mu_0) \delta(\varphi - \varphi_0) + D_g^+(\mu) + \rho_s^+ I_{1,g}(\tau_0, -\mu, \varphi) + \frac{\rho_d^+}{\pi} \int_0^{2\pi} \int_0^1 I_{1,g}(\tau_0, -\mu', \varphi') \mu' d\mu' d\varphi', \quad (2.5a)$$

e

$$I_{R,g}(\tau_R, -\mu, \varphi) = D_g^-(\mu) + \rho_s^- I_{R,g}(\tau_R, \mu, \varphi) + \frac{\rho_d^-}{\pi} \int_0^{2\pi} \int_0^1 I_{R,g}(\tau_R, \mu', \varphi') \mu' d\mu' d\varphi', \quad (2.5b)$$

e condições de interface, para $r = 1, 2, \dots, R-1$, dadas por

$$I_{r,g}(\tau_r, \pm\mu, \varphi) = I_{r+1,g}(\tau_r, \pm\mu, \varphi), \quad (2.6)$$

para $\mu \in (0, 1]$ e $\varphi \in [0, 2\pi]$. O termo $I_{0,g}$ denota a intensidade do feixe incidente, μ_0 e φ_0 são, respectivamente, o cosseno do ângulo polar e o ângulo azimutal do feixe incidente em $\tau = \tau_0$, $D_g^+(\mu)$ e $D_g^-(\mu)$ representam a distribuição da radiação incidente nas superfícies localizadas em $\tau = \tau_0$ e $\tau = \tau_R$, respectivamente. E, por fim, os termos ρ_s^+ , ρ_d^+ , ρ_s^- e ρ_d^- representam a reflexão especular e difusa nas superfícies de contorno em $\tau = \tau_0$ e $\tau = \tau_R$.

Ainda, nesta tese, é considerado que $D_g^+(\mu) = D_g^-(\mu) = \rho_s^+ = \rho_d^+ = \rho_s^- = \rho_d^- = 0$. Dessa forma, o conjunto de Eqs. (2.5) passa a ser escrito como

$$I_{1,g}(\tau_0, \mu, \varphi) = I_{0,g}\delta(\mu - \mu_0)\delta(\varphi - \varphi_0) \quad \text{e} \quad I_{R,g}(\tau_R, -\mu, \varphi) = 0, \quad (2.7)$$

com condições de interface expressas pela Eq. (2.6). O termo $p_r(\cos \Theta)$ será representado por uma expansão finita de polinômios de Legendre (CHANDRASEKHAR, 1950) em termos do cosseno do ângulo de espalhamento Θ , de modo que

$$p_r(\cos \Theta) = \frac{1}{4\pi} \sum_{l=0}^L \beta_{l,r} P_l(\cos \Theta), \quad \beta_{0,r} = 1 \quad \text{e} \quad |\beta_{l,r}| < 2l + 1 \quad \text{para} \quad 0 < l \leq L,$$

em que $\beta_{l,r}$ e P_l são os coeficientes e os polinômios de Legendre de L -ésima ordem, respectivamente.

Portanto, o problema que é resolvido nesta tese é definido pela Eq. (2.4), sujeita às condições de contorno definidas pelas Eqs. (2.7) e condições de interface dadas pela Eq. (2.6). Os coeficientes de espalhamento $b_{r,g}$ e atenuação $c_{r,g}$, bem como os comprimentos de onda adotados, número de regiões espaciais, ângulos de incidência e demais parâmetros, serão tratados no Capítulo 4. Vale lembrar, também, que nesta tese não serão levadas em conta fontes internas de radiação, nem a refração das partículas na superfície da água.

Para resolver o problema definido pela Eq. (2.4), sujeita às condições de contorno definidas pela Eq. (2.7) e condições de interface definidas pela Eq. (2.6), inicialmente, a intensidade de radiação é decomposta (CHANDRASEKHAR, 1950) na parte espalhada (*scattered*), representada por $I_{s,r,g}(\tau, \mu, \varphi)$, e na parte não-espalhada (*unscattered*), representada por $I_{u,r,g}(\tau, \mu, \varphi)$, assim

$$I_{r,g}(\tau, \mu, \varphi) = I_{u,r,g}(\tau, \mu, \varphi) + I_{s,r,g}(\tau, \mu, \varphi). \quad (2.8)$$

A solução para a componente não-espalhada é obtida considerando $b_{r,g} = 0$ na Eq. (2.4), sujeita às condições de contorno definidas pela Eq. (2.7) e às condições de

interface definidas pela Eq. (2.6) e, então, é resolvida a equação diferencial parcial resultante. Já para a componente espalhada $I_s(\tau, \mu, \varphi)$, na qual $b_{r,g} \neq 0$, inicialmente a função de fase é aproximada em sua série finita de polinômios de Legendre. Em seguida, utiliza-se a decomposição de Fourier em cossenos (CHANDRASEKHAR, 1950) a fim de eliminar a integração sobre o ângulo azimutal e, assim, gerar $L + 1$ equações íntegro-diferenciais (ETRs) sem a dependência de φ . A integral angular presente em cada equação íntegro-diferencial é aproximada pelo método da colocação. Esse método consiste em aproximar a integral angular por um esquema de quadratura de Gauss-Legendre. Essa aproximação produz um conjunto de N equações diferenciais ordinárias de primeira ordem para cada ângulo azimutal. A solução analítica para cada um desses conjuntos de equações diferenciais, nesta tese, é obtida através do método AS_N (CHALHOUB, 1997), o qual é baseado na decomposição espectral da matriz de espalhamento. Para a aproximação numérica, é utilizado o código PEESNA (CHALHOUB et al., 2003), o qual implementa o método AS_N .

A escolha do código PEESNA se justifica pelo trabalho de Chalhoub et al. (2003), que fez um estudo comparativo em termos de precisão e tempo de processamento (algoritmos sequenciais) para quatro métodos de solução da ETR: métodos LTS_N , P_N , AS_N e inserção invariante. Em termos de precisão e tempo de processamento, o melhor desempenho foi alcançado pelo método AS_N .

Ao se resolver o problema descrito neste capítulo, conhecendo as propriedades óticas inerentes (coeficientes de absorção e espalhamento e a função de fase), as condições iniciais e/ou de contorno, bem como a existência ou não de fontes internas de radiação, determina-se a radiação emergente na superfície da água (radiância emergente), em direções escolhidas *a priori*, após a interação da radiação incidente com o meio aquático. Esse tipo de abordagem para o problema caracteriza o que é chamado, em geral, de problema direto.

Uma limitação código e do modelo adotados está no fato de que estes não consideram a refração das partículas na entrada e na saída do meio, nem possíveis ondas, bolhas e espuma na superfície da água.

Por fim, os cálculos necessários para a aproximação da ETR em um sistema de equações diferenciais ordinárias (EDO) de primeira ordem são explicados no Apêndice A, e nesse apêndice também é obtida a solução analítica do sistema de EDO através do método AS_N . Os detalhes da expansão da função de fase em sua série de polinômios de Legendre são mostrados no Apêndice B.

3 REDES NEURAIS ARTIFICIAIS

Redes Neurais Artificiais (RNAs) são “sistemas” computacionais que agrupam um determinado número de neurônios artificiais e são inspiradas nas redes neurais biológicas. Os estudos iniciais sobre o cérebro humano datam de 1911, quando Ramón y Cajal introduziram a ideia de neurônios como a estrutura elementar do cérebro humano. Por outro lado, os primeiros estudos sobre neurônios artificiais datam da década de 1940 com o neurofisiologista Warren McCulloch e o matemático Walter Pitts (MCCULLOCH; PITTS, 1943), ambos da Universidade de Illinois. McCulloch e Pitts são considerados os precursores do neurônio artificial.

O passo seguinte ao trabalho inicial de McCulloch e Pitts foi dado por Rosenblatt (1958), quando propôs o Perceptron como uma nova abordagem para o modelo desenvolvido por McCulloch e Pitts. O Perceptron de Rosenblatt foi o primeiro modelo de RNA com aprendizagem supervisionada, e foi utilizado para a classificação de padrões linearmente separáveis. Ainda, o algoritmo de treinamento do Perceptron, proposto por Rosenblatt, converge em um número finito de passos. Para a demonstração do teorema da convergência do Perceptron consulte Haykin (2001).

Posteriormente, Widrow e Hoff (1960) propuseram a Adaline (*Adaptive Linear Network*), a qual utiliza a mesma topologia do Perceptron, no entanto, o treinamento é feito pelo algoritmo de mínimos quadrados, também chamado de regra delta. Em 1962, Widrow e Hoff propuseram a Madaline (*Many Adaptive Linear Network*) (WIDROW, 1962; HOFF, 1962). A Madaline foi uma das primeiras RNAs estruturada em camadas treináveis com múltiplos elementos adaptativos, e consiste em Adalines conectadas, *i.e.*, existem unidades (neurônios) ocultas entre as unidades de entrada e a unidade de saída. Essas unidades ocultas dão à Madaline capacidades computacionais não encontradas em redes sem unidades ocultas. Por outro lado, as unidades ocultas complicam o processo de treinamento (FAUSETT, 1994).

Em 1969, Marvin Minsky e Seymour Papert publicaram o livro *Perceptrons* (MINSKY; PAPERT, 1969), no qual mostraram que existiam limites fundamentais para o que poderia ser calculado por Perceptrons de camada única, afirmando, ainda, que não havia razões para supor que qualquer uma das limitações do Perceptron de camada única pudesse ser resolvida na versão de múltiplas camadas. Minsky e Papert usaram a porta lógica ou-exclusivo (XOR) para mostrar essa deficiência do Perceptron proposto por Rosenblatt.

Vista a influência de Minsky na comunidade científica, desde o lançamento do li-

vro até 1982, pouco se fez a respeito de Perceptrons. Por outro lado, em 1970, surgiram os mapas auto-organizáveis, os quais utilizam a técnica de aprendizagem competitiva. Em 1976, Willshaw e Von der Malsburg publicaram o primeiro artigo sobre a formação de mapas auto-organizáveis (WILLSHAW; VON DER MALSBURG, 1976). Já em 1980, baseado na aprendizagem competitiva, Grossberg estabeleceu um novo princípio de auto-organização conhecido como Teoria da Ressonância Adaptável (*Adaptive Resonance Theory – ART*) (HAYKIN, 2001; FAUSETT, 1994). Kohonen (1982), utilizando a aprendizagem competitiva, propôs um modelo de rede neural chamado de mapas auto-organizáveis (*Self-Organizing Map – SOM*). Devido à aprendizagem competitiva (não supervisionada) e às suas características, os mapas auto-organizáveis são amplamente aplicados em mineração de dados.

As pesquisas relacionadas à abordagem conexionista ressurgiram com o trabalho de Hopfield (1982), e em 1986 foi relatado por Rumelhart, Hinton e Williams o desenvolvimento e a descrição do algoritmo *Backpropagation* (RUMELHART et al., 1986). Com esse trabalho, os autores mostraram que Minsky e Papert tinham um ponto de vista bastante pessimista e equivocado sobre o Perceptron. Nesse mesmo ano, foi publicado um livro em dois volumes, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, editado por Rumelhart e McClelland, o qual teve grande influência na utilização do algoritmo *Backpropagation*, fazendo com que esse algoritmo se tornasse o mais popular método de treinamento para perceptrons de múltiplas camadas (HAYKIN, 2001).

3.1 Neurônio Biológico

Os neurônios são as células que mais se diferenciam das demais células do organismo, e apresentam estruturas que podem variar de acordo com as diferentes partes do cérebro que compõem. A Figura 3.1 mostra a estrutura de um típico neurônio encontrado no córtex motor cerebral. Os sinais chegam nesse neurônio através das sinapses localizadas nos dendritos neuronais, além dos sinais que chegam também pelo corpo celular, se propagam através do axônio até encontrar as sinapses com o neurônio seguinte. A sinapse é o ponto de confluência entre um neurônio e os dendritos do neurônio seguinte. Uma característica especial da maioria das sinapses é que o sinal normalmente se propaga apenas na direção anterógrada (do axônio de um neurônio precedente para os dendritos localizados em neurônios subsequentes), possibilitando, assim, que o sinal trafegue na direção necessária para executar as funções nervosas requeridas (GUYTON; HALL, 2006).

A atividade neuronal no sistema nervoso central é complexa e alguns processos (ex.:

memória) ainda não são bem entendidos pela comunidade científica. Para o escopo desta tese, a descrição dada é suficiente para fazer a analogia entre o funcionamento do neurônio biológico e o neurônio artificial, que é descrito a seguir. Outros detalhes acerca do funcionamento do neurônio biológico são encontrados no Apêndice C. A descrição apresentada nesta seção e no Apêndice C foram retiradas de Guyton e Hall (2006).

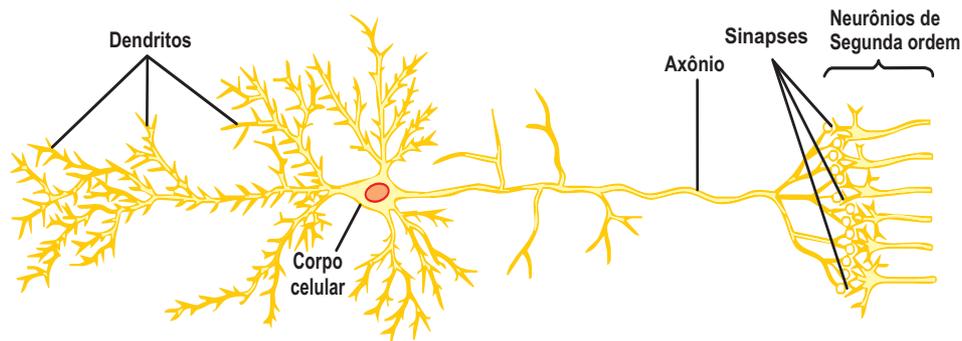


Figura 3.1 - Estrutura de um neurônio presente no encéfalo humano.
Fonte: Guyton e Hall (2006).

3.2 Neurônio Artificial

O modelo de neurônio artificial sugerido por McCulloch e Pitts (1943) é uma representação simplificada do complexo funcionamento do neurônio biológico. No entanto, o modelo matemático guarda o funcionamento básico do neurônio biológico e algumas de suas funções principais, como aprendizado e memória. Como será visto nas páginas seguintes, o aprendizado está associado ao período de treinamento e a memória está associada ao ajuste dos pesos entre as conexões.

O neurônio artificial de McCulloch e Pitts, Figura 3.2, é composto por um conjunto de entradas $\{x_i\}$, representando os estímulos que chegam em um terminal pré-sináptico, e um conjunto $\{w_i\}$, que representa o peso (ou força) da ligação entre os neurônios, em analogia às sinapses. No neurônio artificial, os estímulos que chegam no terminal pré-sináptico são multiplicados pelos respectivos pesos de cada ligação, podendo gerar um potencial excitatório (positivo) ou inibitório (negativo). O combinador linear Σ tem a finalidade de fazer a soma de todos os potenciais (excitatórios e inibitórios) que chegam ao corpo celular, através das sinapses, e do sinal que chega direto no corpo celular, indicado por b , e chamado de nível de *bias*, gerando um potencial de ação v . O valor b pode ser interpretado como um parâmetro que compensa alguma (ou várias) atividade(s) presente(s) no neurônio biológico e não

presente(s) no neurônio artificial. Portanto, durante o treinamento, é feito também o ajuste desse parâmetro a fim de “completar” essa deficiência no modelo artificial de neurônio¹. O potencial de ação v , também chamado de campo local induzido, que chega no neurônio é, representado matematicamente por

$$v = \sum_{i=1}^n x_i w_i + b,$$

e o valor y , que é disparado pelo neurônio através do axônio para o neurônio seguinte, é limitado por uma função de ativação ϕ , tal que,

$$y = \phi(v).$$

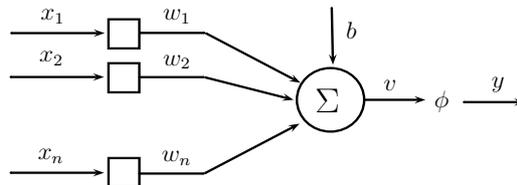


Figura 3.2 - Modelo do neurônio artificial de McCulloch e Pitts.

Durante a “concepção” do neurônio artificial, McCulloch e Pitts se dedicaram mais a descrever o modelo artificial do neurônio e a apresentar suas capacidades computacionais do que desenvolver técnicas de memorização (aprendizagem). Para McCulloch e Pitts, o neurônio artificial teria a capacidade do tudo ou nada, *i.e.*, se o potencial de ação v for positivo, a função de ativação ϕ dispara 1, caso contrário, dispara 0 (HAYKIN, 2001). Dessa forma, a função de ativação ϕ , em termos matemáticos, é expressa por

$$y = \phi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases}, \quad (3.1)$$

e o conhecimento adquirido, por alguma técnica de aprendizagem, ficaria armazenado no conjunto de valores $\{w_i, b\}$.

¹O parâmetro b tem uma grande importância no neurônio artificial, pois sem esse valor o Perceptron proposto por Rosenblatt (1958) não seria capaz de resolver a porta lógica NOR (negação da porta lógica OU). Note que do total (quatro) de possibilidades de pares de entrada para a porta NOR, três delas: (1, 0), (0, 1) e (1, 1) têm como saída 0, no entanto, para o par (0, 0) a saída é 1. Logo, sem o parâmetro b , fica impossível obter a saída 1 para o par de entrada (0, 0), desde que o potencial de ação seja definido como $v = \sum_{i=1}^n x_i w_i + b$ e a função de ativação sendo definida por $\phi(v) = 1$ se $v > 0$ e $\phi(v) = 0$ se $v \leq 0$.

O neurônio artificial guarda a essência do funcionamento de um neurônio biológico, no entanto, como foi comentado no início desta seção, não é um modelo capaz de representar todos os processos (químicos e elétricos) que ocorrem em um neurônio biológico. Desde a publicação do trabalho de McCulloch e Pitts, em 1943, muitos avanços aconteceram no que tange aos tipos de estruturas (topologias) para as RNAs, todavia, o modelo básico de neurônio artificial continua o mesmo.

3.2.1 Funções de Ativação

A função de ativação, como comentado, tem a finalidade de restringir, ou então, disparar um sinal através do axônio para o neurônio seguinte. A primeira função de ativação utilizada em modelos de neurônio artificial foi proposta por McCulloch e Pitts, Eq. (3.1). Com a “generalização” das RNAs e de suas aplicações em diversos tipos de problemas, surgiram outras funções de ativação, cada uma com características específicas. A seguir, são mostradas as funções de ativação que podem ser usadas em projetos de RNAs, lembrando que a escolha da função de ativação está ligada ao tipo de problema que a RNA tem a resolver.

Na classe das funções descontínuas, é possível citar: a função definida pela Eq. (3.1), que foi proposta por McCulloch e Pitts; a função sinal, que é definida matematicamente por

$$\phi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ -1 & \text{se } v < 0 \end{cases},$$

em que v é o potencial de ação; e a função logaritmo natural, definida por

$$\phi(v) = \begin{cases} \ln(1 + \gamma v) & \text{se } v \geq 0 \\ -\ln(1 - \gamma v) & \text{se } v < 0 \end{cases},$$

em que v é o campo local induzido e, $\gamma \in \mathcal{R}_+^*$ é um parâmetro que indica a taxa de crescimento da função. Na classe das funções contínuas, há a função sigmoide (ou logística) e a função tangente hiperbólica, definidas, respectivamente, por

$$\phi(v) = \frac{1}{1 + e^{-\gamma v}},$$

e

$$\phi(v) = \frac{1 - e^{-\gamma v}}{1 + e^{-\gamma v}},$$

em que $\gamma \in \mathcal{R}_+^*$ está associado à inclinação da curva, *i.e.*, a curva é mais íngreme quando $\gamma > 1$ e menos íngreme quando $0 < \gamma < 1$. Ainda há a função linear definida

por

$$\phi(v) = \gamma v,$$

em que $\gamma \in \mathcal{R}^*$ e v é o campo local induzido nas últimas três equações. A Figura 3.3 ilustra cada uma dessas funções de ativação. Observe que as funções logaritmo natural e linear não apresentam limitantes máximo/mínimo no eixo das ordenadas como as demais.

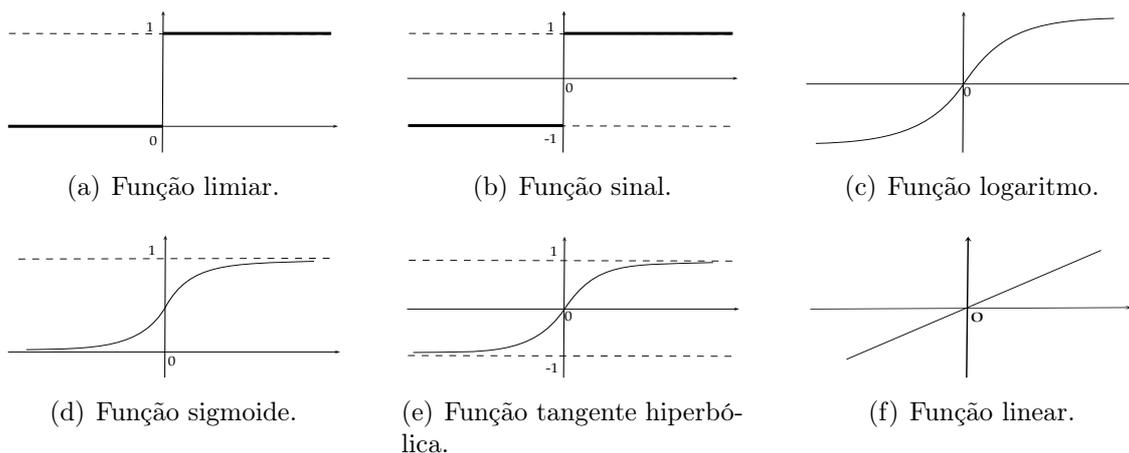


Figura 3.3 - Representação gráfica de possíveis funções de ativação que podem ser utilizadas no neurônio artificial.

3.3 Características das RNAs

O fato de o neurônio artificial ser baseado no funcionamento do neurônio biológico garante que o neurônio artificial apresente algumas características do neurônio biológico. Dessa forma, uma rede de neurônios artificiais (redes neurais artificiais) é capaz de emular, em algum nível de semelhança, o funcionamento do cérebro humano.

Apesar da semelhança entre redes neurais artificiais e as biológicas, a possibilidade de reproduzir o complexo funcionamento do cérebro humano em um computador (pode) está(ar) longe de ser alcançado, mesmo o cérebro trabalhando na ordem dos milissegundos e as portas lógicas de silício (computador) na ordem dos nanosegundos. O diferencial do cérebro reside no fato de que ele é um sistema complexo, distribuído, adaptativo, e com acesso eficiente à memória, propriedades essas difíceis de se reproduzir em (ou por) um computador atual (baseado na arquitetura de von Neumann). No entanto, com essa estrutura baseada no funcionamento do cérebro, as RNAs passam a ter muitas vantagens, como por exemplo: capacidade de aprendizado, auto-organização, tolerância a falhas, flexibilidade e produção de resultados

em tempo real.

A aprendizagem das RNAs é feita por algoritmos específicos, como será visto a seguir, que atuam durante o treinamento (fase escolar) das RNAs, ensinando-as a fazer o mapeamento dos padrões de entradas aos de saídas, para que, posteriormente, elas possam produzir respostas baseando-se no que aprenderam por meio desses algoritmos. A auto-organização ocorre durante o treinamento, pois cada sinapse se ajusta de modo a produzir sua própria representação interna, mas que será passada à próxima conexão, e assim sucessivamente, até obter a saída desejada. A tolerância a falhas ocorre pelo fato de uma RNA armazenar informação redundante, e isso lhe dá a capacidade de continuar a responder mesmo que algum dado de entrada esteja danificado. A flexibilidade é um tanto quanto restrita, pois mesmo sendo tolerante a falhas, uma RNA não consegue responder a dados de entrada que apresentem “grandes” mudanças em relação àqueles com os quais ela foi treinada. Por fim, o fato de as RNAs terem uma estrutura paralela faz com que elas sejam capazes de retornar respostas em tempo real. Outras vantagens e desvantagens podem ser encontradas em Fausett (1994), Bishop (1995), Haykin (2001) e Bishop (2006).

3.4 Perceptron de Múltiplas Camadas

Antes de descrever o Perceptron de Múltiplas Camadas (PMC) (*Multilayer Perceptron – MLP*), é importante apresentar alguns comentários acerca do Perceptron de Rosenblatt (ROSENBLATT, 1958), o qual é o elemento básico na formação do PMC, e a Adaline de Widrow e Hoff (1960).

O Perceptron proposto por Rosenblatt é a forma mais simples de uma RNA (com apenas um único neurônio) capaz de resolver problemas linearmente separáveis, *i.e.*, existe um hiperplano capaz de separar os padrões em duas classes (HAYKIN, 2001). Veja Figura 3.4. Em seu modelo de RNA, Rosenblatt considera apenas um neurônio com pesos sinápticos e nível de *bias* ajustáveis, e apresenta, ainda, um algoritmo para fazer o ajuste dessas variáveis livres², sendo o primeiro algoritmo de aprendizagem supervisionada. A função de ativação utilizada por Rosenblatt segue o modelo de função de ativação do neurônio de McCulloch e Pitts, Eq. (3.1).

A Adaline tem a mesma estrutura do Perceptron, *i.e.*, é composta de um único neurônio com pesos sinápticos e *bias* ajustáveis, no entanto, a diferença entre esses

²Entende-se por variáveis de uma rede o conjunto que contém todos os pesos sinápticos e níveis de *bias* $\{w_j, b\}$ e, eventualmente, como será mostrado na Seção 3.6, o parâmetro γ presente em cada função de ativação.

dois modelos de RNA está no modo como cada um deles é treinado. Enquanto o Perceptron ajusta o valor de suas variáveis livres apenas quando produz uma resposta errada, a Adaline, que em geral utiliza a função linear como função de ativação, atualiza suas variáveis mesmo quando uma resposta certa é obtida. Isso ocorre porque o algoritmo de treinamento da Adaline é baseado na diferença de mínimos quadrados (regra delta) entre as saídas desejadas e as respostas produzidas pela rede.

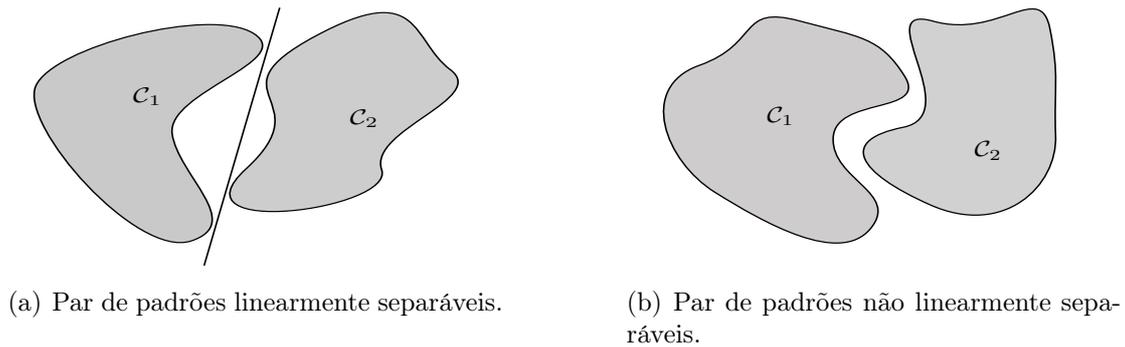


Figura 3.4 - Conjunto de padrões linearmente separáveis e não separáveis.
 Fonte: Adaptada de Haykin (2001).

Um PMC consiste na “união” de vários Perceptrons, de modo a formar uma rede, como ilustra a Figura 3.5. Essa rede é formada por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. No PMC, o sinal de entrada, representado pelo conjunto $\{x_i\}$, é propagado para frente, camada por camada, até que atinja a camada de saída, produzindo um conjunto de respostas $\{y_i\}$. Cada neurônio, seja ele da camada oculta ou da camada de saída, tem uma função de ativação, que pode ser diferente para cada neurônio e/ou camada. Ainda, na Figura 3.5, os quadrados (camada de entrada) não executam nenhuma computação e são utilizados apenas para ilustrar os dados de entrada; os círculos representam cada um dos neurônios que compõem a rede. Os pesos são ilustrados pelas setas que ligam os neurônios entre as camadas.

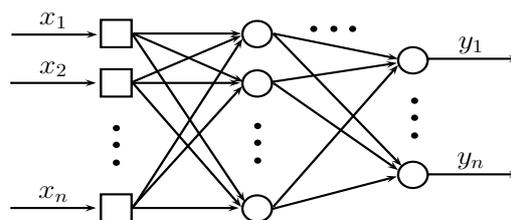


Figura 3.5 - Representação gráfica de um perceptron de múltiplas camadas.

O número de entradas e o número de saídas de uma rede é definido pelo problema que se pretende resolver, no entanto, o número de camadas ocultas e o de neurônios ocultos é definido de maneira empírica (*ad hoc*), e se baseia no conhecimento prévio do problema e na experiência do pesquisador. O número de camadas ocultas é geralmente definido pelas características do problema para o qual se está buscando solução, porém, é interessante levar em conta o trabalho de Cybenko (1989). Nesse trabalho, Cybenko diz que não são necessárias mais do que duas camadas ocultas em um PMC, pois uma camada oculta pode aproximar qualquer função contínua, e duas camadas ocultas são suficientes para aproximar qualquer função matemática. Em seu estudo, Cybenko (1989) considera que a função de ativação de cada neurônio é a função sigmoide.

Metodologias alternativas, a fim de determinar o número de camadas ocultas, número de neurônios em cada camada, funções de ativação, valor da taxa de aprendizagem e da constante de momento, tem sido propostas no trabalhos de Anochi et al. (2012) e Sambatti et al. (2012). Nesses trabalhos, os autores utilizam o Algoritmo de Colisão Múltipla de Partículas (*Multi-Particle Collision Algorithm – MPCA*) (LUZ et al., 2008), a fim de determinar a topologia adequada para a RNA utilizada.

Por fim, o PMC, associado ao algoritmo *Backpropagation*, que será abordado adiante, tem sido aplicado com sucesso na resolução de vários problemas difíceis em diversas áreas da ciência, em especial, na solução de problemas inversos ligados à transferência radiativa, ótica hidrológica e cor do oceano (SCHILLER; DOERFFER, 1999; BOKAR, 1999; GROSS et al., 2000; RICHARDSON et al., 2002; SOEIRO et al., 2004b; SOEIRO et al., 2004a; OSAWA et al., 2005; SOEIRO; SILVA NETO, 2006; CHALHOUB et al., 2007; OLIVEIRA, 2010; OLIVEIRA et al., 2010; CORTIVO et al., 2010; CORTIVO et al., 2012b; CORTIVO et al., 2012c).

3.5 Aprendizagem das Redes Neurais Artificiais

Aprendizagem é o processo pelo qual as RNAs adquirem um determinado conhecimento por meio de um algoritmo de aprendizagem, com ou sem a presença de um professor. Esse é o passo mais importante e difícil na utilização de RNAs para a resolução de problemas.

A aprendizagem, entre outras, é uma das principais características de uma rede neural pois, devido ao modo como as RNAs foram concebidas, elas são capazes de incorporar conhecimento a partir do ambiente em que estão inseridas, por meio do ajuste adequado de suas variáveis. Quando esse ajuste é feito de modo adequado,

é de se esperar que a rede consiga reproduzir esse ambiente de modo eficiente, ou então, pelo menos, produza uma resposta consistente.

Dentre as regras de aprendizagem, é possível citar algumas como: a aprendizagem por correção de erro, aprendizagem baseada em memória, aprendizagem hebbiana, aprendizagem competitiva, aprendizagem de Boltzmann, aprendizagem supervisionada, aprendizagem não-supervisionada, entre outras. A aprendizagem supervisionada, muitas vezes, é conhecida como aprendizagem com professor e suas aplicações e características serão tratadas a seguir. Para as demais regras citadas, sugere-se Kohonen (1989), Fausett (1994), Bishop (1995), Haykin (2001) e Bishop (2006).

3.5.1 Aprendizagem Supervisionada

A aprendizagem supervisionada é caracterizada pela presença de um “professor” que tem o conhecimento do “ambiente” no qual a rede está inserida, e esse professor tem a função de “passar” o conhecimento para a RNA, por meio de algum algoritmo de aprendizagem. Em geral, esse ambiente é representado por um conjunto de entradas (chamado de padrões de entrada) que estão associadas a determinadas saídas (chamado de padrões de saída). Esses conjuntos (de entradas e saídas), em geral, são subdivididos em três outros conjuntos, com uma quantidade de padrões em cada conjunto, geralmente definida pelas características do problema, ou então, seguindo as estratégias sugeridas por Haykin (2001). Esses três conjuntos são chamados de: conjunto de treinamento, conjunto de teste e conjunto de validação.

O conjunto de treinamento, como o próprio nome diz, é utilizado para o treinamento da rede. O conjunto de teste é utilizado a fim de testar o desempenho da rede durante o processo de aprendizagem (treinamento) e, em geral, é utilizado quando se emprega o processo de parada antecipada, também chamado de validação cruzada. E, por fim, o conjunto de validação é utilizado após a rede estar completamente treinada, com o intuito de avaliar seu desempenho em padrões que não estiveram presentes no treinamento e no processo de validação cruzada. Outro fator importante, que afeta o treinamento, é a escolha do conjunto de treinamento, *i.e.*, o conjunto de treinamento deve ser escolhido de modo que seja o mais representativo possível do problema.

Durante o processo de aprendizagem da rede, o professor apresenta uma determinada entrada (vetor) do conjunto de treinamento à rede, a qual processa essa informação e devolve uma saída (resposta da rede). Visto que o professor tem o conhecimento de qual deveria ter sido a resposta gerada por parte da rede, Figura 3.6, ele faz com que a rede compare as respostas e, assim, produza um sinal de erro. Esse valor de

erro (caso tenha ocorrido) é injetado no sistema de aprendizagem que, por sua vez, calcula as atualizações que devem ser aplicadas às variáveis livres da rede, de modo a minimizar esse sinal de erro.

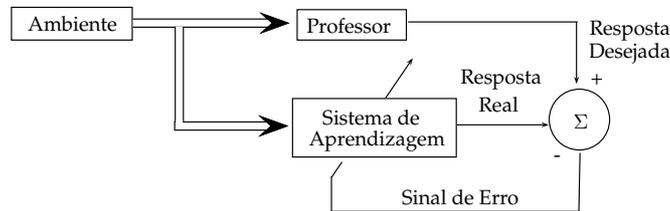


Figura 3.6 - Diagrama de blocos a aprendizagem supervisionada.
 Fonte: Adaptada de Haykin (2001).

A injeção do sinal de erro no sistema de aprendizagem pode ser feita de duas formas: a) modo *online*, o qual consiste na injeção do sinal de erro gerado por cada padrão de treinamento apresentado à rede; e b) modo *batch*, o qual consiste na injeção do sinal de erro total cometido após a apresentação de todos os padrões de treinamento à rede.

Cada vez que todos os padrões de treinamento são apresentados à rede, tanto no modo *online* quanto no modo *batch*, é dito que se passou uma época. Para o treinamento de uma rede são usadas tantas épocas quantas forem necessárias para que a rede aprenda e produza bons resultados. Outro fator importante, e que pode ser utilizado se o treinamento for feito no modo *online*, é a apresentação aleatória dos padrões de treinamento de uma época para outra.

Além da característica de aprendizagem, outras duas características intrínsecas das redes, que são aprimoradas durante o aprendizado, são a tolerância a falhas e a capacidade de generalização. A tolerância a falhas é caracterizada pela habilidade da rede em produzir respostas corretas, mesmo quando são apresentados dados de entrada “incompletos” ou ruidosos. A capacidade de generalização é caracterizada pela habilidade adquirida pela rede, durante o aprendizado, em reconhecer entradas que não lhe foram apresentadas durante o treinamento. Essa capacidade é melhorada a partir do procedimento chamado de validação cruzada.

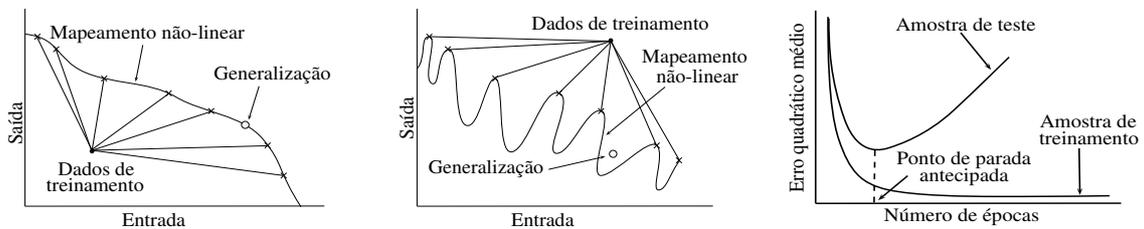
3.5.2 Validação Cruzada

A validação cruzada consiste em utilizar o conjunto de teste, durante o processo de treinamento, *i.e.*, após um determinado número de épocas, o treinamento é pausado

e, então, são apresentados à rede os padrões desse conjunto (teste) a fim de avaliar o seu desempenho. Com o intuito de melhorar e fazer sentido a validação cruzada, é importante que os padrões que estão no conjunto de teste não estejam presentes no conjunto de treinamento.

A validação cruzada é uma boa estratégia a ser adotada para determinar o melhor “momento” do treinamento, pois assegura que a rede terá uma máxima capacidade de generalização, eliminando o *overfitting*. O *overfitting* é um problema que surge quando a rede é treinada em demasia, *i.e.*, ela adquire a capacidade de reconhecer todos os exemplos do treinamento, porém não é capaz de reconhecer um padrão que esteve fora deste.

A Figura 3.7(a) ilustra uma rede em que os dados foram ajustados adequadamente, *i.e.*, a rede apresenta uma excelente capacidade de generalização, o que não ocorre na Figura 3.7(b) devido ao excesso de ajuste dos dados. A Figura 3.7(c) ilustra o comportamento do erro quadrático médio para o conjunto de treinamento e para o conjunto de teste. Observa-se que à medida que o número de épocas avança, o erro do conjunto de treinamento (amostra de treinamento) diminui, no entanto, o erro sobre o conjunto de teste aumenta. O ponto de parada antecipada é escolhido quando o erro sobre o conjunto de teste é mínimo.



(a) Dados ajustados adequadamente (boa generalização).

(b) Dados ajustados em excesso (generalização pobre, *overfitting*).

(c) Parada antecipada baseada na validação cruzada.

Figura 3.7 - Ilustração da generalização de uma rede e o comportamento do erro quadrático médio no procedimento de validação cruzada.

Fonte: Adaptada de Haykin (2001).

3.6 Algoritmos de Aprendizagem

Um algoritmo de aprendizagem é um conjunto de regras bem definidas com a finalidade de resolver um problema de aprendizagem. Não existe apenas um algoritmo de aprendizagem e sim uma variedade deles, cada um para um determinado tipo

de RNA, com suas vantagens e desvantagens específicas. Nesta tese, é dada ênfase aos algoritmos *Backpropagation*, *Backpropagation* generalizado e quasi-Newton, que são algoritmos do tipo supervisionado, os quais serão tratados nas três subseções seguintes.

É claro que cada uma das regras de aprendizagem citadas na Seção 3.5 tem seu algoritmo específico. A descrição de cada um dos algoritmos associados às regras de aprendizagem se tornará longa; além disso, os algoritmos associados à aprendizagem não-supervisionada fogem do escopo desta tese, visto que neste trabalho é utilizado um algoritmo de treinamento do tipo supervisionado.

No que tange aos algoritmos de aprendizagem supervisionada, o mais popular e responsável por alavancar a utilização de RNAs do tipo PMC na busca de soluções para vários problemas, em diversas áreas da ciência e engenharia, foi o algoritmo de treinamento *Backpropagation*, proposto inicialmente por Rumelhart et al. (1986), devido à sua facilidade de implementação computacional. No entanto, esse algoritmo apresenta algumas desvantagens, como: ter uma convergência lenta, ficar preso em mínimos locais, ser sensível aos valores iniciais das variáveis livres (pesos sinápticos e *bias*) (FAUSETT, 1994; BISHOP, 1995; HAYKIN, 2001; BISHOP, 2006).

Visto a importância do treinamento de uma RNA e as desvantagens intrínsecas do *Backpropagation*, muitos cientistas se dedicaram (e se dedicam) à busca de melhorias para o *Backpropagation*, ou, até mesmo, outros algoritmos de treinamento baseados em outros métodos de otimização.

Uma das alternativas encontradas para superar a sensibilidade aos valores iniciais de suas variáveis livres foi proposta por Lee et al. (1991). Esse autor sugere uma reinicialização automática e aleatória para os pesos sinápticos, durante o treinamento, quando a convergência se torna muito lenta. Em Cho et al. (1991), é possível encontrar algoritmos *Backpropagation* que empregam a técnica da reinicialização automática dos pesos, bem como um parâmetro de ganho na função de ativação (uma espécie de *Backpropagation* generalizado). Esses algoritmos, segundo Cho et al. (1991), convergem mais rapidamente quando comparados aos propostos por Lee et al. (1991), no entanto, o uso do parâmetro de ganho na função de ativação pode fazer com que o processo de treinamento não seja convergente em alguns casos.

Tang e Kwan (1993) analisam o efeito do número de parâmetros com relação à velocidade de convergência e à capacidade de generalização e, então, propõem mudanças no algoritmo *Backpropagation* de forma a obter vantagens, tanto na velocidade de

convergência, quanto na capacidade de generalização. A proposta de Tang e Kwan (1993) é considerar que níveis de *bias* e parâmetros das funções de ativação sejam adaptáveis de uma maneira similar aos pesos.

Além desses trabalhos descritos, é possível citar Jacobs (1988), Fahlman (1988), Tallaneare (1990) e Minai e Williams (1990), que propõem modificar o valor da taxa de aprendizagem ou da constante de momento durante o treinamento. Além desses, há também Chun e Bingxue (2000) e Gill et al. (2010), que propõem a hibridização do *Backpropagation* com Algoritmos Genéticos (HOLLAND, 1975).

Algoritmos de treinamento para PMCs baseados em métodos de otimização de segunda ordem, como por exemplo, método de Levenberg-Marquardt (LEVENBERG, 1944; MARQUARDT, 1963), ou então, modificações deste, podem ser encontrados em Hagan e Menhaj (1994), Lera e Pinzolas (1998), Wilamowski et al. (2001) e Wilamowski e Yu (2010). Mohamad et al. (2010) fazem uma comparação entre o algoritmo de Levenberg-Marquardt e o método do Gradiente Conjugado (BURDEN; FAIRES, 2003) utilizando um problema de detecção de nódulos cancerígenos.

Outros métodos determinísticos utilizados para treinar PMCs são os métodos quasi-Newton (DENNIS; MOREÉ, 1977), sendo o mais popular o método quasi-Newton BFGS, desenvolvido por Broyden, Fletcher, Goldfarb e Shanno. Esse método, por ser de segunda ordem, apresenta uma certa robustez quando comparado ao *Backpropagation*. Watrous (1988) faz uma comparação entre BFGS e *Backpropagation* e afirma que o método quasi-Newton requer menos iterações para treinar a rede, *i.e.*, converge mais rapidamente para a solução, no entanto, em cada iteração, é necessário atualizar a matriz inversa da Hessiana. Setiono e Hui (1995) afirmam que o método quasi-Newton é superior em número de iterações e qualidade das respostas quando comparado ao *Backpropagation*.

Modificações no método quasi-Newton, considerando treinamento de PMCs, foram propostas a fim de simplificar a aproximação da matriz inversa da Hessiana a cada iteração, e, dentre os trabalhos acerca dessas modificações, é possível citar: Parker (1987), que desenvolve fórmulas para a atualização da matriz inversa da Hessiana em implementações paralelas; Robitaille et al. (1996), que propõem modificações na aproximação clássica da Hessiana inversa, as quais consistem em uma nova aproximação da Hessiana, que exclui alguns termos de segunda ordem, levando em conta a estrutura da rede; Bortoletti et al. (2003) propõem uma nova classe de métodos quasi-Newtonianos para treinamento de PMCs cuja característica consiste em utilizar uma forma generalizada do método tipo BFGS, envolvendo uma família

adequada de álgebras de matrizes.

A seguir, são apresentados os três algoritmos de aprendizagem para treinamento supervisionado, citados no início desta seção: o *Backpropagation*, o *Backpropagation* generalizado e o método quasi-Newton. Descrições acerca do algoritmo *Backpropagation* e do quasi-Newton podem ser encontrados em Haykin (2001) e Bishop (1995). Descrições resumidas sobre o *Backpropagation* generalizado podem ser encontradas nos artigos de Cho et al. (1991) e Tang e Kwan (1993), no entanto, sem o nível de detalhes apresentado no Apêndice D.

3.6.1 Algoritmo *Backpropagation* Tradicional

No *Backpropagation* “tradicional” são consideradas como variáveis da rede somente os pesos sinápticos e os níveis de *bias*. O parâmetro γ , presente nas funções de ativação mostradas na Seção 3.2.1, é considerado constante e/ou igual a um, *i.e.*, todas as funções de ativação apresentam o “mesmo formato”.

Como comentado anteriormente, o algoritmo *Backpropagation* é um algoritmo de aprendizagem supervisionada e é o algoritmo mais popular para treinamento de PMCs. Basicamente, esse algoritmo consiste em dois passos: a propagação e a retropropagação.

Na propagação, os padrões de entrada (sinais) são aplicados nas entradas da rede, e então, são propagados para frente, camada por camada, até atingirem a camada de saída. Ao chegarem na camada de saída, a rede produz uma resposta e, como o treinamento é do tipo supervisionado, é possível calcular a magnitude do erro cometido pela rede (caso tenha ocorrido). Durante o processo de propagação, nenhuma alteração é feita nas variáveis da rede.

Na retropropagação, daí o nome *Backpropagation*, o erro cometido pela rede que foi “medido” nos neurônios da camada de saída é propagado para trás. Nesse processo de retropropagação é que as variáveis livres da rede são ajustadas de acordo com alguma regra de correção de erro. A regra de correção de erro é baseada no erro global e no erro local de cada neurônio.

Em geral, é utilizada a regra delta ou a regra delta generalizada para efetuar a correção nas variáveis livres da rede. A atualização, pela regra delta na $(n + 1)$ -ésima iteração para os pesos sinápticos que ligam uma camada $\ell - 1$ à camada ℓ , e para os

j -ésimos níveis de *bias* dos neurônios da camada ℓ , é definida, respectivamente, por

$$w_{kj}^{n+1} = w_{kj}^n + \eta_w \delta_j^n y_k^n, \quad (3.2)$$

e

$$b_j^{n+1} = b_j^n + \eta_b \delta_j^n, \quad (3.3)$$

em que w_{kj}^n é o valor do peso sináptico, na n -ésima iteração, que liga o neurônio k da camada $\ell - 1$ ao neurônio j da camada ℓ , δ_j^n é o valor do gradiente local, na n -ésima iteração, do j -ésimo neurônio da camada ℓ , y_k^n é o valor do sinal de saída do neurônio k na camada $\ell - 1$, na n -ésima iteração, e η_w e η_b são as taxas de aprendizagem e que podem ser iguais.

Se as correções sobre os pesos sinápticos e sobre os níveis de *bias* forem dadas pela regra delta generalizada, então, as Eqs. (3.2) e (3.3) passam a ser escritas como

$$w_{kj}^{n+1} = w_{kj}^n + \alpha_w w_{kj}^{n-1} + \Delta w_{kj}^n, \quad (3.4)$$

e

$$b_j^{n+1} = b_j^n + \alpha_b b_j^{n-1} + \Delta b_j^n, \quad (3.5)$$

em que α_w e α_b são as constantes de momento, $\Delta w_{kj}^n = \eta_w \delta_j^n y_k^n$ e $\Delta b_j^n = \eta_b \delta_j^n$. Para os detalhes acerca do formalismo matemático envolvido no algoritmo *Backpropagation* “tradicional”, consulte a Seção D.1 do Apêndice D.

3.6.2 Algoritmo *Backpropagation* Generalizado

A diferença entre o *Backpropagation* “tradicional” e o *Backpropagation* generalizado reside apenas na atualização das variáveis livres da rede durante o treinamento. Enquanto o *Backpropagation* “tradicional” faz a atualização somente dos pesos sinápticos e dos níveis de *bias*, o *Backpropagation* generalizado faz a atualização também do parâmetro γ presente em cada função de ativação.

A vantagem de se fazer a atualização do parâmetro da função de ativação, durante o treinamento, reside no fato de que isso possibilita a cada neurônio ter uma função de ativação “específica” e, com isso, possibilitar que os neurônios se especializem em regiões específicas do conjunto de padrões. Em geral, são necessárias menos épocas para o treinamento da rede. Por outro lado, a atualização do parâmetro γ durante o treinamento pode fazer com que este se torne negativo, ou então, atinja ordens de grandeza elevadas, podendo causar *overflow*. Esse problema é eliminado ao se adotar

restrições, *i.e.*, valores máximos/mínimos que esses parâmetros podem atingir, como será descrito na próxima subseção.

No *Backpropagation* generalizado, as correções feitas nos pesos sinápticos e nos níveis de *bias*, segundo a regra delta, são expressas pelas Eqs. (3.2) e (3.3), respectivamente, e as correções em cada parâmetro das funções de ativação são definidas por

$$\gamma_k^{n+1} = \gamma_j^n + \eta_\gamma \varrho_j^n, \quad (3.6)$$

em que ϱ_j^n é o j -ésimo gradiente local, com relação ao parâmetro γ_j^n , do neurônio j da camada ℓ , e η_γ é a taxa de aprendizagem. Se for utilizada a regra delta generalizada, as atualizações nos pesos sinápticos e nos níveis de *bias* serão definidas pelas Eqs. (3.4) e (3.5). Já as atualizações no parâmetro das funções de ativação são definidas por

$$\gamma_k^{n+1} = \gamma_j^n + \alpha_\gamma \gamma_j^{n-1} + \Delta \gamma_k^n, \quad (3.7)$$

em que α_γ é a constante de momento e $\Delta \gamma_k^n = \eta_\gamma \varrho_j^n$. Para ambas as Eqs. (3.6) e (3.7) podem ser aplicadas as restrições apresentadas na Eq. (3.9) e/ou (3.10). Para detalhes acerca do formalismo matemático envolvido no algoritmo *Backpropagation* generalizado, consulte a Seção D.2 do Apêndice D.

3.6.3 Algoritmo quasi-Newton

O desenvolvimento da estratégia de treinamento utilizando o algoritmo quasi-Newton, apresentada em Cortivo et al. (2012b), surgiu a partir da dificuldade de se treinar RNAs com o *Backpropagation* “tradicional”, visando à resolução de problemas em ótica hidrológica inversa³.

A estratégia original, apresentada em Cortivo et al. (2012b), considera como variáveis apenas os pesos sinápticos e os níveis de *bias*. Posteriormente, essa estratégia foi modificada de modo a considerar como variáveis também o parâmetro de cada função de ativação. O trabalho resultante (CORTIVO et al., 2012a) foi apresentado no WCCI/IJCNN 2012 (*World Congress on Computational Intelligence/International Joint Conference on Neural Networks*) realizado em Brisbane, Austrália. Outro trabalho resultante da aplicação dessa estratégia modificada foi publicado na revista TEMA (Tendências em Matemática Aplicada e Computacional) (CORTIVO et al., 2012c).

³A ideia surgiu durante um seminário do curso de Computação Aplicada (CAP) do INPE, em 2010, ministrado pelo professor Dr. José Demisio Simões da Silva (*in memoriam*).

A estratégia desenvolvida considera o problema de aprendizagem como um problema de otimização, para o qual se define um funcional $\mathcal{J}(\cdot)$ de diferenças quadráticas entre os padrões de saída utilizados no treinamento e as respostas obtidas pela rede durante o treinamento. Assim,

$$\mathcal{J}(\mathbb{E}_t, \mathbb{W}, \mathbb{B}, \Gamma) = \min \frac{1}{2} \sum_{p=1}^{N_p} \sum_{k=1}^{N_s} \|d_{kp}(\mathbb{E}_t) - y_{kp}(\mathbb{E}_t, \mathbb{W}, \mathbb{B}, \Gamma)\|_2^2, \quad (3.8)$$

em que \mathbb{E}_t , \mathbb{W} , \mathbb{B} e Γ são matrizes que contêm, respectivamente, os padrões de entrada para o treinamento, os pesos sinápticos, os níveis de *bias* e os parâmetros das funções de ativação. Ainda, N_s representa o número total de neurônios na camada de saída, N_p representa o número total de padrões de treinamento, $d_{kp}(\cdot)$ e $y_{kp}(\cdot, \mathbb{W}, \mathbb{B}, \Gamma)$ representam, respectivamente, cada entrada das matrizes que contêm os padrões de saída para o treinamento e as saídas calculadas pela rede. Nessa formulação o treinamento é supervisionado e do tipo *batch*.

Visto que os valores γ de cada função de ativação são variáveis, é necessário incluir restrições a estas variáveis a fim de evitar que atinjam ordem de grandeza elevada, ou então, se tornem negativas. Diante disso, durante o processo de otimização (treinamento), deve-se buscar valores de γ que estejam contidos no intervalo definido por

$$0 < \gamma \leq \gamma_{\max}, \quad (3.9)$$

em que γ_{\max} é um valor máximo pré-definido, a fim de evitar que este parâmetro atinja uma ordem de grandeza elevada, tornando a contribuição do termo exponencial desprezível, ou então, gerar um *underflow*, quando utilizadas as funções de ativação sigmoide e/ou tangente hiperbólica, ou ainda, gerar um erro quando utilizada a função logarítmica. Se for utilizada a função de ativação linear, a restrição dada pela Eq. (3.9) pode ser substituída por

$$\gamma_{\min} \leq \gamma \leq \gamma_{\max}, \quad (3.10)$$

em que $\gamma_{\{\min, \max\}} \in \mathcal{R}^*$ são valores mínimos e máximos, respectivamente, pré-definidos a fim de evitar um *overflow*.

Sabe-se que o treinamento de uma rede é um problema de otimização não-linear e pode envolver uma função com múltiplos mínimos locais. Diante disso, é difícil garantir que será encontrado um mínimo global. Portanto, parece ser atraente a ideia de reduzir o espaço solução e, então, buscar por um mínimo local dentro desse

subespaço, desde que seja possível encontrar um conjunto solução capaz de produzir bons resultados. A redução do espaço de busca é obtido colocando restrições nas demais variáveis a serem otimizadas, *i.e.*, nos pesos sinápticos e nos níveis de *bias*. Assim, durante o processo de treinamento, são buscados valores para a matriz \mathbb{W} e para matriz/vetor \mathbb{B} de modo que esses satisfaçam as restrições

$$\begin{aligned} w_{\min} &\leq w \leq w_{\max} \\ b_{\min} &\leq b \leq b_{\max} \end{aligned}, \quad (3.11)$$

em que $w_{\{\min, \max\}} \in \mathcal{R}$ e estão associados às restrições dos pesos, $b_{\{\min, \max\}} \in \mathcal{R}$ e estão associados às restrições dos níveis de *bias*.

Uma vantagem que surge com a inclusão dessas restrições nos pesos e nos níveis de *bias*, além da citada anteriormente, é o fato de que é possível evitar que esses parâmetros também assumam ordens de grandeza elevadas e, assim, evitar uma possível saturação dos neurônios, fato esse não desejado pois prejudica o treinamento. Portanto, durante o treinamento da rede, o método utilizado busca um mínimo para o funcional definido na Eq. (3.8), sujeito às condições expressas pelas Eqs. (3.9) e/ou (3.10) e/ou (3.11).

O algoritmo quasi-Newton, utilizado nos trabalhos de Cortivo et al. (2012b), Cortivo et al. (2012a) e Cortivo et al. (2012c), e que também será utilizado nesta tese, está implementado na sub-rotina E04UCF da biblioteca NAG (*Numerical Algorithms Group*) (NAG, 1995). Optou-se pela utilização desta biblioteca em função da sua robustez para a resolução de problemas de otimização não-lineares e pela possibilidade da inclusão das restrições citadas.

Para o processo de treinamento, as variáveis a serem otimizadas (pesos, *bias* e parâmetro das funções de ativação) são organizadas em um vetor \vec{X} e, então, são atribuídos valores randômicos a essas variáveis. Durante esse processo (treinamento), a biblioteca procura um valor \vec{X}^* para $\vec{X} = [\mathbb{W} \ \mathbb{B} \ \mathbb{A}]^T$, de modo que

$$\nabla \mathcal{J}(\cdot, \vec{X}^*) = \vec{0}, \quad (3.12)$$

e ainda, leva em consideração que \vec{X}^* satisfaz as condições expressas pelas Eqs. (3.9) e/ou (3.10) e/ou (3.11). Agora, se $\vec{X}_{j+1} = \vec{X}_j + \Delta \vec{X}_j$ aproxima o ponto \vec{X}^* na $(j + 1)$ -ésima iteração, então, fazendo a aproximação de primeira ordem em série

de Taylor para a Eq. (3.12) em torno do ponto \vec{X}_{j+1} , obtém-se

$$\vec{X}_{j+1} = \vec{X}_j + \mathbb{H}^{-1} \nabla \mathcal{J}(\cdot, \vec{X}_j),$$

em que \mathbb{H}^{-1} é a matriz inversa da Hessiana, em $\vec{X} = \vec{X}_j$. Na prática, a inversa \mathbb{H}^{-1} não é calculada de forma exata, e sim, de uma forma aproximada. Para isso, considere que \vec{X}_j é um ponto inicial e \mathbb{H}_j^{-1} é uma aproximação da inversa da Hessiana que respeita as condições definidas em Dennis e Moré (1977). Então, primeiramente, o algoritmo calcula $\nabla \mathcal{J}(\cdot, \vec{X}_j)$; em seguida, calcula uma direção de busca definida pela expressão $d_j = -\mathbb{H}_j^{-1} \nabla \mathcal{J}(\cdot, \vec{X}_j)$ e, posteriormente, atualiza \vec{X}_j de modo que $\vec{X}_{j+1} = \vec{X}_j + \alpha_j d_j$, em que α_j é o tamanho do passo. Por fim, atualiza \mathbb{H}_j^{-1} através da expressão

$$\mathbb{H}_{j+1}^{-1} = \mathbb{H}_j^{-1} - \frac{\mathbb{H}_j^{-1} \vec{S}_j \vec{S}_j^T \mathbb{H}_j^{-1}}{\langle \vec{S}_j, \mathbb{H}_j^{-1} \vec{S}_j \rangle} + \frac{\vec{S}_j \vec{S}_j^T}{\langle \vec{Y}_j, \vec{S}_j \rangle},$$

em que $\vec{S}_j = \vec{X}_{j+1} - \vec{X}_j$, $\vec{Y}_j = \nabla \mathcal{J}(\cdot, \vec{X}_{j+1}) - \nabla \mathcal{J}(\cdot, \vec{X}_j)$ e $\langle \cdot, \cdot \rangle$ denota o produto escalar. Detalhes da biblioteca e do método quasi-Newton podem ser encontrados em NAG (1995) e Dennis e Moré (1977), respectivamente. Outras estratégias de aplicações do método quasi-Newton podem ser encontradas em Haykin (2001), Bishop (1995) e Bishop (2006), além das referências citadas no início da Seção 3.6.

A versatilidade da biblioteca utilizada permite ao programador fornecer ou não as equações para o cálculo dos gradientes definidos na Eq. (3.12). O não fornecimento das equações, necessárias para tais cálculos, exige uma maior demanda de tempo computacional, pois a biblioteca utilizada calcula cada um dos gradientes associados às variáveis, por meio de diferenças finitas (NAG, 1995). Visto que o pesquisador conhece as funções de ativação utilizadas em todos os neurônios da rede e, do modo como foi definido o funcional expresso pela Eq. (3.8), é possível calcular as equações associadas aos gradientes definidos na Eq. (3.12) a fim de tornar o processo de treinamento mais rápido. O gradiente do funcional definido pela Eq. (3.8) é dado por

$$\nabla \mathcal{J}(\cdot, \vec{X}) = - \sum_{p=1}^{N_p} \sum_{k=1}^{N_s} \left(d_{kp}(\cdot) - y_{kp}(\cdot, \vec{X}) \right) \frac{\partial}{\partial \vec{X}} y_{kp}(\cdot, \vec{X}), \quad (3.13)$$

em que $\vec{X} = [\mathbb{W} \ \mathbb{B} \ \mathbb{A}]^T$ e $y_{kp}(\cdot, \vec{X})$ é definido em termos das funções de ativação utilizadas na rede e de todas as variáveis livres da rede. Os detalhes para o cálculo das derivadas presente no termo à direita do sinal de igualdade da equação anterior são apresentados na Seção D.3 do Apêndice D.

3.7 Redes Neurais e Problemas Inversos

Esta seção tem o intuito de mostrar, de modo breve, como as RNAs podem atuar na resolução de problemas inversos. Inicialmente, é apresentada um breve descrição sobre problemas inversos, suas aplicações e metodologias de solução. Em seguida, é feita a associação com as RNAs, enfatizando a diferença nas metodologias de solução quando utilizado as RNAs como operadores inversos.

Problemas inversos são difíceis por definição (GROETSCH, 1999). Um problema inverso vem acompanhado de um problema direto, e a escolha de qual é o direto e qual é o inverso está associada à cultura. Em geral, o problema direto é aquele para o qual se conhecem as causas (parâmetros) e o modelo associado (K) e, então, se busca os efeitos (solução) através de algum método matemático (analítico ou numérico). Já o problema inverso, seguindo esse raciocínio, consiste em determinar as causas (parâmetros) a partir do conhecimento completo (ou parcial) do modelo e da observação, em geral parcial, dos efeitos (solução). Veja Figura 3.8. Segundo a “definição” dada do que é um problema direto e um problema inverso, é possível observar que as “informações” que se tem sobre o problema direto são completas, enquanto as do problema inverso não são. Essa observação nos leva à definição dada pelo pesquisador francês Jacques Hadamard para problema bem-posto, e a definição dada pelo pesquisador russo Oleg Mikailivitch Alifanov para problemas inversos.

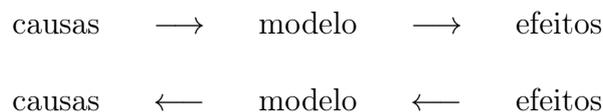


Figura 3.8 - Diagrama representando problema direto e problema inverso.

Hadamard (1902) define que um problema bem-posto deve apresentar três características: a) existência; b) unicidade; c) estabilidade da solução. Isso significa que a solução para o problema deve existir, ser única e, ainda, ter dependência contínua dos dados, *i.e.*, pequenas alterações nos dados de entrada (parâmetros) devem produzir pequenas alterações na solução. Em geral, problemas diretos satisfazem as características de problema bem-posto. Por outro lado, Alifanov (1994) afirma que a solução de um problema inverso consiste em determinar as causas baseando-se na observação de alguns de seus efeitos. A observação “incompleta” dos efeitos, ou então, por propriedades do operador matemático envolvido (operadores compactos), faz com que os problemas inversos apresentem uma característica intrínseca: são mal-postos, *i.e.*, não satisfazem alguma ou todas as condições de problema bem-posto

definidas por Hadamard.

Sob o ponto de vista matemático, para garantir que um problema inverso possa satisfazer as condições impostas por Hadamard, o espaço solução do problema é aumentado a fim de garantir a existência da solução e, impõe-se mais condições sobre o operador, ou então, escolhe-se a solução que mais se adapta ao modelo a fim de garantir a unicidade. Com relação à instabilidade, esta pode ser conseguida através de técnicas de regularização que suavizam as variações bruscas que ocorrem na solução dos problemas inversos (TIKHONOV; ARSEININ, 1977; MOROZOV, 1993). Uma introdução à teoria de problemas inversos, em um nível básico, pode ser encontrada em Groetsch (1999) e, em um nível um pouco mais avançado, em Kirsch (1996). Outras referências interessantes, sob o ponto de vista de aplicações, são Silva Neto e Moura Neto (2005) e Campos Velho (2008).

Em transferência radiativa, metodologias de inversão têm sido estabelecidas como um relevante tema de pesquisa em virtude de suas aplicações à ciência e engenharia (MCCORMICK, 1992). Em McCormick (2001), há uma revisão de métodos e de aplicações de problemas inversos à engenharia, química, medicina, computação e na área espacial, e em Gordon (2002), há uma revisão de técnicas de inversão para problemas ligados à ótica hidrológica.

A formulação de inversão pode ser explícita ou implícita. Veja Figura 3.9. Nas formulações explícitas, as equações utilizadas na modelagem do problema direto são manipuladas de forma a se obter expressões em que as incógnitas do problema apareçam explicitamente. Para o caso em que se conhece a solução do problema direto, Silva Neto e McCormick (2002) propuseram um método baseado nos momentos da radiação emergente do meio. Para o caso em que não se conhece a solução do problema direto, Kauati et al. (2001) propuseram o método fonte-detector.

No caso implícito, que é o mais comum, a busca da solução de um problema inverso se dá pela minimização de um funcional que, por exemplo, pode ser definido como a diferença quadrática entre os valores experimentais e os valores do modelo, *i.e.*,

$$\mathcal{J}(x^{\text{exp}}(t_i), \xi) = \min \sum_{i=1}^M \|x^{\text{exp}}(t_i) - x^{\text{mod}}(t_i, \xi)\|_2^2, \quad (3.14)$$

em que M é o total de pontos amostrais, ξ é um parâmetro (ou um conjunto de parâmetros) a ser determinado, $x^{\text{exp}}(t_i)$ é a solução conhecida (ou experimental) e $x^{\text{mod}}(t_i, \xi)$ é a solução calculada pelo “modelo inverso”, durante o processo de

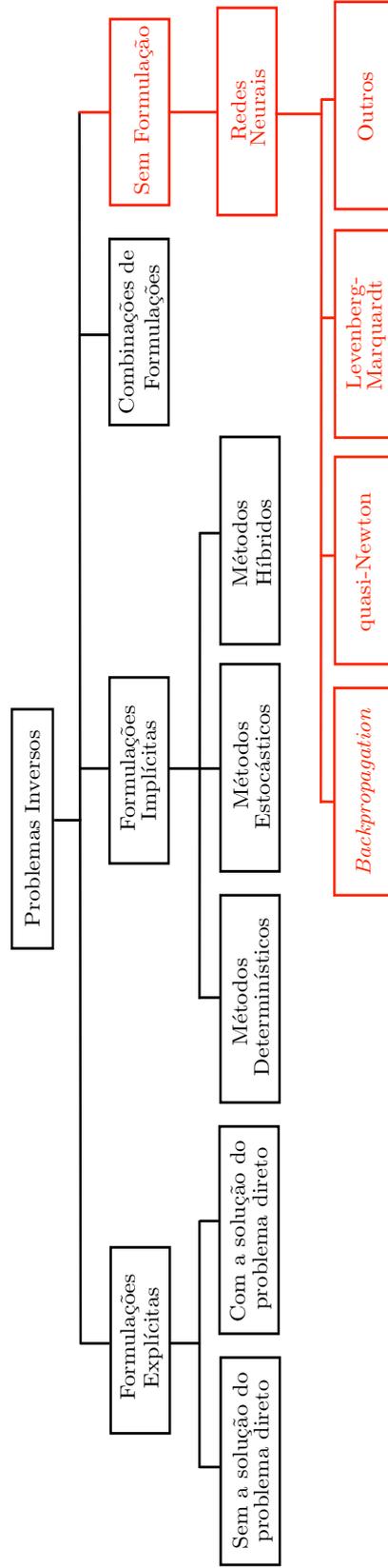


Figura 3.9 - Organograma contendo as possíveis alternativas para buscar a solução de problemas inversos (blocos pretos), adicionada a possibilidade da utilização das RNAs como metodologia de resolução e as alternativas para o treinamento para o treinamento (blocos vermelhos).
 Fonte: Adaptada de Sila Neto e Campos Velho (2009).

otimização, para cada valor de ξ . Para a minimização do funcional existe uma grande variedade de métodos que podem ser divididos em dois grupos: métodos de natureza determinística e métodos de natureza estocástica. Há também um terceiro grupo, que é a combinação desses dois, chamado de métodos híbridos.

No grupo dos métodos determinísticos são encontrados métodos como: máxima descida e gradiente conjugado (BURDEN; FAIRES, 2003), Newton e quasi-Newton (DENNIS; MORÉ, 1977), Levenberg-Marquardt (LEVENBERG, 1944; MARQUARDT, 1963). No grupo dos estocásticos, os mais conhecidos são: Algoritmo Genético (*Genetic Algorithm – GA*) (HOLLAND, 1975); o algoritmo de Recozimento Simulado (*Simulated Annealing – SA*) (METROPOLIS et al., 1953); o método inspirado no comportamento de colônia de formigas (*Ant Colony Optimization – ACO*) (DORIGO, 1992); o Algoritmo de Colisão de Partículas (*Particle Collision Algorithm – PCA*) (SACCO; OLIVEIRA, 2005; SACCO et al., 2006; SACCO et al., 2008), e o Algoritmo de Colisão de Múltiplas Partículas (*Multi-Particle Collision Algorithm – MPCA*) (LUZ et al., 2008). Já os métodos híbridos fazem a combinação dos métodos determinísticos e estocásticos. Por exemplo, é possível iniciar a busca de um ponto ótimo (solução que minimiza o funcional) através de um dos métodos estocásticos e, então, a partir de um determinado número de iterações desse método, iniciar um método determinístico cuja solução inicial seja a última solução “ótima” obtida pelo método estocástico. A Figura 3.9 mostra um organograma com as possíveis formulações a serem adotadas para resolver problemas inversos, e os métodos para se obter a solução.

Além dos problemas inversos serem mal-postos, acrescenta-se o fato de que, em algumas situações, principalmente na modelagem de fenômenos físicos, o modelo direto que descreve o fenômeno pode ser incompleto, não existir, ou então, envolver um operador matemático compacto. Veja Kirsch (1996) para exemplos de operadores compactos. Nessas situações, é de se esperar que a resolução de um problema inverso seja uma tarefa ainda mais complicada ou, talvez, impossível. Diante desses problemas, a utilização de RNAs pode ser uma estratégia interessante a fim de contornar (ou minimizar) essas dificuldades.

Como visto, as RNAs são sistemas computacionais estruturados em uma aproximação à computação baseada em ligações, e a inspiração desses sistemas se deu pela tentativa de modelar o funcionamento do cérebro humano. Diante disso, as RNAs apresentam características como: capacidade de aprendizado, auto-organização, tolerância a falhas, flexibilidade e produção de resultados em tempo real como comentado na Seção 3.3.

A vantagem da utilização desses sistemas na resolução de problemas inversos se dá pelo fato de que não há necessidade de um modelo matemático que represente o problema (veja Figura 3.9), mas sim, apenas os pares (padrões) de entrada (causa) e de saída (efeito). A partir desses padrões, a rede é capaz de “gerar” um modelo matemático, através do ajuste de suas variáveis livres (pesos sinápticos, níveis de *bias* e, eventualmente, parâmetros das funções de ativação), com a habilidade de fazer a representação causa-efeito, *i.e.*, fazer um mapeamento entre os padrões de entrada e os padrões de saída.

A diferença entre a formulação implícita “clássica” apresentada na Eq. (3.14) e a formulação através de RNAs mostrada na Eq. (3.8) reside no fato de que, na primeira, é feita a busca de um valor (ou conjunto de valores) ξ que minimize a distância entre as soluções $x^{\text{exp}}(\cdot)$ e $x^{\text{mod}}(\cdot)$. Enquanto que, na segunda metodologia, são buscados valores $y_{kp}(\cdot)$, calculados pela rede, que minimizem a distância desse valor com cada padrão de treinamento $d_{kp}(\cdot)$ do conjunto de saída. Observe que na formulação clássica é necessário a resolução do problema direto, durante a resolução do problema inverso, pois são comparados os valores de $x^{\text{exp}}(\cdot)$ e de $x^{\text{mod}}(\cdot)$. Dessa forma, é necessário conhecer o modelo matemático associado ao problema. Portanto, a cada iteração do problema inverso um problema direto é resolvido.

Por outro lado, a metodologia de redes dispensa o modelo matemático associado ao problema pelo fato de que a rede “substitui” o modelo (direto), durante a propagação (adiante) dos padrões de entrada. Ainda, a resposta obtida pela rede, ao final da propagação do sinal de entrada, é (são) o(s) valor(es) que se deseja determinar, *i.e.*, os valores $y_{kp}(\cdot)$. Nessa metodologia, esses valores $y_{kp}(\cdot)$ podem ser interpretados como os valores ξ buscados, durante o processo de otimização, pela metodologia “clássica”.

Na formulação “clássica”, o valor ξ é a variável (ou conjunto de variáveis), enquanto que na formulação de redes, as variáveis são os pesos sinápticos, os níveis de *bias* e, eventualmente, o parâmetro de cada função de ativação. Vale lembrar, também, que a metodologia “clássica” é capaz de resolver um único problema, *i.e.*, é encontrado um valor (ou conjunto) ξ que minimiza a distância entre os valores experimentais $x^{\text{exp}}(\cdot)$ e os do modelo $x^{\text{mod}}(\cdot)$ para um problema específico. Porém, quando uma RNA é utilizada, durante o treinamento são apresentadas à rede diversas soluções, e esta aprende a fazer o mapeamento entrada-saída para cada solução apresentada, gerando um modelo capaz de resolver uma série de problemas, desde que essa série esteja contida no “domínio” do conjunto de treinamento.

A utilização de RNAs na busca de solução para problemas inversos em transferência radiativa e/ou ótica hidrológica é pouco utilizada, no entanto, elas se mostram eficientes (sozinhas ou combinadas com outros métodos) na resolução desse tipo de problema (OLIVEIRA, 2010). Uma formulação híbrida, combinando RNA com o método de Levenberg-Marquardt, pode ser encontrada em Soeiro et al. (2004b) e Soeiro et al. (2004a). Nesses trabalhos, os autores utilizaram uma RNA para obter estimativas iniciais para o método de Levenberg-Marquardt e, com isso, conseguiram diminuir o tempo computacional necessário para a inversão do problema por parte do método determinístico. Em Soeiro e Silva Neto (2006), foi utilizada uma RNA com duas camadas para se obter estimativas iniciais para a solução de um problema inverso em transferência de calor. Nesses trabalhos, os autores desenvolveram e implementaram a RNA para a solução dos problemas inversos em transferência radiativa, que envolve a estimativa da espessura ótica, albedo de espalhamento simples e reflexão difusa nas superfícies de contorno do meio.

Em Chalhoub et al. (2007), é resolvido um problema de transferência radiativa no qual foi usado o método de Monte Carlo (BELL; GLASSTONE, 1970), com diferentes níveis de precisão, para gerar soluções a fim de serem utilizadas para treinar a rede. O problema inverso, nesse caso, consistiu na estimativa da espessura ótica e do albedo de espalhamento simples, e, segundo os autores, a rede se mostrou robusta, produzindo bons resultados mesmo quando os dados de entrada (treinamento) foram corrompidos com ruído.

Oliveira et al. (2010) resolve um problema inverso em transferência radiativa em um meio unidimensional, homogêneo, absorvedor e espalhador isotrópico, utilizando o conceito de máquina de comitê (HAYKIN, 2001) de redes neurais artificiais para estimar o albedo de espalhamento simples, a espessura ótica do meio e a reflexividade difusa nas superfícies de contorno. Outros trabalhos que utilizam RNAs para resolver problemas inversos em transferência radiativa são Carvalho (2011) e Oliva Soares (2013).

No próximo capítulo, são detalhados os problemas abordados nesta tese, valores para os parâmetros presentes na ETR e descrita no Capítulo 2, modelos bio-óticos utilizados, entre outras informações necessárias.

4 PROBLEMAS ABORDADOS

Neste capítulo são descritos os detalhes gerais de cada problema e que estão associados aos coeficientes de absorção e de espalhamento, ao perfil de concentração de clorofila, ângulo de incidência e intensidade do feixe. Já os detalhes específicos estão relacionados aos dois parâmetros do perfil adotado: a quantidade total de biomassa e a concentração de clorofila de fundo.

4.1 Detalhes Gerais dos Problemas

No Capítulo 2, foi apresentada uma breve descrição do problema da ótica e do modelo matemático que é capaz de representar tal problema. Na descrição dada naquele capítulo, não foram levadas em conta propriedades específicas do meio aquático que regem os fenômenos de absorção e de espalhamento. Na Eq. (2.1) aparece a razão $b_r(\tau, \lambda)/c_r(\tau, \lambda)$, chamada de albedo de espalhamento simples, geralmente representada por $\varpi_r(\tau, \lambda)$. O numerador da fração anterior representa o coeficiente de espalhamento e o denominador representa o coeficiente de atenuação. O coeficiente de atenuação é definido como a soma dos coeficientes de absorção $a_r(\tau, \lambda)$ e de espalhamento $b_r(\tau, \lambda)$, *i.e.*, $c_r(\tau, \lambda) = a_r(\tau, \lambda) + b_r(\tau, \lambda)$. O índice r indica que cada região espacial poderá ter os valores desses coeficientes distintos. Veja Figura 2.2.

Nesta tese, são utilizados os coeficientes de absorção e de espalhamento definidos por Morel (1991) e Gordon e Morel (1983), respectivamente, e representados matematicamente por

$$a_r(\tau, \lambda) = [a_\lambda^w + 0.06 a_\lambda^c \text{chl}(\tau)^{0.65}] [1 + 0.2 e^{-0.014(\lambda-440)}] \quad (4.1)$$

e

$$b_r(\tau, \lambda) = b_\lambda^w + \frac{550}{\lambda} 0.3 \text{chl}(\tau)^{0.62}, \quad (4.2)$$

em que a_λ^w , b_λ^w e a_λ^c representam, respectivamente, os coeficientes de absorção e de espalhamento da água pura e o coeficiente de absorção específico do fitoplâncton. Os valores de a_λ^w e b_λ^w , utilizados nesta tese, podem ser encontrados em Morel e Prieur (1977), e os valores para a_λ^c podem ser encontrados em Prieur e Sathyendranath (1981). A Figura 4.1 mostra o comportamento desses coeficientes para os comprimentos de onda na faixa do visível, *i.e.*, $\lambda \in [400, 700]$ nm.

Os modelos bio-ópticos apresentados nas Eqs. (4.1) e (4.2) se restringem a águas naturais do Caso 1. Águas do Caso 1 são aquelas em que o fitoplâncton e a matéria orgânica colorida dissolvida (*Colored Dissolved Organic Matter – CDOM*), que é

resultado da degradação do fitoplâncton, são os principais agentes responsáveis pela variação das propriedades óticas da água (MOREL; PRIEUR, 1977; GORDON; MOREL, 1983; SATHYENDRANATH; MOREL, 1983). Segundo esses mesmos autores, águas do Caso 2 são aquelas que são influenciadas não somente pelo fitoplâncton, mas também por outras substâncias opticamente ativas e que variam de forma independente, como as partículas inorgânicas em suspensão e as substâncias amarelas. No entanto, outros estudos (MOREL; AHN, 1991; STRAMSKI; KIEFER, 1991; ULLOA et al., 1992) sugerem que outros pequenos organismos, além do fitoplâncton, também têm um papel importante na determinação de algumas propriedades óticas para águas do Caso 1. Nesse caso, o conceito de águas do Caso 1 leva em conta essas contribuições além das do fitoplâncton. Já nas águas do Caso 2, outros materiais particulados e/ou substâncias amarelas podem ter um influência significativa sobre as propriedades óticas e, nesse caso, não estão necessariamente relacionadas com a concentração de fitoplâncton; portanto, devem ser tratadas como variáveis independentes (KAMPEL; NOVO, 2005).

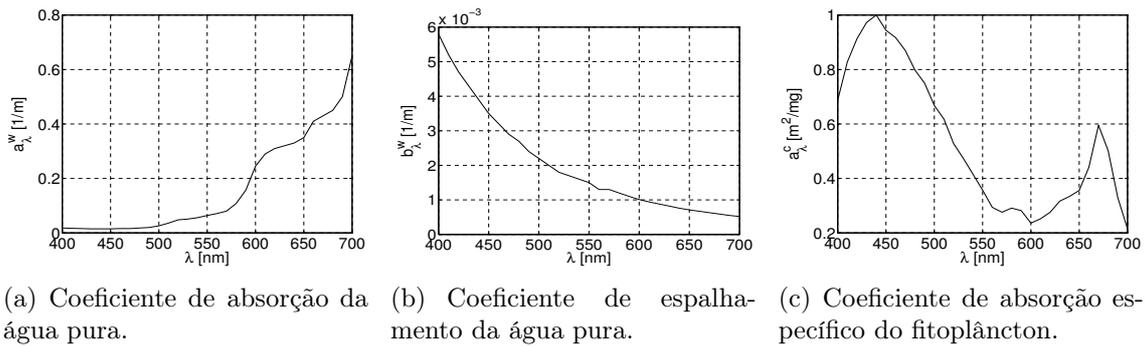


Figura 4.1 - Coeficientes de absorção e de espalhamento da água pura e coeficiente de absorção específico do fitoplâncton.

Para representar o perfil da distribuição da concentração de clorofila, é utilizado o modelo sugerido por Platt et al. (1994). Esse modelo é expresso por uma função gaussiana definida por

$$\text{chl}(\tau) = \text{chl}_0 + \frac{h}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\tau - \tau_m)^2}{2\sigma^2}\right], \quad (4.3)$$

em que chl_0 , h , τ_m e σ são, respectivamente, o valor da concentração de clorofila de fundo, o total de biomassa, a profundidade onde se localiza o pico da concentração e a largura (desvio padrão) da região do pico da concentração. Ainda, segundo Platt e Sathyendranath (1988), esses parâmetros devem ser ajustados a partir de

medições feitas em campo, os quais variam de acordo com a localização e a época do ano. Por exemplo, para o Mar Céltico no mês de maio, Platt e Sathyendranath (1988) sugerem que: $\text{chl}_0 = 0.2 \text{ mg} \cdot \text{m}^{-3}$; $h = 144 \text{ mg} \cdot \text{m}^{-2}$; $\tau_m = 17 \text{ m}$; e $\sigma = 9 \text{ m}$. A Figura 4.2 mostra o perfil da concentração de clorofila, definido pela Eq. (4.3), considerando os valores sugeridos por Platt e Sathyendranath (1988).

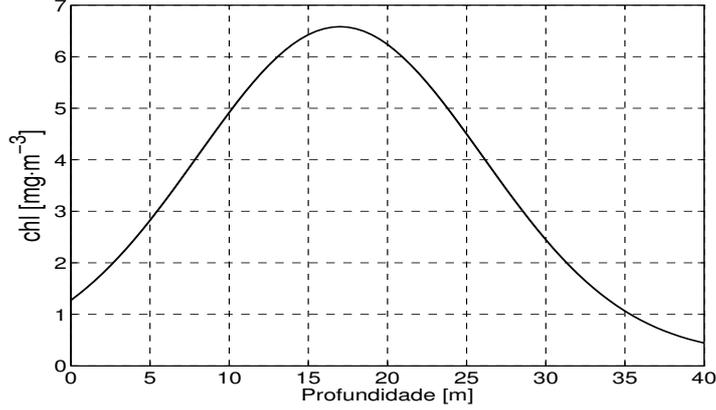


Figura 4.2 - Perfil da concentração de clorofila para o Mar Céltico no mês de maio, segundo Platt e Sathyendranath (1988).

Como pode ser visto, os coeficientes de absorção e espalhamento, definidos nas Eqs. (4.1) e (4.2), apresentam dependência do comprimento de onda λ e da concentração de clorofila, que, por sua vez, tem variação com relação à variável ótica τ . Como no Capítulo 2 é considerado que o comprimento de onda assume valores médios no intervalo (banda) $\Delta\lambda_g$, os coeficientes expressos pelas Eq. (4.1) e (4.2) também assumirão um valor médio λ_g .

A dependência da concentração de clorofila, que apresenta variação contínua com relação à variável ótica τ , torna a resolução do problema complexa. A fim de contornar isso, é comum considerar que a espessura ótica seja dividida em R regiões, e que, em cada região, a concentração de clorofila assuma um valor médio. A partir das considerações feitas, é possível escrever a Eq. (4.3), para cada região espacial, como

$$\text{chl}_r = \text{chl}(\bar{\tau}_{m_r}) = \text{chl}_0 + \frac{h}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\bar{\tau}_{m_r} - \tau_m)^2}{2\sigma^2}\right], \quad (4.4)$$

em que $\bar{\tau}_{m_r}$ é o ponto médio do intervalo $[\tau_{r-1}, \tau_r]$. Como consequência, os coeficientes de absorção e de espalhamento também assumirão valores médios em cada uma dessas regiões. Assim, as Eqs. (4.1) e (4.2), para cada região espacial, passam a ser

escritas como

$$a_{r,\lambda_g} = \left[a_{\lambda_g}^w + 0.06 a_{\lambda_g}^c \text{chl}_r^{0.65} \right] \left[1 + 0.2 e^{-0.014(\lambda_g - 440)} \right],$$

e

$$b_{r,\lambda_g} = b_{\lambda_g}^w + \frac{550}{\lambda_g} 0.3 \text{chl}_r^{0.62}.$$

Para cada valor de λ_g e para cada região espacial, tem-se uma ETR que modela a interação da radiação eletromagnética com o meio, e o acoplamento entre as regiões é feito por meio das condições de interface que foram definidas na Eq. (2.6). Veja Figura 4.3.

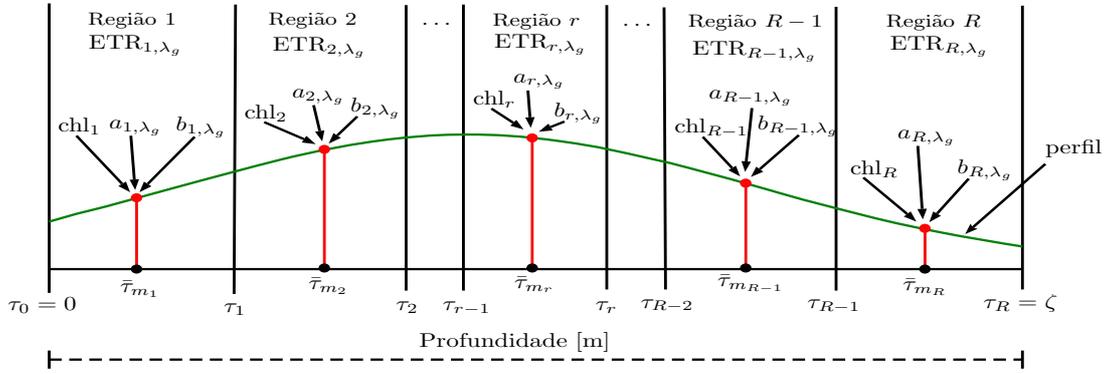


Figura 4.3 - Representação dos valores médios adotados em cada região espacial.

Com relação à função de fase, presente na Eq. (2.1), é usada a função definida por [Henyey e Greenstein \(1941\)](#), representada matematicamente por

$$p(\rho, \Theta) = \frac{1}{4\pi} \frac{1 - \rho^2}{\sqrt{(1 + \rho^2 - 2\rho \cos \Theta)^3}}, \quad (4.5)$$

em que ρ representa o coeficiente de assimetria, o qual é ajustado a fim de controlar as quantidades relativas de espalhamento para frente e para trás ([MOBLEY, 1994](#)), Θ é o ângulo formado entre o vetor que representa a direção incidente do fóton e o vetor que representa a direção do fóton após este ser espalhado. Veja Apêndice B. A função de fase mostrada na Eq. (4.5) será a mesma para todas as regiões espaciais e para todos os problemas considerados nesta tese.

Por fim, os valores para os ângulos polar e azimutal de incidência são $\theta_0 = 60^\circ$ e $\varphi_0 = 0^\circ$, respectivamente, e o valor da intensidade do feixe foi de $I_0 = \pi$. Esses valores foram considerados constantes para todas as simulações realizadas nesta tese.

4.1.1 Detalhes Específicos para os Problemas Abordados

Nos capítulos e seções precedentes, foram definidos os formalismos matemáticos envolvidos com o problema da ótica hidrológica e das RNAs. Como visto, há uma grande quantidade de parâmetros envolvidos na formulação do problema da ótica hidrológica. Alguns desses parâmetros já foram definidos, outros, específicos, serão definidos nesta subseção, e caracterizarão os problemas-teste a serem resolvidos nesta tese. Esses parâmetros específicos estão associados a Eq. (4.4) e dizem respeito à quantidade total de biomassa h e à concentração de clorofila de fundo chl_0 .

Os valores para os parâmetros h e chl_0 , presentes na Eq. (4.4), são apresentados na Tabela 4.1. A escolha desses valores se justifica pelo fato de que o valor mais elevado de h permite que se tenha concentrações de clorofila mais altas, quando comparadas às concentrações de clorofila obtidas com o menor valor de h . Em razão disso, foi modificado o valor da concentração de clorofila de fundo chl_0 de $0.2 \text{ mg} \cdot \text{m}^{-3}$ para $0.01 \text{ mg} \cdot \text{m}^{-3}$. A Figura 4.4 mostra o mapa da distribuição da concentração de clorofila de superfície, resultante da composição anual de 2009, obtido a partir dos dados medidos pelo sensor SeaWiFS, na qual é possível observar essa variação da concentração de clorofila superficial. Ainda, na Tabela 4.1, são apresentados a intensidade do feixe incidente de fótons na superfície da água, os valores dos ângulos de incidência polar e azimutal desse feixe (definidos anteriormente), a profundidade geométrica da camada de água analisada, o número de regiões espaciais consideradas e o valor do coeficiente de assimetria da função de fase, representados respectivamente, por I_0 , θ_0 , φ_0 , ζ , R e ρ . Os valores desses parâmetros serão os mesmos para cada valor de h considerado. Visto que a profundidade é de 40 metros, e o número de regiões espaciais é 20, segue que cada região espacial tem uma espessura de dois metros.

Tabela 4.1 - Valores adotados para os parâmetros I_0 , h , chl_0 , θ_0 , φ_0 , ζ , R e ρ .

h [$\text{mg} \cdot \text{m}^{-2}$]	chl_0 [$\text{mg} \cdot \text{m}^{-3}$]	I_0 [cd]	θ_0 [rad]	φ_0 [rad]	ζ [m]	R	ρ
144	0.20	π	$\pi/3$	0	40	20	0.924
30	0.01						

O passo seguinte consistiu em definir a localização espacial dos picos de concentração de clorofila para os problemas selecionados, *i.e.*, os valores de τ_m presentes na Eq. (4.4). Esses valores seguiram uma distribuição homogênea ao longo da profundidade, com espaçamento de um metro entre os picos, *i.e.*,

$$\tau_{m_i} = [1, 2, \dots, 40], \quad i = 1, 2, \dots, 40, \quad (4.6)$$

tendo, assim, um total de 40 valores para τ_m . Já para o desvio padrão foram adotados oito valores, a saber

$$\sigma_i = [5, 6, 7, 9, 11, 13, 15, 17], \quad i = 1, 2, \dots, 8, \quad (4.7)$$

os quais foram combinados com cada valor de τ_{m_i} . Veja Figura 4.5.

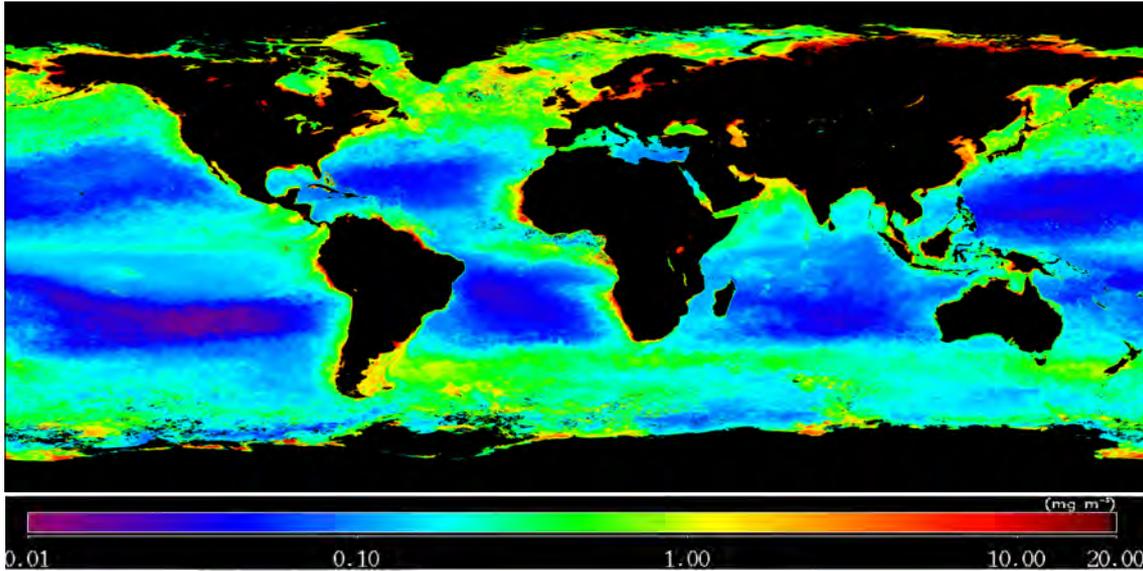


Figura 4.4 - Concentração de clorofila de superfície considerando a composição anual de 2009. Sensor SeaWiFS.

Fonte: Ocean Color (2013c).

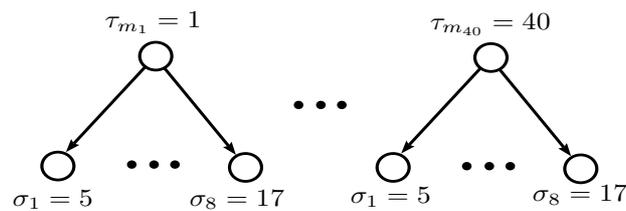


Figura 4.5 - Esquema gráfico da combinação dos valores de τ_m com os valores de σ .

Dessa forma, para cada valor de pico, definido na Eq. (4.6), foram gerados oito perfis gaussianos, resultando em um total de 320 perfis de concentração de clorofila para cada valor de h . A Figura 4.6 mostra os perfis resultantes das combinações feitas para os valores de pico $\tau_{m_5} = 5$ m, $\tau_{m_{20}} = 20$ m e $\tau_{m_{35}} = 35$ m com os valores de σ_i , apresentados na Eq. (4.7), para $h = 144$ mg \cdot m $^{-2}$ (Figuras 4.6(a), 4.6(c) 4.6(e)), e $h = 30$ mg \cdot m $^{-2}$ (Figuras 4.6(b), 4.6(d) e 4.6(f)). Para o primeiro caso chl $\in [0.2, 12)$ mg \cdot m $^{-3}$, já para o segundo chl $\in [0.01, 2.5)$ mg \cdot m $^{-3}$.

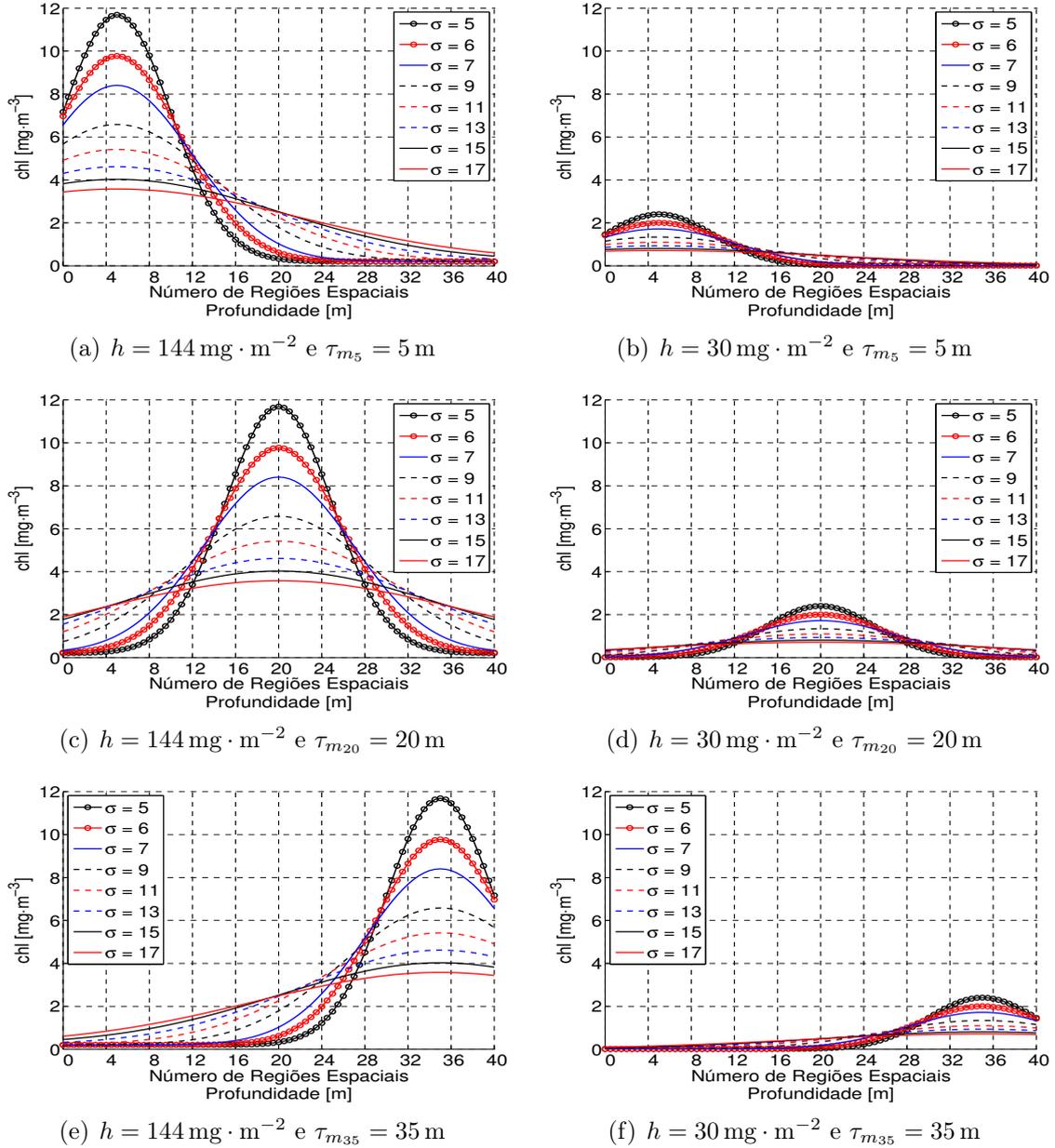


Figura 4.6 - Perfis de concentração de clorofila gerados a partir da combinação de três valores de pico com os valores de desvio padrão adotados.

No final do Capítulo 2, foi comentado que, ao se resolver o sistema de EDOs, gerado a partir da aproximação da ETR, é determinada a radiação emergente na superfície, em direções polares definidas *a priori*, após a interação do feixe de radiação incidente com o meio aquático, para o grupo de energia de interesse. Portanto, com relação aos grupos de energia, ou comprimentos de onda, são adotados dois grupos. O primeiro grupo considera dez valores discretos para λ_g compreendidos no intervalo $\lambda_g \in [412, 678] \text{ nm}$, e esses valores coincidem (de modo intencional) com os valores do sensor MODIS (*Moderate-Resolution Imaging Spectroradiometer*), nessa mesma

faixa do espectro eletromagnético, a bordo do satélite Aqua. Esses valores, segundo o *site* Ocean Color (2013c), são dados por

$$\lambda_{g_i}^A = [412, 443, 469, 488, 531, 547, 555, 645, 667, 678], \quad i = 1, 2, \dots, 10. \quad (4.8)$$

O segundo conjunto é composto por seis comprimentos de onda igualmente espaçados, com tamanho de banda $\Delta\lambda_g = 10$ nm, compreendidos no intervalo $\lambda_g \in [560, 610]$ nm, assim,

$$\lambda_{g_i} = [560, 570, 580, 590, 600, 610], \quad i = 1, 2, \dots, 6. \quad (4.9)$$

A escolha dos valores adotados no segundo conjunto de λ_g está associada a fatores que envolvem características físicas do problema da ótica. Em geral, apesar da radiação emergente na superfície da água ser mais intensa na faixa do azul (dependendo da concentração de clorofila) e no início do verde, esta apresenta pouca variação com relação à concentração de clorofila (principalmente na faixa do verde), *i.e.*, diferentes concentrações de clorofila produzem valores de radiação emergente muito próximas (CORTIVO et al., 2012c). Por outro lado, para os comprimentos de onda na faixa do vermelho, diferentes concentrações de clorofila produzem uma variação significativa na radiação emergente na superfície. No entanto, nessa banda espectral, a intensidade da radiação emergente na superfície decai e, ainda, a radiação incidente não consegue penetrar mais do que alguns poucos metros, devido à alta absorção da luz pela água nessa banda espectral (MOBLEY, 1994). Veja Figura 4.7.

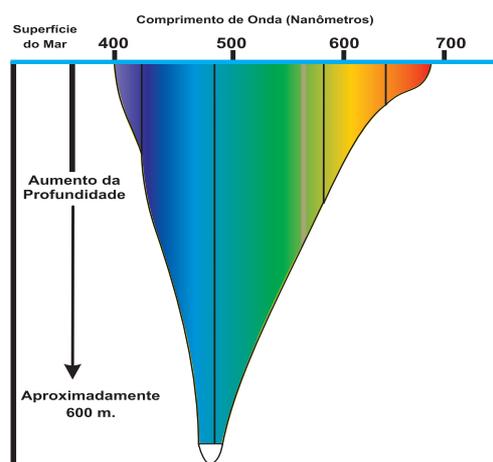
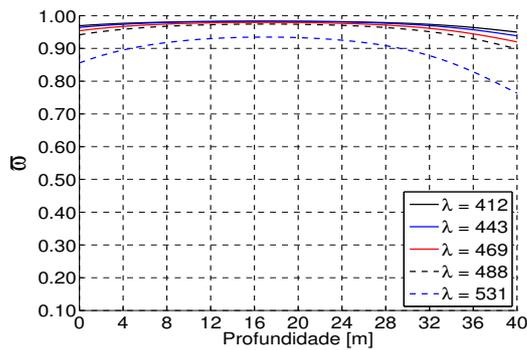


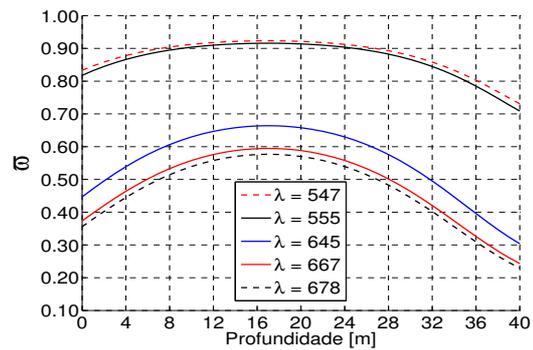
Figura 4.7 - Penetração da radiação solar em águas de oceano aberto para os comprimentos de onda da faixa do visível.

Fonte: Adaptada de Garrison (2011).

As Figuras 4.8 e 4.9 mostram as curvas do albedo de espalhamento simples, obtidas considerando o perfil mostrado na Figura 4.2, para comprimentos de onda definidos pelas Eqs. (4.8) e (4.9), respectivamente. O fato de a concentração de clorofila apresentar variação com relação à variável ótica τ , e influenciar diretamente nos coeficientes de absorção e espalhamento que, por sua vez, definem o albedo de espalhamento simples, faz com que este também apresente variação com relação à profundidade, como pode ser visto nas figuras.

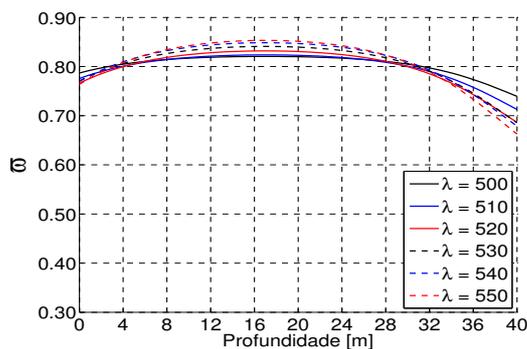


(a) Curvas do albedo de espalhamento simples para comprimentos de onda compreendidos no intervalo $\lambda \in [412, 531]$ nm.

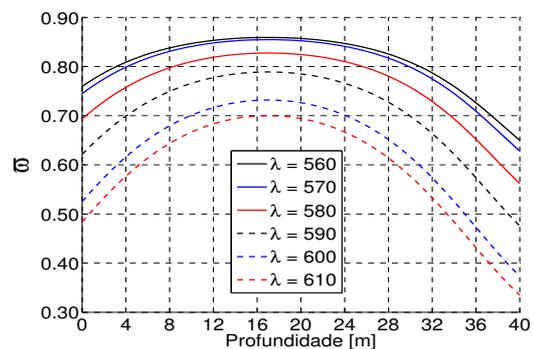


(b) Curvas do albedo de espalhamento simples para comprimentos de onda compreendidos no intervalo $\lambda \in [547, 678]$ nm.

Figura 4.8 - Curvas do albedo de espalhamento simples obtidas considerando o perfil mostrando na Figura 4.2 para diferentes comprimentos de onda compreendidos no intervalo $\lambda \in [412, 678]$ nm.



(a) Curvas do albedo de espalhamento simples para comprimentos de onda compreendidos no intervalo $\lambda \in [500, 550]$ nm.



(b) Curvas do albedo de espalhamento simples para comprimentos de onda compreendidos no intervalo $\lambda \in [560, 610]$ nm.

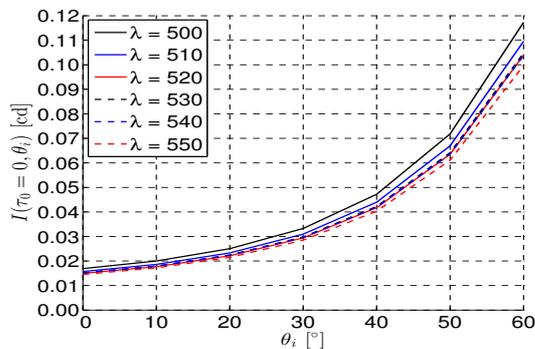
Figura 4.9 - Curvas do albedo de espalhamento simples obtidas considerando o perfil mostrando na Figura 4.2 para diferentes comprimentos de onda compreendidos no intervalo $\lambda \in [500, 610]$ nm.

A partir da Figura 4.8(a) é possível notar que as curvas mostradas praticamente não apresentam variação (curvatura) ao longo da profundidade, principalmente para os comprimentos de onda $\lambda = 412, 443, 469$ e 488 nm. O comportamento se repete na Figura 4.9(a), no entanto a ordem de grandeza do albedo é um pouco menor. Já as curvas de albedo mostradas nas Figuras 4.8(b) e 4.9(b), apresentam uma variação mais significativa com relação à profundidade (curvatura maior) e com relação a diferentes comprimentos de onda. Portanto, a escolha dos valores de comprimento de onda definidos na Eq. (4.9) está associada a essa sensibilidade do albedo para os comprimentos de onda no intervalo $\lambda \in [560, 610]$ nm.

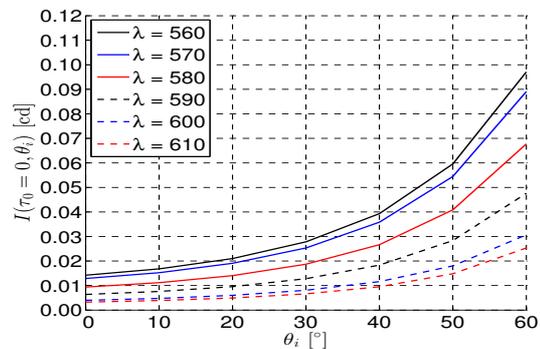
Os últimos parâmetros a serem definidos são as direções polares em que serão feitas as medidas¹ da radiação emergente. Os valores adotados para esses parâmetros são dados por

$$\theta_i = [0^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 50^\circ, 60^\circ], \quad i = 1, 2, \dots, 7. \quad (4.10)$$

As Figuras 4.10(a) e 4.10(b) mostram a radiação emergente na superfície, medidas (calculadas pelo modelo descrito no Capítulo 2) nas direções definidas na Eq. (4.10), correspondentes às curvas de albedo apresentadas nas Figuras 4.9(a) e 4.9(b). Observe que curvas de albedo próximas geram curvas de radiação emergente próximas também.



(a) Radiação emergente na superfície para comprimentos de onda compreendidos no intervalo $\lambda \in [500, 550]$ nm.



(b) Radiação emergente na superfície para comprimentos de onda compreendidos no intervalo $\lambda \in [560, 610]$ nm.

Figura 4.10 - Curvas da radiação emergente na superfície obtidas pela simulação da ETR, levando em conta o perfil mostrado na Figura 4.2, para diferentes comprimentos de onda compreendidos no intervalo $\lambda \in [500, 610]$ nm.

¹O termo “medida” é entendido como o resultado obtido por meio da simulação numérica do problema, e não como uma possível medida *in situ*.

Já as Figuras 4.11(a) e 4.11(b) ilustram, em mesma escala, a variação da radiação emergente para o perfil mostrado na Figura 4.2, com relação a diferentes comprimentos de onda, para medidas feitas em $\theta = 0^\circ$ e $\theta = 60^\circ$, respectivamente. A fim de mostrar que o comportamento da radiação emergente (ao longo de $\lambda \in [400, 700]$ nm) é semelhante nestas duas direções, a Figura 4.11(c) mostra novamente a radiação emergente medida na direção $\theta = 0^\circ$, porém com o eixo das ordenadas expandido. O comportamento (formato) é o mesmo, no entanto a ordem de grandeza é diferente. É importante notar que a região em que há maior variação da radiação emergente, com relação ao comprimento de onda, é justamente o intervalo definido na Eq. (4.9), entre 560 nm e 610 nm.

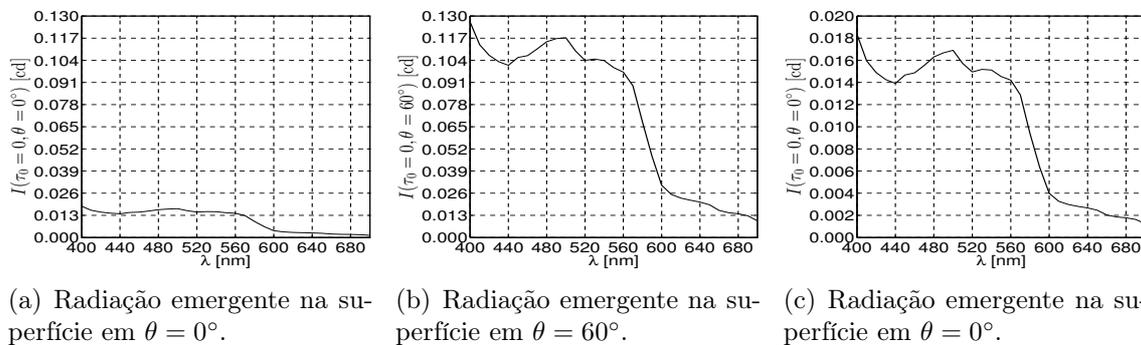
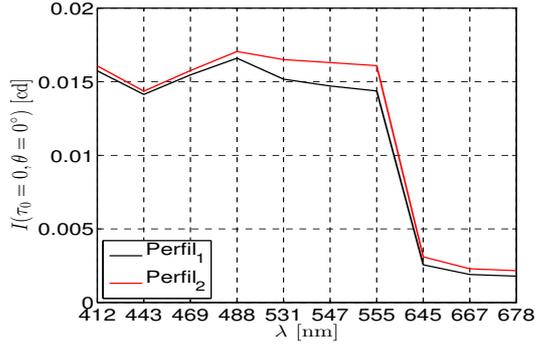


Figura 4.11 - Variação da radiação emergente na superfície, em $\theta = 0^\circ$ e $\theta = 60^\circ$, para diferentes λ para o perfil de concentração de clorofila mostrado na Figura 4.2.

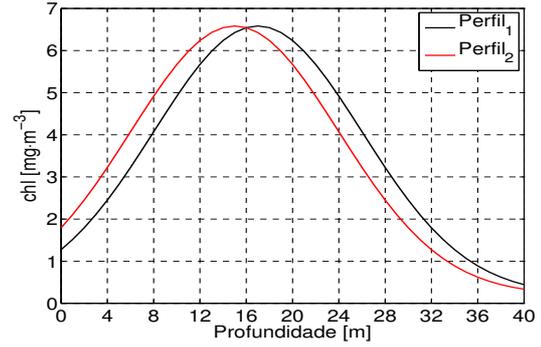
A Figura 4.12(a) ilustra a radiação emergente² calculada pelo modelo adotado nesta tese, levando em conta os comprimentos de onda definidos na Eq. (4.8), para os dois perfis de concentração de clorofila mostrados da Figura 4.12(b). Aqui é possível notar que a região em que há maior variação da radiação emergente, com relação aos comprimentos de onda, é entre 555 nm e 645 nm, e que é semelhante à região referenciada na Figura 4.11.

Definidos os valores necessários para todos os parâmetros, foram gerados dois grupos de problemas a partir dos valores de h . O grupo $\mathbf{G}_{h_{144}}$, que está associado ao valor de $h_{144} = 144 \text{ mg} \cdot \text{m}^{-2}$, e o grupo $\mathbf{G}_{h_{30}}$, que está associado ao valor de $h_{30} = 30 \text{ mg} \cdot \text{m}^{-2}$. Em cada grupo, foram consideradas duas classes de problemas: a classe de problemas $1D_{(6/10)\lambda}$, e a classe de problemas $7D_{1\lambda}$. Na classe de problemas $1D_{(6/10)\lambda}$ foram considerados três tipos de problemas, enquanto que na classe de problemas $7D_{1\lambda}$, foi considerado um único tipo.

²As medidas calculadas pelo modelo adotado, mesmo levando em conta os comprimentos de onda do sensor MODIS, são os valores de radiância e não os de reflectância.



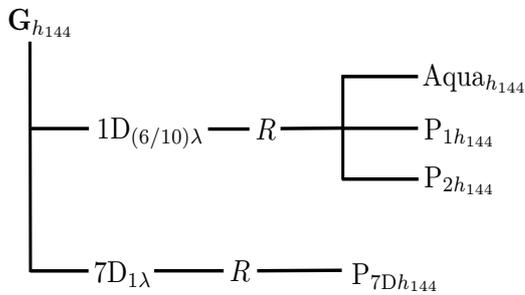
(a) Radiação emergente na superfície, calculada pelo modelo adotado, considerando os comprimentos de onda definidos na Eq. (4.8) para dois diferentes perfis de concentração de clorofila.



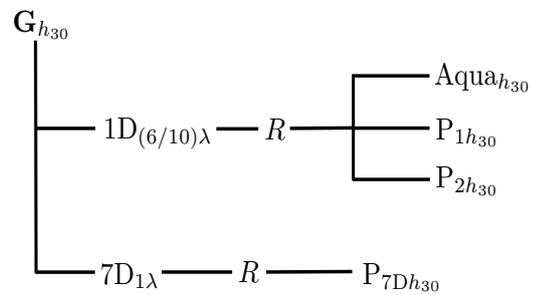
(b) Perfis de concentração utilizados para calcular a radiação emergente mostrada na figura ao lado. O Perfil₁ é o mesmo apresentado na Figura 4.2 e o Perfil₂ considera $h = 144 \text{ mg} \cdot \text{m}^{-3}$, $\sigma = 9 \text{ m}$ e $\tau_m = 15 \text{ m}$.

Figura 4.12 - Radiação emergente na superfície, calculada pelo modelo adotado, para os comprimentos de onda definidos na Eq. (4.8) para dois perfis de concentração de clorofila.

Os três problemas da classe $1D_{(6/10)\lambda}$ são: o problema Aqua_{h_*} ; o problema P_{1h_*} ; e o problema P_{2h_*} . A Figura 4.13 ilustra os problemas nas suas respectivas classes para cada um dos grupos $G_{h_{144}}$ e $G_{h_{30}}$. O parâmetro R se refere as 20 regiões espaciais consideradas ao longo da profundidade ótica.



(a) Grupo de problemas $G_{h_{144}}$.



(b) Grupo de problemas $G_{h_{30}}$.

Figura 4.13 - Diagrama ilustrando os problemas considerados em cada um dos grupos.

As características de cada um dos problemas contidos nas classes $1D_{(6/10)\lambda}$ e $7D_{1\lambda}$ para ambos os grupos são as seguintes:

- na classe de problemas $1D_{(6/10)\lambda}$ é feita a medida da radiação emergente na superfície em apenas uma direção polar, no entanto em seis ou dez comprimentos de onda distintos, como segue:
 - nos problemas Aqua_{h_*} são feitas as medidas da radiação emergente

apenas na direção $\theta = 0^\circ$ e nos comprimentos de onda definidos na Eq. (4.8);

- nos problemas P_{1h_*} também são feitas as medidas da radiação emergente na direção $\theta = 0^\circ$, no entanto para os comprimentos de onda apresentados na Eq. (4.9);
- nos problemas P_{2h_*} são feitas as medidas da radiação emergente na direção $\theta = 60^\circ$ e, também, nos comprimentos de onda apresentados na Eq. (4.9);

- na classe de problemas $7D_{1\lambda}$ são feitas as medidas da radiação emergente nas sete direções definidas na Eq. (4.10), no entanto em apenas um comprimento de onda. O comprimento de onda escolhido foi $\lambda_g = 560$ nm. A justificativa para a escolha desse valor será apresentada adiante.

Em cada um dos problemas há 320 perfis de concentração de clorofila. O problema $Aqua_{h_*}$ considera dez comprimentos de onda, assim, tem-se que o total de problemas diretos é de 3200 (10×320). O problema P_{1h_*} considera seis comprimentos de onda, assim, o total de problemas diretos é de 1920 (6×320). O problema P_{2h_*} também considera seis comprimentos de onda, assim, o total de problemas diretos é de 1920 (6×320) novamente. O problema pertencente à classe $7D_{1\lambda}$ considera um único comprimento de onda, logo, o total de problemas é 320 (1×320). Portanto, a soma de todos os problemas diretos que foram resolvidos, considerando os grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$, é de 14720.

A Tabela 4.2 mostra, além do total de problemas diretos para cada tipo de problema, o tempo computacional³ médio, gasto pelo código PEESNA⁴, na resolução de cada um dos problemas pertencentes a cada grupo. Para os problemas em que a radiação emergente é calculada na direção $\theta = 0^\circ$ (problemas $Aqua_{h_*}$ e P_{1h_*}), o tempo médio para a resolução de cada problema direto é inferior a um segundo, enquanto que, para os demais problemas, esse tempo fica em torno de 30 segundos. Tanto os problemas P_{1h_*} quanto os problemas P_{2h_*} (além dos problemas $Aqua_{h_*}$) consideram uma direção de medida, no entanto, para os problemas P_{1h_*} , o tempo computacional necessário

³O computador utilizado foi um Notebook Dell Vostro 1520 com processador Intel[®] Core[™] 2 Duo P8600 2.40GHz L2 3Mb, 4Gb de memória RAM, sistema operacional GNU/Linux 64 bits (Ubuntu 12.04), compilador Intel[®] Fortran Composer XE 2012. A compilação do código foi feita com a opção -O3.

⁴Visto que, ao final da execução, o código fornece uma série de informações sobre o problema, aquelas que não são de interesse para este trabalho foram eliminadas, a fim de tornar a execução ainda mais rápida.

para resolver um problema é inferior. Isso ocorre porque o código PEESNA, utilizado para a aproximação numérica, encerra o processamento uma vez que a contribuição de uma componente de espalhamento $I_{r,g}^m(\tau, \mu)$, veja Eq. (A.5), se tornar pequena quando compara às demais componentes.

Tabela 4.2 - Tempo médio gasto pelo código PEESNA, em segundos, para a resolução de cada problema direto.

Grupo	Classe	Problemas	$t_{\text{m\u00e9dio}}$	NPD
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	Aqua _{h_{144}}	0.89	3200
		P _{1h_{144}}	0.91	1920
		P _{2h_{144}}	28.43	1920
	7D _{1λ}	P _{7Dh_{144}}	27.83	320
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	Aqua _{h_{30}}	0.85	3200
		P _{1h_{30}}	0.78	1920
		P _{2h_{30}}	30.54	1920
	7D _{1λ}	P _{7Dh_{30}}	29.65	320

NPD = N\u00famero de Problemas Diretos

Com isso, foram apresentados todos os valores adotados para os par\u00e2metros envolvidos na modelagem dos problemas da \u00f3tica hidrol\u00f3gica, os quais s\u00e3o os problemas-teste resolvidos nesta tese, bem como o tempo computacional necess\u00e1rio para a resolu\u00e7\u00e3o desses problemas. De modo resumido (sem se ater a detalhes espec\u00edficos para a resolu\u00e7\u00e3o da ETR) os passos a serem seguidos a fim de se obter as “medidas” da radia\u00e7\u00e3o emergente na superf\u00edcie s\u00e3o ilustrados na Figura 4.14. O primeiro passo consiste em definir o pico τ_m da concentra\u00e7\u00e3o de clorofila e o desvio padr\u00e3o σ , sendo poss\u00edvel, a partir desses dois valores, obter o perfil da concentra\u00e7\u00e3o de clorofila. Com o perfil da concentra\u00e7\u00e3o de clorofila, \u00e9 poss\u00edvel determinar o valor m\u00e9dio chl_r em cada regi\u00e3o espacial, para, ent\u00e3o, a partir do valor de λ_g selecionado, calcular os coeficientes de absor\u00e7\u00e3o e espalhamento, tamb\u00e9m, para cada regi\u00e3o espacial. Em seguida, s\u00e3o definidas as dire\u00e7\u00f5es polares para as quais se deseja obter os valores da radia\u00e7\u00e3o emergente, para, ent\u00e3o, iniciar o processo de solu\u00e7\u00e3o da ETR, a fim de se obter a radi\u00e2ncia emergente na superf\u00edcie $I_{1,\lambda_g}(\tau_0, \mu)$ nas dire\u00e7\u00f5es consideradas.

O diagrama a seguir apresenta uma descri\u00e7\u00e3o resumida de todas as etapas necess\u00e1rias para resolver o problema. A partir do diagrama, fica f\u00e1cil identificar os par\u00e2metros de entrada (τ_m e σ) e de sa\u00edda ($I_{1,\lambda_g}(\tau_0, \mu)$). Visto que a figura apresentada \u00e9 um diagrama em blocos resumido do c\u00f3digo PEESNA, uma descri\u00e7\u00e3o mais detalhada do c\u00f3digo pode ser encontrada em Chalhoub (1997) e Chalhoub et al. (2003).

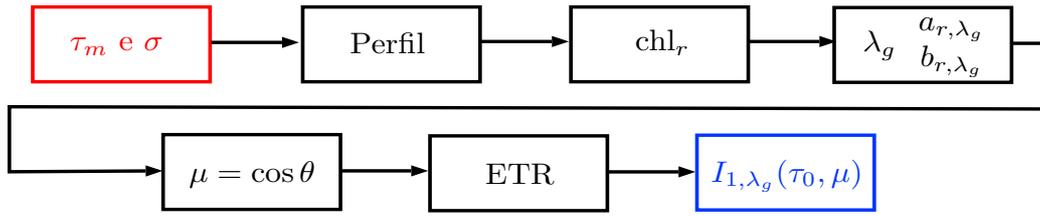


Figura 4.14 - Diagrama simplificado dos passos a serem seguidos para a resolução do problema direto.

4.1.2 Padrões de Entrada e de Saída

No Capítulo 2 foi visto que, ao se resolver a ETR, determina-se a radiação emergente na superfície, após a interação desta com o meio, em direções polares escolhidas *a priori*. Ainda, nas seções precedentes, foram definidos os parâmetros/equações necessários(as) para modelar as características do meio aquático, para, então, se obter a solução numérica da ETR. Nesse processo, é possível observar que a profundidade do pico τ_m e o desvio padrão σ , da concentração de chl, são os primeiros parâmetros a serem definidos e, a partir desses, se desenvolve todo o processo de solução, até se obter a radiação emergente na superfície $I_{1,\lambda_g}(\tau_0, \mu)$ (veja Figura 4.14). Portanto, os parâmetros τ_m e σ podem ser chamados de parâmetros (ou dados ou causas) de entrada, e a radiação emergente na superfície $I_{1,\lambda_g}(\tau_0, \mu)$, de dados de saída (ou efeitos).

Para a RNA, que atua como um operador inverso, a radiação emergente $I_{1,\lambda_g}(\tau_0, \mu)$ passa a representar os padrões (dados) de entrada e os valores de τ_m e σ passam a ser os padrões (dados) de saída. Portanto, a partir da radiação emergente na superfície, a RNA estima a profundidade do pico da concentração de clorofila e o desvio padrão. A partir desses dois parâmetros, é possível reconstruir o perfil da distribuição da concentração de clorofila mostrado na Eq. (4.4), visto que os demais parâmetros já foram definidos anteriormente. A Figura 4.15 ilustra o que foi descrito neste parágrafo.

Os experimentos feitos, considerando cada padrão de entrada composto apenas pela radiação emergente, não apresentaram resultados satisfatórios. A fim de melhorar esse resultados, foi adicionada mais uma posição (entrada) no vetor dos padrões de entrada com a finalidade de fornecer uma informação adicional à rede. Essa nova entrada foi composta pela concentração de clorofila de superfície, denotada por chl_0^S , obtida ao se considerar $\bar{\tau}_{m_r} = 0$ na Eq. (4.4), assim,

$$\text{chl}_0^S = \text{chl}_0 + \frac{h}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(-\frac{\tau_m}{\sigma}\right)^2\right]. \quad (4.11)$$

Portanto, as entradas da rede são constituídas pelas medidas da radiação emergente na superfície e pelo valor de chl_0^S para cada valor de τ_m e σ . Em uma aplicação prática, na qual não é possível calcular o valor da concentração de clorofila de superfície, pois não se conhece os valores do desvio padrão e da profundidade do pico da concentração de clorofila, esse valor de chl_0^S pode ser obtido a partir de medidas feitas *in situ*, ou então, através dos algoritmos razão de banda utilizados para estimativas de mapas de concentração de clorofila de superfície.

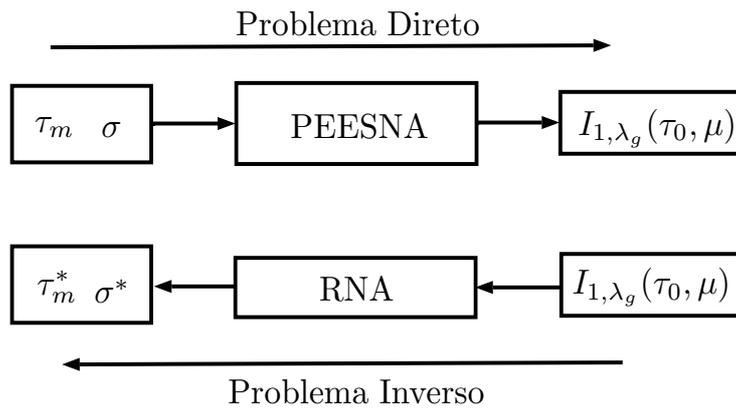


Figura 4.15 - Representação esquemática do problema direto e do problema inverso.

Mesmo com as duas considerações feitas (escolha do(s) comprimento(s) de onda adequado(s) e adição da concentração de clorofila às entradas da rede), uma única rede com duas saídas (uma para cada parâmetro a ser estimado) não foi capaz de resolver o problema de modo satisfatório. Frente a esse novo problema, a solução encontrada foi treinar duas redes: uma rede responsável por determinar o valor de σ e a outra responsável por determinar o valor de τ_m . A primeira rede, denominada RNA_σ , determina o valor do desvio padrão a partir dos padrões de entrada, que são compostos pela radiação emergente e pela concentração de clorofila de superfície. Já a segunda rede, denominada RNA_{τ_m} , determina a profundidade do pico da concentração de clorofila, levando em conta, além da radiação emergente e da concentração de clorofila de superfície, o valor do desvio padrão da concentração de clorofila determinado pela rede RNA_σ . Portanto, as entradas da segunda rede são formadas pela radiação emergente, pelo valor da concentração de clorofila de superfície e pelo valor do desvio padrão estimado pela rede RNA_σ . Veja Figura 4.16.

Essa estratégia tende a facilitar a associação dos padrões de entrada aos padrões de saída, pois, como a primeira rede deve determinar apenas o valor do desvio padrão, esta tem de classificar apenas os valores definidos na Eq. (4.7) a partir de

um determinado vetor do conjunto de padrões. Já a segunda rede, a que determina a profundidade do pico da concentração de clorofila, com a adição do valor de σ^* , determinado pela rede RNA_σ , recebe uma informação adicional e, assim, tende a melhorar seu desempenho na estimativa de um dos valores de τ_m^* definidos na Eq. (4.6).

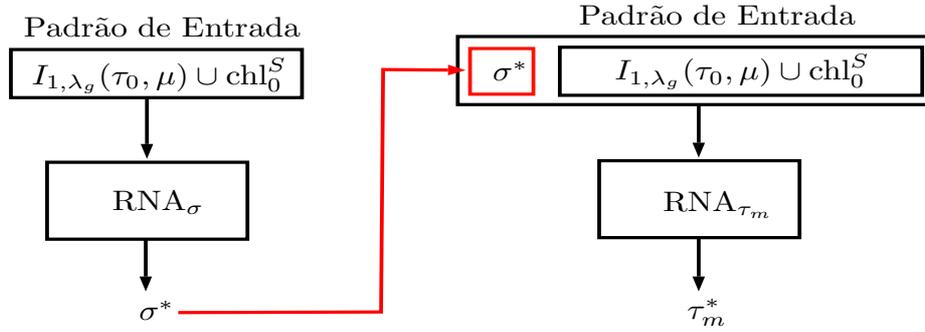


Figura 4.16 - Representação dos padrões de entrada e de saída para cada uma das redes.

Dessa forma, as entradas para cada um dos vetores que compõem o conjunto de treinamento (teste e validação), para os problemas Aqua_{h_*} , contêm o valor de chl_0^S calculado pela Eq. (4.11), e as medidas da radiação emergente na superfície em $\theta = 0^\circ$ nos comprimentos de onda definidos na Eq. (4.8). Assim, somam-se 11 entradas para a rede RNA_σ e 12 entradas para a rede RNA_{τ_m} . Já para os problemas P_{1h_*} , as entradas dos padrões de treinamento são compostas pelo valor de chl_0^S e pelas medidas da radiação emergente na superfície em $\theta = 0^\circ$ nos comprimentos de onda definidos na Eq. (4.9). Para esse tipo de problema, a RNA_σ tem um total de sete entradas, enquanto a RNA_{τ_m} tem oito entradas. A diferença entre as entradas dos problemas P_{1h_*} e P_{2h_*} reside no fato de que, nestes últimos, as medidas da radiação emergente são feitas em $\theta = 60^\circ$. Para os problemas pertencentes à classe $7\text{D}_{1\lambda}$, as entradas para cada vetor do conjunto de treinamento (teste e validação) são compostas, além do valor de chl_0^S , pelas medidas da radiação emergente nas direções polares definidas na Eq. (4.10) para $\lambda_g = 560 \text{ nm}$. Dessa forma, há oito entradas para a rede RNA_σ e nove entradas para a rede RNA_{τ_m} .

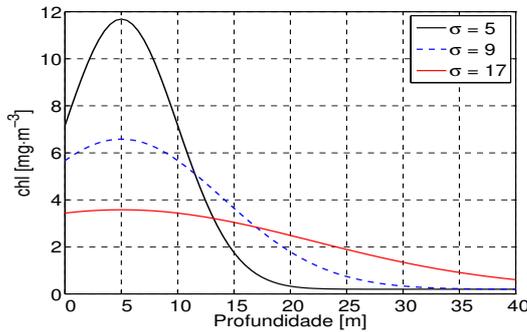
Por fim, vale lembrar que outras estratégias para solução dos problemas foram testadas, no entanto os melhores resultados foram obtidos com a estratégia descrita anteriormente. As estratégias que não apresentaram resultados satisfatórios serão explicadas brevemente no Apêndice E.

4.2 Redes Classificadoras

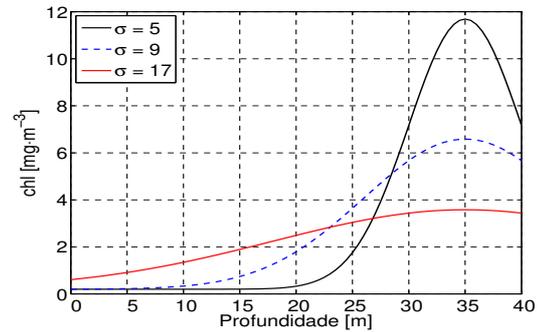
As redes de classificação são RNAs que trabalham em conjunto com as redes RNA_σ e RNA_{τ_m} a fim de classificar aqueles problemas em que é possível determinar os valores do desvio padrão e da profundidade do pico da concentração de clorofila, e aqueles em que não é possível determinar esses dois parâmetros, obtendo apenas um valor médio para o perfil da concentração de clorofila.

Essa classificação se fez necessária pelo fato de que perfis de concentração de clorofila que têm o pico profundo e desvio padrão pequeno não são passíveis de serem determinados a partir da radiação emergente na superfície gerada por eles. Assim, uma classificação foi feita por redes treinadas especialmente para essa tarefa, *i.e.*, identificar aqueles problemas que apresentam essa característica, a fim de facilitar o trabalho das demais redes.

A dificuldade na estimativa desses parâmetros se dá pelo fato de que a radiação emergente, que é obtida a partir dos problemas que apresentam picos de concentração mais profundos, tem ordem de grandeza menor quando comparada à radiação emergente obtida a partir dos problemas que apresentam perfis com o pico da concentração mais próximos à superfície. As Figuras 4.17(a) e 4.17(b) mostram três perfis em 5 e 35 metros, respectivamente, os respectivos albedos de espalhamento simples são mostrados nas Figuras 4.18(a) e 4.18(b), e as radiâncias emergentes, para esses perfis, são mostrados nas Figuras 4.19(a) e 4.19(b). A radiação emergente tem ordem de grandeza maior quando o pico da concentração de clorofila está mais próximo a superfície.

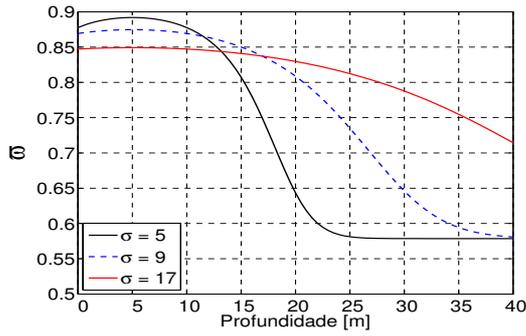


(a) Perfis gerados a partir da Eq. (4.4) para $\tau_m = 5$ m, com $\sigma = 5$ m, 9 m e 17 m, $h = 144 \text{ mg} \cdot \text{m}^{-2}$ e $\lambda_g = 560$ nm.

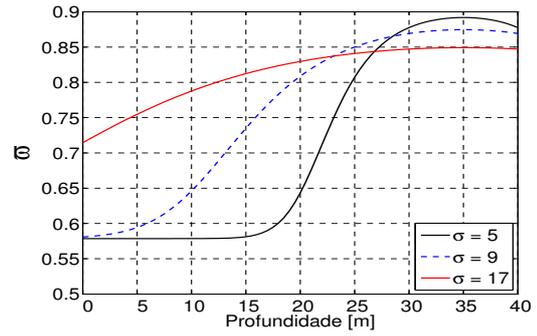


(b) Perfis gerados a partir da Eq. (4.4) para $\tau_m = 35$ m, com $\sigma = 5$ m, 9 m e 17 m, $h = 144 \text{ mg} \cdot \text{m}^{-2}$ e $\lambda_g = 560$ nm.

Figura 4.17 - Perfis de concentração de clorofila em $\tau_m = 5$ m e $\tau_m = 35$ m para $\sigma = 5$ m, 9 m e 17 m.

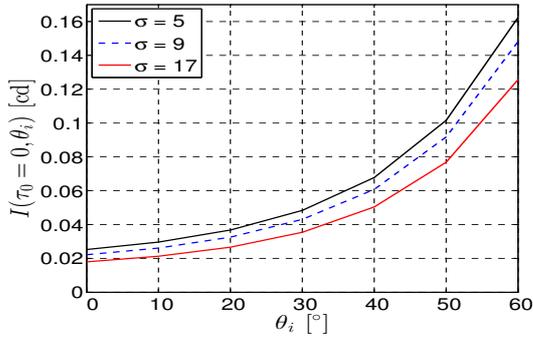


(a) Albedos de espalhamento simples resultantes dos perfis mostrados na Figura 4.17(a).

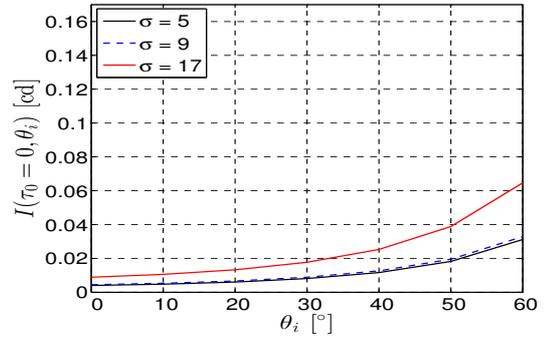


(b) Albedos de espalhamento simples resultantes dos perfis mostrados na Figura 4.17(b).

Figura 4.18 - Curvas do albedo de espalhamento simples geradas a partir dos perfis mostrados nas Figuras 4.17(a) e 4.17(b).



(a) Radiâncias emergentes calculadas pelo modelo, nas direções definidas na Eq. (4.10), para os perfis mostrados na Figura 4.17(a).



(b) Radiâncias emergentes calculadas pelo modelo, nas direções definidas na Eq. (4.10), para os perfis mostrados na Figura 4.17(b).

Figura 4.19 - Radiação emergente medida nas direções definidas na Eq. (4.10) levando em conta os perfis mostrados nas Figuras 4.17(a) e 4.17(b).

Os perfis mostrados nas Figuras 4.17 são simétricos em relação à profundidade de 20 metros, logo, os albedos de espalhamento simples (Figuras 4.18) também são simétricos com relação a essa mesma profundidade. No entanto, a atenuação do valor do albedo a partir dos 20 metros, mostrado na Figura 4.18(a), para o perfil que considera $\sigma = 5$ m, e a atenuação que ocorre nos últimos cinco metros para o albedo correspondente ao perfil obtido com $\sigma = 9$ m, não afetam o comportamento da radiação emergente. Veja Figura 4.19(a). Por outro lado, essa atenuação que ocorre nos metros iniciais para os valores de albedo mostrados na Figura 4.18(b) (observe os mesmos perfis obtidos ao se considerar $\sigma = 5$ m e $\sigma = 9$ m) afeta significativamente a radiação emergente. Observe a Figura 4.19(b), linhas preta e tracejada. Dessa forma, a radiação que deveria emergir na superfície trazendo consigo características

(informações) da distribuição da concentração de clorofila ao fundo, acaba sendo atenuada devido à redução do valor do albedo na metade superior da profundidade. Já o perfil que considera $\sigma = 17$ m, apesar de o pico estar profundo, evita que a radiação seja atenuada como nos casos anteriores.

O fato de as curvas de radiação emergente, que estão associadas aos perfis que consideram $\sigma = 5$ m e $\sigma = 9$ m, estarem muito próximas faz com que possíveis erros de medição, ou então, a adição de ruído, possam tornar impossível a correta associação de cada uma das curvas ao seu respectivo perfil. Veja Figuras 4.20. Por fim, observe que a concentração de clorofila de superfície para os perfis com desvio padrão cinco e sete são as mesmas (Figura 4.17(b)).

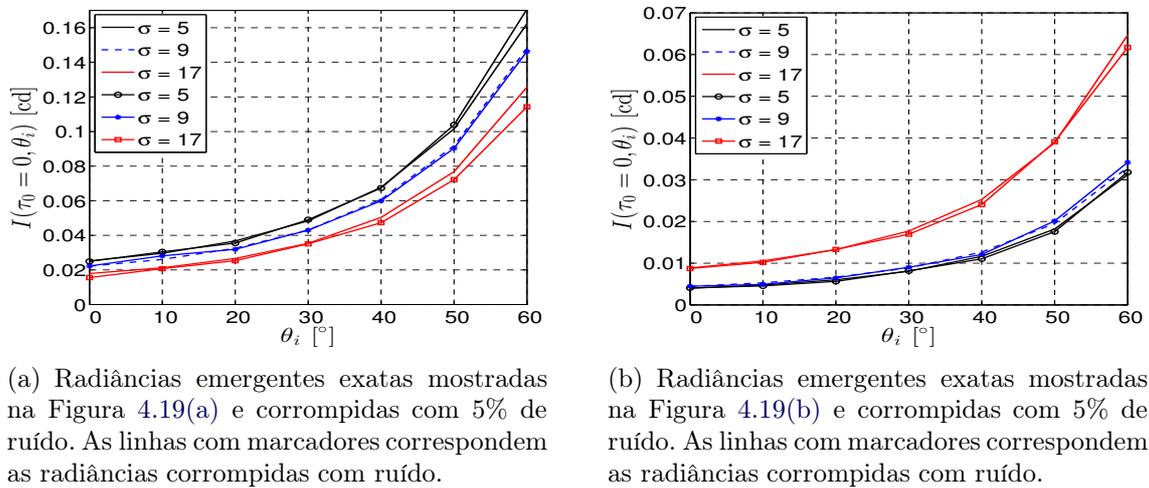


Figura 4.20 - Radiação emergente exata gerada a partir dos perfis mostrados nas Figuras 4.17(a) e 4.17(b), e a radiação emergente corrompida com 5% de ruído.

Portanto, para fazer essa classificação dos problemas, foi considerado um outro conjunto de problemas, semelhante aos já descritos. No entanto, nesse novo conjunto, foi adotado apenas uma região espacial na qual a concentração de clorofila assume um valor médio em todo o domínio (espacial). Veja Figura 4.21. Nesse caso, há apenas uma ETR, válida para toda região espacial, que modela a interação da radiação eletromagnética com o meio aquático. A Figura 4.22 mostra os problemas de cada um dos grupos considerados, para o caso multiregiões com perfil gaussiano, adicionando o caso região única com perfil constante. Note que foram repetidos os problemas já definidos no caso multiregiões para o caso região única, *i.e.*, o novo conjunto de problemas apresenta as mesmas características dos problemas anteriores, exceto pelo fato de que consideram uma única região e um perfil constante.

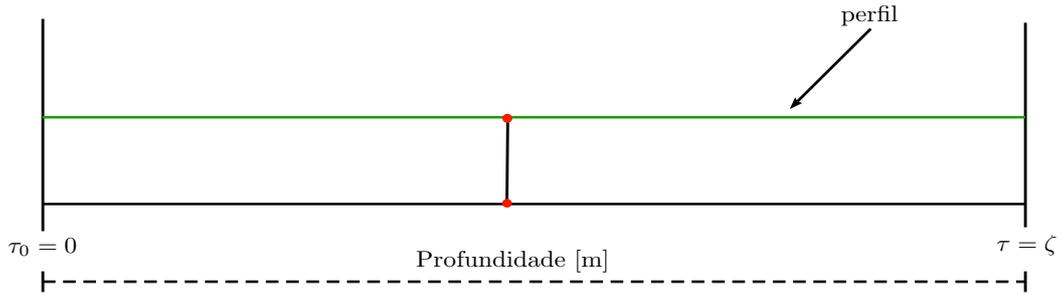


Figura 4.21 - Problema de região única com perfil constante.

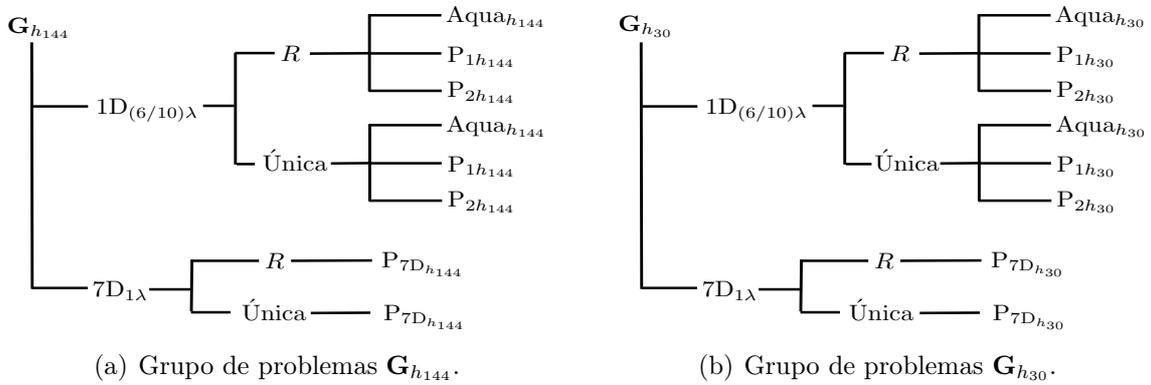


Figura 4.22 - Diagrama ilustrando a combinação dos problemas multiregiões e região única.

Visto que nesse novo conjunto de problemas o perfil de concentração de clorofila é constante (para cada problema), é necessário definir um intervalo de variação para esses valores. Para os problemas do grupo $\mathbf{G}_{h_{144}}$ foi adotado o intervalo $\overline{\text{chl}}_{\mathbf{G}_{h_{144}}} \in [0.01, 2.0] \text{ mg} \cdot \text{m}^{-3}$, e para os problemas do grupo $\mathbf{G}_{h_{30}}$, foi adotado o intervalo $\overline{\text{chl}}_{\mathbf{G}_{h_{30}}} \in [0.01, 1.0] \text{ mg} \cdot \text{m}^{-3}$. Em ambos os intervalos foi adotada uma discretização homogênea com tamanho de passo dado por $\Delta_{\overline{\text{chl}}} = 0.01 \text{ mg} \cdot \text{m}^{-3}$, gerando 200 valores discretos para os problemas do grupo $\mathbf{G}_{h_{144}}$, e 100 valores discretos para os problemas do grupo $\mathbf{G}_{h_{30}}$.

A diferença no processo de resolução (problema direto), entre os problemas desse novo conjunto com os anteriores, reside no fato de que, nos problemas do conjunto recente são eliminadas os dois primeiros “passos”, Veja Figura 4.14. Isso é possível porque a distribuição da concentração de clorofila, ao longo da profundidade, é homogênea e não depende dos valores de σ e τ_m .

As redes classificadoras foram treinadas para determinar a concentração de clorofila média, a partir da radiação emergente na superfície, obtidas a partir dos problemas

de região única. As entradas de cada padrão de treinamento, para essas redes classificadoras, são compostas apenas pela radiação emergente na superfície, seguindo as características de cada problema considerado.

A classificação dos problemas feita por essas redes, treinadas para determinar os valores de clorofila nos intervalos definidos anteriormente, foi realizada por meio de um valor de corte na concentração de clorofila estimada. Esse valor de corte elimina aqueles problemas que apresentam as características que impedem a correta estimativa (ou dentro de uma margem de erro aceitável) dos valores de desvio padrão e profundidade do pico.

A Figura 4.23 ilustra a atuação conjunta das redes. Após treinadas, as redes classificadoras recebem como entrada a radiação emergente na superfície, gerada a partir dos perfis gaussianos, respeitando as características de cada problema, e produzem uma estimativa para o valor médio da concentração de clorofila. Se, para uma determinada radiação emergente, esse valor médio estimado estiver abaixo de um limiar Υ , considera-se que não é possível determinar os valores de σ^* e τ_m^* e a resposta para o problema é o valor médio determinado por essa rede. Por outro lado, se o valor estiver acima do limiar, o processo passa para o passo seguinte, que consiste em fazer a estimativa de σ^* . Para a estimativa de σ^* é utilizado o mesmo padrão apresentado à rede classificadora, no entanto, adicionando o valor da concentração de clorofila de superfície correspondente. Por fim, a estimativa σ^* , feita pela rede RNA_σ , é adicionada ao padrão de entrada para a rede $\text{RNA}_{\tau_m^*}$, que determina o valor do pico τ_m^* .

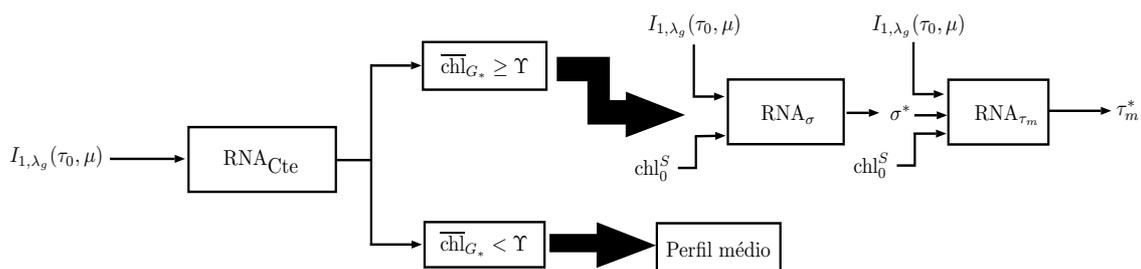


Figura 4.23 - Diagrama da atuação conjunta das redes de classificação e das redes que determinam os valores do desvio padrão e da profundidade do pico da concentração de clorofila.

4.3 Metodologia de Treinamento

Para RNAs do tipo PMC, as quais são utilizadas neste trabalho, o algoritmo de treinamento mais utilizado é o algoritmo *Backpropagation*, tendo em vista a facilidade

de implementação computacional. Por outro lado, esse algoritmo pode apresentar algumas deficiências ligadas à sua convergência, sendo elas: possibilidade de ficar presos em mínimos locais⁵, e com isso convergir para uma solução sub-ótima, ou então, apresentar uma convergência muito lenta (FAUSETT, 1994; HAYKIN, 2001). Estratégias que podem ser utilizadas a fim de minimizar e/ou eliminar essas deficiências é o uso de uma taxa de aprendizagem variável, que, por sua vez, dificulta a implementação computacional do algoritmo, e/ou o uso de uma constante de momento, entre outras técnicas que foram comentadas na Seção 3.6.

Em função das limitações do *Backpropagation*, o algoritmo de treinamento adotado nesta tese é outro. Como o treinamento de uma RNA pode ser interpretado como um problema de otimização com restrições, como comentado na Subseção 3.6.3, foi dada preferência, na realização deste trabalho, para o algoritmo descrito naquela subseção, o algoritmo quasi-Newton. Ainda, uma comparação feita entre os algoritmos quasi-Newton e *Backpropagation*, em um problema de ótica hidrológica, mostrou que o primeiro é superior quando comparado ao segundo (CORTIVO et al., 2012a).

Os passos que são descritos a seguir foram os mesmos para o treinamento de todas as redes envolvidas na busca de solução para os problemas descritos nesta tese. Ainda, o total de padrões em cada conjunto, seja ele de treinamento, teste ou validação, serão definidos no próximo capítulo.

Inicialmente, considere que cada padrão de entrada é representado por um vetor \vec{I}_i . Portanto, defina uma matriz de entrada \mathbb{E} , em que cada coluna é formada por cada padrão de entrada \vec{I}_i , *i.e.*,

$$\mathbb{E} = [\vec{I}_1, \vec{I}_2, \dots, \vec{I}_{N_p}],$$

em que N_p é o total de padrões do conjunto de treinamento. Considere também que cada padrão de saída seja representado por um vetor \vec{d} , *i.e.*,

$$\vec{d} = [\nu_1, \nu_2, \dots, \nu_{N_p}],$$

em que ν_i representa cada padrão de saída do conjunto de treinamento. Para as saídas calculadas pela rede, para cada padrão de entrada \vec{I}_i , é associado o vetor \vec{y} cujas entradas são os valores estimados pela rede, assim,

$$\vec{y} = [\nu_1^*, \nu_2^*, \dots, \nu_{N_p}^*]. \quad (4.12)$$

⁵Em otimização não-linear, não há garantias de que, ao se encontrar um ponto que forneça uma solução ótima, esse ponto seja um mínimo global.

Cada entrada do vetor \vec{d} depende de cada vetor de treinamento \vec{I} , e cada entrada do vetor \vec{y} depende, além de cada entrada do vetor de treinamento \vec{I} , dos pesos sinápticos, dos níveis de *bias*, e de cada parâmetro de cada função de ativação. Dessa forma, seja \mathbb{W} uma matriz que contém todos os pesos sinápticos, \mathbb{B} uma matriz/vetor que contém todos os níveis de *bias*, e Γ uma matriz que contém todos os parâmetros das funções de ativação, tem-se que a Eq. (4.12) é escrita como

$$\vec{y}(\vec{I}_p, \mathbb{W}, \mathbb{B}, \Gamma) = [\nu_1^*, \nu_2^*, \dots, \nu_{N_p}^*], \quad p = 1, 2, \dots, N_p.$$

Diante disso, é formulado o problema de treinamento (otimização), com restrições, baseado na minimização de um funcional de diferenças quadráticas definido entre os padrões de saída e as saídas calculadas pela rede, para cada padrão de entrada, dado por⁶

$$\mathcal{J}(\mathbb{E}, \mathbb{W}, \mathbb{B}, \Gamma) = \min \frac{1}{2} \sum_{p=1}^{N_p} \left\| d(\vec{I}_p) - y(\vec{I}_p, \mathbb{W}, \mathbb{B}, \Gamma) \right\|_2^2, \quad (4.13)$$

e sujeito às restrições

$$\begin{aligned} w_{\min} &\leq w \leq w_{\max} \\ b_{\min} &\leq b \leq b_{\max} \\ 0 &< \gamma \leq \gamma_{\max} \end{aligned} .$$

Quando se utiliza a função de ativação linear, a última restrição anterior pode ser substituída pela restrição definida na Eq. (3.10).

Nessa formulação, o treinamento (busca do valor mínimo para o funcional definido anteriormente) é feito no modo *batch*, *i.e.*, todos os padrões de treinamento são apresentados à rede para depois ocorrer a atualização das variáveis livres da rede. A busca do mínimo foi feita pela rotina E04UCF da biblioteca NAG (*Numerical Algorithms Group*) (NAG, 1995), a qual implementa o método quasi-Newton, e que trabalha em conjunto com a rotina OBJFUN, que é programada pelo usuário. Os detalhes da rotina podem ser encontrados na Subseção 3.6.3 e os detalhes do cálculo dos gradientes do funcional, definido pela Eq. (4.13), podem ser encontrados na Seção D.3 do Apêndice D. Já os detalhes sobre implementação da rotina OBJFUN serão descritos na seção seguinte e no início do próximo capítulo.

⁶No funcional definido neste capítulo não é apresentada a somatória em k como aparece na Eq. (3.8), pois as redes consideradas têm apenas um neurônio na camada de saída.

4.4 Rede Neural Implementada

A RNA utilizada nesta tese foi toda implementada em FORTRAN 90. A rede é dinâmica, sendo assim, é possível definir várias configurações sem a necessidade de “reprogramação” do algoritmo (rede). Foi desenvolvida, também, uma interface gráfica em linguagem C++, utilizando a biblioteca wxWidgets ([WXWIDGETS, 2013](#)), a fim de facilitar a entrada dos parâmetros necessários para a configuração da rede. A Figura 4.24 ilustra a interface criada, e a Figura 4.25 apresenta um diagrama em blocos (simplificado) do funcionamento do algoritmo (rede) desenvolvido(a).

As principais características da rede são: número variável de entradas e de saídas; inclusão de até duas camadas ocultas (sem contar as camadas de entrada e de saída), com número de neurônios em cada camada também variável; escolha da função de ativação para cada camada; escolha da quantidade de padrões para os conjunto de treinamento, teste e validação; pesos, *bias* e parâmetro de cada função de ativação lidos de arquivos ou gerados aleatoriamente (inclui a possibilidade de normalização dos pesos e *bias*); escolha do algoritmo de treinamento; escolha do parâmetro das funções de ativação fixo ou otimizado durante o processo de treinamento; para o algoritmo *Backpropagation*, possibilidade das taxas de aprendizagem para pesos, *bias* e parâmetros serem diferentes entre si (o mesmo é válido para a constante de momento); possibilidade de inclusão de até quatro níveis de ruído, bem como a quantidade de padrões em cada nível; procedimento de validação cruzada; cálculo do gradiente (somente possível para a redes com uma camada oculta com a função sigmoide e função linear na camada de saída); leitura de padrões adicionais para testar desempenho da rede, e possibilidade de normalização dos dados de entrada e/ou saída.

NEURONS AT INPUT AND OUTPUT LAYERS

Neurons Number
 In: 9 Excl. 2 Out: 1

HIDDEN LAYERS, ACTIVATION FUNCTIONS

Hidden Layers: 1
 Neurons Number
 Layer 1: 79 Layer 2:

Activation Functions
 SIG Layer 1: Layer 2: LIN

TRAINING, VALIDATION AND TEST

Total Patterns: 906
 Training: 100.0D0 Test: 0.0D0 Validation: 0.0D0

WEIGHTS AND BIASES OPTIONS

Initial Weights and Biases Random?
 Yes No

Weights Mean Equals 0?
 Yes No

Standard Deviation Equals 1?
 Yes No

Weights and Bias Divisor: 1.0D0

OPTIMIZATION METHOD

Training Algorithm: QN
 Constraints: Weights, Biases and Slopes
 Min. W. and B.: 100.0D0
 Max. W. and B.: 100.0D0
 Maximum Slope: 2.0D0

Initial Slopes Random?
 Yes No

All Initial Slopes Equals 1?
 Yes No

Slope Optimization?
 Yes No

Max. Iterations QN: 300
 Learning, Epochs and Tolerance
 Learning Rate
 Weights: Biases: Slopes:
 Max. Epochs:
 Tolerance:

Random Presentation BP
 Yes No

Learning Rule
 Generalized Delta Rule
 Momentum Constant
 Weights: Biases: Slopes:

NOISE IN DATA TRAINING

Insert Noise:
 Yes No

Noise Levels: 3

Noise Levels
 Value 1: 1.0D0 Value 2: 2.0D0
 Value 3: 5.0D0 Value 4:

Patterns Percentage
 Value 1: 30.0D0 Value 2: 30.0D0
 Value 3: 40.0D0 Value 4:

Hits Tolerance
 Value 1: 1.0D0 Value 2: 2.0D0
 Value 3: 5.0D0 Value 4:

Noiseless: 2.0D0 Tolerance: 0.001D0

CROSS-VALIDATION OPTIONS

Cross-Validation
 Yes No

Test Patterns Extra Patterns

Epochs Number: 10

Use Data Without Noise
 Yes No

FIX SEED

Save Objf Values
 Use Fix Seed? Save Objf?
 Yes No

GRADIENTE

Gradiente?
 Yes No

ADDITIONAL PATTERNS AND NOISE

Extra Patterns to Test
 Yes No

Extra Noise to Test
 Yes No

Patterns: 660 Noise: 3.0D0
 Hit Tolerance Extra: 3.0D0

DATA NORMALIZATION

INPUT DATA

Norm Input
 Yes No Norm Type
 NL NP

Read Input Norms
 Yes No
 [0 1]

X Min: 0.002390 X Max: 0.015855

OUTPUT DATA

Norm Output
 Yes No Norm Type
 NL NP

Read Output Norms
 Yes No
 [0 1]

YY Min: 4.0D0 YY Max: 18.0D0
 Y Min: Y Max:

Name of the Output Neurons
 7d_20_30_zm

Figura 4.24 - Interface gráfica para a configuração da RNA implementada.

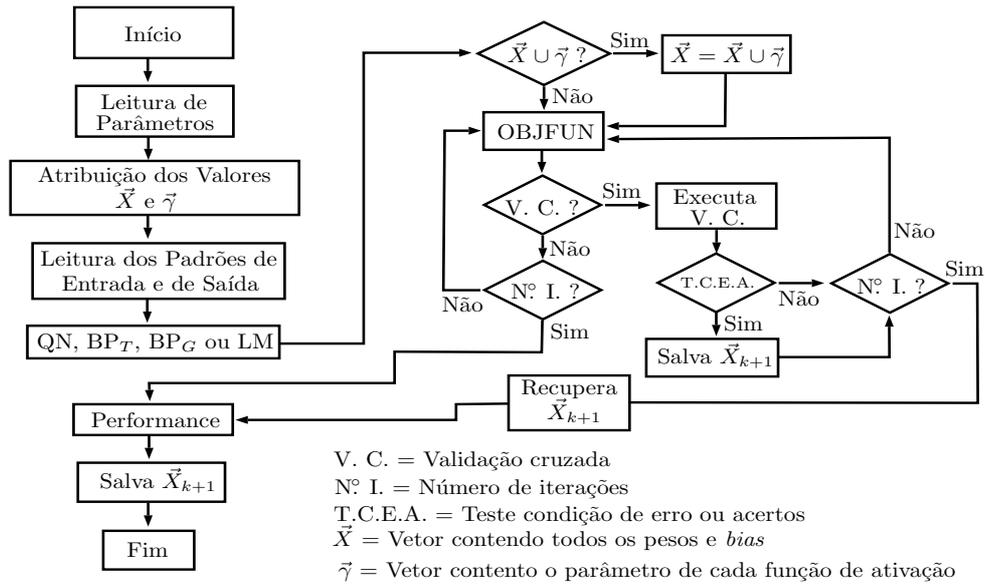


Figura 4.25 - Diagrama em blocos do funcionamento da RNA.

5 RESULTADOS E DISCUSSÕES

Além da metodologia descrita no capítulo anterior, cujos resultados são apresentados a seguir, outras metodologias foram testadas. Basicamente, além da metodologia adotada, foram testadas outras três, no entanto essas três não produziram resultados satisfatórios e, por isso, serão descritas no Apêndice E.

Antes de iniciar a descrição do treinamento e dos resultados obtidos com a metodologia proposta, é importante apresentar algumas informações que são válidas para toda a descrição que virá a seguir. A Tabela 5.1 sumariza os problemas abordados nesta tese, a quantidade de redes utilizadas em cada problema e o valor de corte, denotado por Υ , adotado para a classificação dos problemas para os grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

Tabela 5.1 - Número de redes utilizadas para cada problema de cada classe, total de perfis para multiregiões e região única, e o valor de concentração de clorofila de corte adotado para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

Grupo	Classes	Regiões	Problemas	Redes	Perfis	Υ [$\text{mg} \cdot \text{m}^{-3}$]		
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	20	Aqua _{h_{144}}	2	320	1.2		
			P _{1h_{144}}	2				
			P _{2h_{144}}	2				
		1	Aqua _{h_{144}}	1				
			P _{1h_{144}}	1			200	–
			P _{2h_{144}}	1				
	7D _{1λ}	20	P _{7Dh_{144}}	2	320	1.2		
		1	P _{7Dh_{144}}	1	200	–		
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	20	Aqua _{h_{30}}	2	320	0.11		
			P _{1h_{30}}	2				
			P _{2h_{30}}	2				
		1	Aqua _{h_{30}}	1				
			P _{1h_{30}}	1			100	–
			P _{2h_{30}}	1				
	7D _{1λ}	20	P _{7Dh_{30}}	2	320	0.11		
		1	P _{7Dh_{30}}	1	100	–		

Observe que, para os problemas em que foram consideradas 20 regiões espaciais, cada um apresenta duas redes, pois cada uma delas estima um dos parâmetros do perfil da concentração de clorofila. Nos problemas em que é considerada apenas uma região espacial, uma única rede é suficiente, pois se pretende determinar apenas o valor da concentração de clorofila (médio) para a filtragem dos problemas. Os itens que são listados a seguir apresentam as informações que são comuns aos treinamentos das

redes e/ou aos problemas de cada grupo:

- a) para cada problema com 20 regiões foram treinadas duas redes, *i.e.*, uma para determinar σ^* e outra para determinar τ_m^* . As características dos padrões de entrada de cada problema são diferentes. Veja Subseções 4.1.1 e 4.1.2. Ainda, as redes treinadas para resolver os problemas do grupo $\mathbf{G}_{h_{144}}$ não são capazes de resolver os problemas do grupo $\mathbf{G}_{h_{30}}$ e vice-versa;
- b) a descrição dada para o treinamento das redes é a mesma tanto para as redes que irão atuar no grupo de problemas $\mathbf{G}_{h_{144}}$ como no $\mathbf{G}_{h_{30}}$, sendo assim, os níveis de ruído utilizados, quantidade de padrões para treino, teste e validação serão os mesmos. Há uma pequena diferença no treinamento das redes que farão a classificação dos problemas, no entanto a Subseção 5.1.1 trará os detalhes;
- c) todas as redes foram treinadas pelo método quasi-Newton, com cálculo dos gradientes do funcional de diferenças quadráticas feito de forma analítica. O método quasi-Newton foi descrito na Subseção 3.6.3, a metodologia adotada para o treinamento foi descrita na Seção 4.3 e o cálculo dos gradientes são explicados na Seção D.3 do Apêndice D. O critério de parada do treinamento se deu pelo procedimento de validação cruzada, explicado na Subseção 3.5.2. O processo de validação cruzada foi feito a cada dez chamadas da função objetivo (dez épocas);
- d) em todas as redes, o parâmetro de cada função de ativação foi considerado variável, assim, o algoritmo de treinamento adotado determinou o “melhor” valor para esse parâmetro;
- e) todas as redes utilizam como função de ativação a função sigmoide nos neurônios da camada oculta e a função linear no neurônio da camada de saída. Todas as redes utilizadas têm um neurônio na camada de saída; o número de neurônios na camada oculta foi determinado de modo *ad hoc*;
- f) os padrões de entrada para as redes classificadoras não foram normalizados, no entanto, os padrões de entrada para as redes que determinam o valor do desvio padrão e a profundidade do pico da concentração foram normalizados. Os valores de saída para todas as redes não foram normalizados;
- g) a quantidade de padrões em cada conjunto (treinamento, teste e validação), os níveis de ruído utilizados, bem como a quantidade de padrões em cada um dos níveis de ruído foram definidos experimentalmente (*ad hoc*).

A normalização dos padrões de entrada, quando feita, foi para o intervalo $[0, 1]$ e se deu através da equação normalizadora definida por

$$x_{ij}^{\text{norm}} = \frac{x_{ij} - \min(\mathbb{X})}{\max(\mathbb{X}) - \min(\mathbb{X})}, \quad (5.1)$$

em que \mathbb{X} é uma matriz/vetor que contém todos os valores a serem normalizados, x_{ij} são as entradas, $\min(\mathbb{X})$ e $\max(\mathbb{X})$ são as entradas de valores mínimo e máximo, respectivamente, e x_{ij}^{norm} são as entradas normalizadas.

Os valores mínimos e máximos para cada problema são listados na Tabela 5.2, e foram adotados para normalizar os padrões de entrada (apenas para a radiação emergente) para o treinamento das redes RNA_σ e RNA_{τ_m} . Para obter esses valores, todos os padrões que seriam utilizados para treinar, testar ou validar a rede, foram corrompidos com o nível de ruído $\varepsilon_4 = 5\%$, para, então, determinar os valores apresentados na tabela. Os valores mínimos e máximos adotados para normalizar o desvio padrão são, respectivamente, $\sigma_{\min} = 4$ m e $\sigma_{\max} = 18$ m. Essa normalização no valor de σ foi feita apenas quando esse parâmetro foi considerado como entrada para a rede que determina τ_m^* . Como a concentração de clorofila não depende do comprimento de onda (veja Eq. (4.11)), tem-se um valor mínimo e máximo para os problemas do grupo $\mathbf{G}_{h_{144}}$ e outros dois para os problemas do grupo $\mathbf{G}_{h_{30}}$. Os valores são: $\text{chl}_{0_{\min}}^S = 0.38987 \text{ mg} \cdot \text{m}^{-3}$ e $\text{chl}_{0_{\max}}^S = 12.61890 \text{ mg} \cdot \text{m}^{-3}$ para os problemas do grupo $\mathbf{G}_{h_{144}}$, e $\text{chl}_{0_{\min}}^S = 0.17933 \text{ mg} \cdot \text{m}^{-3}$ e $\text{chl}_{0_{\max}}^S = 2.82201 \text{ mg} \cdot \text{m}^{-3}$ para os problemas do grupo $\mathbf{G}_{h_{30}}$. Esses valores mínimos e máximos foram obtidos de modo similar aos valores mínimos e máximos obtidos para a radiação emergente.

Tabela 5.2 - Valores mínimos e máximos utilizados para a normalização da radiação emergente para os problemas multiregiões.

Grupo	Classes	Problemas	$I_{1,\lambda_q}^{\min}(\tau_0, \mu)$ [cd]	$I_{1,\lambda_q}^{\max}(\tau_0, \mu)$ [cd]
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	Aqua _{h_{144}}	0.00118	0.02918
		P _{1h_{144}}	0.00214	0.02988
		P _{2h_{144}}	0.01827	0.19209
	7D _{1λ}	P _{7Dh_{144}}	0.00984	0.20101
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	Aqua _{h_{30}}	0.00003	0.01872
		P _{1h_{30}}	0.00003	0.01619
		P _{2h_{30}}	0.00325	0.12336
	7D _{1λ}	P _{7Dh_{30}}	0.00239	0.01586

5.1 Treinamento das Redes

O fato de se estar utilizando um modelo matemático para determinar a radiação emergente, após a interação da radiação incidente com o meio aquático, resulta em uma radiação emergente “limpa”, *i.e.*, os valores determinados pelo modelo não contêm possíveis erros de medição e/ou valores espúrios. Em geral, dados calculados por modelos, e utilizados como medidas *in situ*, são chamados de dados sintéticos.

O método adotado pela comunidade científica para tornar esses dados sintéticos próximos às verdades é a adição de ruído, *i.e.*, os dados retornados pelo modelo são corrompidos com algum nível de ruído. Dessa forma, as medidas de radiação emergente, calculadas pelo modelo adotado nesta tese, serão corrompidas com os quatro níveis de ruído mostrados na Tabela 5.3. Desses, três são utilizados no treinamento e teste, e um no processo de validação. O nível de ruído presente no processo de validação não está presente no processo de treinamento e teste das redes. Essa estratégia é utilizada a fim de garantir uma melhor generalização por parte das redes.

Tabela 5.3 - Níveis de ruído adotados para os conjuntos de treinamento, teste e validação.

	ε_1	ε_2	ε_3	ε_4
Ruído	1%	2%	3%	5%
Treinamento	✓	✓	–	✓
Teste	✓	✓	–	✓
Validação	–	–	✓	–

5.1.1 Treinamento das Redes de Filtragem

No treinamento das redes responsáveis pela filtragem dos problemas, foram utilizados todos os padrões do conjunto de treinamento. Nesse caso, os conjuntos de treinamento, teste e validação são iguais, mudando apenas o nível de ruído considerado. Os padrões do conjunto de treinamento foram distribuídos em três partes, e cada uma delas recebeu um dos níveis de ruído, como mostrado na Tabela 5.4. No conjunto de teste, utilizado para o processo de validação cruzada, todos os padrões foram corrompidos com os níveis de ruído indicados nessa mesma tabela. Para a validação, o conjunto de padrões foi corrompido com o nível de ruído ε_3 . Note que todos os padrões estão presentes nos conjuntos, mudando apenas o nível de ruído. No entanto, observe que, por exemplo, 70% dos padrões do conjunto de treinamento não foram corrompidos com os níveis de ruído ε_2 e ε_4 . Portanto, durante o treinamento, um determinado padrão p pode ter sido corrompido com o ruído

ε_1 , mas, no processo de validação cruzada (conjunto de teste), esse padrão p vai ser corrompido com os níveis ε_2 e ε_4 e, então, apresentado à rede. Já na validação, esse mesmo padrão p será corrompido com o nível de ruído ε_3 , o qual não esteve presente no treinamento e teste.

Tabela 5.4 - Percentagens de padrões de cada conjunto em cada nível de ruído para o treinamento das redes responsáveis pela filtragem dos problemas.

Conjunto	Ruído			
	ε_1	ε_2	ε_3	ε_4
Treinamento	30%	30%	–	40%
Teste	100%	100%	–	100%
Validação	–	–	100%	–

A Tabela 5.5 apresenta o número de neurônios nas camadas de entrada e oculta para cada rede de filtragem. O número de neurônios na camada de entrada é definido pelas características dos respectivos problemas, e os neurônios da camada oculta foram determinados de modo *ad hoc*.

Tabela 5.5 - Número de entradas e de neurônios ocultos para cada rede de filtragem.

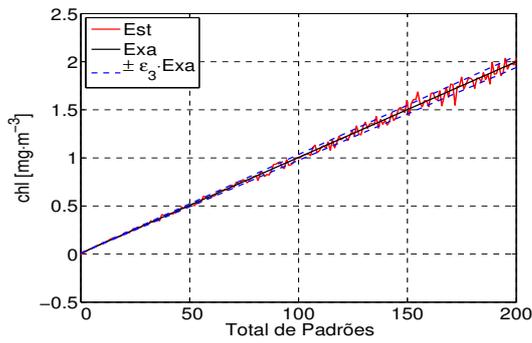
Grupo	Classes	Problemas	NE	NNO
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	Aqua _{h_1}	10	
		P _{1h_{144}}	6	25
		P _{2h_{144}}	6	
	7D _{1λ}	P _{7Dh_{144}}	7	25
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	Aqua _{h_{30}}	10	20
		P _{1h_{30}}	6	
		P _{2h_{30}}	6	25
	7D _{1λ}	P _{7Dh_{30}}	7	25

NE = Número de Entradas

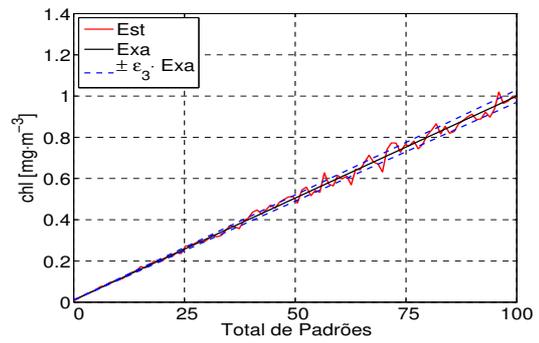
NNO = Número de Neurônios Ocultos

A título de ilustração, a Figura 5.1(a) mostra a estimativa feita pela rede associada ao problema P_{1 h_{144}} , de região única, para padrões de entrada corrompidos com o nível de ruído ε_3 . Já a Figura 5.1(b), mostra a estimativa feita pela rede associada ao problema P_{1 h_{30}} , de região única, também para padrões de entrada corrompidos com ε_3 . Ainda, a Figura 5.1(c) mostra um recorte da Figura 5.1(a), na região em que foi definido o limiar de classificação (filtragem) para os problemas do grupo $\mathbf{G}_{h_{144}}$, e a Figura 5.1(d) mostra um recorte da Figura 5.1(b), na região em que foi definido o limiar de classificação para os problemas para o grupo $\mathbf{G}_{h_{30}}$. Os limiares são $\Upsilon = 1.2 \text{ mg} \cdot \text{m}^{-3}$ para o grupo $\mathbf{G}_{h_{144}}$ e $\Upsilon = 0.11 \text{ mg} \cdot \text{m}^{-3}$ para o grupo $\mathbf{G}_{h_{30}}$.

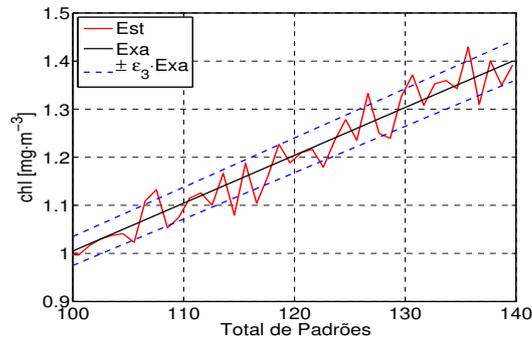
Veja Tabela 5.1. A escolha desses limiares foi baseada no princípio de que a concentração de clorofila média (ao longo da profundidade) deveria ser $1.0 \text{ mg} \cdot \text{m}^{-3}$ para os problemas do grupo $\mathbf{G}_{h_{144}}$, e de $0.1 \text{ mg} \cdot \text{m}^{-3}$ para os problemas do grupo $\mathbf{G}_{h_{30}}$, adicionadas as respectivas concentrações de clorofila de fundo. Vale observar, também, que a modificação desses valores poderá influenciar nos resultados apresentados nesta tese. A linha preta representa o dado exato, a linha vermelha representa a estimativa feita pela rede, e as linhas tracejadas representam o intervalo de tolerância, definido como: $\text{chl}_0^S - \varepsilon_3 \cdot \text{chl}_0^S \leq \text{chl}_0^{S\text{Est}} \leq \text{chl}_0^S + \varepsilon_3 \cdot \text{chl}_0^S$.



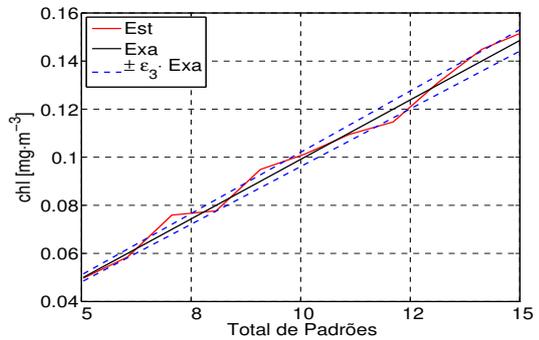
(a) Estimativa feita pela rede associada ao problema $P_{1h_{144}}$



(b) Estimativa feita pela rede associada ao problema $P_{1h_{30}}$



(c) Recorte da figura (a) em torno da região do limiar.



(d) Recorte da figura (b) em torno da região do limiar.

Figura 5.1 - Estimativas obtidas pelas redes de filtragem quando apresentados os padrões de entrada obtidos a partir dos problemas $P_{1h_{144}}$ e $P_{1h_{30}}$ com uma única região espacial e corrompidos com o nível de ruído ε_3 .

Os resultados obtidos pelas redes de classificação, quando aplicados os padrões obtidos a partir dos problemas com 20 regiões e corrompidos com o nível de ruído ε_3 , são mostrados na Tabela 5.6 e 5.7, para os grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$, respectivamente. Essas redes foram treinadas com a radiação emergente (padrões) obtidas(os) a partir dos problemas que consideram uma região espacial, no entanto são aplicados nessas redes a radiação emergente (padrões) obtida(os) a partir dos problemas que

consideram 20 regiões espaciais.

Tabela 5.6 - Quantidade de perfis selecionados e não selecionados pelas redes de filtragem, quando aplicados os padrões pertencentes aos problemas do grupo $\mathbf{G}_{h_{144}}$, que consideram 20 regiões espaciais e corrompidos com o ruído ε_3 .

Regiões	Classes	Problemas	$\Upsilon \geq 1.2$	$\Upsilon < 1.2$	Total
20	1D _{(6/10)λ}	Aqua _{h_{144}}	162 [50.62%]	158 [49.38%]	320
		P _{1h_{144}}	163 [50.93%]	157 [49.07%]	
		P _{2h_{144}}	158 [49.38%]	162 [50.62%]	
	7D _{1λ}	P _{7Dh_{144}}	170 [53.12%]	150 [46.88%]	

Tabela 5.7 - Quantidade de perfis selecionados e não selecionados pelas redes de filtragem, quando aplicados os padrões pertencentes aos problemas do grupo $\mathbf{G}_{h_{30}}$, que consideram 20 regiões espaciais e corrompidos com o ruído ε_3 .

Regiões	Classes	Problemas	$\Upsilon \geq 0.11$	$\Upsilon < 0.11$	Total
20	1D _{(6/10)λ}	Aqua _{h_{30}}	190 [59.38%]	130 [40.62%]	320
		P _{1h_{30}}	198 [61.88%]	122 [38.12%]	
		P _{2h_{30}}	195 [60.94%]	125 [39.06%]	
	7D _{1λ}	P _{7Dh_{30}}	217 [67.81%]	103 [32.19%]	

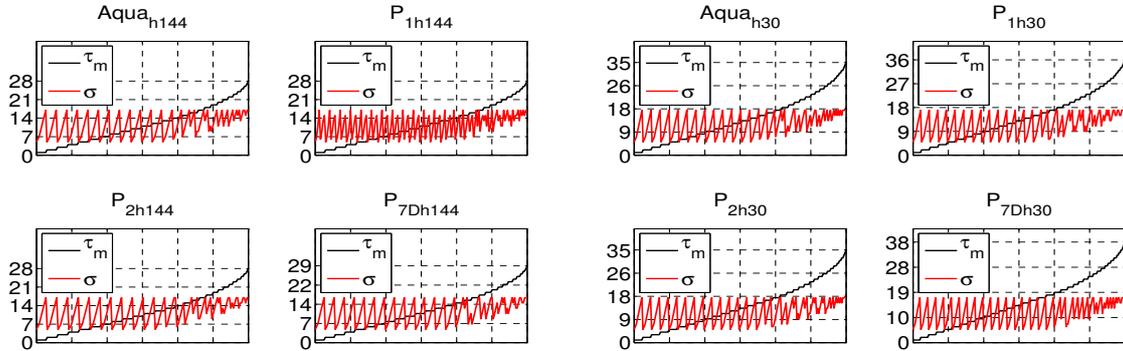
É possível observar que foram classificados em torno de 50% dos padrões (perfis) pertencentes ao grupo de problemas $\mathbf{G}_{h_{144}}$, *i.e.*, a radiação aplicada às redes de filtragem gerou uma concentração de clorofila média acima do limiar definido. Já no grupo de problemas $\mathbf{G}_{h_{30}}$, a quantidade de padrões selecionados ficou em torno de 60%.

As Figuras 5.2(a) e 5.2(b) mostram as combinações dos valores de τ_m com os valores de σ para os problemas classificados dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$, respectivamente. Observe que à medida que o pico da concentração de clorofila se localiza mais ao fundo (eixo das ordenadas), apenas aqueles perfis que contêm o desvio padrão maior são selecionados. Isso ocorre porque, para os valores de desvio padrão menores, a radiação emergente acaba sendo atenuada, contendo informações (em maior parte) da coluna d'água superficial, enquanto que quando o desvio padrão é maior, esse faz com que o perfil seja mais “aberto” e, então, a radiação que emerge da superfície acaba levando consigo informações acerca desse perfil.

5.1.2 Treinamento das Redes RNA $_{\sigma}$ e RNA $_{\tau_m}$

Como foi comentado, há uma diferença entre os padrões de entrada da RNA $_{\sigma}$ e a RNA $_{\tau_m}$, porém o treinamento de ambas as redes foi feito da mesma forma. A

diferença está nas entradas da rede RNA_{τ_m} , pois contêm a estimativa feita pela rede RNA_{σ} . Veja Figura 4.16.



(a) Combinações da profundidade do pico com os valores de desvio padrão classificados para os problemas do grupo $\mathbf{G}_{h_{144}}$.

(b) Combinações da profundidade do pico com os valores de desvio padrão classificados para os problemas do grupo $\mathbf{G}_{h_{30}}$.

Figura 5.2 - Combinações de valores de profundidade do pico com valores de desvio padrão que foram classificados para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

Como visto, os resultados apresentados nas Tabelas 5.6 e 5.7 mostram que, do total de 320 padrões (perfis) de cada problema, em torno de 50% dos problemas do grupo $\mathbf{G}_{h_{144}}$ e 60% dos problemas do grupo $\mathbf{G}_{h_{30}}$ foram classificados. Para o treinamento das redes, esses padrões classificados foram divididos em duas partes. A primeira foi constituída por 70% desses padrões (de 50% ou 60%), e formou o conjunto de treinamento. A segunda, com os 30% restantes, formou os conjuntos de teste e validação. A Tabela 5.8 mostra a quantidade de padrões em cada conjunto para cada um dos problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$. Os padrões do conjunto de teste e de validação não estão presentes no conjunto de treinamento. O que “diferencia” os padrões dos conjuntos de teste e validação é o nível de ruído considerado.

O fato de o número de padrões ser relativamente pequeno no conjunto de treinamento, e menor ainda nos conjuntos de teste e validação e, ainda, levando em conta os três níveis de ruído adotados, faz com que a quantidade de padrões em cada nível de ruído seja pequena¹. Portanto, optou-se por fazer réplicas desses padrões, tanto para o conjunto de treinamento quanto para os conjuntos de teste e validação. Assim, para o conjunto de treinamento, foram feitas seis réplicas, e para os conjuntos de teste e validação, foram feitas dez. Foram feitas mais réplicas no conjunto de

¹Essa quantidade pequena de padrões pode não ser representativa do problema, o que poderá prejudicar o treinamento da rede. Ainda, durante os experimentos, notou-se que o problema apresenta uma certa sensibilidade ao ruído, provavelmente pelo fato de que diferentes valores de σ e τ_m geram radiações emergentes muito próximas e/ou a chl_0^S também próximos.

teste e validação porque estes contêm menos padrões. A Tabela 5.9 mostra o total de padrões considerados em cada problema após as réplicas.

Tabela 5.8 - Total de padrões para os conjuntos de treinamento, teste e validação para as redes RNA_σ e RNA_{τ_m} .

Grupo	Classes	Problemas	Treinamento	Teste e Validação	Total
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	Aqua _{h_{144}}	113	49	162
		P _{1h_{144}}	114	49	163
		P _{2h_{144}}	110	48	158
	7D _{1λ}	P _{7Dh_{144}}	118	52	170
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	Aqua _{h_{30}}	133	57	190
		P _{1h_{30}}	138	60	198
		P _{2h_{30}}	136	59	195
	7D _{1λ}	P _{7Dh_{30}}	151	66	217

Tabela 5.9 - Total de padrões nos conjuntos de treinamento, teste e validação, após as réplicas, para as redes RNA_σ e RNA_{τ_m} .

Grupo	Classes	Problemas	Treinamento	Teste e Validação	Total
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	Aqua _{h_{144}}	678	490	1168
		P _{1h_{144}}	684	490	1174
		P _{2h_{144}}	660	480	1140
	7D _{1λ}	P _{7Dh_{144}}	708	520	1228
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	Aqua _{h_{30}}	798	570	1368
		P _{1h_{30}}	828	600	1428
		P _{2h_{30}}	816	590	1406
	7D _{1λ}	P _{7Dh_{30}}	906	660	1566

Os padrões do conjunto de treinamento foram distribuídos em três níveis de ruído conforme indicado na Tabela 5.4. Para o conjunto de teste, todos os padrões foram corrompidos com os três níveis de ruído indicados, e todo o conjunto de validação foi corrompido com o nível de ruído ε_3 , o qual não esteve presente durante o processo de treinamento e teste da rede. No treinamento dessas redes, foi utilizada a metodologia descrita na Seção 4.3 e o algoritmo descrito na Seção 4.4. Para o treinamento da rede RNA_{τ_m} , a entrada correspondente ao valor do desvio padrão, que foi determinado a partir da radiação emergente e da concentração de clorofila de superfície, pela rede RNA_σ , não foi corrompido com ruído, apenas normalizado para o intervalo $[0, 1]$ segundo a Eq. (5.1). Os resultados obtidos sobre o conjunto de validação para cada um dos problemas são mostrados a seguir.

5.2 Resultados Obtidos

A Tabela 5.10 resume os resultados obtidos pelas redes ao final do treinamento, sobre os conjuntos de teste e de validação, para os grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$. O significado de cada “variável” apresentada na tabela é mostrado na nota de rodapé da tabela.

O término do treinamento se deu por um número máximo de iterações (não são épocas como no caso do algoritmo *Backpropagation*) do algoritmo de treinamento, e está relacionado com um contador de iterações da biblioteca utilizada. O número de iterações foi definido (*ah hoc*) em 300, visto que os experimentos mostram que não havia necessidade de um número maior. Por outro lado, o número de chamadas da função objetivo, definida na Eq. (4.13), é diferente e maior que o número máximo de iterações. O número de chamadas da função objetivo é o número de épocas, pelo fato de que, a cada chamada dessa função, todos os padrões são apresentados à rede (treinamento feito no modo *batch*); sendo assim, o número de épocas é variável e muda do treinamento de uma rede para outra. O total de épocas necessárias para o treinamento de cada rede é mostrado na coluna NET (Número de Épocas Total) da Tabela 5.10.

Mesmo utilizando o procedimento de validação cruzada, todos os treinamentos seguiram até atingir o valor máximo de iterações definido anteriormente. No entanto, a cada dez chamadas (dez épocas) da função objetivo, foi realizado o procedimento de validação cruzada. A cada procedimento de validação cruzada, o erro absoluto total, calculado sobre todos os padrões do conjunto de teste nos três níveis de ruído, e o número de acertos são comparados com os (erro e/ou acertos) do procedimento de validação cruzada anterior. Quando o valor do erro for inferior ao anterior, ou então, o número de acertos for superior, o conjunto de variáveis (pesos, *bias* e parâmetro de cada função de ativação) é salvo e o treinamento prossegue. Ao final do treinamento, aquele conjunto de variáveis que produziu o menor erro absoluto total, ou maior número de acertos, é recuperado e adotado como valores ótimos para as variáveis da rede. O número de épocas necessárias para que a rede apresente a melhor generalização é mostrado na coluna NEV (Número de Épocas para Validação) da Tabela 5.10.

O Total de Padrões para Teste, coluna TPT da Tabela 5.10, é resultado do número de padrões do conjunto de teste, mostrado na Tabela 5.9, multiplicado por três, visto que são utilizados três níveis de ruído.

Tabela 5.10 - Resultados obtidos pelas redes sobre o conjunto de validação para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

G	Classes	P	R	NE	NNO	NET	NEV	TPT	TxT	TPV	TxV	TxSV	TTT	TV	EQM	MPB	
$\mathbf{G}_{h_{144}}$	Aqua $_{h_{144}}$	RNA_{σ}		11	76	1682	570	1470	62.59	490	59.59	47.55	48.40	19.65	1.89	0	
		$\text{RNA}_{\bar{r}_m}$		12	111	2167	260	63.95	65.10	70.41	70.00	55.92	32.31	15.29	1.30	-	
	$1D_{(6/10)\lambda}$	RNA_{σ}		7	80	1470	610	1470	75.31	490	72.45	58.96	37.29	26.61	7.18	1.94	-
		$\text{RNA}_{\bar{r}_m}$		8	78	1457	410	1440	58.13	480	56.46	63.85	42.69	55.54	12.09	2.30	-
	$P_{2h_{144}}$	RNA_{σ}		7	63	1778	420	1560	66.47	520	63.08	63.08	42.69	44.74	27.24	1.93	-
		$\text{RNA}_{\bar{r}_m}$		8	96	1794	220	1560	66.47	520	63.85	63.85	42.69	57.80	27.47	1.87	-
	$7D_{1\lambda}$	RNA_{σ}		8	92	1462	790	1710	73.80	570	71.75	55.61	51.43	40.13	23.94	1.06	-
		$\text{RNA}_{\bar{r}_m}$		9	59	2014	980	1710	72.22	570	72.46	72.46	51.43	51.43	25.93	1.25	-
	$\mathbf{G}_{h_{30}}$	Aqua $_{h_{30}}$	RNA_{σ}		7	59	2030	340	1800	74.00	600	74.17	63.33	35.14	6.52	0.95	0
			$\text{RNA}_{\bar{r}_m}$		8	59	2112	280	1800	80.11	600	78.83	78.83	41.13	7.26	1.06	0
		$1D_{(6/10)\lambda}$	RNA_{σ}		7	48	2077	900	1770	57.01	590	52.71	40.00	28.82	12.80	2.23	-
			$\text{RNA}_{\bar{r}_m}$		8	78	2093	750	1770	56.95	590	57.29	40.00	57.50	57.50	23.94	3.12
$7D_{1\lambda}$		RNA_{σ}		8	78	2388	1260	1980	79.19	660	79.85	63.03	68.80	68.80	39.38	0.80	-
		$\text{RNA}_{\bar{r}_m}$		9	79	2461	830	1980	73.08	660	75.91	63.03	85.16	85.16	37.53	1.32	0
G		- Grupos						TxT	-	Taxa de Acertos no Teste [%]							
P		- Problemas						TPV	-	Total de Padrões para Validação							
R		- Redes						TxV	-	Taxa de Acertos na Validação [%]							
NE		- Número de Entradas						TxSV	-	Taxa de Acertos Simultânea na Validação [%]							
NNO		- Número de Neurônios Ocultos						TTT	-	Tempo Total de Treinamento [s]							
NET		- Número de Épocas Total						TV	-	Tempo de Validação [s]							
NEV	- Número de Épocas para Validação						EQM	-	Erro Quadrático Médio								
TPT	- Total de Padrões para Teste						MPB	-	Média dos Pesos e Bias								

Para a contagem dos acertos, tanto para o desvio padrão como para a profundidade do pico, foi considerado que, se a estimativa feita pela rede estiver no intervalo definido por

$$\nu - 1 \leq \nu^* < \nu + 1, \quad (5.2)$$

em que ν é o valor exato da variável a ser estimada (σ ou τ_m), e ν^* é a estimativa feita pela rede, então o valor estimado pela rede está correto. A escolha do critério definido pela Eq. (5.2) foi baseado no fato de que os valores discretos, para a profundidade do pico, estão a uma distância de uma unidade, logo, não seria adequado adotar um intervalo para a contagem de acertos que aceitasse valores estimados que apresentem um erro maior que uma unidade. Uma interpretação semelhante pode ser feita para os valores do desvio padrão, já que a regra, para contagem de acertos, é a mesma. A taxa de acerto para cada rede é mostrada na coluna TxT (Taxa de Acertos no Teste). Essa coluna mostra a taxa de acerto média (aritmética simples) da rede sobre o conjunto de teste, *i.e.*, são somados os acertos obtidos pela rede, no processo de validação cruzada que apresentou a melhor generalização, para cada nível de ruído utilizado, sendo essa soma dividida pelo total de níveis de ruído. Essa contagem de acertos é independente para cada uma das redes, *i.e.*, não significa que se a RNA_σ produziu uma resposta correta o valor determinado pela rede RNA_{τ_m} também estará correto e vice-versa.

A coluna TPV (Total de Padrões para Validação) mostra a quantidade de padrões presentes no conjunto de validação, em acordo com a Tabela 5.9, e a coluna TxV (Taxa de Acertos na Validação) mostra as taxas de acerto de cada rede sobre esse conjunto corrompido com o ruído ε_3 , levando em conta a Eq. (5.2). Nesse caso, também são considerados os acertos independentes das redes.

A coluna TxSV (Taxa de Acertos Simultânea na Validação) mostra as percentagens de acertos que ocorreram simultaneamente, *i.e.*, para ser considerado um acerto, ambas as redes RNA_σ e RNA_{τ_m} devem produzir uma resposta correta seguindo a regra definida pela Eq. (5.2).

A coluna TTT (Tempo Total de Treinamento) traz o tempo total gasto pelo algoritmo de treinamento para a execução das 300 iterações definidas anteriormente, e a coluna TV (Tempo de Validação) mostra o tempo de treinamento necessário para se obter o Número de Épocas para a Validação (NEV).

A coluna EQM apresenta o erro quadrático médio, definido como

$$EQM = \frac{1}{N_r} \frac{1}{N_P} \sum_{p=1}^{N_P} |d_p - y_p|^2,$$

em que N_r é o número total de níveis de ruído, N_P é número total de padrões em cada nível de ruído, d_p é a saída desejada e y_p é a saída calculada pela rede na apresentação do p -ésimo padrão. Visto que o intervalo para contagem de acertos considera uma tolerância de uma unidade para mais e para menos, tem-se um total de duas unidades como intervalo total. Em razão disso, as redes apresentam uma taxa de acerto de 100% (em média), pois todos os problemas têm um EQM menor que dois (em média), exceto os problemas $P_{2h_{144}}$ (rede RNA_{τ_m}) e $P_{2h_{30}}$.

Por fim, a última coluna da Tabela 5.10 mostra se os valores iniciais para os pesos e níveis de *bias* estão distribuído em torno da média zero. Para os casos nos quais a média não é zero, esse valores iniciais estão distribuídos aleatoriamente no intervalo $(0, 1)$. Para o caso de média zero, os pesos e *bias* estão distribuídos aleatoriamente no intervalo $(-1, 0) \cup (0, 1)$.

5.3 Discussão dos Resultados

Os comentários que serão feitos a seguir se referem aos resultados apresentados na Tabela 5.10, e são produto do treinamento, teste e validação das redes utilizadas.

Ao comparar os valores NET com os valores NEV, é possível observar a eficiência do método de treinamento proposto, pelo fato de que o valor NEV (que representa o ponto ótimo do treinamento) é, em geral, menor que a metade do número total de épocas utilizadas no treinamento das rede RNA_{σ} . Já para as redes RNA_{τ_m} , esse valor tende a ser menor ainda. A redução do número de épocas necessárias para se chegar à melhor generalização possível, para as redes RNA_{τ_m} , é justificado pela adição da estimativa feita pela rede RNA_{σ} às entradas da rede RNA_{τ_m} . Como foi comentado, esse valor proporciona informações adicionais para a rede, facilitando o processo de aprendizagem.

As taxas de acertos sobre o conjunto de teste (coluna TxT), no processo de validação cruzada, indicam que as características do problema $P_{1h_{144}}$ são mais adequadas, pois, foi nesse problema, que as redes apresentaram o melhor desempenho quando comparadas aos resultados obtidos para os demais problemas do grupo $\mathbf{G}_{h_{144}}$. Já para os problemas do grupo $\mathbf{G}_{h_{30}}$, as taxas de acerto estão próximas para todos os problemas, exceto para o problema $P_{2h_{30}}$, no qual a taxa de acerto não chegou aos

60%. Esse mesmo problema ocorreu no problema $P_{2h_{144}}$.

Essa taxa de acerto reduzida para os problemas P_{2h_*} está associada a dois fatores. O primeiro está relacionado ao fato de que a radiação emergente medida em $\theta = 60^\circ$ tem ordem de grandeza maior que a medida em $\theta = 0^\circ$. O segundo (que é consequência do primeiro) está relacionado com a sensibilidade do problema ao ruído. Sendo assim, mesmo que o ruído adicionado seja proporcional à ordem de grandeza da radiação emergente, o ruído tende a ser mais significativo para a radiação emergente medida em $\theta = 60^\circ$.

O melhor desempenho das redes, para os problemas do grupo $\mathbf{G}_{h_{30}}$, está relacionado com a concentração de clorofila, *i.e.*, as redes tendem a ter um desempenho melhor quando a concentração de clorofila é pequena. Isso porque a radiação emergente tem ordem de grandeza menor quando comparada à radiação emergente obtida a partir dos problemas do grupo $\mathbf{G}_{h_{144}}$. Note também que o desempenho das redes se manteve “constante” sobre os conjuntos de validação, visto que as taxas de acerto se mantiveram muito próximas às do processo de validação, mesmo sendo utilizado um nível de ruído e padrões que não estiveram presentes durante o treinamento. Diante disso, é possível concluir que o processo de treinamento adotado é robusto, apesar de as redes do grupo $\mathbf{G}_{h_{144}}$ terem apresentado resultados um pouco inferiores aos das redes do grupo $\mathbf{G}_{h_{30}}$.

As taxas de acertos mostradas nas colunas TxT e TxV não garantem que as duas estimativas estejam dentro do intervalo de tolerância definido pela Eq. (5.2), visto que as taxas apresentadas foram calculadas a partir das estimativas independentes de cada rede. A garantia de que ambas as redes produzam uma resposta dentro do intervalo de tolerância é mostrada na coluna TxSV. Essas taxas de acerto foram obtidas considerando que ambas as redes devem produzir uma resposta dentro do intervalo de contagem de acertos, *i.e.*, para um mesmo padrão de entrada, tanto a RNA_σ quanto a RNA_{τ_m} devem produzir uma resposta dentro do intervalo de tolerância definido pela Eq. (5.2). Como é de se esperar, essas taxas de acerto são inferiores às taxas de acertos independentes, no entanto, novamente, os acertos sobre os problemas do grupo $\mathbf{G}_{h_{30}}$ são superiores aos do grupo $\mathbf{G}_{h_{144}}$. Os problemas P_{2h_*} continuaram tendo as piores taxas de acerto, e os problemas P_{1h_*} continuam com as melhores.

Com relação aos tempos computacionais TTT, é possível observar que estes foram reduzidos, sendo que apenas três (128.30, 68.80 e 85.15) passaram de um minuto. O tempo total de treinamento inclui o tempo gasto para o processo de validação

cruzada. Já os tempos apresentados na coluna TV indicam o tempo total necessário para se atingir o NEV. Nesse caso, apenas três (37.19, 39.38 e 37.53) ultrapassaram a marca dos 30 segundos. Esses tempos também incluem o tempo de treinamento necessário para se chegar ao número de épocas apresentadas na coluna NEV. Os erros quadráticos médios confirmam o que já foi comentado sobre os problemas P_{2h_*} , visto que esses problemas apresentaram o maior erro. Os erros menos expressivos estão relacionados aos problemas P_{1h_*} .

As próximas discussões são acerca os resultados obtidos pelas redes sobre os conjuntos de validação. Em virtude da grande quantidade de problemas considerados, a apresentação de todos os resultados gráficos fica inviabilizada. Diante disso, foram considerados apenas alguns casos que são representativos do conjunto todo.

Observação: Em todas as figuras, as linhas pretas representam os valores exatos, as linhas vermelhas representam os valores estimados pelas redes, e as linhas tracejadas, quando aparecerem, representam os intervalos de tolerâncias.

As Figuras 5.3 – 5.6 mostram as estimativas feitas pelas redes para o desvio padrão e para o pico da concentração de clorofila, para os problemas dos grupos $G_{h_{144}}$. As Figuras 5.7 – 5.10 mostram essas mesmas estimativas para os problemas pertencentes ao grupo $G_{h_{30}}$. Nessas figuras, são mostrados apenas aqueles valores que foram selecionados para o conjunto de validação sem considerar as réplicas. Juntamente com essas figuras, são apresentados os erros cometidos em cada estimativa.

Avaliando o erro das estimativas que estão fora do intervalo de tolerância, é possível perceber que o erro cometido é relativamente pequeno, sendo o maior valor apresentado na Figura 5.10(b), que traz as estimativas para a profundidade do pico de concentração de clorofila. O erro mostrado nessa figura chega na ordem de aproximadamente seis metros. Os maiores erros na estimativa do desvio padrão estão em torno de quatro metros e são mostrados nas Figuras 5.3(a) e 5.5(a) para os problemas $Aqua_{h_{144}}$ e $P_{2h_{144}}$, respectivamente. Note também que uma boa estimativa para o valor do desvio padrão não implica uma boa estimativa para a profundidade do pico da concentração de clorofila.

As Figuras 5.11 – 5.14 e as Figuras 5.15 – 5.18 mostram os erros cometidos nas estimativas do desvio padrão e nas estimativas da profundidade do pico da concentração, para todos os padrões dos conjuntos de validação, para os grupos $G_{h_{144}}$ e $G_{h_{30}}$, respectivamente. As figuras à esquerda mostram os erros associados aos problemas do grupo $G_{h_{144}}$, e as figuras à direita mostram os erros associados aos problemas do

grupo $G_{h_{30}}$. Essas figuras possibilitam fazer uma avaliação geral do desempenho das redes sobre o conjunto de validação para cada problema. Note que, similarmente aos casos anteriores, as respostas fora do intervalo de tolerância apresentam um erro “controlado”. Isso significa que, apesar da sensibilidade à presença de ruído, as redes apresentam a capacidade de produzir respostas “estáveis”.

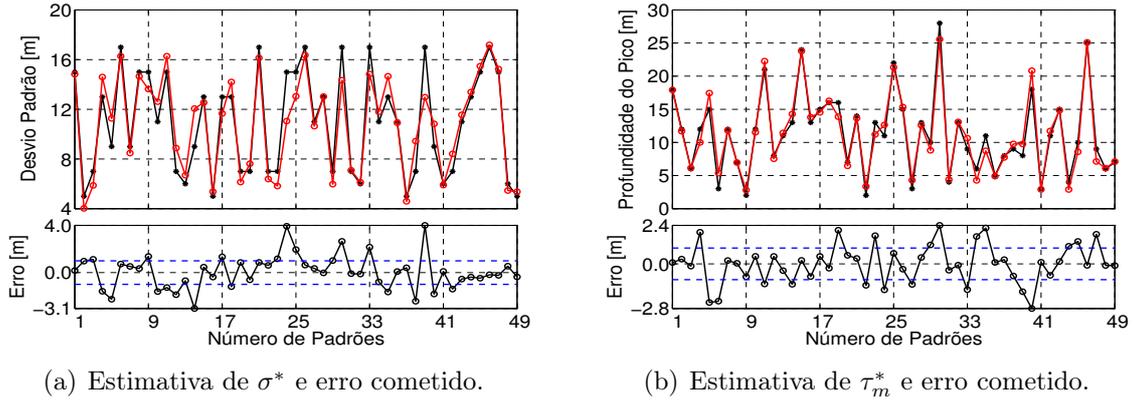


Figura 5.3 - Estimativa do desvio padrão e da profundidade do pico, para o problema $Aqua_{h_{144}}$, e os respectivos erros dessas estimativas.

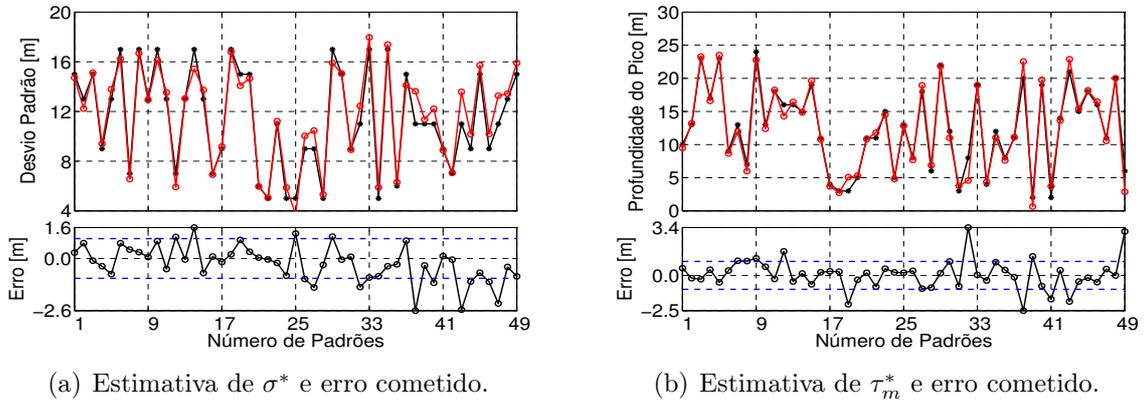
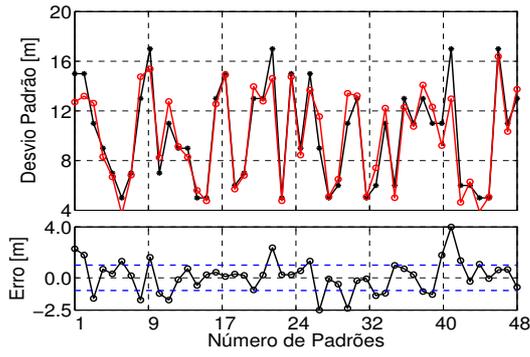
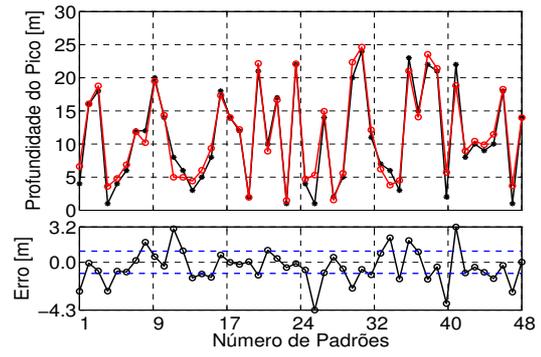


Figura 5.4 - Estimativa do desvio padrão e da profundidade do pico, para o problema $P_{1h_{144}}$, e os respectivos erros dessas estimativas.

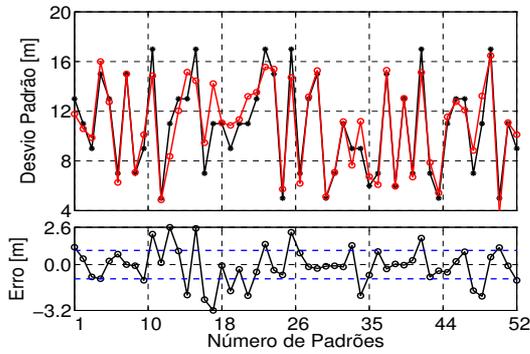


(a) Estimativa de σ^* e erro cometido.

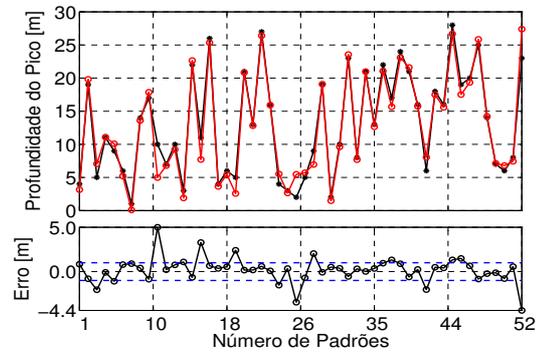


(b) Estimativa de τ_m^* e erro cometido.

Figura 5.5 - Estimativa do desvio padrão e da profundidade do pico, para o problema $P_{2h_{144}}$, e os respectivos erros dessas estimativas.

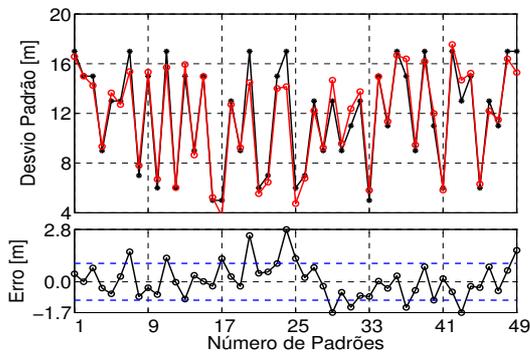


(a) Estimativa de σ^* e erro cometido.

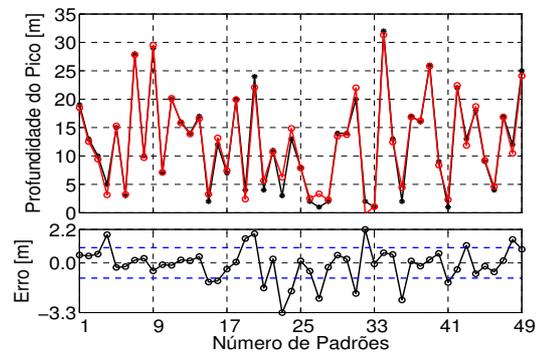


(b) Estimativa de τ_m^* e erro cometido.

Figura 5.6 - Estimativa do desvio padrão e da profundidade do pico, para o problema $P_{7Dh_{144}}$, e os respectivos erros dessas estimativas.

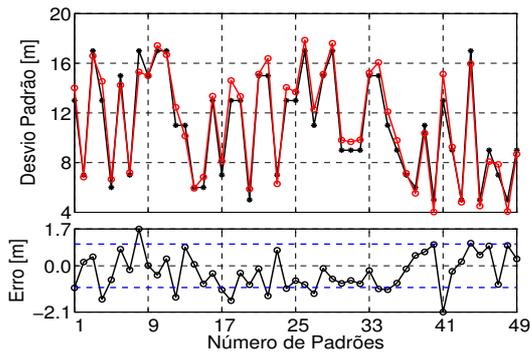


(a) Estimativa de σ^* e erro cometido.

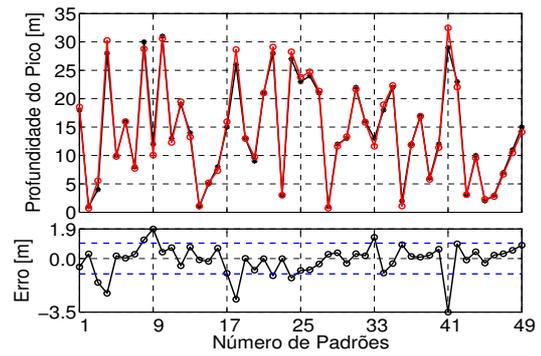


(b) Estimativa de τ_m^* e erro cometido.

Figura 5.7 - Estimativa do desvio padrão e da profundidade do pico, para o problema $Aqua_{h_{30}}$, e os respectivos erros dessas estimativas.

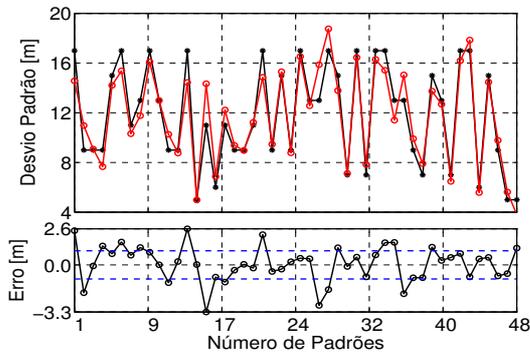


(a) Estimativa de σ^* e erro cometido.

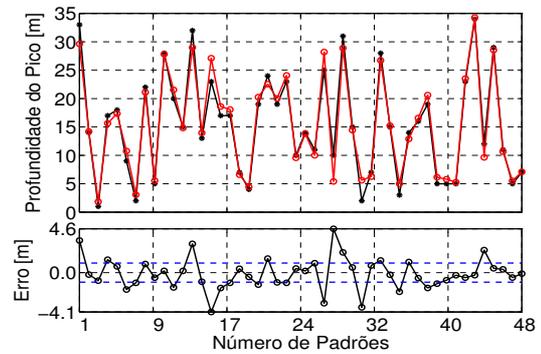


(b) Estimativa de τ_m^* e erro cometido.

Figura 5.8 - Estimativa do desvio padrão e da profundidade do pico, para o problema $P_{1h_{30}}$, e os respectivos erros dessas estimativas.

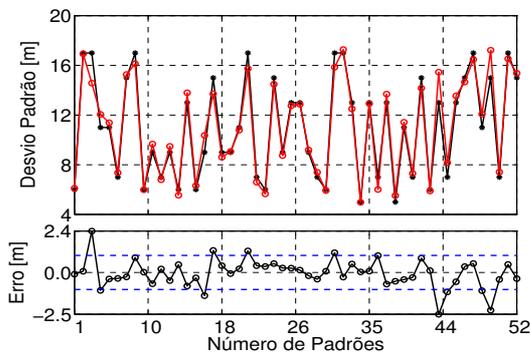


(a) Estimativa de σ^* e erro cometido.

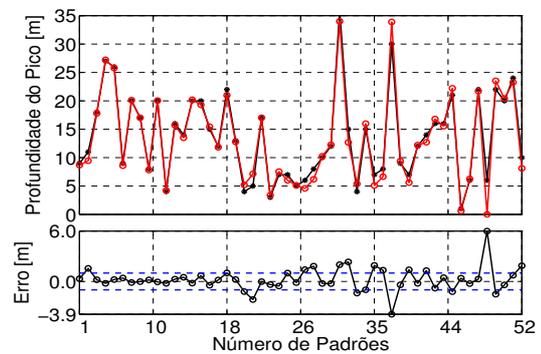


(b) Estimativa de τ_m^* e erro cometido.

Figura 5.9 - Estimativa do desvio padrão e da profundidade do pico, para o problema $P_{2h_{30}}$, e os respectivos erros dessas estimativas.

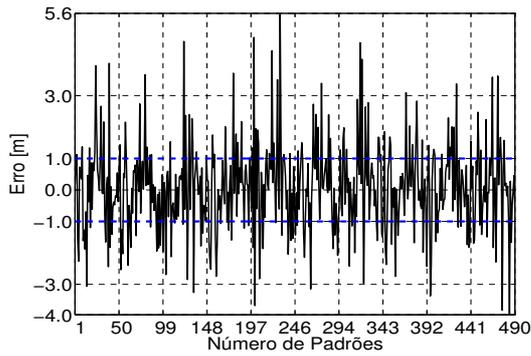


(a) Estimativa de σ^* e erro cometido.

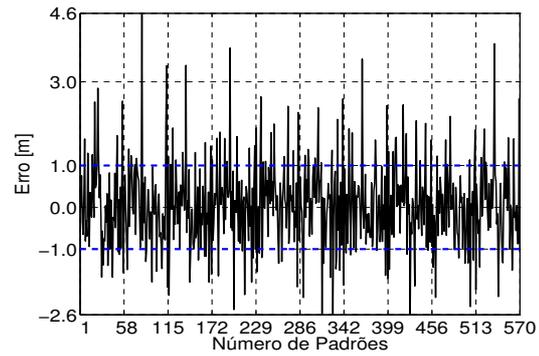


(b) Estimativa de τ_m^* e erro cometido.

Figura 5.10 - Estimativa do desvio padrão e da profundidade do pico, para o problema $P_{7Dh_{30}}$, e os respectivos erros dessas estimativas.

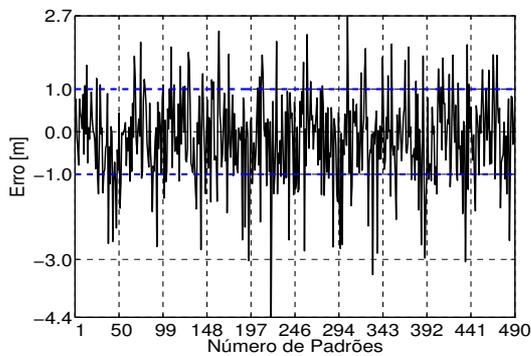


(a) Erro cometido na estimativa de σ^* para o problema $\text{Aqua}_{h_{144}}$.

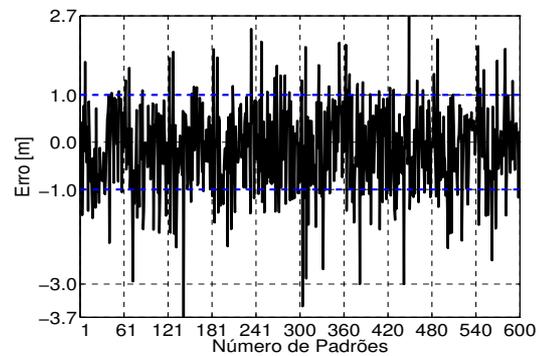


(b) Erro cometido na estimativa de σ^* para o problema $\text{Aqua}_{h_{30}}$.

Figura 5.11 - Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas Aqua_{h_*} .

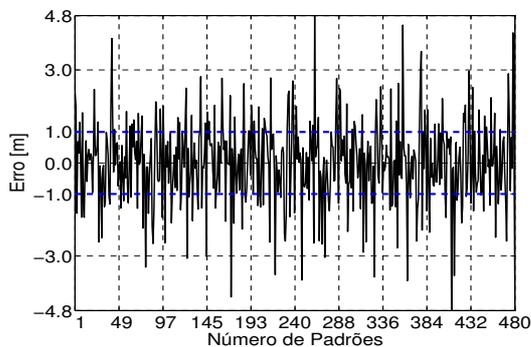


(a) Erro cometido na estimativa de σ^* para o problema $\text{P}_{1h_{144}}$.

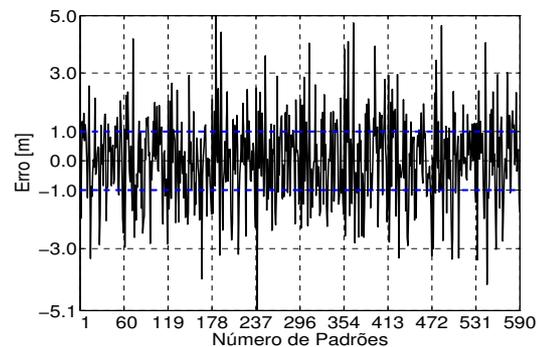


(b) Erro cometido na estimativa de σ^* para o problema $\text{P}_{1h_{30}}$.

Figura 5.12 - Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas P_{1h_*} .

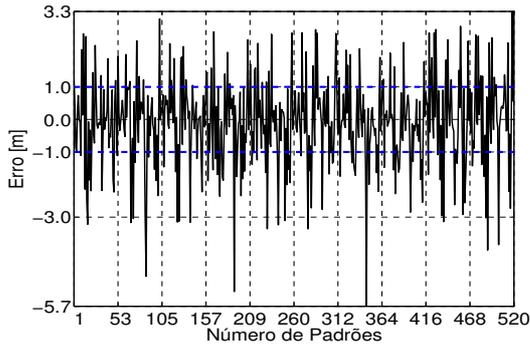


(a) Erro cometido na estimativa de σ^* para o problema $\text{P}_{2h_{144}}$.

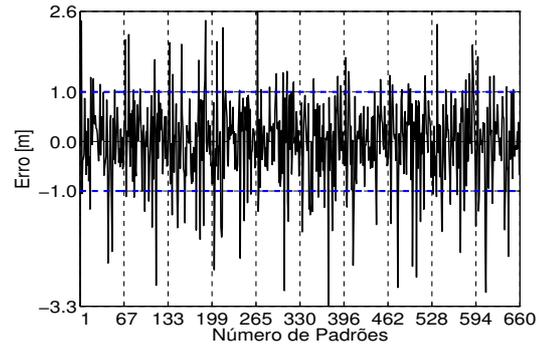


(b) Erro cometido na estimativa de σ^* para o problema $\text{P}_{2h_{30}}$.

Figura 5.13 - Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas P_{2h_*} .

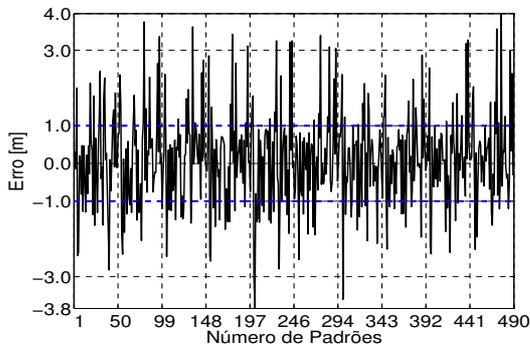


(a) Erro cometido na estimativa de σ^* para o problema $P_{7Dh_{144}}$.

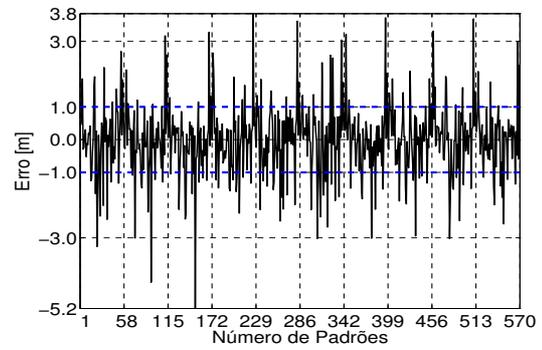


(b) Erro cometido na estimativa de σ^* para o problema $P_{7Dh_{30}}$.

Figura 5.14 - Erros cometidos nas estimativas dos valores de desvio padrão, sobre o conjunto de validação, para os problemas P_{7Dh_*} .

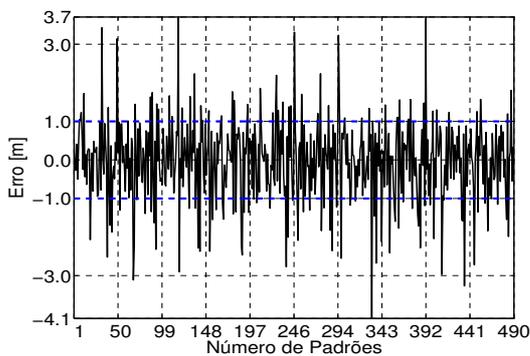


(a) Erro cometido na estimativa de τ_m^* para o problema $Aqua_{h_{144}}$.

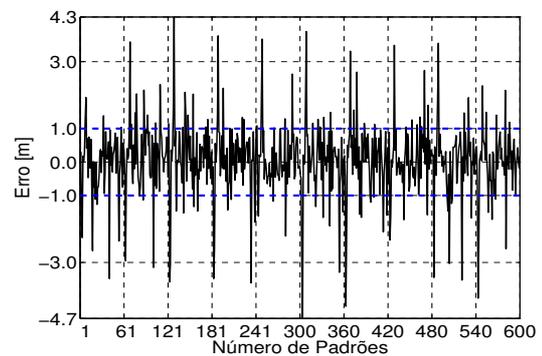


(b) Erro cometido na estimativa de τ_m^* para o problema $Aqua_{h_{30}}$.

Figura 5.15 - Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas $Aqua_{h_*}$.

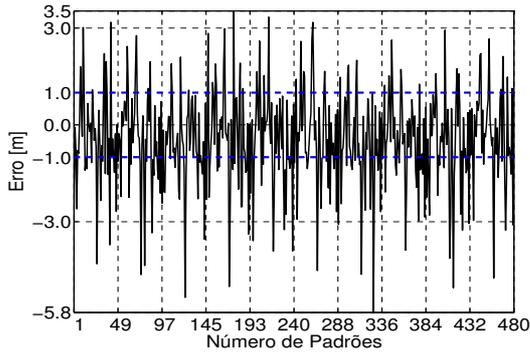


(a) Erro cometido na estimativa de τ_m^* para o problema $P_{1h_{144}}$.

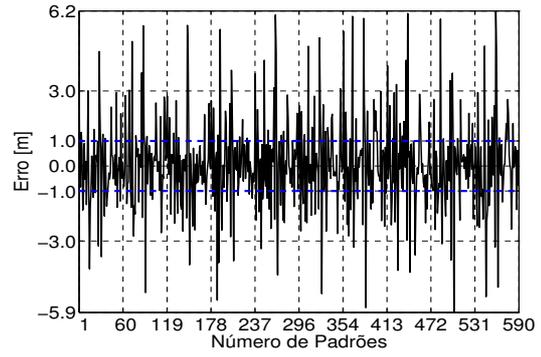


(b) Erro cometido na estimativa de τ_m^* para o problema $P_{1h_{30}}$.

Figura 5.16 - Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas P_{1h_*} .

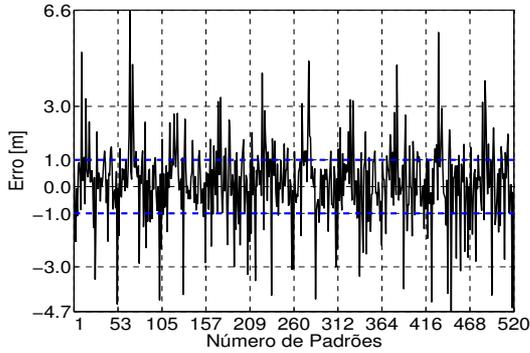


(a) Erro cometido na estimativa de τ_m^* para o problema $P_{2h_{144}}$.

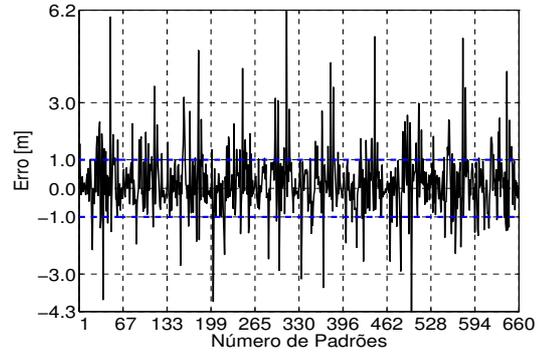


(b) Erro cometido na estimativa de τ_m^* para o problema $P_{2h_{30}}$.

Figura 5.17 - Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas P_{2h_*} .



(a) Erro cometido na estimativa de τ_m^* para o problema $P_{7Dh_{144}}$.



(b) Erro cometido na estimativa de τ_m^* para o problema $P_{7Dh_{30}}$.

Figura 5.18 - Erros cometidos nas estimativas dos valores da profundidade do pico da concentração de clorofila, sobre o conjunto de validação, para os problemas P_{7Dh_*} .

Para as estimativas que foram contadas como acertos, segundo a Eq. (5.2), são apresentadas duas situações que levam em conta os acertos simultâneos das redes sobre o conjunto de validação completo (com réplicas), *i.e.*, ambas as redes devem produzir uma estimativa dentro do intervalo de tolerância. No primeiro caso, são buscadas aquelas estimativas de σ^* que apresentam a menor distância (erro) absoluta(o) em relação aos valores exatos de σ (veja Tabela 5.11), e, posteriormente (veja Tabela 5.12), as estimativas de τ_m^* que também apresentam a menor distância (erro) absoluta(o) em relação aos valores exatos de τ_m . O segundo caso considera as piores estimativas de σ^* e de τ_m^* , no entanto que estejam no conjunto de acertos simultâneos. Nesse caso, são buscados aqueles valores estimados que apresentam a

maior distância (erro) absoluta(o) em relação aos valores exatos. Veja Tabelas 5.13 e 5.14, respectivamente.

Nas tabelas seguintes, os erros cometidos nas estimativas dos valores do desvio padrão e nos valores da profundidade do pico da concentração de clorofila são definidos, respectivamente, por

$$e_\sigma = \sigma - \sigma^*, \quad (5.3)$$

e

$$e_\tau = \tau_m - \tau_m^*. \quad (5.4)$$

Tabela 5.11 - Melhores estimativas para os valores do desvio padrão e os respectivos valores da profundidade do pico da concentração de clorofila para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

Grupo	Classe	Problema	σ [m]	σ^* [m]	τ_m [m]	τ_m^* [m]	e_τ [m]	Fig.
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	Aqua _{h_{144}}	7.00	7.00	10.00	9.46	0.54	5.19(a)
		P _{1h_{144}}	11.00	11.00	15.00	14.50	0.50	5.20(a)
		P _{2h_{144}}	9.00	8.99	6.00	5.15	0.85	5.21(a)
	7D λ	P _{7Dh_{144}}	15.00	15.00	1.00	0.10	0.90	5.22(a)
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	Aqua _{h_{30}}	15.00	14.99	13.00	12.54	0.46	5.19(c)
		P _{1h_{30}}	9.00	9.00	14.00	13.61	0.39	5.20(c)
		P _{2h_{30}}	13.00	12.99	28.00	27.81	0.19	5.21(c)
	7D λ	P _{7Dh_{30}}	11.00	11.00	4.00	4.26	-0.26	5.22(c)

Tabela 5.12 - Melhores estimativas para os valores da profundidade do pico da concentração de clorofila e os respectivos valores do desvio padrão para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

Grupo	Classe	Problema	σ [m]	σ^* [m]	e_σ [m]	τ_m [m]	τ_m^* [m]	Fig.
$\mathbf{G}_{h_{144}}$	1D _{(6/10)λ}	Aqua _{h_{144}}	6.00	6.66	-0.66	7.00	7.00	5.19(b)
		P _{1h_{144}}	15.00	14.24	0.76	6.00	6.00	5.20(b)
		P _{2h_{144}}	15.00	14.90	0.10	14.00	14.00	5.21(b)
	7D λ	P _{7Dh_{144}}	13.00	13.54	-0.54	27.00	26.99	5.22(b)
$\mathbf{G}_{h_{30}}$	1D _{(6/10)λ}	Aqua _{h_{30}}	6.00	6.05	-0.05	8.00	7.99	5.19(d)
		P _{1h_{30}}	11.00	10.25	0.75	6.00	6.00	5.20(d)
		P _{2h_{30}}	9.00	8.82	0.18	15.00	14.99	5.21(d)
	7D λ	P _{7Dh_{30}}	7.00	7.18	-0.18	11.00	11.00	5.22(d)

Como comentado anteriormente, uma ótima estimativa para o valor do desvio padrão não influencia em uma ótima estimativa para a profundidade do pico da concentração, como pode ser visto no problema P_{7D h_{144}} da Tabela 5.11. Nesse problema, o erro na estimativa do desvio padrão foi nulo, enquanto a estimativa da profundi-

dade do pico apresentou um erro significativo. Uma situação semelhante ocorreu para o problema $P_{2h_{144}}$, dessa mesma tabela.

Tabela 5.13 - Piores estimativas para os valores do desvio padrão e os respectivos valores da profundidade do pico da concentração de clorofila para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

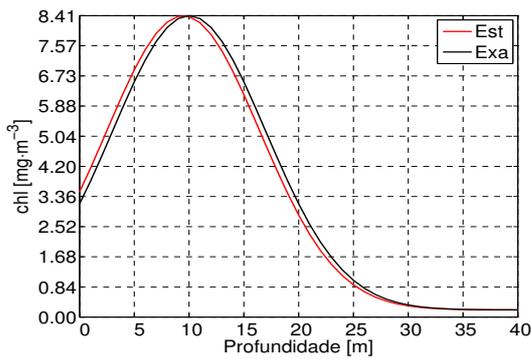
Grupo	Classe	Problema	σ [m]	σ^* [m]	τ_m [m]	τ_m^* [m]	e_τ [m]	Fig.
$\mathbf{G}_{h_{144}}$	1D _(6/10) λ	Aqua _{h_{144}}	13.00	12.01	11.00	10.59	0.41	5.23(a)
		P _{1h_{144}}	13.00	13.99	20.00	20.55	-0.55	5.24(a)
		P _{2h_{144}}	5.00	4.01	9.00	9.89	-0.89	5.25(a)
	7D λ	P _{7Dh_{144}}	15.00	15.99	11.00	11.08	-0.08	5.26(a)
$\mathbf{G}_{h_{30}}$	1D _(6/10) λ	Aqua _{h_{30}}	11.00	11.99	9.00	8.37	0.63	5.23(c)
		P _{1h_{30}}	6.00	6.99	6.00	6.13	-0.13	5.24(c)
		P _{2h_{30}}	11.00	10.02	4.00	4.40	-0.40	5.25(c)
	7D λ	P _{7Dh_{30}}	15.00	14.01	11.00	11.29	-0.29	5.26(c)

Tabela 5.14 - Piores estimativas para os valores da profundidade do pico da concentração de clorofila e os respectivos valores do desvio padrão para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

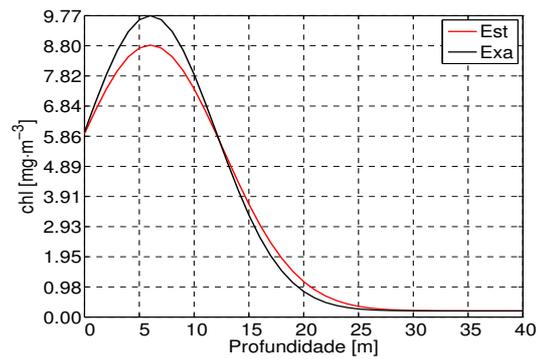
Grupo	Classe	Problema	σ [m]	σ^* [m]	e_σ [m]	τ_m [m]	τ_m^* [m]	Fig.
$\mathbf{G}_{h_{144}}$	1D _(6/10) λ	Aqua _{h_{144}}	11.00	10.04	0.96	5.00	5.98	5.23(b)
		P _{1h_{144}}	15.00	10.05	4.95	12.00	11.01	5.24(b)
		P _{2h_{144}}	5.00	4.77	0.23	11.00	11.98	5.25(b)
	7D λ	P _{7Dh_{144}}	7.00	6.75	0.25	5.00	4.01	5.26(b)
$\mathbf{G}_{h_{30}}$	1D _(6/10) λ	Aqua _{h_{30}}	9.00	8.21	0.79	17.00	16.03	5.23(d)
		P _{1h_{30}}	15.00	14.64	0.36	5.00	5.99	5.24(d)
		P _{2h_{30}}	9.00	8.78	0.22	1.00	1.99	5.25(d)
	7D λ	P _{7Dh_{30}}	11.00	11.11	-0.11	10.00	9.01	5.26(d)

Por outro lado, uma estimativa não tão boa para o desvio padrão pode implicar uma ótima estimativa para a profundidade do pico da concentração de clorofila, como pode ser vista pelos problemas $P_{1h_{144}}$ e $P_{1h_{30}}$ apresentados na Tabela 5.12. Note que nesses problemas, o erro na estimativa da profundidade do pico é praticamente nulo, enquanto o erro na estimativa do desvio padrão é significativo.

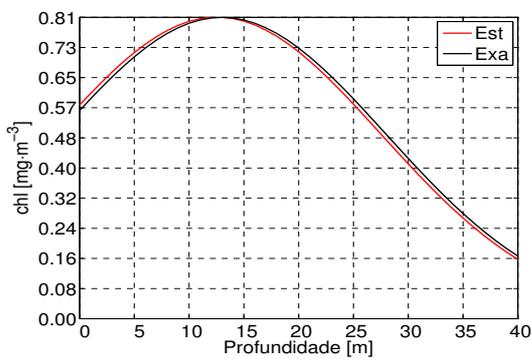
As Figuras 5.19 – 5.22 mostram os perfis recuperados a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para os problemas Aqua _{h_*} , P_{1 h_*} , P_{2 h_*} e P_{7D h_*} , respectivamente, dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$. Já as Figuras 5.23 – 5.26 mostram os perfis recuperados a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para os problemas Aqua _{h_*} , P_{1 h_*} , P_{2 h_*} e P_{7D h_*} , respectivamente, dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.



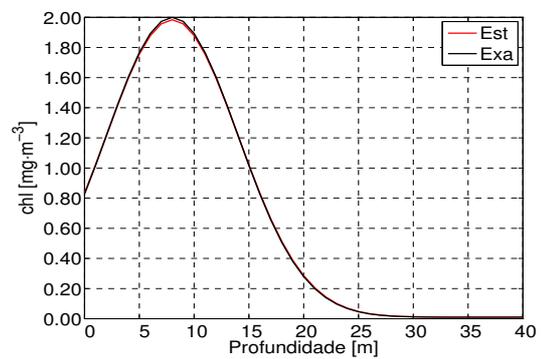
(a) Perfil obtido a partir da melhor estimativa de σ^* para o problema $\text{Aqua}_{h_{144}}$.



(b) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $\text{Aqua}_{h_{144}}$.

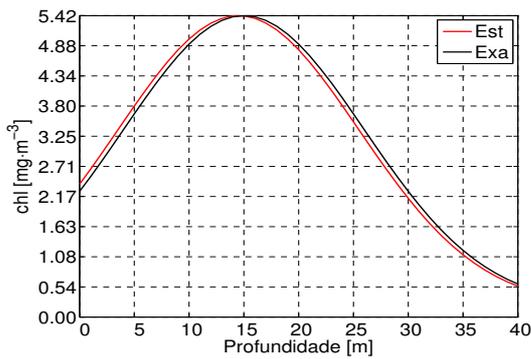


(c) Perfil obtido a partir da melhor estimativa de σ^* para o problema $\text{Aqua}_{h_{30}}$.

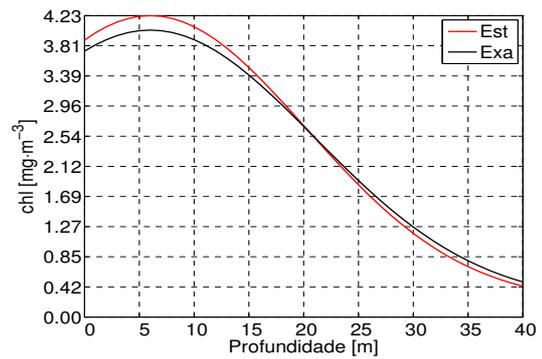


(d) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $\text{Aqua}_{h_{30}}$.

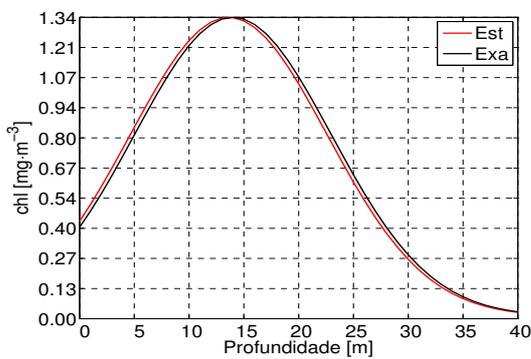
Figura 5.19 - Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema Aqua_{h_*} .



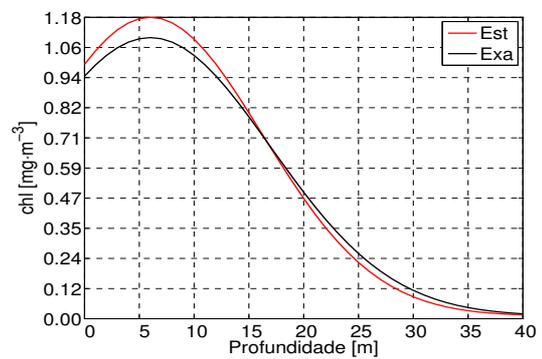
(a) Perfil obtido a partir da melhor estimativa de σ^* para o problema $P_{1h_{144}}$.



(b) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $P_{1h_{144}}$.

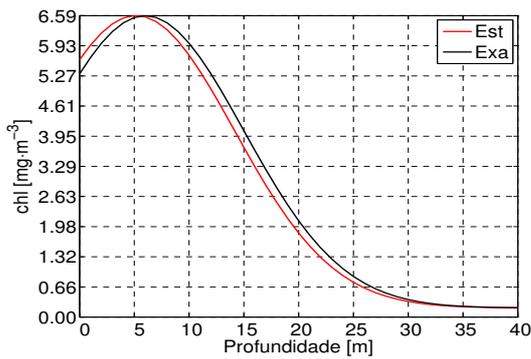


(c) Perfil obtido a partir da melhor estimativa de σ^* para o problema $P_{1h_{30}}$.

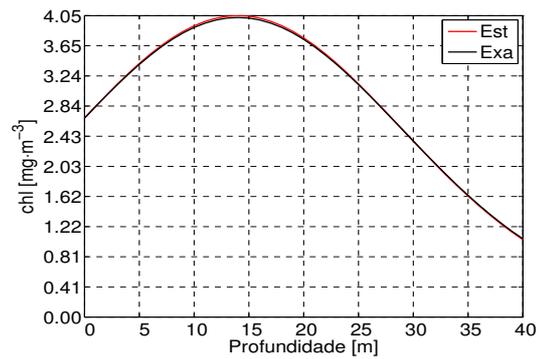


(d) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $P_{1h_{30}}$.

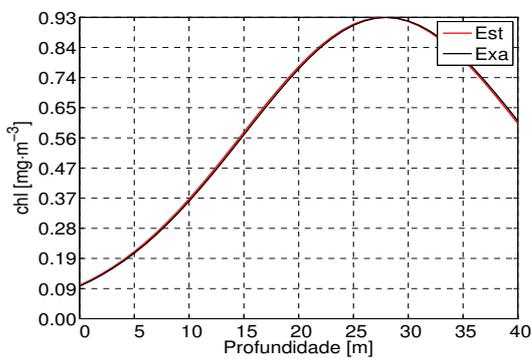
Figura 5.20 - Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P_{1h_*} .



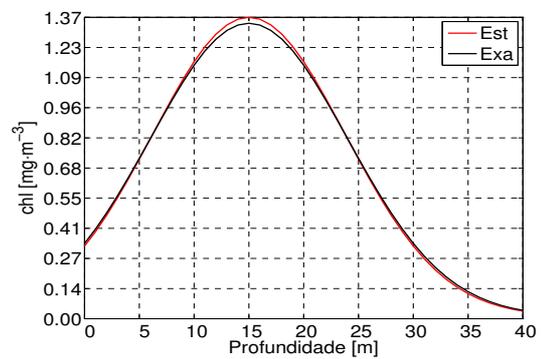
(a) Perfil obtido a partir da melhor estimativa de σ^* para o problema $P_{2h_{144}}$.



(b) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $P_{2h_{144}}$.

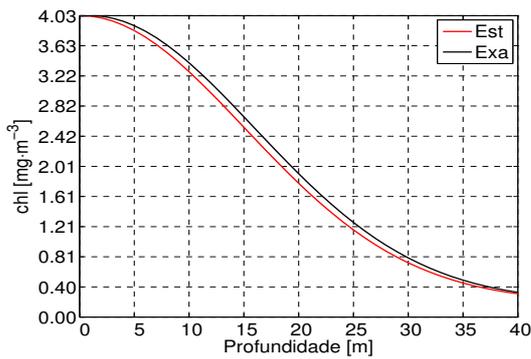


(c) Perfil obtido a partir da melhor estimativa de σ^* para o problema $P_{2h_{30}}$.

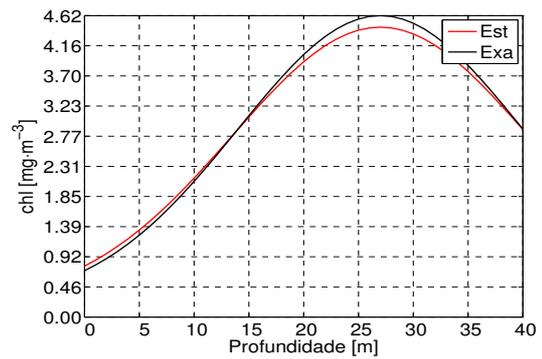


(d) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $P_{2h_{30}}$.

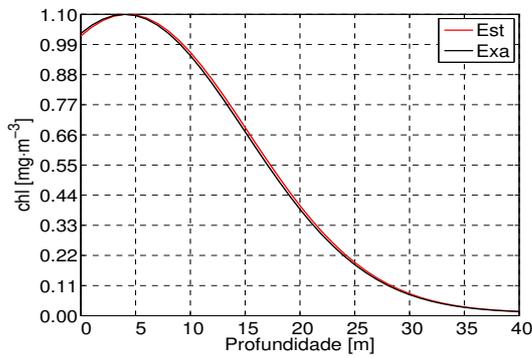
Figura 5.21 - Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P_{2h_*} .



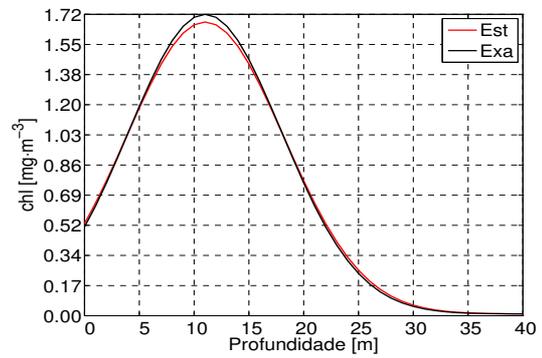
(a) Perfil obtido a partir da melhor estimativa de σ^* para o problema $P_{7Dh_{144}}$.



(b) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $P_{7Dh_{144}}$.

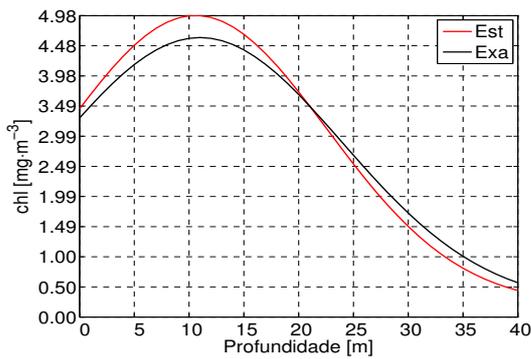


(c) Perfil obtido a partir da melhor estimativa de σ^* para o problema $P_{7Dh_{30}}$.

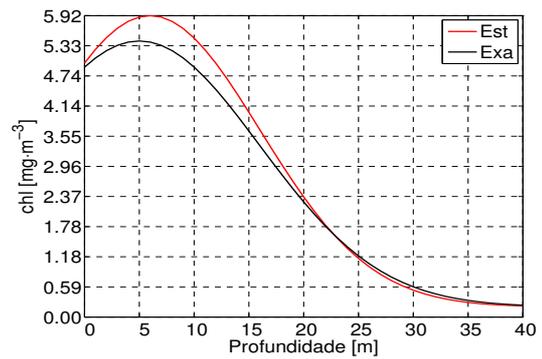


(d) Perfil obtido a partir da melhor estimativa de τ_m^* para o problema $P_{7Dh_{30}}$.

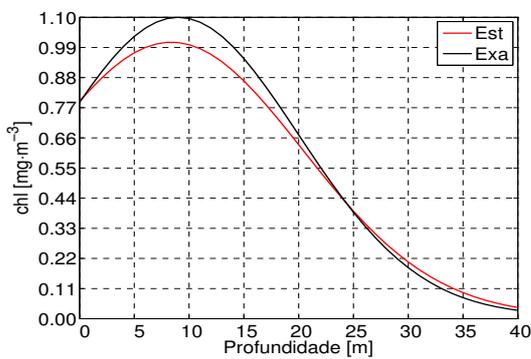
Figura 5.22 - Perfis obtidos a partir das melhores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P_{7Dh_*} .



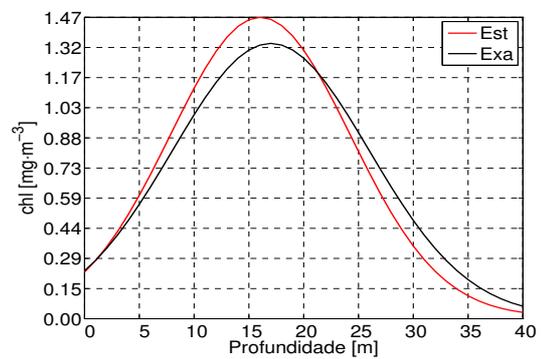
(a) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $\text{Aqua}_{h_{144}}$.



(b) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $\text{Aqua}_{h_{144}}$.

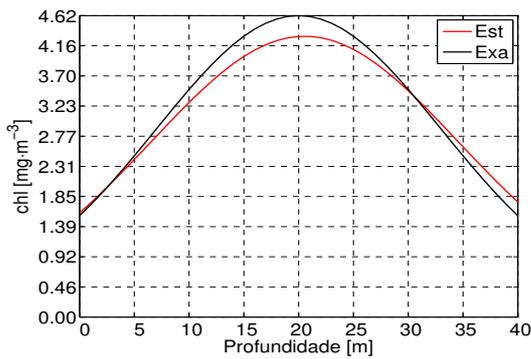


(c) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $\text{Aqua}_{h_{30}}$.

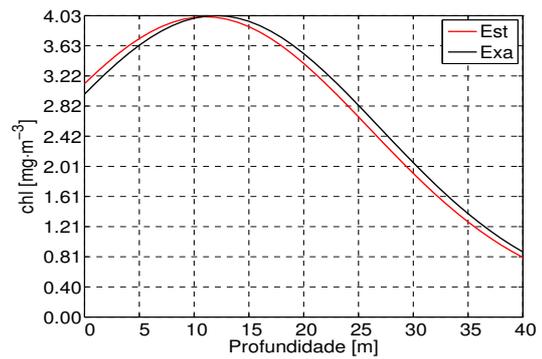


(d) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $\text{Aqua}_{h_{30}}$.

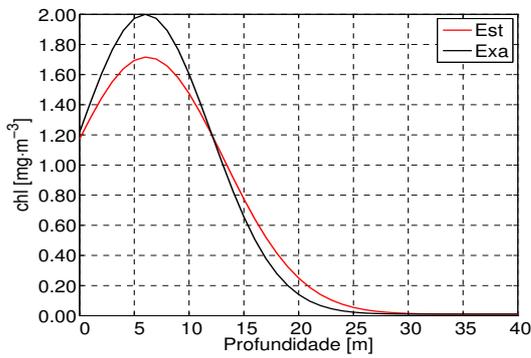
Figura 5.23 - Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema Aqua_{h_*} .



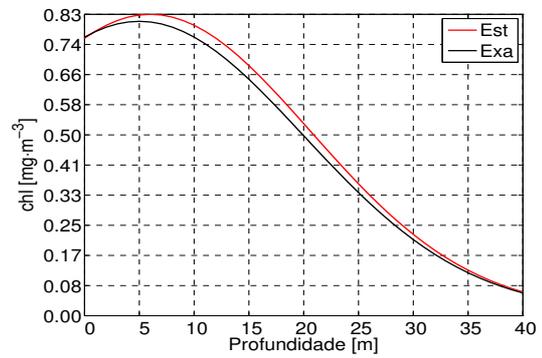
(a) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $P_{1h_{144}}$.



(b) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $P_{1h_{144}}$.

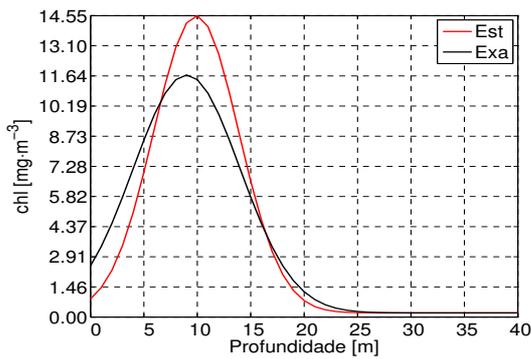


(c) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $P_{1h_{30}}$.

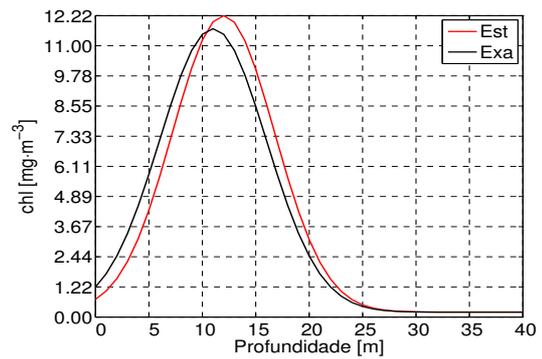


(d) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $P_{1h_{30}}$.

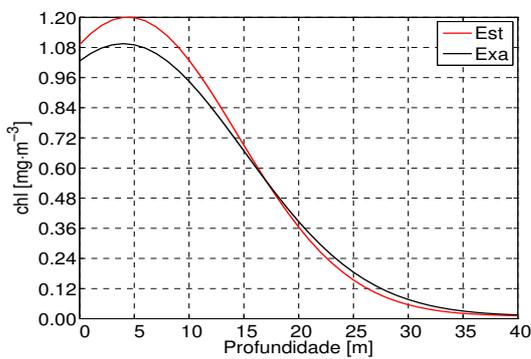
Figura 5.24 - Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P_{1h_*} .



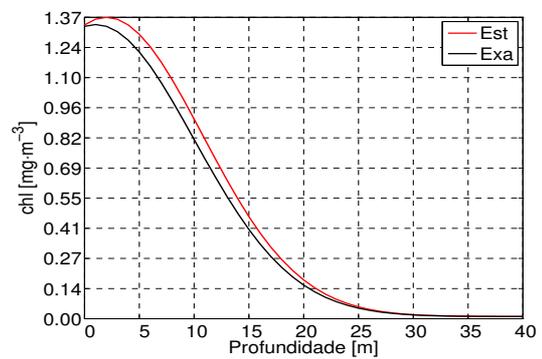
(a) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $P_{2h_{144}}$.



(b) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $P_{2h_{144}}$.

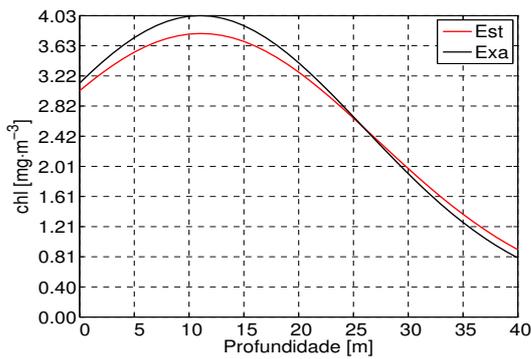


(c) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $P_{2h_{30}}$.

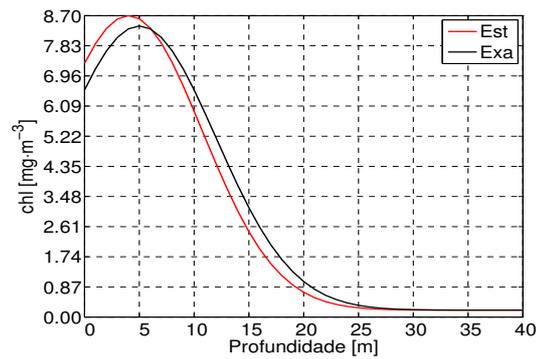


(d) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $P_{2h_{30}}$.

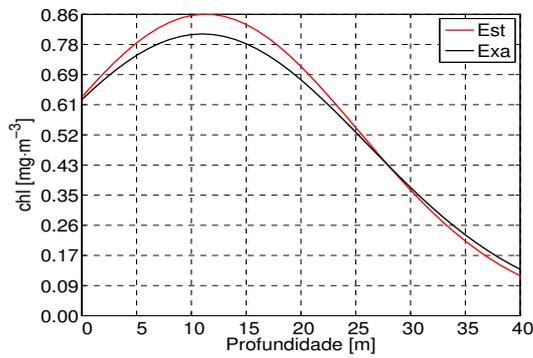
Figura 5.25 - Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problema P_{2h_*} .



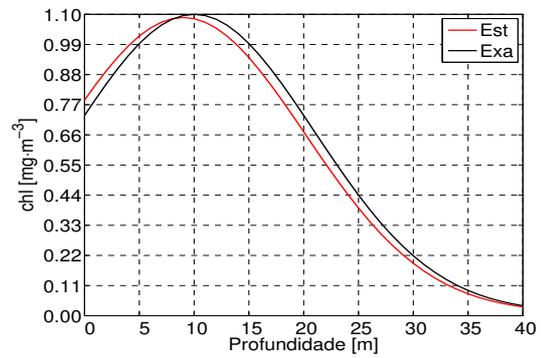
(a) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $P_{7Dh_{144}}$.



(b) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $P_{7Dh_{144}}$.



(c) Perfil obtido a partir da “pior” estimativa de σ^* para o problema $P_{7Dh_{30}}$.



(d) Perfil obtido a partir da “pior” estimativa de τ_m^* para o problema $P_{7Dh_{30}}$.

Figura 5.26 - Perfis obtidos a partir das piores estimativas do desvio padrão e da profundidade do pico da concentração de clorofila para o problemas P_{7Dh_*} .

O último caso considerado é semelhante ao segundo caso anterior. No entanto, a busca dos piores valores é feita sobre todas as respostas obtidas pelas redes, sobre o conjunto de validação, independente se essa resposta está ou não no intervalo de tolerância para a contagem de acertos. Esses resultados são mostrados nas Tabelas 5.15 e 5.16, e os gráficos são mostrados nas Figuras 5.27 – 5.30.

Tabela 5.15 - Piores estimativas gerais, sobre o conjunto de validação, para os valores do desvio padrão e os respectivos valores da profundidade do pico da concentração de clorofila, para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

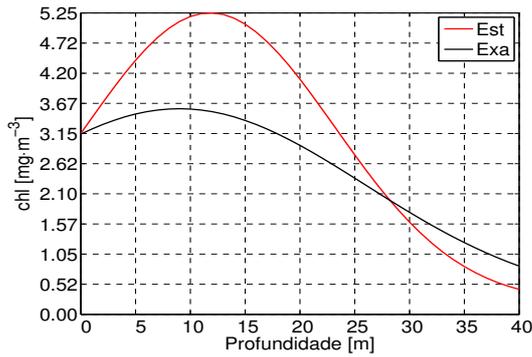
Grupo	Classe	Problema	σ [m]	σ^* [m]	e_σ [m]	τ_m [m]	τ_m^* [m]	e_τ [m]	Fig.
$\mathbf{G}_{h_{144}}$	1D _(6/10) λ	Aqua _{h_{144}}	17.00	11.38	5.62	9.00	11.79	-2.79	5.27(a)
		P _{1h_{144}}	11.00	15.36	-4.36	15.00	16.25	-1.25	5.28(a)
		P _{2h_{144}}	9.00	13.75	-4.75	14.00	16.87	-2.87	5.29(a)
	7D λ	P _{7Dh_{144}}	9.00	14.74	-5.74	21.00	23.51	-2.51	5.30(a)
$\mathbf{G}_{h_{30}}$	1D _(6/10) λ	Aqua _{h_{30}}	17.00	12.36	4.64	13.00	14.47	-1.47	5.27(c)
		P _{1h_{30}}	15.00	18.70	-3.70	28.00	29.52	-1.52	5.28(c)
		P _{2h_{30}}	9.00	14.11	-5.15	14.00	16.16	-2.16	5.29(c)
	7D λ	P _{7Dh_{30}}	13.00	16.32	-3.32	16.00	15.94	0.06	5.30(c)

Tabela 5.16 - Piores estimativas gerais, sobre o conjunto de validação, para os valores da profundidade do pico da concentração de clorofila e os respectivos valores do desvio padrão, para os problemas dos grupos $\mathbf{G}_{h_{144}}$ e $\mathbf{G}_{h_{30}}$.

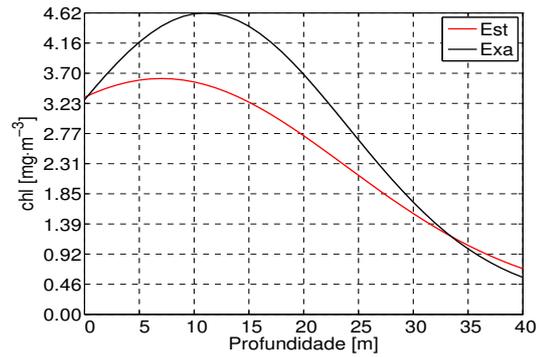
Grupo	Classe	Problema	σ [m]	σ^* [m]	e_σ [m]	τ_m [m]	τ_m^* [m]	e_τ [m]	Fig.
$\mathbf{G}_{h_{144}}$	1D _(6/10) λ	Aqua _{h_{144}}	13.00	16.82	-3.82	11.00	7.03	3.97	5.27(b)
		P _{1h_{144}}	11.00	14.35	-3.35	20.00	24.10	-4.10	5.28(b)
		P _{2h_{144}}	11.00	14.67	-3.67	21.00	26.81	-5.81	5.29(b)
	7D λ	P _{7Dh_{144}}	17.00	15.83	1.17	11.00	4.40	6.60	5.30(b)
$\mathbf{G}_{h_{30}}$	1D _(6/10) λ	Aqua _{h_{30}}	17.00	15.68	1.32	2.00	7.15	-5.15	5.27(d)
		P _{1h_{30}}	13.00	16.47	-3.47	28.00	32.67	-4.67	5.28(d)
		P _{2h_{30}}	17.00	17.82	-0.82	10.00	3.79	6.21	5.29(d)
	7D λ	P _{7Dh_{30}}	11.00	12.21	-1.21	6.00	-0.23	6.23	5.30(d)

Nesse último caso, ocorrem duas situações interessantes: uma na Tabela 5.15 e outra na Tabela 5.16. Na Tabela 5.15, a qual apresenta a pior estimativa feita pelas redes que determinam o valor do desvio padrão, é possível observar que, para o problema P_{7D h_{30}} (Figura 5.30(c)), a estimativa do valor da profundidade do pico da concentração de clorofila foi considerada como resposta correta, segundo a regra definida pela Eq. (5.2). Observe que o erro na estimativa de τ_m^* é de 0.06, satisfazendo o critério. Portanto, a pior resposta para as estimativas do desvio padrão apresenta uma boa

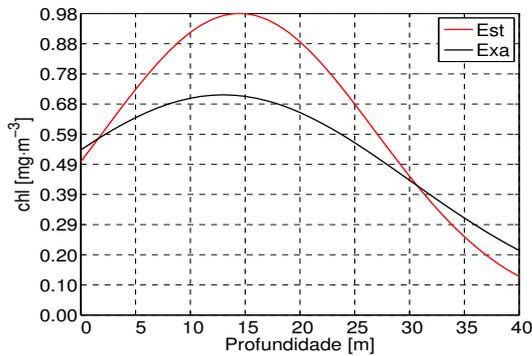
estimativa para o pico da concentração de clorofila. Na Tabela 5.16, o problema que apresenta essa característica é o $P_{2h_{30}}$ (Figura 5.29(d)). A estimativa para a profundidade do pico apresenta um erro significativo, enquanto a estimativa do valor para o desvio padrão está dentro do intervalo de tolerância definido pela Eq. (5.2). Portanto, essa é outra situação que colaborou para a taxa de acerto, mas sob outro ponto de vista, pode ser considerada uma estimativa pobre.



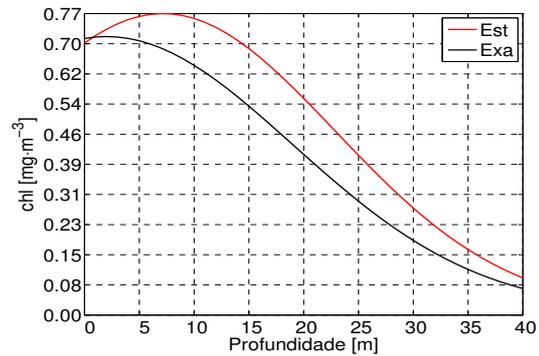
(a) Perfil obtido a partir da pior estimativa de σ^* para o problema $Aqua_{h_{144}}$.



(b) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $Aqua_{h_{144}}$.

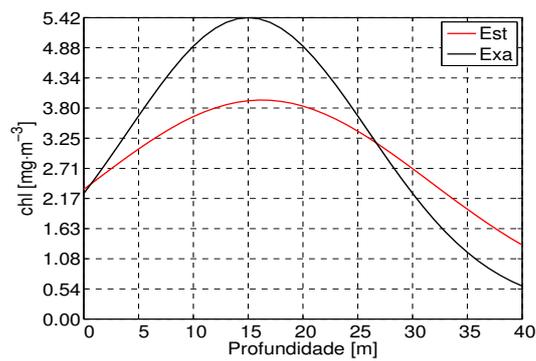


(c) Perfil obtido a partir da pior estimativa de σ^* para o problema $Aqua_{h_{30}}$.

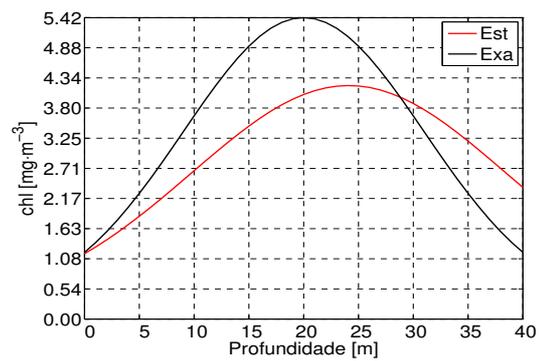


(d) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $Aqua_{h_{30}}$.

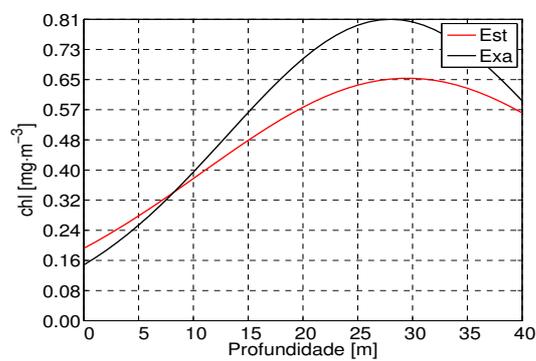
Figura 5.27 - Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para os problemas $Aqua_{h_*}$.



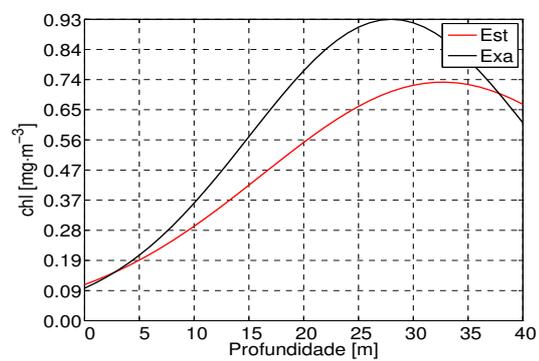
(a) Perfil obtido a partir da pior estimativa de σ^* para o problema $P_{1h_{144}}$.



(b) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $P_{1h_{144}}$.

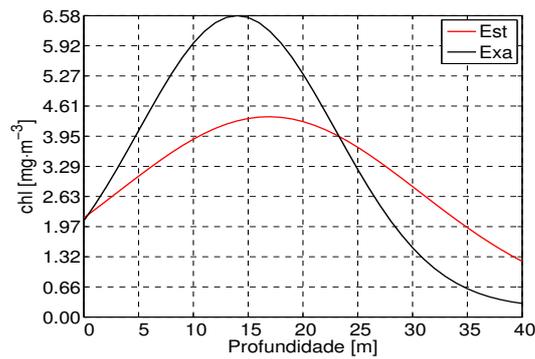


(c) Perfil obtido a partir da pior estimativa de σ^* para o problema $P_{1h_{30}}$.

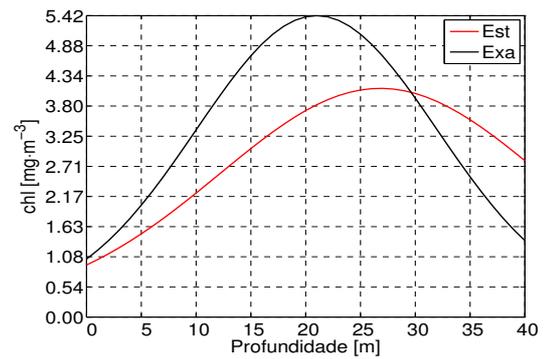


(d) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $P_{1h_{30}}$.

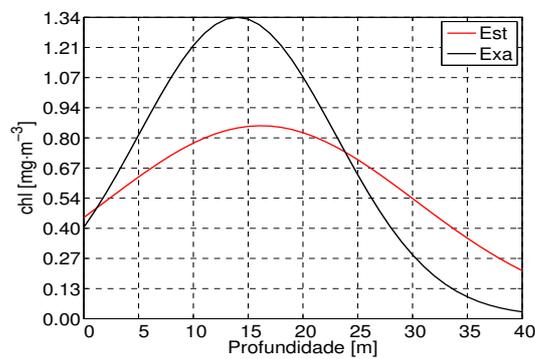
Figura 5.28 - Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para o problemas P_{1h_*} .



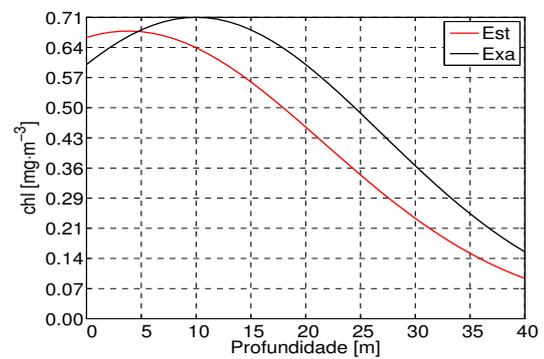
(a) Perfil obtido a partir da pior estimativa de σ^* para o problema $P_{2h_{144}}$.



(b) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $P_{2h_{144}}$.

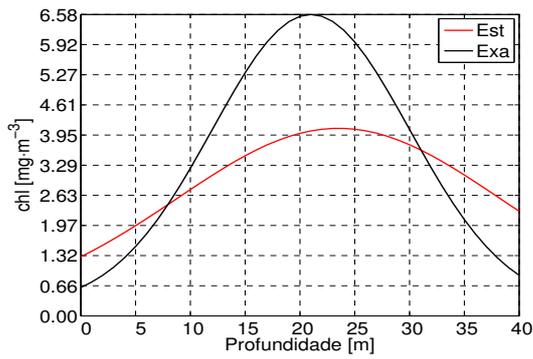


(c) Perfil obtido a partir da pior estimativa de σ^* para o problema $P_{2h_{30}}$.

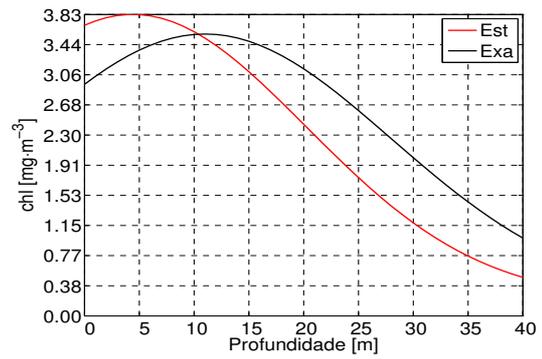


(d) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $P_{2h_{30}}$.

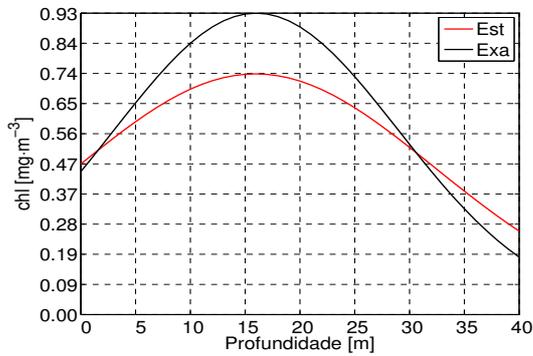
Figura 5.29 - Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para o problemas P_{2h_*} .



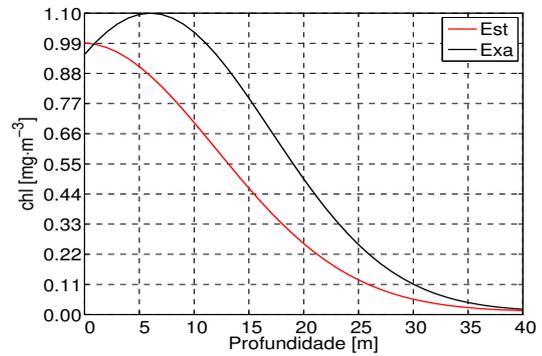
(a) Perfil obtido a partir da pior estimativa de σ^* para o problema $P_{7Dh_{144}}$.



(b) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $P_{7Dh_{144}}$.



(c) Perfil obtido a partir da pior estimativa de σ^* para o problema $P_{7Dh_{30}}$.



(d) Perfil obtido a partir da pior estimativa de τ_m^* para o problema $P_{7Dh_{30}}$.

Figura 5.30 - Perfis obtidos a partir das piores estimativas gerais do desvio padrão e da profundidade do pico da concentração de clorofila para os problemas $P_{7Dh_{*}}$.

A partir das Figuras 5.19 – 5.22, que mostram os perfis obtidos com as estimativas apresentadas nas Tabela 5.11 e 5.12, é possível notar que o perfil estimado é muito próximo (ou sobrepõe) ao perfil exato. Por outro lado, nas Figuras 5.23 – 5.26, que apresentam as piores estimativas (dentro do conjunto de acertos) para os valores de σ^* e de τ_m^* , é possível observar que os perfis não estão tão próximos como no caso anterior, no entanto, há boas estimativas como mostrado nas Figuras 5.24(b), 5.24(d), 5.25(b), 5.25(d), 5.26(b) e 5.26(d). Já os perfis estimados mostrados nas Figuras 5.27 – 5.30 apresentam uma grande “discrepância” com relação aos perfis exatos.

As figuras citadas no parágrafo anterior como boas estimativas, todas elas estão relacionadas à estimativa de τ_m^* , pelo motivo de que este parâmetro está no termo exponencial da função que representa a distribuição da concentração de clorofila ao longo da profundidade. Logo, um erro significativo na estimativa desse parâmetro acaba sendo reduzido pela exponencial. Observe que o mesmo não ocorre para as estimativas do desvio padrão, em especial, para valores pequenos deste, porque esse valor está no denominador do modelo adotado para o perfil.

Talvez a melhor medida para avaliar a qualidade das respostas, sejam elas consideradas acertos ou não, pelo critério definido na Eq. (5.2), seja o erro cometido na estimativa desses parâmetros, *i. e.*, $e^* = \nu - \nu^*$, em que ν e ν^* são, respectivamente, os valores exatos e estimados. Isso pode ser justificado por meio dos problemas $P_{7Dh_{144}}$, apresentados nas Tabela 5.11 e 5.16, respectivamente. O erro na estimativa de τ_m^* no primeiro problema é: $e_1^* = 1.0 - 0.1 = 0.9$ m, e no segundo problema é: $e_2^* = 11.0 - 4.40 = 6.6$ m, logo a estimativa feita no segundo problema é pobre em relação à estimativa feita para o primeiro problema. Por outro lado, se for calculado o erro relativo cometido nessas estimativas, conclui-se para o primeiro problema que

$$e_{r1}^* = \frac{1.0 - 0.10}{1.0} \times 100 = 90.0\%.$$

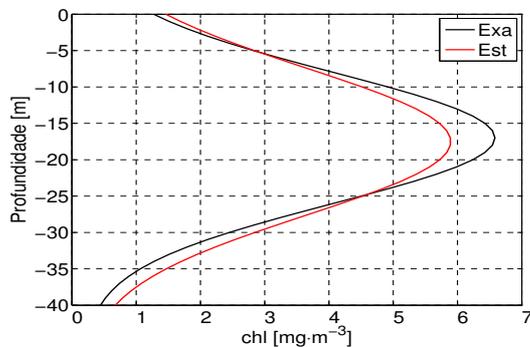
Já para o segundo problema, conclui-se que

$$e_{r2}^* = \frac{11.0 - 4.40}{11.0} \times 100 = 60.0\%.$$

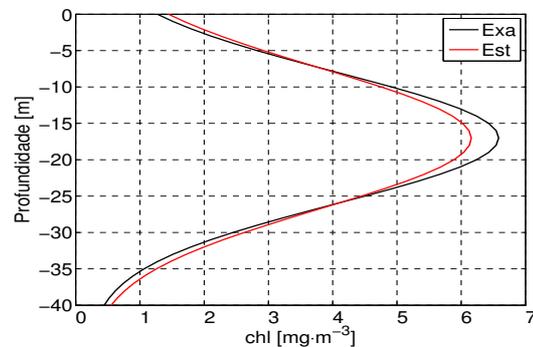
Portanto, é possível concluir que a estimativa feita para o segundo problema seria “melhor” com relação à estimativa feita para o primeiro problema, o que não é verdade. Veja Figura 5.22(a) para o primeiro problema e a Figura 5.30(b) para o segundo.

As Figuras 5.31 apresentam as estimativas do desvio padrão e da profundidade do pico da concentração de clorofila a partir da radiação emergente obtida ao se considerar o perfil mostrado na Figura 4.2. A Figura 5.31(a) apresenta a estimativa para esses parâmetros a partir de dados de entrada (radiâncias) corrompidos com 1% de ruído, e a Figura 5.31(b) apresenta a estimativa para radiâncias corrompidas com 2% de ruído. O problema que está sendo levado em conta é o problema $P_{1h_{144}}$, *i.e.*, as medidas da radiação emergente são feitas em $\theta = 0^\circ$ considerando os comprimentos de onda definidos na Eq. (4.9).

Essas estimativas foram escolhidas de modo proposital a fim de comparar os resultados obtidos com os do trabalho de Souto (2006). Os perfis recuperados por esse autor, para radiâncias corrompidas com 1% de ruído, são mostrados na Figuras 5.32, e para radiâncias corrompidas com 2% de ruído, são mostrados na Figura 5.33. O problema resolvido por Souto (2006) considera dez comprimentos de onda no intervalo de 410 nm e 700 nm, e dez direções polares. A semelhança entre o problema resolvido por este autor e por Souto (2006) está no fato de que, para ambos os problemas, a ETR é sem simetria azimutal, com anisotropia, e ambos levam em conta apenas a radiação emergente na superfície para a resolução do problema.



(a) Perfil gerado a partir das estimativas do desvio padrão e da profundidade do pico da concentração de clorofila levando em conta dados corrompidos com 1% de ruído.



(b) Perfil gerado a partir das estimativas do desvio padrão e da profundidade do pico da concentração de clorofila levando em conta dados corrompidos com 2% de ruído.

Figura 5.31 - Perfis da concentração de clorofila estimados a partir das radiâncias, obtidas levando em conta o perfil mostrado na Figura 4.2, corrompidas com 1% e 2% de ruído.

A fim de melhorar as estimativas mostradas nas Figuras 5.32 e 5.33, Souto et al. (2007) incluem o critério da derivada segunda para a “seleção” dos perfis de concentração de clorofila, melhorando significativamente seus resultados. No entanto, o problema da ótica considerado, para se obter os resultados mostrados nas Fi-

Figuras 5.34, apresenta simetria azimutal e espalhamento isotrópico. A semelhança com o problema resolvido nesta tese se dá pelo fato de serem considerados vários comprimentos de onda (problema multiespectral) e de ser usada apenas a radiação emergente na superfície.

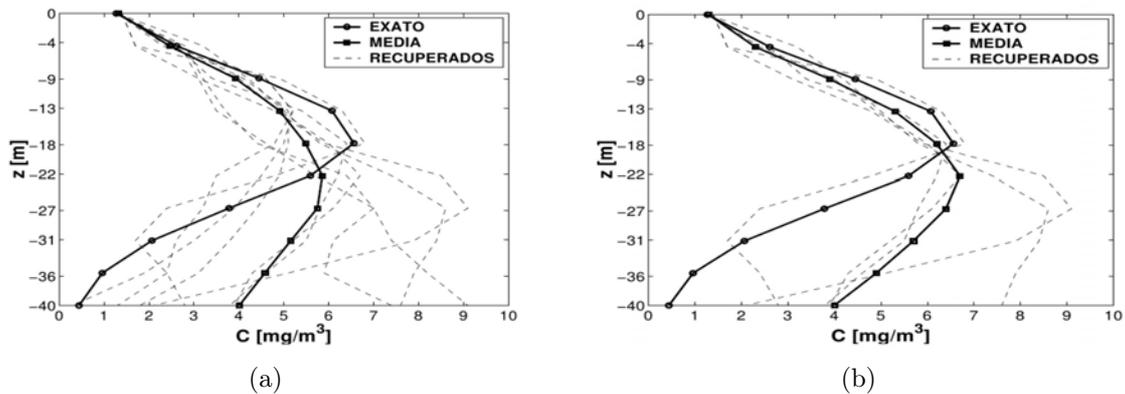


Figura 5.32 - Perfis obtidos com dados (radiâncias) corrompidas com 1% de ruído.
Fonte: Souto (2006), Figuras 4.27(c) e (d) da página 124.

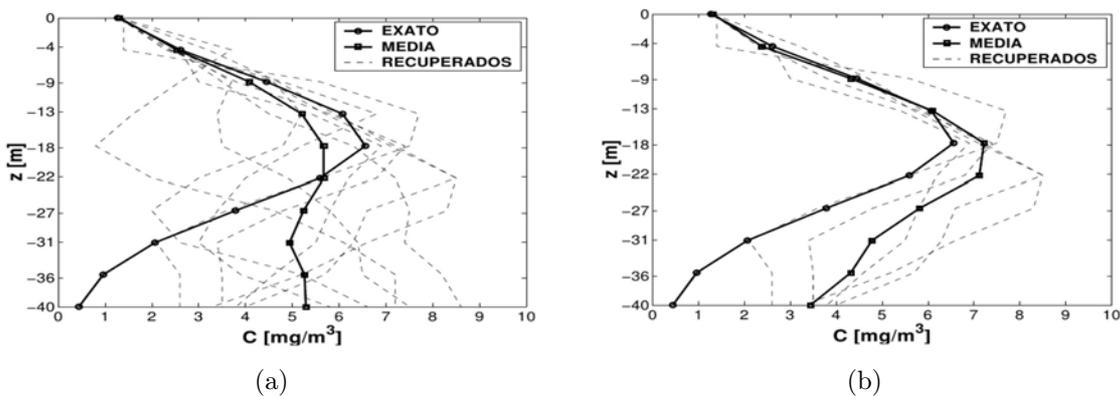
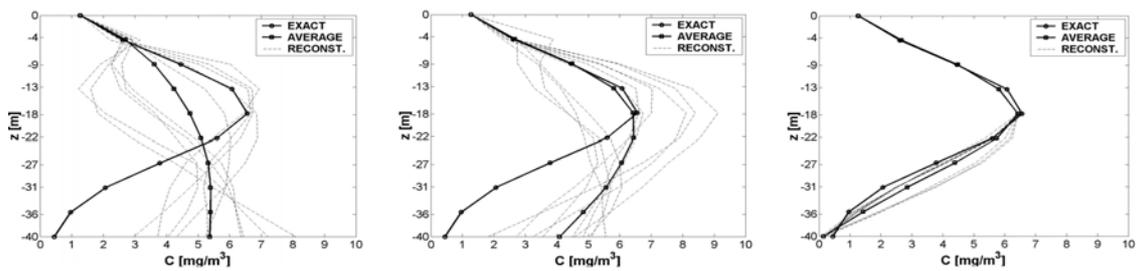


Figura 5.33 - Perfis obtidos com dados (radiâncias) corrompidas com 2% de ruído.
Fonte: Souto (2006), Figuras 4.27(e) e (f) da página 124.

A Figura 5.34(a) apresenta a estimativa obtida a partir da média de dez soluções. A Figura 5.34(b) mostra, também, a estimativa obtida a partir da média de dez soluções, levando em conta, no entanto, o critério da derivada segunda para descartar aquelas soluções que não satisfizessem tal critério. Neste caso, há uma melhora significativa da resposta até onde se localiza o pico máximo da concentração. Visto que a solução até onde se localiza o pico da concentração foi boa, a estratégia adotada por Souto et al. (2007) foi “congelar” essa parte da solução e, então, seguir o processo iterativo até obter a resposta mostrada na Figura 5.34(c).

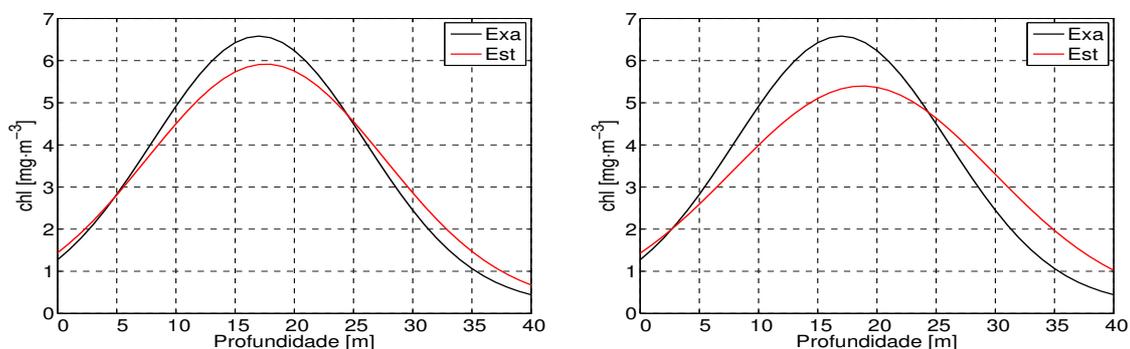


(a) Perfil estimado a partir da média de dez soluções sem o critério da derivada segunda. (b) Perfil estimado a partir da média de dez soluções utilizando o critério da derivada segunda. (c) Perfil médio recuperado a partir de dez soluções utilizando o critério da derivada segunda e fazendo a estimativa em duas etapas.

Figura 5.34 - Perfis recuperados por Souto et al. (2007).

Fonte: Souto et al. (2007), Figura 4(a), 4(b) e Figura 5.

Para encerrar, as Figuras 5.35 mostram as estimativas obtidas a partir de radiâncias corrompidas com 3% e 5% de ruído, utilizando a metodologia proposta nesta tese. Para esses dois níveis de ruído, Souto (2006) e Souto et al. (2007) não apresentaram resultados. A Tabelas 5.17 e 5.18 mostram os valores exatos e estimados pelas redes para o desvio padrão e para a profundidade do pico da concentração de clorofila para cada um dos níveis de ruído. Observe que, para esse exemplo, a metodologia proposta apresentou um acerto, segundo a Eq. (5.2), na estimativa do desvio padrão, e três acertos na estimativa da profundidade do pico da concentração de clorofila. No entanto, as estimativas do desvio padrão, para os níveis de ruído 1% e 3%, estão muito próximas do limite superior para contagem de acertos para esse valor de desvio padrão.



(a) Perfil estimado a partir de radiâncias corrompidas com 3% de ruído. (b) Perfil estimado a partir de radiâncias corrompidas com 5% de ruído.

Figura 5.35 - Perfis da concentração de clorofila estimados a partir das radiâncias, obtidas levando em conta o perfil mostrado na Figura 4.2, corrompidas com 3% e 5% de ruído.

Tabela 5.17 - Estimativas do desvio padrão a partir da radiância emergente, obtida ao se considerar o perfil mostrado na Figura 4.2, corrompida com quatro níveis de ruído.

Problema	σ [m]	σ^* [m]	Ruído	Acerto
P _{1h₁₄₄}	9.0	10.08	1.0%	×
		9.64	2.0%	✓
		10.04	3.0%	×
		11.05	5.0%	×

Tabela 5.18 - Estimativas da profundidade do pico a partir da radiância emergente, obtida ao se considerar o perfil mostrado na Figura 4.2, corrompida com quatro níveis de ruído.

Problema	τ_m [m]	$\bar{\tau}_m^*$ [m]	Ruído	Acerto
P _{1h₁₄₄}	17.0	17.51	1.0%	✓
		17.06	2.0%	✓
		17.58	3.0%	✓
		18.77	5.0%	×

6 CONCLUSÕES

Embora o problema seja sensível à presença de ruído e apresente limitações físicas e/ou limitações no modelo matemático adotado para o problema direto, os resultados obtidos mostraram que a metodologia proposta é promissora para a resolução dos problemas tratados nesta tese, *i.e.*, a estimativa do perfil da distribuição da concentração de clorofila, ao longo da profundidade, em águas oceânicas do Caso 1. Essa afirmação é justificada pela taxa de acerto obtida sobre o conjunto de validação, que ficou entre 60% e 70%; ainda, o erro cometido nas estimativas foi, de certa forma, “controlado”, *i.e.*, ficou em limites “aceitáveis”. Além disso, os problemas P_{1h*} , que utilizam medidas na direção $\theta = 0^\circ$ e nos comprimentos de onda definidos na Eq. (4.9), apresentaram as melhores respostas.

Atribui-se, também, a boa qualidade das respostas à estratégia de treinamento adotada (validação cruzada mais gradientes analíticos) e, principalmente, ao algoritmo de treinamento (quasi-Newton). A inclusão do parâmetro de cada função de ativação no conjunto de variáveis a serem otimizadas aumenta a dimensão do espaço de busca do ponto ótimo (melhor solução sob o ponto de vista da generalização). No entanto, essa inclusão dos parâmetros no conjunto de variáveis permite que se tenha “diferentes formas” para cada função de ativação, permitindo, assim, talvez, que cada neurônio possa se especializar em uma região específica do conjunto de padrões. A inclusão de restrições nos pesos, *bias* e parâmetro de cada função de ativação mantém controlada a ordem de grandeza dessas variáveis e, assim, evita um possível *overflow* devido à presença do termo exponencial, ou então, uma possível saturação dos neurônios.

Outra grande vantagem da metodologia adotada nesta tese está relacionada ao baixo tempo computacional e ao baixo número de épocas necessárias para o treinamento das redes. Como comentado no capítulo anterior, apenas três redes ultrapassaram o tempo de 60 segundos para o treinamento. Essa baixa demanda computacional permite que o problema possa ser resolvido sem a necessidade de processamento de alto desempenho e, portanto, pode ser resolvido em computadores do tipo *desktop* ou *laptop* (como foi o caso desta tese). Outro fator que contribuiu para o baixo custo computacional e baixo número de épocas foi o cálculo dos gradientes, do funcional de diferenças quadráticas definido pela Eq. (3.8), de forma analítica, e a adição dessas expressões (gradientes) à biblioteca utilizada tornou o processo de treinamento rápido. Todos os códigos computacionais utilizados foram executados de modo sequencial.

A cascata de redes (Figura 4.23) se mostrou uma estratégia eficiente para a resolução do problema, pois o fato da primeira rede filtrar (avaliar se é ou não possível fazer a inversão) aqueles problemas em que o sinal (radiação emergente) apresenta os problemas comentados na Seção 4.2 facilita a atuação das redes seguintes. Ainda, mesmo para aqueles problemas em que não é possível determinar o perfil da distribuição da concentração de clorofila, a rede devolve uma resposta com um valor médio dessa distribuição ao longo da profundidade. Portanto, mesmo nessa situação, há uma resposta para o problema.

Uma característica física do problema é o fato que dois diferentes perfis podem gerar a mesma radiação emergente, *i.e.*, um perfil com o pico da concentração mais próximo à superfície, com um desvio padrão “pequeno”, pode gerar a mesma radiação emergente que um perfil com o pico da concentração de clorofila mais profundo, no entanto, com desvio padrão maior. O caminho adotado para contornar esse problema foi considerar a concentração de clorofila de superfície. Todavia, se o perfil estiver profundo e os valores do desvio padrão não tiverem uma diferença significativa entre eles, pode ocorrer que esses dois perfis, além de gerar a “mesma” radiação emergente na superfície, ter a mesma concentração de clorofila de superfície.

Nesta tese o cálculo do valor da concentração de superfície foi feito por meio da Eq. (4.11). No entanto, em aplicações práticas nas quais não se conhece o valor do desvio padrão e a profundidade do pico da concentração de clorofila, esse valor de clorofila de superfície pode ser calculado pelos algoritmos que utilizam a razão de banda (comentados na introdução desta tese), ou então, aplicar alguma outra metodologia para determinar tal valor.

O modelo utilizado considera apenas um tipo de distribuição (gaussiano) para a concentração de clorofila, e duas quantidades de concentração: uma mais elevada e associada ao valor de $h = 144 \text{ mg} \cdot \text{m}^{-3}$, e outra menos elevada e associada ao valor de $h = 30 \text{ mg} \cdot \text{m}^{-3}$. É claro que outras distribuições e/ou quantidades de clorofila devem ser consideradas.

A utilização de RNAs para a resolução de problemas inversos em ótica hidrológica é uma abordagem recente e que se mostrou promissora em virtude dos resultados obtidos. No entanto, visto a dificuldade do problema da ótica, é necessário que, durante a fase de treinamento das redes, o algoritmo de treinamento utilizado seja eficiente.

A possibilidade da implementação de RNAs em dispositivos de *hardware*, como por exemplo FPGA (*Field Programmable Gate Array*), permite que esses sistemas

possam ser embarcados em satélite, a fim de processar os dados de cor do oceano em “tempo real”. Nesse caso é necessário que seja desenvolvido um modelo, ou então, utilizar os existentes para fazer a interação entre oceano-atmosfera.

A seguir são citados alguns possíveis trabalhos futuros que podem vir a colaborar com a metodologia proposta nesta tese.

- As limitações do modelo matemático adotado, citado no capítulo 2, podem ser contornadas com um modelo híbrido. Nesse modelo híbrido, poderia ser utilizado o método de Monte Carlo para simular o comportamento da radiação eletromagnética na camada superficial da água e, após, onde a refração, ondulação, bolhas e espuma não têm mais efeito, poderia ser utilizado o código PEESNA. Nesse caso, a radiação calculada pelo método de Monte Carlo passaria a ser as condições de contorno para o código PEESNA.
- O desenvolvimento de um modelo com a finalidade de fazer a interação entre oceano-atmosfera. Esse modelo é necessário se as RNAs forem operar embarcas em satélites, pois é necessário levar em consideração as informações vindas da atmosfera devido ao espalhamento dos fótons.
- Com relação ao perfil adotado, é claro que um estudo detalhado deve/pode ser feito a fim de determinar o perfil mais adequado para cada região do oceano. Dessa forma, haveria um sistema de inversão (conjunto de RNAs) válido para cada região e a validação desse sistema, por meio de dados medidos *in situ*, por exemplo, garante a aplicação desses sistemas à estimativa dos perfis verticais da distribuição da concentração de clorofila em águas oceânicas do Caso 1.
- Relacionado aos limiares de corte, selecionados para cada um dos grupos de problemas, um estudo mais detalhado, para a escolha desses valores, poderá ser feito a fim de melhorar a escolha desses valores.
- Uma outra estratégia que poderia ser utilizada a fim de contornar o problema (valores de radiação emergente muito próximas para diferentes perfis) citado anteriormente, seria considerar regiões espaciais maiores e, então, determinar um perfil que esteja contido nessa região, sem a preocupação da posição exata, desde que esteja nessa região. Outro caminho seria relacionar “intervalos” da radiação emergente a “intervalos” de concentração de clorofila (ou picos), *i.e.*, determinar intervalos disjuntos de medidas

de radiação emergente e associá-los aos respectivos perfis que os geraram. Dessa forma, um determinado perfil responderia em apenas um intervalo de radiação emergente, facilitando, talvez, sua classificação. No entanto, essa última estratégia apresenta, aparentemente, um grau de complexidade elevado, tendo em vista a quantidade de variáveis envolvidas na modelagem dos coeficientes de absorção, espalhamento e perfil de concentração.

- Por fim, a utilização de neurônios auxiliares poderá contribuir com uma possível melhora nos resultados obtidos. Neurônios auxiliares seriam unidades de processamento presentes apenas na camada de saída, e que aprenderiam informações adicionais sobre os padrões de saída. Dessa forma, talvez, seria possível adicionar o critério da derivada segunda, ou então, alguma outra informação nesse neurônio auxiliar durante o treinamento. Após a rede treinada, esse neurônio pode ser eliminado pelo fato de que suas conexões são independentes das demais conexões dos neurônios da camada de saída, e não influenciam a propagação do sinal para frente.

REFERÊNCIAS BIBLIOGRÁFICAS

ACKER, J. G. **SeaWiFS technical report series: A retrospective on the CZCS NIMBUS Experiment Team (NET) program. The heritage of SeaWiFS.** Maryland: NASA, 1994. (NASA technical memorandum, 21). 6

AIKEN, J.; MOORE, G. F.; HOLLIGAN, P. M. Remote sensing of oceanic biology in relation to global climate change. **Journal of Phycology**, v. 28, p. 579–590, 1992. 4

ALIFANOV, O. M. **Inverse heat transfer problems.** Berlin: Springer-Verlag, 1994. (International series in heat and mass transfer). ISBN 9783540536796. 41

AMBARZUMIAN, B. On the problem of the diffuse reflection of light. **Journal of Physics**, v. 8, n. 2, p. 65–75, 1944. 13, 14

ANOCHI, J. A.; SAMBATTI, S. B. M.; LUZ, E. F. P. d.; CAMPOS VELHO, H. F. Previsão climática de precipitação usando rede neural. In: XVII CBMET CONGRESSO BRASILEIRO DE METEOROLOGIA, 2012, Gramado-RS. **Anais 2012.** Disponível em: <<http://urlib.net/dpi.inpe.br/plutao/2012/11.28.16.48.05>>. Acesso em: 10 abr. 2013. 29

BARICHELLO, L. B.; VILHENA, M. T. A general approach to one group one dimensional transport equation. **Kerntechnik**, v. 58, n. 3, p. 182–184, 1993. 14

BELL, G.; GLASSTONE, S. **Nuclear reactor theory.** New York: Van Nostrand Reinold, 1970. 13, 15, 46

BISHOP, C. M. **Neural networks for pattern recognition.** 1. ed. New York: Oxford University Press, USA, 1995. Paperback. ISBN 9780198538646. 27, 30, 33, 34, 39

_____. **Pattern recognition and machine learning: information science and statistics.** New York: Springer, 2006. Hardcover. ISBN 0387310738. 27, 30, 33, 39

BOKAR, J. C. The estimation of spatially varying albedo and optical thickness in a radiating slab using artificial neural networks. **International Communications in Heat and Mass Transfer**, v. 26, n. 3, p. 359–367, 1999. ISSN 0735-1933. 29

BORTOLETTI, A.; FIORE, C. D.; FANELLI, S.; ZELLINI, P. A new class of quasi-Newtonian methods for optimal learning in MLP-networks. **IEEE**

Transactions on Neural Networks, v. 14, n. 2, p. 263–273, 2003. ISSN 1045-9227. [34](#)

BOYCE, W. E.; DIPRIMA, R. C. **Equações diferenciais elementares e problemas de valores de contorno**. 7. ed. Rio De Janeiro: LTC, 2002. [145](#)

BURDEN, R. L.; FAIRES, J. D. **Análise numérica**. São Paulo: Thomson, 2003. [34](#), [44](#)

CAMPOS VELHO, H. F. **Problemas inversos em pesquisa espacial**. São Carlos: SBMAC, 2008. 128 p. [41](#)

CAMPOS VELHO, H. F.; RETAMOSO, M. R.; VILHENA, M. T. Inverse problems for estimating bottom boundary conditions of natural waters. **International Journal for Numerical Methods in Engineering**, v. 54, n. 9, p. 1357–1368, 2002. [2](#), [5](#)

CAMPOS VELHO, H. F.; STEPHANY, S.; CHALHOUB, E. S.; RAMOS, F. M.; RETAMOSO, M. R.; VILHENA, M. T. New approaches on inverse hydrological optics. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, XXIII., 2000, Santos. **Anais . . .** São Carlos: SBMAC, 2000. p. 100–110. Mini-Symposium on Inverse Problems on Medicine, Engineering and Geophysics. [2](#)

CAMPOS VELHO, H. F.; VILHENA, M. T.; RETAMOSO, M. R.; PAZOS, R. P. An application of the LTS_N method on an inverse problem in hydrologic optics. **Progress in Nuclear Energy**, v. 42, n. 4, p. 457–468, 2003. [2](#)

CARDONA, A. V. **Método genérico de solução analítica para aproximações da equação linear de transporte**. Tese (Doutorado em Engenharia Mecânica) — Programa de Pós Graduação em Engenharia Mecânica, Universidade Federal do Rio Grande do Sul, 1996. [14](#)

CARDONA, A. V.; VILHENA, M. T. A solution of linear transport equation using Chebyshev polynomials and Laplace Transform. **Kernntechnik**, n. 59, p. 278–281, 1994. [14](#)

CARVALHO, A. R. **Uso de redes neurais otimizadas para recuperação do perfil de concentração de gases traço atmosféricos a partir de dados de satélites**. 208 p. Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m19/2011/06.06.12.20>>. Acesso em: 08 abr. 2013. 46

CASE, K. M.; ZWEIFEL, P. F. **Linear transport theory**. Michigan: Addison-Wesley, 1967. 14

CHALHOUB, E. S. **O método das ordenadas discretas na solução da equação de transporte em geometria plana com dependência azimutal**. 99 p. Tese (Doutorado) — Universidade de São Paulo (USP), São Paulo, 1997. Acesso em: 13 maio 2011. 14, 20, 60, 141, 143

CHALHOUB, E. S. Discrete-ordinates solution for uncoupled multi-wavelength radiative transfer problems. **Journal of Quantitative Spectroscopy & Radiative Transfer**, v. 92, p. 335–349, 2005. 15, 17, 141, 146

CHALHOUB, E. S.; CAMPOS VELHO, H. F. Simultaneous estimation of radiation phase function and albedo in natural waters. **Journal of Quantitative Spectroscopy & Radiative Transfer**, v. 69, n. 2, p. 137–149, 2001. 2, 5

_____. Estimation of the optical properties of sea water from measurements of exit radiance. **Journal of Quantitative Spectroscopy & Radiative Transfer**, v. 72, n. 5, p. 551–565, 2002. 2, 5

_____. Multispectral reconstruction of bioluminescence term in natural waters. **Applied Numerical Mathematics**, v. 47, n. 3-4, p. 365–376, 2003. 2, 5

CHALHOUB, E. S.; CAMPOS VELHO, H. F.; GARCIA, R. D. M.; VILHENA, M. T. A comparison of radiances generated by selected methods of solving the radiative-transfer equation. **Transport Theory and Statistical Physics**, v. 32, n. 5-7, p. 473–503, 2003. 20, 60, 143

CHALHOUB, E. S.; CAMPOS VELHO, H. F.; RAMOS, F. M.; CLAEYSSSEN, J. C. R. Phase function estimation in natural waters using discrete-ordinates method and maximum entropy principle. **Hybrid Methods in Engineering**, v. 2, n. 4, p. 373–388, 2000. 2, 5

CHALHOUB, E. S.; GARCIA, R. D. M. The equivalence between two techniques of angular interpolation for the discrete-ordinates method. **Journal of Quantitative Spectroscopy & Radiative Transfer**, v. 64, n. 5, p. 517–535, 2000. 14, 141

CHALHOUB, E. S.; SILVA NETO, A. J.; SOEIRO, F. J. C. P. Estimation of optical thickness and single scattering albedo with artificial neural networks and a Monte Carlo method. **Inverse Problems, Design and Optimization Symposium**, v. 2, p. 576–583, 2007. [29](#), [46](#)

CHANDRASEKHAR, S. **Radiative transfer**. New York: Dover Publications, 1950. [13](#), [14](#), [19](#), [142](#), [143](#), [151](#)

CHO, T. H.; CONNERS, R. W.; ARAMAN, P. A. Fast Backpropagation learning using steep activation functions and automatic weight reinitialization. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. **Proceedings** New York: IEEE Press, 1991. v. 3, p. 1587–1592. [33](#), [34](#)

CHUN, L.; BINGXUE, S. Hybrid back-propagation/genetic algorithm for multilayer feedforward neural networks. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING, 5th., 2000, Beijing, China. **Proceedings** New York: IEEE Press, 2000. v. 1, p. 571–574. [33](#)

CLARK, D. K.; SHERMAN, J. W. Nimbus – 7 Coastal Zone Color Scanner: ocean color applications. **Marine Technology Society Journal**, v. 20, n. 2, p. 43–56, 1986. [4](#)

CORTIVO, F. D. **Reconstrução de termo fonte e condições de contorno em óptica hidrológica com séries de Fourier**. 106 p. Dissertação (Mestrado em Modelagem Computacional) — Fundação Universidade Federal do Rio Grande, Rio Grande, 2008. [15](#), [141](#), [151](#), [152](#)

CORTIVO, F. D.; CHALHOUB, E. S.; CAMPOS VELHO, H. F. A committee of MLP with adaptive slope parameter trained by the quasi-Newton method to solve problems in hydrologic optics. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 2012, Brisbane, QLD, Australia. **Proceedings** New York: IEEE Press, 2012. p. 1–8. ISBN 978-1-4673-1409-9. [2](#), [5](#), [37](#), [39](#), [69](#)

CORTIVO, F. D.; CHALHOUB, E. S.; CAMPOS VELHO, H. F. Comparison of two learning strategies for a supervised neural network. In: INTERNATIONAL SYMPOSIUM ON UNCERTAINTY QUANTIFICATION AND STOCHASTIC MODELLING, 1st., 2012, Maresias, SP, Brazil. **Proceedings** São Paulo, 2012. p. 366–380. ISSN 2238-1007. CD-Rom. [2](#), [5](#), [29](#), [37](#), [39](#)

_____. Estimativa do perfil da concentração de clorofila em águas naturais através de um perceptron de múltiplas camadas. **Tendências em Matemática Aplicada e Computacional**, SBMAC, v. 13, n. 3, p. 233–246, 2012. 3, 5, 29, 37, 39, 54

CORTIVO, F. D.; CHALHOUB, E. S.; SILVA, J. D. S.; CAMPOS VELHO, H. F. Estimativa do albedo de espalhamento simples usando uma rede neural de múltiplas camadas. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, XXXIII., 2010, Águas de Lindóia, SP, Brasil. **Anais** São Carlos: SBMAC, 2010. v. 3, p. 411–417. Disponível em: <http://www.sbmac.org.br/eventos/cnmac/xxxiii_cnmac/pdf/213.pdf>. Acesso em: 5 março 2013. 2, 5, 29

CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals and Systems**, v. 2, n. 4, p. 303–314, 1989. 29

DANDONNEAU, Y.; DESCHAMPS, P.-Y.; NICOLAS, J.-M.; LOISEL, H.; BLANCHOT, J.; MONTEL, Y.; THIEULEUX, F.; BÉCU, G. Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North atlantic, equatorial Pacific and South Pacific. **Deep Sea Research Part II: topical studies in Oceanography**, v. 51, n. 1–3, p. 303–318, 2004. ISSN 0967-0645. 8

DENNIS, J. E.; MORÉ, J. J. Quasi-Newton methods, motivation and theory. **Society for Industrial and Applied Mathematics Review**, v. 19, n. 1, p. 46–89, 1977. 34, 39, 44

DEVAUX, C.; SIEWERT, C. E. The F_N method for radiative transfer problems without azimuthal symmetry. **Zeitschrift fur Angewandte Mathematik und Physik**, v. 31, p. 592–604, 1980. 14

DORIGO, M. **Optimization, learning and natural algorithms**. Tese (PhD Thesis) — Politecnico di Milano, Itália, 1992. 3, 44

DUDERSTADT, J. J.; MARTIN, W. R. **Transport theory**. New York: John Wiley & Sons, Inc., 1975. ISBN 9780608146003. 13

EVANS, R. H.; GORDON, H. R. Coastal Zone Color Scanner “system calibration”: a retrospective examination. **Journal of Geophysical Research: Oceans**, v. 99, n. C4, p. 7293–7307, 1994. ISSN 2156-2202. 6

FAHLMAN, S. E. **Computer science technical report**: An empirical study of learning speed in back-propagation networks. Los Altos CA: Carnegie-Mellon University, 1988. 33

FAUSETT, L. **Fundamentals of neural networks**: architectures, algorithms, and applications. New Jersey: Prentice Hall International, 1994. 21, 22, 27, 30, 33, 69

FELDMAN, G. C.; KURING, N. A.; NG, C.; ESAIAS, W. E.; MCCLAIN, C. R.; ELROD, J. A.; MAYNARD, N.; ENDRES, D.; EVANS, R.; BROWN, J.; WALSH, S.; CARLE, M.; PODESTA, G. Ocean Color: availability of the global data set. **EOS**, v. 70, p. 634–641, 1989. 4, 6

FINKEL, Z. V.; BEARDALL, J.; FLYNN, K. J.; QUIGG, A.; REES, A. V.; RAVEN, J. A. Review – phytoplankton in a changing world: cell size and elemental stoichiometry. **Journal of Plankton Research**, v. 32, p. 119–137, 2010. 4

FIRESTONE, E. R.; HOOKER, S. B. **NASA Technical Memorandum 1998-104566**: SeaWiFS prelaunch technical report series final cumulative index. Maryland: NASA, 1998. v. 43, 4–8 p. 6

GARCIA, C. A. E.; SARMA, Y. V. B.; MATA, M. M.; GARCIA, V. M. T. Chlorophyll variability and eddies in the Brazil–Malvinas Confluence region. **Deep-sea Research Part II: topical studies in Oceanography**, v. 51, p. 159–172, 2004. 7

GARCIA, R. D. M. A review of the Facile (F_N) method in particle transport theory. **Transport Theory and Statistical Physics**, v. 14, p. 391–435, 1985. 14

GARRISON, T. **Essentials of oceanography**. 6. ed. New York: Brooks/Cole, 2011. ISBN 9780840061553. 54

GILL, J.; SINGH, B.; SINGH, S. Training back propagation neural networks with genetic algorithm for weather forecasting. In: IEEE INTERNATIONAL SYMPOSIUM ON INTELLIGENT SYSTEMS AND INFORMATICS, 8th., 2010, Subotica, Serbia. **Proceedings** New York: IEEE Press, 2010. p. 465–469. 33

GONZÁLEZ-SILVERA, A.; SANTAMARÍA DEL, A. E. M.; MILLÁN-NÚÑEZ, R.; MANZO-MONROY, H. Satellite observations of mesoscale eddies in the Gulfs of Tehuantepec and Papagayo (Eastern Tropical Pacific). **Deep Sea Research Part II: topical studies in Oceanography**, v. 51, n. 6, p. 587–600, 2004. ISSN 0967-0645. 7

GORDON, H. R. Inverse methods in hydrologic optics. **Oceanology**, v. 44, n. 1, p. 9–58, 2002. 41

GORDON, H. R.; CLARK, D. K.; BROWN, O. B.; EVANS, R. H.; BROENKOW, W. W. Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparisons between ship determinations and Coastal Zone Color Scanner estimates. **Applied Optics**, v. 22, n. 1, p. 20–36, 1983. 6

GORDON, H. R.; MOREL, A. Y. **Remote assessment of ocean color for interpretation of satellite visible imagery**: a review. New York: Springer-Verlag, 1983. (Lecture notes on coastal and estuarine studies). ISBN 9783540909231. 3, 8, 47

GRANDJEAN, P.; SIEWERT, C. E. The F_N method in neutron-transport theory, Part II: applications and numerical results. **Nuclear Science and Engineering**, v. 69, p. 161–168, 1979. 14

GROETSCH, C. W. **Inverse problems**: activities for undergraduates. New York: Cambridge University Press, 1999. (Classroom Resource Material Series). ISBN 9780883857168. 40, 41

GROSS, L.; THIRIA, S.; FROUIN, R.; MITCHELL, B. G. Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. **Journal of Geophysical Research**, v. 105, n. C2, p. 3483–3495, 2000. 7, 29

GUYTON, A. C.; HALL, J. E. **Fisiologia médica**. 11. ed. Rio de Janeiro: Elsevier, 2006. 22, 23, 155, 156, 158, 159, 160

HADAMARD, J. Sur les problèmes aux dérivés partielles et leur signification physique. **Princeton University Bulletin**, v. 13, p. 49–52, 1902. 40

HAGAN, M. T.; MENHAJ, M. B. Training feedforward networks with the Marquardt algorithm. **IEEE Transactions on Neural Networks**, v. 5, n. 6, p. 989–993, nov 1994. ISSN 1045-9227. 34

HAYKIN, S. **Redes neurais**: princípios e prática. Porto Alegre: Bookman, 2001. 7, 21, 22, 24, 27, 28, 30, 31, 32, 33, 34, 39, 46, 69, 170

HENSON, S. A.; DUNNE, J. P.; SARMIENTO, J. L. Decadal variability in north atlantic phytoplankton blooms. **Journal of Geophysical Research: Oceans**, v. 114, n. C4, p. 1–11, 2009. ISSN 2156-2202. 8

HENYEY, L. C.; GREENSTEIN, J. L. Diffuse radiation in the galaxy.

Astrophysical Journal, v. 93, p. 70–83, 1941. 50, 142

HOFF, M. E. **Learning phenomena in networks of adaptive switching circuits**. Tese (PhD Thesis) — Stanford, Tech. Rep, 1962. 21

HOLLAND, J. H. **Adaptation in natural and artificial systems**. Cambridge, MA, USA: MIT Press, 1975. ISBN 0-262-58111-6. 33, 44

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-Biological Science**, v. 8, n. 79, p. 2554–2558, 1982. 22

IOCCG. **Reports of the international ocean-colour coordinating group: Remote sensing of ocean colour in coastal and other optically-complex waters**. Dartmouth, Canada: IOCCG, 2000. 140 p. Disponível em: <<http://www.ioccg.org/reports/report3.pdf>>. Acesso em: 13 jan. 2013. 7

JACOBS, R. A. Increased rates of convergence through learning rate adaptation. **Neural Networks**, v. 1, n. 4, p. 295–307, 1988. 33

JEANS, J. H. The equations of radiative transfer of energy. **Monthly Notices of the Royal Astronomical Society**, v. 78, p. 28–36, 1917. 14

KAMEDA, T.; MATSUMURA, S. Chlorophyll biomass off Sanriku, Northwestern Pacific, estimated by Ocean Color and Temperature Scanner (OCTS) and vertical distribution model. **Journal of Oceanography**, v. 54, p. 509–516, 1998. 10

KAMPEL, M. **Estimativa da produção primária e biomassa fitoplanctônica através de sensoriamento remoto da cor do oceano e dados *in situ* na Costa Sudeste Brasileira**. Tese (Doutorado em Oceanografia Biológica) — Instituto Oceanográfico, USP, São Paulo, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/21/21131/tde-29042004-165724/>>. Acesso em: 14 jan. 2013. 7

KAMPEL, M.; NOVO, E. M. L. M. O sensoriamento remoto da cor da água. In: SOUZA, R. B. (Ed.). **Oceanografia por satélites**. São Paulo: Oficina de Textos, 2005. Cap. 12, p. 199–218. ISBN 85-86238-48-1. 6, 48

KAUATI, A. T.; SILVA NETO, A. J.; ROBERTY, N. C. A source-detector methodology for the construction and solution of the one-dimensional inverse

transport equation. **Inverse Problems in Science & Engineering**, v. 9, n. 1, p. 45–66, 2001. 41

KAWAMURA, H.; THE OCTS TEAM. OCTS Mission Overview. **Journal of Oceanography**, v. 54, p. 383–399, 1998. 6

KIRSCH, A. **An introduction to the mathematical theory of inverse problems**. New York: Springer-Verlag, 1996. (Applied Mathematical Sciences, v. 120). ISBN 1441984739. 41, 44

KLOSE, A. D.; HIELSCHER, A. H. Optical tomography using the time-independent equation of radiative transfer – Part 2: inverse model. **Journal of Quantitative Spectroscopy & Radiative Transfer**, v. 2, n. 5, p. 715–732, 2002. 13

KLOSE, A. D.; NETZ, U.; BEUTHAN, J.; HIELSCHER, H. A. Optical tomography using the time-independent equation of radiative transfer – Part 1: forward model. **Journal of Quantitative Spectroscopy & Radiative Transfer**, v. 72, n. 5, p. 691–713, 2002. 13

KLOSE, A. D.; NTZIACHRISTOS, V.; HIELSCHER, A. H. The inverse source problem based on the radiative transfer equation in optical molecular imaging. **Journal of Computational Physics**, v. 202, n. 1, p. 323–345, 2005. 13

KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biological Cybernetics**, v. 43, p. 59–69, 1982. 7, 22

_____. **Self-organization and associative memory**. 3. ed. Berlin: Springer-Verlag, 1989. 7, 30

LALLI, C. M.; PARSONS, T. R. **Biological oceanography**: an introduction. Oxford: Oxford: Pergamon-press, 1993. 301 p. 3

LEE, Y.; OH, S.-H.; KIM, M. W. The effect of initial weights on premature saturation in back-propagation learning. In: IEEE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 1991, Singapore. **Proceedings** New York: IEEE Press, 1991. v. 1, p. 765–770. 33

LERA, G.; PINZOLAS, M. A quasi-local Levenberg-Marquardt algorithm for neural network training. In: IEEE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS PROCEEDINGS, 1998, Washington, DC, USA. **Proceedings** New York: IEEE Press, 1998. v. 3, p. 2242–2246. ISSN 1098-7576. 34

- LEVENBERG, K. A method for the solution of certain problems in least-squares. **Quarterly Applied Mathematics**, n. 2, p. 164–168, 1944. 2, 34, 44
- LEWIS, E. E.; MILLER, J. W. **Computational methods of neutron transport**. Illinois: American Nuclear Society, 1993. 13
- LEWIS, M. R.; CULLEN, J. J.; PLATT, T. Phytoplankton and thermal structure in the upper ocean: consequences of nonuniformity in chlorophyll profile. **Journal of Geophysical Research**, v. 88, n. C4, p. 2565–2570, 1983. 7
- LUZ, E. F. P.; BECCENERI, J. C.; CAMPOS VELHO, H. F. A new multi-particle collision algorithm for optimization in a high performance environment. **Journal of Computational Interdisciplinary Sciences**, v. 1, n. 1, p. 3–9, 2008. 29, 44
- MARITORENA, S.; O'REILLY, J. E. **NASA Technical Memorandum 2000-206892**: Oc2v2: Update on the initial operational seawifs chlorophyll a algorithm. Maryland: NASA, 2000. v. 11, 3–8 p. 6
- MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. **Journal of the Society for Industrial and Applied Mathematics**, v. 11, n. 2, p. 431–441, 1963. 2, 34, 44
- MATSUMURA, S.; SHIOMOTO, A. Vertical distribution of primary productivity function $\Phi(\Pi)$: for the estimation of primary productivity using by satellite remote sensing. **Bulletin – National Research Institute of Far Seas Fisheries**, v. 30, p. 227–270, 1993. In Japanese with English abstract. 10
- MCCLAIN, C. R. Review of major CZCS applications: U.S. Case studies: theory and applications in a decade of CZCS experience. In: BARALE, V.; SCHLITTENHARDT, P. M. (Ed.). **Ocean Colour**. Ispra, Italy: Springer Netherlands, 1993, (Eurocourses: Remote Sensing, v. 3). p. 167–188. ISBN 978-94-010-4788-3. 4
- MCCORMICK, N. J. Inverse radiative transfer problems: a review. **Nuclear Science and Engineering**, v. 112, n. 185, p. 185–198, 1992. 41
- _____. Inverse problems: methods and applications. **Tendências em Matemática Aplicada e Computacional**, v. 2, n. 1, p. 1–12, 2001. 41
- MCCULLOCH, W.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115–133, 1943. 21, 23

- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **The Journal of Chemical Physics**, AIP, v. 21, n. 6, p. 1087–1092, 1953. [44](#)
- MINAI, A. A.; WILLIAMS, R. D. Acceleration of Backpropagation through learning rate and momentum adaptation. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 1990, San Diego, California. **Proceedings** New York: IEEE Press, 1990. v. 1, p. 676–679. [33](#)
- MINSKY, M.; PAPER, S. **Perceptrons**. Cambridge: MIT Press, 1969. [21](#)
- MITCHELL, B. G. Coastal zone color scanner retrospective. **Journal of Geophysical Research: Oceans**, v. 99, n. C4, p. 7291–7292, 1994. ISSN 2156-2202. [4](#)
- MOBLEY, C. D. **Light and water: radiative transfer in natural waters**. California: Academic Press, 1994. [13](#), [14](#), [50](#), [54](#), [152](#)
- MODEST, M. F. **Radiative heat transfer**. 2. ed. New York: Academic Press, 2003. [13](#)
- MOHAMAD, N.; ZAINI, F.; JOHARI, A.; YASSIN, I.; ZABIDI, A. Comparison between Levenberg-Marquardt and scaled conjugate gradient training algorithms for breast cancer diagnosis using MLP. In: INTERNATIONAL COLLOQUIUM ON SIGNAL PROCESSING AND ITS APPLICATIONS, 6th., 2010, Malacca City. **Proceedings** New York: IEEE Press, 2010. p. 1–7. [34](#)
- MONGER, B.; MCCLAIN, C.; MURTUGUDDE, R. Seasonal phytoplankton dynamics in the eastern tropical Atlantic. **Journal of Geophysical Research: Oceans**, v. 102, n. C6, p. 12389–12411, 1997. ISSN 2156-2202. [7](#)
- MOREL, A.; AHN, Y.-H. Optics of heterotrophic nanoflagellates and ciliates: a tentative assessment of their scattering role in oceanic waters compared to those of bacterial and algal cells. **Journal of Marine Research**, v. 49, n. 1, p. 177–202, 1991. [48](#)
- MOREL, A. Y. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. **Progress in Oceanography**, v. 26, n. 3, p. 263–306, 1991. [8](#), [47](#)
- MOREL, A. Y.; PRIEUR, L. Analysis of variations in ocean color. **Limnology and Oceanography**, v. 22, n. 4, p. 709–722, 1977. [3](#), [47](#)

MOROZOV, V. A. **Regularization methods for ill-posed problems**. Florida: CRC Press, 1993. 41

NAG. **Fortran library manual**: Mark 17: Library contents. Oxford, UK, 1995. 39, 40, 70

OCEAN COLOR, P. **Ocean color project**. 2013. Disponível em: <<http://oceancolor.gsfc.nasa.gov/>>. Acesso em: 15 jan. 2013. 6

_____. _____. 2013. Disponível em: <<http://oceancolor.gsfc.nasa.gov/REPROCESSING/R2009/ocv6/>>. Acesso em: 16 jan. 2013. 6

_____. _____. 2013. Disponível em: <<http://oceancolor.gsfc.nasa.gov/>>. Acesso em: 15 jan. 2013. 52, 53

OLIVA SOARES, P. **Um problema inverso em dois passos para estimação de perfis de temperatura na atmosfera com nuvens a partir de medidas de radiância feitas por satélite**. 141 p. Tese (Doutorado em Modelagem Computacional) — Instituto Politécnico, UERJ, Nova Friburgo, 2013. Disponível em: <http://www.bdttd.uerj.br/tde_busca/arquivo.php?codArquivo=4776>. Acesso em: 08 abr. 2013. 46

OLIVEIRA, R. C. **Aplicação de máquinas de comitê de redes neurais artificiais na solução de um problema inverso em transferência radiativa**. 135 p. Dissertação (Mestrado em Modelagem Computacional) — Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2010. 29, 46

OLIVEIRA, R. C.; ACEVEDO, N. I. A.; SILVA NETO, A. J.; NETO, L. B. Aplicação de um comitê de redes neurais artificiais para a solução de problemas inversos em transferência radiativa. **Tendências em Matemática Aplicada e Computacional**, SBMAC, v. 11, n. 2, p. 171–182, 2010. 29, 46

O'REILLY, J. E.; MARITORENA, S.; MITCHELL, B. G.; SIEGEL, D. A.; CARDER, K. L.; GARVER, S. A.; KAHRU, M.; MCCLAIN, C. Ocean Color chlorophyll algorithms for SeaWiFS. **Journal of Geophysical Research**, v. 103, n. C11, 1998. 6

O'REILLY, J. E.; MARITORENA, S.; SIEGEL, D. A.; O'BRIEN, M. C.; TOOLE, D.; MITCHELL, B. G.; KAHRU, M.; CHAVEZ, F. P.; STRUTTON, P.; COTA, G. F.; HOOKER, S. B.; MCCLAIN, C. R.; CARDER, K. L.; MULLER-KARGER, F.; HARDING, L.; MAGNUSON, A.; PHINNEY, D.; MOORE, G. F.; AIKEN, J.;

- ARRIGO, K. R.; LETELIER, R.; CULVER, M. **NASA Technical Memorandum 2000-206892**: Ocean color chlorophyll a algorithms for SeaWiFS, OC2 and OC4: version 4. Maryland: NASA, 2000. v. 11, 9–23 p. 6
- OSAWA, T.; ZHAO, C. F.; K., N. I.; SWARDIKA, I. K.; SUGIMORU, Y. Vertical distribution of chlorophyll-a based on neural network. **International Journal of Remote Sensing and Earth Sciences**, v. 2, p. 1–11, 2005. 29
- PARKER, D. B. Optimal algorithms for adaptive networks: second order Backpropagation, second order direct propagation and second order Hebbian learning. In: IEEE FIRST INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, San Diego, CA, USA. **Proceedings** Piscataway, NJ, USA: IEEE Press, 1987. II, p. 593–600. 34
- PLATT, T.; SATHYENDRANATH, S. Oceanic primary production: estimation by remote sensing at local and regional scales. **Science**, v. 241, n. 4873, p. 1613–1620, 1988. xv, 48, 49
- PLATT, T.; SATHYENDRANATH, S.; WHITE, G. N.; RAVINDRAN, P. Attenuation of visible light by phytoplankton in a vertically structured ocean: solutions and applications. **Journal of Plankton Research**, v. 16, n. 11, p. 1461–1487, 1994. 7, 9, 10, 48
- PREISENDORFER, R. W. Generalized invariant imbedding relation. **Proceedings of the National Academy of Science**, v. 47, n. 4, p. 591–594, 1961. 14
- _____. **Hydrologic optics**. California: National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Pacific Marine Environmental Laboratory, 1976. 3
- PRIEUR, L.; SATHYENDRANATH, S. An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials. **Limnology and Oceanography**, v. 26, n. 4, p. 671–689, 1981. 47
- RETAMOSO, M. R.; VILHENA, M. T.; CAMPOS VELHO, H. F.; RAMOS, F. M. Estimation of boundary condition in hydrologic optics. **Applied Numerical Mathematics**, v. 40, n. 1-2, p. 87–100, 2002. 2, 5
- RICHARDSON, A. J.; PFAFF, M. C.; FIELD, J. G.; SILULWANE, N. F.; SHILLINGTON, F. A. Identifying characteristic chlorophyll a profiles in the

- coastal domain using an artificial neural network. **Journal of Plankton Research**, v. 24, n. 12, p. 1289–1303, 2002. 7, 29
- ROBITAILLE, B.; MARCOS, B.; VEILLETTE, M.; PAYRE, G. Modified quasi-newton methods for training neural networks. **Computers & Chemical Engineering**, v. 20, n. 9, p. 1133–1140, 1996. ISSN 00981354. 34
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958. 21, 24, 27
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986. 22, 33
- RYAN, J. P.; YODER, J. A.; TOWNSEND, D. W. Influence of a Gulf Stream warm-core ring on water mass and chlorophyll distributions along the southern flank of Georges Bank. **Deep-sea Research Part II: topical studies in Oceanography**, v. 48, p. 159–178, 2001. 7
- SACCO, W. F.; LAPA, C. M. F.; PEREIRA, C. M. N. A.; ALVES FILHO, H. A. A metropolis Algorithm applied to a Nuclear Power Plant auxiliary feedwater system surveillance tests policy optimization. **Progress in Nuclear Energy**, v. 50, n. 1, p. 15–21, 2008. ISSN 0149–1970. 44
- SACCO, W. F.; OLIVEIRA, C. R. E. A new stochastic optimization algorithm based in a Particle Collision Metaheuristic. In: WORLD CONGRESS OF STRUCTURAL AND MULTIDISCIPLINARY OPTIMIZATION, 6., 2005, Rio de Janeiro. **Proceedings** Rio de Janeiro, 2005. 44
- SACCO, W. F.; OLIVEIRA, C. R. E.; PEREIRA, C. M. N. A. Two stochastic optimization algorithms applied to nuclear reactor core design. **Progress in Nuclear Energy**, v. 48, n. 6, p. 525–539, 2006. ISSN 0149-1970. 44
- SAMBATTI, S. B. M.; ANOCHI, J. A.; LUZ, E. F. P.; SHIGUEMORI, E. H.; CARVALHO, A. R.; CAMPOS VELHO, H. F. Automatic configuration for neural network applied to atmospheric temperature profile identification. In: INTERNATIONAL CONFERENCE ON ENGINEERING OPTIMIZATION, 2012, Rio de Janeiro. **Proceedings** Rio de Janeiro, 2012. p. 1–9. 29
- SANSONE, G. **Orthogonal functions**. New York: Interscience Publishers, 1959. 142, 151

SATHYENDRANATH, S.; COTA, G.; STUART, V.; MAASS, H.; PLATT, T. Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches. **International Journal of Remote Sensing**, v. 22, n. 2-3, p. 249–273, 2001. 3

SATHYENDRANATH, S.; MOREL, A. Light emerging from the sea – interpretation and uses in remote sensing. In: CRACKNELL, A. P. (Ed.). **Remote Sensing applications in marine science and technology**. França: Springer Netherlands, 1983, (NATO ASI Series, v. 106). p. 323–357. ISBN 978-94-009-7165-3. 47

SCHILLER, H.; DOERFFER, R. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. **International Journal of Remote Sensing**, v. 20, n. 9, p. 1735–1746, 1999. 7, 29

SEGATTO, C. F.; VILHENA, M. T. Extension of the LTS_N formulation for discrete ordinates problem without azimuthal symmetry. **Annals of Nuclear Energy**, v. 21, n. 11, p. 701–710, 1994. 14

SETIONO, R.; HUI, L. C. K. Use of a quasi-Newton method in a feedforward neural network construction algorithm. **IEEE Transactions on Neural Networks**, v. 6, n. 1, p. 273–277, jan 1995. ISSN 1045-9227. 34

SIEGEL, D. A.; DONEY, S. C.; YODER, J. A. The spring bloom of phytoplankton in the North Atlantic ocean and Sverdrup's critical depth hypothesis. **Science**, v. 296, p. 730–733, 2002. 8

SIEWERT, C. E. The F_N method for solving radiative transfer problems in plane geometry. **Astrophysics and Space Science**, v. 58, p. 131–137, 1978. 14

SIEWERT, C. E.; BENOIST, P. The F_N method in neutron-transport theory, Part I: theory and applications. **Nuclear Science and Engineering**, v. 69, p. 156–160, 1979. 14

SILA NETO, A. J.; CAMPOS VELHO, H. F. Problemas inversos em transferência radiativa – uma formulação implícita. In: SILVA NETO, A. J.; BECCENERI, J. C. (Ed.). **Técnicas de inteligência computacional inspiradas na natureza: aplicações em problemas inversos em transferência radiativa**. São carlos: SBMAC, 2009. p. 25–34. Volume 41. 43

SILVA NETO, A. J.; MCCORMICK, N. J. An explicit formulation based on the moments of the exit radiation intensity for the one-dimensional inverse radiative transfer problem. In: INTERNATIONAL CONFERENCE ON INVERSE PROBLEMS IN ENGINEERING, 4th., 2002, Angra dos Reis. **Proceedings** Rio de Janeiro, 2002. p. 347–354. 41

SILVA NETO, A. J.; MOURA NETO, F. D. **Problemas inversos**: conceitos fundamentais e aplicações. Rio de Janeiro: EdUERJ, 2005. 41

SOEIRO, F. J. C. P.; OLIVA SOARES, P.; CAMPOS VELHO, H. F.; SILVA NETO, A. J. Using neural networks to obtain initial estimates for the solution of inverse heat transfer problems. In: INVERSE PROBLEMS, DESIGN AND OPTIMIZATION SYMPOSIUM, 2004, Rio de Janeiro , RJ. **Proceedings** Rio de Janeiro, 2004. Proc. in CD-Rom: paper code IPDO-083 (6-pages). Acesso em: 13 maio 2011. 29, 46

SOEIRO, F. J. C. P.; OLIVA SOARES, P.; SILVA NETO, A. J. Solution of inverse radiative transfer problems with artificial neural networks and hybrid method. In: INVERSE PROBLEMS IN ENGINEERING SEMINAR, 13., Cincinnati, USA. **Proceedings** Cincinnati, 2004. 29, 46

SOEIRO, F. J. C. P.; SILVA NETO, A. J. Solution of inverse radiative transfer problems in two-layer media with artificial neural networks. In: INVERSE PROBLEMS IN ENGINEERING SEMINAR, 14., Ames, USA. **Proceedings** Ames, 2006. 29, 46

SOUTO, R. P. **Recuperação de perfis verticais de propriedade óticas inerentes a partir da radiação emergente da água**. 139 p. (INPE-14195-TDI/1097). Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2006. Disponível em: <<http://urlib.net/sid.inpe.br/MTC-m13@80/2006/06.12.14.30>>. Acesso em: 13 maio 2011. 2, 5, 8, 9, 10, 12, 112, 113, 114

SOUTO, R. P.; BARBOSA, V. C. F.; CAMPOS VELHO, H. F.; STEPHANY, S. Determining chlorophyll concentration in off-shore sea water from multi-spectral radiances by using second derivative criterion and Ant Colony Meta-heuristic. In: INVERSE PROBLEMS, DESIGN AND OPTIMIZATION SYMPOSIUM, 2007, Miami, USA. **Proceedings** Miami, 2007. v. 1, p. 341–348. xviii, 2, 5, 8, 9, 12, 112, 113, 114

SOUTO, R. P.; CAMPOS VELHO, H. F.; PAES, F. F.; STEPHANY, S.; NAVAUX, P. O. A.; CHARAO, A. S.; VIZZOTTO, J. K. Grid computing for multi-spectral tomographic reconstruction of chlorophyll concentration in ocean water. In: CONSTANDA, C.; PÉREZ, M. E. (Ed.). **Integral Methods in Science and Engineering**. Boston: Birkhäuser Boston, 2010. v. 2, p. 327–337. ISBN 978-0-8176-4896-1. 8

SOUTO, R. P.; CAMPOS VELHO, H. F.; STEPHANY, S.; KAMPEL, M. Chlorophyll concentration profiles from in situ radiances by Ant Colony optimization. **Journal of Physics: Conference Series**, v. 124, n. 1, 2008. 2, 5, 8

SOUTO, R. P.; DIAS, P. L. S.; CAMPOS VELHO, H. F.; STEPHANY, S.; KAMPEL, M. Reconstruction of chlorophyll-a vertical profiles with enhanced resolution. In: WORLD CONGRESS ON COMPUTATIONAL MECHANICS, 10., 2012, São Paulo. **Proceedings** São Paulo, 2012. 5, 8

SPANIER, J.; GELBARD, E. M. **Monte Carlo principles and neutron transport problems**. New York: Dover Publications, 2008. 15

STAMNES, K.; SWANSON, R. A. A new look at the discrete ordinate method for radiative transfer calculations in anisotropically scattering atmosphere. **Journal of Atmospheric Sciences**, v. 38, p. 387–399, 1981. 14

STAMNES, K.; TSAY, S. C.; WISCOMBE, W.; JAYAWEERA, K. Numerically stable algorithm for discrete ordinate method radiative transfer in multiple scattering and emitting layered media. **Applied Optics**, v. 7, p. 2502–2509, 1988. 14

STEPHANY, S. **Reconstrução de propriedades óticas e de fontes de bioluminescência em águas naturais**. 122 p. Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998. 2, 5, 16

STEPHANY, S.; CAMPOS VELHO, H. F.; RAMOS, F. M.; MOBLEY, C. D. Identification of inherent optical properties and bioluminescence source term in a hydrologic optics problem. **Journal of Quantitative Spectroscopy & Radiative Transfer**, v. 67, n. 2, p. 113–123, 2000. 2, 5

STEPHANY, S.; RAMOS, F. M.; CAMPOS VELHO, H. F.; MOBLEY, C. D. A methodology for internal light sources estimation. **Computer Modeling and Simulation in Engineering**, v. 3, n. 3, p. 161–165, 2000. 2, 5

- STRAMSKI, D.; KIEFER, D. Light scattering by microorganisms in the open ocean. **Progress in Oceanography**, v. 28, p. 343–383, 1991. 48
- TALLANEARE, T. SuperSAB: fast adaptive Backpropagation with good scaling properties. **Neural Networks**, v. 3, p. 561–573, 1990. 33
- TANG, C. Z.; KWAN, H. K. Parameter effects on convergence speed and generalization capability of Backpropagation algorithm. **International Journal of Electronics**, v. 74, n. 1, p. 35–46, 1993. 33, 34
- TIKHONOV, A.; ARSEININ, V. I. A. **Solutions of ill-posed problems**. New York: Winston, 1977. (Scripta series in mathematics). ISBN 9780470991244. 41
- UITZ, J.; CLAUSTRE, H.; MOREL, A.; HOOKER, S. B. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. **Journal of Geophysical Research**, v. 111, n. C8005, p. 1–23, 2006. 10
- ULLOA, O.; SATHYENDRANATH, S.; PLATT, T.; NONES, R. A. Q. Light scattering by marine heterotrophic bacteria. **Journal of Geophysical Research: Oceans**, v. 97, n. C6, p. 9619–9629, 1992. ISSN 2156–2202. 48
- VALÉRIO, L.; KAMPEL, M.; CORTIVO, F. D.; ORTIZ, G. P. Análise temporal da concentração de clorofila-a da superfície do mar na região oceânica Sudeste Brasileira. In: CONGRESSO BRASILEIRO DE OCEANOGRAFIA, V., 2012, Rio de Janeiro (RJ). **Anais** [S.l.], 2012. p. 1321–1325. 8
- VILHENA, M. T.; BARICHELLO, L. B. A new analytical approach to solve the neutron transport equation. **Kerntechnik**, v. 56, n. 5, p. 334–336, 1991. 14
- VILHENA, M. T. M. B.; STRECK, E. An approximated analytical solution for the one-group slab-geometry neutron transport equation. **Kerntechnik**, v. 57, n. 3, p. 196–198, 1992. 14
- WATROUS, R. L. **Learning algorithms for connectionist networks: applied gradient methods of nonlinear optimization**. Philadelphia, PA: [s.n.], 1988. 34
- WICK, G. C. Über ebene diffusionsprobleme. **Zeitschrift für Physik**, v. 121, p. 702–705, 1943. 14
- WIDROW, B. Generalization and information storage in networks of adaline neurons. **Self-Organizing Systems**, Spartan Books, p. 435–461, 1962. 21

- WIDROW, B.; HOFF, M. E. Adaptive switching circuits. **IRE Wescon Convention Record**, p. 96–104, 1960. Part 4. 21, 27
- WILAMOWSKI, B. M.; IPLIKCI, S.; KAYNAK, O.; EFE, M. O. An algorithm for fast convergence in training neural networks. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 2001, Washington. **Proceedings** New York: IEEE Press, 2001. v. 3, p. 1778–1782. 34
- WILAMOWSKI, B. M.; YU, H. Improved computation for Levenberg–Marquardt training. **IEEE Transactions on Neural Networks**, v. 21, n. 6, p. 930–937, 2010. ISSN 1045-9227. 34
- WILLSHAW, D. J.; VON DER MALSBURG, C. How patterned neural connexions can be set up by self-organization. **Proceedings Royal Society B**, n. 194, p. 431–445, 1976. 22
- WXWIDGETS. **wxWidgets Cross-Plataform GUI Library**. 2013. Disponível em: <<http://www.wxwidgets.org/>>. Acesso em: 16 fev. 2013. 71
- YODER, J. A.; ESAIAS, W. E.; FELDMAN, G. C.; MCCLAIN, C. R. Satellite ocean color-status report. **Oceanography**, v. 1, n. 1, p. 18–20, 1988. 4
- YODER, J. A.; MCCLAIN, C. R.; FELDMAN, G. C.; ESAIAS, W. E. Annual cycles of phytoplankton chlorophyll concentrations in the global ocean: a satellite view. **Global Biogeochemical Cycles**, v. 7, n. 1, p. 181–193, 1993. ISSN 1944-9224. 4

APÊNDICE A - APROXIMAÇÃO DA EQUAÇÃO DE TRANSFERÊNCIA RADIATIVA E O MÉTODO AS_N

Neste apêndice, são expostos os detalhes dos passos que devem ser seguidos a fim de transformar a ETR em um sistema de equações diferenciais ordinárias (EDO) de primeira ordem a coeficientes constantes, bem como os passos para se obter a solução desse sistema. Os cálculos expostos nesta seção seguem os trabalhos de [Chalhoub \(1997\)](#), [Chalhoub e Garcia \(2000\)](#), [Chalhoub \(2005\)](#) e [Cortivo \(2008\)](#).

A.1 Aproximação da Equação de Transferência Radiativa

Como foi comentado no final da Seção 2.1, a busca da solução para a ETR é iniciada separando a intensidade de radiação na componente espalhada $I_{s,r,g}(\tau, \mu, \varphi)$ e na componente não-espalhada $I_{u,r,g}(\tau, \mu, \varphi)$. A solução final para a ETR é a soma dessas duas componentes. Para se obter a solução para a componente não-espalhada $I_{u,r,g}(\tau, \mu, \varphi)$, é considerado que $b_{r,g} = 0$ na Eq. (2.4), assim

$$\frac{\partial}{\partial \tau} I_{u,r,g}(\tau, \mu, \varphi) + \frac{1}{\mu} I_{u,r,g}(\tau, \mu, \varphi) = 0,$$

cujas solução é dada por

$$I_{u,r,g}(\tau, \mu, \varphi) = K e^{-\tau/\mu}. \quad (\text{A.1})$$

O valor da constante K é determinado ao se utilizar as condições de contorno expressas pela Eq. (2.7). Aplicando as condições dadas pela Eq. (2.7) na Eq. (A.1) obtém-se

$$I_{u,r,g}(\tau, \mu, \varphi) = \begin{cases} I_{1,g}(\tau_0, \mu, \varphi) e^{-\tau/\mu}, & \text{para } \mu > 0 \\ 0, & \text{para } \mu < 0 \end{cases}, \quad (\text{A.2})$$

em que $\tau \in [0, \zeta]$, $\varphi \in [0, 2\pi]$ e as condições de interface são dadas pela Eq. (2.6). Para a componente espalhada $I_{s,r,g}(\tau, \mu, \varphi)$ é aplicada a igualdade dada pela Eq. (2.8) nas Eqs. (2.4) e (2.7), e levado em conta a solução expressa pela Eq. (A.2). Desse modo, tem-se

$$\begin{aligned} \mu \frac{\partial}{\partial \tau} I_{s,r,g}(\tau, \mu, \varphi) + I_{s,r,g}(\tau, \mu, \varphi) &= \frac{b_{r,g}}{c_{r,g}} \int_{-1}^1 \int_0^{2\pi} p_r(\cos \Theta) I_{s,r,g}(\tau, \mu', \varphi') d\varphi' d\mu' \\ &+ S_{r,g}(\tau, \mu, \varphi), \quad (\text{A.3}) \end{aligned}$$

para $\tau \in [0, \zeta]$, $\mu \in [-1, 1]$ e $\varphi \in [0, 2\pi]$, com as condições de contorno dadas por

$$I_{s,1,g}(\tau_0, \mu, \varphi) = 0 \quad \text{e} \quad I_{s,R,g}(\tau_R, -\mu, \varphi) = 0,$$

e condições de interface definidas pela Eq. (2.6). O termo não homogêneo $S_{r,g}(\tau, \mu, \varphi)$ é expresso por

$$S_{r,g}(\tau, \mu, \varphi) = S_{0,g}(\tau) + \frac{b_{r,g}}{c_{r,g}} \int_{-1}^1 \int_0^{2\pi} p_r(\cos \Theta) I_{u,r,g}(\tau, \mu', \varphi') d\varphi' d\mu'. \quad (\text{A.4})$$

Para esta tese, como comentado na Seção 2.1, não são consideradas fontes internas de radiação, portanto, $S_{0,g}(\tau) = 0$ na Eq. (A.4). A fim de simplificar a notação, o termo $b_{r,g}/c_{r,g}$ será representado por $\varpi_{r,g}$ e chamado de albedo de espalhamento simples.

A solução para a Eq. (A.3) é obtida ao se utilizar a decomposição de Chandrasekhar (CHANDRASEKHAR, 1950). Essa decomposição consiste em aproximar a intensidade de radiação $I_{s,r,g}(\tau, \mu, \varphi)$ em sua série de Fourier em cossenos truncada, em relação a um ângulo de referência φ_0 . Dessa forma, tem-se que $I_{s,r,g}(\tau, \mu, \varphi)$ é expressa por

$$I_{s,r,g}(\tau, \mu, \varphi) = \frac{1}{2} \sum_{m=0}^L (2 - \delta_{0,m}) I_{r,g}^m(\tau, \mu) \cos[m(\varphi - \varphi_0)], \quad (\text{A.5})$$

em que $\delta_{0,m}$ é a função delta de Kronecker. A partir do teorema da adição dos harmônicos esféricos (SANSONE, 1959), é possível expressar a função de fase (genérica), $p_r(\cos \Theta)$, em sua série de polinômios de Legendre (CHANDRASEKHAR, 1950). Portanto,

$$p_r(\cos \Theta) = \frac{1}{4\pi} \sum_{m=0}^L (2 - \delta_{0,m}) \sum_{l=m}^L \beta_{l,r}^m P_l^m(\mu') P_l^m(\mu) \cos[m(\varphi' - \varphi)], \quad (\text{A.6})$$

em que

$$\beta_{l,r}^m = \frac{(l-m)!}{(l+m)!} \beta_{l,r},$$

sendo $\beta_{l,r}$ coeficientes a serem determinados, baseados na função de fase (específica) utilizada, e $P_l^m(\cdot)$ representam as funções de Legendre associadas. Para esta tese, é utilizada a função de fase (específica) de Henyey-Greenstein (HENYEY; GREENSTEIN, 1941) (veja Apêndice B para detalhes). Substituindo as Eqs. (A.6) e (A.5) na Eq. (A.3), e utilizando as propriedades de ortogonalidade das funções trigonométricas, é possível reduzir o problema original em uma série de $L + 1$ problemas desacoplados para os coeficientes $I_{r,g}^m(\tau, \mu)$ da Eq. (A.5). Dessa forma, a Eq. (A.3)

em termos desses novos coeficientes passa a ser escrita como

$$\mu \frac{\partial}{\partial \tau} I_{r,g}^m(\tau, \mu) + I_{r,g}^m(\tau, \mu) = \frac{\varpi_{r,g}}{2} \sum_{l=m}^L \beta_{l,r}^m P_l^m(\mu) \int_{-1}^1 P_l^m(\mu') I_{r,g}^m(\tau, \mu') d\mu' + S_{r,g}^m(\tau, \mu), \quad (\text{A.7})$$

para $m = 0, 1, \dots, L$, sujeita às condições de contorno

$$I_{1,g}^m(\tau_0, \mu) = 0 \quad \text{e} \quad I_{R,g}^m(\tau_R, -\mu) = 0, \quad (\text{A.8})$$

e condições de interface, para $r = 1, 2, \dots, R - 1$,

$$I_{r,g}^m(\tau_r, \pm\mu) = I_{r+1,g}^m(\tau_r, \pm\mu), \quad (\text{A.9})$$

com $\mu \in (0, 1]$. O termo não homogêneo é expresso por

$$S_{r,g}^m(\tau, \mu) = \frac{\varpi_{r,g}}{2} \frac{I_{0,g}}{\pi} \sum_{l=m}^L \beta_{l,r}^m P_l^m(\mu_0) P_l^m(\mu) e^{-\tau/\mu_0}. \quad (\text{A.10})$$

Uma vez resolvidos os problemas definidos pelas Eqs. (A.7), (A.8) e (A.9), para $m = 0, 1, \dots, L$, é possível calcular a componente espalhada através da Eq. (A.5). De posse das soluções espalhada, Eq. (A.5), e não-espalhada, Eq. (A.2), é possível obter a radiação $I_{r,g}(\tau, \mu, \varphi)$, Eq. (2.8), emitida na superfície para as direções $\mu < 0$ e para as direções $\mu > 0$ no fundo.

A.2 Método das Ordenadas Discretas Analítico – AS_N

A diferença entre os métodos S_N (CHANDRASEKHAR, 1950) e AS_N (CHALHOUB, 1997) reside no fato de que o primeiro resolve o conjunto de equações diferenciais, resultante da aproximação da ETR, por diferenças finitas, e o segundo, por um método analítico.

Em ambos os métodos o processo de solução se inicia introduzindo uma quadratura de Gauss Legendre de ordem N para aproximar o termo integral presente na Eq. (A.7). Assim,

$$\int_{-1}^1 P_l^m(\mu') I_{r,g}^m(\tau, \mu') d\mu' \approx \sum_{j=1}^N P_l^m(\mu_j) I_{r,g}^m(\tau, \mu_j) \eta_j,$$

em que $\{\mu_j\}$ e $\{\eta_j\}$ são, respectivamente, os nós (raízes) e os pesos de quadratura. No entanto, diferentemente do método S_N , proposto por Wick-Chandrasekhar, o

método AS_N , implementado em PEESNA (CHALHOUB et al., 2003), aplica uma quadratura dupla de ordem $N = 2n$, sendo n para cada semi-intervalo $[-1, 0]$ e $[0, 1]$. Tomando o conjunto $\mu = \mu_i, i = 1, 2, \dots, N$, na Eq. (A.7), as equações em ordenadas discretas para $r = 1, 2, \dots, R$ são escritas como

$$\mu_i \frac{d}{d\tau} I_{r,g}^m(\tau, \mu_i) + I_{r,g}^m(\tau, \mu_i) = \frac{\varpi_{r,g}}{2} \sum_{l=m}^L \beta_{l,r}^m P_l^m(\mu_i) \sum_{j=1}^N \eta_j P_l^m(\mu_j) I_{r,g}^m(\tau, \mu_j) + S_{r,g}^m(\tau, \mu_i), \quad (\text{A.11})$$

e o termo não-homogêneo dado pela Eq. (A.10) é escrito como

$$S_{r,g}^m(\tau, \mu_i) = \frac{\varpi_{r,g}}{2} \frac{I_{0,g}}{\pi} \sum_{l=m}^L \beta_{l,r}^m P_l^m(\mu_0) P_l^m(\mu_i) e^{-\tau/\mu_0}.$$

Para completar a solução, é necessário redefinir as condições de contorno e as condições de interface. Para isso, considera-se que as raízes de quadratura são ordenadas de forma decrescente, *i.e.*, as primeiras n raízes são positivas e as $N - n$ últimas são negativas. Além disso, considerando que $\mu_{n+i} = -\mu_i$ e $\eta_{n+i} = \eta_i$, para $i = 1, 2, \dots, n$, é possível escrever as condições de contorno expressas pela Eq. (A.8) como

$$I_{1,g}^m(\tau_0, \mu_i) = 0 \quad \text{e} \quad I_{R,g}^m(\tau_R, -\mu_i) = 0, \quad (\text{A.12})$$

e as condições de interface expressas pela Eq. (A.9), para $r = 1, 2, \dots, R - 1$ e $i = 1, 2, \dots, n$, como

$$I_{r,g}^m(\tau_r, \mu_i) = I_{r+1,g}^m(\tau_r, \mu_i) \quad \text{e} \quad I_{r,g}^m(\tau_r, -\mu_i) = I_{r+1,g}^m(\tau_r, -\mu_i). \quad (\text{A.13})$$

Para escrever a Eq. (A.11) de forma matricial, considere que $\vec{I}_{r,g}^m(\tau) = [I_{r,g}^m(\tau, \mu_1), I_{r,g}^m(\tau, \mu_2), \dots, I_{r,g}^m(\tau, \mu_N)]^T$ em que as n primeiras entradas do vetor $\vec{I}_{r,g}^m(\tau)$ estão associadas às raízes $\mu > 0$, e as $N - n$ últimas estão associadas às raízes $\mu < 0$. Portanto, é possível escrever a Eq. (A.11) como

$$\frac{d}{d\tau} \vec{I}_{r,g}^m(\tau) = \mathbb{A}_{r,g}^m \vec{I}_{r,g}^m(\tau) + \vec{S}_{r,g}^m(\tau), \quad (\text{A.14})$$

em que para cada região r , para cada banda de energia g e para cada m , as entradas

da matriz $\mathbb{A}_{r,g}^m$ são expressas por

$$[a_{r,g}^m]_{i,j} = \begin{cases} -\frac{1}{\mu_i} + \frac{\overline{\omega}_{r,g}}{2\mu_i} \sum_{l=m}^L \beta_{l,r}^m P_l(\mu_i) P_l^m(\mu_j) \eta_j, & \text{se } i = j \\ \frac{\overline{\omega}_{r,g}}{2\mu_i} \sum_{l=m}^L \beta_{l,r}^m P_l(\mu_i) P_l^m(\mu_j) \eta_j, & \text{se } i \neq j \end{cases},$$

e as entradas do vetor $\vec{S}_{r,g}^m(\tau)$ para cada região r , para cada banda de energia g e para cada m , são expressas por

$$[s_{r,g}^m]_i(\tau) = \frac{\overline{\omega}_{r,g} I_{0,g}}{2\mu_i \pi} \sum_{l=m}^L \beta_{l,r}^m P_l^m(\mu_0) P_l^m(\mu_i) e^{-\tau/\mu_0}. \quad (\text{A.15})$$

O sistema dado pela Eq. (A.14) pode ser resolvido utilizando a decomposição espectral da matriz $\mathbb{A}_{r,g}^m$, desde que essa matriz tenha um conjunto completo de autovalores linearmente independentes (BOYCE; DIPRIMA, 2002). Aplicando a decomposição espectral na matriz $\mathbb{A}_{r,g}^m$, é possível escrevê-la como o produto de três matrizes: a matriz de autovetores, a matriz diagonal de autovalores e a matriz inversa dos autovetores, *i.e.*,

$$\mathbb{A}_{r,g}^m = \mathbb{P}_{r,g}^m \mathbb{D}_{r,g}^m (\mathbb{P}_{r,g}^m)^{-1}. \quad (\text{A.16})$$

A partir da Eq. (A.16) e da transformação $\vec{I}_{r,g}^m(\tau) = \mathbb{P}_{r,g}^m \vec{Y}_{r,g}^m(\tau)$, o sistema dado pela Eq. (A.14) é reescrito como

$$\frac{d}{d\tau} \vec{Y}_{r,g}^m(\tau) = \mathbb{D}_{r,g}^m \vec{Y}_{r,g}^m(\tau) + \vec{Q}_{r,g}^m(\tau), \quad (\text{A.17})$$

em que $\vec{Q}_{r,g}^m(\tau) = (\mathbb{P}_{r,g}^m)^{-1} \vec{S}_{r,g}^m(\tau)$. A solução para o sistema expresso pela Eq. (A.17) pode ser obtida por meio do método do fator integrante, pois as equações são independentes entre si, e assim

$$\vec{Y}_{r,g}^m(\tau) = e^{\mathbb{D}_{r,g}^m \tau} \vec{K}_{r,g}^m + \int_0^\tau e^{\mathbb{D}_{r,g}^m(\tau-\nu)} \vec{Q}_{r,g}^m(\nu) d\nu. \quad (\text{A.18})$$

Para se obter a solução na variável $\vec{I}_{r,g}^m(\tau)$ basta utilizar, novamente, a transformação $\vec{I}_{r,g}^m(\tau) = \mathbb{P}_{r,g}^m \vec{Y}_{r,g}^m(\tau)$ e, após, determinar os valores do vetor de constantes $\vec{K}_{r,g}^m$ utilizando as condições de contorno. No entanto, em virtude das características dos autovalores da matriz $\mathbb{A}_{r,g}^m$ apresentarem uma ordem de grandeza elevada, à medida que a ordem de quadratura aumenta é interessante reescrever a solução dada pela Eq. (A.18) de uma forma ligeiramente diferente. Esse procedimento é importante

pois ameniza/elimina o problema de *overflow* causado pelos autovalores de ordem de grandeza elevada. Outra característica dos autovalores da matriz $\mathbb{A}_{r,g}^m$ é que estes são simétricos em relação à origem, *i.e.*, n autovalores são negativos e os outros $N - n$ são positivos. Diante disso, admita que os autovalores estejam organizados em ordem decrescente, *i.e.*,

$$[d_{r,g}^m]_1 < [d_{r,g}^m]_2 < \dots < [d_{r,g}^m]_n < 0 < [d_{r,g}^m]_{n+1} < [d_{r,g}^m]_{n+2} < \dots < [d_{r,g}^m]_N \quad (\text{A.19})$$

e que os n primeiros autovalores estão associados às direções $\mu > 0$ e os $N - n$ últimos às direções $\mu < 0$.

Voltando ao processo de transformação, e para facilitar o entendimento, considere que se tenha uma única região espacial como mostrado na Figura A.1. Essa suposição é válida, pois, para cada região tem-se um problema “específico”, assim, ao se obter a solução para um problema de região única, basta generalizar esse resultado para um problema multiregiões, levando em conta as condições de interface. Em função de se considerar apenas uma região, o índice r será omitido (temporariamente). Considere que a Eq. (A.18) deve satisfazer as condições de contorno dadas por

$$Y_g^m(\tau_0, \mu_i) = A_i \quad \text{e} \quad Y_g^m(\zeta, -\mu_i) = B_i, \quad i = 1, 2, \dots, n, \quad (\text{A.20})$$

em que A_i são valores arbitrários e estão associados às direções $\mu > 0$, e B_i são também valores arbitrários e estão associados às direções $\mu < 0$. As condições de interface não são levadas em conta pois se está considerando um problema de região única.



Figura A.1 - Problema com uma única região.

Fonte: Adaptada de Chalhoub (2005).

Reescrevendo as condições dadas pela Eq. (A.20) como $\vec{Y}_g^m(\tau_0) = \vec{A}$ e $\vec{Y}_g^m(\zeta) = \vec{B}$,

e as aplicando na Eq. (A.18), tem-se que

$$\vec{A} = e^{\hat{D}_g^m \tau_0} \check{\mathcal{K}}_g^m + \int_0^{\tau_0} e^{\hat{D}_g^m (\tau_0 - \nu)} \check{\mathcal{Q}}_g^m(\nu) d\nu, \quad (\text{A.21a})$$

e

$$\vec{B} = e^{\hat{D}_g^m \zeta} \hat{\mathcal{K}}_g^m + \int_0^\zeta e^{\hat{D}_g^m (\zeta - \nu)} \hat{\mathcal{Q}}_g^m(\nu) d\nu, \quad (\text{A.21b})$$

em que \check{D}_g^m e \hat{D}_g^m são matrizes diagonais que contêm apenas os autovalores negativos e positivos, respectivamente, $\check{\mathcal{K}}_g^m$ e $\hat{\mathcal{K}}_g^m$ são vetores que contêm as n e $N - n$ entradas do vetor \vec{K}_g^m , respectivamente, e $\check{\mathcal{Q}}_g^m$ e $\hat{\mathcal{Q}}_g^m$ são vetores que contêm as n e $N - n$ entradas do vetor \vec{Q}_g^m , respectivamente. Ao resolver o conjunto de Eqs. (A.21) para $\check{\mathcal{K}}_g^m$ e $\hat{\mathcal{K}}_g^m$ obtém-se, respectivamente,

$$\check{\mathcal{K}}_g^m = e^{-\check{D}_g^m \tau_0} \vec{A} + \int_{\tau_0}^0 e^{-\check{D}_g^m \nu} \check{\mathcal{Q}}_g^m(\nu) d\nu, \quad (\text{A.22a})$$

e

$$\hat{\mathcal{K}}_g^m = e^{-\hat{D}_g^m \zeta} \vec{B} + \int_\zeta^0 e^{-\hat{D}_g^m \nu} \hat{\mathcal{Q}}_g^m(\nu) d\nu. \quad (\text{A.22b})$$

Separando a Eq. (A.18) na parte que contém os autovalores negativos $\check{Y}_g^m(\tau)$ e positivos $\hat{Y}_g^m(\tau)$, e aplicando as Eqs. (A.22) em cada parte, tem-se que

$$\check{Y}_g^m(\tau) = e^{\check{D}_g^m (\tau - \tau_0)} \vec{A} + \int_{\tau_0}^0 e^{\check{D}_g^m (\tau - \nu)} \check{\mathcal{Q}}_g^m(\nu) d\nu + \int_0^\tau e^{\check{D}_g^m (\tau - \nu)} \check{\mathcal{Q}}_g^m(\nu) d\nu, \quad (\text{A.23a})$$

e

$$\hat{Y}_g^m(\tau) = e^{\hat{D}_g^m (\tau - \zeta)} \vec{B} + \int_\zeta^0 e^{\hat{D}_g^m (\tau - \nu)} \hat{\mathcal{Q}}_g^m(\nu) d\nu + \int_0^\tau e^{\hat{D}_g^m (\tau - \nu)} \hat{\mathcal{Q}}_g^m(\nu) d\nu. \quad (\text{A.23b})$$

Definição A.1. Se F é uma função integrável nos intervalos fechados $[a, b]$, $[a, c]$ e $[c, b]$, então,

$$\int_a^c F(x) dx + \int_c^b F(x) dx = \int_a^b F(x) dx,$$

em que $a < c < b$.

Aplicando a Definição A.1 nas Eqs. (A.23), tem-se que

$$\check{Y}_g^m(\tau) = e^{\check{D}_g^m (\tau - \tau_0)} \vec{A} + \int_{\tau_0}^\tau e^{\check{D}_g^m (\tau - \nu)} \check{\mathcal{Q}}_g^m(\nu) d\nu, \quad (\text{A.24a})$$

e

$$\hat{Y}_g^m(\tau) = e^{\hat{D}_g^m(\tau-\zeta)} \vec{B} + \int_{\zeta}^{\tau} e^{\hat{D}_g^m(\tau-\nu)} \hat{Q}_g^m(\nu) d\nu. \quad (\text{A.24b})$$

Ao analisar a Eq. (A.24a) nota-se que $(\tau - \tau_0) \geq 0$ pois $\tau_0 \leq \tau$, e $(\tau - \nu) \geq 0$ pois $\nu \leq \tau$ para todo $\nu \in [\tau_0, \tau]$. Portanto, como $\check{D}_g^m < 0$, segue que as exponenciais presentes na Eq. (A.24a) serão decrescentes. Já a partir da Eq. (A.24b), nota-se que $(\tau - \zeta) \leq 0$ pois $\zeta \geq \tau$, e $(\tau - \nu) \leq 0$ pois $\nu \geq \tau$ para todo $\nu \in [\tau, \zeta]$. Portanto, como $\hat{D}_g^m > 0$, segue que as exponenciais da Eq. (A.24b) também serão decrescentes. Por fim, a integração é feita da direita para a esquerda na equação que contém os autovalores positivos, enquanto a integração é feita da esquerda para a direita, na equação que contém os autovalores negativos, garantindo, assim, que as exponenciais sejam sempre decrescentes, eliminando o problema de *overflow*. Veja Figura A.1.

Para voltar ao problema multiregiões, basta generalizar as Eqs. (A.24) para as R regiões mostradas na Figura 2.2. Sob essa consideração, as Eqs. (A.24) passam a ser escritas como

$$\check{Y}_{r,g}^m(\tau) = e^{\check{D}_{r,g}^m(\tau-\tau_{r-1})} \vec{A}_{r,g}^m + \int_{\tau_{r-1}}^{\tau} e^{\check{D}_{r,g}^m(\tau-\nu)} \check{Q}_{r,g}^m(\nu) d\nu,$$

e

$$\hat{Y}_{r,g}^m(\tau) = e^{\hat{D}_{r,g}^m(\tau-\tau_r)} \vec{B}_{r,g}^m + \int_{\tau_r}^{\tau} e^{\hat{D}_{r,g}^m(\tau-\nu)} \hat{Q}_{r,g}^m(\nu) d\nu,$$

para $r = 1, 2, \dots, R$. Para $r = 1$ e $r = R$ valem as condições de contorno, e para os demais valores de r valem as condições de interface.

Utilizando novamente a transformação $\vec{I}_{r,g}^m(\tau) = \mathbb{P}_{r,g}^m \vec{Y}_{r,g}^m(\tau)$, é possível escrever a solução do problema como

$$\check{I}_{r,g}^m(\tau) = [\mathbb{P}_{r,g}^m]_{1,1} \check{Y}_{r,g}^m(\tau) + [\mathbb{P}_{r,g}^m]_{1,2} \hat{Y}_{r,g}^m(\tau), \quad (\text{A.26a})$$

e

$$\hat{I}_{r,g}^m(\tau) = [\mathbb{P}_{r,g}^m]_{2,1} \check{Y}_{r,g}^m(\tau) + [\mathbb{P}_{r,g}^m]_{2,2} \hat{Y}_{r,g}^m(\tau), \quad (\text{A.26b})$$

em que $[\mathbb{P}_{r,g}^m]_{i,j}$, com $i, j = 1, 2$, são blocos da matriz de autovetores $\mathbb{P}_{r,g}^m$. Na forma

escalar, as Eqs. (A.26) podem ser escritas como

$$I_{r,g}^m(\tau, \mu_i) = \sum_{k=1}^n \left[[p_{r,g}^m]_{i,k} e^{[d_{r,g}^m]_k(\tau-\tau_{r-1})} [A_{r,g}^m]_k + [p_{r,g}^m]_{i,n+k} e^{[d_{r,g}^m]_{n+k}(\tau-\tau_r)} [B_{r,g}^m]_{n+k} \right] \\ + \sum_{k=1}^n \left[[p_{r,g}^m]_{i,k} [\mathcal{A}_{r,g}^m(\tau, \mu_i)]_{i,k} + [p_{r,g}^m]_{i,n+k} [\mathcal{B}_{r,g}^m(\tau, \mu_i)]_{i,n+k} \right], \quad (\text{A.27})$$

com $i = 1, 2, \dots, N$ e $k = 1, 2, \dots, n$. Ainda, o conjunto de coeficientes $\{[\mathcal{A}_{r,g}^m(\tau, \mu_i)]\}$ e $\{[\mathcal{B}_{r,g}^m(\tau, \mu_i)]\}$, da solução particular, são expressos por

$$[\mathcal{A}_{r,g}^m(\tau, \mu_i)]_{i,k} = \int_{\tau_{r-1}}^{\tau} e^{[d_{r,g}^m]_k(\tau-\nu)} [q_{r,g}^m]_k(\nu, \mu_i) d\nu,$$

e

$$[\mathcal{B}_{r,g}^m(\tau, \mu_i)]_{i,n+k} = \int_{\tau_r}^{\tau} e^{[d_{r,g}^m]_{n+k}(\tau-\nu)} [q_{r,g}^m]_{n+k}(\nu, \mu_i) d\nu,$$

com

$$[q_{r,g}^m]_k(\nu, \mu_i) = \sum_{j=1}^N [\tilde{p}_{r,g}^m]_{k,j} [s_{r,g}^m]_j(\nu, \mu_i),$$

em que $[\tilde{p}_{r,g}^m]_{k,j}$ é cada entrada da matriz inversa de autovetores, $[s_{r,g}^m]_j(\nu, \mu_i)$ é dado pela Eq. (A.15), e os autovalores $[d_{r,g}^m]_k$ estão organizados como apresentado na Eq. (A.19).

Já os coeficientes da solução homogênea, $\{[A_{r,g}^m]\}$ e $\{[B_{r,g}^m]\}$, são soluções do sistema de N equações algébricas, obtidas impondo que a solução geral expressa pela Eq. (A.27) satisfaça as condições de contorno, Eqs. (A.12), e as condições de interface, Eqs. (A.13), para $i = 1, 2, \dots, n$, assim

$$\sum_{k=1}^n \left[[p_{r,g}^m]_{i,k} [A_{r,g}^m]_k + [p_{r,g}^m]_{i,n+k} [B_{r,g}^m]_{n+k} e^{[d_{r,g}^m]_{n+k}(\tau_r-\tau_{r-1})} \right] = \\ I_{r,g}^m(\tau_{r-1}, \mu_i) - \sum_{k=1}^n [p_{r,g}^m]_{i,n+k} [\mathcal{B}_{r,g}^m(\tau_{r-1}, \mu_i)]_{i,n+k}, \quad (\text{A.29a})$$

e

$$\sum_{k=1}^n \left[[p_{r,g}^m]_{i,k} [A_{r,g}^m]_k e^{[d_{r,g}^m]_k(\tau_r-\tau_{r-1})} + [p_{r,g}^m]_{i,n+k} [B_{r,g}^m]_{n+k} \right] = \\ I_{r,g}^m(\tau_r, -\mu_i) - \sum_{k=1}^n [p_{r,g}^m]_{i,k} [\mathcal{A}_{r,g}^m]_k(\tau_r, -\mu_i). \quad (\text{A.29b})$$

Definido o sistema de equações algébricas lineares, Eq. (A.29), este é resolvido para $\{[A_{r,g}^m]\}$ e $\{[B_{r,g}^m]\}$ e, assim tem-se todos os valores necessários para avaliar a Eq. (A.27) para qualquer $\tau \in [0, \zeta]$.

APÊNDICE B - TEOREMA DA ADIÇÃO DOS HARMÔNICOS ESFÉRICOS E A APROXIMAÇÃO DA FUNÇÃO DE FASE

Em Transferência Radiativa, a função de fase $p(\cos \Theta)$, presente na Eq. (2.4), representa a probabilidade de uma partícula com direções $d_1 = (\theta', \varphi')$ ser espalhada em direções $d_2 = (\theta, \varphi)$ após a colisão com alguma outra partícula. A posição da partícula com direções d_1 , em coordenadas esféricas, é dada por $P_1 = (\cos \varphi' \sin \theta', \sin \varphi' \sin \theta', \cos \theta')$; a posição da partícula após ser espalhada, *i.e.*, com direções d_2 , é dada por $P_2 = (\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta)$. O ângulo formado entre os vetores \vec{OP}_1 e \vec{OP}_2 , em que O é a origem do sistema de referência, é representado por Θ . O cosseno de Θ é obtido através do produto escalar entre os vetores \vec{OP}_1 e \vec{OP}_2 , assim

$$\cos \Theta = \frac{\langle \vec{OP}_1, \vec{OP}_2 \rangle}{\|\vec{OP}_1\| \|\vec{OP}_2\|} = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\varphi - \varphi'),$$

em que $\|\vec{OP}_1\| = \|\vec{OP}_2\| = 1$.

B.1 Teorema da Adição dos Harmônicos Esféricos

O teorema da adição dos harmônicos esféricos afirma que, se P_1 e P_2 são dois pontos sobre uma esfera e Θ é o ângulo formado entre os vetores \vec{OP}_1 e \vec{OP}_2 , então

$$P_l(\cos \Theta) = P_l(\mu)P_l(\mu') + 2 \sum_{m=1}^l \frac{(l-m)!}{(l+m)!} P_l^m(\mu)P_l^m(\mu') \cos m(\varphi - \varphi'), \quad (\text{B.1})$$

em que $\mu = \cos \theta$, $\mu' = \cos \theta'$, $P_l(\cdot)$ são os polinômios de Legendre e $P_l^m(\cdot)$ são as funções de Legendre associadas.

Detalhes desse teorema são encontrados em Sansone (1959), e em Cortivo (2008) é encontrada uma descrição detalhada desde a construção dos polinômios de Legendre até a demonstração do teorema.

B.2 Aproximação da Função de Fase de Henyey-Greenstein

Chandrasekhar (1950) afirma que a função de fase $p(\cos \Theta)$ pode ser expandida em uma série de polinômios de Legendre. Segundo o autor, essa aproximação para a função de fase é dada por

$$p(\cos \Theta) = \sum_{l=0}^{\infty} \beta_l P_l(\cos \Theta), \quad (\text{B.2})$$

em que β_l são coeficientes a serem determinados e $P_0(\cos \Theta) = 1$. Para esta tese, é utilizada a função de fase de Henyey-Greenstein (MOBLEY, 1994), expressa por

$$p(\rho, \cos \Theta) = \frac{1}{4\pi} \frac{1 - \rho^2}{\sqrt{(1 - 2\rho \cos \Theta + \rho^2)^3}}, \quad (\text{B.3})$$

em que ρ é um parâmetro que pode ser ajustado para controlar o espalhamento para frente e para trás (MOBLEY, 1994). Para se obter os valores para os coeficientes β_l , a Eq. (B.3) é expandida em polinômios de Legendre.

A fim de facilitar os cálculos da expansão da função de fase em polinômios de Legendre, são adotadas as duas definições a seguir, cujos detalhes são encontrados em Cortivo (2008).

Definição B.1. A função

$$\psi(\rho, x) = \frac{1}{\sqrt{1 - 2\rho x + \rho^2}}, \quad \text{com } 0 < \rho < 1 \quad \text{e} \quad -1 \leq x \leq 1, \quad (\text{B.4})$$

é chamada função geratriz dos polinômios de Legendre.

Definição B.2. A expansão em série de potências, em torno de $\rho = 0$, para a função geratriz dos polinômios de Legendre, resulta em

$$\psi(\rho, x) = 1 + x\rho + \left(-\frac{1}{2} + \frac{3x^2}{2}\right)\rho^2 + \left(-\frac{3x}{2} + \frac{5x^3}{2}\right)\rho^3 + \dots = \sum_{l=0}^{\infty} P_l(x)\rho^l, \quad (\text{B.5})$$

em que $P_l(x)$ são polinômios de l -ésima potência e denominados polinômios de Legendre.

Para o cálculo dos coeficientes β_l , inicialmente a Eq. (B.3) é escrita como

$$\frac{1}{4\pi} \frac{1 - \rho^2}{\sqrt{(1 - 2\rho \cos \Theta + \rho^2)^3}} = \frac{1}{4\pi} \left[\frac{1}{\sqrt{1 - 2\rho \cos \Theta + \rho^2}} + 2\rho \frac{d}{d\rho} \frac{1}{\sqrt{1 - 2\rho \cos \Theta + \rho^2}} \right]$$

e, a partir da Eq. (B.4), em que $x = \cos \Theta$, e da Eq. (B.5), é possível escrever a Eq. (B.3) como

$$\frac{1}{4\pi} \frac{1 - \rho^2}{\sqrt{(1 - 2\rho \cos \Theta + \rho^2)^3}} = \frac{1}{4\pi} \left[\sum_{l=0}^{\infty} \rho^l P_l(\cos \Theta) + 2\rho \frac{d}{d\rho} \sum_{l=0}^{\infty} \rho^l P_l(\cos \Theta) \right].$$

Admitindo que o último termo da expressão acima, à direita do sinal de igualdade,

é uniformemente convergente, é possível derivá-lo, e portanto

$$\frac{d}{d\rho} \sum_{l=0}^{\infty} \rho^l P_l(\cos \Theta) = \sum_{l=1}^{\infty} l \rho^{l-1} P_l(\cos \Theta).$$

Dessa forma,

$$\begin{aligned} \frac{1}{4\pi} \frac{1 - \rho^2}{\sqrt{(1 - 2\rho \cos \Theta + \rho^2)^3}} &= \frac{1}{4\pi} \left[\sum_{l=0}^{\infty} \rho^l P_l(\cos \Theta) + 2\rho \sum_{l=1}^{\infty} l \rho^{l-1} P_l(\cos \Theta) \right] \\ &= \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l + 1) \rho^l P_l(\cos \Theta). \end{aligned}$$

Logo, é possível concluir que os coeficientes β_l da Eq. (B.2) são dados por

$$\beta_l = (2l + 1) \rho^l.$$

Para obter a expressão dada na Eq. (A.6), note que a Eq. (B.1) pode ser escrita como

$$P_l(\cos \Theta) = \sum_{m=0}^{\infty} (2 - \delta_{0,m}) \frac{(l - m)!}{(l + m)!} P_l^m(\mu) P_l^m(\mu') \cos m(\varphi - \varphi'), \quad m = 0, 1, \dots \quad (\text{B.6})$$

Substituindo a Eq. (B.6) na Eq. (B.2), obtém-se

$$\begin{aligned} p(\cos \Theta) &= \frac{1}{4\pi} \sum_{l=0}^{\infty} \beta_l \sum_{m=0}^{\infty} (2 - \delta_{0,m}) \frac{(l - m)!}{(l + m)!} P_l^m(\mu) P_l^m(\mu') \cos m(\varphi - \varphi') \\ &= \frac{1}{4\pi} \sum_{m=0}^{\infty} (2 - \delta_{0,m}) \sum_{l=m}^{\infty} \beta_l^m P_l^m(\mu) P_l^m(\mu') \cos m(\varphi - \varphi'), \end{aligned}$$

em que

$$\beta_l^m = \frac{(l - m)!}{(l + m)!} \beta_l,$$

e $\delta_{0,m}$ é a função delta de Kronecker definida por

$$\delta_{0,m} = \begin{cases} 1, & \text{se } m = 0 \\ 0, & \text{se } m \neq 0 \end{cases}.$$

Truncando as somatórias em L termos, obtém-se a expressão dada pela Eq. (A.6).

APÊNDICE C - O NEURÔNIO BIOLÓGICO

Como comentado no Capítulo 3, os neurônios são as células que mais se diferenciam das demais células do organismo e, ainda, se diferenciam de neurônios das diferentes partes do cérebro. Os sinais chegam ao neurônio através das sinapses localizadas nos dendritos neuronais, além dos que chegam também pelo corpo celular, se propagam através do axônio até encontrar as sinapses com o neurônio seguinte. A sinapse é o ponto de confluência entre um neurônio e os dendritos do neurônio seguinte. Veja Figura 3.1. Uma característica especial da maioria das sinapses é que o sinal normalmente se propaga apenas na direção anterógrada (do axônio de um neurônio precedente para os dendritos localizados em neurônios subsequentes) possibilitando, assim, que o sinal trafegue na direção necessária para executar as funções nervosas requeridas (GUYTON; HALL, 2006).

Algumas sinapses transmitem sinais de um neurônio para outro com facilidade, enquanto outras transmitem sinais, mas com dificuldade. Deve-se considerar, também, os sinais facilitatórios e inibitórios vindos de diferentes áreas do sistema nervoso que podem controlar a transmissão sináptica, algumas vezes abrindo as sinapses para a transmissão e, outros momentos, fechando-as (GUYTON; HALL, 2006).

Também segundo Guyton e Hall (2006), há dois tipos principais de sinapses: a sinapse química e a sinapse elétrica. A grande maioria das sinapses, utilizadas para transmissão do sinal no sistema nervoso central da espécie humana, são sinapses químicas. Nesse tipo de sinapse, o “primeiro neurônio” (neurônio pré-sináptico) secreta, no seu terminal, uma substância química chamada de neurotransmissor, e esse neurotransmissor, por sua vez, irá atuar em proteínas receptoras presentes na membrana do neurônio subsequente (neurônio pós-sináptico), para promover a excitação, inibição, ou ainda, modificar, de outra maneira, a sensibilidade dessa célula. Esse é o princípio da condução unidirecional que ocorre nas sinapses químicas. Por outro lado, as sinapses elétricas, em contraste, são caracterizadas por canais que conduzem eletricidade de uma célula para a seguinte. A maior parte destas sinapses consiste em pequenas estruturas proteicas chamadas de junções comunicantes, que permitem o movimento livre de íons do interior de uma célula para o interior de outra e, ao contrário das sinapses químicas, geralmente transmitem os sinais em ambas as direções.

A Figura C.1 ilustra a estrutura básica de uma sinapse. O terminal pré-sináptico é separado do corpo celular do neurônio pós-sináptico pela fenda sináptica. Esse terminal possui dois tipos de estruturas internas importantes para a função excitadora ou inibitória da sinapse: as vesículas transmissoras e a mitocôndria. As vesículas trans-

missoras contêm a substância transmissora que, quando liberada na fenda sináptica, excita ou inibe o neurônio pós-sináptico. A mitocôndria fornece o trifosfato de adenosina que, por sua vez, supre a energia necessária para sintetizar novas moléculas da substância transmissora (GUYTON; HALL, 2006).

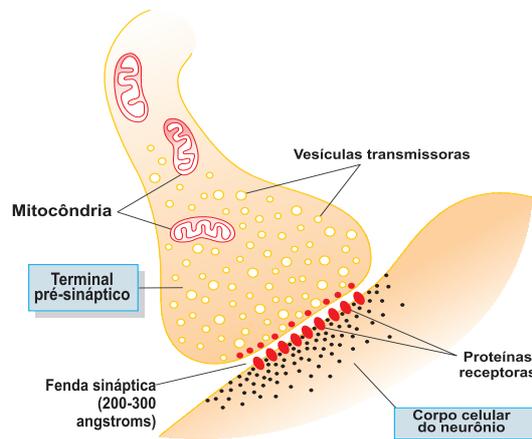


Figura C.1 - Anatomia fisiológica da sinapse.
Fonte: Guyton e Hall (2006).

Quando um potencial de ação chega a um terminal pré-sináptico, a despolarização de sua membrana (pré-sináptica) faz com que um pequeno número de vesículas libere moléculas de neurotransmissores na fenda sináptica. Quando há essa despolarização da membrana, os canais de cálcio se abrem e permitem a passagem de inúmeros íons de cálcio para dentro do terminal pré-sináptico e a quantidade de substância transmissora que é então liberada na fenda sináptica é diretamente proporcional ao número de íons de cálcio que entram (GUYTON; HALL, 2006).

Por outro lado, a membrana do neurônio pós-sináptico contém um grande número de proteínas receptoras, Figura C.1. Segundo Guyton e Hall (2006), as moléculas desses receptores possuem dois componentes importantes: um componente de ligação que se exterioriza, a partir da membrana, na fenda sináptica, e um componente ionóforo, que atravessa toda membrana pós-sináptica até alcançar o interior do neurônio pós-sináptico. A ação da substância transmissora no neurônio pós-sináptico e a função das proteínas receptoras são “eventos” complexos e fogem do escopo desta tese. Veja Guyton e Hall (2006) para detalhes.

Ainda com relação aos processos excitatórios ou inibitórios, Guyton e Hall (2006) dizem que estes são de grande importância, pois permitem uma dimensão adicional à função nervosa, possibilitando a contenção ou a excitação das ações neuronais.

Os mecanismos moleculares e de membrana, utilizados por diversos receptores para induzir excitação ou inibição, incluem:

- na excitação:
 - abertura dos canais de sódio, permitindo o fluxo de um grande número de cargas elétricas positivas para o interior da célula pós-sináptica;
 - condução reduzida através dos canais de cloreto ou potássio, ou de ambos;
 - diversas mudanças no metabolismo interno do neurônio pós-sináptico para excitar a atividade celular ou, em alguns casos, aumentar o número de receptores de membrana excitatórias, ou diminuir o número de receptores de membranas inibitórias.

- na inibição:
 - abertura de canais de íon cloreto na membrana neural pós-sináptica;
 - aumento na condutância de íons potássio para fora dos neurônios;
 - ativação de enzimas receptoras que inibem as funções metabólicas celulares, promovendo o aumento do número de receptores sinápticos inibitórios ou diminuindo o número de receptores excitatórios.

Durante a excitação ou inibição neuronal, ocorrem eventos elétricos que modificam o potencial de repouso da membrana do corpo celular do neurônio. O potencial de repouso da membrana de um neurônio motor espinhal é em torno de 65 milivolts negativos. A diminuição dessa voltagem para um valor menos negativo torna a membrana do neurônio mais excitável, enquanto que o aumento dessa voltagem para um valor mais negativo torna o neurônio menos excitável. Essa é a base para as duas funções (tanto excitação como inibição) do neurônio.

Durante a fase de repouso, o neurônio apresenta um potencial de -65 milivolts; o terminal pré-sináptico não excitado fazendo sinapse sobre sua superfície celular é mostrado na Figura C.2(a). Quando o terminal pré-sináptico secreta um neurotransmissor excitatório na fenda sináptica, Figura C.2(b), entre o terminal e a membrana do corpo celular do neurônio, este neurotransmissor age em um receptor excitatório de membrana, aumentando a permeabilidade da membrana ao Na^+ . O influxo dos íons de sódio carregados positivamente para o interior da célula neutraliza parte da negatividade do potencial de repouso da membrana, assim, o potencial

de repouso da membrana aumenta em direção a um valor mais positivo na voltagem a partir do repouso. Vale lembrar que a descarga de um único terminal pré-sináptico jamais induzirá um aumento do potencial neuronal de -65 milivolts diretamente para -45 milivolts. Para que aconteça um aumento dessa magnitude, é necessário que ocorra uma descarga simultânea a partir de vários terminais ao mesmo tempo e, assim, o corpo celular (veja Figura 3.1) faz a “soma” do íons de sódio elevando o potencial neuronal até os -45 milivolts (GUYTON; HALL, 2006).

Por outro lado, as sinapses inibitórias induzem principalmente a abertura de canais de cloreto e, assim, permitem facilmente que os íons de cloreto, carregados negativamente, se movam do exterior para o interior do neurônio, Figura C.2(c). Dessa forma, o potencial de membrana no interior do neurônio se torna mais negativo que o normal, aproximando-se dos -70 milivolts. A abertura dos canais de potássio também colabora para a diminuição desse potencial, pois os íons de potássio carregados positivamente fluem do interior do neurônio para o exterior. Portanto, tanto o influxo de cloreto quanto o efluxo de potássio aumentam o grau de negatividade intracelular. Nessa explicação, é considerado apenas o efeito das sinapses inibitórias sobre a membrana pós-sináptica, no entanto vale lembrar que pode ocorrer a inibição pré-sináptica também. A inibição pré-sináptica ocorre quando o sinal neural não alcança a sinapse, e é causada pela liberação de uma substância inibitória sobre os terminais nervosos pré-sinápticos (GUYTON; HALL, 2006).

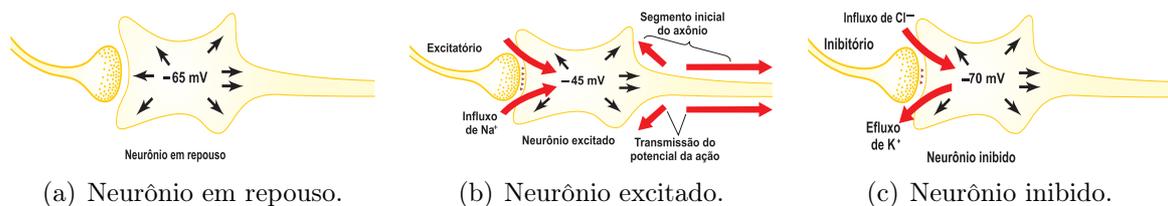


Figura C.2 - Três estados de um neurônio.
Fonte: Guyton e Hall (2006).

Segundo Guyton e Hall (2006), quando uma sinapse excita um neurônio motor, a membrana neuronal se torna altamente permeável aos íons de sódio por um período de 1 a 2 milissegundos. Durante esse intervalo de tempo, uma quantidade suficiente de íons de sódio difunde-se rapidamente para o interior do neurônio motor, aumentando seu potencial em alguns milivolts e, assim, criando um potencial excitatório pós-sináptico, ilustrado na Figura C.3 pelas linhas azul e verde.

O disparo simultâneo de quatro (linha azul) ou de oito (linha verde) sinapses não

foram suficientes para gerar um potencial de ação. Por outro lado, o disparo simultâneo de 16 sinapses elevou o potencial somação até o limiar de excitação e, assim, provocou um potencial de ação. Isso mostra a capacidade de somação do corpo celular à presença de vários disparos sinápticos simultâneos. Observe, também, que após 15 milissegundos do aumento do potencial intraneuronal o neurônio volta ao seu estado de repouso, isso porque esse é o tempo necessário para que o excesso de cargas positivas saiam do neurônio excitado e também para que se restabeleça o potencial de repouso normal da membrana.

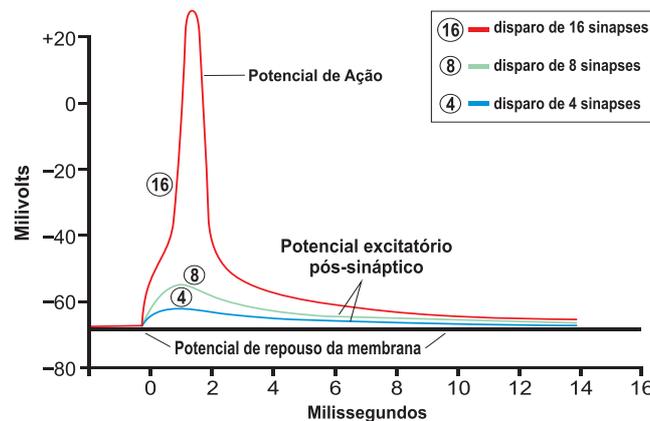


Figura C.3 - Potenciais pós-sinápticos excitatórios.
 Fonte: Guyton e Hall (2006).

Como comentado anteriormente, cada vez que um terminal pré-sináptico dispara, a substância transmissora liberada promove a abertura dos canais de membrana por milissegundos ou pouco mais. No entanto, o potencial pós-sináptico pode durar 15 milissegundos depois de os canais de membrana já terem fechado. Portanto, uma segunda abertura dos mesmos canais pode aumentar o potencial pós-sináptico para um nível ainda maior: quanto mais rápida a velocidade de estimulação, maior se torna o potencial pós-sináptico (GUYTON; HALL, 2006).

O estado excitatório de um neurônio é definido como o impulso excitatório resultante da somação dos potenciais excitatórios e inibitórios de um neurônio. Quando o estado excitatório de um neurônio aumenta acima do limiar de excitação, o neurônio disparará repetitivamente durante o tempo em que o estado excitatório permanecer naquele nível. A Figura C.4 mostra respostas de três tipos de neurônios a vários níveis de estado excitatório. Note também que cada neurônio tem um limiar de disparo diferente e, ainda, cada neurônio tem uma frequência de disparo diferente (GUYTON; HALL, 2006).

Para finalizar, toda a informação que é processada pelo sistema nervoso central, quando relevante, é armazenada para futuro controle das atividades motoras e para uso nos processos cognitivos. O córtex cerebral é a principal região responsável pelo armazenamento dessas informações. Esse armazenamento é chamado de memória, e é, também, executada pelas sinapses. Assim, cada vez que determinados tipos de sinais sensoriais passam através de sequências de sinapses, estas sinapses tornam-se mais capazes de transmitir o mesmo tipo de sinal em outras oportunidades. Esse processo é chamado de facilitação. Segundo [Guyton e Hall \(2006\)](#), pouco se sabe sobre os mecanismos precisos pelos quais a facilitação em longo prazo das sinapses participa no processo de memória. No entanto, uma vez que as informações tenham sido armazenadas no sistema nervoso sob a forma de memória, tornam-se parte do mecanismo de processamento do cérebro para uso futuro sob a forma de “pensamento”.

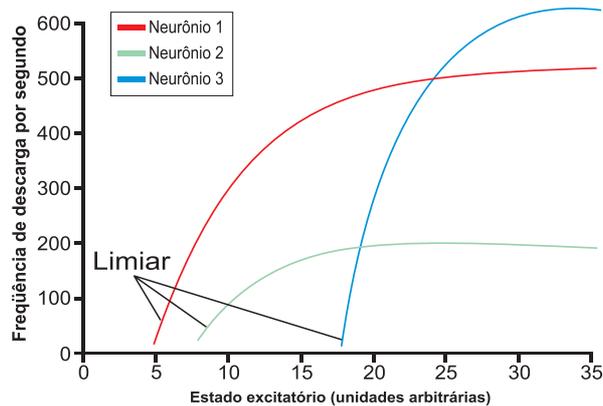


Figura C.4 - Respostas características de diferentes tipos de neurônios a diferentes níveis do estado excitatório.

Fonte: [Guyton e Hall \(2006\)](#).

Como visto neste breve resumo, a atividade neuronal no sistema nervoso central é complexa, e alguns processos (memória) ainda não são bem entendidos pela comunidade científica. Para o leitor interessado nos detalhes omitidos recomenda-se [Guyton e Hall \(2006\)](#).

APÊNDICE D - FORMALISMOS MATEMÁTICOS DOS ALGORITMOS *BACKPROPAGATION*, *BACKPROPAGATION* GENERALIZADO E QUASI-NEWTON

Neste apêndice, são apresentados os formalismos matemáticos envolvidos com os algoritmos citados no título. Alguns dos detalhes matemáticos que serão apresentados aqui podem ser encontrados facilmente nas referências citadas no decorrer desta tese.

D.1 Algoritmo *Backpropagation* Tradicional

No *Backpropagation* “tradicional”, considera-se que o parâmetro de cada função de ativação possa assumir qualquer valor $\gamma \in (0, \infty)$, e que este é constante durante todo o processo de treinamento da RNA. Assim, a função de ativação apresenta apenas a dependência do campo local induzido v , *i.e.*, $y = \phi(v)$. A seguir, mostra-se a dedução do algoritmo *Backpropagation* “tradicional”, que se divide em duas partes: a) uma quando o(s) neurônio(s) está(ão) na camada de saída; b) quando o(s) neurônio(s) está(ão) na(s) camada(s) oculta(s).

D.1.1 O Neurônio está na Camada de Saída

O sinal de erro na saída de um neurônio k na n -ésima iteração, *i.e.*, na apresentação do n -ésimo padrão de treinamento, é dado por

$$e_k^n = d_k^n - y_k^n, \quad (\text{D.1})$$

em que d_k^n representa a resposta desejada no neurônio de saída k , e y_k^n representa a resposta obtida pela rede, por esse mesmo neurônio, para o n -ésimo padrão de treinamento. O erro quadrático médio total cometido por todos os neurônios da camada de saída, na apresentação do n -ésimo padrão de treinamento, é dado por

$$\mathcal{E}^n = \frac{1}{2} \sum_{k \in \Omega} (e_k^n)^2, \quad (\text{D.2})$$

em que Ω representa o conjunto que contém todos os neurônios da camada de saída.

O campo local induzido que chega em um neurônio k na n -ésima iteração é definido por

$$v_k^n = \sum_{j=1}^m w_{jk}^n y_j^n + b_k^n, \quad (\text{D.3})$$

em que m representa o total de entradas aplicadas ao neurônio k , w_{jk}^n é o peso sináptico que liga o neurônio j da camada $\ell - 1$ ao neurônio k da camada de saída ℓ , y_j^n é o sinal de saída do neurônio j da camada $\ell - 1$, e b_k^n é o nível de *bias* do neurônio k da camada ℓ . Portanto, a resposta do neurônio de saída k na n -ésima iteração é expresso por

$$y_k^n = \phi_k(v_k^n). \quad (\text{D.4})$$

As correções que devem ser aplicadas aos pesos sinápticos w_{jk}^n , que ligam a camada oculta $\ell - 1$ à camada de saída ℓ , e aos níveis de *bias* b_k^n dos neurônios da camada de saída devem ser proporcionais às derivadas parciais $\partial \mathcal{E}^n / \partial w_{jk}^n$ e $\partial \mathcal{E}^n / \partial b_k^n$, respectivamente. Portanto, pela regra da cadeia, tem-se que

$$\frac{\partial \mathcal{E}^n}{\partial w_{jk}^n} = \frac{\partial \mathcal{E}^n}{\partial e_k^n} \frac{\partial e_k^n}{\partial y_k^n} \frac{\partial y_k^n}{\partial v_k^n} \frac{\partial v_k^n}{\partial w_{jk}^n} \quad \text{e} \quad \frac{\partial \mathcal{E}^n}{\partial b_k^n} = \frac{\partial \mathcal{E}^n}{\partial e_k^n} \frac{\partial e_k^n}{\partial y_k^n} \frac{\partial y_k^n}{\partial v_k^n} \frac{\partial v_k^n}{\partial b_k^n}. \quad (\text{D.5})$$

As derivadas $\partial \mathcal{E}^n / \partial e_k^n$, $\partial e_k^n / \partial y_k^n$ e $\partial y_k^n / \partial v_k^n$ são obtidas a partir das Eqs. (D.2), (D.1) e (D.4), respectivamente. Portanto,

$$\frac{\partial \mathcal{E}^n}{\partial e_k^n} = e_k^n, \quad \frac{\partial e_k^n}{\partial y_k^n} = -1 \quad \text{e} \quad \frac{\partial y_k^n}{\partial v_k^n} = \frac{\partial \phi_k(v_k^n)}{\partial v_k^n}. \quad (\text{D.6})$$

As derivadas $\partial v_k^n / \partial w_{jk}^n$ e $\partial v_k^n / \partial b_k^n$ são obtidas a partir da Eq. (D.3), logo

$$\frac{\partial v_k^n}{\partial w_{jk}^n} = y_j^n \quad \text{e} \quad \frac{\partial v_k^n}{\partial b_k^n} = 1. \quad (\text{D.7})$$

Substituindo as igualdades dadas pelas Eqs. (D.6) e Eqs. (D.7) nas Eqs. (D.5) tem-se

$$\frac{\partial \mathcal{E}^n}{\partial w_{jk}^n} = -e_k^n \frac{\partial \phi_k(v_k^n)}{\partial v_k^n} y_j^n \quad \text{e} \quad \frac{\partial \mathcal{E}^n}{\partial b_k^n} = -e_k^n \frac{\partial \phi_k(v_k^n)}{\partial v_k^n}. \quad (\text{D.8})$$

As correções que devem ser aplicadas aos pesos sinápticos e aos níveis de *bias* são definidas pela regra delta e são dadas por

$$\Delta w_{jk}^n = -\eta_w \frac{\partial \mathcal{E}^n}{\partial w_{jk}^n} \quad \text{e} \quad \Delta b_k^n = -\eta_b \frac{\partial \mathcal{E}^n}{\partial b_k^n}, \quad (\text{D.9})$$

em que η_w e η_b são as taxas de aprendizagem e podem ser diferentes. O sinal negativo indica a descida do gradiente no espaço das variáveis. Definindo o k -ésimo gradiente

local do k -ésimo neurônio da camada de saída como

$$\delta_k^n = e_k^n \frac{\partial \phi_k(v_k^n)}{\partial v_k^n}, \quad (\text{D.10})$$

e substituindo as Eqs. (D.8) nas Eqs. (D.9), levando em conta a Eq. (D.10), obtém-se

$$\Delta w_{jk}^n = \eta_w \delta_k^n y_j^n \quad \text{e} \quad \Delta b_k^n = \eta_b \delta_k^n. \quad (\text{D.11})$$

Portanto, segue que a atualização que deve ser feita nos pesos sinápticos, que ligam a camada oculta $\ell - 1$ à camada de saída ℓ , e nos níveis de *bias* b_k^n dos neurônios da camada de saída, segundo a regra delta, são dadas por

$$w_{jk}^{n+1} = w_{jk}^n + \Delta w_{jk}^n, \quad (\text{D.12})$$

e

$$b_k^{n+1} = b_k^n + \Delta b_k^n. \quad (\text{D.13})$$

Se a atualização seguir a regra delta generalizada, então as Eqs. (D.12) e (D.13) são expressas por

$$w_{jk}^{n+1} = w_{jk}^n + \alpha_w w_{jk}^{n-1} + \Delta w_{jk}^n,$$

e

$$b_k^{n+1} = b_k^n + \alpha_b b_k^{n-1} + \Delta b_k^n,$$

em que α_w e α_b são chamadas de constantes de momento e podem ser diferentes.

Observação 1. Se a RNA apresentar apenas uma camada de neurônios, *i.e.*, há apenas a camada de saída e a camada de entrada, então os valores y_j^n , mostrados na expressão à esquerda da Eq. (D.11), são substituídos pelas entradas do n -ésimo padrão de treinamento. Nesse caso, o erro e_k^n do k -ésimo neurônio é calculado diretamente pela Eq. (D.1), e o gradiente local desse mesmo neurônio é calculado pela Eq. (D.10) e, então, efetua-se a correção dos pesos sinápticos e dos níveis de *bias* seguindo a regra delta ou a regra delta generalizada.

D.1.2 O Neurônio está em uma Camada Oculta

Quando um neurônio j estiver em alguma camada oculta da rede não há uma resposta específica para esse neurônio. Logo, o sinal de erro desse neurônio oculto deve ser calculado recursivamente, em termos dos sinais de erro de todos os neurônios aos quais esse neurônio oculto estiver diretamente conectado.

Como visto, o erro quadrático médio total cometido pelos neurônios da camada de saída é dado pela Eq. (D.2), e o sinal de erro do k -ésimo neurônio da camada de saída, dado pela Eq. (D.1), pode ser reescrito levando em conta a Eq. (D.4), como

$$e_k^n = d_k^n - \phi_k(v_k^n). \quad (\text{D.14})$$

O campo local induzido do j -ésimo neurônio da camada oculta na n -ésima iteração é definido por

$$v_j^n = \sum_{i=1}^m w_{ij} y_i^n + b_j^n, \quad (\text{D.15})$$

e o sinal de saída desse mesmo neurônio j é dado por

$$y_j^n = \phi_j(v_j^n). \quad (\text{D.16})$$

Se w_{ij}^n é o peso sináptico que liga um neurônio i da camada oculta $\ell - 2$ a um neurônio j da camada oculta $\ell - 1$, e b_j^n é o nível de *bias* do j -ésimo neurônio oculto da camada $\ell - 1$, segue que as atualizações nos pesos sinápticos w_{ij}^n e nos níveis de *bias* b_j^n , na n -ésima iteração, devem ser proporcionais às derivadas parciais $\partial \mathcal{E}^n / \partial w_{ij}^n$ e $\partial \mathcal{E}^n / \partial b_j^n$, respectivamente. Portanto, segue da regra da cadeia que

$$\frac{\partial \mathcal{E}^n}{\partial w_{ij}^n} = \frac{\partial \mathcal{E}^n}{\partial e_k^n} \frac{\partial e_k^n}{\partial v_k^n} \frac{\partial v_k^n}{\partial y_j^n} \frac{\partial y_j^n}{\partial v_j^n} \frac{\partial v_j^n}{\partial w_{ij}^n} \quad \text{e} \quad \frac{\partial \mathcal{E}^n}{\partial b_j^n} = \frac{\partial \mathcal{E}^n}{\partial e_k^n} \frac{\partial e_k^n}{\partial v_k^n} \frac{\partial v_k^n}{\partial y_j^n} \frac{\partial y_j^n}{\partial v_j^n} \frac{\partial v_j^n}{\partial b_j^n}. \quad (\text{D.17})$$

As derivadas $\partial \mathcal{E}^n / \partial e_k^n$, $\partial e_k^n / \partial v_k^n$, $\partial v_k^n / \partial y_j^n$ e $\partial y_j^n / \partial v_j^n$ são obtidas, respectivamente, a partir das Eqs. (D.2), (D.14), (D.3) e (D.16). Assim,

$$\frac{\partial \mathcal{E}^n}{\partial e_k^n} = \sum_k e_k^n, \quad \frac{\partial e_k^n}{\partial v_k^n} = -\frac{\partial \phi_k(v_k^n)}{\partial v_k^n}, \quad \frac{\partial v_k^n}{\partial y_j^n} = w_{jk}^n \quad \text{e} \quad \frac{\partial y_j^n}{\partial v_j^n} = \frac{\partial \phi_j(v_j^n)}{\partial v_j^n}. \quad (\text{D.18})$$

As derivadas $\partial v_j^n / \partial w_{ij}^n$ e $\partial v_j^n / \partial b_j^n$ são obtidas a partir da Eq. (D.15). Logo,

$$\frac{\partial v_j^n}{\partial w_{ij}^n} = y_i^n \quad \text{e} \quad \frac{\partial v_j^n}{\partial b_j^n} = 1. \quad (\text{D.19})$$

Substituindo as igualdades dadas pela Eq. (D.18) e (D.19) nas expressões dadas pela

Eq. (D.17), tem-se que

$$\frac{\partial \mathcal{E}^n}{\partial w_{ij}^n} = - \sum_k e_k^n \frac{\partial \phi_k(v_k^n)}{\partial v_k^n} w_{jk}^n \frac{\partial \phi_j(v_j^n)}{\partial v_j^n} y_i^n \quad \text{e} \quad \frac{\partial \mathcal{E}^n}{\partial b_j^n} = - \sum_k e_k^n \frac{\partial \phi_k(v_k^n)}{\partial v_k^n} w_{jk}^n \frac{\partial \phi_j(v_j^n)}{\partial v_j^n}. \quad (\text{D.20})$$

As correções que devem ser aplicadas aos pesos sinápticos e aos níveis de *bias* são definidas pela Eq. (D.9), assim, efetuando troca dos índices, obtém-se

$$\Delta w_{ij}^n = -\eta_w \frac{\partial \mathcal{E}^n}{\partial w_{ij}^n} \quad \text{e} \quad \Delta b_j^n = -\eta_b \frac{\partial \mathcal{E}^n}{\partial b_j^n}. \quad (\text{D.21})$$

Definindo o j -ésimo gradiente local do j -ésimo neurônio oculto como

$$\delta_j^n = \frac{\partial \phi_j(v_j^n)}{\partial v_j^n} \sum_k \delta_k^n w_{jk}^n, \quad (\text{D.22})$$

em que δ_k^n é expresso pela Eq. (D.10), e substituindo as expressões dadas pela Eq. (D.20) nas expressões dadas pela Eq. (D.21), levando em conta a Eq. (D.22), segue que

$$\Delta w_{ij}^n = \eta_w \delta_j^n y_i^n \quad \text{e} \quad \Delta b_j^n = \eta_b \delta_j^n. \quad (\text{D.23})$$

Portanto, a atualização que deve ser feita nos pesos sinápticos que ligam a camada oculta $\ell - 2$ à camada de oculta $\ell - 1$, e nos níveis de *bias* b_j^n dos neurônios ocultos da camada $\ell - 1$, segundo a regra delta, são dadas por

$$w_{ij}^{n+1} = w_{ij}^n + \Delta w_{ij}^n, \quad (\text{D.24})$$

e

$$b_j^{n+1} = b_j^n + \Delta b_j^n. \quad (\text{D.25})$$

Se a atualização seguir a regra delta generalizada, então as Eqs. (D.24) e (D.25) são expressas por

$$w_{ij}^{n+1} = w_{ij}^n + \alpha_w w_{ij}^{n-1} + \Delta w_{ij}^n,$$

e

$$b_j^{n+1} = b_j^n + \alpha_b b_j^{n-1} + \Delta b_j^n,$$

em que α_w e α_b são chamadas de constantes de momento e podem ser diferentes.

Observação 2. Se a RNA apresentar apenas uma camada oculta, *i.e.*, há a camada

de saída, uma camada oculta e a camada de entrada, então, os valores y_i^n mostrados na expressão à esquerda da Eq. (D.23) são substituídos pelas entradas do n -ésimo padrão de treinamento.

Observação 3. Note que o gradiente δ_j^n de um neurônio oculto na camada oculta $\ell - 1$ é calculado utilizando a derivada da função de ativação em relação ao campo local induzido desse neurônio j , multiplicado pela soma do produto de todos os gradientes dos neurônios da camada de saída pelos respectivos pesos sinápticos. Esse processo “recursivo” é válido para uma rede com qualquer número de camadas ocultas.

D.2 Algoritmo *Backpropagation* Generalizado

A principal diferença entre o *Backpropagation* “tradicional” e o *Backpropagation* “generalizado” reside no fato de que, no primeiro, o parâmetro γ de cada função de ativação é considerado constante, durante o processo de treinamento, e são feitas atualizações apenas nos pesos sinápticos e nos níveis de *bias*. Já no segundo, além das atualizações feitas no primeiro, são feitas atualizações no parâmetro γ de cada função de ativação.

Visto que as correções nos pesos sinápticos e nos níveis de *bias* no *Backpropagation* generalizado são feitas da mesma forma como no *Backpropagation* “tradicional”, esses detalhes serão omitidos, sendo apresentados apenas os detalhes relacionados com a atualização do parâmetro γ .

No *Backpropagation* generalizado considera-se, também, que o parâmetro de cada função de ativação pode assumir qualquer valor $\gamma \in (0, \infty)$, no entanto esse valor é atualizado durante o processo de treinamento da RNA. Assim, a função de ativação apresenta dependência desse parâmetro e do campo local induzido v , *i.e.*, $y = \phi(\gamma, v)$. Como anteriormente, tem-se duas situações: a) o(s) neurônio(s) está(ão) na camada de saída; b) o(s) neurônio(s) está(ão) na(s) camada(s) oculta(s).

D.2.1 O Neurônio está na Camada de Saída

Se o neurônio k está na camada de saída, o sinal de erro na saída desse neurônio na n -ésima iteração do treinamento é dado pela Eq. (D.1), e o erro quadrático médio cometido por todos os neurônios da camada de saída é dado pela Eq. (D.2). O campo local induzido desse neurônio k é dado pela Eq. (D.3), já a resposta do neurônio k na n -ésima iteração é dada por

$$y_k^n = \phi_k(\gamma_k^n, v_k^n). \quad (\text{D.26})$$

As correções que devem ser aplicadas ao parâmetro γ_k^n do k -ésimo neurônio na n -ésima iteração devem ser proporcionais à derivada parcial $\partial\mathcal{E}^n/\partial\gamma_k^n$. Como no caso dos pesos sinápticos e dos níveis de *bias*, essa derivada representa um fator de sensibilidade e determina a direção de busca no espaço dos parâmetros γ_k^n . Portanto, pela regra da cadeia, tem-se que

$$\frac{\partial\mathcal{E}^n}{\partial\gamma_k^n} = \frac{\partial\mathcal{E}^n}{\partial e_k^n} \frac{\partial e_k^n}{\partial y_k^n} \frac{\partial y_k^n}{\partial\gamma_k^n}. \quad (\text{D.27})$$

Note que as derivadas $\partial\mathcal{E}^n/\partial e_k^n$ e $\partial e_k^n/\partial y_k^n$ já foram calculadas e são dadas pelas duas primeiras igualdades apresentadas na Eq. (D.6). A derivada $\partial y_k^n/\partial\gamma_k^n$ é dada por

$$\frac{\partial y_k^n}{\partial\gamma_k^n} = \frac{\partial\phi_k(\gamma_k^n, v_k^n)}{\partial\gamma_k^n}.$$

Portanto, segue que a Eq. (D.27) é escrita como

$$\frac{\partial\mathcal{E}^n}{\partial\gamma_k^n} = -e_k^n \frac{\partial\phi_k(\gamma_k^n, v_k^n)}{\partial\gamma_k^n}. \quad (\text{D.28})$$

As correções que devem ser aplicadas aos parâmetros γ_k^n também são definidas pela regra delta, *i.e.*,

$$\Delta\gamma_k^n = -\eta_\gamma \frac{\partial\mathcal{E}^n}{\partial\gamma_k^n}, \quad (\text{D.29})$$

em que η_γ é a taxa de aprendizagem e o sinal negativo indica a descida do gradiente no espaço dos parâmetros γ_k^n . Definindo o k -ésimo gradiente local, relacionado com o parâmetro γ_k^n , do k -ésimo neurônio de saída como

$$\varrho_k^n = e_k^n \frac{\partial\phi_k(\gamma_k^n, v_k^n)}{\partial\gamma_k^n}, \quad (\text{D.30})$$

segue da substituição da Eq. (D.28) na Eq. (D.29), levando em conta a Eq. (D.30), que

$$\Delta\gamma_k^n = \eta_\gamma \varrho_k^n.$$

Portanto, a atualização nos pesos sinápticos, que ligam a camada $\ell - 1$ à camada ℓ , nos níveis de *bias* e nos parâmetros das funções de ativação da camada de saída, pela regra delta, na n -ésima iteração, são definidas, respectivamente, por

$$w_{jk}^{n+1} = w_{jk}^n + \Delta w_{jk}^n,$$

$$b_k^{n+1} = b_k^n + \Delta b_k^n,$$

e

$$\gamma_k^{n+1} = \gamma_k^n + \Delta\gamma_k^n,$$

em que Δw_{jk}^n e Δb_k^n são definidos pela Eq. (D.11).

Se for utilizada a regra delta generalizada para atualização das variáveis da rede, tem-se que

$$\begin{aligned} w_{jk}^{n+1} &= w_{jk}^n + \alpha_w w_{jk}^{n-1} + \Delta w_{jk}^n, \\ b_k^{n+1} &= b_k^n + \alpha_b b_k^{n-1} + \Delta b_k^n, \end{aligned}$$

e

$$\gamma_k^{n+1} = \gamma_k^n + \alpha_\gamma \gamma_k^{n-1} + \Delta\gamma_k^n,$$

em que α_γ é a constante de momento associada às correções nos parâmetros γ .

Observação 4. Se a RNA apresentar apenas uma camada de neurônios, *i.e.*, há apenas a camada de saída e a camada de entrada, então os valores y_j^n mostrados na expressão à esquerda na Eq. (D.11) são substituídos pelas entradas do n -ésimo padrão de treinamento. Nesse caso, o erro e_k^n do k -ésimo neurônio é calculado diretamente pela Eq. (D.1), e o gradiente local desse mesmo neurônio, com relação aos pesos sinápticos e aos níveis de *bias*, é calculado pela Eq. (D.10). O gradiente local desse neurônio, com relação ao parâmetro de cada função de ativação, é calculado pela Eq. (D.30). Assim, efetua-se a correção dos pesos sinápticos, dos níveis de *bias* e dos parâmetros γ seguindo a regra delta ou a regra delta generalizada.

D.2.2 O Neurônio está em uma Camada Oculta

Quando o neurônio j estiver em uma camada oculta da rede, não há uma resposta específica para esse neurônio. Logo, como visto, o sinal de erro desse neurônio oculto dever ser calculado recursivamente, em termos dos sinais de erro de todos os neurônios aos quais esse neurônio oculto estiver diretamente conectado.

Novamente, o erro quadrático médio total cometido por todos os neurônios da camada de saída é dado pela Eq. (D.2), e a partir da Eq. (D.26), o sinal de erro do k -ésimo neurônio da camada de saída, dado pela Eq. (D.1), pode ser escrito como

$$e_k^n = d_k^n - \phi_k(\gamma_k^n, v_k^n). \quad (\text{D.31})$$

O campo local induzido do j -ésimo neurônio da camada oculta, na n -ésima iteração,

é definido pela Eq. (D.15), enquanto que o sinal de saída do neurônio j é dado por

$$y_j^n = \phi_j(\gamma_j^n, v_j^n). \quad (\text{D.32})$$

Visto que γ_j^n é o parâmetro da j -ésima função de ativação, segue que a atualização sobre esse parâmetro na n -ésima iteração deve ser proporcional à derivada parcial $\partial \mathcal{E}^n / \partial \gamma_j^n$. Assim, pela regra da cadeia, segue que

$$\frac{\partial \mathcal{E}^n}{\partial \gamma_j^n} = \frac{\partial \mathcal{E}^n}{\partial e_k^n} \frac{\partial e_k^n}{\partial v_k^n} \frac{\partial v_k^n}{\partial y_j^n} \frac{\partial y_j^n}{\partial \gamma_j^n}. \quad (\text{D.33})$$

As derivadas $\partial \mathcal{E}^n / \partial e_k^n$ e $\partial v_k^n / \partial y_j^n$ podem ser obtidas a partir das igualdades expressas pela Eq. (D.18), e as derivadas $\partial e_k^n / \partial v_k^n$ e $\partial y_j^n / \partial \gamma_j^n$ são obtidas a partir das Eqs. (D.31) e (D.32), respectivamente, portanto,

$$\frac{\partial e_k^n}{\partial v_k^n} = -\frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} \quad \text{e} \quad \frac{\partial y_j^n}{\partial \gamma_j^n} = \frac{\partial \phi_j(\gamma_j^n, v_j^n)}{\partial \gamma_j^n}.$$

Assim, a Eq. (D.33) é escrita como

$$\frac{\partial \mathcal{E}^n}{\partial \gamma_j^n} = -\sum_k e_k^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} w_{jk}^n \frac{\partial \phi_j(\gamma_j^n, v_j^n)}{\partial \gamma_j^n}. \quad (\text{D.34})$$

As correções que devem ser aplicadas ao parâmetro de cada função de ativação são definidas pela Eq. (D.29) e, assim, efetuando a troca do índice, obtém-se

$$\Delta \gamma_j^n = -\eta_\gamma \frac{\partial \mathcal{E}^n}{\partial \gamma_j^n}. \quad (\text{D.35})$$

Definindo o j -ésimo gradiente local, associado ao parâmetro γ_j^n , do j -ésimo neurônio oculto como

$$\varrho_j^n = \frac{\partial \phi_j(\gamma_j^n, v_j^n)}{\partial \gamma_j^n} \sum_k \delta_k^n w_{jk}^n, \quad (\text{D.36})$$

em que

$$\delta_k^n = e_k^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n},$$

e substituindo a Eq. (D.34) na Eq. (D.35), levando em conta a Eq. (D.36), segue que

$$\Delta \gamma_j^n = \eta_\gamma \varrho_j^n.$$

Portanto, as atualizações que devem ser feitas nos pesos sinápticos, que ligam a camada oculta $\ell - 2$ à camada oculta $\ell - 1$, nos níveis de *bias* b_j^n e nos parâmetros γ_j^n dos neurônios da camada oculta $\ell - 1$, segundo a regra delta, são dadas por

$$w_{ij}^{n+1} = w_{ij}^n + \Delta w_{ij}^n,$$

$$b_j^{n+1} = b_j^n + \Delta b_j^n,$$

e

$$\gamma_j^{n+1} = \gamma_j^n + \Delta \gamma_j^n,$$

em que Δw_{ij}^n e Δb_j^n são dados pela Eq. (D.23).

Se for utilizada a regra delta generalizada para atualização das variáveis da rede, tem-se que

$$w_{ij}^{n+1} = w_{ij}^n + \alpha_w w_{ij}^{n-1} + \Delta w_{ij}^n,$$

$$b_j^{n+1} = b_j^n + \alpha_b b_j^{n-1} + \Delta b_j^n,$$

e

$$\gamma_j^{n+1} = \gamma_j^n + \alpha_\gamma \gamma_j^{n-1} + \Delta \gamma_j^n,$$

em que α_γ é a constante de momento associada às correções nos parâmetros γ .

Observação 5. Se a RNA apresentar apenas uma camada oculta, *i.e.*, há a camada de saída, uma camada oculta e a camada de entrada, então é aplicada a Observação 2.

Observação 6. Note que o gradiente ϱ_j^n de um neurônio oculto na camada oculta $\ell - 1$ é calculado utilizando a derivada da função de ativação em relação ao parâmetro γ_j^n desse neurônio, multiplicado pela soma do produto de todos os gradientes dos neurônios da camada de saída pelos respectivos pesos sinápticos. Esse processo “recursivo” é válido para uma rede com qualquer número de camadas ocultas.

D.3 Cálculo dos Gradientes para o Método quasi-Newton

O termo gradiente, nos algoritmos *Backpropagation*, se refere ao gradiente local de cada neurônio da rede, e a partir desse gradiente, são feitas as correções, proporcionais à taxa (global) de aprendizagem η_* considerada, naquelas variáveis livres da rede que estão diretamente conectadas ao neurônio que gerou tal gradiente (HAYKIN, 2001). Os gradientes calculados para o método quasi-Newton, utilizado neste trabalho, estão relacionados a cada uma das variáveis livres da rede, e não a um determinado neurônio como nos algoritmos *Backpropagation*.

Embora as equações que determinam o $\nabla \mathcal{J}(\cdot, \vec{X})$ são as mesmas que determinam as correções que devem ser aplicadas às variáveis livres da rede pelos algoritmos *Backpropagation*, a diferença entre os métodos está no modo como são feitas as correções nas variáveis livres da rede. Enquanto os algoritmos *Backpropagation* fazem a correção nas variáveis livres da rede se baseando em uma taxa global, a taxa de aprendizagem, o método quasi-Newton calcula essa taxa de correção para cada uma das variáveis livres de modo a buscar que $\nabla \mathcal{J}(\cdot, \vec{X}) \equiv 0$, ou pelo menos, $\nabla \mathcal{J}(\cdot, \vec{X}) \rightarrow 0$. Veja Subseção 3.6.3 do Capítulo 3.

O funcional de diferenças quadráticas foi definido na Subseção 3.6.3, Eq. (3.8), como

$$\mathcal{J}(\mathbb{E}_t, \vec{X}) = \min \frac{1}{2} \sum_{p=1}^{N_p} \sum_{k=1}^{N_s} \left\| d_{kp}(\mathbb{E}_t) - y_{kp}(\mathbb{E}_t, \vec{X}) \right\|_2^2,$$

em que \vec{X} é um vetor que representa todas as variáveis da rede, $d_{kp}(\cdot)$ representa cada uma das saída associadas ao conjunto de treinamento, e $y_{kp}(\cdot, \vec{X})$ representa cada uma das saídas calculadas pela rede. O gradiente da equação anterior foi definido na Seção 3.6.3, Eq. (3.13), como

$$\nabla \mathcal{J}(\cdot, \vec{X}) = - \sum_{p=1}^{N_p} \sum_{k=1}^{N_s} \left(d_{kp}(\cdot) - y_{kp}(\cdot, \vec{X}) \right) \frac{\partial}{\partial \vec{X}} y_{kp}(\cdot, \vec{X}).$$

O cálculo da derivada à direita do sinal de igualdade, da equação acima, depende das funções de ativação utilizadas em cada neurônio e das variáveis livres da rede, e como no caso dos algoritmos *Backpropagation*, há duas situações: a) o(s) neurônio(s) está(ão) na camada de saída; b) o(s) neurônio(s) está(ão) na(s) camada(s) oculta(s).

Observação 7. Os cálculos que serão apresentados a seguir se restringem a uma rede PMC que contém uma camada de entrada, uma única camada oculta e uma camada de saída. O número de neurônios em cada uma dessas camadas é arbitrário. Visto que o treinamento é feito no modo *batch*, a ordem em que as variáveis aparecem na Eq. (D.37) não influencia no resultado final do cálculo dos gradientes.

Como comentado, o vetor \vec{X} é formado pelos pesos sinápticos, níveis de *bias* e, eventualmente, pelos parâmetros das funções de ativação. Portanto, para facilitar o entendimento, admita que o vetor \vec{X} é dado por

$$\vec{X} = [\gamma_k, \gamma_j, w_{jk}, b_k, w_{ij}, b_j]^T, \quad \text{com } k = 1, \dots, N_s, \quad j = 1, \dots, N_o, \quad i = 1, \dots, N_e, \quad (\text{D.37})$$

em que γ_k representa o parâmetro de cada função de ativação do(s) neurônio(s) da camada de saída, γ_j representa o parâmetro de cada função de ativação do(s) neurônio(s) da camada oculta, w_{jk} representa os pesos sinápticos que ligam o(s) neurônio(s) da camada oculta ao(s) neurônio(s) da camada de saída, b_k representa o nível de *bias* de cada neurônio da camada de saída, w_{ij} representa os pesos sinápticos que ligam a camada de entrada ao(s) neurônio(s) da camada oculta, b_j representa o nível de *bias* de cada neurônio da camada oculta, N_s representa o número total de neurônios na camada de saída, N_o representa o número total de neurônios na camada oculta e N_e representa o número total de entradas da rede. Portanto, é necessário calcular cada uma das derivadas parciais relacionadas às variáveis contidas no vetor \vec{X} levando em conta as funções de ativação utilizadas.

D.3.1 O Neurônio está na Camada de Saída

O campo local induzido v_k^n que chega no k -ésimo neurônio da camada de saída, na n -ésima apresentação de um padrão de treinamento, é representado por

$$v_k^n = \sum_{j=1}^{N_o} w_{jk}^n y_j^n + b_k^n, \quad (\text{D.38})$$

em que y_j^n representa o sinal de saída do j -ésimo neurônio da camada oculta, e

$$y_{kp}^n = \phi_k(\gamma_k^n, v_k^n) \quad (\text{D.39})$$

representa o sinal de saída do k -ésimo neurônio da camada de saída.

Portanto, o gradiente dos pesos que ligam a camada oculta à camada de saída, para um determinado padrão de treinamento p , em uma iteração n , é definido por

$$\nabla \mathcal{J}(\cdot, w_{jk}^n) = - (d_{kp}^n(\cdot) - y_{kp}^n(\cdot, w_{jk}^n)) \frac{\partial y_{kp}^n(\cdot, w_{jk}^n)}{\partial w_{jk}^n}.$$

A derivada à direita do sinal de igualdade, pela regra da cadeia e utilizando as Eqs. (D.38) e (D.39), é escrita como

$$\frac{\partial y_{kp}^n(\cdot, w_{jk}^n)}{\partial w_{jk}^n} = \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} \frac{\partial v_k^n}{\partial w_{jk}^n}. \quad (\text{D.40})$$

Visto que $\partial v_k^n / \partial w_{jk}^n = y_j^n$, os gradientes para os pesos sinápticos, que ligam a camada

oculta à camada de saída, são representados por

$$\nabla \mathcal{J}(\cdot, w_{jk}^n) = -e_k^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} y_j^n, \quad (\text{D.41})$$

em que $e_k^n = (d_{kp}^n(\cdot) - y_{kp}^n(\cdot, w_{jk}^n))$.

Para os níveis de *bias*, a Eq. (D.40) é escrita como

$$\frac{\partial y_{kp}^n(\cdot, b_k^n)}{\partial b_k^n} = \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} \frac{\partial v_k^n}{\partial b_k^n},$$

para a qual novamente se utilizou as Eqs. (D.38) e (D.39). Note que $\partial v_k^n / \partial b_k^n = 1$ e, assim, o(s) gradiente(s) para o(s) níveis de *bias* do(s) neurônio(s) da camada de saída é (são) expresso(s) por

$$\nabla \mathcal{J}(\cdot, b_k^n) = -e_k^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n}. \quad (\text{D.42})$$

Os gradientes para cada parâmetro de cada função de ativação são representados por

$$\nabla \mathcal{J}(\cdot, \gamma_k^n) = -e_k^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial \gamma_k^n}. \quad (\text{D.43})$$

Por fim, as Eqs. (D.41), (D.42) e (D.43) fornecem os gradientes para os pesos sinápticos, que ligam a camada oculta à camada de saída, para os níveis de *bias* do(s) neurônio(s) da camada de saída e para cada parâmetro presente na(s) função(ões) de ativação da camada de saída, respectivamente.

D.3.2 O Neurônio está na Camada Oculta

O cálculo dos gradientes para os pesos sinápticos, que ligam os neurônios da camada de entrada à camada oculta, para os níveis de *bias* e parâmetro de cada função de ativação dos neurônios da camada oculta, é um pouco mais complicado, e requer uma interpretação ligeiramente diferente daquela usada nos algoritmos *Backpropagation*.

Como comentado, deve-se tratar a saída da rede $y_{kp}^n(\cdot, \vec{X})$ como uma função que depende das funções de ativação utilizadas e de todas as variáveis da rede. Desse

modo, $y_{kp}^n(\cdot, \vec{X})$ é representado por

$$y_{kp}^n \left(\vec{x}_p, \vec{\gamma}_k^n, \vec{\gamma}_j^n, \vec{w}_{jk}^n, \vec{w}_{ij}^n, \vec{b}_k^n, \vec{b}_j^n \right) = \phi_k \left(\gamma_k^n, \sum_{j=1}^{N_o} \phi_j \left(\gamma_j^n, \sum_{i=1}^{N_e} x_{ip} w_{ij}^n + b_j^n \right) w_{jk}^n + b_k^n \right), \quad (\text{D.44})$$

em que x_{ip} representa a i -ésima entrada do p -ésimo padrão de treinamento, as variáveis associadas ao índice j estão relacionadas aos neurônios da camada oculta, e as variáveis associadas ao índice k estão relacionadas aos neurônios da camada de saída. Note que

$$v_j^n = \sum_{i=1}^{N_e} x_{ip} w_{ij}^n + b_j^n, \quad (\text{D.45})$$

e

$$v_k^n = \sum_{j=1}^{N_o} \phi_j(\gamma_j^n, v_j^n) w_{jk}^n + b_k^n = \sum_{j=1}^{N_o} y_j^n w_{jk}^n + b_k^n. \quad (\text{D.46})$$

Os gradientes para os pesos sinápticos, que ligam a camada de entrada à camada oculta, para um determinado padrão de treinamento p , na n -ésima iteração, são definidos por

$$\nabla \mathcal{J}(\cdot, w_{ij}^n) = - \left(d_{kp}^n(\cdot) - y_{kp}^n(\cdot, w_{ij}^n) \right) \frac{\partial y_{kp}^n(\cdot, w_{ij}^n)}{\partial w_{ij}^n}. \quad (\text{D.47})$$

Visto que a derivada à direita do sinal de igualdade depende de todas as saídas da rede nas quais o neurônio j estiver conectado, essa derivada deve ser calculada k vezes para o mesmo peso w_{ij}^n . Veja Figura D.1. Portanto, para o mesmo peso sináptico w_{ij}^n , a Eq. (D.47) passa a ser escrita como

$$\nabla \mathcal{J}(\cdot, w_{ij}^n) = - \sum_{k=1}^{N_s} e_{kp}^n \frac{\partial y_{kp}^n(\cdot, w_{ij}^n)}{\partial w_{ij}^n}, \quad (\text{D.48})$$

em que $e_{kp}^n = \left(d_{kp}^n(\cdot) - y_{kp}^n(\cdot, w_{ij}^n) \right)$.

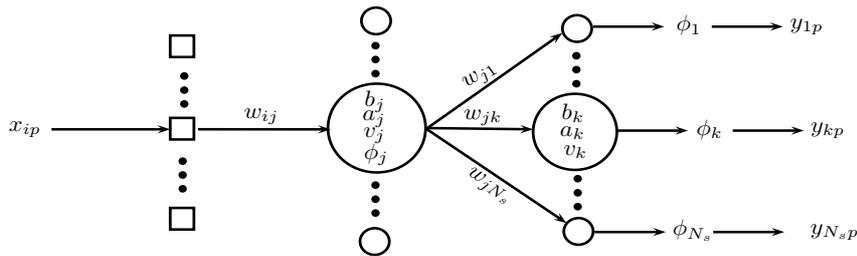


Figura D.1 - Ilustração do cálculo dos gradientes para variáveis da camada oculta.

Para determinar a derivada à direita do sinal de igualdade da Eq. (D.48), aplica-se a regra da cadeia na Eq. (D.44). Dessa forma, após alguns cálculos, obtém-se

$$\frac{\partial y_{kp}^n(\cdot, w_{ij}^n)}{\partial w_{ij}^n} = \frac{\phi_j(\gamma_j^n, v_j^n)}{\partial v_j^n} \frac{\partial v_j^n}{\partial w_{ij}^n} \sum_{k=1}^{N_s} \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} \frac{\partial v_k^n}{\partial y_j^n}. \quad (\text{D.49})$$

Os termos $\partial v_j^n / \partial w_{ij}^n$ e $\partial v_k^n / \partial y_j^n$ podem ser calculados a partir das Eqs. (D.45) e (D.46), respectivamente, assim

$$\frac{\partial v_j^n}{\partial w_{ij}^n} = x_{ip} \quad \text{e} \quad \frac{\partial v_k^n}{\partial y_j^n} = w_{jk}^n. \quad (\text{D.50})$$

Portanto, substituindo as Eqs. (D.50) na Eq. (D.49), e o resultado na Eq. (D.48), obtém-se que o gradiente de um peso w_{ij}^n é representado por

$$\nabla \mathcal{J}(\cdot, w_{ij}^n) = -\frac{\phi_j(\gamma_j^n, v_j^n)}{\partial v_j^n} x_{ip} \sum_{k=1}^{N_s} e_{kp}^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} w_{jk}^n. \quad (\text{D.51})$$

Para os níveis de *bias* dos neurônios da camada oculta, a Eq. (D.48) é reescrita como

$$\nabla \mathcal{J}(\cdot, b_j^n) = -\sum_{k=1}^{N_s} e_{kp}^n \frac{\partial y_{kp}^n(\cdot, b_j^n)}{\partial b_j^n},$$

em que

$$\frac{\partial y_{kp}^n(\cdot, b_j^n)}{\partial b_j^n} = \frac{\phi_j(\gamma_j^n, v_j^n)}{\partial v_j^n} \frac{\partial v_j^n}{\partial b_j^n} \sum_{k=1}^{N_s} \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} \frac{\partial v_k^n}{\partial y_j^n}$$

e, a partir da Eq. (D.45), obtém-se que $\partial v_j^n / \partial b_j^n = 1$. Portanto,

$$\nabla \mathcal{J}(\cdot, b_j^n) = -\frac{\phi_j(\gamma_j^n, v_j^n)}{\partial v_j^n} \sum_{k=1}^{N_s} e_{kp}^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} w_{jk}^n. \quad (\text{D.52})$$

Para os gradientes associados ao parâmetro de cada função de ativação dos neurônios ocultos, a Eq. (D.48) é escrita como

$$\nabla \mathcal{J}(\cdot, \gamma_j^n) = -\sum_{k=1}^{N_s} e_{kp}^n \frac{\partial y_{kp}^n(\cdot, \gamma_j^n)}{\partial \gamma_j^n},$$

em que

$$\frac{\partial y_{kp}^n(\cdot, \gamma_j^n)}{\partial \gamma_j^n} = \frac{\phi_j(\gamma_j^n, v_j^n)}{\partial \gamma_j^n} \sum_{k=1}^{N_s} \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} \frac{\partial v_k^n}{\partial y_j^n},$$

e, portanto, os gradientes para cada parâmetro de cada função de ativação, para os neurônios da camada oculta, são expressos por

$$\nabla \mathcal{J}(\cdot, \gamma_j^n) = -\frac{\phi_j(\gamma_j^n, v_j^n)}{\partial \gamma_j^n} \sum_{k=1}^{N_s} e_{kp}^n \frac{\partial \phi_k(\gamma_k^n, v_k^n)}{\partial v_k^n} w_{jk}^n. \quad (\text{D.53})$$

Por fim, as Eqs. (D.51), (D.52) e (D.53) fornecem os gradientes para os pesos sinápticos que ligam a camada de entrada à camada oculta, para os níveis de *bias* e para cada parâmetro presente na(s) função(ões) de ativação da camada oculta, respectivamente.

D.4 Derivadas das Funções de Ativação Seleccionadas

A seguir, são apresentadas as derivadas de algumas das funções de ativação. As funções consideradas apresentam dependência do parâmetro γ e do campo local induzido v . Dessa forma, são calculadas as derivadas parciais com relação a γ e com relação a v .

D.4.1 Função Sigmoidal

A função sigmoide é definida por

$$\phi(\gamma, v) = \frac{1}{1 + e^{-\gamma v}}, \quad (\text{D.54})$$

em que as derivadas parciais com relação a γ e a v são expressas, respectivamente, por

$$\frac{\partial \phi(\gamma, v)}{\partial \gamma} = \frac{v e^{-\gamma v}}{(1 + e^{-\gamma v})^2} \quad \text{e} \quad \frac{\partial \phi(\gamma, v)}{\partial v} = \frac{\gamma e^{-\gamma v}}{(1 + e^{-\gamma v})^2}.$$

A fim de simplificar essas expressões, é importante notar que

$$1 - \phi(\gamma, v) = \frac{e^{-\gamma v}}{1 + e^{-\gamma v}},$$

em que $\phi(\gamma, v)$ é dada pela Eq. (D.54). Portanto, a derivada parcial da função sigmoide com relação a γ é escrita como

$$\frac{\partial \phi(\gamma, v)}{\partial \gamma} = \frac{v e^{-\gamma v}}{(1 + e^{-\gamma v})^2} = v \phi(\gamma, v) [1 - \phi(\gamma, v)].$$

Similarmente, a derivada parcial da função sigmoide com relação a v é escrita como

$$\frac{\partial\phi(\gamma, v)}{\partial v} = \frac{\gamma e^{-\gamma v}}{(1 + e^{-\gamma v})^2} = \gamma\phi(\gamma, v) [1 - \phi(\gamma, v)].$$

D.4.2 Função Tangente Hiperbólica

A função tangente hiperbólica é definida por

$$\phi(\gamma, v) = \frac{1 - e^{-\gamma v}}{1 + e^{-\gamma v}}, \quad (\text{D.55})$$

em que as derivadas parciais com relação a γ e a v são expressas, respectivamente, por

$$\frac{\partial\phi(\gamma, v)}{\partial\gamma} = \frac{2ve^{-\gamma v}}{(1 + e^{-\gamma v})^2} \quad \text{e} \quad \frac{\partial\phi(\gamma, v)}{\partial v} = \frac{2\gamma e^{-\gamma v}}{(1 + e^{-\gamma v})^2}.$$

Novamente, é importante simplificar essas expressões. Assim,

$$1 - \phi(\gamma, v) = \frac{2e^{-\gamma v}}{1 + e^{-\gamma v}} \quad \text{e que} \quad \frac{1}{2} + \frac{\phi(\gamma, v)}{2} = \frac{1}{1 + e^{-\gamma v}},$$

com $\phi(\gamma, v)$ dada pela Eq. (D.55). Portanto, a derivada da função tangente hiperbólica com relação a γ é escrita como

$$\frac{\partial\phi(\gamma, v)}{\partial\gamma} = \frac{2ve^{-\gamma v}}{(1 + e^{-\gamma v})^2} = \frac{v}{2} [1 - (\phi(\gamma, v))^2].$$

Similarmente, a derivada parcial da função tangente hiperbólica com relação a v é escrita como

$$\frac{\partial\phi(\gamma, v)}{\partial v} = \frac{2\gamma e^{-\gamma v}}{(1 + e^{-\gamma v})^2} = \frac{\gamma}{2} [1 - (\phi(\gamma, v))^2].$$

D.4.3 Função Logaritmo Natural

A função logaritmo natural, como função de ativação nesta tese, é definida por

$$\phi(\gamma, v) = \begin{cases} \ln(1 + \gamma v) & \text{se } v \geq 0 \\ -\ln(1 - \gamma v) & \text{se } v < 0 \end{cases}. \quad (\text{D.56})$$

Apesar de a função ser seccionalmente contínua, ela apresenta a vantagem de não ter seus valores funcionais limitados (como no caso das funções sigmoide e tangente hiperbólica), e esse crescimento é logarítmico.

As derivadas parciais da Eq. (D.56) com relação a γ e a v são expressas, respectivamente, por

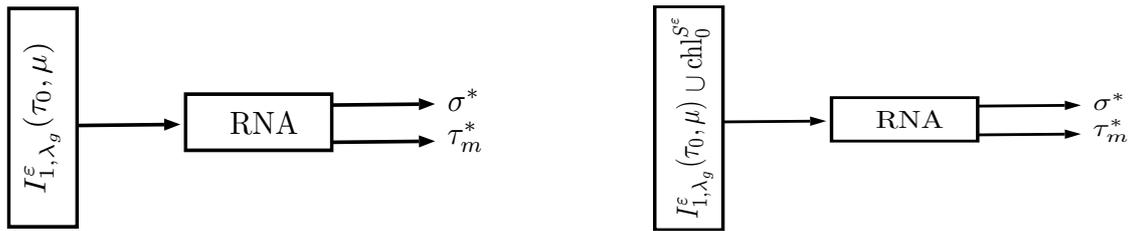
$$\frac{\partial\phi(\gamma, v)}{\partial\gamma} = \begin{cases} \frac{v}{1 + \gamma v} & \text{se } v \geq 0 \\ \frac{v}{1 - \gamma v} & \text{se } v < 0 \end{cases} \quad \text{e} \quad \frac{\partial\phi(\gamma, v)}{\partial v} = \begin{cases} \frac{\gamma}{1 + \gamma v} & \text{se } v \geq 0 \\ \frac{\gamma}{1 - \gamma v} & \text{se } v < 0 \end{cases} .$$

APÊNDICE E - METODOLOGIAS QUE NÃO PRODUZIRAM RESULTADOS SATISFATÓRIOS

Este apêndice apresenta brevemente as três metodologias que foram utilizadas e não produziram resultados satisfatórios.

E.1 Metodologia A: Rede Única

Nessa metodologia, a ideia foi resolver o problema da forma mais simples possível, *i.e.*, uma rede com duas saídas, uma para cada parâmetro. Também foram considerados dois experimentos: um com os padrões de entrada contendo apenas a radiação emergente, conforme características de cada problema, e um segundo experimento, no qual os padrões de entrada foram formados pela radiação emergente, adicionada uma entrada extra para a concentração de clorofila de superfície. A Figura E.1 ilustra as redes utilizadas nesses experimentos. Os padrões de entrada foram corrompidos com ruído.



(a) Rede considerando apenas a radiação emergente como entradas.

(b) Rede considerando a radiação emergente e a concentração de clorofila de superfície como entradas.

Figura E.1 - Redes utilizadas na Metodologia A: Rede Única.

A dificuldade ao resolver o problema dessa forma está associada ao treinamento da rede. Isso ocorre porque os parâmetros a serem estimados são “independentes” entre si, fazendo com que o treinamento da rede convirja para um valor mínimo médio.

E.2 Metodologia B: Comitê de Redes

Nesse experimento, foi treinado um conjunto de redes neurais a fim de construir um comitê de redes especialistas. Cada uma dessas redes especialistas foi treinada para estimar um determinado valor da profundidade do pico da concentração de clorofila τ_m . O conjunto de treinamento para cada rede foi formado pelo total de combinações dos valores de τ_m com os valores do desvio padrão σ . Assim, para todos os padrões do conjunto, a rede deve produzir a mesma resposta, *i.e.*, o mesmo valor de τ_m . Veja

Figura E.2(a). Posteriormente, foi treinada uma rede para estimar os valores de σ que estão associados a esse valor de τ_m . Veja Figura E.2(b). Essas redes treinadas para determinar σ foram incorporadas ao comitê, e atuam em um “segundo estágio”, após as redes treinadas para determinar τ_m .

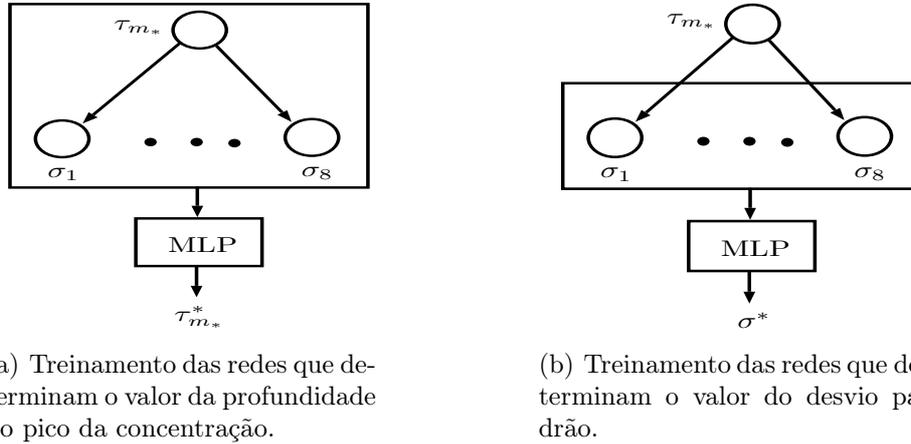


Figura E.2 - Treinamento do comitê utilizado na Metodologia B.

Ao apresentar um padrão de entrada, todas as redes que determinam o valor de τ_m produzem uma resposta. A escolha de qual rede produziu a resposta correta pode ser baseado no seguinte critério: a rede que produzir o menor erro absoluto entre a estimativa feita e sua especialidade é a candidata à resposta correta, e aciona a rede que determina o valor de σ que está associada a ela. Em resumo, o primeiro conjunto de redes determina a profundidade do pico da concentração, e uma segunda rede determina o valor do desvio padrão. Veja Figura E.3. Os padrões de entrada foram corrompidos com ruído e, nessa metodologia, a concentração de clorofila de superfície não foi utilizada como padrão de entrada junto à radiação emergente.

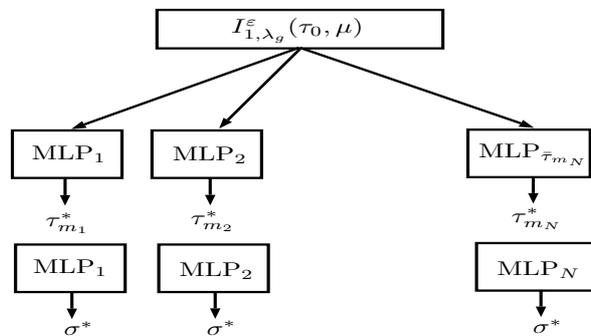


Figura E.3 - Estrutura do comitê utilizado na Metodologia B: Comitê de Redes.

O problema com essa metodologia ocorreu nas redes treinadas para determinar o valor de τ_m , em função das limitações físicas do problema e/ou do modelo matemático adotado, *i.e.*, diferentes valores de τ_m geram valores de radiação emergente muito próximos, o que dificulta a correta classificação dos padrões pelas redes, mesmo sendo especialistas. A Figura E.4 ilustra a radiação emergente obtida para dois valores de τ_m e todos os valores de σ definidos na Eq. (4.7). As curvas pretas estão associadas ao valor $\tau_{m_1} = 1$ m, e as curvas vermelhas estão associadas ao valor $\tau_{m_2} = 2$ m. O fato da distância entre os valores da radiação emergente ser muito pequena (ou não perceptível), inviabiliza a correta classificação por parte das redes. Note que, praticamente não há diferença entre as curvas de radiação emergente geradas a partir de $\tau_{m_1} = 1$ m e $\tau_{m_2} = 2$ m para $\sigma = 17$. Com a adição do ruído, esse quadro se torna mais crítico. Nessa metodologia, também não foi levando em conta a entrada relacionada a chl_0^S nos padrões de entrada.

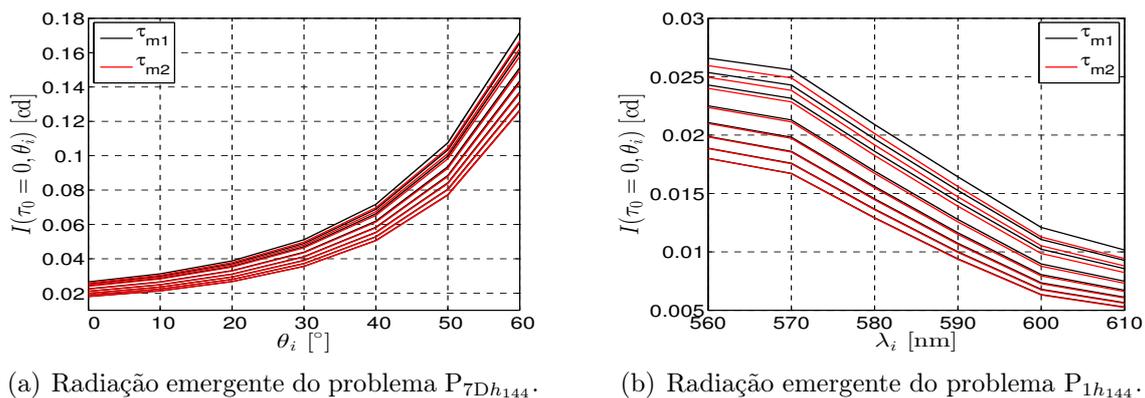


Figura E.4 - Radiação emergente calculada pelo modelo adotado para os problemas $P_{7Dh_{144}}$ e $P_{1h_{144}}$ para $\tau_{m_1} = 1$ m e $\tau_{m_2} = 2$ m.

E.3 Metodologia C: Redes Alternadas

Nesse caso, foram treinadas duas redes independentes, no entanto, após treinadas, trabalham em conjunto de modo alternado. Uma rede foi treinada para estimar o valor de σ , em que os padrões de entrada para essa rede foram a radiação emergente, corrompida com ruído, e o valor de τ_m sem adição de ruído, *i.e.*, valor exato. Uma segunda rede fez a estimativa do valor de τ_m , considerando como padrões de entrada a radiação emergente (corrompida com ruído) e o valor de σ exato associado ao padrão de entrada. Veja Figura E.5.

No processo de validação das redes, em associação à prática, o valor de τ_m não é

conhecido, então esse valor é substituído por um valor aleatório χ , a fim de completar as entradas de cada padrão de treinamento para a rede que determina o valor do desvio padrão. Assim, a primeira rede faz a estimativa do valor de σ^* e, então, esse valor é adicionado ao padrão de entrada, e é processado pela rede que faz a estimativa da profundidade do pico. Em um segundo passo, o valor aleatório χ é substituído pelo valor τ_m^* , determinado pela segunda rede e, então, outro valor de σ^* é encontrado. Esse valor de σ^* é adicionado ao padrão de treinamento (substituindo o valor anterior) e a segunda rede faz uma nova estimativa para o valor de τ_m^* , e assim sucessivamente. O processo iterativo se encerra ao atingir um número máximo de iterações, ou então, quando $|\sigma_k^* - \sigma_{k+1}^*| \leq \delta_1$ ou $|\tau_{m_k}^* - \tau_{m_{k+1}}^*| \leq \delta_2$, ou ainda, $|\sigma_k^* - \sigma_{k+1}^*| + |\tau_{m_k}^* - \tau_{m_{k+1}}^*| \leq \delta_3$, em que δ_* é um valor mínimo predefinido. A Figura E.6 ilustra o processo descrito.

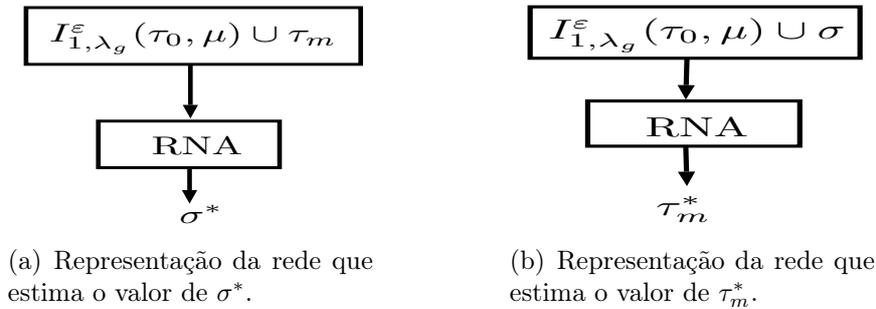


Figura E.5 - Representação das redes utilizadas na Metodologia C: Redes Alternadas.

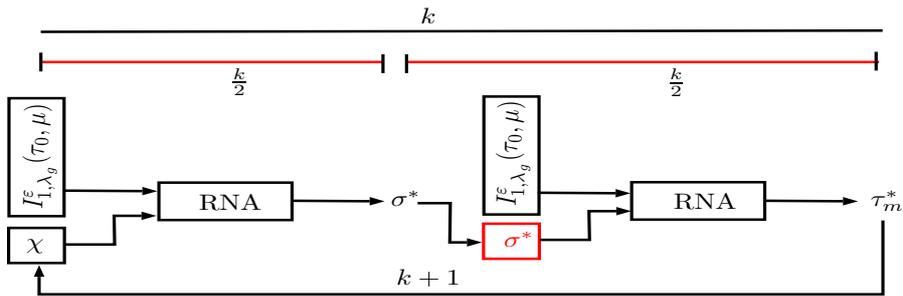


Figura E.6 - Representação do procedimento de validação das redes utilizadas na Metodologia C: Redes Alternadas.

A falha da aplicação dessa metodologia está associada à sobreposição das curvas de radiação emergente para diferentes valores de τ_m . Veja Figura E.4. Essa sobreposição faz com que o processo iterativo não seja convergente, ou então, convirja para respostas erradas, durante o processo de validação, mesmo que cada rede tenha apresentado resultados satisfatórios durante o treinamento.

APÊNDICE F - TRABALHOS PUBLICADOS

F.1 Estimativa do Albedo de Espalhamento Simples Usando uma Rede Neural de Múltiplas Camadas



Estimativa do Albedo de Espalhamento Simples Usando Uma Rede Neural de Múltiplas Camadas*

Fábio Dall Cortivo

Instituto Nacional de Pesquisas Espaciais, INPE,
Programa de Pós-Graduação em Computação Aplicada, CAP,
12227-010, São José dos Campos, SP
E-mail: fabio.cortivo@lac.inpe.br

Ezzat S. Chalhoub, José D. S. da Silva, Haroldo F. de Campos Velho

Instituto Nacional de Pesquisas Espaciais, INPE,
Laboratório Associado de Computação e Matemática Aplicada, LAC,
12227-010, São José dos Campos, SP
E-mail: [ezzat,demisio,haroldo]@lac.inpe.br

Resumo: Identificação de albedo em Óptica Hidrológica através de redes neurais artificiais. O problema direto é modelado pela Equação de Transferência Radiativa (ETR), e a determinação do albedo de espalhamento simples é feita a partir das radiancias medidas. A utilização de duas funções de ativação para o treino da rede, juntamente com o algoritmo de retropropagação do erro, produziu resultados bastante satisfatórios para as estimativas desejadas.

Palavras-chave: Redes Neurais Artificiais, Back-Propagation, Problemas Inversos, Óptica Hidrológica, Equação de Transferência Radiativa

1 Introdução

A solução da ETR pela aplicação do Método de Ordenadas Discretas (S_N) [2] resulta em um sistema de N Equações Diferenciais Ordinárias (EDO) de primeira ordem, onde a matriz associada ao método tem a propriedade (sob certas condições) de possuir todos os autovalores reais, não nulos e simétricos em relação à origem [6]. O sistema de EDO resultante é resolvido com o uso da decomposição espectral da matriz do sistema, que é a base do método AS_N [1].

A estimativa do albedo de espalhamento simples foi efetuada utilizando uma rede neural do tipo Perceptron de Múltiplas Camadas (MLP), com algoritmo de retropropagação do erro como algoritmo de treinamento.

2 Problema Abordado

A ETR é um modelo matemático que descreve a interação do campo de radiação em um ambiente que absorve, emite e espalha essa radiação. A ETR para problemas estacionários, em geometria plano-paralela, em uma determinada faixa de frequência, com dependência azimutal e sem a presença de termo fonte, tem a forma

$$\mu \frac{\partial}{\partial \tau} I(\tau, \vec{\xi}) + I(\tau, \vec{\xi}) = \frac{\varpi_0}{4\pi} \int_{-1}^1 \int_0^{2\pi} \beta(\vec{\xi}; \vec{\xi}') I(\tau, \vec{\xi}') d\vec{\xi}', \quad (1)$$

onde τ representa a espessura ótica, $\vec{\xi} = \vec{\xi}(\mu, \phi)$ representa um vetor unitário com variação polar $\mu = \cos \theta$, ($0 \leq \theta \leq \pi$) e azimutal $0 \leq \phi \leq 2\pi$. $\beta(\vec{\xi}; \vec{\xi}')$ representa a probabilidade de um feixe de radiação incidindo, com intensidade $I(\tau, \vec{\xi})$, na direção $\vec{\xi}'$, sofrer um espalhamento na direção $\vec{\xi} = (\mu', \phi')$, com $d\vec{\xi}' = d\phi' d\mu'$, e ϖ_0 é o albedo de espalhamento simples.

* Ao Curso de Pós-Graduação em Computação Aplicada e a CAPES pelo auxílio financeiro.

F.2 Comparison of Two Learning Strategies for a Supervised Neural Network

ISSN 2238-1007



Proceedings of the 1st International Symposium on Uncertainty Quantification and Stochastic Modeling
February 26th to March 2nd, 2012, Maresias, São Sebastião, SP, Brazil

COMPARISONS OF TWO LEARNING STRATEGIES FOR A SUPERVISED NEURAL NETWORK

Fábio Dall Cortivo¹, Ezzat S. Chalhoub², Haroldo F. de Campos Velho²

¹ Instituto Nacional de Pesquisas Espaciais (INPE), Programa de Pós-Graduação em Computação Aplicada (CAP).
fabio.cortivo@lac.inpe.br

² Instituto Nacional de Pesquisas Espaciais (INPE), Laboratório Associado de Computação e Matemática Aplicada (LAC).
{ezzat.haroldo}@lac.inpe.br

Abstract. Artificial neural networks can be used to solve inverse problems. One relevant problem in hydrologic optics is the estimation of the single scattering albedo from the emitted surface radiation. The multi-layer perceptron (MLP) can be applied to determine the albedo from the measured radiation. The MLP is designed with one hidden layer, where the activation employs the sigmoid function, with backpropagation for set-upping the network parameters. Using the generalized delta rule for the learning process to determine the weight connections, the neural inverse operator (ANN-1) produces good results with 20 inputs (10 incident beams, and 10 emitted beams) and 40 neurons in the hidden layer in two different groups of neurons (30 and 10), with two different parameters for the sigmoid function. The second scheme for training the neural estimator applies the quasi-Newton optimization. For the last strategy, the final neural inverse operator (ANN-2) has 10 inputs (emitted radiation) and 20 neurons in the hidden layer in two different groups of neurons (15 and 5). The measured data were emulated considering five levels of noise. For the generalization test, the ANN-1 and ANN-2 operators obtained 100% of correct answers for the noiseless observational data. For noisy data, the ANN-1 operator obtained 94% of correct answers, while the ANN-2 operator obtained 100% of correct answers. The main difference between these two ANNs is the training method, and the number of neurons in the input and the hidden layer.

Keywords. Inverse Problems, Artificial Neural Networks, Backpropagation, Quasi-Newton Method, Hydrologic Optics

1 INTRODUCTION

Many methodologies of inversion in radiative transfer have been developed in various areas in science and engineering (McCormick, 1992), and a revision of these methods and their applications were performed by McCormick (2001), where the author relates about inverse problems that can explicitly be solved (McCormick, 2004).

A detailed review in inverse Hydrologic Optics can be found in Gordon (2002), and in other works where: Stephany (1998) estimated the bioluminescence internal source using the invariant embedding method (Mobley, 1994); Stephany et al. (2000) estimated the absorption and scattering coefficients also using the invariant embedding method; Chalhoub et al. (2000) estimated the phase function using the S_N method (Chandrasekhar, 1950); Retamóso et al. (2002) estimated the boundary conditions using the $LT S_N$ method (Vilhena & Barichello, 1991); Chalhoub & Campos Velho (2001, 2002, 2003) estimated the phase function, the single scattering albedo, the optical depth and the internal sources, using the AS_N method (Chalhoub, 1997); and finally Souto (2006) estimated the chlorophyll concentration using the $LT S_N$ method.

In the above refereced works, an implicit approach was used to attack the inverse problems, in order to search for the minimal to a functional defined as the quadratic difference sum between the observed data (or synthetic data generated by a computer) and the calculated data using adopted inverse formulations. The search for this minimal was performed by a deterministic method, except for the last work, where the minimal was obtained by the Ant Colony Optimization stochastic technique (Dorigo, 1992).

This type of approach produced good results as noted in the conclusions of the cited references, although, depending on the solved problem, the computational time consumed to find the solution was relatively large due to the extensive and overly refined mathematical formalisms of the direct problem. Furthermore, to obtain the inverse solution, the direct problem must be solved many times, making it necessary to use parallel processing in order to reduce the computational time involved, as emphasized in Souto (2006).

Regarding the consumed computational time, specially for solving problems dealing with the solution of the Radiative Transfer Equation (RTE), a new strategy has lately become dominant: the Artificial Neural Network (ANN) strategy. Due to the fact that the behavior of an ANN is based on the human brain, an ANN strategy presents many advantages over other strategies, such as learning capacity, self-organization, fault tolerance, flexibility and, the most important of all, is its capacity to produce real-time responses. A detailed information of these qualities can be found in Fausett (1994) and Haykin (1999).

In fact, the works (Bokar, 1999; Soeiro et al., 2004; Soeiro & Silva Neto, 2006; Chalhoub et al., 2007; Oliveira, 2010; Cortivo et al., 2010) showed that the utilization of the ANN strategy in solving inverse radiative transfer problems is quite promising. In this case, the ANNs acted as inverse operators, emulating the mathematical formalisms involved in the direct/inverse problems. So, the optimization problem within the inverse problem was exchanged with the network training, which can also be considered as an optimization problem. However, in many cases, the network training can be

F.3 A Committee of MLP with Adaptive Slope Parameter Trained by the quasi-Newton Method to Solve Problems in Hydrologic Optics

WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia

IJCNN

A Committee of MLP with Adaptive Slope Parameter Trained by the quasi-Newton Method to Solve Problems in Hydrologic Optics

Fábio Dall Cortivo
Applied Computing Graduate Program (CAP)
National Institute for Space Research (INPE)
São José dos Campos, São Paulo
Email: fabio.cortivo@lac.inpe.br

Ezzat S. Chalhoub and Haroldo F. Campos Velho
Associated Laboratory for Computing
and Applied Mathematics (LAC)
National Institute for Space Research (INPE)
São José dos Campos, São Paulo
Email: [ezzatz,haroldo]@lac.inpe.br

Abstract—Artificial Neural Networks (ANNs) can be used to solve problems in Hydrologic Optics. A relevant problem is the estimation of the single scattering albedo and the phase function parameters, from the emitted radiation at the surface of natural waters. In this work we use a committee of ANNs of Multilayer Perceptron type to perform the estimation of the two mentioned parameters. The training of each network is formulated as a nonlinear optimization problem subject to constraints. In addition, each activation function has a distinct slope parameter, that is initially chosen by a random number generator function. This set of parameter (slopes) was included within the free variables network set in order to be adjusted to reach “optimal values”, together with the weights and biases, during the network training. This procedure (slope parameters inclusion) makes each one of the activation functions to have a different slope. Each network that composes the committee was trained independently, in order to become expert for the estimation of only one of the hydrologic parameters. For the networks training, we used the quasi-Newton method that is implemented in E04UCF subroutine, in the NAG library, developed by the Numerical Algorithms Group – NAG. The use of the quasi-Newton method to train the networks together with the distinct slope parameters resulted in a network with a fast learning and excellent generalization. Once the networks were trained, they were grouped so to share the input patterns, but remained independent from one another. For the validation/generalization test we used two distinct sets. For all considered noise levels, we obtained 100% of correct answers for the first set, and above 90% of correct answers for the second set.

Index Terms—Artificial Neural Network, Multilayer Perceptron, Backpropagation, quasi-Newton Method, Hydrologic Optics, Phase Function, Single Scattering Albedo, Inverse Problems

I. INTRODUCTION

The popularization of backpropagation training algorithm [1] made Artificial Neural Networks (ANNs) of Multilayer Perceptron (MLP) type to become a computational technique, with a large utilization to approximate highly nonlinear functions, in several areas of science and engineering [2]. To develop a ANN model that produces fast and good solutions, some steps must be taken into account, such as: a) the choice of the training set; b) the accomplishment of the training; c) the choice of the training algorithm. Regarding the training algorithm of MLPs, the most popular methods are the back-

propagation, the Levenberg-Marquardt and the quasi-Newton methods, among others.

In inverse problems solution, the effects (output patterns) and the model/function is known, but the causes (input patterns) are not known. However, due to the ANNs ability to approximate functions, they can be used for the solution of problems where both the causes and the effects are known, but the mathematical model/function that links the causes to the effects, and vice-versa, are not known.

In Hydrologic Optics many inversion techniques were proposed considering the use of a mathematical model that describes the problem. In these methodologies, the algorithm for the resolution is formulated as a nonlinear optimization problem subject to constraints, where the direct model is iteratively solved in successive approximations of the unknown parameters. This process is performed until the objective function, representing the least squares approximation between the model and the experimental data, converges to a minimum. In [3]–[5], details about the resolution of inverse problems in Radiative Transfer and Hydrologic Optics can be found.

To solve problems in Radiative Transfer [6]–[10] and Hydrologic Optics [11], ANNs of MLP type with the backpropagation as the training algorithm were used. In [12] an ANN of MLP type was also used, however the obtained results were better than the ones obtained in [11], due to the use of the quasi-Newton method for the MLP training. In this work, we are proposing additional improvements in the training problem formulation that was adopted in [12].

The problem that is treated in this work consists in solving an inverse problem in Hydrologic Optics, where we perform the joint estimation of the single scattering albedo ϖ and the phase function asymmetry factor g , from the emitted radiation at the surface of natural waters. Due to the inherent difficulty of the considered radiative-transfer problem, we opted for using a committee of ANNs of MLP type, where each one of the networks that composed the committee became a specialist for the estimation of a certain value of g , besides being able to estimate all values of ϖ .

According to [2], the training of ANN can be consid-

F.4 Comparação entre um Algoritmo Genético e uma Rede Neural Artificial na Estimativa de Parâmetros em Oscilações Harmônicas

Comparação entre um Algoritmo Genético e uma Rede Neural Artificial na Estimativa de Parâmetros em Oscilações Harmônicas

Fábio Dall Cortivo, **Leonardo B. L. Santos,**
Instituto Nacional de Pesquisas Espaciais, INPE,
Programa de Pós-Graduação em Computação Aplicada, CAP,
12227-010, São José dos Campos, SP
E-mail: fabio.cortivo@lac.inpe.br, santoslb1@gmail.com.

Resumo: Neste trabalho são utilizados um Algoritmo Genético (AG) orientado a objeto e uma Rede Neural Artificial (RNA) do tipo Perceptron de Múltiplas Camadas, treinada com o algoritmo de retropropagação do erro, para a determinar os coeficientes de amortecimento γ e a rigidez k de um sistema tipo massa-mola-amortecedor. O problema inverso consiste em determinar, a partir da série temporal de deslocamento, os coeficientes γ e k . A comparação entre as técnicas é feita com base nos resultados da estimativa dos parâmetros e custo computacional. Os resultados mostram que no presente estudo o AG apresenta valores de erro médio uma ordem de erro máximo duas ordens de grandeza inferiores aos da RNA, enquanto que a grande vantagem da RNA é sua baixa demanda de tempo de processamento, uma vez treinada.

Palavras-chave: Algoritmos Genéticos, Redes Neurais Artificiais, Oscilações Harmônicas

1 Introdução

Problemas de busca de solução ótima, frequentemente, tratam com conjuntos de cardinalidade muito alta (diversas soluções possíveis), inviabilizando processos exaustivos de procura. Para enfrentar tal questão, um caminho a seguir é a utilização de heurísticas [1]. Essas técnicas não garantem o resultado globalmente ótimo, mas, por usar informações do fenômeno durante a busca, reduzem significativamente o custo computacional, propiciando uma solução razoável a custo reduzido.

As heurísticas utilizadas neste trabalho são baseadas nas técnicas dos Algoritmos Genéticos (AGs) [2] e das Redes Neurais Artificiais (RNAs) [3], ambas advindas da Inteligência Artificial [4].

AGs são heurísticas bio-inspiradas que partem de um conjunto inicial de soluções e via operadores genéticos – seleção, cruzamento e mutação – vão atualizando o conjunto solução até que uma das soluções seja razoável, ou seja, apresente um erro menor que uma tolerância pré-estabelecida – ou um limite de iterações tenha sido alcançado. Essa técnica idealizada pelo americano John Henry Holland, representa uma classe particular de algoritmos evolutivos [5].

RNAs são sistemas computacionais estruturados numa aproximação à computação baseada em ligações. Nós simples, chamados neurônios, são interligados para formar uma rede de nós. Os estudos pioneiros foram feitos pelo neurofisiologista Warren McCulloch, e o matemático Walter Pitts [6]. Nesse trabalho os autores fizeram uma analogia entre células nervosas vivas e o processamento eletrônico. A técnica das RNAs é uma estratégia para solucionar problemas através da simulação do cérebro humano, inclusive em seu comportamento, assim sendo capaz de aprender, errar e fazer descobertas. Isto é, um modelo baseado na estrutura neuronal de organismos inteligentes e que adquirem experiência através do conhecimento.

O objetivo deste trabalho é comparar os resultados obtidos pelas heurísticas (AG e RNA) quando aplicadas na estimativa dos coeficientes de amortecimento e rigidez de um sistema tipo

F.5 Análise Temporal da Concentração de Clorofila-*a* da Superfície do Mar na Região Oceânica Sudeste Brasileira



Congresso Brasileiro de Oceanografia – CBO 2012
13 a 16 de novembro de 2012
Rio de Janeiro – RJ

ANÁLISE TEMPORAL DA CONCENTRAÇÃO DE CLOROFILA-*a* DA SUPERFÍCIE DO MAR NA REGIÃO OCEÂNICA SUDESTE BRASILEIRA

Larissa Patricio Valério; Milton Kampel; Fabio Dall Cortivo; Gustavo Provout Ortiz

larissa@dsr.inpe.br (Instituto Nacional de Pesquisas Espaciais – INPE)

ABSTRACT

In order to analyze the variability of phytoplankton *blooms*, a thirteen-year chlorophyll-*a* time series (1998–2010) was done from GlobColour weekly (8-days) product data set. The maximum chlorophyll-*a* and the increasing biomass beginning year days were computed from the original data and from the original data adjusted with a Gaussian curve model. The time series showed a clear seasonal pattern with increasing amplitude of chlorophyll-*a* value, and a positive trend toward the maximum values were found. The maximum chlorophyll-*a* year day was predominant in July (southern winter), and phytoplankton *blooms* were found to begin in April/May (southern fall) corroborating studies in the North Atlantic and South Pacific. The modeled ones showed the same patterns, indicating a good agreement with the original data set, except the magnitude of the values which were not exactly reproduced by the model, which smoothed the data.

Keywords: timing, spring bloom, time series.

INTRODUÇÃO

O fitoplâncton marinho é responsável pela metade da produtividade primária global, força-motriz da bomba biológica do carbono e da transferência de energia para níveis tróficos superiores nos oceanos (Sommer et al., 2012). A magnitude e o tempo de duração (“*timing*”) da produtividade fitoplanctônica tem impactos significativos no desenvolvimento do zooplâncton e na sobrevivência de larvas de peixes pelágicos (Henson et al., 2009). Seu principal evento, a floração de primavera ou “*spring bloom*” é uma formação anual de sazonalidade marcada, onde as condições fornecidas pelo ambiente regulam o crescimento e a subsequente diminuição da produtividade fitoplanctônica.

A descrição clássica de desenvolvimento inicial do *bloom* foi inserida por Sverdrup (1953), para o Atlântico Norte subpolar. Segundo essa definição, como no inverno predomina uma maior profundidade da camada de mistura e baixos ângulos de incidência solar, o crescimento é limitado. Assim que ocorre o aumento na quantidade de luz incidente e a diminuição da profundidade de mistura a uma “profundidade crítica”, as taxas de crescimento excedem as de perda e o *bloom* se inicia, frequentemente na primavera boreal. Porém nas regiões subtropicais, se houver uma condição persistente ao longo do ano, da camada de mistura rasa e alta intensidade de irradiação solar, o resultado é uma limitação de nutrientes em superfície. Nesse caso, o *bloom* pode se iniciar no outono ou no inverno (Henson et al., 2009).

Como indexador da biomassa fitoplanctônica, um dos produtos do sensoriamento remoto, a concentração de clorofila-*a* tem sido utilizada em estudos acerca do tempo e magnitude dos *blooms* no Atlântico Norte (Siegel et al., 2002; Henson et al., 2009; Kahru et al., 2011) e no Pacífico (Dandonneau et al. 2004). No Atlântico Sul, adjacente à plataforma sudeste brasileira, não existem relatos sobre a duração e magnitude dos *blooms* de maneira temporal. Assim, esse trabalho propõe uma avaliação da variabilidade temporal da concentração de clorofila-*a* superficial, como evidência de ocorrência dos *blooms* fitoplanctônicos no oceano Atlântico Sudoeste.

MATERIAL E MÉTODOS

Foram extraídos e processados dados do projeto Globcolour (<http://www.globcolour.info/>), no setor de 20°S a 30°S e de 39°W a 49°W, adjacente à plataforma sudeste brasileira (Figura). Os produtos resultam da fusão de dados em nível 2 de processamento, obtidos de distintos sensores orbitais independentes que contribuem no fornecimento de parâmetros ambientais. Os sensores

F.6 Estimativa do Perfil da Concentração de Clorofila em Águas Naturais Através de um Perceptron de Múltiplas Camadas

TEMA Tend. Mat. Apl. Comput., **13**, No. 3 (2012), 233-246.

doi: 10.5540/tema.2012.013.03.0233

© Uma Publicação da Sociedade Brasileira de Matemática Aplicada e Computacional.

Estimativa do Perfil da Concentração de Clorofila em Águas Naturais Através de um Perceptron de Múltiplas Camadas¹

F. DALL CORTIVO², Programa de Pós-Graduação em Computação Aplicada, CAP, Instituto Nacional de Pesquisas Espaciais, INPE, Av. dos Astronautas, 1758, 12227-010 São José do Campos, SP, Brasil.

E.S. CHALHOUB, H.F. CAMPOS VELHO, Laboratório Associado de Computação e Matemática Aplicada, LAC, Instituto Nacional de Pesquisas Espaciais, INPE, Av. dos Astronautas, 1758, 12227-010 São José do Campos, SP, Brasil.

Resumo. Estimativa do perfil de concentração de clorofila, em águas naturais, a partir da radiação emergente na superfície de um corpo d'água, com o uso de rede neural artificial do tipo Perceptron de Múltiplas Camadas. A concentração de clorofila está relacionada com os coeficientes de absorção e espalhamento via modelos bio-ópticos. O treinamento da rede é formulado como um problema de otimização, no qual a atualização das variáveis livres da rede (pesos, viés e parâmetros de cada função de ativação) é feita através do método quasi-Newton.

Palavras-chave. Perceptron de múltiplas Camadas, método quasi-Newton, concentração de clorofila, equação de transferência radiativa.

1. Introdução

Neste trabalho a interação de um feixe de fótons com um corpo d'água é representada pela Equação de Transferência Radiativa (ETR). Dadas as condições de contorno, o termo fonte e as propriedades óticas inerentes, é possível resolver a ETR e, assim, determinar a quantidade de partículas (fótons) que estão emergindo na superfície da água após a interação com o meio. Ao abordar o problema dessa forma, caracteriza-se o que é chamado de problema direto. O correspondente problema inverso consiste em determinar uma ou mais propriedades físicas (fontes internas, condições de contorno, propriedades óticas) a partir das medidas radiométricas (radiâncias) emergentes e/ou do interior do corpo d'água. A resolução numérica da ETR é realizada com o código PEESNA [3], o qual implementa o método S_N analítico (AS_N) [1].

¹Agradeço a CAPES pelo auxílio financeiro durante a realização do doutorado.

²E-mail: [fabio.cortivo, ezzat, haroldo]@lac.inpe.br

Recebido em 31 Janeiro 2012; Aceito em 30 Novembro 2012.