TEMA Tend. Mat. Apl. Comput., **13**, No. 3 (2012), 233-246.
doi: 10.5540/tema.2012.013.03.0233

Estimativa do Perfil da Concentração de Clorofila em Águas Naturais Através de um Perceptron de Múltiplas Camadas[1]

F. DALL CORTIVO[2], 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].

[1]Agradeço a CAPES pelo auxilio financeiro durante a realização do doutorado.
[2]E-mail: [fabio.cortivo, ezzat, haroldo]@lac.inpe.br

Recebido em 31 Janeiro 2012; Aceito em 30 Novembro 2012.

Nomenclatura

θ	–	ângulo polar	rad
$I(\tau,\mu,\varphi,\lambda)$	–	radiância	cd
τ	–	variável ótica de profundidade	
μ	–	co-seno do ângulo polar ($\cos\theta$) medido a partir do eixo positivo τ	
φ	–	ângulo azimutal	rad
λ	–	comprimento de onda	nm
ζ	–	profundidade ótica	
Θ	–	ângulo de espalhamento formado pelas direções $\{\mu',\varphi'\}$ e $\{\mu,\varphi\}$	rad
$b(\tau,\lambda)$	–	coeficiente de espalhamento	m^{-1}
$a(\tau,\lambda)$	–	coeficiente de absorção	m^{-1}
$c(\tau,\lambda)$	–	coeficiente de atenuação	m^{-1}
$p(\cos\Theta,\lambda)$	–	função de fase de espalhamento	
$S_0(\tau,\lambda)$	–	fonte interna de radiação	cd
I_0	–	radiância incidente na superfície	cd
μ_0	–	co-seno do ângulo polar de incidência	
φ_0	–	ângulo azimutal de incidência	rad
$\delta(\cdot)$	–	função delta de Dirac	
$I_s(\tau,\mu,\varphi)$	–	componente de radiância espalhada	cd
$I_u(\tau,\mu,\varphi)$	–	componente de radiância não-espalhada	cd
L	–	ordem de espalhamento	
$C(\tau)$	–	concentração de clorofila	$mg \cdot m^{-3}$
a_λ^w	–	coeficiente de absorção da água pura	m^{-1}
a_λ^c	–	coeficiente de absorção específico da $C(\tau)$	
$b(\tau,\lambda)$	–	coeficiente de espalhamento da água pura	m^{-1}
x_j	–	entrada j de um vetor de padrões	
w_{ji}	–	peso sináptico que liga a entrada j ao neurônio i	
b_i	–	viés do neurônio i	
ϕ	–	função de ativação	
v_i	–	campo local induzido do neurônio i	
a	–	coeficiente que indica a inclinação da função de ativação	
$\mathbb{E}_t$	–	matriz que contem todos os padrões de entrada para o treinamento	
$\mathbb{W}$	–	matriz que contem todos os pesos sinápticos	
$\mathbb{B}$	–	matriz/vetor que contem todos os viés	
$\mathbb{A}$	–	matriz que contem todos os coeficientes que indicam a inclinação das funções de ativação	
M	–	número total de neurônios na camada de saída	
T	–	número total de padrões de treinamento	
$O_{t_{ij}}$	–	entrada ij da matriz de padrões de saída para o treinamento	
$O_{n_{ij}}$	–	entrada ij da matriz de saídas calculadas pela rede	

a_*^ℓ	–	valor máximo/mínimo que o coeficiente de inclinação das funções de ativação pode atingir	
w_*^k	–	valor máximo/mínimo que os pesos sinápticos podem atingir	
b_*^l	–	valor máximo/mínimo que os viés podem atingir	
$\boldsymbol{X}$	–	vetor contendo w, b e a iniciais	
∇	–	operador gradiente	
$\mathbb{H}^{-1}$	–	matriz Hessiana inversa	
R	–	nível de ruído	
G	–	grupo de padrões	
TA	–	taxa de acerto	
N_μ	–	número de direções discretas consideradas para medição da radiação emergente	
EQM	–	erro quadrático médio	
N_p	–	número total de padrões no conjunto de validação	
chl_i^e	–	concentração de clorofila exata	$mg \cdot m^{-3}$
chl_n^e	–	concentração de clorofila estimada pela rede	$mg \cdot m^{-3}$

O problema inverso pode ser formulado como um problema de otimização, buscando soluções regularizadas. Essa abordagem, chamada de implícita, foi utilizada para resolver problemas, nos quais se desejou estimar perfis verticais da concentração de clorofila [21]. Apesar dos bons resultados obtidos, a complexidade do problema inverso/direto fez necessário a utilização de processamento paralelo para a resolução [21, 23]. A resolução de ótica hidrológica inversa por esta estratégia requer enorme custo computacional. Gordon [9] faz uma revisão de técnicas de solução em vários problemas inversos em Ótica Hidrológica.

Uma estratégia para reduzir o tempo computacional na resolução deste tipo de problema inverso é a utilização de Redes Neurais Artificiais (RNAs) [7, 19, 5], devido a capacidade intrínseca desses sistemas em aproximar funções [12].

Estimativas da concentração de clorofila de superfície em águas naturais (além de outras propriedades), a partir da reflectância medida por satélites, já tem sido realizadas através de RNAs [14, 13, 11]. Nesses trabalhos, os padrões de entrada da rede são formados pelas reflectâncias medidas nas bandas (comprimentos de onda) do visível, sendo que o valor médio de cada banda corresponde a uma entrada da rede. Em nosso estudo, os padrões de entrada necessários para treinar e validar a rede são formados pela radiação (radiância) emergente na superfície, sendo que cada direção (polar) discreta considerada corresponde a uma entrada da rede. O método de inversão baseado em RNA será testado com medidas de radiâncias sintéticas, em que radiâncias são calculadas a partir da solução da ETR (problema direto). Para os coeficientes de absorção e espalhamento, são empregados modelos bio-ópticos [16, 10], nos quais estes coeficientes são dados em função da concentração de clorofila. A solução inversa é obtida através de uma RNA do tipo Perceptron de Múltiplas Camadas (MLP). O treinamento da rede é feito através da minimização do funcional de diferença quadrática entre as respostas dadas pela rede os padrões de saída conhecidos. A busca mínima do funcional é feita através do método quasi-Newton [5, 6]. O conjunto de parâmetros estimados no processo de treinamento da rede

inclui, além de pesos e viés, o parâmetro das funções de ativação.

2. Formulação do Problema Direto

A ETR é um modelo matemático utilizado para representar a interação de um feixe de luz (fótons) com um corpo d'água. Em muitas aplicações, principalmente aquelas que envolvem o estudo da interação desse feixe de luz com águas oceânicas, é plausível considerar que o corpo d'água apresente variações significativas, com relação aos seus constituintes, apenas com a profundidade. Neste caso, denomina-se o que é chamado de geometria plano-paralela. A ETR para este tipo de geometria, com dependência espectral, azimutal e polar, e espalhamento anisotrópico é dada por

$$\mu\frac{\partial}{\partial\tau}I(\tau,\mu,\varphi,\lambda)+I(\tau,\mu,\varphi,\lambda)= \\ \frac{b(\tau,\lambda)}{c(\tau,\lambda)}\int_{-1}^{1}\int_{0}^{2\pi}\int_{\lambda}p(\cos\Theta,\lambda)I(\tau,\mu',\varphi',\lambda')d\lambda' d\varphi' d\mu'+S_0(\tau,\lambda), \quad (2.1)$$

em que $I(\tau,\mu,\varphi,\lambda)$ representa a intensidade do feixe de radiação, $\tau\in(0,\,\zeta)$, a variável ótica, no qual ζ é a espessura ótica do meio, $\mu\in[-1,\,1]$, $\varphi\in[0,\,2\pi]$ e, representam, respectivamente, o co-seno 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 é o comprimento de onda do fóton. O termo $b(\tau,\lambda)$ é o coeficiente de espalhamento, $c(\tau,\lambda)=a(\tau,\lambda)+b(\tau,\lambda)$ é o coeficiente de atenuação do feixe em que $a(\tau,\lambda)$ é o coeficiente de absorção. Por fim, o termo $p(\cos\Theta,\lambda)$ é chamado de função de fase, e representa o espalhamento de um feixe incidente na direção $\{\mu',\,\varphi',\,\lambda\}$ a ser espalhado na direção $\{\mu,\,\varphi,\,\lambda\}$, e $S_0(\tau,\lambda)$ é uma fonte interna de radiação. As condições de contorno, associadas ao problema abordado, são dadas por

$$I(0,\mu,\varphi,\lambda)=I_0\delta(\mu-\mu_0)\delta(\varphi-\varphi_0) \qquad \text{e} \qquad I(\zeta,-\mu,\varphi,\lambda)=0, \qquad (2.2)$$

em que e I_0 é a radiância incidente na superfície, μ_0 é o cosseno do ângulo polar de incidência, φ_0 é o ângulo azimutal de incidência e $\delta(\,\cdot\,)$ é a função delta de Dirac.

Para resolver o problema definido pela equação (2.1), sujeita às condições dadas pelas equações (2.2), efetua-se uma discretização espacial e espectral [2]. Há um desacoplamento para cada comprimento de onda e aqui a dependência do comprimento de onda será tomado um valor médio, isto é,

$$I_\lambda(\tau,\mu,\varphi)=\frac{1}{\Delta\lambda}\int_{\lambda}^{\lambda+\Delta\lambda}I(\tau,\mu,\varphi,\lambda)d\lambda.$$

A fim de simplificar a notação não será grifado a dependência do comprimento de onda λ. O próximo passo consiste em realizar a decomposição da intensidade de radiação $I(\tau,\mu,\varphi)$ [4] em componentes espalhada $I_s(\tau,\mu,\varphi)$ e não-espalhada $I_u(\tau,\mu,\varphi)$ e, dessa forma, a solução $I(\tau,\mu,\varphi)$ passa a ser expressa como sendo a soma dessas componentes,

$$I(\tau,\mu,\varphi)=I_u(\tau,\mu,\varphi)+I_s(\tau,\mu,\varphi).$$

Para obter a solução para a componente não-espalhada, considera-se $b(\tau, \lambda) = 0$ na equação (2.1), sujeito às condições (2.2) e, então, resolve-se a equação diferencial parcial resultante. Já para a componente espalhada $I_s(\tau, \mu, \varphi)$, inicialmente aproxima-se a função de fase de Henyey-Greenstein [15] em uma série finita de funções de Legendre associadas, e em seguida utiliza-se a decomposição de Fourier em co-senos [4], sobre o ângulo azimutal, de modo a gerar $L + 1$ equações integro-diferencias (ETRs), sem a dependência de φ. O termo integral presente em cada equação íntegro-diferencial é aproximado pelo método da colocação, o qual consiste em substituir 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, é obtida através do método AS_N, que é baseado na decomposição espectral da matriz de espalhamento. Para a aproximação numérica é utilizado o código PEESNA [3], o qual implementa o método AS_N.

A relação entre os coeficientes de absorção e espalhamento com a concentração de clorofila é feita através de modelos bio-ópticos. Segundo [16] e [10], as expressões que relacionam a concentração de clorofila com a absorção e o espalhamento são dadas, respectivamente, por

$$a(\tau, \lambda) = \left[a_\lambda^w + 0.06\, a_\lambda^c\, C^{0.65}(\tau)\right] \left[1 + 0.2\, e^{-0.014(\lambda - 440)}\right],$$

e

$$b(\tau, \lambda) = b_\lambda^w + \frac{550}{\lambda}\, 0.3\, C^{0.62}(\tau),$$

em que a_λ^w e b_λ^w são, respectivamente, o coeficiente de absorção e espalhamento da água pura, cujos valores podem ser encontrados em [17], λ representa o comprimento de onda, a_λ^c é um coeficiente de absorção específico da concentração de clorofila, cujo valor pode ser encontrado em [20], e $C(\tau)$ é a concentração de clorofila, a qual, para o presente trabalho, é considerada constante. Por fim, mais detalhes acerca da solução para o problema definido nessa seção podem ser encontrados em [2].

3. Perceptron de Múltiplas Camadas

Um MLP é obtido através da conexão de vários neurônios entre si, de modo a formarem uma rede. Essa rede é constituída de uma camada de entrada, uma ou mais camadas ocultas, e uma camada de saída. Nesse modelo de rede, a informação é passada adiante camada por camada até atingir a camada de saída. A informação que chega em cada uma das unidades de processamento e que deverá ser processada por essas unidades, é definida matematicamente por $v_i = \sum_j w_{ji} x_j + b_i$, em que x_j representa as informações de entrada, w_{ji} as conexões entre os neurônios (sinapses), e b_i é chamado de nível de viés, e pode ser interpretado como um valor que é ajustado a fim de "complementar" alguma atividade presente no neurônio biológico, mas não presente no neurônio artificial. Essa informação é processada através de uma função de ativação ϕ, a qual tem por finalidade restringir a amplitude do sinal de saída. Assim, o sinal que segue para o próximo neurônio é definido como $y_i = \phi(v_i)$. Comumente são empregadas como funções de ativação: a função sigmóide, a função

tangente hiperbólica e a função linear, definidas, respectivamente, por

$$\phi(v) = \frac{1}{1+e^{-av}}, \quad a > 0, \quad \phi(v) = \frac{1-e^{-av}}{1+e^{-av}}, \quad a > 0 \quad \text{e} \quad \phi(v) = av, \quad a \in \mathcal{R}^*.$$

Redes do tipo MLPs requerem um treinamento do tipo supervisionado, ou seja, com a presença de um professor que tem o conhecimento do ambiente que a rede deverá aprender. Esse ambiente é representado por um conjunto de entradas e saídas conhecidas. Durante a fase de aprendizagem, o professor apresenta uma determinada entrada à rede, que por sua vez processa a informação apresentada e devolve uma saída. A saída calculada pela rede é comparada com a saída desejada, e com isso é gerado um sinal de erro, o qual é propagado para trás (retropropagação) através do algoritmo de treinamento. Durante a retropropagação do erro, o algoritmo de treinamento, baseado em alguma regra, faz uma correção nas variáveis livres da rede de modo a corrigir os valores dos seus parâmetros livres, a fim de minimizar o erro cometido. Dessa forma, espera-se que a rede aprenda a fazer a correta associação entre os padrões de entrada e os padrões de saída.

4. Formulação do Problema Inverso

A formulação para o treinamento, descrita abaixo, segue a adotada em [6], e nessa formulação, o treinamento é supervisionado e do tipo *batch* [12].

Para o treinamento, foi definido um funcional $\mathcal{J}(\,\cdot\,)$ que é dado pela soma das diferenças quadráticas entre os padrões de saída utilizados para o treinamento e as respostas obtidas pela rede para cada padrão de entrada, e para o qual buscou-se um valor mínimo, ou seja,

$$\mathcal{J}\left(\mathbb{E}_t, \mathbb{W}, \mathbb{B}, \mathbb{A}\right) = \min \sum_{j=1}^{T} \sum_{i=1}^{M} \left|\left|O_{t_{ij}} - O_{n_{ij}}\right|\right|_2^2, \tag{4.1}$$

em que $\mathbb{E}_t$, $\mathbb{W}$, $\mathbb{B}$ e $\mathbb{A}$ são matrizes que contem, respectivamente, os padrões de entrada para o treinamento, os pesos sinápticos, os *biases* e os parâmetros das funções de ativação. Já M representa o número total de neurônios na camada de saída, T representa o número total de padrões de treinamento, $O_{t_{ij}}$ e $O_{n_{ij}}$ representam, respectivamente, cada entrada das matrizes que contem os padrões de saída para o treinamento e as saídas calculadas pela rede. Outros detalhes em [6].

Note que é necessário incluir restrições às variáveis que estão sendo otimizadas, especialmente às relacionadas com o parâmetro das funções de ativação. Pela definição das funções sigmóide e tangente hiperbólica, estes parâmetros devem ser positivos e, portanto, os "valores ótimos" para esses parâmetros devem satisfazer

$$0 \;<\; a^{\ell} \;\leq\; a^{\ell}_{\max}\,, \tag{4.2}$$

em que $a^{\ell}_{\max}$ 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*. Se for utilizada a função de ativação

linear, a restrição dada pela equação (4.2) pode ser substituída por

$$-a^{\ell}_{\min} \;\leq\; a^{\ell} \;\leq\; a^{\ell}_{\max}\,, \quad a^{\ell} \neq 0, \tag{4.3}$$

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

O treinamento da rede pode envolver uma função com múltiplos mínimos locais, o que torna difícil garantir que será encontrado um mínimo global. Diante disso, parece ser atraente a idéia de reduzir o espaço de busca e, então, buscar por um mínimo local dentro desse sub-espaço gerado, desde que esse produza bons resultados. A redução do espaço de busca é obtido colocando restrições nas demais variáveis a serem otimizadas, ou seja, nos pesos e nos viés. Portanto, durante o processo de treinamento são buscados valores para a matriz $\mathbb{W}$ e para matriz/vetor $\mathbb{B}$ de modo que estes satisfaçam as restrições

$$\begin{array}{rcccl} -w^{k}_{\min} & \leq & w^{k} & \leq & w^{k}_{\max} \\ -b^{l}_{\min} & \leq & b^{l} & \leq & b^{l}_{\max} \end{array}\,, \tag{4.4}$$

em que w^{k}_{*} está associado às restrições aos pesos, b^{l}_{*} está associado às restrições aos viés. Outra vantagem que surge com a inclusão dessas restrições nos pesos e viés, é o fato que é possível evitar que esses parâmetros assumam ordens de grandeza elevadas e, assim, evitar a 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 equação (4.1), sujeito às condições expressas pelas equações (4.2) e/ou (4.3) e/ou (4.4).

Neste trabalho, para a minimização do funcional definido na equação (4.1), é utilizado o método quasi-Newton implementado na sub-rotina E04UCF da biblioteca NAG, desenvolvida por *Numerical Algorithms Group – NAG* [18]. Optou-se pela utilização da biblioteca citada em função da robustez da mesma para solução de problemas de otimização não-lineares, e pela possibilidade da inclusão das restrições comentadas anteriormente, ver [5, 6].

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

$$\nabla \mathcal{J}\,(\,\cdot\,, \boldsymbol{X}^{*}) = \boldsymbol{0}, \tag{4.5}$$

e ainda leva em consideração que $\boldsymbol{X}^{*}$ satisfaz as condições expressas pelas equações (4.2) e/ou (4.3) e/ou (4.4). Agora, se $\boldsymbol{X}_{k+1} = \boldsymbol{X}_{k} + \Delta\boldsymbol{X}_{k}$ aproxima o ponto $\boldsymbol{X}^{*}$ na $(k+1)$-ésima iteração, então, fazendo a aproximação de primeira ordem em série de Taylor para a equação (4.5) em torno do ponto $\boldsymbol{X}_{k+1}$, obtém-se $\boldsymbol{X}_{k+1} = \boldsymbol{X}_{k} + \mathbb{H}^{-1}\nabla\mathcal{J}\,(\,\cdot\,, \boldsymbol{X}_{k})$, em que $\mathbb{H}^{-1}$ é a matriz Hessiana inversa em $\boldsymbol{X} = \boldsymbol{X}_{k}$. Na prática, a inversa $\mathbb{H}^{-1}$ não é calculada de forma exata, e sim de uma forma aproximada. Para isso, considere que $\boldsymbol{X}_{k}$ é um ponto inicial e $\mathbb{H}^{-1}_{k}$ é uma aproximação da Hessiana inversa, e que respeita as condições definidas em [8]. Então, primeiramente, o algoritmo calcula o gradiente de $\mathcal{J}$ em $\boldsymbol{X}_{k}$, em segundo calcula uma direção de busca definida pela expressão $d_{k} = -\mathbb{H}^{-1}_{k}\nabla\mathcal{J}(\,\cdot\,, \boldsymbol{X}_{k})$, posteriormente atualiza $\boldsymbol{X}_{k}$

de modo que $\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + \alpha_k d_k$, em que α_k é o tamanho do passo e, por fim, atualiza $\mathbb{H}_k^{-1}$ através da expressão

$$\mathbb{H}_{k+1}^{-1} = \mathbb{H}_k^{-1} - \frac{\mathbb{H}_k^{-1}\boldsymbol{S}_k\boldsymbol{S}_k^T\mathbb{H}_k^{-1}}{\langle \boldsymbol{S}_k, \mathbb{H}_k^{-1}\boldsymbol{S}_k\rangle} + \frac{\boldsymbol{S}_k\boldsymbol{S}_k^T}{\langle \boldsymbol{Y}_k, \boldsymbol{S}_k\rangle},$$

em que $\boldsymbol{S}_k = \boldsymbol{X}_{k+1} - \boldsymbol{X}_k$ e $\boldsymbol{Y}_k = \nabla\mathcal{J}(\cdot, \boldsymbol{X}_{k+1}) - \nabla\mathcal{J}(\cdot, \boldsymbol{X}_k)$. Detalhes da biblioteca e do método quasi-Newton podem ser encontrados em [18] e [8], respectivamente.

5. Conjuntos de Treinamento, Validação e Treinamento da Rede

Nesta seção descreve-se como são gerados os conjuntos de treinamento e de validação. Como o formalismo matemático para a formulação do problema de treinamento foi descrito na seção anterior, aqui são abordados apenas os detalhes do procedimento de treinamento.

5.1. Conjuntos de Treinamento e Validação

Ambos conjuntos, de treinamento e de validação, utilizados neste trabalho, são obtidos a partir da resolução do problema exposto na Seção 2.

Para este trabalho, adotou-se uma única região espacial e nessa região (profundidade ótica) considerou-se que o perfil da concentração de clorofila C é constante, ou seja, um valor médio na região. A partir do parâmetro C, é possível resolver o problema definido naquela seção, determinando a radiação emergente na superfície após a interação dessa com o meio, em direções (co-seno do ângulo polar) escolhidas *a priori*. São esses valores da radiação emergente que constituirão os padrões de entrada para o treinamento e validação/teste da rede. As direções discretas adotadas correspondem a um total de dez igualmente espaçadas e pertencentes ao intervalo $\theta \in [90°, 135°]$. Para o ângulo azimutal adotou-se $\varphi_0 = 0°$. Com relação à concentração de clorofila adotou-se $C \in [0.01, 10.0]$.

Para gerar o conjunto de treinamento, esse intervalo foi discretizado como $C = [0.01, 0.02, 0.03, \ldots, 0.1, 0.15, 0.2, 0.3, 0.4, \ldots, 9.8, 9.9, 10.0]$, gerando 110 valores discretos[3]. Já para gerar o conjunto de validação, adotou-se uma discretização igualmente espaçada com um tamanho de passo de $\Delta_C = 0.01$, gerando 1000 valores discretos. Portanto, cada padrão de entrada (solução do problema para cada valor discreto de C) de ambos conjuntos (treinamento e validação) é composto de dez entradas, e cada padrão de saída é composto pelo respectivo valor de C que gerou o padrão de entrada. Note que o número total de padrões no conjunto de treinamento é 110, e no conjunto de validação é 1000.

Por fim, vale lembrar que cada valor de C corresponde a um dado de entrada, e a radiação emergente, calculada pelo modelo, corresponde a saída (problema direto). Já para a rede, a radiação emergente passa a ser a entrada e o valor de C passa a ser a saída (problema inverso).

[3] A diferença no tamanho do passo na discretização é devido a variação significativa da razão $b(\cdot)/c(\cdot)$ para valores baixos de C.

5.2. Treinamento da Rede

Como se está utilizando dados sintéticos para treinar a rede, é necessário que os valores da radiação retornados pelo problema direto, sejam corrompidos com ruído, a fim de simular possíveis erros de medidas. Para isso, são considerados três níveis de ruído gaussiano: $R_1 = 5\%$, $R_2 = 10\%$ e $R_3 = 20\%$. Em seguida, os padrões do conjunto de treinamento são colocados em ordem aleatória e, então, divididos em três grupos (G_1, G_2 e G_3) com a mesma quantidade de padrões em cada grupo. Cada um desses grupos recebeu um dos níveis de ruído, sendo: R_1 para o G_1, R_2 para G_2 e R_3 para G_3. Durante o treinamento esses grupos são apresentados à rede.

O critério de parada adotado é o da validação cruzada [12], sendo que a cada dez iterações da função objetivo o procedimento de treinamento é pausado, e todo o conjunto de validação é apresentado à rede quatro vezes. Na primeira vez, todos os padrões do conjunto de validação são corrompidos com R_1 e, então, apresentados à rede. O mesmo se deu na segunda, terceira e quarta vez, sendo que na última é utilizado um nível de ruído $R_4 = 13\%$. Este nível, não presente no treinamento, é utilizado a fim de agregar à rede uma maior generalização.

Cada vez que o processo de validação é realizado, o desempenho da rede é testado e são armazenadas as variáveis da rede, bem como o número de acertos e o valor total do erro relativo. Ao fim do processo, aquelas variáveis que produziram o maior número de acertos e o menor erro relativo são recuperadas e adotadas como sendo variáveis ótimas.

Para a contagem das respostas corretas da rede, adotou-se o critério de que o valor de C estimado pela rede deve estar no intervalo $C - R \cdot C \leq C^r \leq C + R \cdot C$, em que R é o nível de ruído nos dados de entrada e C^r é o valor da concentração de clorofila estimado pela rede. Além disso, a rede utilizada é constituída de uma camada de entrada com dez neurônios, uma camada oculta com 13 neurônios e uma camada de saída com apenas um neurônio. A função de ativação utilizada em todos os neurônios da camada oculta foi a função sigmóide $\phi(v_i) = 1/(1 + \exp(-a_i v_i))$ e na camada de saída foi utilizada a função linear $\phi(v_i) = a_i v_i$.

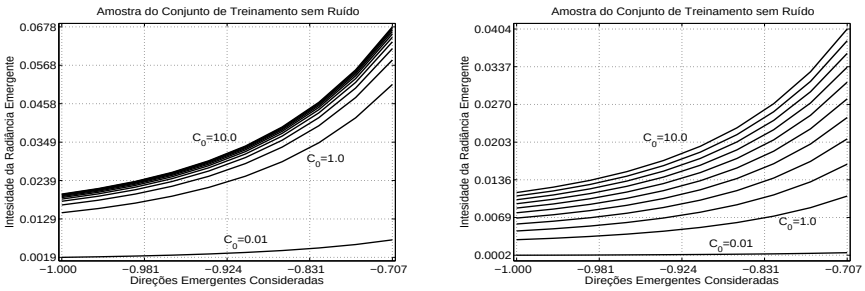
Para as restrições definidas na equação (4.4), adotou-se $w^k_{\min} = b^l_{\min} = w^k_{\max} = b^l_{\max} = 45$, para a restrição (4.2), adotou-se $a^\ell_{\max} = 20$ e, por fim, para a restrição definida pela equação (4.3), adotou-se $a^\ell_{\min} = a^\ell_{\max} = 20$. Esses valores foram determinados durante o treinamento através de experimentos, e são capazes de representar uma região do espaço de busca no qual é possível obter boas respostas.

6. Resultados

Na resolução do problema, foram considerados cinco valores para $\lambda = 500, 550, 600, 650, 700\,\text{nm}$, no entanto, por conveniência, são apresentados apenas os resultados correspondentes à $\lambda = 600\,\text{nm}$. Para $\lambda = 500\,\text{nm}$, os resultados foram extremamente pobres e não expressaram nenhuma confiança, e para $\lambda = 550\,\text{nm}$, os resultados melhoraram um pouco, no entanto, a taxa de acerto ficou abaixo de 71%. Apesar de comprimentos de onda $\lambda > 600\,\text{nm}$ apresentarem bons resultados, estes não são adequados, pois o índice de absorção da água aumenta consideravelmente [15], fazendo com que a radiação eletromagnética não penetre mais do que alguns

poucos metros.

A dificuldade na obtenção de bons resultados para comprimentos de onda abaixo de $\lambda = 600\,\text{nm}$ pode ser analisado na Figura 1. A Figura 1(a) mostra que há uma "separação" entre cada curva de radiação até aproximadamente $C = 3$ e, após esse valor, as curvas de radiação praticamente se sobrepõe. Por outro lado, essa sobreposição não ocorre na Figura 1(b), na qual é possível observar que as curvas da radiação para cada valor de C considerado são distintas.



(a) (b)

Figura 1: Radiação emergente na superfície, sem a adição de ruído, para alguns valores de C, para dois comprimentos de onda: (a) $\lambda = 500\,\text{nm}$; (b) $\lambda = 600\,\text{nm}$.

Uma análise quantitativa acerca da sensibilidade da radiação emergente com a variação da concentração de clorofila é apresentada na Tabela 2. Nessa tabela são apresentados os resultados obtidos através da expressão[4]

$$\Delta I_i \equiv \sum_{k=0}^{N_\mu} \left[I(0, \mu_k, 0^\circ, C + \delta C) - I(0, \mu_k, 0^\circ, C) \right]^2, \qquad i = 1, \ldots 9, \tag{6.1}$$

em que $\delta C = 1$. Como é possível observar pelos resultados, a variação ΔI para $\lambda = 600$ é maior e mais significativa quando comparada a variação ΔI para $\lambda = 500$. A não sobreposição das curvas apresentadas na Figura 1(b), ou então, a maior sensibilidade da radiação emergente com a variação da concentração de clorofila é o fator fundamental que contribui para a melhora no desempenho da rede.

Na Tabela 3, são apresentados, para cada nível de ruído considerado e para os problemas associados aos comprimentos de onda $\lambda = 500, 550$ e $600\,\text{nm}$, as taxas de acerto (TA) e os erros quadráticos médios (EQM), calculados pela expressão

$$EQM = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (chl_i^e - chl_i^n)^2},$$

onde N_p é o número total de padrões no conjunto de validação, e chl_i^e e chl_i^n são, respectivamente, as concentrações de clorofila exatas e estimadas pela rede neural,

[4]Os valores de radiância apresentados das Figuras 1(a) e 1(b) foram normalizados para ter a mesma ordem de grandeza, a fim de facilitar a comparação.

Tabela 2: Soma das diferenças quadráticas expressa pelo equação (6.1) para as radiações emergentes mostradas nas Figuras 1(a) e 1(b).

$\lambda = 500$				$\lambda = 600$			
ΔI_1	0.0375	ΔI_6	0.0004	ΔI_1	0.0634	ΔI_6	0.0146
ΔI_2	0.0083	ΔI_7	0.0002	ΔI_2	0.0399	ΔI_7	0.0122
ΔI_3	0.0030	ΔI_8	0.0002	ΔI_3	0.0287	ΔI_8	0.0104
ΔI_4	0.0013	ΔI_9	0.0001	ΔI_4	0.0221	ΔI_9	0.0090
ΔI_5	0.0006	–	–	ΔI_5	0.0177	–	–

sobre o conjunto de validação. Vale enfatizar que, como os níveis de ruído definem as faixas de acertos, para o intervalo com ruído de 10% considera-se como aceitáveis valores de inversão com desvio de até 10%, diferentemente da faixa de 5% de ruído – onde só se considera aceitável inversões com desvio máximo de 5%. Assim é possível encontrar mais acertos em experimentos com maior nível de ruído (ver Tabela 3).

Tabela 3: Taxas de acerto (TA), em %, e erros quadrático médios (EQM) sobre o conjunto de validação para cada nível de ruído.

λ	Nível de Ruído (%)							
	5		10		15		20	
	Ac	EQM	Ac	EQM	Ac	EQM	Ac	EQM
500	19.7	1.00	31.9	1.55	33.3	1.79	35.7	2.37
550	50.7	0.38	67.2	0.64	70.8	0.76	70.4	1.13
600	83.0	0.21	87.3	0.37	90.2	0.43	91.4	0.66

O menor valor de EQM, mostrado nesta tabela, complementa a análise e ajuda a demonstrar que os melhores resultados são os gerados com $\lambda = 600\,\text{nm}$. Esta melhora está associada à distância (separação) que existe entre cada curva da Fig. 1(b). Esta separação evita a confusão da rede durante a associação do padrão de entrada ao verdadeiro padrão de saída.

A Figura 2 mostra os perfis de clorofila obtidos pela rede para o conjunto de validação, para cada um dos níveis de ruído considerados. O eixo das ordenados representa as concentrações de clorofila e o eixo das abscissas representa cada um dos padrões do conjunto de validação. Nas figuras, as linhas tracejadas representam o limite de erro na estimativa de C pela rede. A linha cheia representa cada perfil recuperado. Optou-se pela divisão dos gráficos a fim de melhorar a visualização dos resultados, principalmente para baixas concentrações de C.

7. Conclusões

A utilização do parâmetro das funções de ativação ajustáveis é uma técnica pouco abordada na literatura, e sua utilização tem se mostrado eficiente, pois cada neurônio pode se especializar em uma determinada região do conjunto de treinamento. Isso melhora a capacidade de generalização da rede e a robustez a dados ruidosos.

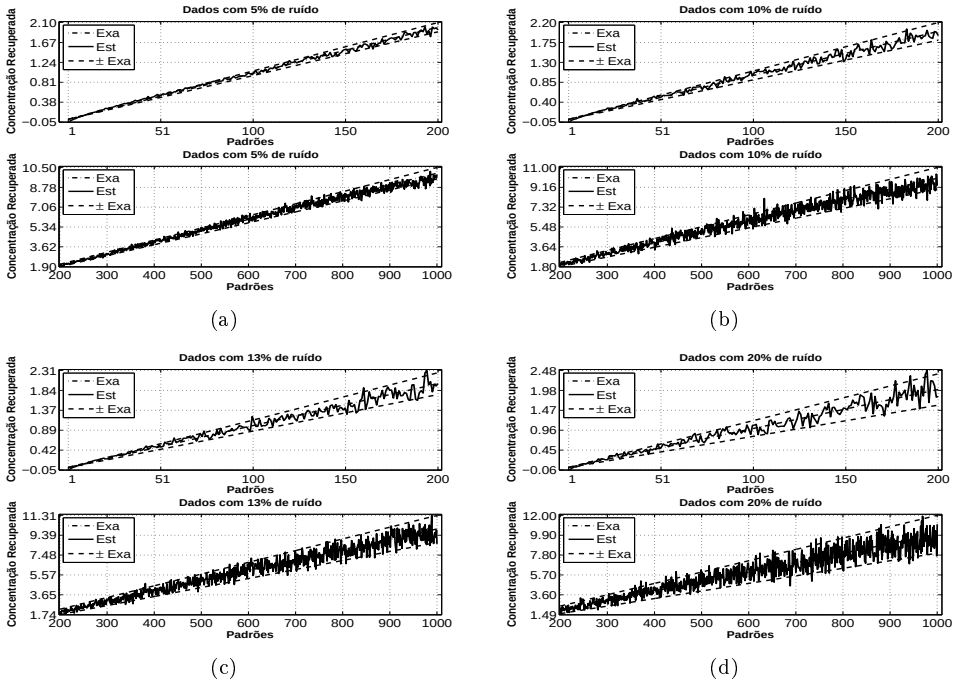


Figura 2: Perfis médios recuperados para dados de entrada gerados com $\lambda = 600$ nm, corrompidos com diferentes níveis de ruído: (a) 5%; (b) 10%; (c) 13%; (d) 20%.

Um treinamento eficiente (de baixo custo computacional), juntamente com os bons resultados obtidos, mostram que a RNA é uma técnica promissora na resolução de problemas inversos em ótica hidrológica.

Fótons com comprimento de onda acima de 650 nm tem baixa penetração no corpo de água (são absorvidos em até 5 m). A radiância para fótons abaixo de 550 nm, para diferentes concentrações de clorofila, colapsam em uma mesma curva (ver Figura 1(a)). A fim de garantir uma certa penetração da radiação, com distintas curvas de radiâncias para diferentes concentrações de clorofila adotou-se o intervalo [550 nm, 650 nm].

O parágrafo acima ilustra a dificuldade intrínseca da estimativa de propriedades óticas em águas naturais. A dificuldade apontada para o comprimento de onda $\lambda = 500$ nm está associado ao colapso, ou ainda, a baixa sensibilidade da radiação emergente com a concentração de clorofila. Esta é uma restrição associada à física do problema e não da metodologia de inversão. Para redes neurais, várias configurações e diferentes valores iniciais para as variáveis foram utilizadas, sendo que, em nenhuma dessas situações, foi possível melhorar os resultados apresentados na Tabela 3. Já, para o problema no qual foi considerado $\lambda = 600$ nm, o aprendizado da rede se deu de forma rápida, sendo pouco sensível ao número de neurônios na camada oculta e aos valores iniciais das variáveis da rede. Em trabalhos futuros, pretende-se retomar o problema utilizando um fator de correção para tentar atenuar

o fator de decaimento exponencial da radiação (efeito da absorção) [22].

Abstract. In this work the average profile of chlorophyll concentration is estimated from the emitted radiation at the surface of natural waters. This is performed through the use an Artificial Neural Network of Multilayer Perceptron type to act as the inverse operator. Bio-optical models are used to correlate the chlorophyll concentration with the absorption and scattering coefficients. The network training is formulated as an optimization problem, in which the update of the free variables of network (weights, viéses and each slope of the activation functions) is performed through the quasi-Newton method.

Keywords. Multilayer perceptron, quasi-Newton method, chlorophyll concentration, radiative transfer equation.

Referências

[1] E.S. Chalhoub, "O Método das Ordenadas Discretas na Solução da Equação de Transporte em Geometria Plana com Dependência Azimutal", Tese de Doutorado, IPEN, USP, São Paulo, SP, 1997.

[2] E.S. Chalhoub, Discrete-ordinates solution for uncoupled multi-wavelength radiative transfer problems, *J. Quant. Spec. Rad. Trans.*, **92** (2005), 335–349.

[3] E.S. Chalhoub, H.F. Campos Velho, R.D.M. Garcia, M.T. Vilhena, A comparison of radiances generated by selected methods of solving the radiative-transfer equation, *Trans. Theory Stat. Phys.*, **32**, No. 5-7 (2003), 473–503.

[4] S. Chandrasekhar, "Radiative Transfer", Dover Publications, New York, 1950.

[5] F. Dall Cortivo, E.S. Chalhoub, H.F. Campos Velho, Comparison of two learning strategies for a supervised neural network, em "1st International Symposium on Uncertainty Quantification and Stochastic Modeling. Uncertainties 2012", 366–380, 2012.

[6] F. Dall Cortivo, E.S. Chalhoub, H.F. Campos Velho, A committee of MLP with adaptive slope parameter trained by the quasi-Newton method to solve problems in hydrologic optics, em "IEEE International Joint Conference on Neural Networks. IJCNN 2012. (IEEE World Congress on Computational Inteligence)", 2159–2166, 2012.

[7] F. Dall Cortivo, E.S. Chalhoub, J.D.S. Silva, H.F. Campos Velho, Estimativa do albedo de espalhamento simples usando uma rede neural de múltiplas camadas, em "Anais do XXXIII CNMAC", 411–417, SBMAC, 2010.

[8] J.E. Dennis, J.J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Review*, **19**, No. 1 (1977), 46–89.

[9] H.R. Gordon, Inverse methods in hydrologic optics, *Oceanology*, **44**, No. 1 (2002), 9–58.

[10] H.R. Gordon, A.Y. Morel, "Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery, A Review", Springer-Verlag, New York, 1983.

[11] L. Gross, S. Thiria, R. Frouin, B.G. Mitchell, Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration, *J. Geophys. Res.*, **105**, No. C2 (2000), 3483–3495.

[12] S. Haykin, "Redes Neurais", Bookman, Porto Alegre, 2001.

[13] S. Helmut, R. Doerffer, Neural network for emulation of an inverse model – operational derivation of Case II water properties from MERIS data, *Int. J. Rem. Sens.*, **20**, No. 9 (1999), 1735–1746.

[14] L.E. Keiner, C.W. Brown, Estimating oceanic chlorophyll concentrations with neural networks, *Int. J. Rem. Sens.*, **20**, No. 1 (1999), 189–194.

[15] C.D. Mobley, "Light and Water", Academic Press, California, 1994.

[16] A.Y. Morel, Light and marine photosynthesis: a spectral model with geochemical and climatological implications, *Prog. Oceanogr.*, **26**, No. 3 (1991), 263–306.

[17] A.Y. Morel, L. Prieur, Analysis of variations in ocean color, *Limnol. Oceanogr.*, **22**, No. 4 (1977), 709–722.

[18] NAG, "Fortran Library Manual", Numerical Algorithms Group, Oxford, 1995.

[19] R.C. Oliveira, N.I.A. Acevedo, A.J. Silva Neto, L. Biondi Neto, Aplicação de um comitê de redes neurais artificiais para a solução de problemas inversos em Transferência Radiativa, *TEMA – Tend. Mat. Apl. Comput.*, **11**, No. 2 (2010), 171–182.

[20] L. Prieur, S. Sathyendranath, 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, *Limnol. Oceanogr.*, **26**, No. 4 (1981), 671–689.

[21] R.P. Souto, "Recuperação de Perfis Verticais de Propriedade Óticas Inerentes a partir da Radiação Emergente da Água", Tese de Doutorado, CAP, INPE, São José dos Campos, SP, 2006.

[22] R.P. Souto, H.F. Campos Velho, S. Stephany, Reconstruction of chlorophyll vertical profiles from in-situ radiances using the ant colony meta-heuristic, "Iberian Latinamerican Congress on Computational Methods", Anais do XXV CILAMCE, 2004. **1**, Recife (PE), Brasil (2004).

[23] R.P. Souto, H.F. Campos Velho, S. Stephany, M. Kampel, Chlorophyll concentration profiles from in situ radiances by ant colony optimization, *J. Phys.: Conf. Series*, **124**, No. 1 (2008).