



20 e 21 de outubro  
Instituto Nacional de Pesquisas Espaciais - INPE  
São José dos Campos - SP

## Uso de uma Rede Neural de Múltiplas Camadas para Estimar o Albedo de Espalhamento Simples

Fábio Dall Cortivo<sup>1</sup>, Ezzat S. Chalhoub<sup>2</sup>, José Demisio S. da Silva<sup>2</sup>,  
Haroldo F. de Campos Velho<sup>2</sup>

<sup>1</sup>Programa de Doutorado em Computação Aplicada – CAP  
Instituto Nacional de Pesquisas Espaciais – INPE

<sup>2</sup>Laboratório Associado de Computação e Matemática Aplicada – LAC  
Instituto Nacional de Pesquisas Espaciais – INPE

{fabio.cortivo, ezzat, demisio, haroldo}@lac.inpe.br

**Abstract.** *The purpose of this work is the identification of the albedo of single scattering in Hydrologic Optic by means of artificial neural networks. The direct problem is modeled by the Radiative Transfer Equation (RTE), where the angular integral is approximated by the Discrete Ordenates Method ( $S_N$ ) and the resultant ordinal differential equations system is solved with the use of spectral decomposition. At the inverse problem we estimated the albedo of single scattering by means of measured radiance. We will adopt as an inverse operator a neural network of Multi Layer Perceptron (MLP) and the back-propagation algorithm for training. The activation function used in the hidden layer was the logistic function with two distinct values for the parameter  $\alpha$ , and this technique has produced satisfactory results for the desired estimatives.*

**Resumo.** *A proposta deste trabalho é a identificação de albedo em Óptica Hidrológica através de redes neurais artificiais. O problema direto é modelado pela Equação de Transferência Radiativa (ETR), onde a integral angular é aproximada pelo Método  $S_N$ , e o sistema de equações diferenciais ordinárias resultante é resolvido com uso da decomposição espectral. No problema inverso é feita a estimativa do albedo de espalhamento simples, a partir das radiâncias medidas. Adotaremos como operador inverso uma rede neural do tipo Perceptron de Múltiplas Camadas (MLP), juntamente com o algoritmo de retro propagação do erro para o treinamento. A função de ativação utilizada na camada oculta foi a função logística com dois valores distintos para o parâmetro  $\alpha$ , e essa técnica 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 ( $S_N$ ) [Chandrasekhar 1960] 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 [Mobley 1994]. O sistema de EDO resultante é resolvido com o uso da decomposição espectral da matriz do sistema, que é a base do método  $AS_N$  [Chalhoub 2003].

O problema inverso consistirá na estimativa do albedo de espalhamento simples, que representa a razão entre o coeficiente de espalhamento e o coeficiente de extinção total, 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

$$\mu \frac{\partial}{\partial \tau} I(\tau, \vec{\xi}) + I(\tau, \vec{\xi}) = \frac{\varpi_0}{4\pi} \int_{-1}^1 \int_0^{2\pi} \beta(\vec{\xi}; \vec{\xi}') I(\tau, \vec{\xi}') d\xi', \quad (1)$$

onde  $\tau$  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  $\varpi_0$  é o albedo de espalhamento simples.

Para completarmos a formulação do problema é necessário que sejam definidas as condições de contorno, ou seja, o fluxo na entrada e na saída do ambiente considerado, dadas por

$$I(0, \vec{\xi}) = G_+(\vec{\xi}) \quad \text{para } \mu > 0, \quad \text{e} \quad I(\tau_0, \vec{\xi}) = G_-(\vec{\xi}) \quad \text{para } \mu < 0. \quad (2)$$

Para que tenhamos uma forma mais simplificada da Eq. (1), precisamos encontrar uma aproximação para a função de fase. O ângulo  $\Theta$  indica o cosseno do ângulo formado por dois vetores com a origem de um sistema de coordenadas espaciais e extremidade sobre uma esfera unitária. Segundo Chandrasekhar [Chandrasekhar 1960], podemos expandir a função de fase em Polinômios de Legendre

$$\beta(\cos \Theta) = \sum_{n=0}^{\infty} \omega_n P_n(\cos \Theta). \quad (3)$$

Utilizando o Teorema da Adição dos Harmônicos Esféricos [Chandrasekhar 1960], (detalhes do teorema são descritos em [Cortivo 2008]), e considerando como função de fase a função de Henyey-Greenstein [Mobley 1994]

$$\beta(g, \Theta) = \frac{1 - g^2}{\sqrt{(1 - 2g \cos \Theta + g^2)^3}}, \quad (4)$$

os coeficientes  $\omega_n$  são expressos [Cortivo 2008], por

$$\omega_n = (2n + 1)g^n, \quad \text{com } \omega_0 = 1, \quad \text{e } P_0(\cos \Theta) = 1. \quad (5)$$

Desta forma, podemos expressar a ETR por

$$\mu \frac{\partial I^m(\tau, \mu)}{\partial \tau} + I^m(\tau, \mu) = \frac{\varpi_0}{2} \sum_{n=0}^{\infty} \omega_n \sum_{m=0}^n \frac{(n-m)!}{(n+m)!} P_n^m(\mu) \int_{-1}^1 P_n^m(\mu') I^m(\tau, \mu) d\mu', \quad (6)$$

na qual  $I(\tau, \mu)$  representa o coeficiente de Fourier, gerado a partir da decomposição de Chandrasekhar [Chandrasekhar 1960]. Nessa última expressão o conjunto de equações íntegro-diferenciais passa a ser uma outra forma de expressar a Eq. (1). Agora, aplicando o processo de Fourier nas condições de contorno (2), passamos a escrever tais condições como

$$I^m(0, \mu) = G_+^m(\mu) \quad \text{para } \mu > 0 \quad \text{e} \quad I^m(\tau_0, \mu) = G_-^m(\mu) \quad \text{para } \mu < 0. \quad (7)$$

O método da aproximação em ordenadas discretas, consiste em aproximar o termo integral da Eq. (6) por um esquema de quadratura de Gauss-Legendre, [Chandrasekhar 1960], onde o polinômio deve ter grau par. Não parece plausível tentar determinar a radiação  $I^m(\tau, \mu_i)$  em um número de direções maior que o número de direções onde é conhecido o espalhamento do feixe. Levando isso em conta, e utilizando o método de colocação para as raízes do polinômio de Legendre de ordem  $N$ , para cada inteiro  $m$ , temos uma nova expressão para a Eq. (6), a saber

$$\mu_i \frac{\partial I_i^m(\tau)}{\partial \tau} + I_i^m(\tau) = \frac{\varpi_0}{2} \sum_{n=0}^{m'} \omega_n \sum_{m=0}^n \frac{(n-m)!}{(n+m)!} P_n^m(\mu_i) \sum_{j=0}^{N-1} \bar{\omega}_j P_n^m(\mu_j) I_j^m(\tau)$$

onde  $m' \leq N - 1$ ,  $I_j^m(\tau) = I_j^m(\tau, \mu_j)$ .

Temos agora para cada  $m$  um sistema de equações diferenciais ordinárias, sujeitas às condições de contorno definidas na Eqs. (7). O problema aqui abordado considerará apenas o caso onde  $m = 0$ . Em forma matricial a equação acima pode ser escrita como

$$\frac{d\vec{I}(\tau)}{d\tau} = \mathbb{A}\vec{I}(\tau) \quad (8)$$

onde a matriz  $\mathbb{A}$  é definida por

$$a_{i,j} = -\frac{\delta_{ij}}{\mu_i} + \frac{\varpi_0}{2\mu_i} \sum_{n=0}^M \omega_n P_n(\mu_i) P_n(\mu_j) \bar{\omega}_j, \quad i, j = 1, 2, \dots, N,$$

sendo  $\delta_{ij} = 1$ , se  $i = j$  e  $\delta_{ij} = 0$ , se  $i \neq j$ . As respectivas condições de contorno são dadas por

$$\vec{I}(0) = \vec{G}_+(\mu_i), \quad \text{se } 1 \leq i \leq N/2, \quad \text{e} \quad \vec{I}(\tau_0) = \vec{G}_-(\mu_i), \quad \text{se } N/2 + 1 \leq i \leq N. \quad (9)$$

Portanto, ao invés de resolver o problema definido pela Eq. (1) sujeito às condições de contorno (2), resolvemos o problema definido pela Eq. (8) sujeito às condições de contorno (9). Adotamos para o problema em questão  $\vec{G}_+(\mu_i) = \vec{1}$  e  $\vec{G}_-(\mu_i) = \vec{0}$ . Resolvemos este problema de geometria plano-paralela com uma camada d'água com espessura de 40 metros. O problema (8) foi resolvido por diagonalização como mostrado em [Chalhoub 2003] e [Cortivo 2008].

### 3. Rede Neural – Perceptron de Múltiplas Camadas

McCulloch e Pitts [McCulloch and Pitts 1943] desenvolveram o neurônio artificial, o qual consiste basicamente em um combinador linear e uma função de ativação, conforme mostrado na Figura 1(a). O vetor de valores  $\vec{x}$  são as entradas das sinapses conectadas a um neurônio  $i$ , através dos  $w_{ij}$ , que são os pesos, representando a “energia” das conexões,  $b$  é um parâmetro de ajuste e  $\varphi$  é a função de ativação, que tem por finalidade restringir a amplitude da saída de um neurônio. A representação em forma de equação para a saída de um neurônio é dada por

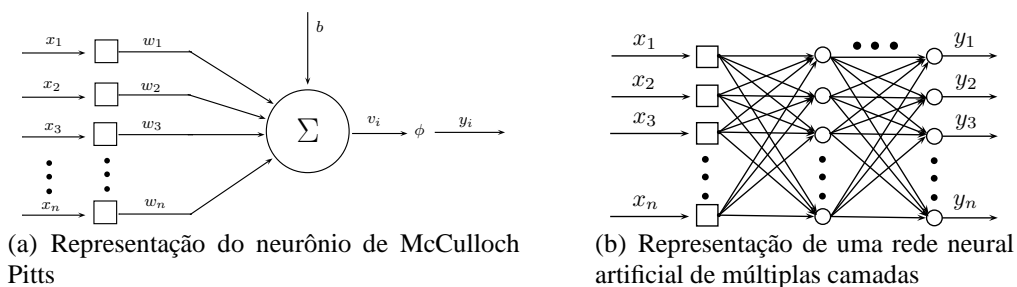
$$y_i = \varphi(v_i) = \varphi \left( \sum_{j=1}^n w_{ij}x_j + b \right),$$

onde  $\varphi$  é a função de ativação,  $x_j$  é o valor de entrada do neurônio  $j$ ,  $b$  é um viés associado ao neurônio  $j$ , e  $w_{i,j}$  é o peso sináptico do neurônio artificial.

Muitas funções de ativação podem ser usadas dependendo do tipo de problema abordado. As redes neurais artificiais podem resolver problemas não lineares se a função de ativação, para as camadas ocultas ou de saída, for não linear. Para o trabalho presente utilizaremos como função de ativação a função logística, a saber

$$\varphi(v) = \frac{1}{1 + e^{-av}}, \quad \text{onde } a > 0. \quad (10)$$

O modelo de rede neural mostrado na Figura 1(a) é chamado perceptron de camada única, e é capaz de resolver problemas linearmente separáveis [Haykin 2001], porém, muitos problemas não são linearmente separáveis. A estratégia para contornar estes problemas foi a “adição” de camadas ocultas de neurônios (nós de fonte), a qual constitui uma camada de entrada, uma ou mais camadas ocultas de nós computacionais e uma camada de saída, onde a função de ativação de cada neurônio/camada oculto(a) pode ser linear ou não [Haykin 2001]. Uma representação gráfica de tal rede é mostrada na Figura 1(b).



**Figura 1. Gráfico do neurônio de McCulloch e Pitts e de uma rede de múltiplas camadas.**

Essa arquitetura de rede, associada com o algoritmo de retro-propagação do erro (treinamento supervisionado), tem sido aplicada com sucesso na resolução de vários problemas difíceis. Tal algoritmo de treinamento, por ser supervisionado, requer o conhecimento dos dados de saída associado aos dados de entrada. Dessa forma, após os dados

de entrada serem propagados adiante, é possível calcular o erro cometido na aproximação feita e os dados de saída reais.

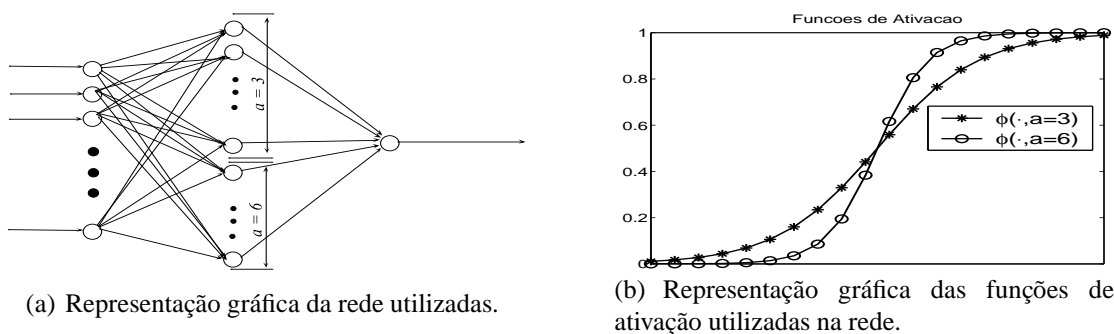
O algoritmo de mínimos quadrados (LMS) aplica uma correção  $\Delta w_{ji}(n)$  ao peso sináptico  $w_{ji}$ , que é proporcional à derivada parcial  $\partial \mathcal{E}(n)/\partial w_{ji}(n)$ , a qual representa um fator de sensibilidade, determinando a direção de busca no espaço de pesos, para o peso sináptico  $w_{ji}$ . O termo  $\mathcal{E}(n)$  é o erro quadrático médio, conforme definido em [Haykin 2001]. Das duas maneiras possíveis de fazer uma correção dos pesos sinápticos, adotaremos a regra delta generalizada, definida por

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha \left[ w_{ji}^{(l)}(n-1) \right] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n), \quad (11)$$

onde  $l$  representa a camada,  $w_{ji}^{(l)}$  é o peso sináptico do neurônio  $j$  da camada  $l$  que é alimentado pelo neurônio  $i$  da camada  $l-1$ ,  $y_i^{(l-1)}(n)$  é o sinal de saída do neurônio  $i$  na camada anterior  $l-1$ , na iteração  $n$ ,  $\eta$  é chamado de taxa de aprendizagem e  $\alpha$  é chamada de constante de momento. Detalhe da dedução de tal regra pode ser encontrado na referência [Haykin 2001].

#### 4. Metodologia

A rede utilizada foi constituída de 20 unidades de entrada, que representa o tamanho do vetor de entrada, uma camada oculta com quarenta neurônios e uma camada de saída com apenas um neurônio. A função de ativação utilizada no neurônio de saída da rede foi a função linear,  $\varphi(v) = v$ . Na camada oculta, foi utilizada a função logística, Eq. (10), com o parâmetro  $a = 3$  para os primeiros 30 neurônios, e para os 10 neurônios restantes foi utilizada a mesma função de ativação, no entanto, com o parâmetro  $a = 6$ , Figura 2.



**Figura 2. Representação gráfica da rede utilizada e das funções de ativação.**

O intervalo físico aceitável para o albedo é  $\varpi_0 \in (0, 1)$ , no entanto, consideramos como valor mínimo e máximo o valor de albedo para águas claras e águas turvas,  $\varpi_0 \in [0.247, 0.833]$ , respectivamente, apresentadas em [Mobley 1994].

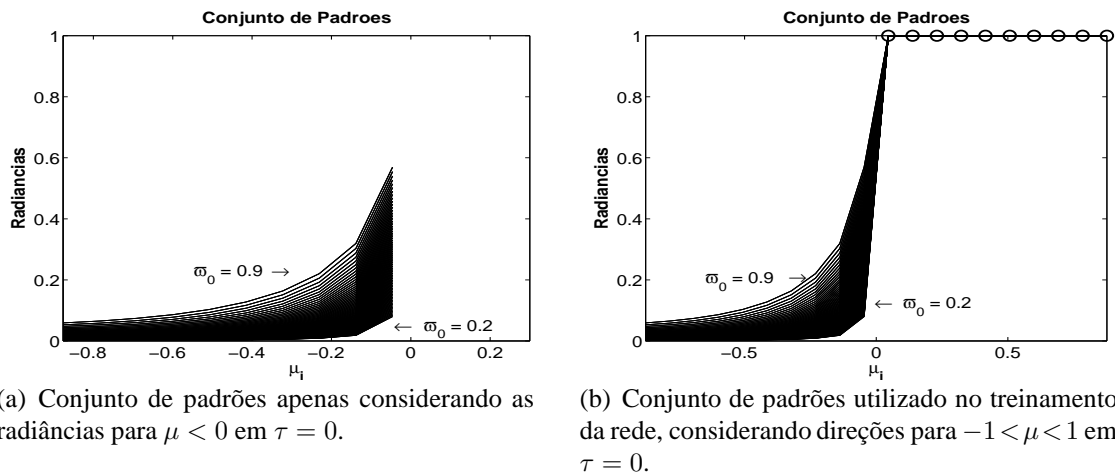
Para testar a metodologia, empregamos dados sintéticos sem ruído (quantidades não representativas da realidade) e dados com ruído.

#### 4.1. Geração de Padrões e Treinamento

##### 4.1.1. Geração de Padrões

Para a geração dos padrões, utilizamos, por conveniência, o intervalo estendido  $\varpi_0 \in [0.20, 0.90]$ , o qual foi discretizado com um tamanho de passo  $\Delta_{\varpi_0} = 0.01$ . Assim,

construímos um vetor com 71 posições  $\vec{\omega}_0 = [0.20, 0.21, 0.22, \dots, 0.89, 0.90]$ . Com estes valores, resolvemos 71 problemas diretos, gerando as radiâncias conforme definido na Seção 2. O número de direções utilizadas na discretização pelo método  $S_N$  foi 150, gerando assim 71 vetores de 150 entradas cada. Ao resolvermos o problema da Seção 2, determinamos as radiâncias na superfície para os valores de  $\mu < 0$ , e no fundo para  $\mu > 0$ . Das 75 direções que se referem aos  $\mu < 0$  na superfície, consideramos apenas 10 no intervalo  $\theta \in (90^\circ, 151.2^\circ]$ , igualmente espaçados, para gerar um dos conjuntos de treinamento da rede, Figura 3(a). O outro conjunto de treinamento gerado, além dessas direções, adicionamos 10 valores de radiância para as direções  $\mu > 0$ , na superfície, tendo assim um vetor com 20 entradas, Figura 3(b).



**Figura 3. Representação gráfica dos conjuntos de padrões (radiâncias).**

Assim formamos o conjunto de treinamento onde os dados de entrada da rede são as radiâncias e os dados de saída são os albedos, associados as radiâncias de entrada.

#### 4.1.2. Treinamento

Observando os dados para treinamento da rede, Figuras 3(a) e 3(b), notamos que as curvas de radiâncias para os valores de albedo próximos a 0.2 são próximas de zero, e a medida que o albedo se aproxima de 0.9 essas curvas tendem a se afastar de zero, para os valores de  $\mu < 0$ . Tendo esse conhecimento, *a priori*, dos padrões de treinamento, podemos escolher funções de ativação que mais se adaptem aos padrões, proporcionando, assim, resultados melhores. De posse dessa informação, adotamos a função logística como função de ativação para os neurônios da camada oculta, com dois valores para o parâmetro  $a$  da Eq. (10).

Para a determinação de tal parâmetro, foram feitos experimentos que levaram a adotar  $a = 3$  para os 30 primeiros neurônios, e  $a = 6$  para os 10 últimos neurônios, na camada oculta, Figura 2(a). Observe, Figura 2, que se  $a = 3$ , a função de ativação é menos íngreme, aproximando-se assim mais das curvas de radiâncias cujos albedos são mais próximos de 0.9, e para  $a = 6$ , a função de ativação é mais íngreme, aproximando-se assim mais das curvas de radiâncias cujos valores do albedo são próximos de 0.2. Dessa

forma, “separamos” o conjunto de treinamento em 2 grupos, e aplicamos uma função de ativação em cada grupo.

O termo “separamos” é empregado aqui no sentido de que, no treinamento, a rede faria essa separação, ou seja, alguns neurônios se especializariam para as radiâncias geradas com albedo próximo a 0.9, e outros neurônios se especializariam para as radiâncias geradas com albedo mais próximo de 0.2.

Já a função de ativação para o neurônio de saída foi linear, pelo fato de que a variação do albedo é linear. Mesmo sabendo que a variação do  $\varpi_0$  em problemas reais não é linear, é possível adotar a linearidade na variação de tal parâmetro, pois seus valores se encontram entre 0.2 e 0.9. Isto facilita a escolha da função de ativação para o neurônio de saída. A apresentação dos padrões de treinamento em ordem aleatória, à rede, faz com que ela aprenda a reconhecer o valor correto.

A taxa de aprendizagem de tal neurônio foi  $8\eta$  para os pesos e  $5\eta$  para o bias, onde os valores de  $\eta$  variaram como apresentado na Tabela 1.

$\eta$	$\alpha$	Épocas
0.01	0.1	1–5000
0.009	0.4	5001–20000
0.009	0.08	20001–30000
0.0005	0.009	30001–35000
0.0001	0.001	35001–70000
0.000005	0.0005	70001–100000

**Tabela 1. Valores de  $\alpha$  e  $\eta$  utilizados no treinamento e o intervalo de épocas.**

O treinamento se deu da seguinte forma: como o número de padrões é limitado, então foram criados mais dois conjuntos de padrões idênticos ao descrito na seção anterior, isto é, o conjunto de padrões de treinamento passou a ter 213 vetores com 20 posições cada, e o mesmo foi feito para o conjunto de treinamento com 10 posições. Adotamos o conjunto de padrões com 20 entradas, Figura 3(b), pelo fato de que tal conjunto, além de apresentar uma maior similaridade com a função de ativação da camada oculta, produziu melhores resultados que o conjunto com apenas dez entradas. Isso demandou um maior tempo computacional para o treinamento da rede, no entanto, os resultados foram satisfatórios o bastante para adotarmos tal conjunto. Para o valor dos pesos iniciais usamos um gerador de números aleatórios, e os valores retornados foram divididos por  $10^9$ , com isso tentamos colocar o mínimo de conhecimento na rede. E, por fim, geramos um vetor contendo 213 números aleatórios entre 1 e 213. Estes valores serviram para embaralhar os padrões de treinamento e após isso, separamos estes em 4 conjuntos e inserimos ruído conforme Tabela 2. Observe que o treinamento da rede foi feito com dados sem ruído para alguns padrões, e para outros com nível de ruído diferentes.

Feito isto, iniciamos o treino da rede, o qual foi feito em quatro etapas utilizando a regra delta generalizada, Eq. (11), onde os valores de  $\alpha$  e  $\eta$  variaram em intervalo de épocas conforme mostrado na Tabela 1.

A primeira etapa consistiu em treinar a rede usando os pesos sinápticos aleatórios, como descrito anteriormente. A segunda, consistiu em um novo treinamento onde o valor

Intervalo	Ruído
1–34	0%
35–71	10%
72–142	5%
143–213	2%

**Tabela 2. Conjunto de padrões para treino da rede na primeira e segunda etapas.**

Intervalo	Ruído
1–10	0%
11–21	10%
22–46	5%
47–71	2%

**Tabela 3. Conjunto de padrões para treino da rede na terceira e quarta etapas.**

dos pesos iniciais foram os pesos retornados do primeiro treinamento, e os valores da taxa de aprendizagem, constante de momento e intervalo de épocas são os mesmos apresentados na Tabela 1. Observe que, quando iniciamos o segundo treinamento, já partimos de um mínimo local, e a partir dele tentamos chegar em um outro ótimo local, variando os valores de  $\eta$  e  $\alpha$ .

Nas etapas 3 e 4 diminuimos o conjunto de padrões para apenas os 71 padrões que são gerados pela discretização do albedo  $\varpi_0$ . O procedimento para treino foi semelhante, ou seja, embaralhamos os dados e então os separamos em 4 conjuntos, conforme descrito na Tabela 3.

Para o treinamento da terceira etapa foram usados os dados apresentados na Tabela 1, e os pesos sinápticos utilizados foram os que retornaram do segundo treinamento, e, para a quarta etapa, os pesos que retornaram da terceira etapa.

Iniciado o treinamento de uma rede neural, é necessário ter critérios para, em algum momento, parar o treinamento. O critério utilizado foi a apresentação do conjunto de teste, conforme descrito na próxima seção, e está apresentação do conjunto de teste se deu no final de cada etapa de treinamento. Paramos o treinamento quando as taxas de acerto começaram a ser inferiores as apresentadas na Seção 5. Como o treinamento foi longo, pode-se esperar que tenha ocorrido *overfitting* [Haykin 2001], no entanto, o presente problema é de domínio fechado<sup>1</sup>, ou seja, os valores para  $\varpi_0$  estão limitados entre 0.2 a 0.9, e como o tamanho do passo utilizado para discretização foi  $\Delta_{\varpi_0} = 0.01$ , torna difícil ocorrer *overfitting*, com isto justificamos o treino em 4 etapas.

Como comentado, se fosse feito mais treinamento na rede, os resultados começariam a ser pobres, talvez, então sim, pelo fato de estar ocorrendo um *overfitting* na rede, fazendo com que ela se especializasse apenas nos padrões de treinamento. Cabe ressaltar também que os parâmetros utilizados no treinamento ( $\alpha$  e  $\eta$ ) foram obtidos através experimentos, nos quais os valores foram sendo variados aleatoriamente.

## 5. Resultados

Após o treinamento da rede, entramos com o processo de ativação que consistiu em primeiramente discretizar o intervalo considerado para o albedo  $\varpi_0$  com um tamanho de passo de  $\Delta_{\varpi_0} = 0.0001$ , gerando assim 7001 valores, e conseqüentemente 7001 problemas diretos, conjunto de teste, com os mesmos parâmetros, na ETR, utilizados na geração dos padrões de treinamento.

<sup>1</sup>Por este motivo que adotamos o próprio conjunto de teste/validação para parar o treinamento.

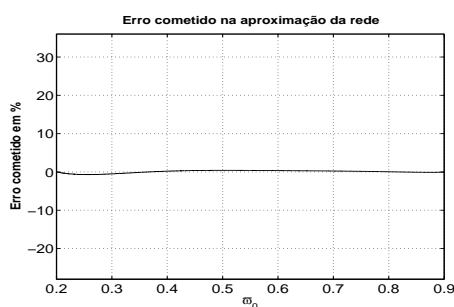


Como primeira validação apresentamos à rede o conjunto de teste sem a presença de ruído. A avaliação do erro cometido pela rede foi baseado no erro relativo, ou seja

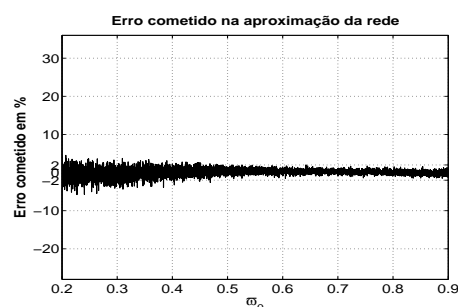
$$\varepsilon = \frac{\varpi_0^{\text{mod}} - \varpi_0^{\text{exato}}}{\varpi_0^{\text{exato}}}, \quad (12)$$

onde  $\varpi_0^{\text{mod}}$  representa o valor de albedo estimado pela rede.

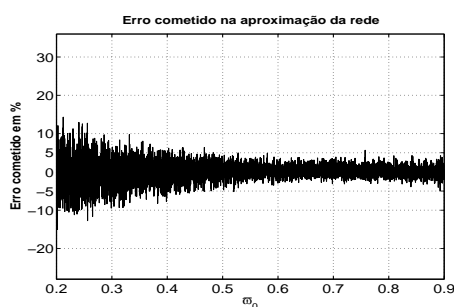
O gráfico do erro cometido pela rede, nos dados sem ruído, é mostrado na Figura 4. Na Figura 5, apresentamos o gráfico do erro relativo quando apresentamos à rede o conjunto de teste corrompido com um nível de ruído de 2%. Na Figura 6, mostramos o mesmo conjunto de teste, no entanto com um nível de ruído de 5%, e, por fim, Figura 7, o conjunto de teste corrompido com um nível de ruído de 10%. Um teste adicional feito, foi utilizar o conjunto de treinamento descrito anteriormente, no entanto com um nível de ruído que não esteve presente no treinamento da rede. O nível de ruído utilizado para este teste foi 7%, e o gráfico que mostra o erro cometido pela rede, neste último teste, é mostrado na Figura 8.



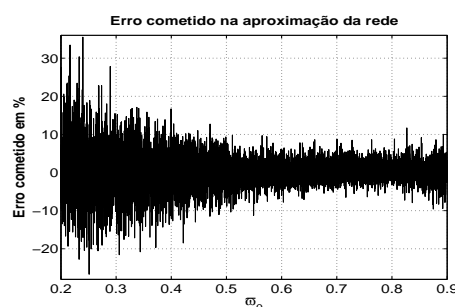
**Figura 4.** Erro para os dados sem ruído.



**Figura 5.** Erro para os dados com 2% de ruído.



**Figura 6.** Erro para os dados com 5% de ruído.

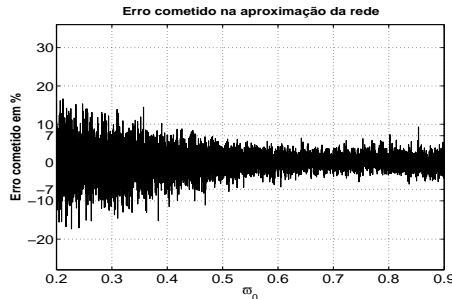


**Figura 7.** Erro para os dados com 10% de ruído.

Observando as figuras percebemos que o erro é maior quando o valor do albedo é inferior a 0.5. Isso deve-se ao fato que, para esses valores de albedo, a “variação” na radiação é pequena, conforme Figura 3, o que torna difícil a classificação correta pela rede.

Para a estimativa da taxa de acerto da rede consideramos todos aqueles valores onde o erro cometido foi inferior ao nível de ruído inserido nos dados. Para o caso onde

não corrompemos os dados com ruído, foi considerado um intervalo de tolerância de  $\epsilon = \pm 0.005$ , ou seja, foi considerado acerto se:  $\varpi_0^{\text{exato}} - \epsilon \leq \varpi_0^{\text{mod}} \leq \varpi_0^{\text{exato}} + \epsilon$ . Na Tabela 4, mostramos as respectivas taxas de acerto para cada caso.



**Figura 8. Erro para os dados com 7% de ruído.**

Ruído %	Acerto
0	100%
2	94,12%
5	94,60%
7	94,93%
10	95,39%

**Tabela 4. Taxa de acerto em cada caso.**

## 6. Conclusões

Ao iniciarmos a busca de solução, usando a rede neural, para o problema em questão, utilizamos apenas uma função de ativação, ou seja, apenas um valor para o parâmetro  $a$ , Eq. (10). Procedendo desta forma, encontramos dificuldade na obtenção de bons resultados. De frente a este problema, buscamos meios alternativos. A idéia que surgiu durante a análise dos padrões de treinamento (informação, *a priori*, disponível no conjunto de padrões), foi a utilização de dois valores distintos para o parâmetro  $a$ , gerando assim, “duas” funções de ativação para a mesma camada de neurônios. O raciocínio utilizado é descrito na Seção 4.1.2.

Cabe ressaltar que: tanto a utilização de “duas” funções de ativação, quanto a utilização de dois valores para a constante  $a$ , é uma metodologia pouco explorada em problemas ligados as redes neurais.

Um comentário importante aqui é em relação aos métodos tradicionais de inversão, por exemplo, Método Quase-Newton, Gradiente Conjugado, Levenberg-Marquardt, entre outros, e à técnica de redes neurais. Para os primeiros métodos é necessário que seja feita uma inversão para cada problema, ou seja, não há uma maneira de fazer uma “generalização”, resolver um problema e a partir deste formar uma estrutura que possa resolver outros problemas. Já a rede neural usada, que apesar de levar um maior tempo para a aprendizagem, após treinada, desde que a formulação do problema direto seja a mesma, conseguiu resolver vários problemas de estimação de albedo em um tempo computacional muito inferior aos métodos tradicionais citados acima. Além disto, os resultados obtidos são satisfatórios, mostrando assim, que a técnica de inteligência computacional pode ser promissora em outros problemas ligados a Óptica Hidrológica.

## 7. Referências

### Referências

Chalhoub, E. (2003). Discrete-ordinates solution for radiative-transfer problems. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 76:193–206.

- Chandrasekhar, S. (1960). *Radiative Transfer*. Dover, New York.
- Cortivo, F. D. (2008). *Reconstrução de Termo Fonte em Ótica Hidrológica Usando Série de Fourier*. Dissertação de mestrado, Programa de Pós-Graduação em Modelagem Computacional (PPMG), Fundação Universidade Federal do Rio Grande, Rio Grande, RS, Brasil.
- Haykin, S. (2001). *Redes Neurais – Princípios e Prática*. Bookman, Porto Alegre, RS, 2ª edição.
- McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. (5):115–133.
- Mobley, C. (1994). *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, Inc, Califórnia.