Learning and Nonlinear Models (L&NLM) – Journal of the Brazilian Neural Network Society, Vol. 10, Iss. 3, pp. 175-180, 2012.

Análise do desempenho de Redes Neurais Artificiais no preenchimento de falhas em séries de precipitação diária

Performance analysis of Artificial Neural Networks in filling gaps in daily precipitation series

Maristela de Salles Silva Almeida [a]
Ricardo Carvalho de Almeida [b]

[a] *Universidade Federal do Paraná*
maristela_salle@yahoo.com
[b] *Universidade Federal do Paraná*
rcalmeida@ufpr.br

Resumo - Uma série de dados de precipitação sem falhas é de grande valia para o desenvolvimento de estudos e modelos em hidrologia. Porém, nem sempre é possível obter estas séries em perfeitas condições. Portanto, são necessários métodos capazes de fazer este preenchimento gerando o resultado mais acurado possível. Este trabalho analisa e compara o desempenho de Redes Neurais Artificiais para o preenchimento de falhas em séries de precipitação diária em duas localidades com regimes de precipitação distintos: em uma região tropical, no Ceará, e em uma região subtropical, no Paraná. Esta comparação é feita através de medidas estatísticas e são utilizados outros dois métodos para a corroboração dos resultados: Método da Distância Inversa e Regressão Linear Múltipla. A pesquisa mostra que as Redes Neurais Artificiais produziram resultados consideravelmente melhores na região Subtropical do que na região Tropical. Esses resultados também são obtidos quando os dados são submetidos ao Método da Distância Inversa e a Regressão Linear Múltipla

Palavras-Chave – redes neurais, precipitação diária, regressão linear, preenchimento de falhas, hidrologia, meteorologia

Abstract - Historical data precipitation with no gaps it's an important piece for the development of models and studies in hydrology, but these series are not always found with this expected condition. Therefore, it's necessary to develop methods with the capacity of filling gaps accurately. This paper analyses and compare the performance of Artificial Neural Networks in filling gaps in daily precipitation series of two different states in Brazil: a tropical region, Ceará, and a tropical region, Paraná. This comparison is made by statistical measures and this research also apply other two methods for filling gaps: Inverse Distance Weighted Method and Multiple Linear Regression. This paper will show that the Artificial Neural Network produces better results in the subtropical region than it produces in the tropical region. These same results are obtained when the Inverse Distance Weighted Method and Multiple Linear Regression are applied in the same data set.

Keywords – artificial neural networks, daily precipitation, linear regression, gap filling, hydrology, meteorology.

1 Introdução

A modelagem de sistemas naturais permite que expressemos o seu comportamento através de fórmulas matemáticas. O sistema de interesse neste trabalho é o sistema hidrológico, que estuda a ocorrência, a distribuição, o movimento e as propriedades das águas da Terra (Serrano, 1997). Mais especificamente, analisaremos a precipitação que, em síntese, é o retorno da água em forma predominantemente líquida, à superfície da Terra. Estações meteorológicas espalhadas pelo mundo são equipadas com pluviógrafos, instrumentos que medem a quantidade total acumulada de chuva durante um certo período de tempo. Estas medições são feitas continuamente, em intervalos que variam de minutos a dias, e esses valores são posteriormente registrados em bancos de dados. Muitos estudos em meteorologia e hidrologia baseiam-se nesses dados e espera-se que estas séries de dados estejam completas e organizadas. Mas, infelizmente, isto não ocorre com todas elas. Muitas vezes estas séries são repletas de falhas que influenciam na acurácia dos resultados obtidos. Muitas podem ser as causas dessa falta de informação seqüencial: a falta de manutenção do medidor; quebra do medidor gerando necessidade de troca; problemas na medição do aparelho; perda de dados; falta de observadores qualificados; e até mesmo falta de fundos para

manter a continuidade das medições. Para que estas séries de dados possam ser utilizadas elas devem ser avaliadas e tratadas para que suas falhas possam ser corrigidas.

Estimar valores para séries de precipitação diária é um grande desafio devido à não linearidade e à irregularidade geralmente observadas nessas séries. Uma solução é encontrar uma função matemática que se adapte ao comportamento da série. Por isso, optamos por testar o desempenho das Redes Neurais Artificiais, já que é demonstrado que uma rede com pelo menos uma camada intermediária é, teoricamente, capaz de aproximar qualquer função não linear (Fausset, 1994). Este trabalho visa comparar o desempenho de Redes Neurais Artificiais para a estimativa de valores faltantes em séries de precipitação de diária de uma estação no Ceará e de outra no Paraná. Após está análise, foram realizadas as mesmas comparações utilizando o Método da Distancia Inversa e a Regressão Linear Múltipla, podendo assim confirmar os resultados obtidos pela RNA.

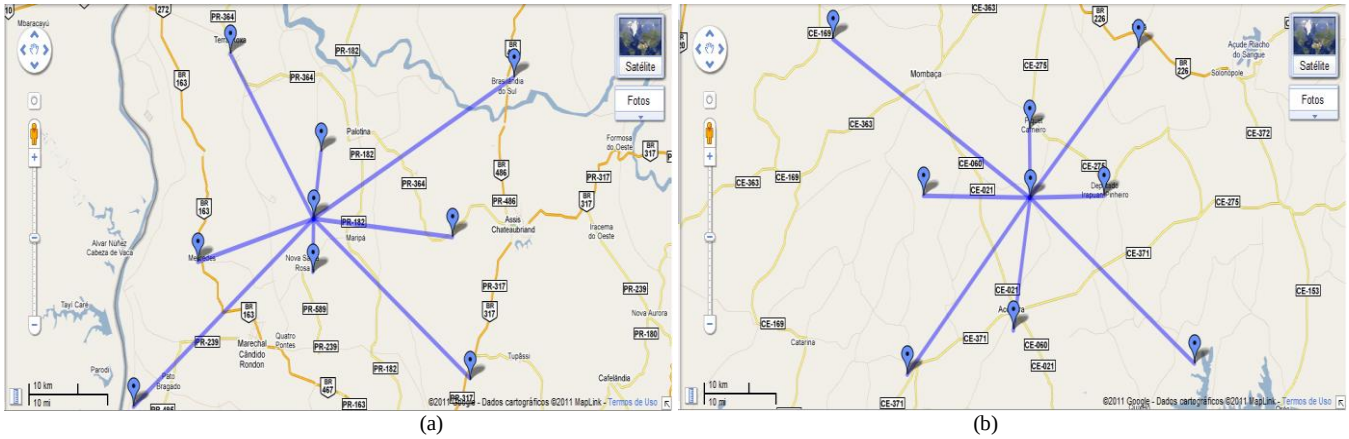
2 Rede Neural Artificial

Uma Rede Neural Artificial (RNA) é um sistema de processamento de informações cujo desempenho tem características semelhantes ao funcionamento de redes neurais biológicas (Barreto, 2002 apud Essenfelder, 2009). As RNAs estão inseridas na Inteligência Artificial, área que busca, através de técnicas inspiradas no comportamento do cérebro, o desenvolvimento de sistemas inteligentes que imitem aspectos do comportamento humano, tais como: aprendizado, percepção, raciocínio, evolução e adaptação (Silva, 2007).

Podemos diferenciar a RNA de outros modelos matemáticos pela sua capacidade de aprender uma determinada tarefa e generalizar para certo problema de mesma natureza. A RNA é uma estrutura de camadas compostas de neurônios artificiais que estão ligados através de conexões. O neurônio artificial simula, através de uma estrutura lógico-matemática, o neurônio biológico. Ele recebe sinais de entrada e multiplica cada um deles por seu respectivo valor de sinapse. É calculada a soma destes valores e neste resultado é aplicada uma função de ativação. Este último valor obtido é enviado como valor de entrada para os neurônios da próxima camada

3 Metodologia

Foram selecionadas duas regiões com regimes de precipitação distintos para a pesquisa: Ceará e Paraná. Uma estação em cada região foi selecionada para ser a estação cuja série possuía dados faltantes, chamada de estação central. Foram utilizadas como variáveis de entrada da RNA dados de precipitação diária de outras estações circunvizinhas. Estas estações preditoras foram escolhidas de tal forma que tivessem distribuições geográficas similares ao redor das estações centrais de cada localidade.



(a) (b)

Figura 1 – Localização geográfica das estações pluviométricas no Paraná (a) e no Ceará (b)

Conforme mostra a Figura 1, as estações estão distribuídas aproximadamente ao Norte (N), Leste (E), Sul (S), Oeste (W), Nordeste (NE), Sudeste (NE), Sudoeste (SW) e Noroeste (NW). Nas quatro primeiras direções se encontram estações mais próximas (10-25 km da estação central) e nas outras quatro estão localizadas estações mais distantes (30-45 km da estação central).

Os dados de precipitação de cada estação foram obtidos na página da Agência Nacional de Águas. Eles foram filtrados separadamente em cada região de tal forma que fossem utilizados somente os dados de dias em que existisse o registro da precipitação em todas as oito estações preditoras e na estação central. Os dados obtidos estavam distribuídos e quantificados da seguinte forma:

- 6580 dias com valores registrados simultaneamente nas 9 estações do Paraná, distribuídos de janeiro de 1976 a dezembro de 1994; e

- 1265 dias com valores registrados simultaneamente nas 9 estações do Ceará, distribuídos de junho de 1973 a abril de 1988;

Devido à grande diferença na quantidade de dados obtida em cada estado, optou-se por fazer uma redução dos dados do Paraná, para que assim os resultados pudessem ser comparáveis. Foram selecionados apenas os dados registrados no intervalo de tempo com dados disponíveis no Ceará, obtendo assim 1265 valores registrados para ambos os estados. Vale ressaltar que os dados não são dos mesmos dias nem dos mesmos meses rigorosamente, mas apenas do mesmo intervalo de anos. Infelizmente não foi possível a utilização de dados mais recentes, pois estes não foram encontrados. Apenas algumas estações continham o registro de dados a partir do ano 2000, porém ao ser realizado o filtro, estes dados foram excluídos por não constarem em todas as estações.

As RNA's foram testadas para quatro conjuntos de variáveis de entrada diferentes como variáveis preditoras: utilizando as oito estações em torno da central; utilizando apenas as quatro estações mais próximas (N, E, S e W); utilizando as quatro estações mais distantes (NE, SE, SW, e NW); e utilizando as três estações mais próximas de cada região (N, E e S no Ceará e N, S e W no Paraná). Esta três estações foram obtidas após a utilização do método de seleção de variáveis screening regression no conjunto de dados, visto que em alguns casos, utilizar todas as variáveis preditoras disponíveis pode resultar em uma estimativa ruim do preditando (Wilks, 2006).

Todos os dados foram normalizados, de forma a variarem entre ±1. Como a quantidade de variáveis preditoras foi diferente em cada conjunto, as RNA's tiveram diferentes arquiteturas na camada de entrada, sendo o número de neurônios nessa camada igual à quantidade de variáveis preditoras para cada caso. Para todos os conjuntos foi utilizada apenas uma camada intermediária, com quatro neurônios com função de ativação tangente hiperbólica, e na camada de saída apenas um neurônio com função de ativação linear.

No treinamento das RNA's deste trabalho foi utilizado o Algoritmo Backpropagation (Algoritmo da Retropropagação do Erro), que busca minimizar a soma dos erros médios quadráticos das diferenças dos valores observados e dos valores estimados pela RNA, utilizando a técnica de busca do gradiente descendente. Para a realização dos experimentos, os dados disponíveis foram aleatóriamente divididos em três conjuntos, de acordo com os seguintes percentuais aproximados: treinamento (70%); teste (20%), para controle de overfitting; e validação (10%), que não eram utilizados no treinamento, simulando os dados faltantes na série de precipitações. O treinamento foi feito por batelada, ou seja, a atualização dos pesos era realizada somente após a realização de toda uma época de treinamento à RNA. Ao final do treinamento, a rede recebia dos dados de entrada do conjunto de validação e comparava as saídas de rede com os dados efetivamente observados. A partir dessa validação, eram obtidas as diversas medidas estatísticas para análise do desempenho da rede.

Devido à alta variabilidade e aleatoriedade do fenômeno de precipitação diária, foram realizados quatro experimentos para cada um dos quatro conjuntos de dados que foram submetidos a RNA. Estes experimentos se diferenciam por serem formados por diferentes seleções aleatórias de registros para os conjuntos de treinamento, teste e validação. Assim, é possível fazer a análise de quanto o resultado da RNA pode ser alterado devido à utilização de um conjunto de dados "ruim" para a estimativa.

Para a análise dos resultados foram utilizadas quatro medidas estatísticas escalares (Wilks, 2006): Coeficiente de Correlação de Pearson (R), Erro Médio (ME), Erro Médio Absoluto (MAE) e Raiz do Erro Médio Quadrático (RMSE). Suas expressões são apresentadas nas Eq. (1) a (4), respectivamente. Em todas as equações x_i representa o valor estimado, y_i representa o valor observado, s_x é o desvio padrão dos valores estimados, s_y é o desvio padrão dos valores observados, $\bar{x}$ é a média dos valores estimados, $\bar{y}$ é a média dos valores observados e n é o número de registros.

$$R = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})(y_i-\bar{y})}{(s_x s_y)} \quad (1)$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) = \bar{x} - \bar{y} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \quad (3)$$

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

Também foi utilizada a razão $RMSE/s_y$, onde seus termos seguem as mesmas definições apresentadas no parágrafo anterior. Esta razão foi utilizada com o objetivo de sintetizar as informações da variabilidade do erro e da variabilidade natural do fenômeno de precipitação de diária em uma única medida. Quanto menor a variabilidade do erro ($RMSE$) em relação à variabilidade do fenômeno (s_y) mais preciso é o método. Quando a variabilidade do erro é muito alta, e até maior que a variabilidade natural do fenômeno, esta razão irá se aproximar ou ultrapassar o valor 1.0.

Com o intuito de analisar comparativamente o desempenho das RNA's, foram testados outros dois métodos para o preenchimento das falhas nos mesmos conjuntos de dados. A Regressão Linear Múltipla (RLM), que é o caso mais geral da Regressão Linear Simples (RLS), foi o primeiro deles. Sua forma geral é apresentada na Eq. (5), onde k é a quantidade de variáveis preditoras, $b_0 \dots b_k$ são os coeficientes da regressão, $x_1 \dots x_k$ são os valores das variáveis preditoras e ε é o erro entre o valor observado e o valor estimado.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + \varepsilon \tag{5}$$

Os coeficientes $b_0 \dots b_k$ representam a mudança esperada na resposta y quando a variável x_i é alterada e as outras variáveis x_j ($i \neq j$) são mantidas constantes. Para o desenvolvimento deste modelo foi utilizado em cada experimento o mesmo conjunto de treinamento gerado e utilizado na RNA, definindo assim os coeficientes da regressão. Após esse procedimento, foram calculadas as estimativas para o mesmo conjunto de validação utilizado na RNA e sobre o resultado gerado foram calculadas as medidas estatísticas definidas.

O segundo método utilizado para a estimativa foi o Método da Distância Inversa (MDI), que é um dos métodos mais utilizado para estimar valores faltantes em séries de precipitação no campo da hidrologia. É necessário conhecer a distância das estações selecionadas como preditoras até a estação central. O cálculo é simples, feito por meio de uma média ponderada pela distância da estação preditora em relação à central, mostrada na Eq. (6), onde n é o número de estações selecionadas para o cálculo, θ_m é o valor de precipitação a ser estimado na estação m (estação central), d_{mi} é a distância da estação m até a estação i e k é a fricção da distância (valor igual a 2.0).

$$\theta_m = \frac{\sum_{i=1}^{n} \theta_i \, {d_{mi}}^{-k}}{\sum_{i=1}^{n} {d_{mi}}^{-k}} \tag{6}$$

Apesar de muito utilizados, deve-se ter cautela ao utilizar métodos geográficos como o MDI, que dependem apenas da disposição relativa das estações, pois essas devem estar em regiões climatologicamente semelhantes, senão os resultados poderão ser seriamente afetados. Por exemplo, consideremos duas estações próximas, porém separadas por um divisor topográfico importante como a Serra do Mar. O comportamento das chuvas em cada lado da serra será bem distinto devido à precipitação orográfica, e utilizar uma estação de um lado do divisor para estimar o valor de chuva em uma estação do outro lado pode produzir erros significativos nas estimativas (Tucci, 2007). Este método foi aplicado diretamente sobre o conjunto de validação utilizado em cada experimento da RNA, tendo suas estatísticas calculadas sobre os resultados gerados.

4 Resultados

Os resultados obtidos para cada conjunto de dados foram resumidos e são apresentados nas Tabelas 1 e 2. O valor de cada medida na tabela é a média aritmética dos quatro valores obtidos em cada experimento realizado. Ao comparar os resultados percebemos que, para todos os conjuntos, o desempenho das RNA's no Paraná foi superior ao Ceará. Nos Conjuntos de dados 1 do Paraná, que tem as oito estações como variáveis preditoras em cada região, obtemos um valor de R claramente superior ao valor obtido no Ceará. O valor obtido na razão $RMSE/s_y$ confirma a conclusão, pois no Ceará ela ficou muito próxima de 1.0, mostrando que a variabilidade do erro ficou muito próxima da variabilidade do fenômeno, e assim percebemos que o método apresenta um baixo desempenho. Apenas o ME obtido no Ceará foi melhor que o seu valor obtido no Paraná. Porém, devido às outras medidas observadas, é possível afirmar que o desempenho no Paraná foi superior ao desempenho no Ceará.

Para os Conjuntos de dados 2, 3 e 4 - formados respectivamente pelas quatro estações mais próximas, pelas quatro estações mais distantes e pelas três estações selecionadas pelo *screening regression* em cada região - podemos fazer as mesmas observações sobre o R, o MAE e a razão $RMSE/s_y$ dos Conjuntos de dados 1. Nota-se que nestes conjuntos o ME obtido no Paraná também foi superior ao obtido no Ceará, tornando todas as medidas melhores naquele estado. Vale destacar que medidas obtidas em cada região ficaram relativamente próximas entre si. Porém, no Conjunto de dados 3 do Ceará, o R obtido ficou bem abaixo dos valores de R obtidos nos demais conjuntos.

Da análise dos resultados, observa-se, também, que a escolha das estação preditoras com o emprego do *screening regression* (Conjuntos de dados 4) produziu resultados comparáveis aos obtidos utilizando-se todas as estações (Conjuntos de dados 1), indicando que este método permite uma otimização no uso dos dados disponíveis, possibilitando economia computacional no treinamento das RNA's, sem perda de qualidade significativa nos resultados.

Tabela 1 – Médias dos valores obtidos nos experimentos de cada conjunto de dados do Paraná utilizando Redes Neurais Artificiais

Medida Estatística	Conjunto de dados 1	Conjunto de dados 2	Conjunto de dados 3	Conjunto de dados 4
R	*0.882986*	*0.852907*	*0.814195*	*0.863048*
ME	*-0.20162*	*-0.14741*	*-0.04793*	*0.268654*
MAE	*2.428510*	*2.534431*	*2.982566*	*2.327763*
RMSE/Sy	*0.474978*	*0.524070*	*0.609202*	*0.520129*

Tabela 2 – Média dos valores obtidos nos experimentos de cada conjunto de dados do Ceará utilizando Redes Neurais Artificiais

Medida Estatística	Conjunto de dados 1	Conjunto de dados 2	Conjunto de dados 3	Conjunto de dados 4
R	*0.466545*	*0.4357937*	*0.2940159*	*0.4980591*
ME	*0.037902*	*0.1546695*	*0.4540216*	*0.3051392*
MAE	*3.175573*	*3.1233486*	*3.2213809*	*3.1972083*
RMSE/Sy	*0.886499*	*0.9114127*	*0.9687764*	*0.9217437*

Ao utilizar a RLM e o MDI nos mesmos conjuntos de dados das duas regiões, obtemos os resultados mostrados nas Tabelas 3 e 4, respectivamente. Nota-se que os resultados obtidos por ambos os métodos seguem o mesmo padrão dos resultados obtidos pelas RNA's, com estimativas consideravelmente superiores para a estação localizada no Paraná em comparação à estação do Ceará. Vale destacar que os valores das medidas para a RLM ficaram bem próximos dos valores para as RNA's, mostrando que os métodos tem desempenhos semelhantes. Já os valores obtidos no MDI ficaram abaixo dos resultados obtidos pelas RNA's na maioria dos conjuntos. Ao analisar os resultados gerados pelo MDI dos Conjuntos de dados 3 e 4 do Ceará, percebe-se que apenas os valores de R foram ligeiramente superiores aos obtidos pela RNA, porém os valores verificados para a razão $RMSE\,/s_y$ foram superiores a 1.0, mostrando que os erros característicos desse método ultrapassam a variabilidade típica da precipitação diária naquela região.

Uma possível razão para o melhor desempenho, em termos gerais, das RNA's e da RLM no Paraná em comparação aos resultados obtidos no Ceará está relacionada às características da precipitação em cada região. No Paraná, onde predomina um regime climatológico subtropical, em que sistemas meteorológicos de escala sinótica (com escala espacial na ordem de 1000 km) afetam a região, a precipitação local é, em geral, organizada pelas circulações de maior escala, associadas a sistemas frontais. Daí, a precipitação em diferentes estações pluviométricas distantes na ordem de até 50 km, como as empregadas neste estudo, apresentarem registros de precipitação mais correlacionados, pois os grandes sistemas meteorológicos afetam essas estações durante um mesmo período de tempo. Por outro lado, no Ceará, onde predomina um regime tropical, raramente afetado por sistemas meteorológicos de grande escala como ocorre no Paraná, a precipitação desenvolve-se, geralmente, forçada por condições locais, com escala espaciais na ordem de poucas dezenas de quilômetros. Dessa forma, na ausência de estruturas de maior escala para organizar a precipitação, cada região desenvolve independentemente sua precipitação, reduzindo, portanto, a correlação entre as precipitações registradas nas diferentes estações pluviométricas. Uma vez que ambos os métodos buscam expressar relações entre as variáveis preditoras e os respectivos preditandos, espera-se que na região onde essas relações sejam estatisticamente mais significativas os métodos tenham um desempenho melhor do que em regiões onde tais relações sejam mais fracas.

Tabela 3 - Médias dos valores obtidos nos experimentos de cada conjunto de dados do Paraná e Ceará utilizando Regressão Linear Múltipla

Medida Estatística	Conjunto de dados 1 - PR	Conjunto de dados 2 - PR	Conjunto de dados 3 - PR	Conjunto de dados 4 - PR	Medida Estatística	Conjunto de dados 1 - CE	Conjunto de dados 2 - CE	Conjunto de dados 3 - CE	Conjunto de dados 4 - CE
R	*0.881403*	*0.850177*	*0.816151*	*0.861242*	**R**	*0.494076*	*0.443011*	*0.298791*	*0.508470*
ME	*-0.178899*	*-0.273225*	*-0.193781*	*0.029812*	**ME**	*-0.20739*	*-0.254011*	*-0.617776*	*0.074454*
MAE	*2.261705*	*2.349586*	*2.687886*	*2.050758*	**MAE**	*2.808660*	*2.74440*	*2.281949*	*2.761740*
RMSE/Sy	*0.472930*	*0.528664*	*0.599078*	*0.524384*	**RMSE/Sy**	*0.868247*	*0.914551*	*0.97910*	*0.925759*

Tabela 4 - Médias dos valores obtidos nos experimentos de cada conjunto de dados do Paraná e Ceará utilizando o Método da Distância Inversa

Medida Estatística	Conjunto de dados 1 - PR	Conjunto de dados 2 - PR	Conjunto de dados 3 - PR	Conjunto de dados 4 - PR	Medida Estatística	Conjunto de dados 1 - CE	Conjunto de dados 2 - CE	Conjunto de dados 3 - CE	Conjunto de dados 4 - CE
R	*0.820314*	*0.742448*	*0.803335*	*0.818300*	**R**	*0.316709*	*0.366203*	*0.329693*	*0.512260*
ME	*0.123385*	*0.193816*	*0.116199*	*0.372130*	**ME**	*0.306744*	*1.010538*	*0.71302*	*1.073060*
MAE	*2.784685*	*2.989362*	*2.841477*	*2.377696*	**MAE**	*3.371459*	*3.598478*	*2.905126*	*3.381495*
RMSE/Sy	*0.582512*	*0.725914*	*0.605635*	*0.625663*	**RMSE/Sy**	*1.053558*	*1.153595*	*1.118819*	*1.140995*

5 Conclusões

Este trabalho utilizou Redes Neurais Artificiais para estimar dados faltantes em séries de precipitação diária de duas estações pluviométricas, uma localizada em uma região subtropical, no estado do Paraná; e de outra em uma região tropical, no estado do Ceará. Através de medidas estatísticas conclui-se que os resultados apresentados para a região subtropical foram consideravelmente superiores aos resultados obtidos na região tropical. Foram utilizados os métodos de Regressão Linear Múltipla e o Método da Distância Inversa nos mesmos conjuntos de dados apresentados às RNA's. Nesses experimentos verificou-se que, para este problema e dados específicos, as RNA's e a RLM tiveram desempenhos comparáveis, enquanto o MDI teve um desempenho bastante inferior ao desses dois métodos. Finalmente, verificou-se que o emprego do método *screening regression* de seleção de variáveis permitiu que com o uso de apenas 3 preditores fossem obtidos resultados de qualidade comparável aos produzidos com 8 preditores.

Referências Bibliográficas

Agência Nacional de Águas – ANA – http://hidroweb.ana.gov.br/

Barreto, J. M., **Introdução às Redes Neurais Artificiais**, Laboratório de Conexionismo e Ciências Cognitivas – UFSC – Departamento de Informática e de Estatística (2002)

Essenfelder, A. H., **Previsão de Curto Prazo da Vazão de um Rio Utilizando Redes Neurais Artificiais**, Trabalho de Conclusão de Curso, UFPR (2009)

Fausset, L. V., **Fundamental of Neural Networks: architectures, algorithms, and applications**, Prentice Hall (1994)

Serrano, S. E., **Hydrology for Engineers, Geologists, and Environmental Professionals**, HydroScience Inc (1997)

Silva, N. B., **Aplicação de Métodos Estatísticos e Redes Neurais Artificiais no Pós-Processamento de Produtos de Previsão Numérica de Tempo**. Tese de Mestrado, COPPE/UFRJ (2007)

Tuccci, C. E. M., **Hidrologia: Ciência e Aplicação**, 4ª Edição, Editora da UFRGS/ABRH (2007)

Wilks, D. S., **Statistical Methods in the Atmosferic Sciences**, 2ª Edição, Elsevier Inc (2006)