



MODELAGEM CHUVA-VAZÃO MENSAL UTILIZANDO REDES NEURAIAS ARTIFICIAIS

Aline Bernarda Debastiani^{1}; Ricardo Dal'Agnol da Silva²; Víctor Luís Padilha³; Eder Alexandre Sá⁴; Sílvio Luís Rafaeli Neto⁵; Thiago Alves Antunes¹*

Resumo – Redes Neurais Artificiais (RNAs) são modelos empíricos de base numérica, empregados como alternativas aos modelos hidrológicos de base física, devido a estes comumente serem complexos e por requerer entendimento da interação entre as variáveis hidrológicas envolvidas. Uma das dificuldades no uso da estrutura Multi-Layer Perceptron (MLP) na modelagem hidrológica, como técnica em RNA, é a identificação do conjunto de variáveis de entrada que melhor representem o comportamento do sistema hidrológico de uma bacia hidrográfica. Por outro lado, há uma carência de estudos sobre quais combinações de vetores de entrada devem ser utilizados na modelagem hidrológica com RNAs, bem como a contribuição individual de cada um. O objetivo do presente estudo foi de avaliar a contribuição relativa de algumas variáveis meteorológicas na representação do processo chuva-vazão mensal, em uma bacia hidrográfica localizada no Planalto Serrano de Santa Catarina. Os resultados demonstram a importância da média mensal da ETP como uma variável relevante ao processo de modelagem na bacia estudada.

Palavras-Chave – simulação mensal, multi-layer perceptron, bacia do Alto Canoas.

RAINFALL-RUNOFF MODELLING USING ARTIFICIAL NEURAL NETWORKS

Abstract – Artificial Neural Networks (ANN) are empirical-based numeric models, used as alternatives to physical-based hydrological models due to its complexity and requirement to understanding the interaction between hydrological variables. One of the difficulties in using the Multi-Layer Perceptron (MLP) structure in hydrological modeling, as an ANN technique, is the identification of a subset of input variables that best represent the hydrological system behavior of a watershed. Moreover, there is a lack of studies on which combinations of input vectors should be used in hydrological modeling with ANNs, as well as the individual contribution of each variable. The aim of this study was to assess the relative contribution of some meteorological variables in the representation of monthly rainfall-runoff process in a watershed located in the Planalto Serrano of Santa Catarina. The results demonstrate the importance of the monthly average ETP as a relevant variable to the modeling process in the studied watershed.

Keywords – Monthly simulation, multi-layer perceptron, Alto Canoas watershed

¹Mestrando do Programa de Pós-graduação em Engenharia Florestal, Universidade do Estado de Santa Catarina – CAV/UEDESC. Av. Luiz de Camões, 2090 – Conta Dinheiro – Lages/SC. Bolsista FAPESC (1º autor). E-mail: aline.debastiani@gmail.com; thiagoaa53@hotmail.com

²Engenheiro Ambiental, Mestre em Sensoriamento Remoto pelo Instituto Nacional de Pesquisas Espaciais (INPE), Bolsista PCI-DC no INPE, São José dos Campos-SP. E-mail: ricds@hotmail.com

³Mestrando do Programa de Pós-Graduação em Engenharia Ambiental, Universidade Federal de Santa Catarina – PPGA/UFSC, Caixa Postal 476, CEP 88040-900, Florianópolis-SC, Brasil. Bolsista CAPES. E-mail: victor.padilha@posgrad.ufsc.br.

⁴Doutorando do Programa de Pós-graduação em Ciência do Solo, Universidade do Estado de Santa Catarina – CAV/UEDESC. Av. Luiz de Camões, 2090 – Conta Dinheiro – Lages/SC. Bolsista CAPES. E-mail: eder.alexandre.ef@hotmail.com.

⁵Universidade do Estado de Santa Catarina – Professor do Departamento de Engenharia Ambiental – CAV/UEDESC. Av. Luiz de Camões, 2090 – Conta Dinheiro – Lages/SC. E-mail: silvio.rafaeli@udesc.br



1. INTRODUÇÃO

Redes Neurais Artificiais (RNAs) são modelos empíricos de base numérica, empregados como alternativas aos modelos hidrológicos de base física, devido a estes comumente ser complexos e por requerer entendimento da interação entre as variáveis hidrológicas envolvidas. Para Solomatine e Ostfeld (2008), as RNAs apresentam soluções mais práticas e rápidas, em um mapeamento direto de dados de entrada em dados de saída, cabendo ao algoritmo determinar as relações entre as variáveis.

Dentre as RNAs, o método *Multi-Layer Perceptron* (MLP) é o mais utilizado, e é composto estruturalmente por uma camada de entrada, uma camada oculta e outra de saída (FU, 1994). É formado por um arranjo de neurônios interconectados que trabalham em paralelo para capturar o padrão de um conjunto de dados de entrada, a fim de reproduzir o padrão do conjunto de dados de saída. Estes padrões são utilizados posteriormente para estimar ou simular os dados de saída a partir de um novo conjunto de dados de entrada (HAYKIN, 2001). As MLPs tem sido cada vez mais utilizadas na hidrologia devido a sua capacidade de modelar sistemas não-lineares, sem a necessidade de fazer quaisquer suposições sobre sua interação (RIAD *et al.*, 2004).

Uma das dificuldades no uso da MLP na modelagem hidrológica é a identificação do conjunto de variáveis de entradas que melhor representem o comportamento do sistema hidrológico de uma bacia hidrográfica, visto que a MLP é um modelo empírico e não tem significado físico. Não existem estudos suficientes na literatura sobre quais combinações de vetores de entrada devem ser utilizados na modelagem hidrológica com redes neurais, bem como a contribuição individual de cada um. Pode-se considerar uma MLP ótima quando esta consegue modelar o sistema utilizando o menor número de vetores de entrada, portanto, eliminar variáveis insignificantes implica em menor custo na aquisição de dados,

Usualmente as MLPs são tratadas como caixas-pretas, em que a contribuição individual de cada variável de entrada não é conhecida. Apesar disso, estudos têm mostrado avanços no entendimento da relação entre os dados de entrada e saída a partir das RNAs (GARSON, 1991; GOZLAN *et al.*, 1999; OLDEN e JAKSON, 2002; SILVA, 2014). O algoritmo de Garson (1991) pondera a contribuição dos vetores de entrada utilizando valores absolutos dos pesos das conexões sinápticas de entrada pelos pesos das conexões sinápticas de saída. Desta forma, é possível analisar a contribuição relativa de um vetor de entrada sobre o vetor de saída.

O objetivo do presente estudo é de avaliar a contribuição relativa de algumas variáveis meteorológicas na representação do processo chuva-vazão mensal, em uma bacia hidrográfica localizada no Planalto Serrano de Santa Catarina.

2. MATERIAL E MÉTODOS

2.1 Área de estudo

O estudo foi realizado na cabeceira principal da bacia hidrográfica do Rio Canoas, denominada bacia do Alto Canoas (Figura 1). Esta área foi escolhida por sua relevância hidrogeológica para a bacia do Rio Canoas, maior bacia hidrográfica do estado de Santa Catarina. Nela se encontra a nascente brasileira do Rio Uruguai e é considerada uma importante Zona de Recarga Direta do Sistema Aquífero Guarani.

A bacia do Alto Canoas compreende uma área de 1979 km², cujas altitudes variam de 839 a 1816m e onde predominam Neossolos Litólicos e Cambissolos Húmicos (EMBRAPA, 2006). Abrange os municípios de Urubici, Rio Rufino e Bom Retiro.

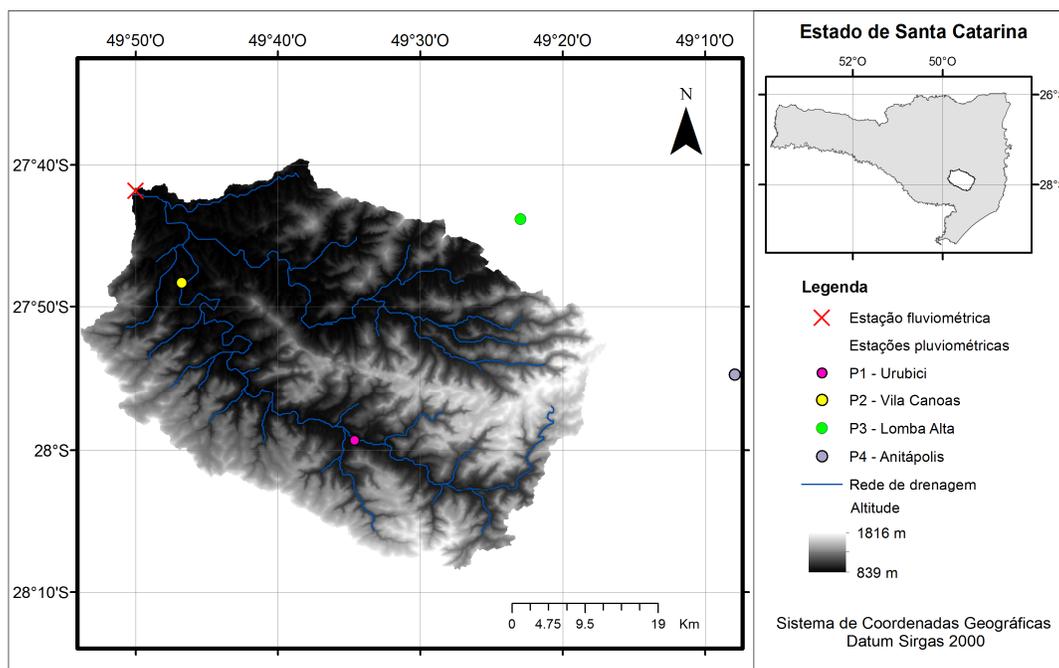


Figura 1 – Localização da bacia hidrográfica do Alto Canoas, no estado de Santa Catarina, e as estações pluviométricas e fluviométrica utilizadas neste estudo.

As estações pluviométricas de Urubici (P1), Vila Canoas (P2), Lomba Alta (P3) e Anitápolis (P4) respondem 47, 34, 16 e 3% de contribuição relativa à bacia (Figura 1). No exutório da bacia está em operação a estação fluviométrica Rio Bonito, gerenciada pela Agência Nacional de Águas (ANA).

2.2 Aquisição e combinação dos vetores de entrada

O período de dados abrangeu 01/01/1980 a 31/12/1989. Utilizaram-se altura precipitada acumulada mensal e vazão média mensal, adquiridos no portal Hidro Web, e ETP média mensal, calculada pelo método de Thornthwaite. A ETP foi calculada com os dados médios mensais da estação climatológica de Lages, localizada cerca de 75 km da bacia, cedidos pela EPAGRI/INMET.

As séries históricas mensais de precipitação, ETP e vazão foram divididas em três subconjuntos amostrais, compostos pelas faixas de 50, 25 e 25% do período, com o objetivo de realizar o treinamento, validação e teste da MLP. O conjunto de vetores de treinamento foi utilizado para o aprendizado propriamente dito, pelo qual ocorre o ajuste dos pesos sinápticos da MLP e implica no sucesso ou fracasso da rede. O processo de validação ocorreu simultaneamente com o processo de treinamento, sobre os vetores de validação. Este processo foi adotado para identificar o momento de parada do processo de aprendizado a fim de evitar o super-ajustamento. O processo de teste realizou a simulação propriamente dita, cujo resultado evidenciou o desempenho da modelagem.

O experimento foi delineado com 10 combinações de vetores de entrada, sendo: 1) P1 e P2; 2) P1, P2, P3, P4; 3) P1, P2, P3, P4, ETP; 4) P1, P2, Q_{T-1} ; 5) P1, P2, P3, P4, Q_{T-1} ; 6) P1, P2, P3, P4, ETP, Q_{T-1} ; 7) PP1, PP2, PP3, PP4, ETP, Q_{T-1} ; 8) P1, P2, ETP; 9) μP -ETP, Q_{T-1} ; 10) ETP. Nestas combinações tem-se que ETP corresponde a evapotranspiração potencial média mensal; PP se refere à precipitação média mensal ponderada pela contribuição relativa de cada estação pluviométrica na área da bacia; Q_{T-1} se refere à vazão média mensal observada no mês anterior (T-



1) e μP -ETP se refere a média da precipitação registrada em todas as estações pluviométricas subtraindo-se o valor da ETP.

2.3 Simulação da resposta hidrológica

A técnica de redes neurais artificiais empregada foi a *Multi-Layer Perceptron* (MLP), empregada no aplicativo MATLAB 2014a. A arquitetura da MLP foi composta por três camadas: camada de entrada, uma camada oculta e a de saída. O algoritmo de aprendizado utilizado foi o *Levenberg-Marquardt* (HAGAN e MENHAJ, 1994), com função de ativação tangente sigmoideal e taxa de aprendizado fixado em 0,01.

O número de neurônios na camada oculta variou conforme sugerido por Heath (2010), onde essa estimativa deve preservar uma relação de 10 vezes mais equações de treino do que de pesos. A partir desse número de neurônios, testou-se uma variação de ± 5 neurônios, buscando uma arquitetura que pudesse produzir a melhor modelagem. Este procedimento foi realizado para as 10 combinações de vetores de entrada. Antes do processo de treinamento efetuou-se a normalização dos dados em Z (média 0 e variância 1).

2.4 Avaliação do modelo

O desempenho da modelagem para as diferentes combinações de vetores de entrada foi avaliado pelo índice de eficiência de Nash e Sutcliffe (NS) (1970), coeficiente de determinação (R^2) e Raiz do Erro Médio Quadrático (RMS), onde as vazões simuladas foram comparadas as vazões observadas.

De acordo com Machado e Vettorazzi (2003) o índice de eficiência NS é um dos critérios estatísticos mais importantes para avaliar o ajuste do modelo hidrológico. Silva e Mello (2005) consideram que NS maior que 0,75 é bom, entre 0,36 e 0,75 é aceitável e inferior a 0,36 inaceitável.

A contribuição relativa individual dos vetores de entrada foi realizada pelo índice CR de Garson (1991) (Equação 1).

$$CR_{ik} = \frac{\sum_{j=1}^L \left(\frac{w_{ij}}{\sum_{r=1}^N w_{rj}} v_{jk} \right)}{\sum_{i=1}^N \left(\sum_{j=1}^L \left(\frac{w_{ij}}{\sum_{r=1}^N w_{rj}} v_{jk} \right) \right)} \quad (1)$$

Na qual CR_{ik} é a contribuição individual em percentagem do vetor de entrada i sobre o vetor de saída k ; $\sum_{r=1}^N w_{rj}$ é a soma dos pesos sinápticos entre a camada de entrada i e o neurônio j ; N corresponde ao total de vetores de entrada; L corresponde ao total de neurônios da camada oculta; v_{jk} é o peso da conexão entre neurônio da camada oculta j e o vetor de saída k .



3. RESULTADOS E DISCUSSÃO

3.1 Simulação da resposta hidrológica

Para cada combinação de vetores de entrada foram realizadas mais de 100 simulações, nas quais foram modificados os valores iniciais dos pesos sinápticos e o número de neurônios na camada oculta. Os resultados do desempenho das simulações pela MLP constam na Tabela 2.

Tabela 2 - Desempenho da simulação da vazão mensal pela MLP em diferentes combinações de vetores de entrada para os conjuntos de treinamento e teste.

Combinação de vetores de entrada	Função objetivo na Etapa de Treinamento		Função objetivo na Etapa de Teste			Número de Neurônios: Entrada/Oculta/Saída
	NS	R ²	NS	R ²	RMS	
					(m ³ .s ⁻¹)	
1 P1 e P2	0,77	0,6898	0,8931	0,5534*	951,5491	2/5/1
2 P1, P2, P3 e P4	0,76	0,7018	0,8698	0,5171*	915,2334	4/5/1
3 P1, P2, P3, P4 e ETP	1,00	0,9982	0,9971	0,9909*	5,638	5/3/1
4 P1, P2 e Q _{T-1}	0,03	0,4207	0,8807	0,6121*	1825,924	3/5/1
5 P1, P2, P3, P4 e Q _{T-1}	0,97	0,8601	0,8752	0,6115*	442,9682	5/3/1
6 P1, P2, P3, P4, ETP e Q _{T-1}	0,99	0,9915	0,9972	0,9937*	31,799	6/5/1
7 PP1, PP2, PP3, PP4, ETP e Q _{T-1}	0,99	0,9916	0,9982	0,9943*	31,4605	6/11/1
8 P1, P2 e ETP	0,99	0,9869	0,9968	0,9811*	41,3518	3/4/1
9 μP-ETP, Q _{T-1}	0,87	0,8396	0,9286	0,8191*	498,6254	2/5/1
10 ETP	1,00	0,9967	0,991	0,9915*	10,052	1/16/4

* p-valor < 0,05

De um modo geral, os desempenhos podem ser considerados bons, com exceção do desempenho na fase de treinamento da combinação 4.

O R² na fase de teste variou entre 0,5171 (combinação 2) a 0,9943 (combinação 7) e NS de 0,8698 e 0,9982, respectivamente. Verifica-se uma divergência entre as funções-objetivos, já que os resultados encontrados utilizando o RMS diferem das demais. No RMS, o melhor resultado é o da combinação 3, usando as precipitações das quatro estações e a ETP (RMS = 5,683). Por outro lado, pelo NS e o R², o melhor resultado é o da combinação 7, que utiliza a precipitação ponderada dos quatro postos, ETP e Q_{T-1} (R² = 0,9943 e NS = 0,9982). De acordo com Machado e Vettorazzi (2003), o valor de NS é o índice mais adequado para avaliação do desempenho, sendo assim, a combinação 7 é a melhor (Figura 2).

Constata-se que, ao utilizar a precipitação das duas estações pluviométricas instaladas dentro da área da bacia (P1 e P2) (combinação 1, NS = 0,8931), o resultado é ligeiramente melhor do que quando utilizadas juntamente com as estações pluviométricas localizadas fora do perímetro da bacia (P3 e P4) (combinação 2, NS = 0,8698).

A combinação 9 foi baseada em Oliveira et al. (2014), o qual considera o balanço de água ou saldo (μP-ETP). No presente estudo, seu desempenho (combinação 9, NS = 0,9286) foi pouco inferior em relação ao uso das precipitações pontuais (combinação 6, NS = 0,9972).

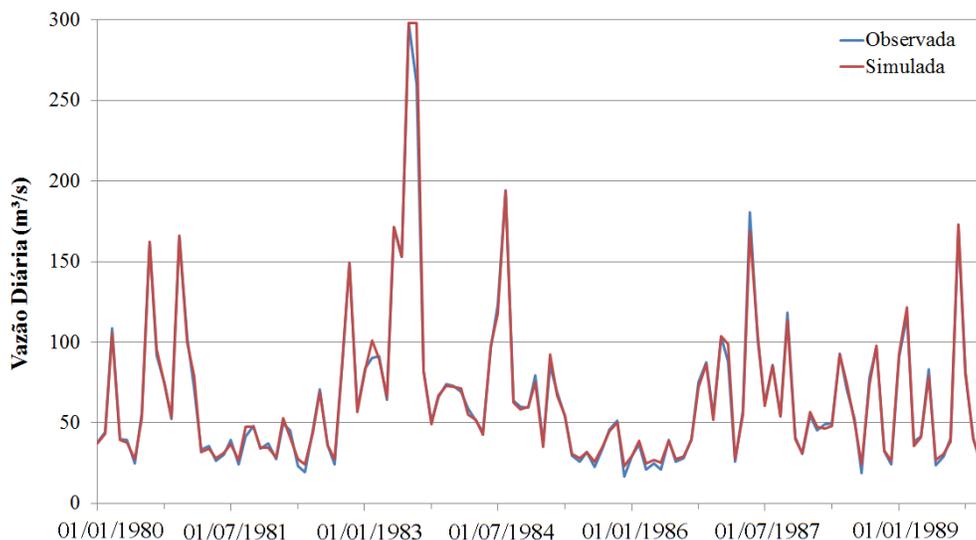


Figura 2 – Hidrograma das vazões médias mensais observadas e simuladas para a combinação 7 dos vetores de entrada.

Nota-se que a utilização da vazão registrada no mês anterior (QT-1) (combinação 5, NS = 0,8752) ocasionou uma melhoria no desempenho da MLP comparada à não utilização desta variável (combinação 2, NS = 0,8698).

Ao utilizar a altura precipitada de todas as estações pluviométricas com influência sobre a bacia (P1, P2, P3 e P4), juntamente com ETP (combinação 3, NS = 0,9971), o resultado é superior ao encontrado nas combinações em que não é utilizada esta variável (combinação 1 com NS = 0,8931; combinação 2 NS = 0,8698; combinação 4 com NS = 0,8807; combinação 5 com NS = 0,8752; combinação 9 com NS = 0,9286). Isto pode ser explicado pela escala mensal dos dados, na qual a ETP sofre sensíveis variações ao longo do ano nesta região, de acordo com o comprimento médio dos dias e da temperatura média mensal, fatores determinantes no método de Thorntwaite. Estes resultados ainda evidenciam a forte relação entre a ETP medida há cerca de 75km da bacia com a ETP local, medida ausente neste trabalho. Isto pode estar relacionado com o fato da Estação Lages estar em latitude e altitude média semelhantes às da bacia.

Portanto, estes resultados evidenciam a importância encontrada pela MLP na variável ETP, como demonstrado na combinação 10, na qual apenas a ETP foi o suficiente para encontrar resultados surpreendentes (NS = 0,991). Isto indica que, para a predição de vazão a um passo de tempo a frente em escala mensal, utilizando-se o modelo MLP, esta variável torna o modelo parcimonioso, além de diminuir os custos de aquisição dos dados de precipitação.

3.2 Avaliação do modelo

A contribuição relativa das variáveis foi calculada para as diferentes combinações de vetores de entrada (Figura 3).

A ETP foi a variável que apresentou maior CR (entre 50 e 60%), o que confirma o ganho no desempenho quando a mesma é acrescida no conjunto de entrada (combinação 3 com NS = 0,9971; combinação 6 com NS = 0,9972; combinação 7 com NS = 0,9982; combinação 8 com NS = 0,9968 e combinação 10 com NS = 0,991).

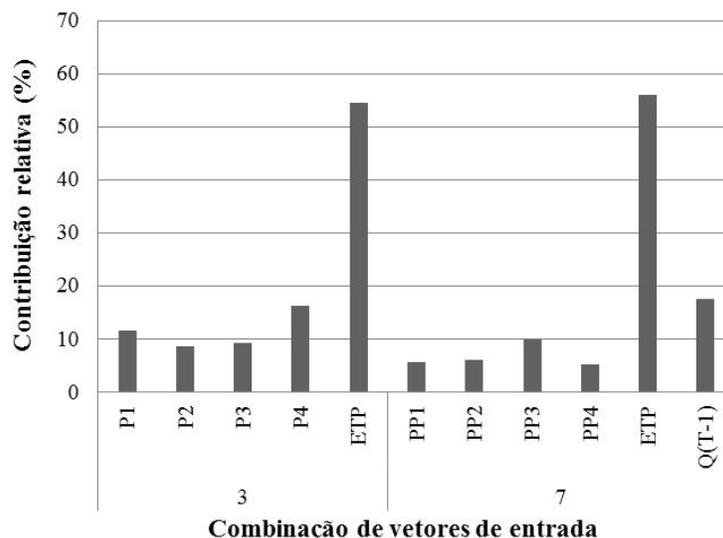


Figura 3 - Análise da contribuição relativa (CR) de cada variável, para as combinações de vetores de entrada com melhor desempenho no teste (combinação 3 e 7).

Na combinação 7 ($NS = 0,9982$), a vazão média do mês anterior (Q_{T-1}) apresentou CR maior que o CR das precipitações ponderadas (PP1, PP2, PP3 e PP4), indicando que esta variável foi importante para o desempenho do modelo nesta combinação.

4. CONCLUSÃO

A ETP se mostrou uma variável essencial para a modelagem, confirmada pelo desempenho obtido utilizando-se a MLP (apenas a ETP alcançou $NS = 0,991$) e pela análise de CR (entre 50 e 60%). O melhor desempenho foi obtido quando combinadas as variáveis de precipitação ponderada para as quatro estações pluviométricas, ETP e Q_{T-1} , com $NS = 0,9982$ e R^2 de 0,9943 na fase de teste. Conclui-se também que a ETP média mensal, individualmente, pode ser utilizada como vetor de entrada na MLP para simular as vazões médias mensais na bacia do Alto Canoas.

Finalmente, estes resultados apontam a RNA como modelo relevante para simular vazões médias mensais de médio prazo, úteis em projetos de engenharia e na gestão dos recursos hídricos.

5. REFERÊNCIAS

EMBRAPA -EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Sistema brasileiro de classificação de solos. 2 ed. Rio de Janeiro, Embrapa Solos, 2006.

FU, L. Neural networks in computer intelligence. New York: McGraw-Hill, 1994.

GARSON, G .D. Interpreting neural-network connection weights. Artif. Intell. Expert 6, p. 47–51, 1991.

GOZLAN, R. E.; MASTRORILLO, S.; COPP, G. H.; LEK, S. (1999). Predicting the structure and diversity of young-of-the-year fish assemblages in large rivers. Fresh. Biol. 41, 809–820.



HAGAN, M. T.; MENHAJ, M. Training feed-forward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks, v.5, n.6, p.989-993, 1994.

HAYKIN, S. Redes neurais princípios e prática. Porto Alegre: Bookman, 2001, 900p.

HEATH, G. E. Training, testing and validating data set in Neural Network.2010. Disponível em: <http://www.mathworks.com/matlabcentral/newsreader/view_thread/295781#917734>. Acesso em: 04 de jul.2013.

MACHADO, R. E.; VETTORAZZI, C. A.; CRUCIANI, D. E. Simulação de escoamento em uma microbacia hidrográfica utilizando técnicas de modelagem e geoprocessamento. Revista Brasileira de Recursos Hídricos, v.8, n.1, p.147-155, 2003.

NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part 1– A discussion of principles. Journal of Hydrology, v.10, n.3.pag.282-290, 1970.

OLDEN, J. D.; JACKSON, D. A. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling, v.154, p.135-150, 2002.

OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. O desempenho das redes neurais artificiais (RNAs) para simulação hidrológica mensal. Revista Brasileira de Recursos Hídricos. v. 19, n. 2, pag. 251-265, 2014.

RIAD, S.; MANIA, J.; BOUCHAOU, L.; NAJJAR, Y. Rainfall-runoffmodelusing na artificial neural network approach. Mathematical and computer modelling. pag. 839-846, n.40, 2004.

SILVA, A. M.; MELLO, C. R. Estudos hidrológicos sobre o regime de escoamento das bacias de drenagem para o reservatório da UHE de Camargos/CEMIG. Lavras: UFLA/CEMIG, 2005. 600p. Relatório de Projeto de Pesquisa e Desenvolvimento (P&D076)

SILVA, R. D. Estudo das variações espectrais e texturais em floresta primária e sucessões secundárias na Flona de Tapajós usando dados ALI/EO-1. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.

SOLOMATINE, D.; OSTFELD, A. Data-driven modelling: some past experiences and new approaches. Journal of Hydroinformatics. n.10.1, p.3-22, 2008.