

# Proposta de um Método para Previsão de Cheias Sazonais Utilizando Redes Neurais Artificiais: Uma Aplicação ao Rio Amazonas

Márcio de M. Rodrigues<sup>1</sup>, Marly G. F. Costa<sup>1</sup>, Cícero F. F. C. Filho<sup>1</sup>

<sup>1</sup>Centro de Pesquisa e Desenvolvimento de Tecnologia Eletrônica e da Informação –  
Universidade Federal do Amazonas (CETELI/UFAM)  
CEP 69077-000 – Manaus – AM – Brazil

marciodemenezes@yahoo.com.br, {marlygfcosta, cffcfilho}@gmail.com

**Abstract.** *This paper proposes a new method for forecasting the maximum seasonal amplitude of rivers, using feedforward neural networks and, as input variables, climatic indices and the river amplitude measured a few months earlier before the maximum amplitude be verified. A new method for selecting the most relevant prediction variables is proposed. For neural networks training, two methods for improving its generalization are used: early stop and regularization. The best prediction result is obtained with two input variables, resulting in a correlation prediction coefficient of  $r_p = 0,755$ .*

**Resumo.** *Esse trabalho discute a aplicação de um novo método para previsão de cheias sazonais de rios, utilizando redes neurais artificiais com as seguintes variáveis de entrada: índices climáticos e o nível do próprio rio, avaliado meses antes da ocorrência do pico da cheia. Um novo método para seleção das variáveis mais relevantes para a predição é proposto. Para o treinamento da rede neural são utilizados dois métodos para melhorar a generalização das mesmas, parada antecipada e regularização. O melhor resultado de predição obtido foi com três variáveis e resultou num índice de correlação de predição de  $r_p = 0,755$ .*

## 1. Introdução

Cheias sazonais, como as que ocorrem no Amazonas, Nilo ou Congo, por exemplo, são caracterizadas por um longo período enchendo e um longo período secando. O ciclo desses rios é de um ano. No rio Amazonas, localizado no nordeste da América do Sul, na maioria das vezes o maior nível do rio ocorre no mês de junho e o menor nível ocorre no mês de outubro ou novembro. Nas Figuras 1(a) e 1(b) mostram-se histogramas das ocorrências dos níveis máximos e mínimos do rio em função do mês em que os mesmos ocorrem, no período que se estende de 1903 a 2013. As medições desses níveis são realizadas desde 1903 pela Autoridade do porto de Manaus (Porto de Manaus, 2015).

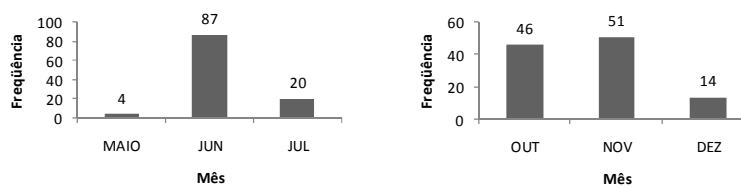


Figura 1. (a) frequência de ocorrência dos picos de cheia em função do mês (b) Imagem com baixa densidade de conteúdo de fundo.

Com uma descarga anual de 6300km<sup>3</sup> anuais [Russel e Miller, 1990], o rio Amazonas disputa com o rio Nilo o título de maior rio do mundo. Na Tabela 1 mostra-se uma comparação entre os três principais rios do mundo [Vörösmarty *et al.*, 1998]. Conforme pode ser visto, em termos de comprimento, o rio Nilo ocupara a primeira posição e o rio Amazonas a segunda posição, porém o inverso ocorre quando se considera o tamanho da bacia. Anualmente, a precipitação pluviométrica na bacia do rio Amazonas situa-se entre 2-3m. Metade desse volume de água é perdida por evapotranspiração, sendo a outra metade despejada no oceano atlântico [Junk and Furch, 1985].

As variações que ocorrem no nível do rio Amazonas anualmente (ou em sua descarga anual) têm uma grande influência nas atividades econômicas da região (pesca, agricultura e extração de madeira) que ocorrem nas regiões de várzea, densamente povoadas, ricas em produtos naturais e com solos produtivos [Junk *et al.*, 2000]. Assim, uma importante tarefa que se impõe é a previsão do nível máximo do rio, ou nível cheia, em cada ano.

**Tabela 1. Comparação entre os 3 principais rios do mundos**

Rio	Continente	Comprimento (Km)	Desembocadura	Área da Bacia STN-30 (Km <sup>2</sup> )
Nilo	Africa	6727	Mar mediterrâneo	3.826.122
Amazonas	América do Sul	6308	Oceano Atlântico	5.853.804
Mississipi	América do Norte	6228	Golfo do México	3.202.959

Diversos autores têm mostrado que a variação inter-anual da descarga e do regime de chuvas no rio Amazonas são correlacionados com as temperaturas da superfície dos oceanos (Sea Surface Temperature - SST) atlântico e pacífico e com o fenômeno climático El Niño South Oscillation (ENSO). Tele conexão é o termo utilizado em meteorologia para descrever variações simultâneas entre clima e processos relacionados em pontos distantes do planeta.

Schongart e Junk (2007), encontraram correlações relativamente altas entre o máximo nível do Rio Negro e o índice de oscilação meridional (Southern Oscillation Index - SOI) de fevereiro ( $r = 0,45, p < 0,01$ ) e o índice SST *Nino*<sub>34</sub> de fevereiro ( $r = -0,49, p < 0,001$ ). Em um importante estudo sobre a descarga de grandes bacias hidrográficas do mundo (bacias com mais de 300.000 km<sup>2</sup>), Potter *et al.* (2004) concluiu que, no período de 1958-1998, existe uma correlação forte e negativa ( $r = -0,5, p < 0,05$ ) entre a descarga do rio e o índice SST do pacífico, *El Nino*<sub>3+4</sub>.

Verifica-se nos trabalhos anteriores uma correlação negativa da descarga do rio com índices SST e uma correlação positiva com o índice SOI.

Alguns trabalhos foram publicados tendo em vista a previsão do nível (ou descarga) do rio Amazonas [Schongart e Junk, 2007, Coe *et al.*, 2002. e Uvo *et al.*, 2000].

Schongart e Junk [2007], utilizando correlação múltipla, e analisando o período de 1903 a 2004, obtiveram uma alta correlação ( $r = 0,71, p < 0,0001$ ) entre o nível máximo do rio Negro em um ano, como variável independente, e o índice SOI de fevereiro e o nível do rio em fevereiro, do mesmo ano, como variáveis dependentes.

Uvo *et al.* [2000], para o período de 1946-1992, para previsão da descarga do rio em Manaus, utilizaram redes neurais de múltiplas camadas, treinadas com o algoritmo de retropropagação, utilizando o método do gradiente descendente com momento adaptativo. O método de treinamento e teste foi o de validação cruzada com 4 pastas. A

arquitetura da rede utilizada foi 8-20-1. Os autores obtiveram os melhores resultados de previsão da descarga, ( $r = 0,86$ ), quando utilizaram as seguintes variáveis de entrada: nível médio do índice SST no pacífico equatorial nos meses de Dez-, Jan e Fev e o total de precipitação nos meses de Set-, Out- e Nov-, somadas em três regiões retangulares a jusante de Manaus: (65-60°W) e (0-5°N), (70-65°W) e (0-5°N) e (65-60°W) e (0-5°S).

No trabalho atual propõe-se uma nova metodologia para previsão das cheias do rio Amazonas utilizando redes neurais de propagação direta. Pretende-se que essa metodologia não apresente viés metodológico, como por exemplo o de se realizar diversas simulações e obter aquela com o melhor resultado para a previsão. Com esse objetivo serão obtidas previsões do nível da cheia com 100 redes treinadas com inicialização aleatória de pesos e polarizações. O nível da cheia será o valor médio das previsões obtidas nas 100 simulações e não o melhor valor obtido para a previsão. Para melhoria da característica de generalização das redes neurais, em adição ao método de parada utilizando o erro médio quadrático, serão utilizados dois métodos alternativos de treinamento: o método de parada antecipada e o método de regularização. Como variáveis de entrada das redes neurais serão utilizados índices climáticos e o nível do rio Negro aferido no porto de Manaus. Como o rio Negro em Manaus encontra-se a poucos quilômetros do rio Amazonas, utiliza-se indistintamente o termo nível do rio Negro ou nível do rio Amazonas. Ambos se referem ao nível de qualquer um dos rios nas imediações de Manaus. A metodologia aqui apresentada, com adaptações nas variáveis de entrada, pode ser aplicada a outros rios com cheias sazonais.

## **2. Materiais e Métodos**

### **2.1. Materiais**

Como informações de entrada para previsão da cheia foram utilizados dois conjuntos de variáveis. O primeiro deles é constituído pelos índices mostrados na Tabela 2: SST *El Niño*<sub>1+2</sub>, SST *El Niño*<sub>3</sub>, SST *El Niño*<sub>3+4</sub>, SST *El Niño*<sub>4</sub> e SOI [Trenberth e Stepaniak, 2001]. Os índices SST são definidos como a temperatura média no pacífico oriental, nas regiões mostradas na Tabela 2. O índice SOI é definido como um Índice de Oscilação Meridional do impacto atmosférico do ENSO, calculado como a diferença padronizada entre a pressão ao nível do mar média entre o Tahiti (17°S, 149°W) e Darwin, Austrália (13°S, 131°E). Os dados desses índices foram obtidos para os anos no período de 1953-2014 [Climate Prediction Center, 2015].

O segundo conjunto de variáveis é constituído pelo nível do rio nos meses que antecedem a mês de ocorrência da cheia. Cada variável é definida por um vetor com 12 valores, em que o primeiro elemento corresponde ao valor do índice no mês de julho do ano anterior (um mês após a ocorrência do pico da cheia do ano anterior) e o último elemento corresponde ao valor do índice no mês de junho do ano em que a previsão é feita.

**Tabela 2: Índices climáticos utilizados para a previsão do nível do rio Negro**

Índice ENSO	Latitude	Longitude	Período
SST El Niño $_{1+2}$	0 a 10 °S	90 a 80°W	1950-2014
SST El Niño $_3$	5°N a 5°S	150 a 90°W	1950-2014
SST El Niño $_{3+4}$	5°N a 5°S	170 a 120°W	1950-2014
SST El Niño $_4$	5°N a 5°S	160°E a 150°W	1950-2014
SOI	Diferença padronizada de pressão entre o Tahiti (17°S a 149°W) e Darwin, Austrália (13°S a 131°E)		

## 2.2. Algoritmos para previsão das cheias

Utilizaram-se dois algoritmos para previsão da cheia do rio Amazonas. O primeiro deles, mais simples foi denominado de algoritmo 1. Esse algoritmo, mostrado a seguir, guarda uma semelhança com o algoritmo *leave-one-out* de treinamento e teste [Sonka and Fitzpratick,2000]. A diferença é que, para se levar em conta a aleatoriedade da rede neural (cujo comportamento é ditado pela inicialização aleatória dos pesos) no cálculo do nível da cheia, são realizadas  $m$  simulações com  $m$  redes iniciadas aleatoriamente. O nível da cheia é então determinado como o valor médio desses  $m$  resultados obtidos. Nesse trabalho foi utilizado um valor de  $m=100$ .

### Algoritmo 1

1. Definir de um conjunto de variáveis de entrada;
2. Selecionar das variáveis com significância estatística;
3. Para um dado ano,  $ano_k$ ;
  - a. Treinar a rede  $m$  vezes com dados de  $N-1$  anos, excetuando-se o  $ano_k$  (em que  $N = \text{número de anos} = 64$ ). Para cada um dos treinamentos obter a previsão da cheia para o  $ano_k$ ;
  - b. Com as  $m$  previsões da cheia para o  $ano_k$ , calcular a previsão média da cheia do  $ano_k$ .

O segundo algoritmo, denominado de algoritmo 2, ao invés de utilizar todas as características, faz uso de um método proposto por nós nesse trabalho e que denominamos seleção escalar de características modificado, para selecionar um conjunto reduzido de  $p$  características a serem utilizados na previsão. A seguir descrevemos esse algoritmo.

### Algoritmo 2

1. Definir de um conjunto de variáveis de entrada;
2. Selecionar das variáveis com significância estatística;
3. Selecionar as  $p$  melhores características utilizando o método de seleção escalar de características modificado;
4. Para um dado ano,  $ano_k$ ;
  - a. Treinar a rede  $m$  vezes com dados de  $N-1$  anos, excetuando-se o  $ano_k$  (em que  $N = \text{número de anos} = 64$ ). Para cada um dos treinamentos obter a previsão da cheia para o  $ano_k$ ;
  - b. Com as  $m$  previsões da cheia para o  $ano_k$ , calcular a previsão média da cheia do  $ano_k$ .

Assim como em Schongart (2007), definiu-se nesse trabalho que a previsão da cheia será feita no mês de fevereiro, ou seja, com 4 meses de antecedência em relação ao pico da cheia.

### 2.3 Método de Seleção Escalar de Características Modificado

Originalmente, o método de seleção escalar de características [Theodoridis and Koutroumbas, 2006], é utilizado em reconhecimento de padrões para selecionar as variáveis que melhor separam um conjunto de amostras em  $n$  classes distintas. Com essa finalidade, utiliza o conceito de Razão Discriminante de Fisher (Fisher Discriminant Ratio - FDR) e índices de correlação cruzada. Como em um trabalho de previsão não temos classes, substituímos, no método proposto nesse trabalho, a Razão Discriminante de Fisher pelo Índice de Correlação de Pearson (ICP). Através da utilização dessa técnica é encontrar um conjunto de características para previsão que não tenham informações redundantes. A seguir descreve-se os passos do método de seleção escalar de características modificado:

- A. Utilizando o coeficiente de correlação de Pearson selecionar a primeira característica, que será aquela com maior valor de correlação com o nível de cheia,  $r_k$ .

Em que:  $k = 1 \dots m$

$m$  – número de características

O valor de  $r_k$  é calculada para cada característica  $x_k$ . A característica  $x_k$  com o maior valor de  $r_k$  é selecionada como a primeira característica,  $x_{s1}$ .

- B. Para selecionar a segunda característica,  $x_{s2}$ , o coeficiente de correlação cruzada entre duas características  $x_i$  e  $x_j$  é utilizado, conforme definido na equação (1)

$$\rho_{ij} = \frac{\sum_{n=1}^N x_{ni}x_{nj}}{\sqrt{\sum_{n=1}^N x_{ni}^2 \sum_{n=1}^N x_{nj}^2}} \quad (1)$$

Em que:  $N = \text{Número total de anos } (63=2014-1951+1)$

$x_{ni}$  and  $x_{nj}$ : valores da  $i$ -ésima e da  $j$ -ésima característica do ano  $n$ .

A segunda característica selecionada,  $x_{s2}$ , é aquela que maximiza a equação (2):

$$\alpha_1 r_{s2} - \alpha_2 |\rho_{s1s2}|, \text{ for all } s2 \neq s1 \quad (2)$$

$\alpha_1$  and  $\alpha_2$  expressam a importância do primeiro e do segundo termo, respectivamente. Nesse trabalho  $\alpha_1 = \alpha_2 = 0.5$ .

- C. Outras características selecionadas,  $x_{sk}$ ,  $k=3, \dots, m$ , são aquelas que maximizam a equação (3):

$$\alpha_1 r_{sk} - \frac{\alpha_2}{k-1} \sum_{r=1}^{k-1} |\rho_{srsk}| \quad (3)$$

### 2.4. Características das Redes Neurais Utilizadas

As redes neurais utilizadas têm uma arquitetura  $n-n-1$ . Dependendo do número de variáveis de entrada, simulações foram feitas para se obter o melhor valor de  $n$ . Para a previsão de um ano, conforme já dito, 100 simulações com pesos aleatoriamente reiniciados foram feitas. O algoritmo de treinamento usado foi o *backpropagation*

associado ao método de otimização *Levenberg-Marquardt* [Cichoki e Unbehauen 1993]. A função de ativação utilizada nos neurônios de todas as camadas foi a tangente hiperbólica.

Com o objetivo de melhorar a generalização das redes neurais utilizadas, além do critério do erro médio quadrático, dois critérios de parada adicionais foram utilizados: regularização [Doan and Liong, 2004] e parada antecipada [Demuth *et al.*, 2015].

Na regularização, com o objetivo de trabalhar-se com redes mais estáveis (com valores de pesos menores), modifica-se o critério de parada do erro médio quadrático, introduzindo-se um fator proporcional à soma dos quadrados dos pesos, conforme mostrado na expressão (4).

$$msereg = \gamma mse + (1 - \gamma)msw \quad (4)$$

em que:

$\gamma$  – taxa de desempenho, varia entre 0 e 1.

$mse$  – erro médio quadrático.

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (5)$$

O critério de parada expresso em (4) faz com que o desempenho da rede tenha pesos menores e isso força a resposta da rede para ser mais suave, melhorando o grau de generalização da rede.

O critério de parada antecipado caracteriza-se por dividir o conjunto de dados em 3 grupos: treinamento, teste e validação. Nesse trabalho, cada um desses conjuntos foi formado por 22-21-21 anos, respectivamente. A principal característica deste método é que quando o erro médio quadrático do conjunto de validação cresce por um número fixo iterações seguidas, o treinamento é finalizado. Nesse trabalho esse número foi feito igual a duas iterações.

### 3. Resultados

#### 3.1 Variáveis de entrada e significância estatística

Na Tabela 3 mostram-se os valores das variáveis definidas nesse trabalho para a previsão da cheia, a correlação das mesmas com o nível da cheia, o valor do teste de significância t-Student e se o mesmo é ou não significativo ao nível de 0.1%.

Conforme pode ser observado na Tabela 3, nenhum dos meses do índice SST *El Nino*<sub>1+2</sub> têm uma correlação significativa com o nível da cheia. O índice SST *El Nino*<sub>3+4</sub> do mês de dezembro tem a maior correlação dos índices SST com o nível da cheia,  $r = -0,428$ . O índice SOI do mês de janeiro tem a correlação mais alta entre todos os índices, no valor de  $r = 0,513$ . Os índices SST apresentam todos uma correlação negativa com o nível da cheia, enquanto que o índice SOI apresenta uma correlação positiva. O nível do rio em fevereiro o maior valor de correlação com o nível da cheia das variáveis estudadas,  $r = 0,761$ . A partir dessa tabela observa-se um total de 14 variáveis com significância estatística. As 9 variáveis com os maiores valores de correlação estão listadas em ordem decrescente de correlação na primeira linha da Tabela 4. Isso conclui o passo 2 dos algoritmos 1 e 2.

### 3.2 Seleção das melhores características de entrada através do método de seleção escalar modificado.

Na segunda linha da Tabela 4 mostra-se a ordem das 9 variáveis com maior valor de  $r$  após a aplicação do método de seleção escalar de características modificado. A partir dessa tabela observa-se que o nível do rio no mês de janeiro, embora seja a variável com o segundo melhor valor de correlação,  $r=0,527$ , ocupa apenas a quarta posição entre na ordem das variáveis selecionadas por esse método. A segunda variável escolhida é o índice SOI do mês de janeiro, enquanto que a terceira, é o índice SOI do mês de setembro. Isso conclui o passo 3 do algoritmo 2.

**Tabela 3: Variáveis de entrada e significância estatística**

Índice / Nível do rio	Medidas	Meses utilizados para previsão								Meses não utilizados para previsão			
		jul	ago	set	out	nov	dez	jan	fev	mar	abr	maio	jun
El Niño <sub>1+2</sub>	r	-0,109	-0,166	-0,255	-0,279	-0,327	-0,298	-0,187	-0,120	-0,132	-0,096	-0,016	0,021
	t - Student	-0,854	-1,315	-2,061	-2,272	-2,700	-2,442	-1,484	-0,946	-1,038	-0,752	-0,128	0,165
	Significante para $p<0,001$ ?	N	N	N	N	N	N	N	N				
SST El Niño <sub>3</sub>	r	-0,280	-0,269	-0,288	-0,311	-0,365	-0,381	-0,380	-0,346	-0,273	-0,165	-0,025	0,024
	t - Student	-2,275	-2,182	-2,352	-2,555	-3,064	-3,217	-3,208	-2,876	-2,213	-1,303	-0,198	0,190
	Significante para $p<0,001$ ?	N	N	N	N	N	S	S	N				
SST El Niño <sub>3+4</sub>	r	-0,371	-0,359	-0,325	-0,351	-0,368	-0,428	-0,444	-0,442	-0,404	-0,284	-0,130	-0,090
	t - Student	-3,119	-3,008	-2,685	-2,924	-3,094	-3,699	-3,872	-3,846	-3,448	-2,313	-1,028	-0,702
	Significante para $p<0,001$ ?	N	N	N	N	N	S	S	S				
SST El Niño <sub>4</sub>	r	-0,260	-0,260	-0,254	-0,327	-0,320	-0,360	-0,378	-0,386	-0,364	-0,189	-0,101	-0,057
	t - Student	-2,104	-2,101	-2,053	-2,700	-2,639	-3,018	-3,194	-3,267	-3,051	-1,501	-0,795	-0,447
	Significante para $p<0,001$ ?	N	N	N	N	N	N	S	S				
SOI	r	0,321	0,243	0,449	0,379	0,502	0,430	0,513	0,290	0,374	0,271	-0,151	0,040
	t - Student	3,521	2,604	5,216	4,261	6,037	4,956	6,217	3,144	4,192	2,926	-1,583	0,421
	Significante para $p<0,001$ ?	N	N	S	S	S	S	S	N				
Nível do rio	r	0,158	0,122	0,069	0,053	0,022	-0,036	0,527	0,761	0,860	0,927	0,970	0,999
	t - Student	1,666	1,277	0,724	0,549	0,230	-0,371	6,437	12,197	17,476	25,714	41,295	324,249
	Significante para $p<0,001$ ?	N	N	N	N	N	N	S	S				

**Tabela 4: Variáveis ordenadas segundo o valor de r e segundo o método de seleção escalar de características modificado**

Critério de ordenação	Variáveis ordenadas								
	Nível do rio em Fev (r=0.761)	Nível do rio em Jan (r=0.527)	SOI Jan (0.513)	SOI Nov (r=0.502)	SOI Set (r=0.449)	El Niño <sub>3+4</sub> Jan (r=-0.444)	El Niño <sub>3+4</sub> Fev (r=-0.442)	SOI Dez (r=0.430)	El Niño <sub>3+4</sub> Dez (r=-0.428)
<b>Seleção Escalar</b>	Nível do rio em Fev	SOI Jan	SOI Set	Nível do rio em Jan	SOI Nov	SOI Dez	El Niño <sub>3+4</sub> Jan	El Niño <sub>3+4</sub> Fev	El Niño <sub>3+4</sub> Dez

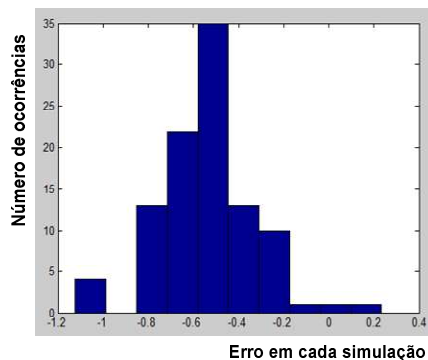
### 3.3 Resultados de previsão utilizando correlação simples

Como pode ser observado a partir da Tabela 3 a correlação do nível do rio no mês de fevereiro com o nível da cheia é  $r = 0,761$ . Para avaliar o poder de previsão dessa variável sobre o nível da cheia fez-se o seguinte teste: para todos os anos de 1951 a 2014 realizou-se 64 experimentos. Em cada experimento, excluiu-se um dos anos e obteve-se a reta com melhor correlação relativa aos 63 anos restantes. Utilizando-se essa reta previu-se a cheia no ano que foi excluído. Após obter as 64 previsões, obteve-se o

coeficiente de correlação de Pearson entre as previsões e o nível da cheia em cada ano. O resultado foi uma correlação de previsão  $r_p = 0,745$ .

### 3.4 Resultados de previsão dos algoritmos 1 e 2

Na Figura 2 mostra-se um histograma dos erros cometidos na previsão do nível da cheia para ao ano de 1975, por uma rede neural com as seguintes condições de treinamento: 3 variáveis de entrada com maiores valores do coeficiente de correlação  $r$ , arquitetura 6-6-1, critério de parada regularização. Foram feitas 100 treinamentos e, para cada um desses treinamentos, uma previsão da cheia. O conjunto de treinamento constituiu-se de dados referentes aos anos de 1951-2014, excetuando-se o ano de 1975. Como pode ser visto, temos valores de erro maiores do que -1. No entanto, o erro médio cometido, que é aquele utilizado como valor da predição da cheia, foi de -0,53. Para a predição da cheia soma-se esse valor médio do erro a cheia ocorrida no ano. Seguindo um procedimento similar, é calculada previsão da cheia para todos os anos e calculado o coeficiente de correlação da previsão.



**Figura 2: Histograma do erro de 100 simulações de uma rede neural utilizando o critério de regularização para o ano de 1975. Erro médio=-0,53**

Para se determinar o melhor conjunto de variáveis e a melhor arquitetura da rede foram feitas várias simulações. A Tabela 5 mostra os resultados obtidos. Dois tipos de resultados são mostrados: valores dos índices de correlação de previsão,  $r_p$  e índices de erro. Os índices de erro são: Ind\_1 - Número de previsões em que a diferença entre o valor previsto para o nível cheia e o valor real ficou abaixo de 0,5; Ind\_2 - Número de previsões em que a diferença entre o valor previsto para o nível cheia e o valor real ficou entre 0,5 e 1; Ind\_3 - Número de previsões em que a diferença entre o valor previsto para o nível cheia e o valor real ficou entre 1 e 1,5; Ind\_4 - Número de previsões em que a diferença entre o valor previsto para o nível cheia e o valor real ficou acima de 1,5. Esses valores foram determinados em diversas condições: três critérios de parada da rede: erro médio quadrático, parada antecipada e regularização; conjuntos de 1, 2, 3 4, 5, 7, 8 e 9 variáveis ordenadas pelo valor do coeficiente de correlação  $r$  da variável com o nível da cheia e pelo método de seleção escalar de características modificado proposto nesse trabalho.

## 4. Discussão e Conclusões

A partir dos resultados mostrados na Tabela 5 observa-se que os melhores resultados foram obtidos com o método de parada regularizado. O melhor valor obtido com regularização para o índice de correlação de previsão,  $r_p$ , foi de 0,755. Para obtenção



desse resultado trabalhou-se com 3 variáveis de entrada selecionadas pelo método de seleção escalar de características e com um arquitetura de rede neural 6x6x1. Para essa entrada da tabela, verifica-se que  $Ind\_1 = 33$ ,  $Ind\_2 = 26$ ,  $Ind\_3 = 4$ ,  $Ind\_4 = 1$ . Assim, em 51,56% dos anos a previsão resultou em um erro menor do que 0,5m. Em 92,18% dos anos a previsão resultou em um erro menor do que 1,0m. Apenas em um ano a previsão resultou em um erro maior do que 1,5m. Em todas as simulações esse ano foi o de 1985.

O método de parada antecipada apresentou o segundo melhor conjunto de resultados, ficando o método do erro médio quadrático com o pior desempenho.

O melhor índice de correlação de previsão obtido com redes neurais,  $r_p = 0,755$ , é superior àquele obtido através de correlação simples mostrado na seção 3.4,  $r_p = 0,745$ .

A técnica de seleção escala de características modificada mostrou-se eficaz apenas com o método de regularização, onde obtivemos os melhores resultados. Para o caso de três variáveis, a diferença entre o algoritmo 1, sem seleção escalar, e o algoritmo 2, com seleção escalar, é de 0,038. Como o conjunto de 5 variáveis e de nove variáveis são iguais no algoritmo 1 e no algoritmo 2, os resultados observados na Tabela 5 para esse dois algoritmos são iguais nesses dois casos.

**Tabela 5: Resultados das simulações dos algoritmos 1 e 2**

Algoritmo	Regularização								Early Stop				Erro médio quadrático					
	Número de variáveis	Arq.	$r_p$	Ind_1	Ind_2	Ind_3	Ind_4	$r_p$	Ind_1	Ind_2	Ind_3	Ind_4	$r_p$	Ind_1	Ind_2	Ind_3	Ind_4	
Algoritmo 1	1	4x4x1	0,723	31	24	8	1	0,698	29	26	8	1	0,689	30	23	9	2	
	2	5x5x1	0,71	28	28	7	1	0,694	29	26	8	1	0,63	23	32	3	6	
	3	6x6x1	0,717	25	32	6	1	0,715	27	31	5	1	0,622	23	27	9	5	
	4	7x7x1	0,706	27	30	6	1	0,727	28	32	3	1	0,677	23	32	6	3	
	5	8x8x1	0,722	26	32	5	1	0,705	30	29	3	2	0,621	23	28	10	3	
	7	9x9x1	0,691	27	28	8	1	0,686	25	34	3	2	0,644	23	34	5	2	
	9	10x10x1	0,704	30	29	3	2	0,705	32	23	6	3	0,634	28	29	5	2	
	Algoritmo 2	1	4x4x1	0,723	32	24	7	1	0,706	27	29	6	2	0,618	30	21	11	2
		2	5x5x1	0,744	31	28	4	1	0,677	28	28	5	3	0,634	27	25	10	2
3		6x6x1	0,755	33	26	4	1	0,719	31	26	6	1	0,582	28	16	15	5	
4		7x7x1	0,716	29	28	6	1	0,677	28	28	5	3	0,6579	29	28	5	2	
5		8x8x1	0,722	26	32	5	1	0,698	25	34	3	2	0,622	22	28	12	2	
7		9x9x1	0,721	30	27	6	1	0,715	29	29	4	2	0,692	27	30	5	2	
9		10x10x1	0,704	30	29	3	2	0,696	27	30	5	2	0,635	28	27	7	2	

O método proposto nesse trabalho para previsão de cheias de rios sazonais oferece duas novas características: uma técnica para seleção escalar de características específica para previsão e um método de previsão isento de viés metodológico, na medida em que apresenta o valor de previsão como o resultado de uma média de valores obtidas através de um grande número de simulações independentes. Essa segunda característica diferencia os resultados do trabalho ora apresentado em relação aos resultados apresentados por Uvo *et al.* [2000], onde os autores, para cada uma das pastas do método de validação cruzada com 4 pastas, utilizaram o melhor resultado de varais simulações de uma rede neural, o que caracteriza, a nosso ver, um viés metodológico.

## Agradecimentos

Parte dos resultados apresentados neste trabalho foram obtidos através do Projeto de Pesquisa e formação de recursos humanos, em nível de graduação e pós-graduação, nas áreas de automação industrial, software para dispositivos móveis e TV Digital, financiado pela Samsung Eletrônica da Amazônia Ltda., no âmbito da Lei no. 8.387 (art. 2º)/91

## Referências

- Cichoki, A., Unbehauen, R. (1993), Neural Networks for Optimization and Signal Processing, John Wiley & Sons Ltd., New York.
- Climate Prediction Center. <http://www.cpc.ncep.noaa.gov/data/indices/>. Acessado em 23/02/2015.
- Coe, M. T., Costa, M. H., Botta, A. and Birkett, C. (2002) Long-term simulations of discharge and floods in the Amazon Basin. In *Journal of Geophysical Research*. art. no. 8044.
- Demuth H; Beale M; Hagan M. Neural Network Toolbox User's Guide. Available from:<http://www.mathworks.com/products/neural-network/description6.html>. Acessado em 24/02/2015.
- Doan, C.D. and Liong, S.Y. (2004). Generalization for multilayer neural network Bayesian regularization or early stopping. In: *Proceedings of 2nd Conference of Asia Pacific Association of Hydrology and Water Resources*. pages 1-8.
- Junk, W.J., Ohly, J.J., Piedade, M.T.F. and Soares, M.G.M. (2000) The Central Amazon Floodplain: Actual Use and Options for a Sustainable Management. Backhuys Publishers B.V, Leiden.
- Junk W. J., Furch, K. (1985). The physical setting and chemical properties of Amazonian waters and their relationships with the biota. In *Amazonia*, Prance GT, Lovejoy TE (eds). Pergamon Press: Oxford
- Porto de Manaus, <http://www.portodemanaus.com.br>. Acessado em 23/02/2015.
- Russell, G., Miller, J. (1990). Global river runoff calculated from a global atmospheric general circulation model. In *Journal of Hydrology*, v.117, pages 241-254.
- Schongart, J and Junk, W. J. (2007). Forecasting the flood-pulse in Central Amazonia by ENSO-indices. In: *Journal of Hydrology*, v. 335, pages 124–132.
- Sonka, M and Fitzpatrick, J. M. (2000) Handbook of Medical Imaging. Volume 2: Medical Imaging Processing and Analysis. Washington: Spie Press.
- Theodoridis, S., Koutroumbas, K. (2006) Pattern Recognition, Elsevier Academic Press, San Diego.
- Trenberth, K.E. and Stepaniak, D.P. (2001). Indices of El Niño evolution. In *Journal of Climate*. v.14, pages. 1697–1701.
- Uvo, C. B., Tolle, U. and Berndtsson, R. (2000). Forecasting Discharge in Amazonia Using Artificial Neural Networks. In: *International Journal of Climatology*, v.20. pages 1495 – 1507.
- Vörösmarty, C.J., Fekete, B. and Tucker B.A. (1998). River Discharge Database, Version 1.1 (RivDIS v1.0 supplement). Institute for the Study of Earth, Oceans, and Space / University of New Hampshire, Durham NH (USA). <http://www.rivdis.sr.unh.edu/>.