

UNIVERSIDADE FEDERAL DO AMAZONAS  
FACULDADE DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MÁRCIO DE MENEZES RODRIGUES

OTIMIZANDO A PREVISÃO DE CHEIAS SAZONAIS PARA O RIO  
NEGRO UTILIZANDO REDES NEURAIS DE PROPAGAÇÃO DIRETA

MANAUS  
2015

UNIVERSIDADE FEDERAL DO AMAZONAS  
FACULDADE DE TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MÁRCIO DE MENEZES RODRIGUES

OTIMIZANDO A PREVISÃO DE CHEIAS SAZONAIS PARA O RIO  
NEGRO UTILIZANDO REDES NEURAIS DE PROPAGAÇÃO DIRETA

Dissertação apresentada ao Programa de Pós  
Graduação *Stricto Sensu* em Engenharia  
Elétrica da Universidade Federal do  
Amazonas, como requisito parcial para  
obtenção de título de Mestre em Engenharia  
Elétrica, área de concentração de Controle e  
Automação de Sistemas.

Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho

Co-orientador: Prof. Dra. Marly Guimarães Fernandes Costa

MANAUS  
2015

Rodrigues, Márcio de Menezes

R696o Otimizando a Previsão de Cheias Sazonais para o Rio Negro Utilizando Redes Neurais de Propagação Direta / Márcio de Menezes Rodrigues. 2015

83 f.: il.; 31 cm.

Orientador: Cícero Ferreira Fernandes Costa Filho

Coorientadora: Marly Guimarães Fernandes Costa Dissertação  
(Mestrado em Engenharia Elétrica) - Universidade

Federal do Amazonas.

1. redes neurais artificias. 2. ferramentas de previsão. 3. cheias

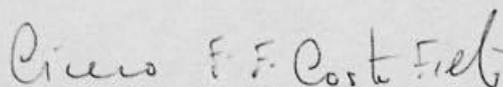
MÁRCIO DE MENEZES RODRIGUES

OTIMIZANDO A PREVISÃO DE CHEIAS SAZONAIS PARA O RIO  
NEGRO UTILIZANDO REDES NEURAI DE PROPAGAÇÃO DIRETA.

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

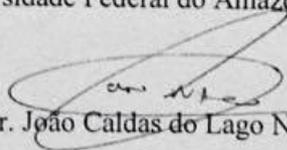
Aprovado em 22 de dezembro de 2015.

BANCA EXAMINADORA



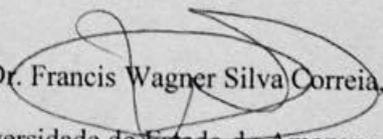
Cícero Ferreira Fernandes Costa Filho, Presidente

Universidade Federal do Amazonas- UFAM



Prof. Dr. João Caldas do Lago Neto, Membro

Universidade Federal do Amazonas- UFAM



Prof. Dr. Francis Wagner Silva Correia, Membro

Universidade do Estado do Amazonas- UEA

Aos meus pais, Arthur Rodrigues *in memoriam* e Maria Socorro Rodrigues, meus irmãos Renata, Eduardo e Jorge, pelo apoio e compreensão durante essa trajetória. À minha esposa Lucienne Rodrigues pela união, carinho, companheirismo e constante apoio.

## AGRADECIMENTOS

Ao meu orientador Prof. Dr. Cícero Ferreira Fernandes Costa Filho pela orientação, apoio, incentivo, amizade, frequente estímulo e ensinamentos úteis no desenvolvimento deste trabalho.

À minha co-orientadora Profa. Dra. Marly Guimarães Fernandes Costa pela colaboração no desenvolvimento desta pesquisa.

Ao amigo Prof. Dr. José Carlos Reston Filho pelo apoio e constante ajuda em minha trajetória.

Aos amigos que sempre me acompanham: Manuel Zumaeta, Vicente Tino, Almir Pacheco, Alfredo Salerno, Djalma Oliveira, Sandro Lucas, Sandro Sales, Silvio Farias, Luiz Augusto, José Rubens, Déric Mourão, Marcelo Paes, Marcelo Reis, Tonny Peixoto, Lúcio Pelinzer e Amarildo Lima.

Aos colegas de turma Endrews, Jonilson, Larissa e Robson pela ajuda e incentivo nessa jornada.

À Universidade Federal do Amazonas em especial ao Centro de Tecnologia Eletrônica e da Informação – CETELI, pela estrutura oferecida e suporte para realização do trabalho.

A Fundação de Amparo à Pesquisa do Estado do Amazonas - FAPEAM pela concessão da bolsa de estudos.

À Samsung. 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, *softwares* para dispositivos móveis e TV Digital, financiado pela Samsung Eletrônica da Amazônia Ltda., no âmbito da Lei no. 8.387 (art. 2º) /91.

## RESUMO

Esse trabalho discute a aplicação de um novo método para previsão de cheias sazonais, 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 previsão 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 previsão obtido foi com três variáveis e resultou num índice de correlação de previsão de  $r_p = 0,755$ .

**Palavras-chave:** Redes Neurais Artificiais; Ferramentas de Previsão; Cheias Sazonais

## **ABSTRACT**

This paper proposes a new method for forecasting the maximum seasonal amplitude, 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  $r_p = 0,755$ .

**Keywords:** Artificial Neural Networks; Forecasting Tools, Seasonal Rivers

## ÍNDICE DE ILUSTRAÇÕES

Figura 1 (a): Frequência de ocorrência dos picos de cheia do Rio Negro em função do mês.	12
Figura 1(b): Frequência de ocorrência dos picos de seca do Rio Negro em função do mês.....	12
Figura 2: Rede Neural de Múltiplas Camadas .....	26
Figura 3: O algoritmo “ <i>Backpropagation</i> ” .....	33
<u>Figura 4</u> : Modelo Matemático de Neurônio.....	33
<u>Figura 5</u> : Coeficiente de Correlação para População e Amostra .....	43
<u>Figura 6</u> : Diagrama do Sistema de Otimização para Previsão de Cheias de Sazonais .....	46
<u>Figura 7</u> : Gráfico mostrando a correlação entre o pico da cheia com a previsão realizada utilizando o nível do rio no mês de fevereiro .....	55
<u>Figura 8</u> : Histograma do erro de 100 simulações de uma rede neural utilizando o critério de regularização para o ano de 1975. Erro médio=-0,53.....	56

## ÍNDICE DE QUADROS

Quadro 1: Resumo da Revisão Bibliográfica.....	20
--	----

## SUMÁRIO

1 INTRODUÇÃO .....	12
1.1 Organização do trabalho .....	14
1.2 Objetivo Geral.....	15
1.3 Objetivo Específico .....	15
2 REVISÃO BIBLIOGRÁFICA .....	15
3 FUNDAMENTOS TEÓRICOS .....	24
3.1 Redes Neurais Artificiais – RNA’s.....	24
3.2 Treinamento das RNMC .....	27
3.3 Características da Aprendizagem .....	29
3.4 O Algoritmo <i>Backpropagation</i> .....	31
3.5 Método de otimização Levenberg – Marquadt.....	36
3.6 Seleção Escalar de Características .....	38
3.7 Coeficiente de Correlação de Pearson .....	41
3.8 Significância Estatística.....	42
3.9 Teste de Significância para o Coeficiente de Correlação de Pearson – t Student.....	43
4 MATERIAIS E MÉTODOS .....	45
4.1 Conjunto de Variáveis.....	45
4.2 Algoritmos de Previsão .....	47
4.3 Método de Seleção Escalar de Características Modificado.....	49
4.4 Configuração das Redes Neurais Utilizadas.....	50
5 RESULTADOS.....	52
5.1 Variáveis de Entrada e Teste de Significância Estatística .....	52
5.2 Seleção das Melhores Variáveis de Entrada Através do Método de Seleção de Características Modificado.....	53
5.3 Índice de Correlação de Previsão <i>rp</i> .....	54
5.4 Resultados de Previsão dos Algoritmos 1 e 2.....	55
5.5 Discussões.....	56
5.6 Conclusões.....	58
REFERÊNCIAS BIBLIOGRÁFICAS.....	61
APÊNDICE I .....	63
APÊNDICE II .....	65
APÊNDICE III.....	75

## 1 INTRODUÇÃO

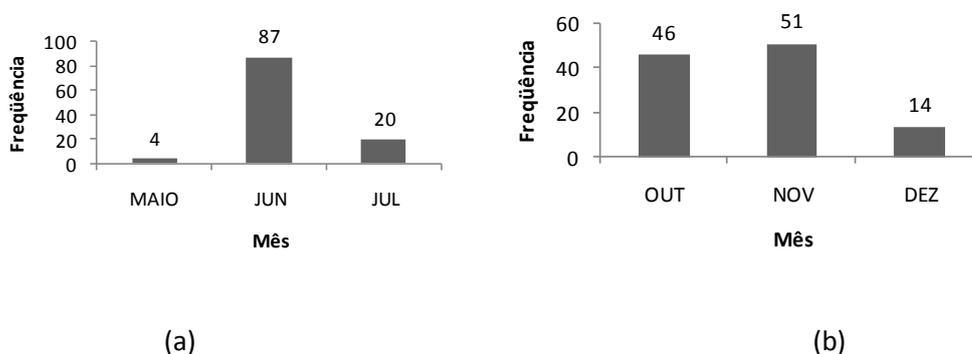
A bacia Amazônica tem área drenada superior a 7 milhões de km<sup>2</sup>, sendo assim a maior bacia hidrográfica do mundo. Deste total, 4 milhões de km<sup>2</sup> estão em território brasileiro (região norte). Também está presente nos territórios da Bolívia, Peru, Venezuela e Colômbia. Começa no território Peruano como rio Vilcanota, e quando entra no território brasileiro recebe o nome de Solimões. Ao encontra-se com o Rio Negro, ganha o nome de Amazonas. O Rio Amazonas é a espinha dorsal da bacia, que conta com grande quantidade de afluentes e canais, criados pelo processo de cheia e vazante.

Todos os anos, com as estações das chuvas na região Amazônica, ocorre o fenômeno das cheias que atingem os municípios que estão nas margens dos rios Solimões, Tapajós, Negro, Juruá, Purus, Japurá, Madeira, entre outros. A cheia de 2012, quando o Rio Negro atingiu 29,97 metros, foi considerada a maior cheia dos últimos tempos. Já no período da estiagem, o fenômeno das secas também atinge a região. Em 2010 ocorreu a maior seca dos últimos 100 anos, tendo o nível do Rio Negro atingido a cota de 13,63 metros.

O período de cheia inicia-se em maio estendendo-se até meados de julho. Neste período ocorreram as maiores cotas já registradas do Rio Negro, bem como a maior frequência de cotas máximas. A seca corresponde ao período situado entre os meses de setembro a novembro. Estes meses registraram as menores cotas já registradas no Rio Negro, bem como a maior frequência de cotas mínimas registradas. As Figura 1(a) e 1(b) mostram a frequência de ocorrência dos picos de cheia e de seca do Rio Negro, respectivamente, em função dos meses do ano, nos últimos 111 anos.

O ciclo sazonal dos rios da região amazônica regula fortemente as atividades de agricultura, pesca e a extração de madeira dos Povos das Águas, os ribeirinhos, pessoas que povoam as margens dos rios de várzeas. As chuvas dificultam fortemente não só as atividades

de agricultura, extrativismo e pesca de subsistência, que fica em grande parte comprometida, mas também a lida com os animais.



**Figura 1. (a) frequência de ocorrência dos picos de cheia do Rio Negro em função do mês, nos últimos 111 anos (b) frequência de ocorrência dos picos de seca do Rio Negro em função do mês, nos últimos 111 anos.**

As atividades econômicas nas regiões de várzea ficam comprometidas com as variações do nível dos rios. A pesca comercial é realizada durante os períodos de cheias, entre os meses de abril e junho. A agricultura é praticada durante a vazante em altitudes mais elevadas, onde em períodos curtos de inundações o cultivo tem tempo suficiente para amadurecer (JUNK et al., 2007). A extração da madeira começa entre março e abril antes que as florestas sejam inundadas entre os meses de junho e julho.

Anualmente, um fator de extrema relevância associada as cheias dos Rios da Bacia Amazônica, que compromete a segurança, a saúde, o índice de produção econômica e qualidade de vida dos municípios do interior do estado do Amazonas, é o nível máximo dos Rios. Nota-se que existe uma grande imprecisão na previsão hidrológica do nível máximo da água, não somente do Rio Amazonas, mas também em todos os rios da bacia Amazônica. O pulso de inundação constitui o fator determinante na maioria dos processos ecológicos na várzea, causando sua transformação periódica de ambientes terrestres a aquáticos (RAMALHO et al., 2009). Essa imprecisão tem como reflexo a ausência de uma política de estado eficiente para a redução dos danos à população, causados pelas enchentes dos rios da região amazônica.

Alguns trabalhos publicados contribuíram para amenizar esses problemas, estabelecendo metodologias mais precisas para a previsão de cheias dos Rios da Bacia Amazônica. Entre esses trabalhos, destacam-se o de Uvo et al. (2000) e Schongart e Junk (2007), que mostram a correlação das chuvas e descargas da bacia Amazônica com os fenômenos climáticos de temperatura da superfície dos oceanos (*Sea Surface Temperature – SST*) atlântico e pacífico e com o fenômeno climático *El Niño South Oscillation* (ENSO) (Potter et al., 2003). Na meteorologia, utiliza-se o termo teleconexão para descrever variações simultâneas entre clima e fenômenos relacionados em pontos distantes no planeta.

Grandes anomalias de precipitação na bacia amazônica são provocados pelo ENSO (SOMBROEK,2001), resultando em níveis altos de inundação do Rio Amazonas em época dos eventos de La Niña e níveis baixos de inundação em época dos eventos de El Niño.

Tendo em vista fornecer subsídios para uma previsão mais precisa das cheias sazonais, este trabalho tem como meta propor uma metodologia eficiente para otimizar a previsão do nível de cheias do Rio Negro. Com tal objetivo, serão utilizadas redes neurais artificiais, associados a métodos de otimização da generalização das mesmas. Outra contribuição importante do trabalho é a proposta de uma técnica original para seleção dos melh

ores índices climáticos que permitem uma previsão precisa do nível a cheia. Essa técnica é denominada nesse trabalho de técnica de seleção escalar de características modificada. Mais adiante, no capítulo de metodologia serão fornecidos mais detalhes sobre a mesma.

## **1.1 Organização do trabalho**

Após a apresentação da motivação desse estudo, feita na Introdução, o capítulo 1 apresenta os objetivos da pesquisa. Nos capítulos seguintes discorre-se sobre o arcabouço teórico da pesquisa, incluindo a revisão bibliográfica, os fundamentos teóricos, a metodologia e resultados.

O Capítulo 2 aborda uma revisão da bibliográfica de artigos com temas relacionados a previsões de cheias de rios, explorando detalhes das metodologias utilizadas pelos autores.

No Capítulo 3 são apresentados os conceitos básicos relacionados às redes neurais, os métodos utilizados para parada do treinamento das mesmas que conferem um melhor poder de generalização, como regularização e parada antecipada. Apresenta-se também o critério tradicionalmente utilizado para parada do treinamento da rede, o critério do erro médio quadrático. Apresenta-se também o método de seleção escalar de característica, utilizado na teoria de reconhecimento de padrões, e que foi utilizado como base para a proposta do método de seleção escalar de características modificado proposto nesse trabalho. Por fim, apresenta-se conceitos relativos aos testes de significância estatística utilizados no capítulo de resultados.

A modelagem do problema e os materiais utilizados para implementação do trabalho são apresentados no Capítulo 4.

Enfim, os resultados e conclusões são apresentados no Capítulo 5.

## **1.2 Objetivo Geral**

O objetivo geral deste trabalho é apresentar um novo método para otimização da previsão de cheias sazonais utilizando redes neurais artificiais.

## **1.3 Objetivo Específico**

São objetivos específicos deste trabalho:

- Propor uma nova técnica de seleção de variáveis para a previsão de cheias sazonais;
- Avaliar o desempenho de métodos de generalização de redes neurais na previsão de cheias sazonais, como regularização e parada antecipada;
- Propor uma nova metodologia de treinamento e teste de redes neurais para previsão de cheias sazonais.

## 2 REVISÃO BIBLIOGRÁFICA

Os trabalhos apresentados nessa revisão bibliográfica foram identificados a partir de uma busca exaustiva nas bases de dados do *Institute of Electrical and Electronic Engineers* (IEEE) e do *Engineering Village*, selecionando aqueles mais relevantes.

Previsões de rios é um tema importante em muitas aplicações da engenharia como construções de barragens para geração de energia, controle de inundações e outros. Existe uma variedade de modelos que têm sido propostos para previsões de cheias sazonais. Os modelos de RNAs têm sido úteis e eficientes quando utilizados para resolver problemas cujos processos envolvidos são difíceis de serem descritos por equações.

Sheta e El-Sherif. (SHETA e EL-SHERIF, 1999), utilizaram uma RNA multicamadas para desenvolver um método de cálculo para prever o fluxo do rio Nilo na estação Dongola, com dados de precipitação, de maneira que se pudesse fazer uma melhor análise sobre o período seco. O algoritmo utilizado para o treinamento da rede neural foi o de *backpropagation*. A arquitetura da rede foi constituída de uma camada de entrada, uma camada oculta e uma camada de saída. A função de ativação utilizada foi a tangente hiperbólica. Neste trabalho os autores analisam um algoritmo de treinamento para a RNA e confrontam os resultados obtidos com os resultados do modelo tradicional linear auto regressivo (AR), utilizando o erro médio quadrático para a comparação. Melhorias significativas foram encontradas utilizando a RNA.

Outros trabalhos de previsões de cheias se preocupam com os alertas emitidos em curtos intervalos de tempo, com o objetivo de reduzir os impactos de inundações. Gopakumar e Taraka (GOPARUMAR e TARAKA, 2009), utilizando RNA, desenvolveram um modelo para previsão de inundações do lago Vembanad, do estado de Kerala, na Índia. As previsões foram feitas com um dia de antecedência, usando dados limitados sobre níveis de água, chuvas e descargas fluviais. Os dados utilizados para a pesquisa eram a média diária de chuvas de duas estações, níveis de água de quatro estações e a média diária de descarga de cinco estações de

medição. Com esses dados foram simulados cinco modelos, utilizando o algoritmo de treinamento *backpropagation*. A função de ativação foi a sigmoide. O desempenho da rede e dos modelos propostos foi avaliado através do erro médio quadrático. O modelo com melhor desempenho para a previsão de inundações com um dia de antecipação do lago Vembanad, foi o que utilizava como entrada as precipitações e os níveis de águas de quatro estações.

Em 2011, Pang e Liang (PANG E LIANG, 2011) desenvolveram um novo modelo para previsão de inundações no baixo Rio Amarelo, na China. O método faz uma comparação com o modelo de perturbação linear (LPM), conhecido como o “modelo híbrido” (NASH e BRASI, 1983). Com o treinamento supervisionado e utilizando o algoritmo *backpropagation*, o modelo mostrou-se mais eficiente que o modelo LPM. Os dados utilizados foram as descargas de duas estações de medição na época das cheias (1 julho – 1 outubro) dos anos de 1998 – 2001, selecionados para treinar a RNA e a série de 2002-2003 para verificação do modelo.

Construções de barragens para usinas hidrelétricas, para serem economicamente viáveis e garantirem que as portas da barragem operem com segurança (protegendo áreas a jusante de inundação), necessitam de precisões de fluxo de água. Egawa *et al* (EGAWA et al., 2011), desenvolveu um método para previsão de fluxo de água para barragem utilizando RNA e regressão linear. Os autores utilizaram dados captados de onze barragens. Para as que possuíam poucos dados, utilizou-se o modelo de regressão linear e para as que possuíam muitos dados, utilizou-se o modelo com RNA. As previsões foram feitas apenas nos meses de junho a novembro, pois em outros meses não era possível medir a chuva, por conta da queda de neve. O método mostrou-se eficiente para previsão de fluxo de água em barragem com três horas de antecedência.

Além das RNA's, outras ferramentas são utilizadas para previsões de cheias, tais como: Máquinas de Vetores de Suporte (*Support Vector Machine* - SVM), que analisam dados e reconhecem padrões, usando classificação e regressão (VAPINIK, 1998). Hu et al. (HU et al.,

2011), utilizou SVM para prever a vazão do rio Fenhe, afluente do rio Amarelo na China. Com um conjunto de dados relativo ao nível de ocorrência de chuvas, coletados diariamente e mensalmente, foi feita a previsão de vazão e escoamento, no período correspondente aos anos de 1990 a 2005. Comparando o modelo SVM com o modelo RNA, os resultados mostraram que o SVM mostrou uma maior capacidade para realizar mapeamentos não-lineares e, assim, representar mais padrões de dados que as RNA's (HU et al. 2011).

Outros métodos de previsão de inundações se preocupam em oferecer uma previsão mais precisa, fornecendo previsões dos níveis de rios em função do dia e da hora. O objetivo desses trabalhos é gerar alertas à população. Chang et al. (CHANG et al., 2011), desenvolveu, para as inundações do rio Lanyoung, em Taiwan, um modelo baseado em Redes Neurais de Base Radial (RBF), previsões de chuva-vazão para 3 horas à frente. A metodologia desenvolvida é constituída por duas fases: A primeira fase consiste de um aprendizado não supervisionado, utilizado para determinar os parâmetros das RBF's. A segunda fase consiste de um aprendizado supervisionado, onde a regressão múltipla é usada para determinar os pesos entre as camadas escondidas e de saída. Trata-se de um rio de montanhas com declive íngreme, que inunda em poucas horas após uma tempestade. Com resultados precisos, a rede RBF mostrou-se eficiente para a previsão de inundações.

Comparando o modelo SVM com o modelo perceptron de multicamada (MLP), Theera et al. (THEERA et al., 2008) utilizou SVM para previsão de inundações na cidade de Chiang Mai, na Tailândia, com 1 hora e com 7 horas de antecedência. Foram utilizados dados coletados do nível do nível do rio de 3 estações a montante de Chiang Mai. Esses dados foram divididos em 7 grupos, formando conjuntos de treinamento, validação e teste. Para as previsões com uma hora de antecedência, os dois modelos apresentaram o mesmo desempenho na previsão do nível do rio. Já nos conjuntos de validação para previsões com 7 horas de antecedência, o modelo SVM mostrou melhores resultados quando comparado ao modelo MLP.

Novos trabalhos têm mostrado correlações entre as precipitações na Amazônia e os fenômenos que ocorrem nos oceanos pacíficos e atlântico, fazendo assim, a previsão de cheias do rio Amazonas (Schongart e Junk, 2007, Coe *et al.*, 2002. e Uvo *et al.*, 2000).

Uvo et al. (UVO et al., 2000), constatou uma correlação entre as temperaturas da superfície do mar dos oceanos Pacífico e Atlântico (TSMs) com os níveis de precipitações na bacia Amazônia. Para previsão da descarga do rio Amazonas os autores utilizaram RNA com arquitetura 8x20x1. A RNA foi treinada com o algoritmo de *backpropagation*. Para o treinamento e teste foi utilizada o método de validação cruzada com 4 pastas. Como variáveis de entrada da RNA foram utilizados o nível médio do índice SST no pacífico equatorial, nos meses de dezembro, janeiro e fevereiro e o total de precipitações nos meses de setembro, outubro e novembro, somadas em três estações a jusante de Manaus. Com essas variáveis de entrada foram obtidos os melhores resultados de previsão de descarga. A qualidade da previsão foi então estimada por meio da correlação entre os valores observados e estimados da série dos níveis do rio e também pela soma dos quadrados dos erros. Assim, com uma correlação de 0,76 nessas 3 estações atingiu o melhor resultado.

Utilizando correlação múltipla e fazendo uma análise do período de 1903 a 2004, Schongart e Junk (SCHONGART e JUNK, 2007), obtiveram uma alta correlação entre o nível máximo do Rio Negro em um ano com as seguintes variáveis: índice SOI (*Southern Oscillation Index*) de fevereiro e o nível do rio em fevereiro do mesmo ano ( $r = 0,45$ ,  $p < 0,001$ ). Houve uma boa congruência entre os valores observados e os previstos para os níveis de cheia do Rio Negro. Em 46,5% dos casos, as diferenças entre os níveis de cheia observadas e previstas são inferiores a 50 cm. Em 81,2 % dos casos, inferiores a 100 cm, e em 97,0% dos casos, inferiores a 150 cm.

Coe et al. (COE et al., 2002), para simular a descarga as variações sazonais da inundação no Rio Amazonas, no período de 1939 a 1998, utilizaram o algoritmo de roteamento hidrológico

(HYDRA - *Hydrological Routing Algorithm*). Trata-se de um simulador complexo que faz uso de uma modelagem com equações diferenciais e diversos tipos de informações de entrada: a topografia da região, dados de evaporação, dados climáticos de precipitações e informações de escoamento do solo. A conclusão dos mesmos, no entanto, é bastante limitada, e diz respeito apenas ao período 1992-1998. Para esse período, os autores simularam os desvios do nível do rio em relação à média do período. Particularmente, na confluência dos rios Negro e Solimões, os autores obtiveram uma boa correlação com os dados reais, conseguindo um coeficiente de correlação de  $r^2 = 0,79$  entre a descarga simulada e a observada.

O Quadro 1 apresenta um resumo geral dos artigos revisados, destacando o ano do trabalho, título, sua aplicação, características, dados de entrada, materiais e resultados.

**Quadro 1: Resumo da Revisão Bibliográfica**

Ano	Autor	Título	Aplicação	Características	Variáveis de Entrada	Materiais e Métodos	Resultados
1999	Alaa F. Sheta, Mohammed S. El-Sherif	Optimal Prediction of the Nile River Flow Using Neural Networks - Previsão Optimal do Nilo Fluxo Rio Utilizando Redes Neurais	Comparação entre os modelos de regressão linear (AR) e Redes Neurais (RNA)	Previsão do fluxo do Rio nilo	(Volume médio do fluxo diário para cada período de 10 dias em uma estação no Egito) Cj de dados coletados de eventos anteriores dos anos de 1975 a 1993 - os dados de 1975 a 1984 foram utilizados para treinamento e o dados de 1984 a 1993 foram utilizados para teste	Redes Neurais, utilizando o algoritmo BP. Usando um par para treinamento (1975 - 1983) e outro para para teste (1984 - 1993). Com arquitetura 5x5x1	Através do NRMSE, a comparação entre os dois modelos o RNA mostrou-se mais eficiente que o auto regressão AR diminuindo o erro de forma significativa
2000	Cíntia Bertacchi Uvo*, ute to8 lle and ronny berndtsson	Forecasting discharge in amazonia using artificial neural networks (previsão de descarga na amazônia usando Redes Neurais)	Modelo RNA para previsões	Previsão de descargas na Amazonia	Anomalias TSM, SST, séries mensais de precipitações de 3 estações a jusante de Manaus	Arquitetura de 8-20-1, treinamento da RN feito usando gradiente descendente e taxa de aprendizagem adaptativa Back-Propagation	as RNA's teve seu melhor desempenho tendo sua entrada as anomalias SST e os níveis de precipitação como alvo. E variação da correlação SST e níveis nas estações. Com um coeficiente de correlação de 0,86

2002	Michael T. Coe, Marcos Heil Cost, Aurélie Botta, and Charon Birkett	Long-term simulations of discharge and floods in the Amazon Basin (Simulações de longo prazo de descargas e inundações na Bacia Amazônica)	Algoritmo de roteamento hidrológico para simular a descargas e idundações na bacia Amazônica	Previsão de descargas e inundações na Amazônia	Índices Climáticos de precipitações; topografia da região; dados de evaporação; escoamento do solo; níveis do rio nas estações de medição dos períodos de 1939 a 1998.	Simulador complexo com modelagem com equações diferenciais.	Para o período de 199-1998 conseguiram um coeficiente de correlação entre a descarga simulada e observada de $r^2=0,79$
2007	Schongart and Junk 2007	Forecasting the flood-pulse in Central Amazonia by ENSO-indices (Previsão do nível de água no Rio Negro, localizado na Manaus, Brasil.)	Water level 4 months before the maximum water level that occurs in the second half of June (O nível de água quatro meses antes que o nível máximo de água que ocorre na segunda quinzena de junho)	Statistical correlation	Pacific South Oscillation Index (SOI) and water level at february	Regressão linear	Um coeficiente de correlação de 0,71, $p < 0,0001$ , no período de 1903-2004.
2008	Sansanee Auephanwirivakul, Sitawit Suteepohnwirroj Nipon Theera-Umpon,	River basin flood prediction using support vector machines (Previsão de inundações da bacia do rio usando máquinas de vetores de suporte)	Modelo para previsão de inundações da cidade de Chiang Mai, Na Tailândia	Previsão de inundações do rio Ping na cidade Chiang Mai, Tailândia em 1 hora e 7 horas de antecedência	Dados de níveis de rio coletados nas 3 estações a montante (medidores flutuantes) da e de 2006) estação de predição - 7 eventos cruciais de inundações (5 do ano de 2005 e 2 do na	Utiliza Máquina de Suporte de Vetores (SVM), fazendo uma comparação com o preceptron de multicamada (MLP)	Através da percentagem de erro médio absoluto (MAPE), observou-se que o modelo SVM é mais eficiente que o MLP

2009	R. Gopakumar; K. Takara	Water Level Forecasting in the Vembanad Water System Using Artificial Neural Networks (Previsão de nível de água nas estações do rio Vembanad Usando Redes Neurais Artificiais)	Modelo para previsão de inundações	Previsão do nível de água em 11 estações do rio Vembanad com 1 dia de antecedência	Conjunto de dados de precipitações e níveis de água coletados em 11 estações no período de 1996 a 1999	Redes Neurais, utilizando o algoritmo BP. Com a função de ativação sigmoide com a verificação do erro médio quadrático	os melhores resultados foram obtidos quando o modelo utilizava as precipitações e os níveis de água em quatro estações
2011	T. Egawa, K. Suzuki, Y. Ichikawa, T. Iizaka, T. Matsui and Y. Shikagawa	A Water Flow Forecasting for Dam using Neural Networks and regression models ( Previsão de fluxo de água para barragem utilizando redes neurais e modelos de regressão)	Utilização de RNA e regressão linear paara previsão de inundações em barragens	Previsão de inundação do rio Tadami- Agano com 3 horas de antecedência apenas para os meses de junho a novembro.	Dados de precipitações coletados em onze barragem, utilizando os com poucos dados para o modelo de Regreção Linear e os com maiores dados para os modelos com RNA	Regressão linear para as barragens com poucos dados de precipitações e RNA para as Barragens com maiores dados	as RNA's mostrou-se eficnete da previsão em estações com afluentes de outros rios e o modelo de regressão para as barragens que não tinham rios afluentes
2011	Cai-hong Hu, Ze-ning Wu, Ji-jun Wang, Lina-Liu,	Application of the Support Vector Machine on Precipitation-Runoff Modelling in Fenhe River (Aplicação do Support Vector Machine na modelagem chuva - vazão no rio Fenhe)	Utilização de SVM para previsões de chuva e vazão do Rio Fenhe	Modelagem para previsão de iundação do rio Fenhe com até 6 horas de antecedência	cj de dados coletados diariamente e mensalmente de chuvas, vazão, escoamento dos anos de 1990 - 2005	Utiliza Maquina de Suporte de Vetores (SVM)	Através da comparação dos coeficientes de Nash R2 e o erro relativo RE, o Modelo SVM mostrou-se mais eficiente que o RNA

2011	Fi - John Chang, Jin-Ming Liang, and Yen-Chang Chen	Flood forecasting using radial basis function neural networks (Previsão Flood Usando Radial Basis Function Redes Neurais)	Utilização de RBF para previsão de inundações	Modelagem para previsão de inundações do rio Lanyoung com 3 horas de antecedência	Níveis de rios das estações de medição	RNA com a Função Base Radial (RBF) - modelo híbrido	Mostrou-se eficiente para previsões de inundações
2011	Bo Pang, Yuan Liang	Application of an improved neural network to flood forecasting of the Lower Yellow River (Aplicação de uma rede neural melhorada a previsão das inundações do Baixo Rio Amarelo)	Comparação entre os modelos NLPM-ANN LPM e ANN - (RNA - BP - séries temporais-regressão linear) para previsão de inundações do Rio Amarelo	Previsão de inundações do Rio Amarelo para alertas em 36 horas	Chuvas, conforme a sazonalidade - dados coletados 1998-2001 (treinamento) e 2002-2003 (verificação)	Utilização de um modelo híbrido para comparar redes neurais com a perturbação linear	o modelo NLPM-ANN foi o mais eficiente que os modelos LPM e ANN

### 3 FUNDAMENTOS TEÓRICOS

Neste capítulo são apresentados tópicos teóricos fundamentais relacionados ao trabalho proposto. Será destacado o uso de RNA's como modelos para predições, empregando o algoritmo de retropropagação do erro (*backpropagation*) associado ao método de otimização *Levenberg-Marquadt* (CICHOKI e UNBEHAUEN, 1993). Tradicionalmente, como critério de parada, as RNA's utilizavam apenas o erro médio quadrático. Nesse trabalho, no entanto, serão utilizados outros métodos de parada das RNA's, com o objetivo de melhorar a poder de generalização das mesmas, métodos de regularização e de parada antecipada. Assim sendo, nesse capítulo também serão revisados conceitos pertinentes a esses dois métodos. Com o objetivo de fundamentar o novo método apresentado nesse trabalho para seleção das variáveis mais relevantes para a predição, o método de seleção escalar de características modificado, será apresentado nessa seção o método de seleção escalar de características (THEODORIDIS AND KOUTROUMBAS, 2006), utilizado em reconhecimento de padrões, que motivou a proposição do primeiro. O capítulo será finalizado com a apresentação dos testes de hipótese estatísticas utilizadas para avaliação dos resultados obtidos nesse trabalho.

#### 3.1 Redes Neurais Artificiais – RNA's

“As RNA's são processadores massivamente paralelos e distribuídos, que têm uma força natural para armazenar o conhecimento proveniente da experiência e torna-lo útil. Desta forma, assemelhando-se ao cérebro humano em dois aspectos:

1. O conhecimento é adquirido pela rede através de um processo de aprendizado e
2. As intensidades das conexões entre neurônios, conhecidos como pesos sinápticos, são usados para armazenar o conhecimento”. (HAYKIN, 1994)

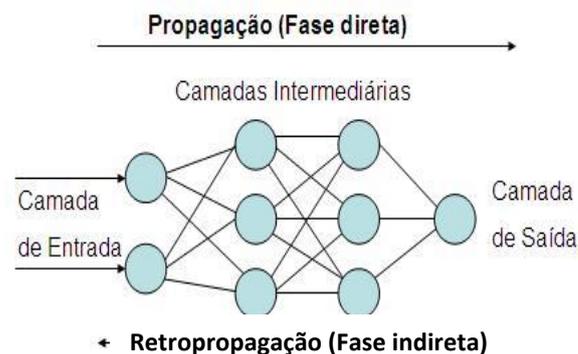
De acordo com Braga et al. (BRAGA et al, 2000), RNA's são modelos matemáticos que se assemelham às estruturas biológicas e que tem capacidade computacional adquirida por meio de aprendizado e generalização.

Muitos problemas da ciência e da engenharia tem sido resolvido com as aplicações de RNA's. Demonstra potencialidade em diversas aplicações, superando expectativas e gerando resultados até então não alcançados com outra técnica, seja computacional ou convencional.

A arquitetura das Redes Neurais de Múltiplas Camadas (RNMC) é a mais encontrada na literatura (SKAPURA, 1996), embora existam inúmeras arquiteturas. Pode-se citar como razão de sua popularidade a capacidade de aproximação universal e a flexibilidade para formar soluções de qualidade para uma ampla classe de problemas, a partir de um mesmo algoritmo de aprendizado.

Tipicamente, as arquiteturas do tipo RNMC consistem em um conjunto de unidades sensoriais que formam uma camada de entrada, uma ou mais camadas intermediárias (ocultas) de unidades computacionais e uma camada de saída. A camada de entrada propaga os sinais de entrada na direção da camada de saída. As camadas intermediárias funcionam como extratoras de características. Seus pesos são uma codificação de padrões presentes nos dados de excitação de entrada e permitem que a rede crie sua própria representação, mais rica e complexa. São necessárias, no máximo, duas camadas intermediárias (com um número suficiente de nodos por camada) para se produzir qualquer mapeamento não linear. Apenas uma camada intermediária é suficiente para aproximar qualquer função contínua (CYBENCO, 1988). Na Figura 2 apresenta-se a arquitetura de uma RNA com duas camadas intermediárias.

Para os vários problemas envolvendo altos graus de não-linearidade, as redes do tipo RNMC têm sido utilizadas como solução com sucesso. Na sua forma mais conhecida e utilizada as RNA's empregam treinamento supervisionado e utilizam o algoritmo, como algoritmo de treinamento, o algoritmo de retropropagação de erro, ou uma de suas variações.



**Figura 2. Rede Neural de Múltiplas Camadas**

A capacidade de generalização é resultado da habilidade de emitir respostas aceitáveis e coerentes, mesmo para pontos que não compuseram o conjunto de treinamento da rede. Em destaque se coloca a habilidade em extrair informação a partir de dados de entrada ruidosos e mapear relações não lineares complexas entre entradas e saídas, com excelente capacidade de generalização.

Estas características transformam as RNMC's em aproximadores universais de funções com resultados surpreendentes e, na maioria das vezes, melhores que os obtidos por procedimentos estatísticos convencionais.

A principal limitação das redes neurais é não ter capacidade de explicação. Os dados entram na rede e uma previsão sai, mas o tipo de relacionamento entre as variáveis não é revelado (FRANCIS, 2001), assim como detalhes de como são processados os dados para se chegar aos mapeamentos de resposta.

Por este motivo, as redes neurais são recomendadas para serem aplicadas em áreas de conhecimento cujas teorias ainda não conseguem explicar adequadamente o comportamento dos fenômenos observados ou em áreas de grande complexidade que não necessitam de

modelos precisos da realidade física do problema. Ou seja, a maior limitação das redes neurais não representa um problema para a sua aplicação na ciência e na engenharia.

Pode-se sintetizar o entendimento das RNA em três palavras: neurônio, arquitetura e aprendizagem (HAGAN *et al*, 1995).

O neurônio é a unidade computacional básica, a arquitetura é a estrutura topológica de como os neurônios são conectados e a aprendizagem é o processo que adota a rede de modo a computar uma função desejada ou realizar uma tarefa.

### **3.2 Treinamento das RNMC**

As RNAs caracterizam-se pelo aprendizado por meio de exemplos. Para um determinado conjunto de dados, o algoritmo de aprendizado deve ser responsável pela adaptação dos parâmetros da rede, de maneira que, em um número finito de iterações do algoritmo, haja convergência para uma solução. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas (BARRETO, 1995).

Considera-se que o processo de aprendizado tem como pressuposto a ocorrência de um estímulo externo para a rede, advindo do meio externo. Trata-se da apresentação de um conjunto de treinamento à camada de entrada. Como consequência deste estímulo, o algoritmo de aprendizado provoca mudança nos parâmetros da rede e, finalmente, a mudança nos parâmetros acarreta em mudança no comportamento da RNA. A modificação do processamento ou da estrutura de conhecimento de uma rede neural envolve a alteração do seu padrão de interconexão. Isto pode ocorrer de três maneiras: desenvolvimento de novas conexões; perda de conexões existentes na rede ou modificação dos pesos das conexões já existentes (KOVACS, 1996).

O objetivo principal do aprendizado em Redes Neurais é a obtenção de modelos com boa capacidade de generalização, tendo como base o conjunto de dados (BRAGA *et al*, 2000). O objetivo dos algoritmos de treinamento deve ser o de aproximar as funções geradoras dos dados e não simplesmente o de minimizar o erro do conjunto de treinamento.

Os procedimentos de treinamento que levam as RNA's a aprenderem determinadas tarefas podem ser classificados em três classes:

a) Treinamento supervisionado – Esse tipo de treinamento necessita de um par de vetores composto do vetor de entrada e do vetor alvo que se deseja como saída. Juntos, estes vetores são chamados de pares de entrada-saída, sendo interessante ressaltar que geralmente a rede é treinada com vários pares entrada-saída. O procedimento de treinamento funciona da seguinte forma: o vetor de entrada é aplicado. A saída da rede é calculada e comparada com o correspondente vetor alvo. O erro encontrado é então realimentado através da rede e os pesos são atualizados de acordo com um algoritmo determinado, a fim de minimizar este erro. Este processo de treinamento é repetido até que o erro para os vetores de treinamento tenha alcançado níveis bem baixos.

b) Treinamento não supervisionado – Esse tipo de treinamento não requer vetor alvo para as saídas e, obviamente, não faz comparações para determinar a resposta ideal. O conjunto de treinamento modifica os pesos da rede de forma a produzir saídas que sejam consistentes, isto é, tanto a apresentação de um dos vetores de treinamento, como a apresentação de um vetor que é suficientemente similar, irão produzir o mesmo padrão na saída. O processo de treinamento extrai as propriedades estatísticas do conjunto de treinamento e agrupa os vetores similares em classes. A aplicação de um vetor de uma determinada classe à entrada da rede irá produzir um vetor de saída específico, mas não existe maneira de se determinar, antes do treinamento, qual o padrão que será produzido na saída para um vetor de entrada de uma determinada classe. Desta forma, a saída de RNA's treinadas com algoritmo não supervisionado

deve ser transformada em uma forma compreensível após o processo de treinamento, o que é um simples problema de identificação das relações entrada-saída estabelecidas pela rede.

c) Aprendizagem híbrida – Esse tipo de treinamento mescla os conceitos apresentados anteriormente de treinamento supervisionado e não supervisionado. Parte dos pesos é determinada através da aprendizagem supervisionada, enquanto outra parte é obtida através da aprendizagem não-supervisionada. Em um processo de aprendizagem, os pesos dos neurônios são ajustados através de um algoritmo de aprendizagem.

### 3.3 Características da Aprendizagem

Tipicamente dois critérios de paradas são mais utilizados para se parar o treinamento de uma RNMC: número de interações ou valor do erro médio quadrático. Dessa maneira, cessa-se o treinamento ao se atingir um número pré-definido de iterações ou ao se atingir um erro médio quadrático pré-definido.

A experiência com redes neurais revelou que, após um determinado ponto crítico, a RNMC treinada com o algoritmo de retropropagação, ou uma de suas variações, o erro cometido pela rede no conjunto de treinamento continuará a diminuir, porém o desempenho da rede no conjunto de teste tende a piorar. Este fenômeno é conhecido como *overtraining* (PRINCIPE et al, 2000).

Dois métodos são mais utilizados para melhorar o desempenho da rede neural no conjunto de teste: regularização e parada antecipada.

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 (3.1).

$$msereg = \gamma mse + (1 - \gamma) msw \quad (3.1)$$

Em que:

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

mse – erro médio quadrático.

$$m_{sw} = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (3.2)$$

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.

No método de parada antecipada, o conjunto de dados disponível é dividido em três subconjuntos: o conjunto de treinamento, o conjunto de teste e o conjunto de validação. O conjunto de validação possui geralmente 10% do tamanho do conjunto de treinamento. Na entrada da rede é apresentado apenas o conjunto de treinamento, que é utilizado para ajustar os pesos da rede através do algoritmo de *backpropagation*.

O conjunto de validação é empregado para parar o treinamento. O treinamento cessa quando o erro atinge um valor mínimo no conjunto de validação. Verifica-se, experimentalmente, que o erro no conjunto de validação aumenta quando as interações do treinamento prosseguem após esse ponto de mínimo ser alcançado. O treinamento deve, portanto, cessar no ponto de menor erro dentro do conjunto de validação.

O tamanho do conjunto de treinamento influencia diretamente a performance de qualquer RNMC. Devido ao total desconhecimento das correlações existentes no conjunto de dados, as RNMC requerem um grande conjunto de dados para o treinamento. O tamanho do conjunto de treinamento é uma função que depende do tamanho da RNMC e da exatidão necessária ao mapeamento.

O número de padrões de treinamento requeridos para classificar exemplos de teste com um erro é aproximadamente dado por  $N > W / \delta$ , onde  $W$  é o número de pesos da rede e  $\delta$  o erro permitido sobre os dados de teste (BRONSON, 1995). Esta equação expressa o fato de que

o número de padrões de treinamento requeridos aumenta linearmente com o número de parâmetros livres da RNMC. Uma regra de aproximação é a de que  $N \approx 10W$ , o que representa que o tamanho do conjunto de treinamento deve ser 10 vezes maior do que o número de pesos da rede, para uma maior exatidão nos resultados.

A escalabilidade é a característica das RNMC em atribuir um fator de escala às suas propriedades, proporcional ao crescimento do tamanho dos problemas. A literatura é cheia de exemplos de sistemas que funcionam bem para pequenos problemas, mas são ineficazes em estender o mesmo desempenho para problemas maiores. Um achado importante na análise do erro médio quadrático para problemas de várias dimensões mostra que, para um grande conjunto de treinamento, o erro da RNMC é independente do tamanho do espaço de entrada e é escalável ao inverso do número de neurônios presentes na camada escondida (HSU et al., 1996). Isto é muito melhor do que os métodos de aproximação baseados em polinômios, onde o erro decai geometricamente com a dimensão do espaço de entrada. As RNMC são excelentes ferramentas para problemas com grandes espaços de entrada. Isto explica a excelente performance das RNMC em problemas grandes de classificação de padrões.

O tempo de treinamento de uma RNMC treinada com o algoritmo *backpropagation* cresce exponencialmente de acordo com o tamanho do problema. Na prática, esta escalabilidade exponencial no tempo de treinamento sugere como mais interessante iniciar com RNMC pequenas e aumentá-las até que os resultados sejam satisfatórios.

### **3.4 O Algoritmo *Backpropagation***

O algoritmo *backpropagation*, ou algoritmo de retropropagação do erro, é o mais conhecido e utilizado algoritmo para aprendizado das RNA's. Foi desenvolvido por

Rummelhart, Hinton e Williams em 1986 (RUMMELHART *et al*, 1986) e consiste numa retropropagação de erro através das diversas camadas de uma RNMC.

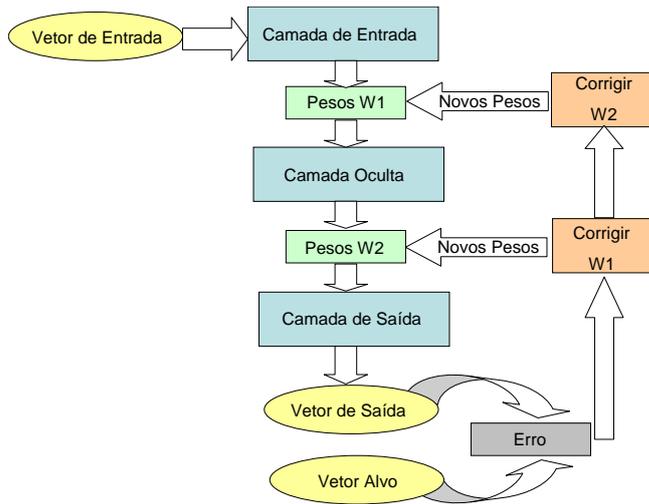
O algoritmo *backpropagation*, em sua forma mais simples, é um algoritmo de gradiente descendente, assim como a regra de aprendizagem de Widrow-Hoff. Para o caso específico de aproximação de funções, uma RNMC é treinada com pares de vetores de entrada e saída. Durante o treinamento há o ajuste dos pesos sinápticos.

Basicamente, o processo de retropropagação do erro é constituído de duas fases: uma fase de propagação do sinal funcional (direta) e uma de retropropagação do erro (propagação reversa). Na fase positiva, os vetores de dados são aplicados às unidades de entrada, e seu efeito se propaga pela rede, camada a camada.

Finalmente, um conjunto de saídas é produzido como resposta da rede. Durante a fase de propagação direta, os pesos das conexões são mantidos fixos. Na retropropagação do erro, ou propagação reversa do erro, por outro lado, os pesos são ajustados de acordo com uma regra de correção do erro.

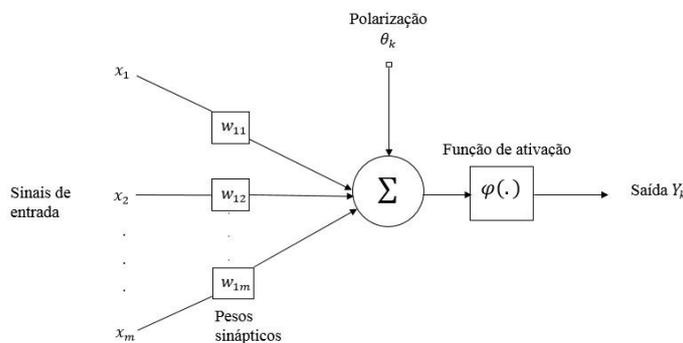
Especificamente, a resposta da rede em um instante de tempo  $s(t)$  é subtraída da saída desejada  $s_d(t)$  para produzir um sinal de erro na saída rede. Este sinal de erro é propagado da saída para a entrada, camada a camada, originando o termo “retropropagação do erro”. Os pesos são ajustados de forma que a “distância” entre a resposta da rede e a resposta desejada seja reduzida.

De forma simplificada o algoritmo de retropropagação pode ser expresso conforme o fluxo de processos apresentado na Figura 3.



**Figura 3. O algoritmo *Backpropagation***

O modelo atual para representar um neurônio artificial tem origem no neurônio formal proposto por Warren McCulloch e Walter Pitts (MC CULLOCH e PITTS, 1943). Na Figura 4 mostra-se esse modelo. O  $k$ -ésimo neurônio artificial mostrado nessa figura é modelado com os seguintes elementos: entradas ( $x_i$  representa a saída do neurônio  $i$ ); peso  $w_{ki}$  (ligação entre a saída do neurônio  $i$  e a entrada do neurônio  $k$ ), somador (soma ponderada,  $v_k$ ). Cada termo da soma corresponde ao produto entre uma entrada do neurônio  $k$  pelo respectivos peso:  $x_i w_{ki}$ ); polarização ( $\theta_k$ ), constante adicionada à soma ponderada  $v_k$ ; função de ativação ( $\varphi$ ). Para simplificar a notação, a polarização  $\theta_k$  pode ser considerado como o peso  $w_{k0}$  associado a uma entrada constante  $x_0=1$ , conforme Equações (3.5) e (3.7).



**Figura 4. Modelo Matemático de Neurônio**

O treinamento supervisionado para uma RNMC consiste em determinar valores adequados para os pesos (incluindo a polarização) de todos os neurônios que compõem as camadas intermediárias e de saída. A atualização desses pesos pode acontecer de duas formas: incremental ou em lote. No primeiro caso a atualização dos pesos se dá à medida que cada padrão entrada-saída do conjunto de treinamento é apresentado. No segundo, a atualização se dá apenas depois de finalizada uma apresentação completa do conjunto de treinamento. Os pesos são ajustados no sentido de que atendem a um critério ou índice de desempenho, que visa minimizar a diferença entre os valores de saída previstos pela rede e os valores de saída dos pares entrada-saída do conjunto de treinamento.

Seja o erro no instante  $t$  em um neurônio  $k$ , dado por:

$$e_k(t) = s_d(t) - s(t) \quad (3.3)$$

Onde  $s_d(t)$  é a saída desejada e  $s(t)$  é a saída do neurônio.

A soma dos erros quadráticos de todos os neurônios de uma RNMC é dado por:

$$\varepsilon(t) = \frac{1}{2} \sum_{k \in C} e_k^2(t), \quad (3.4)$$

Em que o conjunto  $C$  inclui todos os neurônios da camada de saída.

De acordo com o modelo do perceptron apresentado na Figura 4, a soma do produto dos pesos da camada de entrada pelas entradas é dada por:

$$v_k(t) = \sum_{i=1}^M w_{ki} x_i(t) + \theta(t), \quad (3.5)$$

Em que  $M$  é o total de entradas aplicadas ao neurônio  $k$ . E também:

$$y_k(t) = \varphi_k(v_k(t)), \quad (3.6)$$

O processo de busca de parâmetros é similar ao algoritmo LMS, ou seja, calculando-se as derivadas parciais  $\frac{\partial \varepsilon(t)}{\partial w_{ki}(t)}$ . Aplicando-se a regra da cadeia:

$$\frac{\partial \varepsilon(t)}{\partial w_{ki}(t)} = \frac{\partial \varepsilon(t)}{\partial e_k(t)} \bullet \frac{\partial e_k(t)}{\partial s(t)} \bullet \frac{\partial s(t)}{\partial v_k(t)} \bullet \frac{\partial v_k(t)}{\partial w_{ki}(t)}, \quad (3.7)$$

Em que as derivadas parciais de cada um dos termos é:

$$\begin{aligned} \frac{\partial \varepsilon(t)}{\partial e_k(t)} &= e_k(t) \\ \frac{\partial e_k(t)}{\partial s(t)} &= -1 \\ \frac{\partial s(t)}{\partial v_k(t)} &= \varphi'_k(v_k(t)) \\ \frac{\partial v_k(t)}{\partial w_{ki}(t)} &= x_i(t) \end{aligned} \quad (3.8)$$

Logo, a regra de aprendizagem será:

$$\frac{\partial \varepsilon(t)}{\partial w_{ki}(t)} = -\lambda e_k(t) \cdot \varphi'_k(v_k(t)) \cdot x_i(t), \quad (3.9)$$

Em que  $e_k(t) \cdot \varphi'_k(v_k(t)) = \zeta_k(t)$ .  $\lambda$  é a taxa de aprendizagem e  $\zeta_k(t)$  é o gradiente local.

As equações de adaptação para as polarizações dos neurônios são análogas as equações para adaptação dos pesos. Dessa maneira, o algoritmo de aprendizagem dos  $\theta_k$  obedecem também ao gradiente descendente local. As equações de correção da polarização podem ser escritas como:

$$\begin{aligned} \theta_k(t) &= -\lambda \frac{\partial \varepsilon(t)}{\partial \theta_k(t)} \\ \theta_k(t) &= -\lambda \zeta_k(t) \end{aligned} \quad (3.10)$$

Existem algumas otimizações do algoritmo básico de retropropagação que visam acelerar o processo de convergência do mesmo. Na seção seguinte será abordada em detalhes a otimização proposta por Levenberg-Marquadt (CICHOKI E UNBEHAUEN, 1993), que é utilizada nesse trabalho.

### 3.5 Método de otimização Levenberg – Marquadt

O algoritmo de retropropagação de erro possui uma otimização chamada de algoritmo de Levenberg-Marquardt. O treinamento de redes neurais com esse algoritmo é mais rápido do que com a aplicação do algoritmo padrão de retropropagação de erro e suas variantes.

Esta seção introduz conceitos básicos sobre algoritmos de aprendizagem baseados no gradiente da superfície do erro e os utiliza na descrição do algoritmo de Levenberg-Marquardt.

O método Levenberg-Marquardt utiliza treinamento em lote e é uma extensão do método de Gauss-Newton, que é uma variante do método de Newton.

O método de Newton usa a informação da derivada parcial de segunda ordem do índice de performance utilizado para corrigir os pesos  $\bar{w}$ . Isso permite que além da informação do gradiente  $\nabla$  seja usada informação sobre a curvatura da superfície do erro (RANGANATHAN, 2004).

O método de Newton pode ser usado para resolver a equação:

$$\nabla V(\bar{w}) = 0 \quad (3.11)$$

Expandindo o gradiente de V em uma série de Taylor em torno dos pesos  $w_0$ , obtemos:

$$\nabla V(\bar{w}) = \nabla V(\bar{w}_0) + (\bar{w} - \bar{w}_0)^T \nabla^2 V(\bar{w}_0) + \dots \quad (3.12)$$

Desprezando os termos de ordem superior a 2 e resolvendo a equação (3.12) em relação a  $\nabla \bar{w}$  obtemos a regra de atualização de  $\bar{w}$ , dada pela equação (3.13) e de maneira mais geral pela equação (3.12).

$$\Delta w = -[\nabla^2 V(w_0)]^{-1} \nabla V(w_0) \quad (3.13)$$

$$\Delta \bar{w} = -[\nabla^2 V(\bar{w})]^{-1} \nabla V(\bar{w}) \quad (3.14)$$

Ao desprezarmos os termos de ordem superior, assumimos a aproximação quadrática para o gradiente. Embora a convergência seja rápida, pode ser bastante sensível a posição inicial

na superfície do erro, função dos valores dos pesos iniciais em  $w_0$ . O índice de performance  $V$  pode ser definido como a soma do quadrado do erro  $e_i$ , associado ao  $i$ -ésimo padrão de treinamento da rede, conforme a equação (3.15):

$$V(\bar{w}) = \sum_{i=1}^N e_i^2(\bar{w}) \quad (3.15)$$

Então o gradiente e o hessiano podem ser expressos como funções do jacobiano  $J$  (derivadas parciais) do índice de performance  $V$ , como:

$$\nabla V(\bar{w}) = J^T(\bar{w})\bar{e}(\bar{w}) \quad (3.16)$$

$$\nabla^2 V(\bar{w}) = J^T(\bar{w})J(\bar{w}) + S(\bar{w}) \quad (3.17)$$

$$S(\bar{w}) = \sum_{i=1}^N e_i(\bar{w})\nabla^2 e_i(\bar{w}) \quad (3.18)$$

A avaliação do hessiano pode ser extremamente complexa. Para contornar esse problema, foram propostos métodos que utilizam aproximações, denotados Quasi-Newton, dentre os quais se situam o método de Gauss-Newton e Levenberg-Marquardt. No método de Gauss-Newton assume-se  $S(\bar{w}) \approx 0$ , então a regra de atualização de Newton passa a ser:

$$\Delta\bar{w} = -[J^T(\bar{w})J(\bar{w})]^{-1} J^T(\bar{w})\bar{e}(\bar{w}) \quad (3.19)$$

O problema com esse método é que a matriz  $H = [J^T(\bar{x})J(\bar{x})]$  pode não ter inversa. Para contornar essa situação, Levenberg propôs somar a parcela  $\mu I$  a essa matriz, onde  $\mu$  é um escalar e  $I$  é a matriz identidade, resultando na seguinte regra de atualização (HAGAN *et al*, 1995):

$$\Delta\bar{w} = [J^T(\bar{w})J(\bar{w}) + \mu I]^{-1} J^T(\bar{w})\bar{e}(\bar{w}) \quad (3.20)$$

A matriz  $[J^T(\bar{w})J(\bar{w}) + \mu I]$  sempre possuirá inversa (RANGANATHAN, 2004).

Após uma atualização, se o valor do índice de performance diminui,  $\mu$  deve ser diminuindo para reduzir a influência do gradiente descendente. Se ao contrário, o valor do índice de performance aumenta, então seguir a direção do gradiente descendente é a melhor escolha e o valor de  $\mu$  deve ser aumentado.

Se  $\mu$  tornar-se muito grande, a informação dada pela aproximação H para o hessiano não será útil no cálculo da atualização de  $\bar{w}$ . Para contornar esse problema, Marquardt propôs substituir a matriz identidade pela matriz diagonal de H, resultando na seguinte regra de atualização (ROWEIS, 2003):

$$\Delta \bar{w} = -\left[ J^T(\bar{w})J(\bar{w}) + \mu \text{Diag}[J^T(\bar{w})J(\bar{w})] \right]^{-1} J^T(\bar{w})e(\bar{w}) \quad (3.21)$$

É fundamental que o cálculo da matriz  $J^T(\bar{w})$  seja realizado de forma eficiente. Essa matriz é formada pelo cálculo dos gradientes, para cada padrão de treinamento apresentado à rede. Menhaj e Hagan (1994) propuseram uma modificação no algoritmo de retropropagação que permite utilizar esse algoritmo para calcular o jacobiano.

O método Levenberg-Marquardt pode ser resumido como:

- 1) Obtenha uma atribuição inicial de pesos;
- 2) Calcule uma atualização dos pesos de acordo com a Eq. (3.21) e avalie o erro, propagando os padrões na rede;
- 3) Se o erro aumentou desfaça a atualização de  $(\bar{w})$  e diminua  $\mu$ ; senão valide a atualização de  $(\bar{w})$  e incremente  $\mu$ ;
- 4) Se convergiu encerre; senão volte para 1.

### 3.6 Seleção Escalar de Características

Utilizado em reconhecimento de padrões para selecionar as variáveis que melhor separam um conjunto de dados de amostras em  $n$  classes distintas, o método de seleção escalar de características (THEODORIDIS e KOUTROUMBAS, 2006) utiliza o conceito de Razão Discriminante de Fisher (FDR) e índices de correlação cruzada.

Supondo que existam duas classes equiprováveis,  $w_1$  e  $w_2$ , para se saber quão bem uma variável  $x$  separa essas duas classes, pode-se utilizar a Razão Discriminante de Fisher (*Fisher's Discriminant Ratio* - FDR), dada pela expressão (3.22).

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

(3.22)

Em que:

$\mu_1$  – média da variável  $x$  na classe  $w_1$ ;

$\mu_2$  – média da variável  $x$  na classe  $w_2$ ;

$\sigma_1^2$  – variância da variável  $x$  na classe  $w_1$ ;

$\sigma_2^2$  – variância da variável  $x$  na classe  $w_2$ .

Para múltiplas classes, o valor de FDR para uma variável  $x$  pode ser calculada através da expressão (3.23).

$$FDR = \sum_i^M \sum_{j \neq i}^M \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2}$$

(3.23)

Para a seleção escalar de características seguimos os seguintes passos:

1. Utiliza uma das medidas de separação de classe (FDR) para ranquear as características. O Valor da medida escolhida,  $C(k)$  é calculado para cada uma das características,  $k = 1, 2, \dots, m$ .

2. Selecionar a primeira característica baseada no maior valor de  $C(k)$ . Seja essa característica  $x_{i_1}$

3. Para selecionar as demais características seguimos:

- Seja  $x_{nk}, n = 1, 2, \dots, N$  e  $k = 1, 2, \dots, m$ , a característica  $k$  do padrão  $n$ . o coeficiente de correlação cruzada de uma característica é dado por:

$$\rho_{ij} = \frac{\sum_{n=1}^N x_{ni} x_{nj}}{\sqrt{\sum_{n=1}^N x_{ni}^2 \sum_{n=1}^N x_{nj}^2}} \quad (3.24)$$

- Calcular o coeficiente de correlação cruzada  $\rho_{ij}$  entre a primeira característica selecionada  $x_{i_1}$  e as outras  $m - 1$  características, isto é  $\rho_{i_1 j}, j \neq i_1$

- Selecionar a segunda característica  $x_{i_2}$  para a qual:

$$i_2 = \operatorname{argmax}_j \{ \alpha_1 C(j) - \alpha_2 |\rho_{i_1 j}| \}, \forall j \neq i_1 \quad (3.25)$$

Em que:

$\alpha_1$  e  $\alpha_2$  são fatores que determinam a relativa importância dada aos dois termos.

- Para seleção da variável  $k, k > 2$  selecionar  $x_{i_k}$  de tal forma que:

$$i_k = \operatorname{argmax}_j \left\{ \alpha_1 C(j) - \frac{\alpha_2}{k-1} \sum_{r=1}^{k-1} |\rho_{i_r j}| \right\}, \forall j \neq i_r \quad (3.26)$$

Dessa forma, a correlação média com todas as características previamente selecionadas é levada em conta.

### 3.7 Coeficiente de Correlação de Pearson

Entende-se por coeficiente de correlação de Pearson ( $r$ ), ou coeficiente de correlação produto-momento, ou r-de-Pearson a medida do grau de associação entre duas variáveis aleatórias. Esse coeficiente é dado por:

$$r = \frac{Cov(X, Y)}{S_X \cdot S_Y}, \quad r \in [-1, 1] \quad (3.27)$$

Em que:

$Cov(X, Y)$  – Covariância ou variância conjunta das variáveis X e Y;

$S_X$  – Desvio padrão da variável X;

$S_Y$  – Desvio padrão da variável Y.

Na prática, o coeficiente  $r$  é interpretado como um indicador que descreve a interdependência entre as variáveis X e Y, com a forma  $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ , onde  $\hat{\alpha}$  e  $\hat{\beta}$  são constantes.

É um índice adimensional com valores situados entre -1,0 e 1,0, que reflete a intensidade de uma relação linear entre dois conjuntos de dados, com as seguintes condições:

- Para  $r = 1$  significa uma correlação positiva perfeita entre duas variáveis;
- Para  $r = -1$  significa uma correlação negativa perfeita entre duas variáveis - Isto é, se uma aumenta, a outra sempre diminui;
- Para  $r = 0$  significa que as duas variáveis não dependem linearmente uma da outra. No entanto, pode existir uma outra dependência que seja "não linear". Assim, o resultado  $r=0$  deve ser investigado por outros meios.

Nesse trabalho, utiliza-se o índice de correlação de Pearson  $r$  em duas situações: A primeira para se determinar a importância, ou ranqueamento das variáveis de entrada em relação

a previsão do nível da cheia do rio; A segunda, para substituir a razão do discriminante de Fisher no método de seleção escalar de característica modificado, que será apresentado no capítulo 4.

### 3.8 Significância Estatística

Nesse trabalho testes de significância estatística são utilizados com o propósito de avaliar se a dependência das variáveis utilizadas para a previsão com o nível da cheia do rio, expressa através do coeficiente de correlação r-de-Pearson, é estatisticamente significativa, ou não.

Em estatística, aborda-se a questão da significância de um resultado usando-se o conceito de hipótese nula. A hipótese nula ( $H_0$ ) simplesmente assume que um dado resultado estatístico foi obtido apenas por acaso, devido a flutuações probabilísticas dos eventos medidos, e não devido a uma causa real que influencia o resultado. Sempre que se trabalha com uma hipótese para explicar um dado fenômeno, considera-se a possibilidade de pelo menos uma hipótese concorrente a ela. Essa hipótese é chamada de hipótese alternativa ( $H_A$ ).

A hipótese alternativa é testada com base em resultados amostrais, sendo aceita ou rejeitada. O teste pode levar a aceitação ou rejeição de  $H_0$  que corresponde, respectivamente à negação ou afirmação de  $H_A$ .

Em testes de significância é necessário definir-se o nível de significância, que é a probabilidade máxima de rejeitar  $H_0$ , sendo a mesma verdadeira. Tipicamente, esse nível de significância pode assumir valores de 5%, 1% ou 0,1%.

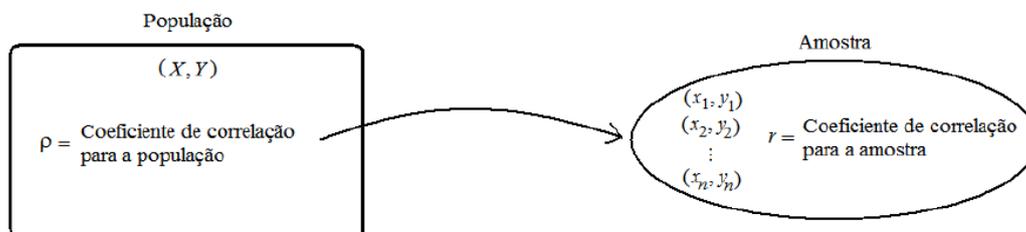
Nesse trabalho testaremos o coeficiente r-de-Pearson utilizando um nível de significância de 0,1%, ou  $p < 0,001$ .

Para avaliação da significância estatística dos coeficientes r-de-Pearson das diversas variáveis utilizadas para previsão do nível de cheia, foi utilizado o teste t-Student (STUDENT, 1908; LEVINE e BERENSON, 2000).

### 3.9 Teste de Significância para o Coeficiente de Correlação de Pearson – t Student

Quando se coleta uma amostra de  $n$  pares de valores das variáveis  $(X, Y)$  e se calcula o seu coeficiente de correlação  $r$ , o que se quer saber é se esse valor de  $r$  é significativo. Para isso, deve-se assumir como hipótese nula  $H_0$  que, na população, não existe correlação entre as variáveis  $X$  e  $Y$ , o que implicará que o valor obtido para  $r$  ocorreu por mero acaso.

Costuma-se denotar o coeficiente de correlação para a população das variáveis  $X$  e  $Y$  por  $\rho$ . Portanto, temos dois coeficientes de correlação: um para a população de valores  $X$  e  $Y$ , denotado por  $\rho$ , que é puramente teórico; e outro para uma amostra dos  $N$  pares  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , retirada da população, denotado por  $r$ . A figura 5 ilustra a situação.



**Figura 5. Coeficiente de Correlação para População e Amostra**

O valor de  $r$  é usado para estimar o coeficiente de correlação  $\rho$  e o teste de significância desse valor consiste em assumir como  $H_0$  que  $\rho = 0$ , para verificar se sob tal hipótese o valor obtido para  $r$  é muito ou pouco provável.

Em geral, para as amostras de  $n$  pares de valores  $(x, y)$ , para  $n \geq 30$ , a condição de distribuição normal para as duas variáveis é satisfeita.

Nesse trabalho o teste de significância foi feito da seguinte forma:

- Considerar  $H_0: \rho = 0$  e  $H_A: \rho \neq 0$ ;
- Escolha do valor crítico para significância, nesse caso  $p < 0,001$ ;
- Considerando que a amostra contém  $n$  pares de dados, consultar a tabela da distribuição t-Student e obter o valor de  $t(gl)$  para o valor de  $\alpha$  escolhido, onde  $gl = n - 2$ ;

- Calcular a variável  $t_0$ ,

$$t_0 = \frac{r}{\sqrt{\left[\frac{1-r^2}{n-2}\right]}}$$

(3.28)

- Se  $t_0 > t(gl)$  ou  $t_0 < -t(gl)$ , rejeita-se  $H_0$ . Caso contrário, não se rejeita  $H_0$ .

Assim, conclui-se o teste de significância com as seguintes considerações:

- Se  $H_0$  for rejeitada, o valor de  $r$  obtido para amostra é significativo e que existe correlação  $r$  entre as variáveis  $X$  e  $Y$  com o nível de significância igual a  $\alpha$ ;
- Se  $H_0$  não for rejeitada, o valor obtido de  $r$  para amostra não é significativo, o que podemos dizer que tanto pode haver correlação  $r$  como não haver correlação.

## 4 MATERIAIS E MÉTODOS

Para realizar a otimização da previsão de cheias sazonais através de redes neurais de propagação direta, utilizamos um microcomputador com processador Intel Core i3 @ 2.13 GHz e 4 GB de RAM, operando sob a plataforma Microsoft Windows 8. O ambiente de simulação utilizado foi o Matlab2014®.

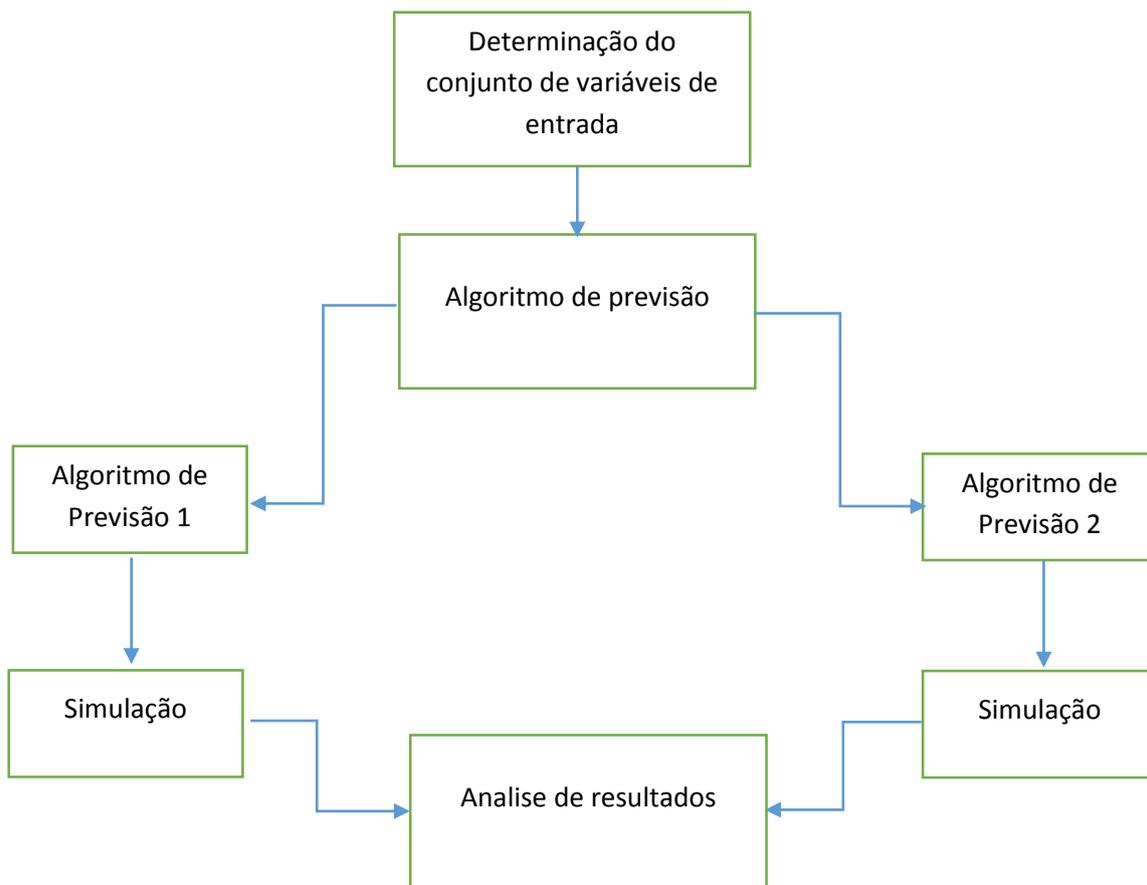
A figura 6 mostra o diagrama do sistema criado para otimização da previsão de cheias sazonais. Conforme mostrado nessa figura, foram desenvolvidos dois algoritmos de previsão, o algoritmo 1 e o algoritmo 2. O algoritmo 1 faz a seleção das variáveis através do valor do coeficiente e Pearson, enquanto que o algoritmo 2 faz a seleção das variáveis através de uma técnica nova proposta nesse trabalho para seleção de variáveis, a técnica de seleção escalar de características modificada.

Na sequência descreveremos cada uma das etapas descritas na Figura 6.

### 4.1 Conjunto de Variáveis

Como no estudo feito por Schongart e Junk (SCHONGART E JUNK, 2007), utilizou-se dois conjuntos de variáveis e efetuou-se a previsão no mês de fevereiro, ou seja, com 4 meses de antecedência do pico da cheia.

O primeiro conjunto é constituído pelos índices climáticos *El Niño South Oscillation* (ENSO):  $SST El Niño_{1+2}$ ,  $SST El Niño_3$ ,  $SST El Niño_{3+4}$ ,  $SST El Niño_4$  e SOI [Trenberth e Stepaniak, 2001], e o índice de Oscilação Sul (SOI), índice 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), obtidos na página oficial do *Climate Prediction Center* (Climate Prediction Center, 2015). Os dados coletados foram para os anos do período de 1953-2014. Essas variáveis são mostradas na Tabela 1.



**Figura 6. Diagrama do Sistema de Otimização para Previsão de Cheias Sazonais**

A propósito, sobre a escolha desses índices, registra-se que Potter et al. (POTTER et al, 2004), no período de 1958-1998, relatou uma forte correlação negativa ( $r = -0,5, p < 0,05$ ) entre a descarga do rio e o índice SST do pacífico e uma correlação positiva do índice SOI e mesma descarga.

O segundo conjunto é constituído pelo nível do Rio Negro, medido diariamente na estação do posto fluviométrico de Manaus (*Roadway*), nos meses que antecedem ao mês de ocorrência da cheia. O nível do rio também foi coletado no período de 1953-2014.

Para o ciclo de um ano, cada variável foi definida por um vetor de 12 valores, em que o primeiro elemento corresponde ao valor do índice no mês de julho do ano anterior a cheia que será prevista e o último elemento corresponde ao valor do índice do mês de junho do ano da

previsão. Os dados de níveis do Rio Negro foram coletados no site da Agencia Nacional de Águas – ANA (ANA, 2015) e Porto de Manaus (Porto de Manaus, 2015).

**Tabela 1: Índices climáticos utilizados para a previsão do nível do Rio Negro**

Índice ENSO	Latitude	Longitude	Período
SST El Niño <sub>1+2</sub>	0 a 10°S	90 a 80°W	1950-2014
SST El Niño <sub>3</sub>	5°N a 5°S	150 a 90°W	1950-2014
SST El Niño <sub>3+4</sub>	5°N a 5°S	170 a 120°W	1950-2014
SST El Niño <sub>4</sub>	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)		

## 4.2 Algoritmos de Previsão

O primeiro algoritmo, chamado de algoritmo 1, mantém uma semelhança com a metodologia de treinamento e teste *leave-one-out* (SONKA e FITZPRATICK, 2000). Nesse algoritmo, um determinado ano  $k$  é removido do conjunto de  $N$  anos. Faz-se então o treinamento da rede neural com os dados de  $N-1$  anos. Depois, faz-se a previsão do ano que foi removido. A diferença para o método *leave-one-out* é que, para cada ano  $k$ , treina-se uma rede neural  $m$  vezes com o mesmo conjunto de  $N-1$  anos e faz-se  $m$  previsões para o ano  $k$ . O valor de previsão para o ano  $k$  será, então a média dessas  $m$  previsões. Com esse procedimento, espera-se obter um valor para a previsão da cheia para cada ano sem viés metodológico e que leva em conta o comportamento aleatório das redes neurais, em que a otimização obtida depende muito dos valores iniciais escolhidos para os pesos e polarizações. Para este trabalho foi utilizado um valor de  $m = 100$ . Segue a sequência dessa metodologia:

### Algoritmo 1

1. Escolher o conjunto de variáveis de entrada;
2. Aplicar o teste de significância estatística t-Student ( $p < 0,001$ ) para o coeficiente de Pearson das variáveis;

3. Selecionar as variáveis com coeficientes de Pearson estatisticamente significantes ao nível de 0,1% e ranqueá-las pelo maior valor desse coeficiente;
4. Para um dado  $ano_k$ :
  - a. Treinar a rede  $m$  vezes com dados de  $N - 1$  anos, excetuando-se o  $ano_k$ , onde  $N = \text{número de anos} = 64$ . Obter a previsão da cheia do  $ano_k$ ;
  - b. Com as  $m$  previsões da cheia para o  $ano_k$ , obter uma média de previsão da cheia do  $ano_k$ .

No segundo algoritmo, chamado de algoritmo 2, seguimos a mesma metodologia. Porém, ao invés de ranquear as variáveis pelo valor de  $r$  e selecionar todas aquelas cujos coeficientes de Pearson são significantes ao nível de 0,001, utilizou-se a técnica de seleção escalar de características modificado, para selecionar um conjunto reduzido de características a serem utilizados na previsão. Segue a sequência dessa metodologia:

#### Algoritmo 2

1. Escolher o conjunto de variáveis de entrada;
2. Aplicar o teste de significância estatística t-Student ( $p < 0,001$ ) para o coeficiente de Pearson das variáveis;
3. Selecionar as variáveis com coeficientes de Pearson estatisticamente significantes ao nível de 0,001 e ranqueá-las utilizando o método de seleção escalar de características modificado;
4. Para um dado  $ano_k$ :
  - a. Selecionar um conjunto reduzido de variáveis com maior valor de ordenamento;

- b. Treinar a rede  $m$  vezes com dados de  $N - 1$  anos, excetuando-se o  $ano_k$ , onde  $N = \text{número de anos} = 64$ . Obter  $m$  valores de previsão da cheia do  $ano_k$ ;
- c. Com as  $m$  previsões da cheia para o  $ano_k$ , obter uma média do coeficiente de previsão da cheia do  $ano_k$ .

### 4.3 Método de Seleção Escalar de Características Modificado

No Capítulo 3, mostrou-se o método original de seleção escalar de características, utilizado em reconhecimento de padrões para selecionar as variáveis que melhor separam um conjunto de amostras com  $n$  classes distintas. Porém, como em um trabalho de previsão não temos classes, fizemos uma modificação na metodologia.

A diferença da técnica original para a que é proposta nesse trabalho, é a substituição da Razão Discriminante de Fisher (FDR) pelo Índice de Correlação de Pearson (ICP). Dessa forma, pode-se selecionar as melhores características ou variáveis que não tenham informações redundantes. Segue a descrição dos passos do método de seleção de características modificado:

- a. Seleciona-se inicialmente o conjunto  $C$  de variáveis que apresentam coeficientes de correlação de Pearson com o pico da cheia estatisticamente significantes. A partir desse conjunto, selecionar a primeira variável  $x_{s1}$ , que será aquela com maior valor de correlação com o pico da cheia.  $r_k$ .

Em que:  $k = 1 \dots m$

$m$  – número de variáveis do conjunto  $C$

- b. Para selecionar a segunda variável,  $x_{s2}$ , o coeficiente de correlação cruzada entre duas variáveis  $x_i$  e  $x_j$  é utilizado, conforme definido na equação (4.1)

$$\rho_{ij} = \frac{\sum_{n=1}^N x_{ni}x_{nj}}{\sqrt{\sum_{n=1}^N x_{ni}^2 \sum_{n=1}^N x_{nj}^2}}$$
(4.1)

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

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

A segunda variável selecionada,  $x_{s2}$ , é aquela que maximiza a equação (4.2):

$$\alpha_1 r_{s2} - \alpha_2 |\rho_{s1s2}|, \quad \text{for all } s2 \neq s1$$
(4.2)

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

- c. Outras variáveis selecionadas,  $x_{sk}$ ,  $k=3, \dots, m$ , são aquelas que maximizam a equação (4.3):

$$\alpha_1 r_{sk} - \frac{\alpha_2}{k-1} \sum_{r=1}^{k-1} |\rho_{srsk}|$$
(4.3)

#### 4.4 Configuração das Redes Neurais Utilizadas

Utilizou-se redes neurais com arquitetura  $n - n - 1$ . Experimentos foram feitos variando-se o valor de  $n$ . Utilizou-se função de ativação tangente hiperbólica em todos os

neurônios de todas as camadas. O Algoritmo utilizado para o treinamento foi o *backpropagation* associado ao método de otimização *Levenberg-Marquardt* (CICHOKI e UNBEHAUEN, 1993), para melhorar a velocidade de convergência do treinamento.

Cada arquitetura é treinada com três critérios de parada: erro médio quadrático, regularização e parada antecipada. Todos os critérios de paradas são testados com o algoritmo 1 e com o algoritmo 2.

## 5 RESULTADOS

Neste capítulo, inicialmente são mostradas tabelas contendo as variáveis de entrada, os valores dos coeficientes de Pearson, os valores dos testes de significância *t-Student* para esses coeficientes, o ordenamento das variáveis pelos valores de *t-Student* e o ordenamento das variáveis utilizando o método de seleção escalar de características modificado.

Para os algoritmos 1 e 2 apresentados no capítulo de metodologia são apresentados resultados obtidos com os critérios de paradas com regularização, parada antecipada e erro médio quadrático. Uma comparação é efetuada entre esses diversos resultados e aqueles obtidos no estudo de Schongart (SCHONGART e JUNK, 2007).

### 5.1 Variáveis de Entrada e Teste de Significância Estatística

Na Tabela 2 mostram-se as variáveis utilizadas nesse estudo, os valores dos coeficientes de Pearson e os valores dos testes de significância *t-Student* para esses coeficientes, com um nível de significância de 0,1%.

Pode-se observar na Tabela 2, que nenhum dos meses do índice SST *El Nino*<sub>1+2</sub> têm uma correlação significativa com o nível de 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 apresenta 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.

Na Tabela 3 são mostradas as nove variáveis ordenadas pelos maiores valores de correlação, primeira linha da Tabela 3, e pelo método de seleção escalar de característica modificado, segunda linha da Tabela 3. A Tabela 3 operacionaliza os passos 2 e 3 dos algoritmos 1 e 2, respectivamente.

**Tabela 2: 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	-0,86062	-1,32532	-2,07761	-2,29023	-2,72179	-2,46183	-1,4963	-0,95417	-1,04686	-0,75861	-0,12933	0,166379
	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	-2,29396	-2,19968	-2,37084	-2,57544	-3,0893	-3,24371	-3,23368	-2,89974	-2,2312	-1,31399	-0,19933	0,191117
	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	-3,14405	-3,03293	-2,70724	-2,94751	-3,1189	-3,72967	-3,90321	-3,8778	-3,47595	-2,33176	-1,03617	-0,7079
	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	-2,12166	-2,11794	-2,06927	-2,72202	-2,66068	-3,04305	-3,21963	-3,29356	-3,07618	-1,51355	-0,80198	-0,45115
	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	2,667744	1,973303	3,951974	3,228589	4,574221	3,755255	4,710428	2,381835	3,175864	2,217314	-1,19924	0,318709
	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	1,262399	0,967415	0,548234	0,415976	0,173943	-0,28144	4,877472	9,241579	13,24095	19,48327	31,28799	245,6756
	Significante para p<0,001?	N	N	N	N	N	N	S	S				

## 5.2 Seleção das Melhores Variáveis de Entrada Através do Método de Seleção de Características Modificado

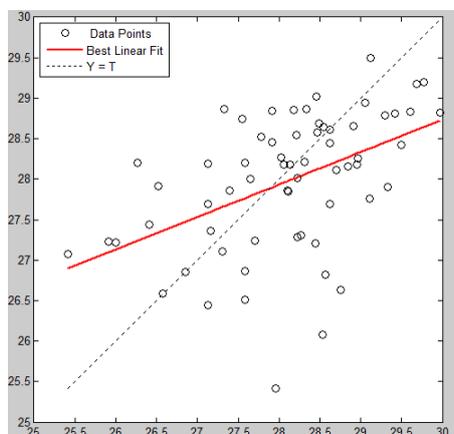
Conforme dito, a segunda linha da Tabela 3 mostra o ordenamento das 9 variáveis após a aplicação do método de seleção escalar de características modificado. A partir dessa tabela observa-se que o nível do rio no mês de janeiro, embora seja a variável com o segundo melhor valor de correlação,  $r=0,527$ , ocupa apenas a quarta posição no ordenamento das variáveis realizado pelo método de seleção escalar de características modificado. A segunda variável da ordenação 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 ordenadas segundo o valor de  $r$  e segundo o método de seleção escalar de características modificado**

Critério de ordenação	Variáveis ordenadas								
	$r$	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 ( $r=0.430$ )
Seleção Escalar	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

### 5.3 Índice de Correlação de Previsão $r_p$

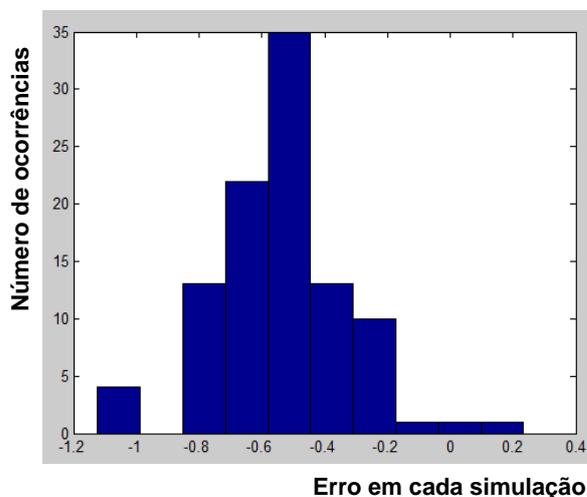
A partir da Tabela 3 pode ser observado que a correlação do nível do rio no mês de fevereiro com o nível da cheia é  $r = 0,761$ . Para avaliar o poder de previsão dessa variável sobre o nível da cheia adotou-se o seguinte procedimento: para todos os anos de 1951 a 2014 realizou-se 64 experimentos. Em 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 valor desse coeficiente foi de  $r_p = 0,745$ . A figura 7 mostra a reta formada entre a cheia observada com a média das cheias previstas.



**Figura 7. Gráfico mostrando a correlação entre o pico da cheia com a previsão realizada utilizando o nível do rio no mês de fevereiro**

#### 5.4 Resultados de Previsão dos Algoritmos 1 e 2

A Figura 8 mostra 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 feitos 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 previsão da cheia, foi de -0,53. Para a previsão 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 8: 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 4 mostra os resultados obtidos. Dois tipos de resultados são mostrados: valores dos índices de correlação de previsão,  $r_p$  e três índices de erro. Esses índices

de erro são definidos como: 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 associando as condições listadas a seguir:

- 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;
- conjuntos de 1, 2, 3 4, 5, 7, 8 e 9 variáveis ordenadas pelo método de seleção escalar de características modificado proposto nesse trabalho.

## 5.5 Discussões

A partir dos resultados mostrados na Tabela 4 observa-se que os melhores resultados foram obtidos com o método de parada de regularização. 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 modificada e com uma 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.

A técnica de seleção escalar de características modificada mostrou-se eficaz apenas com o método de regularização, onde se obteve os melhores resultados. Para o caso de três variáveis, a diferença entre o algoritmo 1, sem seleção escalar de característica modificada, e o algoritmo 2, com seleção escalar característica modificada, 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 4 para esses dois algoritmos são iguais nesses dois casos.

**Tabela 4: Resultados das simulações dos algoritmos 1 e 2 com os três métodos de parada do treinamento: regularização, parada antecipada e erro médio quadrático**

Algoritmo	Regularização							Early Stop					Erro médio quadrático				
	Número de variáveis	Arq.	r <sub>p</sub>	Ind_1	Ind_2	Ind_3	Ind_4	r <sub>p</sub>	Ind_1	Ind_2	Ind_3	Ind_4	r <sub>p</sub>	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
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

A seguir fazemos uma comparação dos resultados obtidos nesse trabalho com os resultados obtidos por Schongart e Junk (SCHONGART e JUNK, 2007).

Primeiramente, observa-se que o melhor índice de correlação de previsão obtido nesse trabalho,  $r_p = 0,755$ , é superior ao índice de correlação obtido por Schongart e Junk (SCHONGART e JUNK, 2007), que foi  $r = 0,714$ .

Na Tabela 5 mostra-se uma comparação de três faixas de erros, obtidas através do melhor resultado desse trabalho, com o resultado apresentado no trabalho de Schongart (SCHONGART e JUNK, 2007). Podemos observar que, nas três faixas mostradas, os resultados obtidos nesse trabalho apresentam percentuais superiores àqueles obtidos pelos autores citados.

**Tabela 5: Comparação dos índices de erros**

<b>Erro</b>	<b>(SCHONGART e JUNK, 2007). (<math>r = 0,714</math>)</b>	<b>Melhor resultado desse trabalho (<math>r_p = 0,755</math>)</b>
Ind_1( <i>erro</i> ≤ 0,5)	46,50%	51,56%
Ind_1+Ind_2 ( <i>erro</i> ≤ 1)	81,20%	92,18%
Ind_1+Ind_2+Ind_3 ( <i>erro</i> ≤ 1,5)	97,00%	98,43%

## 5.6 Conclusões

Um novo método de previsão de cheias sazonais, utilizando redes neurais artificiais, é proposto nessa dissertação. Esse método apresenta duas novas características a serem destacadas: uma nova técnica para seleção de características a serem utilizadas na previsão, a técnica de seleção escalar de características modificada, 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 obtidos através de um grande número de simulações independentes.

Essa segunda característica diferencia dos trabalhos ora apresentados no capítulo de referências bibliográficas, principalmente 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 obtido em várias simulações de uma rede neural, o que caracteriza, a nosso ver, um viés metodológico, pois adapta a rede neural aos dados.

Assim, o objetivo geral desse trabalho, posto na introdução como: “O objetivo geral deste trabalho é apresentar um novo método para otimização da previsão de cheias sazonais utilizando redes neurais artificiais”, foi inteiramente alcançado.

A proposição da técnica de seleção escalar de características modificada fez com que o primeiro objetivo específico desse trabalho fosse atendido. Esse objetivo específico foi posto na introdução como: “Propor uma nova técnica de seleção de variáveis para a previsão de cheias sazonais”, fosse atendido.

Os resultados apresentados na Tabela 4 fez com que o segundo objetivo específico desse trabalho fosse atendido. Esse objetivo específico foi posto na introdução como: “Avaliar o desempenho de métodos de generalização de redes neurais na previsão de cheias sazonais, como regularização e parada antecipada”.

Na medida em que se propôs um método para determinação do valor da previsão da cheia sazonal como o resultado de uma média de valores obtidos através de um grande número de simulações de redes neurais inicializadas de forma independente e aleatória, atendeu-se ao terceiro objetivo específico desse trabalho. Esse objetivo específico foi posto na introdução como: “Propor uma nova metodologia de treinamento e teste de redes neurais para previsão de cheias sazonais”.

Dessa forma, todos os objetivos traçados na introdução foram cumpridos. A comparação com resultados obtidos por outros autores, por outro lado, mostrou que os melhores resultados obtidos nesse trabalho foram satisfatórios.

Como trabalhos futuros, com o objetivo de melhorar os resultados ora obtidos, vislumbra-se:

- A utilização de outras técnicas de previsão, como máquinas de vetores de suporte, lógica fuzzy e pseudo inversa;
- Utilização de um número maior de variáveis de previsão, como por exemplo o nível de precipitação pluviométrica em estações a montante da cidade de Manaus;
- Ao invés de se utilizar o valor da média para previsão, utilizar-se outras medidas estatísticas, como a mediana.

## REFERÊNCIAS BIBLIOGRÁFICAS

BARRETO, JORGE MUNIZ. Redes Neurais - Fundamentos e Aplicações. 2o. Simpósio Brasileiro de Automação Inteligente, CEFET-PR, 1995.

BO PANG; YUAN LIANG. Application of an Improved Neural Network to Flood Forecasting of the Lower Yellow River, in Computational Intelligence and Design (ISCID), 2011 Fourth International Symposium on , vol.2, no., pp.43-46, 28-30 Oct. 2011

C. POTTER, P. ZHANG, S. KLOOSTER, V. GENOVESE, S. SHEKHAR AND V. KUMAR. "Understanding Controls on Historical River Discharge in the World's Largest Drainage Basins", Earth Interactions, v. 8, pp. 1-20, 2004.

CAI-HONG HU; ZE-NING WU; JI-JUN WANG; LINA LIU. Application of the Support Vector Machine on precipitation-runoff modelling in Fenhe River, in Water Resource and Environmental Protection (ISWREP), 2011 International Symposium on , vol.2, no., pp.1099-1103, 20-22 May 2011

CHANG, F.-J.; JIN-MING LIANG; YEN-CHANG CHEN. Flood forecasting using radial basis function neural networks, in Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on , vol.31, no.4, pp.530-535, Nov 2001

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.

CYBENKO, G. Continuous valued Neural Networks with two Hidden Layers are Sufficient. Technical report, Department of Computer Science, Tufts University. 1988.

DOAN, C.D. e LIONG, S.Y. 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.

EGAWA, T.; SUZUKI, K.; ICHIKAWA, Y.; IIZAKA, T.; MATSUI, T.; SHIKAGAWA, Y. A water flow forecasting for dam using neural networks and regression models, in Power and Energy Society General Meeting, 2011 IEEE , vol., no., pp.1-6, 24-29 July 2011

FRANCIS, L. The Basics of Neural Networks Demystified. Contingencies, p. 56-61, nov/dez. 2001.

GOPAKUMAR, R.; TAKARA, K. Water Level Forecasting in the Vembanad Water System Using Artificial Neural Networks, in Advances in Recent Technologies in Communication and Computing, 2009. ARTCom '09. International Conference on , vol., no., pp.948-953, 27-28 Oct. 2009

HAGAN, M. T; DEMUTH, H.; BEALE, M. Neural network design. Boston.1995

- HAYKIN, S. Redes neurais artificiais: princípios e prática. 2.ed. – Porto Alegre, RS : Bookman, 2001
- MCULLOCH, W. & PITTS, W. A Logical Calculus and the Ideas Immanent in the Nervous Activity. *Bulletin of Mathematical Biophysics* 5, 115-133 (1943).
- PRINCIPE, J.C.; EULIANO, N.R.;LEFEBVRE, W.C. *Neural and Adaptive Systems*. John Wiley & Sons, INC. 2000.
- RAMALHO, E. E.; MACEDO, J.; VIEIRA, T. M.; VALSECCHI, J.; CALVIMONTES, J.; MARMONTEL, M.; QUEIROZ, H. L. O ciclo hidrológico na RDSM, 1990 a 2008, *UAKARI*, v.5, n.1, p. 61-87, jun. 2009
- RANGANATHAN, Ananth. *The Levenberg-Marquadt Algorithm*. 2004.
- RUMELHART, D. E; HINTON, G. E.; WILLIAMS, R.J., *Learning Representations by Back-propagation Errors*. *Nature*, 1986.
- 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.
- SHETA, A.F.; EL-SHERIF, M.S. Optimal prediction of the Nile River flow using neural networks, in *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, vol.5, no.,pp.3438-3441vol.5,1999
- STUDENT. The Probable Error of a Mean. *Biometrika*, Vol. 6, No. 1 (Mar., 1908), pp. 1-25  
Published by: Biometrika Trust
- THEODORIDIS,S., KOUTROUMBAS, K. (2006) *Pattern Recognition*, Elsevier Academic Press, San Diego.
- 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.
- VAPNIK, V. N. 1998. *Statistical Learning Theory*. John Wiley & Sons, New York.

## APÊNDICE I

**Distribuição t-Student a primeira coluna indica o número de graus de liberdade (gl), os cabeçalhos das outras colunas dão a probabilidade (p)**

gl/p	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
02	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
03	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,541	12,924
04	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
05	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
06	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
07	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,365	3,499	5,408
08	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
09	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015

<b>17</b>	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
<b>18</b>	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
<b>19</b>	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
<b>20</b>	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
<b>21</b>	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
<b>22</b>	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
<b>23</b>	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
<b>24</b>	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
<b>25</b>	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,726
<b>26</b>	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
<b>27</b>	0,127	0,256	0,389	0,531	0,684	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
<b>28</b>	0,127	0,256	0,389	0,530	0,683	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
<b>29</b>	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
<b>30</b>	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
<b>40</b>	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
<b>60</b>	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
<b>120</b>	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
<b>i</b>	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

## APÊNDICE II

### XXXV CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO - CSBC

---

#### PROPOSTA DE UM MÉTODO PARA PREVISÃO DE CHEIAS SAZONAIS UTILIZANDO REDES NEURAS 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

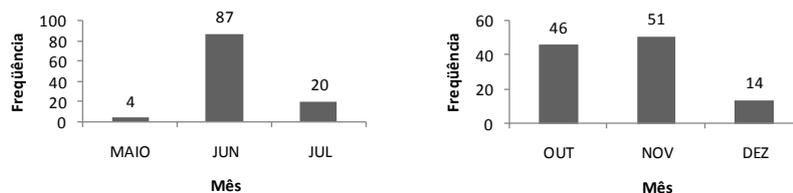
marciodemenezes@yahoo.com.br, {marlygfcosta, cffcfilho}@gmail.com

**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 2. (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  $6300\text{km}^3$  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  $Nino_{34}$  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  $300.000\text{ km}^2$ ), 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,  $El\ Niño_{3+4}$ .

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 termos 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 <sub>1+2</sub>	0 a 10°S	90 a 80°W	1950-2014
SST El Niño <sub>3</sub>	5°N a 5°S	150 a 90°W	1950-2014
SST El Niño <sub>3+4</sub>	5°N a 5°S	170 a 120°W	1950-2014
SST El Niño <sub>4</sub>	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 Liang, 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								
	$r$	Nível do rio em Fev ( $r=0,761$ )	Nível do rio em Jan ( $r=0,527$ )	SOI Jan ( $r=0,513$ )	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 ( $r=0,430$ )
Seleção Escalar	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> Feb	El Niño <sub>3+4</sub> Dez

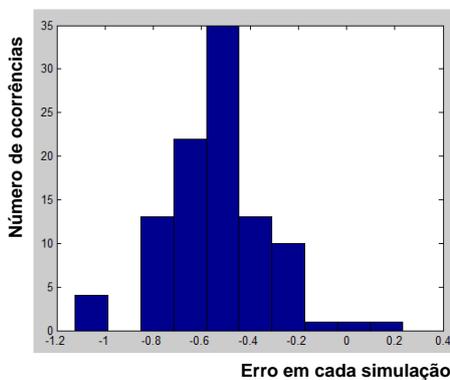
### 3.3 Resultados de previsão utilizando correlação simples

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
Algoritmo 2	9	10x10x1	0,704	30	29	3	2	0,705	32	23	6	3	0,634	28	29	5	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 vies 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 vies 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.portodemanau.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/>.

## APÊNDICE III

EUROPEAN SIMULATION AND MODELLING CONFERENCE - ESM'2015

### COMPARING NEURAL NETWORKS AND PSEUDO-INVERSE MATRIX TO FLOOD FORECASTING OF THE AMAZON RIVER

Marcio Menezes

Marly Costa

Cicero Costa Filho

Technological and Information Center

Federal University of Amazonas

Av. General Rodrigo Otavio Jordão Ramos, 3000

Aleixo, Manaus, AM, Brazil. 69077-000

E-mail: cffcfilhol@gmail.com

#### ABSTRACT

Forecasting the maximum peak flood of rivers with reasonable behavior is an important task from an economic and social point of view. In this paper, we compare the performance of two forecasting methods for the peak flood of the Amazon River: feedforward neural networks and pseudo inverse. To improve the generalization of feedforward neural networks, the regularization training stopping criteria was employed. The neural network performance was calculated as a mean value of 100 simulations. As input variables for both forecasting methods, Pacific Ocean climatic indices and the river level measured at the port of Manaus. The best forecasting coefficient value was obtained with pseudo inverse network,  $r_f = 0.756$ .

#### INTRODUCTION

Some important watersheds, such as those of the Amazon, Congo and Nile rivers, are characterized by seasonal yearly floods. Forecasting the maximum annual flood level of one of these rivers is a difficult task, since the precipitation-runoff relationships are among the most complex hydrological phenomena to

understand, due to the huge spatial and temporal variability of watershed characteristics, precipitation, as well as a number of characteristics that must be taken into account in modeling the physical processes.

In this paper we are concerned with forecasting the peak flood levels (maximum annual water level) of the Amazon River. The variations that occur annually in the Amazon River have a tremendous impact on economic activities (fishery, agriculture and logging) that takes place in floodplains along sediment-loaded, nutrient-rich white-water rivers (várzea). According to Junk et al. (2000), these regions have the highest human population densities in rural Amazonia due to abundant natural resources, including highly-productive nutrient-rich soils, with easy accessibility.

With an annual discharge of 6300Km<sup>3</sup>, the Amazon River is the largest river in the world in terms of the basin size. The peak flood usually occurs in June and the lower water level, in October or November. Figure 1 shows a histogram of maximum and minimum water level in the period of 1903-2013. The river level has been measured since 1903 by the Manaus Port of Authority (Manaus Port of Authority Port 2015).

Different methods were used in the literature for forecasting hydrologic phenomena. Table 1 shows a

brief summary of studies published in recent years concerned with these phenomena.

The forecasting applications shown in Table 1 varies from hours to months. The main tools used for forecasting are: feedforward neural networks (FNN), RBF neural networks (RBFNN), support vector machines (SVM), statistical correlations and genetic program. The input variables used for forecasting are of different types: previous rainfall and streamflow, water level in previous moments, the Southern Pacific oscillation index, Tropical Pacific Sea Surface Temperature, daily discharge time series and seasonal discharge time series.

Two studies listed in Table 1 address the issue of forecasting the peak flood of The Amazon River. Schongart and Junk (2007), employing multiple correlation, and analyzing the period of 1903-2004, obtained a high correlation ( $r=0.71$ ,  $p<0.0001$ ), between the peak flood of The Amazon River and the South Oscillation Index of the South Pacific Ocean.

Uvo et al. (2000), for the period 1946-1992, used FNN to forecast the Amazon River discharge. The training-test method was a 4-cross validation one. The FNN architecture was 8-20-1. The best performance of the neural network ( $r=0.86$ ) was obtained using the following input variables: Sea Surface Temperature (SST) index of the equatorial Pacific sea and the rainfall of three regions downstream of the city of Manaus.

In this paper, we compare the performance of two forecasting methods for the peak flood of The Amazon

River: FNN and Pseudo-Inverse (PI) methods. To improve the generalization of FNN, the regularization training stopping criteria was employed. Pacific Ocean climatic indices and the river level measured at the port of Manaus were used as input variables for both methods.

## MATERIALS AND METHODS

### Materials

Two variable sets were used as input variables for both forecasting methods used in this paper.

The first variable set comprises the Pacific Ocean climatic indices shown in Table 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> and the Southern Oscillation Index (SOI) of El Niño South Oscillation (ENSO) (Trenberth and Stepaniak 2001).

The SST indices are defined as the mean temperature of the West Pacific Ocean in the regions shown in Table 2. The SOI index is calculated as the standard difference between the atmospheric pressure in the sea level measured in Tahiti (17°S 149°W) and Darwin, Australia (13°S 131°E). These index data were obtained for the period 1951-2014 (Climatic Prediction Center 2015).

Table 1: Summary of studies concerning forecasting of hydrologic phenomena published in recent years

Reference	Application	Forecasting	Tools	Input variables	Performance
Chang 2001	A rainfall-runoff model for Lanyoung River, Taiwan flood forecasting during typhoons.	Water level three-hours ahead	Three layers RBF Neural Network	Rainfall and the streamflow information for up to 3 h before the forecast	Correlation coefficients between 0.817 and 0.999 at t+1, t+2 and t+3
Ruslan 2013	Flood water level prediction at Kelang river located at Petaling Bridge.	Water level 10 minutes ahead	Three layers RBF Neural Network and Inverse Model	water level in previous moments	Root mean square error of 0,038m in a 10 minutes forecasting from 06/03/2011 to 11/03/2011
Schongart and Junk 2007	Water level prediction at Negro River, located at Manaus, Brazil.	Water level 4 months before the maximum water level that occurs in the second half of June	Statistical correlation	Pacific South Oscillation Index (SOI) and water level at february	A correlation coefficient of 0.71, p<0.0001 in the period of 1903-2004.
Hu 2011	Runoff forecast in Fenhe River, the branch of the Yellow River (China)	Daily runoff forecasting of water level	Support Vector Machine	Antecedent runoff at the observed station, past rainfall, and the average rainfall on the intervening area.	Runoff relatively error of 0.126.
Uvo 2000	Discharge forecast at ten river sites in Amazonia.	Discharge one and two seasons in advance	Three layers (8-20-1) neural network	Pacific Sea Surface Temperature indice (SST indice) and precipitation.	A correlation coefficient of 0.86 in the period of 1946-1994.
Khatibi 2011	Discharge routing of Kizilirmak River, Turkey.	Daily mean discharge.	Artificial Neural Network, Adaptive Nero-Fuzzy Inference System and Genetic Programming.	Daily mean discharge time series between 1999-2003.	Genetic Programming model performs better in predicting single-peaked series and multi-peaked series.
Pang 2011	Flood forecasting of the yellow river	Seasonal discharge.	Artificial Neural Network (ANN), Linear Perturbation Model (LPM), and Artificial Neural Network with Non Linear Perturbation Model (ANN - NLPM).	Seasonal discharge time series between 1998 and 2003.	Nash-Sutcliffe model efficiency index of 91.13 for ANN-NLPM, 88.43 for ANN and 84.84 for LPM.

Table 2: Pacific Ocean climatic indices

ENSO indice	Latitude	Longitude	Period
SST El Niño <sub>1+2</sub>	0 a 10°S	90 a 80°W	1950-2014
SST El Niño <sub>3</sub>	5°N a 5°S	150 a 90°W	1950-2014
SST El Niño <sub>3+4</sub>	5°N a 5°S	170 a 120°W	1950-2014
SST El Niño <sub>4</sub>	5°N a 5°S	160°E a 150°W	1950-2014
SOI	Standard pressure difference at sea level between Tahiti (17°S a 149°W) and Darwin, Australia (13°S a 131°E)		

Table 3: Input variables and statistical significance

Indice/ river level	Measures	Months used for forecasting								Months not used for forecasting			
		Jul	Ago	Set	Oct	Nov	Dez	Jan	Feb	Mar	Apr	May	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
	Significant for 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
	Significant for 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
	Significant for 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
	Significant for 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
	Significant for p<0.001?	N	N	S	S	S	S	S	N				
River level	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
	Significant for p<0.001?	N	N	N	N	N	N	S	S				

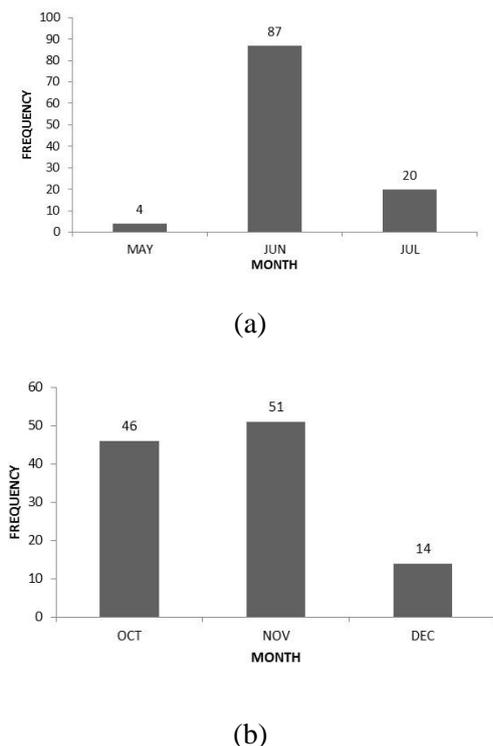


Figure 1: (a) Occurrence frequency of The Amazon River flood maximum level in function of the month. (b) occurrence frequency of The Amazon River flood minimum level in function of the month.

The second variable set comprises the river level in months prior to the month of maximum water level. Each variable is defined as a vector with 12 coordinates. The first coordinate corresponds to the variable value in July of the previous year (usually one month after the occurrence of the previous year's peak flood). The last coordinate corresponds to the variable value in June of the year that the forecasting was done. Table 3 shows these variables, their correlations with the maximum flood level, the value of t-Student statistical significance level and if this test is not significant at 0.1% level. As shown in Table 3,  $SST\ El\ Nino_{1+2}$  does not have a significant correlation with the maximum flood level.  $SST\ El\ Nino_{3+4}$  in December has the high correlation with the maximum flood level,  $r = -0.428$ . The SST indices present a negative correlation, while the SOI index presents a positive correlation. In this study the peak flood forecasting was calculated in February. The same procedure was adopted by Schongart and Junk (2007).

## Methods

### Feedforward Neural Networks

The algorithms employed for forecasting the peak flood of The Amazon River is like the *leave-one-out* method of training and testing (Sonka and Fitzpatrick 2000). In the *leave-one-out* method, the FNN is trained with  $N-1$  patterns and tested with pattern  $I$ , This process is repeated  $N$  times and a mean performance value is obtained. In algorithm 1, the only difference is that for calculating the peak river flood,  $m$  neural networks, with randomized weights and bias initializations, are trained and used to  $m$  forecasting of the peak flood. A mean forecasting value is then calculated. This method takes into account that the neural network behavior depends on the initialization sets of weights and bias. In this study  $m = 100$  and  $N = 64$  (1951-2014) were used. The algorithm is listed as follows:

4. Define an input variable set;
5. Select statistically significant  $n$  variables;
6. For each year ( $year_k$ )
  - a. Train  $m$  neural networks with data of  $N-1$  years ( $N$  years -  $year_k$ ). For each trained FNN, calculate a forecast value for the peak flood,  $pf_i$ ;
  - b. With the  $m$  forecast values of  $year_k$  calculate a mean value:  $f_k = \frac{\sum_{i=1}^m pf_i}{m}$ .

In this study, the number of simulations  $m = 100$  and the number of  $n$  variables varied from 1 to 9. The architecture of FNN used was s-s-1. The best value of  $s$  was determined for each value of  $n$ . The back propagation algorithm was associated with the optimization method of Levenberg-Marquardt (Cichoki and Unbehauen 1993). The neuron activation function was the hyperbolic tangent. To improve the neural network generalization, the regularization training stop criteria was employed (Doan and Liang 2004). The objective of regularization is work with lower values of FNN weights and bias, resulting in a FNN with a smoother behavior. With this goal, the mean quadratic error criteria usually employed for FNN training stop is modified to the one shown in equation (1).

$$\text{msereg} = \gamma \cdot \text{mse} + (1 - \gamma) \text{msw} \quad (1)$$

The linear system shown in (4) can be represented as:

Where:

$$AX = P \quad (5)$$

$\gamma$  – Performance ratio. Varies between 0 and 1.

mse – mean square error.

$$\text{msw} = \text{mean square weight} = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (2)$$

Where:

In this study  $\gamma=0.5$ .

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \dots & \dots & \dots & \dots \\ a_{Q1} & a_{Q2} & \dots & a_{Qp} \end{pmatrix} \quad (6)$$

*Pseudo Inverse*

Suppose that we use  $p$  variables to forecast the peak flood and that the forecast of a given year be calculated using a linear equation, as shown in equation (3):

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{pmatrix} \quad (7)$$

$$a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kp}x_p = pf_k \quad (3)$$

$$P = \begin{pmatrix} pf_1 \\ pf_2 \\ \dots \\ pf_Q \end{pmatrix} \quad (8)$$

Where:

$a_k$  – variable<sub>k</sub> value

$x_k$  – variable<sub>k</sub> coefficient

$pf_k$  – peak flood value of year<sub>k</sub>

If a number of variables,  $p$  was equal to  $Q$ ; the linear system shown in equation (5) could be solved using the inverse of  $A$ , as shown in equation (9):

$$X = A^{-1} \cdot P \quad (9)$$

For  $Q$  years, we have a linear system with  $Q$  equations:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p = pf_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p = pf_2$$

.

.

$$a_{Q1}x_1 + a_{Q2}x_2 + \dots + a_{Qp}x_p = pf_Q$$

(4)

Nevertheless, the number of variables usually used for forecasting,  $p$  is much lower than  $Q$ , so the linear system can be solved using equation (9). It can be shown (Albert 72) that a solution

$$X = Y \cdot P \quad (10)$$

that minimizes the quadratic error given in equation (11) is given by the pseudo-inverse matrix  $Y = A^+$ .

$$E = \sum_{i=1}^{64} \|pf_i - A_i \cdot X\|^2 \quad (11)$$

Where:

$$A_i = (a_{i1} \quad a_{i2} \quad \dots \quad a_{ip}) \quad (12)$$

$$A^+ = (A^T \cdot A)^{-1} A^T \quad (13)$$

The following algorithm for forecasting with the PI is proposed:

1. Define an input variable set;
2. Select statistically significant  $n$  variables;
3. For each year ( $year_k$ )
  - a. Using equation (4) calculate a pseudo-inverse matrix  $A^+$  with data of  $N-1$  years ( $N$  years -  $year_k$ ):  $Q = N-1=63$ .
  - b. Calculate a coefficient matrix  $X = A^+ \cdot P$ .
  - c. Calculate a forecast value  $pf_k = A_k \cdot X$ .

### Pearson Forecast Correlation Coefficient

In both forecasting algorithms previously described (FNN and PI),  $N$  forecast values, one for each year, are calculated. In this study, using these  $N$  forecast values and the corresponding maximum flood level of each year, a Pearson forecast correlation coefficient,  $r_f$  is calculated for both FNN and PI methods.

## Results

### Results using simple correlation and river level in February

As shown in Table 3, the correlation of the river level in February with the peak flood is  $r = 0.761$ . To evaluate the power prediction of this variable, a Pearson forecast correlation coefficient,  $r_f$ , was calculated considering the following procedure: for all the years between 1951 and 2014, 64 linear correlation curves were obtained. For obtaining each curve, the forecasted year is excluded, so only 63 years are considered. Using these correlation curves the peak flood of

the forecasted year is then calculated. With the 64 forecasted values the  $r_f$  coefficient is calculated. The obtained result is  $r_f = 0.745$ .

### Neural Network Results

To illustrate the method used for forecasting the peak flood of one year with FNN, Figure 2 shows a histogram of the observed errors of 100-peak flood forecasting for the year 1975. One hundred neural networks were trained with random weights and bias initialization. In this example, the following conditions were used for neural networks training: 3 input variables (3 variables with maximum  $r$  values in Table 3: river level in February, river level in January and SOI index in January), 6-6-1 architecture and regularization training stop criterion. The training data correspond to the years 1951-2014, except for 1975. As shown, there are errors with absolute values larger than 1. Nevertheless, the mean square error of these 100 predictions is -0.53.

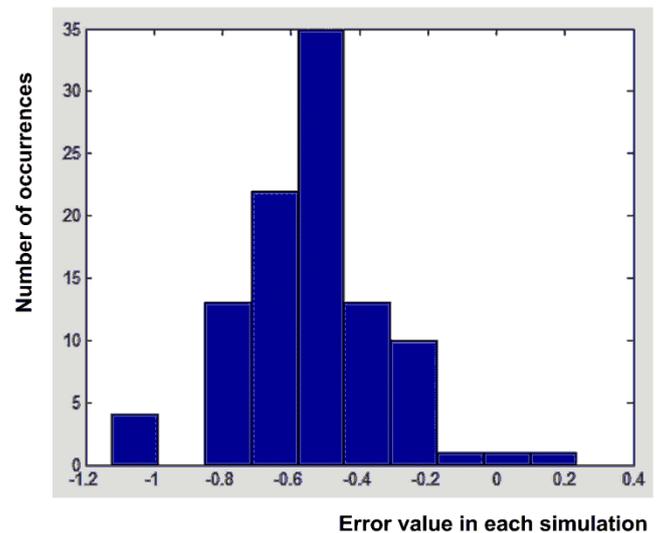


Figure 2: histogram of forecasting errors observed in 100 FNN simulations for 1975 year, with three input variables, architecture 6-6-1 and regularization training stop criterion. The mean error is -0.53.

Some experiments were done to determine the best neural network architecture and size of the input variable set. Table 5 shows the obtained results. For each neural network architecture and input variable set, two types of parameters are shown in the results: forecasting correlation,  $r_f$  and error indices. The error indices are: Ind\_1 - number of cases where the difference between the peak flood and the forecasting value is lower than 0.5m; Ind\_2 - number of cases where the difference between the peak flood and the forecasting value is situated between 0.5m and 1.0m; Ind\_3 - number of cases where the difference between the peak flood and the forecasting value is situated between 1.0m and 1.5m. Ind\_4 - number of cases where the difference between the peak flood and the forecasting value is situated above 1.5m. These

parameters were determined in different conditions: two neural network stopping criteria: mean square error below 0.1 and regularization criterion; input variable sets with 1, 2, 3, 4, 5, 6, 7, and 9 variables orderly by absolute value of Pearson correlation coefficient value,  $r$ . From Table 3, these orderly list of variables are: river level in February, river level in January, SOI index in January, SOI index in November, SOI index in September, El Nino<sub>3+4</sub> index in January, El Nino<sub>3+4</sub> index in February, El Nino<sub>3+4</sub> index in December and SOI index in December.

#### Pseudo Inverse Results

Table 5 shows the results obtained with a Pseudo Inverse method. Input variable sets with 1, 2, 3, 4, 5, 6, 7, 8 and 9 variables; orderly by absolute value of Pearson correlation coefficient were employed.

Table 4: Results for Feedforward Neural Networks

Regularization						
Number of variables	Architec.	$r_f$	Ind_1	Ind_2	Ind_3	Ind_4
1	4x4x1	0.723	31	24	8	1
2	5x5x1	0.710	28	28	7	1
3	6x6x1	0.717	25	32	6	1
4	7x7x1	0.706	27	30	6	1
5	8x8x1	0.722	26	32	5	1
7	9x9x1	0.691	27	28	8	1
9	10x10x1	0.704	30	29	3	2
Mean square error						
Number of variables	Architec.	$r_f$	Ind_1	Ind_2	Ind_3	Ind_4
1	4x4x1	0.689	30	23	9	2
2	5x5x1	0.630	23	32	3	6
3	6x6x1	0.622	23	27	9	5
4	7x7x1	0.677	23	32	6	3
5	8x8x1	0.621	23	28	10	3
7	9x9x1	0.644	23	34	5	2
9	10x10x1	0.634	28	29	5	2

Table 5: Results for Pseudo Inverse method

Number of variables	$r_f$	Ind_1	Ind_2	Ind_3	Ind_4
1	0.745	32	26	5	1
2	0.737	31	26	6	1
3	0.753	32	26	5	1
4	0.756	37	22	4	1
5	0.746	37	22	4	1
6	0.741	35	22	6	1
7	0.736	36	21	6	1
8	0.725	35	22	6	1
9	0.716	35	20	8	1

## Discussion and Conclusion

Table 4 shows that the best forecasting coefficient obtained with neural networks was  $r_f = 0.723$ . This result was obtained for the regularization training stop method. This result was obtained with only one input variable, the river level in February and with a neural network architecture 4x4x1. The error indices for this table input is: Ind\_1 = 31, Ind\_2 = 24, Ind\_3 = 8 and Ind\_4 = 1. So, in 51.6% of the years, the peak flood forecasting error was less than 0.5m, and in 91.67% of the years, the peak flood forecasting error was less 1.0m. The peak flood forecasting was higher than 1.5m in only one year: 1985. For all input set variables, the regularization training stop method shows better results than the mean square error method.

Table 5 shows that the best forecasting coefficient obtained with pseudo inverse method,  $r_f = 0.756$ . This result was obtained with four input variables: river level in February, river level in January, SOI index in January and SOI index in November. The error indices for this table input were: Ind\_1 = 37, Ind\_2 = 22, Ind\_3 = 4 and Ind\_4 = 1. So, in 61.7% of the years, the peak flood forecasting error was less than 0.5m, and in 98.33% of the years, the peak flood forecasting error was less than 1.0m. The peak flood forecasting was higher than 1.5m in only one year: 1985. For all input set variables the Pseudo Inverse method shows better results than neural networks.

In both proposed methods, neural networks and pseudo inverse matrix, there is no methodological bias. In neural networks, the forecasting value is obtained as a mean of 100 different simulations. This is different from other studies that use cross folder validation methods for training and testing neural networks, where the authors, for each folder, try to obtain the neural network with the best performance.

In this work, the variable sets were constructed with variables with higher Pearson correlation coefficients. This approach might select variables with redundant information. In future studies, we will address the problem of selecting the best set of variables.

## Acknowledgment

Part of the results presented in this paper were obtained through the project for research and human resources qualification, at the undergraduate and graduate level, in the areas of industrial automation, mobile software and Digital TV devices, sponsored by Samsung Eletronica da Amazonia Ltda, under the terms of Brazilian Federal Law number

8.387/91. We like to thank AcademicEnglishSolutions.com for revising the text

## REFERENCES

- Albert, A 1972. "Regression and the Moore-Pseudo Inverse", New York, Academic Press.
- Cichoki, A. and Unbehauen, R. 1993. "Neural Networks for Optimization and Signal Processing", John Wiley & Sons Ltd., New York.
- Chang, F-J; Liang, J-M and Chen, Y-C 2001. " Flood Forecasting Using Radial Basis Function Neural Networks." *IEEE Transactions on Systems, Man, and Cybernetics*, part c: applications and reviews, No. 31, 4, 530-535.
- Doan, C.D. and Liang, S.Y. 2004. "Generalization for multilayer neural network Bayesian regularization or early stopping." *2nd Conference of Asia Pacific Association of Hydrology and Water Resources*, 1-8.
- Hu, C-h and Wu, Z-n 2011. " Application of the Support Vector Machine on Precipitation-Runoff Modelling in Fenhe River." *International Symposium on Water Resource and Environmental Protection*, No. 4, 1099-1103.
- Khatibi, R.; Ghorbani, M. A.; Kashani, M. H. and Kizi, O. 2011. " Comparison of three artificial intelligence techniques for discharge routing." *Journal of Hydrology*, No. 403, 201-212.
- 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.
- Pang, B and Liang, Y. 2011. " Application of an improved neural network to flood forecasting of the Lower Yellow River", *Fourth International Symposium on Computational Intelligence and Design*, 43-46.
- Ruslan, F. A.; Samad, A. M.; Zain, Z. M and Adnam, R. 2013. " Modelling Flood Prediction Using Radial Basis Function Neural Network (RBFNN) and Inverse Model: A Comparative Study." *IEEE International Conference on Control System, Computing and Engineering*, 577-581.
- Schongart, J and Junk, W. J. 2007. "Forecasting the flood-pulse in Central Amazonia by ENSO-indices." *Journal of Hydrology*, No. 335, 124-132.
- Trenberth, K. E. and Stepaniak, D.P. 2001. "Indices of El Niño evolution." *Journal of Climate*, No. 14, 1697-1701.
- Uvo, C. B.; Tolle, U. and Berndtsson, R. 2000. "Forecasting Discharge in Amazonia Using Artificial Neural Networks." *International Journal of Climatology*, No. 20, 1495 - 1507.

## WEB REFERENCES

Climate Prediction Center,  
<http://www.cpc.ncep.noaa.gov/data/indices/>. Accessed in 02/23/2015.

Manaus Port of Authority, <http://www.portodemanaus.com.br>. Accessed in 02/23/2015.

## BIOGRAPHY

**MARCIO MENEZES RODRIGUES** was born in Manaus, Amazonas, Brazil, where he studied at the Federal University of Amazonas and obtained his degrees in 2005. After graduating, he worked in the private sector with banking security, and since 2013, he has worked at the Technological and Information Center of Amazonas Federal University.