

Preenchimento de Falhas de Dados Horários de Precipitação Utilizando Redes Neurais Artificiais

Haline Depiné¹, Nilza Maria dos Reis Castro¹, Adilson Pinheiro², Olavo Pedrollo¹

hali_ne@hotmail.com; nilza@iph.ufrgs.br; pinheiro@furb.br; olavopedrollo@gmail.com

Recebido: 29/06/12 - revisado: 03/06/13 - aceito: 24/10/13

RESUMO

A disponibilidade de séries contínuas de chuva pode viabilizar a execução de muitos estudos que não seriam possíveis com séries falhadas, porém estas normalmente fazem parte da realidade dos dados disponíveis, devido a dificuldades técnicas, operacionais e orçamentários. O objetivo desta pesquisa é testar uma metodologia com redes neurais artificiais (RNAs) para o preenchimento de falhas em séries horárias de dados de precipitação. Utilizou-se uma série de dados pluviométricos horários de nove estações, em uma bacia de 78 km², localizada na região noroeste do Rio Grande do Sul, a qual possui clima temperado com quatro estações climáticas anuais bem definidas. Foram experimentadas três diferentes alternativas, com respeito à extensão e disposição das séries horárias utilizadas: 1. Séries longas de treinamento e verificação, com aproximadamente 365 dias; 2. Séries curtas de treinamento com aproximadamente 12 a 30 dias, e várias séries também curtas de verificação, independentes da estação climática; 3. Séries curtas (também de 12 a 30 dias), semelhantemente à alternativa anterior, porém com verificação do preenchimento somente em períodos próximos ao treinamento (menos de 30 dias de diferença). Obteve-se os melhores resultados com a terceira alternativa, cujos coeficientes de Nash-Sutcliffe (NS) foram superiores a 0,9, tanto no treinamento quanto na verificação, em todos os postos analisados, reproduzindo bem os picos de verão. Isto demonstra que, mesmo em bacias onde as precipitações são consideradas homogêneas, as relações entre postos pluviométricos não são constantes, requerendo treinamento para cada processo atmosférico específico, para preenchimento de falhas em períodos curtos e sucessivos ao treinamento.

Palavras chave: serie temporal contínua, dados horários de chuva, treinamento sazonal.

INTRODUÇÃO

O conhecimento do regime hídrico em uma bacia hidrográfica é fundamental nos estudos hidrológicos que servem como base para projetos de diferentes usos de água, sendo indispensável à gestão sustentável dos recursos hídricos (SANTOS *et al.*, 2009). Os dados pluviométricos constituem elemento importante para o planejamento estratégico e a gestão dos recursos hídricos. A disponibilidade de séries temporais de precipitação em uma bacia é determinante na estimativa, entre outros, da capacidade de um manancial utilizado para abastecimento público de água potável e industrial (CARVALHO, 2007).

A caracterização da variabilidade temporal e espacial das precipitações é imprescindível para quantificar adequadamente os efeitos ocasionados, de modo especial, ao controle do escoamento superficial em áreas urbanas e rurais (CRUCIANI *et al.*, 2002).

A Agência Nacional de Águas disponibiliza os dados hidrológicos, através do Portal SNIRH (www.ana.gov.br/portalsnrh), para a escala de tempo diário. Em geral, os dados de chuva e vazão são medidos em bacias com áreas de drenagem superiores a 100 km².

Em geral, os métodos estatísticos e os modelos hidrológicos requerem séries longas e contínuas de dados, tanto de chuva quanto de vazão. Além disso, para pequenas bacias, que têm tempo de concentração menor que um dia, os modelos hidrológicos, quando utilizam dados diários, não apresentam boa eficiência (SOARES *et al.*, 2010; SILVA, 2011), requerendo frequência horária.

Dados com frequência horária, coletados para estudos em bacias pequenas, quando disponíveis, geralmente não constituem séries contínuas

¹ – Instituto de Pesquisas Hidráulicas/UFRGS

² – Fundação Universidade Regional de Blumenau/FURB

longas. O padrão mais frequentemente encontrado para séries de dados horários, além da pequena extensão, é a ocorrência de períodos de falhas ocasionados por problemas de falta de precisão do equipamento, deficiências de aferição e de calibração de instrumentos, indisponibilidade de pessoal qualificado, falta de recursos para manutenção contínua, controle de qualidade insuficiente, entre outros.

Muitas vezes, as falhas nos dados são provenientes de erros nas medições, que foram identificados durante o processo de consistência, não sendo passíveis de correção, resultando em lacunas. Segundo Castro *et al.* (2000), os erros mais frequentes são relacionados a falhas no sensor, de defasagem do relógio e de erros no processamento dos dados. Além desses erros, outros fatores podem comprometer o monitoramento contínuo através de pluviógrafo, como, por exemplo, aparelhos sem manutenção adequada, vandalismos, e problemas de entupimento do coletor, ocasionado por restos de vegetais, sedimentos ou excrementos de pássaros.

Nesse contexto, quando se dispõe de um conjunto de postos pluviométricos com frequência horária, é desejável que se possam transferir informações dos demais postos para um posto que contenha falhas. A precisão do procedimento dependerá da qualidade dos modelos utilizados para o preenchimento de falhas e também da distribuição espacial dos postos na bacia e da dinâmica dos fenômenos climatológicos da região. Porém, apesar das incertezas geradas pelo preenchimento de falhas, a disponibilidade da série contínua pode viabilizar a execução de muitos estudos que não seriam possíveis com séries falhadas.

A aplicação de modelagem no preenchimento de falhas, além de ser uma ferramenta útil nos estudos de gestão de recursos hídricos, facilita o conhecimento do comportamento hidrológico da bacia (BARNETCHE & KOBAYAMA, 2006).

Dentre os métodos disponíveis para preenchimento de falhas de dados hidrometeorológicos pode-se destacar a utilização de médias de dados observados ou de dados sintéticos obtidos com geradores de dados (CHIBANA *et al.*, 2005), o método da ponderação regional e da regressão linear (BERTONI & TUCCI, 2007), regressão linear múltipla, com utilização de ferramentas SIG (SANTANA *et al.*, 2003; SILVA, 2011), aplicação de redes neurais artificiais (GARDNER & DORLING, 1998; MAIER & DANDY, 2001; MAIER & DANDY, 2000; LUCIO *et al.*, 2006).

Carvalho *et al.* (1989) consideram que o uso de equações de regressão utilizando dados climáti-

cos constitui uma boa alternativa para o preenchimento das falhas em séries históricas.

Bertoni e Tucci (2007) analisaram duas metodologias empregadas no preenchimento de falhas (regressão linear e ponderação regional), concluindo que nenhuma destas foi satisfatória ao preenchimento de falhas diárias e horárias, embora sejam aplicáveis ao preenchimento de falhas mensais ou anuais. A complexidade física dos processos de precipitação e as escalas de tempo-espaço envolvidas em tais processos, ainda não podem ser resolvidas satisfatoriamente pelos modelos numéricos. Mesmo os modelos mais sofisticados (alta resolução) de previsão numérica de tempo (PNT) não conseguem, por meio de suas parametrizações, estimar satisfatoriamente o campo de chuva horário (CUO *et al.*, 2011). Em uma escala de tempo horária, os padrões espaço-temporais do processo de chuva são complexos, não lineares e altamente variáveis, o que dificulta a aplicação de métodos clássicos. Mesmo quando se consegue coincidência de precipitação diária em diferentes postos, existem diferenças de horas de ocorrência, devido ao deslocamento das massas de ar, o que dificulta a reconstrução do campo horário de chuvas.

A utilização de redes neurais vem sendo proposta por vários autores, para a modelagem chuva-vazão (OLIVEIRA *et al.* 2013, DORNELES *et al.* 2013; BRAVO *et al.* 2009; BRAVO *et al.* 2008); previsão (WEIGEND *et al.*, 1990; CONTRELL *et al.*, 1995; LACHTERMARCHER & FULLER, 1995; ELSHORBAGY *et al.*, 2000; SOUSA & SOUSA, 2010; SILVA *et al.*, 2006) e preenchimento de falhas de séries temporais (LUCIO *et al.*, 2006; CARVALHO, 2007), mostrando a viabilidade da utilização destes modelos.

A RNA pode ser definida como uma ferramenta estatística, que, por meio do processamento de informações, gera uma saída a partir de uma ou mais entradas. Através de sucessivas apresentações dos dados de entrada e saída previamente conhecidas, a RNA aprende a relação entre estes, através de um algoritmo de minimização de erros, procurando reduzir o erro quadrático médio em cada iteração do treinamento. Assim, esta relação pode ser aplicada para novos valores de entrada, os quais produzirão um valor de saída associado a uma nova previsão.

As Redes Neurais Artificiais (RNAs) têm a capacidade de extrair dependências implícitas entre os dados, mesmo quando não existe entendimento da natureza dessas dependências (NEVES & CORTEZ, 1997). Além disso, as RNAs possuem a capacidade de generalizar uma resposta mesmo ao receber

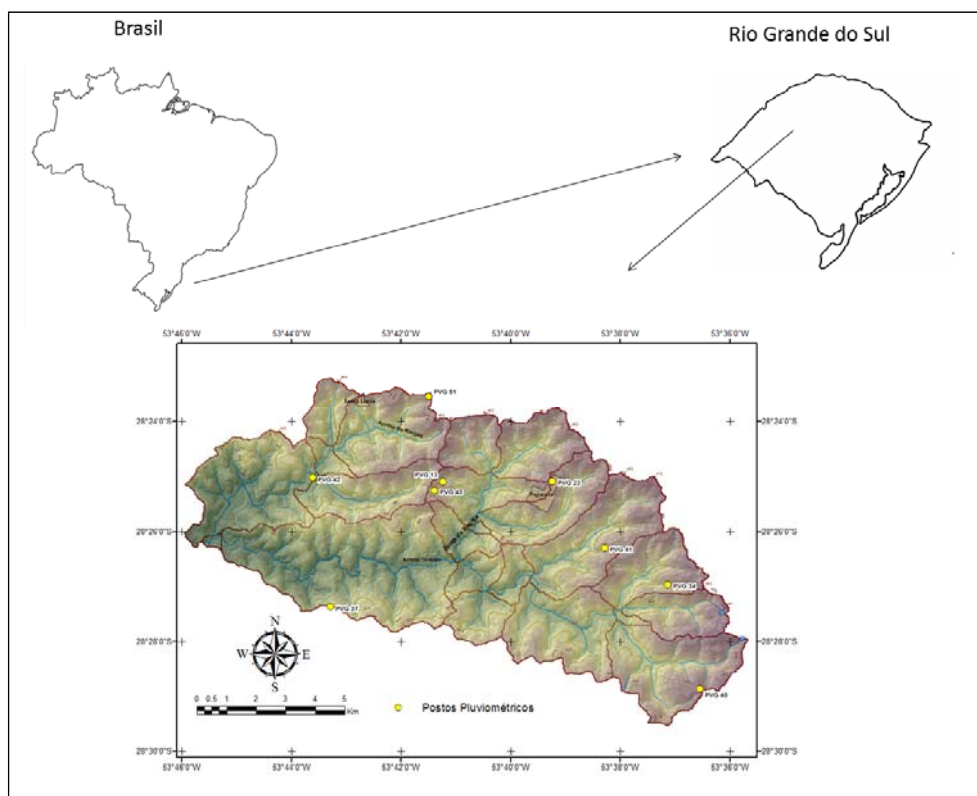


Figura 1 – Localização da Bacia do arroio Taboão e da rede pluviográfica

Tabela 1 – Número total de registros durante período de funcionamento de cada pluviógrafo no período de 01/01/2000 a 05/06/2011.

Registros	Pluviógrafos								
	P13	P23	P34	P37	P40	P41	P42	P43	P51
Total	45.184	53.901	81.929	83.082	53.408	21.526	53.365	53.371	60.228
Com dados	26.263	49.628	71.434	71.101	34.610	8.209	27.750	30.302	20.889
Sem dados	18.921	4.273	10.495	11.981	18.798	13.317	25.615	23.069	39.339
Falhas (%)	42	8	13	14	35	62	48	43	65

uma entrada incompleta ou com erro, a capacidade de adaptação diante de uma situação nova, e a capacidade de previsão de dados com base em informações anteriores.

O objetivo desta pesquisa é testar uma metodologia com redes neurais artificiais (RNAs) para o preenchimento de falhas em séries horárias de dados de precipitação. A pesquisa será aplicada para o preenchimento de séries históricas de dados pluviométricos horários monitorados na bacia do arroio Taboão – RS, no período de janeiro de 2000 a junho de 2011. Isto permitirá o fornecimento de subsídios que viabilizarão diversas pesquisas que

dependem da utilização de dados mais completos, bem como a avaliação de disponibilidade hídrica para o planejamento agrícola e a defesa ambiental, por exemplo.

MATERIAL E MÉTODOS

Área de estudo

A área objeto deste estudo é a bacia do arroio Taboão, o qual é um afluente do rio Potiribu (CASTRO *et al.*, 2000). A bacia do rio Potiribu,

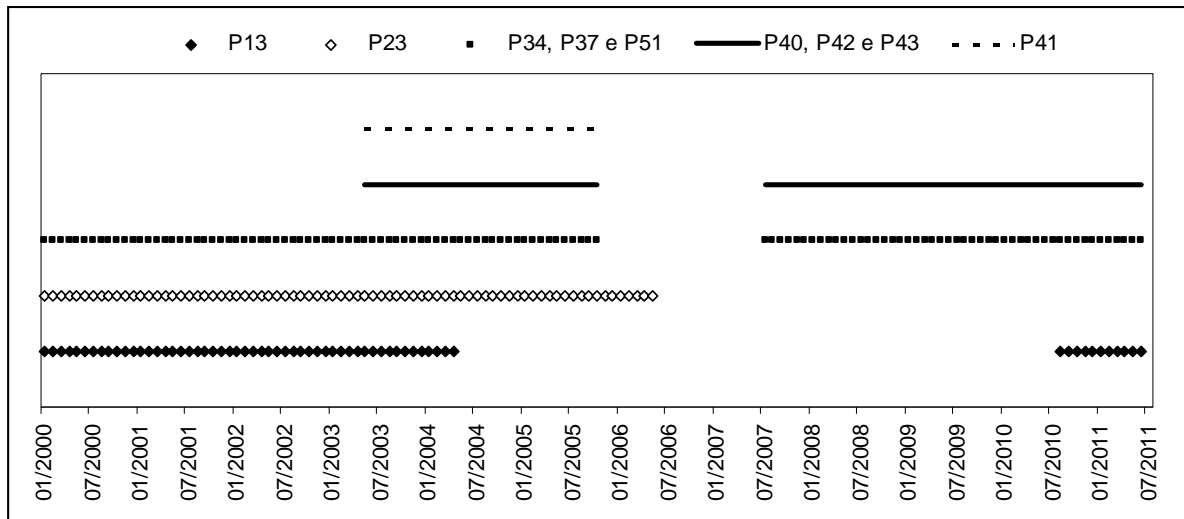


Figura 2 – Período de funcionamento de cada pluviógrafo durante janeiro de 2000 a junho de 2011.

situada no nordeste do Estado do Rio Grande do Sul, abrangendo uma área total de 664 km², ocupando parte dos municípios de Ijuí, Cruz Alta e Pejuçara, é formada por várias sub-bacias embutidas, que vem sendo monitoradas desde 1989 (CASTRO *et al.*, 1999). Na Figura 1 é apresentada a localização bacia do arroio Taboão, com área de drenagem de 78 km² bem como as estações pluviográficas.

As precipitações médias anuais na região são da ordem de 1.800 mm, bem distribuídas entre todos os meses do ano, não sendo possível identificar um período de estiagem (CHEVALLIER & CASTRO, 1991). O clima é classificado como mesotérmico brando superúmido, do tipo temperado (Cfa) (CASTRO *et al.*, 1993). Esta região possui quatro estações climáticas bem definidas. A temperatura média anual do ar oscila entre 18 e 19 °C, sendo julho o mês mais frio (13 a 14 °C) e janeiro o mais quente (24 °C). A média das temperaturas máximas é 32 °C e a média das mínimas fica em torno de 8 °C (BELTRAME, 2000).

Dados de Precipitação

Para a realização do estudo foram obtidos dados de precipitação de estações pluviográficas existentes na bacia, no período compreendido ente 01/01/2000 e 05/06/2011.

A bacia possui dezesseis estações equipadas com pluviômetros automáticos de balsa. Porém, durante o período estudado, apenas nove estavam

em funcionamento, P13, P23, P34, P37, P40, P41, P42, P43 e P51, conforme Tabela 1.

O período que cada aparelho esteve em funcionamento está ilustrado na Figura 2. De janeiro a julho de 2006 somente a estação meteorológica (P23) esteve em funcionamento. No período de julho de 2006 a junho de 2007 nenhum aparelho permaneceu em funcionamento, por falta de recursos financeiros. A estação pluviométrica que mais apresentou períodos com falhas foi a P43 (112 períodos). A P51, apesar de não ser a estação com maior número de períodos com falhas, é a que apresenta o maior número de dados faltantes (65%), pois o período de funcionamento da estação foi menor que o dos outros postos (ver Figura 2).

O pluviógrafo automático de balsa consiste de um funil que coleta a chuva e a transporta para um sistema constituído de uma haste apoiada em seu centro com dois recipientes de armazenamento de água em suas extremidades. Quando a quantidade de chuva acumulada em uma das básculas atinge 0,2 mm, o peso desta quantidade de líquido aciona o mecanismo, que automaticamente descarta a água e prepara a outra balsa para receber mais uma quantidade de chuva (BRAGA & FERNANDES, 2006). Quando esse mecanismo é acionado fecha-se um relé magnético, produzindo um pulso que é encaminhado a uma entrada contadora de pulsos em um datalogger. A precipitação é registrada em loggers, que totalizam os dados acumulados da altura de precipitação em milímetros, em um intervalo de tempo programado. Na

bacia em estudo, este intervalo de tempo é de dez minutos.

Redes Neurais Artificiais

As redes neurais artificiais são modelos matemáticos, estruturados de forma a reproduzir o funcionamento esquemático dos sistemas nervosos biológicos, sendo capazes de aprendizado, quando submetidos a treinamento, para produzir resposta a estímulos.

O elemento fundamental destes modelos é o neurônio artificial, o qual é uma evolução do modelo de McCulloch e Pitts (1943), formulado a partir de observações do neurônio natural. Sua composição básica é uma função de soma ponderada das entradas (estímulos) e uma função de ativação que atua sobre esta, resultando um sinal correspondente à resposta desejada, à medida que os parâmetros (chamados pesos sinápticos, em analogia aos neurônios biológicos) sejam adequadamente treinados.

A estruturação da rede neural artificial constitui-se de camadas sucessivas de neurônios artificiais, sendo a primeira correspondente aos nós de entrada, uma ou mais camadas intermediárias e uma camada de saída. Na figura 3 apresenta-se o esquema de uma rede neural artificial com i entradas, j neurônios na camada intermediária (h), e k neurônios na camada de saída (s). Os parâmetros w_{hj} e w_{sk} são os pesos sinápticos, respectivamente, das camadas, intermediária e de saída, enquanto que os parâmetros b_{hj} e b_{sk} são os respectivos termos de tendência.

Diversos autores enunciaram simultaneamente teoremas da aproximação universal com redes neurais, para relações de interesse prático, entre os quais o de Hornik et. al. (1989), o qual estabelece que qualquer relação mensurável $r: R^n \rightarrow R^m$ pode ser aproximada uma única camada interna, desde que adequadamente treinada e com um número compatível de neurônios artificiais nesta camada.

O número de neurônios na camada interna, na prática, depende da aplicação e da quantidade de amostras a ser utilizada no treinamento. O método mais utilizado para a sua determinação é a experimentação.

O método de treinamento retropropagativo consiste em um método de procura dos pesos sinápticos (a partir de valores iniciais, sorteados aleatoriamente) o qual minimiza, em ciclos sucessivos de aplicação à série de registros de treinamento, a soma do quadrado dos erros, com o uso da chama-

da regra delta (WIDROW & HOFF, 1960), após a determinação dos erros na(s) camada(s) intermediária(s), pela retropropagação destes a partir da camada de saída (RUMELHART *et al.*, 1986).

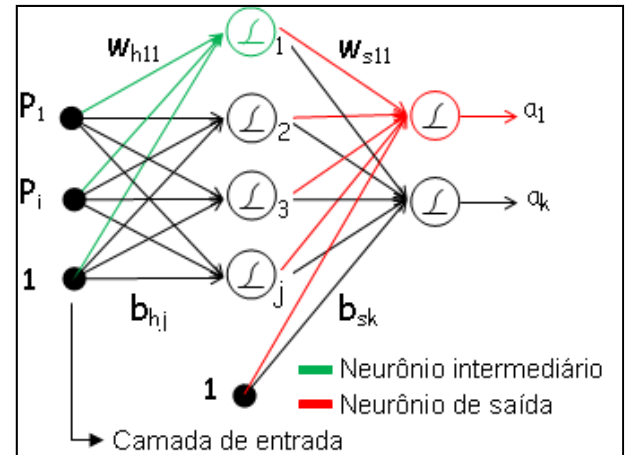


Figura 3 – Esquema de uma rede neural artificial

Para a aplicação deste método, as funções de ativação devem ser contínuas e deriváveis em todo o seu domínio. Outra limitação deste método de treinamento é a dependência das condições iniciais, a partir das quais o treinamento corre o risco de estacionar em mínimos locais. Este fenômeno é tanto mais frequente quanto mais complexa é a superfície de desempenho, a qual depende das complexidades da rede e das relações que se pretende aproximar com esta.

As condições das funções de ativação foram atendidas, neste estudo, pelo uso da função sigmoide bipolar na camada intermediária e da função linear na camada de saída.

A limitação relativa à aleatoriedade das condições iniciais foi contornada com o uso de diversas inicializações dos pesos sinápticos, constituindo iterações sucessivas, e com a adoção do modelo de melhor desempenho no treinamento.

O número de neurônios nas camadas internas foi obtido através de um processo de tentativa e erro, no qual foram analisados os índices de desempenho da RNA, escolhendo-se a arquitetura com o melhor desempenho no treinamento.

Foram treinadas, com o método retropropagativo, e de forma sistemática, redes neurais com duas variáveis de entrada, correspondentes às séries de precipitação mais próximas da série de saída, que continham dados para o treinamento e a verifi-

cação. O número de entradas escolhido resultou da necessidade de sistematização e da observação de que a restrição a períodos com ocorrência simultânea de mais de três séries de dados horários sem falhas, para o treinamento e a verificação, seria uma limitação importante.

Os testes para a escolha do número de ciclos e de iterações, bem como do número de neurônios na camada intermediária, foram efetuados com séries de dados de 2010 e de 2011, mantendo-se as definições obtidas para o restante do estudo. Uma vez que os resultados foram semelhantes com uso de duas e de trinta iterações, e que o número de 20.000 ciclos não promoveu melhoramentos na qualidade dos resultados em relação aos obtidos com 10.000 ciclos, embora tenha consumido tempo adicional considerável de processamento, adotaram-se para o restante do estudo, sistematicamente, duas iterações e 10.000 ciclos.

Porém, o aumento do número de neurônios na camada intermediária, de quatro para oito, proporcionou melhoria nos resultados de coeficiente de NS, aumentando a proporção da variância explicada, da ordem de 5,4%, e sem aumento significativo do tempo de processamento, o que justificou a escolha deste número de oito neurônios na camada intermediária.

Foram utilizadas séries de dados variáveis, mas nunca sequências com menos de trezentos valores, os quais foram separados em partes para treinamento e verificação da rede, correspondentes a 70% e 30% respectivamente. A adoção destas proporções visa proporcionar uma quantidade de amostras abrangentes para o treinamento, para que este produza redes neurais com suficiente capacidade de generalização (ANMALA *et al.*, 2000).

Para efeito deste trabalho, considera-se cada RNA treinada um novo modelo, mesmo que tenham a mesma arquitetura. Cada novo treinamento constitui a modelagem de um processo atmosférico diferente, o que se reflete nos parâmetros resultantes.

O processo de criação, treinamento e simulação da RNA, foi desenvolvido no software MATLAB® 7.12.0, R2010a.

Séries de treinamento e de verificação

Foram avaliadas três alternativas de treinamento e verificação para as RNAs, em crescente ordem de dificuldade e complexidade, sendo elas:

- I. modelos com séries longas de treinamento e verificação compostas de 8.760 registros,

que representam aproximadamente 365 dias, sendo os primeiros 70% dos registros horários para o treinamento e os seguintes 30% para a verificação.

- II. modelos com séries curtas de treinamento e verificação compostas por 300 a 800 registros, que representam 12 a 30 dias, sendo a verificação realizada em diferentes períodos ao longo do mesmo ano independente da estação sazonal na qual o modelo foi treinado;
- III. modelos com séries curtas de treinamento e verificação compostas por 300 a 800 registros, que representam 12 a 30 dias, sendo a verificação realizada sempre na sequência imediata da série de treinamento, ou seja, considerando o mesmo período sazonal do mesmo ano em que foi realizado o treinamento.

Em todas as três alternativas foi selecionado um período sem falhas com pelo menos três pluviógrafos, sendo dois deles utilizados como entradas e um como saída da rede, para o treinamento. Diversas arquiteturas de modelo no que diz respeito à quantidade de neurônios da camada intermediária foram experimentadas, sendo selecionado o modelo que resultou no melhor desempenho. A verificação para identificar se o modelo escolhido pode ser utilizado para o preenchimento, foi feita com o restante da série de precipitação que possuía dados tanto de entrada quanto de saída. A comparação, para cada série pluviográfica de saída, dos dados calculados com os observados, permitiu avaliar a capacidade do modelo para ser posteriormente utilizado para o preenchimento das falhas desta série, utilizando para isso índices de desempenho. Cada série pluviográfica funcionou como série de saída em pelo menos um modelo que utilizou outras duas séries como entradas, assim como funcionou como série de entrada para os modelos das demais.

Nas três alternativas, para cada ano foi feita uma sequência de treinamento e verificação diferentes, ou seja, não foi utilizado o treinamento de um ano para outros anos da série histórica. As séries de entrada e de saída foram inseridas completas, e não por eventos de chuva, ou seja, em dias nos quais não choveu, há um registro de chuva zero a cada hora. Para esta pesquisa foram utilizados dez anos de dados horários, ou seja, 87.672 registros para cada um dos nove pluviógrafos, sendo que parte destes registros são lacunas (+ de 20%).

Índices de desempenho

Para se avaliar o desempenho das RNAs, foram utilizadas duas estatísticas: o coeficiente de *Nash-Sutcliffe* (NS), como usado, por exemplo, em Zhang *et al.* (1998) e Silva *et al.* (2008), e o quantil 95% dos erros (E_{95}), como em Dornelles *et al.* (2013).

O coeficiente de *Nash-Sutcliffe* (NS) é uma medida da proporção dos dados de saída que é explicada pelo modelo. O coeficiente NS representa a capacidade de ajuste da RNA, sendo também a proporção com que o modelo explica os dados melhor do que estes o seriam adotando-se sistematicamente a média da série como previsão. O NS pode variar de negativo infinito ($-\infty$) a um, sendo que valores altos indicam que a rede apresenta elevada capacidade preditiva, sendo calculado pela equação 1:

$$NS = 1 - \frac{\sum(P_{obs_t} - P_{calc_t})^2}{\sum(P_{obs_t} - \overline{P_{obs}})^2} \tag{1}$$

Onde: P_{obs_t} é a precipitação registrada pela estação pluviométrica no intervalo de tempo t ($mm\ h^{-1}$); P_{calc_t} é a precipitação calculada no intervalo de tempo t ($mm\ h^{-1}$); e $\overline{P_{obs}}$ é a precipitação média registrada pela estação pluviométrica em todo o período simulado ($mm\ h^{-1}$).

Foram considerados com bom desempenho, para efeito desta pesquisa, os modelos que resultaram valores de NS superiores ou iguais a 0,90 no treinamento e na verificação da sua aplicação.

O uso de um quantil associado a uma frequência, como estatística de desempenho, é mais robusto do que seria a do erro máximo, para avaliar os erros nas piores condições, o qual poderia estar sendo afetado por erros nos dados ou valores atípicos. Portanto, neste trabalho, foi utilizado o E_{95} , o qual representa o valor com que o erro não é ultrapassado em 95% das amostras preenchidas, ou seja, apenas 5% dos erros dos valores calculados serão superiores ao E_{95} .

RESULTADOS E DISCUSSÃO

Como primeira alternativa, adotou-se um modelo geral, com séries longas de treinamento e verificação (com 8.760 dados, que representam aproximadamente 365 dias) sem que os períodos de treinamento e verificação pertencessem ao

mesmo período sazonal. No entanto, dificuldades foram encontradas. Observou-se que a verificação não teve a eficiência adequada (O NS foi inferior a 0,85). Na Figura 4 são apresentados os dados diários dos postos 34, 37, 40 e os valores calculados para as lacunas do posto 40 no intervalo de 09 a 19 de janeiro de 2011. Foram utilizados os dados dos postos 34 e 37 para análise da coerência do valor calculado do posto 40. No período em que se dispunha de dados nos três postos, de 09 à 13/01, os valores observados foram bastante semelhantes. O resultado calculado pelo modelo foi compatível com os valores observados nos postos vizinhos, conforme ocorrido no dia 19/01. Porém, em outros três momentos, o resultado calculado pelo modelo foi muito diferente dos valores observados nos postos vizinhos, pois ora o valor de precipitação simulado era muito além do observado (16/01), ora ficava muito abaixo (14/01 e 17/01).

Na segunda alternativa, os modelos foram treinados por séries curtas (variando de 12 a 30 dias de dados horários que constituem de 300 a 800 valores), independente de a estação climatológica verificada ter sido a mesma do treinamento. Para os testes desta segunda alternativa, foram usados os mesmos postos do treinamento e verificação da alternativa anterior. Observou-se uma melhora nos índices de desempenho da verificação em relação à primeira alternativa (NS superior a 0,9), portanto suficiente, em termos gerais, para os padrões desejados.

Porém, ao analisar-se a Figura 5, na qual são apresentados os resultados de um pequeno período simulado no verão (de 16 a 19/01/2010), observa-se que ocorrem diferenças muito grandes entre os valores calculados para o posto 51 e os valores observados dos postos vizinhos 37 e 42, que foram usados como entrada na RNA para o cálculo das chuvas horárias do posto 51 (Figura 5A). Para o treinamento deste modelo, foram utilizados dados do período compreendido entre 20 de setembro e 09 de outubro de 2009, não sendo o treinamento e a verificação na mesma estação climatológica. Observa-se que os valores calculados do posto 51, do dia 17/01 são muito inferiores aos observados nos postos 37 e 42.

Na terceira alternativa, além de séries mais curtas, levou-se em consideração, para elaboração do modelo, as estações do ano em que se encontravam a série de treinamento e a de verificação, tendo por objetivo minimizar possíveis efeitos dos diferentes períodos sazonais. Observou-se uma melhora nos índices de desempenho da verificação em relação à primeira alternativa e desempenho igualmente

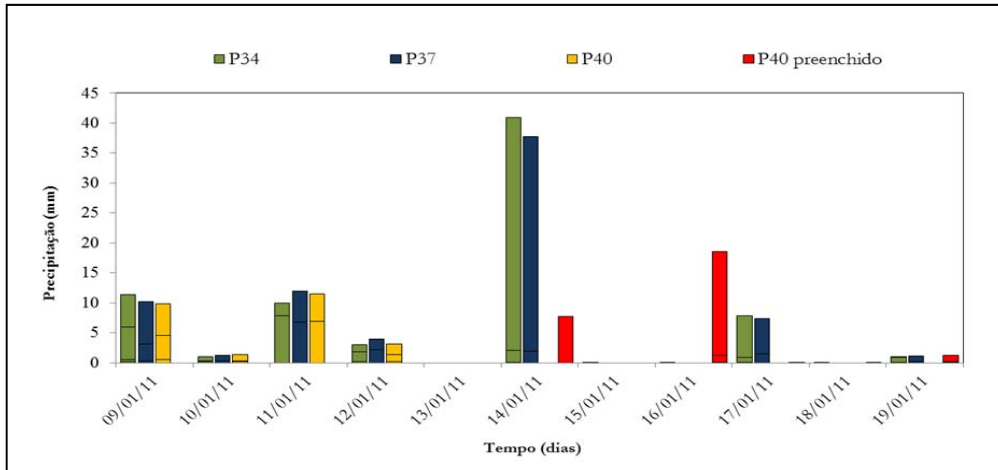


Figura 4 – Preenchimento do P40, a partir dos dados das estações P34 e P37, no período de 2010 e 2011.

te satisfatório, como na segunda alternativa (NS superior a 0,9).

A Figura 5B apresenta os dados observados e calculados dos mesmos postos e no mesmo período (16 a 19/01/2010) da alternativa anterior, porém com o modelo treinado usando séries de dados de dezembro de 2009 e janeiro de 2010. Nota-se que o modelo utilizado com as séries treinadas em períodos que possuem as mesmas características climatológicas do período preenchido conseguiu produzir picos de precipitação muito próximos aos observados nos postos vizinhos. Estes resultados corroboram com os obtidos por Gomes *et al.* (2010), que trabalharam com redes neurais artificiais para previsão de vazões naturais médias diárias em uma sub bacia do Rio São Francisco, e observaram melhores resultados quando as redes neurais foram ajustadas considerando os diferentes períodos sazonais.

Essas diferenças sazonais que influenciam a precipitação em determinados períodos estão fortemente relacionadas com fenômenos meteorológicos regionais. Minuzzi *et al.* (2005), descrevem que durante eventos do El Niño há uma ocorrência de irregularidades na distribuição de chuvas no Rio Grande do Sul (região em que se localiza a bacia em estudo). Segundo os autores, o El Niño produz precipitação acima da média climatológica em quase todos os meses do ano, mas o período de maior influência é na primavera e no início de verão. Portanto, modelos treinados com séries de verão, quando, por exemplo, ocorreu o fenômeno El Niño, podem não reproduzir bem as chuvas de inverno.

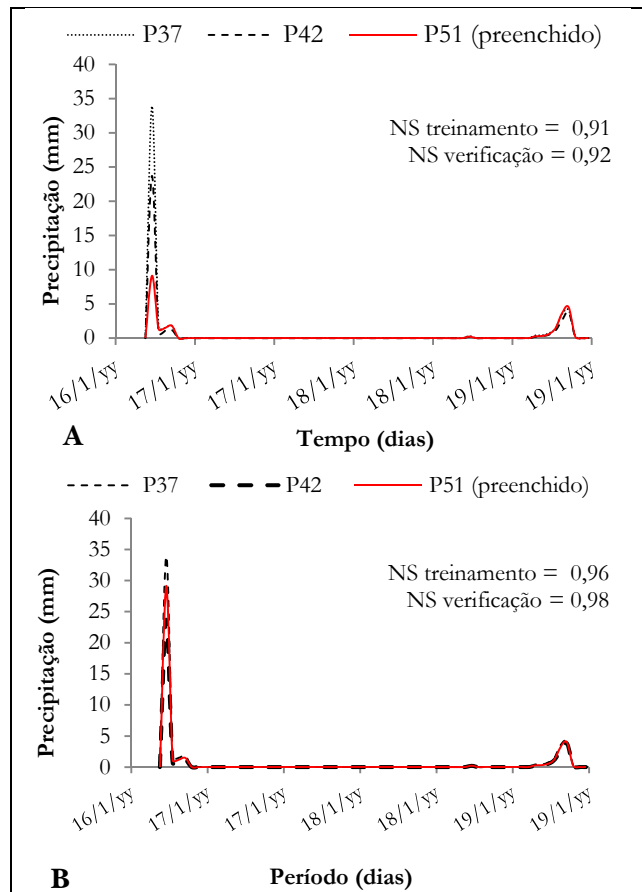


Figura 5 – Preenchimento do P51, a partir das séries P37 e P42, utilizando modelo de treinamento de setembro e outubro (A) e dezembro e janeiro (B).

Tabela 2 – Pluviógrafos preenchidos e usados no preenchimento e estatísticas médias do treinamento e da verificação.

Pluviógrafo preenchido	Pluviógrafos utilizados	Número de modelos	Estatística de treinamento				Estatística de verificação			
			Er95% (mm)		NS		Er95% (mm)		NS	
			Mín	Máx	Mín	Máx	Mín	Máx	Mín	Máx
13	23,34,37,40,42	41	0,709	1,261	0,97	0,99	0,765	1,004	0,92	0,97
23	13, 34, 37, 40	12	0,591	0,785	0,99	0,98	0,432	1,083	0,96	0,98
34	13, 23, 37,40, 41, 42, 43,51	48	0,007	1,076	0,94	0,99	0,002	1,170	0,9	0,99
37	13, 23, 34, 40, 41, 42, 43	51	0,03	1,430	0,91	0,99	0,005	1,420	0,9	0,99
40	23, 34, 37, 41, 42, 43, 51	55	0,003	1,500	0,93	0,99	0,005	1,260	0,9	0,97
41	23, 34, 37, 40, 43, 51	37	0,012	0,582	0,98	0,99	0,094	0,547	0,9	0,99
42	23, 34, 37, 40, 41, 43, 51	93	0	0,990	0,91	0,99	0,002	1,070	0,91	0,99
43	23, 34, 37, 40, 41, 42, 51	112	0,014	0,908	0,94	0,99	0,030	1,130	0,9	0,96
51	13, 23, 34, 37, 40, 41, 42, 43	106	0	0,652	0,97	0,99	0,015	0,840	0,91	0,98

Er95% - Erro não ultrapassado em 95% das amostras (mínimo e máximo), por aparelho (mm) e NS - coeficiente de Nash-Sutcliffe (mínimo e máximo) no treinamento e na verificação dos modelos.

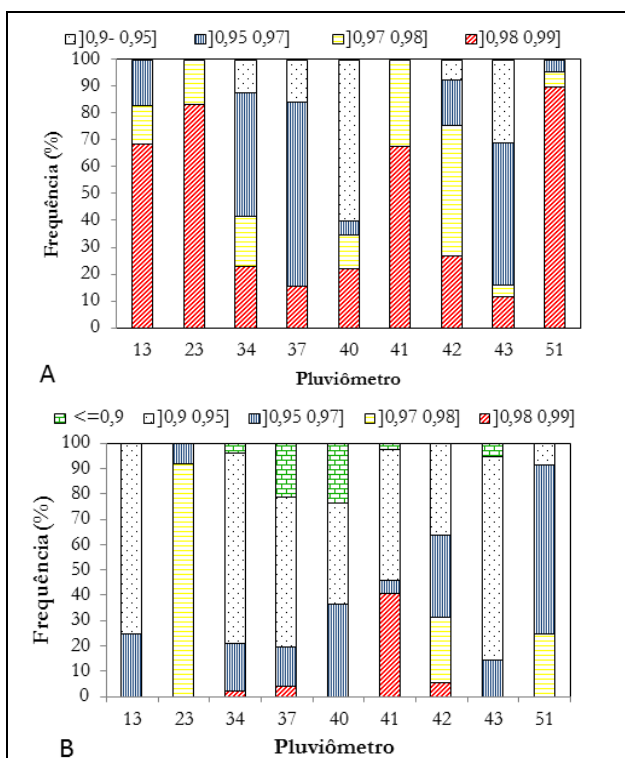


Figura 6 – Frequência de ocorrência do coeficiente de Nash-Sutcliffe, no treinamento (A) e na verificação (B) dos modelos.

As estatísticas resultantes do treinamento e da verificação da terceira alternativa, o número de modelos feitos para preencher todas as falhas de cada posto, e os postos utilizados como entradas em cada modelo são apresentadas na Tabela 2.

Analisando-se os treinamentos da RNA, observa-se que o erro não ultrapassado com 95% de frequência para cada série de dados variou de 0,000004 e 1,5 mm, sendo que o erro máximo ocorreu com o modelo para o posto P40. Enquanto que, para a verificação, o erro não ultrapassado com 95% de frequência para cada série de dados variou de 0,002 mm a 1,42 mm.

De maneira geral, o coeficiente de Nash-Sutcliffe (NS) mínimo encontrado no treinamento e na verificação dos modelos foi de 0,91 e de 0,90, respectivamente, e o máximo foi de 0,99.

As frequências com que, no treinamento (Figura 6A), os coeficientes NS pelo menos igualaram o valor de 0,99 foram, para cada pluviógrafo (13, 23, 34, 37, 40, 41, 42, 43 e 51), de, respectivamente, 68%, 83%, 23%, 16%, 22%, 68%, 27%, 12% e 90%. Os NS que igualaram ou ultrapassaram o valor de 0,97, por sua vez, ocorreram, respectivamente, com frequências 100%, 100%, 88%, 84%, 40%, 100%, 92%, 69% e 100%. As frequências com que, na verificação (Figura 6B), os coeficientes NS pelo menos igualaram o valor de 0,99 foram, para cada pluviógrafo (13, 23, 34, 37, 40, 41, 42, 43 e 51),

de, respectivamente, 0%, 0%, 2%, 4%, 0%, 41%, 5%, 0% e 0%. Os NS que igualaram ou ultrapassaram o valor de 0,97, por sua vez, ocorreram, respectivamente, com frequências 24%, 100%, 21%, 20%, 36%, 46%, 63%, 14% e 92%.

CONCLUSÃO

Este trabalho apresentou três alternativas de modelos para o preenchimento de falhas das séries pluviométricas horárias da bacia do arroio Taboão – RS utilizando Redes Neurais Artificiais (RNAs). A primeira proposição, com utilização de modelos com séries longas (mais de 850 valores), apresentou no período de verificação coeficiente de Nash-Sutcliffe menor que 0,85. A segunda proposição, utilizando séries curtas para treinamento da RNA (de 300 a 850 valores) com a série de treinamento e de verificação realizadas em diferentes períodos sazonais, apresentou um coeficiente de Nash-Sutcliffe melhor que o da primeira alternativa ($NS > 0,9$), porém nem sempre conseguiu reproduzir as chuvas de verão (alta intensidade e curta duração), considerando-se a correspondência espacial esperada. A terceira proposição, com séries curtas e considerando a mesmo período sazonal para a série de treinamento e de verificação, apresentou, assim como a segunda, resultados satisfatórios ($NS > 0,9$). Além disso, observou-se visualmente a capacidade de reprodução dos picos de verão, demonstrando a capacidade do método em produzir soluções regulares e consistentes, com 100% dos testes apresentando coeficientes de Nash-Sutcliffe entre 0,91 e 0,99 para o treinamento dos modelos e 0,90 e 0,99 para a verificação.

Desta forma, consideramos mais prudente, para o preenchimento de falhas de dados horários nesta bacia, a utilização de modelos que tenham sido treinados com dados do mesmo período sazonal, na sequência anterior ou posterior ao período a ser preenchido, evitando assim os possíveis efeitos da variação temporal do processo representado pelo modelo.

Os resultados demonstram que a escolha da série de treinamento na mesma estação climática que os dados a serem preenchidos foram o fator mais importante para a reprodução adequada das chuvas horárias, inclusive dos picos de verão. Quando o modelo foi treinado no verão, com série curta (12 a 30 dias), o preenchimento de falhas deve ser realizado nesta mesma estação climatológica, no mesmo ano, na sequência imediata, anterior

ou posterior ao treinamento. Se o modelo foi treinado com uma série no verão, as séries não poderão ser preenchidas no inverno, por exemplo. Quando o modelo é treinado com uma série curta para aplicação na mesma estação climática, tendo captado e incorporado as variações específicas daquela estação, consegue reproduzir bem esta chuva em uma frequência horária.

Os resultados obtidos nesta pesquisa foram promissores, e podem ser muito úteis, na medida em que permitirão a reconstrução de uma série histórica, na escala horária. Porém, deve-se ressaltar que a eficiência das redes neurais artificiais para preenchimento de falhas de dados horários é dependente das condições locais de cada região (tamanho da bacia, distribuição espacial de chuva na bacia), qualidade dos dados de entrada e da escolha e da configuração da rede a ser utilizada.

AGRADECIMENTOS

Agradecemos ao MCT/FINEP/AÇÃO TRANSVERSAL - Previsão de Clima e Tempo 04/2008, convênio 1406/08, projeto 01.08.0568.00 pelo financiamento da pesquisa e ao CNPq pela bolsa de doutorado do primeiro autor e pela bolsa de produtividade do terceiro autor.

REFERÊNCIAS

- ANMALA, J., ZHANG, B., GOVINDARAJU, R.S. Comparison of ANNs and empirical approaches for predicting watershed runoff. *Journal of Water Resources Planning and Management*, Reston: ASCE. v.126, n.3, p.156 – 166. 2000.
- BARNETCHE, D., KOBAYAMA, M. Aplicação do hycymodel no preenchimento de falhas de monitoramento de vazões. *Geosul*, v.21, n42, p. 185 – 194, 2006.
- BELTRAME, L. F. de S. (coord.) Consistência de Dados Hidrológicos da Bacia Hidrográfica do Alto Uruguai, Sub-Bacia 75. Porto Alegre, Instituto de Pesquisas Hidráulicas da UFRGS. 2000.
- BERTONI, J. C., TUCCI, C. E. M. Precipitação. In: TUCCI, C. E. M. *Hidrologia: Ciência e Aplicação*. Porto Alegre, UFRGS. p. 177 – 241. 2007.

- BRAGA, S. M., FERNANDES, C. V. S. Performance de Sensores de Precipitação do Tipo “Tipping Bucket” (Báscula) — Um Alerta para a Ocorrência de Erros. *Revista Brasileira de Recursos Hídricos* v.12, n.1, p.197-204, 2007.
- BRAVO, J. M.; COLLISCHONN, W.; UVO, C. B. ; PEDROLLO, O. C.; CHOU, S. C. Incorporating Forecasts of Rainfall in Two Hydrologic Models Used for Medium-Range Streamflow Forecasting. *Journal of Hydrologic Engineering*, v. May, p. 1-11, 2009.
- BRAVO, J.M.; PEDROLLO,O.C.; COLLISCHONN, W.; PAZ, A. R. ; UVO, C. B. Previsões de curto prazo de vazão afluente ao reservatório de Furnas utilizando redes neurais artificiais. *Revista Brasileira de Recursos Hídricos*, v. 13, p. 77-88, 2008.
- CARVALHO, A. C. A. Caracterização da rede pluviométrica do Estado de Sergipe e aplicação das redes neurais para preenchimento de falhas. Programa de Pós-graduação em Desenvolvimento e Meio Ambiente. Universidade Federal do Sergipe. Dissertação de Mestrado. 117p. 2007.
- CARVALHO, M.P.E., LOMBARDI N, F.; VASQUES F, J. Erosividade da chuva de Mococa (SP) analisada pelo índice EI_{30} . *Revista Brasileira de Ciência do Solo*, Campinas,v.13, n.2, p.243-249, 1989.
- CASTRO, N. M. R., CHEVALLIER, P. Funcionamento hidrossedimentológico de uma Terra Roxa Estruturada. In: Simpósio Brasileiro de Recursos Hídricos, Gramado, 1993. Anais, Porto Alegre: AB-RH, v. 2, pg. 456-465.1993.
- CASTRO, N. M. R., AUZET, A.V., CHEVALLIER, P., LEPRUN, J. C. Land use change effects on runoff and erosion from plot to catchment scale on the basaltic plateau of Southern Brazil.. *Hydrological Processes*, Chichester, Inglaterra, v. 13, n. 11, p. 1621-1628, 1999.
- CASTRO, N.M., CHEVALIER, P., GOLDENFUM, J.A. Projeto Potiribu, atualização 1989-1998: dados básicos de fluviometria e pluviometria. Porto Alegre. Instituto de Pesquisas Hidráulicas da UFRGS. (*Recursos Hídricos* Publicação, n. 35). 2000.
- CHEVALLIER, P., CASTRO, N. M. R. As precipitações na região de Cruz Alta e Ijuí RS-Brasil. *Recursos Hídricos*, Porto Alegre, v. 24, p. 1-24, 1991.
- CHIBANA, E. Y., FLUMIGNAN, D., MOTA, R. G., VIEIRA, A. S., FARIA, R. T. Estimativa de falhas em dados meteorológicos. In: Congresso Brasileiro de Agroinformática, 9. Londrina, 2005.
- CONTRELL, M., GIRARD, B., T., MANGUEAS, M.; MULLER, C. Neural Modeling for time series: A statistical stepwise method for weight elimination. *IEE Transaction on Neural Networks* 6(6), p. 1355-1364. 1995.
- CRUCIANI, D. E., MACHADO, R. E., SENTELHAS, P. C. Modelos da distribuição temporal de chuvas intensas em Piracicaba, SP. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.6, n.1, p.76-82, 2002.
- CUO, L., PAGANO, T. C., WANG,Q. J. A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting. *J. Hydrometeor*, 12, p. 713 – 728, 2011.
- DORNELLES, F., GOLDENFUM, J. A., PEDROLLO, O. C. Artificial neural network methods applied to forecasting river levels. *RBE. Revista Brasileira de Engenharia. Caderno de Recursos Hídricos* (Cessou em 1995. Cont. ISSN 1414-381X RBRH). , 2013. Aceito para publicação no Vol. 18 n° 4, outubro/dezembro/ 2013.
- ELSHORBAGY, A., SIMONOVIC, S. P., PANU, U. S. Performance evaluation of artificial neural networks for runoff prediction. *Journal of Hydrologic Engineering*, v.5, n.4, p.424-427, 2000.
- GARDNER, M. W., DORLING, S. R. Artificial neural networks (the multilayer perceptron) – A review of applications in the Atmospheric Sciences. *Atmospheric Environment*, v. 32, n.1 4-15, p. 2627-2636. 1998.
- GOMES, L. F. C., MONTENEGRO, S. M. G. L., VALENÇA, M. J. S. Modelo baseado na técnica de redes neurais para previsão de vazões na bacia do rio São Francisco. *RBRH-Revista Brasileira de Recursos Hídricos*, v. 15 n.1 p.05-15, 2010.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. 1989. Multilayer feedforward networks are universal approximators. *Neural networks*, v. 2, n. 5, p. 359-366.

- LACHTERMARCHER, G., FULLER, J. D. Backpropagation in time-series forecasting. *Journal of Forecasting* 14(4) p. 381-393, 1995.
- LUCIO, P. S., CONDE, F. C., CAVALCANTI, I. F. A.; RAMOS, A. M., CARDOSO, A. O. Reconstrução de séries meteorológicas via redes neurais artificiais. In: *Anais XIV CBMET, Florianópolis – SC.* 2006.
- MAIER, Holger R.; DANDY, Grame C. Neural network based modelling of environmental variables: A systematic approach. *Mathematical and Computer Modelling*, v. 33, n. 6, p. 669-682, 2001.
- MAIER, H., DANDY, G. Neural networks for the predictions and forecasting of water resources variables: review of modeling issues and applications. *Environmental Modelling & Software*, v. 15, n. 1, p. 101-124, 2000.
- MCCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics.*, Vol. 5, pp. 115-133, 1943.
- MINUZZI, R. B., SEDIYAMA, G. C., RIBEIRO, A., COSTA, J. M. N. El Niño: ocorrência e duração dos veranicos do Estado de Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 9(3), p. 364-371, 2005.
- NEVES, J., CORTEZ, P. An artificial neural network – Genetic based approach for time series forecasting. In: *Brazilian Symposium on Neural Networks (SBNR) IEEE.* p. 9 – 13. 1997.
- OLIVEIRA, G., PEDROLLO, O. C., CASTRO, N. M. R., BRAVO, J. M. Simulações hidrológicas com diferentes proporções de área controlada na bacia hidrográfica. *RBE. Revista Brasileira de Engenharia. Caderno de Recursos Hídricos (Cessou em 1995. Cont. ISSN 1414-381X RBRH).*, 2013. Aceito para publicação nº Vol. 18, n.3 (jul/set).
- RUMELHART, D.E., HINTON, G.E., WILLIAMS, R.J. Learning representations by back-propagating errors. *Nature*, vol. 323, pp. 533-536, 1986.
- SANTANA, M.O., RIBEIRO, A., SEDIYAMA, G. C. Sistema de geoespacialização da demanda de irrigação suplementar para o Estado de Minas Gerais I: Descrição. *Campina Grande. Rev. bras. eng. agríc. ambient.* v.7, n.1, 2003.
- SANTOS, G. G., FIGUEIREDO, C. C. DE, OLIVEIRA, L. F. C. DE., GRIEBELER, N. P. Intensidade-duração-frequência de chuvas para o Estado de Mato Grosso do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.13, p.899-905, 2009.
- SILVA, V. S. V. Estimativa de precipitação pontual em diferentes escalas para uso em modelo concentrado chuva-vazão. (Dissertação) Programa Pós-graduação em Recursos Hídricos e Saneamento Ambiental. UFRGS. Porto Alegre. 2011.
- SILVA, P.M.O., MELLO, C.R., SILVA, A.M., COELHO, G. Modelagem da hidrografia de cheia em uma bacia hidrográfica da região Alto Rio Grande. *R. Bras. Eng. Agric. Amb.*, v.12, p.258-265, 2008.
- SILVA, A. F., COSTA, L. C., SEDIYAMA, G. C. Previsão da Evapotranspiração de referência utilizando redes neurais. *Engenharia na Agricultura*, v. 14, p. 93-99, 2006.
- SOARES, P. A., PINHEIRO, A., SOARES, K.H., ZUCCO, E. Estimativa da disponibilidade hídrica em pequenas bacias hidrográficas com escassez de dados fluviométricos. *Revista de estudos ambientais (Online)*v.12, n. 1, p. 29-38, 2010.
- SOUSA, W. dos S., SOUSA, F. de A. S. Rede neural artificial aplicada à previsão de vazão da Bacia Hidrográfica do Rio Piancó. *Revista Bras. de Eng. Agrícola e Ambiental* v.14, n.2, p.173–180, 2010.
- WEIGEND, A. S., RUMERLHART, D., HUBERMAN, B. A. Predicting the future: A connectionist approach. *International Journal Neural System* 1:3. 1990.
- WIDROW, B.; HOFF, M. E. 1960. Adaptive switching circuits. *IRE WESCON Convention Record*, New York: IRE Part 4, pp. 96-104.
- ZHANG, G., PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, [S.l.], v.14, Issue 1, p.35-62, 1998.

**Filling Hourly Data Gaps In Rainfall Time Series
Using Artificial Neural Networks**

ABSTRACT

The availability of continuous rainfall time series can enable the execution of many studies that were not possible using series with gaps, but these are usually part of the reality of the data available due to technical, operational and budget difficulties. The objective of this research is to test a methodology with artificial neural networks (ANN) to fill gaps in hourly series of rainfall data. We used a series of rainfall data from nine stations in a 78 km² basin, located in the northwestern region of Rio Grande do Sul, which has a temperate climate with four clearly defined annual seasons. Three different alternatives have been tried with regard to the length and the arrangement of the time series used: 1. Long series, with approximately 365 days, for training and verification; 2. Short training series, with approximately 12 to 30 days, and also several verification series, each one independent of the weather station; 3. Short series (also 12-30 days), similarly to the previous alternative, except that the verification is done only in periods close to training. The best results were obtained with the third alternative, whose Nash-Sutcliffe (NS) coefficients were higher than 0.9, with both the training and verification series. This demonstrates that even in basins where rainfall is considered homogeneous, relations between rain gauges are not constant, requiring training for each specific atmospheric process to fill the gaps in short, successive periods.

Key-words: continuous time series; rain hourly data; neural network, seasonal training.