

METODOLOGIA DE ANÁLISE DE SENSIBILIDADE E EXCLUSÃO DE VARIÁVEIS DE ENTRADA EM SIMULAÇÃO HIDROLÓGICA POR REDES NEURAIS ARTIFICIAIS (RNAs): resultados preliminares.

Guilherme Garcia de Oliveira¹; Olavo Correa Pedrollo²; Nilza Maria dos Reis Castro³

RESUMO --- Em hidrologia, muitas pesquisas obtiveram excelentes resultados com a aplicação de modelos em RNAs, embora estas forneçam poucas informações sobre a influência das entradas na vazão de saída. Existe uma carência na literatura científica de metodologias para exclusão das variáveis menos influentes em RNAs. Assim, o objetivo deste trabalho é apresentar um método de análise de sensibilidade e exclusão de variáveis de entrada em uma simulação hidrológica por RNAs. A análise de sensibilidade consistiu em provocar perturbações nos registros das entradas e verificar a redução no desempenho do modelo. A cada rodada de exclusão foram descartadas as duas variáveis menos influentes. A simplificação da RNA com a redução de suas dimensões melhorou o desempenho das simulações hidrológicas e reduziu o processamento computacional. A capacidade da RNA foi potencializada ao eliminar variáveis desnecessárias, até um determinado limite, quando o desempenho foi prejudicado. Com o método proposto foi possível compreender a contribuição das variáveis de entrada na saída da RNA. Verificou-se que a exclusão de uma ou duas variáveis da RNA pode alterar significativamente a contribuição das outras entradas no modelo. Algumas entradas, embora não apresentem elevados índices de contribuição, conferem estabilidade à simulação, pois quando excluídas reduzem o desempenho da simulação.

ABSTRACT --- In hydrology, many studies have obtained excellent results with the application of ANN models, although these provide little information on the influence of the entries in the output flow. There is a dearth of scientific literature on methodologies for deletion of the least influential variables in ANNs. The objective of this paper is to present a method of sensitivity analysis and exclusion of input variables in a simulation hydrological of ANNs. The sensitivity analysis was to provoke disturbances in the records of inputs and verify the reduction in model performance. Each round of exclusion was eliminated two variables less influential. The simplification of the ANN with the reduction of its dimensions improved performance of the hydrological simulations and reduced computational processing. The ANN ability was enhanced by eliminating unnecessary variables, to a certain limit, when performance was impaired. With the proposed method was possible to understand the contribution of input variables in the output of the ANN. It was found that the exclusion of one or two variables of ANN can significantly change the contribution of other inputs in the model. Some entries, while not having high rates of contribution, provide stability to the simulation, because when deleted reduce simulation performance.

Palavras-chave: Simulações hidrológicas; redes neurais artificiais; análise de sensibilidade.

¹ Doutorando em Recursos Hídricos e Saneamento Ambiental UFRGS – IPH, E-mail: g.g.oliveira10@gmail.com

² Professor Adjunto UFRGS – IPH, Av. Bento Gonçalves, 9.500, 91501-970, CP 15029, Porto Alegre/RS. E-mail: olavopedrollo@gmail.com

³ Professora Adjunta UFRGS – IPH, Av. Bento Gonçalves, 9.500, 91501-970, CP 15029, Porto Alegre/RS. E-mail: nilza@iph.ufrgs.br

1 - INTRODUÇÃO

O surgimento das redes neurais artificiais (RNAs) e o crescente desenvolvimento tecnológico em computação tem proporcionado resultados promissores em simulações hidrológicas (Jain *et al.*, 2004). As RNAs são extremamente eficazes para modelar os complexos relacionamentos que ocorrem na natureza, destacando-se a sua aplicação no desenvolvimento de simulações, modelos de previsão e classificação (Valença & Ludermir, 2007).

A criação das RNAs está ligada à publicação de McCulloch & Pitts (1943) com inspiração puramente biológica. A rede possuía capacidade de representar funções lógicas simples, de adaptação e aprendizado de forma semelhante ao cérebro humano. Porém, é somente a partir da publicação do estudo de Rosenblatt (1958) que surge a primeira RNA com aplicabilidade aos problemas práticos observados na natureza, o Perceptron. Esse tipo de rede possui capacidade de aprendizado com os erros a partir de pesos aleatórios, sendo a base para a compreensão das RNAs mais complexas que surgiriam nas décadas seguintes.

Em seguida, Widrow & Hoff (1960) idealizaram uma técnica de aprendizado tipo gradiente-descendente com aplicações em diversas áreas, tais como o reconhecimento de padrões e os modelos adaptativos para previsão. O modelo adaptativo (Regra Delta) possui um potencial ampliado por ser constituído de funções de ativação contínuas e não-lineares.

No entanto, somente com o algoritmo retropropagativo, desenvolvido por Rumelhart *et al.* (1986), é que as RNAs passaram a ser difundidas e receber investimentos em pesquisa de forma contínua e crescente. O algoritmo retropropagativo é uma generalização da Regra Delta aplicado às redes multicamadas. Os erros da camada oculta de neurônios (intermediária) são calculados a partir dos erros da camada de neurônios de saída. A atualização dos pesos de um neurônio depende apenas dos erros, entradas e derivada do próprio neurônio. Com apenas uma camada intermediária e a adoção do número correto neurônios nessa camada e de variáveis de entrada consegue-se resolver qualquer problema prático.

Atualmente, no âmbito das simulações hidrológicas, diversas pesquisas obtiveram excelentes resultados com a aplicação de modelos em RNAs, como por exemplo, os trabalhos de Smith & Eli (1995), Shamseldin (1997), Dawson & Wilby (1998), Sajikumar & Thandaveswara (1999), Chang & Chen (2001), Sudheer *et al.* (2002), Jain *et al.* (2004) e Jain & Kumar (2007).

Entretanto, apesar dos resultados promissores, estas são comumente rotuladas como modelos “caixa preta”. Essa crítica se baseia no fato das RNAs fornecerem poucas informações sobre a influência das variáveis de entrada sobre a variável de saída, pois esta influência é difícil de ser compreendida através dos pesos sinápticos da rede (Bishop, 1995; Valença & Ludermir, 2007). No

caso das simulações hidrológicas, as RNAs não explicam os processos físicos inerentes à dinâmica dos rios e bacias hidrográficas que originam o escoamento (Hsu *et al.*, 1995; Jain *et al.*, 2004).

Com o intuito de analisar o relacionamento entre as variáveis de entrada e saída das RNAs, foram sugeridos alguns algoritmos que realizam a análise de sensibilidade para determinar a influência de cada variável nos resultados. Como exemplos, pode-se citar Garson (1991), Goh (1993), Oh & Lee (1995), Mak & Blanning (1998), Guha *et al.* (2005) e Valença & Ludermir (2007). Além disso, esses algoritmos permitem que se descarte alguns dados irrelevantes à RNA para reduzir a dimensão dos dados e processamento computacional.

Na análise da sensibilidade gera-se uma perturbação nas variáveis de entrada para medir o impacto nos pesos sinápticos ou na saída da RNA (Goh, 1993; Oh & Lee, 1995). Algumas abordagens foram concebidas para permitir uma medida de correlação entre as entradas e saídas da rede, como o método descrito por Mak & Blanning (1998). No entanto, a obtenção explícita de elasticidades nesse tipo de modelo é de difícil implementação, em função do elevado grau de não-linearidade que pode surgir na RNA (Silva, 2002). No estudo de Goh (1993), após a realização das perturbações nas entradas da rede, o índice de sensibilidade é o somatório das diferenças na saída do modelo entre os dados originais e os “perturbados”.

No entanto, existe uma carência na literatura científica de propostas metodológicas para exclusão das variáveis menos influentes na RNA. Nos trabalhos mencionados acima, são propostos apenas métodos para verificar a influência, mas não para a redução da dimensionalidade da RNA, tão pouco é realizada uma análise do desempenho da simulação ao longo do processo de exclusão de entradas.

Nesse contexto, o objetivo deste trabalho é apresentar uma metodologia para análise de sensibilidade e exclusão de variáveis de entrada em uma simulação hidrológica por RNAs, com a finalidade de reduzir a dimensionalidade dos dados, minimizando o processamento computacional e melhorando o desempenho da RNA.

2 - MATERIAIS E MÉTODOS

A simulação foi realizada na bacia hidrográfica do rio Ijuí (Figura 1), no Rio Grande do Sul, no posto fluviométrico Santo Ângelo, operado pela Companhia de Pesquisa em Recursos Minerais (CPRM). A área de contribuição da bacia até o posto é de 5.414 km². A escolha desta área de estudo se deve ao fato do monitoramento realizado pelo Instituto de Pesquisas Hidráulicas (IPH) da Universidade Federal do Rio Grande do Sul desde 1989 em pequenas bacias da região, com banco de dados hidrológicos de mais de 20 anos.

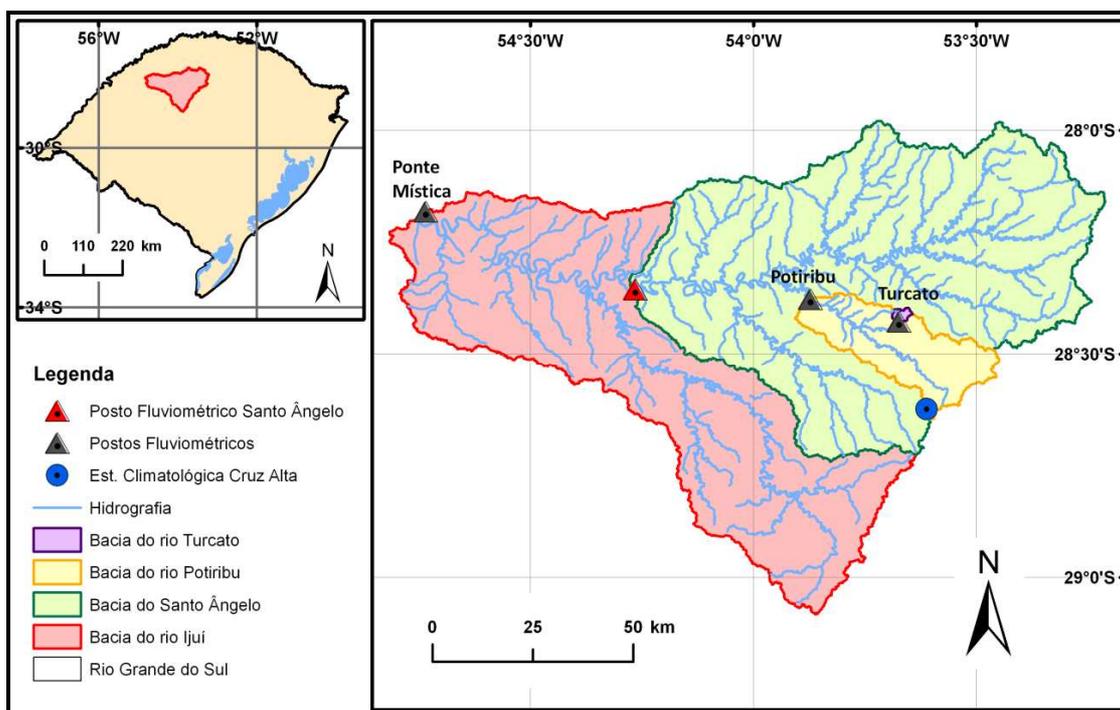


Figura 1. Mapa da área de estudo: postos fluviométricos e climatológico, hidrografia e bacias.

2.1 - Materiais

Este estudo foi desenvolvido no aplicativo MATLAB R2010a e os materiais utilizados foram os seguintes: a) dados de precipitação média interpolada pelo método do vizinho natural para a bacia do rio Santo Ângelo, disponibilizados pelo IPH, cuja metodologia está descrita em Silva (2011); b) dados de vazão do posto fluviométrico do rio Turcato, disponibilizados pelo IPH; c) dados de vazão dos postos fluviométricos Santo Ângelo, Ponte Mística e Ponte Nova do Potiribu, acessíveis pelo portal Hidroweb, da Agência Nacional de Águas (ANA); d) dados de evaporação obtidos pelo Tudo de Piché na estação climatológica Cruz Alta, obtidos junto ao banco de dados do Instituto Nacional de Meteorologia (INMET).

2.2 - Métodos

Neste item serão descritos os procedimentos metodológicos aplicados à simulação hidrológica com RNAs, à análise de sensibilidade e à exclusão de variáveis de entrada para a redução da dimensionalidade dos dados.

2.2.1 - Simulação hidrológica com RNAs

Para este estudo, foi desenvolvido um programa para simulação com RNA, utilizando o algoritmo retropropagativo com validação cruzada, no qual a simulação compreende um determinado número de iterações, e cada iteração compreende um certo número de ciclos. Com base em testes realizados previamente, a configuração da RNA foi definida da seguinte maneira: i)

número de iterações: 20; ii) número de ciclos máximo em cada iteração: 5.000; iii) número de neurônios na camada oculta igual ao de variáveis de entrada. Logo, à medida que são excluídas as variáveis de entrada, após a análise de sensibilidade, o número de neurônios na camada oculta também é reduzido, simplificando a RNA.

Foram testadas três combinações de variáveis de entrada, levando em consideração a disponibilidade de dados e uma análise exploratória dos dados com extração de estatísticas básicas. Nas duas primeiras combinações para simulação da vazão Q, no tempo T, no rio Ijuí (Posto Santo Ângelo), foram utilizados apenas os dados de precipitação interpolados, de vazão do posto Ponte Mística e de evaporação do posto Cruz Alta. O período utilizado para treinamento da RNA nessas combinações foi de 12/01/02 a 30/12/03, para validação de 26/05/90 a 31/01/92 e verificação de 06/01/89 a 25/05/90.

Na terceira combinação, foi acrescida a vazão nos postos Potiribu e Turcato, entretanto, em um período menor da série histórica e mais fragmentado temporalmente, tendo em vista as falhas observadas nessas séries. O período de treinamento foi de 11/09/93 a 07/12/95, para validação de 02/03/93 a 10/09/93 e para verificação de 01/12/90 a 01/03/93.

A primeira combinação compreende 17 variáveis de entrada (17 neurônios na entrada e na camada oculta e um na saída – 17x17x1): a) precipitação média interpolada nos tempos T, T-1, T-2, T-3 e T-4, denominadas de P(T), P(T-1), P(T-2), P(T-3) e P(T-4); b) precipitação acumulada nas últimas 72 h, 96 h e 120 h, denominadas de P(72h), P(96h) e P(120h); c) precipitação ponderada pelo coeficiente de correlação entre a precipitação nos tempos T, T-1, T-2, T-3 e T-4 e a vazão no tempo T, denominada de P(r); d) precipitação filtrada por função de decaimento exponencial (Equação 1), denominada de P(f); e) evaporação nos tempos T e T-1, denominadas de E(T) e E(T-1); f) vazão no posto Ponte Mística nos tempos T, T-1, T-2, T-3 e T-4, denominadas de Qo(T), Qo(T-1), Qo(T-2), Qo(T-3) e Qo(T-4). Já na segunda combinação, foram selecionadas as variáveis que obtiveram coeficiente de correlação (r) superior a 0,5. Com base nesse critério, foram descartadas cinco das 17 variáveis iniciais, sendo elas: P(T), P(T-4), E(T), E(T-1) e Q(T-4), isto é, dimensão 12x12x1.

$$P(f)_t = (1 - \alpha).P(f)_{t-1} + \alpha.P_t \quad (1)$$

Onde: $P(f)_t$ é a chuva filtrada com decaimento exponencial no tempo t; α é um coeficiente com valores entre 0 e 1; $P(f)_{t-1}$ a chuva filtrada com decaimento exponencial no tempo t-1; P_t é a chuva observada no tempo t.

Na terceira combinação, com o uso das séries de vazão de Potiribu e Turcato, foram utilizadas somente variáveis com r superior a 0,5, totalizando 17 variáveis (RNA com dimensões 17x17x1): i)

P(T-1), P(T-2) e P(T-3); ii) P(72h), P(96h) e P(120h); iii) P(r) e P(f); iv) Qo(T), Qo(T-1) e Qo(T-2); v) vazão do rio Potiribu nos tempos T, T-1, T-2 e T-3, sendo denominadas de Qo2(T), Qo2(T-1), Qo2(T-2) e Qo2(T-3); vi) vazão do rio Turcato nos tempos T-1 e T-2, sendo denominadas de Qo3(T-1) e Qo3(T-2). A Tabela 1 sintetiza as configurações utilizadas.

Tabela 1 – Síntese das combinações de variáveis de entrada das RNAs.

	Combinação 1	Combinação 2	Combinação 3
Amostras	1.835	1.835	907
Treinamento	714	714	483
Validação	616	616	154
Verificação	505	505	270
Variáveis de entrada	17	12	17
Precipitação	P(T), P(T-1), P(T-2), P(T-3), P(T-4), P(72h), P(96h), P(120h), P(r), P(f)	P(T-1), P(T-2), P(T-3), P(72h), P(96h), P(120h), P(r), P(f)	P(T-1), P(T-2), P(T-3), P(72h), P(96h), P(120h), P(r), P(f)
Evaporação	E(T), E(T-1)	-	-
Vazão	Qo(T), Qo(T-1), Qo(T-2), Qo(T-3), Qo(T-4)	Qo(T), Qo(T-1), Qo(T-2), Qo(T-3)	Qo(T), Qo(T-1), Qo(T-2), Qo2(T), Qo2(T-1), Qo2(T-2), Qo2(T-3), Qo3(T-1), Qo3(T-2)

2.2.2 - Análise de sensibilidade

A análise de sensibilidade consistiu em provocar pequenas perturbações nos registros das variáveis de entrada na série de verificação e averiguar a contribuição de cada uma no cálculo da vazão Q(T) no posto Santo Ângelo. O índice que define a contribuição de cada variável de entrada na saída da RNA é obtido pela diferença no desempenho da simulação (com e sem perturbação) através do coeficiente Nash-Sutcliffe (R^2). Quanto maior a diferença entre o R^2 da simulação com perturbação nos valores de uma variável de entrada x e o R^2 da simulação normal, maior a contribuição da variável x no cálculo da vazão Q(T). O R^2 é definido pela Equação 2 e o índice de contribuição de cada variável (Cont) é obtido pela Equação 3.

$$R^2 = 1 - \frac{\sum (Qo_i - Qc_i)^2}{\sum (Qo_i - \overline{Qo})^2} \quad (2)$$

Onde: Qo_i é a vazão observada; Qc_i é a vazão calculada; \overline{Qo} é a média das vazões observadas.

$$Cont_x = R^2 - R^2' \quad (3)$$

Onde: R^2 é o coeficiente de Nash-Sutcliffe obtido na simulação com perturbação de uma variável x ; $Cont_x$ é o índice de contribuição de uma variável x na simulação da vazão $Q(T)$.

As perturbações consistem na soma do valor observado com o produto de um número aleatório com distribuição normal ($\mu=0$, $\sigma=1$) com o valor médio da variável de entrada (Equação 4). A utilização do valor médio dos registros da variável de entrada garante que as perturbações terão magnitudes compatíveis com os valores observados em cada entrada.

$$x'_i = x_i + (nr \cdot \bar{x}) \quad (4)$$

Onde: x'_i é o valor perturbado da variável x no registro i ; x_i é o valor original da variável x no registro i ; nr é um número aleatório que segue a distribuição normal com $\mu=0$ e $\sigma=1$; \bar{x} é valor observado médio da variável x na série de verificação.

Ao todo, foram geradas 100 séries de números aleatórios com a mesma dimensão das amostras de verificação para a obtenção de um valor médio de contribuição para cada variável de entrada, evitando assim erros de interpretação nos resultados em virtude da aleatoriedade das perturbações. Além disso, para impedir que as perturbações resultem em valores negativos de precipitação, evaporação ou vazão, foi aplicada a inversão de sinal no valor perturbado sempre que necessário.

2.2.3 - Exclusão das variáveis de entrada

Após a análise de sensibilidade, são descartadas as duas variáveis com menor índice de contribuição. Em seguida, é realizada uma nova simulação da vazão $Q(T)$ sem as variáveis de entrada excluídas e o coeficiente R^2 da simulação é armazenado.

Em virtude dos pesos sinápticos das RNAs serem inicializados aleatoriamente, não se pode garantir que as variáveis excluídas são, de fato, as menos importantes para explicar a vazão $Q(T)$. Ao repetir todo o procedimento, podem-se obter outras variáveis para exclusão.

Logo, como alternativa para reduzir falhas na exclusão das variáveis, foi implantado um *looping* no qual se realiza a simulação, a análise de sensibilidade, a exclusão parcial de duas variáveis de entrada e uma nova simulação, cujo R^2 é armazenado. Ao final de x repetições do procedimento são excluídas permanentemente as duas variáveis de entrada que, ao serem descartadas, originaram uma nova simulação com o maior R^2 . Quanto mais repetições, maior a certeza de que as variáveis excluídas são as que menos explicam a vazão $Q(T)$. Neste trabalho, a exclusão de variáveis considerou 10 repetições.

Ao fim do procedimento, são descartadas permanentemente duas variáveis dando fim à primeira rodada de exclusão. Em seguida iniciam-se novas rodadas para reduzir a dimensionalidade do modelo até que a RNA tenha seu desempenho otimizado. Neste trabalho, na 1ª e 3ª combinações (com 17 entradas) foram definidas seis rodadas de exclusão (resultando em uma RNA de cinco entradas) e na 2ª combinação (com 12 variáveis) foram definidas apenas três rodadas (resultando em uma RNA de seis entradas). A Figura 2 ilustra a metodologia do trabalho de forma simplificada.

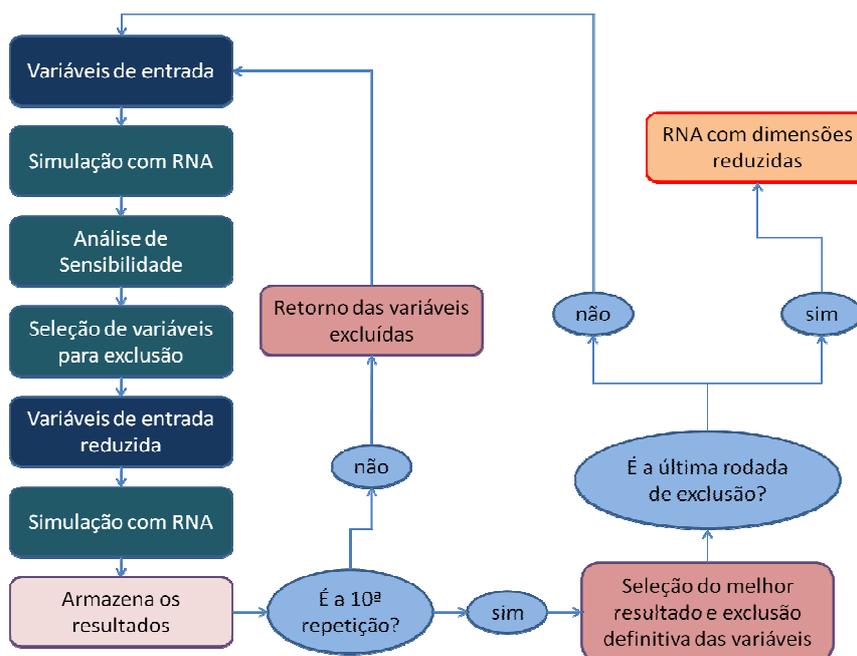


Figura 2. Fluxograma simplificado da metodologia.

3 - RESULTADOS E DISCUSSÕES

Os resultados e discussões do trabalho serão apresentados em dois tópicos. No primeiro será abordado desempenho da simulação hidrológica durante as rodadas de exclusão até a obtenção da RNA com dimensões reduzidas. No segundo, a análise de sensibilidade com a discussão sobre as variáveis de entrada que se destacaram na simulação, bem como os aspectos mais interessantes da exclusão de variáveis.

3.1 - Desempenho da simulação

Na primeira combinação de simulação, utilizando 17 variáveis de entrada e excluindo variáveis menos importantes até obtenção de uma RNA com cinco entradas, foram obtidos resultados muito satisfatórios, como pode ser observado nas Figuras 3 e 4. O coeficiente de Nash-Sutcliffe evoluiu de 0,9102 na primeira rodada (17x17x1) para 0,9472 na última (5x5x1). O

coeficiente de correlação também apresentou um crescimento de 0,9678 para 0,9768 com a redução da complexidade da rede pela exclusão de variáveis menos importantes.

Os outros indicadores do desempenho do modelo hidrológico com RNAs também apresentaram evolução ao longo do procedimento. O RMSE na primeira rodada (17 neurônios) foi de 64,4 m³/s e reduziu para 49,3 m³/s no final da simulação (5 neurônios). O erro médio absoluto diminuiu de 34,3 m³/s para 29,6 m³/s na RNA final.

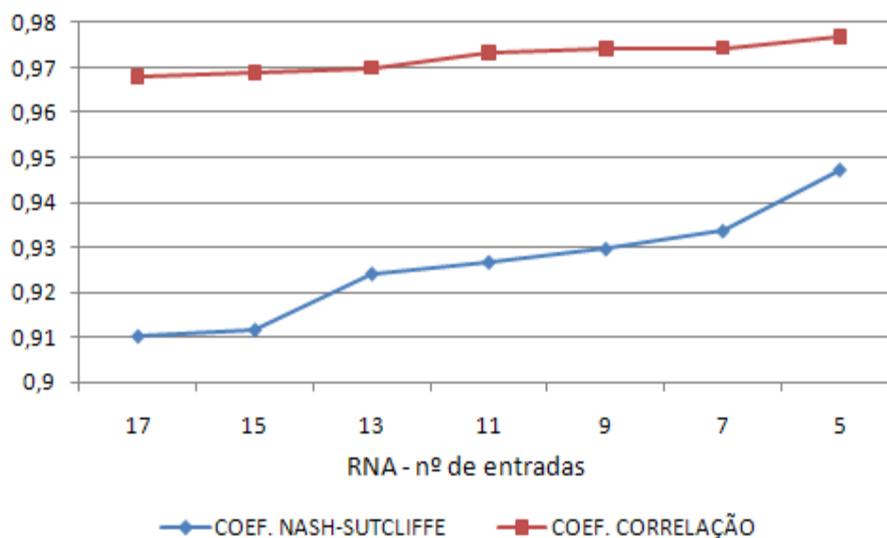


Figura 3. Evolução dos coeficientes de Nash-Sutcliffe e de correlação com a redução da dimensionalidade da RNA pela exclusão de variáveis (Combinação 1).

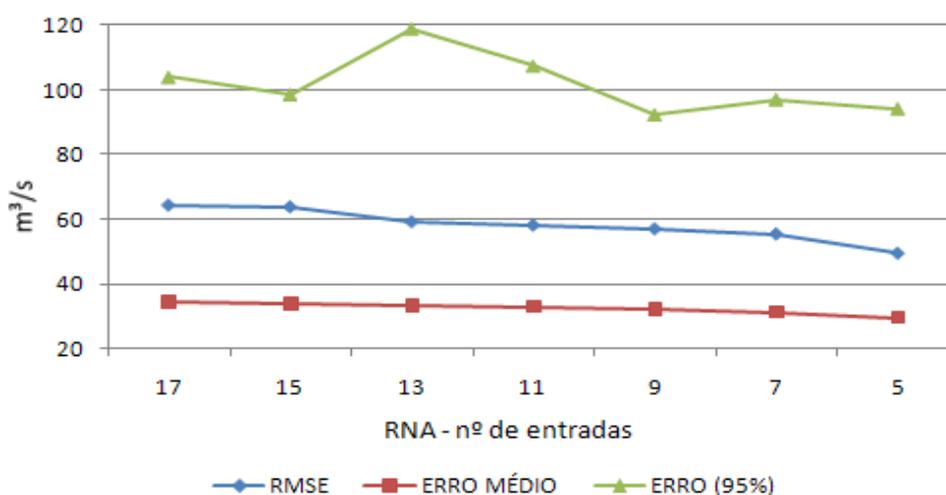


Figura 4. Redução das imprecisões do modelo (RMSE, erro médio e o valor aos quais 95% dos erros são inferiores).

Na segunda combinação de simulação, utilizando 12 variáveis de entrada (somente com $r > 0,5$) e excluindo variáveis menos importantes até obtenção de uma RNA com seis entradas, foram obtidos resultados ainda mais satisfatórios (Figuras 5 e 6): a) o coeficiente de Nash-Sutcliffe evoluiu de 0,9097 na rodada inicial (12 neurônios) para 0,9488 no final da simulação (6 neurônios); b) o coeficiente de correlação evoluiu de 0,9695 para 0,9773; c) o RMSE reduziu de 64,5 m³/s para 48,6

m^3/s e o erro médio de $34,3 m^3/s$ para $29 m^3/s$. Somente o erro com 95% de confiança não obteve melhora, variando de $95,5 m^3/s$ (12 neurônios) para $95,8 m^3/s$ (6 neurônios).

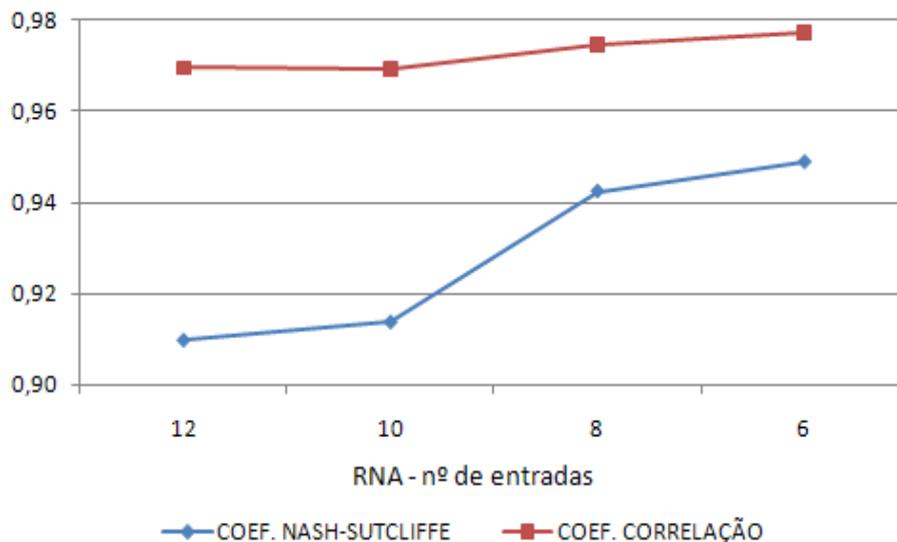


Figura 5. Evolução dos coeficientes de Nash-Sutcliffe e de correlação com a combinação 2.

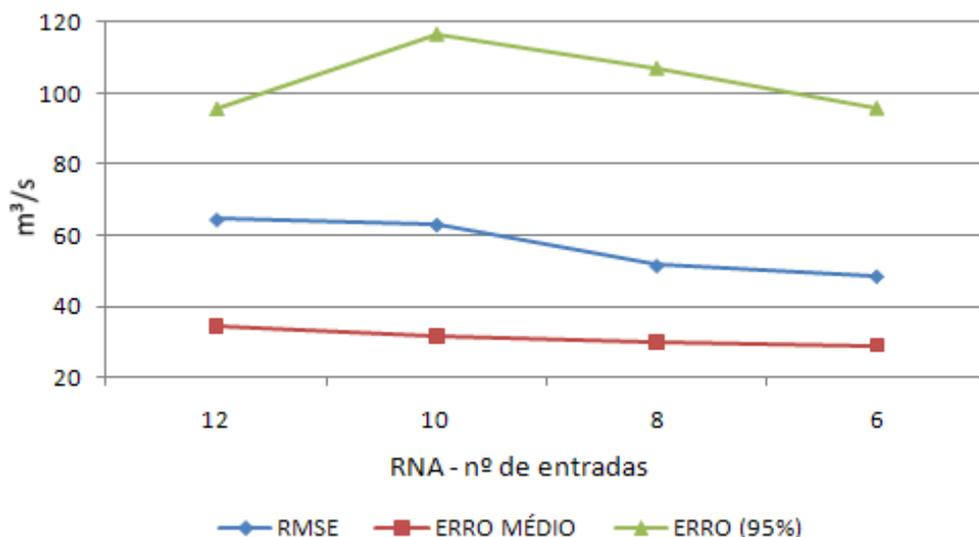


Figura 6. Redução das imprecisões do modelo na simulação com a combinação 2.

Ao utilizar a terceira combinação de variáveis, com dados de vazão de três postos fluviométricos, totalizando 17 variáveis de entrada, foram observados resultados diferentes dos verificados nas duas combinações anteriores, como ilustram as Figuras 7 e 8. O coeficiente R^2 se manteve estável (com pequenas oscilações) entre a primeira rodada (17 neurônios) e a penúltima (7 neurônios), entre 0,9491 e 0,9276, sendo que o melhor resultado foi obtido na segunda rodada. No entanto, ao excluir as últimas duas variáveis de entrada, na última rodada, houve um decréscimo significativo do coeficiente para 0,8535. Esse resultado indica que a RNA mínima para

interpretação desses dados de entrada exige mais de cinco neurônios, ao contrário do verificado nos outros casos.

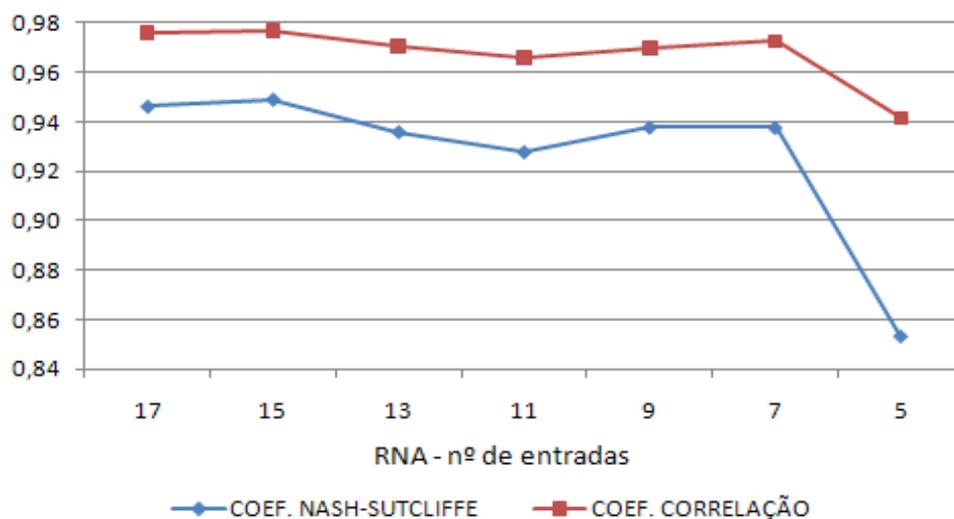


Figura 7. Desempenho da RNA nas rodadas de exclusão usando a combinação 3.

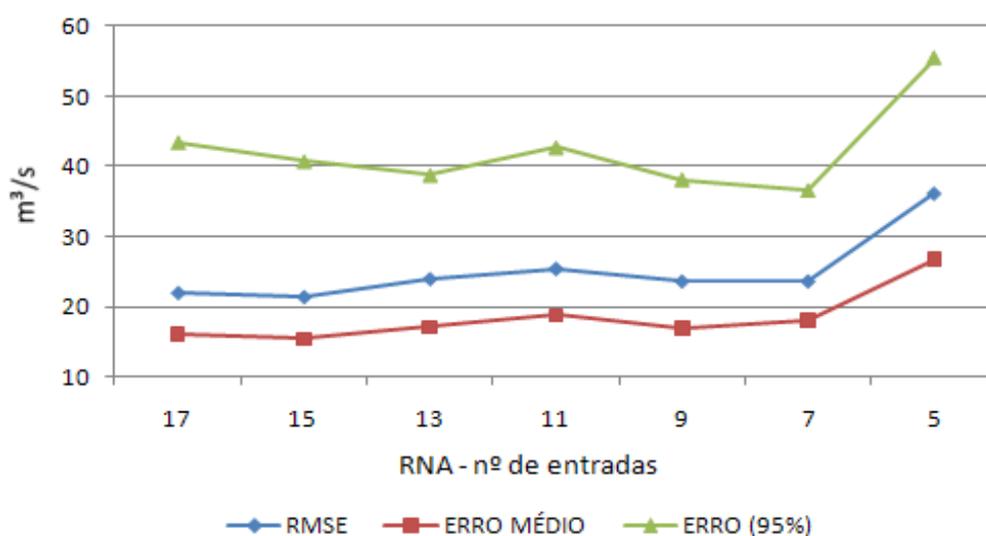


Figura 8. Imprecisões do modelo na simulação com a combinação 3.

Com os outros indicadores observou-se o mesmo: desempenho estável até a penúltima rodada, tendo a RNA com 15 neurônios os melhores resultados. Na última rodada de exclusão o desempenho sofreu forte queda, a qual será discutida com maior detalhe no tópico seguinte, de análise de sensibilidade.

De modo geral pode-se dizer que a simplificação da RNA com a redução de suas dimensões (neurônios de entrada e na camada oculta) melhora o desempenho da simulação hidrológica, além de reduzir o tempo de processamento computacional. A capacidade da RNA é potencializada ao eliminar variáveis desnecessárias ou redundantes, até um limite mínimo, no qual o desempenho tem brusca queda (Figura 8).

3.2 - Análise de sensibilidade e exclusão de variáveis de entrada

Na primeira combinação, apenas três variáveis de entrada apresentaram significativos índices de contribuição na simulação da vazão $Q(T)$: a chuva filtrada por um decaimento exponencial, $P(f)$, a vazão no posto Ponte Mística, $Qo(T)$ e a chuva ponderada pelo coeficiente de correlação, $P(r)$. A Figura 9 ilustra o comportamento do índice de contribuição dessas variáveis ao longo das rodadas de exclusão.

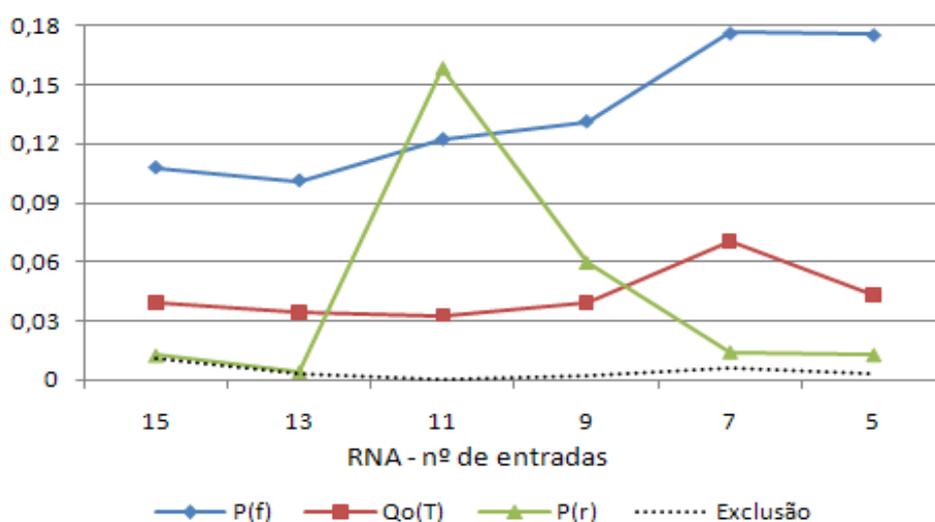


Figura 9. Índice de contribuição das variáveis mais importantes na combinação 1.

A variável $P(f)$ foi considerada a mais importante para a simulação da vazão, com índice de contribuição evoluindo de 0,108 (1ª rodada) para 0,176 (6ª rodada). Essa variável representa eficientemente o armazenamento de água no solo, sendo de fato, uma variável muito importante em qualquer simulação hidrológica (mesmo que expressa de outras formas). Além disso, na análise exploratória dos dados, foi observado um coeficiente de correlação alto entre a $P(f)$ e a vazão $Q(T)$. Ainda assim, esse resultado pode ser considerado surpreendente, pois outras entradas eram melhores correlacionadas com a saída da RNA.

A segunda variável mais importante foi a $Qo(T)$, com índice de contribuição de modo geral estável, entre 0,033 e 0,071. A elevada influência dessa variável era esperada, pois se trata da entrada com maior coeficiente de correlação com a vazão $Q(T)$. Na última rodada o índice dessa variável foi de 0,044, indicando uma contribuição significativa na vazão de saída, mas bem inferior à $P(f)$.

No entanto, um dos aspectos mais interessantes pode ser observado ao analisar a variável $P(r)$, terceira no índice de contribuição ao final do procedimento de exclusão, com valor de 0,013. Essa

variável representa a parcela da chuva que efetivamente está escoando no tempo T , no posto Santo Ângelo, uma vez que é obtida pela ponderação do coeficiente de correlação.

Nas duas primeiras rodadas a contribuição dessa variável esteve muito próxima do limiar de exclusão das duas variáveis menos importantes. No entanto, com a exclusão de quatro entradas, já na terceira rodada, a variável teve sua contribuição potencializada significativamente, chegando a ser a mais importante na RNA com 11 neurônios (índice de 0,158).

Esse resultado permite realizar uma consideração importante: a quantidade de exclusões por rodada pode influenciar na eficiência do procedimento de otimização da RNA. Ao excluir muitas variáveis de uma só vez pode-se excluir uma entrada importante em outras configurações de rede. Dependendo da configuração inicial da RNA, uma variável qualquer (bem correlacionada) pode obter resultados bem distintos de contribuição na saída da rede. O exemplo da variável $P(r)$ ilustra isso, já que a mesma foi uma das cinco menos importantes no início do procedimento.

A exclusão de uma ou duas variáveis da RNA pode alterar significativamente a contribuição das outras variáveis no modelo. Logo, recomenda-se a exclusão de poucas variáveis por rodada para evitar erros grosseiros no processo de redução das dimensões da RNA.

Outro aspecto importante que justifica a realização de algumas rodadas de exclusão é que na primeira rodada, com exclusão da $P(f)$ e da $Qo(T)$, todas variáveis de entrada obtiveram índices de contribuição muito semelhantes, isto é, muito próximos do limiar de exclusão. Logo, se a RNA fosse reduzida em apenas uma rodada, o risco de se obter um modelo limitado seria potencializado.

Além disso, outro resultado destacado diz respeito às entradas de chuva acumuladas: quando as três variáveis utilizadas estão sendo consideradas no modelo, nenhuma delas apresenta contribuição significativa no modelo. No entanto, ao excluir a $P(72h)$ na 3ª rodada e a $P(120h)$ na 4ª rodada, a outra variável, a $P(96h)$, tem sua contribuição aumentada, ficando entre as cinco variáveis finais da simulação na combinação 1.

Provavelmente isso ocorre em virtude da redundância dos dados de chuva acumulada, ou seja, da alta correlação entre as três entradas. A RNA “divide” a contribuição da chuva nos últimos dias entre três entradas. Possivelmente, ao eliminar duas delas, a contribuição da entrada restante é potencializada.

Já na segunda combinação, iniciando com 12 neurônios e finalizando com seis, além das três variáveis destacadas na simulação anterior, $P(f)$, $P(r)$ e $Qo(T)$, outra entrada obteve contribuição significativa na simulação da vazão $Q(T)$: a chuva acumulada dos últimos cinco dias, a $P(120h)$. A Figura 10 ilustra o comportamento do índice de contribuição dessas variáveis ao longo das rodadas de exclusão.

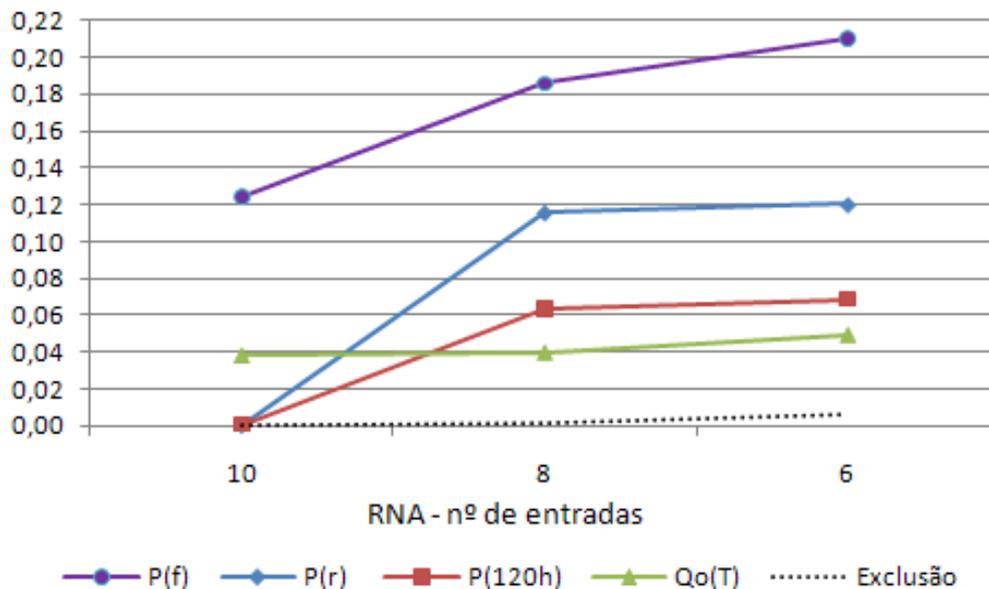


Figura 10. Índice de contribuição das variáveis mais importantes na combinação 2.

A variável P(f) novamente foi considerada a mais importante para a simulação da vazão, com índice de contribuição crescendo de 0,125 (1ª rodada) para 0,211 (3ª rodada). A segunda variável mais importante foi a P(r), com índice de contribuição evoluindo de 0,001 para 0,12 na RNA com seis neurônios de entrada. A variável P(120h) também apresentou grande evolução na análise de sensibilidade, sobretudo quando eliminada outra entrada de chuva acumulada, a P(72h). A quarta variável mais importante, a Qo(T) novamente apresentou índice de contribuição estável ao longo da simulação (entre 0,038 e 0,049).

Nesta combinação de variáveis de entrada, a P(r) e a P(120h) iniciaram com baixos índices de contribuição, muito próximos dos índices observados nas variáveis excluídas. Aliás, na primeira rodada, pode-se afirmar que somente a P(f) e a Qo(T) influenciaram na simulação. Isso fortalece a afirmação realizada anteriormente de que é arriscado eliminar muitas entradas de uma só vez sob pena de excluir uma amostra relevante.

Outro aspecto importante a ser destacado diz respeito à análise preliminar dos dados de entrada. Embora o uso de estatísticas seja útil na escolha das variáveis de entrada, o resultado da simulação hidrológica com RNAs e a influência das entradas na vazão Q(T) não são explicados unicamente por coeficientes como o de correlação. Muitas vezes, dados altamente correlacionados com o que se deseja modelar podem influenciar pouco na saída da RNA, geralmente em virtude da semelhança com outras entradas.

Na terceira combinação de dados de entrada, com uso de três postos fluviométricos (um a jusante e dois a montante), cinco variáveis de entrada se destacaram: P(f), Qo(T), P(r), Qo2(T-1), representando a vazão no posto do rio Potiribu no tempo T-1 e a Qo3(T-1), representando a vazão

no posto do rio Turcato no tempo T-1. A Figura 11 ilustra o comportamento do índice de contribuição dessas variáveis ao longo das rodadas de exclusão.

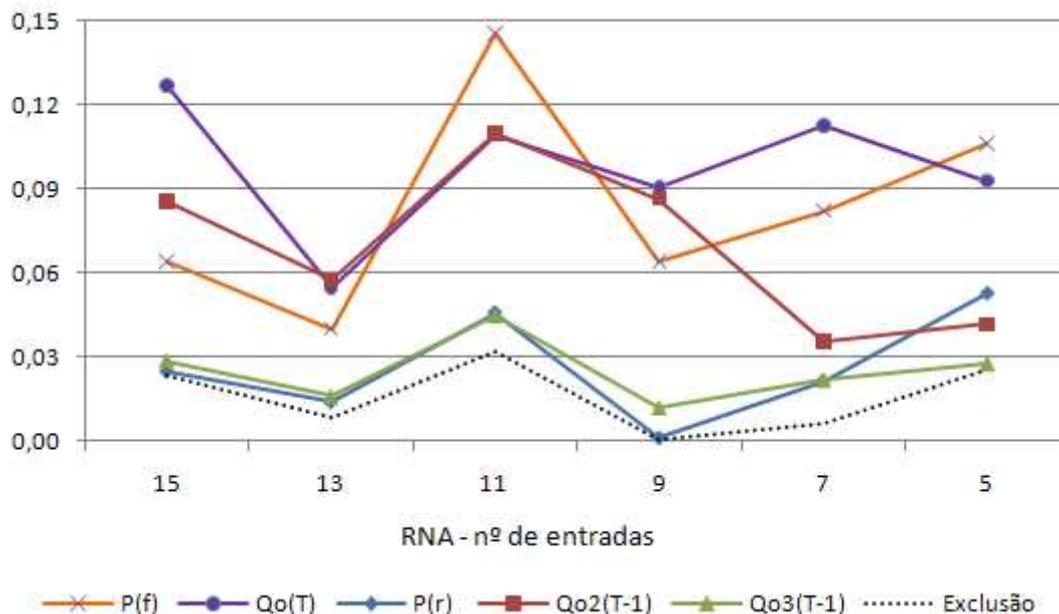


Figura 11. Índice de contribuição das variáveis mais importantes na combinação 3.

Nesta simulação, primeiramente, é importante destacar o “revezamento” entre três variáveis no ranking das mais importantes: a P(f), a Qo(T) e a Qo2(T-1). No final, a variável P(f) foi considerada a mais importante para a simulação da vazão, com índice de contribuição de 0,107, inferior aos observados nas outras duas combinações. Além disso, os limiares de exclusão ao longo das seis rodadas foram em geral bem superiores aos observados nas outras duas combinações de simulação. Isso indica que a maioria das variáveis de entrada apresentou maior sensibilidade na RNA.

Com base nessas primeiras afirmações, pode-se observar que nesta simulação, a RNA é menos dependente da variável P(f), havendo uma descentralização no modelo. Ou seja, mais variáveis ajudam a explicar a vazão do rio Ijuí no posto Santo Ângelo.

Esse comportamento diferenciado é explicado pela inserção de novas variáveis de entrada que compreendem informações que até então o modelo não dispunha, como as vazões no Potiribu e no Turcato. Mesmo que a inclusão de mais variáveis para explicar o comportamento da vazão não represente necessariamente uma melhora no desempenho da simulação hidrológica, pode-se afirmar que aumenta a credibilidade do modelo, uma vez que outros aspectos físicos da simulação do escoamento alimentam a RNA.

É importante também recordar que na última rodada de exclusão, houve uma significativa queda no desempenho do modelo, com decréscimo do R^2 de 0,938 para 0,854. Essa perda ocorreu

quando foram excluídas duas variáveis de chuva acumulada, a $P(72h)$ e a $P(120h)$. Pode-se afirmar que, embora não apresentem elevados índices de contribuição, estas entradas conferem estabilidade à simulação, isto é, quando excluídas reduzem a capacidade da RNA de explicar a vazão $Q(T)$.

Ressalta-se que as entradas de precipitação comumente utilizadas em simulações hidrológicas com redes neurais (chuva nos tempos T , $T-1$ e $T-2$, por exemplo) não apresentaram grande contribuição nas simulações realizadas neste trabalho. Em métodos de simulação hidrológica estritamente numéricos e sem base física, como nas RNAs, fica evidente a necessidade de extrair informações mais correlacionadas com a vazão que se deseja estimar, como a $P(f)$ e a $P(r)$. Essas variáveis potencializam a capacidade da RNA de simular o escoamento observado nos rios, sendo mais importantes que as variáveis de precipitação “brutas”.

4 - CONSIDERAÇÕES FINAIS

De modo geral, a metodologia de exclusão de duas variáveis por rodada até obter uma configuração de RNA mais compacta apresentou resultados muito satisfatórios, com melhora no coeficiente de Nash-Sutcliffe e redução constante no tempo de processamento computacional. Além disso, é importante destacar que a redução no número de neurônios na camada oculta na mesma proporção dos neurônios de entrada, não prejudicou o desempenho da RNA. Concluiu-se que a capacidade da RNA é potencializada ao eliminar variáveis desnecessárias, até um certo limite, no qual o desempenho pode decair drasticamente.

Foram testadas três combinações de variáveis de entrada, obtendo evolução no coeficiente R^2 de 0,9102 para 0,9472 na primeira combinação de simulação e de 0,9097 para 0,9488 na segunda combinação. Somente na terceira simulação que o coeficiente R^2 não apresentou evolução. Ainda assim, foi possível identificar uma configuração mínima de sete neurônios para obtenção de uma simulação com bom desempenho.

Na primeira combinação, apenas três variáveis de entrada apresentaram significativos índices de contribuição na simulação da vazão $Q(T)$: a chuva filtrada por um decaimento exponencial, $P(f)$, a vazão no posto Ponte Mística, $Q_0(T)$ e a chuva ponderada pelo coeficiente de correlação, $P(r)$. Na segunda combinação, além das três variáveis destacadas acima, a chuva acumulada dos últimos cinco dias, $P(120h)$, também influenciou significativamente nos resultados. Em ambas as combinações, a variável $P(f)$ foi a que mais contribuiu na simulação da vazão $Q(T)$.

Já na terceira combinação de dados de entrada, cinco variáveis de entrada se destacaram: $P(f)$, $Q_0(T)$, $P(r)$, $Q_02(T-1)$, representando a vazão no posto do rio Potiribu no tempo $T-1$ e a $Q_03(T-1)$, representando a vazão no posto do rio Turcato no tempo $T-1$. Nesta simulação, houve uma redução da dependência da RNA pela entrada $P(f)$.

Além da identificação das principais variáveis de entrada, pode-se observar que a exclusão de uma ou duas variáveis da RNA altera a contribuição das outras entradas no modelo. Logo, recomenda-se a exclusão de poucas variáveis por rodada para evitar erros grosseiros no processo de redução das dimensões da RNA, assim como no método apresentado nesse estudo.

Uma consideração importante que pode ser diagnosticada neste trabalho diz respeito à redundância de dados. Quando duas ou mais entradas são muito correlacionadas entre si, pode-se interpretar erroneamente a contribuição destas na saída da RNA. Ao eliminar uma das entradas, a outra pode ter sua contribuição potencializada na rodada seguinte, chegando ao verdadeiro índice de contribuição da variável.

Outra conclusão ressaltada no estudo revela que as entradas de precipitação geralmente utilizadas em simulações hidrológicas com RNAs (chuva nos tempos T, T-1 e T-2, por exemplo) não apresentaram grande contribuição nas simulações realizadas. Ficou evidente a necessidade de extrair informações mais correlacionadas com a vazão a ser estimada, como a P(f) e a P(r). Essas variáveis potencializaram a capacidade da RNA de simular o escoamento observado, sendo mais importantes que as variáveis de precipitação “brutas”.

Por fim, salienta-se que a simplificação da RNA ajuda a entender com maior clareza como a rede está funcionando e identificar quais variáveis são mais significativas para a composição do hidrograma simulado. A redução da complexidade da RNA permite que o pesquisador acompanhe mais facilmente como as alterações dos valores medidos de poucas variáveis influenciam na vazão, esclarecendo de certa forma os processos internos da rede.

BIBLIOGRAFIA

- BISHOP, C. M. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford. 1995.
- CHANG, F. & CHEN, Y. *A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction*. Journal of Hydrology, n. 245, 2001. pp. 153-164.
- DAWSON, C. W. & WILBY, R. *An artificial neural network approach to rainfall-runoff modeling*. Hydrological Sciences Journal, n.43(1), 1998. pp. 47-66.
- GARSON, G.D. *Interpreting neural-network connection weights*. Artificial Intell. Expert 6, 1991. pp. 47-51.
- GOH, T. H. *Semantic Extraction Using Neural Network Modelling and Sensitivity Analysis*. Proceedings of International Joint Conference on Neural Networks. Nagoya, Japão, 1993. pp. 1031-1034.

- GUHA, R.; STANTON, D. T. & JURIS, P. C. *Interpreting Computational Neural Network Quantitative Structure – Activity Relationship Models: A Detailed Interpretation of the Weights and Biases*. Journal of Chemical Information and Modeling, N° 45, 2005. pp. 1109-1121.
- HSU, K.; GUPTA, H. V. & SOROOSHIAN, S. *Artificial Neural Network Modeling of the rainfall-runoff process*. Water Resources Research, vol. 31, n. 10, 1995. pp. 2517-2530.
- JAIN, A. & KUMAR, A. M. *Hybrid neural network models for hydrologic time series forecasting*. Applied Soft Computing, n.7, 2007. pp. 585-592.
- JAIN, A.; SUDHEER, K. P. & SRINIVASULU, S. *Identification of physical processes inherent in artificial neural network rainfall runoff models*. Hydrological Processes, N. 18, 2004. pp. 571-581.
- MAK, B. & BLANNING, R. W. *An empirical measure of element contribution in neural networks*. IEEE Trans. Syst., Man Cybernetics, n. 28, 1998. pp. 561-564.
- MCCULLOCH, W. & PITTS, W. *A logical calculus of the ideas immanent in nervous activity*", *Bulletin of Mathematical Biophysics*. vol. 5, 1943.
- OH, S. H. & LEE, Y. *Sensitivity Analysis of Single Hidden-Layer Neural Networks with Threshold Functions*. IEEE Transactions on Neural Networks, Vol. 6, N° 4, 1995. pp. 1005-1007.
- ROSENBLATT, F. *The Perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review, Vol. 65, N. 6, 1958.
- RUMELHART, D. E.; HINTON, G. E. & WILLIAMS, R. J. *Learning representations by back-propagating errors*. Nature, vol. 323, 1986. pp. 533-536.
- SAJIKUMAR, N. & THANDAVESWARA, B. S. *A non-linear rainfall-runoff model using an artificial neural network*. Journal of Hydrology, n. 216, 1999. pp. 32-55.
- SHAMSELDIN, A. Y. *Application of a neural network technique to rainfall-runoff modelling*. Journal of Hydrology, n.199, 1997. pp. 272-294.
- SILVA, A. B. M. *Redes Neurais Artificiais, análise de sensibilidade e o comportamento de funções de comércio exterior do Brasil*. Tese de Doutorado – Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
- SILVA, V. S. V. *Estimativa de precipitação pontual em diferentes escalas para uso em modelo concentrado chuva-vazão*. Dissertação de Mestrado – Instituto de Pesquisas Hidráulicas da Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

SMITH, J. & ELI, R. N. *Neural-network models of rainfall-runoff process*. Journal of Water Resources Planning and Management, 1995. pp. 499-508.

SUDHEER, K. P.; GOSAIN, A. K. & RAMASASTRI, K. S. *A data-driven algorithm for constructing artificial neural network rainfall-runoff models*. Hydrological Processes, n. 16, 2002. pp. 1325-1330.

VALENÇA, M. J. S. & LUDERMIR, T. B. *Explicando a relação entre as variáveis de uma rede neural – iluminando a “caixa preta”*. XVII Simpósio Brasileiro de Recursos Hídricos – São Paulo, 2007.

WIDROW, B. & HOFF, M. E. *Adaptive switching circuits*. In: 1960 IRE WESCON Convention Record, New York: IRE Part 4, 1960. pp. 96-104.