



XV Encontro de Recursos Hídricos em Sergipe

APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA MODELAR O PROCESSO CHUVA-VAZÃO NA BACIA DO RIO PIAUITINGA (SE)

Luiz Antônio Muñoz Dantas¹; Alcigeimes Batista Celeste²

RESUMO: *Este trabalho apresenta a aplicação de redes neurais artificiais para modelar o processo chuva-vazão na bacia hidrográfica do rio Piauitinga, em Sergipe. Os dados de entrada utilizados na modelagem foram os do posto fluviométrico de Estância e do pluviométrico de Salgado, os quais foram divididos entre calibração e validação, selecionados de maneira randômica. A execução das simulações ocorreu através do software de aprendizado de máquina Weka, e foram testadas quatro formulações diferentes para a situação mensal. A eficiência das simulações realizadas foi verificada a partir do coeficiente de Nash-Sutcliffe, conforme o qual obtiveram-se resultados consideráveis. A formulação contendo como atributos chuva e vazão de até três períodos anteriores foi a que apresentou os melhores resultados, atingindo índices de eficiência classificados como muito bons. Modelagens alternativas foram realizadas para comparar o modelo usando apenas um posto de chuva com outra usando a chuva média na bacia, e notou-se que o modelo que utiliza apenas o posto de precipitação de Salgado apresentou resultados melhores ou iguais. Por fim, foram feitas novas simulações que utilizaram os períodos iniciais para calibração e os finais para validação, e percebeu-se que as simulações que classificaram os dados de forma randômica apresentaram resultados superiores.*

Palavras-chave: Aprendizado de máquina. Weka. Modelagem Guiada por Dados.

¹Estudante de Engenharia Civil, Departamento de Engenharia Civil, Universidade Federal de Sergipe, Avenida Marechal Rondon, s/n, Jardim Rosa Elze, São Cristóvão, SE, CEP: 49100-000, luiz_munoz@hotmail.com.

²Professor, Departamento de Engenharia Civil, Universidade Federal de Sergipe, Avenida Marechal Rondon, s/n, Jardim Rosa Elze, São Cristóvão, SE, CEP: 49100-000, geimes@yahoo.com..

INTRODUÇÃO

Com o intuito de auxiliar na gestão adequada dos diferentes usos da água, como abastecimento, irrigação, geração de energia e indústria, além da previsão de eventos extremos de cheias e períodos de estiagem, é empregado comumente a modelagem chuva-vazão, baseada no complexo processo de transformação da precipitação em vazão (MACHADO, 2005). Muitos desses modelos já elaborados são do tipo conceitual, entretanto modelos empíricos do tipo 'caixa-preta', ou guiados por dados (DDM: *data driven models*) começaram a aparecer gradativamente na literatura nos últimos anos (CELESTE; MENEZES; FONTES, 2014).

Para a aplicação dos modelos conceituais é necessário a compreensão dos processos hidrológicos e exige-se profundo conhecimento das características físicas da bacia hidrográfica. Por outro lado, para os DDMs não é preciso ter conhecimento dessas características (SOUSA; SOUSA, 2006).

Existem diversos DDMs que podem ser empregados para solucionar o problema da modelagem chuva-vazão. As redes neurais artificiais (RNAs) são um dos modelos mais populares (CELESTE; MENEZES; FONTES, 2014), e têm sido vastamente utilizadas para modelar sistemas não-lineares complexos com grande quantidade de dados, incluindo informações ruidosas (SOUSA; SOUSA, 2006).

Neste estudo, foi proposta a aplicação de RNAs para modelar o processo chuva-vazão na bacia hidrográfica do rio Piauitinga, em Sergipe, através do software de aprendizado de máquina Weka (<<https://www.cs.waikato.ac.nz/ml/weka/>>).

MATERIAL E MÉTODOS

O levantamento de dados de precipitação e vazão é uma etapa fundamental para modelagem chuva-vazão, pois constituem os atributos de entrada e os dados de saída observados.

Dessa forma, foram utilizados os dados disponíveis no portal HidroWeb, uma ferramenta da Agência Nacional de Águas e Saneamento Básico (ANA), que disponibiliza os dados coletados pelas estações hidrometeorológicas. Através do portal foi possível ter acesso às séries históricas de vazão do posto fluviométrico de Estância (50230000), e de chuva do pluviométrico de Salgado (1137001). Esses postos foram selecionados baseados em suas localizações.

Neste trabalho, para a realização da modelagem foi decidido utilizar uma abordagem mensal. Desse modo, foi utilizado o período de janeiro de 1966 até junho de 2006 (40 anos e 5 meses), pois não foram observadas falhas no posto fluviométrico de Estância e os dados consistidos de chuva importados do posto de Salgado datam até o ano de 2006. Contudo, no posto de Salgado foram verificadas falhas no intervalo de outubro de 1992 até março de 1993. Sendo assim, foi feita a substituição das informações faltosas pelos respectivos dados do posto pluviométrico de Estância (1137017).

Para a realização da modelagem chuva-vazão, os DDMs foram submetidos a quatro tipos diferentes de formulações, da maneira:

$$h(x) = h(x_1, x_2, \dots, x_n) \quad (1)$$

onde, $h(x)$ é a hipótese para vazão $Q(t)$ do mês t em função dos atributos x_1, x_2, \dots, x_n . Esses atributos são constituídos por precipitações do mês atual ($P(t)$) ou de períodos passados ($P(t-1), P(t-2), \dots, P(t-n)$) e/ou vazões de períodos passados ($Q(t-1), Q(t-2), \dots, Q(t-n)$). As formulações foram escolhidas de maneira empírica, e podem ser observadas no Quadro 1.

Quadro 1. Formulações investigadas

1	$Q(t) = h(Q(t - 1), Q(t - 2), Q(t - 3), Q(t - 4), Q(t - 5))$
2	$Q(t) = h(P(t), P(t - 1), P(t - 2), P(t - 3), P(t - 4), P(t - 5))$
3	$Q(t) = h(P(t), P(t - 1), P(t - 2), P(t - 3), Q(t - 1), Q(t - 2), Q(t - 3))$
4	$Q(t) = (P(t - 1), P(t - 2), P(t - 3), P(t - 4), Q(t - 1), Q(t - 2), Q(t - 3), Q(t - 4))$

Para o desenvolvimento da modelagem, os conjuntos de dados foram divididos entre calibração e validação. Os dados de calibração são utilizados para treinar o modelo, enquanto que os de validação servem para validar o modelo. Neste trabalho, foram utilizados 70% dos dados para calibração e 30% para validação, selecionados de maneira randômica, i.e., tanto dados de períodos passados quanto de presentes podem fazer parte dos dois conjuntos.

Para a utilização do software Weka, é necessário realizar a exportação dos dados para o programa. Dessa maneira, é preciso converter os dados da planilha para o formato ARFF, que é o padrão utilizado pelo Weka.

Esse formato consiste em uma lista de instâncias, e os valores de atributo para cada instância são separados por vírgula. Além disso, o arquivo ARFF deve conter indicações (marcações) para que o programa realize a interpretação dos dados de maneira apropriada (FRANK; HALL; WITTEN, 2016). Para configurar o arquivo utiliza-se um editor de texto, e posteriormente basta salvá-lo no formato ARFF.

Desse modo, para realização da modelagem foram criados dois tipos de arquivos ARFF, um denominado *trainData*, constituído pelo conjunto de dados de calibração, e outro chamado de *allData*, que abrange todo o conjunto de dados (calibração e validação).

No Weka é possível selecionar qual DDM será empregado, além disso para as RNAs o programa é capaz de gerar automaticamente uma configuração para a(s) camada(s) oculta(s), baseado na formulação utilizada. Contudo, o Weka também permite que o usuário configure sua própria topologia de RNA, pois possibilita a edição do número de neurônios e a quantidade de camadas ocultas.

O Weka sugeriu dois tipos de configurações diferentes para as camadas intermediárias das formulações investigadas (Quadro 1). Para as formulações 1 e 2, o programa utilizou três neurônios artificiais arranjados em uma única camada oculta. Já nas formulações 3 e 4, usou quatro neurônios, também dispostos em uma única camada.

Foram testadas diversas configurações diferentes para RNA. Decidiu-se que para as formulações 1 e 2 seriam utilizadas as configurações do Quadro 2, e para as formulações 3 e 4 as do Quadro 3. Nesses esquemas, os valores numéricos significam a quantidade de neurônios que pertencem a uma camada oculta, e a adição de novas camadas é representada pela utilização de vírgulas.

Quadro 2. Configurações das camadas ocultas na RNA para as formulações 1 e 2

RNA-3	RNA-4	RNA-6
RNA-3,1	RNA-4,2	RNA-6,3
RNA-3,3	RNA-4,4	RNA-6,6
RNA-3,1,1	RNA-4,2,1	RNA-6,3,3
RNA-3,3,1	RNA-4,4,2	RNA-6,6,3
RNA-3,3,3	RNA-4,4,4	RNA-6,6,6

Quadro 3. Configurações das camadas ocultas na RNA para as formulações 2 e 3

RNA-4	RNA-6	RNA-8
RNA-4,2	RNA-6,3	RNA-8,4
RNA-4,4	RNA-6,6	RNA-8,8
RNA-4,2,1	RNA-6,3,3	RNA-8,4,4
RNA-4,4,2	RNA-6,6,3	RNA-8,8,4
RNA-4,4,4	RNA-6,6,6	RNA-8,8,8

Após selecionar o DDM a ser aplicado, deve-se utilizar o arquivo que contém os dados de calibração (*trainData*) como conjunto de treinamento, com isso o programa treinará o modelo. Em seguida, é inserido o conjunto de dados *allData*, contendo dados de calibração e validação. Dessa forma, o Weka é submetido a um novo conjunto de dados (dados de validação), que não foram utilizados durante o processo de treinamento. Assim, pode-se testar a modelagem realizada pelo software. A partir do arquivo *allData*, o programa fornece todos os resultados obtidos ao fim da simulação, tanto para os dados de calibração como os de validação.

Para avaliar o desempenho da simulação gerada pelo Weka, foi utilizada a eficiência de Nash-Sutcliffe:

$$E = \frac{\sum_{i=1}^m (y_i - Y_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (2)$$

em que y_i é o valor observado, Y_i é o modelado, e \bar{y} é a média de y_i ($i = 1, \dots, m$). Neste trabalho, os valores dos coeficientes de Nash-Sutcliffe foram calculados com o auxílio de uma planilha eletrônica, e foram classificados de acordo com Baltokoski et al. (2010, apud Lima, 2022) da seguinte forma:

- Muito Bom: $E > 0,75$;
- Bom: $0,65 < E \leq 0,75$;
- Satisfatório: $0,50 < E \leq 0,65$;
- Insatisfatório: $E \leq 0,50$.

RESULTADOS E DISCUSSÃO

Nas Tabelas 1 e 2 é possível observar todos os coeficientes de eficiência que foram encontrados nas simulações para os dados de validação, calibração e completo.

Tabela 1. Eficiências de Nash-Sutcliffe nas Formulações 1 e 2.

Configuração	Formulação 1			Formulação 2		
	Completo	Calibração	Validação	Completo	Calibração	Validação
RNA-3	-0,04	-0,05	-0,03	0,77	0,81	0,61
RNA-3,1	0,00	0,02	-0,03	0,74	0,78	0,59
RNA-3,3	-0,01	0,01	-0,04	0,74	0,78	0,60
RNA-3,1,1	-0,48	-0,52	-0,41	0,73	0,76	0,61



RNA-3,3,1	-0,48	-0,52	-0,41	-0,05	-0,05	-0,06
RNA-3,3,3	-0,48	-0,52	-0,41	0,75	0,78	0,62
RNA-4	-0,05	-0,06	-0,03	0,76	0,81	0,56
RNA-4,2	-0,01	0,01	-0,04	0,76	0,79	0,63
RNA-4,4	-0,01	0,00	-0,04	0,76	0,79	0,63
RNA-4,2,1	-0,48	-0,52	-0,41	-0,05	-0,05	-0,06
RNA-4,4,2	-0,48	-0,52	-0,41	0,76	0,78	0,65
RNA-4,4,4	0,00	0,02	-0,03	0,76	0,78	0,66
RNA-6	-0,05	-0,06	-0,04	0,76	0,81	0,58
RNA-6,3	0,00	0,01	-0,03	0,76	0,79	0,64
RNA-6,6	-0,01	0,01	-0,04	0,77	0,80	0,64
RNA-6,3,3	-0,48	-0,52	-0,41	0,76	0,79	0,66
RNA-6,6,3	0,05	0,07	0,02	0,76	0,78	0,66
RNA-6,6,6	0,02	0,04	-0,01	0,75	0,78	0,64

Tabela 2. Eficiências de Nash-Sutcliffe nas Formulações 3 e 4.

Configuração	Formulação 3			Formulação 4		
	Completo	Calibração	Validação	Completo	Calibração	Validação
RNA-4	0,82	0,83	0,77	-0,49	0,52	-4,21
RNA-4,2	0,81	0,82	0,78	0,38	0,55	-0,22
RNA-4,4	0,81	0,82	0,81	0,34	0,45	-0,06
RNA-4,2,1	0,00	0,00	-0,01	-0,05	-0,05	-0,07
RNA-4,4,2	0,82	0,82	0,83	0,26	0,27	0,24
RNA-4,4,4	0,82	0,82	0,81	0,31	0,31	0,29
RNA-6	0,72	0,76	0,55	-0,48	0,52	-4,15
RNA-6,3	0,82	0,82	0,82	0,38	0,49	-0,03
RNA-6,6	0,82	0,82	0,81	0,36	0,51	-0,21
RNA-6,3,3	0,82	0,82	0,83	0,32	0,33	0,29
RNA-6,6,3	0,79	0,78	0,81	0,29	0,28	0,29
RNA-6,6,6	0,82	0,82	0,81	0,27	0,27	0,27
RNA-8	0,72	0,76	0,50	-0,44	0,53	-4,00

RNA–8,4	0,81	0,81	0,79	0,32	0,35	0,22
RNA–8,8	0,81	0,81	0,80	0,31	0,34	0,21
RNA–8,4,4	0,82	0,82	0,82	0,32	0,35	0,23
RNA–8,8,4	0,82	0,82	0,83	0,29	0,29	0,29
RNA–8,8,8	0,82	0,82	0,81	0,29	0,30	0,26

É interessante destacar os valores de Nash-Sutcliffe obtidos a partir dos dados de validação, pois estes não são utilizados durante o processo de treinamento do modelo. Sendo assim, contribuem para testar e ratificar a eficiência da modelagem desenvolvida. As configurações que apresentaram o melhor valor de coeficiente de Nash-Sutcliffe para os dados de validação, com a menor quantidade de neurônios, são apresentadas abaixo:

- Formulação 1: RNA–6,6,3 ($E = 0,02$);
- Formulação 2: RNA–4,4,4 e RNA–6,3,3 ($E = 0,66$);
- Formulação 3: RNA–4,4,2 ($E = 0,83$);
- Formulação 4: RNA–4,4,4 e RNA–6,3,3 ($E = 0,29$).

A formulação 3 foi a que mais se destacou dentre todas as simulações realizadas, com o resultado de 83% para o índice de eficiência para validação, sendo considerado muito bom.

A formulação 1, que somente possui atributos de vazão, foi a que retratou os piores resultados, com valores considerados insatisfatórios ($E = 0,02$). Já a formulação 2, composta apenas por atributos de precipitação, apresentou resultados classificados como bom ($E = 0,66$). A formulação 4 não apresentou nenhum resultado expressivo, mesmo sendo composta por atributos de vazão e precipitação, e teve somente resultados superiores sobre formulação 1. Ademais, nota-se que as formulações que apresentam o atributo de precipitação atual ($P(t)$) na sua composição (formulações 2 e 3) foram as que apresentaram os melhores desempenho na modelagem.

Celeste, Menezes e Fontes (2014) utilizaram o modelo conceitual *Tank Model* e a modelagem empírica baseada em RNAs para simular o processo chuva-vazão na bacia hidrográfica do rio Japarutuba, em Sergipe. Foi observado que, assim como neste trabalho, as RNAs que não possuem a precipitação do mês atual como neurônio de entrada apresentaram menores eficiências, e que os modelos que utilizaram somente vazões como atributos de entrada apresentaram os piores desempenho entre as RNAs. Os autores verificaram que a vazão modelada tem uma alta correlação com a chuva atual, e que vazões passadas tem pouca correlação com a vazão atual.

Com o intuito de comparar a modelagem elaborada, testou-se novas simulações com diferentes dados de precipitação, mas ainda utilizando o posto fluviométrico de Estância nos dados de vazão. Para isso, foi aplicada a formulação 3, visto que esta foi a que apresentou os melhores resultados obtidos. Assim, para a análise utilizou-se a configuração automática gerada pelo Weka, e a melhor configuração obtida na situação mensal. Além disso, foi aplicado o mesmo intervalo de tempo para as novas modelagens.

Na primeira simulação, para a obtenção dos dados de chuva é utilizado o posto pluviométrico de Estância, e foram obtidos $E = 0,47$ para RNA–4 e $E = 0,62$ para RNA–4,4,2. Na segunda simulação, para a obtenção dos dados de chuva, foi empregado a média ponderada aplicada no estudo de Lima (2022), em que o autor utilizou o método dos polígonos de Thiessen e encontrou que o posto pluviométrico de Salgado (01137001) possui influência de 77,54% da área total da bacia do rio Piauitinga, enquanto que o posto de Estância (01137017) tem influência de 22,46%. Dessa maneira, esses pesos foram aplicados nos dados de precipitação e foi encontrado

$E = 0,75$ para a configuração RNA-4 e $E = 0,83$ para a configuração RNA-4,4,2.

Dessa maneira, pode-se notar que neste estudo a utilização da modelagem utilizando o posto pluviométrico de Salgado e o fluviométrico de Estância, é um opção viável que apresentou resultados consideráveis (Figura 1), e sua modelagem é mais simplificada que a simulação que utiliza a chuva média na bacia. A modelagem obteve resultados superiores às demais para as configurações propostas, com exceção da configuração RNA-4,4,2 na segunda simulação, em que foram encontrados valores idênticos ($E = 0,83$).

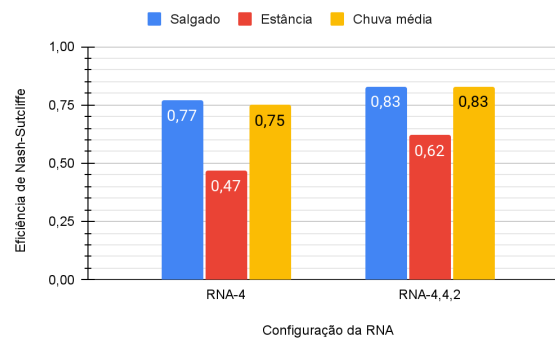


Figura 1. Comparação entre as modelagens.

Para observar neste estudo a influência da escolha randômica da classificação dos dados com 70% para calibração e 30% validação, realizou-se novas simulações em que os primeiros 70% dos dados foram classificados como calibração e os 30% finais como validação. Para isso, foi aplicada RNA para a formulação 3, e novamente utilizou-se a configuração automática gerada pelo Weka e a melhor configuração obtida na abordagem mensal.

Dessa forma, para a simulação que utiliza 70% dos dados iniciais para calibração e 30% dos finais para a validação, obteve-se índices de eficiência para validação de $E = 0,70$ e $E = 0,67$ para RNA-4 e RNA-4,4,2, respectivamente. Por outro lado, na simulação que utiliza 70% dos dados de calibração e 30% para validação selecionados de maneira randômica, foram encontrados $E = 0,77$ para RNA-4 e $E = 0,83$ para RNA-4,4,2.

Na Figura 2, fica evidente que os resultados obtidos utilizando a classificação randômica foram bastante superiores para as configurações analisadas, quando comparados com a classificação inicial/final.

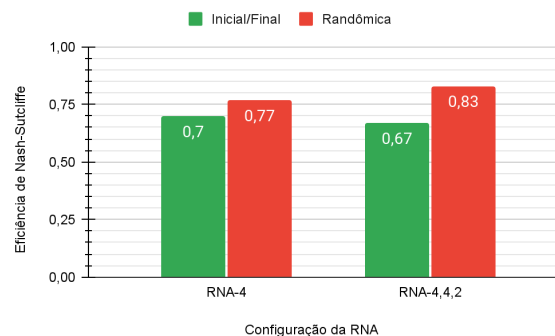


Figura 2. Influência da seleção dos dados de calibração e validação.

CONCLUSÕES

O acesso de maneira adequada aos recursos hídricos é imprescindível para o desenvolvimento das regiões, entretanto o gerenciamento eficiente desses recursos pode ser uma



tarefa desafiadora. A aplicação da modelagem chuva-vazão pode ser utilizada para auxiliar na gestão adequada do uso da água.

Neste trabalho, observou-se que é possível realizar a aplicação da modelagem chuva-vazão utilizando modelos baseados em redes neurais artificiais na bacia hidrográfica do rio Piauitinga, em Sergipe, através do software Weka.

Por meio do índice de eficiência para os dados de validação, foi possível verificar o desempenho das simulações para cada configuração empregada neste estudo. Foram utilizadas quatro tipos de formulações diferentes e notou-se que a formulação 3, constituída por atributos de chuva e vazão de até três períodos anteriores, foi a que mais se sobressaiu.

Com finalidade de comparar os resultados das simulações geradas a partir dos dados de precipitação provenientes do posto pluviométrico de Salgado, realizaram-se novas modelagens com outros dados de chuva. Com isso, observou-se que a modelagem com apenas o posto pluviométrico de Salgado apresentou resultados superiores ou iguais às outras simulações. Constatou-se, assim, que esta é uma atrativa opção, devido aos seus bons resultados e por apresentar uma modelagem mais simples quando comparada com a da chuva média.

Para investigar a influência da escolha randômica da classificação dos dados de calibração e validação neste estudo, foram feitas novas simulações utilizando os primeiros dados para calibração e os últimos para validação. Dessa forma, percebeu-se que a aplicação da classificação dos dados de forma randômica para este trabalho foi positiva, em que foram apresentados resultados superiores em relação a classificação inicial/final.

Todas as simulações realizadas neste trabalho foram avaliadas a partir do coeficiente de Nash-Sutcliffe, de modo que foi necessário testar de maneira individual cada configuração empregada nos modelos gerados pelo Weka, utilizando planilhas eletrônicas. Consequentemente, uma das limitações dessa pesquisa foi a quantidade de simulações que o usuário pode efetuar.

Para novas pesquisas, é recomendável a aplicação de RNAs acopladas a um algoritmo genético para realização do treinamento, comparando os resultados com os obtidos neste trabalho, e observando se ocorre melhora no desempenho da modelagem.

Também é recomendado o emprego de outros DDMs, verificando se é possível atingir melhores resultados para o coeficiente de Nash-Sutcliffe.

REFERÊNCIAS

CELESTE, A. B.; MENEZES, T.; FONTES, V. Modelagem do processo chuva-vazão pelo tank model e por redes neurais artificiais na bacia do rio Japarutuba, Sergipe. *Scientia Plena*, v. 10, n. 7, 2014.

FRANK, E.; HALL, M. A.; WITTEN, I. H. *The WEKA Workbench*. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques". Fourth edition. Burlington: Morgan Kaufmann, 2016.

LIMA, R. G. S. Modelo chuva-vazão baseado na aplicação do método dos resíduos ponderados ao balanço hídrico. São Cristóvão – SE, 2022. 33 f. Trabalho de Conclusão de Curso (TCC) – Departamento de Engenharia Civil. Universidade Federal de Sergipe

MACHADO, F. W. Modelagem chuva-vazão mensal utilizando redes neurais artificiais. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, Universidade Federal do Paraná, 2005.

SOUSA, F. A. S.; SOUSA, W. S. Rede neural aplicada ao processo chuva-vazão da Bacia Hidrográfica de Caraúbas - PB. In: Congresso Brasileiro de Meteorologia, 14, 2006. Anais... Florianópolis. SBMET. CD Rom