



Ambiente & Água - An Interdisciplinary Journal
of Applied Science

ISSN: 1980-993X

ambi-agua@agro.unitau.br

Universidade de Taubaté
Brasil

Santana Santos, José Rafael; Batista Celeste, Alcigeimes

Avaliação de estratégias de modelagem guiada por dados para previsão de vazão em rio sergipano

Ambiente & Água - An Interdisciplinary Journal of Applied Science, vol. 9, núm. 3, julio-septiembre,

2014, pp. 488-498

Universidade de Taubaté

Taubaté, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=92831698011>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal

Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto



Avaliação de estratégias de modelagem guiada por dados para previsão de vazão em rio sergipano

doi: 10.4136/ambi-agua.1356

Received: 04 Apr. 2014; Accepted: 24 May 2014

José Rafael Santana Santos; Alcigeimes Batista Celeste*

Universidade Federal de Sergipe (UFS), São Cristóvão, SE, Brasil

Departamento de Engenharia Civil

*Autor correspondente: e-mail: geimes@yahoo.com,
jrafassantos@yahoo.com.br

RESUMO

Dentre as várias técnicas de previsão de vazão, os modelos guiados por dados (DDMs: *data-driven models*) estão sendo muito utilizados. Estes se baseiam num banco de dados formado pelos registros históricos das variáveis de entrada (precipitação e vazão) e saída (vazão) para realizar a previsão. Redes neurais artificiais (ANNs: *artificial neural networks*) são os tipos de DDMs mais comuns e se mostram normalmente mais precisas do que outros modelos empíricos, mas possuem a desvantagem de não serem suficientemente transparentes. Um dos métodos de aprendizado de máquina que não possui esse problema é o aprendizado baseado em instâncias (IBL: *instance-based learning*). O modelo *k-nearest neighbor* (KNN) é um exemplo de IBL. Neste trabalho, variações do modelo KNN são utilizadas e propostas a fim de realizar previsão de vazão em rio do estado de Sergipe. Os resultados são comparados aos de simulações feitas com o uso de redes neurais artificiais e indicam superioridade das ANNs, mas também previsões satisfatórias com o KNN.

Palavras-chave: previsão hidrológica, inteligência artificial, mineração de dados.

Assessment of data-driven modeling strategies for streamflow forecasting in a river of Sergipe

ABSTRACT

Among several streamflow forecasting techniques, data-driven models (DDMs) are widely used. They employ a database formed by historical input (precipitation and streamflow) and output (streamflow) variables to perform the prediction. Artificial neural networks (ANNs) are the most common types of DDMs, and are typically more accurate than other empirical models, but have the disadvantage of not being sufficiently transparent. One of the machine-learning methods that do not have this problem is the so-called “instance-based learning” (IBL). The KNN algorithm is an example of IBL. This research applies and proposes variations of the KNN model in order to forecast streamflows in a river of the state of Sergipe, Brazil. The results are compared to simulations carried out by the use of artificial neural networks and indicate the superiority of the ANNs, but also show satisfactory forecasts of the KNN.

Keywords: hydrological forecasting, artificial intelligence, data mining.

1. INTRODUÇÃO

O conhecimento prévio de informações hidrológicas, sobretudo de vazões, em diferentes escalas de tempo e espaço é extremamente necessário para a tomada de decisão na operação de reservatórios e de hidrelétricas como também no planejamento e gerenciamento de vários outros sistemas de recursos hídricos.

Modelos de previsão de vazão do tipo conceitual levam em consideração a descrição detalhada dos processos físicos e precisam usualmente de informações sobre parâmetros hidrológicos, alguns dos quais não podem ser medidos. Por causa desse tipo de restrição, modelos empíricos, ou guiados por dados (DDMs: *data-driven models*), vêm se tornando cada vez mais populares (Solomatine et al., 2008). Dentre os vários tipos de DDMs, redes neurais artificiais (ANN: *artificial neural networks*) são os mais populares (Cruz et al., 2010; Gomes et al., 2010; Farias et al., 2007; Valença, 2005; Galvão e Valença, 1999).

Modelos de predição numérica baseados em ANNs e outros que reconstroem dependências não lineares complexas são tipicamente mais precisos do que alguns modelos empíricos mas, como observado por Solomatine et al. (2008), sofrem por ficarem encapsulados em códigos de software e, dessa forma, não serem transparentes o suficiente, o que é um problema para sua aceitação pelo usuário final uma vez que necessitam, também, de alguma experiência do usuário para definição da arquitetura da rede, do número de camadas e de neurônios, e do algoritmo de otimização para calibração dos parâmetros.

Uma das técnicas de aprendizado de máquina que tem potencial para resolver a questão da não transparência é o chamado aprendizado baseado em instâncias (IBL: *instance-based learning*), que simplesmente armazena todos os exemplos históricos (instâncias) na memória (processo conhecido como *easy learning*) sem fazer qualquer tipo de generalização dos dados fornecidos (*eager learning*, usado pelas ANNs) e sem necessitar de uma fase de calibração de parâmetros. Todo o trabalho do algoritmo é feito durante a previsão, utilizando-se alguma métrica para encontrar os exemplos passados mais similares ao da situação atual de modo a prever o que acontecerá no futuro com base no que ocorreu no passado.

Exemplos de modelos baseados em IBL são o algoritmo do vizinho mais próximo (NN: *nearest neighbor*) e sua variante *k-nearest neighbor* (KNN). O modelo KNN classifica um novo exemplo calculando a média entre os *k* exemplos mais próximos a ele. O NN unicamente escolhe a instância mais próxima.

Na literatura de recursos hídricos existem algumas aplicações de modelos IBL, com a grande maioria das pesquisas utilizando-se do modelo KNN (Toth et al., 2000; Shamseldin e O'Connor, 1996; Galeati, 1990; Karlsson e Yakowitz, 1987). No campo da previsão de vazão, o vetor de exemplos de treinamento (atributos) é geralmente composto pelos valores de chuva e vazão de períodos anteriores.

Este trabalho teve como objetivo implementar estratégias de aprendizado baseado em instâncias, especificamente o algoritmo KNN, e aplicá-las na previsão de vazão em escala mensal de um curso d'água localizado em uma bacia hidrográfica do estado de Sergipe, com o intuito de dar subsídios para o processo de planejamento e gerenciamento dos recursos hídricos da região. Além da metodologia tradicional do KNN, foi proposta e aplicada outra variação deste modelo bem como a técnica das redes neurais artificiais.

2. MATERIAL E MÉTODOS

2.1. Procedimento de previsão

De maneira bem simples, o procedimento de previsão de vazão visa estimar o valor da vazão Q no período $t + H$, onde t é o mês atual e H é o horizonte de previsão, em função de valores de chuva (P) e deflúvio em períodos passados ($t - 1$, $t - 2$, etc.). Seja $y = Q(t + H)$ a

vazão a ser prevista e $\mathbf{X} = [P(t-1), P(t-2), \dots, P(t-\tau_p), Q(t-1), Q(t-2), \dots, Q(t-\tau_q)]$ o chamado vetor de atributos de entrada, i.e., os valores passados de chuva e vazão usados para prever o deflúvio futuro. O objetivo é encontrar uma função f que mapeie a entrada \mathbf{X} com a saída y da forma $y = f(\mathbf{X})$ de modo que:

$$Q(t+H) = f \left\{ P(t-1), P(t-2), \dots, P(t-\tau_p), Q(t-1), Q(t-2), \dots, Q(t-\tau_q) \right\} \quad (1)$$

A função f pode ser qualquer estratégia de aprendizado de máquina como uma ANN ou um modelo KNN, utilizados neste trabalho. Esta pesquisa usou $H = 0$, i.e., a vazão do mês atual foi prevista com base em chuvas e vazões passadas.

Cada elemento x_j do vetor $\mathbf{X} = [x_1, \dots, x_n]$ é um atributo. As séries históricas de chuva e vazão permitem formar um conjunto de treinamento com vários vetores de atributos (Quadro 1).

Quadro 1. Vetores de atributos para treinamento.

Vetores de Entrada	Atributos				Valores de Saída
	A_1	A_2	...	A_n	
\mathbf{X}_1	$x_{1,1}$	$x_{1,2}$...	$x_{1,n}$	y_1
\mathbf{X}_2	$x_{2,1}$	$x_{2,2}$...	$x_{2,n}$	y_2
\vdots			\ddots		\vdots
\mathbf{X}_m	$x_{m,1}$	$x_{m,2}$...	$x_{m,n}$	y_m

Nota: Os valores de y_i ($i = 1, 2, \dots, m$) são as vazões para um horizonte de H meses correspondentes aos vetores de atributos \mathbf{X}_i obtidos com base nos dados históricos.

As defasagens (*lags*) limites τ_p e τ_q determinam a quantidade de valores passados de chuva e vazão a serem utilizados na previsão futura. Podem ser escolhidas a partir da análise hidrológica da bacia hidrográfica e de estudos de correlação entre entradas e saídas.

Nesta pesquisa, foram utilizados o coeficiente de correlação de Pearson e a chamada informação mútua média para a determinação dos *lags* máximos. O coeficiente de correlação de Pearson indica a magnitude e a direção do relacionamento linear entre duas variáveis. Seus valores situam-se dentro do intervalo $[-1, 1]$, com -1 indicando correlação linear inversa perfeita e $+1$ indicando correlação linear direta perfeita. A informação mútua média (AMI: *average mutual information*), por sua vez, mede a correlação não linear entre duas variáveis. Seu valor ótimo varia de acordo com os dados. Neste estudo, este valor foi o encontrado correlacionando a série fluviométrica com ela mesma para *lag* 0 (i.e., sem defasagem de tempo).

2.2. Definição da área de estudo e coleta de dados

O estudo de caso para realização deste trabalho foi a bacia hidrográfica do rio Japaratuba, localizada no estado de Sergipe (Figura 1). Essa bacia foi escolhida pelo fato de existirem sequências razoavelmente longas de dados pluviométricos e fluviométricos obtidos através do portal HidroWeb (hidroweb.ana.gov.br) da Agência Nacional de Águas (ANA).

A bacia hidrográfica do rio Japaratuba apresenta um regime hidrológico homogêneo, situando-se entre as latitudes $10^\circ 14' 00''$ e $10^\circ 48' 00''$ S, e longitudes $37^\circ 18' 00''$ e $36^\circ 48' 00''$ W, drenando uma área de 1.664,61 km². O curso d'água principal é o rio Japaratuba com uma extensão de 113,21 km, com nascente no município de Feira Nova, próxima ao município de

Graccho Cardoso, desembocando no oceano Atlântico entre os municípios de Pirambu e Barra dos Coqueiros. A rede de drenagem tem como principais afluentes pela margem direita os rios Siriri, Riachão e Lagarto e, pela margem esquerda, o Japarutuba Mirim, Cajueiro e Calomba. O rio Japarutuba corta o estado de Sergipe no sentido Oeste-Leste (SEMARH-SE, 2010).

Foi adotado um período de dez anos para o estudo, sendo oito anos destinados para a fase de calibração da ANN e aprendizado do KNN e dois anos para teste, i.e., uso de ambos os modelos para prever vazões já conhecidas durante estes dois anos a fim de comparar os valores observados com os previstos pelos modelos, estratégia conhecida como *hindcasting* ou previsão inversa. O período de estudo escolhido foi de janeiro de 1969 a dezembro de 1978. O critério para a escolha desse período foi adotar uma faixa que compreendesse o maior número possível de séries históricas de chuva que tivessem dados no mesmo intervalo de tempo das séries históricas de vazão. Por este motivo, algumas estações existentes na bacia foram descartadas do estudo. A Tabela 1 mostra as estações efetivamente empregadas. Portanto, a estação fluviométrica para a qual a modelagem de previsão foi feita é a estação de Japarutuba (código ANA: 50040000).

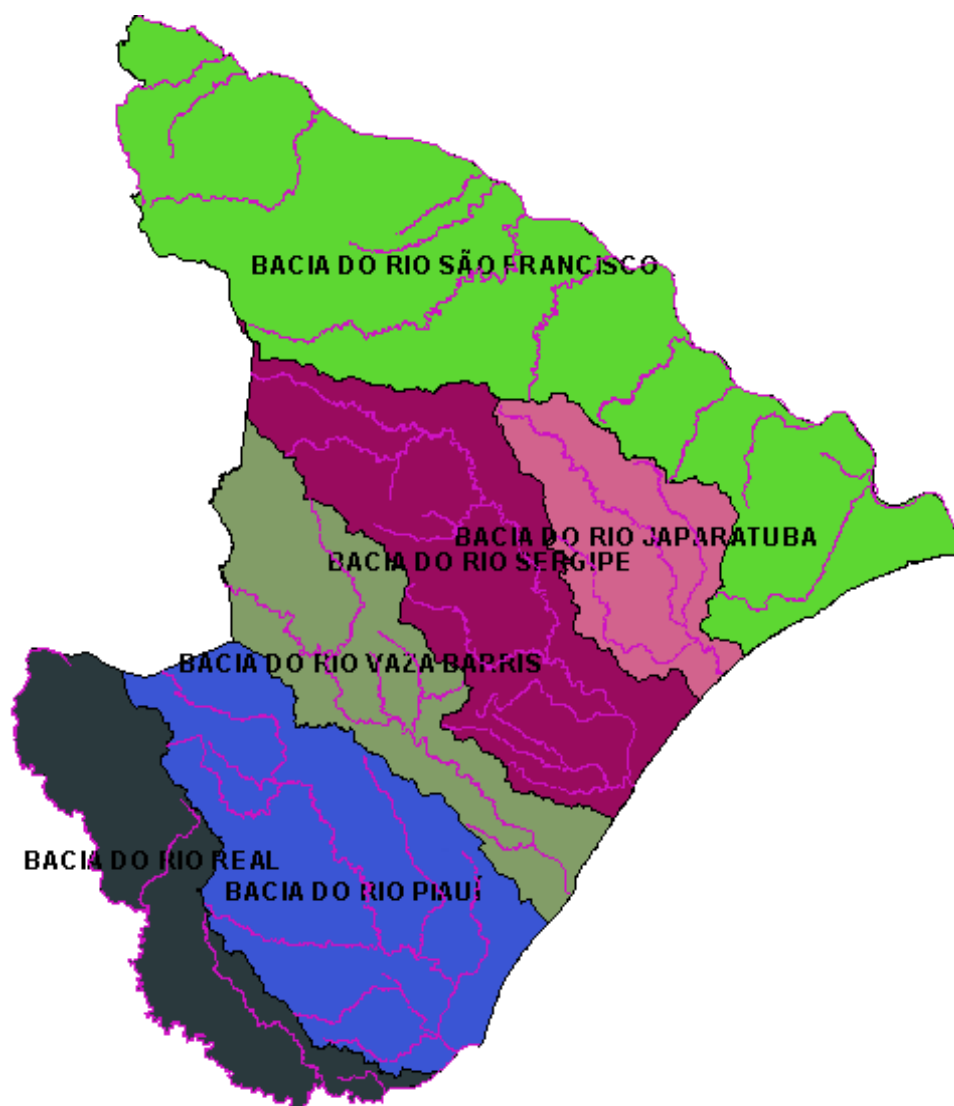


Figura 1. Bacias hidrográficas do estado de Sergipe.

Fonte: Atlas Digital Sobre Recursos Hídricos de Sergipe (SEMARH-SE, 2012).

Tabela 1. Estações fluviométricas e pluviométricas utilizadas.

Nome da Estação	Tipo	Código (ANA)
Japaratuba	Fluv.	50040000
Japaratuba	Pluv.	01036014
Japaratuba	Pluv.	01036016
Usina Outeirinhos	Pluv.	01036020
Curral do Meio	Pluv.	01036033
Capela	Pluv.	01037009
Cumbe	Pluv.	01037013
Siriri	Pluv.	01037047
Nossa Senhora das Dores	Pluv.	01037060

2.3. Modelo ANN

Redes neurais artificiais são dispositivos computacionais capazes de estabelecer relações não lineares entre dados de entrada e saída de um sistema. Esse mapeamento é conduzido por elementos que funcionam como neurônios artificiais, pois recebem uma série de informações, processam-nas e produzem uma saída, assim como as células do tecido nervoso humano. Existem muitos tipos de ANNs e, neste estudo, foi utilizada uma do conhecido tipo *feedforward* com três camadas: entrada, intermediária (ou oculta) e saída.

Cada neurônio da camada oculta coleta os valores de entrada ponderados para formar o valor de saída ou campo local induzido. Cada campo local induzido é submetido a uma função de ativação para que se torne o valor de entrada do neurônio da camada de saída. O neurônio da camada de saída possui o mesmo esquema dos neurônios da camada oculta e o seu campo local induzido, depois de submetido à função de ativação, torna-se a saída final da ANN. O número de neurônios da camada de entrada depende da quantidade de atributos de períodos passados (chuva, vazão, etc.) usados para prever a vazão futura. O número de neurônios da camada oculta é determinado fazendo simulações e verificando a melhor combinação de resultados. Por último, a camada de saída tem um único neurônio representando a vazão a ser prevista. Na rede *feedforward*, os neurônios de uma mesma camada não são conectados entre si.

Neste trabalho foi utilizada a função sigmóide tangencial como função de ativação para a camada intermediária, e para a camada de saída foi utilizada uma função linear. A calibração da rede foi conduzida pelo algoritmo de otimização Levenberg-Marquadt. Dos oito anos de dados usados na calibração, seis foram empregados no treinamento (que ajusta os parâmetros de acordo com o erro entre valores observados e simulados) e os outros dois na validação (que mede a generalização da rede e interrompe o treinamento quando a generalização parar de melhorar). No processo de otimização para ajuste dos parâmetros da rede, foram realizados vários treinamentos com diversas combinações de neurônios a fim de minimizar o erro entre valores observados e simulados. Finalmente, a melhor performance obtida foi com o uso de 50 neurônios na camada central.

2.4. Modelo KNN

De acordo com o modelo *k-nearest neighbor*, dado um novo vetor de atributos \mathbf{X}_q não incluído no conjunto de treinamento, também chamado de vetor de consulta (*query vector*), a

vazão correspondente y_q é determinada combinando os k valores de saída dos vetores \mathbf{X}_i (seção 2.1) mais próximos de \mathbf{X}_q .

Uma norma Euclidiana ponderada é geralmente usada para medir a proximidade (similaridade) dos vetores \mathbf{X}_i e \mathbf{X}_q :

$$d(\mathbf{X}_q, \mathbf{X}_i) = \sqrt{\sum_{j=1}^n w_j^A (x_{ij} - x_{qj})^2} \quad (2)$$

em que:

w_j^A é o peso do j -ésimo atributo. Com os valores de saída y_i dos k pontos mais próximos (vizinhos), o valor previsto para y_q é calculado por:

$$y_q = \frac{\sum_{i=1}^k w_i^V y_i}{\sum_{i=1}^k w_i^V} \quad (3)$$

em que:

w_i^V é o peso do i -ésimo vizinho. Na sua forma mais simples, o modelo KNN usa $w_i^V = 1$ e a estimativa é o valor médio dos k vizinhos mais próximos. Na versão melhorada, entretanto, cada vizinho possui um peso baseado na distância $d(\mathbf{X}_q, \mathbf{X}_i)$. Um vizinho mais longe recebe um peso menor, o que reduz seu efeito sobre a previsão em comparação com outros vizinhos mais próximos. Desta forma, algumas funções *kernel*, que decrescem monotonicamente à medida que a distância aumenta, têm sido utilizadas (Akbari et al., 2011):

$$\text{Linear:} \quad w_i^V = 1 - d(\mathbf{X}_q, \mathbf{X}_i) \quad (4)$$

$$\text{Inversa:} \quad w_i^V = [d(\mathbf{X}_q, \mathbf{X}_i)]^{-1} \quad (5)$$

$$\text{Inversa quadrática:} \quad w_i^V = \{[d(\mathbf{X}_q, \mathbf{X}_i)]^2\}^{-1} \quad (6)$$

$$\text{Exponencial:} \quad w_i^V = \exp[-d(\mathbf{X}_q, \mathbf{X}_i)] \quad (7)$$

$$\text{Gaussiana:} \quad w_i^V = \exp\{-[d(\mathbf{X}_q, \mathbf{X}_i)]^p\} \quad (8)$$

Não existem evidências de que uma função é melhor do que as demais. Além disso, também não há um método totalmente estabelecido para selecionar o valor ótimo de k no algoritmo KNN.

2.5. Modelo RKNN: KNN com extrapolação

Em princípio, o modelo KNN nunca poderá prever uma vazão superior às vazões registradas nas instâncias de aprendizado. Para corrigir esse problema, este trabalho propõe um modelo denominado *rainstorming* KNN (RKNN). O RKNN apresenta uma pequena diferença em relação ao KNN, em que um chamado fator de tormenta R incrementa a vazão prevista pelo KNN. Esse fator é calculado da seguinte forma:

$$R = 1 + \frac{\sum_{i=1}^n \max(\mathbf{X}_q(i) - \mathbf{X}_{NN}(i), 0)}{\sum_{i=1}^n \mathbf{X}_{NN}(i)} \quad (9)$$

em que:

$\mathbf{X}_q(i)$ é o i -ésimo componente do vetor de consulta e $\mathbf{X}_{NN}(i)$ é o i -ésimo componente do vizinho mais próximo. Logo, a vazão a ser prevista pelo algoritmo será dada pelo membro do lado direito da Equação 3 multiplicado por R .

Quando ocorrer um vetor de consulta com chuvas ou vazões superiores aos valores do vizinho mais próximo, $R > 1$ e a vazão prevista será superior à do modelo KNN. Caso isso não aconteça, $R = 1$ e a previsão será a mesma do modelo KNN padrão.

2.6. Medidas de desempenho

As medidas de desempenho empregadas para comparar vazões observadas e previstas por cada modelo foram a raiz do erro quadrático médio (*RMSE: root mean square error*), o erro absoluto médio (*MAE: mean absolute error*) e o critério de eficiência de Nash-Sutcliffe (*E*). As equações a seguir mostram como calcular essas medidas:

$$RMSE = \sqrt{\frac{\sum_{t=1}^T [\mathcal{Q}_{\text{obs}}(t) - \mathcal{Q}_{\text{prev}}(t)]^2}{T}} \quad (10)$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |\mathcal{Q}_{\text{obs}}(t) - \mathcal{Q}_{\text{prev}}(t)| \quad (11)$$

$$E = 1 - \frac{\sum_{t=1}^T [\mathcal{Q}_{\text{obs}}(t) - \mathcal{Q}_{\text{prev}}(t)]^2}{\sum_{t=1}^T [\mathcal{Q}_{\text{obs}}(t) - \bar{\mathcal{Q}}_{\text{obs}}(t)]^2} \quad (12)$$

em que:

$\mathcal{Q}_{\text{obs}}(t)$ e $\mathcal{Q}_{\text{prev}}(t)$ são, respectivamente, as vazões observada e prevista para o mês t ($t = 1, \dots, T$), e $\bar{\mathcal{Q}}_{\text{obs}}(t)$ é a média das vazões observadas durante o horizonte de simulação T . A interpretação para esses índices é que o *MAE* e o *RMSE* representam o quanto o valor previsto se desvia do observado. Para ambos, quanto menor o valor melhor é a previsão. A eficiência de Nash-Sutcliffe pode variar de $-\infty$ a 1. Uma eficiência de $E = 1$ corresponde a um ajuste perfeito dos dados modelados aos observados. Uma eficiência de $E = 0$ indica que as previsões do modelo são tão precisas quanto a média dos dados observados, enquanto que uma eficiência inferior a zero ocorre quando a média observada é um indicador melhor do que o modelo.

3. RESULTADOS E DISCUSSÃO

3.1. Correlações para definição dos atributos

Inicialmente, foi calculada a correlação de Pearson e a AMI entre os dados da estação pluviométrica Japarutuba com ela mesma e com as estações pluviométricas para vários *lags*.

De acordo com os valores de correlações calculados foram feitos vários testes via modelo KNN cada um deles com uma combinação diferente de atributos. O Quadro 2 mostra as séries históricas que foram utilizadas em cada teste com o respectivo *lag*. As estações descritas como Japarutuba-1 e Japarutuba-2 são as estações pluviométricas de códigos 01036014 e 01036016, respectivamente.

3.2. Resultados

Para cada um dos testes listados no Quadro 2, foram conduzidas várias simulações com o modelo KNN combinando todas as funções *kernel* das Equações 4 a 8 e os valores de vizinhos $k = 1, \dots, 8$. Ao todo, foram 240 simulações (5 funções \times 8 valores de vizinhos \times 6 testes). Por questões de espaço, apenas os melhores resultados são mostrados neste artigo.

Quadro 2. Atributos utilizados nos testes de aplicação do modelo KNN.

Teste	Atributos de Chuva	Atributos de Vazão
Teste 1	Japaratuba-1 (<i>lag</i> 1) Japaratuba-2 (<i>lag</i> 1) Capela (<i>lag</i> 1)	—
Teste 2	Japaratuba-1 (<i>lag</i> 1)	—
Teste 3	Japaratuba-2 (<i>lag</i> 1)	—
Teste 4	Outeirinhos (<i>lag</i> 1)	—
Teste 5	Japaratuba-1 (<i>lag</i> 1) Japaratuba-2 (<i>lag</i> 1) Capela (<i>lag</i> 1) Curral do Meio (<i>lag</i> 1)	Japaratuba (<i>lag</i> 1)
Teste 6	Usina Outeirinhos (<i>lag</i> 1) Usina Outeirinhos (<i>lag</i> 2)	Japaratuba (<i>lag</i> 1) Japaratuba (<i>lag</i> 2)

A melhor configuração para os modelos KNN e RKNN foi a do Teste 6 com função exponencial e com $k = 2$ vizinhos. A Tabela 2 apresenta o resumo das melhores simulações realizadas pelos modelos KNN, RKNN e ANN levando em consideração os índices de desempenho. As Figuras de 2 a 4 mostram, respectivamente, a comparação entre vazões observadas e previstas a partir dos modelos KNN, RKNN e ANN.

Tabela 2. Melhores medidas de desempenho para cada modelo.

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>E</i>
KNN	2,99	2,12	0,71
RKNN	2,70	1,97	0,76
ANN	1,94	1,30	0,88

3.3. Discussão dos resultados

De acordo com os resultados, a ANN mostrou-se superior ao KNN e RKNN com uma eficiência de $E = 0,88$ e com os menores valores de *RMSE* e *MAE*. Porém, os modelos baseados no algoritmo do vizinho mais próximo também forneceram resultados satisfatórios com o KNN obtendo eficiência de $E = 0,71$ e o RKNN fornecendo um valor de $E = 0,76$.

Um fator limitante para o KNN nesta pesquisa foi a pouca quantidade de dados utilizados para a fase de aprendizado dos modelos (8 anos para a fase de treinamento) o que afeta negativamente o desempenho do modelo.

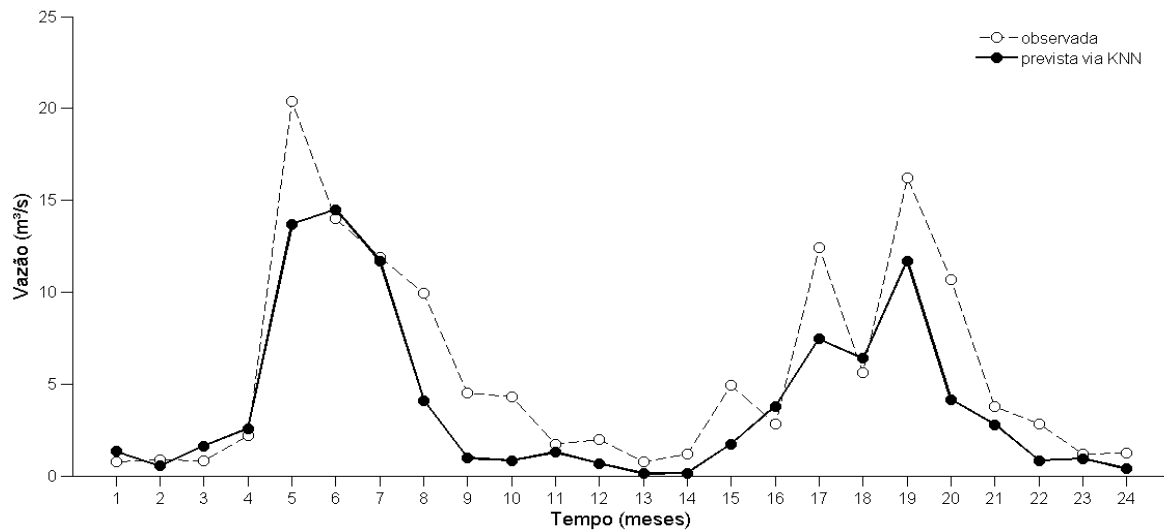


Figura 2. Resultados de previsão via modelo KNN.

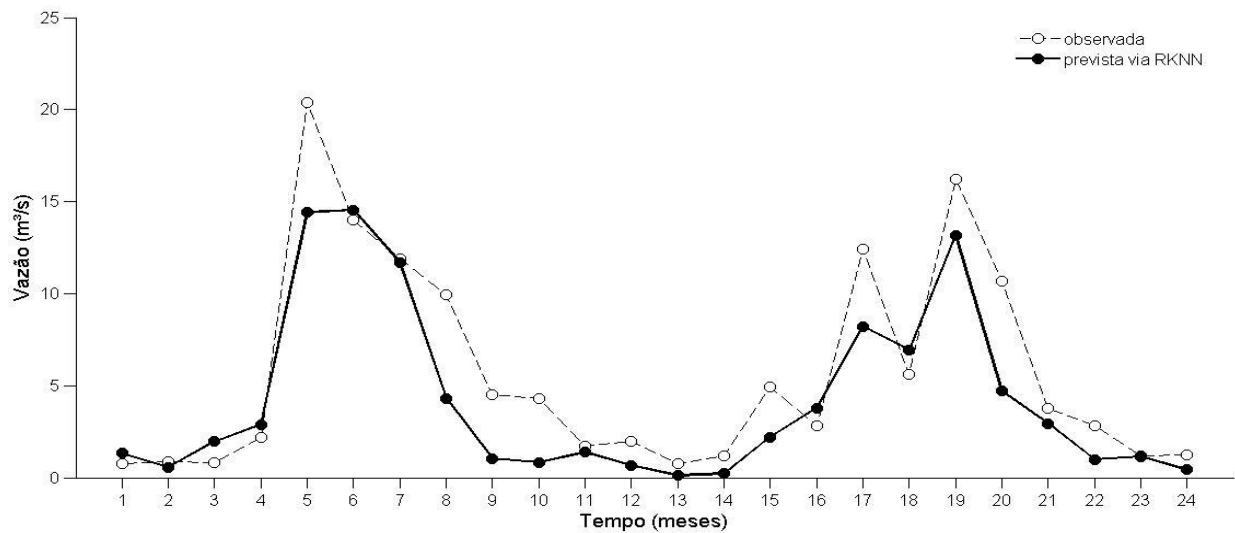


Figura 3. Resultados de previsão via modelo RKNN.

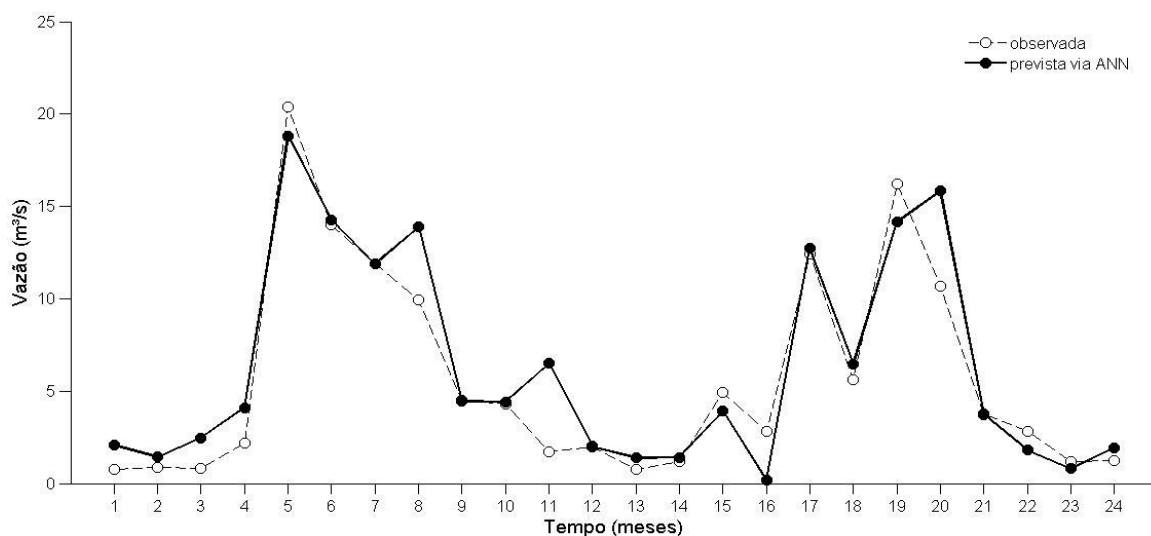


Figura 4. Resultados de previsão via modelo ANN.

O objetivo do modelo RKNN foi alcançado nas simulações para a bacia do Japarutuba uma vez que foi evidenciado que esse procedimento corrige, em parte, o defeito do KNN em prever vazões acima dos registros utilizados para o aprendizado.

A ordem de grandeza das eficiências foi semelhante à encontrada em algumas bacias analisadas por Guilhon et al. (2007) que utilizaram outros modelos empíricos para previsão de vazão. Levando em consideração a simplicidade do algoritmo do vizinho mais próximo em relação à técnica mais sofisticada de redes neurais artificiais, conclui-se que o KNN pode ser útil para previsão de vazão.

4. CONCLUSÃO

Neste trabalho, a metodologia KNN de aprendizado baseado em instâncias foi aplicada para a previsão de vazão no rio Japarutuba, estado de Sergipe. Também foi proposto o modelo RKNN, um refinamento no KNN para que o mesmo fosse capaz de prever vazões maiores do que as históricas (o que não acontece com o KNN original). O desempenho destes modelos foi comparado com os de simulações feitas com o uso de redes neurais artificiais.

As ANNs obtiveram um melhor desempenho em relação aos modelos baseados no algoritmo do vizinho mais próximo. Todavia, ambos o KNN e o RKNN apresentaram bons resultados apesar da pouca quantidade de dados para o aprendizado.

Uma recomendação para trabalhos futuros é testar os modelos em períodos mais longos que o usado neste trabalho (apenas oito anos para calibração/aprendizado e dois anos para validação) e também para horizontes de previsão mais distantes.

5. REFERÊNCIAS

- AKBARI, M.; OVERLOOP, P.; AFSHAR, A. Clustered k nearest neighbor algorithm for daily inflow forecasting. **Water Resources Management**, v. 25, n. 5, p. 1341–1357, 2011. <http://dx.doi.org/10.1007/s11269-010-9748-z>
- 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. **Revista Brasileira de Recursos Hídricos**, v. 15, n. 1, p. 05-15, 2010.
- CRUZ, M. F. M.; RODRIGUES, L. D.; VERSIANI, B. R. Previsão de vazões com a metodologia DPFT e com redes neurais artificiais. **Revista Brasileira de Recursos Hídricos**, v. 15, n. 1, p. 121-132, 2010.
- FARIAS, C. A. S. et al. RNN-based inflow forecasting applied to reservoir operation via implicit stochastic optimization. In: BOEGH, E. et al. (Ed.). **Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management**. [S.l.]: IAHS, 2007. p. 452–462. (IAHS Publi., 313)
- GALEATI, G. A comparison of parametric and non-parametric methods for runoff forecasting. **Hydrological Sciences Journal**, v. 35, n. 1, p. 79–94, 1990. <http://dx.doi.org/10.1080/02626669009492406>
- GALVÃO, C. O.; VALENÇA, M. J. S. (Orgs.). **Sistemas inteligentes: aplicações a recursos hídricos e ciências ambientais**. Porto Alegre: Editora da UFRGS, 1999. 247 p.
- GUILHON, L. G. F.; ROCHA, V. F.; MOREIRA, J. C. Comparação de métodos de previsão de vazões naturais afluentes a aproveitamentos hidroelétricos. **Revista Brasileira de Recursos Hídricos**, v. 12, p. 13–20, 2007.

- KARLSSON, M.; YAKOWITZ, S. Nearest neighbour methods for nonparametric rainfall-runoff forecasting. **Water Resources Research**, v. 23, n. 7, p. 1300–1308, 1987. <http://dx.doi.org/10.1029/WR023i007p01300>
- SERGIPE. Secretaria de Estado do Meio Ambiente e Recursos Hídricos - SEMARH-SE. Elaboração do plano estadual de recursos hídricos: RE-6, Volume 1, Tomo I (Bacia Hidrográfica do Rio Japaratuba). Aracajú, 2010. 70 p.
- SERGIPE. Secretaria de Estado do Meio Ambiente e Recursos Hídricos - SEMARH-SE. Atlas digital sobre recursos hídricos de Sergipe. Aracajú, 2012. 1 CD-ROM.
- SHAMSELDIN, A. Y.; O'CONNOR, K. M. A nearest neighbour linear perturbation model for river flow forecasting. **Journal of Hydrology**, v. 179, p. 353–375, 1996. [http://dx.doi.org/10.1016/0022-1694\(95\)02833-1](http://dx.doi.org/10.1016/0022-1694(95)02833-1)
- SOLOMATINE, D. P.; MASKEY, M.; SHRESTHA, D. L. Instance-based learning compared to other data-driven methods in hydrological forecasting. **Hydrological Processes**, v. 22, p. 275-287, 2008.
- TOTH, E.; BRATH, A.; MONTANARI, A. Comparison of short-term rainfall prediction models for real-time flood forecasting. **Journal of Hydrology**, v. 239, p. 132–147, 2000. [http://dx.doi.org/10.1016/S0022-1694\(00\)00344-9](http://dx.doi.org/10.1016/S0022-1694(00)00344-9)
- VALENÇA, M. J. S. **Aplicando redes neurais**: um guia completo. Olinda: Ed. do Autor, 2005. 284 p.