

Previsão de consumo diário de água: estudo de caso de Nobres (Brasil)

Forecasting daily water consumption: case study of Nobres (Brazil)

Silva, W.T.P.^{a1}, Campos, M.M.^{a2}, Santos, A.A.^{a3}

^aDepartamento de Engenharia Sanitária e Ambiental, Universidade Federal de Mato Grosso, campus Cuiabá, Avenida Fernando Corrêa da Costa, No. 2367, Cuiabá, Brasil. E-mail: ^{a1} welitom@ufmt.br, ^{a2} medinasmarcel@gmail.com, ^{a3} aldecy_allmeida@yahoo.com.br

Recebido: 13/09/2015

Aprovado: 11/03/2016

Publicado: 29/04/2016

RESUMO

Para obter eficiência operacional em sistema de abastecimento de água (SAA) a previsão de consumo de água em curto prazo (para o próximo dia) é necessária. Esse trabalho teve o objetivo de desenvolver um modelo matemático de previsão do consumo diário de água da cidade de Nobres, Estado de Mato Grosso, no Brasil. As etapas metodológicas realizadas foram: (1) revisão de literatura; (2) coleta e análise de dados (consumo de água e clima); (3) proposição de modelo de previsão de consumo diário de água; (4) calibração e verificação do modelo; e, (5) aplicação de modelo. As técnicas de modelagem matemática empregadas foram Regressão Linear, Séries de Fourier e Sistema Especialista. Os resultados indicaram um erro médio percentual do modelo inferior a 10% indicando que apresentou bom ajuste e que pode ser utilizado para prever o consumo de água. Como principal conclusão apresenta-se que o modelo desenvolvido pode ser utilizado para o planejamento operacional do SAA estudado.

Palavras chave | Demanda de água; Regressão Linear; Séries de Fourier; Sistema Especialista.

ABSTRACT

In order to move operational efficiency from the water system supply (Waterworks, i.e. a combined water purification plant and pumping station) forecast water consumption values 24 h ahead are required. The objective of this paper is to develop mathematical model forecast water consumption 24 h ahead for Nobres city, Mato Grosso State, in Brazil. The methodology developed comprises the following steps: (1) literature review; (2) gathering and data analysis (water consumption and climatic); (3) proposal of a model forecast water consumption; (4) calibration and verification of the proposed model; and, (5) application of model. The mathematical modelling techniques employed were Linear Regression, Fourier Series and Expert System. The results indicated that there is error average percentage of less than 10% of model indicating that provided a good fit and can be used to predict water consumption. It can be concluded that the model development which may be used for operational planning the Waterworks study.

Key words | Water demand; Lineal Regression; Fourier Analysis; Expert System.

INTRODUÇÃO

O sistema de abastecimento de água (SAA) consiste do conjunto de estruturas (captação, adutoras, estação de tratamento de água, reservação e distribuição) destinado a ofertar água em quantidade e qualidade adequadas a população (Tsutiya, 2006). Falhas operacionais em SAA podem ter reflexos negativos sobre a qualidade de vida e saúde da população, alguns exemplos incluem reclamações da população, baixa de eficiência operacional, falta de água, entre outros problemas. Com o intuito de contribuir com a resolução desses reflexos negativos ou de evitar que eles venham a ocorrer, vários estudos têm sido desenvolvidos, especialmente na área de planejamento operacional em SAA. Uma das ferramentas de auxílio a minimização de falhas operacionais em SAA são os modelos de previsão de curto prazo, que apresentam entre as principais finalidades a resposta instantânea dos sistemas produtores e distribuidores a picos prolongados e flutuações acentuadas de demanda e o consequente subsídio para a decisão sobre manobra operacional (Silva e Rocha, 1999). Na sequência, alguns estudos realizados como esse intuito são brevemente apresentados.

Lertpalangsunti *et al.* (1999), com o objetivo de auxiliar a tomada de decisões gerenciais no SAA da cidade de Regina, no Canadá, desenvolveram um sistema híbrido (SH), para previsão de consumo de água, que combina técnicas de Lógica Fuzzy, conhecimento baseado em casos e de sistema baseado em conhecimento. León *et al.* (2000), motivados pelas necessidades de redução de custos com energia elétrica, de otimização operacional e de ações de gestão de demanda da água no SAA da cidade de Seville, na Espanha, desenvolveram um SH denominado EXPLORE, os resultados indicam redução de 25% nos custos com energia e benefícios adicionais como qualificação dos operadores menos experientes. Para realizar, eficientemente, a distribuição diária de água, Zhou *et al.* (2000) formularam um modelo matemático computacional capaz de prever o consumo de água para o próximo dia, os resultados indicaram um modelo com performance satisfatória. O modelo proposto por Zhou *et al.* (2000) se fundamentou na hipótese de que a série temporal de consumo é igual ao somatório do consumo de base e o consumo sazonal (componentes sazonal, climático e de persistência).

Esse trabalho teve por objetivo contribuir com o planejamento operacional do SAA da cidade de Nobres, Estado de Mato Grosso, no Brasil, por meio do desenvolvimento de um modelo para previsão do consumo diário de água. O estudo é estruturado em cinco seções. A seção 1 se refere à introdução, a seção 2 apresenta fundamentação teórica, a seção 3 a área de estudo, a seção 4 a metodologia, a seção 5 os resultados e a seção 6 as conclusões do trabalho.

FUNDAMENTAÇÃO TEÓRICA

Nessa seção, uma breve descrição dos fundamentos teóricos acerca das técnicas (Regressão Linear, Séries de Fourier e Sistema Especialista) utilizadas na pesquisa são apresentadas. Também, uma análise comparativa sucinta entre abordagens tradicionais e não-tradicionais para previsão do consumo diário de água é realizada.

Prioritariamente, a Regressão Linear (RL) é uma técnica estatística comumente usada com o propósito de prever e modelar relações entre variáveis de entrada e saída. De acordo com Gujarati (2000) a RL se ocupa do estudo da dependência de uma variável (variável dependente) em relação a uma ou mais variáveis (variáveis independentes) com vistas a estimar e/ou prever o valor médio de uma população de dados. Como fundamento a RL apresenta a utilização do método dos mínimos quadráticos e proposição de equações lineares para descrição da relação entre as variáveis independente e dependentes. Maiores detalhes acerca do método RL podem ser encontrados nos trabalhos de Draper e Smith (1981) e Gujarati (2000).

A ideia central da teoria das Séries de Fourier (SF) é a de representar uma dada função f , periódica e de período 2π , como a soma de uma série trigonométrica (Gandulfo, 1990). De modo geral, esta é uma das ideias básicas da realização de análise matemática: decompor funções arbitrárias em termos de outras mais simples, como o objetivo de encontrar propriedades das funções componentes que as representam. A SF permite decompor sinais periódicos nas suas componentes e tem muitas aplicações no processamento de sinal, no processamento de imagem, na física em várias aplicações, na probabilidade e estatística assim como em muitas outras áreas. Informações adicionais podem ser encontradas nos trabalhos de Santos (2014), Stein e Shakarchi (2003) e James (2011).

Artero (2009) definiu Sistema Especialista (SE) como um sistema computacional destinado a representar o conhecimento de um ou mais especialistas humanos, sobre um domínio bem específico e, a partir do processamento da base de conhecimento, buscar soluções para problemas que, em geral, requerem grande conhecimento especializado. Internamente, o SE consiste de dois componentes principais que são a base de conhecimento e a engenharia de inferência. A base de conhecimento armazena o conhecimento e a engenharia de inferência usa o conhecimento armazenado para construir as conclusões. Alguns conceitos básicos referem-se ao conceito de domínio do problema, conceito de domínio do conhecimento e conceito de engenharia de inferência. O domínio de um problema refere-se a um problema específico de uma área (medicina, finanças, ciência ou engenharia) que o especialista pode resolver. O conhecimento do especialista, sobre como resolver um problema específico é chamado de domínio do conhecimento. A engenharia de inferência diz respeito à habilidade que o SE possui para inferir da mesma maneira que o especialista humano deveria inferir frente a um problema. Na prática, o SE é um programa executável que busca um arquivo, em separado, o conhecimento sobre o seu domínio e isso significa que a base de conhecimento pode ser completamente alterada e, mesmo assim, o programa funcionará normalmente, adotando o conhecimento da nova base (Artero, 2009). Algumas referências sugeridas acerca do assunto são: Klahr e Waterman (1986); Rezende *et al.* (2005); Artero (2009); Giarratano e Riley (2004).

A abordagem tradicional utiliza para previsão do consumo diário de água um modelo matemático idealizado a partir de seu respectivo processo. Para obtenção de modelos matemáticos simplificados, certas limitações são comumente incorporadas ao modelo idealizado, por exemplo, a linearidade da função entrada-saída. Ao se assumir a propriedade de linearidade, pode-se utilizar técnicas extremamente poderosas e conhecidas na área de engenharia e tecnologia, como soluções analíticas ideais (Simões e Shaw, 2007). Devido a essa limitação, sérias dificuldades são encontradas no desenvolvimento e na aplicação do modelo matemático idealizado, o que é especialmente válido para companhias de saneamento com problemas de qualidade de dados e deficiência de qualificação técnica. Por outro lado, a abordagem não-tradicional se fundamenta no método heurístico e possibilita a construção de uma função entrada-saída não matemática (Simões e Shaw, 2007). Em outras palavras, a restrição de linearidade não é mais necessária ou mesmo relevante, uma vez que a função entrada-saída pode ser descrita ponto a ponto, como no método experimental. Logo, a abordagem não-tradicional pode ser recomendada sempre que um modelo matemático equivalente for difícil ou complexo de se obter, sendo essa uma importante vantagem da abordagem não-tradicional, quando se trata de previsão do consumo diário de água em companhias de saneamento que enfrentam problemas na qualidade de dados e deficiência de qualificação técnica.

ÁREA DE ESTUDO

Na descrição da área de estudo foi realizada consulta a informações, principalmente sobre os problemas relacionados com a água, às instituições como Instituto Brasileiro de Geografia e Estatística (IBGE), a administração local e a companhia de saneamento local – Empresa de Saneamento de Nobres Ltda. (ESAN).

O município de Nobres, no Estado de Mato Grosso, no Brasil, está localizado a 151 km da capital Cuiabá (SEPLAN, 2008). A cidade de Nobres possui uma população de 15001 habitantes, distribuídos em uma área de 3859 km² (SEPLAN, 2008). Também, segundo dados do IBGE CIDADES (2016), se observa pequena variação, com tendência a estagnação, da população do município de Nobres (ano: 1991, população: 15174; ano: 1996, população: 15207; ano: 2000, população: 14983; ano: 2007, população: 14862; ano: 2010, população: 15001). O clima em Nobres é equatorial quente e úmido, o período de seca compreende cinco meses (de maio até setembro), os demais meses constituem o chamado período chuvoso. A precipitação média anual de 2000 mm, com intensidade máxima em janeiro, fevereiro e março. A temperatura média anual de 24°C, maior máxima 38°C. O turismo e a indústria de mineração são à base da economia local (SEPLAN, 2008). Em relação as condições de saneamento, um relatório do MCidades acerca desse tema, indica que: (1) a cidade de Nobres apresenta um serviço de abastecimento de água razoavelmente adequado, embora necessite de ajustes para que possa cumprir integralmente com a sua função social; (2) o município não dispõe de um sistema de esgotamento sanitário para atendimento à população o que pode ser considerada um problema de saúde pública; e, (3) em termos financeiros, o serviço prestado no município é autossuficiente para a cobertura das despesas operacionais e apresenta um índice de inadimplência de 2%. Informações complementares acerca das condições de saneamento do município podem ser encontradas no trabalho de MCidades (2008).

A Figura 1 apresenta a localização da cidade de Nobres.

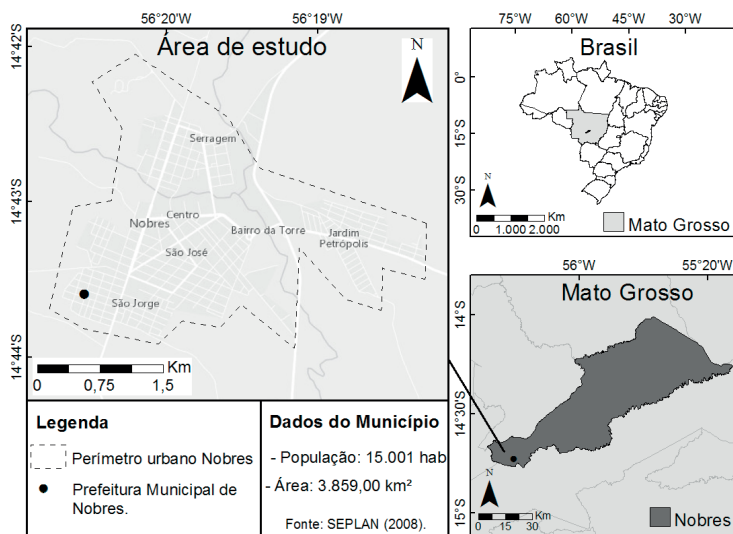


Figura 1 | Localização município de Nobres.

METODOLOGIA

A metodologia proposta compreendeu as seguintes etapas: (1) revisão de literatura; (2) coleta dados; (3) proposição de modelo para previsão de consumo diário de água; (4) calibração e verificação; e, (5) aplicação.

Na etapa de revisão de literatura foram consultados uma série de artigos, dissertações e teses acerca do tema previsão do consumo de água em cidades.

Para essa pesquisa, definiu-se como dados relevantes o volume de água distribuído e os dados climáticos conforme o realizado por Lertpalangsunti *et al.* (1999), Zhou *et al.* (2000), Falkenberg (2005), Odan (2010) e Tian *et al.* (2016). Também, ao considerar o objetivo dessa pesquisa (modelo para previsão do consumo diário de água) verifica-se que informações como indicadores sociais (taxa de alfabetização, população, IDH, outros) e de atividades econômicas (industrial, comercial, turismo, outros), com influência mensal ou anual no consumo de água, não foram considerados. Os dados de volume de água distribuído foram obtidos junto a ESAN, os dados climáticos foram obtidos junto ao Instituto Nacional de Meteorologia (INMET), ambos de frequência diária e no período de 2008 até 2014.

Para proposição de modelo para previsão de consumo diário de água adotou-se a abordagem híbrida (não-tradicional), como justificativa apresenta-se dois pontos principais, o primeiro se refere a existência de recomendações na literatura (Lertpalangsunti *et al.*, 1999; León *et al.*, 2000; Zhou *et al.*, 2000; e, Nikolopoulos, 1997) e o segundo trata-se da complexidade das relações entre as variáveis envolvidas na predição do consumo de água, conforme reportado no trabalho de Tsutiya (2006). Uma vantagem da abordagem adotada é que ela pode integrar diferentes técnicas para formar sistemas mais robustos (Rezende *et al.*, 2005). No caso, as técnicas selecionadas, com base nos trabalhos de Zhou *et al.* (2000), Liao (2005) e Odan (2010), para composição do modelo híbrido foram a de Regressão Linear (RL), a de Séries de Fourier (SF) e a de Sistema Especialista (SE).

Para a etapa de calibração e verificação do modelo fez-se uso do método dos mínimos quadráticos e dos seguintes indicadores de qualidade de ajuste: coeficiente de correlação (r); coeficiente de determinação (R^2); erro médio percentual absoluto (MAPE); e, análise gráfica (valores observados *versus* valores estimados e análise residual). Também, fez-se uso do algoritmo de otimização, acessível e de fácil utilização, hospedado no MS Excel 2010, desenvolvido pela Microsoft Inc. e fundamentado em técnica de otimização não linear do Gradiente Reduzido Generalizado (GRG2). Mais informações sobre processos internos de resolução do algoritmo de otimização Excel-Solver podem ser obtidas no site <http://www.solver.com/>. Para verificação, foram separados do

banco de dados, 25% dos dados, naturalmente, 75% do banco de dados foi utilizado para calibração. Por fim, na etapa de aplicação realizou-se a comparação entre os dados estimados pelo modelo e os observados no SAA da cidade de Nobres, o período estudado foi de 1 de janeiro até 30 de março de 2015 (três meses), período considerado como adequado segundo o realizado por Zhou *et al.* (2000).

RESULTADOS

Na revisão da literatura, constatou-se inúmeras pesquisas realizadas com objetivo similar (previsão de consumo de água para o próximo dia), sendo que as pesquisas mais recentes, fazem uso, de técnica de Redes Neurais Artificiais e sistemas híbridos como metodologia (Lertpalangsunti *et al.*, 1999; León *et al.*, 2000; Falkenberg, 2005). Essas pesquisas, também apontam importante ganho econômico, para companhia de saneamento (via medidas operacionais, como a modificação nos tempos de funcionamento de bombas e consequente redução de custos com energia elétrica) que desenvolvem e implementam modelos de previsão, como principal resultado.

Na Tabela 1 estão apresentados os dados coletados (y é o volume diário consumido; x_1 é a temperatura máxima diária; x_2 é média diária da umidade relativa do ar), análise estatística descritiva e fonte de dados.

Tabela 1 | Resultados da análise estatística descritiva dos dados coletados.

Variável	Unidade	Média	Mediana	Moda	DP	Variância	Curtose	Assimetria	Intervalo	Mínimo	Máximo	Fonte
y	$m^3 \cdot dia^{-1}$	1127.76	1124.42	1147.51	131.36	17255.55	5.68	-0.45	1899.82	60.96	1960.77	ESAN
x_1	$^{\circ}C$	33.25	33.60	33.40	3.32	11.05	2.36	-0.97	25.80	14.90	40.70	INMET
x_2	%	77.28	79.75	84.50	11.59	134.23	0.20	-0.86	57.00	40.25	97.25	INMET

DP: desvio padrão

Como resultados da etapa de proposição de modelo para previsão de consumo diário de água são apresentados, na sequência, descrições acerca do modelo proposto. A ideia básica do modelo híbrido (HI) consiste na hipótese de que o consumo diário de água pode ser previsto a partir de um modelo de Regressão Linear (RL) e correções sucessivas de seus erros, usando para isso as técnicas de Séries de Fourier (SF) e de Sistema Especialista (SE), conforme nas Equações (1), (2) e (3).

$$y = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_i \cdot x_i + \varepsilon_1 \quad (1)$$

$$\varepsilon_1 = a_0 + \sum a_n \cdot \cos(n \cdot \omega_n \cdot t) + b_n \cdot \text{sen}(n \cdot \omega_n \cdot t) + \varepsilon_2 \quad (2)$$

$$\varepsilon_2 = f_{SE}(x_{1D,k}, x_{2D,k}, \dots, x_{iD,k}) + \varepsilon_3 \quad (3)$$

Nas quais: y é o consumo diário de água (m^3); $\beta_0, \beta_1, \dots, \beta_i$ são coeficientes de Regressão Linear; x_1, \dots, x_i são variáveis independentes; ε_1 são os erros do modelo RL; a_0, a_n e b_n são coeficientes da SF; $n\{1, 2, \dots, \infty\}$ é o conjunto de números harmônicos da série; ω_n é o coeficiente de frequência fundamental; t é a variável tempo; ε_2 são os erros do modelo SF; f_{SE} é modelo cognitivo do SE; ε_3 são os erros do modelo SE; x_{1D}, \dots, x_{iD} são a variáveis independentes discretizadas; $x_{1D,k}, \dots, x_{iD,k}$ são valores das variáveis independentes discretizadas. Os coeficientes $\beta_0, \beta_1, \beta_n, a_0, a_n, b_n, \omega_n$ foram obtidos pelo método dos mínimos quadráticos e do algoritmo GRG2. A variável tempo t foi obtida por meio das Equações (4), (5) e (6). Para o último número de harmônicos (∞) adotou-se o valor de 7 ($\infty=7$), uma vez que adotou-se o ciclo semanal, conforme trabalho de Odan (2010). Para discretização das variáveis fez-se uso do método de discretização por intervalos uniformes (U), um caso ($U=0.01$), meramente afim de exemplificação do processo de discretização, é apresentado nas Equações (7), (8) e (9).

$$t_1 = \begin{cases} 1 & \text{se o dia da semana for domingo} \\ 2 & \text{se o dia da semana for segunda - feira} \\ 3 & \text{se o dia da semana for terça - feira} \\ 4 & \text{se o dia da semana for quarta - feira} \\ 5 & \text{se o dia da semana for quinta - feira} \\ 6 & \text{se o dia da semana for sexta - feira} \\ 7 & \text{se o dia da semana for sábado} \end{cases} \quad (4)$$

$$t_2 = j \cdot 365^{-1}, \forall j = 1, \dots, 365 \quad (5)$$

$$t = t_1 + t_2 \quad (6)$$

$$x_{iD} = x_{iD,k}, \forall k = 1, \dots, m \quad (7)$$

$$x_{iD} = \begin{cases} x_{iD,1} = 2.72 & \text{se } x_i \approx 2.72 \\ x_{iD,2} = 2.73 & \text{se } x_i \approx 2.73 \\ \dots & \\ x_{iD,m} = 2.78 & \text{se } x_i \approx 2.78 \end{cases} \quad (8)$$

$$f: x_i = 2.73253 \rightarrow x_{iD} = x_{iD,2} = 2.73 \quad (9)$$

Nas quais: t_1 é a parcela semanal da variável tempo t ; t_2 é a parcela anual da variável tempo t ; j é o j -ésimo dia do ano; $k \{1, \dots, m\}$ é o conjunto de valores discretos da variável x_{iD} . Para determinação de m adotou-se U de 1.00, 2.00 e 0.20 para as variáveis x_1 , x_2 e t , respectivamente. Para discretização dos valores de ε_2 optou-se por sua transformação, conforme Equação (10) e posteriormente adotou-se o mesmo processo de discretização (anteriormente utilizado) com U de 0.05.

$$\varepsilon_{2T} = \ln(\varepsilon_2 - \min_{\varepsilon_2} + 1) \quad (10)$$

Nas quais: ε_{2T} são os erros do modelo SF transformados; \min_{ε_2} é o mínimo valor de ε_2 . Como justificativa para transformação apresenta-se o fato da existência de valores negativos, a elevada quantidade de m possíveis e a necessidade de precisão nos resultados. Na Tabela 2 estão apresentados valores de $x_{1D,k}$, $x_{2D,k}$, $x_{iD,k}$ e $\varepsilon_{2D,k}$ após processo de discretização.

Tabela 2 | Discretização de variáveis ($x_{iD,k}$, ..., $x_{iD,m}$).

Variável	U	m	k=1	k=2	...	k=13	k=28	k=30	k=36
$x_{1D,k}$	1.00	28	14.00	15.00	...	26.00	41.00		
$x_{2D,k}$	2.00	30	40.00	42.00	...	64.00	94.00	98.00	
$x_{iD,k}$	0.20	36	1.00	1.20	...	3.40	6.40	7.00	8.00
$\varepsilon_{2D,k}$	0.05	13	6.60	6.65	...	7.20			

O modelo cognitivo do SE, com estrutura apresentada na Figura 2, foi obtido por meio da realização das atividades tradicionalmente utilizadas para desenvolvimento do SE, e incluem (1) planejamento, (2) explicitação do conhecimento, (3) codificação do conhecimento e (4) avaliação (Giarratano e Riley, 2004).

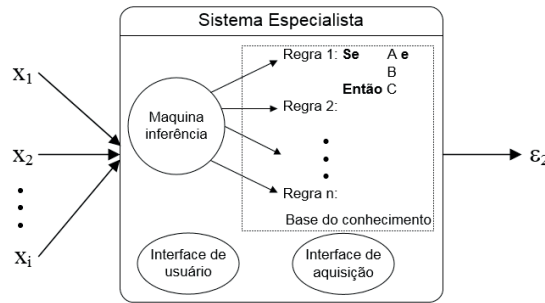


Figura 2 | Estrutura do Sistema Especialista.

Na atividade de planejamento a viabilidade de desenvolvimento e a identificação de recursos (recursos humanos, tempo, recursos financeiros, software e hardware necessários) foram avaliados. Para isso, foi realizado o levantamento, via revisão bibliográfica, dos recursos empregados no desenvolvimento de SE precusores com objetivos similares e a comparação desses recursos com os recursos disponíveis para desenvolvimento do SE aqui proposto.

Para atividade de explicitação do conhecimento, realizou-se a aplicação da técnica de mineração de dados, que busca extrair informações, de modo automático, a partir da Base de Dados de Treinamento (BDT). A BDT, comumente, resulta da consulta (aplicação de questionários) a especialistas em uma dada área do conhecimento, nesse trabalho, a BDT foi constituída por um banco de dados com as variáveis do problema estudado (\$x_{1D,k}, x_{2D,k}, x_{iD,k}, \epsilon_{2D,k}\$). A ideia, aqui considerada, é de que, independentemente da fonte do BDT, a técnica de SE consegue modelar de forma adequada o conhecimento necessário a resolução do problema estudado.

A explicitação do conhecimento se fundamentou na obtenção do domínio de conhecimento, usando o modelo de representação de conhecimento de Regras de Produção. Um exemplo de Regra de Produção, conforme Equação (11), indica a existência de algumas premissas (\$x_{1D} = x_{1D,k}; x_{2D} = x_{2D,k}; \dots; x_{iD} = x_{iD,k}\$) para que uma inferência seja emitida (\$\epsilon_2 = \epsilon_{2D,k}\$).

$$\begin{array}{ll}
 \text{Se} & x_{1D} = x_{1D,k} \quad \text{e} \\
 & x_{2D} = x_{2D,k} \quad \text{e} \\
 & \dots \quad \text{e} \\
 & x_{iD} = x_{iD,k} \\
 \text{Então} & \epsilon_{2D} = \epsilon_{2D,k}
 \end{array} \tag{11}$$

Para a obtenção das Regras de Produção optou pelo emprego do algoritmo NNGE (Non-Nested Generalized Exemplars), mais robusto em casos em que existem informações pouco importantes no conjunto de dados (Mendes, 2011). Situação esperada para o caso estudado, uma vez que há pesquisas prévias indicando a baixa correlação entre consumo de água e variáveis climáticas (Silva *et al.*, 2008). Informações adicionais acerca do algoritmo NNGE podem ser encontradas no trabalho de Zaharie *et al.* (2011). Para utilização do algoritmo NNGE, utilizou-se a ferramenta de mineração de dados Weka (Waikato Environment for Knowledge Analysis). Como técnica para resolução de conflito entre regras optou-se pela estratégia Random (seleção ao acaso de uma regra), conforme recomendado por Giarratano e Riley (2004). Para avaliação do modelo de representação de Regras de Produção, fez-se uso da Matriz de Confusão e da Estatística Kappa (\$\kappa\$), seguindo o recomendado por Rezende *et al.* (2005).

Para atividade codificação do conhecimento, foi utilizado um microcomputador Pentium 2.13GHz, com 4GB de memória RAM em Plataforma Windows, utilizando o VBA (Visual Basic for Applications), hospedado no MS Excel 2010. O VBA foi selecionado devido a facilidade de compartilhamento com usuários e outros programas. Na atividade de avaliação optou-se pela validação por simples estudo de caso, que consiste basicamente na comparação dos resultados apresentados pelo SE e os resultados observados ao analisarem o mesmo problema (Spring, 1997). Para isso, adotou-se o percentual de mínimo de acerto do SE de 75% para que fosse considerado adequado, no caso contrário (percentual inferior a 75%) adequações no SE deveriam ser realizadas.

Os resultados da etapa de calibração e verificação do modelo estão apresentados nas Tabelas 3 e 4, na Figura 3 e nas Equações (12), (13) e (14).

Tabela 3 | Resultados dos coeficientes (modelos RL e SF).

Coefficientes	i=0	i=1	i=2	n=0	n=1	n=2	n=3	n=4	n=5	n=6	n=7
β_i	997.88	10.44	-2.81								
a_n				0.00	-21.01	-7.56	14.96	-3.91	4.82	22.07	38.11
b_n					14.61	-20.24	24.64	38.86	-36.65	7.42	-32.82
ω_n					1635.65	3270.92	2947.10	-1801.50	-458.61	1586.35	-326.73

Tabela 4 | Resultados dos indicadores de qualidade de ajuste (modelos RL, SF, SE e HI).

Modelo	Calibração			Verificação		
	r	R ²	MAPE	r	R ²	MAPE
RL	0.4489	0.2015	9.3	0.4023	0.1619	8.1
SF	0.5157	0.2659	313.6	0.4570	0.2089	165.3
SE	0.7418	0.5502	50.6	0.8499	0.7223	17.6
HI	0.8473	0.7179	3.9	0.8995	0.8090	3.0

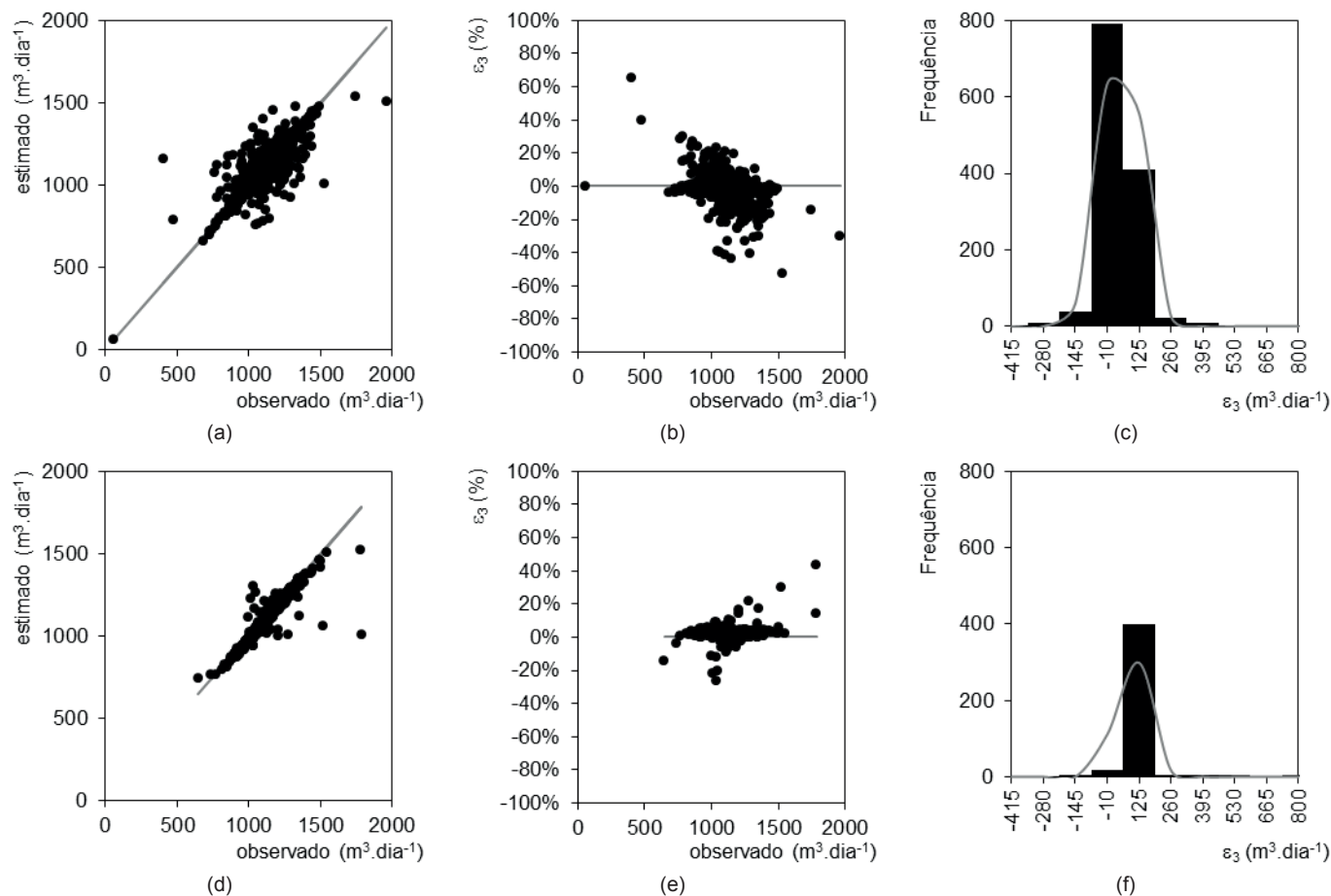


Figura 3 | Análise gráfica: (a), (b) e (c) são resultados de calibração; (d), (e) e (f) são resultados da verificação.

Da Tabela 3 observou-se a existência de uma relação positiva de proporcionalidade entre temperatura (x_1) e o volume diário consumido (y) e também uma relação negativa de proporcionalidade entre umidade relativa do ar (x_2) e o volume diário consumido (y), relações esperadas. Da Tabela 4 notou-se fracos ou moderados valores de r , R^2 e MAPE, com tendência crescente, quando se analisam os modelos (RL, SF e SE) individualmente, tanto no momento de calibração quanto no momento de verificação, porém ao se analisar o modelo HI encontrou-se um bom ajuste (valores de r , R^2 e MAPE adequados). Ajustes semelhantes foram encontrados nos trabalhos de Lertpalangsunti *et al.* (1999), Zhou *et al.* (2000) e Altunkaynak *et al.* (2005), o indica que o modelo HI pode ser aplicado. Ao se analisar a Figura 3 pode notar: (1) forte relação linear entre valores observados e estimados, especialmente no momento de verificação (Figuras 3a e 3d); (2) valores de erro percentual de $\pm 20\%$ e de $\pm 10\%$ nos momentos de calibração e verificação, respectivamente (Figuras 3b e 3e); e, (3) distribuição razoavelmente normal dos resíduos, em ambos os momentos, calibração e verificação (Figuras 3c e 3f). Tais resultados reforçam a indicação de que o modelo HI pode ser aplicado a situação real.

As Equações (12), (13) e (14) apresentam três das 650 regras de produção encontradas após aplicação do algoritmo NNGE. A interpretação literal da regra 1, apresentada na Equação (12), válida, por analogia, para as demais regras, menciona que: se a temperatura máxima do dia for 23°C e umidade média relativa do ar for de 68% ou 80%, e o dia da semana for quarta-feira ou quinta-feira, e mês do ano for março ou agosto então o erro do modelo SF será de -100 . Nas Equações (12), (13) e (14), pode-se constatar uma das grandes vantagens do emprego da técnica de SE, a possibilidade de obtenção de respostas iguais para entradas diferentes (ou, obtenção de respostas diferentes para entradas iguais), confirmando as vantagens já apresentadas por Giarratano e Riley (2004).

$$\begin{aligned} \text{Se} \quad & x_{1D} = x_{1D,10} \quad e \\ & x_{2D} = x_{2D,15} \quad \text{ou} \quad x_{2D} = x_{2D,21} \quad e \\ & x_{1D} = x_{1D,17} \quad \text{ou} \quad x_{1D} = x_{1D,24} \end{aligned} \quad (12)$$

$$\begin{aligned} \text{Então} \quad & \varepsilon_{2D} = \varepsilon_{2D,8} \\ \text{Se} \quad & x_{1D} = x_{1D,10} \quad e \\ & x_{2D} = x_{2D,20} \quad \text{ou} \quad x_{2D} = x_{2D,22} \quad \text{ou} \quad x_{2D} = x_{2D,26} \quad e \\ & x_{1D} = x_{1D,9} \quad \text{ou} \quad x_{1D} = x_{1D,20} \end{aligned} \quad (13)$$

$$\begin{aligned} \text{Então} \quad & \varepsilon_{2D} = \varepsilon_{2D,8} \\ \text{Se} \quad & x_{1D} = x_{1D,10} \quad \text{ou} \quad x_{2D} = x_{2D,19} \quad e \\ & x_{2D} = x_{2D,16} \quad \text{ou} \quad x_{2D} = x_{2D,25} \quad \text{ou} \quad x_{2D} = x_{2D,27} \quad e \\ & x_{1D} = x_{1D,15} \quad \text{ou} \quad x_{1D} = x_{1D,16} \quad \text{ou} \quad x_{1D} = x_{1D,26} \end{aligned} \quad (14)$$

$$\text{Então} \quad \varepsilon_{2D} = \varepsilon_{2D,8}$$

Como resultados da avaliação do modelo de representação de conhecimento são apresentados a Matriz de Confusão (Figura 4) e a Estatística Kappa (Tabela 5).

Na Matriz de Confusão, encontrou-se boa exatidão global (85.6%) do modelo de representação de conhecimento. Da mesma forma, a Estatística Kappa ($\kappa=0.83$) indicou a concordância quase perfeita entre os dados e o modelo de representação de conhecimento. Ambos indicadores de avaliação do modelo de representação de conhecimento apresentaram bons resultados, o que indica que o domínio de conhecimento foi adequadamente representado pelo modelo escolhido.

A atividade de codificação do conhecimento, em ambiente VBA, resultou em um SE capaz de relacionar de forma satisfatória o banco de dados, a máquina de inferência, a interface de aquisição e a interface do usuário. A Figura 5 apresenta a interface do modelo SE, que naturalmente foi integrado aos modelos RL e SF, constituindo assim o modelo híbrido (HI).

=== Matriz de confusão ===

	a	b	c	d	e	f	g	h	i	j	k	l	m	<-- Classificado como
a	3	0	0	0	0	0	0	1	1	0	0	0	0	a = $\epsilon_{2D,1}$
b	0	3	0	0	0	0	1	0	1	0	0	0	0	b = $\epsilon_{2D,2}$
c	0	0	6	0	0	0	0	1	1	0	0	0	0	c = $\epsilon_{2D,3}$
d	0	0	0	21	2	1	0	0	0	1	0	0	0	d = $\epsilon_{2D,4}$
e	0	0	0	0	37	0	4	1	0	2	1	0	0	e = $\epsilon_{2D,5}$
f	0	0	0	2	1	93	6	4	5	0	4	0	0	f = $\epsilon_{2D,6}$
g	0	0	1	1	0	4	206	8	5	4	2	0	0	g = $\epsilon_{2D,7}$
h	0	0	0	1	2	3	12	253	17	12	0	0	0	h = $\epsilon_{2D,8}$
i	0	1	1	0	0	5	9	6	260	6	3	1	0	i = $\epsilon_{2D,9}$
j	0	0	1	1	1	2	6	8	7	153	1	0	0	j = $\epsilon_{2D,10}$
k	0	0	0	0	0	1	2	1	4	1	47	0	0	k = $\epsilon_{2D,11}$
l	0	0	0	0	0	0	1	0	1	0	0	9	0	l = $\epsilon_{2D,12}$
m	0	0	0	0	0	0	0	0	1	0	1	0	4	m = $\epsilon_{2D,13}$

Figura 4 | Matriz de Confusão da avaliação do modelo de representação de conhecimento (RC).

Tabela 5 | Estatística Kappa da avaliação do modelo de representação de conhecimento (RC).

Estatística Kappa (κ)						
Nenhuma concordância ⁽¹⁾	Leve concordância ⁽¹⁾	Concordância regular ⁽¹⁾	Concordância moderada ⁽¹⁾	Concordância substancial ⁽¹⁾	Concordância quase perfeita ⁽¹⁾	Modelo de RC escolhido
< 0.00	0.00-0.20	0.21-0.40	0.41-0.60	0.61-0.80	0.81-1.00	0.83

⁽¹⁾ Landis e Koch (1977).

Figura 5 | Interface do modelo híbrido (HI) desenvolvido.

O modelo HI facilmente pode ser utilizado por usuários, sua utilização consiste basicamente na inserção de dados no campo “Dados de entrada do modelo” e ativação das opções “Enviar dados” e “Calcular”. O resultado da avaliação do SE indicou um percentual de acerto de 86%, que segundo a metodologia adotada indica que o SE pode ser considerado como adequado (percentual mínimo de acerto $\geq 75\%$).

Os resultados da comparação entre os valores de consumo de água estimados e observados (etapa de aplicação do modelo HI) estão apresentados na Figura 6. De modo geral, o modelo HI apresentou performance satisfatória o valor de erro médio percentual foi inferior $\pm 10\%$. Também, se notou subestimação do modelo, especialmente no período de 1 até 19 de janeiro de 2015 (período de comemorações de início de ano e de férias), no restante do período estudado (20 de janeiro até 30 de março de 2015) o modelo apresentou performance bastante satisfatória (erro médio percentual inferior a $\pm 8\%$). Assim, o modelo HI pode ser utilizado pela companhia de saneamento local para suas atividades operacionais.

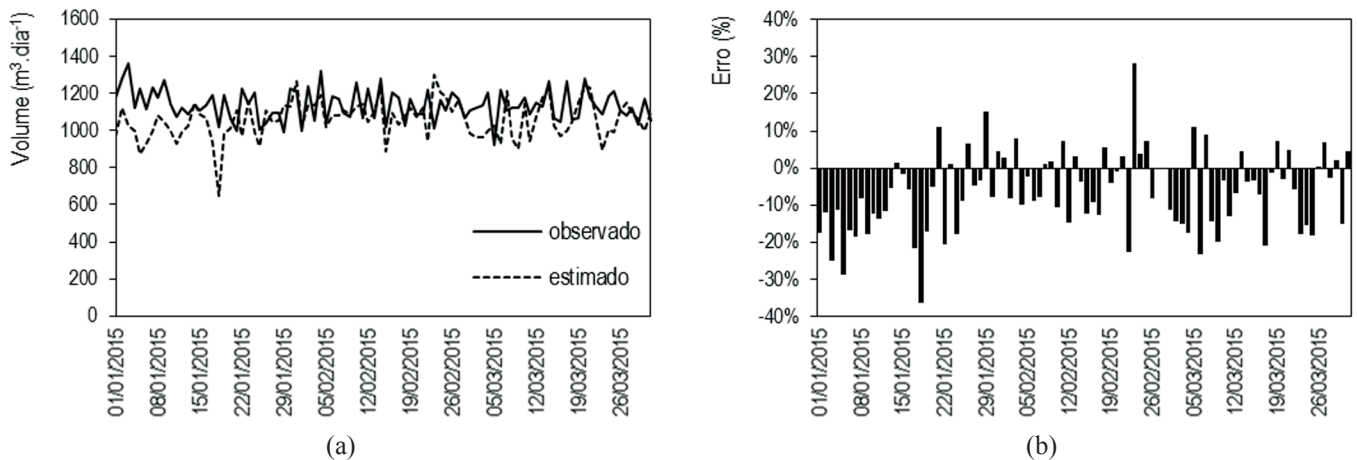


Figura 6 | Aplicação do modelo HI: (a) Série temporal de observado e estimados; (b) Série temporal de Erro (%).

CONCLUSÃO

Um modelo de previsão do consumo diário de água para a cidade de Nobres, Estado de Mato Grosso, no Brasil, foi desenvolvido. O modelo se fundamentou na integração de três técnicas de modelagem distintas (Regressão Linear, Séries de Fourier e Sistema Especialista) e foi implementado em ambiente VBA (hospedado em MS Excel 2010). O modelo foi aplicado (testado em situação real) e apresentou bons resultados, indicado que pode ser utilizado para o planejamento operacional do sistema de abastecimento de água. Do ponto de vista prático, agora, de posse do modelo de previsão do consumo diário, a companhia de saneamento local poderá responder de forma adequada a picos e flutuações acentuadas de demanda, o que implicará na redução de custos operacionais e na redução de indicadores como o percentual de reclamações da população.

A integração entre técnicas de modelagem distintas (abordagem híbrida) se mostrou de grande valia a essa pesquisa, sugere-se a utilização dessa técnica para previsão de consumo de água em outras cidades. Também, sugere-se a utilização dessa técnica para a previsão de consumo de água a curtíssimo prazo (próxima hora) com a finalidade de contribuição com modelos de eficiência energética em sistema de abastecimento de água.

AGRADECIMENTOS

Os autores agradecem à Empresa de Saneamento de Nobres Ltda. (ESAN) e ao Instituto Nacional de Meteorologia (INMET), pelo fornecimento de dados e pela colaboração com o desenvolvimento da pesquisa.

REFERÊNCIAS

- Altunkaynak, A., Özger, M., Çakmakci, M. (2005). Water consumption of Istanbul City by using logic fuzzy. *Water Resources Management*, 19(5), 641-654. <http://dx.doi.org/10.1007/s11269-005-7371-1>
- Artero, A. O. (2009). *Inteligência artificial: teórica e prática*. Editora Livraria da Física, São Paulo, BRA.
- Draper, N.R., Smith, H. (1981). *Applied regression analysis*. John Wiley and Sons, New York, USA.
- Falkenberg, A. V. (2005). *Previsão de consumo urbano de água em curto prazo*. Dissertação de Mestrado, Publicação Mestrado em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, BRA.
- Gandulfo, R. O. (1990). Séries de Fourier e convergência. *Matemática Universitária*, 11, 27-52.
- Giarratano, J. C., Riley, G. D. (2004). *Expert system: principles and programming*. PWS Publishing Company, Boston, USA.
- Gujarati, D. (2000). *Econometria Básica*. 4. ed. Elsevier, Rio de Janeiro, BRA.
- IBGE CIDADES. (2016). *Infográficos: evolução populacional e pirâmide etária*. IBGE. Disponível em: <http://cidades.ibge.gov.br/painel/populacao.php?lang=&codmun=510590&search=mato-grosso|nobres|infogr%E1ficos:-evolu%E7%E3o-populacional-e-pir%E2mide-et%E1ria> Acesso em: 27 fev. 2016
- James, J. F. (2011). *A student's guide to Fourier Transforms*. 3. ed. Cambridge University Press, Cambridge, UK. <http://dx.doi.org/10.1017/CBO9780511762307>
- Klahr, P., Waterman, D. A. (1986). *Expert systems techniques, tools and applications*. Addison-Wesley Pub. Co., New York, USA.
- Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <http://dx.doi.org/10.2307/2529310>
- León, C., Martín, S., Elena, J. M., Luque, J. (2000). EXPLORE – Hybrid expert system for water networks management. *Journal of Water Resources Planning and Management*, 126(2), 65-74. [http://dx.doi.org/10.1061/\(ASCE\)0733-9496\(2000\)126:2\(65\)](http://dx.doi.org/10.1061/(ASCE)0733-9496(2000)126:2(65))
- Lertpalangsunti, N., Chan, C. W., Mason, R., Tontiwachwuthikul, P. (1999). A toolset for construction of hybrid intelligent forecasting systems: application for water demand prediction. *Artificial Intelligence in Engineering*, 13(1), 21-42. [http://dx.doi.org/10.1016/S0954-1810\(98\)00008-9](http://dx.doi.org/10.1016/S0954-1810(98)00008-9)
- Liao, S. H. (2005). Expert system methodology and applications – a decade review from 1995 to 2004. *Expert Systems with Applications*, 28(1), 93-103. <http://dx.doi.org/10.1016/j.eswa.2004.08.003>
- MCidades. (2008). *Municipalização dos serviços de abastecimento de água e de esgotamento sanitário no Estado do Mato Grosso: Diagnóstico, lições e perspectivas*. Ministério das Cidades/Secretaria Nacional de Saneamento Ambiental/Programa de Modernização do Setor de Saneamento (PMSS)/Unidade de Gerenciamento do Programa (UGP). Disponível em: <http://www.pmss.gov.br/index.php/apoio-a-estados/166-municipalizacao-dos-servicos-de-abastecimento-de-agua-e-de-esgotamento-sanitario-no-estado-do-mato-grosso-diagnostico-licoes-e-perspectivas-continuacao> Acesso em: 13 jul. 2015.
- Mendes, D. R. (2011). *Reconhecimento de Orador em Dois*. Dissertação de Mestrado, Publicação Mestrado Integrado em Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Porto, PRT.
- Nikolopoulos, C. (1997). *Expert systems: introduction to first and second generation and hybrid knowledge based system*. Marcel Dekker, New York, USA.
- Odan, F. K. (2010). *Previsão de demanda para sistema de abastecimento de água*. Dissertação de Mestrado, Publicação PPG-SHS, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, BRA.
- Rezende, S. O., Evsukoff, A. G., Garcia, A. C. B., Carvalho, A. C. P. L. F., Braga, A. P., Monard, M. C., Ebecken, N. F. F., Almeida, P. E. M., Ludermir, T. B. (2005). *Sistemas inteligentes: fundamentos e aplicações*. Editora Manole Ltda, Barueri, BRA.

- Santos, F. J. (2014). *Introdução às Séries de Fourier*. PUC Minas. Disponível em: http://www.matematica.pucminas.br/profs/web_fabiano/calculo4/sf.pdf Acesso em: 23. abr. 2015
- SEPLAN. (2008). *Informativo Socioeconômico de Mato Grosso 2005*. Central de Texto. Disponível em: http://www.seplan.mt.gov.br/sitios/indicador/informativo_populacional_%20e_economico_2008.pdf Acesso em: 17 jul. 2012
- Silva, R. T., Rocha, W. S. (1999). *Caracterização da demanda urbana de água*. BRASIL, SEDU, SPU, PNCDA, Brasília, BRA.
- Silva, W. T. P., Silva, L. M., Chichorro, J. F. (2008). Gestão de recursos hídricos: perspectivas do consumo per capita de água em Cuiabá. *Engenharia Sanitária Ambiental*, 13(1), 8-14. <http://dx.doi.org/10.1590/S1413-41522008000100002>
- Simões, M. G., Shaw, I. S. (2007). *Controle e modelagem fuzzy*. Blucher: FAPESP, São Paulo, BRA.
- Spring, G. S. (1997). Critical Review of Expert System Validation in Transportation. *Journal of the Transportation Research Board*, 1588, 104-109. <http://dx.doi.org/10.3141/1588-13>
- Stein, E. M., Shakarachi, R. (2003). *Fourier analysis: an introduction*. Princeton University Press, New Jersey, USA.
- Tian, D., Martinez, C. J., Asefa, T. (2016). Improving Short-Term Urban Water Demand Forecasts with Reforecast Analog Ensembles. *Journal of Water Resources Planning and Management*, 04016008 , [http://dx.doi.org/10.1061/\(ASCE\)WR.1943-5452.0000632](http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000632)
- Tsutiya, M. T. (2006). *Abastecimento de água*. Escola Politécnica de São Paulo, São Paulo, BRA.
- Zaharie, D., Perian, L., Negru, V., Zamfirache, F. (2011). Evolutionary Pruning of Non-Nested Generalized Exemplars. *Proceedings of the 6th IEEE International Symposium on Applied Computational Intelligence and Informatics*, May 19-21, Timișoara, Romania, 57-62. <http://dx.doi.org/10.1109/saci.2011.5872973>
- Zhou, S. L., McMahon, T. A., Walton, A., Lewis, J. (2000). Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology*, 236(3-4), 153-164. [http://dx.doi.org/10.1016/S0022-1694\(00\)00287-0](http://dx.doi.org/10.1016/S0022-1694(00)00287-0)