

## ANALISE DE SINAL DE DEMANDA PARA DETECÇÃO DE EVENTOS ANÔMALOS

*Daniel A. Sierra<sup>1\*</sup>; Lucas Bindi<sup>2</sup>; Victor Petrin<sup>3</sup>; Daniel Manzi<sup>4</sup>; Bruno M. Brentan<sup>5</sup>; Gustavo Meirelles<sup>6</sup> & Edevar Luvizotto Jr<sup>7</sup>.*

**Resumo** – Os sistemas de abastecimento de água estão frequentemente expostos a problemas de cunho operacional que levam a comportamentos não esperados (anomalias). A detecção de eventos anômalos pode ser de grande importância para as companhias de abastecimento por que permitem a rápida restituição do serviço em condições adequadas, reduzindo então possíveis perdas pela ocorrência de vazamentos e aumentando a satisfação dos usuários. Tendo como base um estudo de caso de um distrito de medição em Santa Barbara do Oeste, este trabalho apresenta uma metodologia para detecção de falhas e anomalias na operação de redes de abastecimento de água utilizando-se da análise de dados de demanda de entrada do setor. Para tal, um modelo de previsão de demanda em tempo real é implementado, baseado em redes neurais recorrentes. Com o modelo desenvolvido, é possível estimar a curva de demanda prevista e compara-la de forma gráfica com os dados reais. A comparação permite avaliar o erro cometido pela rede neural, de onde se identifica os eventos anômalos e a presença de vazamento.

**Palavras-Chave** – Sistemas de abastecimento de água, Detecção de eventos anômalos em redes, Previsão de demanda de água.

## ANALYSIS OF DEMAND SIGNAL FOR DETECTION OF ANOMALOUS EVENTS

**Abstract** – Water supply systems are often exposed to operational problems that lead to unexpected behaviors (anomalous). Detection of anomalous events can be of great importance for water supply companies because this allow the prompt return of service under proper conditions, thereby reducing possible leakage losses and increasing the satisfaction users. Based on a case study of a measurement district in Santa Barbara do Oeste, this paper presents a methodology for detecting faults and anomalies in the operation of water supply networks using analysis of water demand data of the input of sector. For this, a real-time demand prediction model is implemented, based on neural networks. With the model developed, it is possible to estimate the expected demand curve and compare it graphically with the real data. The comparison allows evaluating the error committed by the neural network, from where anomalous events and the presence of leakage are identified.

**Keywords** – Water supply systems, Detection of anomalous events in networks, Forecasting of water demand.

---

<sup>1</sup> Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Unicamp, Campinas, São Paulo. [dasierra@unal.edu.co](mailto:dasierra@unal.edu.co)

<sup>2</sup> Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Unicamp, Campinas, São Paulo. [lucascsbindi@gmail.com](mailto:lucascsbindi@gmail.com)

<sup>3</sup> Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Unicamp, Campinas, São Paulo. [victorpetrin@gmail.com](mailto:victorpetrin@gmail.com)

<sup>4</sup> Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Unicamp, Campinas, São Paulo. [dmanzi@gmail.com](mailto:dmanzi@gmail.com)

<sup>5</sup> Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Unicamp, Campinas, São Paulo. [brunocivil08@gmail.com](mailto:brunocivil08@gmail.com)

<sup>6</sup> Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Unicamp, Campinas, São Paulo. [limameirelles@gmail.com](mailto:limameirelles@gmail.com)

<sup>7</sup> Faculdade de Engenharia Civil, Arquitetura e Urbanismo, Unicamp, Campinas, São Paulo. [edevar@fec.unicamp.br](mailto:edevar@fec.unicamp.br)

## 1. INTRODUÇÃO

Desde o século XVIII, a sociedade tem se concentrado em cidades. O processo de mecanização da agricultura reduziu a necessidade de trabalhadores no campo, levando-os à procura de melhores condições, o que resultou no adensamento dos núcleos urbanos. Menos de 5% da população morava nas cidades até o século XVIII, em 2007 cerca de 50% vivia em cidades e é provável que a final do presente século mais do 80% das pessoas morarão em grandes cidades (Harrison e Donnelly, 2011). Este processo de adensamento demanda uma maior oferta de serviços e infraestruturas. O desenvolvimento das cidades orientado às *Smart Cities*, conceito originado no marco do *Smart Growth* (Bollier, 1998), aplicando tecnologia de informação para automatizar funções rotineiras de serviços e realizar monitoramento em tempo real, é uma resposta às maiores exigências em serviços e tem benefícios como: redução do consumo de energia e emissões de CO<sub>2</sub>, exploração de meios de transporte multimodal e otimização no uso da infraestrutura como os sistemas de abastecimento.

A identificação de anomalias em redes de distribuição de água tem ganhado grande destaque na literatura atual, principalmente na rápida identificação de incidentes de terrorismo, como os ciberataques (Rasekh *et al.* 2016). Pasqualetti *et al.* (2013) mostra um marco dos estudos sobre ataques virtuais, analisando as limitações da teoria de grafos aplicada a este problema e apresentando o desenho de identificadores para ataques. Ostfeld e Salomons (2004) apresenta uma aproximação para mitigar os ataques a partir do monitoramento dos sistemas e da geração de alarmes imediatos (Early Warning System, EWS). Agências federais, comunidades acadêmicas e companhias privadas estão trabalhando para identificar eventos de contaminação de baixa probabilidade e alto impacto para evitar os impactos adversos (Hasan e Deininger, 2010). Na pesquisa da presença de ataques dirigidos à perda de água e à contaminação, tem-se usado técnicas de aprendizagem de máquinas como redes neurais artificiais para a análise de padrões de dados (pressão, demanda, qualidade, etc) em períodos de operações normais e se compararam com um eventual ataque.

O reconhecimento de padrões por meio da análise de desvio entre séries previstas e medidas tem sido um caminho preferencial dos recentes trabalhos que visam identificar anomalias no sistema. Ye e Fenner (2010) propuseram um método para identificação de grandes vazamentos em tubulações utilizando a diferença obtida entre série histórica medida e série ajustada a partir de filtro de Kalman adaptativo. Os autores apontam para a alta correlação entre a capacidade do método em classificar um evento anômalo e a magnitude de tal evento, mostrando que a redução gradual do vazamento implica em aumento significativo no erro de classificação. Recentemente, Bakker *et al.* (2014) propuseram a análise de sinal de vazão na entrada de setores para identificação e geração de alarme no caso de ruptura de tubulações ou grandes vazamentos. A comparação entre a série prevista e a série medida através da análise via carta de controle de soma cumulativa, um método de controle estatístico que compara médias e desvios, foi aplicada para a geração de alarmes de ruptura.

Diversos trabalhos têm sido apresentados na literatura para tratar de uma ferramenta importante para a detecção de eventos anômalos, a previsão de demanda (Brentan *et al.* 2017; Odan e Reis 2012 e Herrera *et al.* 2010). Isso porque a partir da comparação entre o comportamento do normal do consumo dos usuários, estimado pela previsão de demanda, e os dados reais pode-se identificar problemas como erros na medição ou eventos anômalos (fugas, vazamentos, rupturas).

Tendo em conta o desafio da detecção de eventos anômalos e os benefícios do rápido tratamento das anomalias, este trabalho apresenta uma análise dos sinais de demanda de entrada de um setor real, identificando os padrões normais de consumo e as possíveis irregularidades. Uma rede neural recorrente foi treinada para prever a demanda da área de estudo, o resultado da previsão foi comparado com os dados reais e se identificaram trechos do padrão com uma grande variação.

## 2. METODOLOGIA

### A. Criação de banco de dados

O banco de dados para o treinamento e a aplicação da rede neural foi construído com dados de consumo do distrito de medição de Santa Barbara do Oeste, no estado de São Paulo, fornecidos pela Agência Reguladora ARES-PCJ. Tal banco fornece consumos medidos a cada 15 minutos, com duração de 15 meses, iniciando em dezembro de 2015.

Diversos autores têm estudado a correlação entre as variáveis climáticas e os hábitos de consumo (Adamowski e Karapataki, 2010; Tiwari e Adamowski, 2017) usando aproximações empíricas ou matemáticas. Por tanto, além dos dados de demanda, construiu-se um segundo banco de dados com variáveis climáticas como temperatura, umidade e velocidade relativa do vento, tomadas no posto meteorológico de Piracicaba o qual corresponde ao ponto de medição mais próximo da rede de estudo. Os dados climáticos são medidos a cada hora desde o mês de março de 2016 até março de 2017.

O primeiro passo no tratamento dos dados foi encontrar o período de tempo onde exista a mesma disponibilidade de dados de demanda e climáticos. Este período de tempo corresponde aos 12 meses desde março de 2016 até março de 2017. Tendo o período de tempo de estudo definido, eliminaram-se os dados com falhas de medição maiores que seis horas, tanto para variáveis climáticas quanto para a demanda. Esse período de seis horas foi determinado como limite para impedir que o preenchimento estatístico do banco de dados possa prejudicar as aplicações futuras desses dados.

Considerando que as medições climáticas e a demanda têm periodicidade de 1 hora e 15 minutos, respectivamente, foi necessária uma interpolação das variáveis climáticas. Com os dados horários, interpolou-se linearmente para obter os dados cada 15 minutos, construindo assim um banco de dados contínuo e pronto para desenvolver o modelo de previsão de demanda.

Uma rede neural necessita de um banco de dados bastante robusto para que se obtenha um bom desempenho na sua aplicação. Para isso, um processo de filtragem dos dados foi realizado, eliminando eventos que visualmente fossem anomalias e ainda períodos em que a vazão mínima noturna fosse superior a 1 l/s. Por tanto o banco de dados usado para o treinamento só incluiu os dados de demanda sem vazamento e sem eventos anômalos.

Logo para o processo de teste para detecção de anomalias, dois cenários foram criados: 1) Teste sem eventos anômalos com dados de vazamento e 2) Teste com eventos anômalos e com dados de vazamentos. No primeiro cenário, os dados de teste correspondem aos dados com vazamento supõem-se que o erro de previsão de demanda será menor em comparação com o cenário 2, uma vez que o vazamento translada a curva de demanda normal. No cenário 2, espera-se que a diferença entre a previsão de demanda e os dados reais seja maior, isso porque esse cenário apresenta as mudanças bruscas ou eventos anômalos.

### B. - Redes neurais não lineares auto-regressivas com entradas exógenas - NARX

O uso de redes neurais artificiais tem sido observado na literatura como caminho preferencial para a modelação de séries temporais e, no caso específico do abastecimento de água, a modelação de demanda de água (Bowden *et al.* 2005; Ghiassi *et al.* 2008; Adamowski, 2008). Uma rede neural artificial (RNA) consiste em unidades computacionais (neurônios artificiais) interconectadas que recebem dados de entrada (*inputs*), que são multiplicados por pesos (*weights*) e então computadas por uma função matemática que determinará a ativação do neurônio, para só então uma outra função computar os dados de saída (*outputs*). Entre os neurônios de entrada e os de saída pode haver inúmeros

neurônios intermediários conhecidos como “hidden nodes”, ou simplesmente camadas ocultas, que contribuem para a otimização das redes em problemas mais complexos. As redes podem contar com os atrasos nos sinais de envio dos dados de entrada e de *feedback* dos dados de saída, conhecido como *delays*, que influenciam diretamente na resposta final da rede. É possível dividir uma RNA em três camadas, sendo elas a camada de *input* (*input layer*), a camada escondida (*hidden layer*) e a camada de *output* (*output layer*).

Com o entendimento da RNA e tomando a alta dependência temporal da demanda de água, as redes neurais recorrentes são ideais para tal modelação. Dentre as redes recorrentes, as redes NARX (Auto regressiva não-linear com *entrada* exógena) (Lin, 1996) tem se mostrado potente ferramenta de previsão de demanda, dado caráter sequencial do problema de previsão de séries de demanda.

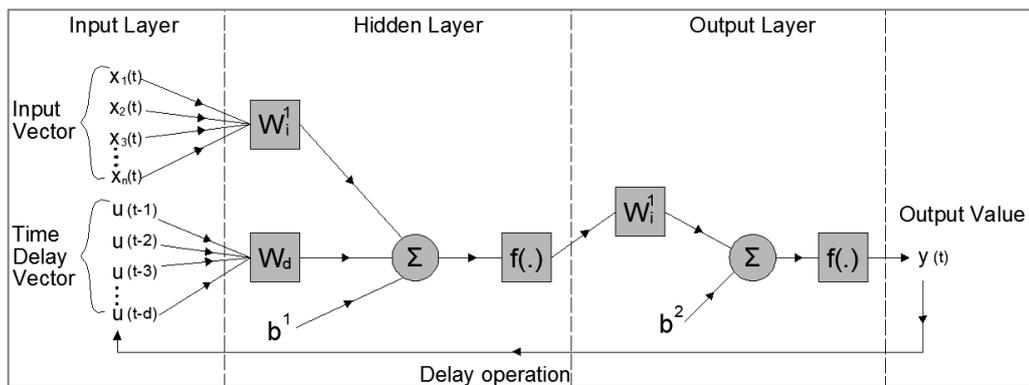


Figura 1– Arquitetura Geral da NARX com uma camada oculta de passos no tempo de atraso, fonte: Brentan *et al.* 2017.

### 3. ESTUDO DE CASO: SANTA BARBARA DO OESTE

#### A. Descrição da Rede

O distrito de medição da cidade de Santa Barbara do Oeste, no estado de São Paulo, tem um comprimento de tubulações aproximado de 2,5 km com diâmetro de duas polegadas (2”) em PVC. As vazões de demanda variam de 0 até 7 l/s ao longo do dia e durante o período de análise. A rede abastece um total de 222 usuários residenciais e a elevação média dos nós é de 546 m, variando desde 530 m até 557 m.

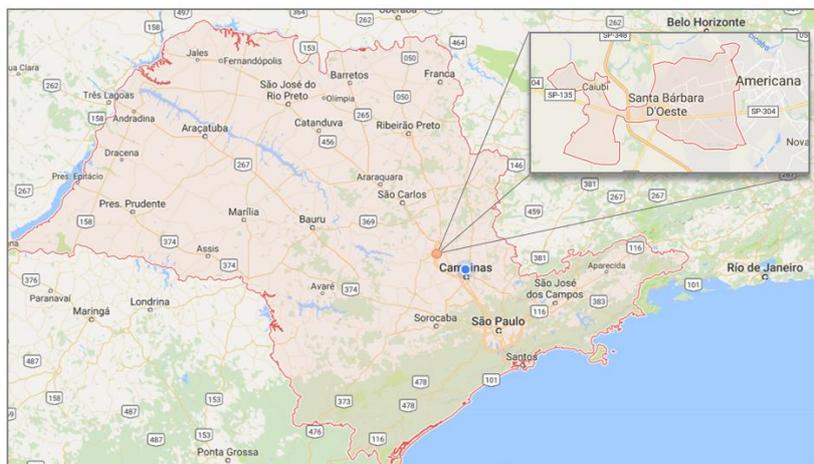


Figura 2– Localização de Santa Bárbara do Oeste no Estado de São Paulo, fonte: Google Maps 2017.



Figura 3– Área de Estudo em Santa Barbara do Oeste, fonte: Google Maps 2017.

## B. Parâmetros da rede Neural (NARX)

Como mencionado na metodologia, a rede neural tem parâmetros que devem ser ajustados para melhorar seu desempenho, neste caso minimizar o erro entre a demanda medida e a prevista. A NARX utilizada nesse trabalho teve a configuração de 10 neurônios, foi treinada como o método de Gradiente Escalonado Conjugado, sendo que o banco de dados de treinamento foi dividido entre 80% dos dados para o treinamento e 20% para a validação. O número de *delay* (sinais de atraso) foi alterado várias vezes e o erro médio absoluto (*e*, em unidades de l/s) foi calculado pela equação 1:

$$e = \sum_{i=1}^n \frac{|d_i - y_i|}{n} \quad (1)$$

Em que  $d_i$  corresponde aos dados de demanda reais,  $y_i$  aos resultados do previsor de demanda e  $n$  é o número total de dados utilizados para a validação.

Na Tabela 1 nota-se os diferentes valores para o número de *delay* e os erros correspondentes. Considerando que o passo de tempo seja de 15 minutos, um *delay* de 24 corresponde a um período de 6 horas. Observa-se que até um *delay* de 72 o erro médio absoluto tende a aumentar, voltando a reduzir em 96. Também é possível observar que o erro menor foi obtido como um valor de  $e = 0,1945$  l/s como um *delay* de 24. Com este número de *delay*, o modelo de previsão de demanda foi desenvolvido para a análise dos sinais de demanda.

Tabela 1. Número de Delay e erro no previsor de demanda

Delay	Erro (l/s)
24	0,1945
48	0,1999
72	0,3157
96	0,2002

## C. Resultados da previsão e detecção

Após a elaboração do modelo de previsão de demanda baseado na NARX e da aplicação aos dois cenários de teste, as curvas da demanda prevista foram obtidas. No primeiro cenário (validação

com dados sem eventos anômalos, mas com vazamentos) a curva de previsão de demanda comparada com os dados reais não mostra informação útil para detectar algum evento.

No cenário 2 (teste com eventos anômalos) apresentam-se variações as quais são mostradas na Figura 4 e na Figura 5. Na Figura 4 um trecho da curva de previsão de demanda e dos dados de demanda reais é apresentado. Pode-se observar de forma clara um evento anômalo. Este corresponde com a mudança instantânea na curva de demanda e a curva de previsão apresenta um erro consideravelmente maior neste ponto.

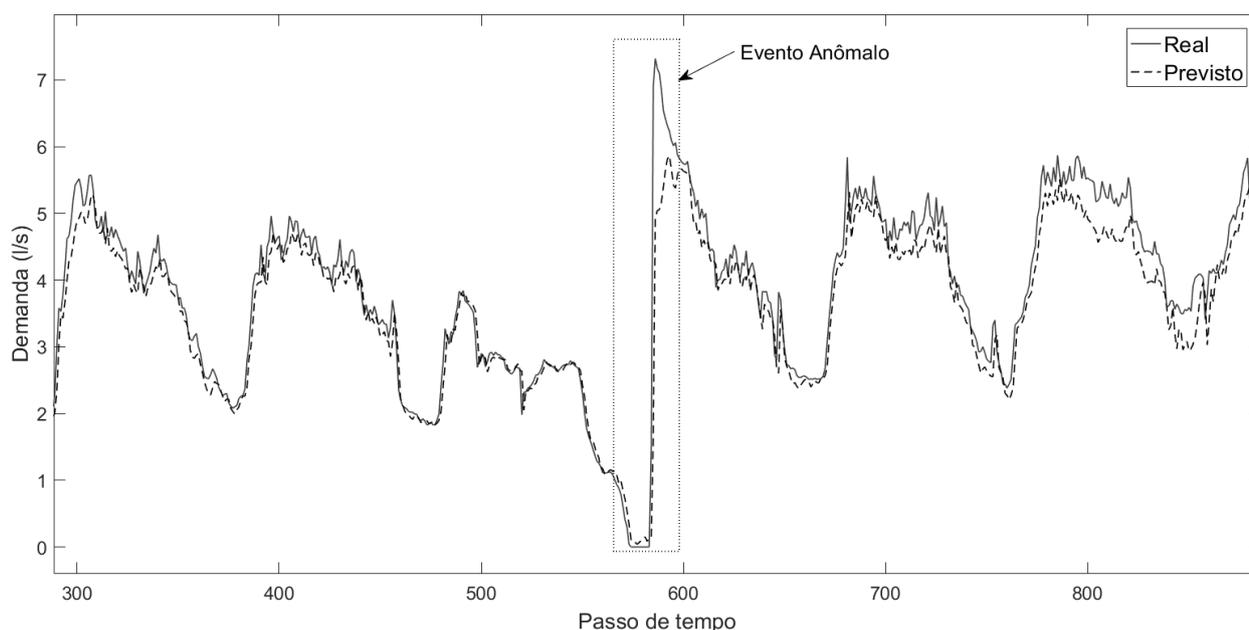


Figura 4 – Trecho da curva de previsão de demanda – Evento Anômalo (Cenário 2), fonte: Elaboração própria

Na Figura 5 apresenta-se outro trecho da curva de previsão de demanda e os dados de demanda reais. Pode-se olhar uma mudança nas vazões mínimas, das horas da noite, que tende a aumentar. Esta mudança, onde se deveria ter uma demanda perto de zero, corresponde com um vazamento. A curva de previsão não tem uma diferença significativa pelo qual ela por se só não consegue detectar o evento anômalo relacionado como o começo do vazamento.

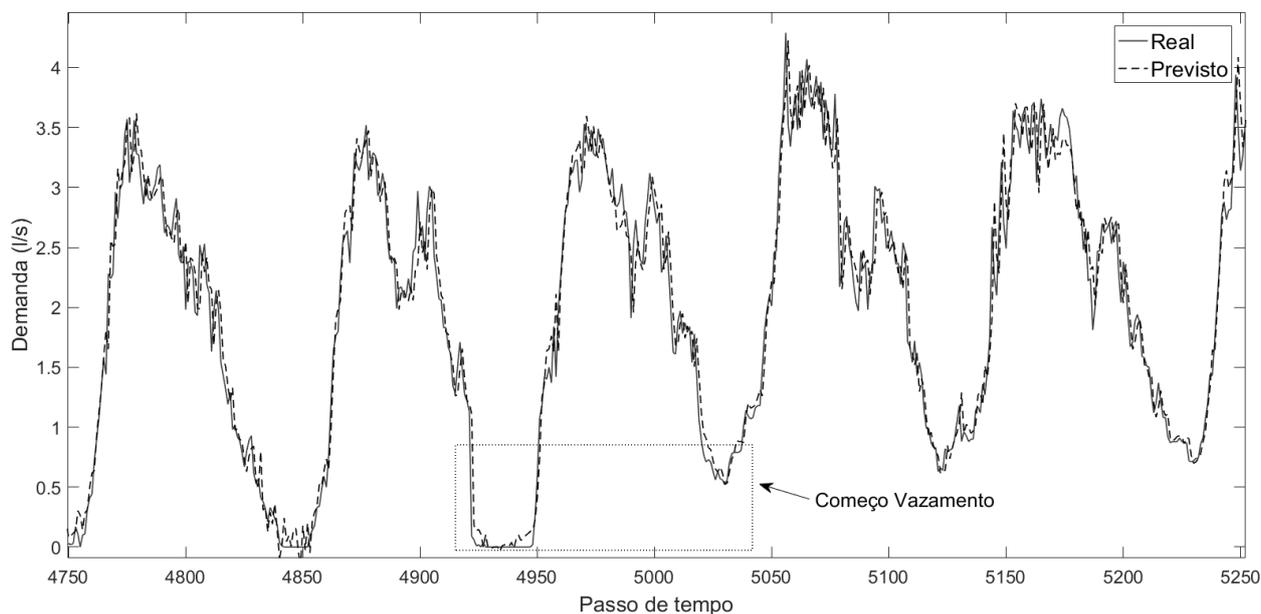


Figura 5 – Trecho da curva de previsão de demanda – Começo Vazamento (Cenário 2), fonte: Elaboração própria

#### D. Conclusões

Os eventos anômalos em redes de abastecimento de água, como falhas de medição, ruptura de tubulações e aparecimentos de vazamentos diminuem a eficiência do serviço, levando a custos operacionais elevados. Após a elaboração do modelo de previsão de demanda com os dados de treinamento sem vazamento e os dados de teste incluindo os eventos anômalos é possível observar na Figura 4 que o erro fica consideravelmente maior quando se apresenta um evento anômalo.

Para a geração do banco de dados, a informação contínua no tempo com o menor número de ausências ou erros na medição é de fundamental importância, isso porque a qualidade da previsão de demanda e consequentemente a precisão da detecção de anomalias depende da consistência do banco de dados.

Ainda que a correlação entre os maiores erros de previsão e as anomalias detectadas graficamente seja evidente, uma análise estatística do erro entre medido e estimado pode ajudar a identificar eventos anômalos que não são evidentes na primeira olhada. Os vazamentos não são identificados de forma clara pela só previsão de demanda. A abordagem via análise estatística do erro ou da modelagem de outras variáveis hidráulicas, como a pressão, pode ajudar detectar as anomalias relacionadas ao começo de um vazamento.

#### REFERÊNCIAS

- ADAMOWSKI, J.; KARAPATAKI, C. (2010). Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering* 15.10, 729-743.
- ADAMOWSKI, J F. (2008). Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management* 134.2, 119-128.

- BAKKER, M.; JUNG, D.; VREEBURG, J.; VAN DE ROER, M.; LANSEY, K.; RIETVELD, L. (2014). Detecting pipe bursts using Heuristic and CUSUM methods. *Procedia Engineering*, 70, 85-92.
- BOLLIER, D. (1998). *How smart growth can stop sprawl: a fledgling citizen movement expands*. Essential Books.
- BOWDEN, G J.; DANDY G C.; MAIER H R. (2005). Input determination for neural network models in water resources applications. Part 1 — background and methodology. *Journal of Hydrology* 301(1), 75-92.
- BRENTAN, B. M.; LUVIZOTTO JR, E.; HERRERA, M.; IZQUIERDO, J.; PÉREZ-GARCÍA, R. (2017). Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*, 309, 532-541.
- GHIASSI, M.; ZIMBRA, D. K.; SAIDANE, H. (2008). Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management*, 134(2), 138-146.
- HARRISON, C.; DONNELLY, I. A. (2011). A theory of smart cities. In Proceedings of the 55th Annual Meeting of the ISSS-2011, Hull, UK (Vol. 55, No. 1).
- HASAN, J.; DEININGER, R. (2010). Safeguarding the security of public water supplies using early warning systems: a brief review. *Journal of Contemporary Water Research and Education*, 129(1), 7.
- HERRERA, M.; TORGO, L.; IZQUIERDO, J.; PÉREZ-GARCÍA, R. (2010). Predictive models for forecasting hourly urban water demand. *Journal of hydrology*, 387(1), 141-150.
- LIN, T.; HORNE, B. G.; TINO, P.; GILES, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks*, 7(6), 1329-1338.
- ODAN, F. K.; REIS, L. F. R. (2012). Hybrid water demand forecasting model associating artificial neural network with Fourier series. *Journal of Water Resources Planning and Management*, 138(3), 245-256.
- ODAN, F. K.; FERRERO, C. A.; REIS, L. F. R.; MONARD, M. C. (2009). Análise comparativa dos modelos kNN-TSP e Série de Fourier para previsão de demanda horária para abastecimento de água. *Anais do XVIII Simpósio Brasileiro de Recursos Hídricos*, Campo Grande, 1-30.
- OSTFELD, A.; SALOMONS, E. (2004). Optimal layout of early warning detection stations for water distribution systems security. *Journal of Water Resources Planning and Management*, 130(5), 377-385.
- PASQUALETTI, F.; DÖRFLER, F.; BULLO, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11), 2715-2729.
- RASEKH, A.; HASSANZADEH, A.; MULCHANDANI, S.; MODI, S.; BANKS, M. K. (2016). *Smart Water Networks and Cyber Security*.
- TIWARI, M. K.; ADAMOWSKI, J. F. (2017). An ensemble wavelet bootstrap machine learning approach to water demand forecasting: A case study in the city of Calgary, Canada. *Urban Water Journal*, 14(2), 185-201.
- YE, G.; FENNER, R. A. (2010). Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *Journal of pipeline systems engineering and practice*, 2(1), 14-22.