ARTIFICIAL NEURAL NETWORK BASED WATER NETWORK STATE ESTIMATION TOOL FOR BANGALORE INFLOW SYSTEM

Anjana G R[1*], K R Sheetal Kumar[1], Justin S Payan[2], and M S Mohan Kumar[3]

ABSTRACT: The main aim of this work is to develop a software tool for water distribution system modeling by coupling Artificial Neural Network (ANN) with real time flow and pressure data from the system. This tool helps in predicting the future state of the system, i.e. the flow in every pipe, pressure at every node and reservoir levels, for a given set of sensor readings at the current time step. The study helps to perform an ANN based sensitivity analysis of the network, and it can be extended to sensor placement optimization and demand prediction. Here, we have utilized feed forward artificial neural network with three hidden layers to predict water level in Bangalore inflow model. To compensate for practical sensor error, random noises were added in the training data set. The objective of this work is to create a collection of ANN if there is a well know question (state) we can instantaneously answer with help of the model. Genetic algorithm was used to optimize the network architecture. Gradient descent, and resilient back-propagation were used as training algorithms. In this research work, it was observed that the computational cost of the ANN based model is less than that of classical modeling approaches and hence can be used to replace hydraulic based tools for system state estimation. In addition, the normalized root mean squared error of our best model is around 0.05, meaning that little information would be lost by replacing a classical model with the neural network model.

Keywords: ANN based network model, Alternate scheme for modeling, Water network analysis.

[1] Research Scholar, Department of Civil Engineering, Robert Bosch Center for Cyber Physical Systems (RBCCPS), Indian Institute of Science, Bangalore -560012, Indiaanjanagr@civil.iisc.ernet.in, sheetal@civil.iisc.ernet.in
[2] M S Student, Institute for Artificial Intelligence, University of Georgia, 200 D.W. Brooks Drive, Athens, GA, USA 30602, jpay17@uga.edu
[3] Professor, Department of Civil Engineering, Indo-French cell for Water Sciences, Robert Bosch Center for Cyber Physical Systems (RBCCPS) and Inter-disciplinary Centre for Water Research (ICWaR), Indian Institute of Science, Bengaluru-560012, msmk@civil.iisc.ernet.in

INTRODUCTION

Water Distribution Systems (WDS) modelling has played a substantial part in system management and daily operational practices. Modelling of WDS using software like EPANET, WaterGEMS, Tandler etc. in the course of time enabled the water authorities to make informed decisions regarding water allocation. Since the advent of low cost sensors and flow measurement devices, there have been enormous amounts of real time data available with water authorities for decision making. Integrating this real time flow and pressure data with the hydraulic models will help us in real time even detection and localization like leak location, contaminant source detection, demand prediction etc. This work presents a novel approach for integrating the real time data for state estimation of the WDS using Artificial Neural Networks (ANNs).

Artificial Neural Networks (ANNs) are a type of machine learning models inspired from central nervous systems. In the past it was observed that ANNs are efficient in many modelling environmental systems and their underlying phenomenon. They have been found to accurately model the export of nutrients in surface water bodies, forecast salinity, ozone levels, air pollution, land use patterns, forecast loading in power systems, groundwater depth fluctuations etc. (Maier and Dandy, 2001; Almeida et.al, 2008; Park et.al, 1991; Coulibaly et.al, 2001). In WDS, ANN models were used for design optimization, event detection, parameter forecast and prediction. Rodriguez et.al (1997) presents an ANN based approach for predicting residual chlorine in WDSs. Milot et.al (2002) compared ANN models with regression models (logistic and multivariate) for predicting THM levels in water systems. It was observed that ANN model performed better than regression models. ANN was also used for developing consumer demand forecast models for predicting long term and short term water demands (Jain and Ormsbee, 2001; Romano and Kapelan, (2014a). Broad et.al (2006) used ANN based meta-models for WDS design and optimization and it was found to perform better than/ comparable to Genetic Algorithm (GA) based optimization methodology. An ANN-GA based meta-model with less runtime was proposed in the study. Goncalves et.al (2011) proposed a hybrid model using ANN for WDS optimization in a case study of a Portugal WDS. ANN based methodology have been proved to be useful in leak event detection as well in the recent years, provided that a time series data of leak events in the network are available. Romano et.al (2014b) coupled real time data from the networks, ANN and hydraulic model to detect leaks events in WDS. But in most of the studies, the authors have used ANN methods coupled with classical WDS modelling tool for analysis. In this study, we are introducing a data driven methodology for modelling the WDS, there by replacing the conventional modelling tools.

STUDY AREA

Bangalore inflow system supplies about 910 MLD of water to a vast population of 10 million dispersed across the 760 km^2 area of Bangalore. This 910 MLD of water is pumped from Torekadanahalli (TKH) Water Treatment Plant water treatment plant about 100 kms away from the city and at an elevation of about 400m in 4 stages of pumping and is distributed to 52 reservoirs of varying capacity across different locations in the city, from which the water is supplied to the consumers. Cauvery Water Supply Scheme (CWSS) which supplies water to the Bangalore city has four stages of supply and the capacities of these stages are shown in Table 1 (Manohar and Mohan Kumar, 2014).

Table 1 Capacities of different stages of Cauvery Water Supply Scheme

Source (supply scheme)	Established (Year)	Supply (MLD)
Stage-I	1974	140
Stage-II	1983	140
Stage-III	1993	315
Stage-IV – Phase 1	2002	315
Total Supply		910

The Bangalore inflow network consists of 500 junctions, 55 Ground Level Reservoirs (GLRs), 516 pipes, 85 pumps and 156 valves (Figure 1). The inflow system is instrumented with 216 flow meters; the main objective of this work is to develop an ANN based tool for state estimation of Bangalore inflow system using these 216 flow meter readings. Here, state of the system denotes the water levels in the GLRs.

ARTIFICIAL NEURAL NETWORKS

ANNs operate as a parallel processor, where there are many processing units interconnected to give the outputs. It consists of an input layer, one or more hidden layers and an output layer, which are connected to each other. Each layer is formed with one or more neurons or Processing Elements (PE) and output of the PE in one layer forms the input to the PE in the next layer. Figure 2 is a schematic of the widely used model called the multi-layered perceptron (MLP) of ANN. The MLP type ANN consists of one input layer, one or more hidden layers and one output layer. The output of the neuron i, h_i in the hidden layer is

$$h_i = \sigma\left(\sum_{j=1}^{N} w_{ij}x_j + \theta_i\right) \quad (1)$$

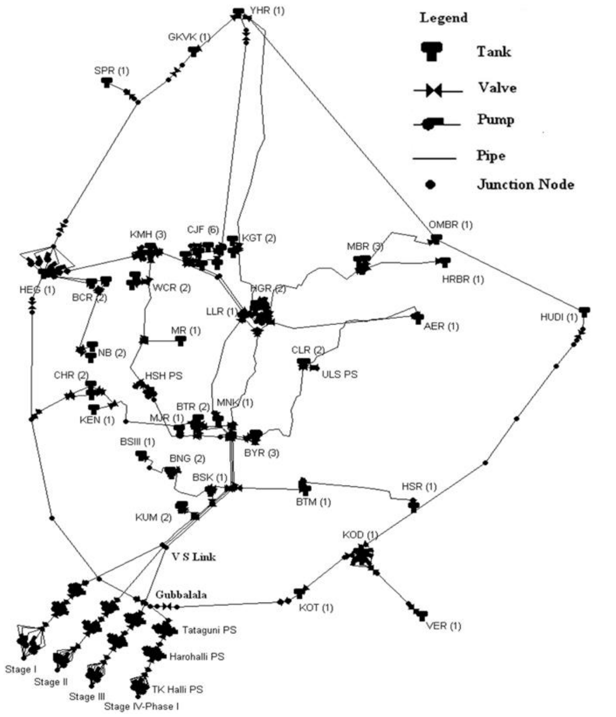

Fig. 1 Schematic of Bangalore Water Supply inflow system (55 GLRs)

Here, **s** is the activation (transfer) function, $\boldsymbol{N}$ is the number of neurons, $\boldsymbol{w_{ij}}$ the weights, $\boldsymbol{x_j}$ inputs to the neuron, and θ_i the threshold term of the hidden neuron $\boldsymbol{i}$. ANNs are well suited for environmental modelling as they are nonlinear in nature, relatively insensitive to data noise, and can perform well even when limited data are available (Maier and Dandy, 2001). In this study, resilient back propagation based ANN was used for modelling WDS with different combination of selection methodology and optimization techniques.

METHODOLOGY

In this study the authors have chosen n number of pipes randomly from the WDS and the flow in the pipes are used for training the ANN model to predict the water levels in the GLRs. EPANET is used for generating the training data set for ANN model. In this model, the hyperbolic tangent function is used as the transfer function **s**. And the results are compared to that of classical modelling technique.

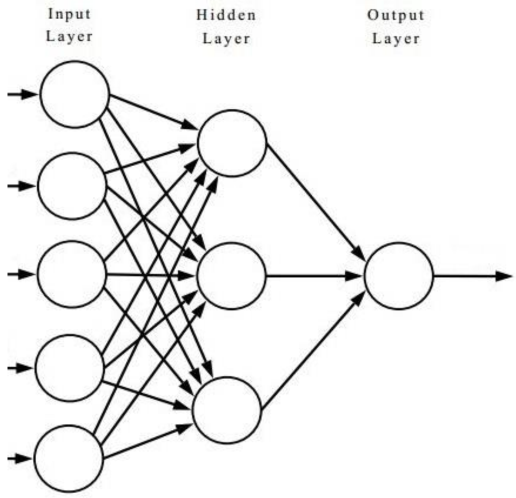

Fig. 2 Schematic of an ANN (MLP)

CLASSICAL MODELING METHOD

Traditional modelling techniques in WDS are based on the principles of conservation of mass and energy. According the theorem of conservation of mass, the fluid mass that enters the pipe will be equal to the fluid mass leaving the pipe (fluid is neither created nor destroyed in hydraulic systems). This can be represented as

$$\sum_{i=1}^{N_i} Q_j - D_i = 0 \quad (2)$$

where Q_j = flow rate in pipe j, N_i = number of pipes connected to node i, D_i = demand at node i.

Conservation of energy states that the difference in energy at any two points connected in a network is equal to the energy gained from pump and energy lost in pipes and fittings that occur in the path between them, i.e.

$$\sum_{j=1}^{Nk}(\mathit{\Delta} H_j + \mathit{\Delta} M_j - E_j) - \mathit{\Delta} E_k = 0 \quad (3)$$

where, $\mathit{\Delta} H_j$ = head loss in pipe j, $\mathit{\Delta} M_j$ = minor loss in pipe j, Ej = energy added by a pump if there is a pump in pipe j, N_k = the number of pipes within the loop k, $\mathit{\Delta} E_k$ = difference in water surface elevation between the two constant head boundaries if k represents a path between the two boundaries. For calculating the head-loss along a pipe, Hazen-Williams head loss equation is utilized:

$$h_{loss(f)} = \left(\frac{10.71 l_p}{C^{1.852}\, D_p^{4.87}}\right) Q_p^{1.852} \quad (4)$$

EPANET (Rossman, 2000), is an open-source software developed by EPA for network analysis. It uses Todini-Pilati (1987) method for calculating the pressure at every node, flow in pipes, tank levels etc. given boundary conditions. In this study, EPANET model for Bangalore Inflow network was built using the data given by Bangalore Water Supply and Sewerage Board (BWSSB). And the data from the EPANET model simulation was used for training the ANN model. Details of the EPANET model (GLR locations, Valves etc.) can be inferred from Figure 1.

ANN BASED MODELING

A total of 48 sets of training data are available; the training data is generated from EPANET simulation of Bangalore inflow network. The inputs imposed are the pipe flow, and it is trained to predict the water levels in the 55 GLRs around the city. The MLP network in this study has three hidden layers with 23, 26 and 28 neurons respectively. Genetic algorithm was used to select the architecture of the ANN model (Arifovic and Gencay, 2001). Two different training algorithms are used in this study: Gradient descent method (Battiti,1992) and Resilient back-propagation method (Riedmiller and Braun,1993).

GRADIENT DESCENT METHOD

The gradient descent algorithm is used to minimize the error (difference between the target and actual values), through the manipulation of a weight vector w. The error function should be a linear combination of the weight vector (*w*) and an input vector x.

$$w_{ij}(t+1) = w_{ij}(t) + \eta g\left(w_{ij}(t)\right) \tag{5}$$

Here, g is the error function or the cost function in this case, and η is known as the step-size parameter, and affects the rate of convergence of the algorithm. If the step size is too small, the algorithm will take a long time to converge. If the step size is too large the algorithm might miss the optimal solution.

RESILIENT BACK-PROPAGATION METHOD

Resilient back-propagation algorithm is similar to the regular back propagation algorithm, but it is faster than the latter since it does not require the value of the learning rate. In this method, an individual update value is designated to each weight, which solely determines the size of the weight-update. This adaptive update-value evolves during the learning process based on its local sight on the error function, according to a specific learning rule. Every time the partial derivative of the corresponding weight w_{ij} changes its sign, which indicates that the last update was too big and the algorithm has jumped over a local minimum, then the update-value Δ_{ij} is decreased by the factor η-. If the derivative retains its sign, the update-value is slightly increased in order to accelerate convergence in shallow regions. Once the update-value for each weight is computed, the weight-update itself follows a very simple rule: if the derivative is positive (increasing error), the weight is decreased by its update-value and if the derivative is negative, the update-value is added to the weight.

PIPE SELECTION ALGORITHMS

In this study, the authors have used four different methods for selecting the flow data to be used in training the ANN model. They are i) Random method ii) Principal Component Analysis (PCA) iii) Pearson's Correlation Coefficient (PCC) iv) Shannon Entropy (SE).

In the Random method, a random set of 20 pipe flow data are selected from the 516 pipes in the system. In the PCA method, initially a random set of 200 pipes are selected and then PCA analysis is carried out on these 200 pipes to form 20 linearly uncorrelated principal components. Then the strongest contributing pipe to each of the 20 principal components with the highest coefficient in the linear combination is selected. The data from these 20 pipes are used in further modelling.

In the PCC method, initially both the input and the output time series are normalized. Then the Pearson's correlation coefficient between all 55 GLR water level time series and all 200 pipe flow time series is calculated. Pearson's correlation coefficient ρ, between two random variables X and Y:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \, \sigma_Y} \tag{6}$$

Where, *Cov(X,Y)* is the covariance of *X* and *Y*, and *s* is the standard deviation of the random variable. P value is summed for all the GLRs in the network and 20 pipes with the highest sum of squared Pearson's coefficient are selected.

Shannon Entropy (SE) for a random variable s is defined as:

$$H(S) = \int P(s) \log\left(P(s)\right) ds \tag{7}$$

Where, *P(s)* is the probability density function of *s*. The theory behind this method of pipe selection is that the least predictable pipe flows will provide the largest amount of information to the artificial neural network. Entropy is a measure of the unpredictability of the variable, so the 20 most unpredictable pipe flows were selected with this method.

Each of these methods was iterated 500 times, selecting 200 random pipes each time, and whittling it down to 20. The final selected set for any of the four methods consists of the 20 (distinct) most commonly selected pipes.

PERFORMANCE INDICATORS

Three different performance indicators are used in this study. They are i) Mean Absolute Percentage Error (MAPE) ii) Root Mean Squared Error (RMSE) and iii) Normalized Root Mean Square Error (NRMSE). All the above mentioned indices are calculated for the GLR water levels (the difference between simulated and ANN output water level). Here, NRMSE is the normalized value of RMSE,

$$NRMSE = \frac{RMSE}{\max(w_n) - \min(w_n)} \tag{8}$$

Here, w_n denotes the water level in GLR n.

RESULTS AND DISCUSSION

It was found that the ANN based modelling of Bangalore inflow network was able to reproduce the EPANET model results with an acceptable range of error. The error rates for the neural networks derived from the four pipe selection methods is found in Table 2. For the ANN model training, resilient back-propagation is more accurate than gradient descent. Earlier literature on ANN network training comply with the above observation for modelling environmental systems.

The ANN models derived from the four pipe selection methods were tested against an ANN model derived from the already existing flow meters in the WDS and it was found that the models are performing well in comparison. For the existing flow meter locations, the model obtains an RMSE of 0.6859, NRMSE of 0.0722, and MAPE of 14.23%.

PCA method for pipe selection performs better than the other methods. All three pipe selection methods (PCA, SE, and PCC) consistently produce more accurate neural network models than randomly selecting pipes. Furthermore, it appears that these pipe selection methods produce more accurate neural networks than simply using all 58 of the sensor locations that are currently in the actual network. Thus, our experiments lead us to believe that real-life sensor placement could benefit from our pipe selection model.

Table 2: Comparison of Resilient Back Propagation and Gradient Descent algorithms for ANN Model training

	Resilient Back-Propagation		
	NRMSE	RMSE	MAPE
Random	0.0783	0.7443	15.99%
PCA	0.0544	0.5172	11.01%
PCC	0.0523	0.4965	11.23%
SE	0.0538	0.5110	11.49%
	Gradient Descent		
Random	0.1352	1.2844	45.53%
PCA	0.1190	1.1307	38.29%
PCC	0.1388	1.3185	48.35%
SE	0.1911	1.8157	74.28%

In addition, the time taken for prediction is quite fast. Taking the total runtime for prediction for the 150 test data and averaging, we find that the average runtime for predicting all 55 tank levels in a given time period is 0.1 milliseconds. In addition, both resilient back-propagation and gradient descent train the full model in under ten seconds. This improvement in speed could justify use of an ANN model in place of a traditional state estimation model. In cases of very complex networks, the state estimation may in fact be impossible without the computational efficiency of the ANN model. Even in simple water networks, the amount of data generated is massive in a real-time SCADA system. Dealing with this amount of real-time sensor data requires a fast-responding model. The speed of the ANN also allows constant retraining of the neural network in a real-time system. Retraining a system daily, or perhaps weekly, could help reduce drift in prediction accuracy caused by new leaks, non-revenue water demand, variations in temperature, and various other changes.

One other benefit of using an ANN model is that there are relatively few parameters to optimize. The genetic algorithm optimizes the number of hidden layers as well as the number of nodes in the hidden layers, and there are a number of possible methods for pipe selection. The other parameters are the training algorithm and the transfer function for the hidden layers. This is much smaller than the plethora of parameters that must be initialized in an EPANET model.

Multiple training functions were explored. However, as speed of training was considered to be of high importance, the only two functions that achieved reasonable accuracy in a short amount of time were the resilient back propagation and the gradient descent methods.

Despite its promise, there are a few observations to make about the limitations of the system. Firstly, PCA and PCC are limited because they only measure linear variation. PCA reflects which pipes are most important when linear combinations of pipes are chosen as principal components. However, there may be a pipe which gives a nonlinear contribution to the flow readings. PCC likewise measures the linear correlation between the pipe flows and the GLR levels. Replacing PCC with a nonlinear measure of correlation has the potential to drastically reduce error. One possible option would be to run a nonlinear regression on the dataset, then select the pipes with the largest absolute value coefficients. Nonlinear regression is of course a challenging task in itself, and many papers could be devoted to the use of nonlinear regression as the entire water level prediction model. Another option might be to use a metric typically employed in time series analysis to measure causality, such as Granger causality. Such a metric could more effectively select which pipes truly drive variations in tank levels, rather than selecting pipes that seem to be correlated with the tank levels. Graph theoretical methods may also be very useful in pipe selection. From Figure.4 it is observed that the developed ANN model works well, since the NRMSE plot for the model with actual sensor location data is similar to Figure.3.

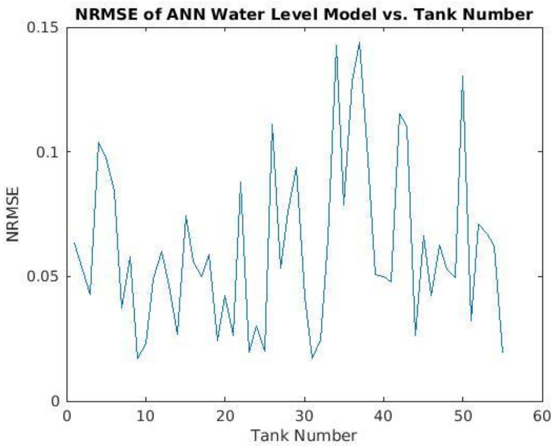

Fig. 3 NRMSE for GLRs water level

It is apparent that while the neural network model achieves relatively low error on most of the 55 GLRs, the average error rate is dragged up by its performance on a handful of GLRs. The GLRs with the top five NRMSEs are 34, 36, 42, 50, and 26. The tanks with the top five highest standard deviation in tank level are 37, 33, 34, 4, and 53. The tanks with the top five highest average levels are 35, 12, 51, 41, and 5. The tanks with the top five tank level ranges (max – min) are 35, 42, 33, 53, and 4. Finally, the tanks with the largest Shannon entropy are 35, 12, 51, 33, and 41. No clear pattern emerges, though it appears that the tanks with the highest error have a combination of large range and large standard deviation. Strangely, entropy of the tank seems to have little effect on the error rate. The simplicity of these metrics does not compare with the complexity of the artificial neural network, so maybe there is no way of knowing why the neural network does not perform well on certain tanks. Perhaps the best way to mitigate these areas of poor performance is to train multiple neural networks and use them for different subsets of the tanks. Multiple neural networks are likely to perform more effectively on different subsets of the tanks.

Of course, the neural network models must be validated against field data in order to truly trust their predictive powers. However, collection of field data is quite expensive and time-consuming. In addition, the neural network is useful precisely because we often cannot obtain real-time field data. Thus further empirical validation of the model will be quite challenging.

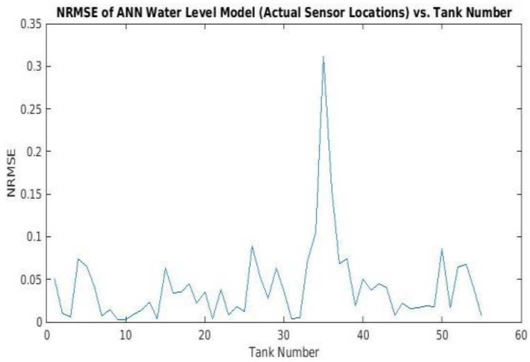

Fig. 4 NRMSE for GLRs water level (ANN using actual flow meter)

In the end, this neural network could be a single piece in a larger smart water distribution system designed to perform demand prediction, leak localization, online adjustment of flows, and more.

CONCLUSIONS

This work aims at developing an ANN based software tool for WDS modelling. In this study, two different training algorithms and 4 different pipe selection methods are being compared. It was found that the resilient back-propagation method for ANN training performs better than gradient descent method. Also, it was observed that PCA method is efficient in pipe selection for model building and it performs better than PCC and SE methods. It is to be noted that this approach can also be extended for finding out the optimal location of flow meters for WDS modelling in real time. This model was validated using the already existing flow meters locations in the system and it was found to perform well. Hence, this ANN based model can be used to replace the classical model for Bangalore inflow network for real time state estimation.

REFERENCES

Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. International Journal of Geographical Information Science, 22(9): 943-963.

Arifovic, J., & Gencay, R. (2001). Using genetic algorithms to select architecture of a feedforward artificial neural network. Physica A: Statistical mechanics and its applications, 289(3): 574-594.

Battiti, R. (1992). First-and second-order methods for learning: between steepest descent and Newton's method. Neural computation, 4(2): 141-166.

Broad, D. R., Dandy, G. C., Maier, H. R., & Nixon, J. B. (2006, July). Improving metamodel-basd optimization of water distribution systems with local search. In 2006 IEEE International Conference on Evolutionary Computation: 710-717.

Coulibaly, P., Anctil, F., Aravena, R., & Bobée, B. (2001). Artificial neural network modeling of water table depth fluctuations. Water Resources Research, 37(4): 885-896.

Gonçalves, F. V., Costa, L. H., & Ramos, H. M. (2011). Best economical hybrid energy solution: Model development and case study of a WDS in Portugal. Energy policy, 39(6): 3361-3369.

Jain, A., & Ormsbee, L. E. (2002). Short-term water demand forecast modeling techniques—Conventional methods versus AI. Journal (American Water Works Association), 94(7):64-72.

Maier, H. R., & Dandy, G. C. (2001). Neural network based modelling of environmental variables: a systematic approach. Mathematical and Computer Modelling, 33(6):669-682.

Manohar, U., & Mohan Kumar, M. S. (2013). Modeling equitable distribution of water: Dynamic inversion-based controller approach. Journal of Water Resources Planning and Management, 140(5): 607-619.

Milot, J., Rodriguez, M. J., & Sérodes, J. B. (2002). Contribution of neural networks for modeling trihalomethanes occurrence in drinking water. Journal of water resources planning and management, 128(5): 370-376.

Park, D. C., El-Sharkawi, M. A., Marks, R. J., Atlas, L. E., & Damborg, M. J. (1991). Electric load forecasting using an artificial neural network. IEEE transactions on Power Systems, 6(2): 442-449.

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster back propagation learning: The RPROP algorithm. In Neural Networks: 586-591.

Rodriguez, M. J., West, J. R., Powell, J., & Sérodes, J. B. (1997). Application of two approaches to model chlorine residuals in Severn Trent Water Ltd (STW) distribution systems. Water Science and Technology, 36(5): 317-324.

Romano, M., & Kapelan, Z. (2014a). Adaptive water demand forecasting for near real-time management of smart water distribution systems. Environmental Modelling & Software, 60: 265-276.

Romano, M., Kapelan, Z., & Savić, D. A. (2012 b). Automated detection of pipe bursts and other events in water distribution systems. Journal of Water Resources Planning and Management, 140(4): 457-467.

Rossman, L. A. (2000). EPANET 2: Users manual.

Todini, E., and Pilati, S. (1987). A gradient algorithm for the analysis of pipe networks. Computer Applications in Water Supply: Volume 1 - Systems Analysis and Simulation, B. Coulbeck and C.H. Orr, eds., John Wiley and Sons: 1- 20.