A Novel Computationally Efficient Asset Management Framework Based on Monitoring Data from Water Distribution Networks

Ahmad Momeni[1]; Kalyan R. Piratla[2]; and Kapil Chalil Madathil[3]

[1]Ph.D. Student, Glenn Dept. of Civil and Environmental Engineering, Clemson Univ., Clemson, SC, USA. E-mail: amomeni@clemson.edu
[2]Liles Associate Professor, Glenn Dept. of Civil and Environmental Engineering, Clemson Univ., Clemson, SC, USA. E-mail: kpiratl@clemson.edu
[3]Dean's Assistant Professor, Glenn Dept. of Civil and Environmental Engineering, Clemson Univ., Clemson, SC, USA. E-mail: kmadath@clemson.edu

ABSTRACT

Drinking water infrastructure in the U.S. is in a deteriorated state needing immediate intervention that is sustainable. Although many technologies are being developed to inspect buried pipeline assets, they are still expensive and human-dependent to use for comprehensive condition assessment and prioritization of the most critical assets for immediate rehabilitation and replacement planning. This paper presents a novel system-level condition assessment framework where monitoring data from distribution infrastructure is leveraged to predict the condition of assets using evolutionary optimization and machine learning algorithms. Pipeline roughness values and effective hydraulic diameters (given the possibility of graphitization/corrosion) are two parameters that would reveal their overall condition, and therefore these two parameters will be used to demonstrate the framework presented in this paper. In this respect, a modified benchmark water distribution network is used to represent an ageing, deteriorated network by randomly reducing effective pipe diameters and roughness coefficient values. Subsequently, a novel reverse engineering optimization method is leveraged to minimize the mean square errors of operational parameters (e.g., pressure and flow) via both predicted (through optimization) and modeled data obtained from a given set of monitoring stations. Roughness values and effective hydraulic diameters are the decision variables in this optimization framework that are to be predicted. EPANET 2.0 software is used for modeling the water distribution network performance in this study. Faster convergence is achieved through fine-tuning of genetic algorithm properties. Specifically, the computational efficiency and prediction accuracy benefits derived from appropriately narrowing down on the upper and lower bounds of the decision variables through multiple runs of the optimization process will be demonstrated in this paper. The framework proposed in this study offers great analytical capability to predict the condition of various assets in a water distribution network without having to undertake expensive inspection procedures.

KEYWORDS: Genetic Algorithms, Effective Hydraulic Diameters, Roughness Coefficients, EPANET, Water Distribution Network

INTRODUCTION

Water distribution systems (WDSs) are as complex as they are critical to societies. Timely and proper monitoring, rehabilitation and replacement policies of the WDS assets play an essential role in maintaining the reliability targets of water utilities within budgetary constraints (Mazumder et al. 2018; Pietrucha-Urbanik and Tchórzewska-Cieślak 2018). The conventional methods of WDS asset management are expensive and time taking for large systems due to operator dependency. They are also based on human-dependent manual methods that may be prone to errors (Frangopol and Liu 2007; Momeni and Piratla 2019; van Riel et al. 2016). The methodologies presented in the literature, despite including numerical algorithms, are rather intertwined with municipal intervention and on-site inspections, thus likely making them more time-consuming (Bonthuys et al. 2019; Chen et al. 2017). Thus, a cyber-monitoring method has been presented in the literature that deals with reverse engineering practices to predict the conditions of a deteriorated pipeline using hydraulic monitoring data (Momeni and Piratla 2019). Pipeline roughness coefficients were predicted using hydraulic monitoring data derived from a modeled WDS that is deteriorated (Piratla and Momeni 2019). This study extends the previous work by predicting the effective hydraulic diameters of pipelines in addition to the roughness coefficients based on hydraulic monitoring data. The prediction accuracy is evaluated using numerical metrics such as mean absolute error (MAE) and mean absolute percentage error (MAPE). The optimization algorithm employed in this study is also improved compared to the previous study. While the previous studies (Momeni and Piratla 2019; Piratla and Momeni 2019) attempted to present the fundamental framework and demonstrate it for predicting pipeline roughness values, this paper specifically offers the following advancements in comparison: (i) the betterment of the optimization algorithm through tapering function of the boundaries, (ii) adding a new parameter (effective hydraulic diameters) for prediction, and (iii) demonstration on a new WDS network.

METHODOLOGY

In this study, in order to assess the conditions of a deteriorated pipeline, actual and predicted values of hydraulic parameters (herein roughness coefficient and effective hydraulic diameters) are measured and compared using reverse engineering to evaluate their proximity and accuracy. GoYang network (Kim et al. 1994) depicted in Figure 1 is modified to characterize an ageing WDS network through reduction in pipe roughness (C) values as well as effective internal pipe diameters. Flow and pressure monitoring meters are quasi-randomly placed in GoYang modified network to be a representation of real-time data collection. Five pressure monitoring locations and seven flow monitoring locations are selected across the network. Base demands are varied with time to represent the alterations in system behavior through time. Assuming there is no failure in the network such as pipe failure or pump outage, nodal base demands are the representatives of inputs whereas pressure heads and flow rates are the outputs. The physics-based input-output relationship is leveraged to predict deteriorated WDS characteristics in this study. A total of 200 scenarios of demand variation are used in the optimization algorithm to characterize the dynamic behavior of the WDS (Piratla and Momeni, 2019). Base nodal demands in the deteriorated network are varied randomly within ±20% of the original GoYang network (Piratla and Momeni 2019). Subsequently, an optimization framework is developed in MATLAB by incorporating EPANET 2.0 software simulator to account for the hydraulic simulations (Piratla and Momeni 2019). In each of these scenarios, nodal pressure heads and pipeline flow values at the 12 monitoring locations are generated for the randomly varied nodal demands. Ultimately, the optimization platform is created using genetic algorithm tool by assuming the decision variables are the pipe roughness coefficients and effective hydraulic diameters, thus minimizing the absolute difference, quantified through mean squared error (Allen, 1971), between the predicted and actual (i.e., synthetic) values for pressure heads and flow rates utilizing the 200 demand scenarios.

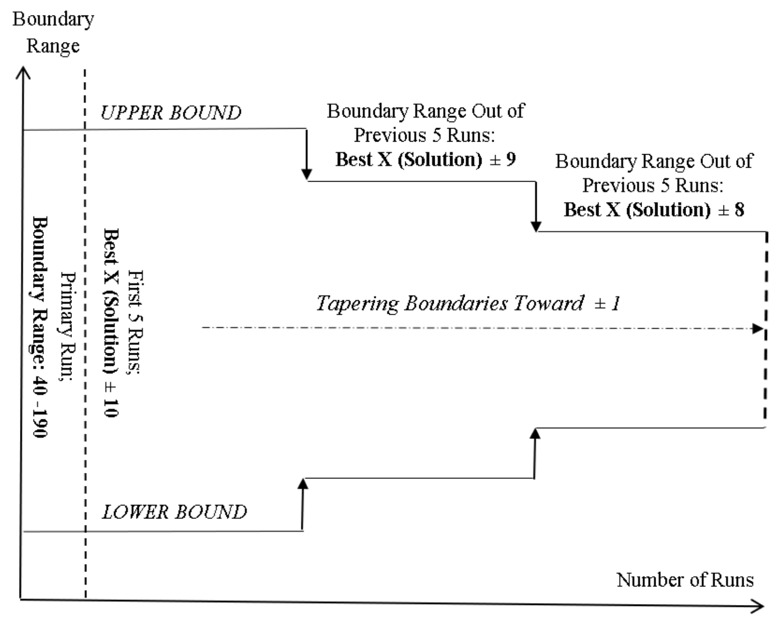

Figure 1. Schematic illustration of the optimized boundary setting

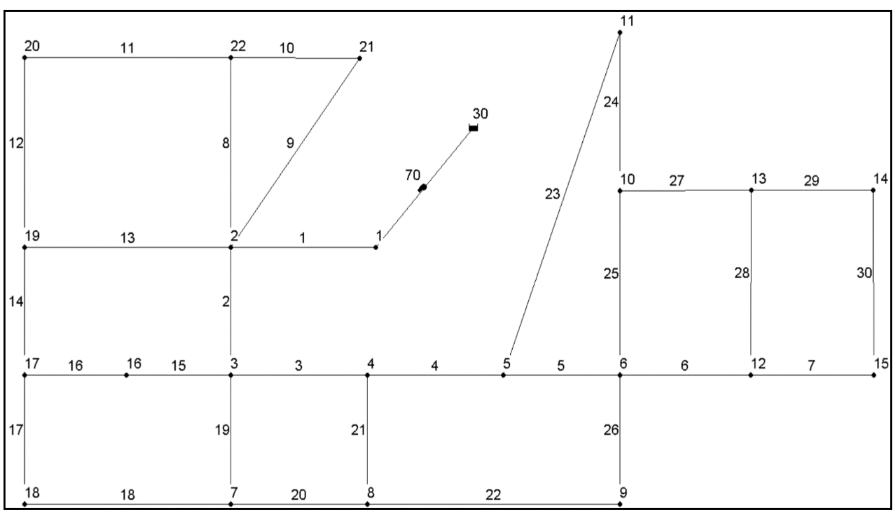

Figure 2. GoYang Network to represent a deteriorated network for the framework

To optimize the convergence agility of the optimization process, an auxiliary function is added to the original optimization scheme to fine-tune the boundary conditions of the search span at each iteration. Particularly, the approach depicted in Figure 1 is leveraged to make the process of narrowing down more intelligent by running the optimization process five times at each set of chosen boundary conditions. Then, the best objective function value (minimum of mean square errors) out of these 5 runs is selected and fed into the next set of boundary conditions while it is reduced within one increment of the whole upper and lower boundary ranges one at a time starting from ± 10 to ± 1 of the best solution of the previous five runs; therefore, the search span (the middle closed area in Figure 1) would be smartly reduced symmetrically from both top and bottom bounds. This procedure goes on until the minimum value is achieved (Only three steps are shown in Figure 1). Automatically, the entire primary range of randomization of 40 to 190 (i.e., roughness coefficient and effective pipe diameter values) will be back in play again and the whole process will be repeated without overlapping unless the preferred MSE value of zero, and thus MAE of zero, is obtained yet.

Table 1. GoYang original and modified (deteriorated) network parameters

Pipe index	Original network		Modified network	
	Pipe diameter (mm)	Pipe roughness (C)	Pipe diameter (mm)	Pipe roughness (C)
1	200.0	100.0	182.0	80.0
2	200.0	100.0	187.0	56.0
3	150.0	100.0	133.0	62.0
4	150.0	100.0	136.0	56.0
5	150.0	100.0	130.0	73.0
6	100.0	100.0	89.0	79.0
7	80.0	100.0	68.0	71.0
8	100.0	100.0	90.0	46.0
9	80.0	100.0	65.0	59.0
10	80.0	100.0	61.0	67.0
11	80.0	100.0	65.0	43.0
12	80.0	100.0	64.0	73.0
13	80.0	100.0	65.0	79.0
14	80.0	100.0	61.0	74.0
15	100.0	100.0	84.0	51.0
16	80.0	100.0	66.0	75.0
17	80.0	100.0	64.0	63.0
18	80.0	100.0	70.0	78.0
19	80.0	100.0	70.0	66.0
20	80.0	100.0	61.0	60.0
21	80.0	100.0	68.0	54.0
22	80.0	100.0	64.0	76.0
23	80.0	100.0	69.0	80.0
24	80.0	100.0	68.0	46.0
25	80.0	100.0	60.0	55.0
26	80.0	100.0	68.0	75.0
27	80.0	100.0	62.0	69.0
28	80.0	100.0	61.0	74.0
29	80.0	100.0	67.0	76.0
30	80.0	100.0	68.0	78.0

Iteratively, it is evident that the near-optimal solution can be achieved in less time and at finer accuracy rate using this approach. In this case, through trial and error, less than a hundred iterations with increasing number of generations have been conducted. Elitism is inherently embedded in the body of the genetic algorithm framework in that the function “gamultiobj” in MATLAB leverages an elitist form of a controlled evolutionary algorithm (a variant of NSGA-II). A controlled elitist evolutionary algorithm also contributes to the increase in diversity of population through elite individuals even if they possess a lower fitness value. Subsequently. the mean absolute error (MAE) (Willmott and Matsuura 2005) and mean absolute percentage error

(MAPE) (de Myttenaere et al. 2016) play an integral part in determining both the accuracy of the procedure and the correlation between actual and predicted values.

CASE STUDY

A benchmark WDS – modified GoYang network (Kim et al., 1994) – is chosen to characterize a deteriorated network by reducing both pipeline roughness coefficients and effective diameters. Majority of WDS pipelines in the real word are metallic and they get rougher with age leading to higher energy consumption and accelerated deterioration. Similarly, aged metallic pipelines could exhibit graphitization that may essentially reduce the effective hydraulic diameter. Figure 2 depicts the modified version of GoYang (Kim et al. 1994) network, which consists of one reservoir, one pump, 30 pipes and 22 nodes. First, the roughness coefficients of the original network are reduced to be in the range of 40 to 80 from the original value of 100 and then the pipe diameters are randomly reduced by a maximum of 20mm, as can be seen from Table 1. This reduction characterizes the deteriorated WDS network. A minimum pressure head of 5m has been maintained to avoid violating pressure constraints.

Table 1 demonstrates the characteristics of the original and deteriorated versions of GoYang network. As is shown, both diameters and roughness coefficients are reduced in the deteriorated version, since these parameters are on the wane as pipeline ages.

DEMONSTRATION

Formulation of the Optimization Framework

The proposed optimization algorithm is set out to determine the combined set of roughness coefficients and hydraulic diameters as decision variables. The objective function is to minimize the differences (mean square errors - MSE) between the predicted and actual pressure and flow values at the chosen monitoring locations for a total of 200 randomized demand scenarios. The actual pressure and flow values are obtained using the deteriorated WDS characteristics presented in Table 1 whereas the predicted values are obtained from the optimization algorithm where the reduced pipe diameters and roughness coefficients are the decision variables. Equation 1 below represents the calculation procedure for MSE between actual and predicted values (Piratla and Momeni 2019).

A. *Decision variables: {x1, x2, …., x60}* →where, x_1 is the roughness coefficient of pipe 1 up to x_{30} which is the roughness coefficient of pipe 30, and x_{31} is the diameter of pipe 1 up to x_{60} which is the diameter of pipe 30. The decision variables are constrained to vary between 40 and 80 for roughness values and between 60 and 190 for hydraulic diameters.

B. *Objective:* Minimize the following

$$\text{Minimum of } [(a_m - P2_m)^2 + (b_m - P6_m)^2 + (c_m - P15_m)^2 + (d_m - P18_m)^2 + (e_m - P22_m)^2]$$

for all m

$$+ \text{ Minimum of}[(f_m - F1_m)^2 + (g_m - F3_m)^2 + (h_m - F5_m)^2 + (i_m - F10_m)^2 + (j_m - F17_m)^2 \quad (1)$$

$$+(k_m - F24_m)^2 + (l_m - F30_m)^2]$$

for all m

Where, m is the simulation number (i.e., the scenario number ranging from 1 to 200); a_m, b_m, c_m, d_m, e_m, f_m, g_m, h_m, i_m, j_m, k_m are estimated pressures and flows during optimization; a_m is the pressure at node 2 in simulation m; b_m is the pressure at node 6 in simulation m, c_m is the pressure at node 15 in simulation m; d_m is the pressure at node 18 in simulation m; e_m is the pressure at node 22 in simulation m; f_m is the flow in pipe 1 in simulation m; g_m is the flow in pipe 3 in simulation m; h_m is the flow in pipe 5 in simulation m; i_m is the flow in pipe 10 in simulation m; j_m is the flow in pipe 17 in simulation m; k_m is the flow in pipe 24 in simulation m; l_m is the flow in pipe 30 in simulation m;

Where, $P2_m$, $P6_m$, $P15_m$, $P18_m$, $P22_m$, $F1_m$, $F3_m$, $F5_m$, $F10_m$, $F17_m$, $F24_m$, $F30_m$ are actual pressures and flows; $P2_m$ is the pressure at node 2 in simulation m, $P6_m$ is the pressure at node 6 in simulation m, $P15_m$ is the pressure at node 15 in simulation m; $P18_m$ is the pressure at node 18 in simulation m, $P22_m$ is the pressure at node 22 in simulation m; $F1_m$ is the flow in pipe 1 in simulation m; $F3_m$ is the flow in pipe 3 in simulation m; $F5_m$ is the flow in pipe 5 in simulation m; $F10_m$ is the flow in pipe 10 in simulation m; $F17_m$ is the flow in pipe 17 in simulation m; $F24_m$ is the flow in pipe 24 in simulation m; $F30_m$ is the flow in pipe 30 in simulation m;

C. *Constraint Function:* Although several parameters can end up contributing to the constraint function in the optimization framework, only pressure heads at all nodes have been considered to be the constraint that needs to be satisfied throughout the optimization process. Since the physical and operational values are rather low in the network, the minimum pressure head has been set to 5 meters.

Prediction Measures

As mentioned before, the actual and predicted values of roughness coefficients and diameters are assessed using mean absolute error (MAE) and mean absolute percentage error (MAPE). These metrics can be found in Equations 2 and 3 below that summarize the calculation of the MAE and MAPE for both predicted and actual results (Momeni and Piratla, 2019). Lower values of MAPE or MAE indicate greater accuracy of the prediction model.

$$Mean\ Absolute\ Error = \left(\frac{\sum_{i=1}^{n} abs\left(pr_i - sim_i\right)}{n} \right) \quad (2)$$

$$Mean\ Absolute\ Percentage\ Error = \left(\frac{\sum_{i=1}^{n} \frac{abs\left(pr_i - sim_i\right)}{sim_i}}{n} \right) * 100 \quad (3)$$

Where pr is the predicted value and sim is the simulated value, i is the associated node or link for a specific scenario as outputs. Also, n is the number of inputs, which is the number of decision variables.

Results

The novelty herein lies in the fact that multiple initial runs have been conducted to incorporate different initial random populations as well as fine-tuned boundary conditions into the optimization. The purpose briefly centers on evaluating the tapering nature of the auxiliary boundary-tuning function that helps both with the accuracy and speed of the optimization procedure. Table 2 demonstrates the optimization characteristics for all the runs and boundary conditions.

Table 2. Optimization Characteristics

Number of Generations	Number of Population Size	Boundary Tuning Function	Crossover Factor
500 (Ultimately)	135	Yes/Searching for optimal solution over a span of a difference of five between lower and upper bounds	0.85

Tables 3 and 4 show optimization outputs (optimal solution for predicted roughness coefficients and diameters respectively), where MSE has acted as the objective function in the optimization process for five pressure head stations and seven flow rate stations. After tapering boundaries from within ± 10 of the best solution of five repetitious runs to ± 1 and tuning the number of generations at each iteration, the best results are fine-tuned and produced.

Table 3. Output Correlation for Simulation-based Results for Actual and Predicted Roughness Coefficients

Pipe Number	Actual Values	Predicted Values	Pipe Number	Actual Values	Predicted Values	Prediction Accuracy	
						Mean Absolute Error	Mean Absolute Percentage Error
1	80.0	73.0	16	75.0	68.0		
2	56.0	58.0	17	63.0	51.0		
3	62.0	56.0	18	78.0	74.0		
4	56.0	49.0	19	66.0	57.0		
5	73.0	61.0	20	60.0	45.0		
6	79.0	73.0	21	54.0	56.0		
7	71.0	54.0	22	76.0	72.0		
8	46.0	47.0	23	80.0	71.0	7.2	11.01%
9	59.0	53.0	24	46.0	58.0		
10	67.0	50.0	25	55.0	54.0		
11	43.0	49.0	26	75.0	69.0		
12	73.0	53.0	27	69.0	76.0		
13	79.0	72.0	28	74.0	71.0		
14	74.0	74.0	29	76.0	74.0		
15	51.0	48.0	30	78.0	71.0		

In this regard, Tables 3 and 4 demonstrate that the predictions are accurately approximated by 7.2 in terms of mean absolute error (MAE) and 11.01% in terms of mean absolute percentage error (MAPE) for roughness coefficients and 7.0 for MAE and 9.95% for MAPE regarding effective hydraulic diameters. Evidently, Tables 3 and 4 demonstrate that pipe diameter predictions are more accurate than those of roughness coefficients – 9.95% to 11.01%. Also, by closely analyzing the results, there are little variations between actual and predicted results in some of the values of roughness and pipe diameters whereas there are more of variation between other predictions. This accuracy gap could likely fade away through increasing the number of generations and more diverse initial population runs as well as conducting sensitivity analysis to determine the efficient number and position of monitoring locations and number of actual scenarios. Also, Figure 3 depicts the performance behavior of fitness function over time (20 minutes per iteration), since iterative runs have been carried out to select the best set of feasible solution through elitism.

Table 4. Output correlation for simulation-based results for actual and predicted effective hydraulic diameters

Pipe Number	Actual Values	Predicted Values	Pipe Number	Actual Values	Predicted Values	Correlation Value: Mean Absolute Error	Correlation Value: Mean Absolute Percentage Error
1	182.0	186.0	16	66.0	77.0		
2	187.0	185.0	17	64.0	73.0		
3	133.0	140.0	18	70.0	80.0		
4	136.0	130.0	19	70.0	71.0		
5	130.0	130.0	20	61.0	62.0		
6	89.0	80.0	21	68.0	81.0		
7	68.0	75.0	22	64.0	72.0		
8	90.0	87.0	23	69.0	70.0	7.0	9.95%
9	65.0	71.0	24	68.0	67.0		
10	61.0	67.0	25	60.0	77.0		
11	65.0	79.0	26	68.0	80.0		
12	64.0	69.0	27	62.0	64.0		
13	65.0	80.0	28	61.0	66.0		
14	61.0	67.0	29	67.0	79.0		
15	84.0	70.0	30	68.0	72.0		

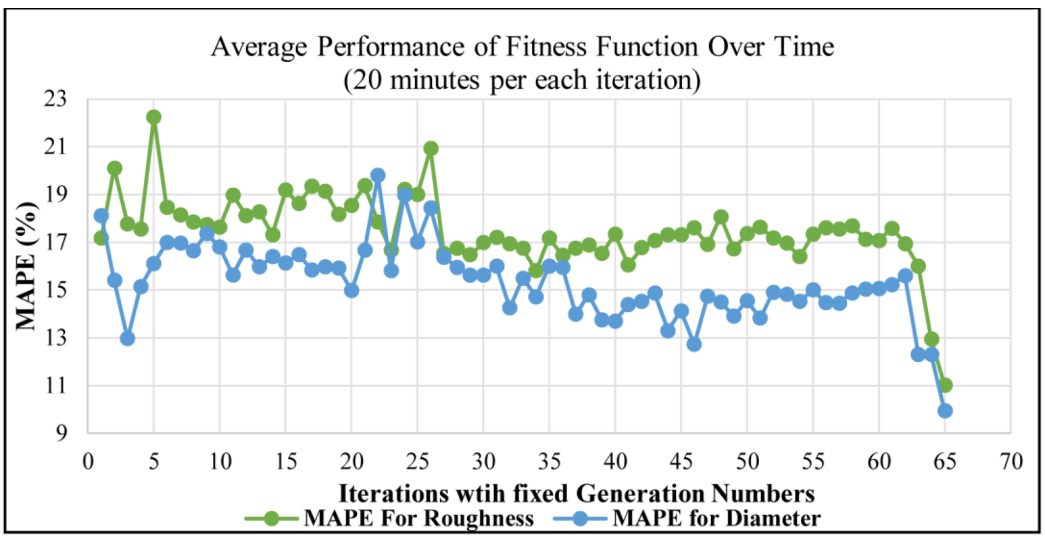

Figure 3. Average performance of fitness function through Elitism

CONCLUSIONS AND FUTURE WORK

First off, fine-tuning the algorithmic parameters in this study has contributed to the convergence trend to an acceptable extent. Also, the ad-hoc function adjusting the boundary conditions using a quasi-self-learning tapering feature helps narrow down the search span, thus tending to increase the probability of reaching the near-optimal solution. Furthermore, the proposed scheme tends to offer a promising future auxiliary add-on means to the prevalent manual inspection for higher accuracy and faster results. Also, not only will this method help with the time-consuming functionality of conventional asset management to a good extent, but it also tries to put forth a proportionally reliable data-driven scheme. In terms of future work, this study adds to the notion that the optimization framework attempts to work toward a promising scheme and that new ideas in terms of the lengthy duration of optimization can be improved using machine-learning approaches like neural networks. Also, there is an algorithmic limitation on the scale of the network used to be optimized, as hydraulic simulation through EPANET 2.0 is time-consuming and by increasing the search span of the optimization framework, it will not be time-efficient. So, further research should be geared toward circumventing the time-consuming EPANET 2.0 in MATLAB and making the scheme more accurate and faster, one way of which is to train neural networks to bypass the EPANET toolkit to make the framework both more applicable to larger networks and more condition assessment parameters, more convergent.

ACKNOWLEDGEMENT

This study is supported and funded by National Science Foundation (NSF) under Award Number of 1638321. Any opinions, findings and conclusions, or recommendations expressed in this material are those of author(s) and do not necessarily reflect the views of National Science Foundation.

REFERENCES

Allen, D. M. (1971). “Mean square error of prediction as a criterion for selecting variables.” *Technometrics*, 13(3), 469–475.

Bonthuys, G. J., van Dijk, M., and Cavazzini, G. (2019). “Leveraging water infrastructure asset management for energy recovery and leakage reduction.” *Sustainable Cities and Society*, 46.

Chen, T. Y. J., Beekman, J. A., and Guikema, S. D. (2017). “Drinking Water Distribution Systems Asset Management: Statistical Modelling of Pipe Breaks.” *Pipelines 2017: Condition Assessment, Surveying, and Geomatics - Proceedings of Sessions of the Pipelines 2017 Conference*, 173–186.

Frangopol, D. M., and Liu, M. (2007). “Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost.” *Structure and Infrastructure Engineering*, 3(1), 29–41.

Kim, J. H., Kim, T. G., Kim, J. H., and Yoon, Y. N. (1994). “A study on the pipe network system design using non-linear programming.” *J. Korean Water Resour. Assoc*, 27(4), 59–67.

Mazumder, R. K., Salman, A. M., Li, Y., and Yu, X. (2018). “Performance Evaluation of Water Distribution Systems and Asset Management.” *Journal of Infrastructure Systems*, 24(3), 03118001.

Momeni, A., and Piratla, K. R. (2019). “A Novel Cyber-Monitoring Based Asset Management Scheme For Water Distribution Networks Through Fine-Tuning Genetic Algorithm

Parameters." *International No-Dig 2019 37th International Conference and Exhibition; Florence, Italy 30th September – 2nd October 2019*.
de Myttenaere, A., Golden, B., Le Grand, B., and Rossi, F. (2016). "Mean Absolute Percentage Error for regression models." *Neurocomputing*, 192, 38–48.
Pietrucha-Urbanik, K., and Tchórzewska-Cieślak, B. (2018). "Approaches to failure risk analysis of the water distribution network with regard to the safety of consumers." *Water (Switzerland)*, 10(11).
Piratla, K. R., and Momeni, A. (2019). "A Novel Water Pipeline Asset Management Scheme Using Hydraulic Monitoring Data." *Pipelines 2019: Multidisciplinary Topics, Utility Engineering, and Surveying*, 190–198.
van Riel, W., van Bueren, E., Langeveld, J., Herder, P., and Clemens, F. (2016). "Decision-making for sewer asset management: Theory and practice." *Urban Water Journal*, 13(1), 57–68.
Willmott, C. J., and Matsuura, K. (2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance." *Climate Research*, 30(1), 79–82.