International No-Dig 2019
37th International Conference and Exhibition

Florence, Italy
30th September – 2nd October 2019

Paper Ref # 2326

A NOVEL CYBER-MONITORING BASED ASSET MANAGEMENT SCHEME FOR WATER DISTRIBUTION NETWORKS THROUGH FINE-TUNING GENETIC ALGORITHM PARAMETERS

Ahmad Momeni[1], and Kalyan R. Piratla[2]

[1] PhD Student, Glenn Department of Civil and Environmental Engineering, Clemson University, Clemson, SC USA
[2] S.E. Liles, Jr. Distinguished Associate Professor, Glenn Department of Civil and Environmental Engineering, Clemson University, and Clemson, SC USA

ABSTRACT:

There has been an urgent need to monitor, model, and tackle the extreme deterioration of water pipeline systems in the United States and other parts of the world to the quickest and best extent. Since manual pipeline inspection takes excessive costs, effort and time to consider most uncertainties, a novel and swift scheme is required to engage in asset management, which should ensure the accuracy as well as the quickness of the inspection at minimum effort and cost. This paper presents and validates a framework to predict the condition of critical pipeline assets using water distribution monitoring data. A benchmark water distribution network is first operationally optimized, and its pipe roughness values and diameters are subsequently reduced in a random manner in order to create a representation of an old and deteriorated water distribution network. The operational data (i.e., pipe flow and pressure) for such deteriorated network as obtained through simulations using EPANET 2.0 is leveraged to predict the pipe roughness and effective hydraulic diameter values. Evolutionary optimization algorithms are used to predict the roughness and diameter values. The novelty of this study entails: (i) employing modified genetic algorithms in MATLAB interface where multiple attempts are made towards finding the optimal solution by running on several distinct sets of initial population in conjunction with fine-tuning of mutation and crossover parameters; (ii) including an ad hoc function to cut down on the temporal expensiveness of the convergence; and (iii) assessing the validity and accuracy of the simulation-based optimization framework using mean absolute error (MAE) and mean absolute percentage error (MAPE). Successful outcomes of this study offer a great potential in predicting condition of critical water infrastructure assets based on the operational monitoring data that is increasingly being collected in the recent times.

1. INTRODUCTION

Asset management in water distribution networks has always been entwined with labor-intensive inspection as well as limited budget allocation. Particularly, conventional methods as to how data collection is carried out along with time-consuming nature of classic manual inspection lend themselves to an obstacle to the speed and accuracy of asset management in water distribution networks (Bastian et al. 2019; Mutikanga et al. 2012; Newton and Christian 2015). To put the criticality of novel asset management into perspective, in the United States, the annual cost replacement of pipelines through manual inspection is approximated at over $2.5 billion for a rough estimation of 7,000 km of pipeline (US EPA National Center for Environmental Assessment and Shaw 2007). More importantly, considering uncertainties like roughness and hydraulic diameters in aged pipelines, shedding light on how these parameters behave through time has turned out to be a challenge in this field (Meirelles et al.

2017; Sivakumar et al. 2016). This paper aims at presenting a novel data-driven SCADA-based scheme for predicting the condition of pipeline assets. Specifically, being tough to be delineated and monitored, pipeline roughness is deemed as a critical parameter whose behavior contributes to the overall wellbeing of the water pipeline system. More particularly, owing to internal corrosion and biofilm growth as well as nitrification that results in scaling, metal pipelines are more prone to higher deterioration and poor quality along with tougher delineation and monitoring capability as they age, so this adds more difficulty to how a pipeline may behave through time (Liu et al. 2016). Moreover, a conventional utilization of data-driven platforms using metaheuristic approaches like genetic algorithms has been proven to be extremely time-consuming as it takes several repetitions through simulators, which adversely affect the time it takes to achieve the near-optimal solution (Altarabsheh et al. 2018). Also, there have been preliminary steps to addressing the criticality and importance of reverse engineering data-driven asset management using monitoring data in the literature (Piratla and Momeni 2019). However, this needs improvement in terms of how accurate the framework is in case of temporal aspects as well as its convergence agility to avoid local optima and thus making the framework more temporally and accurately intelligent. Therefore, this paper doubles down on a newfangled data-driven framework to predict pipeline roughness in a three-looped benchmark network using an ad hoc function which narrows down on the lower and upper bounds of the search span for decision variables in MATLAB that will both speed up the time-consuming optimal-solution convergence process and avoid the local optima. Also, in comparison to the contribution presented in the literature for cyber-monitoring asset management (Piratla and Momeni 2019), the current study specifically closes in on the fine-tuning of algorithmic parameters like crossover and mutation factors through longitudinal literature review and reverse engineering as well as the fixed number of monitoring stations for all decision variables in the optimization problem, consideration of MAE and MAPE measures, and the temporal and accuracy improvement in case we consider multiple initial population/runs. This will pave the way for coming up with a universal platform that can be validated and utilized under various circumstances for asset management considering high accuracy and speed.

2. METHODOLOGY

In this paper, reverse engineering is utilized in a MATLAB-based genetic algorithm framework to predict the roughness values using hydraulic monitoring data. A modified version of Hanoi network (Fujiwara and Khang 1990) shown in Figure 2 is developed to represent an actual aged water pipeline system by reduction in pipe roughness (C) values as well as effective internal pipe diameters, which is assumed to result from corrosion-related scaling through time. This modified version of the network known hereby as Alt#1 is used to characterize the condition assessment prediction platform (Piratla and Momeni 2019). Monitoring locations of flow rate and pressure heads are placed on Alt#1 network to characterize the real-time data acquisition. As per Figure 1, where there are 34 links and 31 nodes, pressure head monitoring stations are placed at eight nodes and flow monitoring stations in seven pipes. Admittedly, these monitoring stations would make possible the collection of pressure and flow data by assuming these locations are provided with smart monitoring devices in reliance to the consumption rate at any time scale. To characterize the real-world scenario, the water network behavior is meant to vary with time including nodal demands and the operability of water network components. The base demands at all the nodes are supposed to be considered the inputs and the outputs would include pressure heads and flow rates at the monitoring locations respectively, assuming there will be no failure or pipe breaks in the system. The principle idea is to leverage the inherent relationship between these input and output data, as measured through smart water meters and smart monitoring devices placed in the distribution system, to predict the pipeline roughness values. A total of 200 datasets of inputs (i.e., nodal demands) and corresponding pressure heads and flow rates at the predetermined nodes and links are produced to represent the synthetic monitoring data to capture the dynamics of the water distribution network (Piratla and Momeni 2019). The nodal demands are randomized within ±20% of the base nodal demands of the original Hanoi network (Piratla and Momeni 2019). Conventionally, EPANET 2.0 software simulator is coupled with MATLAB interface by using open-source EPANET 2.0 extension toolkit library to leverage the hydraulic simulations. Ultimately, considering the decision variables are the pipe roughness coefficients, a genetic-algorithm optimization framework is formed. This algorithm is meant to minimize the absolute difference (quantified through mean squared error – MSE) between the calculated and actual (i.e., synthetic) values for pressure heads and flow rates in the predetermined nodes and links mentioned above over the 200 demand scenarios. However, in this study, a newfangled idea is utilized to pare down the time-consuming convergence process of optimization in MATLAB. In this case, a function is added to the conventional genetic algorithm framework that narrows down the search span boundaries on each iteration of the optimization to make sure optimal solution can be achieved. It has shown that the probability required for obtaining the optimal solution rises significantly using adjusted boundary conditions.

Consequently, correlation between the predicted pipe roughness coefficients and the actual pipe roughness coefficients will assess the accuracy of the proposed approach using correlation function including mean absolute error (MAE) (Willmott and Matsuura 2005) and mean absolute percentage error (MAPE) (de Myttenaere et al. 2016). Figure 1 shortly represents the steps as to how the approach works all the way through the results.

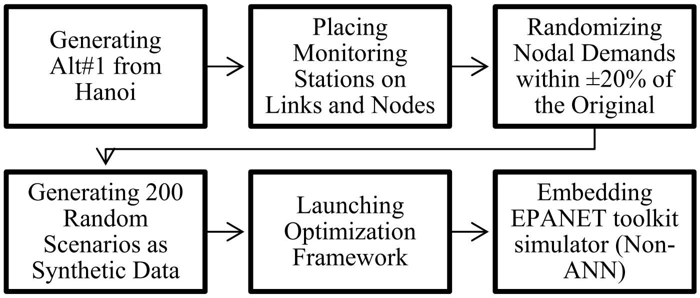

Figure 1. Flowchart for the proposed methodology

Table 1. Hanoi Original and Alternative Network Alterations (Piratla and Momeni 2019)

Pipe Index	**Original Network**		**Alt #1**	
	Pipe Diameter(mm)	Pipe Roughness (C)	Pipe Diameter (mm)	Pipe Roughness (C)
1	1066.8	130.0	1023.8	82.0
2	1524.0	130.0	1482.0	90.0
3	1066.8	130.0	1036.8	89.0
4	1066.8	130.0	1029.8	85.0
5	1066.8	130.0	1029.8	69.0
6	914.4	130.0	870.4	66.0
7	762.0	130.0	716.0	88.0
8	914.4	130.0	869.4	70.0
9	762.0	130.0	722.0	63.0
10	762.0	130.0	752.0	68.0
11	609.6	130.0	598.0	81.0
12	609.6	130.0	599.2	80.0
13	508.0	130.0	473.0	72.0
14	609.6	130.0	569.6	87.0
15	508.0	130.0	474.0	70.0
16	914.4	130.0	879.4	69.0
17	1066.8	130.0	1030.8	65.0
18	914.4	130.0	878.4	75.0
19	914.4	130.0	884.4	80.0
20	1066.8	130.0	1028.8	65.0
21	508.0	130.0	476.0	76.0
22	762.0	130.0	722.0	88.0
23	914.4	130.0	869.4	80.0
24	508.0	130.0	467.0	73.0
25	508.0	130.0	466.0	64.0
26	457.2	130.0	423.2	77.0
27	609.6	130.0	573.6	64.0
28	762.0	130.0	719.0	88.0
29	762.0	130.0	728.0	85.0
30	914.4	130.0	873.4	82.0
31	914.4	130.0	865.4	74.0
32	508.0	130.0	471.0	67.0
33	914.4	130.0	864.4	68.0
34	609.6	130.0	559.6	75.0

3. BENCHMARK PROPERTIES

In the Alt#1 network, the diameters are randomly reduced by values between 30 and 50 mm whereas the roughness values of each pipe are randomly reduced to be in the range of 60 and 90 from the original value of 130 (Piratla and Momeni 2019). Table 1 above shows the pipe diameters and roughness coefficient values in the original Hanoi network as well as the Alt#1 network. It is made sure that the minimum pressure required for the network is met during the reduction in pipe sizes and roughness coefficient values to well represent a real-world network. As shown in Figure 2 below, the Alt#1 network comprises of 31 nodes, 34 pipelines and one reservoir (Fujiwara and Khang 1990).

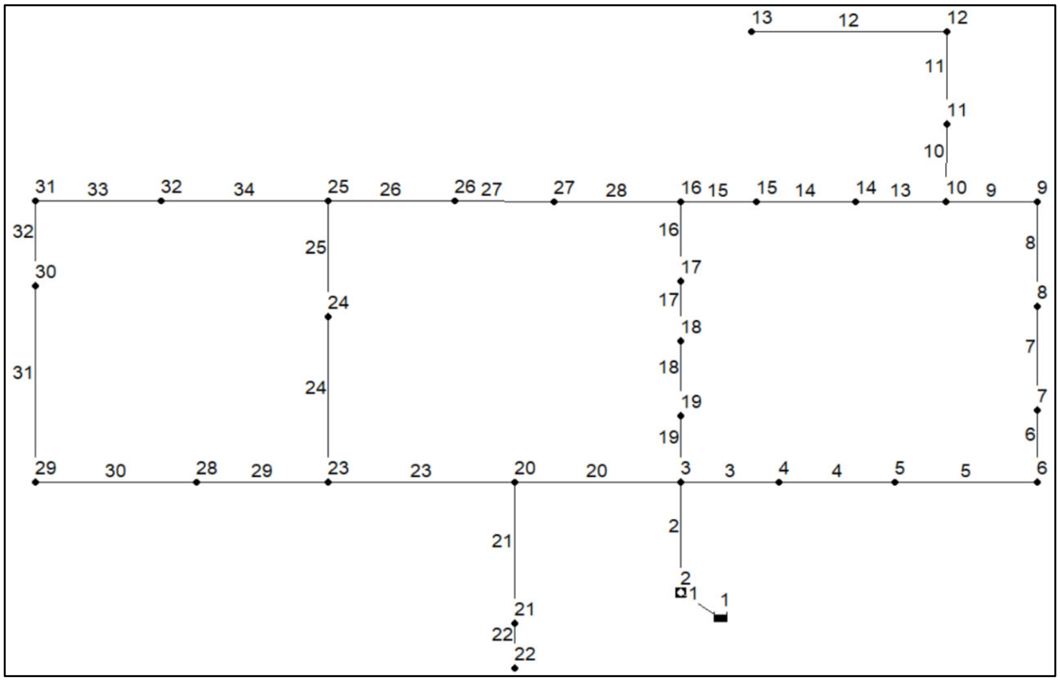

Figure 2. Alt#1 (Hanoi) Water Distribution Network Layout

3.1 Optimization Formulation and Calculations

The algorithm randomizes a set of roughness coefficients as decision variables into the 200 scenarios and minimizes the results by calculating the difference of predicted and actual roughness values. Then, hydraulic simulation is carried out using EPANET toolkit in MATLAB, and the outputs for the calculation of the objective function would be flow rate and pressure at the monitoring locations. Ultimately, the mean squared error (MSE) between the actual and predicted flow and pressure head has been formulated as the objective in Equation 1 below (Piratla and Momeni 2019).

A. **<u>Decision variables:</u>** $\{x1, x2, \ldots, x34\}$ →where, x1 is the roughness coefficient of pipe 1 and so on. The decision variables are constrained to vary between 50 and 130.
B. **<u>Objective:</u>** Minimize the following

Minimum of $[(a_i - P16_i)^2 + (a_i - P23_i)^2 + (a_i - P27_i)^2]$ for all i + Minimum of $[(d_i - F5_i)^2 + (e_i - F27_i)^2 + (f_i - F29_i)^2]$ for all i [1]

Where, i is the simulation number (i.e., the scenario number ranging from 1 to 200); a_i, b_i, c_i, d_i, e_i, f_i are estimated pressures and flows during optimization; a_i is the pressure at node 16 in simulation i; b_i is the pressure at node 23 in simulation i, c_i is the pressure at node 25 in simulation i; d_i is the flow in pipe 5 in simulation i; e_i is the flow in pipe 27 in simulation i; f_i is the flow in pipe 29 in simulation i;

Where, $P16_i$, $P23_i$, $P25_i$, $F5_i$, $F27_i$, $F29_i$ are actual pressures and flows; $P16_i$ is the pressure at node 16 in simulation i, $P23_i$ is the pressure at node 23 in simulation i, $P25_i$ is the pressure at node 25 in simulation i; $F5_i$ is the flow in pipe 5 in simulation i, $F27_i$ is the flow in pipe 27 in simulation i; $F29_i$ is the flow in pipe 29 in simulation i;

C. **Constraint Function:** The only constraint that might cause trouble halfway through the optimization process has turned out to be the pressure values that intermittently violate the minimum pressure limits. So, by considering penalty functions, the optimization model is secured to yield reliable results. The minimum pressure head value to consider has been 10 meters.

4. RESULTS

The output of the optimization framework is a set of 34 roughness values that need to be compared to the actual one. The outcome of the simulation-based optimization, whose characteristics can be seen in Table 2, is assessed through MAE and MAPE. These metrics are shown in Equations 2 and 3 below that represent the calculation process of the MAE and MAPE for both predicted and simulated values. The lower the value of MAPE and/or the closer the value of MAE to one, the more accurate the results of the optimization will be.

$$Mean\ Absolute\ Error = \frac{\sum_{i=1}^{m}\left(\frac{\sum_{j=1}^{n} abs\left(pr_{ij}-sim_{ij}\right)}{n}\right)}{m} \quad [2]$$

$$Mean\ Absolute\ Percentage\ Error = \left(\frac{\sum_{i=1}^{m}\left(\frac{\sum_{j=1}^{n}\frac{abs(pr_{ij}-sim_{ij})}{abs(sim_{ij})}}{n}\right)}{m}\right) * 100 \quad [3]$$

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, and *j* represents the number of scenarios. Also, *m* is associated with number of scenarios and *n* is the number of inputs, which is the number of decision variables in this matter.

1.1. Simulation-based Results

Using multiple initial population to resolve the problem with local optima, the model narrows down on the boundary conditions to shrink the vast search area that will be extremely time-consuming. Table 2 represents the optimization properties.

Table 2. Optimization Properties

Number of Generations	Number of Population Size	Boundary Tuning Function	Crossover Factor
85	100 - 600 (each new run will increase the population)	Yes/ Searching for optimal solution over a span of a difference of five between lower and upper bounds	0.85

Optimization results (optimal solution for predicted roughness coefficient) are listed in Table 3 where MSE for 8 pressure head monitoring stations and 7 flow rate monitoring stations has functioned as the objective function. Table 3 also shows that our model predictions are off by approximately 2.47 in terms of mean absolute error and 3.19% in terms of mean absolute percentage error (MAPE).

Table 3. Results of Simulation-based Analysis for actual and predicted roughness values

Pipe Number	Actual Values	Predicted Values	Pipe Number	Actual Values	Predicted Values	Correlation Value	
						Mean Absolute Error	Mean Absolute Percentage Error
1	82	84	**18**	75	76	2.470588	3.19%
2	90	90	**19**	80	80		
3	89	87	**20**	65	65		
4	85	86	**21**	76	77		
5	69	70	**22**	88	74		
6	66	72	**23**	80	80		
7	88	88	**24**	73	73		
8	70	80	**25**	64	64		
9	63	63	**26**	77	77		
10	68	68	**27**	64	70		
11	81	81	**28**	88	77		
12	80	80	**29**	85	85		
13	72	72	**30**	82	84		
14	87	75	**31**	74	67		
15	70	70	**32**	67	73		
16	69	69	**33**	68	68		
17	65	66	**34**	75	74		

5. DISCUSSION AND FUTURE WORK

Primarily, the fine-tuning of genetic algorithm parameters as well as re-initializing the optimization process through multiple initial populations assist the convergence trend in both avoiding getting stuck in local optima and covering better spans of search areas. Also, the tuning function that accounts for lowering the boundaries of the search span contributes to taking care of all feasible solutions, thus increasing the probability of achieving the near-optimal solution. Furthermore, apart from the behavior of optimization model, the novelty in this study suggests that the scheme could be an acceptable substitute for expensive and time-consuming methods of asset management. Also, since these methods rely on the operational and physical alterations and variations of the water distribution network, an accurate model for asset management can be developed by only relying on cyber-monitoring data that would both cut down on expenses and time that conventional methods would take. In terms of future work, in order for this study to be applicable to all water systems, different methods are needed to decrease the time it takes for the model to optimize the problem by, for instance, adding neural networks that can be trained to eliminate the lengthy simulator-based optimization, thus speeding up the process. Also, a comprehensive sensitivity analysis on different criteria such as the number of scenarios or number and location of monitoring stations is required to make certain the model is as accurate as possible.

6. 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

7. REFERENCES

Altarabsheh, A., Ventresca, M., and Kandil, A. (2018). "New Approach for Critical Pipe Prioritization in Wastewater Asset Management Planning." *Journal of Computing in Civil Engineering*, 32(5), 04018044.

Bastian, B. T., N, J., Ranjith, S. K., and Jiji, C. V. (2019). "Visual inspection and characterization of external corrosion in pipelines using deep neural network." *NDT and E International*, 107.

Fujiwara, O., and Khang, D. B. (1990). "A two-phase decomposition method for optimal design of looped water distribution networks." *Water Resources Research*, 26(4), 539–549.

Liu, J., Chen, H., Yao, L., Wei, Z., Lou, L., Shan, Y., Endalkachew, S. D., Mallikarjuna, N., Hu, B., and Zhou, X. (2016). "The spatial distribution of pollutants in pipe-scale of large-diameter pipelines in a drinking water distribution system." *Journal of Hazardous Materials*, 317, 27–35.

Meirelles, G., Manzi, D., Brentan, B., Goulart, T., and Luvizotto, E. (2017). "Calibration Model for Water Distribution Network Using Pressures Estimated by Artificial Neural Networks." *Water Resources Management*, 31(13), 4339–4351.

Mutikanga, H. E., Sharma, S. K., and Vairavamoorthy, K. (2012). "Methods and Tools for Managing Losses in Water Distribution Systems." *Journal of Water Resources Planning and Management*, 139(2), 166–174.

de Myttenaere, A., Golden, B., Le Grand, B., and Rossi, F. (2016). "Mean Absolute Percentage Error for regression models." *Neurocomputing*, 192, 38–48.

Newton, L. A., and Christian, J. (2015). "Challenges in asset management - a case study." (February), 13.

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.

Sivakumar, P., Prasad, R. K., and Chandramouli, S. (2016). "Uncertainty Analysis of Looped Water Distribution Networks Using Linked EPANET-GA Method." *Water Resources Management*, 30(1), 331–358.

US EPA National Center for Environmental Assessment, O. of the D. D., and Shaw, D. (2007). "EPA's 2007 Report on the Environment: Science Report (SAB Review Draft)."

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.