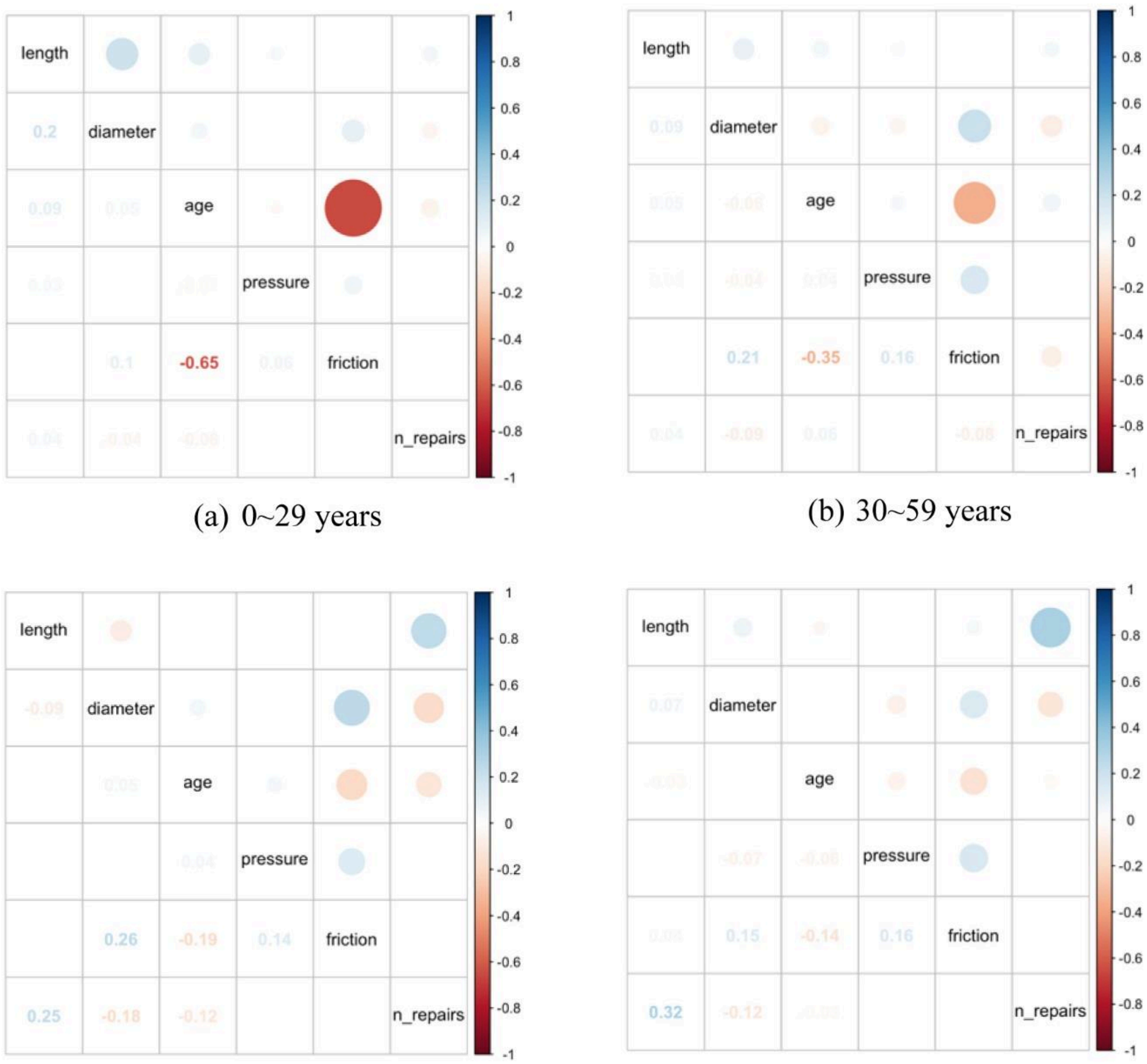
number of clusters, such as Elbow method, Average Silhouette method and Gap Statistic Method. Here the average Silhouette method is used, and the result is shown in Figure 8. Both 2 and 3 clusters seem to be acceptable.



(a) 0~29 years

(b) 30~59 years

(c) 60~89 years

(d) > 90 years

Figure 5. Correlograms for Different Age Groups of Pipes

SURVIVAL ANALYSIS

Using the whole water pipe dataset of Utility A, the non-parametric survival curves of all water pipes can be developed. This kind of proportional hazards model can be visualized with a Kaplan-Meier plot. Survival curves can also be fitted with parametric distributions. The most commonly used distributions are Weibull distribution, exponential distribution, log-normal distribution and log-logistic distribution. Such kinds of models are parametric, and thus belong to accelerated failure time (AFT) models. When the pipe dataset is partitioned into different groups, the influence of a grouping covariate on the survival probability can be illustrated via a categorized Kaplan-Meier plot.

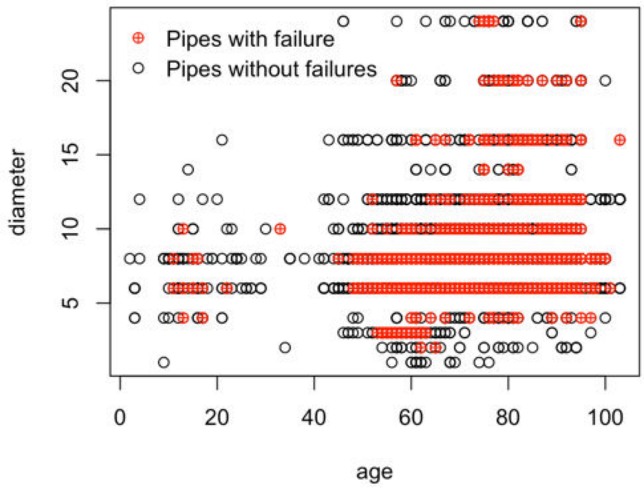


Figure 6. Distribution of Pipes in terms of Age and Diameter

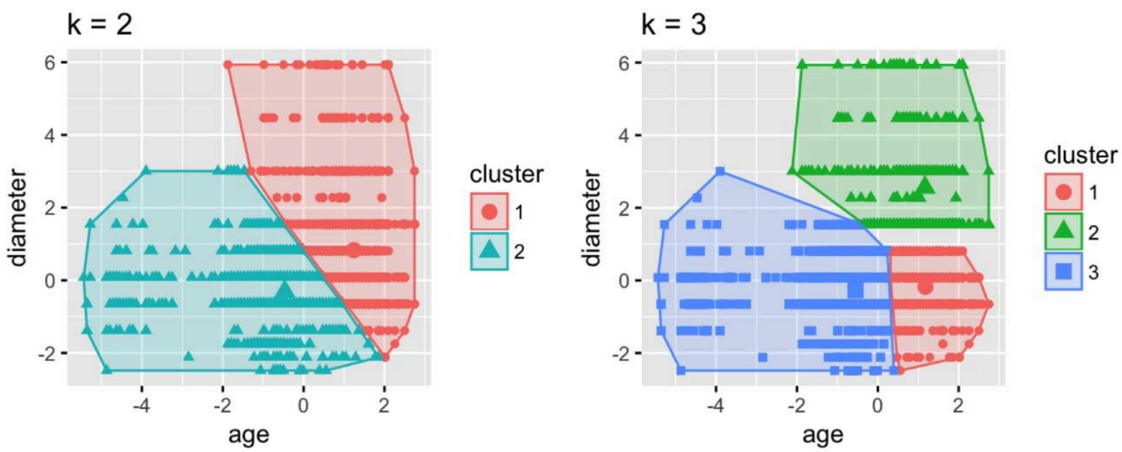


Figure 7. Cluster Results with Different Numbers of Clusters

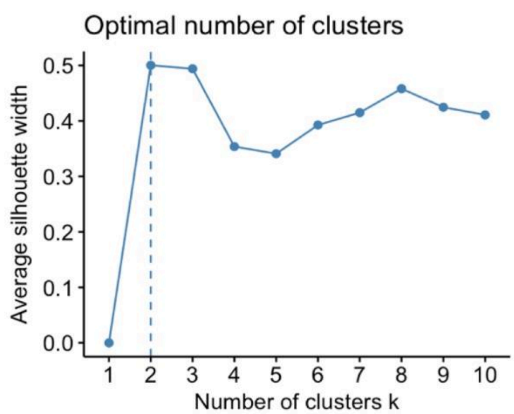
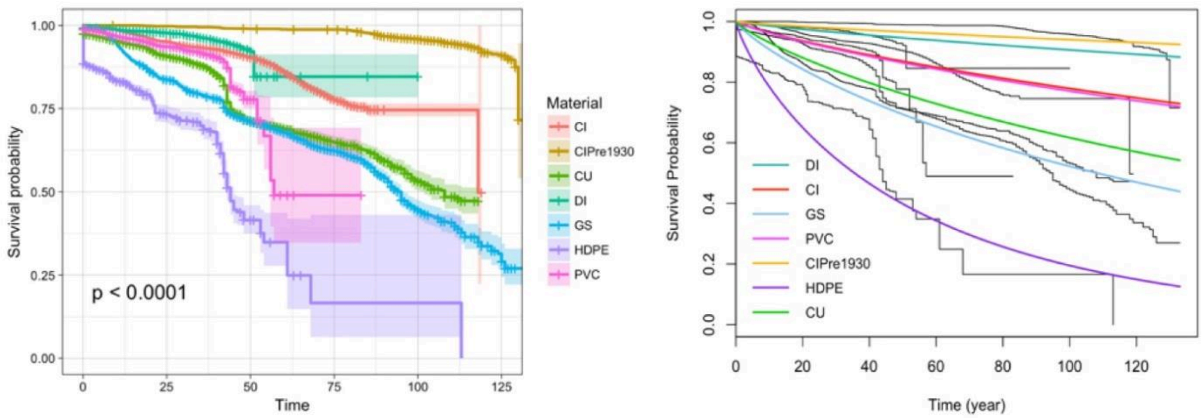


Figure 8. Optimal Number of Clusters

In the dataset of Utility A, the major material types include CI, DI, CU, GS, HDPE and PVC. It should also be noted that the CI pipes installed prior to 1930 are specially sorted out as 'CIPre1930' in the dataset of Utility A. In this analysis, we adopt the representation and notation systems of the utility. Based on the stratification by pipe materials, Kaplan-Meier plots for different material types are created separately, as are shown in Figure 9 (a). CI pipes installed prior to 1930 are the most durable among different materials' pipes, while HDPE pipes are least likely to survive. The parametric survival curves by subgroups are also developed. To make the plot legible, only the survival models fitted with Weibull distribution are shown in Figure 9 (b).



(a) Kaplan-Meier plot (b) Weibull fitting of survival curves

Figure 9. Survival analysis stratified by pipe materials

CONCLUSION

This paper presents a comprehensive analytical framework for statistical analysis of water pipeline field performance data. The proposed framework incorporates three objectives (i.e. failure, performance and risk) and employs both exploratory and predictive statistical modelling approach.

Some knowledge and findings are discovered and verified in the preliminary analysis. For utilities that do not record pipe types, pipes with diameters equal to or larger than 16 inches can be treated as transmission pipes and those with diameter less than 16 inches can be considered distribution pipes. In the initial stage (e.g. first 30 years) pipe's internal surface gets rougher with the increase of pipe age significantly, while afterwards the friction factor will not be influenced with pipe age as significantly. This finding is not only about trends but also the trend of the strength of trends. Premature failures can be detected and investigated. Both 2 and 3 clusters are appropriate for cluster analysis. According to survival analysis of Utility A, CI pipes installed prior to 1930 are the most durable among different materials' pipes, while HDPE pipes are least likely to survive. In the survival analysis, most of the recorded breaks in HDPE pipes occurred in the early stage, while most of the breaks in CI pipes occurred in the late stage. There are few recorded breaks for HDPE pipes after the age of 75 years. This is the main reason why the survival curve of HDPE pipes exhibits a steep downward trend.

REFERENCES

Chen, T. Y. J., Beekman, J. A., & Guikema, S. D. (2017) "Drinking Water Distribution Systems Asset Management: Statistical Modelling of Pipe Breaks." In Pipelines 2017 (pp. 173-186).

Clair, A.M., & Sinha, S. (2012). "State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models." Urban Water Journal, 9(2), 85-112.

Clair, A.M. (2013). "Development of Drinking Water Pipe Performance Index and Prediction." Doctor of Philosophy in Civil Engineering, Virginia Polytechnic Institute and State

University, Blacksburg, VA.
Farmani, R., Kakoudakis, K., Behzadian, K., & Butler, D. (2017). “Pipe failure prediction in water distribution systems considering static and dynamic factors.” Procedia Engineering, 186, 117-126.
Sinha, S., & Sears, L. (2017) Collection and Compilation of Water Pipeline Field Performance Data. In Pipelines 2017 (pp. 124-135).
Yamijala, S., Guikema, S. D., and Brumbelow, K. (2009). “Statistical models for the analysis of water distribution system pipe break data.” Reliability Engineering and System Safety, 94(2), 282–293.

A Novel Water Pipeline Asset Management Scheme Using Hydraulic Monitoring Data

Kalyan R. Piratla, A.M.ASCE[1]; and Ahmad Momeni[2]

[1]S. E. Liles, Jr. Distinguished Associate Professor, Glenn Dept. of Civil Engineering, Clemson Univ., PO Box 29634, Clemson, SC 29634. E-mail: kpiratl@clemson.edu
[2]Ph.D. Student and Graduate Research Assistant, Glenn Dept. of Civil Engineering, Clemson Univ., PO Box 29634, Clemson, SC 29634. E-mail: amomeni@clemson.edu

ABSTRACT

Many water pipeline systems in the United States and several other countries are in a deteriorated state needing immediate intervention. One of the critical challenges to such interventions is the dearth of economical and reliable inspection tools to assess the condition of the pipeline assets and prioritize their rehabilitation. Although few innovative inspection techniques have been developed and demonstrated in the last few years, they are reserved to be used on a limited number of seemingly failing assets due to high inspection costs. The fields of embedded sensing and artificial intelligence techniques offer unique opportunities to predict asset conditions based on hydraulic monitoring data from water pipeline systems. This paper proposes a novel asset-management scheme for water pipeline systems where pipeline flow and pressure data streamed in through the SCADA systems are leveraged to deduce uncertain asset parameters such as reduced pipe diameters and roughness values. To demonstrate the proposed scheme, a well-known water distribution network is used in this paper to show that pressure and flow data monitored at three different locations each can inform the roughness values of all the pipelines in the system. Firstly, the roughness coefficients of all the pipelines in the chosen water network are randomly reduced within a certain reasonable range in order to characterize the real-world system behavior. Subsequently, synthetic monitoring data for pipeline pressure and flow is generated using hydraulic simulations for 200 scenarios with varied nodal demands. Finally, an optimization algorithm is developed based on a reverse engineering approach to predict pipeline roughness coefficients using the synthetic monitoring data. Least of the minimum squared error between the modeled and synthetic flow (and pressure) data for all the 200 scenarios are considered the objectives in the optimization process. MATLAB programming interface is used in this study in conjunction with EPANET hydraulic modeling tools. The proposed scheme and the included results of this paper will offer a promising and powerful approach towards asset management where system-wide monitoring data could inform the condition of the assets and subsequently support prioritization for rehabilitation.

INTRODUCTION

Asset management entails gathering knowledge about the condition of various assets in order to make informed decisions on systematic maintenance, rehabilitation, and capital improvement planning. Asset management is deemed more critical when it comes to the limited budget dedicated to pipeline rehabilitation as well as the emergence of various ageing pipelines around the world (Moglia et al. 2006). Particularly, what hinders the fast and accurate asset management for stakeholders is the tough nature of condition assessment and record keeping of data by relying on conventional inspection using outdated tools and time-consuming methods (Newton and Christian 2015). Also, the labor-intensiveness associated with inspection and condition assessment is another challenge that needs to be addressed. Furthermore, even with the best of the available technologies, it may not be possible to accurately assess the condition of all the assets in the system. This paper presents a novel framework to predict the condition of pipeline assets based on hydraulic monitoring data that could be available through SCADA systems. Given the large percentage of metal pipelines currently in service in our drinking water systems, pipeline roughness is considered in this study to be a critical parameter that needs to be predicted. Metal pipelines tend to get rougher with age and deterioration, and this phenomenon may be aggravated by corrosion and scaling inside the metal pipelines. Furthermore, the remaining hydraulic diameter in older metal pipelines may be much smaller than the actual diameters of the pipelines. Determining how rough a pipeline has become with age and the effective internal diameter will be useful information for asset management purposes. This paper demonstrates the prediction of pipeline roughness based on SCADA data.

Table 1. Hanoi Original and Alternative Network Variations

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

METHODOLOGY

The approach into how asset management is carried out in this paper relies on reverse engineering. A deteriorated water distribution network representative of a typical water system is first developed. The deterioration is characterized by reduced pipe roughness (C) values as well as effective internal pipe diameters (assumed to have been reduced by corrosion-related scaling). A popular benchmark water distribution network – Hanoi network (Fujiwara and Khang 1990) depicted in Figure 1 – is used in this study to demonstrate the proposed condition assessment approach. The diameter of each pipeline is randomly reduced by some amount ranging between 30 to 50 mm, whereas the roughness coefficient (C) of each pipeline is randomly reduced to a value between 60 and 90 from its original value of 130. Several alternate Hanoi networks were generated by randomly reducing the pipe diameters and pipe roughness coefficients within the previously stated ranges. One such alternate network (which is referred hereafter as Alt #1) is used to demonstrate the proposed condition assessment prediction framework. Table 1 shows the pipe diameters and roughness coefficient values in the original Hanoi network as well as the Alt#1 network. It is ensured that the reduction in pipe sizes and roughness coefficient values did not affect the ability of Alt#1 network in meeting nodal demands.

Once the Alt#1 network, which is assumed to be representative of a deteriorated network, is developed, monitoring locations are identified for SCADA type real-time hydraulic data acquisition from the Alt#1 network. It is assumed that pressure monitoring stations are initially located at nodes 16, 23 and 25 in Alt#1 network and flow monitoring stations are located in pipelines 5, 27 and 29; however, more monitoring locations at other nodes and pipes are assumed for sensitivity analysis purposes. Essentially, these monitoring stations would allow acquisition of pressure and flow data from these locations. It is also assumed that the Alt#1 network has smart water meters at the demand nodes that would relay the consumption data at any time point that would correspond to the hydraulic monitoring parameters (i.e., pressures at nodes 16, 23 and 25 and flows in pipes 5, 27, and 29) at the same time point. Nodal demands are expected to vary with time on a given day and correspondingly the network hydraulic parameters vary. In a real-world scenario, the water network performance varies in response to change in water demands and change in the availability of water network components. Assuming there are no failures in the water system, the inputs for modeling the performance of the water network would be nodal demands and the outputs would be pressures at different locations in the system and flows in all the pipes. In order to capture the hydraulic system dynamics of the Alt#1 network, 200 sets of inputs (i.e., nodal demands) and corresponding pressures at nodes 16, 23 and 25 and flows in pipelines 5, 27 and 29 are generated as synthetic monitoring data. The nodal demands are varied in ±20% of the base nodal demands of the original Hanoi network. EPANET 2.0 software is used to perform the hydraulic simulations and MATLAB programming interface is used to develop the algorithm in conjunction with an open-source EPANET 2.0 extension toolkit library.

Finally, an optimization algorithm is formulated where the decision variables are the pipe roughness coefficients. The algorithm attempts to minimize the variation (quantified through mean squared error – MSE) between the calculated and actual (i.e., synthetic) values for pressures at nodes 16, 23 and 25 and flows in pipelines 5, 27 and 29 over the 200 demand scenarios. Genetic algorithm is used to solve this optimization problem which would result in the most optimal set of pipe roughness coefficients. Comparison of the optimized pipe roughness coefficients with the actual pipe roughness coefficients would validate the proposed approach.

Demonstration on Alt#1 Network

As shown in Figure 1 below, the Alt#1 network comprises of 31 nodes, 34 pipelines and one reservoir (Fujiwara and Khang 1990).

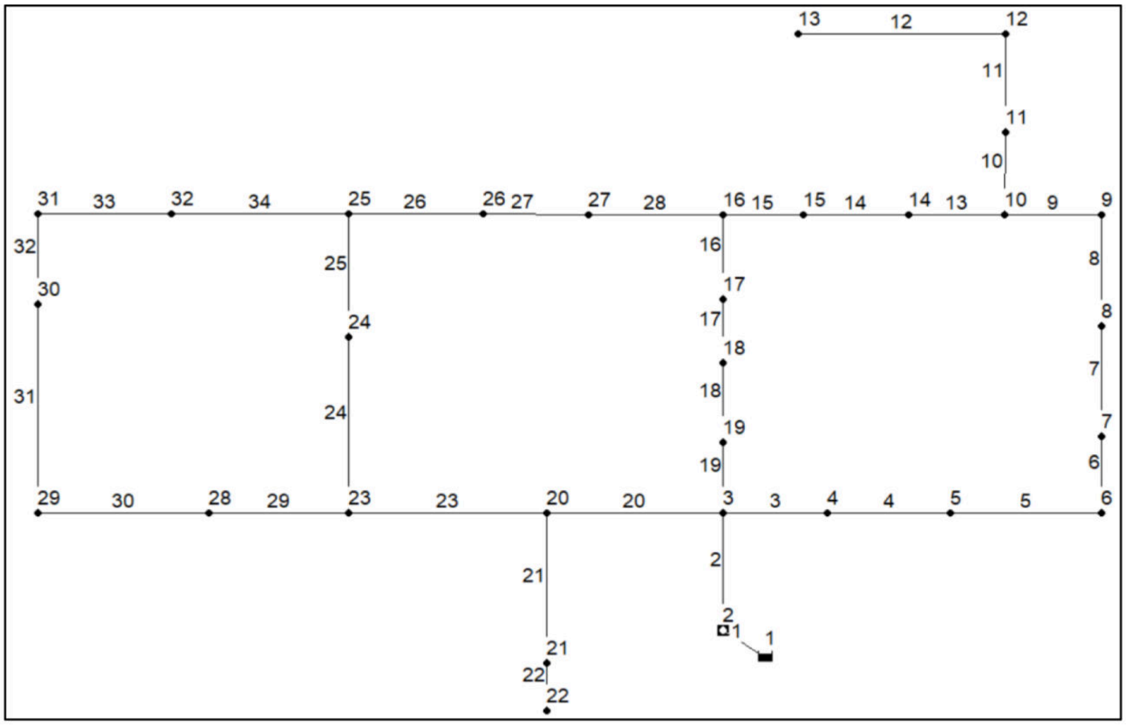


Figure 1. Alt#1 (Hanoi) Water Distribution Network Layout

Considerations for the Optimization Formulation

The algorithm simply aims at feeding a set of randomized pipe roughness values to the previously provided synthetic monitoring data (200 scenarios of nodal demands and network hydraulic performance data) in the study. Then, a 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. In this case, the mean squared error (MSE) between the modeled and synthetic flow and pressure data has been formulated as the objective as shown in Equation 1 below.

A. **Decision variables:** {*x1, x2,…., and 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. **Objective:** Minimize the following

$$\text{Minimum of } \left[(a_i - P16_i)^2 + (a_i - P16_i)^2 + (a_i - P16_i)^2\right] \text{for all i } + \text{ Minimum of } [(d_i - F5_i)^2 + (e_i - F27_i)^2 + (f_i - F29_i)^2] \text{ for all } i \quad \text{(Eq. 1)}$$

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 value to consider has been 10 meters.

Table 2. Results for simulations associated with 4 scenarios of predicting roughness coefficients

Scenarios	Generation/ Population Size for GA	Number of Monitoring Station	Roughness (Actual Values)	Roughness (Actual Values)	Roughness (Actual Values)	Flow and Pressure MSE	Roughness MSE [sum((R_a-R] o)^2)
		Pressure	Flow				
Scenario 1	200/200	3	3	82	82	0	0
				90	90		
Scenario 2	300/600	5	5	82	84	14.693	17
				90	89		
				89	87		
				85	83		
				69	71		
Scenario 3	500/500	8	5	82	90	1.3085	80
				90	86		
				89	89		
				85	85		
				69	69		
Scenario 4	200/200	3	3	82	87	52.118	105
				90	88		
				89	83		
				85	86		
				69	71		
				66	71		
				88	85		
				70	69		

Results and Discussion

Table 2 represents four scenarios where multiple iterations have been simulated to analyze the behavior of roughness coefficient prediction model through mean square errors of roughness values (actual, constant values (Ra) of roughness that were acquired through Alt#1 and obtained values (Ro) that are the results of the optimization model for each simulated scenario) based on flow and pressure monitoring data. As can be seen in Table 2, Scenario 1 was simply meant to solve for the roughness values of only two pipes. This scenario was run using the GA optimization code for 200 population size and generations, and thus yielded perfectly accurate results with actual and predicted roughness coefficients being the same and the mean square error (MSE) of zero for both flows and pressure values at three monitoring stations. However, Scenarios 2 and 3 have been devised for the first five pipes of Hanoi Alt#1 network; Scenario 3 where the number of monitoring stations was increased to eight for pressure and five for flow yielded more accurate MSE for roughness in comparison to Scenario 2 with five stations for both flow and pressure. However, the MSE-based objective functions in this optimization problem probably look to be more dependent on the smaller mutation shrink factor (0.01 – 0.02) as well as optimized numbers of both population size (approximately twice the number of decision variables) and generations, insofar as the number of decision variables are on the increase, scenario by scenario. This means that the MSE for roughness in Scenario 2, accounting for 600 of population size, turned out to be slightly smaller than that in Scenario 3 where the population size equals 500. To compare the results more critically, Scenario 4 along with the conventional 3-station, 200-generation-and-population-size status yielded somewhat much less accurate results both on MSE for Flow/Pressure and roughness in presence of eight decision variables in comparison to scenario 1 with only two decision variables yet similar GA parameters, which makes the operation of the optimization model computationally expensive to an excessive extent.

Table 2 aims at clarifying that as the number of decision variables increases, higher number of generations and populations will be needed to extend the search span. Besides, the accuracy and sensitivity with which the optimization model attempts to converge to the optimal solution appear to be improved as the number of monitoring stations of flow and pressure rises. However, accurate sensitivity analysis as to how many monitoring stations as well as where to place monitoring locations should be the integral part of associated future work in this study. For instance, by comparing Scenario 2 and Scenario 3, it is evident that the obtained set of roughness values in Scenario 3, where higher monitoring locations and more generations and population sizes are considered, looks to be in better proximity to the actual roughness values than that in Scenario 2.

Table 3 includes results of two more scenarios associated with predicting roughness values for all 34 pipes utilizing 8 pressure stations and 5 flow stations accommodated on Alt#1 network.

It apparently demonstrates that possibly as the number of decision variables increase drastically, the accuracy with which the GA method should yield results drops considerably, as the MSEs in Table 3 equal much higher values, thus less accurate results.

Table 4 suggests Scenarios 7, 8, and 9, where the first 12, 14, and eight pipes of Alt#1 Hanoi network have been attempted to be optimized for their roughness values. Thanks to the tuning of GA parameters including mutation shrink and crossover factors as well as considering 3 discrete objective functions rather than the summation of all the MSE values and ultimately tightening the lower and upper bounds of each decision variable, the accuracy of these sets in comparison to those in Table 2 appears to have increased proportional to the number of decision variables (number of pipes involved in optimization), contributing to smoother convergence to the optimal solutions and thus more accurate results. This demonstrates that both actual and obtained roughness values in these scenarios, especially in scenario 9, tend to be similar.

Observations and Future Work

Due to the high number of decision variables (34 pipe roughness values) and rather low numbers of generations, it is probably predicted that the actual search span for accurate results is much more extensive than the maximum 500 generation and population size which was utilized in this study. It was observed during the optimization process of at least 12 scenarios that after specific numbers of generations prior to the termination of optimization process, the optimization outcome tended to converge on some obtained values and no more mutation or newly generated set of decision variables was carried out anymore. This apparently demonstrates that owing to the extremely vast search span, the optimization algorithm might have stuck and trapped in local optima and fail to search for the global optima. Therefore, part of the future work in this study