International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 SJIF Rating: 7.185 ISSN: 2582-3930

An Innovative Geostatistical Framework for Prioritizing Inspections in Water Distribution Networks

Tanay Kulkarni
Graduate Student, USA
Email: tanaykulkarni[at]outlook.com

Devashri Karve
Graduate Student, USA
Email: devashrikarve[at]outlook.com

Abstract: Aging water infrastructure networks face significant challenges, with the United States experiencing frequent pipe failures and substantial water loss. This study leverages geospatial and geo-statistical analysis to prioritize pipeline inspection in Watertown, Connecticut. By integrating Geographic Information Systems (GIS) with machine learning, various factors influencing pipeline health—such as pipe age, material, length, diameter, soil pH, and groundwater depth—are analyzed. Historical pipe break data between 2015 and 2021 are mapped to identify spatial correlations with environmental and operational attributes. A K-Means clustering algorithm classifies pipelines into high, medium, and low inspection priority categories based on critical attributes. Findings reveal that corrosive soil pipes with low diameters and high break frequencies pose significant risks. The study demonstrates the effectiveness of GIS-driven data integration and analysis in optimizing maintenance strategies, reducing system-wide impacts, and improving resilience. These insights provide a replicable framework for water utilities worldwide to address aging infrastructure challenges.

Keywords: water distribution network, pipe inspection, risk prioritization, geospatial statistics, unsupervised learning

1. INTRODUCTION

Aging water infrastructure poses significant challenges to utilities worldwide. In the United States alone, over 6 billion gallons of treated potable water are lost daily due to leaks and breaks, with an average pipe failure occurring every two minutes (ASCE, 2021). The country's water systems, some of which date back over a century, require extensive repairs and replacements to maintain reliability. In 2020, approximately 12,000 miles of water pipelines were planned for replacement, demonstrating the scale of this issue (ASCE, 2021). These challenges underscore the urgency for innovative strategies to optimize maintenance and inspection schedules.

Effective inspection prioritization is critical for mitigating risks associated with aging infrastructure. Water distribution networks are complex systems influenced by diverse factors, including pipe age, material, soil conditions, and operational pressures. Traditional approaches to pipeline maintenance often need more precision to allocate resources efficiently. When integrated with Geographic Information Systems (GIS), geostatistical analysis offers a robust framework for identifying high-risk pipelines, improving resilience, and minimizing water losses.

This study demonstrates how geostatistical analysis can prioritize pipeline inspections in a water distribution network. Using Watertown, Connecticut, as a case study, the research explores the influence of factors such as pipe material, age, soil pH, and groundwater depth on pipeline failures. A K-Means clustering algorithm classifies pipes into priority categories, providing actionable insights for utility managers.

The study analyzes a case-study system and provides a replicable framework for other municipalities facing similar challenges.

2. LITERATURE REVIEW

Aging Water Infrastructure in the United States: The aging water infrastructure in the United States poses a critical challenge for utilities and policymakers. Many pipelines installed in the mid-20th century have surpassed their designed service lives, leading to frequent failures and inefficiencies. The American Society of Civil Engineers (ASCE) reported that over 2.2 million miles of underground pipelines deliver water nationwide, yet significant portions of this network are over 75 years old (ASCE, 2021).

GIS for Infrastructure Management: GIS has emerged as a transformative tool for managing water distribution networks by integrating spatial and non-spatial data for analysis and visualization. Combined with geostatistical methods, GIS enables utilities to assess pipeline conditions and prioritize interventions effectively. Additionally, GIS-based decision-making models provide utilities with insights for optimizing maintenance schedules, reducing downtime, and enhancing resource allocation. This approach is increasingly adopted globally, bridging the gap between traditional management practices and data-driven decision-making. Geographic Information Systems (GIS) have become indispensable in water resource management, offering data collection, storage, analysis, and visualization tools. By integrating spatial and non-spatial data, GIS facilitates comprehensive analyses of water distribution networks, enabling informed decision-making for maintenance and resource allocation. Geostatistical methods, such as kriging and variography, are employed to interpolate measured data points, creating continuous spatial maps that enhance the understanding of hydrological variables (Atkinson & Lloyd, 2021). These techniques are particularly valuable in groundwater monitoring, providing detailed insights into aquifer characteristics and contamination risks. Additionally, GIS supports managing water supply and sewer systems by mapping infrastructure components, analyzing spatial relationships, and optimizing maintenance schedules (Shamsi, 1996). This integration enhances the efficiency of water distribution and wastewater management, ensuring reliable service delivery. Wang and Xie (2018) highlight that remote sensing facilitates the mapping of water resources and monitoring of hydrological fluxes, while GIS offers robust tools for managing water resources and assessing flood risks. Integrating these technologies enables precise mapping and management of water resources, thereby enhancing disaster preparedness and response strategies. The utilization of GIS in hydrological and water quality modeling has advanced the understanding of water flow and pollutant transport processes. Kim (2018) discusses how GIS efficiently parameterizes input data for various models, representing the spatial and temporal characteristics of factors affecting hydrologic components and pollutant generation. This integration is crucial for sustainable water resource management, allowing for the simulation and analysis of complex hydrological systems. The application of GIS and spatial analysis has become increasingly prevalent in hydrological modeling and water resource system analysis. Chang (2018) notes that GIS technology is instrumental in managing and understanding water resources across various scales, facilitating spatially explicit analyses essential for effective water resource management. This approach enables the visualization and analysis of spatial patterns and relationships within hydrological data, enhancing the ability to address complex water resource challenges. The advancement of GIS technologies has introduced new perspectives in water resource management. Yang (2021) explores how GIS, combined with emerging spatial models, depicts the spatiotemporal patterns of water resources and related risks. This approach enhances the evaluation of the impacts of water resource engineering on ecosystems, promoting environmental sustainability. The ability of GIS to integrate various data sources and provide comprehensive analyses makes it a valuable tool in addressing contemporary water resource challenges.

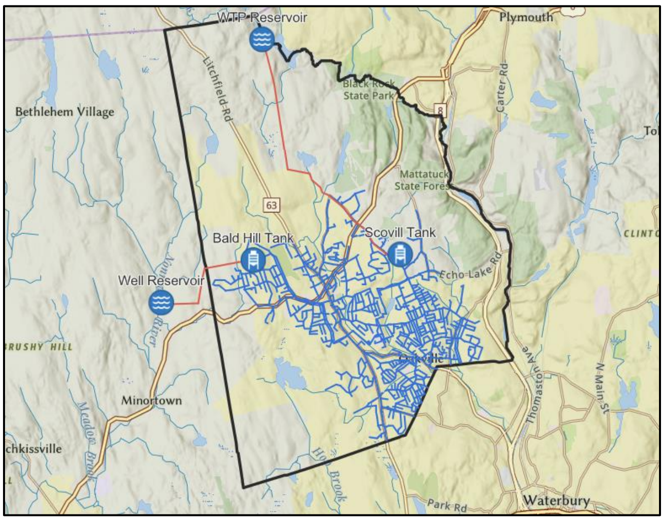
Factors Influencing Pipeline Health and Failures: Pipeline failures are influenced by a complex interplay of material, environmental, and operational factors. Corrosion, often accelerated by low soil pH and high moisture content, is a leading cause of pipeline degradation (Zhang et al., 2020). Environmental conditions, such as soil chemistry, groundwater levels, and climate variations, significantly impact pipeline health. Corrosive soils with low pH levels and high moisture content accelerate external corrosion, particularly in metallic pipes (Rajani & Kleiner, 2001). Seasonal temperature changes can also induce thermal stress, leading to cracks and fractures in older pipelines (Kleiner & Rajani, 2002). Additionally, extreme weather events like floods and earthquakes increase the risk of physical damage and pipeline displacement (Deb et al., 2010). The material composition of pipelines determines their durability and resistance to external stressors. Cast iron and ductile iron pipes, widely used in legacy systems, are prone to corrosion and mechanical wear over time (Rajani & Kleiner, 2001). While corrosion-resistant, Polyvinyl chloride (PVC) pipes may degrade under prolonged UV exposure and mechanical stress. Material selection based on environmental compatibility is essential for improving pipeline longevity. Operational factors, such as pressure, flow rates, and maintenance practices, directly influence pipeline performance. High-pressure fluctuations and operational errors, such as over-pressurization, can cause structural failures, particularly in aging pipelines (Kleiner & Rajani, 2002). Inadequate maintenance further exacerbates risks, leading to increased leakages and failures. Third-party activities, including unauthorized excavations and vandalism, threaten pipeline integrity. These activities account for many accidental pipeline failures (Deb et al., 2010). Implementing stricter monitoring and public awareness programs can mitigate these risks effectively.

Geographic and Demographic Profile of Watertown Watertown is a small town in Litchfield County, Connecticut, with an area of approximately 30 square miles, of which 29.5 square miles are land and 0.5 square miles are surface water. Its proximity to Waterbury, a city in New Haven County, adds regional significance. The town's water distribution network, which serves water to approximately 21,000 residents (U.S. Census Bureau, 2020), consists of pipes with significant variations in age, material, and condition. Watertown's blend of residential, institutional, and commercial areas creates a diverse demand for water distribution services, making it an ideal study area for analyzing infrastructure dynamics.

3. CASE STUDY

Water Distribution Infrastructure Overview: Watertown's water distribution system consists of reservoirs, elevated storage tanks, and an extensive pipeline network (Figure 1).

Figure 1: Watertown's Water Distribution Network in GIS

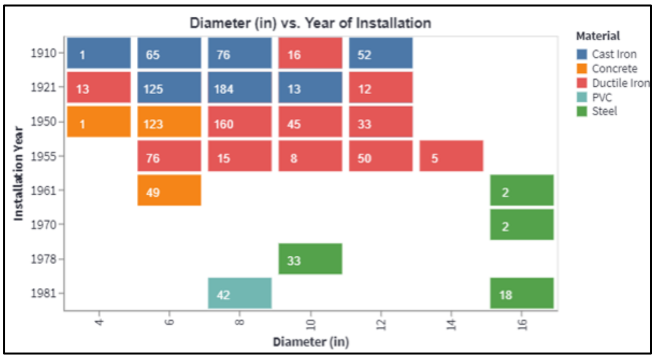


- *Reservoirs and Storage Tanks:* The Nonnewaug Falls Reservoir and Wigwam Reservoir are Watertown's primary potable water sources. The Nonnewaug Falls Reservoir has a pumping capacity of 800 gallons per minute (GPM), while the Wigwam Reservoir pumps at a rate of 900 GPM. Elevated storage tanks, such as the Bald Hill Tank and Scovill Tank, are critical in maintaining water pressure and supply. The Bald Hill Tank has a capacity of 1.5 million gallons and a hydraulic grade line (HGL) at 821 feet, while the Scovill Tank holds 1 million gallons with an overflow elevation of 999.5 feet.
- *Pipeline Network:* The pipeline network comprises 1,219 segments, spanning a total length of 7,042 feet. The network includes pipes made from ductile iron, cast iron, concrete, steel, and PVC, with 4 to 16-inch diameters. The age of the pipes varies significantly, reflecting the historical development of the town's infrastructure.

Data Collection and Sources: The study utilizes a range of datasets to analyze pipeline health and prioritization.

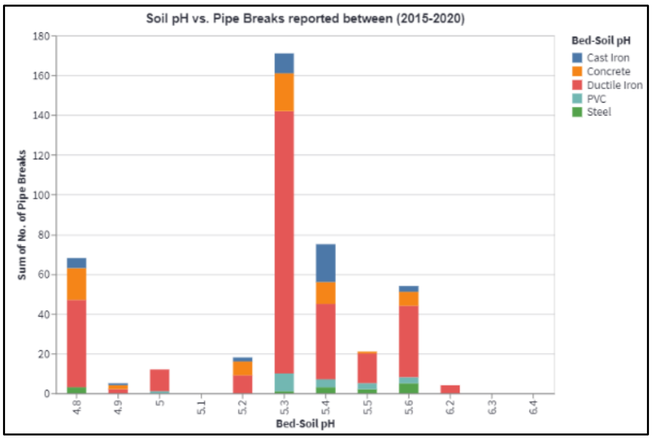
- *Pipe Characteristics:* Data included material composition, diameter, length, installation year, and operational attributes of the pipelines (Figure 2).

Figure 2: Diameter (in) vs. Install Year



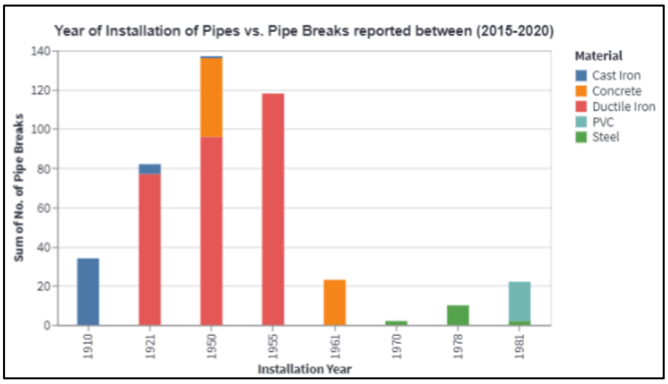
- *Environmental Factors:* Soil pH and groundwater depth data are sourced from the Soil Survey Geographic Database (SSURGO) and the U.S. Geological Survey (Figure 3). These parameters are critical for understanding external stresses on pipelines.

Figure 3: Bed-soil pH vs. Pipe Breaks (2015-2021)



- *Historical Pipe Break Data:* Records from 2015 to 2021 provide insights into the frequency, location, and causes of pipe failures, aiding in identifying patterns related to pipeline deterioration (Figure 4).

Figure 4: Install Year vs. Pipe Breaks (2015-2021)



4. METHODOLOGY

To facilitate spatial analysis, all data underwent preprocessing to ensure accuracy and compatibility. Missing values were addressed using statistical imputation methods, and categorical variables were standardized. Geospatial data, including pipeline locations and environmental attributes, were georeferenced using a consistent coordinate system. A relational database was created to integrate spatial data with pipeline characteristics and failure records, enabling seamless analysis in the GIS platform.

Statistical and Machine Learning Techniques Applied

K-Means Clustering: The study applied K-Means clustering to segment pipelines into three priority groups: high, medium, and low. Input variables such as pipe material, diameter, length, age, soil pH, and break frequency were used.

Correlation Analysis: Preliminary statistical analyses included Pearson and Spearman correlation tests to identify relationships between environmental and operational factors, such as soil pH, break density, groundwater depth, and material corrosion.

Mapping and Visualization Techniques

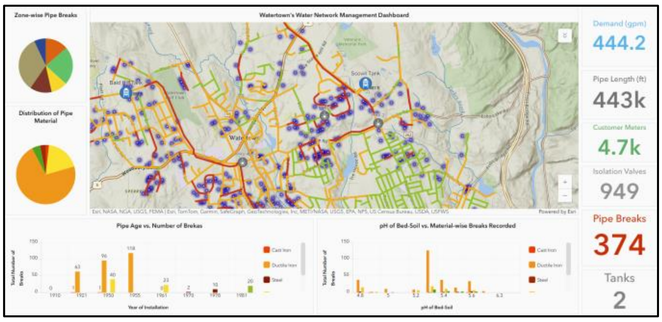
Visualization techniques within GIS were utilized to communicate effectively.

Thematic Maps: Maps depicting individual attributes, such as material distribution and break frequency, highlighted spatial vulnerabilities in the pipeline network.

Inspection Priority Map: Results from clustering were visualized as an inspection priority map, categorizing pipeline segments by risk levels to inform maintenance scheduling.

Interactive Visualizations: A web-based dashboard was developed for stakeholders to interact with layered maps to explore relationships between attributes and prioritize interventions (Figure 5).

Figure 5: Water Network Interactive Dashboard



5. RESULTS AND ANALYSIS

Statistical Insights from Pipe Characteristics

The analysis of Watertown's water distribution network revealed significant insights into the characteristics and performance of its pipelines. Most pipelines were constructed using ductile iron and cast iron, which are widely used in mid-20th century water infrastructure. These materials have exhibited high vulnerability to corrosion and failure, especially when exposed to low-pH soils. Diameter-wise, pipes ranged between 4 and 16 inches, with smaller-diameter pipes (4–6 inches) showing a higher frequency of failures. Age emerged as a critical determinant, with the majority of breaks occurring in

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 SJIF Rating: 7.185 ISSN: 2582-3930

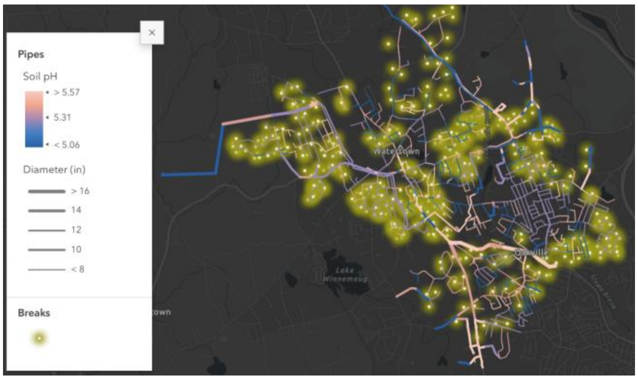
pipes installed between 1946 and 1981. These older pipes, particularly ductile iron segments, exhibited substantial deterioration due to prolonged exposure to environmental and operational stresses.

Spatial Analysis of Pipe Failures

Spatial patterns of pipeline failures highlighted correlations with environmental and operational variables:

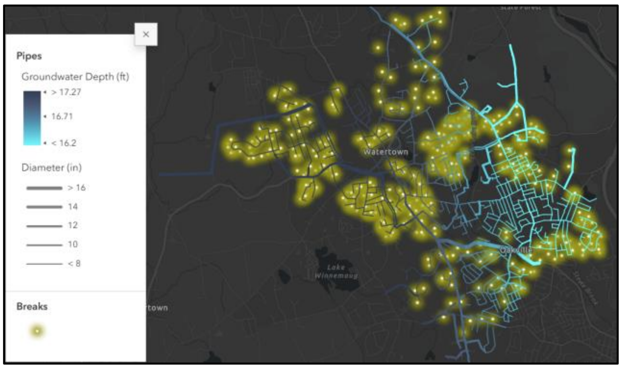
- *Soil pH:* Low-pH soils (below 6) were strongly associated with higher failure rates, indicating accelerated pipe corrosion in these areas. (Figure 5)

Figure 5: Correlation between Soil pH and Breaks



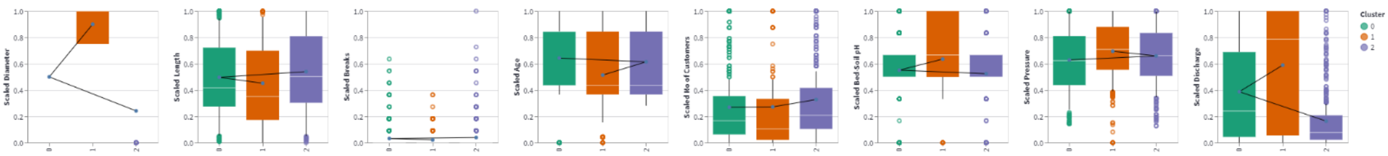
- *Groundwater Depth:* Although no direct correlation was observed between failure rates and groundwater depth, pipes closer to the groundwater table often showed accelerated corrosion, particularly in metallic materials. (Figure 6)

Figure 6: Correlation between Groundwater and Breaks



- *Pressure and Discharge:* Failures were more frequent in areas of high-pressure zones, particularly in smaller-diameter pipelines. Conversely, pipes with higher discharge rates were less prone to breaks.

Figure 7: Multi-variate K-means Clusters



Inspection Priority Clustering

The K-Means clustering algorithm classified the pipeline network into three priority clusters based on attributes such as diameter, material, age, soil pH, and break frequency (Figure 7). Inspection priorities were determined based on the clusters (Figure 8).

Figure 8: Inspection Priority Map

- High Priority (Cluster 2): Pipes in this cluster had the highest failure rates and smallest diameters and were predominantly constructed of ductile iron. These pipes were also located in areas with corrosive soils and moderate groundwater levels.

Due to their systemic impact, this cluster represented the segments most critical for immediate inspection and maintenance.

- Medium Priority (Cluster 1): This cluster consisted of pipes with intermediate attributes, including moderate failure rates and average diameters. Environmental and operational factors for these pipes were within tolerable ranges but required periodic monitoring.
- Low Priority (Cluster 0): These segments comprised larger-diameter pipes, predominantly constructed of non-corrosive materials such as PVC, and exhibited the lowest failure rates. These pipes were often located in non-corrosive soils and low-pressure zones, making them least susceptible to immediate risks.

Key Findings

The results underscored the significance of integrating geospatial and statistical analyses for prioritizing pipeline inspections. Ductile iron pipes, particularly those installed in the mid-20th century, emerged as the most vulnerable segments due to their age, material properties, and exposure to corrosive environments. Smaller-diameter pipes in high-pressure zones demonstrated higher failure frequencies, necessitating focused interventions. Spatial correlations with soil pH reaffirmed the impact of environmental factors on pipeline health. The clustering results provided actionable insights, enabling utility managers to allocate resources effectively toward high-risk pipeline segments while maintaining the resilience of the overall network. This comprehensive analysis highlights the critical interplay between environmental, material, and operational factors in managing water infrastructure effectively.

6. DISCUSSION

Implications of Findings for Pipe Asset Management

The findings of this study have significant implications for pipeline management. The prioritization framework derived from geostatistical and machine learning analyses allows water utilities to focus their resources on the most vulnerable pipeline segments. Utility managers can plan targeted inspections and maintenance activities by identifying high-priority pipes based on material, age, diameter, and environmental factors. For example, the high frequency of failures in ductile iron pipes located in low-pH soils suggests the need for proactive corrosion mitigation measures, such as cathodic protection or replacement with corrosion-resistant materials. Moreover, identifying high-pressure zones and small-diameter pipes as risk factors can inform operational strategies like pressure management, which reduces stress on the network and minimizes failure rates.

Benefits of Integrating GIS and Machine Learning

Integrating Geographic Information Systems (GIS) and machine learning techniques, such as K-Means clustering, proved to be a powerful approach in pipeline management. GIS facilitated the spatial visualization of pipeline attributes and failure data, enabling utility managers to detect patterns and correlations not evident in tabular datasets. Machine learning algorithms added predictive capabilities, allowing for the classification of pipelines into risk-based clusters. This combination of spatial and analytical tools enhances decision-making and reduces the subjectivity inherent in traditional maintenance planning methods. Furthermore, the inspection priority maps generated through GIS visualization ensure the insights are easily interpretable, fostering better stakeholder communication.

Limitations of the Study

Despite its strengths, the study has certain limitations. First, the analysis relies on the availability and accuracy of historical failure data and environmental attributes, which may need to be uniformly detailed across the entire network. Second, the clustering approach does not account for temporal changes, such as the evolving impact of climate on soil conditions or future pipeline degradation. Finally, the study focuses primarily on technical and environmental factors, potentially overlooking economic and social considerations, such as the costs of repairs or the criticality of certain service zones. Addressing these limitations in future studies could further enhance the robustness of the methodology and its applicability to diverse urban settings.

7. FUTURE RESEARCH

Future research should focus on expanding the current framework by integrating additional variables and refining the analytical techniques. For instance, incorporating real-time data from Internet of Things (IoT) sensors, such as pressure, flow, and temperature readings, could enhance the accuracy of pipeline condition assessments. Including socio-economic factors, such as the costs of repairs and the criticality of service areas, would provide a more holistic prioritization model. Advanced machine learning algorithms, such as gradient boosting or neural networks, could be explored to improve prediction accuracy for failure risks. Additionally, longitudinal studies that analyze temporal trends in pipeline performance under changing environmental conditions, such as climate impacts on soil pH and groundwater levels, could offer valuable insights. Finally, extending the study to include multiple municipalities would validate the replicability of the methodology and its scalability for larger water distribution networks. These enhancements would contribute to more resilient and sustainable infrastructure management strategies.

8. CONCLUSION

This study demonstrated the effectiveness of integrating geospatial analysis and machine learning for prioritizing pipeline inspections in Watertown, Connecticut. By leveraging Geographic Information Systems (GIS) and a K-Means clustering algorithm, the analysis identified high-risk pipeline segments based on critical factors such as material, age, diameter, soil pH, and historical failure data. The results highlighted the vulnerabilities of older ductile iron pipes, particularly those located in corrosive soils and high-pressure zones, emphasizing the importance of targeted maintenance to enhance system reliability. The study underscores the value of combining spatial data with advanced analytical tools to optimize resource allocation and decision-making in water distribution network management. While the methodology offers a robust framework for assessing pipeline health, the limitations, including reliance on historical data and the exclusion of socio-economic factors, present opportunities for refinement. Overall, this research provides a replicable and scalable approach to addressing the challenges of aging water infrastructure, paving the way for more resilient and sustainable systems.

9. ACKNOWLEDGEMENTS

The authors express our sincere gratitude to Bentley Systems, Inc. for providing the comprehensive dataset and Carnegie Mellon University for providing the ESRI ArcGIS Pro licenses that formed the foundation of this academic study. The authors also thank Professor Kristen Kurland of Carnegie Mellon University for her invaluable guidance and support throughout the research process. The authors would like to acknowledge Dr. Tom Walski and Mr. Shar Govindan from Bentley Systems, Inc., for their mentorship and technical insights, which were instrumental in shaping the direction and outcomes of this study.

REFERENCES

- American Society of Civil Engineers. (2021). 2021 report card for America's infrastructure: Drinking water. Retrieved from https://infrastructurereportcard.org
- U.S. Census Bureau. (2020). Census data for Watertown, Connecticut. Retrieved from https://www.census.gov
- Zhang, W., Smith, R., & Chen, H. (2020). Corrosion risk assessment for underground pipelines: A geostatistical approach. *Journal of Pipeline Infrastructure Management, 12*(3), 45-60.
- Chang, H. (2018). Geographic information science and spatial analysis in water resources. *ISPRS International Journal of Geo-Information*, Special Issue.
- Kim, J. (2018). GIS-based hydrology and water quality modeling. *Water*, Special Issue.
- Wang, X., & Xie, H. (2018). A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management. *Water*, 10(5), 608. https://doi.org/10.3390/w10050608
- Yang, X. (2021). GIS and water resources: New perspectives and applications. *Water*, Special Issue.
- Deb, A. K., Hasit, Y. J., & Grablutz, F. M. (2010). *Quantifying future rehabilitation and replacement needs of water mains*. Denver, CO: Water Research Foundation.
- Kleiner, Y., & Rajani, B. (2002). Forecasting variations and trends in water-main breaks. *Journal of Infrastructure Systems*, *8*(4), 122–131. https://doi.org/10.1061/(ASCE)1076-0342(2002)8:4(122)
- Rajani, B., & Kleiner, Y. (2001). Comprehensive review of structural deterioration of water mains: Physically based models. *Urban Water*, *3*(3), 151–164. https://doi.org/10.1016/S1462-0758(01)00032-2