Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Understanding the Societal Impacts of Water Distribution Failures: Implications for Capital Planning and Operations

Joseph Toland[1], Kareem Mostafa[2], Laurie Baise[3] and Lauryn Spearing[4]

[1] Research Assistant, Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA; joseph.toland@tufts.edu

[2] Postdoctoral researcher, Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada; Senior Engineer, Strategic Asset Management, City of Ottawa, Ontario, Canada; kareem.mostafa@ottawa.ca

[3] Professor, Civil and Environmental Engineering Chair, Civil and Environmental Engineering, Tufts University, Medford, MA, USA; laurie.baise@tufts.edu

[4] Assistant Professor, Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA; lauryn.spearing@tufts.edu

ABSTRACT: Water distribution failures and service outages present significant challenges for vulnerable communities. For example, boil water notices are often issued to protect public health when there are water quality concerns. In turn, residents must rely on alternative water sources, which adds to existing social and economic challenges (e.g., financial burdens, mobility). Water system operations are tied to community resilience, however, there is limited understanding of how these connections emerge. Consequently, this gap in understanding limits the development of evidence-based policies in water system capital planning and operations that could address such interconnections. Here, we propose a multi-scale approach to inform water system capital planning and maintenance projects based on data about community vulnerabilities and water infrastructure asset inventories from a major metropolitan area. We use a linear regression to investigate pipe failure rates, which serves as a foundation for the application of geographically weighted regression to uncover spatial patterns of community vulnerability. Results show that technical factors related to service disruptions are often collocated with challenging community and household-level conditions. For example, disruptions may be more severe in communities that rely heavily on public transit and where a significant portion of the population is not participating in the workforce. Further, findings show that some elderly community members could be more vulnerable during construction activities. Our findings provide insights into the intersection of community vulnerabilities and aging infrastructure, identifying key leverage points for capital investments that could strengthen community resilience.

1. INTRODUCTION

Water distribution failures and service outages create significant challenges, especially in communities facing socio-economic challenges [1-3]. For example, boil water notices are commonly issued as a precautionary measure to safeguard public health during system outages. These advisories can force residents to rely on alternative water sources, such as purchasing bottled water, which may increase household expenses [4]. Additionally, individuals with limited transportation options or mobility challenges may struggle to obtain water or face difficulties when repairs disrupt transport networks [5-7]. Water systems are interdependent with transportation systems and the communities they serve, forming complex socio-technical systems [8-11]. Consequently, water system operations are integral to community resilience—the ability of a community to anticipate, respond to, and recover from a crisis [12-13]. While it is well-established that water system operations are closely linked to community resilience [12], the mechanisms behind these connections are often poorly understood [13].

The socio-technical perspective on infrastructure systems [8-10] offers a framework for investigating interconnected social and technical subsystems. Within this framework, social-social interactions refer to the exchange of information between social actors [9], while technical-technical interactions involve the physical connections between infrastructure components. Social-technical (or 'socio-technical') interactions describe how social actors (e.g., operators or households) manage and adapt to the service levels provided by technical infrastructure. On the other hand, technical-social interactions refer to the flow of information from technical components to social actors. In this study, we focus on socio-technical relationships, connecting pre-existing community and household-level vulnerabilities to disruptions in infrastructure services [3]. A deeper understanding of these socio-technical interactions could be used to improve capital planning and management policies that strengthen community resilience.

Here, we investigate records of water distribution pipeline failures (n = 1,818) from 2013 to 2023 in a major metropolitan area, analyzing break rates per kilometer of pipe. We apply Ordinary Least Squares (OLS) regression alongside Geographically Weighted Regression (GWR) to identify the community and household-level vulnerabilities associated with the impacts of these failures [8-9]. To assess vulnerabilities at the census tract scale, we use 28 independent variables derived from a 2019 Neighborhood Equity Index (NEI), which is based on the World Health Organization's (WHO) Urban HEART (Health Equity Assessment and Response Tool) framework [14-17]. By uncovering spatial patterns of vulnerability, our findings help identify areas that may require targeted resilience investments and expedited response to service outages.

2. METHODS

In this section, we overview the infrastructure inventory data and community/household-level vulnerability indicators used in the analysis, as well as the application of statistical methods.

2.1 The Water Distribution Network

A major metropolitan water utility in North America provided water distribution pipeline data, break records (2013–2023), and transportation datasets under a restricted use agreement. The infrastructure inventory included 4,148 kilometers of pipe, with 1,818 recorded breaks and associated total outage duration in hours. To protect proprietary information, the data was aggregated and anonymized at the census tract level (2019) for analysis (Figure 1).

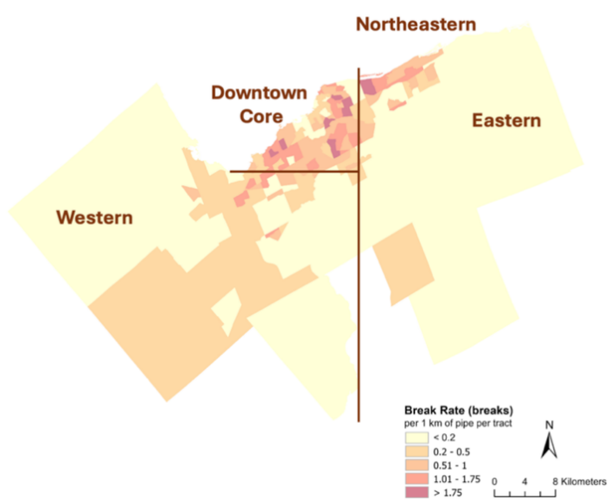

Figure 1: Break rates per kilometer of pipe from 2013 to 2023 in the study area. The study area is divided into geographic regions for discussion.

Breaks were randomly generated based on the recorded number of breaks per pipe, following a uniform distribution, $U(0, L_i)$, where L_i represent the length of each of the i pipes in which breaks occurred [1,11]. Breaks were aggregated to census tracts, and pipeline lengths for each tract were calculated for analysis. For the purposes of this study, census tracts with no population or with no water distribution pipeline inventory data were excluded. Final break rates (per kilometer of pipe, per census tract; n = 182), were calculated as the dependent variable 'Break rate (per 1-km of pipe)'. We include 'Total outage (hours)' as a control variable to account for the impact of service disruptions on the model.

2.2 Socio-Technical Interactions and Community Resilience

To capture community and household-level vulnerabilities at the census tract scale, we use 28 independent variables from a 2019 Neighborhood Equity Index (NEI) [14-15]. We classify these variables into a socio-technical systems framework in Table 1, including relevant variables from the infrastructure inventory data, along with descriptive statistics from the study area.

Table 1: Dependent and independent variables in a socio-technical systems framework

Type	Variable	Min/Max*	Mean*	Std. Dev*	Median*
Technical	Break rate (per 1-km of pipe)	[0,3.10]	0.52	0.60	0.29
	Total outage (hours)*	[0,386]	51	65	27
Social (Household)	Median employment income	[0,0.93]	0.50	0.20	0.50
	No secondary education (20-24 year olds)	[0,1]	0.18	0.17	0.15
	Working part-time	[0,0.80]	0.32	0.15	0.31
	Housing affordability (renters)	[0,1]	0.60	0.17	0.61
	Non-mortgage consumer debt	[0,1]	0.28	0.19	0.26
	Housing tenure (moved in last 5 years)	[0,1]	0.33	0.20	0.30
	Not participating in the workforce	[0,0.79]	0.25	0.14	0.20
	Financial assets	[0,0.99]	0.52	0.19	0.53
	Working poor (18-65 years old)	[0,0.88]	0.36	0.19	0.30
	Low income	[0,0.80]	0.18	0.17	0.12
	Seniors living alone	[0,1]	0.30	0.21	0.24
Social (Community)	Housing affordability (owner)	[0.05,1]	0.28	0.12	0.24
	Rental housing (availability)	[0,1]	0.66	0.26	0.76
	Diabetes health issues	[0,1]	0.23	0.09	0.23
	Proximity to childcare	[0,1]	0.91	0.10	0.94
	Community meeting spaces	[0,0.99]	0.53	0.25	0.56
	Public transit availability	[0,0.98]	0.48	0.19	0.50
	Employment in the community	[0,1]	0.78	0.28	0.90
	Commute times	[0,0.91]	0.39	0.19	0.39
	Senior (65+) ER visits	[0.27,1]	0.34	0.16	0.31
	Crime again persons	[0,1]	0.07	0.19	0.03
	Crime again property	[0,1]	0.09	0.14	0.04
	Usable green space	[0,0.95]	0.59	0.19	0.60
	Pedestrian and cyclist collisions	[0,1]	0.08	0.11	0.05
	Walkability score	[0,1]	0.48	0.24	0.50
	Mental health and substance abuse	[0,1]	0.10	0.10	0.08
	Early childhood development index	[0,1]	0.53	0.21	0.47
	Post-secondary education (25-29 years old)	[0,0.79]	0.39	0.15	0.39

*All community and household-level variables are standardized to a [0,1] scale. A value closer to 1 indicates a stronger negative effect (i.e., variables where higher values correspond to worse outcomes), while values closer to 0 indicate a stronger positive effect (i.e., variables where lower values correspond to better outcomes). For technical variables, break rates are in units of break rate, per 1-km of pipe (B_r) and outages are in hours (h). Total outage (hours) refers to the total number of outage hours reported per break, summarized by generated break per census tract for the period between 2013 to 2023.

2.3 Regression Analysis

Recent studies on pipe failure analysis in water distribution systems often employ statistical models to identify key factors associated with failures [5-6,11,18]. To investigate community and household-level vulnerabilities, we first employed an OLS regression to specify the global model, then used GWR to explore spatial variations in vulnerability across the study area [19-21]. These methods were chosen due to their interpretability and suitability for spatial data [19-20]. OLS identifies global trends, while GWR captures spatially varying relationships, which are central to our research questions on community resilience. While more advanced statistical methods or machine learning approaches could improve accuracy, capture non-linear patterns, and reduce bias, OLS and GWR provide transparency and a solid foundation for initial inquiry.

2.3.1 Ordinary Least Squares Regression

OLS estimates the parameters of a linear regression model, assuming a linear relationship between the dependent and independent variables, with constant coefficients (Equation 1):

$$(1)\ \mathrm{y} = \beta_0 + \sum_{\mathrm{k}=1}^{\mathrm{p}} \beta_{\mathrm{k}} \mathrm{x}_{\mathrm{k}} + \varepsilon$$

In Equation 1, y is the dependent variable (break rates), β_0 is the intercept, x_{k} are the independent variables, β_{k} are the regression coefficients, ε is the error term, and $p = 29$ is the number of predictors from Table 1. For variable selection, we applied backward selection, iteratively removing variables with p-values greater than the marginal significance threshold of p = 0.1 [22]. We chose a marginal significance threshold of p = 0.1 for backward selection to retain more features at each step, ensuring that variables were not excluded prematurely [23]. This threshold has been used in other research focused on human-infrastructure interactions, striking a balance between retaining relevant variables while avoiding the overfitting that can occur with stricter thresholds [24-26]. The final model was chosen using the corrected Akaike Information Criterion (AICc), with lower AICc values indicating a more optimal model [27]. To assess multicollinearity, we calculated the Variance Inflation Factor (VIF), setting a threshold of 10, consistent with prior studies [18,28]. We then examined the correlations between the final selected independent variables using Pearson's correlation coefficient. Model performance was evaluated using the coefficient of determination (R^2), which measures the proportion of variance in the dependent variable explained by the model, and the F-statistic, which tests the overall significance of the model [20]. To assess spatial autocorrelation in the residuals, we employed Moran's I test, which analyzes both continuous edges and corners within the neighborhood structure of the census tract data [19].

2.3.2 Geographic Weighted Regression

GWR (Equation 2) is an extension of the OLS regression method that allows for varying parameters across locally defined neighborhoods, accounting for nonstationary spatial relationships between independent and dependent variables in the study area [19-21].

$$(2)\ y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \epsilon_i$$

In Equation 2, the coordinates (centroid) or unique index (FID) of each census tract location (u_i, v_i) are incorporated into the regression model to estimate location-specific parameters. The y_i represents the dependent variable at census tract location i, for $1 \leq i \leq n$ $(n = 182)$, u_i and v_i are the spatial coordinates (centroid) or index of census tract location i, $\beta_0\ (u_i, v_i)$ is the intercept specific to location i, β_k (u_i, v_i) denotes the local estimate for the k-th predictor x_{ik} at location i, and ϵ_i is the error term. The golden section search method was used to find the optimal number of nearest census tract neighbors (g = 74) for the model, based on AICc [20-21]. We applied the Gaussian local weighting scheme to ensure that closer observations are given more influence. Consequently, the weighting distance was adaptively scaled according to the size and distribution of neighboring census tracts. We evaluated the final model performance using the coefficient of determination (R^2) and AICc for comparison with OLS regression results.

3. RESULTS

The OLS model was well-specified, explaining 83% of the variance in the dependent variable ($R^2 = 0.83$). To address issues of non-normality and heteroscedasticity, we applied log transformations to the break rate and outage variables, improving model fit. Standard scaling was applied to all independent variables. The community meeting spaces variable was excluded from the model due to its high VIF. We performed backward stepwise selection, improving the Akaike Information Criterion (AIC) from 423.85 to 394.98, indicating a more parsimonious model with fewer predictors at each step. The final corrected AIC (AICc) was 398, where AICc was chosen to account for the small sample size and the number of predictors in the model.

The final model was found to be significant, with an F-statistic of 119 ($p = 0.00$) and included seven predictors and the intercept (Table 2). Diagnostic tests confirmed that the residuals were normally distributed (Jarque-Bera test = 1.77, $p = 0.41$) and there was no evidence of heteroscedasticity (Breusch-Pagan test = 9.22, $p = 0.24$). Further, there was no evidence of multicollinearity in the final model with all VIFs below 3, and no pairwise correlation exceeding 0.7 among the predictors. We controlled for the total outage variable and found that removing it significantly reduced the explanatory power of the model, which then explained only 39% of the variance in the dependent variable ($R^2 = 0.39$), with the AICc increasing to 623. Similarly, the strength of the coefficients for public transit availability, commute times, and housing tenure were noticeably weaker, while the variable for not participating in the workforce was stronger without the outage variable. However, the direction and statistical significance of these factors remained the same in the model without control, and the F-statistic remained significant (16.5, $p = 0.00$). As expected, this shows the importance of the outage variable in its role as a moderating factor in these socio-technical relationships. Finally, there was no evidence of spatial autocorrelation in the residuals (Moran's I = 0.03, $p = 0.46$).

Table 2: Summary of OLS model results and GWR coefficient ranges

Type	Variable	Coef.	Std. Error	t-stat.	p-value	95% CI Lower	95% CI Upper	Coef. Range (GWR)*
Technical	Total outage (hours)	1.26	0.06	21.31	0.000	1.14	1.38	[1.16, 1.49]
Social (Household)	Not participating in the workforce	0.22	0.06	3.82	0.000	0.10	0.33	[0.15, 0.20]
	Housing tenure*	-0.13	0.07	-1.88	0.062	-0.27	0.01	[-0.16, 0.01]
Social (Community)	Commute times*	-0.30	0.07	-4.55	0.000	-0.43	-0.17	[-0.36, -0.19]
	Employment in the community	0.31	0.06	5.34	0.000	0.19	0.42	[0.24, 0.36]
	Public transit availability*	-0.24	0.07	-3.23	0.001	-0.39	-0.09	[-0.30, -0.15]
	Senior (65+) ER visits*	0.10	0.06	1.82	0.070	-0.01	0.21	[0.06, 0.11]
	Constant	-1.55	0.05	-29.91	0.000	-1.66	-1.45	[-1.60, -1.50]

* Senior (65+) ER visits refer to the rate of emergency department visits due to falls among seniors. Housing tenure is represented by the percentage of the population who moved in the last 5 years. Commute times refer to the percentage of people spending over 45 minutes on their commute. Public transit availability is represented by a transit score that measures availability and "usefulness" to the community from high to low transit access. Coefficient range is from the result of geographically weighted regression. The strength of coefficients (Figure 2) is shown as the standard deviations of these values. Total outage (hours) is used as a control variable to account for service disruptions in the model.

After the OLS model was well-specified, we ran the GWR model using the same seven predictors. The GWR model outperformed the OLS model, explaining 85% of the variance in the dependent variable ($R^2 = 0.85$), and the AICc improved to 392, indicating a better fit. These results suggest that, despite the absence of global spatial autocorrelation (Moran's I = 0.03, $p = 0.46$), the local spatial variations captured by GWR provided a more accurate representation of the data than the global OLS model. By visualizing the GWR coefficients (Figure 2), clear spatial trends emerge, revealing clusters and separations in areas with variation in coefficient strength.

Figure 2: Spatial distribution of coefficient strengths from GWR

The coefficient strengths shown in Figure 2 illustrate that the statistically significant predictors vary across the study area. The range of coefficient strengths from the GWR model is provided in Table 2. In this context, the OLS coefficient can be viewed as the global average across the entire study area. The final spatial distribution of residuals in both models, shown in Figure 3, exhibit a pattern of complete spatial randomness (CSR), except for a linear trend observed in the residual plot (Figure 4; lower left). This suggests that non-linear relationships may exist in the data that are not being captured by the model (Ramsey RESET test: $F = 16.80$, $p = 0.00$). However, the impact of this nonlinearity on the model was minimal (Durbin-Watson test: 2.06, which shows no significant autocorrelation in the residuals).

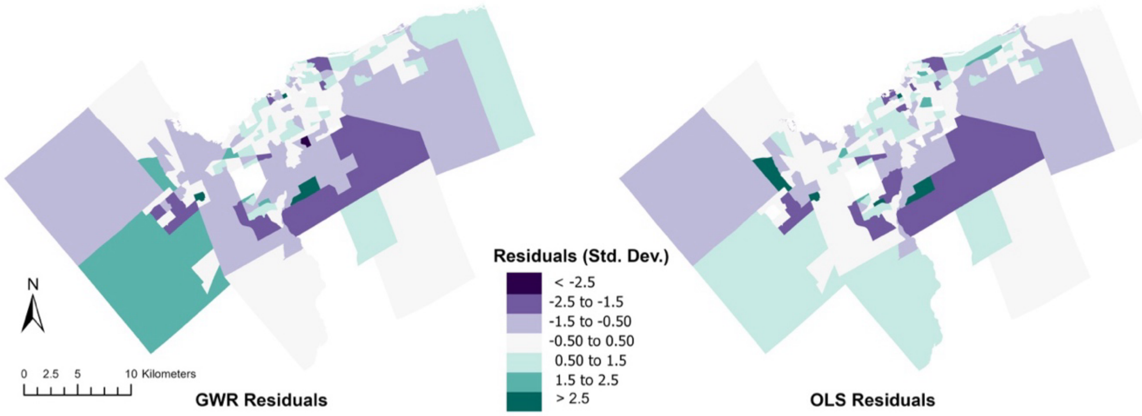

Figure 3: Spatial distribution of residuals from GWR (left) from OLS (right)

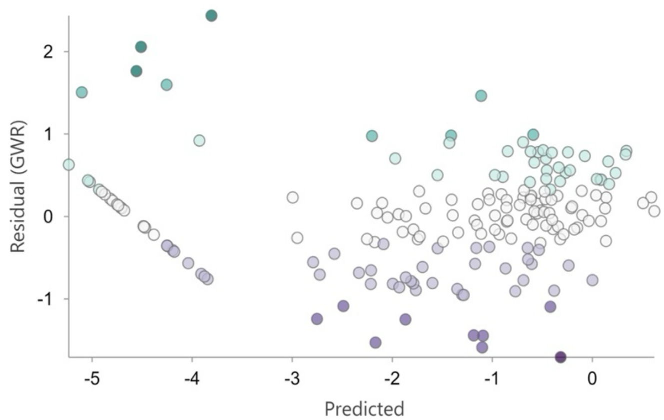

Figure 4: Residuals plot with color-coded GWR residuals. Dark purple indicates residuals more than 2.5 standard deviations below the mean from GWR, while dark green represents residuals more than 2.5 standard deviations above the mean from GWR.

4. DISCUSSION

The OLS model results reveal several associations between community or household-level vulnerabilities and break rates. Unsurprisingly, the length of outage is associated with increased break rates ($p = 0.00$). These findings align with existing literature that suggests pipe failures are closely associated with the technical dimensions of infrastructure systems [29-30]. In the downtown core, where infrastructure is older and population density is higher in this specific city, disruptions in public transit or water service may disproportionately affect a larger number of residents. Results show that communities that have higher public transit availability tend to experience higher break rates. Additionally, communities with shorter commute times show an association with increased break rates. Areas with shorter commute times may experience more immediate impacts from service interruptions, as residents rely heavily on these systems for day-to-day activities. In contrast, neighborhoods on the outskirts of the city, such as those in the eastern parts, may be less reliant on public transit and have greater access to personal transportation. This pattern may be attributed to the stronger effects observed in census tracts within the downtown core, where aging infrastructure is more likely to be present. Previous research has shown that water main breaks are often collocated with transportation disruptions [1,5, 31-32], which further supports the importance of addressing the connections between transportation and water for capital planning decisions.

Similarly, the strength of the relationship between higher break rates and employment is stronger in census tracts within the eastern parts of the city ($p = 0.00$). This suggests that disruptions to water service may contribute to socio-economic challenges in these areas when they occur. Additionally, areas with a higher percentage of the population aged 25 to 65 who are not participating in the labor force tend to have higher break rates. Together, the factors of transportation disruptions, business interruptions, and socio-economic strains—such as lower labor force participation—could compound the challenges these communities face in accessing emergency resources during prolonged water service outages [4].

Communities with more stable housing tenure (i.e., a greater proportion of the population has remained in place over the past five years) tend to have higher break rates (marginally significant; $p = 0.06$). This stability may be explained by the fact that downtown core areas are highly desirable to residents—leading to less frequent turnover—but also tend to have older infrastructure. Finally, communities impacted by high break rates show a marginally significant association with increased emergency department visits among seniors (age 65+) due to falls ($p = 0.07$). These findings draw attention to the increased vulnerability of elderly community members. For example, additional safety measures may be needed to reduce the risk of falls in construction zones set up for water pipeline repairs and maintenance in census tracts within the eastern parts of the city [7].

The practical implications of our findings can be further explored through the spatial variability captured in the GWR model (Figure 3). For example, water service outages are associated with break rates in areas with older infrastructure, particularly in census tracts within the northeastern part of the city. Capital improvement projects in these areas could reduce the relative impacts from disruptions in comparison with other areas. Similarly, the availability of public transit and increased commute times, both statistically significantly associated with higher break rates, show stronger effects in the downtown core (Figure 2). Special considerations for transportation interdependencies should be made for capital improvement planning in these areas. By prioritizing transportation system upgrades, improvements to the water system operations may also be achieved. In summary, our findings reveal spatial patterns of vulnerability, identifying priority areas where targeted resilience investments and faster responses to service outages could provide the greatest benefit to community members.

4.1 Limitations

Our modeling approach has some limitations. While backward selection in OLS is simple and interpretable, it can introduce biases like overfitting and exclusion of important variables. Methods like the Least Absolute Shrinkage and Selection Operator (LASSO) and forward selection address some of these issues but also have limitations, such as model complexity in LASSO and similar bias in forward selection due to the order of variable inclusion [19]. Here, we chose backward selection for its simplicity and transparency, while acknowledging these limitations.

A theoretical limitation of GWR is its focus on local parameter estimates. For instance, the Golden Section algorithm identifies a local optimum for the number of nearest neighbors, which may not be the global optimum. While GWR functions as an ensemble of local regressions, it does not account for the spatial interconnections between coefficients across locations. Additionally, some variation in coefficient estimates may occur due to randomness in the model results. The observed lack of spatial autocorrelation (Moran's $I = 0.03$, $p = 0.46$) also suggests no significant global pattern. However, further exploration is needed to determine whether local clusters or anomalies might exist that affect the GWR results. Consequently, although GWR captures spatial heterogeneity effectively, it is better suited for descriptive analysis than for causal inference. To address these challenges, we map the standard deviation of coefficient strengths to reveal any underlying spatial trends for each variable. Geographic terms like 'eastern,' 'northeastern,' and 'downtown core' describe these trends, but interpretations should be made within the context of the census tract scale and neighboring census tracts with a comparable standard deviation range of coefficient strength.

Here, we apply GWR to investigate the socio-technical relationships between water service disruptions and impacts to communities, which should be understood as reflecting a descriptive analysis of localized relationships rather than as establishing definitive causal effects. It is important to note that increased break rates have been causally connected in the literature to aging infrastructure, pipe diameter, materials, time of year, and average monthly temperatures [29-30]. However, investigation of these technical-technical relationships [9-10] is beyond the scope of the current study.

5. CONCLUSIONS

Our results suggest that technical factors, such as break rates, are closely linked to select community and household-level challenges. Specifically, we found that communities that are likely dependent on public transit, with jobs located within the area, and with a significant portion of the population not participating in the workforce tend to experience higher water system break rates. Similarly, our findings show the increased vulnerability of elderly community members, who face greater challenges during service disruptions. Given these findings, it is essential to consider the safety and specific needs of seniors when planning infrastructure maintenance and capital improvement projects. Our findings offer valuable insights into the intersection of socio-economic vulnerabilities and aging infrastructure, helping to identify key leverage points that could improve community resilience. Future work will examine the socio-technical factors influencing community vulnerability at a more granular level, using advanced statistical methods and machine learning techniques. Ultimately, this preliminary work lays the groundwork for more integrated approaches to infrastructure management. Water utilities and city planners can use the results of this study to integrate socio-technical perspectives and findings related to senior populations into resilience strategies, as well as inform policy decisions that link water and transportation capital planning.

ACKNOWLEDGMENTS

Special thanks to Maggie Roberts for her GIS analysis and mapping advice.

REFERENCES

[1] Toland, J., and L. A. Spearing. 2024. “A transportation disruption metric for emergency household food and water access after earthquakes.” In Vol. 496 of Proc., Canadian Society for Civil Engineering, Annual Conf., edited by S. Desjardins and G. J. Poitras. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-031-60419-5_10.

[2] Soden, R., Lallemant, D., Kalirai, M., Liu, C., Wagenaar, D., Jit, S. 2023. “The importance of accounting for equity in disaster risk models”. *Commun Earth Environ* 4, 386. https://doi.org/10.1038/s43247-023-01039-2

[3] Coleman, N., Esmalian, S., and Mostafavi, M. 2020. "Equitable Resilience in Infrastructure Systems: Empirical Assessment of Disparities in Hardship Experiences of Vulnerable Populations during Service Disruptions." *Natural Hazards Review* 21, no. 4 (June 18, 2020). https://doi.org/10.1061/(ASCE)NH.1527-6996.0000401.

[4] Yang, E. and K. Faust. 2019. "Human–Water Infrastructure Interactions: Substituting Services Received for Bottled and Filtered Water in US Shrinking Cities." *Journal of Water Resources Planning and Management* 145, no. 7: 04019043. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001126.

[5] Rahimi-Golkhandan, A., B. Aslani, and S. Mohebbi. 2022. "Predictive Resilience of Interdependent Water and Transportation Infrastructures: A Sociotechnical Approach." *Socio-Economic Planning Sciences* 80: 101166. https://doi.org/10.1016/j.seps.2022.101166.

[6] Mazumder, R., A. Salman, Y. Li, and X. Yu. 2021. "Asset Management Decision Support Model for Water Distribution Systems: Impact of Water Pipe Failure on Road and Water Networks." *Journal of Water Resources Planning and Management* 147, no. 3: 04021004. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001365.

[7] Toland, J., A. Wein, A-M. Wu, and L. Spearing. 2023. "A Conceptual Framework for Estimation of Initial Emergency Food and Water Resource Requirements in Disasters." *International Journal of Disaster Risk Reduction* 90: 103661. https://doi.org/10.1016/j.ijdrr.2023.103661.

[8] Baxter, G. and I. Sommerville. 2011. "Socio-technical Systems: From Design Methods to Systems Engineering." *Interacting with Computers* 23: 4–17. https://doi.org/10.1016/j.intcom.2010.07.001.

[9] Ottens, M., M. Franssen, P. Kroes, and I. van de Poel. 2005. "Modelling Infrastructures as Socio-technical Systems." *International Journal of Critical Infrastructures*. https://doi.org/10.1504/IJCIS.2005.X.

[10] Manny, L., M. Angst, J. Rieckermann, and M. Fischer. 2022. "Socio-Technical Networks of Infrastructure Management: Network Concepts and Motifs for Studying Digitalization,

Decentralization, and Integrated Management." *Journal of Environmental Management* 318: 115596. https://doi.org/10.1016/j.jenvman.2022.115596.
[11] Toland, J., and A. Wein. 2021. "A Simplified Method for Rapid Estimation of Emergency Water Supply Needs after Earthquakes." *Water* 13: 2635.
[12] National Institute of Standards and Technology (NIST). 2016. “Community Resilience Planning Guide for Buildings and Infrastructure Systems, Volume I”.
[13] Toland, J. and L. Spearing. 2023. "Water Utilities and Equity in Disasters: A Systematic Literature Review." In ASCE Inspire 2023, 17. https://doi.org/10.1061/9780784485163.042.
[14] World Health Organization. 2010. Urban Health Equity Assessment and Response Tool. World Health Organization.
[15] Social Planning Council of Ottawa. 2019. “Neighborhood Equity Index: Project Report”. https://neighbourhoodequity.ca/wp-content/uploads/2021/02/NEI-Report-Feb-22-2001.pdf. Accessed September 10, 2024.
[16] City of Toronto. 2014. “TSNS 2020 Neighbourhood Equity Index: Methodological Documentation”. City of Toronto.
[17] City of Toronto. 2014. “Urban HEART @ Toronto: Technical Report / User Guide”. City of Toronto.
[18] Karadirek, I., E. Kaya-Basar, and T. Akdeniz. 2024. "A Study on Pipe Failure Analysis in Water Distribution Systems Using Logistic Regression." *Water Supply* 24, no. 1: 176.
[19] Wheeler, D., and A. Páez. 2010. "Geographically Weighted Regression." In *Handbook of Applied Spatial Analysis*, edited by M. Fischer and A. Getis, 317-331. Berlin: Springer.
[20] Fotheringham, A, C. Brunsdon, and M. Charlton. 2009. "Geographically weighted regression." *The Sage Handbook of Spatial Analysis* 1: 243-254.
[21] Koh, E-H., E. Lee, and K-K Lee. 2020. "Application of Geographically Weighted Regression Models to Predict Spatial Characteristics of Nitrate Contamination: Implications for an Effective Groundwater Management Strategy." *Journal of Environmental Management* 268: 110646.
[22] Derksen, S. and H. Keselman. 1992. "Backward, Forward and Stepwise Automated Subset Selection Algorithms: Frequency of Obtaining Authentic and Noise Variables." *The British Journal of Mathematical and Statistical Psychology* 45, no. 2: 265–282.
[23] Bursac, Z., C. Gauss, D. Williams and D. Hosmer. 2008. "Purposeful Selection of Variables in Logistic Regression." *Source Code for Biology and Medicine* 3: 17.
[24] Toland, J., A. Whelton, C. Wukich and L. Spearing. "What Drives Household Protective Actions in an Industrial Crisis? Insights from the East Palestine Train Derailment." Sustainable Cities and Society 115 (2024): 105867.
[25] Haque, M., A. Rahman, D. Hagare and R. Chowdhury. A Comparative Assessment of Variable Selection Methods in Urban Water Demand Forecasting. Water. (2018): 10(4):419.
[26] Gedefaw, M., Hao, W., Denghua, Y., & Girma, A. (2018). Variable selection methods for water demand forecasting in Ethiopia: Case study Gondar town. Cogent Environmental Science, 4(1). https://doi.org/10.1080/23311843.2018.1537067
[27] Bozdogan, H. 1987. "Model Selection and Akaike’s Information Criterion (AIC): The General Theory and Its Analytical Extensions." *Psychometrika* 52, no. 3: 345–370. Springer; The Psychometric Society.
[28] Diress, F., Y. Negesse, D. Worede, D. Ketema, W. Geitaneh, and H. Temesgen. 2024. "Multilevel and Geographically Weighted Regression Analysis of Factors Associated with Full Immunization among Children Aged 12–23 Months in Ethiopia." *Scientific Reports* 14, no. 22743. https://doi.org/10.1038/s41598-024-22743-3.
[29] Zamenian, H., F. Mannering, D. Abraham, and T. Iseley. 2016. "Modeling the Frequency of Water Main Breaks in Water Distribution Systems: Random-Parameters Negative-Binomial Approach." Journal of Infrastructure Systems 23, no. 2: Article 04016009.
[30] Folkman, S. 2018. *Water Main Break Rates in the USA and Canada: A Comprehensive Study*. Utah State University, Buried Structures Laboratory.
[31] Chang, S., A. Brown, R. Costa, B. Dobson, H. Dowlatabadi and T. Haukaas. 2017. “A community resilience approach to assessing transportation risk in disasters”, 16th World Conference on Earthquake, WCEE 2017, Santiago.
[32] Soga, K., R. Wu, B. Zhao. and C. Wang. 2021. “City-Scale Multi-Infrastructure Network Resilience Simulation Tool; PEER Report No. 05”. Pacific Earthquake Engineering Research Center Headquarters at the University of California: Berkeley, CA, USA.