4th Construction Specialty Conference
4e Conférence spécialisée sur la construction

Montréal, Québec
May 29 to June 1, 2013 / *29 mai au 1 juin 2013*

Estimation of Flushing Duration for Preventive Maintenance of Wastewater Collection System

Hamid Zaman[1], Ahmed Bouferguène[2], Mohamed Al-Hussein[1], Carl Lorentz[3], Derek Melmoth[3]
[1]Department of Civil and Environmental Engineering, University of Alberta
[2]Campus Saint-Jean, University of Alberta
[3]Drainage Services - Operations, City of Edmonton

Abstract: Preventive maintenance of drainage networks is an essential component of urban infrastructure management. Large cities require significant expenditures of capital and other resources to perform the necessary prescheduled cleaning and flushing activities at various locations around the city at regular intervals. However, planning and scheduling of these activities can be challenging because of the wide variation in actual on-site flushing time, which depends upon such factors as location, pipe properties, frequency of flushing, time of day, and season. This study develops a model for estimating on-site high-pressure flushing (HPF) duration based on such predictor variables. The model is developed using historical data from the operations division of the Drainage Services Branch at the City of Edmonton, where a 5,500-kilometre drainage network is maintained through more than 1,400 prescheduled preventive maintenance routes for HPF. The panel dataset utilized in this study has been obtained by integrating several databases, one of which is the historical data (spanning over 4 years) collected using the automatic vehicle locators (AVLs) installed in the flushing trucks. Preliminary analysis shows that the actual on-site flushing duration varies from 10 minutes to 5 hours, primarily depending on the number and length of pipes to be flushed at a particular location; however, the panel dataset provides the opportunity to incorporate the effect of other factors in the model (such as those mentioned above). The outcome of this study is a flushing duration estimation model that can be used for resource optimization, route scheduling, and performance evaluation.

1 Introduction

A wastewater collection system consists of sanitary, storm, and/or combined pipelines, lift stations, force mains, and other facilities to collect wastewater from residential, industrial, and commercial sources and convey it to facilities that provide treatment prior to discharge to the environment (Poltak 2003). Preventive maintenance (PM) of the collection system refers to pre-scheduled inspection, flushing, and regular cleaning of various components (pipes, catch basins, manholes, etc.) in order to maintain a standard level of service. (Note that these activities differ from typical maintenance or repair works; however, "maintenance" is the widely accepted industry term used to refer to these cleaning and flushing works, and thus has been used in this paper.) Large cities require a significant budget and the resources to carry out the necessary preventive maintenance work across the city. Although technological innovation has led to advancements in maintenance tools and techniques that make operation activities more effective, the aging of the system coupled with urban growth necessitates continuous improvement in PM performance (Gaudreault and Lemire 2006).

Recent literature on wastewater collection systems focuses on various aspects, such as inspection technologies, flow modeling, design and construction of sewer systems, operation and maintenance, environmental protection, pollution control, pumping, rehabilitation, overflow, and odour control (Hollenbeck 2004, Clark et al. 2007, Vallabhaneni 2012). However, a thorough review of this literature has revealed that only a few studies, such as Mohamed et al. 2002, Tafuri et al. 2002, Bowen et al. 2003, Miles et al. 2004, Knapp et al. 2004, and Agbulos et al. 2006, have focused on the performance measurement and productivity improvement of maintenance activities from the management perspective. An ongoing collaborative research and development program between the authors' respective organizations primarily focuses on productivity improvement for drainage operations. Some research outcomes of this study were presented at the CSCE Annual Conference in 2012, where an optimization algorithm which minimizes the travel distance of maintenance vehicles while maximizing the daily effective work time was presented (Zaman et al. 2012). From the preliminary results, the algorithm was found to be useful for the planning and scheduling of infrastructure maintenance work where multiple locations, each requiring a certain maintenance duration, are covered within a specific time interval (e.g., an 8-hour shift). However, it was found that such an algorithm cannot be applied effectively without accurate estimates of on-site work duration for each location. This paper, therefore, focuses on developing models for estimation of on-site flushing duration for PM activities.

2 Preventive Maintenance of Collection System

Edmonton has a large drainage infrastructure with a replacement value of $14.9 billion (as of 2010), with a collection system comprising 5,500 km of pipes and 332,000 service connections (City of Edmonton 2012). In order to maintain a high standard of service by keeping this large network running efficiently, comprehensive PM activities are carried out throughout the city on a regular basis. Established techniques and advanced equipment are used for various PM activities, such as visual inspection (VI), low-pressure flushing (LPF), high-pressure flushing (HPF), catch-basin cleaning (CBC), and mainline televising (MTV). This study focuses particularly on the development of an on-site flushing duration model for HPF, as this activity consumes the greatest number of man-hours among all PM activities. The process of scheduled HPF is briefly described below.

As part of the annual PM program, scheduled HPF is performed at 1,400 pre-designated locations within the city. These locations are referred to as "routes", each containing one or more pipe sections. Since some routes require more frequent flushing than others, each of the routes is pre-scheduled for periodic HPF at a particular frequency (every 1 month, 3 months, 6 months, or 12 months). At the beginning of each month, a database generates the list of HPF job orders for the routes that are due that month. These job orders, grouped by location, are passed on to the drainage supervisor, who then assigns a set of jobs to each of the individual crews. Each day, the crews travel to the assigned route locations and perform HPF using state-of-the-art flushing equipment (Figure 1). The number of routes flushed in a typical 8-hour shift varies widely, depending on the flushing duration at each location and travel time between locations. Although the on-site flushing duration at a particular location depends primarily on the number and length of pipes within the route, it is also affected by such factors as season, time of day, crew performance, and pipe properties. This study thus develops the on-site flushing duration model using data from various sources in order to capture the wide variability.

Figure 1: PM vehicle (combo unit) performing high-pressure flushing

3 Data for Model Development

3.1 Data Collection

The dataset used in this study has been collected and merged from several databases. Figure 2 presents the four (4) databases linked to create the modeling dataset, as well as the variables drawn from each data source. Drain data provides the physical properties of the pipes, such as diameter, length, slope, material, year of construction, and location. SAP databases have been used to collect the scheduled route information (route number, pipes within the route, route frequency, location, scheduled flushing date, etc.) and crew information (the crew assigned for a particular route, vehicle ID, flushing date). The actual on-site flushing duration data have been collected from the automatic vehicle locator (AVL) database, which records the historical location, time, and speed data for all vehicles involved in PM activities. The on-site flushing durations for all scheduled routes are then linked with the parameters obtained from the other databases. It should be noted that the drainage department and SAP databases are inter-connected through unique IDs (Pipe ID, Route ID, Crew ID). However, the AVL database does not share any primary/foreign keys with the other databases, necessitating that potential connections between the AVL database and the other databases (flushing date, location, and vehicle IDs) be utilized in order to look up manually the vehicle used by a particular crew on a particular day and then search for that vehicle's stop durations near the given job location. This process involves the assumption that a crew is performing flushing activities when its vehicle is found to be idle (at a stationary position with engine running) at the location of the route scheduled for that day.

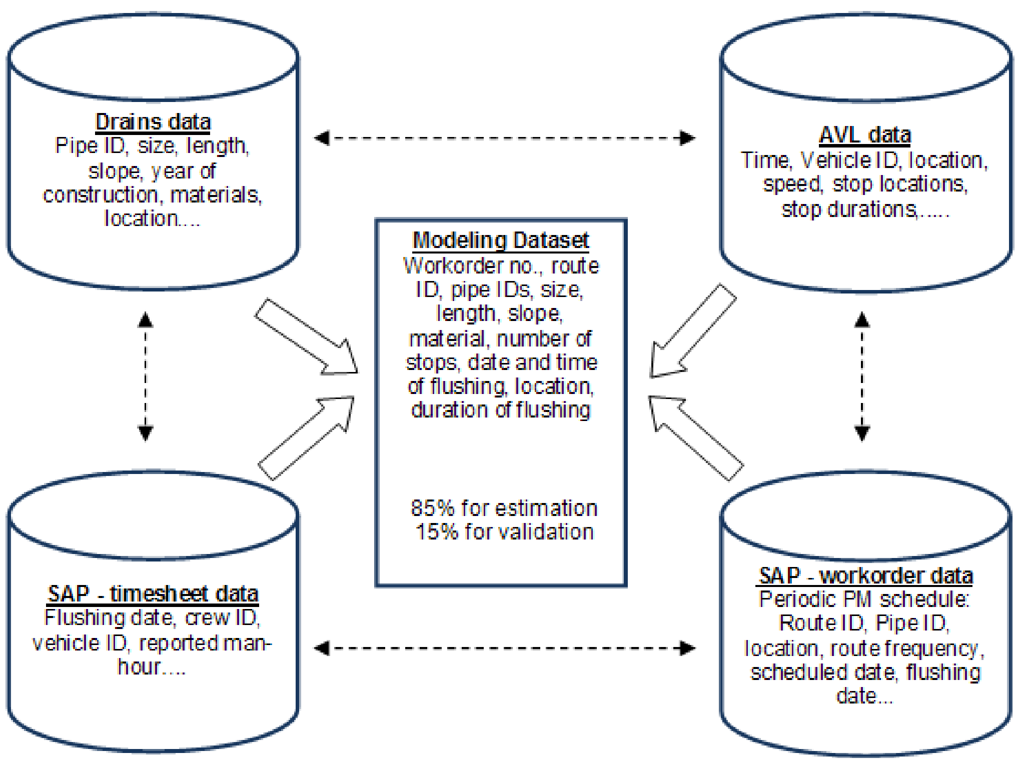


Figure 2: Data collection schematic

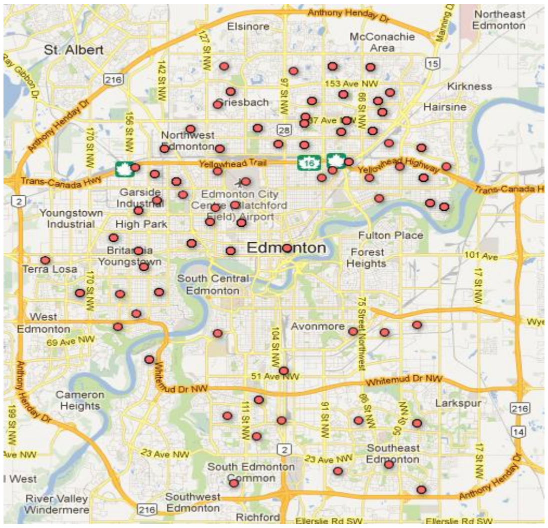


Figure 3: Spatial distribution of the routes in the modeling dataset

The collected dataset contained observations from the years 2009-2012 at various locations throughout the city (Figure 3). Following necessary cleansing of the database, the final dataset contained 448 observations. Among these, 85% (381) of observations were randomly selected for model estimation, while the remainder (67) were used for model validation. The list and descriptions of the variables in the dataset are presented in Table 1. It should be noted that the dataset is an unbalanced panel dataset in which each of the four (4) 1-month routes has multiple observations, while most of the other routes do not.

Table 1: List and description of variables in the modeling dataset

Variable Name	Description	Range	Variable Type
Flushing_duration	Total time taken to flush the route	10 ~ 339 minutes	Continuous
Number_of_pipes	Total number of pipes in the route	1 ~ 18 nos.	Discrete
Total_length	Total length of pipes in the route	3 ~ 1132 meters	Continuous
Number_of_stops	Number of locations the vehicle stops at to finish flushing the route	1 ~ 14 nos.	Discrete
Average_diameter	Average diameter of all the pipes in the route	15 ~ 67.5 cm	Continuous
Average_depth	Average depth of the downstream manholes of the pipes in the route	2 ~ 10 m	Continuous
Age_of_pipes	Average age of all the pipes in the route	14 ~ 105 years	Continuous
Flush_per_year	Number of flushing per year = (12/route_frequency)	12, 4, 2, 1	Discrete
Month	Month of flushing	Jan ~ Dec	Discrete
Day	Day (of the week) of flushing	Mon ~ Sun	Discrete
Time	Time (of the day) of flushing	Morning, Midday, Afternoon, Evening, Night	Categorical
Neighbourhood_type	Neighbourhood type for the route	Residential, Commercial, Industrial	Categorical

3.2 Descriptive Statistics of the Data

Initial descriptive statistical analyses have been performed in order to support a complete understanding of the dataset and the correlation between the variables. Preliminary results show that the on-site flushing duration varies from 10 to 339 minutes, with an average value of 70.93 minutes. The standard deviation, median, and mode of the data are 58.42, 51, and 29 minutes, respectively. To explicate the reasons behind such wide variation, flushing duration was plotted against the predictor variables (a portion of which is shown in Figure 4). As expected, flushing duration has a strong linear correlation with the number of pipes and total length of the routes; however, the pipe diameter and depth do not seem to affect flushing duration.

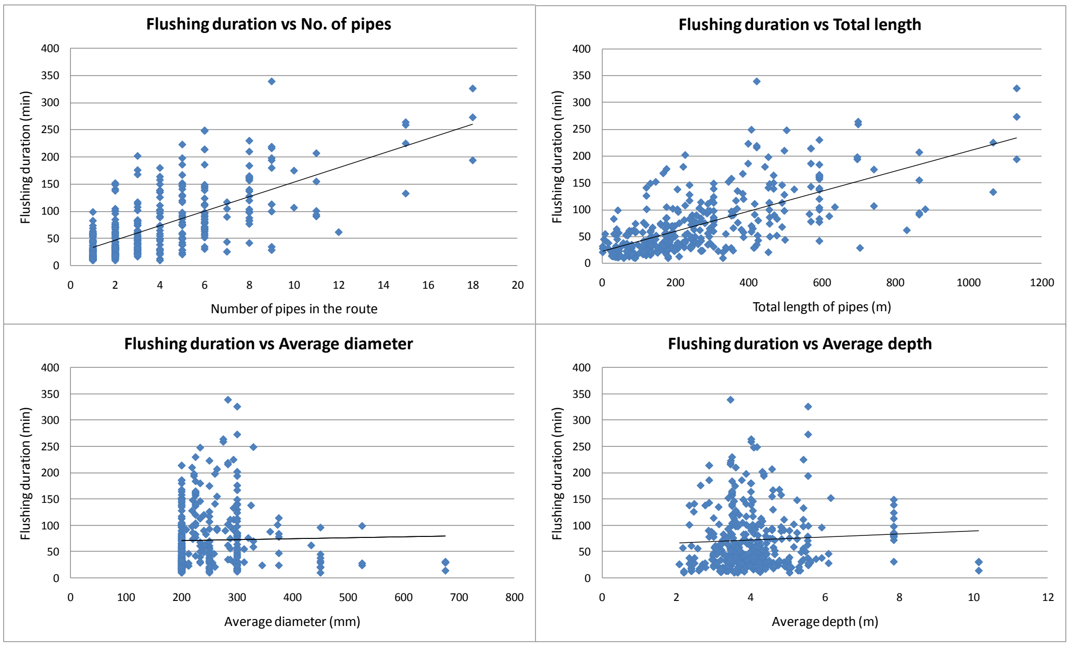
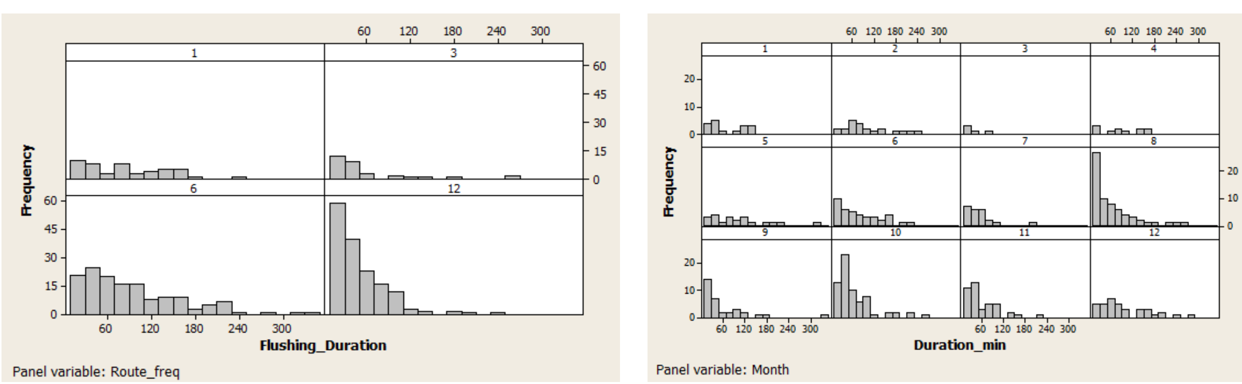


Figure 4: Flushing_duration vs. number_of_pipes, total_length, average_diameter, and average_depth

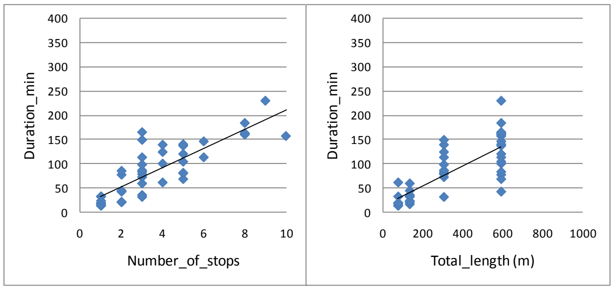
At this point it is of interest to explore the effect of the other predictor variables by analyzing subsets of the data. When the flushing durations are grouped by route frequency, different patterns for 1-, 3-, 6-, and 12-month routes can be observed. Similar variations are observed when the dataset is grouped by month (Figure 5), which suggests that route frequency and month affect the variation of flushing duration.



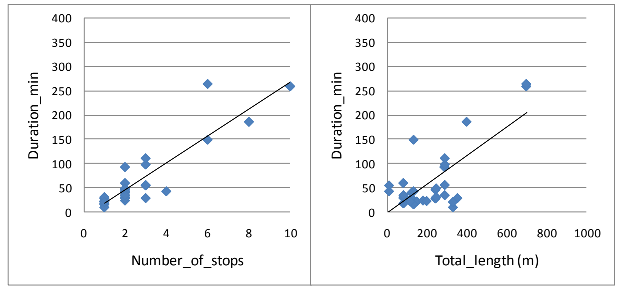
(a) (b)

Figure 5: Histogram of flushing_duration by (a) route frequency and (b) month

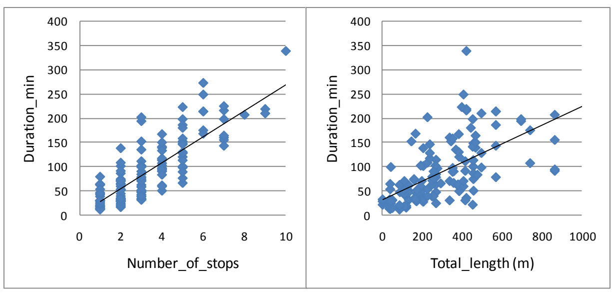
Figure 6 presents the scatterplots showing relationships between flushing duration and number of stops and total length, where strong linear relationships can be observed. Interestingly, “number of stops” shows a stronger correlation with flushing duration than does the “number of pipes”. One explanation for this is that, theoretically, the crews should stop at every manhole to access all the pipes in the route; however, in practice, experienced crew members flush two stretches of pipe from the same manhole when possible. This allows the crew to finish their job with fewer stops.



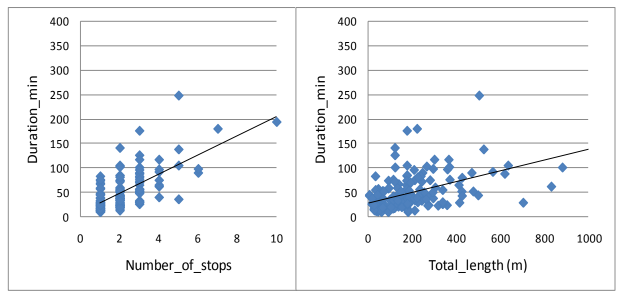
(a) 1-month routes



(b) 3-month routes



(c) 6-month routes



(d) 12-month routes

Figure 6: Scatterplots showing flushing_duration vs. number_of_stops and total_length

4 Model Development, Results and Discussion

It is quite clear from the preliminary analysis that the utilization of a multiple linear regression model should be sufficient to capture the majority of variations. However, owing to the fact that route frequency has a considerable effect on flushing duration in the estimation dataset, separate models for each route frequency are to be developed first. Moreover, the 1-month route subset of data has panel observations which need to be modeled separately in order to capture the temporal variation of flushing duration.

The general form of linear multiple regression is expressed as Equation 1 (Neter et al.1996):

$$[1] \qquad Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad \text{for } X_{i0} \equiv 1$$

Where,

Y_i = Flushing duration for route i
X_{ik} = Predictor variable k for route i
β_k = Parameter for variable k
ε_i = Independent $N(0,\sigma^2)$ Error term for route i
i = 1,2,3……..,n; where n is the total number of observation
p = number of predictor variables

However, for $E\{\varepsilon_i\} = 0$, the response function of Equation 1 for a particular route becomes:

$$[2] \quad E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_{p-1} X_{p-1}$$

Models in the form of Equation 2 have thus been developed with sequential addition of variables. The inclusion or deletion of each factor is performed based on its sign, T-stat, and its impact on the model's goodness-of-fit value (adjusted R-squared). It should be noted that all analyses and model estimations in this study have been conducted in Minitab® Statistical Software.

Table 2: Model results

Model	Predictor Variable	Coefficient	T-stat	P-value
1-month Routes	Constant	-41.26	-3.26	0.002
	Number_of_stops	12.966	6.53	0.000
	Total_length	0.12044	5.26	0.000
	Dia_square	0.05585	3.26	0.002
	Jan	23.988	2.59	0.013
	Feb	19.117	1.95	0.058
	Jul	-54.51	-3.41	0.001
		Adjusted R-squared = 85.3%		
3-month Routes	Constant	-27.14	-3.57	0.001
	Number_of_stops	21.42	8.69	0.000
	Total_length	0.121	3.77	0.001
	Midday	16.858	1.91	0.067
		Adjusted R-squared = 87.9%		
6-month Routes	Constant	-5.873	-1.09	0.277
	Number_of_stops	20.194	11.24	0.000
	Total_length	0.0595	3.85	0.000
	Midday	23.925	3.9	0.000
		Adjusted R-squared = 76.0%		
12-month Routes	Constant	13.17	3.39	0.001
	Number_of_stops	13.506	6.62	0.000
	Total_length	0.1057	3.95	0.000
	No_of_pipes	-4.654	-2.18	0.031
		Adjusted R-squared = 54.4%		
Final Model (All Routes)	Constant	-12.106	-2.15	0.032
	Number_of_stops	19.015	17.10	0.000
	Total_length	0.05711	5.69	0.000
	Flushing_per_year	-0.8724	-1.94	0.053
	Age_of_pipe	0.1792	1.94	0.053
	Midday	19.35	5.63	0.000
	Feb	14.82	2.22	0.027
	Dec	11.087	2.07	0.039
		Adjusted R-squared = 73.6%		

As can be inferred from Table 2, which presents the model results, the 1-month route model contains temporal variables with significant T-stats (greater than 1.64). Both the 1-month and 3-month route models have R-squared values greater than 0.85, implying that more than 85% of the variability in flushing duration has been captured in the models. The 6-month model also has a reasonable goodness-of-fit value (76%). However, the 12-month model has a poor fit; it contains two of the highly correlated variables identified previously—"number_of_stops" and "number_of_pipes", which both have significant

T-stats but have opposite signs. Careful investigation of the 12-month subset of data has revealed that the number of stops has weak correlations with number of pipes and total length. This is due to the fact that the average pipe lengths for 12-month routes are much shorter than those for other routes. Another possible explanation may be that the layout of the pipes is such that the crews can access multiple pipes from the same manhole.

However, based on the observations from the separate models, a final route-independent model has been developed which has a reasonable goodness-of-fit value (73.6%). The model contains seven statistically-significant parameters with expected signs and values. For example, the “midday” coefficient (used as a dummy variable in the model) implies that the flushing operation takes about 20 minutes longer than usual when performed between 11:00 a.m. and 1:00 p.m. This observation has to do with the fact that the crews usually take a short break for lunch around this time of day. The model also captures the variability of all four route frequencies by virtue of the “flushing_per_year” parameter. This result is complemented by a stepwise regression model developed in the statistical software, which produces very similar results with an adjusted R-squared value of 73.98%.

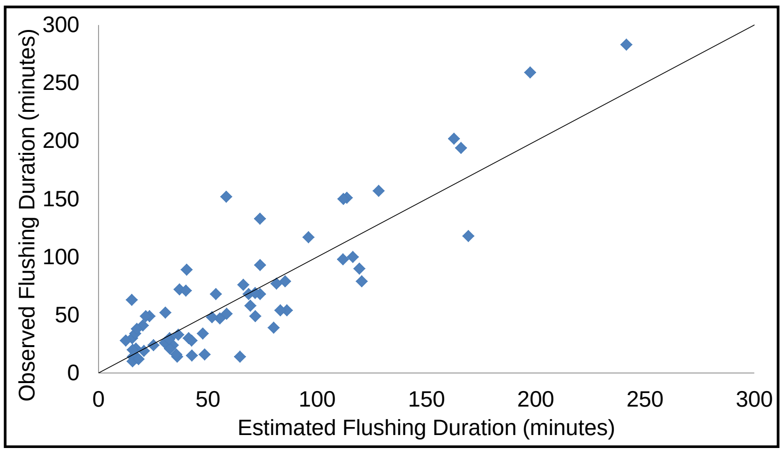


Figure 7: Estimated vs. observed flushing_duration

The final model is validated using the validation dataset (see Figure 7 for the resulting plot), from which it can be seen that the estimated and observed points are reasonably close to the 45° reference line. It is to be noted that more than 50% of the observations in the validation dataset are from the 12-month route subset, yet the model estimates the flushing duration fairly accurately. Nevertheless, the model errors (estimated minus observed) for each observation in the validation dataset have been calculated. The probability distribution function (PDF) of the errors, which is given in Figure 8, follows a slightly skewed normal distribution with parameter values of μ = -3.83 and σ = 28.12. This is the un-captured portion of the variability of on-site flushing duration, which is believed to be due to crew-specific errors. The nature of the flushing works is such that crew members use their judgment to perceive the cleanliness of the pipe during flushing in order to determine when to stop. Flushing duration thus depends upon the experience and the judgement of the crew. Incorporation of such crew-specific variables can potentially improve the model.

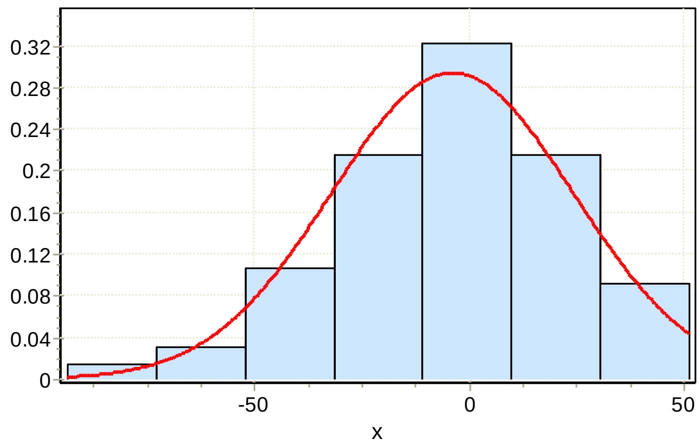


Figure 8: Probability Distribution Function of error

5 Conclusion

This study develops an on-site flushing duration estimation model for high-pressure flushing of collection system infrastructure, with a simple linear model estimated and validated using actual data from a Canadian city. In addition to the common route-specific variables (length of pipe, age of pipe), it captures temporal variations through the inclusion of variables representing time of day, as well as month. The goodness-of-fit value of 73.6% suggests that the model can estimate the flushing time with reasonable accuracy. This model can be used effectively for monitoring and benchmarking of on-site productivity, sensitivity analysis, resource optimization, and route scheduling. A limitation of the model is that it does not reflect variation related to crew judgment, especially when the most influential predictor variable, “number of stops”, is sometimes contingent upon the discretion of the crew. It is also believed that the model can be further improved by incorporating random effects derived from crew characteristics.

Acknowledgments

This research has been conducted by the Construction Engineering & Management research group at the University of Alberta, in collaboration with the Drainage Services Branch of the City of Edmonton. The authors would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support, and the City personnel for logistical and technical cooperation.

References

Agbulos, A., Mohamed, Y., Al-Hussein, M., AbouRizk, S., and Roesch, J. 2006. Application of Lean Concepts and Simulation Analysis to Improve Efficiency of Drainage Operations Maintenance Crews. *Journal of Construction Engineering and Management*, ASCE, 132(3): 291-299.

Bowen, A. Roesch, J., Mohamed, Y., AbouRizk, S., and Al-Hussein, M. 2003. Development of Productivity Standards for Collection Systems Maintenance Activities. *2003 Water Environment Federation Technical Exhibition and Conference*, Session 48: 803-813.

City of Edmonton. 2012. Capital Budget: Drainage Facilities. <http://www.edmonton.ca/city_government/budget_taxes/2012budget-drainage-facilities.aspx> (Jan., 2012).

Clark, S., Pitt, R., Burian, S., Field, R., Fan, E., Heaney, J., and Wright, L. 2007. Annotated Bibliography of Urban Wet Weather Flow Literature from 1996 through 2006. *U.S. Environmental Protection Agency, Urban Watershed Management Branch*, Edison, NJ, USA.

Gaudreault, V., Lemire, P. 2006. The Age of Public Infrastructure in Canada. Catalogue no. 11-621-MIE – No. 035, Statistics Canada.

Hollenbeck, A. 2004. Wastewater Collection Systems: Literature Reviews. *Water Environment Research,* 76(6): 668-691.

Knapp, T. and Eising, T. 2004. The Use of Sophisticated Analysis Tools to Determine Collection System Capacity and to Prioritize Capital Improvement Programs at the City of Folsom. *2004 Water Environment Federation Technical Exhibition and Conference*, New Orleans, LA, USA, Session 86: 378-393.

Miles, W., Moyer, J., and Ridge, J. 2004. Building a Better Business Case for Funding Sewer Rehabilitation. *2004 Water Environment Federation Technical Exhibition and Conference*, New Orleans, LA, USA, Session 7: 558-566.

Mohamed, Y., AbouRizk, S., Al-Hussein, M., Roesch, J., and Bowen, A. 2002. Using Simulation for Improving Productivity of Drainage Operations. *Canadian Society for Civil Engineering (CSCE) Annual Conference*, Montréal, QC, Canada.

Neter, J., Kutner, M.H., Nachtsheim, C., and Wasserman, W. 1996. Applied Linear Statistical Models. 4^{th} ed., Times Mirror Higher Education Group Inc., Burr Ridge, IL, USA.

Poltak, R.F. 2003. Optimizing Operation, Maintenance, and Rehabilitation of Sanitary Sewer Collection System. *New England Interstate Water Pollution Control Commission*. <http://www.mass.gov/dep/water/laws/omrguide.pdf> (Jan., 2012).

Tafuri, A. and Selvakumar, A. 2002. Wastewater Collection System Infrastructure Research Needs in the USA. *Urban Water,* 4(1): 21-29.

Vallabhaneni, S. 2012. Wastewater Collection Systems: Literature Review. *Water Environment Research*, 84(10): 845-860.

Zaman, H., Bouferguène. A., Al-Hussein, M., Lorentz, C., and Girard, L. 2012. Sustainable Practice in Infrastructure Management through Route Optimization for Preventive Maintenance of Wastewater Collection System. *1^{st}* International *Specialty Conference on Sustaining Public Infrastructure, CSCE*, Edmonton, AB, Canada.