A FRAMEWORK FOR IMPROVING THE PRODUCTIVITY OF OPERATIONAL PREVENTIVE MAINTENANCE ACTIVITIES FOR WASTEWATER COLLECTION SYSTEM

By:

Hamid Zaman

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Construction Engineering and Management

Department of Civil and Environmental Engineering

University of Alberta

ABSTRACT

Operational maintenance of the wastewater collection system is an important part of urban infrastructure management. It involves various activities, such as visual inspection, low-pressure flushing (LPF), high-pressure flushing (HPF), catch basin cleaning, and hydro-mechanized cleaning. Large cities require significant budget and resources to perform the necessary cleaning activities at various locations around the city at regular intervals. For instance, the collection system in Edmonton, Canada, comprises over 5,500 km of sewer pipes, and as of 2014 there are over 1,400 prescheduled HPF locations. However, planning and scheduling these activities can be challenging because of the wide variation of actual on-site flushing duration, which depends on a number of factors such as length and diameter of the pipes, frequency of flushing, structural condition, age, and season. Moreover, travelling between these locations results in a large amount of unproductive time. Reviews of the literature and of current industry practice reveal that the existing models and algorithms do not specifically address these issues. This research, therefore, develops a framework for improving the productivity of these activities by optimizing operational maintenance schedule. The research consists of two primary modules: (*i*) developing a forecasting model to estimate the on-site duration of activities, and (*ii*) developing an optimization algorithm to maximize productivity. The models are developed and tested using historical data of HPF activity from the Drainage Operations group at the City of Edmonton. The forecasting model captures the majority of the variations in on-site flushing duration and provides useful insight into the factors affecting on-site productivity. For optimization, this research formulates the drainage operations scheduling problem (DOSP) as a special class of the stochastic and capacitated vehicle routing problem (VRP), where the objective is to maximize value-added on-site flushing time while minimizing travel. Alongside existing algorithms (such as integer programming, genetic algorithm), a heuristic algorithm is developed to meet the specific needs of this complex combinatorial problem. The proposed optimization algorithm is tested and compared with other algorithms by simulating a monthly HPF schedule. The results show that accurate estimation of on-site duration, coupled with schedule optimization can improve daily productivity by a considerable margin. The outcome of this research makes significant academic, economic, and environmental contributions by proposing a systematic approach to planning and scheduling operational maintenance for wastewater collection systems.

DEDICATION

To my mother, Ayesha Khatun, PhD

Inspiration for my fascinating journey through this research!

ACKNOWLEDGEMENTS

I would like to start with expressing my humble gratitude to Almighty Allah for making this come true!

My deepest appreciation goes to my co-supervisors, Dr. Mohamed Al-Hussein and Dr. Ahmed Bouferguene, for their constant guidance and inspiration throughout this research. I thoroughly enjoyed working with these two amazing individuals and I consider myself very fortunate to have known them closely. I can't thank you enough!

I would also like to thank my friends and colleagues at the University of Alberta for the wonderful times we spent together. There are so many, I can only mention a few here: Sang, Lei, Sadiq, Jon, Monjur, Dony, Mana, Sherry, Rita, Mona, Jacek, Saiful, Ahmad, Gurjeet, Christy, Claire… (the list continues)…thanks for all the good times!

This research would not be possible without continuous support from the drainage operations section of the City of Edmonton. Among others, special thanks go to Carl Lorentz, Kelly Davis, and Darren Kirstiuk for their kind cooperation.

Last but not the least, my wonderful family—Farhana, Farhan, and Aisha! I couldn't have made it without your love and support!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION (in alphabetic order)

AVL	Automatic vehicle locator
CCTV	Closed circuit televising inspection of sewer pipe
CHF	Hydro-mechanized cleaning of sewer pipe (also known as Chain failing)
CMMS	Computerized maintenance and management system
COV	Coefficient of variation
DOSP	Drainage operations scheduling problem
GA	Genetic algorithm
GIS	Geographic information system
GPS	Global positioning system
HPF	High-pressure flushing of sewer pipe
LPF	Low-pressure flushing of sewer pipe
MTV	Mainline televising (same as CCTV)
NVA	Non value added (time or activity)
OR	Operations research
O&M	Operations and Maintenance
PM	Preventive maintenance
PMX	Partially matched crossover
RTM	Regional travel model
TSP	Travelling salesman problem
TT	Total travel time
TW	Total waste time (refers to unused time at the end of work shift)
VA	Value added (time or activity)
VI	Visual Inspection of asset component (such as manhole, catchbasin)
VRP	Vehicle routing problem

1 INTRODUCTION

1.1 Background/Motivation

The sewer network is a critical component of urban infrastructure, consisting of sanitary, storm, and/or combined pipelines, lift stations, force mains, and other elements 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). The construction, condition assessment, repair, rehabilitation, operations and maintenance of this particular type of infrastructure are difficult and expensive due to the fact that substantial portions of these large and complex systems are buried in the ground (Vanier 2001; Tafuri et al. 2002). However, like other asset types, wastewater collection systems in many developed countries are now facing age-associated deterioration, which is a growing concern among professionals and the researchers. According to the 2013 Report Card published by the American Society of Civil Engineers (ASCE), America's wastewater and stormwater systems received an average grade of 'D'. Improving the conditions of that nation's wastewater system alone will require $298 billion over the next 20 years, of which the pipes account for three-quarters of total needs (ASCE 2013). In Canada, about 40% of wastewater plants, pump stations, and storage tanks are in "fair" to "very poor" condition, and 30.1% of pipes are in "fair" to "very poor" condition. The replacement cost of this infrastructure is $39 billion, or $3,136 per Canadian household (Canadian Infrastructure 2012). Although the water and wastewater sector in the UK currently stands in a slightly better condition with an average grade of 'B', regular preventive maintenance of the system is required to prevent rapid deterioration (ICE 2010).

Because the performance of a wastewater collection system directly affects the quality of urban life and the environment, the entire system needs to be maintained at the required level of service at all times. However, drainage infrastructure has a very high replacement value; hence, the rehabilitation and replacement of even a small portion of the system requires extensive capital investment. For instance, the replacement value of Edmonton's drainage system was estimated at $15.1 billion (as of 2012), which includes the large collection system comprising over 5,500 km of pipes (2,365 km of storm, 2,180 km of sanitary, and 946 km of combined sewer lines); 332,128 service connections; 68,496 manholes; and 74 pump stations (City of Edmonton 2013). Understandably, the entire system is not at optimum structural or operational condition at all times, and hence the City spends significant budget and resources to carry out the necessary operational maintenance work to ensure that the system is operationally functional. These operational activities mainly involve pre-scheduled periodic inspection and cleaning of various components, such as pipes, catch basins, manholes, and lift stations.

In recent years, many researchers have emphasized the importance of proactive and improved operation and maintenance activities (Pleau et al. 2005; Strauch & Wetzel 2006; Halfawy & Hengmeechai 2013). Although the effectiveness of these activities has increased significantly over the years with advancements in maintenance tools and techniques, aging and the associated deterioration of the system, coupled with urban growth, necessitates continuous improvement in O&M performance (Gaudreault & Lemire 2006). This research, therefore, aims to develop a framework to improve the productivity of operational activities for wastewater collection systems

.

1.2 Problem Statement

Today, most municipalities (interchangeably referred to as WWC operators) around the world carry out comprehensive operational activities throughout their jurisdictions on a regular basis in order to maintain the hydraulic functionality of the system. Established techniques and state-of-the-art equipment are used for various O&M activities, such as visual inspection (VI), low-pressure flushing (LPF), high-pressure flushing (HPF), catch-basin cleaning (CBC), televising (CCTV), and hydro-mechanized cleaning (also known as chain flail, or CHF). The selection of the necessary inspection or cleaning activities and their frequencies mainly depend on the operational conditions of the system components. It is evident that pipes in poor operational condition require more frequent flushing than those in good condition. For example, as of 2014, Edmonton has over 1,400 scheduled HPF locations across the city, selected based on their conditions and trouble history. Each of the locations is pre-scheduled for periodic HPF at a particular frequency, such as every 1 month, 3 months, 6 months, or 12 months. At the beginning of each month, a query to a central database generates a list of HPF work orders for the locations that are due that month. These work orders, grouped by neighbourhood, are then assigned to the individual crews.

The on-site durations of these flushing activities typically range from 10 minutes to several hours per location, and are stochastic in nature. Hence, the daily work flow for the operational crews follows a pattern where the crews start from a yard at the beginning of work shift, travel to the pre-scheduled locations and perform flushing one after another, and then return to the yard by the end of the shift. The number of locations completed in a typical 8-hour shift varies widely, depending on the flushing duration at each location and travel time between the locations. Because there is no reliable model currently available to estimate the flushing durations, the crews are often unable to predict how long the next location may actually take to flush. Hence, the crews can either return to the yard prior to the end of the shift time, which leads to un-used (or waste) time, or they can run overtime. Moreover, in large cities, travel can account for a significant amount of unproductive time. It is easily conceivable that reducing these non-value-added time can improve the daily productivity of operational activities.

This can be accomplished by optimizing the sequence of the scheduled flushing locations, which is conceptually similar to the well-known vehicle routing problem (VRP). However, unlike traditional route optimization problems where the objective is to minimize travel time or distance only, the "Drainage Operations Scheduling Problem" (DOSP) should also minimize the un-used time at the end of work shifts. Hence, there is a need to formally describe and formulate DOSP as a combinatorial optimization problem which maximizes daily productivity for a given schedule.

1.3 Research Objectives

The primary goal of this research is to improve daily productivity of drainage operations activities by achieving the following specific objectives:

- Reducing travel time (or distance)
- Reducing un-used time (or overtime) at the end of work shift

This is accomplished by developing two primary modules:

1. A forecasting model to estimate the on-site flushing duration at various locations

2. Suitable Optimization algorithms for DOSP

The efficacy of the optimization algorithms and the sensitivity of the on-site duration model are tested using simulation. In addition, the general framework proposed in this research includes the following supplementary modules to support the development of the primary modules.

- Review of operational preventive maintenance practice and productivity analysis
- Data collection and descriptive analysis
- Implementation strategy

1.4 Organization of the Thesis

This thesis is organized as follows: chapter 1 (Introduction) presents the background and motivation, and a brief description of the problem studied in this research. The specific objectives of this study are also listed here. Chapter 2 (Literature Review) presents a summary of the existing literature that has helped identify the research need, as well as to provide the theoretical basis of this research. The next chapter (Methodology) describes the development of the on-site duration estimation model and the schedule optimization algorithms for DOSP. In this chapter, the algorithms are developed for hypothetical graphs and their performances are compared through simulation. The proposed methodology is then applied to a case study in the Drainage Operations group at the City of Edmonton, which is presented in chapter 4 (Case Study). The detailed data collection, analysis, and results are also presented and discussed in this chapter. Finally, chapter 5 (Conclusion) summarizes the research findings and provides recommendations for future studies.

2 LITERATURE REVIEW

A thorough literature review has been conducted during this research in order to obtain a good understanding of the current practice, existing studies, recent advancements, and to identify the research needs. The literature review mainly focuses on the following areas, with particular emphasis on wastewater collection systems and related fields: infrastructure asset management; productivity analysis and improvement; and optimization algorithms for vehicle routing and scheduling.

2.1 Infrastructure Asset Management

Infrastructure asset management has been a widely studied field in recent decades. Recent literature related to water and wastewater systems addresses various important issues such as condition assessment and failure modelling, rehabilitation planning, design and construction, environmental protection, and pollution control (Hollenbeck 2004; Halfawy et al. 2006; Clark et al. 2007; Vallabhaneni 2010; Vallabhaneni 2011; Vallabhaneni 2012).

Hao et al. (2012) have conducted a review of condition assessment technologies and their relative advantages and challenges. Chughtai & Zayed (2011) developed classification protocol and condition assessment models for sewer pipelines that can be useful for standardizing sewer condition assessment. Studies both by Guo et al. (2009) and Halfawy & Hengmeechai (2014) developed visual pattern recognition techniques for defect reporting and condition assessment of drainage pipes, while Khazraeializadeh et al. (2014) compared different existing condition assessment protocols.

For operational condition of sewer network, Chughtai & Zayed (2008) developed a regression model based on pipe attributes such as age, length, diameter, slope, and roughness coefficient. The model was later simulated by Khan et al. (2009) in a study which explores the effect of these parameters on operational condition. In a recent study in Austria, Plihal et al. (2014) presented a case study where manhole zoom camera inspection is effectively used to optimize the schedule of operational activity (e.g., cleaning).

For failure analysis, Achim et al. (2007) and Moselhi & Fahmy (2008) developed Neural Network (NN) models for the prediction of water pipe failure, while studies both by Salman & Salem (2012), and Younis & Knight (2008) used ordinal regression models for the same purpose. Rostum (2000) developed statistical models for pipe failure in water networks, and Hoffman et al. (2010) performed statistical analysis to predict sewer blowouts during high-velocity jet cleaning operations. These studies identified important parameters (such as diameter, slope, and flow capacity) that affect wastewater collection system deterioration.

For rehabilitation planning, Selvakumar & Tafuri (2012) addressed the key challenges and issues in rehabilitation of aging water infrastructure, providing useful recommendations in terms of standardizing and improving management practice for wastewater collection systems. Becker et al. (2009) developed a decision support system for rehabilitation planning for sewers, where the reliability analysis-based system makes use of inspection results and considers important factors such as pipe material properties, loading system, and soil properties. Abraham (2003) used life cycle cost analysis for prioritizing the rehabilitation of wastewater infrastructure. Sample &

Kilpatrick (2006) presented an integrated "find and fix" approach for prioritizing the repair and maintenance schedule in a manner which makes effective use of condition assessment data integrated with GIS. Caldwell (2007) also addressed issues related to prioritization of rehabilitation and replacement of wastewater collection systems.

For overall operation and maintenance (O&M) aspects, Hassanein & Khalifa (2008) performed a comparative analysis for the performance of water and wastewater utilities in different countries. The study focused on operational indicators such as number of staff per thousand connections, labour cost, and operational cost. Ugarelli et al. (2010) discussed the concepts, methodologies, and limitations of current asset management practice with a particular focus on wastewater collection systems. Their paper addressed important aspects, such as asset management (AM) methodologies, condition assessment, level of service (LOS), life cycle cost analysis (LCCA), risk estimation, and different PM approaches (reactive, proactive, and predictive). Hannan (2000) addressed maintenance issues, with a particular focus on recent maintenance trends, program developments, and challenges, along with design, construction, and rehabilitation aspects of wastewater collection system. Ratliff et al. (2008) discussed improvements to O&M planning and management to reduce sanitary sewer overflow.

Many researchers have also emphasized improving maintenance operations through the development of an intelligent system. For instance, Hammond & Horton (1997) presented an automated system for lift station maintenance using bar-codes, while Loucks & Stahr (2007) implemented an integrated data management system for optimal use of the Little Rock Wastewater Utility in terms of its capacity. Several other researchers have also underscored the importance of improved and integrated planning and optimization of maintenance activities (Dekker 1996; Fenner 2000; Gamisch et al. 2010).

2.2 Sewer Asset Maintenance and Productivity Analysis

Oxford defines preventive maintenance (PM) as maintenance that is performed regularly to prevent or detect incipient failures. For wastewater collection system operation, PM can be described as the activities performed, on a predetermined schedule, on select components of the system to prevent operational failure (blockage), and to ensure that the operating components meet the system goals. Operational PM, which, according to the Water Environment Federation (WEF), refers to proactive flushing and cleaning of pipes, manholes, catch basins, and other system components, is performed to address general collection system function and recurring problems where rehabilitation or reconfiguration is not immediately feasible (WEF 2009). The effectiveness of PM has been well established through many studies; however, it is imperative for the purpose of this research to have a clear understanding of current O&M practice for wastewater collection systems.

There are a number of different operation and maintenance (O&M) approaches for sewer asset management. Plihal et al. (2014) categorized these approaches into the three following strategies:

(1) A reactive strategy, where a sewer line is cleaned only when necessary due to an operational failure such as a blockage caused by excessive grease or debris. While this approach reduces the cost of PM, it may result in poor system performance, blockage and sewer overflow, environmental pollution, and expensive repairs. Hence, this

approach is not considered an efficient and feasible strategy, and therefore is omitted from the scope of this research.

(2) A proactive strategy, where all the sewer lines in the jurisdiction are cleaned at pre-defined regular intervals. This approach requires adequate budget and resources to perform the PM activities on a regular basis, especially for large systems.

(3) A selective strategy, where a sewer line is cleaned based on its problem history and operational condition. This approach involves proactive inspection and condition assessment of the system, and analyzing data related to pipe properties (such as diameter, age, and material), spatial properties (such as industrial/residential zone, restaurant/carwash, type and amount of debris, grease, or tree roots), historical maintenance, and failure records in order to prioritize the PM schedule.

In addition to the above-mentioned approaches, recent research has focused on predictive maintenance, which involves failure analysis to forecast the likelihood of failure of an asset component. A periodic maintenance schedule is developed based on such probabilistic models.

The Water Environment Federation (WEF 2009) provided useful guidelines on establishing O&M strategies for wastewater collection systems, recommending the use of computerized maintenance and management system (CMMS) and geographic information system (GIS) to record and analyze five categories of data: structural, maintenance, inspection, hydraulic capacity, and condition assessment. Analysis of periodic (weekly, monthly, yearly) productivity reports is explicitly mentioned in the manual; however, no specific directions with respect to schedule optimization can be found.

The preliminary review of existing literature and current industry practice reveals that no well-documented objective rule or strategy exists that an operator can follow. Considering the environmental, economic, policy, and legal aspects at play, it is understood that this is a decision problem that involves strategic and subjective judgement. Nonetheless, as mentioned in the introduction of this thesis, age-associated deterioration and the expansion of the system lead to increased maintenance needs every year, which can result in either backlog in the PM schedule or an increase in resource requirements.

Various activities are performed as operational maintenance measures for a wastewater collection system. The United States Environmental Protection Agency summarizes inspection and cleaning activities (also referred to as methods or techniques) applicable for different system components such as pipes, manholes, and catch basins. The effectiveness, advantages, and limitations of the various methods are also discussed (EPA 1999). Some of the common inspection techniques include visual inspection, CCTV inspection, while some of the common cleaning activities are high-pressure flushing (HPF) (or jetting), and hydro-mechanized cleaning.

Among the studies focused on the performance measurement and productivity improvement of O&M activities, Bowen et al. (2003) have developed productivity standards for various maintenance activities for collection systems, where daily accomplishments and reported man-hours are used to determine the daily productivity (expressed as metres/man-hour for flushing activities). Mohamed et al. (2002) and Agbulos et al. (2006) have applied simulation and lean concepts to break down operational activities into smaller value added and non-value added tasks in order to improve productivity by reducing time waste. Navab-Kashani et al. (2015) have shown the potential of improving the productivity of CCTV inspection through time study and route optimization.

2.3 Route and Schedule Optimization

As for optimization, several studies have applied state-of-the-art optimization tools and techniques in planning and design of wastewater collection systems. For example, Boomgaard et al. (2004) applied Genetic Algorithm (GA) to optimize the cost and capacity of a wastewater collection system. Gupta et al. (1983) and Botrous et al. (2000) used dynamic programming to optimize the design of a wastewater collection network, while Maier et al. (2003) used ant colony optimization for design of a water distribution system, and Yeh et al. (2011) applied tabu search algorithm for optimization of sewer networks.

The drainage network maintenance scheduling problem at hand is a variant of the well-known Travelling Salesman Problem (TSP) or Vehicle Routing Problem (VRP), widely studied problems in the field of operations research (OR). Numerous articles are available in the literature which tackle various types of optimization problems, using different tools and/or developing various algorithms according to the given problem type. Among earlier studies, Laporte (1992) presented an overview of algorithms developed over the years, where the author discussed the methodologies of several exact and approximate route optimization algorithms (direct tree search method, dynamic programming, integer linear programming, heuristic approach, tabu search algorithm, etc.) and their suitability for solving various problems.

The increasing complexity of route networks and the quest for optimal solutions (especially under specific constraints) have led researchers to improve previous algorithms or to propose new approaches. For instance, Desrochers et al. (1992) presented VRP with time window, which is particularly effective for school bus service. Vijay et al. (2008) presented a study on GIS-based location analysis of collection bins for municipal solid waste management, where the results suggested improvement in the efficiency of collection and transport of solid waste towards bins and disposal sites. Apaydin & Gonullu (2007) developed a shortest path model (integrated with GIS elements) for solid waste collection which reportedly reduces time, travel distance and cost by a significant margin. In another study, they also presented the benefits of route optimization from the sustainability point of view; i.e., minimizing harmful gas emissions by optimizing travel distances for solid waste collection (Apaydin & Gonullu 2008).

Thus, wide applications of various route optimization techniques can be found in the existing literature. However, the efficacy of a particular optimization technique depends largely on the characteristics of the problem at hand. As opposed to some other VRPs, such as mail delivery service, school bus service, or solid waste collection, drainage operations has a unique characteristic where the operational crews perform maintenance activities at pre-scheduled locations one by one and then return to the yard by the end of the work-shift. Hence, optimizing the sequence of the locations to be flushed (or destinations) depends on the travel time as well as the activity duration at each location, which is stochastic in nature. Therefore, this problem becomes similar to the stochastic VRP described by Bertsimas (1992). Although several techniques of stochastic VRP are described in the literature (Huang & Louks 2000; Kleywegt et al. 2001; Bertsimas et al. 2011; Lei et al. 2011), their suitability for drainage operations scheduling and routing has not been studied.

2.4 Summary of Literature Review

In summary, it can be concluded that although the planning, design, failure, cost, operation and maintenance aspects of wastewater collection systems have been covered in numerous studies, the improvement potential in O&M productivity through optimized scheduling has not been studied in detail. This is a specific optimization problem encountered in every municipality and the existing optimization algorithms do not serve its specific needs. However, with the advancements in real-time data collection technology and integrated management system, there exists an excellent opportunity to develop a systematic approach to improving the scheduling of operational activities for wastewater collection systems.

3 METHODOLOGY

3.1 Introduction

This chapter describes the methodology of this research, which is built upon the following hypothesis: *"Schedule optimization based on accurate estimation of on-site duration can improve the productivity of operational maintenance activities for wastewater collection systems"*. The primary components of this research are the development of an on-site duration estimation model, and the formulation and development of a schedule optimization model. These components are complemented by data collection, analysis, and simulation modules. Fig. 3-1 illustrates the methodological flowchart and the components (or modules) of this study, showing that the analysis and modelling modules utilize inputs from existing databases, perform the necessary analyses, and then provide outputs to achieve outcomes aligned with the research objectives.

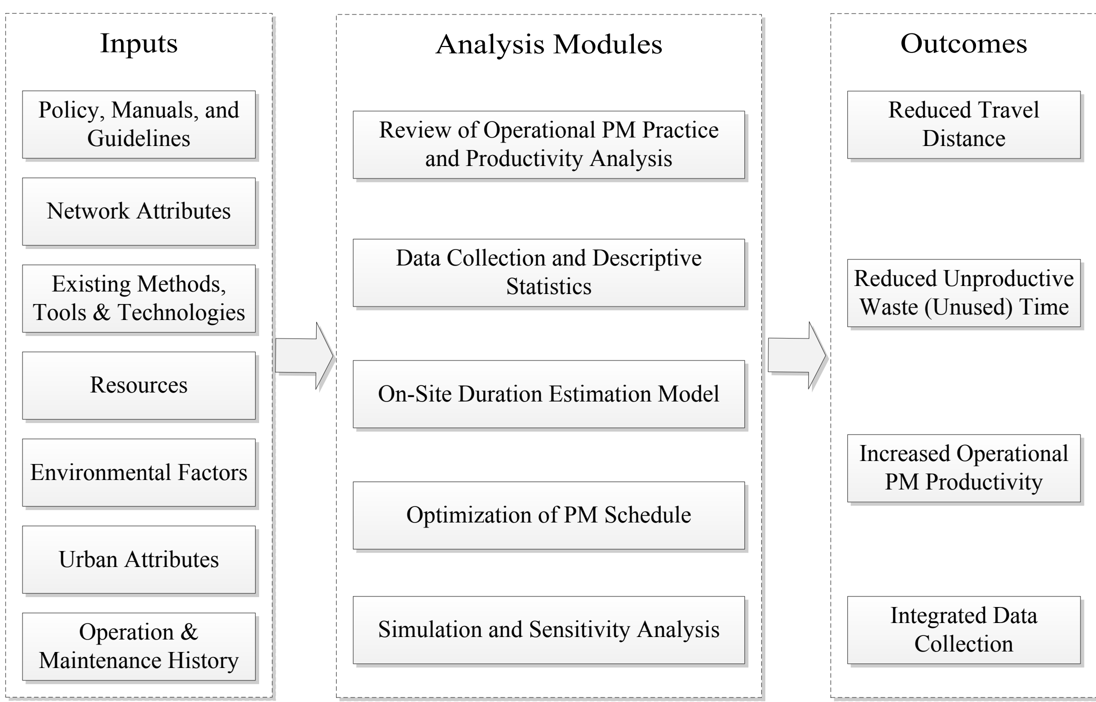


Fig. 3-1. Overall methodology of the research

The first step in the methodology is to study the current practice and guidelines in order to gain an understanding of operational preventive maintenance (PM) strategies, and to analyze the current performance trends of PM crews in order to identify the factors that affect productivity. Then, the data collection and preliminary analysis module examines the existing databases maintained by the operators, and establishes the data requirements for the estimation and optimization models. The next component, the on-site productivity estimation model, develops statistical models to predict on-site flushing duration, which is stochastic in nature. The output from this model becomes a vital input in the schedule optimization, which develops algorithms to minimize travel time and waste time. Because of the stochastic nature of the on-site duration, the algorithms are then tested for their robustness by running simulation models. Each of these modules is discussed separately in the following sub-sections.

It is to be noted that some of the material in this chapter has been presented at CSCE conferences (Zaman et al. 2012; Zaman et al. 2013), and published in the *Journal of Infrastructure Systems* (Zaman et al. 2014) and *Urban Water Journal* (Zaman et al. 2015). However, more elaborate discussion on the methodology of this research is provided here.

3.2 Review of Operational PM Practice and Productivity Analysis

The PM strategy adopted by an operator depends on the size and age of the network, availability of data, capital and operational budget, and resources. However, regardless of the method used for selecting the pipes and their frequencies for operational maintenance, the operators implement flushing/cleaning activities for a pre-scheduled list of pipes within a given period (e.g., month), and efficient productivity of these activities is critical in achieving the necessary performance indices.

In general, productivity is expressed as a ratio between input and output (Park et al. 2005). However, Bowen et al. (2003) defined productivity of drainage operations activities as accomplished per unit time: for example, *locations/day* for visual inspection, or *metres/man-hour* for flushing activities. Based on that definition, the productivity of flushing activity can be expressed as Eq. (3.1):

$$P_d = \frac{\sum_{i=1}^{n} L_i}{T_d \times C} \tag{3.1}$$

where

P_d = daily productivity;

n = number of locations flushed in a day;

L_i = total length of pipes (metres) in location i;

T_d = man-hours consumed to flush n locations; and,

C = size of crew (typically one operator and one assistant).

T_d in Eq. (3.1) can be broken down into four parts: (*i*) start-up time at the beginning of work shift; (*ii*) transportation time from the yard to the first location and, subsequently, between locations; (*iii*) on-site flushing and cleaning at locations; and (*iv*) clean-up time at the end of the work shift. The sequence of these activities during a typical work shift follows the pattern expressed in Eq. (3.2):

$$T_d = t_s + t'_{y\to 1} + t_1^f + t'_{1\to 2} + t_2^f + t'_{2\to 3} + t_3^f + \cdots + t_n^f + t'_{n\to y} + t_e \tag{3.2}$$

where

t_s = beginning-of-shift start-up time;

$t'_{y\to 1}$ = transportation time from yard to location 1;

t_1^f = on-site flushing time at location 1;

$t'_{1\to 2}$ = transportation time from location 1 to location 2;

t_n^f = on-site flushing time at location n;

$t'_{n\to y}$ = transportation time from location n to yard; and,

t_e = end-of-shift clean-up time.

Now, Eq. (3.1) and Eq. (3.2) can be combined as:

$$P_d = \frac{\sum_{i=1}^{n} L_i}{\left[t_s + t'_{y\to 1} + t_1^f + \sum_{i=1}^{n-1} \left(t'_{i\to i+1} + t_{i+1}^f \right) + t'_{n\to y} + t_e \right] \times C} \tag{3.3}$$

where t_i^f is the on-site flushing duration at location i. In this equation, t_s and t_e can be assumed to be constant, while the transportation times ($t'_{i\to i+1}$) can be derived from the travel distance between the consecutive locations. And, since the total length of pipes for a specific location is constant, the only unknown variable on the right-hand side of Eq. (3.3) is the on-site flushing duration (t^f), which varies across locations depending on various factors. For productivity improvement, it is imperative to identify the factors causing such variations, which can be accomplished by collecting historical data for productivity and related factors.

3.3 Data Collection and Descriptive Statistics

Today, many municipalities use computerized maintenance and management systems (CMMS) or comprehensive business solution packages to maintain a variety of information related to system attributes, condition of assets, maintenance schedule, work order history, human resources, and equipment. This wealth of data is useful for many purposes such as budgeting, accounting, scheduling, and performance measures; however, not all of the data or information is used for productivity modelling and schedule optimization. For instance, the daily reported man-hour data may be useful for accounting purposes, but not for performance measurement, since the total man-hours includes transportation time, which varies across flushing locations. The necessary dataset can be created either by collecting the necessary information from the field, or by combining existing databases, or a combination of both. Either way, it is important to consult industry experts and existing literature to comprehend and plan the collection process. For example, collecting input from field operators may be valuable in ascertaining how oil and grease in the flow affects on-site productivity, which types of tree roots are more likely to penetrate the pipes, or how climate affects productivity. The level of impact of these factors varies across geographical locations and seasons, which makes it imperative to incorporate the appropriate factors in order to build efficient models.

Because database management systems and data availability vary across municipalities, this chapter provides only a general overview of the data requirements for productivity analysis, on-site flushing duration estimation, optimization, and simulation modelling. However, the next chapter presents a detailed summary of data collection steps followed during the case study implementation of the methodology.

Fig. 3-2 presents the four types of databases and the variables drawn from each data source necessary for this study. The “Drainage Pipes Properties” database provides the physical properties of the pipe segments, such as diameter, length, slope, material, year of construction, and location coordinates. The “Work Order” database is used to collect the scheduled maintenance information, such as pipe segments, flushing frequency, location, and scheduled flushing date. The “Field Environment Data” includes useful feedback from field crews regarding any blockage or presence of fats, oil, grease, tree roots, or debris in the pipes during flushing. Correlating this information with geographical locations of specific types of trees or grease sources (e.g., restaurant) will allow the collection and documentation of the effects these items have on on-site flushing duration. In addition, climate- or weather-related data such as historical rainfall or spring runoff may be useful in terms of exploring the effect of weather on storm and combined sewers.

The actual on-site flushing duration data can be collected either by field time study or from Automatic Vehicle Locators (AVL) sensors connected with the flushing trucks. Travel-time data may be available from Google maps, geographic information system (GIS) software, or a regional travel model. All of these databases can be merged by using primary and secondary keys such as pipe ID and vehicle ID.

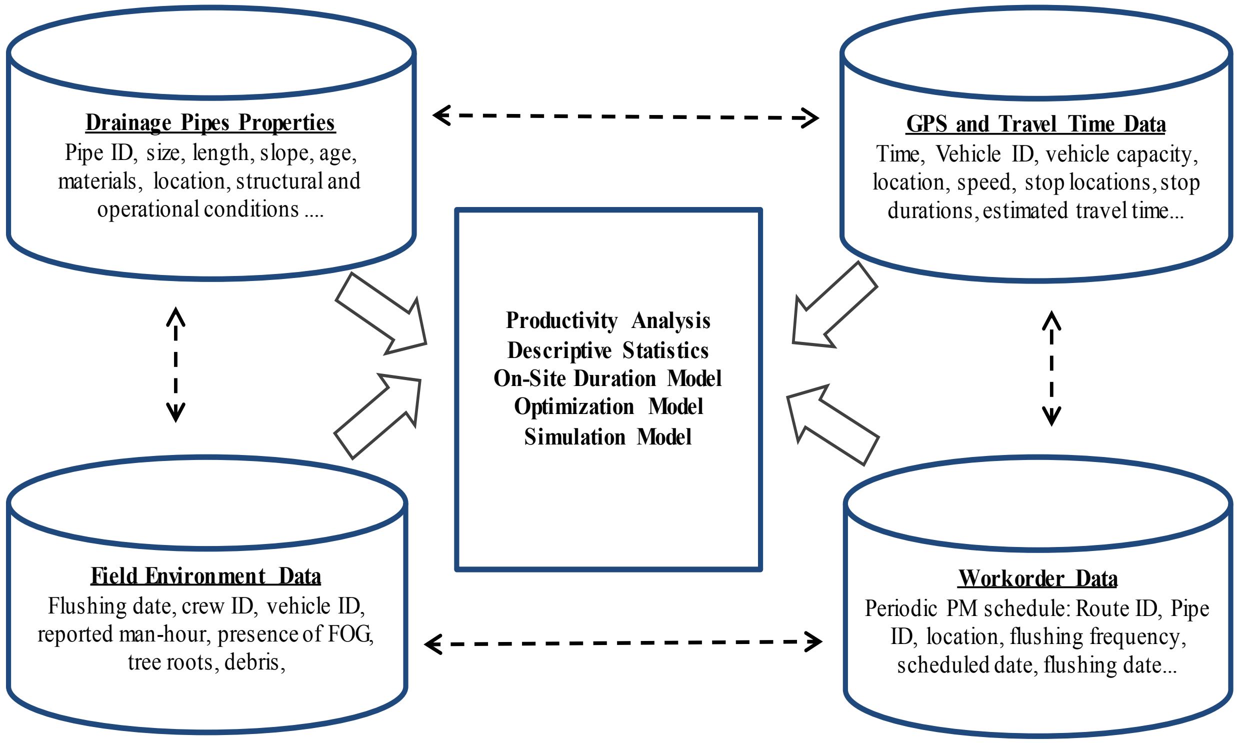


Fig. 3-2. Data collection schematic

Chughtai & Zayed (2008) developed structural and operational condition prediction models for sewer pipes, where various factors that affect the deteriorations are discussed (Fig. 3-3), and it is fairly reasonable to assume that the on-site flushing duration is directly proportional to a pipe's operational condition. Thus, the factors shown in the figure can be contributory factors for variation in daily productivity, and a reliable model for estimating t^f for a given location can thus be used to calculate the productivity of the flushing activity.

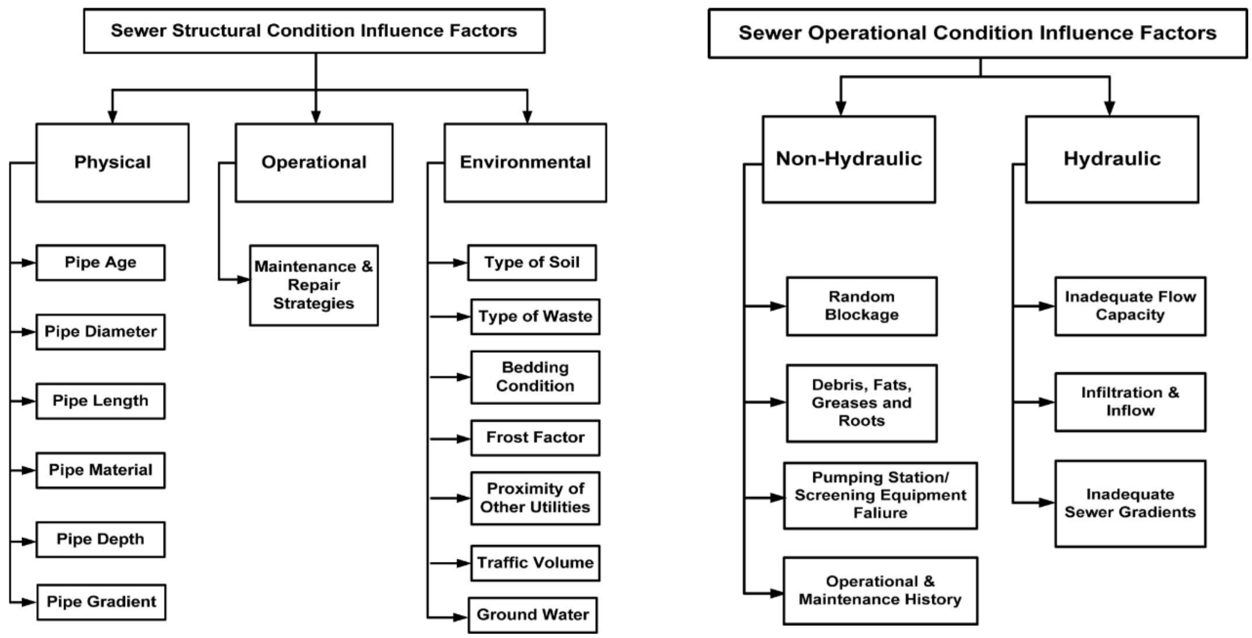


Fig. 3-3. Factors affecting the structural and operational conditions of sewer pipes (Chughtai & Zayed 2008)

The relationship between on-site flushing duration and these factors (predictor variables) can be obtained from descriptive statistics of collected data. Previous studies have shown that the productivity (or on-site flushing duration) has both spatial and temporal variations. To investigate that, field data of actual on-site duration is to be collected by means of direct field observation or automated process (such as AVL connected with flushing motor). Then the collected on-site duration data can be plotted against different factors to explore the relationship. For example, Fig. **3-4** shows linear relationship between flushing duration and length of pipes flushed.

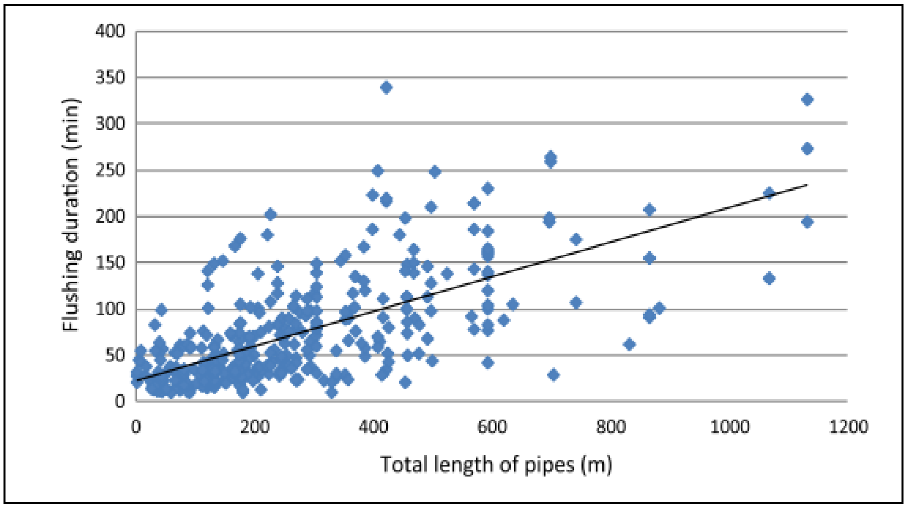


Fig. 3-4. Length of pipes vs. duration of high-pressure flushing

3.4 On-Site Duration Estimation Model

The following procedure lays out the formulation used to estimate the on-site flushing duration for a given location, *i*. To begin with, it is important to obtain a complete understanding of the factors that influence the on-site flushing duration. Once the factors influencing on-site duration are identified, the relationships between predictor variables (such as length of pipe, diameter, slope, location, and age) and the response variable (on-site duration in this case) are obtained from descriptive statistics. These relationships aid in the selection of the appropriate model to capture the variation in the response variable. For example, Zaman et al. (2013) have shown that the on-site duration has a linear relationship with the predictor variables, and thus that a multiple regression can capture the majority of the variation.

3.4.1 Multiple Regression Analysis

The general form of a multiple linear regression model can be expressed satisfying Eq. (3.4) (Kutner et al. 2005):

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad \text{for } X_{i0} \equiv 1 \qquad (3.4)$$

where

Y_i = flushing duration for location i;

X_{ik} = predictor variable k for location i;

β_k = parameter for variable k;

$\mathcal{E}_i$ = independent $N(0,\sigma^2)$ error term for location i;

i = 1,2,3…….., n; where n is the total number of observations; and,

$p-1$ = number of predictor variables.

However, since $E(\varepsilon_i) = 0$, the response function of Eq. (3.4) for a particular location becomes:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_{p-1} X_{p-1} \qquad (3.5)$$

Therefore, models in the form of Eq. (3.5) are developed with the sequential addition of variables. The inclusion or exclusion of each factor is determined based on its sign, t-stat, and its impact on the model's overall goodness-of-fit value (adjusted R^2). Previous research (Agbulos et al. 2006; Navab-Kashani et al. 2015) show that the actual on-site duration can be affected by the number of stops, which refers to the number of times a crew stops to flush a given set of pipes. This variable has a significant effect on the on-site duration of flushing activities, and it differs from the number of pipes depending on many factors such as the location of pipes, their upstream facility types, and crew judgement. Because this variable is unknown to the planner at the time of scheduling, it should be modelled separately rather than being included in the model as a predictor.

3.4.2 Ordered Probit Analysis

Let, a crew is to flush a number of consecutive sewer pipes (referred to as "number of pipes", or N_p) at a particular location, where each of the pipes has an access manhole. Under ideal conditions, the number of times the crew should stop to flush all the given pipes (referred to as "number of stops", or N_s) should be equal to N_p. However, analysis shows that the difference between the number of stops and number of pipes varies within as well as across locations, and follows a normal distribution pattern. For example, the number of stops for a 3-pipe location may range from 1 to 5 (resulting in $N_s - N_p$ varying between −2 and +2), while the same for a 7-pipe location may range from 4 to 11 (resulting in $N_s - N_p$ varying between −3 and +4). If this difference between N_s and N_p is assumed to be the error (ε), then, owing to the ordered nature of N_s and normally distributed ε, an ordered probit model can be used to estimate the number of stops for a given location (McKelvey and Zavoina 1975; Baik et al. 2006). Therefore, the probability of the number of stops is first estimated using an ordered probit model, and is then used in the primary multiple regression model to estimate the on-site flushing duration.

The basic idea of the probit model is that there is a latent continuous metric underlying the observed ordinal responses, partitioned by a series of regions corresponding to the various ordinal categories (Baik et al. 2006). This can be mathematically expressed as follows:

$$(N_s)_i = j \iff \mu_{j-1} < y_i^* \leq \mu_j \tag{3.6}$$

where *j=1,2,.....m*; y_i^* is a latent continuous variable having linear relations with some predictors, and μ defines the range ($\mu_0 = -\infty$, $\mu_m = +\infty$) for y_i^* between which N_s takes the value of j. Therefore, y_i^* satisfies an equation similar to Eq. (4), and can be written as:

$$y_i^* = \beta X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \ \forall\, i = 1,2, \ldots .n \tag{3.7}$$

In turn, the probability of a given ordinal outcome (in this case, number of stops) is expressed as:

$$P[(N_s)_i = 1] = P[\mu_0 < y_i^* \leq \mu_1] = P[-\infty < y_i^* \leq \mu_1] = P[y_i^* \leq \mu_1] \tag{3.8}$$

Substituting from Eq. (7),

$$P[(N_s)_i = 1] = P[\beta X_i + \varepsilon_i \leq \mu_1]$$

$$= P[\varepsilon_i \leq \mu_1 - \beta X_i]$$

Therefore, $$P[(N_s)_i = 1] = \Phi(\mu_1 - \beta X_i) \tag{3.9}$$

Similarly, $$P[(N_s)_i = 2] = \Phi(\mu_2 - \beta X_i) - \Phi(\mu_1 - \beta X_i) \tag{3.10}$$

And, $$P[(N_s)_i = 3] = \Phi(\mu_3 - \beta X_i) - \Phi(\mu_2 - \beta X_i) \tag{3.11}$$

The model can be estimated using maximum likelihood estimation (MLE), satisfying the following equation:

$$\ln \mathcal{L}\,[\beta, \mu | X] = \sum_{i=1}^{n} \sum_{j=1}^{m} Z_{ij} \ln[\Phi_{ij} - \Phi_{ij-1}], \ Z_{ij} = \begin{cases} 1 \ for \ (N_s)_i = j \\ 0 \ otherwise \end{cases} \tag{3.12}$$

The selection of variables for the "number of stops" estimation model requires careful consideration. Stepwise regression is a systematic procedure to select variables based on their statistical significance; however, because the output from the probit model is used as a predictor variable in the multiple regression, the factors used in the former model may also be statistically significant for the latter. For example, "work interruption" affects both "number of stops" and "flushing duration", either directly or indirectly. However, if this factor is included in the probit model, it should not be used in the multiple regression, which would imply that the factor indirectly affects the duration. Therefore, the selection of variables for the two models is to be done not only based on their statistical significance but also on the work procedure, experience, and judgement. Hence, trials are performed to select the optimum variables for the probit model in such a manner that it provides reliable output without compromising the goodness-of-fit of the multiple regression model.

3.5 Schedule Optimization

This section presents the formal description and formulation of the Drainage Operations Scheduling Problem (DOSP) as a schedule or route optimization problem, and then compares existing exact and meta-heuristic algorithms with a proposed heuristic algorithm.

3.5.1 Problem Statement

From a theoretical perspective, DOSP can be viewed as a fully connected graph, ***G***, with *n* vertices that represent the scheduled O&M locations across the city, where each closed tour represents the locations visited by a crew in a single work shift. If the location of the yard is denoted by *0*, there exist *n!* possible tours (sequences of locations to visit) among which the optimal tour (representing the least travel time) is to be determined. In addition, DOSP requires partitioning of the feasible tours for each day. Hence, if *n* locations are scheduled to be flushed in *j* days, there exist $\binom{n-1}{j-1}$ splitting possibilities for each combination of TSP. For example, if 5 locations are to be flushed in (a maximum of) 2 days, with location *0* denoting the yard, there can be {*5! * (4C1)*} = 480 different combinations of routing sequences. Fig. 3-5 illustrates examples of three different combinations for flushing 5 locations in 2 days, each of which may have different travel time and unused time.

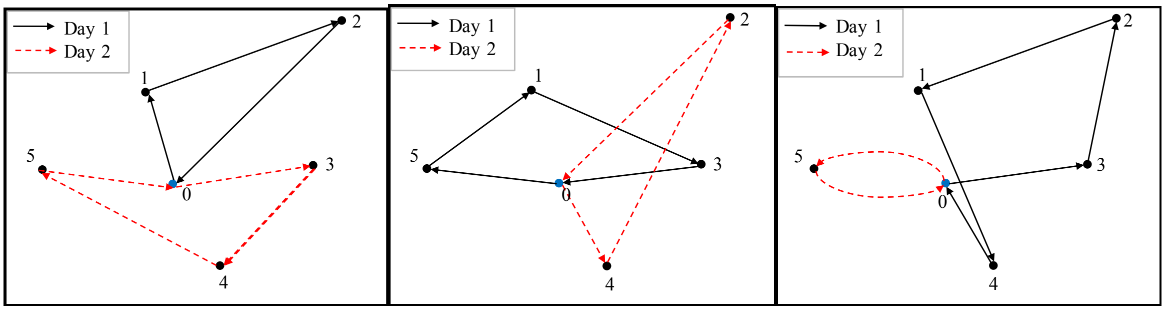


Fig. 3-5. Examples of different combinations of 5 locations flushed in 2 days

This complexity can be reduced by some margin based on the following consideration: the following two partitions result in the same amount of travel and unused time, provided that different days do not affect the travel and flushing time, and that unused time is considered at the end of both days.

- *Partition 1 Day 1: {0-1-2-0}, Day 2: {0-3-4-5-0}*
- *Partition 2 Day 1: {0-3-4-5-0}, Day 2: {0-1-2-0}*

However, depending on the travel distance and on-site flushing duration, these locations could also be flushed in 1 day. In this context, the number of days could be either 1 or 2, which results in calculating the total feasible combinations using the following expression:

$$\sum_{j=l}^{k} n! \frac{\binom{n-1}{j-1}}{(j)!} \tag{3.13}$$

where

n = number of locations;

k = maximum expected number of days to flush all locations;

l = minimum expected number of days to flush all locations; and,

$l \leq k \leq n$, assuming that no single location takes longer than one day.

From Eq. (3.1), it is easily understood that the total feasible combinations grows rapidly (faster than exponentially) as the number of locations increases, making the optimization problem difficult to solve within a reasonable amount of time. For instance, if 150 locations (n) can be flushed in 20 to 30 crew-days (k), the total possible combinations calculated using Eq. (3.1) is found to be 1.62E+268, which is a large problem to solve in reasonable time.

Many exact and approximate algorithms (such as integer programming, evolutionary algorithms, and heuristic algorithms) have been proposed and implemented over the years for similar problems (Laporte 1992). However, some practical situations may not have the sufficient time available to run lengthy algorithms. For DOSP, municipalities can run the schedule optimization model daily, weekly, or monthly. In any case, the model should be fast enough to quickly generate an optimal (or near-optimal) sequence of locations to be visited by the crews. The advantage of running a monthly schedule optimization model (say, for n = 150) at the beginning of the month is that it can produce a better result because the splitting portion offers more options from which to choose. However, the actual work may deviate from the planned sequence as the month progresses due to the stochastic nature of flushing duration. On the other hand, running the model every day for fewer locations can generate a sequence fairly quickly; this, however, compromises the quality of the results. Either way, it may be necessary to re-optimize the sequence intermittently in order to avoid propagation of schedule deviation. Therefore, it is more realistic from a practical perspective to use an optimization algorithm that can quickly generate a solution that is reasonably close to optimality.

3.5.2 Problem Formulation

The primary objective of this study, to improve the productivity of drainage network maintenance activities, can be realized by facilitating efficient utilization of daily effective work time (i.e., minimizing travel and end-of-shift unused times). The integer programming formulation of the problem can be written as follows:

Let,

$N = \{1,2,3,4,.....,n\}$ is the set of locations scheduled to be flushed in a given month;

$S_k \subseteq N$, containing the locations flushed on the k^{th} day;

K = number of days taken to flush all the scheduled locations;

T = shift length (typically 8 hours);

$T^w{}_k$ = end-of-shift unused time for k^{th} day;

T_{ij} = travel time between location i to j, $i \neq j$;

γ = relative importance of saving end-of-shift unused time vs. travel time, $0 \leq \gamma \leq 1$; and

$E\left(T_i^f\right)$= expected on-site flushing duration at location i.

Thus, the objective function can be expressed as follows:

$$\min\left[\gamma\sum_{k=1}^{K}T_k^w+(1-\gamma)\left\{\sum_{k=1}^{K}\left(T_{0i}x_{0ik}+T_{j0}x_{j0k}+\sum_{i=1}^{n}\sum_{j=1}^{n}T_{ij}\,x_{ijk}\right)\right\}\right];\ i\neq j \tag{3.14}$$

Subject to,

$$x_{ijk} = \begin{cases} 1 & \text{if the crew goes from location } i \text{ to } j \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

$$x_{oik} = \begin{cases} 1 & \text{if location } i \text{ is the first location on } k^{th} \text{ day} \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

$$x_{j0k} = \begin{cases} 1 & \text{if location } j \text{ is the last location on } k^{th} \text{ day} \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

$$0 \le x_{ijk} \le 1\,; \quad i,j \in S_k \text{ and } \notin S_k^{'} \tag{3.18}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{n} x_{ijk} = 1; \quad i \ne j; \quad \forall j \tag{3.19}$$

$$\sum_{k=1}^{K}\sum_{j=1}^{n} x_{ijk} = 1; \quad i \ne j; \quad \forall i \tag{3.20}$$

$$\sum_{i=1}^{n} x_{0ik} = \sum_{j=1}^{n} x_{j0k} = 1; \quad \forall k \tag{3.21}$$

And, the disjoint sub-tour elimination constraint within each day,

$$u_i - u_j + nx_{ij} \le n - 1; \quad \forall k \tag{3.22}$$

where u_i and u_j are alternate variables; $u_i = t$ if location i is the t^{th} location visited on the tour (Papadimitriou and Steiglitz 1982). Understandably, the number of disjoint sub-tour elimination grows rapidly as the number of locations increases. However, an important practical assumption that may reduce this number significantly is that an 8-hour shift can typically accommodate 5-6 high-pressure flushing (HPF) locations. Under this consideration, it may be feasible (in most cases) to eliminate disjoint sub-tours totalling only up to 3 or 4 locations.

In addition, a flusher truck has certain capacities for its debris and water tanks. The water tank empties as water is used to flush the pipes, and the debris tank fills as debris is pumped out of the pipes. Both of these rates vary across locations depending on many factors such as pipe length, diameter, and the amount and type of debris; however, it is fairly reasonable to assume that these rates are proportional to the flushing duration. In any case, the trucks have to travel to the nearest dumping yard if the debris tank reaches its capacity, or travel to the nearest water filling station if the water tank requires refilling, each case resulting in additional travel. These capacity constraints can be added to the formulation by the following equation:

$$\sum_{i=1}^{n} \omega_i x_i \leq \hat{\mathrm{W}} \; and \; \sum_{i=1}^{n} \delta_i x_i \leq \dot{\mathrm{D}} \qquad (3.23)$$

where

ω_i = amount of water used at location i;

δ_i = amount of debris collected from location i;

$\hat{W}$ = capacity of water tank; and

$\dot{D}$ = capacity of debris tank.

It is to be noted that, although this is a non-linear problem due to the daily split of tours and the capacity constraints, the above formulation allows the problem to be solved using a simplex algorithm, resulting in a guaranteed optimal solution. Also, due to the stochastic nature of the on-site flushing duration, the formulation uses expected duration during the optimization process. The actual routing and work flow is then simulated against the optimized schedule, which is described in section 3.6 of this chapter.

3.5.3 Tour Splitting

The daily split of tours can be executed in several different ways, depending on the following two conditions: (*i*) overtime policy—whether the crews are allowed to work beyond the end of the shift,; and (*ii*) location splitting—whether the crews are allowed to leave an incomplete location at the end of the shift and return to complete the job the following day. Based on these conditions, the daily tour splitting is performed by one of three methods:

1) In the "no overtime, no location split" option, the crews will return to the yard if the expected travel plus the on-site duration for the next location is greater than the time available (i.e., time remaining before the end of the shift). The other condition to consider in this option is that, once a location is visited, the crew must complete the job before going to the next location or the yard. This condition restricts crews from leaving an incomplete location at the end of the shift and returning the next day to finish the job. Understandably, this will result in more end-of-shift waste time.
2) The second option allows overtime when necessary; however, no location splitting is allowed. Hence, if a location with a long expected duration is visited near the end of the

shift, the crew must finish the job before returning to the yard. This option will result in increased overtime.

3) The third option consists of a mix of options one and two, where both overtime and location splitting are allowed, depending on the situation. For this option, the crews utilize anticipated duration of the locations, as well as their experience and judgement to decide when to return to the yard. If a location has not been completed by the end of the shift, the crew may return to the yard and return to that location to complete it the following day. This option may reduce the amount of end-of-shift waste time or overtime; however, it will result in more travel due to visiting the same location twice. Moreover, the on-site duration model presented in the previous section reveals that "work interruption" results in a higher number of stops, which will eventually increase the on-site duration.

3.5.4 Integer Linear Programming (ILP)

Although this formulation guarantees optimal solution, it is not feasible to apply this method for a large number of locations due to runtime or space limitation. Therefore, the formulation is tested on a small dataset, comprising 12 locations (and the yard as 0). The locations and their expected on-site durations are presented in Table 3-1 and Fig. 3-6. As can be seen in the figure, the locations are spread across four quadrants to resemble different zones or neighbourhoods within an operator's jurisdiction.

Table 3-1. Randomly generated test locations (with coordinates and expected flushing durations)

Location ID	X	Y	Expected Duration (min)
0	0	0	0
1	4	5	27
2	18	9	81
3	13	17	115
4	17	−10	59
5	7	−19	94
6	3	−13	38
7	−5	18	68
8	−19	11	71
9	−3	1	32
10	−17	−5	86
11	−6	−16	84
12	−10	−18	23

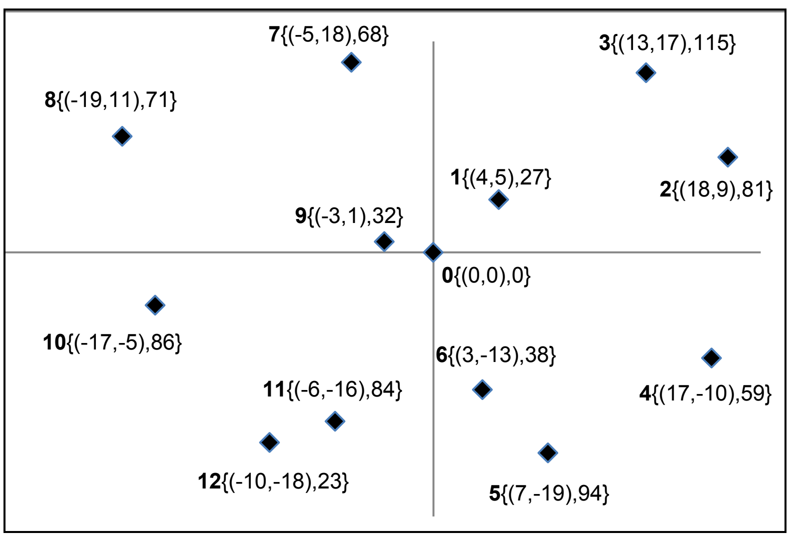


Fig. 3-6. Graph showing randomly generated locations {(X,Y), duration}

For routing, it is assumed that a single given crew has been assigned to flush all 12 locations, and the effective length of shift is 6 hours. Each day, the crew will start from the yard, flush as many locations as possible and return to the yard before the end of the shift. A common approach to accomplishing this is neighbourhood-based routing, where the crew starts with a specific neighbourhood, finishes all the locations within it, and then moves to the next neighbourhood. If overtime is not allowed, the crew must return to the yard when the remaining time within a shift is not sufficient to perform the cleaning activity at the next location. Fig. 3-7 shows the routing for this approach, where the different colours represent different days.

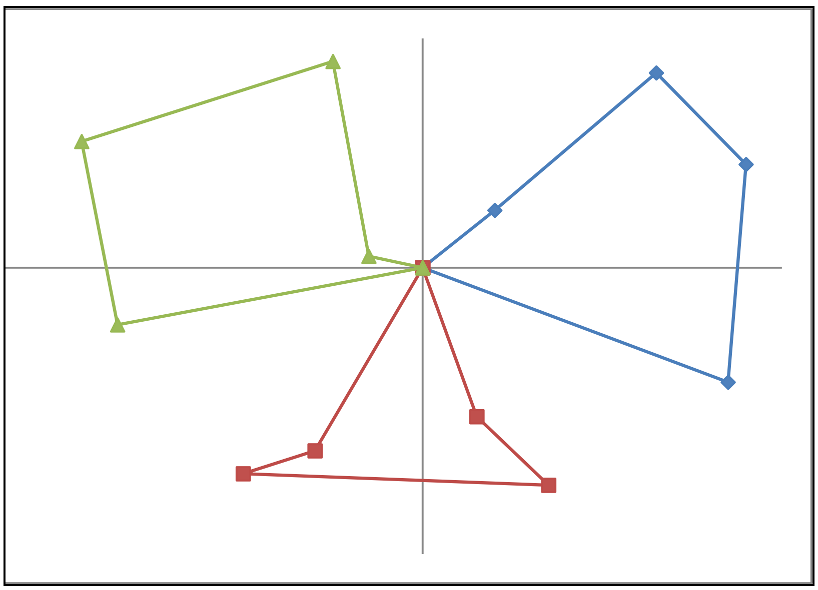


Fig. 3-7. Neighbourhood-based routing of scheduled locations

It can be seen that the crew starts from location 1 (Fig. 3-1), and follows a routing to the next location within the neighbourhood. When a neighbourhood is completed and there is still sufficient time remaining to start the next neighbourhood, the nearest location to the neighbourhood just completed is selected. However, because the on-site duration of the locations is not taken into account in this approach, it can result in either overtime or unused time. For instance, as presented in Table 3-2, the unused time at the end of day 2 is 62 minutes, because the next location within the neighbourhood (location 10) has a lengthy expected on-site duration, and the crew cannot finish before the end of the day. Therefore, this routing option results in a total of 199 minutes of travel, and 70 minutes of unused time; adding these together, the total non-value added time (NVA) is 269 minutes. As a general expression, this NVA time can be expressed as a percentage of total value-added time for the scheduled period, which equals the total flushing time of all 12 locations (778 minutes). Thus, the *NVA/VA* ratio is 34.58%. This ratio is used in further analysis of optimization results, as it can serve as a representative comparative measure for all of the randomly generated graphs.

It is to be noted that the last day's unused time is not added to the calculation of unused time, as O&M activities are a continuous process and more locations can be added to the schedule before the end of the period. Moreover, it is evident that this effect diminishes for longer schedule periods (e.g., 1 week or longer), where the optimization is performed for a larger number of locations (e.g., over 30 locations).

Table 3-2. Daily travel time and unused time for neighbourhood-based routing option

Day	Daily Flushing Time (minutes) (*DFT*)	Daily Travel Time (minutes) (*DTT*)	Daily Unused Time 360-*DFT*-*DTT*
1	282	70	8
2	239	59	62
3	257	70	-
TOTAL	778	199	70
Total VA	778	-	-
Total NVA	-	269	
NVA/VA%		269/778 = 34.58%	

Now, the same dataset is optimized by means of integer linear programming (ILP) formulation using the Premium Solver software platform, which employs simplex algorithm to solve linear problems. The results of the optimized routing are presented in Fig. 3-8 and Table 3-3. It can be seen that the overall non-value added time (222 minutes) is reduced significantly by the optimization, although the travel time has increased. This is because of the fact that no relative weighting between travel and unused times has been assigned to the optimization. Therefore, the algorithm reduces the overall NVA time.

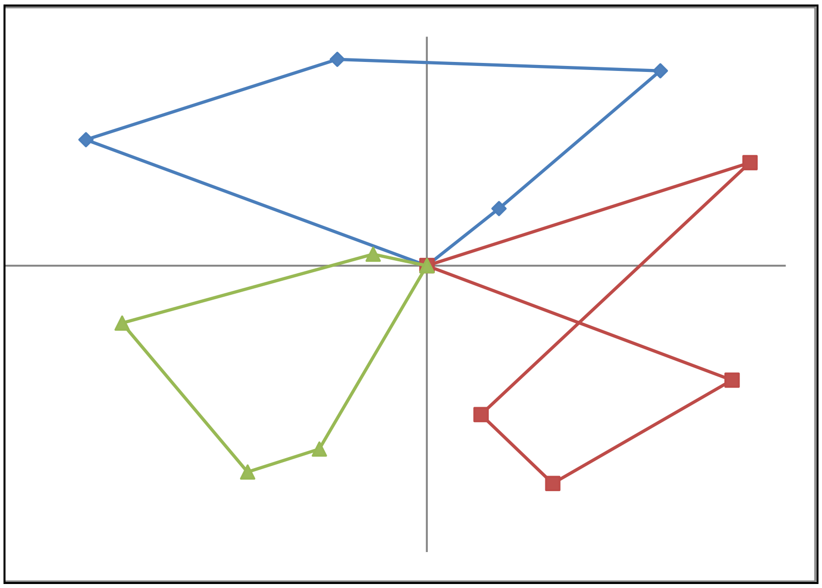


Fig. 3-8. Optimized routing by ILP formulation using simplex algorithm

Table 3-3. Daily travel time and unused time for optimized routing using ILP

Day	Daily Flushing Time (minutes) (*DFT*)	Daily Travel Time (minutes) (*DTT*)	Daily Unused Time 360-*DFT*-*DTT*
1	281	77	2
2	272	87	1
3	225	55	-
TOTAL	778	219	3
Total VA	778	-	-
Total NVA	-	222	
NVA/VA%		222/778 = 28.53%	

Assigning a higher importance on travel time saving (i.e., $\gamma > 0.5$ in Eq. [3.14]) can improve the solution in terms of travel. Thus, the ILP provides a different solution, where the overall NVA is the same as before; however, travel is reduced relative to unused time. The solution results are presented in Fig. 3-9 and Table 3-4.

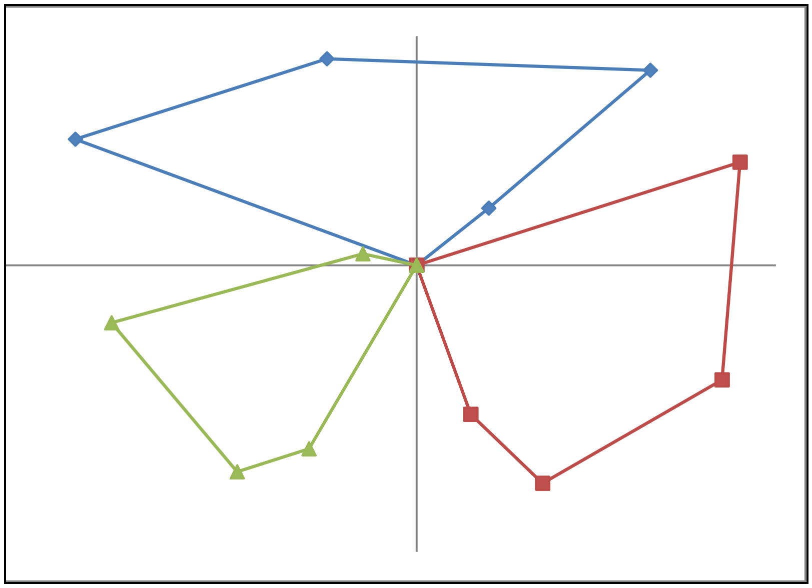


Fig. 3-9. Optimized routing by ILP formulation (with higher relative weighting on travel)

Table 3-4. Daily travel time and unused time for optimized routing using ILP (with higher relative weighting on travel)

Day	Daily Flushing Time (minutes) (*DFT*)	Daily Travel Time (minutes) (*DTT*)	Daily Unused Time 360-*DFT*-*DTT*
1	281	77	2
2	272	73	15
3	225	55	-
TOTAL	778	205	17
Total VA	778	-	-
Total NVA	-	222	
NVA/VA%	222/778 = 28.53%		

This provides the proof of concept of the ILP formulation of DOSP. However, as discussed above, this formulation can only be realistically applied to small datasets. This research, therefore, explores the suitability of a meta-heuristic algorithm to solve the problem.

3.5.5 Genetic Algorithm

Genetic Algorithm (GA) is a widely used meta-heuristic search algorithm inspired by the process of evolution (Holland 1992). Although GA does not guarantee optimality, it can be easily applied to a variety of large, practical problems where quick runtime is important and a near-optimal solution is acceptable. A GA model has thus been developed for the DOSP, as per the flowchart presented in Fig. 3-10. The process involves generating several "parents", each having a randomly generated sequence of all the locations. Thus, starting from day 1, the sequence is followed until the day's remaining time is less than the time necessary to travel to and flush the next location. The end-of-shift unused time is calculated as the unused time at the end of the shift after the crew has returned to the yard, and the daily travel time (or distance) is recorded. The next day begins with travelling to and flushing the next location in the sequence, and so on, until all the scheduled locations are completed. The total unused time is calculated by adding the daily wastes for *k-1* days, since the time remaining after completing all the given locations is not considered unused time. Thus, each parent has total travel and total unused times for the scheduled period, and the fitness of the parent is calculated by adding these together. However, in this case, a lower value of the fitness function yields more desirable results. After repeating the previous steps for all the parents, they are sorted according to their fitness values. Then, elitist selection is used to preserve the best solution, and parents with high fitness are selected for crossover.

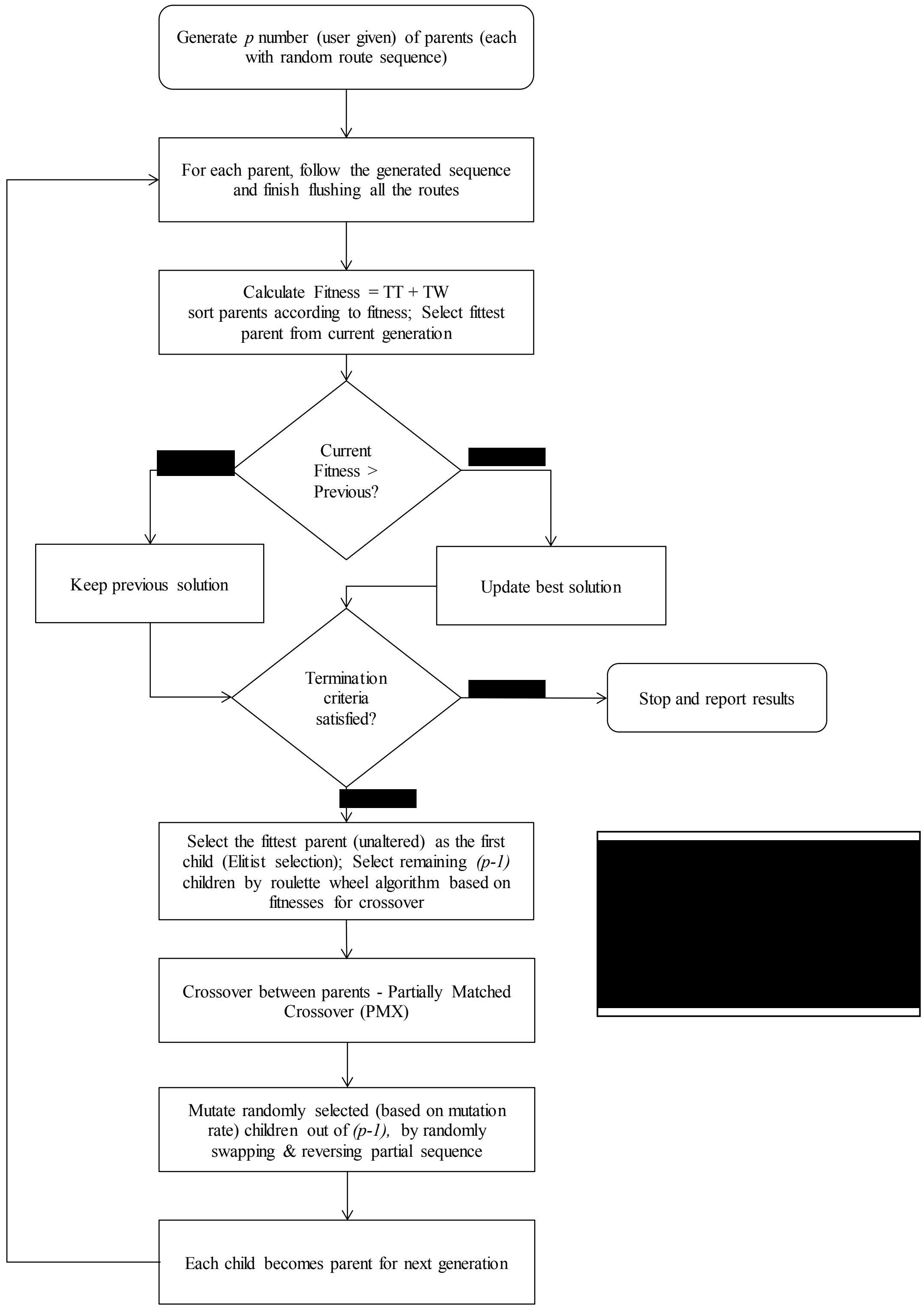


Fig. 3-10. Genetic algorithm flowchart for DOSP

For combinatorial problems such as DOSP, a partially-matched crossover (PMX) technique is used to avoid the repetition of locations in the schedule, and thus a new generation of children is obtained (Fig. 3-11). Mutation operation is performed for a few randomly selected children to prevent the algorithm from becoming stuck in the local optimal. Following this step, the fitness of all children are calculated again, the best one is kept unaltered, and the remaining children (ordered according to their fitness) are carried over as parents to the next generation (or trial). Thus, progress toward an improved solution is expected over the course of the trials (Fig. 3-12).

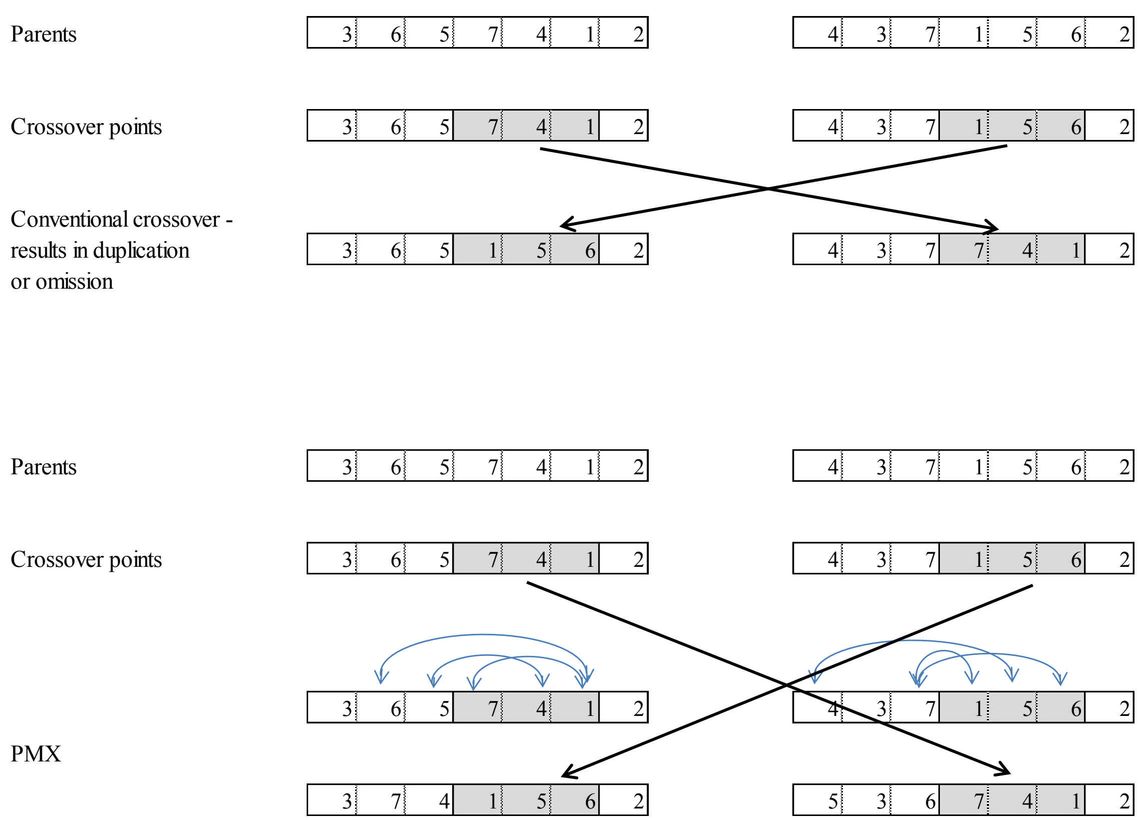


Fig. 3-11. Partially-matched crossover (PMX) process

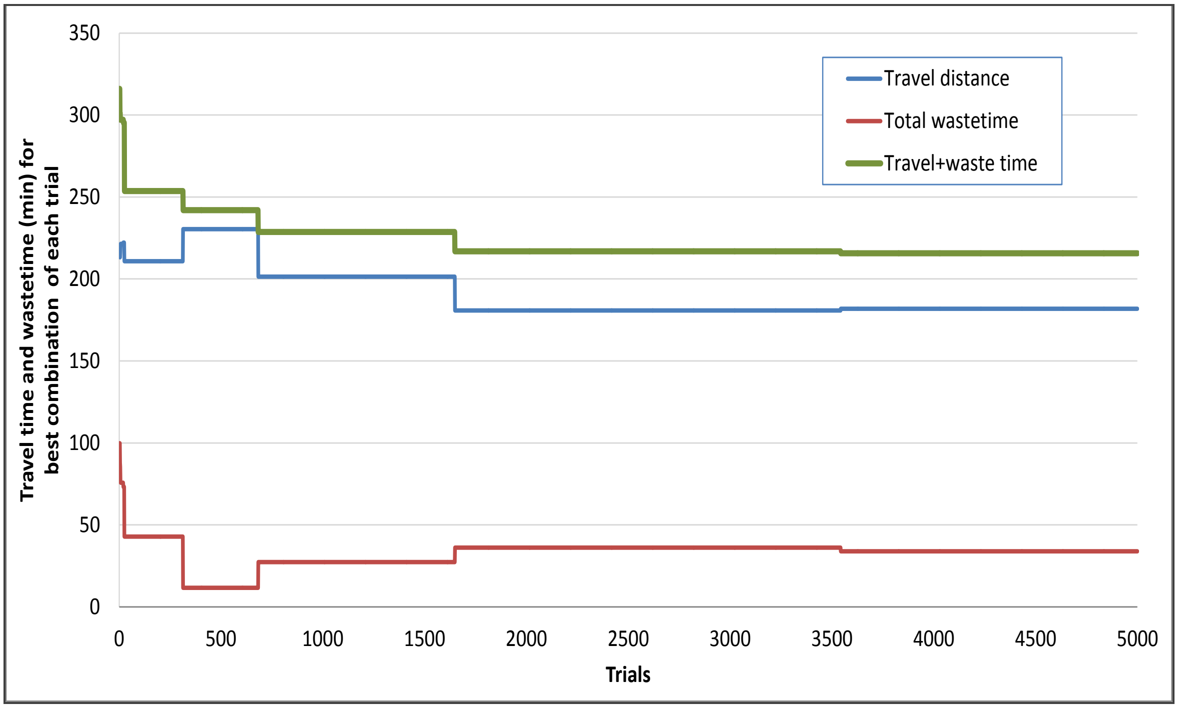


Fig. 3-12. GA convergence

The algorithm is coded using MS Visual Studio, and the same graph ***G*** (Fig. 3-6) is used for optimization. Fig. 3-13 and Table 3-5 present the results of the optimization using GA. Equal weighting is applied to travel and unused times, and the results show that GA provides a greater reduction in NVA than the neighbourhood-based approach. As expected, it does not provide the optimal solution; however, at 30.38%, the NVA/VA ratio is close to the optimal value obtained from ILP. The performance of GA largely depends on the number of trials (generations), population size, cross-over rate, and mutation rate. Understandably, conducting more trials can lead to a better solution; however, more trials also necessitates longer runtime. Nonetheless, the main advantage of this algorithm is that it can be applied for a large number of locations, and a near-optimal solution can be obtained within a reasonable time.

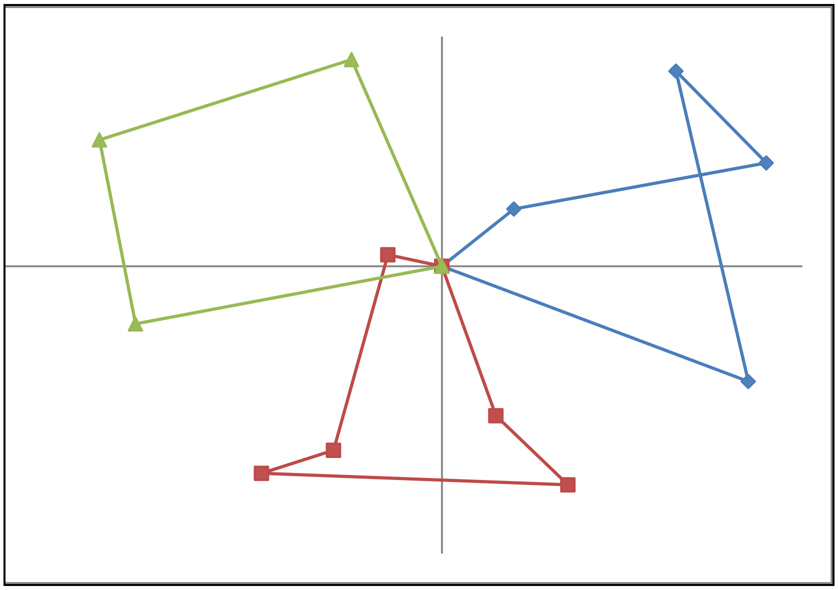


Fig. 3-13. Optimized routing using GA

Table 3-5. Daily travel time and unused time for optimized routing using GA

Day	Daily Flushing Time (minutes) (*DFT*)	Daily Travel Time (minutes) (*DTT*)	Daily Unused Time 360-*DFT*-*DTT*
1	282	70	8
2	271	62	27
3	225	68	-
TOTAL	778	200	35
Total VA	778	-	-
Total NVA	-	235	
NVA/VA%	235/778 = 30.21%		

However, several researchers have stated the shortcomings of traditional GA in solving combinatorial problems similar to VRP, as binary representations and classical crossover techniques can lead to invalid tours with missing or duplicate locations (Prins 2004; Potvin 2009). The limitation is observed for DOSP as well, when the model has been tested for randomly generated locations. Although the algorithm works reasonably well for a small number of locations (e.g., fewer than 50), its performance declines as the number of locations increases. Prins (2004) recommended several techniques to improve the performance of GA, such as efficient splitting procedure, or inclusion of a good heuristic for initial population. This study thus develops a heuristic algorithm, which is described in the following section.

3.5.6 Heuristic Algorithm

For the integer programming formulation of DOSP, it is evident that the problem size grows rapidly with increasing number of locations in the schedule. However, since the objective value in Eq. (2) is equivalent to $\left[T \times K - \sum_{i=1}^{n} E\left(T_i^f\right)\right]$, and assuming γ = *0.5*, a simplified expression of the objective function can be written as:

$$max\left[\sum_{k}\left\{\sum_{i} E\left(T_{ik}^{f}\right) + T_{0ik} + T_{j0k} + \sum_{i,j} T_{ijk}\right\}\right] \tag{3.24}$$

Subject to:

$$\begin{cases} min\left[\sum_{i,j}\left\{T_{ij} + T_{0i} + T_{j0}\right\}\right], and \\ \left[\sum_{i} E\left(T_i^f\right) + \sum_{i,j}\left\{T_{ij} + T_{0i} + T_{j0}\right\}\right] \le T, \quad \forall k \end{cases} \tag{3.25}$$

According to this formulation, it is possible to maximize the work time for daily subsets while minimizing the total travel for the scheduled period. Because of rapid runtime requirements for daily operations, this study proposes to use a meta-heuristic or heuristic algorithm for this problem. It should be noted that another assumption which simplifies DOSP formulation is that the productivity does not vary significantly across crews. This allows for the development of the model for a single vehicle, and each daily subset (or closed tour) of the optimized sequence can then be assigned to an individual crew.

From Eq. (3.24) and Eq. (3.25), it is clear that the objective of DOSP is essentially to maximize flushing time while minimizing travel time, which can be achieved by prioritizing the locations with longer flushing durations and shorter travel. Hence, a greedy approach is applied where the algorithm prioritizes the next location by the proportion of its flushing duration to the travel time. The algorithm flowchart is presented in Fig. 3-14, which shows that $\frac{E(Dur_{i+1})}{T_{i \to i+1} + T_{i+1 \to 0}}$ is used as the priority function (PF) at each step in order to rank all of the remaining locations. However, the constraint function is also checked for each remaining location to determine whether or not the remaining time within the shift is greater than its travel and flushing time. Once the location is flushed, it is assigned a negative priority value so that it appears at the bottom of the list of next possible locations. It should be mentioned that the expected flushing duration at location $i+1, E(Dur_{i+1})$ can be obtained from an on-site flushing duration estimation model (Zaman et al. 2015). Although the algorithm does not guarantee an optimal solution, the primary advantage of this algorithm is that it has a brief runtime (as it does not require any trials) and can be easily updated along the progress of the day if necessary.

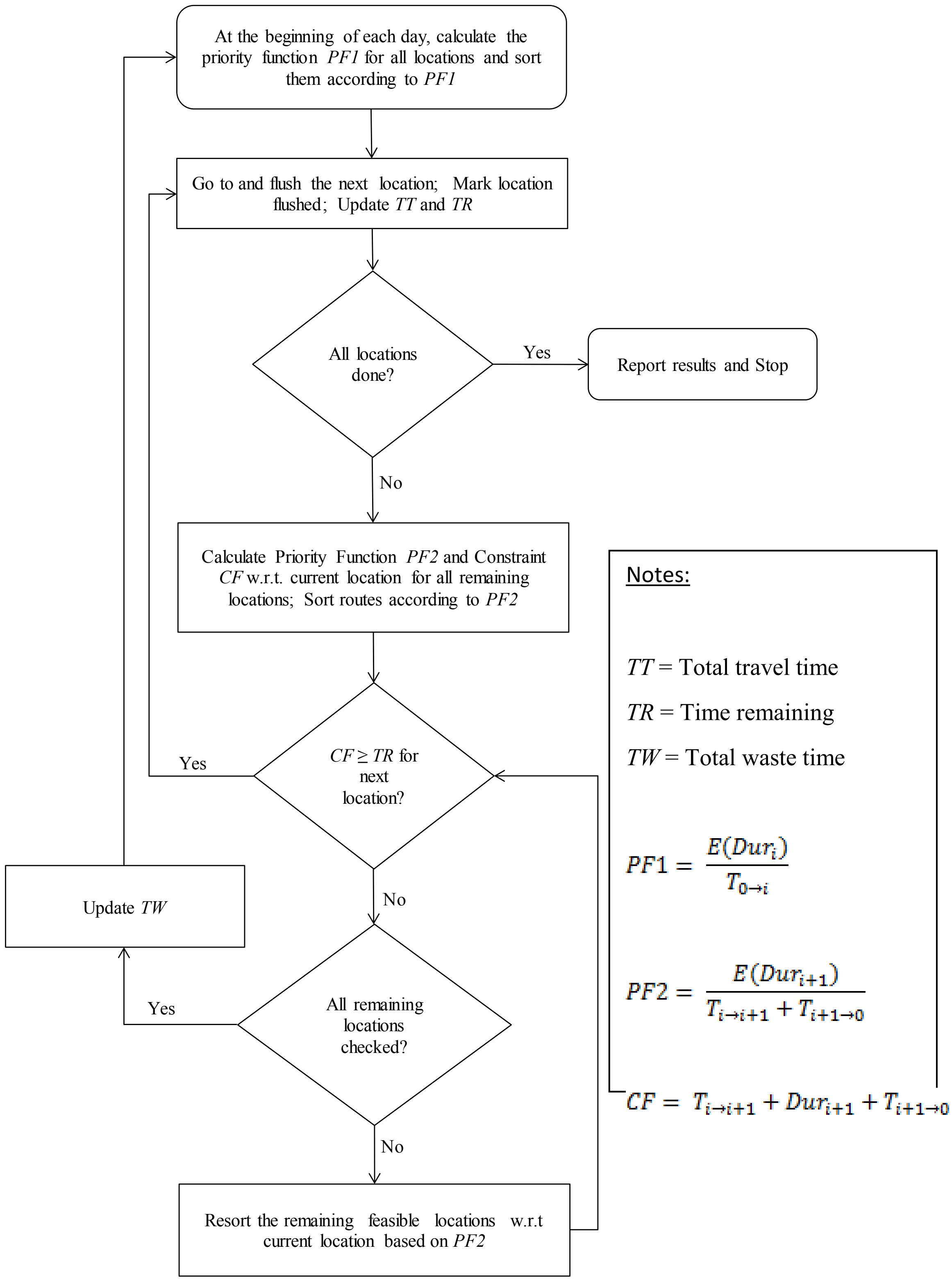


Fig. 3-14. Heuristic algorithm for DOSP

When the heuristic algorithm is used to optimize the above-mentioned set of locations, it is found that the result is not satisfactory. The NVA/VA ratio is found to be 38.1%, which is higher than neighbourhood-based, genetic algorithm, and ILP optimization (the results are presented in Fig. 3-15 and Table 3-6). This happens due to the fact that the algorithm attempts to prioritize the locations with higher value (longer on-site duration) and lower NVA time (travel time). In other words, there is a higher likelihood of choosing larger (longer-duration) locations at the beginning of a shift, and leaving the smaller locations to fill the bins toward the end. The small dataset used to test the algorithm does not offer sufficient options to fill the bins, and therefore the algorithm fails to realize its full potential in terms of performance. However, in reality, large municipalities have hundreds of locations in the schedule for which the heuristic is expected to perform better. Hence, the performance of the heuristic algorithm is evaluated for an increasing number of locations.

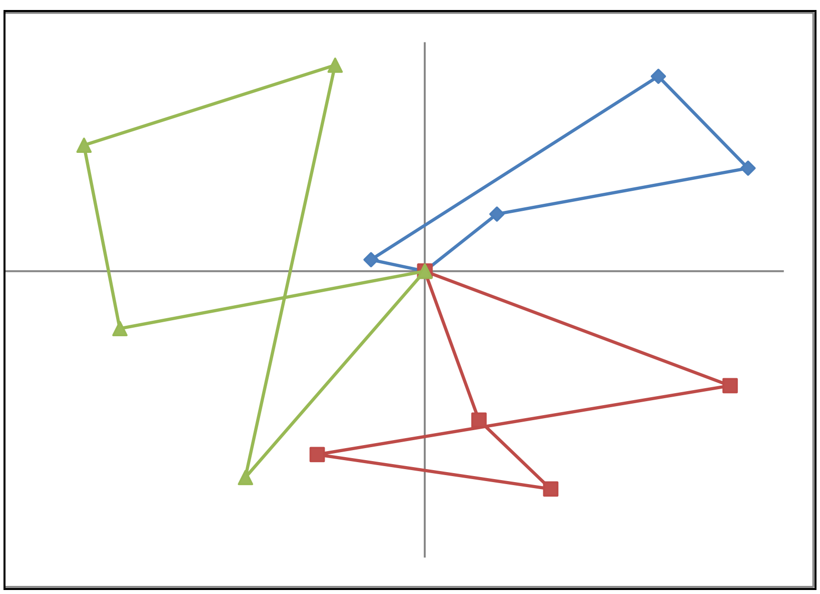


Fig. 3-15. Optimized routing using heuristic algorithm

Table 3-6. Daily travel time and unused time for optimized routing using heuristic algorithm

Day	Daily Flushing Time (minutes) (*DFT*)	Daily Travel Time (minutes) (*DTT*)	Daily Unused Time 360-*DFT*-*DTT*
1	255	56	49
2	275	77	8
3	248	106	-
TOTAL	778	240	56
Total VA	778	-	-
Total NVA	-	296	
NVA/VA%	296/778 = 38.10%		

3.5.7 Performance of Heuristic Algorithms for Increasing Number of Locations

This experiment entails creating randomly generated graphs with increasing number of locations. Different graphs are created, from a small 10-location graph to a 100-location graph, with the number of locations increasing by 10 for each consecutive graph. The size of the graph is kept the same to avoid the bias of excessive travel time with larger graphs. However, the expected durations vary between 20 and 200 minutes, following uniform distributions. This represents the situation where optimization is done for a longer period (e.g., 1 month). The graphs are then optimized by the greedy heuristic, and the traditional GA, and the expected NVA/VA% ratios are compared. The results are presented in Fig. 3-16 and Fig. 3-17, which are obtained from two different datasets. The effective shift lengths are 7 hours and 6 hours for datasets 1 and 2, respectively.

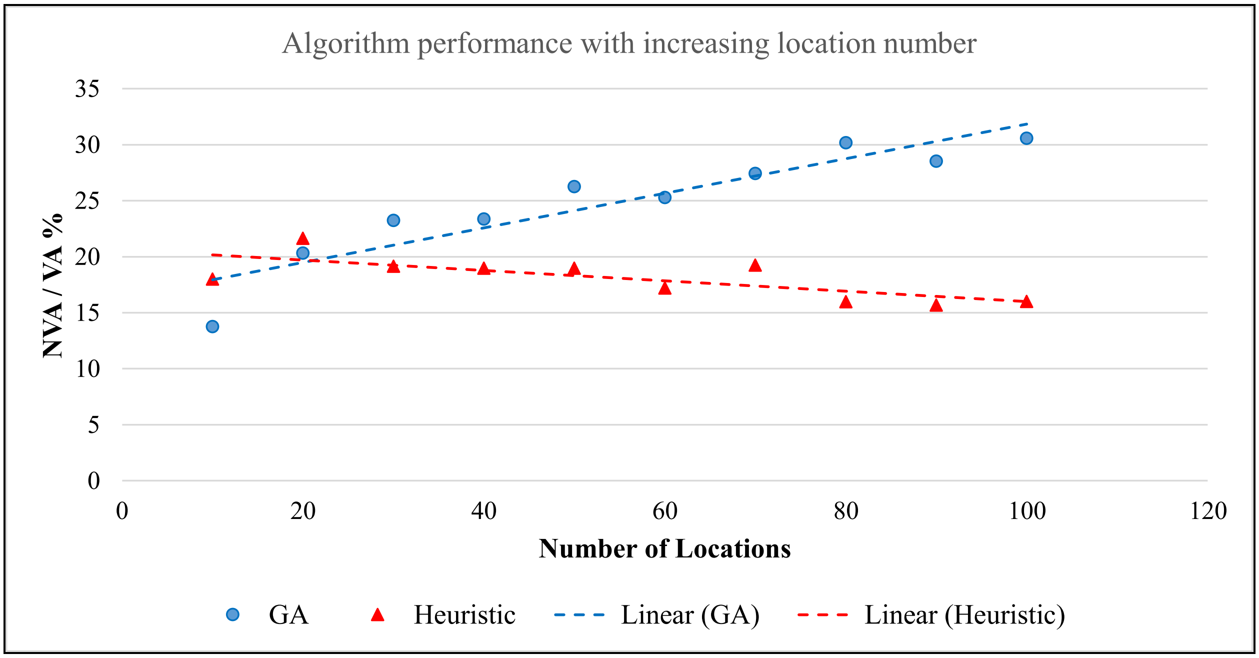


Fig. 3-16. Algorithm performance with increasing number of locations (random dataset 1)

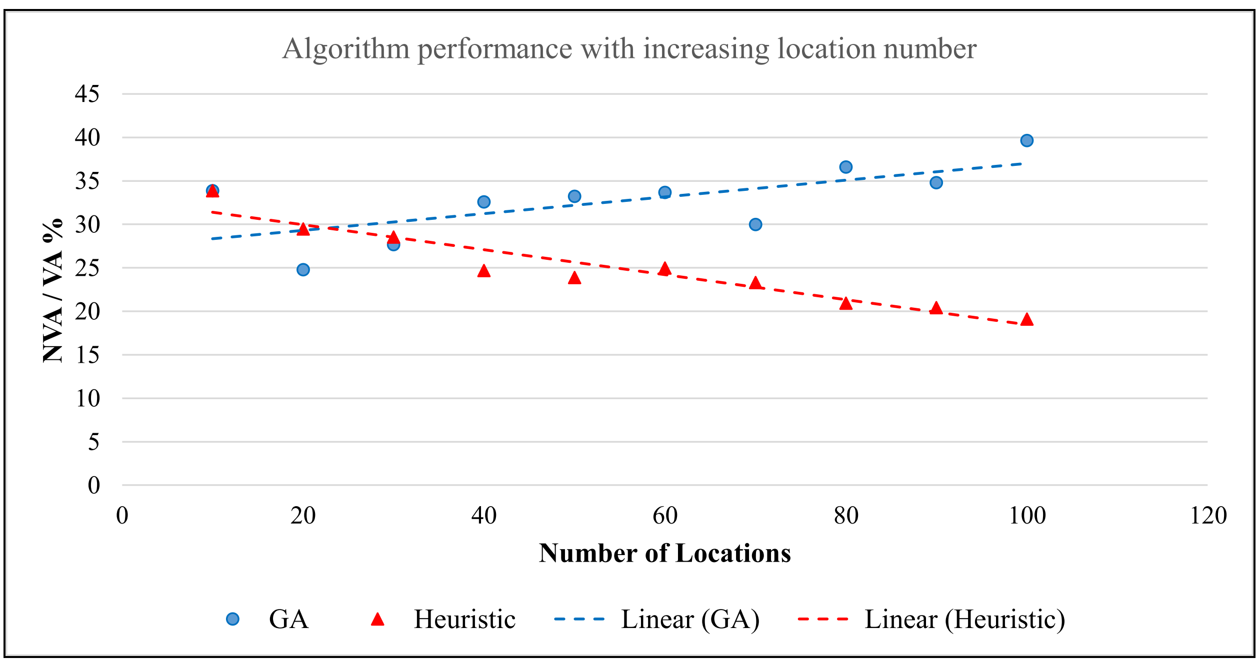


Fig. 3-17. Algorithm performance with increasing number of locations (random dataset 2)

It is found that the performance of traditional GA decreases for larger graphs, while the performance of the heuristic shows improvement with increasing number of locations. This is clearly observed by reduction in expected NVA/VA% ratio for heuristic algorithm. This complies with the expectation that the heuristic algorithm improves when it has a greater selection of locations to fill the smaller timeslots toward the end of a shift. On the other hand, the NVA/VA% ratios increase for larger graphs when traditional GA is used. This occurs because the same number of iterations (20,000) is allowed for all the datasets. It is expected that the performance of GA is likely to improve by allowing more trials and larger population size, although that would necessitate longer runtime. However, as suggested by Prins (2004), the performance of GA to solve combinatorial problems can also be improved by incorporating a strong heuristic as the initial population. To do so, a hybrid GA can be developed where the optimized sequence obtained from the heuristic algorithm is used as the initial population. The following section describes the hybrid algorithm.

3.5.8 Hybrid Genetic Algorithm

Genetic algorithm is a meta-heuristic stochastic search approach that can produce reasonably good solutions for a variety of problems. The algorithm can also be refined to improve its performance for particular needs. Over the years, researchers have developed many different variants of GA. In some cases, it has been combined with other search algorithms to develop hybrid GA (El-Mihoub et al. 2006). In the present study, a hybrid GA is developed by combining GA with the heuristic algorithm, which provides the generation of the initial population. The remaining steps of the algorithm are kept the same.

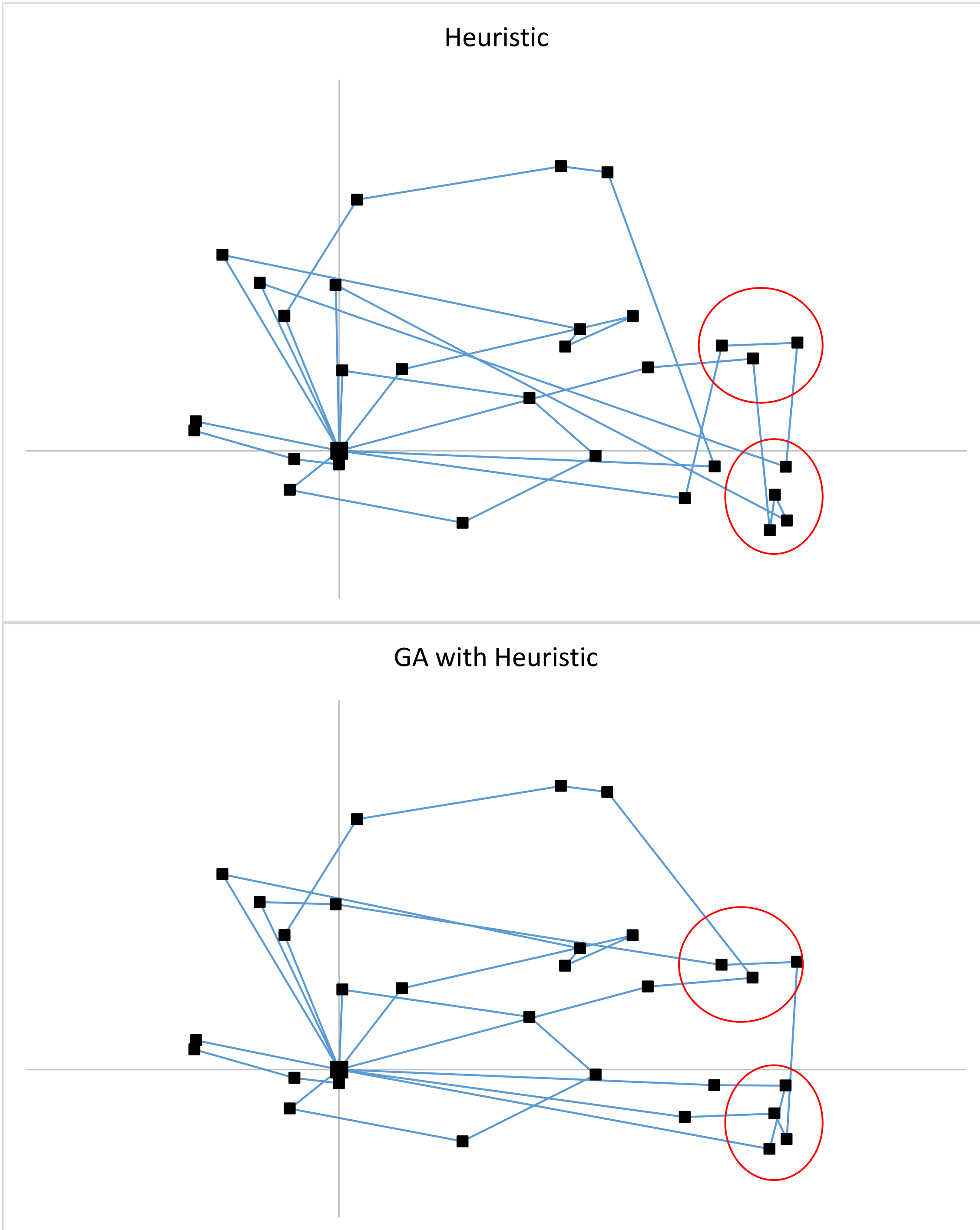


Fig. 3-18. Test case 1 - Optimum routing by (top) heuristic, and (bottom) hybrid GA

Table 3-7. Results for test case 1

Number of locations = 30		
Total on-site duration = 2039.22		
	Heuristic	Hybrid (Heuristic + GA)
Total Travel Time	274	232
Total Unused Time	28	13
Total NVA	302	245
NVA/VA	14.79%	12.02%

The hybrid GA is tested on several randomly generated graphs (test cases), and the improvement is compared with the heuristic algorithm. Results from two different datasets are presented here. Fig. 3-18 and

Table **3-7** present the routing results for the heuristic algorithm and hybrid GA for test case 1, where it can be seen that the hybrid GA makes a few modifications in the heuristic routing, which reduces the NVA/VA ratio from 14.79% to 12.02%. A similar result is observed for test case 2, where the NVA/VA ratio is reduced from 18.50% to 15.11% (Fig. 3-19 and Table 3-8).

This provides a clear indication that it is possible to improve the GA by using the heuristic sequence as the initial population. However, all of the random datasets used for testing the hybrid algorithm are relatively small (30 locations), with 100,000 trials (generations) required to reach this level of improvement; it is evident that achieving similar improvement for larger datasets will necessitate longer runtime.

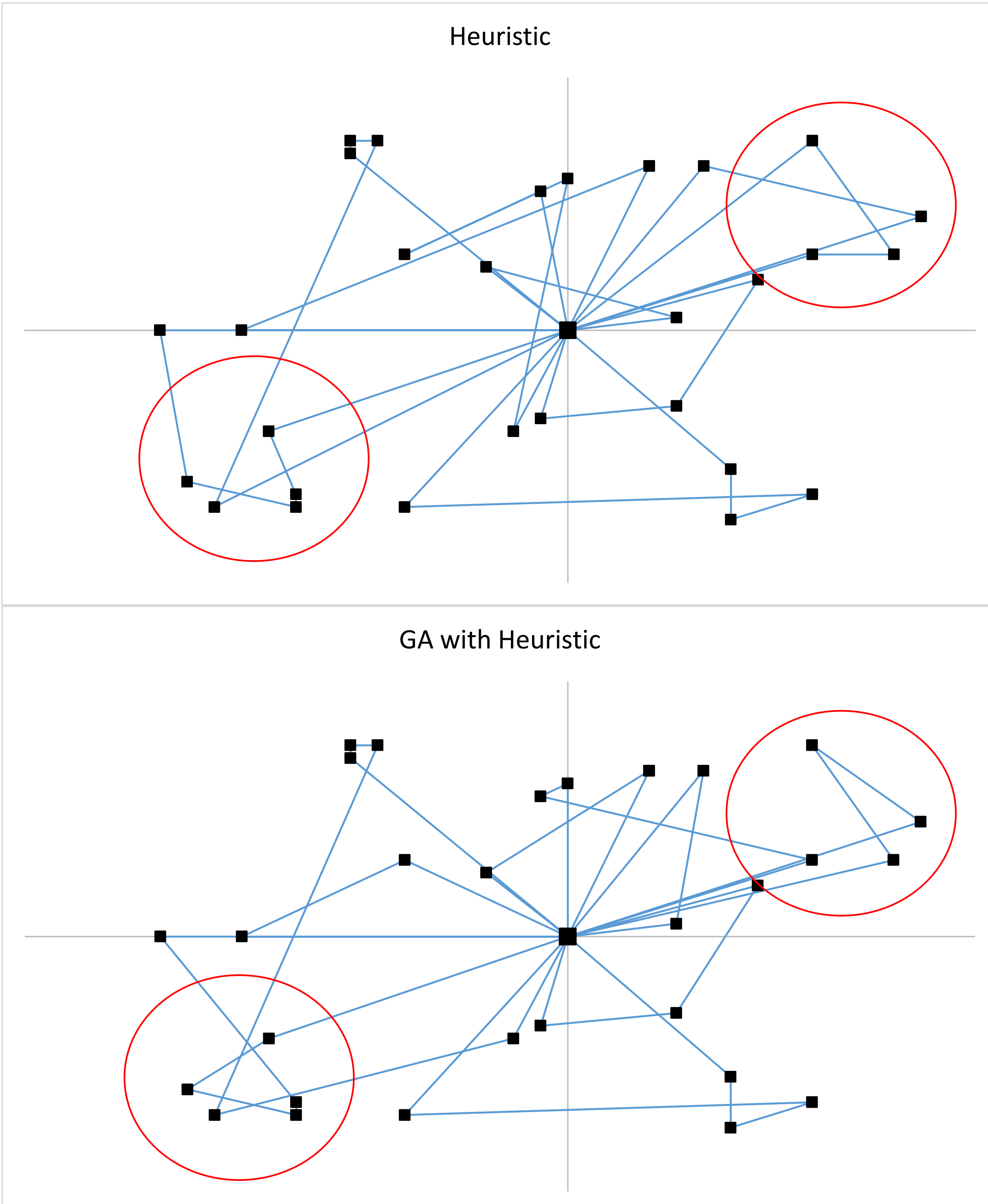


Fig. 3-19. Test case 2 - Optimum routing by (top) heuristic, and (bottom) hybrid GA

Table 3-8. Results for test case 2

Number of locations = 30		
Total on-site duration = 3123		
	Heuristic	Hybrid (Heuristic + GA)
Total Travel Time	397	372
Total Unused Time	181	100
Total NVA	578	472
NVA/VA	18.50%	15.11%

An interesting finding while producing improved solution in hybrid GA is the combination of cross-over and mutation probabilities. It is known that the success of evolutionary algorithm largely depends on these operators, and previous studies have reported a wide range of values for crossover and mutation probabilities (Nagata & Kobayashi 2013). A common combination is a relatively high crossover rate and a low mutation rate, and the traditional GA described in Section 3.5.5 in this study used values in the ranges of 0.8 to 0.9 and 0.1 to 0.2 for the two operators respectively. However, the optimum values for the hybrid GA is found to be 0.5 for both. A possible explanation for this is that when an already good solution (obtained from heuristic) is used as the initial population, it is not likely to be improved by crossover with a randomly generated inferior solution. Instead, setting a higher mutation probability increases its chances to improve by swapping locations within itself. The length of the mutation operator also plays an important role, which, in this case was done for a small portion of the sequence. This results in small changes in the route sequence which is clearly observable from the Figures. However, this aspect can be studied in further detail in future studies.

3.6 Simulation and Sensitivity Analysis

As described above, one of the unique properties of DOSP is the variation in on-site duration, which occurs across locations as well as across observations within a location. The on-site duration estimation model described in Section 3.4 is expected to capture the majority of the variations across locations, which is considered during the optimization process by means of expected value for each location. However, the uncaptured stochasticity for each location in the optimized schedule may still vary from the expected values. When multiple locations are scheduled for a shift, the duration variations in each location may compensate for each other, and the crews may still be able to finish the shift's work on time. However, large variations from expected values may cause deviation from the schedule and lead to overtime or unused time at the end of the shift. It is thus necessary to check the robustness of the proposed algorithms for variation in the on-site durations.

In the literature, several strategies exist to optimize stochastic process, which falls under the field of stochastic programming. Some examples of available methods include sample average approximation method (Kleywegt et al. 2001), two stage stochastic programming (Huang and Loucks 2000), a priori strategies (Bertsimas 1992), and robust optimization (Bertsimas et al. 2011). The basic idea underlying these methods is the combination or integration of optimization and simulation. In this study, the optimization is performed using expected on-site duration values for the locations to be scheduled, and then simulation is performed to evaluate the robustness of the proposed algorithm.

3.6.1 Effect of Variation in On-site Duration

Once the optimized schedule is available, the first step in simulating the effect of variation in on-site duration is to sample data from the appropriate distribution. For this purpose, it is assumed that the on-site duration for a given location is normally distributed with a known variance. Thus, it is possible to sample on-site duration for each location in the schedule using the inverse transform method.

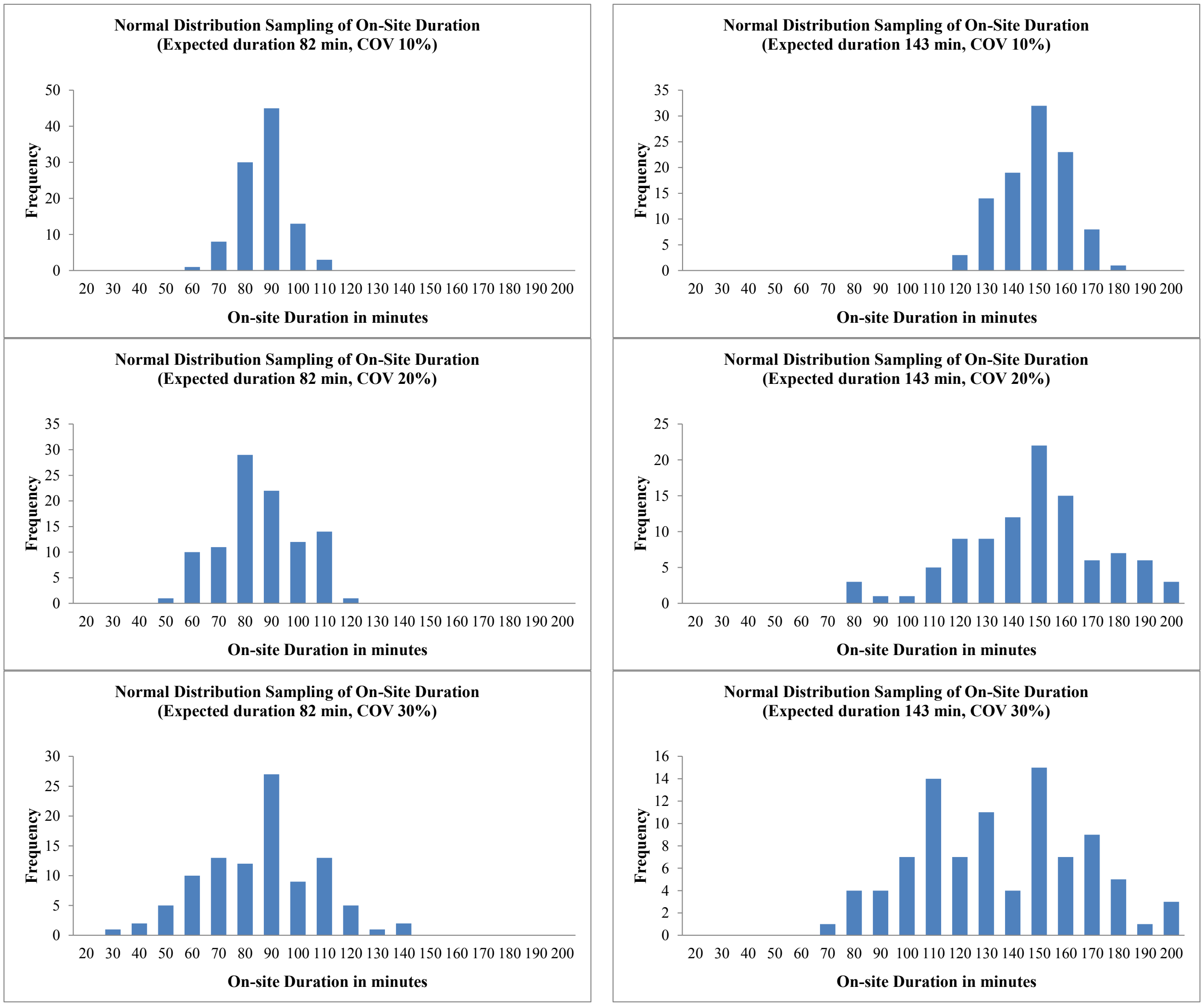


Fig. 3-20. Normal distribution sampling of on-site duration for two different locations

In reality, each location would have its own mean and variance; however, the expected duration (used during the optimization process) can be utilized as the mean for the simulation. And, then sampling is performed for each location using three different coefficient of variation (COV) levels—10%, 20%, and 30%. This allows exploring the sensitivity of schedule deviation on the variation in on-site duration. Fig. **3-20** presents normal distribution sampling results for two different locations for varying coefficient of variation. 100 runs are performed for sampling each location for each COV value. Here, the charts on the left of the figure represent a location having an expected duration of 82 minutes. The top chart shows the distribution for 10% COV, which can be seen to vary between 60 and 110 minutes. As expected, this range becomes wider with increasing COV. The center-left shows the sampled duration for 20% COV which ranges from 50 to 120 minutes. The bottom-left shows a variation between 30 to 140 minutes, which is highly unlikely for an 82-minute expected-duration location. However, majority of samples stay within the range between 70 and 110 minutes. The charts on the right show the distribution of sampled data for a location having an expected duration of 143 minutes. It is clearly observed that using the same COV ranges (10% to 30%) results in even wider ranges, where the data range from 70 to 200 minutes for 30% COV.

The next step is to simulate an actual shift's work by following the given optimized sequence by using the sampled on-site duration. Travel times are calculated by Euclidean distances between locations in the graph, and elapsed times are calculated at the end of each shift. If the elapsed time is less than the effective shift length, it is considered as unused time, while overtime time is considered when elapsed time exceeds the effective shift length. This way, unused or overtimes are recorded for each simulation run. This method is applied for a graph containing 100 locations, which has first been optimized using the heuristic algorithm. The result from the optimization provided with closed routes for 37 shifts. Travel times and unused/overtimes are recorded for each of these 37 shifts for each run of simulation. It is assumed during the simulation that the crews will adhere to the given sequence regardless of variations observed during the progress of the shift.

Fig. 3-21 presents the scheduled deviations obtained from 100 simulation runs. It can be seen that the variations in on-site durations may cause unused time (positive values in the figure) or overtime (negative values in the figure) at the end of the shift. As expected, the deviations are more probable for larger variations (higher COV value, e.g., 30%) in on-site duration. For 10% COV, the majority of the observations are within acceptable limits. However, an interesting finding at this point is that the mode of the schedule deviation data occurs on the positive side (unused time).

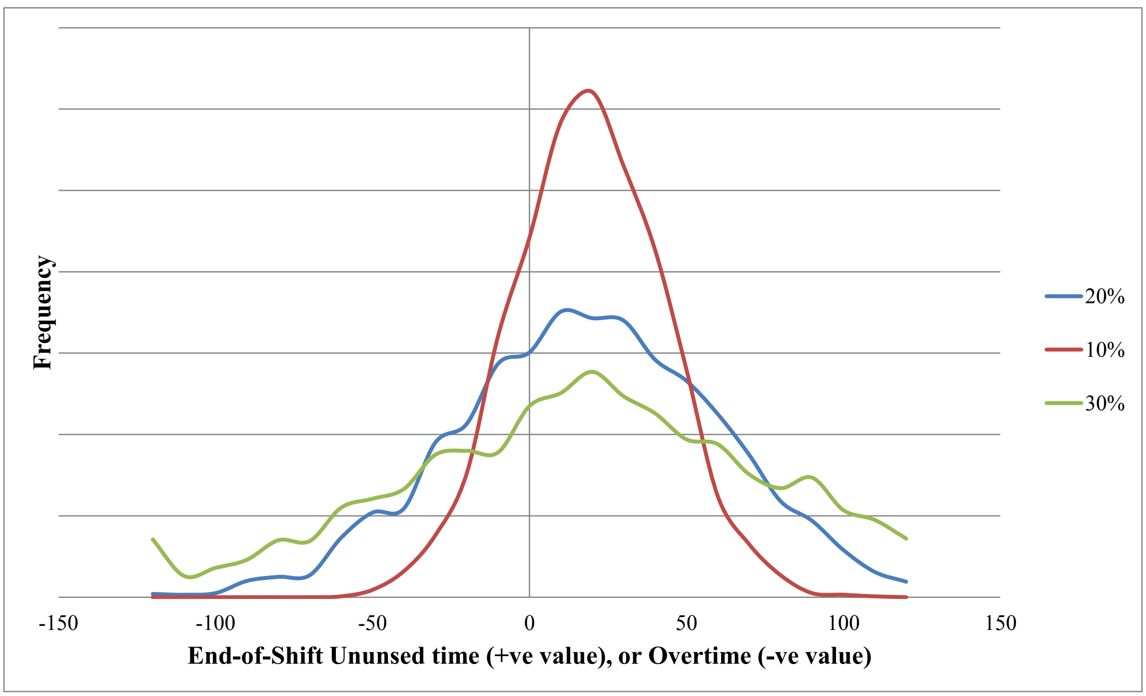


Fig. 3-21. Schedule deviation due to variations in on-site duration

The reason for this is the shift-length constraint used in the optimization algorithm. Because the effective shift length is used as a hard constraint, each optimized sequence has an expected unused time. Hence, a certain amount of unused time is always generated at the end of a shift, which causes the mode to appear on the positive side. It is common practice in operation research to assign penalties (as soft constraints) to reduce such occurrences. However, allowing a certain amount of leeway during the optimization can also shift the mode toward zero (Fig. 3-22). For example, it is found from Fig. 3-21 that the modes of all three distributions appear around the value of +20, which represents 20 minutes of unused time.

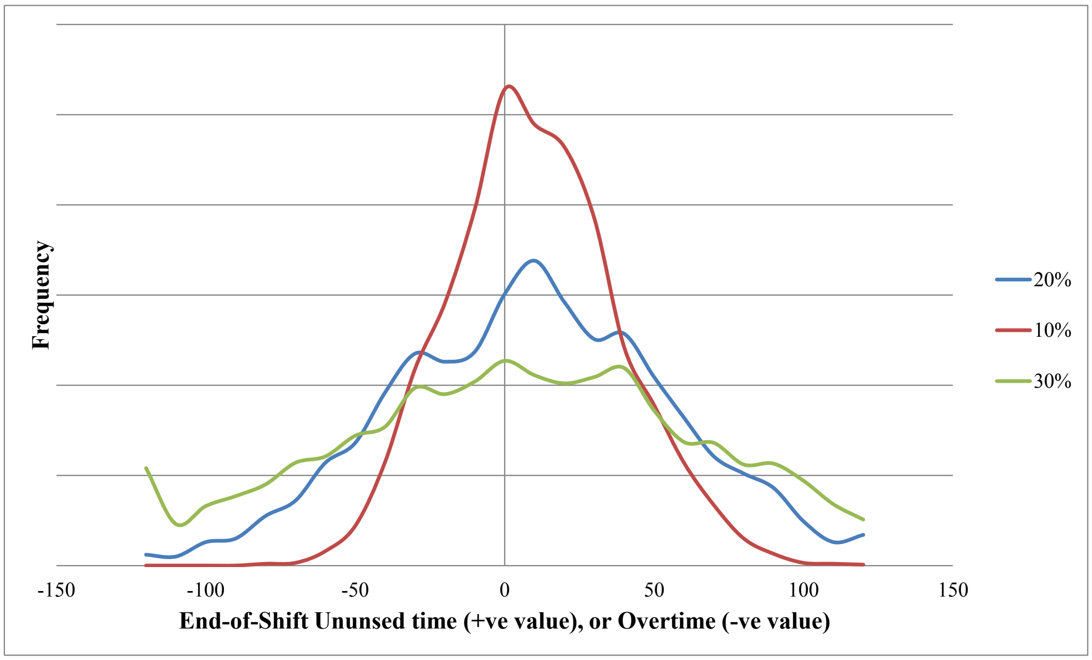


Fig. 3-22. Schedule deviation due to variation in on-site duration (with 20 minute leeway during optimization)

Therefore, the schedule is re-optimized using an effective shift length of 6 hours and 20 minutes, which provides a new optimized schedule. This schedule is then simulated for an effective shift length of 6 hours. The results are presented in Fig. 3-22, which clearly shows that the modes have shifted to the left and appear closer to zero. Hence, allowing leeway during the optimization process can effectively reduce the likelihood of having more unused time at the end of the shift.

However, on the contrary, this may result in more overtime. Therefore, the amount of leeway to be considered in the optimization depends on the operator's overtime policy, length of shift, size of municipality, and distribution of on-site duration. Notwithstanding, the probability of schedule deviation can be significantly reduced if the work process is standardized and a reliable on-site duration model is available.

3.6.2 Effect of Shift Length

In this section, the effect of shift length on expected daily productivity is explored. The analysis is performed for 10 different randomly generated graphs, each having 100 locations. The schedules are optimized using the heuristic algorithm for three different shift lengths: 8 hours, 9 hours, and 10 hours.

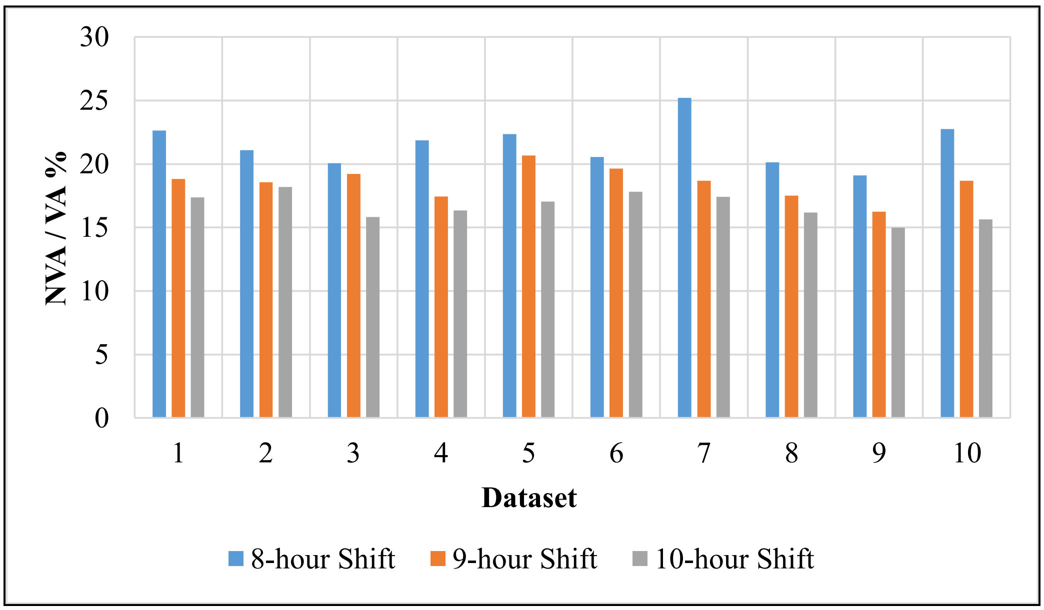


Fig. 3-23. Effect of shift length on non-value added time

For all cases, effective shift lengths are considered to be 2 hours less than total shift length to allow sufficient time for start-up, cleanup, lunch, and other breaks. As daily productivity is inversely proportional to NVA/VA ratio, it is expected that a lower ratio will result in higher productivity. For all 10 datasets, the results show that longer shifts can produce higher productivity (Fig. 3-23). However, this does not consider reduced productivity as a consequence of longer working hours.

3.7 Summary of Methodology

The basic framework of this research is presented in this chapter, including productivity analysis, data collection and descriptive statistics, development of an on-site duration estimation model, formulation and development of algorithms for schedule optimization, and simulation. The optimization and simulation is performed for randomly generated, fully-connected graphs. It is to be noted that the values of NVA or NVA/VA ratios are not to be taken as representative or conclusive values, as they depend on the properties of a randomly generated graph and Euclidean distance. However, the simulation results provide useful insight into the comparative performance of different algorithms that can be used to solve the drainage operations scheduling problem.

4 CASE STUDY

4.1 Introduction

The framework described in the Methodology chapter is applied to a case study in Edmonton, Alberta, Canada. Along with the results of the on-site duration model, schedule optimization using a heuristic algorithm, and simulation, this case study includes a review of operation and maintenance (O&M) practice at the City of Edmonton, productivity analysis, data collection, and descriptive statistics. Portions of this chapter have been presented at *CSCE Annual Conferences* (Zaman et al. 2012; Zaman et al. 2013), and published in the *Journal of Infrastructure Systems* (Zaman et al. 2014), and *Urban Water Journal* (Zaman et al. 2015). However, a more elaborate discussion of the research and case study is provided here.

4.2 Review of O&M Practice and Productivity Analysis

The Drainage Operations group at the City of Edmonton performs various O&M activities to operate and maintain its large collection system. These include inspection activities such as visual inspection, channel inspection, and CCTV inspection, as well as cleaning activities such as low-pressure flushing (LPF), high-pressure flushing (HPF), catch basin cleaning, and hydro-mechanical cleaning. Historical data reveals that, among these activities, HPF consumes the largest amount of time and resources; therefore, this case study particularly focuses on HPF (Fig. 4-1).

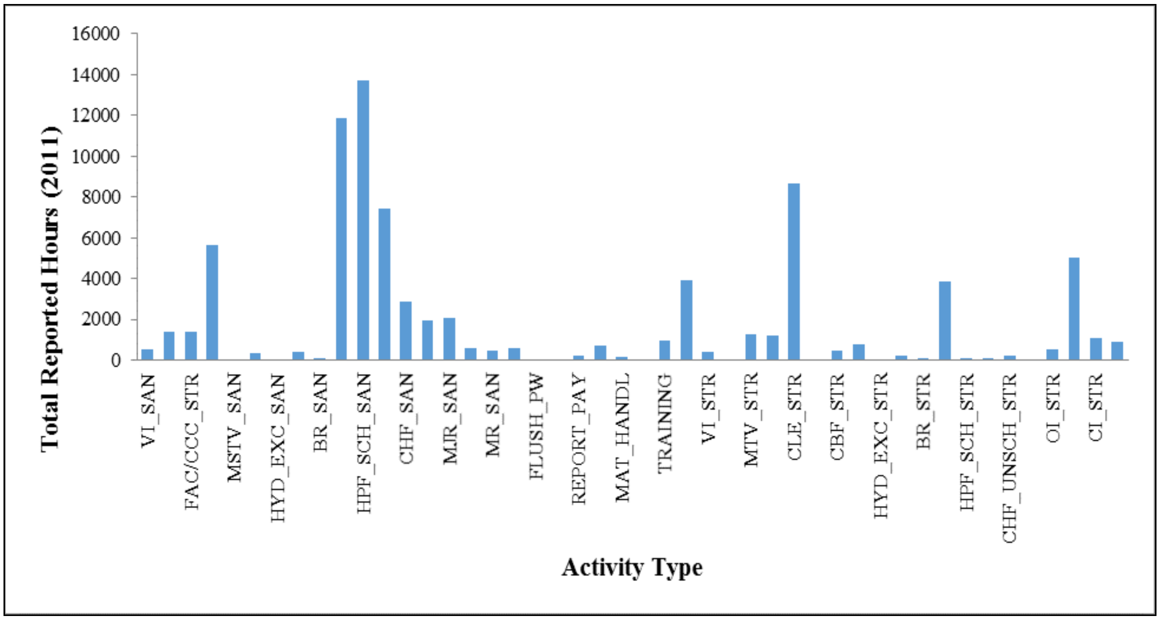


Fig. 4-1. Reported man-hours for various O&M activities at the City of Edmonton (2011)

4.2.1 Operational Preventive Maintenance Strategy

For large networks such as the one in Edmonton, it is not always feasible to proactively flush all of the pipes in the network on a regular basis. Hence, the City currently employs a selective approach where a pipe is selected for scheduled flushing based on its age, condition, relative importance, and problem history. Several inspection and judgment decisions are performed in order to select the cleaning activity and frequency necessary to maintain the operational condition of a pipe. Fig. 4-2 presents the process of selecting the scheduled operational maintenance for a pipe, where the activities are shaded in grey. The flowchart also shows the degrees of effectiveness for various activities, the interactions among them, and the continuous evaluation cycle. The development of this flowchart provides useful understanding of the selection process that establishes the total maintenance volume (i.e., demand) for the operator.

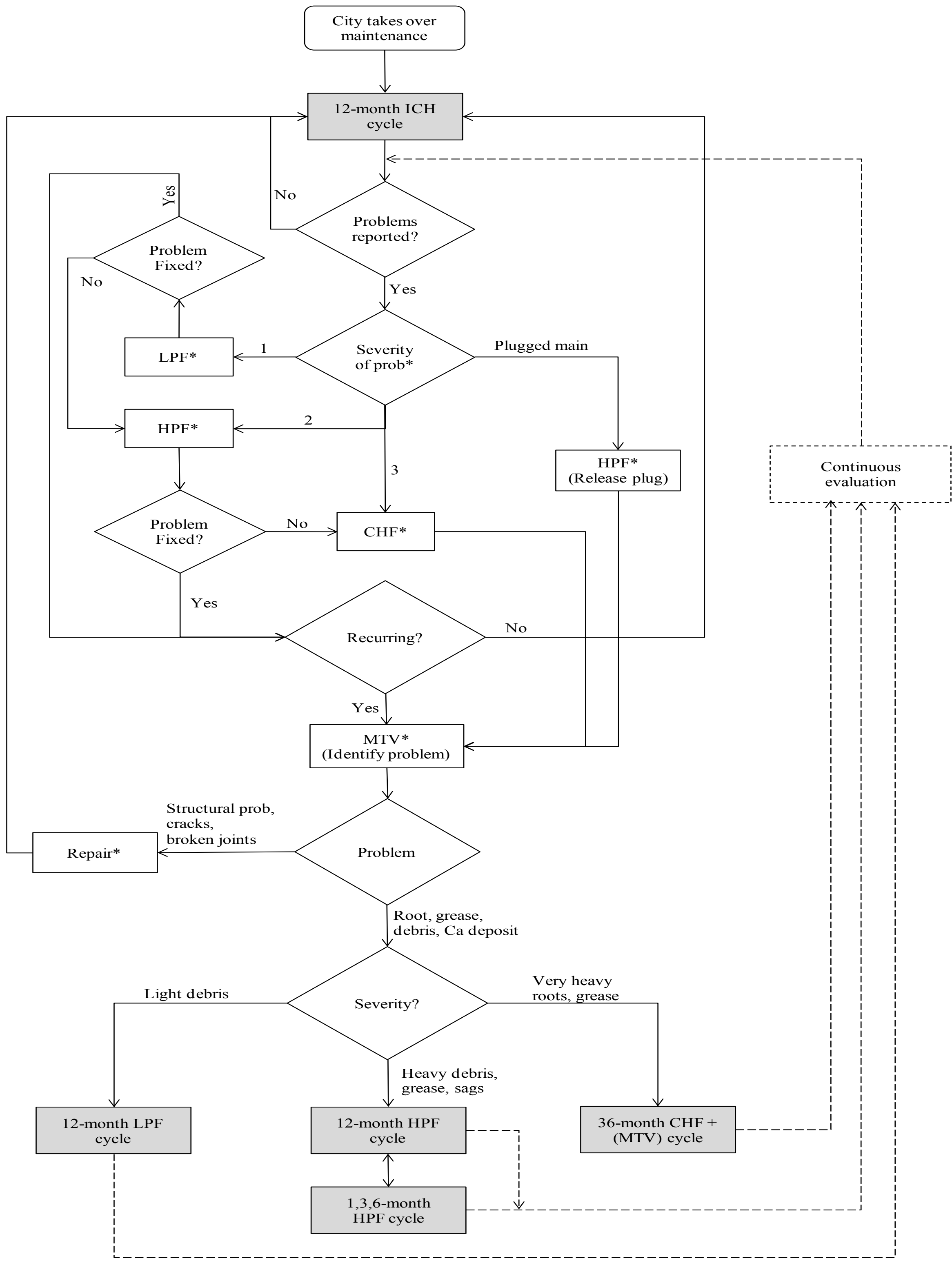


Fig. 4-2. Flowchart showing selection of pipes for operational preventive maintenance

4.2.2 High-Pressure Flushing (HPF)

High-pressure flushing (HPF) (also known as jetting) is employed to remove debris, grease, calcium deposit, and small roots from inside sewer pipes using water blasted at a pressure over 2,000 psi (EPA 1999; Bowen et al. 2003). State-of-the-art flushing vehicles (also known as flusher trucks, or combo units), equipped with various sizes of nozzles, pump, water tank, and debris tank, are used to perform flushing operations (Fig. 4-3).

Fig. 4-3. High-pressure flushing truck

HPF requires a crew of two people, who begin their shift at the yard (there is only one yard in Edmonton), travel to the pre-scheduled locations in a flusher truck, and perform the necessary tasks at each location. Upon completion of the flushing at one location, the crew travels to the next location, and so on, until the shift ends. On a typical day, 5-6 HPF crews work simultaneously at different locations throughout the city. The number of locations flushed in a typical 8-hour shift varies greatly, depending on the flushing duration at each location and travel time between them. Thus, the effective on-site maintenance time is significantly reduced by the large amount of travel time. Furthermore, the on-site flushing durations at different locations vary from less than 10 minutes to several hours, and are stochastic in nature (Zaman et al. 2013). Because there is no reliable model currently in practice by which to estimate this variable flushing duration, the crews are unable to predict the flushing duration of the next location. Hence, crews often return to the yard prior to the end of the shift time, which leads to unproductive unused time (also referred to as end-of-shift waste time in this thesis).

In 2014, the City of Edmonton performed scheduled HPFs at over 1,400 pre-designated locations across the city as part of its annual preventive maintenance (PM) regimen. Because some locations require more frequent flushing than others, each is pre-scheduled for periodic HPF at a particular frequency such as every 1 month, 3 months, 6 months, or 12 months. Fig. 4-4 shows the locations of scheduled HPF, where red, purple, green, and yellow dots represent the above-mentioned frequencies, respectively. As seen in the figure, older neighbourhoods and those with combined sewer systems (near the center of the city) require a higher frequency of scheduled flushing than do newer neighbourhoods. At the beginning of each month, a query to a central database generates a list of HPF work orders for the locations that are due that month. These work orders, grouped by neighbourhood, are passed on to the crew supervisor, who then assigns sets of locations to the individual crews. Monthly scheduled HPF locations for four different months are shown in Fig. 4-5. It is found that some months can have scheduled locations well distributed across the city, while some other months can have locations more concentrated in particular neighbourhoods. Either way, each of the monthly HPF schedules has over a hundred locations where there is an opportunity to apply the proposed framework.

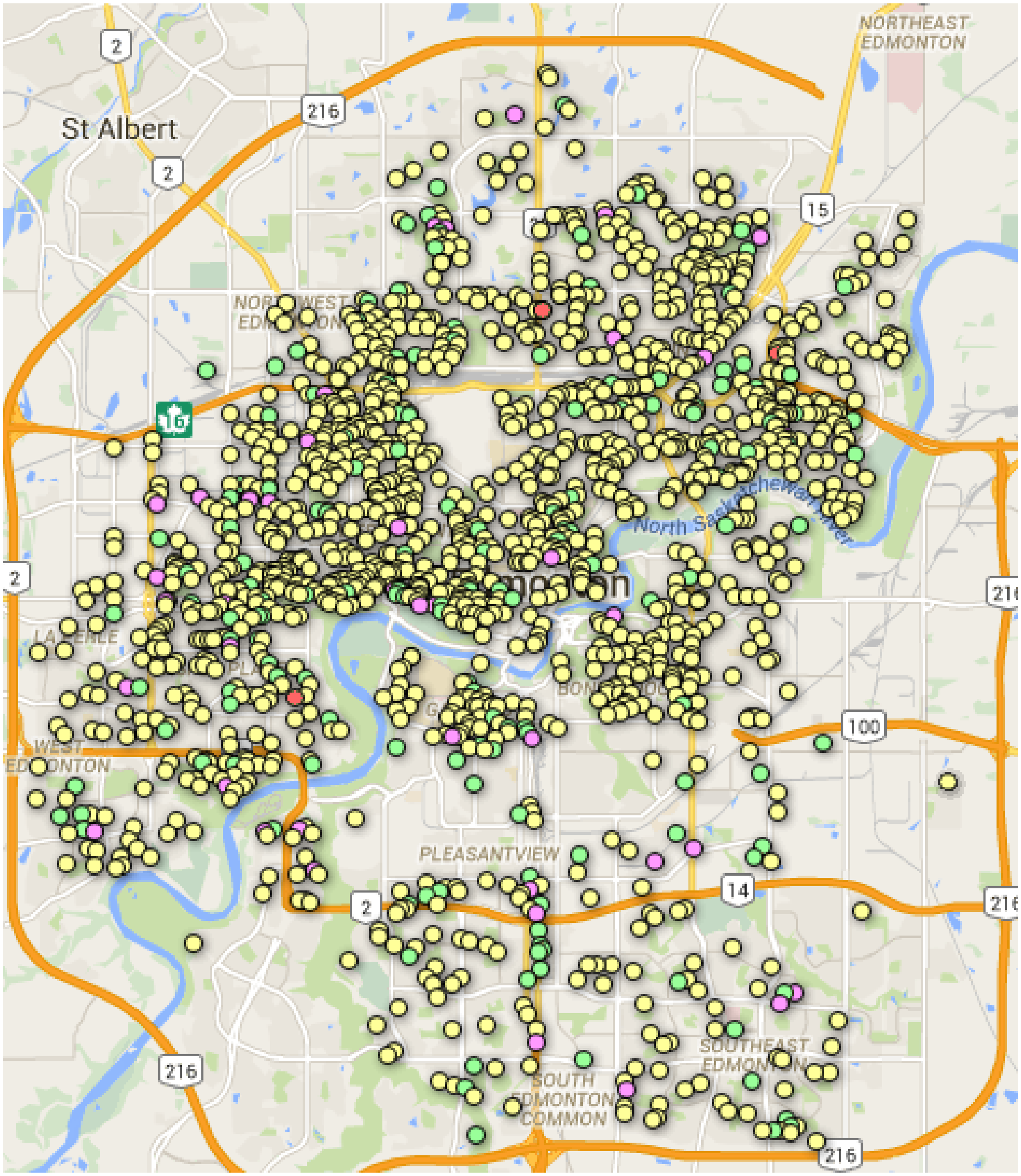


Fig. 4-4. Scheduled HPF locations in the City of Edmonton (as of 2014)

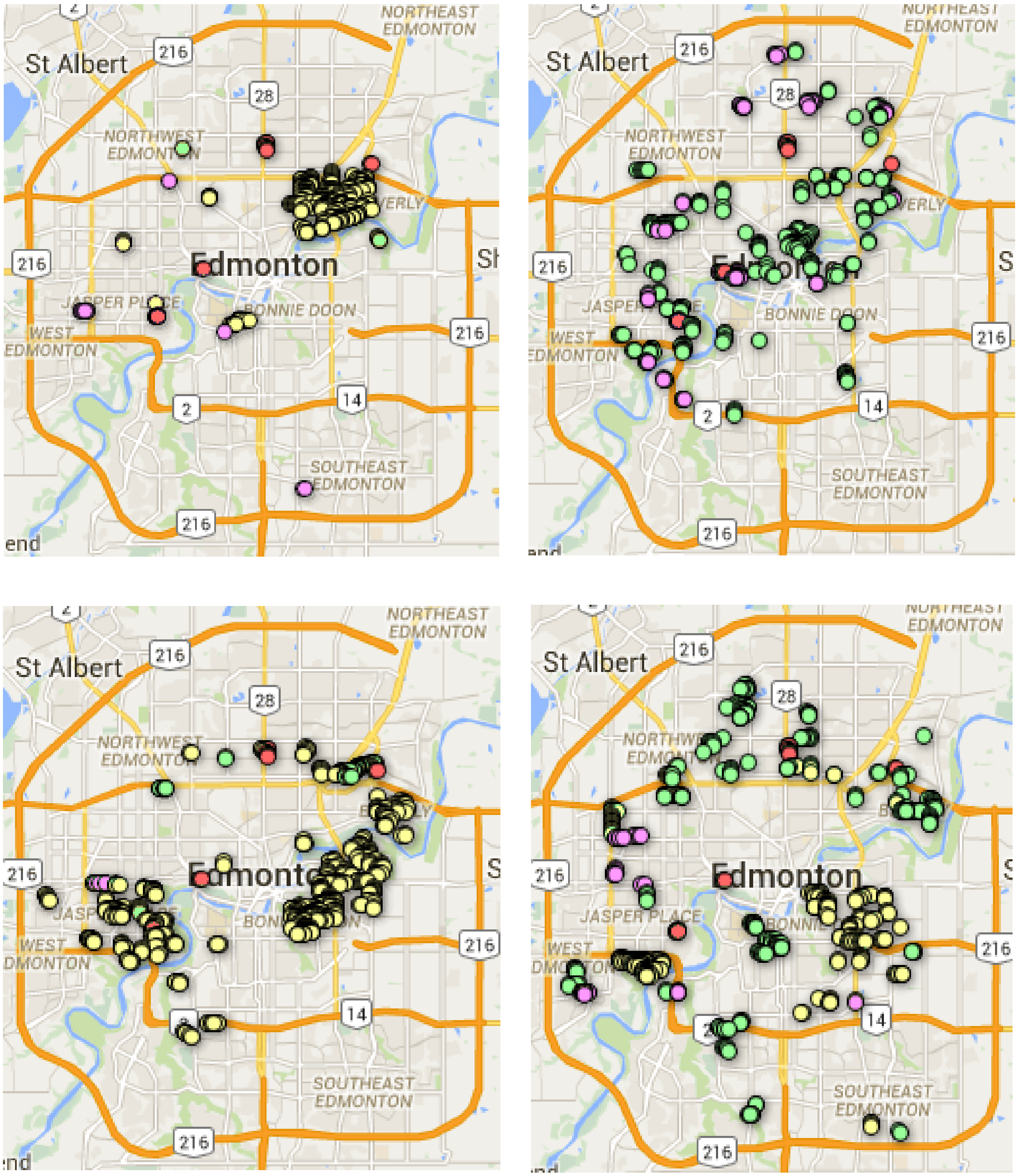


Fig. 4-5. Monthly scheduled HPF locations in four different months

4.2.3 Productivity Analysis

The productivity of these O&M activities is expressed as accomplishments per unit time, e.g., *locations/day* for VI, and *m/man-hour* for LPF, HPF, and CHF. In a benchmarking study at the City of Edmonton, Bowen et al. (2003) have developed daily productivity standards for each activity. Fig. 4-6 presents year-long daily productivity charts for HPF, which have been created using historical data from the City of Edmonton. It can be observed that although the average annual productivity exceeds the target (33 m/man-hour), there are wide-ranging variations across observations.

The primary cause of such variation is that each productivity observation is calculated based on the reported man-hours, which include start-up preparation, end-of-shift cleaning, lunch and coffee breaks, and travel time. Agbulos et al. (2006) have applied lean principles in order to improve the efficiency of drainage operations, having identified travel as a non-value added task that consumes a high proportion of daily work time. However, it is easily understood that travel time for a particular maintenance location is dependent upon its distance from the yard and from other locations. For a large city, the flushing locations that are nearer to the yard require much less travel time than more distant locations. The on-site flushing duration also varies across maintenance locations depending on several factors (Chughtai & Zayed 2008; Zaman et al. 2013). Thus, both on-site flushing duration and travel time contribute to variations in daily productivity, and it is evident that consideration of these two factors during maintenance scheduling leads to more efficient use of resources.

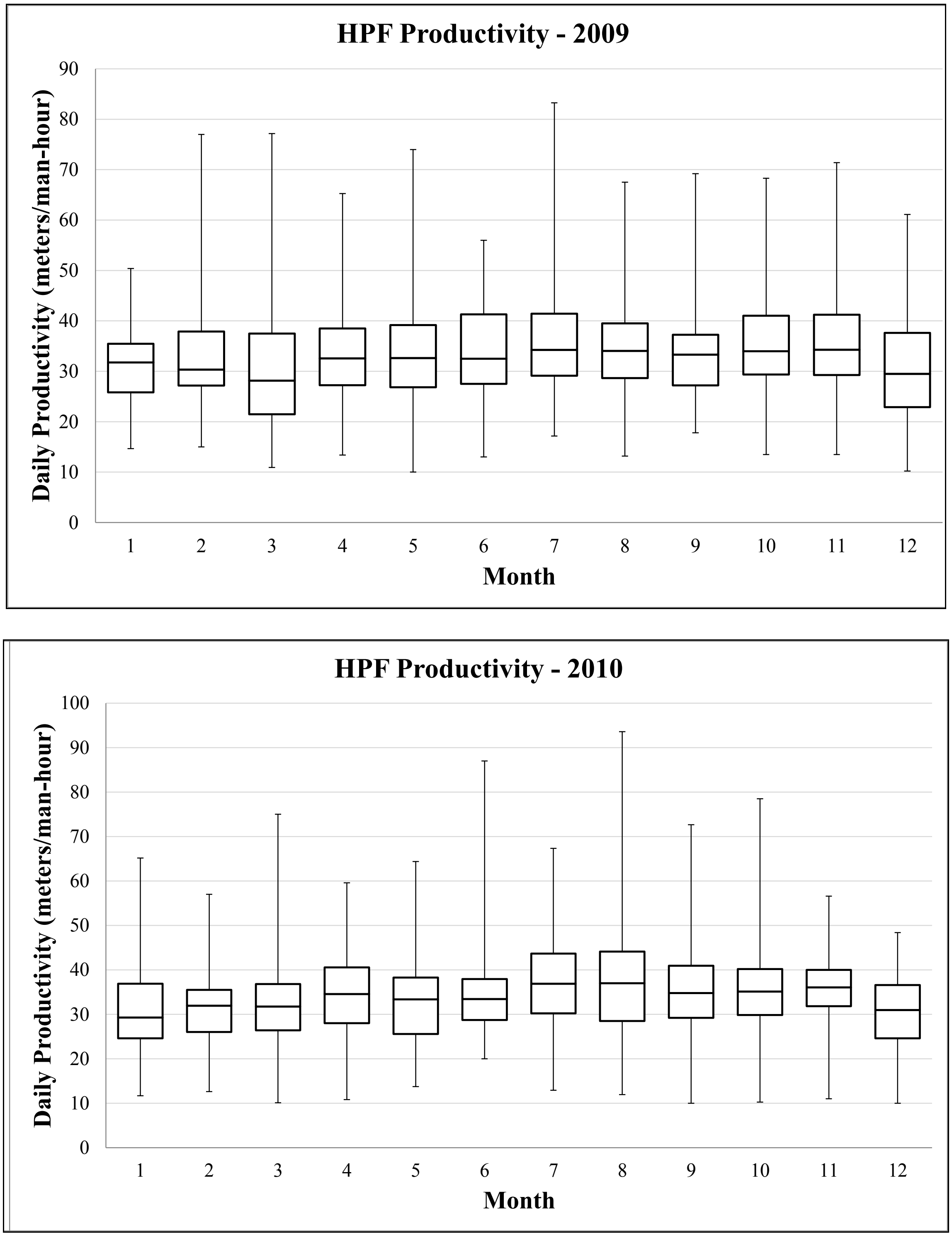


Fig. 4-6. Daily productivity charts for HPF at the City of Edmonton

4.3 Data Collection and Descriptive Statistics

4.3.1 Data Collection

The dataset required for the case study is created by merging several databases maintained by the City of Edmonton, as well as by collecting necessary data. Fig. 4-7 illustrates the data collection schematic for the on-site duration estimation model. The "Drainage Pipes Properties" database provides the physical properties of the pipe segments—diameter, length, slope, material, year of construction, and location coordinates. The "Timesheet" and "Workorder" databases are used to collect the scheduled location information (location number, pipe segments within the location, flushing frequency, location, scheduled flushing date, etc.) and crew information (the specific crew assigned to a particular location, vehicle ID, and flushing date).

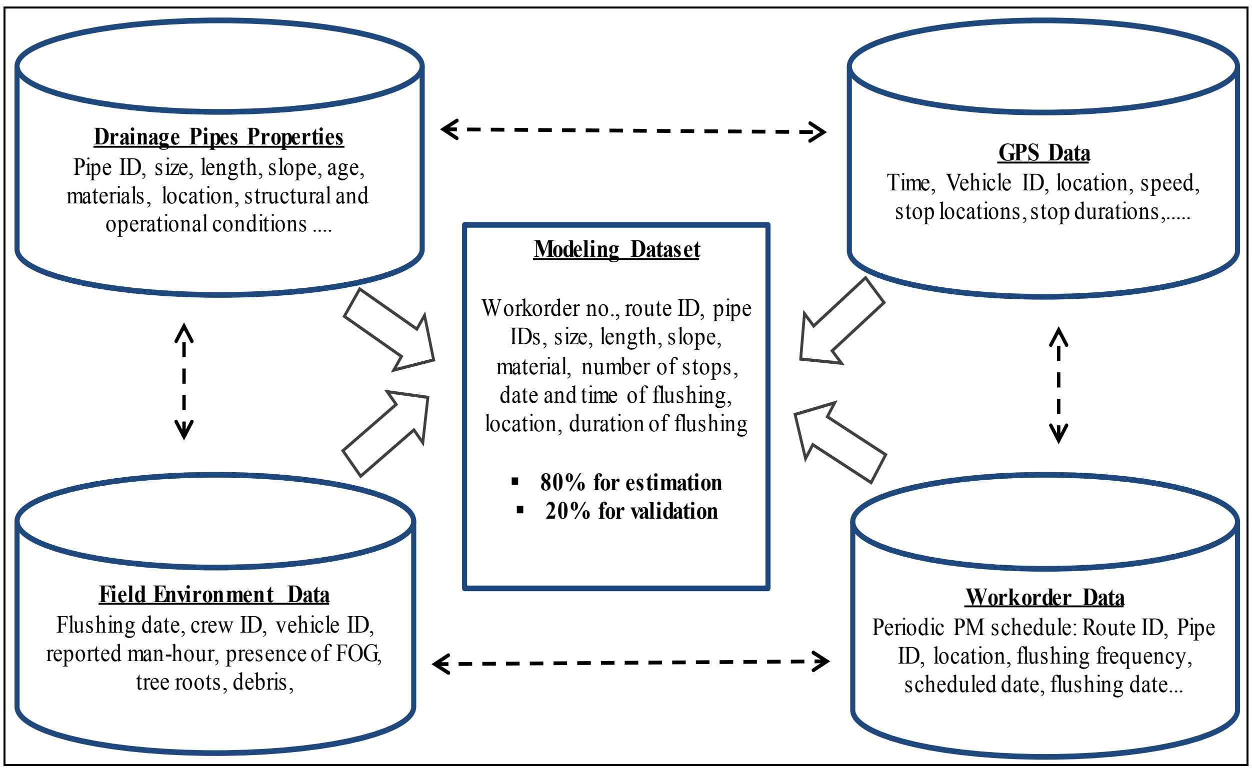


Fig. 4-7. Dataset preparation schematic for on-site duration model

The actual on-site flushing duration data is collected from the "GPS" database, which records the location, time, and speed for each of the flushing vehicles. Then, the on-site flushing durations for all scheduled locations are linked with the parameters obtained from the other databases. The challenge during this phase is that the GPS database does not share any primary or foreign key with the other databases. Accordingly, connection of the GPS data with other data repositories (flushing date, location, and vehicle ID) is carried out manually and consists of searching for the vehicle used by a particular crew on a given day and identifying that vehicle's stop near the job location. This process assumes that a crew is performing flushing activities when its vehicle is found to be idle (in a stationary position with the engine running) at a location scheduled for that day. In addition, the modelling dataset contains information such as crew experience, (with regular HPF crews marked as 1 and used in the model as a dummy variable), types of trees located near the pipes, location type (e.g., residential, commercial, institutional), and the presence of restaurant or carwash facility in the vicinity.

The dataset collected for the on-site flushing duration model contains observations at various locations across the city between 2009 and 2012. After necessary noise cleaning, the final dataset contains 448 observations. Among these, 85% of the observations (381) are randomly selected for model estimation, while the remainder (67) are used for model validation. The list and descriptions of the variables in the dataset are presented in Table 4-1.

Table 4-1 List and description of variables in the modelling dataset

Variable Name	Description	Range	Variable Type
Flushing_duration	Total time taken to flush	10 ~ 339 minutes	Continuous
Number_of_pipes	Total number of pipes	1 ~ 18 nos.	Discrete
Total_length	Total length of pipes	3 ~ 1132 m	Continuous
Number_of_stops	Number of times the vehicle stops in order to complete flushing	1 ~ 14 nos.	Discrete
Average_diameter	Average diameter of all pipes	15 ~ 67.5 cm	Continuous
Average_depth	Average depth of downstream manholes of the pipes	2 ~ 10 m	Continuous
Age_of_pipes	Average age of all pipes	14 ~ 105 years	Continuous
Flush_per_year	Number of flushes per year = (12/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	Residential, Commercial, Industrial	Categorical
Material	Material of pipes	Concrete, Clay	Categorical
Crew Experience	Regular HPF Crew	1 or 0	Categorical

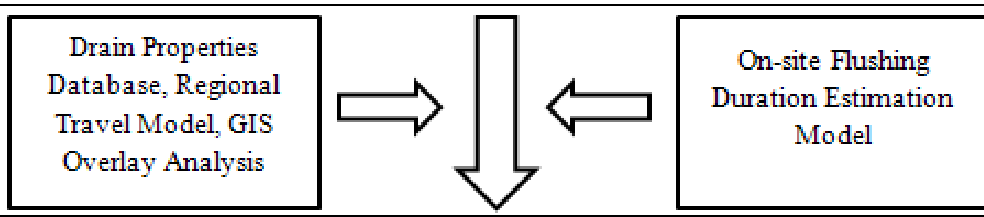
It should be noted that the dataset is an unbalanced panel dataset where each of the five 1-month-frequency locations contains multiple observations, which provides the opportunity to explore the seasonal variation in on-site productivity. Because the 1-month-frequency locations are flushed more frequently, it is easier to collect a large amount of data for these locations. Nevertheless, the data is taken from independent flushing instances that have taken place at different points in time (i.e., different months); hence, the responses are independent from one another.

In addition to these data, the optimization model requires travel time information. As mentioned above, this study uses randomly generated coordinates for flushing locations, and Euclidean distances as travel time (assuming travel time is proportional to distance for constant speed) during the development stage of the optimization algorithms. However, in practical scenarios, the travel time between two locations is dependent upon a number of variables, including existing road network, route assignment, speed limits, traffic volume, and time of day. Moreover, the drainage maintenance vehicles must follow the designated truck routes for most of the trips.

This study collects these travel time data from two different sources: the first data source is Google API, which provides reliable travel times and distances between two locations. The advantage of using Google API is that the input can be a physical address, latitude-longitude coordinates, or even a neighbourhood (for cases in which accurate location data is not available). However, implementation of this method reveals that the optimization runtime increases significantly due to the time necessary for extraction of this data from the Internet. Also, there is a daily limit of data extraction, which restricts its applicability to problems of a moderate to large size. This can be overcome using regional travel model data obtained from the city's transportation authority.

The City of Edmonton's Regional Travel Model (RTM) is developed and maintained by the city's transportation department. The model divides the city into 666 transportation zones and provides the travel time matrix. If the zone number of each maintenance location, which can be obtained by performing geographic information system (GIS) overlay analysis, is known, it is possible to easily extract the travel time from the matrix. Although this method uses approximate locations (through zoning) instead of actual, there are several advantages of using this source: (*i*) rapid extraction of data during optimization run; (*ii*) inter-zonal, as well as intra-zonal, travel times; (*iii*) different travel times for AM/PM peaks, and off-peak hours; and (*iv*) explicit travel time data for trucks that considers the use of designated truck routes. RTM is thus used in order to optimize the following case study. The data preparation steps for the optimization model are presented in Fig. 4-8.

Order#	Pipe ID	Length (meter)	Latitude	Longitude
80690208	39682	16.57	53.50976042	-113.6517788
80690208	307983	4.94	53.50973942	-113.6515315
80690209	308045	38.75	53.5037715	-113.6525181
80690209	39435	30.32	53.50376619	-113.6519341
80690214	41884	30.17	53.52030022	-113.6236939
80690214	41883	99.08	53.52004525	-113.6238486
80690214	311028	7.62	53.51917273	-113.6241458
80690215	41558	90.74	53.51622663	-113.623434
80690216	41891	77.11	53.51939982	-113.6187567
80690216	310953	8.81	53.51867519	-113.618718
80690217	55224	36.58	53.58013711	-113.5643459
80690217	55223	36.51	53.57980724	-113.5637965
80690217	55193	110.3	53.58013533	-113.5637935
80690217	55194	109.73	53.58013002	-113.5621281
80690217	55195	52.72	53.58012468	-113.5604713
80690218	55197	80.72	53.57939936	-113.560478
80690219	62187	127.4	53.52267424	-113.5904382
80690219	62188	118.81	53.52152947	-113.5904384
80690219	62189	101.46	53.52024416	-113.590438
80690219	71241	121.2	53.51933251	-113.5904344



Unique Order (Location ID)	Number of Pipes	Total Length (meters)	Transportation Zone	Estimated Duration (min)
80690208	2	21.51	609	27
80690209	2	69.07	609	30
80690214	3	136.87	608	53
80690215	1	90.74	608	12
80690216	2	85.92	608	31
80690217	5	345.84	802	103
80690218	1	80.72	802	12
80690219	4	468.87	511	91

Fig. 4-8. Dataset preparation steps for schedule optimization

4.3.2 Descriptive Statistics

Initially, descriptive statistical analyses are performed in order to obtain a complete understanding of the dataset and the correlation between the variables. Preliminary results show that the on-site flushing duration varies from 10.00 to 339.00 minutes, with an average value of 70.93 minutes. The standard deviation, median, and mode of the data are 58.42, 51.00, and 29.00 minutes, respectively. In order to investigate the cause of such wide variance, flushing duration is plotted against the system attributes (an excerpt of which is shown in Fig. 4-9). As expected, flushing duration has a strong linear correlation with number of pipes and total length of location; however, the pipe diameter and depth do not seem to affect the flushing duration. It is of interest at this juncture to explore the effect of the other predictor variables by analyzing subsets of the data. When the flushing durations are grouped by frequency, different patterns for 1-, 3-, 6-, and 12-month-frequency locations can be observed. Analysis of the means (of on-site duration) for each of these location frequencies exhibits different variance, revealing heteroskedasticity in the data, especially for 12-month-frequency locations. The cause of such variation is that the frequent locations are flushed more regularly, which results in more consistent on-site durations. Moreover, many of the 12-month-frequency locations are performed by temporary crews during the summer months, causing similar variations when the dataset is grouped by month (Fig. 4-10). This heteroskedasticity violates the assumption of linear regression, and thus suggests that either separate models should be developed for each frequency, or that the variations should be captured in the model by including these factors.

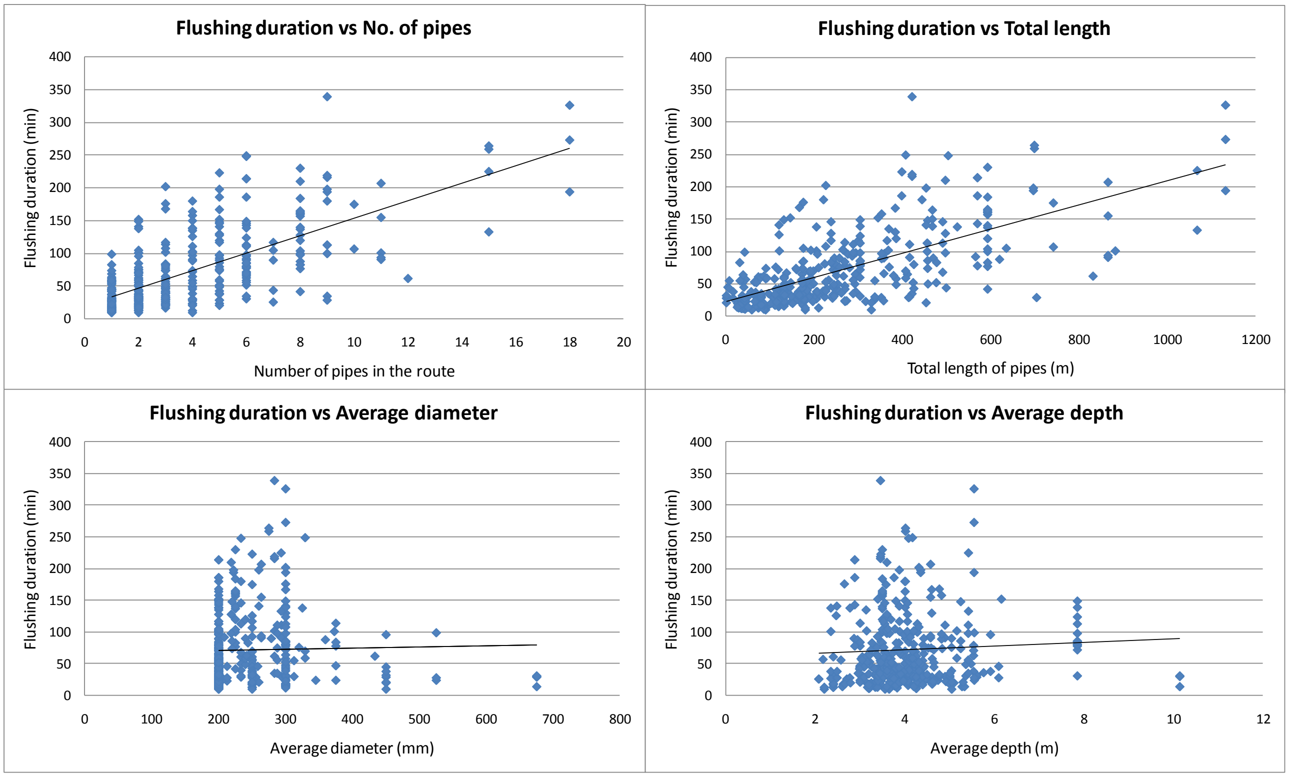
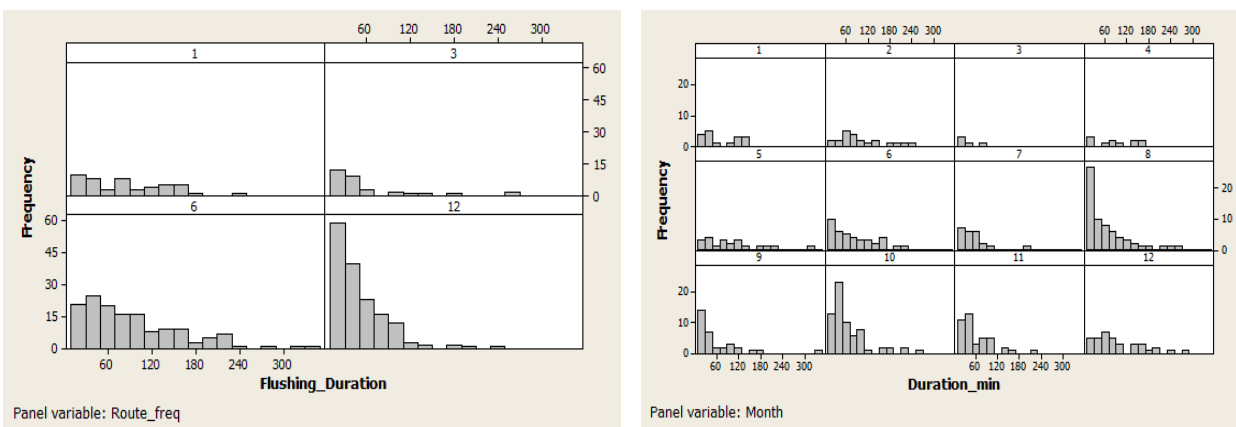


Fig. 4-9. Flushing duration vs number of pipes, total length, average diameter, and average depth

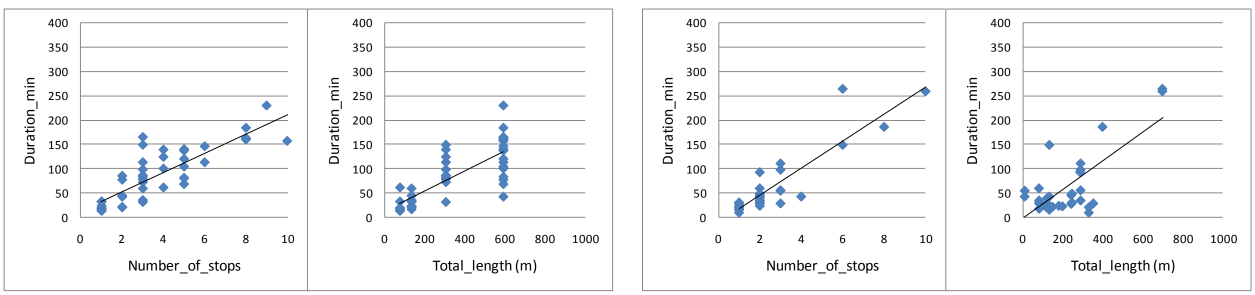


(a) (b)

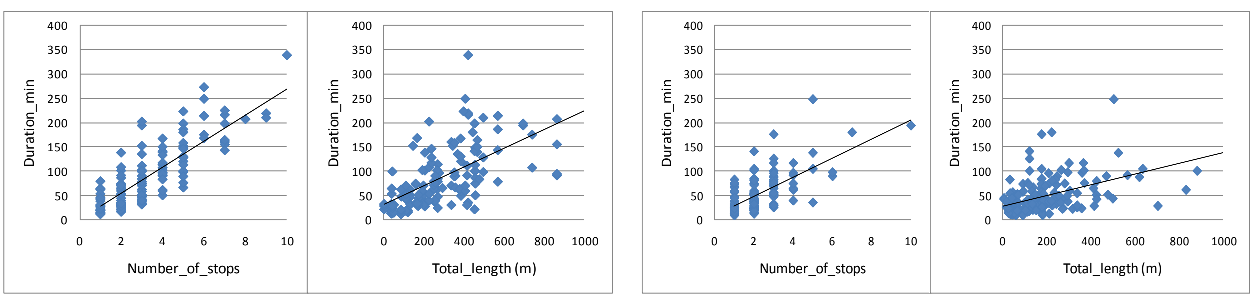
Fig. 4-10. Histogram of Flushing_duration by (a) frequency and (b) month

Fig. 4-11 presents the scatterplots for flushing duration versus the “number of stops” and “total length”, whereby strong linear relationships can be observed. Interestingly, the “number of stops” has a stronger correlation with flushing duration than does the “number of pipes”.

Theoretically, the crews are expected to stop at every manhole to access all of the pipes in that location; however, in practice, experienced crew members flush two stretches of pipes from the same manhole whenever possible, which allows them to complete their job with fewer stops. On the other hand, crews must occasionally make additional stops in order to check the map for manhole locations, for a short break during flushing of a large location, (e.g., a location of five or more pipes), or when work is interrupted due to an emergency task arising which requires the crew's attention, (i.e., the crew must leave a location unfinished and return later to complete the flushing task).



(a) 1-month-frequency locations (b) 3-month-frequency locations



(c) 6-month-frequency locations (d) 12-month-frequency locations

Fig. 4-11. Scatterplots showing Flushing_duration versus No_of_stops and Total_length

All of these situations can lead to variation in "number of stops" for a given location. Because the on-site duration data collected from GPS is based on the number of stops at a particular location, this factor is more significant than the number of pipes. Thus, it is important to introduce factors such as crew experience and location layout (whether each of the pipes in a location has an upstream channel that provides continuous access to the next pipe) in order to estimate the probability of “number of stops” for a particular location, and then to input that value into the duration estimation model. Another important consideration during GPS data collection is to avoid the bias caused by interdependent multiple observations. When work orders are issued, they are often grouped together by location so that the designated crew is able to flush multiple adjacent locations consecutively. Fig. 4-12 shows two adjacent 12-month-frequency locations that were flushed on the same day by the same crew, making it difficult to identify the number of stops (and, hence, the duration) for individual locations. This kind of bias should be avoided during data collection.

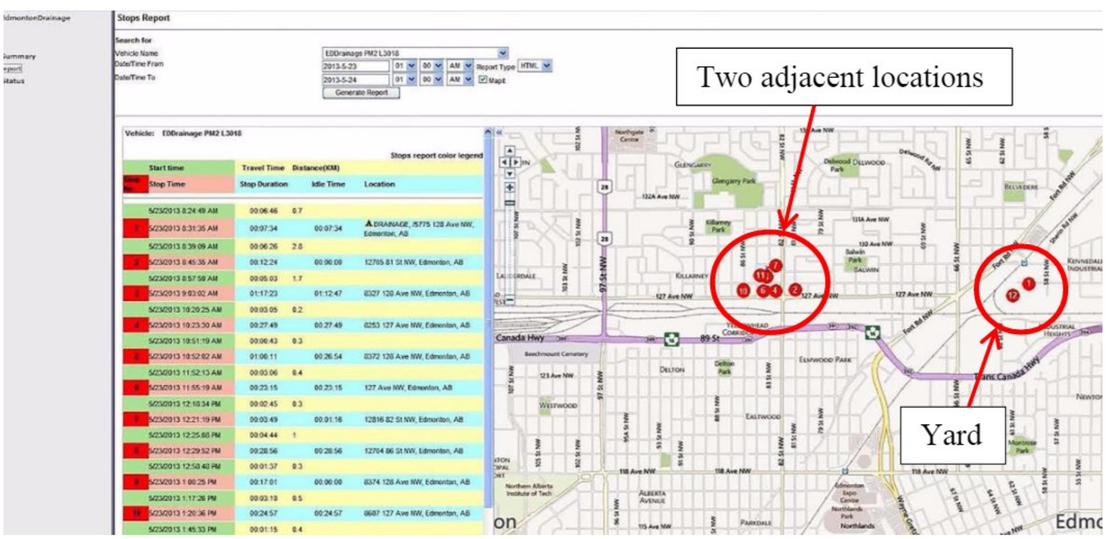


Fig. 4-12. GPS data collection interface at the City of Edmonton (Interfleet.com)

4.4 On-site Flushing Duration Estimation Model

From the descriptive statistics, it is evident that the on-site duration has a strong linear relationship with the predictor variables, and thus that a multiple regression model should be able to capture the majority of the variation. However, owing to the fact that location frequency has a considerable effect on the flushing duration in the estimation dataset, separate models for each frequency must first be developed. Moreover, the 1-month-frequency location subset of data contains panel observations, which must be modelled separately in order to explore the temporal variation of flushing duration.

The results for individual frequency models as well as a combined model are presented in Table 4-2. It can be seen from the table that the 1-month-frequency location model contains temporal variables with p-values < 0.05. Both 1-month and 3-month-frequency location models have R^2 values greater than 0.85, which implies that more than 85% of the variability of flushing duration is captured by the models. The 6-month and 12-month models have reasonable goodness-of-fit values (76% and 71.6%, respectively). It is also observed that the frequent locations have higher R^2 values, which underscores the fact that these locations are flushed on a regular basis, resulting in consistency in on-site productivity.

While separate models may be more efficient for estimating purposes, the planner may be interested in the application of a single model for all frequencies in the case where significantly higher annual man-hours are consumed by the locations having a certain frequency. At the City of Edmonton, the total lengths (of pipe) flushed under 1-, 3-, 6-, and 12-month frequencies are 12,180 m, 36,289 m, 99,083 m, and 280,272 m, respectively; this represents a situation which, in the interest of convenience, warrants the use of a single, frequency-independent model. A frequency-independent model is thus developed which has a reasonable goodness-of-fit value of 73.6%. The model contains seven statistically significant parameters with expected signs and values (see Table 4-2). For example, the "midday" coefficient (used as a dummy variable in the model) implies that the flushing takes approximately 20 minutes longer than usual when an operation takes place between 11:00 a.m. and 1:00 p.m. This captures the fact that crews usually take a short break for lunch within this timeframe. The model also captures the variability of all four location-frequencies by virtue of the "flushing_per_year" parameter. The negative coefficient for this factor supports the previously-mentioned finding that more frequent locations have shorter flushing times if all other factors remain constant.

One major limitation of this model is that it includes the "number_of_stops" as a predictor variable, which refers to the number of times a crew stops to flush a given set of pipes. This variable has a significant effect on the on-site duration of flushing activities, and it differs from the number of pipes depending on a number of factors, such as location of pipes, their upstream facility types, and crew judgment. Because this variable is unknown to the planner at the time of scheduling, an ordered probit model is developed in order to estimate the number of stops for a given location.

Table 4-2 Multiple linear regression 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
		Adjusted R^2 = 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^2 = 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^2 = 76.0%		
12-month Routes	Constant	11.624	1.38	0.178
	Number_of_stops	7.085	2.019	0.052
	Total_length	0.147	4.051	0.000
	Splits	22.290	2.774	0.009
	Regular_crew	15.160	2.396	0.022
		Adjusted R^2 = 71.6%		
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^2 = 73.6%		

4.4.1 Ordered Probit Model Results

As described in the methodology chapter, under ideal conditions the "number of stops" (N_s) for a given location should be equal to the "number of pipes" (N_p). However, analysis shows that the difference between the number of stops and number of pipes varies within as well as across locations, and follows a normal distribution pattern. For example, the number of stops for a 3-pipe location may range from 1 to 5 (resulting in $N_s - N_p$ varying between −2 and +2), while the same for a 7-pipe location may range from 4 to 11 (resulting in $N_s - N_p$ varying between −3 and +4). If this difference between N_s and N_p is assumed to be the error (ε), then, owing to the ordered nature of N_s and normally distributed ε, an ordered probit model can be used to estimate the number of stops for a given location (McKelvey & Zavoina 1975; Baik et al. 2006). Therefore, the probability of "number of stops" is first estimated using an ordered probit model, and it is then used in the primary multiple regression model to estimate the on-site flushing duration (as illustrated in Fig. 4-13).

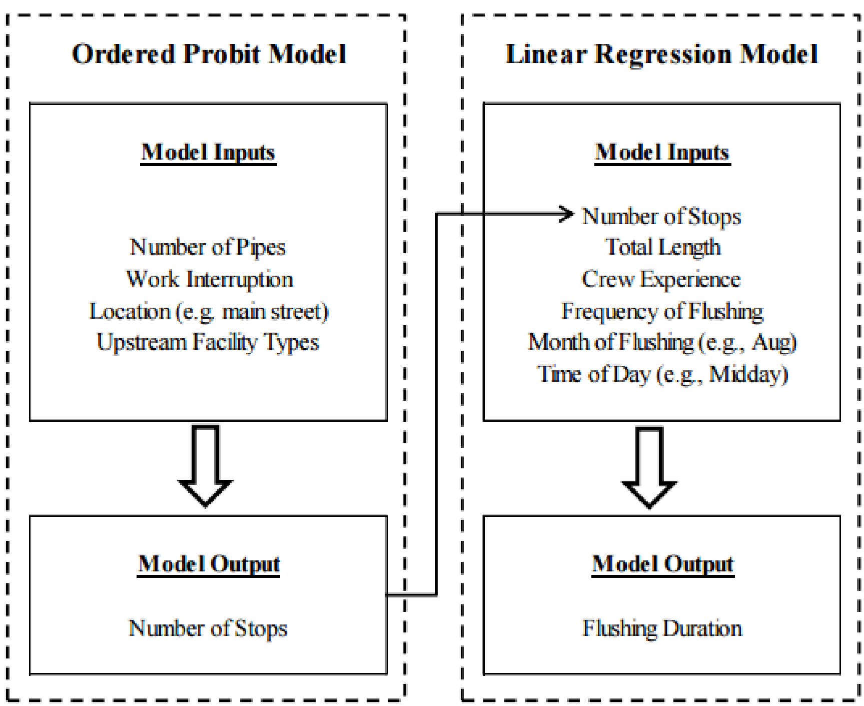


Fig. 4-13. On-site flushing duration estimation modelling schematic

The results of the model are presented in Table 4-3, from which it can be observed that the number of stops for a specific location depends on the number of pipes, work interruptions (described above), location type (internal road or main street), and layout type (described above). The Pearson Chi-square goodness-of-fit test is significant and the Pseudo R^2 values are acceptable (see Greene & Hensher 2008); however, it is found during the trials that the model does not yield satisfactory results when the number of stops differs considerably from the number of pipes. Discussion with the operators reveals that, in theory, the "number of stops" for a specific location should be equal to the "number of pipes". Although it varies in practice, however, the difference $(Ns - Np)$ in most cases should not exceed the range of ±4. For example, if a certain location has 7 pipes, flushing may be completed in a minimum of 3 stops (by flushing more than one consecutive stretch of pipes from the same manhole, if the pipe layout allows), or it may be completed with a maximum of 11 stops (considering 7 stops for accessing 7 pipes, plus additional stops for map checking, initial inspection, discussion, and work preparation). Therefore, any observations with wider variation are considered non-standard, and are excluded from further analysis. This potentially leads to standardization of the operating procedure, which, if implemented, could reduce the likelihood of any future observations similar to those already excluded from the analysis. Hence, this study uses smaller subsets of data for model estimation (249 observations) and validation (45 observations), where the absolute value $(Ns - Np)$ is limited to 4.

Table 4-3 Ordered probit model results

Parameter	Estimate	Significance
[stops = 1]	−3.536	0.000
[stops = 2]	−2.522	0.000
[stops = 3]	−1.361	0.014
[stops = 4]	−0.592	0.280
[stops = 5]	0.708	0.185
[stops = 6]	1.398	0.009
[stops = 7]	2.097	0.000
[stops = 8]	2.785	0.000
[stops = 9]	3.896	0.000
[pipes = 1]	−2.932	0.000
[pipes = 2]	−2.780	0.000
[pipes = 3]	−1.921	0.000
[pipes = 4]	−1.325	0.002
[pipes = 5]	−1.007	0.012
[pipes = 6]	−0.221	0.622
[pipes = 8]	0.867	0.035
[pipes = 9]	0[a]	-
[Work_interruption = 0]	−1.889	0.000
[Work_interruption = 1]	0[a]	-
[Main_street = 0]	0.780	0.006
[Main_street = 1]	0[a]	-
[U/S_other_than_manhole = 0]	0.680	0.002
[U/S_other_than_manhole = 1]	0[a]	-

Notes:

[a] Parameters set to zero

-2 Log Likelihood: Intercept only model = 528.869; Final model = 205.435

Goodness-of-Fit: Pearson Chi-Square = 105.066; Significance = 1.0

Pseudo R^2: Cox and Snell = 0.706; McFadden = 0.321

4.4.2 Multiple Regression Model Results

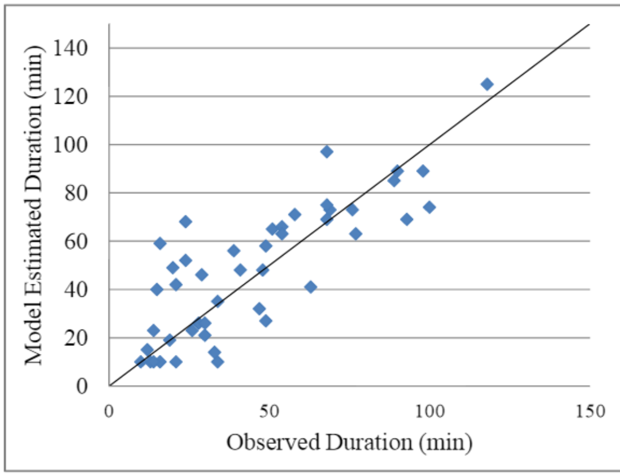
Once the number of stops for a particular location is forecasted by taking the stop number corresponding to the highest probability, this value can be used as a predictor variable in the multiple regression model described above. The results of this model are presented in Table 4-4, which shows that the model has a reasonable adjusted R^2 value of 0.703. The estimated number of stops possesses the highest *t*-stat value among the predictors, suggesting the efficacy of the probit model. An interesting observation from the model results is that "regular crew" (denoting crews that regularly perform HPF), is included as a predictor, and has a positive coefficient. This counter-intuitively implies that a regular, experienced crew would take longer to flush a location. A possible cause may be the difference in perception between regular and non-regular crews. The nature of the work is such that the crews use their judgment to perceive the cleanliness of the pipe during flushing in order to determine when to stop, and perhaps the regular crew members flush the pipes more meticulously. The result also suggests that there is seasonal variation in on-site productivity of HPF, where the winter months are found to correspond to an increase in flushing duration. Further investigation of this finding reveals that in some cases the manholes are covered with ice and snow during the winter, which necessitates additional time for clearing the snow and ice in order to open them. Another possible cause is a data collection bias: on a cold day, the truck engines are usually left running even during breaks in order to ensure that the water used for flushing does not freeze; however, this break time cannot be identified and separated from actual flushing time data.

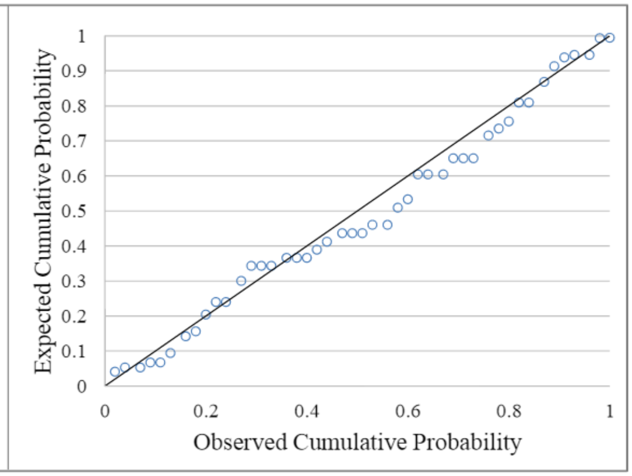
Table 4-4 Multiple regression using estimated number of stops

Parameter	**Coefficient**	***t*-stat**	**Significance**
Constant	−3.761	−0.764	0.446
Tot_length	0.117	5.132	0.000
Estimated_number_of_stop	13.849	6.794	0.000
Regular_crew	10.514	2.568	0.011
Midday	25.814	6.172	0.000
Flush_per_year	−0.860	−2.004	0.046
Aug	−27.427	−2.819	0.005
Feb	10.288	1.585	0.114

Note: Model Goodness-of-Fit: R^2 = 0.711; Adjusted R^2 = 0.703

The model is validated using the validation dataset. It can be observed from the resulting plot (Fig. 4-14a) that the estimated versus observed points are moderately close to the 45º reference line, although some individual points deviate from this line. It is understood that the model may not be appropriate for micro-level estimation, i.e., the use of this model to estimate the on-site productivity for an individual location may lead to errors. However, the model can conveniently be used for weekly or monthly schedules. Here, it is of interest to check whether or not the model errors follow the assumed normal distribution. To do so, the probability density function (PDF) of the model errors (estimated minus observed) from each observation is determined, based upon which it is found to have a slightly skewed normal distribution pattern (μ = 2.64, and σ = 16.46), which satisfies the regression assumption. The P-P plot for errors is shown in Fig. 4-14b.





(a) (b)

Fig. 4-14. (a): Estimated vs observed flushing duration; (b): Normal P-P plot of error

4.4.3 Duration Estimation Model using SAP Enterprise Mobile Data

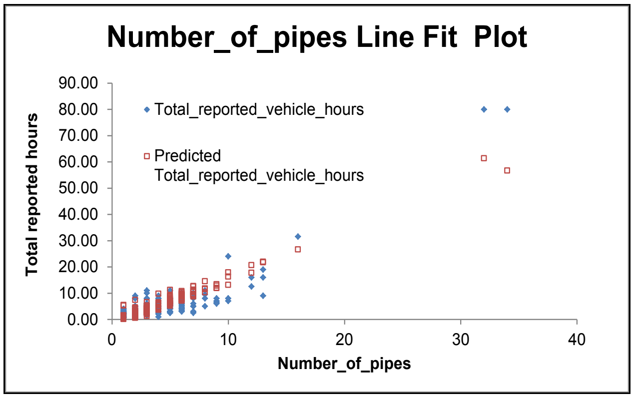
As mentioned before, the accuracy of the estimation model largely depends on the quality of data. The models described in the preceding section were developed based on the data available at the time. However, in 2014, drainage operations group at the City of Edmonton implemented mobility in SAP ERP environment (SAP 2015), where the field crews receive work orders (one for each scheduled location) electronically, and have the ability to charge time to individual orders. Thus, the flushing duration charged against the orders can be directly linked with inventory databased (Drains properties data) through unique pipe IDs. This eliminates the need for extraction of AVL data and model the number of stops for each location.

Therefore, multiple regression model has been developed to estimate flushing duration using data from SAP mobility system, and the results are presented in Table 4-5. It is found that the model goodness of fit is improved significantly, and the model also captures important attributes such as material and diameter of pipes as predictor variables. Fig. 4-15 presents the line plots for flushing duration vs number and length of pipes for each location. Linear relation between the variables confirms the observation found the preceding section of this research.

Table 4-5 Flushing duration estimation model using SAP Mobile data

Regression Statistics	
Multiple R	0.887
R Square	0.786
Adjusted R Square	0.777
Standard Error	3.377
Observations	279

	Coefficients	Standard Error	t Stat	P-value
Intercept	1.304	1.717	0.759	0.448
Work_split	1.151	0.560	2.057	0.041
Number_of_pipes	1.285	0.166	7.721	0.000
Total_length_of_pipes	0.005	0.002	2.125	0.034
Avg_dia_of_pipes	0.012	0.004	2.938	0.004
PVC	-5.576	1.546	-3.607	0.000



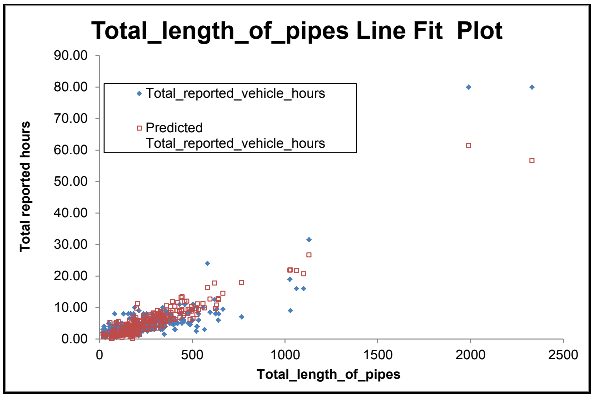


Fig. 4-15 Line plots showing relations between duration and number-of-pipes and pipe-length

It is to be noted that the mobility data is based on hours reported by field crews and may be verified using AVL data if necessary. Also the charged times include morning start-up time, end-of-shift cleaning time, travel and break times, which are distributed to all the locations flushed in a day. Therefore, the model does not represent the actual on-site duration for each location; however, the model can be effectively used to estimate daily productivity. Moreover, electronic work order system helps in better management of accounting data charged against different asset components, which is valuable in terms of asset management perspective. This also eliminates the hard copy paper orders, and the duplication in data entry for work order assignments and timesheet recording.

4.5 Schedule Optimization

The optimization models are applied to a case study from the City of Edmonton, which involves optimizing a monthly HPF schedule to improve productivity. For this purpose, the optimization models are developed in a .NET framework, and are connected with the necessary databases. The case study comprises 179 scheduled HPF locations, represented by the red dots in Fig. 4-16. The maintenance locations are well distributed throughout the city, while the yard is located in the north-east area of the city (represented by the blue circle in the figure). Select pertinent statistics and assumptions of the case study are outlined below:

Number of locations = 179
Total shift length = 8 hours
Preparation, cleaning, break time = 2 hours
Effective shift length = 6 hours (used for optimization)
Source of travel data: Edmonton regional travel model
On-site flushing duration data: Expected value derived from estimation model

	Minimum	Maximum	Average	Sum of all locations
Number of pipes at each location	1	13	3.73	669
Length of pipes at each location (m)	13.35	1,100.55	245.75	43,989.00
Expected value of on-site flushing duration at each location (minutes)	20	204	73	13,117

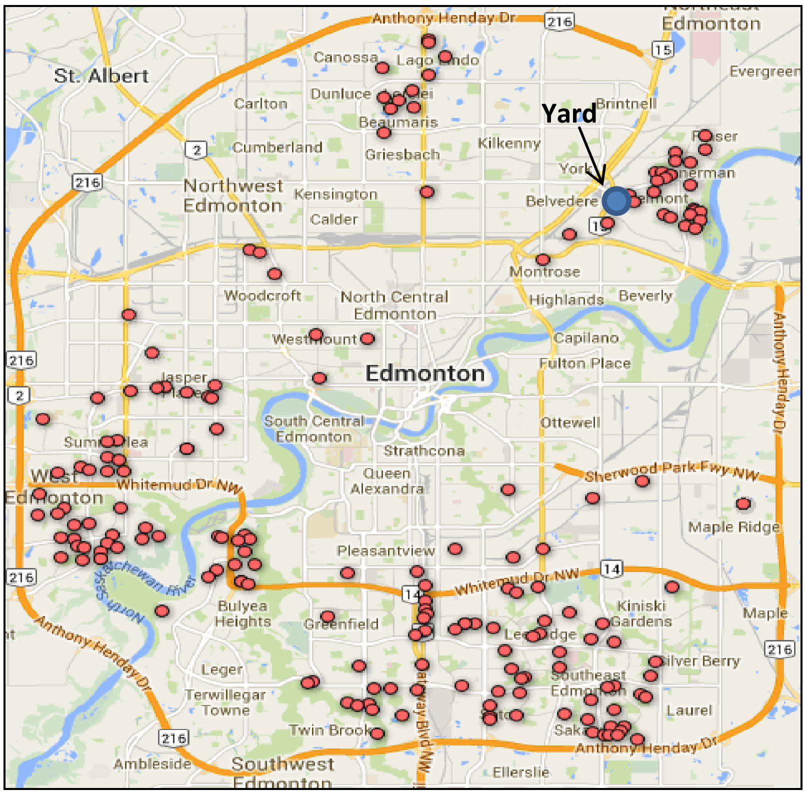


Fig. 4-16. Monthly scheduled HPF locations for the case study

4.5.1 Neighbourhood-based Routing

At first, expected daily productivities for the given schedule are estimated using neighbourhood-based routing, which is the current practice at the City of Edmonton. For this purpose, expected on-site duration of each location (derived from the on-site duration model described in the previous section) and travel time matrix obtained from Edmonton's regional travel model are used. It is assumed that these expected durations and travel times are known to the planner/scheduler as well as to the flushing crews prior to implementation. The scheduled locations are grouped by neighbourhood, and the routing follows the sequence one by one. When all locations in a neighbourhood are completed, the crews move to the next (nearest) neighbourhood, and so on. However, it may not be possible to complete all the locations in some neighbourhoods within the 6-hour effective shift length, while other neighbourhoods may not have a sufficient amount of work to fill an entire shift. Therefore, the monthly schedule is split into multiple closed tours, each representing a daily route that is expected to be completed within a shift.

Daily expected productivities for each of these overtime and location-splitting options are determined by running a simple discrete-event simulation (DES) model, where an entity (crew) travels to and flushes each of the locations one by one. Both the elapsed and remaining time are recorded after each event—leaving from the yard, arriving at location 1, completing flushing for location 1, arriving at location 2, and so on. The crew covers as many location as possible while remaining within the effective shift length and the given conditions (i.e., the options mentioned above), and then returns to the yard. Daily travel time, end-of-shift waste time, overtime, and daily accomplishments (total length of pipes flushed) are recorded at the end of each day. If the total shift length is 8 hours and one crew comprises 2 persons, the daily total man-hours is 16. Thus, the unused times do not directly affect the productivity calculation. However, overtimes (in terms of man-hours) are two times costlier than regular time, and therefore are doubled and added to the man-hours. And, finally, the productivity for each day is calculated by dividing the accomplishments by the total incurred man-hours. For example, if 800 m of pipe is flushed in an

8-hour shift, the productivity would normally be calculated as $\frac{800}{8\times2}$ = 50 m/man-hour. However, in the case of 20 minutes of overtime for the same job, the productivity becomes $\frac{800}{(8+0.33\times1.5)\times2}$ = 47.05 m/man-hour.

Results of neighbourhood-based routing using the above-mentioned splitting options are presented in Table 4-6. Interestingly, it is found that the average expected daily productivity is approximately 50 m/man-hour for all three options, each case exceeding the current productivity target. This is due to the assumption that the expected on-site durations for all locations are known to the crew. It is evident that availability of this information will enable the crews to manage their times more efficiently, and thus the productivity will be increased; (however, the model was not yet available when the benchmarks were developed). This finding strongly supports the importance of having a reliable duration estimation model for productivity improvement.

Table 4-6 Results from neighbourhood-based routing

	Option 1	Option 2	Option 3
Overtime Allowed?	No	Yes	No
Location Splitting Allowed?	No	No	Yes
Total Expected On-site Duration (minutes)	13,117	13,117	13,631
Total Crew-days	55	41	53
Total Travel Time (minutes)	4,173	3,714	4,406
Total Expected Unused/Overtime (minutes)	2,727	(2,071) OT	1,259
Min. Expected Daily Productivity	24.31	37.44	32.14
Max. Expected Daily Productivity	79.01	86.15	84.91
Avg. Expected Daily Productivity	49.42	55.56	50.62
Standard Deviation	11.65	10.99	10.92

Comparing the results from different overtime and location-splitting options, it can be observed that option 2 (i.e., overtime allowed when necessary, but location splitting not allowed) produces the highest average expected daily productivity (55.56 m/man-hour). This option also allows the completion of all the scheduled locations in fewer crew-days, (where "crew days" refers to the number of days required for one crew to finish the given schedule). For instance, if four crews are available, this work can be finished in 10 working days (plus one additional day for one crew). However, as expected, this option results in a great amount of overtime. On the other hand, allowing location splitting (option 3) provides the crews the flexibility to leave an incomplete location at the end of the shift, which significantly reduces the end-of-shift waste time. Furthermore, returning to the same location the following day not only leads to more travel but also increases the total on-site duration due to activities such as parking, securing the job-site, and opening of manholes. This is evident in the increase in total on-site duration for option 3, which is approximately 4% more than the time necessary to flush the same locations without interruption.

4.5.2 Schedule Optimization by Heuristic Algorithm

The given schedule is now optimized by the proposed heuristic algorithm, with the results presented in Table 4-7. Tour splitting has been performed by allowing no overtime or location splitting. It is found that the expected average productivity is improved, while the total travel and unused times are reduced by significant margins. As explained above, because the case study contains 179 locations of various sizes, the heuristic algorithm has a number of options to fill the bins, which improves the tour splitting. This is clearly observed when the total expected unused time is compared with those from the neighbourhood-based routing. Fig. 4-17 presents the distribution of expected daily productivities for the three options using neighbourhood-based routing. It is found that options 1 and 3 produce similar results in terms of productivity distribution, while option 2 shows slightly higher productivity. When the heuristic algorithm is applied for option 1, it results in marked improvement from neighbourhood-based routing for the same option (Fig. 4-18). The 95% confidence intervals of daily productivities for the two routing algorithms are calculated using Eq. (4.1), and the results are as follows:

$$\bar{X} \pm t_{\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}} \tag{4.1}$$

- Neighbourhood-based routing: 49.2 ± 3.15 = 46.3 ~ 52.6 metres/man-hour
- Heuristic algorithm: 58.1 ± 3.22 = 54.9 ~ 61.3 metres/man-hour

Table 4-7 Results from optimized schedule using heuristic algorithm

Overtime Allowed?	No
Location Splitting Allowed?	No
Total Expected On-site Duration (minutes)	13,117
Total Crew-days	47
Total Travel Time (minutes)	3,451
Total Expected Unused/Overtime (minutes)	597
Min. Expected Daily Productivity	31.24
Max. Expected Daily Productivity	79.28
Avg. Expected Daily Productivity	58.10
Standard Deviation	10.99

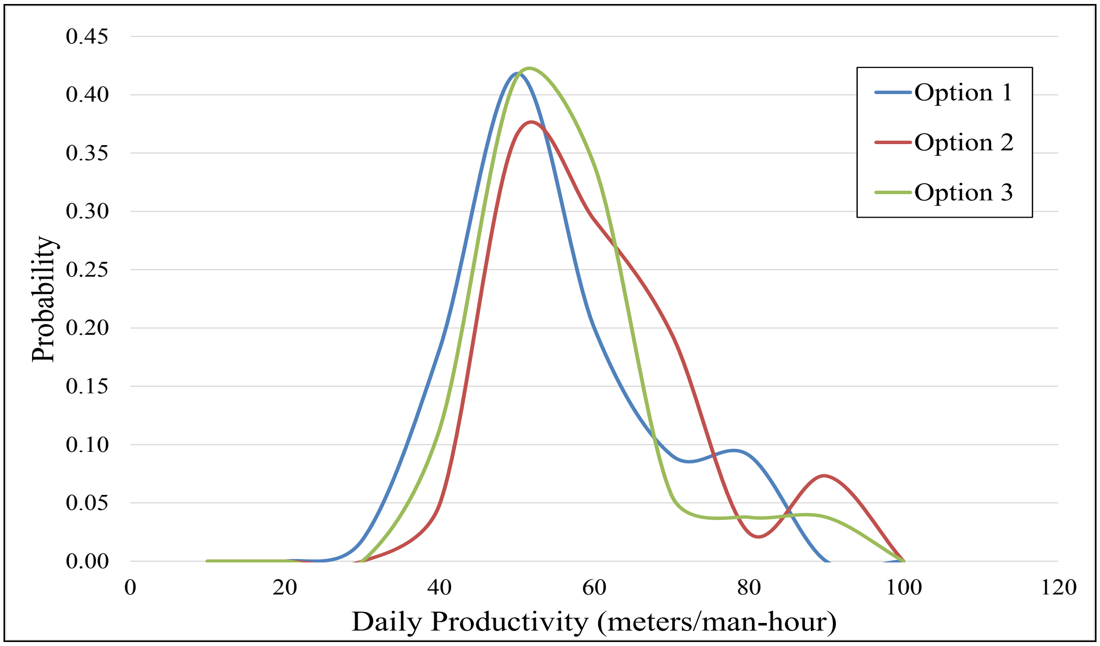


Fig. 4-17. Distribution of expected daily productivities for neighbourhood-based routing

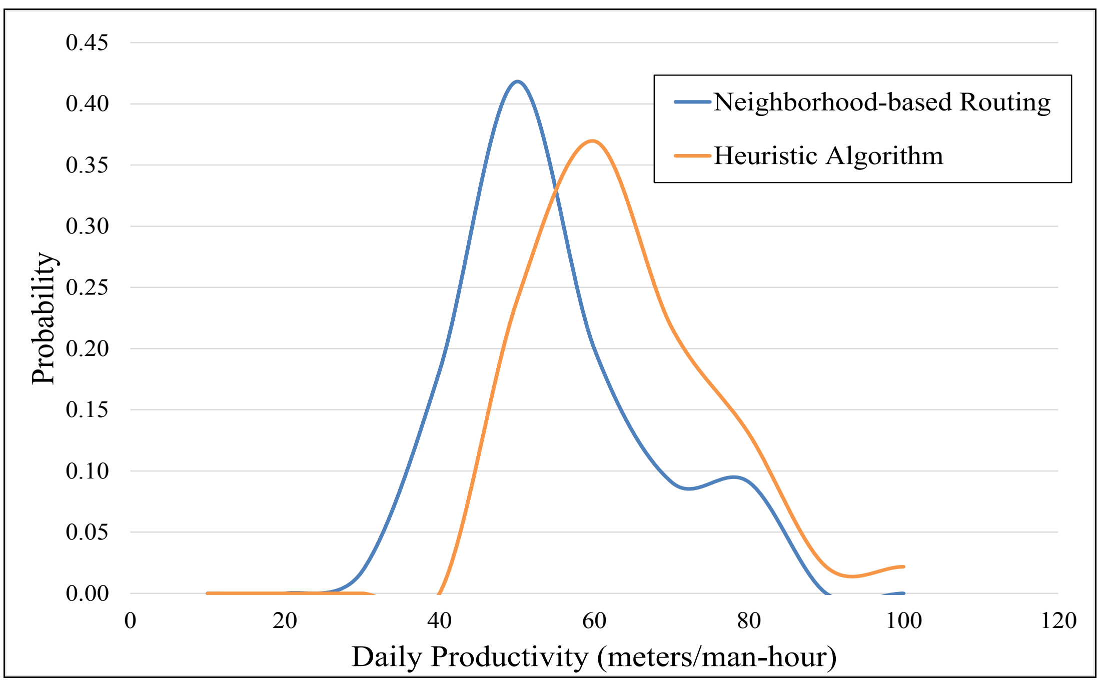


Fig. 4-18. Distribution of expected daily productivity for neighbourhood-based routing and heuristic algorithm (splitting option 1)

4.6 Simulation

The results presented in the previous section are based on expected on-site durations, as well as on the assumption that is no deviation from the optimized schedule with the progress of each day. However, deviation from the planned schedule is very likely to occur due to stochasticity in the on-site flushing duration as well as in travel time. It is thus necessary to simulate the actual work flow to test the impact of schedule deviation and the robustness of the optimization algorithm. For this purpose, the on-site flushing duration at each location is simulated using the probit model (described in section 4.4). Simulation is then performed by following the given optimized sequence of locations, and the end-of-shift waste times are recorded for each day. Fig. 4-19 presents the results from 100 simulation runs, where positive values indicate unused time at the end of the shift, and negative values represent overtime.

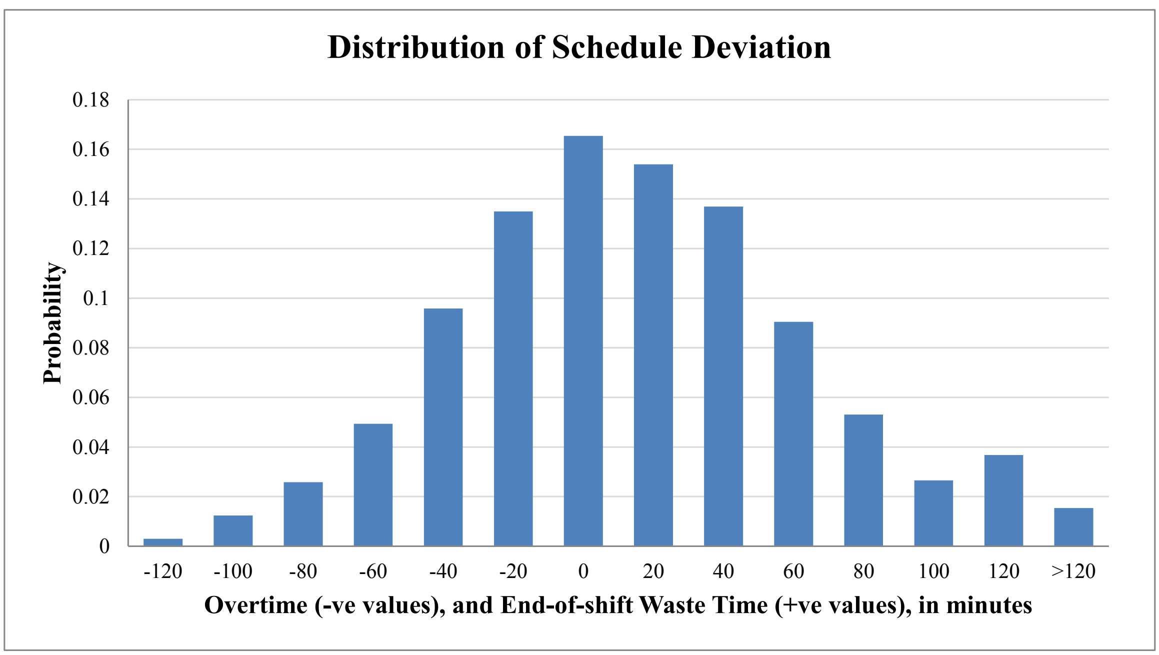


Fig. 4-19. Distribution of schedule deviation due to variation in on-site duration

As expected, the distribution of schedule deviation is normally distributed and is slightly skewed toward positive values (representing unused times at the ends of shifts). However, most of the observations show 0 to 20 minutes of unused time. In any case, it is of interest to explore the impact of such variations on daily productivity. The productivity of each simulated day is illustrated in Fig. 4-20. Quite interestingly, the schedule deviations result in only a negligible impact on daily productivity. This is due to the fact that the effect of a small overtime value is diminished during productivity calculation. Nevertheless, it is not desirable to allow the daily work time to extend to overtime, so further improvement can be made by introducing schedule overrun penalties during optimization.

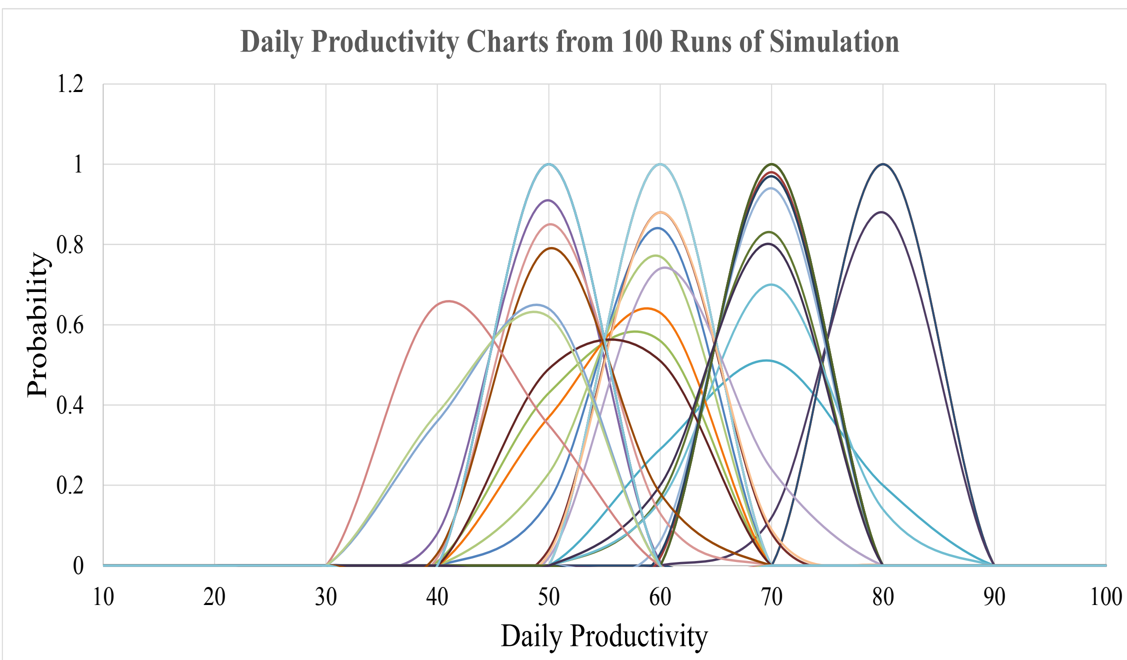
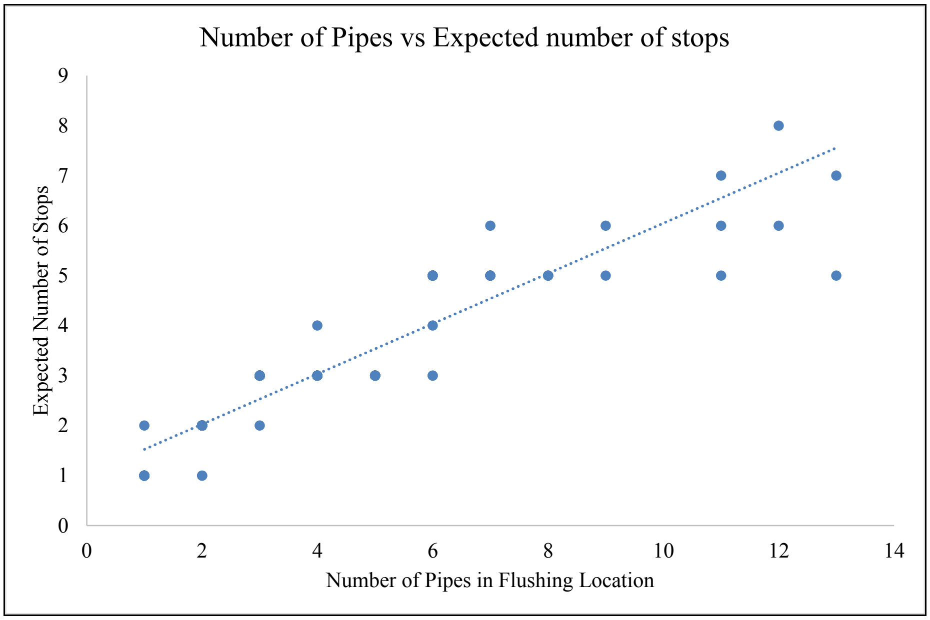


Fig. 4-20. Distribution of simulated daily productivities

4.6.1 Standardization of Flushing Activity

Another way to reduce the likelihood of schedule deviation is to standardize the flushing process. For this particular case study, it is evident that the stochasticity in the on-site duration comes from the probit mode, which estimates the number of stops for the given locations. The probit model provides probabilities of each possible stop option for each location, from which the options with the highest probabilities are chosen to represent the expected number of stops.

However, in reality the number of stops varies significantly, which is observed from the historical data (Fig. 4-21). When the probit model is simulated, it generates probabilities for all possible values for the number of stops, and hence the large variations in on-site duration are created. If the process is standardized, the field crews will be advised to reduce the variability in their number of stops. For instance, the allowable stops for a 5-pipe location may range from 2 to 5, depending on various factors, whereas historical data confirms between 2 and 8 observed stops. It is evident that creating boundaries (marked as green and red lines in Fig. 4-22) for the number of stops can significantly reduce the variability.



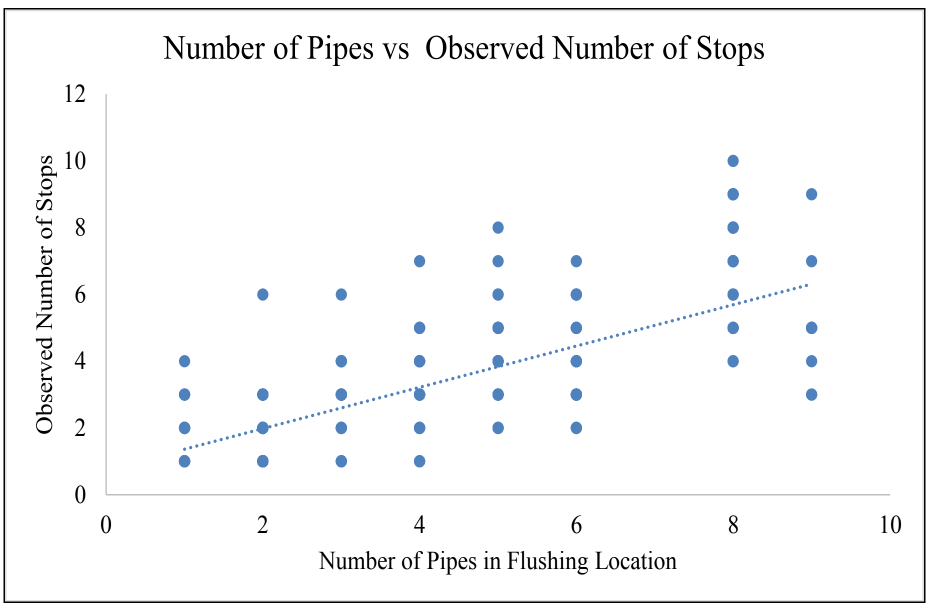


Fig. 4-21. Number of pipes versus expected (top) and observed (bottom) number of stops

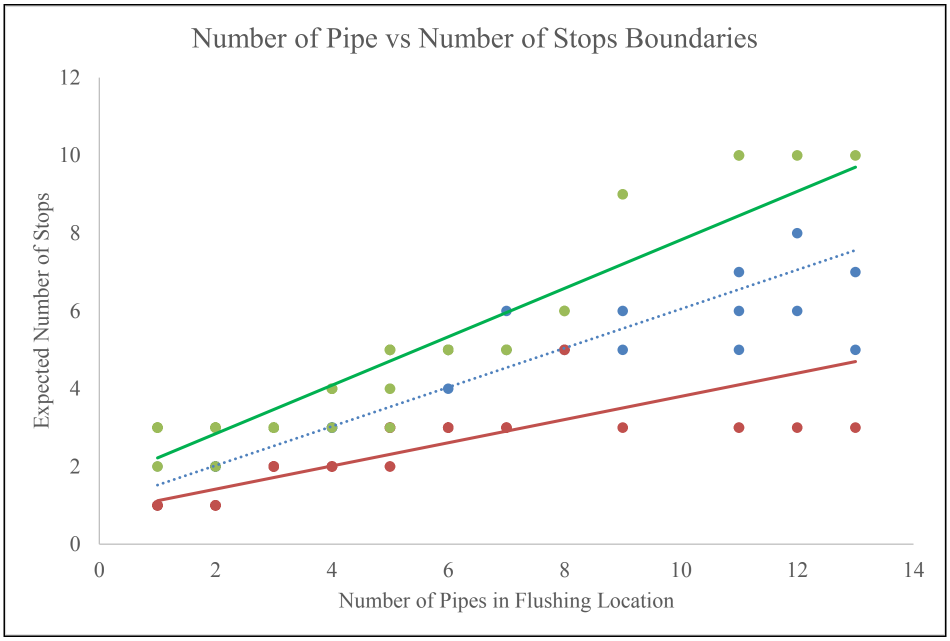


Fig. 4-22. Number of stops possibilities after standardizing process

To test the effect of process standardization, the on-site duration estimation model is simulated again, this time using boundary conditions for possible number of stops. The results from 100 runs of simulation are presented in Fig. 4-23, which shows a marked reduction in schedule deviation. The standardized process reduces variability in on-site duration and thus reduces the likelihood of end-of-shift unused time or overtime. As can be seen from the distribution, possibilities of large deviations are eliminated. Now, it is of interest to see how this affects the daily productivity distribution. Productivity simulation reveals that some wide variations in productivity distributions are reduced by standardizing the process (Fig. 4-24).

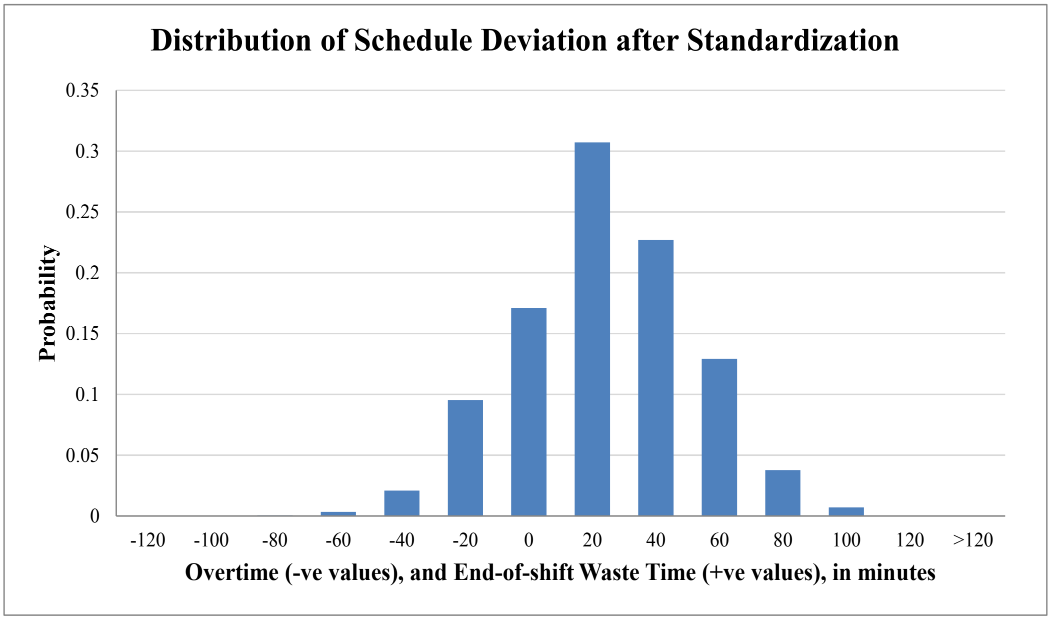


Fig. 4-23. Distribution of schedule deviation after process standardization

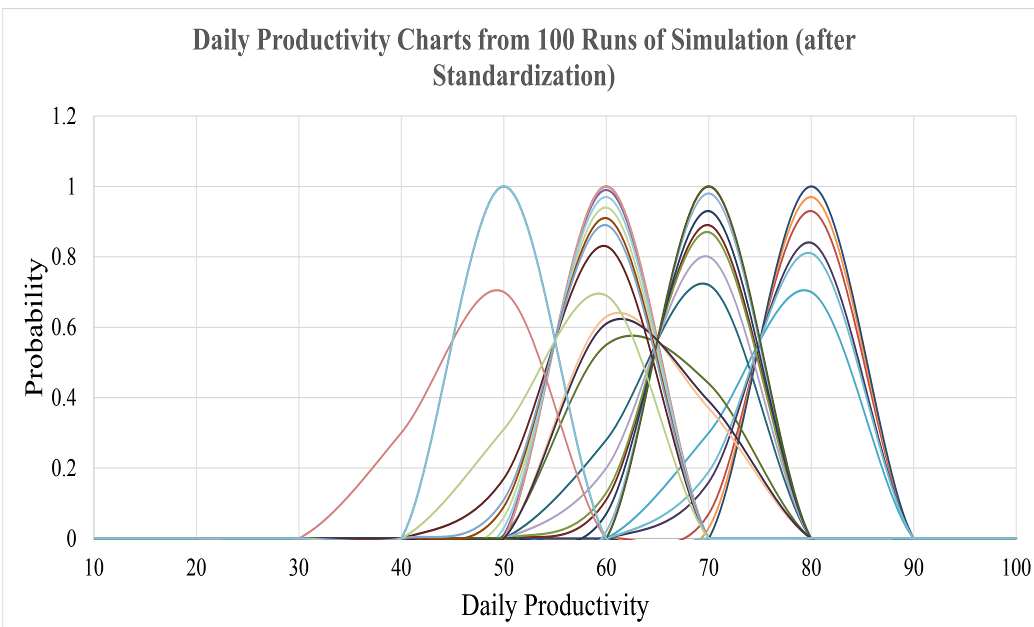


Fig. 4-24. Distribution of simulated daily productivities after process standardization

4.6.2 Implementation Strategies

From the case study results, it is understood that there exist many different ways of optimizing the routing, splitting the tours, or standardizing the process, and all of these contribute to improved productivity. However, implementation strategies for the above-mentioned models or options are subject to a specific context and existing policies. A review of industry practice and of recent advancements in technology and software solutions reveals that integrated data and O&M management systems are currently available and have been implemented by many operators at various capacities. However, the implementation approach depends on the accuracy of the estimation model and the efficiency of the optimization model. For example, if the estimation of on-site flushing duration is fairly accurate and the daily schedule optimization can be performed within a reasonable amount of time, it is expected that the crews may only deviate slightly from the optimum sequence. In such a case, real-time dynamic sequence updating may not be required. On the contrary, a less reliable (or conservative) estimation model may lead to time deviation, which warrants the real-time updating of optimum sequence.

This raises the issue of how to develop a reliable estimation model. The model formulation described in the methodology chapter is based on established statistical techniques. However, the accuracy of the forecast is largely dependent on the availability of a large quantity of high quality historical data. Therefore, an automated data collection process is crucial to improving efficiency. As mentioned above, the flushing vehicles in many cities are now equipped with automatic vehicle locator (AVL) devices. Connected the sensors with flushing and pumping motors can eliminate manual work in the data collection process and thereby improve the quality of data. The data collected from the flushing and pumping motors can also be used in the development of water usage estimation and sediment deposition models respectively, which can provide valuable information for developing operational schedules.

The selection of the appropriate algorithm for schedule optimization requires a trade-off between optimality and run-time. While ILP may not be a feasible option for municipalities due to longer run-time, the hybrid GA presented in the methodology can be reasonably effective. However, the case study results show that the quick heuristic also will improve the daily productivity by a significant margin. The advantage of the heuristic algorithm is its fast run-time, which allows the planner to re-optimize the schedule even during working hours if large schedule deviation is experienced by a crew. However, the real-time updating of a PM schedule during flushing operation can be challenging, especially when multiple trucks are simultaneously operating at different locations across the city. For such cases, all the vehicles should have access to a central server that has the capacity to run an optimization algorithm every time a crew completes the flushing of their current location. Although it is possible to do so using existing technology, the concern is that generating an updated sequence for one vehicle may affect the current sequence of the other vehicles.

4.7 Summary of Case Study

This chapter has presented a case study from the drainage operations group at the City of Edmonton, where a monthly HPF schedule has been optimized. The expected on-site durations for the flushing locations have been estimated using an ordered probit model combined with multiple regression. Simulation reveals that the proposed methods effectively improve daily productivity compared to the existing neighbourhood-based routing. However, the variations produced by on-site duration may result in schedule deviations. Nonetheless, accurate data collection and forecasting of on-site duration and standardization of the process can reduce such variability.

5 CONCLUSION

This research has developed a framework for improving the productivity of drainage operations, with a primary focus on preventive operational activities such as high-pressure flushing (HPF). These activities are carried out at regular intervals at various prescheduled locations across the given jurisdiction. For large municipalities, travelling between these locations results in a large amount of non-value added travel time. Moreover, the operational activities are typically of short duration, depending on several factors, and are stochastic in nature. This may lead to unused time or overtime at the end of work shifts. Therefore, this research proposes to optimize operational schedules by taking these factors into consideration. The framework includes two primary components: (*i*) developing a statistical model for the estimation of on-site duration, and (*ii*) developing suitable optimization algorithms to minimize travel time and end-of-shift unused/overtime.

5.1 Conclusion

The methodology presented in this thesis has been applied to a case study taken from the Drainage Operations group at the City of Edmonton. On-site duration data for HPF activity have been collected from historical automatic vehicle locator (AVL) records, and have been used to develop the on-site duration estimation model. For optimization, this study has formally described and formulated the Drainage Operations Scheduling Problem (DOSP) as a special case of the stochastic vehicle routing problem. Along with applying several established optimization algorithms (such as integer linear programming and genetic algorithm), this study develops a greedy heuristic and a hybrid genetic algorithm which explicitly serve the needs. As a general rule of thumb, the daily schedule can be optimized by prioritizing the locations with larger work/travel ratios at the beginning of the shift and leaving the smaller ones to fill the bins toward the end. The algorithms have been tested by optimizing a monthly HPF schedule, and their performances have been verified by simulation.

Results from the optimization models show that the proposed heuristic algorithm perform reasonably well within a very quick runtime. The hybrid GA uses the optimum schedule obtained from the heuristic and can further improve the results; however, it requires longer runtime. Although the expected daily productivity can be improved significantly through schedule optimization, results from simulation suggest that the stochasticity in on-site duration can still cause deviation from the planned schedule and lead to unused time or overtime. The likelihood of such deviation can be reduced by increasing the accuracy of the on-site duration estimation model and/or by standardizing the on-site process.

5.2 Research Contributions

The framework described in this thesis can be used for other drainage operations activities, such as visual inspection or low-pressure flushing (LPF). However, additional criteria for specific processes may be required to be incorporated in the models. The framework is also applicable for other jurisdictions; however, the on-site duration estimation model is not readily transferrable, as the tools & techniques and factors affecting the duration may vary from one jurisdiction to another. In addition, the framework presented in this thesis can be used for O&M of other infrastructure assets, provided there are scheduled short-duration activities at different locations across a jurisdiction.

This research makes the following academic contributions:

1) describes and formulates the drainage operations scheduling problem as a combinatorial optimization problem, thereby minimizing travel time and unused/overtime—or, in other words, maximizing effective work time;
2) develops heuristic and hybrid GA algorithms in order to quickly obtain near-optimal solutions; and
3) applies ordered probit analysis, combined with multiple regression model, in order to forecast the on-site flushing duration for a given set of pipes.

In addition, this research makes the following contributions to industry practice:

1) reduces travel and unused/overtime, resulting in improved productivity of operational activities;
2) reduces travel distance, resulting in savings in fuel and carbon emissions; and
3) provides guidelines to improve data collection and integration.

5.3 Limitations of the Study

This research has the following limitations:

1) The on-site duration estimation model does not capture the variations due to some expected predictor variables (such as slope, diameter, and operational conditions) due to the lack of sufficient amount of quality data.
2) The proposed heuristic and hybrid GA do not guarantee an optimal solution.

3) The runtime comparison between the algorithms is only relative, and does not report the actual runtimes. The runtimes may be improved by more efficient coding and data connection.
4) The optimization algorithms have been verified by simulation only and not through field tests.
5) This study focuses only on the preventive (or proactive) operational schedule. Interruptions in scheduled work due to emergency reactive work (if any) are not considered.

5.4 Recommendations for Future Work

Based on the findings and limitations of this research, the recommendations for future work are as follows:

1) There is a need to collect high quality on-site duration data by connecting sensors with flushing and pumping motors. This will permit the capture not only of the actual on-site duration for the flushing of a set of pipes, but also the amount of water usage or debris collected during flushing. This information can be effectively used for future operational planning.
2) There is a great potential to further improve productivity by applying or developing other existing or improved algorithms.
3) Further productivity improvement can be possible if two different activities sharing the same resources (i.e., a combination unit is used for both catch basin cleaning and high-pressure flushing activities) are combined in one schedule.

BIBLIOGRAPHY

Abraham, D. M. (2003). "Life Cycle Cost Integration for the Rehabilitation of Wastewater Infrastructure." *Proceedings, Construction Research Congress*, Honolulu, HI, USA, Mar. 19-21.

Achim, D., Ghotb, F., McManus, K. (2007). "Prediction of Water Pipe Asset Life using Neural Networks." *Journal of Infrastructure Systems*, 13(1), pp. 26-30.

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

Apaydin, O., Gonullu, M. (2007). "Route Optimization for Solid Waste Collection: Trabzone (Turkey) Case Study." *Global NEST Journal*, 9(1), pp. 6-11.

Apaydin, O., Gonullu, M. (2007). "Emission Control with Route Optimization in Solid Waste Collection Process: A Case Study." *Sadhana*, 33(2), pp. 71-82.

Applegate, D., Bixby, R., Chvatal, V., Cook, W., Espinoza, D., Goycoolea, M., Helsgaun, K. (2009). "Certification of an Optimal TSP Tour through 85,900 cities." *Operations Research Letters*, 37(1), pp. 11-15.

ASCE (2013). "2013 Report Card for America's Infrastructure." *ASCE*.

Baik, H., Jeong, H., Abraham, D. (2006). "Estimating Transition Probabilities in Markov Chain-Based Deterioration Models for Management of Wastewater Systems." *Journal of Water Resources Planning and Management*, 132(1), pp. 15-24.

Becker, G., Boduroglu, H., Camarinopoulos, S., Frondistou-Yannas, S., Gedikli. A., Kallidromitis, V., Kampranis, D., Sanna, C. (2009). "Structural Assessment and Upgrading of Sewer Beds on Inspection Results." *Journal of Infrastructure Systems*, 15(4), pp. 321-329.

Bertimas, D. J. (1992). "A Vehicle Routing Problem with Stochastic Demand." *Operations Research*, 40(3), pp. 574-585.

Bertsimas, D., Brown, D., Caramanis, C. (2011). "Theory and applications of robust optimization." *SIAM Review*, 53(3), pp. 464-501.

Boomgaard, M., Langeveld, J., Clemens, F. (2001). "Wastewater System Optimization using Genetic Algorithm." *Proceedings, Bridging the Gap Conference*, Orlando, FL, USA, May 20-24.

Botrous, A., El-Hattab, I., Dahab, M. (2000). "Design of Wastewater Collection Networks using Dynamic Programming Optimization Technique." *Proceedings, Environmental and Pipeline Engineering*, Kansas City, MO, USA, Jul. 23-26, pp. 503-512.

Bowen, A., Roesch, J., Mohamed, Y., AbouRizk, S., Al-Hussein, M. (2003). "Development of Productivity Standards for Collection Systems Maintenance Activities." *Proceedings, Water Environment Federation 76th Annual Technical Exhibition and Conference*, Los Angeles, CA, USA, Oct. 11-15, Sessions 41 through 50, pp. 803-813.

Caldwell, J. T. (2007). "Wastewater Collection System Rehabilitation and Replacement (R&R) Program Prioritization." *Proceedings, Pipelines: Advances and Experiences with Trenchless Pipeline Projects*, Boston, MA, USA, Jul. 8-11.

Canadian Infrastructure (2012). "Canadian Infrastructure Report Card 2012 – Highlights."

City of Edmonton (2012). “Capital Budget: Drainage Facilities.” City of Edmonton, Canada.

City of Edmonton (2012). “State of the City Report 2012.” City of Edmonton, Canada.

City of Edmonton (2012). “Drainage Services Annual Report, Leadership 2012.” City of Edmonton, Drainage Services.

City of Edmonton (2012). “Drainage Services Business Plan 2014-2016.” City of Edmonton, Drainage Services.

Chu, P., Beasley, J. (1998). “A Genetic Algorithm for the Multidimensional Knapsack Problem.” *Journal of Heuristics*, 4(1), pp. 63-86.

Chugtai, F., Zayed, T. (2008). “Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines.” *Journal of Performance of Constructed Facilities*, 22(5), pp. 333-341.

Chughtai, F., Zayed, T. (2011). “Integrating WRc and CERIU Condition Assessment Models and Classification Protocols for Sewer Pipelines.” *Journal of Infrastructure Systems*, 17(3), pp. 129-136.

Clark, S., Pitt, R., Burian, S., Field, R., Fan, E., Heaney, J., 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.

Dekker, R. (1996). “Applications of Maintenance Optimization Models: A Review and Analysis.” *Reliability Engineering and System Safety*, 51(3), pp. 229-240.

Desrochers, M., Desrosiers, J., Solomon, M. (1992). “A New Optimization Algorithm for the Vehicle Routing Problem with Time Windows.” *Operations Research*, 40(2), pp. 342-354.

El-Mihoub, T., Hopgood, A., Nolle, L., and Battersby, A., (2006). "Hybrid genetic algorithms: A review." *Engineering Letters*, 13(2), EL_13_2_11.

Fenner, R. A. (2000). "Approaches to Sewer Maintenance: A Review." *Urban Water*, 2(4), pp. 343-356.

Gamisch, T., Girmscheid, G. (2010). "Maintenance Life Cycle Cost Model for Drainage Systems of Infrastructure." *Challenges, Opportunities and Solutions in Structural Engineering and Construction,* Tailor and Francis Group, London UK.

Gaudreault, V., Lemire, P. (2006). "The Age of Public Infrastructure in Canada." Statistics Canada, Catalogue 11-621-ME, No. 35.

Guo, W., Soibelman, L., Garett Jr., J. (2009). "Visual Pattern Recognition Supporting Defect Reporting and Condition Assessment of Wastewater Collection Systems." *Journal of Computing in Civil Engineering*, 23(3), pp. 160-169.

Gupta, A., Mehndiratta, S., Khanna, P. (1983). "Gravity Wastewater Collection Systems Optimization." *Journal of Environmental Engineering*, 109(5), pp. 1195-1209.

Halfawy, M., Newton, L., Vanier, D. (2006). "Review of Commercial Municipal Infrastructure Asset Management Systems." *Journal of Information Technology in Construction*, 11, pp. 211-224.

Halfawy, M., Hengmeechai, J. (2013). "Efficient Algorithm for Crack Detection in Sewer Images from Closed-Circuit Television Inspection." *Journal of Infrastructure Systems*, 20(2), 10.1061/(ASCE)IS.1943-555X.0000161, 04013014.

Hammond, D., Horton, K. (1997). "Automatic Collection System Maintenance with Bar-coding." *Florida Water Resources Journal*, Issue March 1997, pp 31-32.

Hannan, P. (2000). “Collection System Practices for the New Age.” *Proceedings, Environmental and Pipeline Engineering*, Kansas City, MO, USA, Jul. 23-26, pp. 156-165.

Hassanein, A., Khalifa, R. (2008). “Financial and Operational Performance Assessment of Water/Wastewater Utilities: Comparative Study.” *Journal of Infrastructure Systems*, 14(4), pp. 283-292.

Hoffman, D., Buchberger, S., Flanders, M. (2010). “Preventing Sewer Blowouts during High-Velocity Jet Cleaning Operations.” *Journal of Infrastructure Systems*, 16(4), pp. 273-281.

Huang, G. and Loucks, D. (2000). “An inexact two-stage stochastic programing model for water resources management under uncertainty.” *Civil Engineering and Environmental Systems*, 17(2), 95-118.

ICE (2010). “The State of the Nation’s Infrastructure 2010.” Institution of Civil Engineers, London, UK.

Khan, Z., Zayed, T., Moselhi, O. (2009). “Simulating Impact of Factors Affecting Sewer Network Operational Condition.” *Proceedings, CSCE Annual Conference*, Calgary, AB, Canada.

Khazraeializadeh, S., Gay, L., Bayat, A (2014). “Comparative analysis of sewer physical condition grading protocols for the city of Edmonton.” *Canadian Journal of Civil Engineering*, 41(9), pp. 811-818.

Kleywegt, A., Shapiro, A., and Homem-De-Mello, T. (2001). “The sample average approximation method for stochastic discrete optimization.” *SIAM Journal on Optimization*, 12(2), 479-502.

Laporte, G. “The Vehicle Routing Problem: An Overview of Exact and Approximate Algorithms.” *European Journal of Operational Research*, 59(3), pp. 345-358.

Lei, H., Laporte, G., and Guo, B. (2011). “The capacitated vehicle routing problem with stochastic demands and time windows.” *Computers & Operations Research*, 38(12), 1775-1783.

Liu, G., Ge, Y., Qiu, T., Soleymani, H. (2013). “Vehicle Routing of Urban Snow Plowing Operation: A Case Study for the City of Edmonton.” *Proceedings, 92nd Annual Meeting of Transportation Research Board (TRB)*, Washington DC, USA, Jan. 13-17.

Loucks, E., Stahr, L. (2007). “Optimal Operation of the Little Rock Wastewater Collection System.” *Proceedings, World Environmental and Water Resource Congress*, Tampa, FL, USA, May 15-19.

Lu, M. (2001). “Productivity Studies using Advanced ANN Models.” Doctoral thesis, University of Alberta, Edmonton, AB, Canada.

Maier, H., Simpson, A., Zecchin, A., Foong, W., Plang, K., Seah, H., Tan, C. (2003). “Ant Colony Optimization for Design of Water Distribution Systems.” *Journal of Water Resources Planning and Management*, 129(3), pp. 200-209.

McKelvey, R., Zavoina, W. (1975). “A Statistical Model for the Analysis of Ordinal Level Dependent Variables.” *Journal of Mathematical Sociology*, 4(1), pp. 103-120.

Mohamed, Y., AbouRizk, S., Al-Hussein, M., Roesch, J., Bowen, A. (2002). “Using Simulation for Improving Productivity of Drainage Operations.” *Proceedings, CSCE Annual Conference*, Montréal, QC, Jun 5-8.

Moselhi, O., Fahmy, M. (2008). "Discussion of 'Prediction of Water Pipe Asset Life using Neural Networks' by D. Achim, F. Ghotb, and K.J. McManus." *Journal of Infrastructure Systems*, 14(3), pp. 272-273.

Nagata, Y., Kobayashi, S. (2013). "A Powerful Genetic Algorithm using Edge Assembly Crossover for the Traveling Salesman Problem." *Informs Journal on Computing*, 25(2), pp. 346-363.

Navab-Kashani, R., Gay, L., Bayat, A. (2015). "Productivity Improvement of Sewer CCTV Inspection through Time Study and Route Optimization." *Journal of Construction Engineering and Management*, 141(6).

Neter, J., Kutner, M., Nachtsheim, C., Wasserman, W. (1996). *Applied Linear Statistical Models, 4th Edition*. Times Mirror Higher Education Group Inc., IL, USA.

Pleau, M., Colas, H., Lavallee, P, Palletier, G., Bonn, R. (2005). "Global Optimal Real-time Control of the Québec Urban Drainage System." *Environmental Modeling and Software*, 20(4), pp. 401-413.

Plihal, H., Kretschmer, F., Schwarz, D., Ertl, Th. (2014). "Innovative Sewer Inspection as a Basis for an Optimised Condition-based Maintenance Strategy." *Water Practice & Technology*, 9(1), pp. 88-94.

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. Accessed: Jan., 2012.

Ralph, T., Kopman, L., Pulleyblank, W., Trotter Jr., L. (2003). "On the Capacitated Vehicle Routing Problem." *Mathematical Programming*, 94, pp. 343-359.

Ratliff, A., Donahue, R., Lewandowski, L. (2008). "Collection System Improvements to Effectively Reduce SSOs." *Proceedings, International Pipelines Congress*, Atlanta, GA, USA, Jul. 22-27.

Rostum, J. (2000). "Statistical Modelling of Pipe Failures in Water Networks." Doctoral Thesis, Norwegian University of Science and Technology, Norway.

Salman, B., Salem, O. (2012). "Modeling Failure of Wastewater Collection Lines using Various Section-Level Regression Models." *Journal of Infrastructure Systems*, 18(2), pp. 146-154.

Sample, D., Kilpatrick, S. (2006). "Prioritizing Collection System Repairs with Limited Information." *Proceedings, World Environmental and Water Resources Congress*, Omaha, NE, USA, May 21-25.

SAP (2015). "SAP Mobile Technology." http://go.sap.com/solution/mobile-technology.html. Accessed: Dec. 2015.

Selvakumar, A., Tafuri, A. (2012). "Rehabilitation of Aging Water Infrastructure Systems: Key Challenges and Issues." *Journal of Infrastructure Systems*, 18(3), pp. 202-209.

Solomon, M. M. (1987). "Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints." *Operations Research*, 35(2), pp. 254-265.

Song, L., AbouRizk, S. (2008). "Measuring and Modeling Labor Productivity using Historical Data." *Journal of Construction Engineering and Management*, 134(10), pp. 786-794.

Strauch, J., Wetzel, D. (2006). "Collection System Asset Management – Benchmarking Performance through Annual Reporting." *Proceedings, Pipeline Division Specialty Conference*, Chicago, IL, USA, Jul. 30-Aug. 2.

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

Ugarelli, R. (2010). "Asset Management for Urban Wastewater Pipeline Networks." *Journal of Infrastructure Management*, 16(2), pp. 112-121.

Vallabhaneni, S. (2010). "Wastewater Collection Systems: Literature Review." *Water Environment Research*, 82(10), pp. 931-940.

Vallabhaneni, S. (2011). "Wastewater Collection Systems: Literature Review." *Water Environment Research*, 83(10), pp. 983-993.

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

Vanier, D. (2001). "Why Industry Needs Asset Management Tools." *Journal of Computing in Civil Engineering*, 15(Special Issue), pp. 35-43.

Vijay, R., Gautam, A., Kalamdhad, A., Gupta, A., Devotta, S. (2008). "GIS-based Location Analysis of Collection Bins in Municipal Solid Waste Management Systems." *Journal of Environmental Engineering and Science*, 7(1), pp. 39-43.

Water Environment Federation (2010). *Wastewater Collection System MOP No. 7*, Sixth Ed., Water Environment Federation Press, Alexandria, VA, USA.

Yeh, S., Chu, C., Chang, Y., Lin, M. (2011). "Apply Tabu Search and Simulated Annealing to the Optimization of Sewer Networks." *Engineering Optimization*, 43(2), pp. 159-174.

Younis, R., Knight, M. (2008). "An Empirical Model for the Prediction of Structural Behavior of Wastewater Collection System." *Proceedings, International Pipelines Congress*, Atlanta, GA, USA, Jul. 22-27.

Younis, R., Knight, M. (2010). "A Probability Model for Investigating the Trend of Structural Deterioration of Wastewater Pipelines." *Journal of Tunnelling and Underground Space Technology*, 25(6), pp. 670-680.

Zaman, H., Bouferguene, A., Al-Hussein, M., Lorentz, C., Girard, L. (2012). "Sustainable Practice in Infrastructure Management through Route Optimization for Preventive Maintenance of Wastewater Collection System." *Proceedings, CSCE Annual Conference*, Edmonton, AB, Canada, Jun. 6-9.

Zaman, H., Bouferguene. A., Al-Hussein, M., Lorentz, C., Melmoth, D. (2013). "Estimation of Flushing Duration for Preventive Maintenance of Wastewater Collection System." *Proceedings, Annual CSCE Conference*, Montréal, QC, Canada, May 29-Jun. 1.

Zaman, H., Bouferguene. A., Al-Hussein, M., Lorentz, C. (2015). "Framework for modeling on-site productivity of preventive maintenance activities for wastewater collection system." *Journal of Infrastructure Systems*, 21(4), 10.1061/(ASCE)IS.1943-555X.0000252, 04015003.

Zaman, H., Bouferguene. A., Al-Hussein, M., Lorentz, C. (2015). "Improving the productivity of drainage network maintenance activities through schedule optimization." *Urban Water Journal,* Tailor & Francis online, 10.1080/1573062X.2015.1112409.