Predictive Analytics for Planning Inspections of Linear Water and Wastewater Infrastructure

by

R. Richard Harvey

A Thesis
Presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy

In
Engineering

Guelph, Ontario, Canada

© R. Richard Harvey, May, 2014
ABSTRACT

PREDICTIVE ANALYTICS FOR PLANNING INSPECTIONS OF LINEAR WATER AND WASTEWATER INFRASTRUCTURE

R. Richard Harvey  
University of Guelph, 2014

Advisor:  
Professor Edward McBean

Techniques currently available for modeling the deterioration of aging water and wastewater linear infrastructure tend to focus on providing municipalities with generalized estimates of condition at the network-wide level. These models have often been found incapable of reliably predicting the condition of individual pipes within a larger network.

The primary goal of this research was to utilize existing data mining tools to make predictions of individual pipe condition that could effectively direct the inspection, maintenance and rehabilitation of critical infrastructure on an asset-by-asset basis. The municipalities of Guelph, Ontario, Canada and Scarborough, Ontario, Canada are provided as case studies.

Portions of the sanitary sewer and stormwater networks in Guelph were inspected from 2008 to 2011 using closed circuit television (CCTV) technology. A combined application of predictive and spatial analytics effectively leverages information contained within the existing inspection dataset so that the potential threat posed by uninspected pipes can be suitably assessed. Methods are described for using
class-imbalanced inspection datasets to train, tune and test support vector machines, decision tree classifiers and random forests. Decision tree classifiers were found to be a useful first step for extracting information from existing datasets as they illustrate the influence of pipe-specific attributes (e.g. year of installation and length) on structural condition. Support vector machines were outperformed by random forests that achieved excellent levels of predictive accuracy for what is, in reality, a difficult classification task. Ultimately, the proposed modeling methodology has the potential to significantly reduce the time and money spent identifying bad condition, uninspected pipes.

An analysis of historical water main failures within Scarborough, Ontario, Canada indicates the majority of failures occur during the very cold winter months. Extensive installation of cement mortar lining and cathodic protection has extended the lifespan of aging water mains. Artificial neural networks are found capable of predicting the time to failure for individual pipes. Simulated failure scenarios indicate a return to high failure rates if cement mortar lining and cathodic protection are not extended to all candidate pipes within the distribution network.
“Knowing is not enough; we must apply.

Willing is not enough; we must do.”

– Johann Wolfgang von Goethe
Acknowledgements

The work in this thesis was funded by student and Canada Research Chair funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Guelph.

I left Newfoundland to come to Ontario largely because of the respect and admiration I have for my supervisor Dr. Edward McBean. You were an amazing mentor and a constant source of inspiration.

Thank you to Dr. Bahram Gharabaghi for your guidance, support and valuable feedback over the past four years. I am constantly amazed by your drive and creativity.

I would also like to thank my other committee members, Dr. Richard Zytner and Dr. Wanhong Yang for providing their valuable feedback throughout the course of my research.

Thank you also to the School of Engineering for providing a warm and welcoming place of study.

This thesis is dedicated to two of the funniest, thoughtful, caring and prettiest girls in the world: my wife Rebecca and my daughter Nell.

I am a fortunate man.
Contents

List of Tables xiii

List of Figures xvi

List of Equations xx

List of Abbreviations xxi

1 Introduction 1

1.1 Research Motivation . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.1.1 Towards Proactive Sewer Infrastructure Management . . . . . 3

1.1.2 Towards Proactive Water Distribution Infrastructure Man-
agement . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

1.2 Research Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . 6

1.3 Organization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

2 An Overview of Pipe Inspection Technology 14

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

2.2 The Role of GIS in Proactive Management . . . . . . . . . . . . . . 15

2.3 Screening Technology . . . . . . . . . . . . . . . . . . . . . . . . . 17

2.3.1 Zoom Camera Inspection . . . . . . . . . . . . . . . . . . . . . 17
### 4 Comparing the Utility of Decision Trees and Support Vector Machines when Planning Inspections of Linear Sewer Infrastructure

#### 4.1 Introduction

#### 4.2 Background Information on Sewer Pipe Condition Modeling

#### 4.3 Case Study

#### 4.4 Data Mining Methodology

4.4.1 Selection of a Classification Target

4.4.2 Implementing the Algorithms

4.4.3 Evaluating Predictive Performance

#### 4.5 Results

4.5.1 Support Vector Machine Model

4.5.2 Decision Tree Classifier

#### 4.6 Discussion

4.6.1 Comparing the SVM and Decision Tree Predictive Models

4.6.2 Screening Pipes for Future Inspection

4.6.3 The Impact of Dataset Size

#### Conclusion
4.6.4 Improving Overall Accuracy and Future Directions of Research

4.7 Conclusion

5 Predicting the Structural Condition of Individual Sanitary Sewer Pipes with Random Forests

5.1 Introduction

5.2 Sewer Deterioration Modeling

5.3 Predictive Modeling

5.3.1 Random Forests

5.4 The City of Guelph Case Study

5.5 Evaluating Predictive Performance

5.6 Results

5.7 Discussion

5.7.1 A Novel Approach to Enhancing Predictive Power

5.7.2 Predicting Uninspected Pipe Condition

5.7.3 Gauging the Success of Pipe-Level Models

5.8 Conclusion

6 Predictive and Spatial Analytics for Planning Inspections of Sewer Infrastructure

6.1 Introduction

6.2 Case Study

6.3 Assessing Predictive Capability

6.4 Results

6.5 Discussion
6.5.1 Planning Future Inspections through Combined Spatial and Predictive Analytics ........................................ 160
6.5.2 Sensitivity Analysis .......................................................... 163
6.5.3 Inherent Restrictions on Predictive Capability for Individual Pipes .......................................................... 164
6.6 Conclusions ........................................................................ 165

7 Understanding Stormwater Pipe Deterioration Through Data Mining ............................................................... 168

7.1 Introduction ........................................................................ 168
7.1.1 The Threat Posed by Deteriorating Stormwater Pipes ........................................................................ 168
7.1.2 Inspecting the Threat ...................................................... 170
7.1.3 Modeling Pipe Deterioration ........................................ 171
7.2 Data Mining Methodology .................................................. 173
7.2.1 The C5.0 System .............................................................. 173
7.3 Case Study ........................................................................... 174
7.3.1 CCTV Inspections of Guelph’s Stormwater Pipes .......................................................... 174
7.3.2 Evaluating Predictive Performance ...................................... 180
7.4 Results ................................................................................ 183
7.5 Discussion ........................................................................... 187
7.5.1 Planning Future Inspections ........................................ 187
7.5.2 Improving Predictive Performance ...................................... 190
7.6 Conclusion ........................................................................... 190

8 Predicting the Timing of Water Main Failure Using Artificial Neural Networks .................................................. 193
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>193</td>
</tr>
<tr>
<td>8.2</td>
<td>Case Study Scenario</td>
<td>198</td>
</tr>
<tr>
<td>8.3</td>
<td>A First Look at Pipe Failure in Scarborough</td>
<td>199</td>
</tr>
<tr>
<td>8.4</td>
<td>Artificial Neural Network Model Development</td>
<td>204</td>
</tr>
<tr>
<td>8.5</td>
<td>Results</td>
<td>208</td>
</tr>
<tr>
<td>8.6</td>
<td>Discussion</td>
<td>212</td>
</tr>
<tr>
<td>8.6.1</td>
<td>Planning Future Rehabilitation Activities</td>
<td>212</td>
</tr>
<tr>
<td>8.6.2</td>
<td>The Influence of Pipe-specific Attributes on ANN Prediction</td>
<td>215</td>
</tr>
<tr>
<td>8.7</td>
<td>Conclusions</td>
<td>218</td>
</tr>
<tr>
<td>9</td>
<td>Conclusions and Recommendations</td>
<td>220</td>
</tr>
<tr>
<td>9.1</td>
<td>The Role of GIS and Inspection Technology</td>
<td>221</td>
</tr>
<tr>
<td>9.1.1</td>
<td>Inspection Technology</td>
<td>221</td>
</tr>
<tr>
<td>9.1.2</td>
<td>The Role of GIS</td>
<td>221</td>
</tr>
<tr>
<td>9.1.3</td>
<td>Future Work Related to Pipe Inspection</td>
<td>222</td>
</tr>
<tr>
<td>9.2</td>
<td>Predicting Sanitary Sewer Pipe Condition</td>
<td>223</td>
</tr>
<tr>
<td>9.2.1</td>
<td>A Summary of Findings</td>
<td>223</td>
</tr>
<tr>
<td>9.2.2</td>
<td>Future Work Related to Sanitary Sewers</td>
<td>230</td>
</tr>
<tr>
<td>9.3</td>
<td>Predicting Stormwater Pipe Condition</td>
<td>231</td>
</tr>
<tr>
<td>9.3.1</td>
<td>A Summary of Findings</td>
<td>231</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Future Work Related to Stormwater Pipes</td>
<td>231</td>
</tr>
<tr>
<td>9.4</td>
<td>Predicting Water Main Failure</td>
<td>232</td>
</tr>
<tr>
<td>9.4.1</td>
<td>A Summary of Findings</td>
<td>232</td>
</tr>
<tr>
<td>9.4.2</td>
<td>Future Work Related to Water Main Failure</td>
<td>232</td>
</tr>
<tr>
<td>9.5</td>
<td>Concluding Remarks</td>
<td>233</td>
</tr>
</tbody>
</table>
# List of Tables

2.1 Case studies using zom camera inspection .................................. 19
2.2 WRc Structural Defect Scores .................................................. 22
2.3 WRc Scoring Thresholds ......................................................... 24
2.4 ICG assigned to clay ware, concrete and plastic sewers. ................. 25
2.5 CCTV inspection costs. ............................................................ 27

3.1 Contingency table for each split point in the tree ......................... 54
3.2 Confusion matrix for a binary classification task .......................... 55
3.3 The inspection dataset used for model development ...................... 61
3.4 Test Set Confusion Matrix for the Default Classification Threshold (50%) ............................................................. 63
3.5 Test Set Confusion Matrix for Optimal Classification Threshold (32%) 65
3.6 Test Set Confusion Matrix for a Classification Threshold of 20% ...... 65

4.1 Contingency table for each split point in the tree ......................... 81
4.2 Attributes available for data mining ......................................... 87
4.3 A summary of internal condition grade (ICG) by pipe material ....... 88
4.4 Confusion matrix for a binary classification task .......................... 89
4.5 SVM model confusion matrix for the test set (cutoff of 0.50) ........ 93
4.6 SVM model confusion matrix for the test set (cutoff of 0.11) ....... 94
List of Figures

2.1 The unique integration capability of GIS allows disparate data sets to be brought together/integrated to create a complete picture of an infrastructure situation. .................................................. 15

2.2 CCTV operators remotely control the robot from the inside of the survey truck and review the video on a display screen. .......... 21

3.1 Inspections are typically carried out using robots equipped with cameras that are inserted into a pipe and remotely operated from the road surface. .................................................. 46

3.2 A decision tree predictive model illustrates the knowledge gained from a hypothetical dataset in a tree-like format. In this example, the values of the input predictors (weather and wind speed) can be used to determine the likelihood of a game being played. ............... 50

3.3 The receiver operating characteristic (ROC) curve can be used to evaluate the trade-offs between true positives and false positives. . 59

3.4 The CART model for predicting sewer pipe condition. ............ 62

3.5 The evaluation set receiver operating characteristic (ROC) curve indicates an alternative threshold could reduce the false negative rate of the tree. .................................................. 64
3.6 Sanitary sewer pipes predicted to be in poor condition by the decision tree model that are also located in close proximity to a defective stormwater pipe are candidates for immediate inspection.

4.1 The evaluation set ROC curves developed for the SVM and decision tree models.

4.2 A visual depiction of the decision tree predictive model for sewer pipe condition.

4.3 Gain chart for the decision tree model.

5.1 Examples of structural defects observed during CCTV inspection.

5.2 Map of inspected pipes in Guelph, Ontario, Canada.

5.3 The Receiver Operating Characteristic (ROC) Curve.

5.4 The ROC curve for Random Forest Model 1.

5.5 The ROC Curve for Random Forest Model 2 (with upstream and downstream condition information).

5.6 Gain chart for the random forests (the shaded grey region indicates the lift that would be achieved with a perfect model).

6.1 CCTV technology is used to inspect the internal condition of sewer pipes currently in operation.

6.2 Change in average sanitary sewer pipe age over time.

6.3 Distribution of structural defects present in various sewer pipe materials.

6.4 Density estimation of inspected pipes found in ICG4-5 condition.
Predictive and spatial analytics can be combined to identify uninspected pipes that should be targeted for immediate inspection. . . 162

Length of Guelph stormwater system by year. . . . . . . . . . . . . 175
Inspected stormwater pipes. . . . . . . . . . . . . . . . . . . . . . . 177
Stormwater pipe condition classification tree. . . . . . . . . . . . . . 184
Distribution of condition for pipes installed after 1968. . . . . . . . . . 187
Location of poor condition stormwater pipes. . . . . . . . . . . . . . 189
Impact of CML and CP on the total number of water main failures
each year in Scarborough. . . . . . . . . . . . . . . . . . . . . . . . . . . 201
Impact of CML and CP on the total number of days each year with
a water main failure. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 202
Seasonal variation in the total number of water main failures. . . . . 203
Optimized cast iron artificial neural network architecture. . . . . . . . 210
Scatterplot of cast iron artificial neural network predictions of the
time to failure. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 211
Histogram of the difference between observed and predicted time to
failure for cast iron water mains with (a) zero previous failures, (b)
one previous failure, and (c) two or more previous failures. . . . . . . 212
Predicted number of yearly failures for cast-iron pipes installed post-
1946 without CML or CP. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 214
List of Equations

3.1 ....................................................... 54
3.2 ....................................................... 54
3.3 ....................................................... 54
3.4 ....................................................... 56
3.5 ....................................................... 57
3.6 ....................................................... 57
3.7 ....................................................... 57
4.1 ....................................................... 81
4.2 ....................................................... 82
4.3 ....................................................... 88
4.4 ....................................................... 90
4.5 ....................................................... 90
4.6 ....................................................... 90
4.7 ....................................................... 90
4.8 ....................................................... 90
4.9 ....................................................... 99
5.1 ....................................................... 113
5.2 ....................................................... 114
5.3 ....................................................... 115
List of Abbreviations

CCTV  closed circuit television
FN    false negative
FNR   false negative rate
FP    false positive
FPR   false positive rate
ICG   internal condition grade
OPG   operational performance grade
ROC   receiver operating characteristic
SPG   structural performance grade
SVM   support vector machine
TN    true negative
TNR   true negative rate
TP   true positive

TPR  true positive rate

WRc  Water Research Council
Chapter 1

Introduction

A large portion of water and wastewater infrastructure installed during the post-
World War II era of rapid economic growth and development is approaching the end
of its effective design life. Municipalities have entered into now what is commonly
referred to as the era of infrastructure replacement (ASCE, 2013). Over the coming
years, a massive investment will be required to maintain current levels of access to
clean drinking water and prevent widespread environmental contamination from
leaking sewers.

The most recent report card for American infrastructure indicates the more than
one million kilometers of publicly owned sewer mains currently in operation are
in poor condition (ASCE, 2013). Three-quarters of the estimated $298 billion
of capital investment in wastewater infrastructure that will be made over the
next twenty years will be dedicated exclusively to sewer pipes (ASCE, 2013).
Approximately $1 trillion will be spent over the next 25 years to restore existing
water distribution systems as they reach the end of their useful lives and expand
existing systems to meet the needs of increased population (AWWA, 2012).
Canada’s first report card on the condition of municipal infrastructure indicates 30% of sanitary sewer pipes and 23.4% of stormwater pipes are in less than good condition and require some form of rehabilitation or replacement (CIRP, 2012). An immediate investment of $25.9 billion ($2,082 per household) is required to replace the nation’s drinking water infrastructure currently in fair to very poor condition (CIRP, 2012).

The cracks, fractures, holes, and broken joints commonly found in aging sanitary sewer pipes allow significant amounts of raw wastewater to leak into the surrounding soil and nearby sources of groundwater. In municipalities with separate sanitary and stormwater networks, there are inevitable consequences when leaking raw sewage (containing pathogens, industrial toxins and endocrine disrupting compounds) migrates into broken stormwater pipes located in close proximity. Sewage contamination of a dedicated stormwater network poses a serious risk to human health whenever stormwater pipes release their untreated contents to surface waters, such as lakes and rivers commonly used for recreational pursuits.

The leaks that form in crumbling water distribution infrastructure are responsible for a staggering amount of water waste as there is a considerable difference in the amount of water currently going into distribution networks and the amount reaching consumers (referred to as non-revenue water). Up to 30% of municipal piped water in Canada is currently lost due to pipe leaks (Environment Canada, 2013). In America, anywhere from 14 - 18% of municipal water is lost through pipe leakage, which is equivalent to approximately 8 trillion liters of annual water loss. In the states surrounding the Great Lakes, 252 billion liters of water is lost annually - enough to meet the water needs of 1.9 million Americans for a full year.
Overseas in the United Kingdom, an estimated 3.3 billion liters of treated water is lost every day due to pipe leaks, which is approximately 20% of the nation’s total water supply (BBC News, 2012). Increased quantities of non-revenue water inevitably leads to increased water rates or taxes. As an example, water rates increased 90% in the United States from 1996 to 2010 and have been rising faster than any other utility rates (CNT, 2013). Rates will likely continue to rise as previous generations of pipe continue to break down.

1.1 Research Motivation

1.1.1 Towards Proactive Sewer Infrastructure Management

Sewage contamination of the urban environment is nearly ubiquitous and there is a strong likelihood that continued deterioration of municipal sewer systems may result in the loss of decades of progress that have been made in public health and environmental protection since the 1970s (EPA, 2002). Recognition of the potential for widespread environmental contamination in tandem with recently introduced provincial and federal legislation has caused a major shift in municipal sewer infrastructure management.

Many Canadian municipalities have long adopted an out-of-sight and out-of-mind approach to sewer infrastructure. Historically, tangible assets such as sewer pipes were recorded as expenditures in their year of installation and the value of existing assets did not appear on any further annual financial reports (i.e. pipes were installed and then forgotten). With the introduction of the Canadian Institute of
Chartered Accountants Public Sector Accounting Board (PSAB) Statement 3150 in 2009, capital asset inventory is now reported using accounting methods that indicate assets have a value long after their initial cost of construction is incurred, but this value depreciates over time. PSAB Statement 3150 requires municipalities to amortize the value of their existing pipes over their useful life - with depreciation recorded as an expense on the municipality’s statement of operations (OMBI, 2007). Similar legislation also exists outside of Canada - e.g. Governmental Accounting Standards Board (GASB) Statement 34 in America (GASB, 2012) and the Australian Accounting Research Foundation Standard 27 (Howard, 2001).

The data required for accurate estimation of sewer-related capital expenditure required for compliance with PSAB Statement 3150 can be obtained through visual inspection of sewer pipes currently in operation. Visual inspection of pipe condition is typically determined using expensive closed-circuit television (CCTV) technology, which consists of recording the internal condition of a pipe using a small video camera mounted on a robot. Canada’s municipal infrastructure deficit currently sits at approximately $123 billion ($31 billion for water and sewer systems) (Mirza, 2007), and the resulting budgetary restrictions tend to limit most municipalities to inspecting small portions of their entire network of pipes. These municipalities are in need of tools that can maximize the value of their existing inspection data so that accurate predictions of condition can be made for uninspected pipes in the sewer system. The majority of existing techniques that have been proposed for modeling sewer pipe deterioration and learning from existing inspection datasets are of limited utility as they often require excessive amounts of data manipulation, assumption-checking, and typically result in generalized predictions of condition.
that are only of utility at the network-wide level. As such, efficient approaches need to be made available to municipalities seeking to learn from existing inspection datasets so that location-specific predictions of condition can be made for individual pipes that have not yet been inspected.

1.1.2 Towards Proactive Water Distribution Infrastructure Management

Approximately $3 can be saved for every $1 spent on the proactive prevention of water distribution system pipe breaks (Environment Canada, 2013). As such, research into the proactive management of aging water distribution infrastructure is essential for the preservation of both municipal economic vitality and the conservation of vast amounts of water. Utility managers need access to location-specific information on water distribution pipe failure that can be of significant utility when planning inspection and rehabilitation-related work before pipe breaks occur. This work is of critical importance for conserving water on a global scale as more than 32 trillion liters of treated water is leaking out of urban water supply systems worldwide - with a total cost of $14 billion a year (Kingdom et al., 2006). The situation is particularly dire in developing countries, where approximately 45 million cubic meters of water is lost daily through water leakage in distribution networks - enough to meet the water needs of 200 million people (Kingdom et al., 2006). Reducing pipe breaks would generate $2.9 billion in cash every year for the water sector in developing countries (though increased revenue and reduced costs) (Kingdom et al., 2006).
1.2 Research Objectives

The goal of this research was to evaluate the suitability of various data mining techniques for leveraging information from existing records of water and wastewater pipe condition maintained by municipalities. This research will add to the existing state of knowledge regarding proactive management of linear water and wastewater infrastructure. The key objectives of the research were as follows:

- Using the City of Guelph, Ontario, Canada as a case study area, evaluate the suitability of data mining as a means of extracting information contained within existing datasets of pipe condition. No previous work of this nature had been carried out for the municipality.

- Enhance municipal-level understanding of the factors that influence sewer pipe deterioration.

- Develop efficient tools capable of predicting the condition of individual sewer pipes using information learned from previous inspections of a sewer system. A significant portion of both the case study municipality’s sanitary sewer and stormwater system remains uninspected and the groundwater supply has been indicated as being under threat of contamination (as shown in Allen (2013) - where 45% of 22 drinking water wells exhibited human enteric viruses derived from the exfiltration of domestic sewage flows).

- Develop novel methods of planning future rounds of sewer pipe inspection.

- Using the City of Scarborough, Ontario, Canada as a case study area, develop a tool capable of reducing water loss by predicting the time to failure of
individual pipes (i.e. provide the municipality with the capability to predict failure before it actually happens). Although substantial efforts have been made to model water pipe failure over the past 15 years, most studies have predicted generalized failure rates, condition scores or the total number of failures per year. Those three predictions do not provide info on when failures are most likely going to occur.

- Analyze historical records of water main failure to evaluate the utility of techniques aimed at extending the life of existing water distribution pipes (e.g. cement mortar lining and cathodic protection).

1.3 Organization

This thesis is organized in a manuscript format according to the University of Guelph 2013-2014 Graduate Calendar “Thesis Format” section. Following this format, chapters 2 through 8 are written as completely separate articles. The chapters are outlined here:

Chapter 1 - Introduction
Serves as an overarching introduction to the research covered by the thesis.

Chapter 2 - A Guide to Sewer Inspection Technology
Provides an introduction to sewer pipe inspection with particular emphasis on the technology currently available to North American municipalities - including zoom-cameras and closed-circuit television (CCTV). Addresses the potential benefits of inspecting pressurized forcemains using acoustic monitoring technology.
Chapter 3 - A Data Mining Tool for Planning Sanitary Sewer Condition Inspection

Introduces the data mining system known as the decision tree classifier as a means of extracting information from existing pipe inspection records. Decision trees are shown to be a useful method of gaining insight into sewer pipe deterioration in the municipality of Guelph, Ontario, Canada and are capable of enhancing industry understanding of the factors responsible for the deterioration of individual sewer pipes. This application is novel in that it relies exclusively on attributes available prior to pipe inspection to predict pipe condition (whereas some existing applications of decision trees used inspected condition as an input and would therefore not be suitable for predicting condition of uninspected linear assets). The model is developed with minimal data preprocessing effort and illustrates the influence of pipe-specific attributes on pipe condition in a format that can be easily shared with those unfamiliar with the data mining process. The decision tree model is used to predict individual pipe condition (Good vs. Poor). A novel approach to screening pipes for inspection using proximity to structurally defective stormwater pipes is also discussed. This paper has been accepted for publication in the peer-reviewed, edited volume Conflict Resolution in Water Resources and Environmental Management:

Chapter 4 - Comparing the Utility of Decision Trees and Support Vector Machines when Planning Inspections of Linear Sewer Infrastructure

Continues the investigation into the utility of using decision tree algorithms to extract knowledge from existing inspection datasets so that the condition of individual pipes in a sewer system can be predicted with certainty. In this manuscript the target used for prediction is whether or not pipes are in Good vs. Bad structural condition. This manuscript represents a shift towards predicting the condition of pipes in critical condition (whereas the previous chapter focused on a minority class of interest that may contain pipes with significant service life remaining).

The predictive capabilities of decision trees are compared against support vector machines using a case study of sanitary sewer pipe inspection data collected by the municipality of Guelph, Ontario, Canada. Both modeling algorithms are tuned to counteract the negative impact on predictive performance resulting from class imbalance common within pipe inspection datasets. The decision tree classifier outperforms support vector machines for this classification task - achieving an acceptable area under the receiver operating characteristic curve of 0.77 and an overall accuracy of 76% on a stratified test set. This paper has been accepted for publication in an upcoming issue of the peer-reviewed IWA Publishing Journal of Hydroinformatics:

Chapter 5 - Predicting the Structural Condition of Individual Sanitary Sewer Pipes with Random Forests

Introduces the potential benefits of predicting individual sewer pipe condition using the data mining algorithm known as the random forests classifier - combining the predictive power of hundreds of individual decision tree classifiers. A novel approach of considering neighboring pipe condition is described for improving the predictive power of the developed model. The random forest predictive models are found to have potential to significantly reduce the time and money allocated to identification of bad structural condition, uninspected pipes in a sanitary sewer network. This paper has been published in the peer-reviewed Canadian Journal of Civil Engineering:


Chapter 6 - Predictive and Spatial Analytics for Planning Inspections of Linear Sewer Infrastructure

Combines the predictive benefits of random forests with network spatial analytics. The utility of a tool for identifying clusters of bad condition pipes within the case study municipality is discussed. This paper has been accepted for publication in the peer-reviewed International Journal of Environmental Protection:

Chapter 7 - Understanding Stormwater Pipe Deterioration through Data Mining

Discusses the potential benefits of using decision tree classifiers (such as those indicated in Chapters 3 and 4) to predict the condition of individual stormwater pipes - using the municipality of Guelph, Ontario as a case study. The predictive model, developed using 107 km of inspection data, illustrates the influence of construction year, diameter, length and slope on individual stormwater pipe condition. Year of construction was found to strongly influence structural condition, with 82% of those pipes placed after 1968 in good structural condition and 45% of those placed prior to 1968 in poor structural condition. Pipe diameter, slope and length were found to strongly influence the resulting condition of the older pipes in the stormwater system. This paper has been published in the peer-reviewed Journal of Water Management and Modeling:

Chapter 8 - Predicting the Timing of Water Main Failure Using Artificial Neural Networks

Data mining analysis of historical records of failure obtained from the municipality of Scarborough, Ontario indicates the installation of cement mortar lining and cathodic protection on candidate-pipes can extend the lifespan of aging assets through reduction in the annual number of water pipe failures. Novel applications of artificial neural networks are used to determine the time to future failure for individual assets in water distribution systems - which is of critical importance when pro-actively planning inspection and rehabilitation of individual assets before a leak forms and enormous quantities of treated water is wasted. This paper has been published in the peer-reviewed ASCE Journal of Water Resources Planning and Management:

- Harvey, R., McBean, E., and Gharabaghi, B. 2014. Predicting the Timing of Water Main Failure Using Artificial Neural Networks. Journal of Water Resources Planning and Management, 140(4), 425-434 (with permission from the copyright holders, ASCE Publishing)

Chapter 9 - Conclusions and Recommendations

Serves as an overarching conclusion that emphasizes the contributions made by this research to the field of engineering as well as recommendations for future work.

Author’s note: Appendix A contains permission forms for including previously published manuscripts in this thesis. Appendix B contains a list of general references that would prove useful for those looking to learn more about the data mining algorithms presented in this thesis.
Transition to Chapter 2

The following chapter provides an introduction to pipe inspection - with an emphasis on the technology currently being used in North America to inspect sewer pipes and the standardized methodology used to establish pipe condition ratings.
Chapter 2

An Overview of Pipe Inspection Technology

2.1 Introduction

A proactive approach to the maintenance and rehabilitation of existing sanitary sewer, stormwater, and water distribution pipes requires accurate data on their condition. This chapter provides an overview of sewer pipe inspection technology - with particular emphasis on technology that is currently available to North American municipalities (e.g. zoom-cameras, closed-circuit television, and digital scanning). A look beyond these traditional inspection techniques is provided with an indication of the potential to use focused electrode leak location, sonar and laser inspection methods to improve municipal knowledge of pipe condition. Acoustic monitoring, now commonly used for detecting water main breaks, is shown to be a potentially useful method of detecting leaks in pressurized sewer forcemains.
Figure 2.1: The unique integration capability of GIS allows disparate data sets to be brought together/integrated to create a complete picture of an infrastructure situation.

2.2 The Role of GIS in Proactive Management

The ability for a municipality to effectively manage its wastewater infrastructure will largely be controlled by its ability to organize large amounts of data. A proactive management system can help organize information on asset location, properties, maintenance history, inspection records and typically consists of the following components:

A municipality can organize all of its geographic information (i.e. sewer location, water mains, roads, etc.) into one seamless environment (Figure 2.1). The
unique integration capability of GIS allows disparate data sets to be brought together/integrated to create a complete picture of a situation. This makes GIS an ideal tool to enable more intuitive and efficient mechanisms to query, explore, visualize and analyze infrastructure data in its spatial context (Halfawy et al., 2007).

Infrastructure assets need to be classified according to a capability to perform an intended function (e.g. transporting sewage, stormwater or treated drinking water). GIS can be used as an excellent document management system that saves time spent in locating, organizing and confirming the accuracy of disparate asset attribute information and data collected from field-inspections. If inspection data is not integrated it will largely be under-utilized as accessing the information from conventional video records and paper inspection log sheets is difficult and time consuming. Most industry experts believe utilities can significantly reduce their inspection and maintenance costs by implementing GIS-based asset management practices. Municipalities are recognizing that GIS is the perfect platform to design and create an integrated GIS-centric asset management system using spatial relationships as a way to manage, coordinate and analyze public assets and work activities.

The challenges of managing buried assets are substantial (Lemer, 2001). The functional components of a region’s infrastructure are managed by a myriad of agencies and at several jurisdictional levels. Software specifically designed to assist with managing buried assets can improve the rate of return on public investment. To consider a Canadian example: the City of Guelph, Ontario integrated their asset databases with their GIS, making it easier for municipal staff to find the data they
need to make effective decisions, perform their work efficiently and better manage community assets.

As part of this thesis, existing pipe inspection data was integrated with existing GIS asset inventory data. It is now possible for engineers, city planners and other staff to access data layers on thousands of municipal assets including road networks, buildings, and property base maps as well as any previously carried out inspections on water and sewer infrastructure. The municipality now has all of its asset information in one place and has eliminated redundant data. Incorporating inspection data into GIS has made it possible for the municipality to learn the “whole story” of an asset and its condition over time.

2.3 Screening Technology

A systematic approach to sewer evaluation and rehabilitation is critical for the efficient use of limited financial resources (WEF, 2007). Screening technologies let cash-strapped municipalities devote the majority of their time and funding to detailed assessments of problematic sections of sewer pipe.

2.3.1 Zoom Camera Inspection

Zoom camera inspection involves the generation of still imagery and/or recorded video imagery of a pipe using a camera mounted on a pole that is lowered into a manhole. Because the camera remains stationary, the task of imaging the pipe is quite quick - up to 1609 meters of inspection can be carried out per day,
whereas typical CCTV inspections usually cover only 300-460 meters per day. Zoom cameras should not be viewed primarily as a replacement for more thorough inspection using CCTV - but rather as a tool to screen and prioritize pipes for more comprehensive follow-up inspection (EPA, 2010).

The defects most likely to be missed by zoom camera inspection are those located in the middle portion of the pipe (i.e. those that are farthest away from the manholes). Brainbridge and Krinas (2008) explored this issue and determined that on average 59.44% of pipe defects are located within 20 meters of manholes, and 76.12% are within 30 meters of manholes. Joseph and DiTullio (2003) noted that 80% of the pipe defects are usually within 15-20 meters of the manhole.

Most zoom cameras have a site distance of 10-30 meters inside a pipe - suggesting the cameras are quite capable of detecting a large percentage of the defects within a pipeline. Defects tend to be located closer to manholes for a number of reasons including, vibrations from surface traffic and void areas created by infiltration around the manhole (Joseph and DiTullio, 2003). Large changes in direction of a pipe through a manhole increase the potential for turbulence, which increases the rate of transfer of sulphides from the liquid phase to gas phase above the sewer flow and accelerates the rate of corrosion of the exposed pipe wall (Mashford et al., 2011).

A brief summary of some field applications of zoom camera equipment is presented in Table 2.1. The cost of zoom camera inspection is reported to be one-half to two-thirds less than the cost of cleaning and conventional CCTV inspection, based on case study reports.
Table 2.1: Case studies using zoom camera inspection

<table>
<thead>
<tr>
<th>Location</th>
<th>Technical Performance</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aqua-zoom pilot project in Houston, Texas, USA</td>
<td>Approximately 1 mile of pipe inspected per day. 2% of the system needed repairs and 28% needed cleaning and further CCTV. No cost provided.</td>
<td>Renfro et al. (2005)</td>
</tr>
<tr>
<td>Fairfax, Virginia, USA</td>
<td>85 miles of sewer inspected. 66% of pipes needed further CCTV. Cost of the zoom camera combined with CCTV was $3.33/ft. Relying solely on CCTV would cost $4.89/ft.</td>
<td>Batman et al. (2008)</td>
</tr>
</tbody>
</table>

2.4 Inspecting Sewer Pipes

2.4.1 The WRc Inspection Methodology

One of the most commonly used systems of carrying out a structural investigation of a sewer pipe was developed by the Water Research Center (WRc) in the United Kingdom. The WRc investigation involves identifying, inspecting, and assessing the structural condition of sewers where a proactive approach to structural rehabilitation is justified. A WRc structural investigation comprises the following six stages: prepare the inspection program, carry out investigations, assess structural performance, compare with performance criteria, identify structural deficiencies, and identify the causes of deficiencies.
2.4.2 Step 1 - Prepare the Inspection Program

The principle criterion for selecting sewers is to select a group of pipes such that the total costs of failure over time would be significantly higher than the costs of inspection and rehabilitation before failure (WRc, 2001). In countries where inspection of sewers is legislated, the legal requirement of inspection overrides any other considerations. A master plan is then prepared to indicate which sewers will be inspected.

2.4.3 Step 2 - Carry Out Investigations

Sewers identified for inspection in the first step are inspected and the results collated. The most commonly used method for inspecting sewers are closed-circuit television (CCTV) systems. This nondestructive inspection technique is conducted by recording the internal condition of a pipeline using a small inspection camera mounted on a robot so that humans do not have to physically enter the pipe. A section of sewer to be inspected will generally start at one manhole and then run downstream to another manhole. A service truck is parked above one access point of the pipe and the robot (with a flexible cable attached to the rear) is then lowered into the pipeline. The camera operator remotely controls the camera from inside of the survey truck and reviews the video on a display screen (Figure 2.2). This video record can be used to identify longitudinal cracks, circumferential cracks, collapsed sections, damaged pipe walls, defective and displaced joints, evidence of abrasion or corrosion, siltation, encrustation, root penetration, loss of mortar, deformations, infiltration and the degree of penetration of all lateral connections.
Upon completion of the inspection, the camera tractor is put into reverse and the cable is wound up at the same time.

Figure 2.2: CCTV operators remotely control the robot from the inside of the survey truck and review the video on a display screen.

The results of the CCTV inspections normally contain the following:

- a written report containing factual information about the sewers.
- coded inspection data - where all visual inspection data is coded in accordance
with the WRc Manual of Sewer Condition Classification (WRc, 1996).

- A video recording of the inspection.

## 2.4.4 Step 3 - Assess Structural Performance

The initial stage of assessing structural performance is to put the defect codes assigned in step 2 in order of severity and then use a software package to assign scores to each of these defects according to their relative severity (Table 2.2).

<table>
<thead>
<tr>
<th>Defect</th>
<th>MSCC Code</th>
<th>Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinally displaced joint</td>
<td>OJM</td>
<td>Medium &lt; 1* pipe thickness</td>
<td>1</td>
</tr>
<tr>
<td>joint or open joint.</td>
<td>OJL</td>
<td>Large &gt; 1* pipe thickness</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If soil visible grade as hole</td>
<td>165</td>
</tr>
<tr>
<td>Radially displaced joint</td>
<td>JDM</td>
<td>Medium &lt; 1* pipe thickness</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>JDL</td>
<td>Large &gt; 1* pipe thickness</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; 10% diameter and soil visible</td>
<td>80</td>
</tr>
<tr>
<td>Surface crack</td>
<td></td>
<td>Circumferential</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longitudinal*</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex*</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Helical*</td>
<td>2</td>
</tr>
<tr>
<td>Cracked</td>
<td>CC</td>
<td>Circumferential</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>CL</td>
<td>Longitudinal*</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex*</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>CM</td>
<td>Helical*</td>
<td>40</td>
</tr>
<tr>
<td>Fractured</td>
<td>FC</td>
<td>Circumferential</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>FL</td>
<td>Longitudinal*</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex*</td>
<td>80</td>
</tr>
</tbody>
</table>

Continued on Next Page...
Table 2.2 – Continued

<table>
<thead>
<tr>
<th>Defect</th>
<th>MSCC Code</th>
<th>Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>FM Helical</td>
<td>B</td>
<td>Radial extent &lt; 1/4</td>
<td>80</td>
</tr>
<tr>
<td>Broken B</td>
<td></td>
<td>Radial extent 1/4 +</td>
<td>80</td>
</tr>
<tr>
<td>Hole H Radial extent &lt; 1/4</td>
<td></td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Radial extent 1/4 +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collapsed X</td>
<td></td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Spalling Slight</td>
<td>SSS</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>SSM</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>SSL</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Wear Slight</td>
<td>SWS</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>SWM</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>SWL</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Intruding sealing ring</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Defective repair Radial extent &lt; 1/4</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radial extent 1/4 +</td>
<td></td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Weld failure (plastic)</td>
<td>Longitudinal*</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Circumferential</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helical</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weld failure (steel)</td>
<td>Longitudinal*</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Circumferential</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helical</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deformation D</td>
<td>D</td>
<td>0 - 5%</td>
<td>20</td>
</tr>
<tr>
<td>6 - 10%</td>
<td></td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>greater than 10%</td>
<td></td>
<td>165</td>
<td></td>
</tr>
</tbody>
</table>

For peak score calculation assume longitudinal defects extend for 1m (unless the continuous defect facility is in use). Longitudinal defects are indicated by *. If a number of circumferential defects appear at the same chainage, only the most severe single defects is included, regardless of the radial extent. Source: WRc (2001)
The software then scans the record for each inspected pipe and calculates:

- Peak score for the manhole-to-manhole length. Any sewer with a reasonable risk to collapse in the short term due to a random event is immediately identified by the peak score.

- Total score for the manhole-to-manhole length. Pipes with significant deterioration are quickly highlighted by a review of their total score.

- Mean score for the manhole-to-manhole length (i.e. the total score divided by the length). Pipes with short reach lengths but significant deterioration are quickly highlighted by a review of the mean score.

The pipe is then assigned an internal condition grade (ICG) according to where the peak score falls within a set of threshold values (Table 2.3). The WRc recommends only considering the peak score as it is the worst defect along the manhole length (though some engineers also consider the total scores and mean scores for planning purposes). The ICG is then an assessment of expression of failure risk (Table 2.4).

Table 2.3: WRc Scoring Thresholds.

<table>
<thead>
<tr>
<th>Computed ICG</th>
<th>Peak Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Less than 10</td>
</tr>
<tr>
<td>2</td>
<td>10 - 39</td>
</tr>
<tr>
<td>3</td>
<td>40 - 79</td>
</tr>
<tr>
<td>4</td>
<td>80 - 164</td>
</tr>
<tr>
<td>5</td>
<td>more than 165</td>
</tr>
</tbody>
</table>
Table 2.4: ICG assigned to clay ware, concrete and plastic sewers.

<table>
<thead>
<tr>
<th>ICG</th>
<th>Typical defect descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td><em>Collapsed or collapse imminent.</em></td>
</tr>
<tr>
<td></td>
<td>Deformation greater than 10% and broken; Extensive areas of fabric missing; Fracture with deformation.</td>
</tr>
<tr>
<td>4</td>
<td><em>Collapse likely in the foreseeable future.</em></td>
</tr>
<tr>
<td></td>
<td>Broken; Deformation up to 10% and broken; Fracture with deformation 6 - 10%; multiple fracture; serious loss of level; Serious joint defects with voids or soil visible (open joint with greater than 50 mm soil or void visible or joint displacement greater than 25% of diameter); Surface damage - spalling large; Surface damage - wear large.</td>
</tr>
<tr>
<td>3</td>
<td><em>Collapse unlikely in the near future but further deterioration likely.</em></td>
</tr>
<tr>
<td></td>
<td>Fracture with no deformation or deformation less than 5%; Longitudinal cracking or multiple cracking; Minor loss of level; Severe joint defects (i.e. open joint large or joint displaced large); Surface damage - spalling medium; Surface damage - wear medium.</td>
</tr>
<tr>
<td>2</td>
<td><em>Minimal collapse likelihood in the short term but potential for further deterioration.</em></td>
</tr>
<tr>
<td></td>
<td>Moderate joint defects (i.e. open joint medium or joint displaced medium); Surface damage - spalling slight (breaking away of small fragments from the surface); Surface damage - wear slight (increased roughness).</td>
</tr>
<tr>
<td>1</td>
<td><em>Acceptable structural condition.</em></td>
</tr>
<tr>
<td></td>
<td>No structural defects.</td>
</tr>
</tbody>
</table>

2.4.5 Step 4 - Compare with Performance Criteria

The structural performance of the inspected sewers should then be compared with any performance criteria set out during the initial planning stages of the inspection program.

2.4.6 Step 5 - Identify Structural Deficiencies

In general terms, this step involves identifying sewers where the risk of failure (from step 3) is higher than the accepted criteria established in step 4. The inspection results should be examined thoroughly at this point to identify any urgent structural deficiencies posing an immediate threat to system integrity.

2.4.7 Step 6 - Identify Causes of Deficiencies

At this point, it is important to examine the causes of the performance deficiencies to plan effective rehabilitation or upgrading options for the structurally defective pipe.

Author's note: a WRc procedure exists for identifying defects posing a threat to the operational performance of an inspected pipe. The focus in this thesis is on understanding and predicting structural condition - consequently, operational performance grades will not be discussed, but further information can be found in WRc (2001).
2.5 CCTV Inspection Costs

Although CCTV has been a mainstay of sewer condition assessment over the past three decades, the amount of publicly available inspection cost data is very limited (EPA, 2010). Table 2.5 summarizes the average inspection cost and project size for small, medium and large utilities from a recent costing survey. There is a considerable amount of variability in the cost associated with carrying out CCTV inspections. The cost of an inspection will vary depending on the type of work being completed and how the work is accomplished (contractors vs. utility-owned equipment and in-house staff). Projects conducted by contractors may cost more in the short term than projects conducted by utilities that have the necessary in-house human resources. In other cases, providing a contractor with a steady supply of inspection work can reduce the project cost (i.e. like in the Denver, Colorado suburb of Westminster where extremely low CCTV costs were achieved by providing a private contractor with system-wide inspections over a five-year cycle). In either case, inspection costs will likely reach into the million dollar range - so careful planning of any inspection program is a necessity.

Table 2.5: CCTV inspection costs.

<table>
<thead>
<tr>
<th>Utility size (n = sample size)</th>
<th>Average cost per ft</th>
<th>Range of costs per ft</th>
<th>Average Length of Inspection (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small (n = 8)</td>
<td>$0.84 ±$0.37</td>
<td>$0.35 - $1.47</td>
<td>513,755 ±160,293</td>
</tr>
<tr>
<td>Medium (n = 8)</td>
<td>$0.62 ±$0.29</td>
<td>$0.28 - $1.21</td>
<td>1,370,572 ±377,091</td>
</tr>
<tr>
<td>Large (n = 4)</td>
<td>$0.76 ±$0.44</td>
<td>$0.33 - $1.17</td>
<td>5,586,947 ±3,657,579</td>
</tr>
</tbody>
</table>

All monetary amounts are in American dollars. Source: RedZone (2009)
2.6 Criticisms of CCTV

CCTV surveys are widely used throughout the wastewater industry but are not without fault. CCTV inspections miss certain types of defects, especially those that are hidden from the camera by obstructions. Slight deformations of the sewer may go unnoticed and any defect hidden beneath water inside the sewer will definitely not be found. In some cases, cracks hidden by mud may be missed and harmless glare may be recognized as damage. This inspection method is also particularly unsuited to surveying pipes with varying diameters, materials (including brick, concrete, ductile iron, and clay), odd shapes, sumps, and angle entries (Tafuri et al., 2002).

The usefulness of a CCTV inspection is tied to operator interpretation of the survey video. Defects may go unnoticed for any number of reasons including operator inattention, distractions, environment, fatigue and equipment capability. Correct diagnosis of failures or defects will depend on the reliability of the image being displayed on the operator’s computer monitor. It is also difficult to estimate the field production rate as the time required to inspect a pipe will depend on the number and degree of deterioration of the defects in the pipe. The camera operator needs to stop, assess and record the condition of the pipe at every defect and extensive surveys of sewer systems can take years to complete (Zhao et al., 2001; Wirahadikusumah et al., 1998).

As for the actual coding of the defects there is the potential problem that the location of defects will be incorrectly marked if cameras equipped with zoom capabilities are used in a survey. Even experienced operators tend to place the
camera in a static position and zoom-in on a defect that is coded anywhere from 1.5 - 6 meters in advance of its true position. This will certainly cause trouble when defects need to be located during follow-up repair activities. One possible solution to the problem of subjective image interpretation is the automatic assessment of conventional CCTV video using image processing techniques. Research into this area has been on-going for more than 10 years - many consider the technique to be unfeasible because of the lack of geometric references, unsteady camera movement and the overall poor quality of the images (Kirkham et al., 2000). Guo et al. (2009) illustrates that there has been some progress in the area but that the method for object detection and recognition is still an active research area in the fields of image processing and computer vision. A significant amount of further research is needed to adapt and improve the existing techniques in image reasoning and computer vision before automatic assessment of pipe inspection data becomes commonplace in the industry.

2.7 Alternative Inspection Technology

2.7.1 Digital Scanning

While CCTV inspections have traditionally been used in the assessment of internal sewer conditions, digital scanning is emerging as a potentially more effective inspection alternative. Digital scanning represents the state-of-the-art in camera inspection technology. Like conventional CCTV, digital scanning equipment is transported through pipelines using self-propelled crawlers. Unlike conventional
CCTV camera, digital scanning uses 360-degree high-resolution fish-eye digital cameras to scan the pipe wall, millimeter by millimeter, to form unfolded, 360 degree spherical images of the pipe. The unfolded view of the inner pipe surface provides an excellent view of pipe condition and permits computer-aided measurement of defects (Feeney et al., 2010). These unfolded images also make the comparison between one image and another collected at a later date very easy. Furthermore, digital scans allow for a second level of quality-control as other individuals involved in the review process (e.g. designers, rehabilitation contractors and utility owners) can quickly review the unfolded images to gain insight pipe condition and defect coding (unlike CCTV, where a complete video inspection would need to be reviewed). Digital scanning makes it possible for the data to be assessed independently of the real-time sewer inspection (unlike CCTV, where the camera operator has to pant, tilt and zoom into damaged areas of the pipe for further review). Digital scans are however more expensive than CCTV and they can only rarely be used to inspect laterals. Like CCTV, digital scanning equipment only provides images of the pipe surface above the waterline.

2.7.2 Laser Profiling

Camera-based technologies are limited to detecting defects in the pipe wall and cannot be used to identify wall thickness or demonstrate the potential for infiltration when there is no visible infiltration at the time of inspection. Alternative inspection techniques are required when the integrity of the pipe wall is to be evaluated (EPA, 2010). Laser profiling can be used to generate a 2-D profile of the pipe interior wall above the waterline so that changes in the shape of a sewer pipe that may
be caused by deformation, deflection, corrosion or siltation can be identified. The concept is fairly simple - mount laser profiling equipment on the front of a CCTV unit, project a ring of laser light onto the internal pipe surface, and analyze the resulting ring of light as the camera moves through the pipe. Currently, laser data analysis does not rely on a standard defect coding system - but instead relies on the judgment of the operator to determine the severity of any defects in the pipe wall. The tool is particularly useful for determining the true pipe condition prior to and after rehabilitation. The generated 2-D profile provides accurate information on reductions in cross-sectional area, corrosion, grease build-up, etc.

Laser profiling is often used in conjunction with CCTV inspection for large diameter pipes - where the additional cost of using both inspection technologies is offset by the cost savings associated with better defined rehabilitation projects. In the same way that airplane pilots rely on a combination of vision, instrumentation and experience to make good decision, a CCTV operator’s ability to interpret the video image is greatly enhanced by laser information. While both sensors have their own distinct strengths and weaknesses, their combined use provides a superior level of inspection. For example, lasers can detect minute changes in pipe geometry such as ovality that would be impossible to detect with CCTV (RedZone, 2011).

2.7.3 Sonar for Detecting Debris

Sonar inspections of pipelines are accomplished by passing a sonar head (mounted on either a raft, skid or robotic tractor) through a sewer. As the sonar head moves through the pipeline it sends out high frequency ultrasonic signals which are reflected by the pipe walls and then received by the sonar head (EPA, 2010). The
technology can be applied to gravity sewers and sewage forcemains made of any pipe material, and it can be deployed in pipes with diameters greater than 4 inches. Multi-frequency sonar inspections are increasingly being used to discriminate between the variety of materials that can be found within a sewer pipeline. All materials conduct sound in different ways - so multi-frequency sonar units can distinguish between debris, grease and other build-up on the pipe walls. This is particularly useful when estimating the volume of sediment in a pipeline and identifying points of collapse that are unreachable for CCTV inspection. This ensures scope of work packages for the procurement of sewer pipe cleaning are accurate and helps to avoid costly time and material cost overruns that occur when decisions are made with lacking information (RedZone, 2012). Sonar can also be used to identify corrosion, pits, voids and cracks in the pipe wall below the water surface (extremely advantageous as most imaging techniques fail to provide a detailed view of the pipe below the waterline). However imaging below the waterline can be limited by the amount of sediment on the pipe wall. As with laser, sonar data analysis does not use a standard defect coding system but relies on engineering judgment to assess the severity of a defect.

2.7.4 Electrical Methods for Identifying Leaks

The results from traditional visual inspection methods like CCTV are known to be only been marginally successful in locating defects in sewer pipe walls and pipe connections responsible for infiltration and exfiltration. Electrical leak detection systems have shown great promise in identifying the location and magnitude of defects in sewer pipes by measuring the electrical resistance of the pipe wall. The
basic premise is that while electricity will not pass through non-metallic pipe walls, it will pass through any defect that can cause water leaks. Thus, this technique works for all non-ferrous pipe - as most common sewer pipe materials (clay, plastic, concrete, reinforced concrete and brick) are electrical insulators that have a high resistance to electrical current. Electro-scanning measures the current flowing between an electrode in the pipe (referred to as a sonde) and an electrode kept in a fixed position on the surface (usually a metal stake driven into the ground). The water in the pipe is maintained at a level that ensures the pipe is completely full at the location of the sonde (as water is needed to conduct the current). A sliding plug is usually used to hold the water in the pipe as the sonde is moved through the pipe at a speed of approximately 10 meters per minute.

A microprocessor and other electronic devices in the sonde undertake continuous measurement of the current between the sonde electrode and the surface electrode, and send these data along the winch cable to the recording system on the surface. The cable passes over an electronic measuring pulley that sends data to the recording system. When the center of the sonde comes within close proximity of a defect in the pipe the electric current between the sonde and the surface electrode increases, attaining a maximum when the center of the sonde is radially aligned with the defect. Variations in the electric current flowing between the sonde and the electrode on the surface provide information on the pipe condition. Results of electro-scanning are typically graphed to show spikes or elevated levels of the measured electrical current that indicate the location of potential leaks, pipe defects or pipe features. It is the shape and amplitude of these anomalies that are interpreted to define the type and severity of the defect. Operator experience and
previous studies are used to distinguish between electrical currents that represent normal conditions (i.e. no defect) versus an anomaly (i.e. cracks and defective joints). The magnitude of the anomaly is estimated based on a comparison of electro-scanning results with pressure testing results for pipe joints (Harris and Tasello, 2004).

Gokhale and Graham (2003) used the Focused Electrode Leak Locator (FELL-41) in a 2001 pilot testing program carried in Louisville, Kentucky for gravity sewers (vitrified clay, PVC, and high-density polyethylene). Eight different sanitary sewer lines ranging from 200-300 mm diameter were tested as part of the project. While the FELL-41 system located numerous defects in the pipes, this inspection method was not considered to be a replacement for CCTV. Instead, the FELL-41 system can be used to prioritize or eliminate sections that do not need to be inspected with CCTV. They determined that this inspection method is not cost-effective for larger pipes (i.e. up to 1.5 meters in diameter) because of the time and cost of flooding the pipe. Harris and Tasello (2004) described the use of electro-scanning in the city of Redding, California - where the municipality’s wastewater collection system had a large infiltration problem when during periods of high rainfall, treatment plants received flows in excess of 300% of their daily dry flows. The costs to operate the systems were considerably higher during these high volume wet weather flows. Excessive amounts of infiltration also resulted in greater capital expenditure to increase pipe sizes and treatment plant capacity. The city had used flow monitoring, CCTV inspection, joint air pressure testing and smoke testing but had little success locating points of infiltration and peak wet weather flows were still a problem after 20 years of inspection. They figured a large number of pipe defects were not
being located with the commonly used methods. An electro-scan pilot study was conducted with the objective of collecting more accurate and reliable information on the location, nature and size of potentially leaky pipe defects: 7.6 km of 150 mm and 200 mm sewer pipe were scanned over a period of seven days using the FELL-41 system. Integrating the electro-scan surveys into regular jet cleaning operations turned out to be a straightforward process. Twelve sites for spot repair using the defect position provided by the electro-scan data. Prior to these repairs the storm event flows in the collection sub-basin were in excess of 450,000 gallons per day and the system often overflowed during these storm events. After the repairs, the sub-basin maximum flow during similar storm events was a more manageable 250,000 gallons per day. Overall, the sewer rehabilitation program formulated using the electro-scan data was much more cost effective than any other wet weather flow reduction program that had been previously carried out by the city. Harris and Dobson (2006) inspected 2230 feet of gravity sewer pipe using joint pressure testing, CCTV and electro-scanning. The number of defective joints detected by joint-pressure testing were within 4% of those detected by electro-scan, and both the joint-pressure testing and electro-scanning independently detected more than three times the number of defective joints detected by CCTV. The comparison between faulty service connections reported by CCTV and electro-scan showed that electro-scan detected four times more defective service connections than CCTV. The electro-scan also suggested that these defects were generally considerably larger than those at the pipe joints. Comparison of costs showed that joint pressure tests are 4 to 5 times more expensive than electro-scan or CCTV and the all-up cost of electro-scan is significantly less than CCTV.
Very few inspection techniques are capable of inspecting sewer pipe laterals - as a result, the contribution of aging service laterals to excessive inflow and infiltration is largely unknown. Electro Scan Inc. out of California has recently announced the availability of the ES-38TM for the inspection of sewer laterals. The product is the first in a series of next generation sewer evaluation products that can be used to identify and quantify pipe leaks in pipes that range from 3 to 8 inches in diameter.

2.7.5 Electromagnetic Inspection of Sewer Pipes

Many utilities use large diameter pre-stressed concrete cylinder pipe (PCCP) for the pressurized pipes in their sewer system. PCCP consists of a concrete core, a thin steel cylinder, high tensile pre-stressing wires and a mortar coating. The concrete core serves as the main structural load-bearing component with the steel cylinder providing the water water barrier between the concrete layers. The pre-stressing wires produce a uniform compressive pressure in the core that offset the tensile stresses in the pipe, and the mortar coating protects the pre-stressing wires from physical damage. Large diameter PCCP are a significant investment for many water and wastewater agencies. Assessing and monitoring the condition of these pipes is becoming an increasingly important and challenging task.

One of the main forms of failure in PCCP is breakage of pre-stressing wires due to corrosion. Electromagnetic inspection technology functions in a similar way to a radio transmitter and receiver - where a transmitter produces an electromagnetic field and the pre-stressing wires in the pipe amplify the signal that is recorded by the receiver. If some of the prestressing wires are broken, the signal will be distorted - measuring this distortion can therefore be used to quantify the number of broken
wires. While relatively new to the wastewater industry, electromagnetic inspection is being increasingly used to inspect the large diameter PCCP pipes that represent a significant investment for many wastewater agencies. Assessing and monitoring the condition of these pipes is becoming an increasingly important and challenging task. The results of electromagnetic testing can be used to highlight sections of a pipe needing repair and serve as a baseline condition assessment of the PCCP. These tests are usually the first step in a PCCP pipe management program - where follow-up testing with acoustic monitoring technology is used to determine the rate of deterioration of the pipe.

### 2.7.6 Acoustic Monitoring of PCCP

Acoustic monitoring systems can be used to detect the acoustic signals produced by breaking or broken pre-stressed wire within the PCCP pipes. One acoustic monitoring system available to Canadian municipalities dealing with aging PCCP forcemains is the Soundprint AFO acoustic monitoring technology manufactured by Pure Technologies. Soundprint AFO is a fiber optic monitoring system that provides continuous monitoring of the acoustic activity in PCCP to identify the acoustic events associated with the failure of pre-stressing wires. Fiber optic sensors installed inside the pipe where the entire fiber is acoustically sensitive is used to continuously track the condition of each pipe in the line and report on the time and location of breaking pre-stressing wires. This data makes it possible to identify actively deteriorating pipe sections and determine the rate of deterioration.
2.8 Comparing Utility and Cost of Sewer Pipe Inspection Methods

There are a wide variety of inspection technologies available to a municipality in this current market, but there is limited information on the suitability of the technology for field work in the urban setting. The EPA recently carried out three weeks of field-testing in Kansas City, Missouri to determine the suitability of a variety of technologies for assessing the condition of wastewater collection systems (Martel et al., 2011). A summary of the findings of the EPA field work can be found in Appendix C.

2.9 Discussion

While the EPA field test results presented in Martel et al. (2011) provides much needed information on the suitability of some innovative pipe inspection technologies for gravity sewer pipe inspection, there is still a considerable need to address the usefulness of a number of other advanced inspection technologies on pressurized sewer forcemains used to convey pumped wastewater. Camera-based technologies (i.e. zoom camera, CCTV, digital profiling, etc.) are suitable for inspecting gravity sewer pipelines but cannot effectively be used to inspect forcemains that lack the access points (manholes) required for camera inspection. Attempting to inspect forcemains using camera technology is undesirable as the process of taking the pipe out of service and draining it is time consuming and expensive. Thus, inspection of forcemains calls for alternative technology. Acoustic leak detectors are devices
used to detect the sound or vibration produced by leaks in pressurized sewer pipes. Leak detection tools are divided into two groups: those suitable for small diameter pipe and those suitable for large diameter pipe. Leak detection technology for small diameter forcemains include:

- **Acoustic Correlators** - acoustic sensors attached to pipeline appurtenances. If they are able to detect acoustic activity from a leak occurring between the sensors, the leak can be located. The ability of the sensor to detect a leak depends on the acoustic energy of the leak and the pipe material itself. Correlators tend to be effective in small-diameter pipes made of cast-iron, ductile-iron or steel - but are less effective in pipes made of concrete, PVC and larger diameter pipes. The acoustic sensors need to be moved frequently to inspect long lengths of pipeline.

- **Acoustic Data Loggers** - similar to correlators, data loggers are attached to pipeline appurtenances for a period of hours (usually overnight - when the flow in the pipe is at its lowest point and traffic activity above ground is at a minimum. Data loggers can acquire acoustic information about leaks but cannot be used to identify the location and only tend to perform well on small diameter cast-iron, ductile iron and steel pipes.

Larger diameter pipes require a sensor to be placed inside the pipe to successfully identify leaks. These in-line systems are a cross-over technology from the oil-and-gas industry that are now widely used for water distribution pipe inspection and are becoming more common for sewer pipe inspection.
• Tethered Systems - work by inserting a hydrophone attached to a cable into a pipeline. PureTechnologies’ Sahara system uses a parachute and the water flow in the pipe to deploy the hydrophone along the length of a pipe. This technology is capable of inspecting pipes with a diameter greater than 4 inches (100 mm). Because the sensor passes directly over the leak location, the sensitivity of the system is far greater than any acoustic correlators or data loggers. While the sensor has been shown to be suitable for use in the wastewater industry - survey length tends to be limited by drag on the cable, bends on the pipe and in-line appurtenances (e.g. butterfly valves). A typical survey length is approximately 1.6 kilometers - though complex pipeline configurations can significantly reduce the range of the system.

• Autonomous In-Line Devices - free-swimming, non-tethered devices capable of detecting the acoustic activity associated with leaks and pockets of trapped gas in pressurized pipelines. The presence of gas pockets in forcemains can indicate a build-up of hydrogen sulphide gas which can turn into sulphuric acid and attack the inner lining of the main.

Pure Technologies’ has recently made their SmartBall sensor available for sewer force main inspection. This free-swimming foam ball has an instrumented aluminum core capable of detecting the acoustic activity associated with leaks and pockets of trapped gas in pressurized pipelines greater than 150 mm in diameter. A typical SmartBall survey is performed as follows:

• Site review - drawings, flow characteristics and pipe configurations reviewed.

• Install tracking sensors on appurtenances to track the position of the Smart-
Ball as it rolls through the pipe.

- Install the extraction net into the flow to capture the SmartBall. In sewer forcemains, the point of capture is usually a point of discharge.

- Compress and insert the SmartBall into an operational pipeline.

- Once inside the pipe, the sensor rolls with the flow in the pipeline and obtains a high quality recording of the acoustic activity. The device also sends out ultrasonic pulses so that technicians can track the position of the ball during the survey.

- Once the SmartBall rolls into the extraction net, it is compressed and pulled up through the outlet.

- Proprietary software is then used to evaluate the data and provide a report that lists the size and location of leaks and pockets of trapped gas.

Given that a significant portion of North American forcemains have been in use for several decades but have never been assessed or pro-actively managed it is highly likely that municipalities would benefit from incorporating the equipment into inspections of existing sewer infrastructure. The condition of these pressurized forcemains needs to be immediately determined to ensure there are no locations that are failed or susceptible to failure.

2.10 Conclusion

Municipalities typically use less costly inspection methods for initial evaluations, then progress to more comprehensive and more costly techniques as warranted
(Feeney et al., 2010). For example, a municipality may start with zoom camera inspections of manholes and connecting pipes to identify pipes within their network that need more comprehensive inspection with CCTV. Video inspection through CCTV can then be used to assign a 1 - 5 internal condition grade to the pipe. Even though they have their faults, grading systems and CCTV technology play a crucial role in developing a cost estimate for maintenance and rehabilitation efforts.

While municipalities have traditionally used CCTV to visualize the condition of the many kilometres of pipelines beneath their streets, many are using advanced inspection technology like digital scanning in their fight against sewer deterioration. Some inspection technologies are not aimed at replacing CCTV, but instead complement the data collected through visual inspection. Innovative technologies like sonar, laser scanning, and electro-scanning are becoming more commonplace in the industry and make a more well-rounded approach to sewer condition assessment possible. While North American municipalities are beginning to implement comprehensive condition assessment practice for gravity mains, condition assessment of forcemains has been minimal at best.
Transition to Chapter 3

This next chapter begins the investigation into leveraging information from existing sanitary sewer pipe inspection datasets using data mining tools and techniques. A decision tree classifier model is developed using a set of basic pipe attributes that many municipalities have readily accessible within an existing linear asset inventory: trunk sewer (yes or no), material of construction, year of installation, diameter, length, slope and burial depth. The model is found to be a useful tool for knowledge discovery. It has been accepted for publication in the peer-reviewed, edited volume Conflict Resolution in Water Resources and Environmental Management:

Chapter 3

A Data Mining Tool for Planning
Sanitary Sewer Condition
Inspection

3.1 Introduction

Approximately 800,000 miles of sewer pipe are currently in operation across the United States of America (ASCE, 2013). The majority of this infrastructure is approaching the end of its useful design life and it appears decades of neglect have taken a toll on this important national capital asset. Aging sewers discharge an estimated 900 billion gallons of untreated wastewater into American waterways each year (ASCE, 2013). These sewers often carry pathogenic microorganisms, industrial toxins, and endocrine disrupting compounds capable of causing immediate and long-term damage to the environment. As such, leaking sewers place decades
of progress in public health and environmental protection at risk (EPA, 2002).

Many municipalities rely on sewers that were installed more than 75 years ago to transport their wastewater. Pipes in these systems were initially designed to serve populations half their current size and continued population growth is expected to stress pipes beyond their design capacity and accelerate the process of pipe deterioration over time. It is expected that an investment of $390 billion will be required from 2002-2022 to rehabilitate defective American sewers and expand existing systems to meet increasing demand (EPA, 2002).

The threat posed by deteriorating sewers is typically determined using closed-circuit television (CCTV) inspection, where a small camera mounted on a robot is driven along the length of a pipe (Figure 3.1). Defects inside the pipe (e.g. cracks, fractures and collapsed sections) are identified and assessed using systems designed to minimize subjective evaluation of the camera footage. Each inspected pipe is typically assigned an ordinal grade to reflect its structural condition (e.g. 1: no defects, 2, 3, 4 or 5: collapsed or collapse imminent).

CCTV inspections are time-consuming and expensive, posing an average cost to a small municipality of $0.84 per linear foot of inspected pipe (EPA, 2010). Budgetary restrictions may limit inspection to small portions of an entire sewer system. Consequently, the threat posed by individual pipes that have not yet been inspected remains unquantified, creating a potentially dangerous and costly failure scenario for the municipality. As an example, the sudden failure of a 35-year old, 1.1 meter diameter sewer pipe in Tucson, Arizona caused two large sink holes on a major five lane roadway. The county was served with 11 environmental violations as an estimated 50 million gallons of raw sewage flowed into the nearby Santa Cruz River.
Figure 3.1: Inspections are typically carried out using robots equipped with cameras that are inserted into a pipe and remotely operated from the road surface. (Carlson and Urquhart, 2006). A systematic approach to predicting the condition of uninspected pipes is critical for the efficient use of limited financial resources. Sewer deterioration rates are highly dependent on a variety of factors (e.g. original design, material of construction, environmental conditions, extreme loading events, and asset age), which serves to make the characterization of time-dependent deterioration of individual pipes a challenging task. As a result, the majority of sewer deterioration models currently available focus on the provision of generalized estimates of condition at the system/network level. Models capable of predicting individual sewer pipe condition are still in their infancy as few studies have been devoted to model development and even fewer to model validation (Kley and Caradot, 2013). Hence, a framework is proposed for implementing an interpretable data mining system that is capable of learning from existing inspection datasets so that reliable conditions of prediction can be made for individual pipes in a sewer.
system that have not yet been inspected. Open-source software is utilized, making the data mining approach accessible to municipalities, regardless of pre-existing financial circumstance. The predictive capabilities of the model are illustrated using a case study representative of condition data typically available after a municipality inspects a portion of their sewer system. The modeling approach relies on a set of basic pipe-specific attributes when predicting pipe condition, ensuring it can be easily reproduced by other municipalities.

3.2 An Overview of Sewer Condition Modeling

Proactive management of aging sewer pipes can be supported by the use of pipe deterioration modeling strategies, the majority of which have their foundations in statistical theory (Ana and Bauwens, 2010). Markov models developed for predicting the deterioration of stormwater pipes in Australia were not initially designed with the intention of predicting individual pipe condition (Micevski et al., 2002). A Markov process developed using inspections from more than 5,000 concrete sewer pipes in Germany successfully predicted the condition of individual pipes, but it was conceded that the approach involved heavy and cumbersome data manipulations (LeGat, 2006). Markov models and ordinal regression models developed using 27 kilometers of stormwater pipe condition in Australia were only useful at the system-level (predictive accuracy of the ordinal regression model for individual pipes was 42%) (Tran et al., 2008). A Markov process proposed for evaluating the suitability of condition assessment technologies for water pipes in Hamilton, Ontario was not intended for predicting pipe condition (Atef et al., 2012). The predictive
capabilities of binary logistic regression to categorize the condition of PVC pipe segments in Phoenix, Arizona were not provided at the individual-pipe level (Koo and Ariaratnam, 2006). A variety of regression models developed to predict the deterioration of sewers in Cincinnati, Ohio were either unsuitable (e.g. the available data violated assumptions necessary to perform ordinal regression) or achieved less than desirable levels of performance (e.g. 46% of poor structural condition pipes were correctly predicted using a binary logistic regression model) (Salman, 2010). Ordinal regression models provided insight into sewer pipe deterioration in Niagara Falls, Ontario but the predictive capabilities of the modeling approach were not provided at the individual pipe level (Younis and Knight, 2010). Survival analysis models were used to predict structural condition at the system-level for cohorts of sewer pipes in Quebec City, Quebec (Duchesne et al., 2012).

Predictive models have also been developed using data mining techniques drawn from the intersection of statistics, artificial intelligence and machine learning. These techniques are capable of extracting information from large inspection datasets so that predictions of pipe condition can be obtained without the parametric assumptions that tend to restrict the utility of statistical approaches. Neural networks were superior to Markov models and ordinal regression when predicting the deterioration of stormwater pipes in Australia (Tran, 2007). Neural networks were also used to investigate sewer pipe deterioration in Pierrefonds, Quebec (Khan et al., 2010). Support vector machines successfully predicted the structural condition of individual sewer pipes in Australia (Mashford et al., 2011).

The literature generally indicates no single modeling strategy works best for all predictive modeling tasks, a phenomenon that can be explained by the no free
lunch theorem of optimization (Wolpert and Macready, 1997). Although some approaches have provided high levels of predictive performance for some municipalities, predictions of condition made for another municipality using the same modeling strategy have been poor. As an example, Markov, survival, regression and neural network models were all found incapable of reliably predicting individual pipe condition in two Belgian municipalities (e.g. the neural networks often predicted pipe condition would improve over time, which would be impossible without some sort of rehabilitation) (Ana, 2009).

Many municipalities will be unable to devote the development time necessary for a modelling strategy requiring extensive data manipulation, pre-processing and assumption making/validation. Municipalities need access to efficient alternatives capable of learning from existing inspection datasets so that location-specific pipe-level models can be obtained in a relatively short-period of time. Hence, an efficiently implemented decision tree modeling strategy is described for predicting individual pipe condition. In general terms, decision tree algorithms learn from an existing dataset so that a flowchart-like tree structure can be obtained that illustrates relationships between input-predictors and a target class embedded within a dataset (Figure 3.2). The resulting image of the tree is interpretable, even by those unfamiliar with the data mining process (unlike neural networks, which are black boxes). The topmost node, called the root of the tree, contains all the instances/observations contained in a dataset used for training. Each internal node specifies a test on an input predictor and each branch in the tree represents an outcome of the test. An instance is classified by starting at the root of the tree and then, depending on the values of the attributes, tracing a path down through the
branches of the tree. Eventually, a leaf node is reached at the bottom of the tree and a classification can be obtained using the distribution of instances observed in the leaf.

Will the baseball game be played today - Yes or No?

![Decision Tree Diagram]

**Figure 3.2:** A decision tree predictive model illustrates the knowledge gained from a hypothetical dataset in a tree-like format. In this example, the values of the input predictors (weather and wind speed) can be used to determine the likelihood of a game being played.

Decision trees are among the most popular data mining systems used within the fields of financial analysis, medicine and marketing. They have previously been used to model sewer collapse rates (Heywood et al., 2007) and to identify important attributes associated with high-density regions of defective pipes in a mid-sized city (Jung et al., 2012). These models effectively handle a wide variety of continuous data, categorical data, sparse data and skewed data and require very little in the way of data pre-processing and manipulation (unlike traditional statistical
techniques). Feature selection (where the most useful input predictors are used to make predictions) is also implicitly conducted as part of the algorithm.

3.3 Data Mining Methodology

In data mining terms, a decision tree classifier needs to be trained using each instance/inspected pipe in the dataset so that a relationship can be made between input predictors and the target condition class. The input predictors are the pipe-specific attributes obtained from a facilities inventory maintained in an asset management system such as GIS or through field-investigation (e.g. each pipe in the inspection dataset has a unique material of construction, age, diameter and length). The initial targets are condition grades assigned to each inspected pipe (e.g. the Water Research Council Sewer Rehabilitation Manual (WRc SRM) internal condition grades (ICG) of 1, 2, 3, 4 or 5) (WRc, 2001).

Inspection datasets tend to be imbalanced, with significantly more pipes in some condition classes than others. As a hypothetical example, inspecting 1,000 pipes may reveal 550 pipes in excellent condition (ICG 1), 200 in good condition (ICG 2), 150 in fair condition (ICG 3), 75 in bad condition (ICG 4) and only 25 that have failed (ICG 5). This imbalance is the natural result of collecting condition data over a limited time period (if inspections are performed over 5 years, any pipes in the sewer system prone to failure will have already failed and been replaced prior to inspection). Traditional data mining systems aim to minimize the total number of errors made during classification and tend to be most effective when there is a balanced distribution of observations across all classes in the dataset.
Consequently, predictive models developed using imbalanced inspection datasets will be biased towards the majority classes (i.e. ICG 1 and 2) and will show poor classification rates for minority classes of interest (i.e. ICG 3, 4 and 5). The majority classes dominate the learning process and the resulting model will almost always predict a pipe belongs to the majority classes ICG 1 or 2.

Improving the classification accuracy of minority classes within an imbalanced multi-class dataset is a difficult task and remains an active area of research within the data mining community (Han et al., 2006). The majority of class-imbalance learning techniques currently available for implementation have been designed for two-class problems, necessitating transforming pipe condition into a two-class format. The transformation of pipe condition is typically the choice of the municipality and will reflect any local intentions to rehabilitate structurally damaged pipes. One option would be to assign ICG of 1-2 to a *good* condition class as these pipes have no defects or contain very few defects of concern. Pipes with an ICG of 3-5 can be assigned to a *poor* condition class as their structural defects pose a potential threat to the integrity of the sewer system and they would be potential candidates for rehabilitation. Although the distribution of instances across classes may still be imbalanced after this recoding of pipe condition (for the previously mentioned hypothetical 1,000 inspections, 75% of the pipes would be *good* and 25% would be *poor*), the binary format reduces the imbalance and facilities implementation of strategies designed to improve the ability to correctly identify bad pipes representing the positive, minority class of interest.

After transformation into a two-class format, the dataset is subdivided into training, evaluation and test sets using a 70-10-20 splitting ratio. The training set is
used to train the model, the evaluation set is used during model tuning and the test set is used to provide an unbiased indication of the predictive capabilities of the model on previously unseen observations. Stratified random sampling should be used to ensure the class distribution in the three sets match (Kuhn and Johnson, 2013).

A variety of techniques are currently available for constructing decision tree classifiers. One of the oldest and most widely used is the classification and regression tree (CART) methodology developed in the mid-1980s (Breiman et al., 1984). CART models in this research were trained and tuned using the caret (Kuhn and Johnson, 2013) and rpart (Therneau et al., 2014) packages developed for the open-source R software environment. The CART methodology constructs decision tree classifiers using a top-down divide-and-conquer approach - where instances of pipe condition in the training dataset are partitioned into small, homogeneous groups that are pure (with a larger proportion of one class than another).

*Author’s note - refer to Appendix D for a sample of the R code that can be used to train, tune and evaluate a decision tree classifier.*

Optimal splits for numeric input predictors (e.g. a pipe’s length in meters) are determined by sorting all the available instances in the training set based on their predictor value. Any categorical input predictors considered for splitting (e.g. pipe material of construction) are first decomposed into binary dummy variables prior to their presentation to the CART algorithm, forcing binary splits of the categories. The potential split points are then the midpoints between each unique predictor value and a contingency table is then generated for each potential split point (Table 3.1):
The purity of a potential split is determined using the Gini index, which can be calculated before and after a split for a two-class problem using the equations:

\[ Gini\ (prior\ to\ split) = 2 \left( \frac{n_{1+}}{n} \right) \left( \frac{n_{2+}}{n} \right) \]  (3.1)

\[ Gini\ (after\ the\ split) = 2 \left[ \left( \frac{n_{11}}{n_{1+}} \right) \left( \frac{n_{12}}{n_{1+}} \right) + \left( \frac{n_{21}}{n_{2+}} \right) \left( \frac{n_{22}}{n_{2+}} \right) \right] \]  (3.2)

The CART algorithm evaluates all the potential split points and seeks out the one that minimizes the Gini purity criterion as it results in the least amount of randomness/impurity out of all the possible splits. This evaluation process then continues within each newly created partition of the tree and the tree grows larger as new splits are created. Trees may eventually grow so large that the splits being added only reflect anomalies in the training data due to noise. Unreliable branches can be removed from the fully-grown tree using a complexity parameter that is a function of the number of leaves in the tree and the error rate (the percentage of instances misclassified by the tree):

\[ Accuracy_{cp} = Accuracy + (c_p) (number\ of\ leaves) \]  (3.3)
The algorithm uses the complexity parameter when evaluating the impact on predictive performance when various sub-trees at a node are pruned away and the node itself is replaced by a terminal leaf (Kuhn and Johnson, 2013). Given a particular value of the complexity parameter, the algorithm seeks out the tree having the best predictive performance on cross-validation partitions of the training set. A complexity parameter of zero results in a fully grown tree (potentially weak predictive capabilities) and values close to one might result in a tree with only one split (limited utility for knowledge extraction purposes).

The evaluation and testing datasets can be used to gauge the predictive capabilities of a pruned decision tree model. The leaf nodes in the tree are used to generate a continuous-valued class membership probability between 0 and 1 established based on the distribution of condition classes at that leaf. By default, the tree uses a probability threshold of 50% to assign a class label to any instances that reach that leaf. As an example, an individual pipe with attributes that lead it to a leaf in the tree with a distribution of 60% good pipes and 40% poor pipes, then the tree will predict that pipe belongs to the good condition class. These discrete class predictions can then be used to evaluate the performance of the model on a dataset with known class labels using the confusion matrix shown in Table 2.

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor (ICG 3-5)</td>
<td>Poor (ICG 3-5)</td>
</tr>
<tr>
<td>True Positive (TP)</td>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>Good (ICG 1-2)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td></td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

Table 3.2: Confusion matrix for a binary classification task
When pipes in poor condition are considered the positive minority class of interest the following outcomes are shown in the confusion matrix:

- **True Positive** = a pipe known to be in poor condition correctly predicted to be in poor condition by the decision tree.
- **False positive** = a pipe known to be in good condition incorrectly predicted to be in poor condition.
- **True negative** = a pipe that is actually in good condition correctly predicted to be in good condition.
- **False negative** = a pipe that is actually in poor condition incorrectly predicted to be in good condition.

Ideally, there will be very few false positives and false negatives (i.e. highly accurate). Accuracy may not provide a reliable indicator of predictor performance for models trained using imbalanced datasets as it may provide a false impression of capabilities for the minority class of interest. Considering the hypothetical dataset of 750 pipes in good condition and 250 pipes in poor condition, assigning every pipe to the good condition class would achieve an accuracy of 75%. Accuracy assumes the costs of false positive and false negative errors are the same. In reality false negatives have a much higher cost than false positives when predicting sewer condition. If an uninspected pipe was leaking raw sewage into the ground but the model predicted the pipe was actually in good condition (and therefore did not need to be inspected) there could be serious environmental and economic consequences. As a result, a series of alternative metrics should be used understand model utility when working with imbalanced datasets:
True Positive Rate = sensitivity = \frac{TP}{TP + FN} \quad (3.4)

True Negative Rate = specificity = \frac{TN}{FP + TN} \quad (3.5)

False Positive Rate = 1 - TNR = \frac{FP}{FP + TN} \quad (3.6)

False Negative Rate = 1 - TPR = \frac{FN}{TP + FN} \quad (3.7)

Class imbalance inevitably results in a trade-off between the true positive rate and the false positive rate. A useful tool for evaluating this trade-off is the receiver-operating characteristic (ROC) curve (Figure 3.3). The ROC curve, first used during the Second World War to help radar operators correctly distinguish enemy targets from noise on their screens, is a plot of the true positive rate versus the false positive rate achieved when different probability thresholds are used to establish classifications for instances in a dataset. For each candidate threshold (e.g., 50%) the true positive rate and false positive rate are plotted against each other. The area under the ROC curve can be used to indicate model performance. Perfect models have an area under the ROC curve of 1 and random models have an area under the ROC curve close to 0.5 (Fawcett, 2006). As a rough guide, an area under the ROC greater than 0.7 on a stratified test set would be considered acceptable (Hosmer and Lemeshow, 2000).
A threshold-moving approach can be implemented to improve the identification capability for poor condition pipes representing the minority class of interest. This technique involves moving the classification threshold away from the baseline of 50% so that minority classes are more frequently predicted. For example, a leaf in the tree may contain 35% poor condition pipes and 65% good pipes and any instance that reaches that leaf will be classified as being good. Changing the threshold down from 50% to 30% would then mean any instance that reaches that leaf will be classified as being in poor condition to reduce the likelihood of a costly false negative error. A new, optimal cutoff can be determined by finding the point on an ROC curve developed for the evaluation dataset that is closest to the top left corner as it is closest to the perfect model. This new threshold can then be used to reclassify pipes in the test dataset. Threshold moving has been empirically shown to outperform some of the other commonly used techniques to accommodate data imbalance (e.g. oversampling and under-sampling) (Han et al., 2006). In addition to simplicity, the approach is advantageous as it does not alter the original tree structure of the trained model.

3.4 Case Study

The City of Guelph is located in southwestern Ontario, Canada. The 120,000 residents of the municipality are served by a 515 kilometer long sanitary sewer system composed of more than 7,000 gravity sewer pipes.

A total of 221 kilometers of Guelph’s gravity sewer pipes were CCTV inspected from 2009 to 2011 by AECOM Canada Ltd. These inspections were carried out
Figure 3.3: The receiver operating characteristic (ROC) curve can be used to evaluate the trade-offs between true positives and false positives.
using the third edition of the Water Research Council Manual of Sewer Condition Classification (WRc MSCC) (WRc, 1996). Defects observed inside each pipe were then assigned severity scores using the fourth edition of the Water Research Council Sewerage Rehabilitation Manual (WRc SRM). The peak score (highest defect value accumulated in any one meter length of a pipe) was used to assign an internal condition grade of 1 - 5 to each inspected pipe based on peak score threshold guidelines contained within the WRc SRM.

Comprehensive quality assurance/quality controlled was carried out by AECOM to ensure accuracy of the CCTV inspection records. The modeling dataset used for data mining purposes in this research consisted of 1,825 pipes (123 km) inspected by the primary sub-contractor (Table 3.3). There are significantly more pipes in ICG 1 than in classes 2-5 for the algorithm to learn from. A binary reclassification of condition (ICG 1-2 vs. 3-5) results in a 3:1 ratio of good to poor condition pipes. Splitting the data using a 70-10-20 division created 1278 instances (325 poor and 953 good condition), 183 instances for evaluation (47 poor and 136 good) and 364 instances for testing (92 poor and 272 good). The following input predictors were considered during the process of training the decision tree: trunk sewer (categorical: yes or no), material of construction (categorical: asbestos cement, concrete, PVC, reinforced concrete, or vitrified clay), year of installation (numeric: 1902 - 2008), diameter (numeric: 200 - 900 mm), length (numeric: 3 - 200 m), slope (5 - 16 m per 100 m), and burial depth (numeric: 0.8 - 8.8 m).
Table 3.3: The inspection dataset used for model development

<table>
<thead>
<tr>
<th>Internal Condition Grade (ICG)</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1    2    3    4    5</td>
<td>Good (ICG 1-2)</td>
</tr>
<tr>
<td>Asbestos Cement</td>
<td>373 14 24 5 0</td>
</tr>
<tr>
<td>Concrete</td>
<td>364 79 73 36 7</td>
</tr>
<tr>
<td>PVC</td>
<td>110 9 1 2 1</td>
</tr>
<tr>
<td>Reinforced Conc.</td>
<td>56 8 0 0 0</td>
</tr>
<tr>
<td>Vitrified Clay</td>
<td>213 135 165 121 29</td>
</tr>
<tr>
<td>Total</td>
<td>1116 245 263 164 37</td>
</tr>
</tbody>
</table>

3.5 Results

The decision tree developed using the training data is shown in Figure 3.4. Tree-pruning was carried out using an optimal complexity parameter of 0.0136 determined using caret (maximum ten-fold cross-validation area under the ROC curve = 0.76). The model achieved an area under the ROC curve of 0.77 for both the evaluation and test sets, which can be considered an acceptable result for a binary classification task. The test set confusion matrix developed for the default probability threshold of 50% is given in Table 3.4. The table indicates the default threshold needed tuning to account for class imbalance. Although accuracy is high (0.79), the high FNR (0.49) indicates many poor condition pipes are misclassified.
The root of the decision tree contains 100% of the inspected pipes in the training set (325 “poor” condition pipes & 953 “good” condition pipes).

The leaves provide insight into sewer deterioration. For example, Leaf 7 indicates a long, small diameter pipe that is more than 50 years old, has a 64% chance of being in “poor” condition. The numbers at the bottom of Leaf 7 indicate that this 64% chance is based on the condition of 216 pipes in the training set.

Figure 3.4: The CART model for predicting sewer pipe condition.
Table 3.4: Test Set Confusion Matrix for the Default Classification Threshold (50%)

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Poor (ICG 3-5)</th>
<th>Good (ICG 1-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor (ICG 3-5)</td>
<td>48</td>
<td>44</td>
</tr>
<tr>
<td>Good (ICG 1-2)</td>
<td>33</td>
<td>239</td>
</tr>
</tbody>
</table>

Accuracy = \( \frac{48 + 239}{48 + 44 + 33 + 239} = 0.79 \)

TPR = \( \frac{48}{48 + 44} = 0.52 \) (correctly classified poor pipe)

FPR = \( \frac{33}{33 + 239} = 0.12 \) (incorrectly classified good pipe)

TNR = \( \frac{239}{33 + 239} = 0.88 \) (correctly classified good pipe)

FNR = \( \frac{44}{48 + 44} = 0.48 \) (incorrectly classified poor pipe)

Area under the ROC curve = 0.77

The evaluation set ROC curve (Figure 3.5) was used to identify a new, optimal threshold of 32.4% (closest to the top left corner, therefore closest to a perfect model). The test set confusion matrix shown using this new threshold is shown in Table 3.5. Although this new threshold does slightly reduce the overall accuracy of the model from 0.79 to 0.76, the false negative rate drops significantly from 0.48 to 0.32 (with the consequences that the false positive rate slightly increased from 0.12 to 0.22). A municipality could also choose to tune using a more severe classification threshold to further reduce the likelihood of a false negative. As an example, a threshold of 20% would result in an overall accuracy of 0.68 but would achieve a false negative rate of 0.21 and a false positive rate of 0.36 (Table 3.6). The desirability of this alternative threshold is dependent on the municipality’s desire to reduce the risk of missing poor condition pipes.
Figure 3.5: The evaluation set receiver operating characteristic (ROC) curve indicates an alternative threshold could reduce the false negative rate of the tree.
### Table 3.5: Test Set Confusion Matrix for Optimal Classification Threshold (32%)

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Poor (ICG 3-5)</th>
<th>Good (ICG 1-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor (ICG 3-5)</td>
<td>63</td>
<td>29</td>
</tr>
<tr>
<td>Good (ICG 1-2)</td>
<td>59</td>
<td>213</td>
</tr>
</tbody>
</table>

Accuracy = (63 + 213) / (63 + 29 + 59 + 213) = 0.76

TPR = (63) / (63 + 29) = 0.68 (correctly classified *poor* pipe)

FPR = (59) / (59 + 213) = 0.22 (incorrectly classified *good* pipe)

TNR = (213) / (59 + 213) = 0.78 (correctly classified *good* pipe)

FNR = (29) / (63 + 29) = 0.32 (incorrectly classified *poor* pipe)

Area under the ROC curve = 0.77

### Table 3.6: Test Set Confusion Matrix for a Classification Threshold of 20%

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Poor (ICG 3-5)</th>
<th>Good (ICG 1-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor (ICG 3-5)</td>
<td>73</td>
<td>19</td>
</tr>
<tr>
<td>Good (ICG 1-2)</td>
<td>97</td>
<td>175</td>
</tr>
</tbody>
</table>

Accuracy = (73 + 175) / (73 + 19 + 97 + 175) = 0.68

TPR = (73) / (73 + 19) = 0.79 (correctly classified *poor* pipe)

FPR = (97) / (97 + 175) = 0.36 (incorrectly classified *good* pipe)

TNR = (175) / (97 + 175) = 0.74 (correctly classified *good* pipe)

FNR = (19) / (73 + 19) = 0.21 (incorrectly classified *poor* pipe)

Area under the ROC curve = 0.77
3.6 Discussion

The branches and leaves of the tree provide insight into local pipe deterioration processes occurring in the municipality. The first major split of the tree partitions pipes in the training set according to their year of installation. In Guelph, 91% of the pipes installed after 1966 (i.e. less than 50 years old) are in good condition (leaf 3). Although there are some newer pipes in poor condition, no further splits after leaf 3 would increase purity of the dataset. Alternatively, 47% of all pipes in the training dataset installed prior to 1966 are in poor condition. Nodes 4 and 5 of the tree indicate installation year can further subdivide the dataset as 33% of pipes installed between 1956 and 1966 are in poor condition (node 5) compared to the 55% of the pipes installed prior to 1956 that are in poor condition (node 4).

Pipes constructed between 1956 and 1966 having a slope greater than 2m/100m have a 67% chance of being in poor condition (leaf 10) compared to a 28% chance of being in poor condition when slope is less than 2m/100m (leaf 11). A possible explanation for this phenomenon is that pipe sections with steep slopes tend to have higher deterioration rates as flows are generally faster in these steep sections, causing erosion and surface wear inside the pipe.

The influence of pipe length on the condition of pipes installed prior to 1956 is shown in node 8 and leaf 9. It has been suggested in the literature that longer sections of pipe have more joints and tend to deteriorate at faster rates than shorter ones (Ana and Bauwens, 2010). Node 6 indicates pipes with a length greater than 53 m have a 60% likelihood of being in poor condition, which is significantly higher than the 37% likelihood for pipes with a length less than 53 m (leaf 9).
Pipe diameter is also found to be related to pipe structural condition. Leaf 7 indicates there is a 64% chance of being in poor condition if the diameter of older, longer pipes is less than 238 mm. Pipes with a diameter greater than 238 mm have only a 42% chance of being in poor condition (leaf 8). One possible explanation for the reduced likelihood of larger diameter pipes being in poor condition could be the tendency of larger diameter pipes to be installed by experienced personnel, reducing the likelihood of defects related to installation error (Davies et al., 2001).

### 3.6.1 Planning Future Inspections

The City of Guelph has a significant portion of their sanitary sewer system that has not yet been inspected (although the majority of these pipes are newer PVC pipes installed within the past 50 years, approximately 800 pipes were installed prior to 1966). The unique attributes of the uninspected pipes can be presented to the decision tree and the predicted probability of being in poor condition can be used to identify pipes that should be scheduled for the next round of inspections. Using the optimal probability threshold of 32%, a total of 624 pipes are predicted to be in poor condition.

A municipality may not have the time or the budget required to immediately inspect every pipe identified by the decision tree as being in poor condition. An alternative to the immediate inspection of all poor predicted pipes would be the selection of a smaller subset of pipes that are of concern due to their proximity to other deteriorated municipal infrastructure. A number of recent studies indicate defective stormwater pipes are a conduit for raw sewage when leaking sanitary pipes are located in close proximity (Doshi, 2012) (Sercu et al., 2011). Raw sewage
passing untreated from broken sewer pipes into broken stormwater pipes has the potential to cause wide-spread contamination of municipal waterways and beaches. The City of Guelph inspected one-third of their stormwater system from 2008 to 2011. Of the 2,460 inspected stormwater pipes, 903 were found to be in poor structural condition with an ICG of 3-5. Of the 624 sanitary sewer pipes predicted by the decision tree to be in poor condition, 25 have a probability of being in poor condition of at least 50% and are located within one meter of an ICG 3-5 storm-water pipe (Figure 3.6). Securing the budgetary allocation for the immediate inspection of 25 pipes should prove to be an easier task than inspecting 624 pipes.

3.6.2 Future Directions of Research

The CART predictive model achieved an acceptable area under the ROC curve of 0.76, but improvements in predictive performance should be pursued in future research. Advanced algorithms such as random forests may increase the capability of a predictive model to accurately identify pipes in poor structural condition. Random forests use the combined predictions of hundreds of trained CART models to make predictions and have been proven to outperform single decision trees on a variety of data mining tasks (Kuhn and Johnson, 2013).

An alternative encoding can represent pipe condition for municipalities seeking alternative predictions of condition. It may be beneficial to encode pipes as being in either ICG 123 vs. 45 as pipes with an ICG $\geq 4$ have serious defects and can be expected to collapse or will collapse within a short time frame. One consequence of this alternative encoding will be more severe class imbalance.
Figure 3.6: Sanitary sewer pipes predicted to be in poor condition by the decision tree model that are also located in close proximity to a defective stormwater pipe are candidates for immediate inspection.
While the suggestion of targeting pipes for inspection based on their proximity to poor condition stormwater pipes is novel, a more elaborate risk of failure scenarios would allow municipalities to further assess the suitability of candidate pipes for inspection. As an example, there would be utility in assessing the consequences of failure associated when pipes are located within clusters of poor condition pipes identified through hotspot analysis of network infrastructure.

3.7 Conclusion

Sewer pipe deterioration is a major concern as broken pipes pose a threat to the environment and to the financial well-being of municipalities. There have been very few validated predictive models made available to municipalities across North America that are capable of identifying the location of individual, poor condition pipes in a sewer system. The popular classification and regression tree (CART) system was used to construct a predictive model capable of identifying individual pipes in a sewer system that are most likely to be in poor structural condition. Imbalance common within inspection dataset was accommodated by using a technique of adjusting the baseline probability threshold used by the CART model when making binary classifications of pipe condition.

The City of Guelph, Ontario inspected a portion of their sanitary sewer system from 2008-2011. The CART algorithm extracted information from Guelph’s existing record of CCTV inspection. The model provides a visual guide to the influence of pipe-specific attributes on structural condition. The majority of sanitary sewer pipes built after 1966 are in good structural condition as they have not yet reached
the end of their 50-year design life. The tendency for steeper slopes to be in poor structural condition is depicted for pipes installed between 1956 and 1966. Pipes installed prior to 1956 with a length greater than 53 m have a 55% chance of being in poor condition, compared to a 37% chance for shorter pipes. Older pipes with lengths greater than 53 m have a 64% chance of being in poor condition if they have diameters less than 238 mm, compared to a 42% chance for larger diameter pipes.

A significant portion of the Guelph sanitary system has not yet been inspected. The model predicts 624 of these pipes are in poor condition. A subset of these pipes can be scheduled for immediate inspection using their proximity to other failed infrastructure. In Guelph, 25 of the 624 pipes are within one meter of an inspected stormwater pipe that was already found to be in a failed condition state. It is possible that sewage is leaking from these 25 sanitary sewer pipes and then entering into the stormwater system, where it inevitably contaminates outfall areas (i.e. rivers, lakes and beaches).

The data mining methodology was designed with an intention for efficient implementation. Localized predictive models can be developed for any municipality with inspection data in a similar format to Guelph’s. These predictive models can enhance a municipality’s understanding of life-cycle degradation of sanitary sewer pipes and can serve as a consensus-building tool for the allocation of municipal funds towards future pipe inspection.
Transition to Chapter 4

This next chapter further investigates the suitability of decision tree classifiers for predicting sewer pipe condition. Instead of predicting a condition of *good* (ICG 1-2) vs. *poor* (ICG 3-5) a switch is made to focus on pipes that pose a greater threat to system integrity. As such, individual pipes are now considered to be in either *good* (ICG 1-3) or *bad* (ICG 4-5) structural condition.

The predictive capability of decision trees compared to support vector machines is demonstrated using a case study of sanitary sewer pipe inspection data collected by the municipality of Guelph, Ontario, Canada. An expanded set of candidate predictor inputs is considered for model development (with invert elevation, road coverage and the presence of nearby watermain breaks added to the inputs introduced in the previous chapter). These additional attributes were developed using tools for spatial analytics within GIS (with an intention to identify the possible influence of surrounding infrastructure on sewer pipe condition).

A novel analysis of the location of structural defects indicates future pipe inspection efforts in Guelph may benefit from zoom-camera inspection technologies that are faster and cheaper than traditional CCTV.

The decision tree classifier outperforms support vector machines for this classification task - achieving an acceptable area under the receiver operating characteristic curve of 0.77 and an overall accuracy of 76% on a stratified test set.
A novel application of gain chart analysis is described for screening pipes for future inspection. A sensitivity analysis indicates reliable models can be developed using datasets significantly smaller than those that were available in the case study area.

This paper has been accepted for publication in the Journal of Hydroinformatics:

Chapter 4

Comparing the Utility of Decision Trees and Support Vector Machines when Planning Inspections of Linear Sewer Infrastructure

4.1 Introduction

Sewers play a crucial role in the day-to-day operations of modern cities as they convey enormous volumes of wastewater to treatment centres for safe processing and disposal. Individual pipes within a sewer system often remain in operation
well beyond their typical design life of 50-75 years and a growing list of evidence suggests unchecked deterioration of this infrastructure is causing significant harm to the natural environment.

The cracks, holes and fractures that form in aging sewer pipes allow raw, untreated wastewater to leak/exfiltrate into the surrounding environment. As an example, the exfiltration rate for the approximately 850,000 km of sewer pipe currently in operation in Germany is reportedly 6% of the average daily flow, equating to 300 million cubic meters of leakage per year (Scheyer et al., 1991). Exfiltration rates in the United Kingdom have been reported to be in the range of 1 - 5% of average daily flow (Anderson et al., 1995; Bishop et al., 1998; Yang et al., 1999; Ellis et al., 2004). Exfiltration rates in North America are also reportedly high as a study carried out by Amick and Burgess (2000) indicates 56% of the average flow was leaking out of sewer pipes studied in California, 49% in Baltimore, 35% in Kentucky and 16% in Washington.

Leaking sewers pose a significant environmental threat as a range of contaminants can be found in typical sewer flows (e.g. bacteria and pathogenic microorganisms) and it has generally been concluded that the poor quality of urban groundwater aquifers in the United Kingdom is due to elevated rates of sewer exfiltration (Price and Reed, 1989; Reynolds and Barrett, 2003). In a study of sandstone aquifers underlying the cities of Birmingham and Nottingham in the United Kingdom, sewage-derived contaminants were found capable of reaching depths of 60-91 m (Powell et al., 2003). These findings were supported by the work of Morris et al. (2005) where faecal indicators were detected in an aquifer 60 m below ground level. Leaking sewers were the source of X-ray contrast media and boron in groundwater
studied in Wolf et al. (2004) and the presence of viruses in 33 municipal supply wells studied in Hunt et al. (2010) was attributed to the close proximity of each well to aging sanitary sewer pipes. Approximately 90% of 22 drinking water wells recently studied in Guelph, Ontario, Canada contained at least one sewage-derived contaminant and 45% of the wells exhibited human enteric viruses derived from the exfiltration of domestic sewage flows (Allen, 2013).

Infiltration of groundwater into sanitary sewer pipes through defective joints, cracks and fractures is also cause for concern as these pipes are only designed to convey wastewater. Excess amounts of infiltration can overload the sanitary sewer system during periods of wet weather, thereby forcing a system overflow. Sanitary sewer overflows (SSOs) are a widespread problem in North America, with an estimated 75,000 occurring every year in the United States, resulting in an annual discharge of several billion gallons of untreated wastewater into the environment (EPA, 2004). An increase in wastewater treatment costs will also be directly related to the amount of water that enters a sanitary sewer system as a result of infiltration. The pumps involved with wastewater treatment will need to work harder to handle the increased load which puts unneeded strain on these expensive pumps and shortens their life expectancy. Infiltration is also known to increase the failure probability of adjacent infrastructure, such as paved roads (Kuo et al., 2005; Karpf and Krebs, 2011). The potential damage from the collapse of an aging sewer pipe can therefore be significant, as in addition to environmental pollution, the collapse may also cause severe interruptions to service and traffic. Furthermore, an unexpected sewer pipe collapse can result in expensive emergency repairs having an average unit cost 3.6 times higher than the average unit cost of non-emergency rehabilitation (Zhao and
The general consensus among the wastewater industry is that an "out-of-sight out-of-mind" approach to sewer management only serves to accelerate the inevitable deterioration of sewer pipes and exacerbates exfiltration and infiltration-related issues. For example, the portion of American sewer pipes in “poor,” “very poor,” and “life-elapsed” condition has been projected to increase from 10% of total system size in 2000 to 44% by 2020 if existing sewers are extended to meet increased population growth but there is no renewal or replacement of the existing pipes (EPA, 2002). Alternatively, a more proactive approach to management based on assembling a dataset of pipe condition using visual inspection techniques can help maintain the effectiveness of existing sanitary sewer systems.

Closed-circuit television (CCTV) inspection can be used to visually investigate the condition of individual pipes in a sanitary sewer system. Although CCTV inspections provide valuable information required for rehabilitation planning purposes, they are time-consuming and expensive. As a result, most municipalities are forced to limit inspection-related work to portions of their entire system. Proactive asset management can be supported by modeling techniques that extract information from existing inspection datasets so that predictions of condition can be made for pipes that have not yet been inspected. Efficient approaches need to be made available to municipalities seeking to learn from existing inspection datasets as many existing modelling techniques (e.g. multiple linear regression, logistic regression and Markov-chains) are often incapable of reliably predicting individual pipe condition.

The capabilities of support vector machines and decision tree classifiers for predict-
ing individual sewer pipe condition are evaluated in this paper. A sanitary sewer pipe condition dataset collected by the City of Guelph, Ontario, Canada from 2008 to 2011 is used to demonstrate the process of implementing tools for predictive analytics. The described modeling framework represents a simple, yet powerful approach for gaining knowledge of the sewer pipe deterioration process that is novel as it provides a solution to class imbalance problems common in sewer inspection datasets. The concept of a decision tree classifier is general and the resulting model is interpretable by those unfamiliar with the data mining process, a characteristic lacking when other complex statistical or machine learning approaches are applied.

4.2 Background Information on Sewer Pipe Condition Modeling

The majority of sewer deterioration models have been foundationally based on statistically theory. Binary logistic regression models were developed to predict deficiency probability for sewers in Edmonton, Alberta but no goodness-of-fit tests were carried out (Ariaratnam et al., 2001). Binary logistic regression categorized the condition of sewer pipe segments in Phoenix, Arizona, but no indication of predictive capability for individual pipes was provided (Koo and Ariaratnam, 2006). Bayesian logic and expert opinion were used to develop a sewer cataloging retrieval and prioritization system presented in Merrill et al. (2003). Wright et al. (2006) indicate a linear regression model greatly under-estimated the length of deficient pipe and a logistic discriminant model was questionable on a pipe-by-pipe basis in a California sewer system (87% of acceptable pipes misclassified as being deficient).
Multiple regression models were trained to predict the condition of concrete, asbestos cement and PVC sewers in Pierrefonds, Quebec and Niagara Falls, Ontario using a small dataset (Chugthai, 2007; Chugthai and Zayed, 2008). Ordinal regression models developed in Younis and Knight (2010) predicted network-level sewer condition in Niagara Falls, Ontario. Statistical deterioration models developed in Opila (2011) had low R-squared values in the range of 0.12 - 0.17. Binary logistic regression models developed in Ens (2012) were incapable of reliably predicting sewer condition.

Deterioration curves for cohorts of sewer pipes in Germany were developed using survival models (Baur and Herz, 2002). Survival models tend to underestimate the number of pipes in the poorest condition states (Ana and Bauwens, 2010). Markov methods described in Micevski et al. (2002) were used to model condition for groups of stormwater pipes in Australia. Similar models developed by Baik et al. (2006) using the results of a condition survey of 90 km of sewer pipe in San Diego, California were found to be unsatisfactory as goodness-of-fit scores were low. Markov models and ordinal regression models developed in Tran et al. (2008) were deemed unsuitable for pipe-level predictions.

A case-based reasoning approach is described in Fenner et al. (2007) - where information on condition, performance (e.g. number of previous complaints) and management outcomes (e.g. intervene or non-intervene) for a small number of pipes can be used to proactively manage other pipes in a network. Deterioration models have also been developed using data mining techniques derived from the fields of artificial intelligence and machine learning. Neural networks have been used to predict the frequency and timing of failure in water distribution systems.
(Tabesh et al., 2009; Harvey et al., 2014) and the condition of pipes in sewer and stormwater systems (Najafi and Kulandaivel; Tran, 2007; Khan et al., 2010). A historical database of customer complaints related to pipe blockages was used in Arthur et al. (2009) to develop a methodology that incorporated consequence and likelihood of pipe failure to prioritise sewerage maintenance for system in the United Kingdom. A novel data mining technique, Evolutionary Polynomial Regression (EPR), was developed in Savic et al. (2006) to identify sewer pipes most likely to fail for a confidential location. EPR was also used in Berardi et al. (2008) to predict pipe bursts in a water distribution system and in Berardi et al. (2009) to plan sewer pipe inspections based on pipe-specific attributes, condition predicted by a Markov model and an expected cost of failure. Generalized pipe failure prediction models were developed in Savic et al. (2009) using an EPR approach that considered data from a number of individual water and sewer systems. Support vector machines (SVM) were presented as an alternative to neural networks for predicting sewer condition in South Australia in the work of Mashford et al. (2011).

SVMs, first developed in the 1970s (Vapnik, 1979, 1999), are based on an algorithm that constructs a linear model called the maximum margin hyperplane, which is a line (in two dimensions) or a flat plane (in multiple dimensions) that provides the greatest separation between instances with different values of the target variable. Datasets containing instances that cannot be easily separated with a straight line are projected into a higher-dimensional space using a kernel function.

Models developed using neural networks or SVMs are inherently black boxes, where relationships between inputs and outputs are deeply embedded within the model. A more transparent approach to predicting pipe condition would be decision tree
classifiers, such as those developed using the classification and regression tree (CART) algorithm developed by Breiman et al. (1984). The CART algorithm is capable of extracting information from mixed datasets (i.e. those datasets containing numerical, categorical and missing data) and requires very little in the way of data pre-processing (Han et al., 2006) characteristics that make the algorithm of potential utility when extracting information hidden within existing pipe inspection datasets. In general terms, the CART algorithm extracts information embedded within an existing dataset for knowledge discovery purposes. The resulting model is presented in a tree-like structure consisting of a root node (containing all the instances in the dataset used to train the model) and branches (illustrating the influence of various input predictors on the target class). The CART algorithm constructs a decision tree classifier using a divide-and-conquer approach, where the branches of the tree are optimally selected so that they lead to homogeneous subsets of data that have a larger proportion of one class than another. Optimal splits for the decision tree are determined by first sorting all the available instances based on their predictor value. A contingency table is generated for each split point (Table 4.1), and a Gini purity criterion is determined for each potential split for a two-class problem:

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; split</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>≤ split</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
</tr>
<tr>
<td></td>
<td>$n_{1+}$</td>
<td>$n_{2+}$</td>
</tr>
</tbody>
</table>
\[ Gini \text{ (prior to split)} = 2 \left( \frac{n_{1+}}{n} \right) \left( \frac{n_{2+}}{n} \right) \] (4.1)

\[ Gini \text{ (after the split)} = 2 \left[ \left( \frac{n_{11}}{n} \right) \left( \frac{n_{12}}{n+1} \right) + \left( \frac{n_{21}}{n} \right) \left( \frac{n_{22}}{n+2} \right) \right] \] (4.2)

- \( n_{11} \) = count of class 1 instances greater than the split point,
- \( n_{21} \) = count of class 1 instances less than or equal to the split point,
- \( n_{1+} \) = total number of instances in class 1,
- \( n_{12} \) = count of class 2 instances greater than the split point,
- \( n_{22} \) = count of class 2 instances less than or equal to the split point,
- \( n_{2+} \) = total number of instances in class 2,
- \( n_{+1} \) = total number of instances greater then the split point,
- \( n_{+2} \) = the total number of instances less than or equal to the split point, and
- \( n \) = total number of instances.

The CART algorithm evaluates all possible split points and partitions the dataset using the split point resulting in the minimum Gini purity criterion (Kuhn and Johnson, 2013). The tree grows larger as this process continues for each newly created partition until the data cannot be split any further. Eventually, branches of the tree are pruned away using a complexity parameter that is a function of the number of leaves in the tree and the resulting accuracy. Decision tree algorithms are popular within the fields of marketing, finance and medicine and but have not yet been used to predict sewer pipe condition using attributes only available.
prior to condition inspection. Sewer collapse rate and blockage rate deterioration models were developed in Heywood et al. (2007) using decision trees. Decision trees were used in Oliveira et al. (2007) as part of an exploratory analysis of the relationship between defect severity and pipe attributes. Decision trees in Syachrani et al. (2012) were used to visualize the relationship between operational condition and pipe material on an estimated real age, reflecting an adjusted pipe age based on location and operational conditions. Jung et al. (2012) used decision trees to identify important attributes associated with high-density regions of defective pipes in a mid-sized city.

4.3 Case Study

The municipality of Guelph, Ontario, Canada relies on a 515 km long sanitary sewer system consisting of 7,446 gravity pipes, 43 siphons and 33 pressurized forcemains. The average sewer pipe age in Guelph is 38 years old, including the approximately 1,900 pipes in operation for more than 50 years and the approximately 300 pipes transporting sewage for more than 100 years. Pipe diameters in the sanitary sewer system range from 100 to 1650 mm and approximately 82% of all pipes have a diameter between 200 to 300 mm. Burial depths range from 0.4 to 10 m (average burial depth is 3.2 m). The proportion of total system length grouped according to pipe material is as follows: asbestos cement (11%), concrete (16.1%), PVC (44.6%), reinforced concrete (3.1%), and vitrified clay (25.2%). In general, the oldest pipes in the system are vitrified clay and the majority of pipes constructed within the past 30 years are made of PVC.
The City of Guelph retained an engineering consultancy from 2008 to 2011 to CCTV inspect a portion of their sanitary sewer system and assist in the development of a capital rehabilitation/replacement program for their linear wastewater infrastructure. Pipes were selected for inspection using expert opinion, where pipes older than 50 years received the majority of inspection effort as it was expected these would be in the poorest structural condition. Structural defects were identified using the third edition of the Water Research Center Manual of Sewer Condition Classification (WRc MSCC) (WRc, 1996) and severity scores were assigned to the defects using the fourth edition of the Water Research Center Sewerage Rehabilitation Manual (WRc SRM) (WRc, 2001). The engineering consultancy assigned each inspected pipe an internal condition grade (ICG) of 1, 2, 3, 4 or 5 based on thresholds established in the WRc SRM for the highest severity scores accumulated in any one meter length of the pipe. Comprehensive quality assurance/quality control (QA/QC) was carried out to ensure accuracy of inspection data.

An analysis of detailed CCTV inspection records indicates the average inspected pipe length was 68 m and 33%, 59% and 80% of all structural defects occur within 10 m, 20 m, and 30 m of the nearest manhole, respectively. These findings have implications for future inspection-related activity as there is potential to reduce inspection costs by implementing zoom-camera technology. Zoom cameras capture video of the pipe interior using a camera mounted on a pole that is lowered into a manhole but have an effective usable distance of approximately 30 m. Zoom cameras are capable of inspecting up to 1,800 m of pipe per day at a cost of approximately $0.90/m (EPA, 2010). CCTV on the other hand can only cover 500
m of pipe per day with an average cost for utilities that outsource inspection of $3.96/m (EPA, 2010). While the technology is incapable of providing the same detailed visual evaluation as CCTV, zoom cameras may potentially expedite investigations of uninspected pipe given most structural defects are within the sight distance limit of the technology.

4.4 Data Mining Methodology

4.4.1 Selection of a Classification Target

Data mining analysis presented herein deals exclusively with 123 km of gravity sanitary sewer inspection data collected from 2008 to 2011 (1,825 individual pipes) performed solely by the primary inspection sub-contractor. A complete list of modeling inputs/attributes contained within the inspection dataset used for model development is presented in Table 4.2. Stratified random sampling was used to partition the inspection dataset into separate training, evaluation and test sets using a 70-10-20 split ratio. The inspection dataset is class imbalanced, with the majority of inspected pipes assigned an ICG of 1 (acceptable condition), 2 (minimal collapse risk but potential for further degradation) or 3 (collapse unlikely but further deterioration likely). Very few inspected pipes were assigned an ICG of 4 (collapse likely in the near future) or 5 (collapsed or collapse imminent) (Table 4.3). Class imbalance is common within inspection datasets, as pipes prone to failure tend to have already failed and been replaced prior to CCTV inspection, causing the number of observations for poor condition pipes to be underestimated (Ana...
and Bauwens, 2010). Class imbalance significantly compromises the ability of most algorithms to construct useful models for all condition classes, as model-building efforts focus on correctly classifying pipes in the majority class. As an example, an initial SVM model trained to assign pipes to one of the five ICGs assigned every pipe with an ICG 4 or 5 to the majority classes ICG 1, 2 or 3 and was therefore unsuitable for planning future CCTV work. Similar complications posed by class imbalance were reported in the work of Salman (2010), when a variety of statistical models were investigated for sewers in Cincinnati, Ohio using pipe-specific attributes (e.g. size, length, slope, age, burial depth, and material). The available data violated the proportional odds assumption necessary for ordinal regression model development. The validation set overall accuracy of a multinomial logistic regression model was 53% and 66% for binary logistic regression. A study carried out using class imbalanced datasets from two Belgian municipalities indicated Markov models were inaccurate at the individual pipe level, binary logistic models were 20% accurate for pipes in a failed condition state, and neural network models were unsuitable for forecasting pipe deterioration (Ana, 2009).

Techniques currently available for accommodating class-imbalance in data mining applications have primarily been developed for two-class problems. Dealing with class imbalance when more than two classes are involved is considerably more difficult and remains an active area of research within the data mining community (Han et al., 2006). This necessitates transforming pipe condition into a two-class format. Guelph considers the relatively rare categories of ICG 4 and 5 to be of great interest as their defects pose an immediate threat to the environment and surrounding infrastructure (although ICG 4 pipes may have some remaining
life, it is unwise from both an environmental and economic perspective to delay their rehabilitation). An alternative classification task can therefore be established whereby pipes are classified as being in either a good (ICG 1-3) or bad (ICG 4-5) structural condition state. The rate of bad pipes in the dataset after this transformation is approximately 11% (training set: 141 bad and 1,137 good pipes, evaluation set: 20 bad and 163 good pipes and the test set: 40 bad and 324 good pipes).

Table 4.2: Attributes available for data mining

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Material</td>
<td>Nominal</td>
<td>Material being either asbestos cement, concrete, PVC, reinforced concrete or vitrified clay</td>
</tr>
<tr>
<td>2. Age</td>
<td>Numeric</td>
<td>Age (range: 1 – 108 years, mean = 50.1 years)</td>
</tr>
<tr>
<td>3. Type</td>
<td>Nominal</td>
<td>Type of sewer (Trunk or branch)</td>
</tr>
<tr>
<td>4. Diameter</td>
<td>Numeric</td>
<td>Diameter (150 – 900 mm, mean = 270 mm)</td>
</tr>
<tr>
<td>5. Length</td>
<td>Numeric</td>
<td>Length (1.70 – 198.57 m, mean = 68 m)</td>
</tr>
<tr>
<td>6. Slope</td>
<td>Numeric</td>
<td>Slope (0 – 9.97 m/100 m, mean = 1.47 m/100 m)</td>
</tr>
<tr>
<td>7. Down Elevation</td>
<td>Numeric</td>
<td>Downstream pipe invert elevation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(304.40 – 355.95 m, mean = 329 m)</td>
</tr>
<tr>
<td>8. Depth</td>
<td>Numeric</td>
<td>Burial depth (0.65 – 8.88 m, mean = 3.2 m)</td>
</tr>
<tr>
<td>9. Road Coverage</td>
<td>Numeric</td>
<td>Portion of the pipe covered by a roadway</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0 – 100%, mean = 52%)</td>
</tr>
<tr>
<td>10. Watermain breaks (3m)</td>
<td>Numeric</td>
<td>Number of historical watermain breaks within 3 m of the pipe (0 – 7 breaks, mean = 0.2)</td>
</tr>
<tr>
<td>Condition</td>
<td>Nominal</td>
<td>Target: good (ICG 1–2–3) or bad (ICG 4–5)</td>
</tr>
</tbody>
</table>

Note: Attributes 1-10 represent a predictor dataset with no irrelevant predictors (i.e. no predictors have near-zero variance, nor are there any between predictor correlations)
Table 4.3: A summary of internal condition grade (ICG) by pipe material

<table>
<thead>
<tr>
<th>Material</th>
<th>Diameter (mm)</th>
<th>ICG</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Asbestos Cement</td>
<td>200-400</td>
<td>373</td>
<td>14</td>
</tr>
<tr>
<td>Concrete</td>
<td>200-900</td>
<td>364</td>
<td>79</td>
</tr>
<tr>
<td>PVC</td>
<td>200-450</td>
<td>110</td>
<td>9</td>
</tr>
<tr>
<td>Reinf. Concrete</td>
<td>300-900</td>
<td>56</td>
<td>8</td>
</tr>
<tr>
<td>Vitrified Clay</td>
<td>200-450</td>
<td>213</td>
<td>135</td>
</tr>
<tr>
<td>Total</td>
<td>200-900</td>
<td>1116</td>
<td>245</td>
</tr>
</tbody>
</table>

4.4.2 Implementing the Algorithms

The SVM model was developed using the *kernlab* package (Karatzoglou et al., 2004) and tuned using the *caret* package Kuhn (2013) within the open source software environment *R*. The general purpose radial basis function (RBF) kernel was used to project the data:

\[ k(a, b) = \exp\left(-\sigma \|a - b\|^2\right) \]  

(4.3)

where \( k \) is the kernel function, \( a \) and \( b \) represent two instances of pipe condition and \( \sigma \) is a hyper parameter that is automatically determined using an algorithm within the *kernlab* package. An additional cost parameter \( (C) \) included as part of the SVM optimization objective function for the RBF kernel was optimally determined using three repeats of ten-fold cross-validation of the training dataset. Predictor inputs presented to the SVM were centered and scaled to have a mean of zero and standard deviation of one. Recursive feature elimination was implemented.
to identify the predictive benefit of removing any non-informative predictors.

The CART algorithm was implemented using the \textit{rpart} (Therneau et al., 2014) and \textit{caret} (Kuhn, 2013) packages for \textit{R}. Optimal tree size was determined using three repeats of ten-fold cross-validation of the training dataset. Decision tree algorithms are largely insensitive to the characteristics of the predictor data, and no data pre-processing was required to improve predictive performance for the given set of input predictors. Whereas the predictive success of SVMs may hinge on presenting an appropriate subset of features to the algorithm, decision trees implicitly perform feature selection (ensuring predictors that do not contribute to the predictive power of the model are ignored).

### 4.4.3 Evaluating Predictive Performance

Continuous valued predictions in the form of a class membership probability between 0 and 1 generated by the models are used to establish predictions of pipe condition (\textit{good} vs. \textit{bad}) using a default classification threshold/cutoff of 0.50 and predictive performance on a dataset with known class labels can be evaluated using the confusion matrix shown in Table 4.4.

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Bad} (ICG 4-5)</td>
<td>\textit{Bad} (ICG 4-5)</td>
</tr>
<tr>
<td>\textit{Good} (ICG 1-3)</td>
<td>True Positive (TP)</td>
</tr>
<tr>
<td>\textit{Bad} (ICG 4-5)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td>\textit{Good} (ICG 1-3)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

When pipes in \textit{bad} condition are considered the positive class of interest, correctly
classified pipes are represented by TP = true positive (pipe actually in bad condition correctly predicted to be in bad condition) and TN = true negative (pipe actually in good condition correctly predicted to be good). Incorrect classifications are represented by FP = false positive (pipe predicted to be bad, when in fact, it is not), and FN = false negative (pipe predicted to be good, when it is actually bad essentially, saying there is nothing to worry about, when there actually is). Using these definitions, a model’s predictive accuracy is defined by:

\[
\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}
\]  

(4.4)

When dealing with class imbalance, accuracy on its own can be misleading as a trivial classifier that predicts every pipe as belonging to the majority good class can achieve high accuracy. The accuracy metric assumes false positives and false negative errors have the same costs, when in reality false negatives will have more severe consequences. As a result, a series of alternative metrics should be used when evaluating predictive capability:

\[
\text{True Positive Rate} = TPR = \text{sensitivity} = \frac{TP}{TP + FN}
\]  

(4.5)

\[
\text{True Negative Rate} = TNR = \text{specificity} = \frac{TN}{FP + TN}
\]  

(4.6)

\[
\text{False Positive Rate} = FPR = 1 - TNR
\]  

(4.7)
A useful visual tool for contrasting the predictive performance of two models is the receiver operating characteristic (ROC) curve, which is a plot of the TPR and FPR achieved across the full continuum of probability thresholds that could be used to make discrete class predictions. The area under the ROC curve can be used to gauge model performance, where perfect models have an area under the ROC curve of 1 and random models have an area under the ROC curve close to 0.5 (Fawcett, 2006). An area under the ROC curve greater than 0.7 on a stratified test set would be considered acceptable (Hosmer and Lemeshow, 2000). The model with the largest area under the ROC curve can be considered to be most effective for the classification task at hand.

The default threshold of 0.50 used to determine discrete class predictions tends to be unsuitable when working with class imbalance. The models were tuned to enhance predictive accuracy on the positive minority class using an alternative classification threshold that effectively changed the definition of a predicted event. The evaluation set ROC curve can be used to derive a new cut-off, where the threshold closest to the upper left corner of the ROC curve is optimal (as it is closest to a perfect model). This threshold moving approach has been shown to outperform some other popular class-imbalance learning techniques (e.g. down-sampling and up-sampling) (Han et al., 2006).
4.5 Results

4.5.1 Support Vector Machine Model

The SVM model achieving the highest area under the ROC curve during three repeats of ten-fold cross-validation of the training dataset had the following parameters: full-set of input predictors, $\sigma = 0.09$, $\gamma = 256$, support vectors = 381, and area under ROC curve = 0.69. The test set confusion matrix using this optimal SVM design and the default classification threshold of 0.50 is shown in Table 4.5. Although overall accuracy was 89%, the confusion matrix indicates the algorithm focused entirely on correctly classifying the majority class, resulting in a true positive rate of 0%. An optimal classification threshold derived from the point closest to the top left of the evaluation set ROC curve was 0.11 (Figure 4.1). The test set confusion matrix obtained when using this optimal threshold to reclassify predictions of condition in the test set is shown in Table 4.6. This optimal threshold results in the following performance metrics on the test set: TPR = 83%, TNR = 54%, accuracy = 58% and an area under the ROC curve = 0.72.
Table 4.5: SVM model confusion matrix for the test set (cutoff of 0.50)

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>0</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>1</td>
<td>323</td>
<td></td>
</tr>
</tbody>
</table>

True Positive Rate = (0) / (0 + 40) = 0%
True Negative Rate = (323) / (1 + 323) = 99.9%
False Positive Rate = (1) / (1 + 323) = 0.01%
False Negative Rate = (40) / (0 + 40) = 100%
Accuracy = (0 + 323) / (363) = 89%
Area under the ROC curve = 0.73

Figure 4.1: The evaluation set ROC curves developed for the SVM and decision tree models.
<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Bad (ICG 4-5)</th>
<th>Good (ICG 1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>33</td>
<td>7</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>148</td>
<td>176</td>
</tr>
</tbody>
</table>

True Positive Rate = \( \frac{33}{33 + 7} = 83\% \)  
True Negative Rate = \( \frac{176}{148 + 176} = 54\% \)  
False Positive Rate = \( \frac{148}{148 + 176} = 46\% \)  
False Negative Rate = \( \frac{7}{33 + 7} = 17\% \)  
Accuracy = \( \frac{33 + 176}{363} = 58\% \)  
Area under the ROC curve = 0.73

### 4.5.2 Decision Tree Classifier

The decision tree classifier achieving the highest area under the ROC curve during three repeats of ten-fold cross-validation of the training dataset had a complexity parameter for pruning of 0.002 and achieved a cross-validated area under the ROC curve of 0.71. A visual depiction of the decision tree is provided in Figure 4.2. The test set confusion matrix using this optimal design and the default classification threshold of 0.50 is shown in Table 4.7. Although overall accuracy was 89%, the confusion matrix indicates the algorithm focused on correctly classifying the majority class, resulting in a true positive rate of only 5%. The evaluation set area under the ROC curve for the decision tree model was 0.75 and an optimal classification threshold derived from the point closest to the top left of the evaluation set ROC curve was 0.05 (Figure 4.1). The test set confusion matrix obtained when using this optimal threshold (Table 4.8) indicates the following performance metrics: TPR =
78%, TNR = 76%, accuracy = 76% and area under the ROC curve = 0.77.

**Table 4.7:** Decision tree model confusion matrix for the test set (cutoff of 0.50)

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>0</td>
<td>324</td>
</tr>
</tbody>
</table>

True Positive Rate = (1) / (1 + 39) = 3%
True Negative Rate = (324) / (0 + 324) = 100%
False Positive Rate = (0) / (0 + 324) = 0%
False Negative Rate = (39) / (1 + 39) = 97%
Accuracy = (1 + 324) / (363) = 89%
Area under the ROC curve = 0.78

**Table 4.8:** Decision tree model confusion matrix for the test set (cutoff of 0.05)

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>79</td>
<td>245</td>
</tr>
</tbody>
</table>

True Positive Rate = (31) / (31 + 9) = 78%
True Negative Rate = (245) / (79 + 245) = 76%
False Positive Rate = (79) / (79 + 245) = 24%
False Negative Rate = (9) / (31 + 9) = 22%
Accuracy = (31 + 245) / (363) = 76%
Area under the ROC curve = 0.78
Figure 4.2: A visual depiction of the decision tree predictive model for sewer pipe condition.
4.6 Discussion

4.6.1 Comparing the SVM and Decision Tree Predictive Models

In terms of the area under the ROC curve metric, the decision tree model performed better than the SVM model during cross validation of the training set (0.71 vs. 0.69) and during model testing (0.77 vs. 0.72). The decision tree model also achieved a higher level of accuracy on the test dataset than the SVM model (76% vs. 58%).

In terms of knowledge discovery, the branches and leaves of the decision tree provide insight into the role of various pipe-specific attributes on determining pipe condition. The root of the decision tree shown in Figure 4.2 indicates 11% of the pipes in the training set were in bad condition and 89% were in good condition. The first split of the tree off of this root node illustrates the influence of time on pipe condition, with 5% of inspected pipes less than 50 years old were in bad condition (node 2) compared to 24% of those more than 50 years old found to be in bad condition (leaf 3).

Nodes 4 and 5 of the decision tree indicate pipe burial depth influences the likelihood of older pipes being in a structurally deteriorated condition state. Pipes more than 50 years old with burial depths less than 1.9 m have a 42% chance of being in bad condition (node 4) compared to a 21% chance for pipes buried at greater depths (node 5). The literature indicates pipe defect rates and the likelihood of collapse tend to increase as burial depths decrease (Cullen, 1982; Jones, 1984; O’Reilly...
et al., 1989; Fenner et al., 2007; Savic et al., 2006; Berardi et al., 2009).

Water and wastewater infrastructure deterioration in the municipality are related, as sewer pipes more than 50 years old, with a burial depth less than 1.9 m that have experienced at least 1 nearby water main break have an 80% likelihood of being in bad condition (leaf 6), compared to a 33% chance for similar sewer pipes with no nearby water main breaks (leaf 7). Author’s note - An unexpected failure of a drinking water distribution pipe may result in damage to surrounding infrastructure via destruction of overlying roadways, flooding, and transport of substantial sediment/debris loads into sewers. The heavy equipment activity and excavation work typically required during an emergency water main break repair may be an additional source of structural damage to any existing sanitary sewer pipes buried within the area of construction.

Structural condition is influenced by the length of an individual pipe, where 24% of older pipes with a burial depth greater than 1.9 m and a length greater than 33 m are in bad condition (node 8). This is significantly higher than the 5% of similar pipes with a length less than 33 m that are in bad condition (leaf 9). Longer pipe are known to be more vulnerable to differential settlement and have more possible locations of pipe failure (e.g. joints) (Ana et al., 2008).

Node 10 and leaf indicate the influence of pipe diameter on pipe condition, where smaller diameter pipes have a 26% chance of being in bad condition, compared to a 12% chance for larger diameter pipes. It is possible that larger diameter pipes have a reduced likelihood of being in bad condition due to the tendency for larger diameter pipes to be installed by experienced personnel, reducing the likelihood of defects related to installation error (Davies et al., 2001; Savic et al., 2006).
An additional split of smaller diameter pipes (leaves 12 and 13) again indicates a
history of nearby watermain failures can increase the likelihood of an older sewer
pipe being in \textit{bad} condition.

4.6.2 Screening Pipes for Future Inspection

The decision tree classifier can be used to predict the condition of uninspected pipes
in the Guelph system. Of the more than 4,000 uninspected pipes (the large majority
of which are less than 50 years old), a total of 876 pipes are classified by the decision
tree as being in \textit{bad} condition. For the next round of inspection, the municipality
can choose to inspect all 876 predicted \textit{bad} condition pipes, knowing that some
of them will actually be in \textit{good} condition. Alternatively, the municipality can
use the class probability scores output by the decision tree to select a subgroup of
predictions that have a higher proportion of \textit{bad} pipes than in the original dataset.

This approach is similar to one taken when advertising firms, looking to decrease
marketing costs, use predictive analytics to identify customers that are most likely
to respond to mail-out campaigns. Considering the 365 instances in the test set,
if all predictions of pipe condition are sorted in order of decreasing probability of
being \textit{bad}, the gain for a given sample size (percentile) is:

\[
Gain = \left( \frac{\text{\# of pipes in the sample that are actually bad}}{\text{Total \# of bad pipes}} \right) \times 100\% \tag{4.9}
\]

Figure 4.3 presents the gain associated with various sampling scenarios, indicating
the top 20\% of pipes sorted by decreasing propensity (likelihood of being in \textit{bad}
condition) contains approximately 68\% of all \textit{bad} pipes in the data set. The top 73
predictions (20% x 365) in the test set contain 27 pipes that were actually observed to be *bad*. The entire test set only contained 40 *bad* pipes, so 20% of the decision tree’s top predictions capture $27/40 = 68\%$ of all the *bad* pipes. The diagonal line indicates the expected response when inspections are carried out without the predictive model (i.e. using the same expert opinion used when planning the first round of inspections), indicating 248 or 68% of all 365 pipes in the test set would need to be inspected to capture 68% of all the *bad* pipes. The model provides an opportunity to carry out significantly fewer inspections to identify the same number of *bad* pipes, resulting in considerable time and cost-savings.

![Figure 4.3: Gain chart for the decision tree model.](image)
4.6.3 The Impact of Dataset Size

A sensitivity analysis can be carried out to examine the potential to develop decision tree models using less extensive datasets as such models would be useful for those municipalities in the early-stages of a planned inspection program. Guelph’s original inspection dataset can be split into two - the first containing 30% of the original dataset (representing a smaller, hypothetical inspection program of 547 pipes) and the second containing the other 70% of the original dataset (representing a hypothetical set of 1278 uninspected pipes). Given that the condition of the pipes in the latter dataset is available for analysis (141 bad and 1,137 good pipes), it can be used to evaluate the predictive utility for smaller datasets.

The hypothetical inspection of program of 547 pipes would have identified 347 good pipes and only 35 bad pipes, and the CART algorithm can use these instances of pipe condition to construct a decision tree classifier. Although there were fewer instances of bad pipes available for learning purposes, which inevitably lowered the predictive capabilities of the model when compared to a model developed with a larger dataset (three repeats of ten-fold cross validation area under ROC = 0.65), the trained model was found to be an effective tool for screening pipes for condition inspection. The decision tree model achieved a true positive rate of 70% and a true negative rate of 70% for the 1,278 pipes in the hypothetical set of uninspected pipes (for which, in this sensitivity analysis, the condition state is known already). Rather than inspecting every pipe predicted to be bad, the top 20% x 1278 pipes = 256 inspections would have identified 73 pipes that were actually in bad condition. There were only 141 bad pipes in the dataset, therefore inspecting only 20% of all the pipes captured 73/141 = 52% of all the remaining bad pipes in the network that
have not yet been inspected. The decision tree provides an opportunity to perform 256 inspections instead of $52\% \times 1,278 = 665$ inspections (the number required to achieve the same result if data mining was not carried out), indicating the algorithm continues to be valuable even for smaller datasets. Given that the average CCTV inspection cost in North America is approximately $2/m and the average inspected pipe length in Guelph was approximately 70 m, the opportunity to capture more information on bad condition in fewer inspections can result in considerable cost savings to the municipality. For this hypothetical example, implementing predictive analytics after the first round of inspections are performed so that 73 bad condition pipes can be identified in 256 inspections instead of 665 inspections, representing an inspection cost difference of approximately $35,800 vs. $93,100.

4.6.4 Improving Overall Accuracy and Future Directions of Research

The decision tree classifier was developed without information related to soil condition in Guelph as it was unavailable at the time of model development. This likely had an impact on predictive accuracy as soil corrosion potential (a product of particle size, uniformity, organic content, etc.) has been known to affect the rate of exterior pipe corrosion. Municipalities looking to implement decision-tree based predictive analytics should include detailed soil records if they are available as this information may increase the predictive capability of a developed model. Information on maintenance activities for individual assets would also potentially improve predictive performance for any developed model. As class-imbalance within inspection datasets provides an obstacle for model development,
municipalities should direct inspection efforts towards identifying pipes that are most likely in bad condition. Beyond providing information required for planning necessary rehabilitative action, the new bad pipe condition data would help to alleviate class-imbalance and increase the likelihood the data mining algorithm can learn to accurately predict pipe condition.

Although the developed decision tree model is easier to interpret than many existing deterioration modeling techniques, the use of a single decision tree results in a general trade-off between predictive performance and the simplicity of the overall model structure. As such, future modeling efforts can be directed towards other data mining systems (e.g. random forests that combine the predictive power of many individual decision trees) that may potentially improve predictive accuracy as they are known to be effective when dealing with imbalanced datasets. Investigations into the utility of decision-tree based predictive models developed within a risk assessment framework are recommended as a future direction for research - where predicted condition paired to risk-based concepts (e.g. WRc critical sewers (WRc, 2001) or other utility-specific asset importance criteria) might enhance the process of screening pipes for inspection based on both the likelihood and consequence of failure.

4.7 Conclusion

An examination of CCTV inspections carried out in Guelph, Ontario indicates the most common structural defects within Guelph’s sanitary sewer pipes are cracks, fractures and defective joints, representing 42%, 38% and 11% of all recorded
structural defects, respectively. Inspection records indicate 33.2%, 59.3% and 80.3% of all recorded structural defects occur within 10 m, 20 m, and 30 m of the nearest manhole, respectively. These findings suggest future inspection efforts in Guelph may benefit from zoom-camera inspection technologies that are faster than CCTV but often have their utility limited unless sight distance is less than 30 m.

Many proposed approaches to modeling sewer pipe deterioration are unsuited to predicting individual pipe condition and may provide municipalities with a spurious impression of their true predictive power. The predictive capabilities of support vector machines (SVMs) and decision tree classifiers are evaluated. Decision trees are found to be a simple and effective method of gaining deeper insight into the influence of pipe-specific parameters on the structural condition of individual pipes. Transforming the classification task into a binary format of good vs. bad pipe condition and then adjusting thresholds used to make classification decisions accommodates class imbalance common within pipe inspection datasets.

The predictive capabilities of the models were evaluated using a variety of metrics suitable for assessing imbalanced datasets. The developed decision tree had a higher area under the ROC curve on a stratified test set than the SVM model. The decision tree achieved an acceptable cross-validated area under the ROC curve of 0.71 during training and a test set accuracy of 76%, true positive rate = 78% and a true negative rate = 76%. Overall the model obtained a high prediction rate for bad pipes without sacrificing a reasonable level of accuracy for the good pipes, suggesting the model can guide inspection future work towards uninspected pipes that have a high likelihood of being in bad structural condition.
The trained decision tree classifier indicates pipe age is important for determining the structural condition of sanitary sewers in Guelph, with pipes more than 50 years old having a significantly greater chance of being in bad structural condition than newer pipes in the same sanitary system. The decision tree also illustrates the increased likelihood of pipe failure associated with shallow burial depths, longer pipe lengths, smaller diameters, and the presence of nearby water main failures for select subsets of sanitary pipes. The developed model provides the municipality with an opportunity to learn from an existing inspection dataset so that bad pipe yield can be significantly higher during future inspection programs. Overall, the combination of inspection and efficiently implemented, open-source data mining techniques presented in this paper may result in significant cost-savings for municipalities looking to ensure publicly owned sewer systems are being effectively managed.
Transition to Chapter 5

This next chapter builds on the predictive success of data mining systems by introduction of the concept of predicting sanitary sewer pipe condition using the random forests data mining algorithm. The City of Guelph is again used as a case study area. The random forests algorithm is found to be more capable at predicting pipe condition than a single decision tree classifier. This paper has been peer-reviewed and accepted for publication in the Canadian Journal of Civil Engineering:

Chapter 5

Predicting the Structural Condition of Individual Sanitary Sewer Pipes with Random Forests

5.1 Introduction

Canada’s first report card on the condition of municipal infrastructure indicates 30% of sewer pipes across the country are in very poor, poor, or fair condition (CIRP, 2012). The cracks, fractures, holes, and broken joints present in these defective pipes allow raw wastewater to leak into the surrounding soil. The exfiltration of raw wastewater poses an environmental threat as a range of groundwater pollutants are found in typical municipal sewer flows including pathogenic microorganisms, industrial toxins and endocrine disrupting compounds (EDCs). EDCs are dangerous as they interfere with the behavior of natural hormones in the body that
are responsible for the maintenance of homeostasis, reproduction, and development. As an example, concentrations of ethinyl estradiol (a synthetic estrogen found in birth control pills and present at biologically relevant levels in wastewater) were the cause of the feminization and near extinction of minnows in an experimental lake in northwestern Ontario (Kidd et al., 2007).

Studies have shown that groundwater in the direct vicinity of a leaking sewer pipe and deeper aquifers are both under threat of contamination as sewage-derived contaminants are capable of reaching depths of more than 60 meters when geological heterogeneities such as cracks and fissures are present in the subsurface (Heberer and Stand, 1997; Morris et al., 2005; Powell et al., 2003). A recent field study by Sercu et al. (2011) provides multiple lines of evidence that storm drains may also be contaminated directly by sewer exfiltration, even during periods of dry weather. These contaminated storm drains provide an additional pathway for raw, untreated sewage to enter the natural environment. All things considered, North America’s deteriorating sewers may potentially bring water pollution levels back in-line with those observed in the 1970s - placing decades of progress in public health and environmental protection at risk (EPA, 2002).

Structural defects also allow groundwater surrounding defective pipes to infiltrate into sanitary sewers. This infiltrating groundwater may overload sanitary sewer systems and wastewater treatment plants, particularly during periods of extreme wet weather. Once the sewer network is overloaded, excess wastewater may migrate into basements, streets, rivers, and lakes. The economic fallout caused by the resulting sanitary sewer overflows (SSOs) may be severe - one extreme rainfall event in southern Ontario in the summer of 2005 resulted in $247 million of insurance
claims related to sewer backups in homes (ICLR, 2013). According to the Insurance Bureau of Canada, basement flooding caused by defective sewer infrastructure is one of the fastest growing causes of extreme damage in Canada, annually costing upwards of $1.7 billion in insurance payments (IBC, 2012).

A 2003 survey of 150 Canadian municipalities found significant variations in asset characterization by municipalities. For the Province of Ontario, 17% of municipalities did not visually inspect their sewers and 33% of those who did have inspections performed failed to record inspection results (PWC, 2003). At the time, most municipalities were not performing the activities necessary to determine sewer system condition and the general attitude towards sewer management across Canada could be described as out-of-sight and out-of-mind. A decade later, there is still considerable room for improvement as 33% of Canadian municipalities still rely primarily on the opinion of qualified individuals when determining rehabilitation and renewal requirements, not on comprehensive visual inspection (CIRP, 2012).

Visual inspection of buried sewer pipes may serve as a foundation for pro-actively managing sewers experiencing exfiltration and infiltration issues. The backbone of the sewer condition assessment industry is closed-circuit television (CCTV) inspection, which consists of recording the internal condition of a pipe (Figure 5.1) using a small video camera mounted on a robot. Although CCTV provides valuable information on pipe condition, it is expensive, with an average cost for medium-size municipalities of approximately $2 per meter of inspected pipe (EPA, 2010). Consequently, budgetary restrictions tend to limit most municipalities to inspecting small portions of their entire network of pipes. As a result, these municipalities are in need of tools that can maximize the value of their existing inspection data so
that accurate predictions of condition can be made for those pipes in the network that have not yet been inspected.

Figure 5.1: Examples of structural defects observed during CCTV inspection.

5.2 Sewer Deterioration Modeling

A variety of approaches have been proposed in recent years for modeling sewer pipe deterioration using data collected through CCTV inspection. Ariaratnam et al. (2001) modeled sewer pipe deterioration in Edmonton, Alberta using binary logistic regression (no goodness-of-fit tests were provided). Linear regression mod-
els developed in Wright et al. (2006) greatly under-estimated the length of deficient sewer pipes in a California sewer system. Multiple linear regression models were developed by Chugthai (2007) using small sewer condition datasets collected in two Canadian municipalities. Salman (2010) found binary logistic regression models were capable of achieving a correct prediction rate of 46% for structurally unsound sewer pipes in Cincinnati, Ohio. Younis and Knight (2010) used 45 km of inspection data collected in Niagara Falls, Ontario to develop network-level cumulative logistic models for predicting the probability of concrete and vitrified clay pipes being in various condition states, given material and age. Ens (2012) found binary logistic regression was incapable of reliably predicting the condition of individual sewers in an unspecified Canadian municipality (R-squared values below 0.02).

A number of authors have used survival and Markov-chain methods for modeling pipe deterioration (e.g. Tran (2007)). Markov-chain models for sewer deterioration in San Diego, California were useful for developing generalized network-level deterioration curves but goodness-of-fit tests were low, particularly for pipes in poor condition (Baik et al., 2006). LeGat (2006) used inspection results from more than 5,000 pipes in Germany to calibrate a Markov-chain model that was effective in detecting the most deteriorated pipes in the sewer network (Le Gat conceded the approach involved heavy and cumbersome data manipulations). Tran et al. (2008) found Markov models were only useful at the network-level and ordinal regression was only capable of achieving an overall accuracy of 42% when predicting storm sewer condition in Australia. Duchesne et al. (2012) used survival analysis to predict structural condition at the network-level for cohorts of pipes in a Quebec City sewer network. A Markov process presented in Atef et al. (2012) and Osman
et al. (2012) was not developed for pipe condition prediction, but evaluated the suitability of various condition assessment technologies (e.g. low accuracy vs. high accuracy inspection techniques) for water pipes in Hamilton, Ontario.

Machine learning and artificial intelligence techniques represent an alternative modeling approach when the inherent bias and sparseness of inspection datasets prevent statistical model development. Khan et al. (2010) investigated the influence of parameters related to sewer pipe deterioration in Pierrefonds, Quebec using artificial neural networks. Support vector machines were used with success for individual pipe prediction in Australia (Mashford et al., 2011). Machine learning techniques have not always proven successful - e.g. Ana (2009) who reported no approach was appropriate for forecasting individual pipe deterioration as the generated aging profiles were problematic (some pipes characterized as improving in condition over time and a tendency to predict pipes as being in better condition than would be observed upon inspection).

Although some approaches have proven useful for predictions of sewer pipe condition at the network-level (e.g. cohort survival, Markov, etc.), they have generally been found incapable of predicting individual pipe condition with any reasonable level of confidence. Network-level approaches have merit when establishing budgets for future rehabilitative actions but may provide municipalities with only limited information on individual pipe condition. In this paper, a novel application of the random forests algorithm effectively extracts information related to pipe deterioration from CCTV inspection surveys in a manner facilitating individual pipe condition prediction. Although not previously used for sewer condition prediction, random forests are an increasingly popular data mining technique that have been
used in recent years on a variety of classification tasks, including remote sensing (Gislason et al., 2006), ecological prediction (Prasad et al., 2006), and microarray-based cancer classification (Statnikov et al., 2008). Random forests are appealing for the task of predicting individual sewer pipe condition as they implicitly carry out feature selection, are robust to outliers, often achieve high levels of performance with a minimal amount of data pre-processing and tuning (unlike neural networks), and have compared favorably or have outperformed support vector machines on a variety of classification tasks (Nappi et al., 2012).

5.3 Predictive Modeling

Each instance (or observation) of pipe condition from an inspection dataset is generally labeled according to an ordinal condition grade, such as the Water Research Center (WRc) internal condition grades 1 - 5. In data mining terms, a classification model needs to be developed to investigate the relationship between the target class (pipe condition) and a set of input predictors (various pipe-specific attributes). In general terms, the resulting model would be an approximation of some unknown function:

$$Y = f(X_1, X_2, \ldots X_p)$$  \hspace{1cm} (5.1)

where $Y$ is the condition and pipe-specific attributes are predictors $X_1, X_2, \ldots X_p$.  

113
5.3.1 Random Forests

Decision trees (the traditional building blocks of the data mining process) have previously been used to extract information from sewer inspection datasets (Jung et al., 2012; Oliveira et al., 2007; Syachrani et al., 2012). Although useful, a more powerful approach to modeling pipe deterioration is the random forests system (Breiman, 2001) as it has proven to have better predictive performance than single trees for a variety of data mining tasks (Kuhn and Johnson, 2013). The random forest approach to predictive modeling consists of growing hundreds of unpruned decision tree classifiers and then combining them into a single ensemble of models, or forest of trees. The algorithm constructs individual unpruned trees using bootstrap aggregation where instances are first randomly sampled (with replacement) from the instances comprising the original training dataset. This bootstrap sample of instances is used to grow the first tree and each tree in the forest is constructed using a different bootstrap sample obtained from the original dataset. The tree is then grown to its maximum size (with no pruning) using an algorithm similar to CART (Breiman et al., 1984) where for each binary split in the tree, \( m \) predictors are specified that will be randomly selected out of all \( M \) possible predictors. The best predictor among the available \( m \) predictors used to make the split is determined using the so-called Gini index, which is calculated for each potential split point of a predictor in a dataset containing two classes using the equation:

\[
Gini = p_1 (1 - p_1) + p_2 (1 - p_2) = 2p_1p_2 \quad (5.2)
\]
where \( p_1 \) is the relative frequency/probability of Class 1 in the dataset and \( p_2 \) is the probability of Class 2 (these two probabilities sum to one). When a two-class dataset is split into two subsets based on a potential split point of an attribute, a contingency table is developed (Table 5.1).

Table 5.1: Contingency table obtained after a potential split is made in the tree

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; split</td>
<td>( n_{11} )</td>
<td>( n_{12} )</td>
</tr>
<tr>
<td>( \leq ) split</td>
<td>( n_{21} )</td>
<td>( n_{22} )</td>
</tr>
<tr>
<td></td>
<td>( n_{1+} )</td>
<td>( n_{2+} )</td>
</tr>
</tbody>
</table>

The Gini index before and after the split are determined using the equations:

\[
Gini \ (prior \ to \ split) = 2 \left( \frac{n_{1+}}{n} \right) \left( \frac{n_{2+}}{n} \right) \tag{5.3}
\]

\[
Gini \ (after \ split) = 2 \left[ \left( \frac{n_{11}}{n} \right) \left( \frac{n_{12}}{n_{1+}} \right) + \left( \frac{n_{21}}{n} \right) \left( \frac{n_{22}}{n_{2+}} \right) \right] \tag{5.4}
\]

The best input predictor is the one providing the smallest Gini index after the split (Breiman et al., 1984). A total of \( K \) decision trees are grown in a similar fashion and the resulting trees make up the random forest. The resulting class prediction for new instances is obtained by having each tree cast a class vote (the mode of all the votes is the resulting prediction of the random forest) (Kuhn and Johnson, 2013). Class probabilities for individual instances are determined by counting the number of votes for each individual class and dividing by the total number of trees in the forest (e.g. if there are 1,000 trees and, for a particular instance, 600 trees
vote for class 1, then the probability of class 1 for that instance is $600/1000 = 60\%$.

Random forests for this research were developed and tuned using the \texttt{randomForest} (Liaw and Wiener, 2002) and \texttt{caret} (Kuhn, 2013) packages developed for \textit{R}.

## 5.4 The City of Guelph Case Study

The City of Guelph, Ontario, Canada relies on a 515 km sanitary sewer system (7,446 gravity pipes, 43 siphons, and 33 pressurized forcemains) to transport their wastewater. Gravity pipes (the focus of this research paper) have installation dates ranging from 1902 to 2012, diameters of 200 - 1,650 mm, lengths of 1 - 600 m, and are made of PVC (44.6\% of all pipes in operation), vitrified clay (25.2\%), concrete (16.1\%), asbestos cement (11\%), and reinforced concrete (3.1\%).

Guelph retained an engineering consultancy in the spring of 2008 to assist in the development and management of a sewer condition assessment program for their linear wastewater infrastructure. The consultancy (by way of sub-contractor) inspected 221 km of sanitary sewer pipe from 2008 to 2011 (Figure 5.2). Pipes were selected for inspection based on expert opinion provided by the consultancy and the City (with the intention of focusing inspection work on older pipes with an increased likelihood of being in poor condition). Structural defects inside inspected pipes were reviewed using the third edition of the Water Research Council Manual of Sewer Condition Classification (WRc MSCC). Each defect observation was assigned a severity score using the fourth edition of the Water Research Center Sewerage Rehabilitation Manual (WRc SRM). These severity scores were tabulated to determine the peak (highest defect values accumulated in any one meter length
of the pipe), mean (sum of all defect values along the pipe divided by pipes length) and total (sum of all defect values along the full pipe length) scores. The results of these three different scores were then reviewed so that pipes with a risk of collapse in the short term were identified by their peak score, pipes with a significant amount of general deterioration were highlighted by a review of their total scores and pipes with short lengths but significant deterioration were highlighted by a review of their mean score. After this initial screening, an internal condition grade (ICG) of 1 (no defects), 2, 3, 4 or 5 (collapsed or collapse imminent) was assigned to each inspected pipe based on the set of peak score thresholds defined in the WRc SRM. Guelph considers any pipe with an ICG of 4 or 5 to be in bad structural condition as these pipes have structural defects that are potentially acting as points of exfiltration or infiltration and are in need of rehabilitation.

The consultancy carried out comprehensive quality assurance/quality control to ensure accuracy of the defects recorded during pipe inspection. The existing linear asset database contained a considerable amount of missing or erroneous information related to individual pipes in the Guelph system (e.g. missing construction year and material of construction). Pipes missing these essential pieces of information were excluded from the modelling dataset. Analysis presented herein deals exclusively with 1,825 gravity sanitary sewer pipes (123 km) inspected by the primary sub-contractor (Table 5.2). The modeling dataset is class imbalanced as structurally defective pipes are heavily under-represented (i.e. there are only 201 pipes with an ICG of 4-5 but 1624 pipes with an ICG of 1-3). Class imbalance is common when sewer condition data are collected within a limited time period as pipes prone to failure would have already failed and been replaced prior to CCTV inspection,
Figure 5.2: Map of inspected pipes in Guelph, Ontario, Canada.
causing the number of observations for poor condition pipes to be underestimated (Ana and Bauwens, 2010). The various pipe-specific attributes used for data mining purposes are summarized in Table 5.3.

**Table 5.2: A Summary of Internal Condition Grades by Pipe Material**

<table>
<thead>
<tr>
<th></th>
<th>Internal Condition Grade (ICG)</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Asbestos Cement</td>
<td>373</td>
<td>14</td>
</tr>
<tr>
<td>Concrete</td>
<td>364</td>
<td>79</td>
</tr>
<tr>
<td>PVC</td>
<td>110</td>
<td>9</td>
</tr>
<tr>
<td>Reinforced Concrete</td>
<td>56</td>
<td>8</td>
</tr>
<tr>
<td>Vitrified Clay</td>
<td>213</td>
<td>135</td>
</tr>
<tr>
<td>Total</td>
<td>1116</td>
<td>245</td>
</tr>
<tr>
<td>Attribute</td>
<td>Type</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1. Material</td>
<td>Nominal</td>
<td>Asbestos cement, concrete, PVC, reinforced concrete or vitrified clay</td>
</tr>
<tr>
<td>2. Age</td>
<td>Numeric</td>
<td>Age at the time of inspection</td>
</tr>
<tr>
<td>3. Installation Era</td>
<td>Nominal</td>
<td>Pre-WWI (&lt;1914), WWI (1914-1918), Inter-war (1919-1938), WWI (1939-1945), Post-WWI (1946-1966), or Modern (1967-Present)</td>
</tr>
<tr>
<td>4. Type</td>
<td>Nominal</td>
<td>Type of sewer (trunk or branch)</td>
</tr>
<tr>
<td>5. Diameter</td>
<td>Numeric</td>
<td>Pipe diameter</td>
</tr>
<tr>
<td>6. Length</td>
<td>Numeric</td>
<td>Pipe length</td>
</tr>
<tr>
<td>7. Slope</td>
<td>Numeric</td>
<td>Pipe slope</td>
</tr>
<tr>
<td>8. Slope change</td>
<td>Numeric</td>
<td>Maximum change in slope at either the upstream or downstream connection</td>
</tr>
<tr>
<td>9. Down elevation</td>
<td>Numeric</td>
<td>Downstream pipe invert elevation</td>
</tr>
<tr>
<td>10. Orientation change</td>
<td>Numeric</td>
<td>Maximum change in pipe orientation at either the upstream or downstream connection.</td>
</tr>
<tr>
<td>11. Depth</td>
<td>Numeric</td>
<td>Pipe burial depth</td>
</tr>
<tr>
<td>12. Road Coverage</td>
<td>Numeric</td>
<td>Portion of the pipe covered by a roadway</td>
</tr>
<tr>
<td>13. Watermain breaks (3m)</td>
<td>Numeric</td>
<td>Number of historical watermain breaks within 3 m of the pipe (0 – 7 breaks,)</td>
</tr>
<tr>
<td>14. Land use</td>
<td>Nominal</td>
<td>Agricultural, commercial, industrial, institutional, park or residential.</td>
</tr>
<tr>
<td>15. Census Tract</td>
<td>Nominal</td>
<td>Pre-determined census tracts (1-27 tracts)</td>
</tr>
</tbody>
</table>

*Census tract refers to one of 27 districts within the municipality that were pre-determined by the Government of Canada for census purposes - see Figure 5.2
Slope is the percent change in elevation and is determine using the equation
(upperstream elevation - downstream elevation)/100m x 100%*
5.5 Evaluating Predictive Performance

The random forest algorithm generates two types of predictions: (i) a continuous-valued prediction in the form of a class membership probability between 0 and 1 and (ii) a discrete class prediction based on the class membership probabilities (using a baseline probability cut-off of 0.50). Based on the discrete class prediction, the predictive capability of a classification model is typically evaluated using a confusion matrix. A test set confusion matrix obtained for a random forest (1500 trees, \( m = 4 \) predictors considered when making tree splits, trained using 80% of the inspection data, and tested using 20% of the data) developed for the prediction of ICG 1-5 is presented in Table 5.4. Instances along the diagonal of the matrix are correctly classified and off-diagonal instances are incorrectly classified. In this case, the trained model is incapable of reliably assigning pipes to ICG 4 and 5 (e.g. of the seven ICG 5 pipes in the test set, three are incorrectly predicted to be ICG 1, one is incorrectly predicted to be ICG 2 and three are incorrectly predicted to be ICG 3). In its current form, the model would be unsuited to the task of identifying uninspected pipes that pose a threat to the structural integrity of the sewer network. Fortunately, strategies can be implemented that significantly improve the capabilities of the random forest algorithm to correctly predict the location of ICG 4-5 pipes that are of primary concern (due to exfiltration and infiltration-related threats).
Table 5.4: Test set confusion matrix for random forest trained for ICG 1-5 classification task

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>1</td>
<td>201 8 11 3 0</td>
</tr>
<tr>
<td>2</td>
<td>25 5 10 9 0</td>
</tr>
<tr>
<td>3</td>
<td>28 5 9 10 0</td>
</tr>
<tr>
<td>4</td>
<td>15 2 5 9 1</td>
</tr>
<tr>
<td>5</td>
<td>3 1 3 0 0</td>
</tr>
</tbody>
</table>

Correctly classified pipes are located along the diagonal of the matrix (e.g. 201 pipes that were actually ICG 1 are correctly predicted by the model to be ICG 1, 5 pipes actually in ICG 2 are correctly predicted by the model to be ICG 2, etc.)

Incorrectly classified pipes are located off-diagonal (e.g. 25 pipes that were actually ICG 2 incorrectly predicted to be ICG 1, 28 pipes that were actually ICG 3 incorrectly predicted to be ICG 1, etc.)

The imbalanced distribution of instances of pipe condition across the five target classes (ICG 1 - 1116 pipes, ICG 2 - 245 pipes, ICG 3 - 263 pipes, ICG 4 - 164 pipes and ICG 5 - 37 pipes) poses a problem as data mining systems tend to be most effective when instances are balanced across all classes (i.e. approximately the same). Classifiers developed from an imbalanced dataset will generally predict the majority class more frequently than the minority class. Therefore, instances belonging to the minority class are misclassified more often than those in the majority class (Nguyen et al., 2009). For the Guelph dataset, the random forest algorithm focuses the learning effort on the majority ICG 1 class, resulting in 201 out of 223 correctly classified instances of ICG 1 pipes in the test dataset. This consequently biases the model towards the ICG 1 class (i.e. approximately half
of the ICG 4 and ICG 5 pipes in the test dataset are incorrectly predicted to be ICG 1). A variety of strategies are available for neutralizing the negative impact of class imbalance on predictive performance (e.g. adjusting predicted probability cut-offs). The majority of class-imbalance learning strategies have been designed for two-class problems, where predictive models are encouraged to be more aggressive when predicting the minority class of interest. Consequently, implementation of a class-imbalance learning strategy required categorizing pipe condition using guidelines established by the municipality, where pipes with an ICG of 4 or 5 are of concern due to their elevated potential for exfiltration and infiltration. In this binary format there are 201 bad (ICG 4-5) pipes and 1,624 good (ICG 1-3) pipes in the inspection dataset available for model development. Predictive performance for this imbalanced classification task can be evaluated using the confusion matrix shown in Table 5.5.

**Table 5.5:** Confusion matrix for a binary classification task

<table>
<thead>
<tr>
<th></th>
<th>Predicted Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bad (ICG 4-5)</td>
</tr>
<tr>
<td>Actual Condition</td>
<td>Bad (ICG 4-5)</td>
</tr>
<tr>
<td></td>
<td>Good (ICG 1-3)</td>
</tr>
</tbody>
</table>

With bad pipes defined as the minority class of interest (as they pose an environmental threat), TP is a true positive (correctly predicted bad pipe), FP is a false positive (when the model incorrectly predicts good and the actual is bad), TN is a true negative (correctly predicted good pipe) and FN is a false negative (when a prediction of good is made when it is actually bad). Using this confusion matrix, predictive performance on a binary classification task can be assessed
using modeling accuracy, true positive rate (sensitivity), and true negative rate (specificity):

\[
Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{5.5}
\]

\[
True Positive Rate = TPR = sensitivity = \frac{TP}{TP + FN} \tag{5.6}
\]

\[
True Negative Rate = TNR = specificity = \frac{TN}{FP + TN} \tag{5.7}
\]

Using these metrics, the false negative rate (FNR = 1 - sensitivity) and the false positive rate (FPR = 1 - specificity) can also be used for assessment of model results.

When imbalance exists in a modeling dataset, a reliance solely on accuracy as an overall indicator of performance is unsuitable as it is inevitably biased towards the majority class (i.e. models that predict only the majority class will appear to have predictive power, when this is not actually what is desired). Furthermore, class imbalance causes an inevitable trade-off between the FNR and the FPR and it is the modeler’s task to consider the consequences of both types of misclassification when determining which one is of greater importance. In the case of sewer condition prediction, a false positive leading to the inspection of a pipe that is actually in good condition is less of an issue than a false negative (where a pipe leaking raw wastewater into the ground is missed in the next round of inspections as it was predicted to be in good condition). One technique for determining the true
predictive power of the model and for evaluating the resulting trade-off between the two types of misclassification caused by data imbalance is the receiver operating characteristic (ROC) curve, which is a plot of the true positive rate as a function of the false positive rate (Figure 5.3). A perfect model (with a 100% true positive rate and 100% true negative rate) would have an area under the ROC curve of 1 and would be a line that passes through the upper left corner of the plot. An ineffective model (one no better than randomly guessing pipe condition) would have an area under the ROC curve of approximately $0.5 \times 1 \times 1 = 0.50$ and is represented by the diagonal line in Figure 5.3. Hosmer and Lemeshow (2000) indicate a model that achieves an area under the ROC above 0.9 is outstanding, between 0.8-0.9 is excellent and between 0.7-0.8 is acceptable. Hence, if comparisons are to be made between models, the most effective model is the one with the largest area under the ROC curve (Kuhn and Johnson, 2013).

![Figure 5.3: The Receiver Operating Characteristic (ROC) Curve.](image-url)
As previously indicated, the random forest algorithm uses a baseline probability cut-off for classification, where for this binary classification task, any pipe with a predicted probability of being in bad condition greater than 50% will be assigned to the bad condition class. The ROC curve can be used to establish alternative probability cut-offs (e.g. any pipe with a predicted probability of being in bad condition greater than 15% is assigned to the bad condition class) that can potentially overcome complications posed by data imbalance. One option is to locate the cut-off on the ROC curve that is closest to the upper left corner of the plot (i.e. closest to the optimal model). Alternatively, a new cut-off can be determined using Youden's J index (Youden, 1950):

\[ J = \text{Sensitivity} + \text{Specificity} - 1 = TPR + TNR - 1 \]  (5.8)

Youden's J index is computed for each cut-off used to create the ROC curve, and the cut-off associated with the largest value of J may be superior to the baseline cut-off of 50%. A new cut-off that is closest to the optimal model or established using Youden's J index does not change the developed model (as the same model parameters are being used) and is only used to increase the sensitivity of the model to the minority class of interest.

The available inspection dataset was partitioned into training, evaluation and test sets using a 70-10-20 split ratio and stratified random sampling (to preserve the distribution of class instances in each dataset). The training set was used to train the model, the evaluation set was used during the process of evaluating alternative probability cut-offs, and the test set was used to provide an unbiased source of
predictive capability on previously unseen instances. An additional measure of predictive performance was obtained by performing ten-fold cross-validation during the training process.

5.6 Results

The test set confusion matrix for Random Forest Model 1 (1500 trees and \( m = 3 \) predictors considered when making tree splits) developed using a training dataset consisting of 138 bad pipes and 1117 good pipes is presented in Table 5.6. Although the model achieved a test set overall accuracy of 90\%, the confusion matrix indicates the data imbalance has forced the algorithm to concentrate on correctly classifying good pipes in the majority class with the consequence that very few bad condition pipes are correctly predicted. The area under the ROC curve obtained during ten-fold cross-validation of the training set was 0.81. During model testing, the following metrics were obtained: true positive rate = 11\% (4 out of 38 bad pipes correctly predicted as being bad) and true negative rate = 99\% (316 out of 318 good pipes correctly predicted as being good). The model achieved an excellent area under the ROC curve of 0.81, suggesting it would be capable of accurately predicting pipe condition once a new probability cut-off is optimally determined using the ROC curve.
Table 5.6: Test set confusion matrix for random forest model 1 using cutoff of 0.50

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>4</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>2</td>
<td>316</td>
<td></td>
</tr>
</tbody>
</table>

True Positive Rate = Sensitivity = \(\frac{4}{34 + 4} = 11\%\)

False Negative Rate = \(\frac{34}{4 + 34} = 89\%\)

True Negative Rate = Specificity = \(\frac{316}{316 + 2} = 99\%\)

False Positive Rate = \(\frac{2}{2 + 316} = 1\%\)

Accuracy = \(\frac{4 + 316}{4 + 34 + 2 + 316} = 90\%\)

Area under the ROC curve = 0.81

The ROC curve shown in Figure 5.4 presents the true positive and false positive rates achieved when various probability cut-offs are used by the random forest when classifying instances in the evaluation set. Table 5.7 presents the confusion matrix when the baseline cutoff of 0.50 is used to assign pipes in the evaluation set to a condition class (i.e. a pipe is predicted to be bad if the probability of being bad is greater than 50%). The baseline cutoff achieves an unacceptably low true positive rate of 10% for the evaluation set (only 2 of 20 bad pipes in the evaluation set correctly predicted as being bad).
Figure 5.4: The ROC curve for Random Forest Model 1.

Table 5.7: Random Forest Model 1: baseline cutoff (0.50) used to classify evaluation set

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
<th>True Positive Rate</th>
<th>False Negative Rate</th>
<th>True Negative Rate</th>
<th>False Positive Rate</th>
<th>Accuracy</th>
<th>Area under the ROC curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>2</td>
<td>10%</td>
<td>90%</td>
<td>99%</td>
<td>1%</td>
<td>89%</td>
<td>0.81</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Accuracy = (2 + 159) / (2 + 18 + 1 + 159) = 89%
Alternatively, a probability cut-off of 0.140 (closest to the upper left hand corner of the ROC curve and therefore closest to the optimal model) achieves a true positive rate of 85% (17 out of 20 bad pipes in the evaluation set correctly predicted as being bad) and a false positive rate of 27% (43 out of 160 good pipes in the evaluation set incorrectly predicted as being bad) (Table 5.8).

**Table 5.8:** Random Forest Model 1: cutoff closest to the upper left hand corner of the ROC curve (0.140) used to classify the evaluation set.

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Bad (ICG 4-5)</th>
<th>Good (ICG 1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad (ICG 4-5)</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>43</td>
<td>117</td>
</tr>
</tbody>
</table>

True Positive Rate = Sensitivity = (17) / (17 + 3) = 85%

False Negative Rate = (3) / (17 + 3) = 15%

True Negative Rate = Specificity = (117) / (43 + 117) = 73%

False Positive Rate = (43) / (43 + 117) = 27%

Accuracy = (17 + 117) / (17 + 3 + 43 + 117) = 74%

Area under the ROC curve = 0.81

The cut-off maximizing Youden’s index in the evaluation set is 0.125:

\[ J = TPR + TNR - 1 = 0.95 + 0.69 - 1 = 0.64 \]  \hfill (5.9)

Using this cut-off, any pipe with a predicted probability of being bad greater than 12.5% will be classified as being bad. This cut-off achieved a TPR of 95% for the evaluation set (19 out of 20 bad pipes correctly predicted as being bad) and a FPR of 31% (50 out of 160 good pipes incorrectly predicted as being bad) (Table 5.9).
**Table 5.9**: Random Forest Model 1: cutoff maximizing Youden’s Index (0.125) used to classify the evaluation set.

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Bad (ICG 4-5)</th>
<th>Good (ICG 1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad (ICG 4-5)</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>50</td>
<td>110</td>
</tr>
</tbody>
</table>

- True Positive Rate = Sensitivity = \( \frac{19}{19 + 1} = 95\% \)
- False Negative Rate = \( \frac{1}{19 + 1} = 5\% \)
- True Negative Rate = Specificity = \( \frac{110}{50 + 110} = 69\% \)
- False Positive Rate = \( \frac{50}{50 + 110} = 31\% \)
- Accuracy = \( \frac{19 + 110}{19 + 1 + 50 + 110} = 72\% \)
- Area under the ROC curve = 0.81

Of the available options, the cutoff that maximizes Youdens index is best suited to the task of identifying bad condition pipes as it achieved the highest evaluation set true positive rate. The 0.125 cutoff established using the evaluation curve can then be used to predict the condition of pipes in the stratified test set (Table 5.10).

The random forest is now significantly better at predicting bad condition pipes as it achieves a test set:

- TPR = 82% (31 out of 38 bad pipes correctly predicted as being bad),
- FNR = 18% (7 out of 38 bad pipes incorrectly predicted as being good),
- TNR = 73% (232 out of 318 good pipes correctly predicted as being good),
- FPR = 27% (86 out of 318 good pipes incorrectly predicted as being bad),
• Overall accuracy = 74% (263 correctly classified pipes out of the 356 in the test set) and,

• an area under the ROC curve = 0.81.

Although the overall accuracy of the model is lower when using this new cutoff (the result of the inevitable increase in the false positive rate that occurs when implementing a class imbalance learning strategy aimed at improving the true positive rate), this reduction in accuracy is acceptable as false negatives (and their consequences) are of greater concern than a false positive.

**Table 5.10:** Test Set Confusion Matrix for Random Forest Model 1 using maximum Youden’s index cut-off of 0.125.

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Actual Condition</th>
<th>Bad (ICG 4-5)</th>
<th>Good (ICG 1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>31</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>86</td>
<td>232</td>
<td></td>
</tr>
</tbody>
</table>

True Positive Rate = Sensitivity = \( \frac{31}{31 + 7} = 82\% \)  
False Negative Rate = \( \frac{7}{7 + 31} = 18\% \)  
True Negative Rate = Specificity = \( \frac{232}{232 + 86} = 73\% \)  
False Positive Rate = \( \frac{86}{86 + 232} = 27\% \)  
Accuracy = \( \frac{31 + 232}{31 + 7 + 86 + 232} = 74\% \)  
Area under the ROC curve = 0.81
5.7 Discussion

5.7.1 A Novel Approach to Enhancing Predictive Power

As demonstrated below, gains in predictive power of random forest models can be achieved by including additional pipe-specific attributes reflecting the condition of nearby pipes that have already been inspected. This novel approach to the task of improving individual pipe condition prediction can be achieved by obtaining the following information for each pipe in the dataset:

- WRc peak, mean, and total structural condition scores of an attached upstream pipe (e.g. Pipe X is attached to Pipe Y at its upstream manhole. Pipe Y already was inspected and was found to have a WRc peak score = 60, a mean score = 2.2, and a total score of 105).

- WRc peak, mean, and total structural condition scores of any inspected pipe that is attached at a downstream manhole (e.g. Pipe X is attached to Pipe Z at its downstream manhole. Pipe Z has already been inspected and was found to have a peak score = 23, a mean score = 0.9, and a total score of 82).

On the basis as outlined above, three numeric attributes are then added to the modeling dataset to reflect neighboring pipe condition by obtaining the maximum WRc peak, mean and total structural condition scores from the pipes attached at either the upstream or downstream manhole (e.g. the maximum neighboring WRc scores that will be included in the modeling dataset as attributes of Pipe X are neighbor peak = 60, neighbor mean = 2.2, and neighbor total = 105).
Consideration of either the upstream or downstream manhole reduces the likelihood that neighboring pipe information will be missing for any pipes. For the modeling dataset under consideration, a total of 34 pipes were not attached to any upstream or downstream pipe that had already been inspected (the mean neighboring WRc condition scores obtained from the modeling dataset were substituted for these 34 pipes). Figure 5.5 provides the evaluation set ROC curve for Random Forest Model 2, developed using the expanded dataset containing the three new attributes describing the maximum peak, mean and total structural condition scores of any neighboring pipe. The poor predictive performance of the baseline cut-off (0.50) can be overcome by using a cut-off closest to the optimal model (0.145) as it achieves an evaluation set true positive rate of 90%. A slightly higher evaluation set true positive rate of 100% for the evaluation set can be achieved using the maximum Youden’s index cut-off of 0.134. The test set confusion matrix, with the cut-off of 0.134, is shown in Table 5.11. It indicates the model achieved the following performance metrics:

- True positive rate = 89% (increased from 82% achieved for a random forest model developed using a dataset that did not include upstream or downstream pipe condition information),
- False negative rate = 18% (decreased from 18%),
- True negative rate = 75% (increased from 73%),
- False positive rate = 25% (decreased from 27%),
- Overall accuracy = 76% (increased from 74%) and,
- an area under the ROC curve = 0.85 (increased from 0.81).
Figure 5.5: The ROC Curve for Random Forest Model 2 (with upstream and downstream condition information).

Table 5.11: Test set confusion matrix for Random Forest Model 2 with upstream and downstream condition information (probability cut-off maximizing Youden’s index).

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Bad (ICG 4-5)</th>
<th>Good (ICG 1-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad (ICG 4-5)</td>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>81</td>
<td>237</td>
</tr>
</tbody>
</table>

True Positive Rate = Sensitivity = \( \frac{34}{34 + 4} = 89\% \)

False Negative Rate = \( \frac{4}{34 + 4} = 11\% \)

True Negative Rate = Specificity = \( \frac{237}{237 + 81} = 75\% \)

False Positive Rate = \( \frac{81}{81 + 239} = 25\% \)

Accuracy = \( \frac{34 + 237}{34 + 4 + 81 + 237} = 76\% \)

Area under the ROC curve = 0.85
The reduced false negative rate combined with the increase in the area under the ROC curve suggests the inclusion of neighboring pipe condition information has improved the predictive capabilities of the model for the bad condition class of interest.

5.7.2 Predicting Uninspected Pipe Condition

The performance metrics achieved for both of the random forest models suggest they will provide accurate predictions of structural condition for uninspected pipes in the Guelph sanitary sewer network. Of the remaining pipes in the Guelph network that have yet to be inspected, 3949 have no directly upstream or downstream pipes and can be modeled by Random Forest Model 1 for condition prediction. The model predicts 3357 of these pipes will be in good condition and 592 will be bad. The pipes in this dataset have an average age of 24 years and the majority (2499) were installed after 1991. As indicated by the predictions of Random Forest Model 1, most of these pipes are likely to be in good condition as they are quite new to the sewer network. An additional 708 pipes that have not yet been inspected have neighboring pipe inspection information available and can have their condition predicted by the more powerful Random Forest Model 2. The average age of these pipes is 49 years and it appears advanced age has left many of them in an advanced state of disrepair as 349 are predicted by the random forest to be in bad condition.

As both random forest models achieved high true positive rates during model testing, it can be expected that very few bad uninspected pipes in the network will be missed if every predicted bad uninspected pipe is targeted during the next scheduled rounds of inspection.
5.7.3 Gauging the Success of Pipe-Level Models

Beyond basic characteristics such as material of construction and age, sewer pipe deterioration rates are likely highly dependent on a variety of factors that are often not available for analysis including original design quality control, local environmental conditions, history of extreme loading events, operating context, and maintenance history since installation. An absence of this information serves to make the characterization of individual pipe condition rather challenging.

Given these restrictions on predictive performance, the primary gauge of modeling success at the individual pipe level should be whether or not a model is capable of improving how municipalities plan future inspections. The cumulative gains chart provided in Figure 5.6 is commonly used within the field of data mining to gauge the effectiveness of predictive models. The chart plots the percentage of the total number of bad pipes that would be gained (or detected) by targeting a percentage of the top predictions made by the random forest models. As an example, if the 356 pipes in the test are ranked from highest to lowest in terms of likelihood of being in bad condition (as predicted by the random forest), the top 30% of the predictions made by Random Forest Model 2 (using neighboring condition information) would capture 83% of all the bad pipes in the dataset. In other words, 0.3 x 356 = 107 inspections would identify 32 bad pipes. There are only 38 bad pipes in the entirety of the testing dataset, and hence 32/38 = 83% of all the bad pipes have been found in only 107 inspections. In a similar manner, the top 30% of predictions made by Random Forest Model 1 (without neighboring pipe condition) would identify 73% of all the bad pipes in the dataset (i.e. 107 inspections would identify 0.73 x 38 = 28 of the bad pipes).
Figure 5.6: Gain chart for the random forests (the shaded grey region indicates the lift that would be achieved with a perfect model).

The diagonal line in Figure 5.6 indicates that without the random forests, 83% of the entire dataset, or 0.83 x 356 = 295 pipes would need to be inspected to find 32 bad pipes. Both random forests provide the municipality with an opportunity to capture significantly more information on bad condition pipes with less effort. As such, the models can significantly reduce the time and money spent during the next round of inspections. Using the test set as an example, without the model, the CCTV inspection cost would be $41,300 but only $14,980 if the random forest model was used (considering an average pipe length of 70 m and CCTV cost of $2 per inspected meter of pipe). In the same manner, inspections of larger sets of candidate pipes (i.e. 3,000 vs. 300) guided by the random forest model would result in considerably larger cost savings.
5.8 Conclusion

Aging sanitary sewers are out-of-sight and out-of-mind and decades of neglect have had a negative impact on the reliability of this important national capital asset. A long list of evidence suggests defective sewers across Canada pose a major environmental and economic threat. This threat can be addressed by adapting proactive management practices within the wastewater industry. A variety of modeling approaches have been proposed that use information gained from CCTV inspections to model the deterioration of sanitary sewer pipes. While some proposed approaches are suitable for achieving an overall estimation of network condition, very few validated approaches exist for predicting individual pipe condition and there is room for improvement in the way municipalities learn from existing inspection datasets.

A random forest predictive model developed to predict individual sanitary sewer pipe condition in Guelph, Ontario, Canada achieved a stratified test set true positive rate = 82%, false negative rate = 18%, true negative rate = 71%, false positive rate = 29% and an “excellent” area under the ROC curve of 0.81. These performance metrics indicate the random forest approach is highly capable of learning from existing inspection datasets so that reliable predictions of individual pipe condition are available for asset management purposes.

A novel approach of including upstream and downstream condition information for individual pipes in the modeling dataset enhances the predictive power of the random forest model for bad pipes representing the minority class of interest (true positive rate = 89%, false negative rate = 11%, true negative rate = 75%, false
positive rate = 25%, overall accuracy = 76% and an area under the ROC curve of 0.85).

Use of the random forest predictive models has the potential to significantly reduce the time and money allocated to identification of bad condition, uninspected pipes in a sanitary sewer network. Gain chart analysis indicates significantly more bad condition pipes can be found in a shorter amount of time than was previously possible prior to data mining. The dataset reserved for testing the predictive capabilities of the model indicates the CCTV inspection cost of identifying 32 bad pipes in the sewer network without data mining would be $41,300. The cost of locating the same number of bad pipes using inspections guided by the random forest model is reduced to $14,980 as the significantly higher target response rate drastically reduces the number of required inspections. The combination of random forest predictive models with gain chart analysis provides municipalities with an opportunity to accumulate savings as larger sets of uninspected pipes are considered during future rounds of CCTV inspection.
Transition to Chapter 6

This next chapter further explores the use of random forests for predicting sewer pipe condition. The case study area (Guelph, Ontario) is used to demonstrate the predictive capabilities when class-imbalance learning strategies (down-sampling and threshold adjustment) are used to optimally tune the model to predict the minority class of interest. An analysis of inspection records indicates cracks, fractures and defective joints are the most commonly recorded structural defects - which is in-line with other defect observations across Canada. Model predictions are combined with network spatial analytics as a method of screening pipes for future inspection based on their likelihood of being located within regions of the sewer network that are experiencing clusters of structurally deteriorated sewer pipes.

This paper has been peer-reviewed and is published in the International Journal of Environmental Protection:

Chapter 6

Predictive and Spatial Analytics for Planning Inspections of Sewer Infrastructure

6.1 Introduction

The majority of North American sewer infrastructure was installed during the period of rapid economic expansion that followed the conclusion of the Second World War. The first phase of this infrastructure is rapidly approaching the end of its useful life and a growing list of evidence suggests many sewers are in an advanced state of disrepair. In the 2013 Report Card for American Infrastructure, a poor condition grade was assigned to the 700,000 miles of publicly owned sewer mains currently in operation (ASCE, 2013). Aging pipes represent the largest capital investment need comprising three-quarters of the estimated $298 billion
of capital investment required over the next twenty years to address wastewater infrastructure (ASCE, 2013).

Structural defects in sewer pipes allow raw, untreated wastewater to leak into the surrounding environment. This phenomenon is referred to as exfiltration and is of great concern to municipalities as a range of groundwater pollutants are found in typical municipal wastewater flows (Bishop et al., 1998). There are numerous historical instances of contamination to validate this concern - e.g. a leaking sewer contaminated an aquifer in the Bath region of the United Kingdom in 1928, resulting in a typhoid outbreak affecting 50 people (Halliday, 1992) and a total of 54 separate incidents of sewer-related groundwater contamination have been recorded in the United Kingdom alone (Bishop et al., 1998). Although published data on the rates of exfiltration of raw sewage from aging sewers is rather limited, estimated rates have been established for some international locations e.g. 18% of average daily flow in Munich, Germany (Mull et al., 1992), 10% of dry weather flow in Rome, Italy (Cardoso et al., 2005), and 8% of total pumped volume in Hong Kong (Lerner, 1997).

In municipalities with separate sanitary and stormwater networks, the issue of sewage exfiltration is exacerbated when sewage leaking from broken sanitary sewers passes untreated into any broken stormwater pipes located nearby. Sewage contamination of a dedicated stormwater system poses a significant risk to human health as stormwater pipes release their untreated contents to surface waters, such as lakes used for recreational pursuits. A number of recent studies indicate sewage contamination of stormwater systems is nearly ubiquitous in the urban environment. Approximately half of 18 stormwater outfalls studied in Milwaukee,
Wisconsin contained more than 25% sanitary sewage by volume and were linked to poor water quality of receiving waters in the area (Sauer et al., 2011). Individuals swimming in recreational waters located near a stormwater outfall studied in Haile and Alamillo (1996) were 50% more likely to experience adverse health effects (e.g. gastroenteritis) than those swimming further away from the same outfall. Sewer leaks located near building connections were a significant source of faecal contamination in stormwater systems recently studied in central Singapore (Doshi, 2012). A study conducted in California indicates stormwater pipes act as conduits for raw sewage leaking from failed sanitary sewers, even during periods of dry weather (Sercu et al., 2011).

Raw sewage that flows untreated into surface waters is potentially very dangerous as it may contain a wide range of bacteria, pathogenic microorganisms and endocrine disrupting compounds (EDCs - commonly found in human pharmaceuticals, detergents and chemical disinfectants). These EDCs threaten long-term environmental and human health as they interfere with natural processes in the body responsible for reproduction, development and behavior. The pharmaceuticals carbamazepine (an anticonvulsant and mood-stabilizing drug) and sulfamethoxacol (an antibiotic commonly used to treat urinary tract infections) were not completely eliminated in soil columns studied in Cordy et al. (2004), suggesting they are capable of passing from leaking sanitary sewers to nearby stormwater pipes. Ultimately, chronic faecal contamination of recreational waters ultimately reduces the potential economic benefit of water-based recreation and tourism activities as municipalities are forced to close recreational waters to protect users from illness and disease (NRDC, 2012). Wastewater leaking out of structurally damaged sewer pipe is also
capable of contaminating drinking water distribution systems during transient low or negative pressure events, further compounding the potential consequences of aging infrastructure (Teunis et al., 2010).

Until quite recently, most municipalities in North America carried out sanitary sewer-related work primarily as part of basement flooding relief studies (generally limited to qualitative inspections of select sewer pipes representing a small portion of the entire sewer system), as part of roadway rehabilitation projects (where older pipes are dug-up and replaced when new roads are constructed), or in response to sewer failure or blockages. A major shift in municipal accounting practice has had a dramatic impact on the way sewers are now managed in North America. Historically, tangible assets such as sewer pipes were only recorded as expenditures in their year of installation and the value of existing assets did not appear on annual financial reports. With the introduction of GASB Statement 34 in the United States in 1999 and PSAB Statement 3150 in Canada in 2009, capital asset inventory is now reported using accrual accounting methods reflecting the fact that assets such as sewer pipes have a value long after their initial cost of construction is incurred, but this value depreciates over time (GASB, 2012; OMBI, 2007). These changes in accounting practice have improved the accountability of local government to their citizens as the value of existing sewers are now amortized over their useful life with depreciation recorded as an expense on the municipality’s statement of operations.

Accurate information needed to satisfy these new legislative requirements is typically obtained by determining the condition of individual pipes in a sewer system using closed-circuit television (CCTV) inspection technology where a video record
of condition is captured using a camera mounted on a remote-controlled robot driven along the length of the pipe. Although CCTV provides valuable information on defects present inside each inspected pipe (Figure 6.1), these inspections are expensive and budgetary restrictions force most municipalities to limit inspection-related work to portions of their entire network of sewer pipes.

**Figure 6.1**: CCTV technology is used to inspect the internal condition of sewer pipes currently in operation.

In this paper, the random forests data mining algorithm is proposed as an efficient means of leveraging information from existing inspection datasets so that predictions of condition can be made for the remaining pipes that have not yet been inspected. As the case study described herein demonstrates, predictions of pipe condition made available by random forest models can be combined with network spatial analytics that seek out clusters of bad condition pipes. The new information made available by this novel application of predictive and spatial analytics can serve as a valuable screening tool for planning future inspection-related work.

The majority of sewer pipe deterioration models have been based on statistical theory, with the output typically being a binary response (e.g. a yes or no answer to the question of whether or not a pipe will fail), categorical responses, or a matrix
of probabilities for the transition of pipes between condition states. No goodness-of-fit tests were provided for logistic regression models developed in Ariaratnam et al. (2001). Multiple regression models were developed in Chugthai and Zayed (2008) using a small dataset of pipe condition (22 training and 7 for testing asbestos cement; 79 training and 30 for testing concrete; 34 training and 16 for testing for PVC). Fitting data-sensitive sample means to a full population mean is inherently problematic, particularly when very few poor condition pipes are available for model development and validation (Tran et al., 2005). Salman (2010) explored a variety of regression models for predicting the condition of sewer pipes in Cincinnati, Ohio - binary logistic regression had the highest prediction accuracy with a correct prediction rate of 46% for structurally poor pipes. Younis and Knight (2010) used 45 km of inspection data collected in Niagara Falls, Ontario to develop network level logistic models for predicting the probability of concrete and vitrified clay pipes being in various condition states. A linear regression model developed in Wright et al. (2006) was believed to greatly under-estimate the length of deficient pipe in a California sewer system. Wirahadikusumah et al. (1998) used Markov-chain models to describe the deterioration of large combined sewers in Indianapolis, Indiana. Limited data made model development problematic. Markov models were used in Sinha and McKim (2007) to estimate surface condition deterioration curves for groups of concrete sewer pipes in Waterloo, Ontario, Canada. Markov models developed to describe the deterioration of groups of stormwater pipes in Australia could only reasonably predict the future condition for cohorts of pipes (Micevski et al., 2002). Markov models and ordinal regression models developed using 27 km of stormwater pipe inspection data collected by an Australian municipality were useful at the system-level (predictive accuracy of the ordinal regression model for
individual pipes was 42%) (Tran et al., 2008). Survival analysis was used to predict structural condition at the network-level for sewer pipes in Quebec City, Quebec in Duchesne et al. (2012). As noted in the state-of-the-art of statistical deterioration methods presented in Ana and Bauwens (2010), cohort survival models are difficult to develop (computationally tedious and they require extensive datasets) and tend to underestimate the number of pipes in the poorest condition states.

While statistical techniques have their own unique sets of advantages, they often require assumptions that limit utility of the developed predictive tool. Data mining techniques, drawing on the fields of artificial intelligence and machine learning, can serve as an alternative when the inherent bias and sparseness of inspection datasets prevent statistical model development. Khan et al. (2010) investigated the importance of parameters related to sewer pipe deterioration using artificial neural networks. Neural networks were found to be superior to both Markov models and ordinal regression models when modeling stormwater pipe deterioration in Tran (2007). Mashford et al. (2011) used support vector machines to predict individual pipe condition in South Australia. Jung et al. (2012) and Syachrani et al. (2012) used decision tree classifiers to investigate sewer pipe condition.

In general, there is still a significant need to develop efficient methods for learning from existing inspection datasets so that location-specific predictions of condition can be made for individual pipes that have not yet been inspected. In the majority of cases, existing modelling techniques (e.g. multiple linear regression, logistic regression and Markov-chains) are of limited utility at the individual-pipe level. As such, this manuscript investigates the possibility of reliably predicting individual pipe condition using the “random forests data mining system. The random
forests algorithm consists of growing hundreds of decision tree classifiers and then combining the predictive capabilities of these individual trees into an ensemble, or forest of trees. Each individual tree is grown to a maximum size by first selecting a random subset of predictor inputs (i.e. pipe attributes) to split the dataset on and then calculating the best split based on the CART algorithm first described in Breiman et al. (1984). Hundreds of these trees are grown in a similar fashion, and classification predictions are made by having each tree in the forest cast a vote and then determining the mode of these votes.

Models consisting of a single decision tree classifier may exhibit high variance and tend to be unstable as their structure will change depending on the instances available for model training. A more powerful approach to modeling deterioration is the random forests system as it has proven to be one of the best performing classification algorithms on a variety of tasks (Kuhn and Johnson, 2013). Random forests are a logical choice for modeling sewer pipe condition as they are efficient for large databases, require minimal parameter tuning, are insensitive to outliers in a dataset, capable of high levels of performance even when faced with class imbalanced datasets, require no assumptions of pipe behavior over time and are inherently capable of identifying the most important input predictors out of a larger candidate set. Some example applications in other fields of research have included ecology (classification of invasive plant species in Cutler et al. (2007)), marketing (predicting customer retention in Lariviere and Poel (2005)) and genetics (analysis of genome wide association datasets in Goldstein et al. (2010)). Further information on random forests can be found in Breiman (2001) and Cutler et al. (2012).
6.2 Case Study

A municipality located in south-western Ontario, Canada was used as a case study area. The 120,000 residents of the municipality rely on a fractured bedrock aquifer as the source of their drinking water. A recent study indicates shallow over-burden thickness and fast groundwater velocities make the aquifer vulnerable to the deterioration of the municipality’s aging sanitary sewer system, with 90% of 22 sampled drinking water wells found to contain at least one sewage-derived contaminant and 45% of the wells exhibited human enteric viruses derived from the exfiltration of domestic sewage flows (Allen, 2013).

The municipality’s sanitary system increased by approximately 100 km every 15 years after WWII to satisfy the demands of increased residential development. Although the average sanitary sewer pipe in the municipality has only been in operation for 38 years (Figure 6.2), approximately 1,900 pipes were installed more than 50 years ago. Pipes are made of a variety of materials, including asbestos cement (11% of system length), concrete (16.1%), PVC (44.6%), reinforced concrete (3.1%), and vitrified clay (25.2%). In general, the oldest pipes in the network are made of vitrified clay and most pipes installed within the past 30 years are PVC.

An engineering consultancy was retained from 2008 - 2011 to help the municipality prepare for compliance with PSAB Statement 3150 by CCTV inspecting a portion of the municipality’s more than 7,000 gravity sanitary sewer pipes. A total of 221 km of inspections were completed over this time period, with pipes selected for inspection based on expert opinion that determined those pipes would have an increased likelihood of being in poor structural condition. Structural defects inside
each pipe were identified using the third edition of the Water Research Center Manual of Sewer Condition Classification (WRc MSCC) and severity scores were assigned to defects using the fourth edition of the Water Research Center Sewerage Rehabilitation Manual (WRc SRM). An internal condition grade (ICG) of 1, 2, 3, 4 or 5 was assigned to each inspected pipe using thresholds established in the WRc SRM for the highest severity scores accumulated in any one meter length of the pipe.

The consultancy carried out comprehensive quality assurance and quality control (QA and QC) to ensure accuracy of the collected inspection data. Approximately 90% of the inspected pipes were assigned an ICG of 1, 2, or 3 (Table 6.1). As a result, the inspection dataset is class imbalanced, with significantly more pipes
in some condition classes than others. Class imbalance is common within pipe inspection datasets (as indicated in Salman (2010) and Ana et al. (2008) and is problematic for predictive modeling as most algorithms will pursue high levels of accuracy by focusing their learning effort on correctly predicting instances belonging to the majority class at the expense of misclassifying minority instances. Fortunately, the data mining community offers a variety of class imbalance learning techniques the majority designed with the intention of solving two-class problems, as accounting for class imbalance is exceedingly complex beyond two classes.

Table 6.1: Internal condition grades assigned to inspected pipes.

<table>
<thead>
<tr>
<th>Internal Condition Grade (ICG)</th>
<th>Condition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td>Good (ICG 1-3)</td>
<td>Bad (ICG 4-5)</td>
</tr>
<tr>
<td>Asbestos Cement 373 14 24 5 0</td>
<td>411</td>
<td>5</td>
</tr>
<tr>
<td>Concrete 364 79 73 36 7</td>
<td>516</td>
<td>43</td>
</tr>
<tr>
<td>PVC 110 9 1 2 1</td>
<td>120</td>
<td>3</td>
</tr>
<tr>
<td>Reinforced Concrete 56 8 0 0 0</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>Vitrified Clay 213 135 165 121 29</td>
<td>513</td>
<td>150</td>
</tr>
<tr>
<td>Total 1116 245 263 164 37</td>
<td>1624</td>
<td>201</td>
</tr>
</tbody>
</table>

Pipes were reclassified as being in *good* (ICG 1 - 3) or *bad* (ICG 4 - 5) structural condition to accommodate the following class-imbalance learning strategies:

- Down-sampling (i.e. balanced random forests): where for each tree grown in the random forest, a bootstrap sample of instances is drawn from the minority class and the same number of instances are randomly drawn (with replacement) from the majority class.

- Alternative classification cut-offs: where a random forest is grown but the
model is tuned to improve performance on the minority class by identifying a classification cutoff that is more effective than the baseline threshold of 0.50 typically used to convert class probabilities to discrete class predictions (Kuhn and Johnson, 2013).

Pipes with an ICG of 4-5 are the minority class of interest as their defects make them susceptible to failure and exfiltration-related issues. An analysis of the inspection data indicates reduced structural integrity was most often the result of cracks, fractures and defective joints as they represent 42%, 38% and 11% of all recorded structural defects, respectively (Figure 6.3). This appears to be the case for many Canadian sewer systems, as cracks were previously found to be the most common structural defect in a study of inspection datasets obtained from six other Canadian municipalities in (Newton and Vanier, 2006).

CCTV inspection records were integrated with existing spatial data in GIS to prevent under-utilization of the data. ESRI ArcMap was found to be an excellent document management system that saved considerable time in locating, organizing and confirming the accuracy of field-inspection information. Pipe-specific attributes obtained from GIS used as input predictors for model development include: material of construction (asbestos cement, concrete, PVC, reinforced concrete or vitrified clay), age at the time of inspection (years), installation era (pre-WWI (< 1914), WWI (1914-1918), inter-war (1919-1938), WWII (1939-1945), post-WWII (1946-1966), or modern (1967-present)), sewer type (trunk or branch), diameter (mm), length (m), slope (%), slope change (maximum % change in slope at either the upstream or downstream pipe connection), upstream invert elevation (m), orientation change (maximum change in pipe orientation at either the upstream
or downstream connection), burial depth (m), road coverage (portion of a pipe’s total length covered by an overlying roadway), number of nearby water main breaks that have occurred within 3 m of the sewer pipe, land use (agricultural, commercial, industrial, institutional, park or residential), and census tract (one of 27 districts within the municipality pre-determined by government for census purposes). This set of predictor inputs was analyzed to ensure there were no near-zero variance predictors, nor were there any strong correlations between predictor variables. Further identification of important input predictors (e.g. feature elimination) was unnecessary due to capabilities inherent in the random forests algorithm for such purposes.

Integration of inspection data within a GIS environment afforded an opportunity
to identify clusters/hotspots of bad condition inspected pipes. Network spatial analysis techniques can be used for cluster identification that establish proximity using the shortest-path distance between events located on the network. A recently introduced third party application for ESRI ArcMap known as SANET (Okabe and Sugihara, 2012) can be used to efficiently implement network spatial analysis without the burden of excessive mathematical manipulations of the dataset.

Figure 6.4 presents a hotspot map of bad condition inspected pipes developed using SANET (equal-split continuous kernel density function, midpoint of a bad pipe considered to be a point of failure, and a shortest path distance between events of 200 m). Hotspot severity can be gauged using threshold values of standard deviation from the mean network density - e.g. pipes having a density greater than 155.
than 2.5 standard deviations above the mean can be considered to be a hotspot of serious concern due to the number of structurally poor pipes located within close proximity.

6.3 Assessing Predictive Capability

Random forests generate both a continuous-valued prediction (in the form of a class membership probability between 0 and 1) and a discrete class prediction based on the class membership probability (using a baseline probability cut-off of 50%). Based on the discrete class predictions, predictive capability was evaluated using the confusion matrix in Table 6.2 - where bad condition pipes are defined as the positive, minority class of interest for inspection planning purposes. The inspection dataset was partitioned using stratified random sampling into training, evaluation and test sets using a 70-10-20 split ratio.

**Table 6.2:** Confusion matrix used to evaluate predictive performance on a binary classification task

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bad</em> (ICG 4-5)</td>
<td><em>Bad</em> (ICG 4-5)</td>
</tr>
<tr>
<td><em>Good</em> (ICG 1-3)</td>
<td><em>Good</em> (ICG 1-3)</td>
</tr>
</tbody>
</table>

In addition to overall accuracy (which can be an unreliable indicator of predictive capability when dealing with class imbalanced data), model performance was assessed using the metrics true positive rate, true negative rate, false positive rate, and false negative rate. The area under the receiver operating characteristic
(ROC) curve provided further indication of predictive power (with excellent models generally defined as those achieving an area under the ROC > 0.8). The ROC curve, a commonly used tool for evaluating the success of a data mining investigation, is a plot of the true positive rate versus the false positive rate achieved when different probability thresholds are used to establish classifications for instances in a dataset. For each candidate threshold (e.g. 50%) the true positive rate and false positive rate are plotted against each other. More information on the use of these metrics for gauging model performance can be found in Han et al. (2006).

6.4 Results

Random forests were trained, tuned and tested using the randomForest (Liaw and Wiener, 2002) and caret (Kuhn, 2013) packages developed for R.

Table 6.3 contains the test set classification results for Random Forest Model 1 (using the entire set of available input-predictors, no class-imbalance learning strategies, the baseline classification cutoff of 0.50, and maximum area under the ROC curve used during model tuning to establish the optimal number of input predictors randomly select to grow each tree in the forest). The confusion matrix indicates class imbalance had the negative consequence of a predictive model that focused almost entirely on correctly classifying good pipes in the majority class in the pursuit of high overall accuracy. The area under the ROC curve achieved during three-repeats of ten-fold cross-validation of the training set was 0.81. Area under the ROC curve for the evaluation and tests sets was 0.82 and 0.81, respectively. These ROC results indicated the model would be capable of being an excellent
classifier for bad pipes if the model could be optimally tuned using an alternative
cutoff that is more appropriate than the baseline classification cutoff.

Table 6.3: Test set confusion matrix for random forest model 1: baseline
classification cutoff of 0.50

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>34</td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>316</td>
</tr>
</tbody>
</table>

True Positive Rate = Sensitivity = (4) / (34 + 4) = 11%
False Negative Rate = (34) / (4 + 34) = 89%
True Negative Rate = Specificity = (316) / (316 + 2) = 99%
False Positive Rate = (2) / (2 + 316) = 1%
Accuracy = (4 + 316) / (4 + 34 + 2 + 316) = 90%
Area under the ROC curve = 0.81

Table 6.4 contains the test set classification results for Random Forest Model 2 (the
baseline Random Forest Model 1 with an optimal cutoff of 0.125 derived from the
evaluation set ROC curve). Using this new optimal cut-off, any instance predicted
by Model 1 to have a probability of being bad > 0.125 is classified as being bad.
Using this new cutoff, the model achieved a true positive rate of 0.82 (up from 0.11),
a true negative rate of 0.73 (down from 0.99, but still acceptable), and an overall
accuracy of 0.74. As the existing structure of Random Forest Model 2 is the same
as Model 1, it has the same excellent area under the ROC curve characteristics.
Table 6.4: Test Set Confusion Matrix for Random Forest Model 2: optimally tuned classification cutoff of 0.125.

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad (ICG 4-5)</td>
<td>31</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Good (ICG 1-3)</td>
<td>86</td>
<td>232</td>
<td></td>
</tr>
</tbody>
</table>

True Positive Rate = Sensitivity = \( \frac{31}{31 + 7} \) = 82%

False Negative Rate = \( \frac{7}{7 + 31} \) = 18%

True Negative Rate = Specificity = \( \frac{232}{232 + 86} \) = 73%

False Positive Rate = \( \frac{86}{86 + 232} \) = 27%

Accuracy = \( \frac{31 + 232}{31 + 7 + 86 + 232} \) = 74%

Area under the ROC curve = 0.81

Table 6.5 indicates the test set classification results for Random Forest Model 3: where instances in the majority class were down-sampled during model training. This balanced random forest model was found to be less suited to correctly identifying bad condition pipes than Model 2, achieving a test set true positive rate of 0.71 and a slightly lower test set area under the ROC curve of 0.80.
Table 6.5: Test Set Confusion Matrix for Random Forest Model 3: down-sampling of the majority class.

| Actual Condition | Predicted Condition | | | |
|------------------|---------------------|---|---|
| Bad (ICG 4-5)    | 27                  | 11 | |
| Good (ICG 1-3)   | 75                  | 243|

True Positive Rate = Sensitivity = (27) / (27 + 11) = 71%
False Negative Rate = (11) / (11 + 27) = 29%
True Negative Rate = Specificity = (243) / (243 + 75) = 76%
False Positive Rate = (75) / (75 + 243) = 24%
Accuracy = (27 + 243) / (27 + 11 + 75 + 243) = 75%
Area under the ROC curve = 0.80

6.5 Discussion

6.5.1 Planning Future Inspections through Combined Spatial and Predictive Analytics

The increased true positive rate and slightly higher area under the ROC curve indicate Random Forest Model 2 (optimal probability cut-off) performs better than Random Forest Model 3 (down-sampling) when predicting the minority class of interest in the test set. Overall, the performance metrics obtained for Model 2 indicate it would be capable of providing reliable predictions of condition for uninspected pipes in the sanitary sewer network.

Of the more than 4,000 pipes in the system that have not been inspected, Model 2 predicts 1073 will be in bad structural condition. Although the immediate
inspection of these 1073 would be preferred, budgetary restrictions will most likely limit the number of pipes that can be inspected within the coming year. A variety of approaches could be used to plan the next round of inspections. One option would be to consider the individual predicted probability of belonging to the bad pipe class in tandem with risk-of-failure concepts (e.g. WRc critical sewers). It would also be possible to select a portion of the most likely bad pipes as the focus for the next round of inspection (i.e. gain-chart analysis as indicated in Harvey et al. (2014)). Alternatively, the municipality can use the combined power of the validated predictive model with the information learned from network spatial analytics to establish a subset of pipes for immediate inspection. In essence, this establishes a risk assessment framework whereby future inspections are based on a predicted probability of being bad and a proximity to existing network hotspots, making the consequence of their failure greater than pipes located outside the hotspot. As an example, of the 1073 uninspected pipes predicted to be bad, 52 are within a hotspot established according to a network density value greater than 2.5 standard deviations above the mean density value for the entire network (Figure 6.5).
Figure 6.5: Predictive and spatial analytics can be combined to identify uninspected pipes that should be targeted for immediate inspection.
6.5.2 Sensitivity Analysis

A sensitivity analysis was performed to evaluate the impact on the predictive capabilities of the random forests algorithm when using a dataset containing significantly fewer inspections than the one available for analysis by the case study municipality. A scenario was established by using one-third of the inspection dataset (consisting of 531 good pipes and 61 bad pipes) to train a random forest model and then keeping the remaining two-thirds of the inspection dataset as a hypothetical set of 1,195 pipes that have not yet been inspected (but would actually contain 1064 good and 131 bad pipes if they were eventually inspected).

The model trained using one-third of the dataset achieved a cross-validated area under the ROC curve of 0.76 and a test set true positive rate and accuracy of 75% and 72%, respectively. Using this trained model to predict the condition of the 1195 pipes in the hypothetical set of uninspected pipes would have resulted in an overall accuracy of 73% (866 pipes correctly classified) and a true positive rate of 79% (104 out of 131 bad pipes correctly identified). This result indicates that although models developed using a significantly a smaller dataset may not have as high a level of performance, trained predictive model can still be of significant utility to a municipality.
6.5.3 Inherent Restrictions on Predictive Capability for Individual Pipes

Beyond basic characteristics such as material of construction and age, sewer pipe deterioration rates are likely highly dependent on a variety of factors that are often unavailable for analysis including original design quality control, local environmental conditions, history of extreme loading events, operating context, and maintenance history since installation. An absence of this information serves to make the characterization of any time-dependent relationship between asset condition and failure difficult. Although drawn from experience within the medical field, Henderson and Keiding (2005) provide some explanation as to why predictions of individual pipe condition will, most likely, always be associated with some degree of uncertainty. Individual pipe models are similar to models developed in the medical field for individual patients, where even the best models are of little use when predicting survival times for terminally ill patients (with expected error rates usually 50% at best as human survival is so uncertain). In much the same way, the behaviour of individual sewer pipes over time is proving to be difficult to model with the same levels of accuracy as generalized network level models. The individual pipe condition modeling task is further complicated by the nature of the data used for condition classification. CCTV inspections are carried out using standardized systems of defect detection but there is still considerable potential for operator subjectivity or experience to introduce noise into the modeling dataset (i.e. some operators may miss certain types of defects resulting in a less severe ICG score for a pipe).
6.6 Conclusions

Widespread evidence of environmental contamination caused by sewer deterioration and higher expectations for sewer performance both call for a more proactive approach to sewer inspection and management. An analysis of CCTV inspections carried out by a municipality in Ontario, Canada indicates the most common structural defects are cracks, fractures and defective joints, representing 41%, 38% and 11% of all recorded structural defects, respectively. The inspection dataset was found to be class imbalanced - with approximately 90% of all pipes in good structural condition. Network spatial analytics in GIS was found to be a useful tool for identifying clusters/hotspots of inspected sanitary sewer in bad structural condition.

The random forests data mining algorithm was used to predict the structural condition of individual sanitary sewer pipes using a set of input predictors derived using GIS. Two class imbalance learning techniques were utilized to improve the predictive capabilities of the algorithm for bad condition pipes representing the minority class of interest. Although down-sampling (balanced random forests) performed well on the classification task at hand, it was outperformed by the simpler approach of establishing a new classification cut-off for the predictive model - an approach that achieved an excellent area under the receiver operating characteristic (ROC) curve of 0.81, an overall accuracy of 74% and a true positive rate of 82%.

Although accurate predictions of condition of individual pipes within a sewer network are often difficult to achieve (due to class imbalance common in inspection
datasets, the uncertain nature of the inspection data, and the considerable variability that can exist in environmental conditions and loading events experienced by individual pipes), the combined application of the random forest algorithm with network predictive analytics was found to be a useful tool for efficiently screening uninspected pipes for future rounds of inspection.
Transition to Chapter 7

This next chapter transitions away from sanitary sewer inspection and explores the potential to use data mining to further municipal understanding of stormwater pipe deterioration. The City of Guelph is again used as a case study area - with decision trees used as a means of visualizing the influence of various pipe-specific attributes on individual pipe condition. This paper has been peer-reviewed and is published in the Journal of Water Management Modeling.

Chapter 7

Understanding Stormwater Pipe Deterioration Through Data Mining

7.1 Introduction

7.1.1 The Threat Posed by Deteriorating Stormwater Pipes

Effective stormwater system management is a challenging task, particularly for older municipalities dealing with pipes that have been in operation for more than 50 years. The replacement value of the 23% of Canadian stormwater pipes currently estimated to be in fair to very poor condition is an estimated $15.8 billion, or $1.270/household (CIRP, 2012). Many of these pipes cannot effectively transport stormwater runoff from impervious surfaces to oceans, lakes or artificial ponds.
as originally intended. An estimated one in ten stormwater pipes are unfit for handling rainfall during extreme weather events and the resulting floods can have significant financial consequences (CIRP, 2012). As an example, 150 mm rainfall over a 3 hour period in Toronto, Ontario in 2005 flooded basements, washed away roads, overloaded treatment plants and destroyed water, wastewater and stormwater pipes. The storm was responsible for more than $500 million in damages and is considered to be the most expensive natural disaster to ever occur in Ontario (Kessler, 2011).

The infiltration of groundwater into stormwater pipes through defective joints, cracks and holes accelerates the process of pipe aging and increases the failure probability of adjacent infrastructure due to the flushing of backfill material around the leak (Karpf and Krebs, 2011). The resulting failure of overlying roadways can be unexpected, as was the sinkhole caused by the failure of a 50 year old stormwater pipe in Ottawa, Ontario that swallowed the car of a motorist and disrupted traffic flow along a major highway for weeks (Hurley, 2012). The failure of a 120 year old brick stormwater pipe in Baltimore, Maryland created multiple sinkholes along a busy street near Johns Hopkins hospital. Repairs to the pipe and other surrounding infrastructure were expected to last several months, with an estimated cost of $7 million (Reutter, 2012).

An indication of further issues associated with stormwater pipe integrity exists because these pipes often carry high levels of fecal indicator bacteria and are a major contributor to poor surface water quality in urban areas (EPA, 2009). Although faecal matter in the stormwater system has traditionally been assumed to originate from the excrement of domestic pets and wildlife, growing evidence suggests
wastewater leaking from defective sanitary sewers is another major contributor to stormwater pipe faecal contamination. The passage of untreated wastewater from defective sanitary sewers into structurally unsound stormwater pipes poses a potentially serious threat to the environment. Stormwater outfalls studied in Milwaukee, Wisconsin were found to be highly contaminated with sewage, suggesting that the pipes were major contributors to poor beach, river and stream water quality (Sauer et al., 2011). Field experiments carried out in California indicate sanitary sewer leaks are responsible for severe sewage pollution in stormwater pipes, even during dry weather flow (Sercu et al., 2011). An estimated 30% of Canadian sanitary sewers are in fair to very poor condition (CIRP, 2012) and it can be reasonably assumed that Canadian urban stormwater systems are experiencing levels of sewage contamination similar to those observed in the United States. Sanitary sewers carry a wide range of chemicals and endocrine disruptors capable of causing long term damage to environmental and human health if they pass untreated into stormwater pipes and thence to surface water. Endocrine disruptors are particularly dangerous as these chemicals interfere with the natural hormones in the body responsible for reproduction, development and behavior (Holtz, 2006). Kolpin et al. (2002) identified a variety of emerging contaminants and other waste-associated chemicals in 80% of 139 rivers and streams surveyed across the United States.

7.1.2 Inspecting the Threat

Many municipalities have recognized the threat posed by deteriorating stormwater pipes and are actively determining pipe condition using CCTV. This non-destructive inspection technique is conducted by recording the internal condition
of a pipe using a small camera mounted on a robot. An operator controls the camera from the surface and reviews the transmitted video on a display screen. The operator is trained to identify defects (e.g. cracks and collapsed sections) and to set defect severity scores according to systems designed to minimize subjective evaluation of the camera footage. One of the most commonly used rating systems is the Water Research Center Manual of Sewer Condition Classification (WRc MSCC) and Sewerage Rehabilitation Manual (SRM). According to this system each inspected pipe is assigned an ordinal internal condition grade (ICG) from 1 (no structural defects) to 5 (collapsed section or collapse imminent) based on defect severity.

Although CCTV inspections provide insight into pipe condition, they are time consuming and expensive. As a result, most municipalities are limited to performing inspections of small portions of their entire stormwater network. Advanced asset management processes capable of extracting valuable information pipe deterioration from existing condition records are therefore required to enhance the value of their inspection efforts.

### 7.1.3 Modeling Pipe Deterioration

Proactive stormwater management can be supported by the use of deterioration models that use information gained from inspection to predict the current and future condition of pipes. Such models are still in their infancy, as models devoted exclusively to stormwater pipes have received significantly less attention than those for water mains and sanitary sewers (Ana and Bauwens, 2010).
Micevski et al. (2002) used Markov models to predict changes in structural condition for stormwater pipes grouped according to factors such as diameter and length in Newcastle, Australia. While their models performed well at the network level, they were not well suited to predicting the condition of individual pipes. Tran (2007) compared the performance of statistical Markov and ordinal regression models against neural networks when predicting stormwater pipe condition in Dandenong, Australia. These models were developed using CCTV data collected from 417 pipe segments, representing 27 km pipe or 2.2% of the total length of the stormwater system. Neural networks were better suited for predicting structural condition (ordinal values ranging from 1 to 3), although they suffered from a number of serious limitations related to model development (e.g. time consuming process of defining the optimal network architecture). Tran and Ng (2010) revisited the Dandenong dataset and support vector machines were found to be capable of marginally outperforming neural networks when predicting stormwater pipe condition. The overall success rate for instances in their dataset reserved for testing performance was 76% for the support vector machine model, compared to 74% for the neural network.

While predictions of condition provided by support vector machines and neural network models are useful for asset management purposes, these models are essentially black boxes. As such, information learned about pipe deterioration is deeply embedded within the models and is not easily interpreted. These black boxes do not explicitly indicate the relationship between attributes and condition and cannot answer questions along the lines of: What pipe diameters have a greater likelihood of being in poor structural condition? Alternatively, pipe deterioration models can
be developed using classification tree algorithms so that knowledge extracted from inspection datasets is represented in an easily understood format.

Classification tree algorithms are among the most powerful approaches for knowledge discovery and data mining (Rokach and Maimon, 2008). They produce interpretable white box models, are robust to missing values and are insensitive to skewed data distributions - making data preparation for modeling less cumbersome than with neural networks and support vector machines (Berry and Linoff, 2004). As a result, classification trees are very popular within the data mining community and are commonly used in the fields of remote-sensing (Tooke et al., 2009), finance (Sahin and Duman, 2011), marketing (Moore and Carpenter, 2010) and medicine (Podgorelec et al., 2002).

### 7.2 Data Mining Methodology

#### 7.2.1 The C5.0 System

The C5.0 system can be used to develop classification tree models capable of predicting individual stormwater pipe condition. The C5.0 system is the successor of the C4.5 system first developed by Australian computer scientist Ross Quinlan (Quinlan, 1993). This new system is significantly faster than C4.5 and can achieve similar results but with smaller, easier to interpret trees.

The C5.0 system constructs classification trees in two phases:

- *Grow the initial classification tree using a divide-and-conquer approach: training instances are split using an attribute that provides the maximum infor-*
motion gain. One branch is created for each possible value of the specified attribute and the dataset is split into two subsets. This process continues for each subset defined by the first split until the subsets cannot be split any further (Witten et al., 2011); and

- **Prune the tree**: remove splits that make an insignificant contribution to the tree’s predictive power (IBM, 2011).

The developed classification tree provides a map of the classification process. Instances are classified by starting at the top (root) node of the tree containing the entire training dataset. The attributes specified by the root node are evaluated and a path is then followed downwards corresponding to the specified attribute splits. Each instance belongs to only one leaf in the tree and the distribution of instances at the leaf determines the resulting classification. By default, a probability threshold of 0.50 is used so that instances are labeled with the most probable class (McCarthy et al., 2005).

### 7.3 Case Study

#### 7.3.1 CCTV Inspections of Guelph’s Stormwater Pipes

The city of Guelph is located in southwestern Ontario and has a population of 120,000. This fifth fastest growing city in Canada relies on a 432 km stormwater network, consisting of 8,073 pipes constructed from 1903 to 2011. The stormwater network has grown at a rate of 100 km every 15 years since 1945 (Figure 7.1).

The average stormwater pipe section is 33 years old and 22% of the pipes have
been in operation for more than 50 years. The average pipe length is 54 m, average diameter is 540 mm, and average pipe slope is 1.45%. Stormwater pipes are predominantly made of concrete (2,853), extra strength concrete (756) and reinforced concrete (2,184). Other pipe materials include PVC (491), polyethylene (162), corrugated steel (168), miscellaneous (55) and unknown (1,421).

![Graph showing the population and total length of stormwater pipe in operation over time.]

**Figure 7.1:** Length of Guelph stormwater system by year.

CCTV inspections of approximately one-third of Guelph’s stormwater pipes were carried out from 2008 to 2011 (Figure 7.2). Defects inside each pipe were reviewed using the third edition of the WRc MSCC. These defects were then assigned a severity score using the fourth edition of the WRc SRM. An internal condition grade (ICG) was then assigned to each pipe based on severity score thresholds established in the WRc SRM.
A dataset of 1,699 instances of stormwater pipe condition (total inspection length 107 km) was used for model development. A total of 759 inspections were excluded from data mining for the following reasons:

- 552 inspected pipes were of an unknown construction material,
- recorded material of construction for 179 inspected pipes did not match the material of construction observed during inspection,
- 23 PVC pipe inspections with construction dates ranging from 1985 to 2003 (every PVC pipe that was inspected had an ICG of 1), and
- 5 vitrified clay pipe inspections.

Of the 1,699 inspected pipes used for model development, 1,073 were made of concrete, 287 were extra strength concrete and 339 were reinforced concrete. Condition grades associated with each pipe material are presented in Table 7.1.

Guelph considers pipes with an ICG of 1 or 2 to be in good condition while those with an ICG of 3, 4 or 5 are in poor condition as they have already failed (holes, broken sections or collapse) or have a reasonable risk of failure in the near future. This creates the following binary modeling task: To develop a classification tree model that is easy to interpret and capable of reliably predicting whether individual pipes belong to the good or poor condition class.
Figure 7.2: Inspected stormwater pipes.
Pipe-specific attributes considered for model development were material of construction, year of construction, diameter, length, slope, upstream invert elevation and downstream invert elevation (Table 7.2). These attributes were obtained using an ArcGIS stormwater inventory database maintained by the municipality. Burial depth, soil conditions around the pipe, and wall thickness were not included in the database and therefore could not be used as inputs for the predictive model. While the database contained information on the strength of the concrete pipes (e.g. class III reinforced concrete), this information was largely incomplete and was considered unfit for modeling purposes.
Table 7.2: Attributes used for model development.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>Nominal</td>
<td>Concrete, extra-strength concrete, or reinforced concrete.</td>
</tr>
<tr>
<td>Construction year</td>
<td>Numeric</td>
<td>1910-2008</td>
</tr>
<tr>
<td>Diameter</td>
<td>Numeric</td>
<td>200 mm - 2100 mm</td>
</tr>
<tr>
<td>Length</td>
<td>Numeric</td>
<td>5 m - 275 m</td>
</tr>
<tr>
<td>Slope</td>
<td>Numeric</td>
<td>0.01 m - 5.58 m rise for every 100 m run</td>
</tr>
<tr>
<td>Upstream invert elevation</td>
<td>Numeric</td>
<td>305 m - 356 m above mean sea level</td>
</tr>
<tr>
<td>Downstream invert elevation</td>
<td>Numeric</td>
<td>305 m - 356 m above mean sea level</td>
</tr>
</tbody>
</table>

Attributes not available for analysis: burial depth, soil condition and wall thickness.

Attributes not suitable for modeling purposes: material strength.

The original dataset (1,699 instances of pipe condition: 1,144 pipes in good condition and 555 in poor condition) was randomly divided into a training set (75% of all instances: 858 good, 416 poor) and a stratified test set (25% of all instances: 286 good, 139 poor). The inspection dataset is class imbalanced, in that there are significantly more pipes in one condition class than the other. The stratified sampling strategy of proportional allocation ensures that the predictive performance of any developed tree is properly evaluated as the class distribution in the test set is the same as the distribution in the original dataset. Attributes were not transformed to a standardized scale as classification trees are largely unaffected by differences in scale between numeric attributes and any transformation would have negatively impacted the resulting interpretability of any developed tree model.
7.3.2 Evaluating Predictive Performance

Predictive performance on a binary classification task is typically evaluated using the confusion matrix shown in Table 7.3 (with poor pipes being designated the positive class of interest). True positives and true negatives represent correctly classified stormwater pipes. False positives occur when a pipe that was actually in good condition is incorrectly predicted to be poor. False negatives occur when a poor pipe is misclassified as being good.

Table 7.3: Condition prediction outcomes.

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Predicted Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor (ICG 3-5)</td>
<td>Poor (ICG 3-5)</td>
</tr>
<tr>
<td></td>
<td>True Positive (TP)</td>
</tr>
<tr>
<td></td>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>Good (ICG 1-2)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td></td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

Note - classification results are often presented using the above format, known as a confusion matrix. The number of instances/examples for which the actual pipe condition observed via CCTV inspection is the column and the condition predicted by the classification tree is the row.

The confusion matrix can be used to determine the overall success rate (OSR - also known as: accuracy) and the false negative rate (FNR) for both the training set and stratified test set:

\[
OSR = \text{Accuracy} = \frac{\text{correctly classified instances}}{\text{total number of instances}} \quad (7.1)
\]

\[
FNR = \frac{\text{poor pipes incorrectly classified as good}}{\text{number of pipes actually in poor condition}} \quad (7.2)
\]
The OSR indicates the predictive capabilities of the developed tree for the two pipe condition classes and the FNR indicates the tendency of the model to misclassify poor condition pipes. A high OSR in combination with a low FNR in a stratified test set ensures accurate and reliable predictions.

Each prediction made by the classification tree is associated with a propensity score. If a prediction of poor condition is made, propensity is the probability of predicting poor at the leaf node ($P$) and is determined by the distribution of instances at the leaf (i.e. if 70% of the instances that make it to the leaf are poor and only 30% are good, then $P = 0.70$). If the predicted condition at the leaf is good, then propensity is $1 - P$.

Data mining algorithms tend to be most effective when there are equal numbers of instances in both classes. Any class imbalance in the training set of instances can have a negative impact on the predictive performance for the minority class as the algorithm will tend to focus model building efforts on correctly classifying instances in the majority. Instances in the minority class will usually be assigned to the majority class as the algorithm seeks to minimize the total number of misclassification errors. As an example, consider a dataset made up of 80% good pipes and 20% poor pipes. The data mining algorithm might choose to assign every pipe to the good class as this would achieve an OSR of 80%. Assigning every pipe to the majority class would mean every poor pipe was incorrectly classified and this would be of limited utility when planning future inspections and understanding stormwater pipe deterioration.

One approach to dealing with data imbalance is to use the cost sensitive learning capabilities of the C5.0 system. The C5.0 system treats misclassification costs as
being equal using the default cost matrix shown in Table 7.4. The combination
of imbalanced data and equal misclassification costs will result in a classifica-
tion tree that focuses on correctly predicting *good* pipes at the expense of misclassifying
*poor* pipes. The cost matrix can be adjusted to force the algorithm to pay greater
attention to the classification of the *poor* pipes in the minority class. If, for
example, the modified cost matrix shown in Table 7.4 is used, misclassified *poor*
pipes cost twice as much as misclassified *good* pipes and the 2:1 cost ratio changes
the probability threshold at each leaf from 0.50 to 0.33. With this cost ratio,
a distribution of 60% good and 40% poor pipes at a leaf node will result in a
classification of *poor* (as the algorithm wants to make the less costly decision
when the distribution of instances are approximately equal). As the true cost
of stormwater pipe condition misclassification is not explicitly known, a variety
of cost ratios can be evaluated. The C5.0 system takes this modified cost matrix
into account when constructing the tree so that expected misclassification costs are
minimized (IBM, 2011).

**Table 7.4:** A comparison of the original and modified cost matrix.

<table>
<thead>
<tr>
<th>Original Cost Matrix</th>
<th>Modified Cost Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Condition</td>
<td>Predicted Condition</td>
</tr>
<tr>
<td>Poor</td>
<td>Good</td>
</tr>
<tr>
<td>Actual Poor</td>
<td>0</td>
</tr>
<tr>
<td>Condition Good</td>
<td>1</td>
</tr>
</tbody>
</table>
7.4 Results

A classification tree predictive model developed without taking into account imbalance in the training set of instances had an OSR of 63% (268 out of 425 pipes correctly classified) and a FNR of 53% when predicting conditions in the stratified test set. The algorithm focused more on accurately classifying good pipes than it did on the poor pipes in the minority class. Its poor predictive performance made it unsuitable as a tool for planning future inspections.

The classification tree developed using the 2:1 modified cost ratio shown in Table 7.4 is shown in Figure 7.3. Classification results for the training and stratified test set are shown in Table 7.5. With a stratified test set OSR of 71% (301 out of 425 pipes correctly classified) and a FNR of 37% the classification tree represents a considerable improvement to the one constructed without accounting for data imbalance. The 2:1 cost ratio proved to be the most effective as larger ratios resulted in poor overall performance for both condition classes (e.g. 3:1 cost ratio resulted in an OSR of 61% with 263 out of 425 pipes in the test set correctly classified.).

Table 7.5: Results for the classification tree.

<table>
<thead>
<tr>
<th></th>
<th>Training Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted Condition</td>
<td>Predicted Condition</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>Good</td>
</tr>
<tr>
<td>Actual Condition</td>
<td>Poor</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>217</td>
</tr>
<tr>
<td>Accuracy</td>
<td>903/1274 = 71%</td>
<td>Accuracy = 301/425 = 71%</td>
</tr>
<tr>
<td>FNR</td>
<td>154/416 = 37%</td>
<td>FNR = 52/139 = 37%</td>
</tr>
</tbody>
</table>

183
Figure 7.3: Stormwater pipe condition classification tree.
Each split in the tree provides insight into the influence of pipe-specific attributes on stormwater pipe structural conditions. The first split divides pipes according to their construction year, where 82% of the stormwater pipes placed after 1968 are in *good* condition. Pipes placed prior to 1968 are rapidly approaching or have already passed 50 years of operation and 47% of them are now in *poor* structural condition (node 1). Although some pipes placed after 1968 are structurally unsound, 82% of these newer pipes are still in *good* structural condition (node 10).

A subsequent split divides older pipes into those with diameter ≥ 450 mm and those with diameter < 450 mm. Approximately half of the older, small diameter stormwater pipes in Guelph are in *poor* structural condition (node 2). Older, large diameter pipes on the other hand are predominantly in *good* condition (node 5). The influence of diameter on pipe condition has been previously identified in the literature. Micevski et al. (2002) found that smaller diameter stormwater pipes in Australia are aging faster than bigger diameter pipes as designers often underestimate the traffic loads and cover requirements for small pipes. Davies et al. (2001) note that large diameter pipes are often put in place by more experienced personnel, reducing the likelihood for damage to occur during the installation process. While older, small diameter pipes have a 55% chance of being in *poor* condition if their lengths are greater than 35 m (node 4), shorter pipes have only a 26% chance of being in *poor* condition (node 3). This is likely caused by an increased number of joints in longer pipes leading to more possible points of failure than in shorter pipes.

Pipes constructed prior to 1968 with diameter > 450 mm and steeper slopes are more likely to be in *poor* condition than those with flatter slopes (nodes 6 and 7).
Pipe sections with steeper slopes tend to have the highest deterioration rates as flow rates are generally faster in steep sections. This creates a greater possibility for erosion and surface wear of the wall inside of the pipe (Baik et al., 2006). The influence of length on pipe condition is once again apparent in nodes 8 and 9, where 55% of those pipes with a length > 63 m are in poor condition, compared to only 29% of those with a shorter length.

The classification tree does not provide an indication of the influence of material on condition, as splitting the instances by that attribute did not improve predictive performances. Of the 651 pipes in the training set constructed prior to 1968 (node 1), 627 are made of concrete and division according to pipe material is therefore unnecessary. Although there is a more uniform distribution of material for pipes constructed after 1968 (168 concrete, 224 extra strength concrete and 231 reinforced concrete), any tree branches incorporating material on the right side of the tree were pruned away by the C5.0 system as they did not improve predictive performance. For purposes of knowledge discovery, the modeling software can be used to drill down further into any relationship between material and condition for these newer pipes in the training set. As illustrated in Figure 7.4, very few of the inspected reinforced concrete pipes were found to be in poor structural condition.
7.5 Discussion

7.5.1 Planning Future Inspections

The developed classification tree can be used to predict the condition of the 1,596 concrete, 422 extra strength concrete and 1,764 reinforced concrete pipes that have not been inspected. The tree predicts the large majority of these pipes are in good condition as 3,360 pipes were constructed after 1968. Of these newer pipes, 2,733 are less than 25 years old and are likely of limited threat to the overall integrity of the stormwater network. Resources for future inspection should instead be dedicated to the 265 uninspected pipes the tree classifies (predicts) as being in poor structural condition.
A total of 859 uninspected pipes have an unspecified material type listed in the construction records. These can still be presented to the classification tree to gain a general idea of their potential condition. The tree classifies 247 out of the 859 unspecified material pipes as being in poor condition. These 247 pipes can be added to the 265 pipes previously designated for future condition inspection.

A total of 941 uninspected pipes are made from a variety of materials other than concrete and cannot be presented to the classification tree. Of these pipes:

- 466 are made of PVC and are < 25 years old,
- 161 pipes are made of corrugated steel, of which 93 have diameter >1 m, and
- 311 pipes are made from materials such as asbestos cement and corrugated metal pipe, of which 243 are < 25 years old.

It is up to the discretion of utility managers to add any of these 941 pipes to the list of pipes that are targeted for immediate inspection.

A total of 2,458 out of 8,073 stormwater pipes were inspected from 2008 to 2011. Of these inspected pipes, 901 were found to be in poor condition. The classification tree identified an additional 512 pipes that have not yet been inspected but are likely to be in poor condition. The location of the 1,413 (901 inspected + 512 predicted) poor condition pipes can be visualized using ESRI ArcMap (Figure 7.5).
Figure 7.5: Location of poor condition stormwater pipes.
7.5.2 Improving Predictive Performance

The dataset available for model development lacked information on burial depth and this may have negatively impacted the predictive performance of the developed models as the influence of depth pipe condition has been identified in the literature. Lester and Farrar (1979) note that the frequency of (sewer) pipe defects decreases as burial depths increase. Shallow sewers failed faster than deeper pipes in Cullen (1982). Obtaining reliable data on soil conditions around each pipe would likely improve the predictive capabilities of the classification tree. Soil type determines the rate of infiltration around the pipe and the concrete corrosion potential. The dataset available for model development was imbalanced, with more good condition pipes than poor. Future inspection of pipes predicted by the classification tree to be in poor condition would restore balance to the dataset available for model development and may improve predictive performance.

7.6 Conclusion

CCTV inspections are commonly used to determine the condition of aging stormwater pipes. This work is expensive and most municipalities are limiting their inspections to small portions of their stormwater network. Data mining with classification trees enhanced the value of existing inspection records by gaining insight into the stormwater pipe deterioration process.

A classification tree was developed using 107 km inspection data obtained in Guelph, Ontario (representing one third of the stormwater network). The classi-
fication tree clearly illustrates the influence of construction year, diameter, length and slope on pipe condition. Year of construction was found to strongly influence stormwater pipe structural condition, with 82% of those pipes placed after 1968 in good structural condition and 45% of those placed prior to 1968 in poor structural condition. Pipe diameter, slope and length were found to strongly influence the resulting condition of the older pipes in the stormwater system.

The tree reliably predicts pipe condition (with 301 out of 425 correctly classified instances in a stratified test set) and can therefore be used to identify the location of structurally unsound pipes that have not yet been inspected. The tree predicts 512 pipes that have not yet been inspected are likely in poor structural condition. These pipes should be targeted for the next round of inspection as there is a basis to assess their threat to the integrity of the stormwater network.

Data mining provides an opportunity to go beyond simply inspecting pipes and assigning condition grades. Classification trees are a logical choice for extracting information from inspection datasets as they are straightforward in their implementation and are easy to understand. Any municipality with a pipe inspection dataset can adopt the presented methodology to develop their own powerful classification tree models. The combination of inspection and data mining can guide future inspections and will be valuable when dealing escalating levels of stormwater pipe deterioration.
Transition to Chapter 8

This next chapter explores the potential to use data mining to enhance the way municipalities manage their water distribution systems. Analysis of historical records maintained by the City of Scarborough, Ontario indicates CML and CP significantly reduce the number of water main failures occurring in the water distribution system. A novel approach to extracting instances of time to failure from a pipe failure dataset is described. These instances are then used to train an artificial neural network capable of predicting the time to failure (in years) for individual water mains. This paper has been peer-reviewed and published in the American Society of Civil Engineers Journal of Water Resources Planning and Management.

Chapter 8

Predicting the Timing of Water Main Failure Using Artificial Neural Networks

8.1 Introduction

Water distribution pipes placed in North America following World War II are rapidly approaching the end of their effective design life. As these critical pieces of buried infrastructure age, they deteriorate and lose resilience to imposed stresses, which leads to their failure. The consequences of water main failure include reduced revenue from water lost during transmission, interrupted access to safe drinking water, compromised water quality after the ingress of contaminants into the pipeline, and possible damage to surrounding infrastructure (i.e. overlying roadways, buried electrical utilities, and nearby building foundations). Frequent
occurrences of water main failure place tremendous financial strain on cities such as Toronto, Ontario, where there are 1,300 water main failures each year (Schuster and McBean, 2008). Toronto Water’s recommended capital plan for 2012 to 2020 acknowledges their responsibility to carry out water main rehabilitation work in a timely manner to ensure clean and accessible drinking water is provided for generations to come. With a planned investment of $1.13 billion in water main rehabilitation over the next 10 years, they aim to reduce the number of annual water main failures, restore revenue from lost water sales, and eliminate Toronto’s backlog of rehabilitation work by 2020 (Toronto, 2012).

Utility managers need access to location-specific information on pipe failure if rehabilitation programs are to be successfully implemented. Given this, substantial efforts have been made to model pipe failure in a manner amenable to the planning of rehabilitation activities. Shamir and Howard (1978) used regression analysis to predict failure rates (number of failures per unit length per year) for homogeneous groups of pipes installed under similar conditions. Walski and Pelliccia (1982) further explored this approach and incorporated data on pipe age, diameter, and the number of previous breaks. Clark et al. (1982) used regression to predict the time to first failure and the number of failures after the first event, although predictive capabilities of the models were not demonstrated. Kettler and Goulter (1985) found failure rate to be inversely proportional to pipe diameter in 43 km of cast-iron pipes studied in Winnipeg, Manitoba. Goulter and Kazemi (1988) used regression to determine the probability of subsequent pipe breaks given that at least one failure had occurred. Malandain et al. (1999) used regression to predict failure rates for pipes in Lyon, France, grouped according to structural and environmental
factors. Boxall and O’Hagan (2007) used regression to examine the influence of diameter, age, length, and soil conditions on annual burst rates of cast-iron and asbestos cement water pipes in the United Kingdom. Regression models developed in Wang et al. (2009) for pipes in a large water distribution system provided insight into the influence of material, age, and length on annual break rate but could not be used to predict the timing of future failures.

A number of studies have developed time-related failure risk relationships using a branch of statistics known as survival analysis. Andreou et al. (1987b) found early stages of pipe failure could be characterized by a semi-parametric proportional hazards model and a Poisson-type model could be used to predict failure later in an asset’s life. A coefficient of determination of 0.46 was achieved when predicting failure in older pipes (Andreou et al., 1987a). A similar approach was described in Eisenbeis (1994) and Bremond (1997) where proportional hazards models were used to investigate the influence of failure history on the expected number of failures in large water distribution networks in France. Gustafson and Clancy (1999) used a semi-Markov process to model failure of thick wall, medium wall and thin wall water mains in Saskatoon, Saskatchewan. While the model was found to be adequate for predicting inter-break times for water mains in a historical dataset, it could not effectively predict the timing of future failure. Mailhot et al. (2000) focused on the development of statistical failure models for municipalities with short recorded histories of pipe breaks. Rostum (2000) developed proportional hazards and non-homogeneous Poisson process models to determine the influence of pipe length, diameter, age, soil conditions, and the number of previous failures on cast-iron and ductile-iron water main failure times in Norway. Pelletier et al.
(2003) used survival analysis to predict the total annual number of failures for three Quebec municipalities. While the coefficient of determination for the model was low (0.39 - attributed to the variation of natural processes involved as pipes grow older), the model was useful for exploring the impact of different pipe replacement scenarios on the annual number of pipe failures. Park et al. (2008) analyzed 92 pipes (each having at least five failures) to develop failure rate models with coefficient of determination values of 0.48 to 0.59. Kleiner and Rajani (2010) used a non-homogeneous Poisson process to predict the expected number of breaks for groups of cast-iron pipe using 40 years of historical break data from western Canada. While predictions of the total number of breaks over a 5 year period were accurate, predictions for individual pipes were not as valid. Survival analysis presented in Christodoulou (2011) provided insight into the change in hazard rate over time for pipes grouped by pipe diameter and type of failure (i.e. corrosion or tree root penetration).

Data mining represents an alternative approach to water main failure modeling that draws on the fields of artificial intelligence, machine learning, and statistics so that valuable information can be extracted in a manner that circumvents assumptions of time-related failure behavior required by statistical approaches. Sacluti et al. (1999) used artificial neural networks (ANNs) to predict the number of failures in a distribution system based on a 7-day weather forecast. While the model was incapable of predicting failure for individual pipes in the network, its utility lay in the short-term planning of cold-weather maintenance. Babovic et al. (2002) modeled break potential using self-organizing feature maps and Bayesian networks trained with 3,175 repair jobs carried out from 1928 to 1995 on water pipes in
Copenhagen, Denmark. Ahn et al. (2005) used neural networks to examine the influence of water and soil temperature on water pipe failure using a failure record collected over one year. Neural network analysis was used in Christodoulou et al. (2007) to examine the influence of various attributes on pipe life-cycle behavior in Limassol, Cyprus. The number of previous failures was found to have the greatest influence on failure, followed by pipe material, diameter, and traffic loading. Al-Barqawi and Zayed (2008) predicted water main condition ratings (0 - 10) for individual water mains in three Canadian municipalities using an ANN with physical, environmental and operational inputs. Fares and Zayed (2010) developed a hierarchical fuzzy expert system to determine water main condition ratings (0 - 10) using a knowledge base of failure established through a literature survey of municipal experts and a database of 544 water main failures collected in Moncton, New Brunswick. Ho et al. (2010), Tabesh et al. (2009), and Jafar et al. (2010) used neural networks to predict the total number of water main failures in a distribution system. Asnaashari et al. (2011) examined changing water main failure rates in Etobicoke, Ontario, and found neural networks to be better suited to the prediction of failure rate than multiple regression methods.

At this point, data mining approaches have focused on either the prediction of failure rate, condition score, or the total number of failures. These predictions are useful when planning rehabilitation activities (i.e., water mains with the highest rate of failure should be a priority for replacement), but do not provide information on when failures are likely to occur. In response, a novel approach to the prediction of the time to failure for individual water mains using artificial neural networks is described herein. Background information on the case study area of
Scarborough, Ontario, is first presented - with a focus on the positive impact of cathodic protection and cement mortar lining on the annual number of water main failures. Unique instances extracted from historical records are used to train neural networks capable of predicting the time to failure in years, for individual water mains, using attributes such as diameter, length, year of construction, and the number of previous failures. Models developed for asbestos cement, cast-iron, and ductile-iron water mains have correlation coefficients of 0.70 to 0.82 and have utility on the planning of inspection, maintenance, and rehabilitation activities on an asset-by-asset basis.

8.2 Case Study Scenario

The 5.5 million residents of the Greater Toronto Area (GTA) rely on an extensive network of piping infrastructure for their drinking water. The average pipe in the 5,850 km distribution network is 55 years old - with 17% of the water mains older than their effective design life of 80 years and 6.5% more than 100 years old (Toronto, 2012). The historical database of water main failures for the district of Scarborough used in this research consists of year of construction, pipe material, length, diameter, soil type, year of cement mortar lining (if done), year of cathodic protection (if done), and pipe failure dates recorded from 1962 to 2005 (up to the tenth break of an individual pipe). The dataset contains 6,346 water mains installed from 1905 to 2000 with a system length of 1,021 km. These water mains are constructed from asbestos cement, cast iron, ductile iron, or PVC with lengths ranging from 0.28 to 1,634.31 m and diameters ranging from 32 to 1,500 mm. A
total of 3,500 water mains have never failed, while 2,846 have failed at least once. A brief summary of the failure dataset is presented in Table 8.1.

Table 8.1: Summary of the Scarborough water main failure dataset.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Asbestos Cement</th>
<th>Cast Iron</th>
<th>Ductile Iron</th>
<th>PVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of pipes</td>
<td>463</td>
<td>3793</td>
<td>1747</td>
<td>343</td>
</tr>
<tr>
<td>Pipe length (m)</td>
<td>1-831</td>
<td>0.3-978</td>
<td>0.3-1634</td>
<td>0.6-731</td>
</tr>
<tr>
<td>Avg. pipe length (m)</td>
<td>194</td>
<td>153</td>
<td>171</td>
<td>149</td>
</tr>
<tr>
<td>Total pipe length (km)</td>
<td>90</td>
<td>582</td>
<td>299</td>
<td>50</td>
</tr>
<tr>
<td>Diameter (mm)</td>
<td>150-400</td>
<td>50-1500</td>
<td>100-600</td>
<td>150-300</td>
</tr>
<tr>
<td>Number in sand/gravel</td>
<td>173</td>
<td>589</td>
<td>209</td>
<td>85</td>
</tr>
<tr>
<td>Number in silt/clay</td>
<td>290</td>
<td>3204</td>
<td>1538</td>
<td>258</td>
</tr>
<tr>
<td>Number of pipes with CP</td>
<td>9</td>
<td>514</td>
<td>785</td>
<td>10</td>
</tr>
<tr>
<td>Number of pipes with CML</td>
<td>0</td>
<td>1801</td>
<td>137</td>
<td>0</td>
</tr>
<tr>
<td>Number with zero failures</td>
<td>337</td>
<td>1530</td>
<td>1300</td>
<td>333</td>
</tr>
<tr>
<td>Number with at least one failure</td>
<td>126</td>
<td>2263</td>
<td>447</td>
<td>10</td>
</tr>
<tr>
<td>Number of failures in total</td>
<td>248</td>
<td>8204</td>
<td>1454</td>
<td>12</td>
</tr>
</tbody>
</table>

8.3 A First Look at Pipe Failure in Scarborough

Figure 8.1 depicts the significant reduction in annual pipe failures following the implementation of cathodic protection (CP) in 1986 and cement mortar lining (CML) in 1988. The CML process removes rust build-up on the inside of the water main and then lines the internal pipe surface with a thin layer of cement. The CP technique involves the attachment of zinc or magnesium sacrificial anodes to the
water main. While typically used for metallic pipe, CP can also be used to protect metallic fittings and connections on nonmetallic pipes such as asbestos cement and PVC. The trend line in Figure 8.1 illustrates the decrease in the number of yearly failures that began in the mid-1980s, which is directly correlated to Scarborough’s vigorous implementation of pipe protection. The 117 failures in 2005 are directly comparable to the 84 - 127 annual failures that occurred over the 1962 - 1965 period, even though the length of the distribution system had doubled from 558 km in 1962 to 1,021 km in 2005.

CP has been shown to be effective for extending the lifespan of aging pipes and reducing pipe failures in a number of other studies. Kleiner et al. (2003) illustrated a decrease in the total annual number of failures following the implementation of CP for 1,750 km of cast-iron and ductile-iron pipes in Ottawa, Ontario. A 17% reduction in break frequency was observed after the implementation of CP in Winnipeg, Manitoba (Rajani and Kleiner, 2003). Case studies in Des Moines, Iowa, indicate CP extends the lifespan of existing water mains by 20 years at less than 10% of the replacement cost of the piping (Schramuk and Klopper, 2005). A study of water main failure in the GTA (Schuster and McBean, 2008) indicates significant cost savings can be achieved through CP - where repair costs for 150-mm diameter ductile iron pipes with CP over 10 years represented one-third of the non-CP repair costs (CAD $100,000 versus CAD $300,000).

Figure 8.2 illustrates the reduction in the number of days per year with a recorded failure that occurred after the implementation of CML and CP. Most water mains in the early 1960s were relatively new and there were only 100 days per year with a single failure and 50 days with multiple failures. As pipes began to deteriorate
Figure 8.1: Impact of CML and CP on the total number of water main failures each year in Scarborough.
Figure 8.2: Impact of CML and CP on the total number of days each year with a water main failure.
in the 1980s, the number of days in a year with at least one break rose to more than 200. At that point in time there were approximately 100 days each year with multiple failures occurring at different locations in the distribution system and mobilization of repair equipment and construction crews must have been challenging and expensive. Days with multiple failures increase the likelihood that members of the general public will be without safe water as they wait for repairs in another part of the City to be completed before construction crews reach their neighborhood. CML and CP implementation brought the annual failure days back to the more manageable levels of 50 years ago.

![Box-whisker plot of monthly failures during 1962 - 2005](image)

**Figure 8.3:** Seasonal variation in the total number of water main failures.

Scarborough water mains fail at variable rates within the year. Figure 8.3 displays a box-whisker plot of monthly failures during 1962 - 2005 and shows how failures have historically been highest during the colder winter months (November - 25, December - 36, January - 43, and February - 32). Approximately 60% of all pipe failures occur during the colder winter months and this seasonal variation in water
main failure frequency has been previously documented in Lackington and Large (1980), Newport (1981), Kettler and Goulter (1985), Chambers (1994), Habibian (1994), and Saegrov et al. (1999). Blair et al. (2011) note that pipe bursts in Scotland are highest during November to March when minimum air temperatures remain below zero for long periods of time.

8.4 Artificial Neural Network Model Development

Artificial neural networks (ANNs) mimic the information processing capabilities of the human brain and have revolutionized the way complex, real-world problems can be modeled in engineering, science, and finance. ANNs are capable of approximating any nonlinear relationship between inputs and outputs in a way that is robust against noise in the training data (Mitchell, 1997). Although a variety of architectures have been developed over the years, the archetypal network is the feed-forward multilayer perceptron (MLP) - consisting of an input layer, an output layer and one hidden layer of neurons (Hinton, 1992). While each neuron on its own is only capable of performing simple computations, the hierarchical organization of interconnected neurons with variable weights allows the network to learn to make predictions based on historical instances (El-Din and Smith, 2002).

In general terms, neural networks are trained to perform a task by learning from a set of training instances or examples. Each instance presented to a network takes the form \( \{ \vec{x}, t \} \) where \( \vec{x} \) is a vector of attribute values and \( t \) is a target/observed value. The instances of water main failure used to train the neural networks were extracted from Scarborough’s historical dataset of failure. The average time to first
failure for pipes constructed post-1962 was found to be 16 years, so instances of pipe failure were extracted using pipes installed from 1946 to 2005. Although some individual pipes constructed from 1946 to 1961 may have experienced a failure prior to the first logged failures in 1962 (as variations in break rate may occur due to such factors as variations in construction practice), expansion of the modeling dataset in this manner reduces the likelihood of missing failures in older pipes.

Unique instances of failure were extracted using the recorded dates for each individual failure. To illustrate the process of developing the modeling dataset, consider a cast iron pipe installed in 1955 with a diameter of 300 mm, length of 80 m, buried in sand/gravel, cement mortar lining installed in 1991, cathodic protection installed in 1990, and two failures recorded in 1985 and 2002. The first instance of failure would have the following attributes: construction year = 1955, diameter = 300, length = 80, soil = 0 (sand/gravel), CML = 0 (failure occurred prior to CML installation in 1991), CP = 0 (failure occurred before CP installed in 1990), number of previous failures = 0 and time to failure = (date of break 1 - installation date) = (1985 - 1955) = 30 years. The attributes for the second instance of failure would be similar, with the exception of: CML = 1 (date of the second break came after CML installation), CP = 1 (date of second break came after CP installation), number of previous failures = 1 and time to failure = (date of break 2 - date of break 1) = (2002 - 1985) = 17 years. Inclusion of the number of previous failures is a necessity as it separates instances of failure extracted from a pipe’s failure history that would otherwise have identical attributes but different observed time to failure. Extraction of individual failure instances from each of the 2,699 pipes that have failed at least once created 9,508 unique instances of pipe failure.
Neural networks for predicting the time to failure in years for asbestos cement, cast iron and ductile iron water mains were trained using the online back-propagation with momentum algorithm available in IBM SPSS Statistics 20. The algorithm searched through the \( n \)-dimensional Euclidean space of \( n \) network weights for the best hypothesis (i.e. the weights that give the best prediction of the observed value). PVC water mains failed 12 times over the 1982-2005 period and this small sample size prevented the development of a model for PVC water main failure. The pipe-specific inputs for each ANN model are described in Table 8.2. The attributes and observed time to failure for each instance presented to the neural network were standardized by subtracting the mean and dividing by the standard deviation to prevent differences in magnitude among the variables from influencing network weights. The instance datasets available for each water main material were randomly partitioned into training, validation and testing sets using a 70-15-15 split ratio. Division into these datasets ensured convergence on a globally optimal solution through use of the early-stopping cross-validation technique (Mitchell, 1997). The training set was used to train the model, the validation set was used to track errors during training and prevent over-training, and the testing set was used to evaluate the predictive capabilities of the ANN. A double-loop technique was used to search for the optimal network architecture, where the outer loop ran through a range of 1-20 hidden neurons configurations and the inner loop ran through a series of 20 random weight initializations to ensure evaluation of the full range of network performance. The best-performing/optimal network architecture had the lowest sum-of-squares error (SSE) between the predicted and observed time to failure. All networks were trained using a hyperbolic tangent transfer function for each hidden neuron and the identity function for the output layer.
Table 8.2: Attributes and targets used for neural network model development.

<table>
<thead>
<tr>
<th>Material</th>
<th>Attribute</th>
<th>Type</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asbestos</td>
<td>Diameter</td>
<td>Numeric</td>
<td>150-400 mm</td>
</tr>
<tr>
<td>cement</td>
<td>Length</td>
<td>Numeric</td>
<td>18.79-828.55 m</td>
</tr>
<tr>
<td>248</td>
<td>Construction year</td>
<td>Numeric</td>
<td>1952-1984</td>
</tr>
<tr>
<td>instances</td>
<td>Soil Type</td>
<td>Numeric</td>
<td>0-sand/gravel (109) or 1-silt/clay (139)</td>
</tr>
<tr>
<td>of failure</td>
<td>Previous failures</td>
<td>Numeric</td>
<td>0-9 failures</td>
</tr>
<tr>
<td></td>
<td>Time to failure</td>
<td>Numeric</td>
<td>0-48 years</td>
</tr>
<tr>
<td>Cast</td>
<td>Diameter</td>
<td>Numeric</td>
<td>50-600 mm</td>
</tr>
<tr>
<td>iron</td>
<td>Length</td>
<td>Numeric</td>
<td>1.82-970.29 m</td>
</tr>
<tr>
<td>7795</td>
<td>Construction year</td>
<td>Numeric</td>
<td>1946-1985</td>
</tr>
<tr>
<td>instances</td>
<td>Soil type</td>
<td>Numeric</td>
<td>0-sand/gravel (598) or 1-silt/clay (7197)</td>
</tr>
<tr>
<td>of failure</td>
<td>CML protection</td>
<td>Nominal</td>
<td>0-no CML (7414) or 1-CML (381)</td>
</tr>
<tr>
<td></td>
<td>Cathodic Protection</td>
<td>Nominal</td>
<td>0-no CP (7744) or 1-CP (51)</td>
</tr>
<tr>
<td></td>
<td>Previous failures</td>
<td>Numeric</td>
<td>0-9 failures</td>
</tr>
<tr>
<td></td>
<td>Time to failure</td>
<td>Numeric</td>
<td>0-56 years</td>
</tr>
<tr>
<td>Ductile</td>
<td>Diameter</td>
<td>Numeric</td>
<td>100-400 mm</td>
</tr>
<tr>
<td>iron</td>
<td>Length</td>
<td>Numeric</td>
<td>1.98-1260.10 m</td>
</tr>
<tr>
<td>1453</td>
<td>Construction year</td>
<td>Numeric</td>
<td>1957-1985</td>
</tr>
<tr>
<td>instances</td>
<td>Soil type</td>
<td>Numeric</td>
<td>0-sand/gravel (99) or 1-silt/clay (1354)</td>
</tr>
<tr>
<td>of failure</td>
<td>CML protection</td>
<td>Nominal</td>
<td>0-no CML (1446) or 1-CML (7)</td>
</tr>
<tr>
<td></td>
<td>Cathodic Protection</td>
<td>Nominal</td>
<td>0-no CP (1377) or 1-CP (76)</td>
</tr>
<tr>
<td></td>
<td>Previous failures</td>
<td>Numeric</td>
<td>0-9 failures</td>
</tr>
<tr>
<td></td>
<td>Time to failure</td>
<td>Numeric</td>
<td>0-32 years</td>
</tr>
</tbody>
</table>
8.5 Results

Table 8.3 presents the network architectures and performance for the asbestos cement, cast iron and ductile iron water main models. The relative error is the ratio of the sum-of-squares error for the dependent variable (time to failure) to the sum-of-squares error for the null model (the model where the mean value of the time to failure is used as the predicted value for each instance). The relative errors were constant across the three datasets, indicating successful prevention of over-training.

Table 8.3: Predictive performance of ANNs on training, validation and testing datasets.

<table>
<thead>
<tr>
<th>Material</th>
<th>Architecture</th>
<th>Dataset</th>
<th>Instances</th>
<th>SSE</th>
<th>Relative error</th>
<th>R-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asbestos cement</td>
<td>5 - 4 - 1</td>
<td>Training</td>
<td>174</td>
<td>39.60</td>
<td>0.46</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Validation</td>
<td>37</td>
<td>6.293</td>
<td>0.42</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Testing</td>
<td>37</td>
<td>–</td>
<td>0.56</td>
<td>0.70</td>
</tr>
<tr>
<td>Cast iron</td>
<td>7 - 6 - 1</td>
<td>Training</td>
<td>5457</td>
<td>1387.34</td>
<td>0.51</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Validation</td>
<td>1169</td>
<td>309.75</td>
<td>0.53</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Testing</td>
<td>1169</td>
<td>–</td>
<td>0.50</td>
<td>0.70</td>
</tr>
<tr>
<td>Ductile iron</td>
<td>7 - 4 - 1</td>
<td>Training</td>
<td>1017</td>
<td>172.40</td>
<td>0.34</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Validation</td>
<td>218</td>
<td>37.86</td>
<td>0.32</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Testing</td>
<td>218</td>
<td>–</td>
<td>0.34</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The optimal asbestos cement ANN consisted of 5 input nodes (length, diameter, soil, year of construction, and number of previous failures) and 4 hidden neurons. While the training set of data available for developing the asbestos cement model
was the smallest of the three materials, the correlation coefficient for the 37 instances reserved for evaluating the trained ANN was 0.70, indicating the model is capable of providing a realistic estimate of the time to failure for the 463 asbestos cement water mains in the distribution system.

The ANN best-suited for cast iron failure prediction consisted of 7 input nodes (length, diameter, soil, CML, CP, year of construction, and number of previous failures) and 6 hidden neurons. The architecture of the model is provided in Figure 8.4, where thick lines between neurons indicate a strong connection. The correlation coefficient for the 1,169 instances in the holdout set was 0.70. The scatter plot of the observed versus predicted time to failure in Figure 8.5 shows that while the network is not perfect in its predictions, it is capable of providing a reasonable estimate of the observed time to failure based on attribute information available for training. Figure 8.6 illustrates the difference between the observed and predicted times to failure in the holdout set. The average absolute value of the difference between observed and predicted time to failure for instances in the holdout set with 0 previous failures, 1 previous failure and at least 2 previous failures are 8.2, 6.1, and 3.5 years, respectively.

The optimal ductile iron ANN architecture consisted of 7 input nodes (length, diameter, soil, CML, CP, year of construction and number of previous failures) and 4 hidden neurons. The correlation coefficient for the 218 instances in the holdout dataset was 0.81, which is the highest of all three ANNs and indicates the model is capable of providing a reasonable estimate of the time to failure for the 1,747 ductile iron water mains in Scarborough.
Figure 8.4: Optimized cast iron artificial neural network architecture.
Figure 8.5: Scatterplot of the cast iron artificial neural network predictions of the time to failure.
8.6 Discussion

8.6.1 Planning Future Rehabilitation Activities

The predictive performance of the developed neural networks indicates they have “learned to provide a reasonable estimate of the time to failure for an individual pipe given a set of pipe-specific attributes. As such, predictions of the timing of future failure for an individual pipe can be made if that pipe has similar attributes to instances used to train the neural network. Consider an asbestos cement water main installed in 1961, with a length of 52 m, diameter of 150 mm, buried in silt/clay, and one previous break in 2005. Presenting this information to the asbestos cement ANN results in a predicted time to failure of 10 years. Because the first failure occurred in 2005, this second failure is expected to occur in 2015.
The ANN can be used to iteratively predict the timing of failures 1 through 10 for pipes in the distribution system. As an example, Figure 8.7 illustrates the predicted number of future failures for the 1,528 cast iron water mains installed post-1946 that were left without CML or CP as of 2005. Historical records indicate the total number of pipe failures for all pipes in 2005 was 117 (after approximately half the pipes in the distribution system were protected by CML or CP). Simulations obtained from the cast iron ANN indicate the number of annual failures solely for this subset of pipes left unprotected will likely be in the range of 300 per year by 2015 with more than 50 cast iron pipes per year reaching 10 total failures. While predictions of the year of future failure for some pipes in the distribution system pre-date 2006 (a contradiction to the historical dataset of observed failure that indicates the failure in question had not yet happened as of 2006), the ANN has based its prediction on what was learned about the time to a failure from other similar pipes that actually failed. In other words, if a prediction is to be made for a pipe with two recorded failures, the predicted time to the third failure will be based on what was learned from other similar pipes with an observed time between second and third failure. While there is some uncertainty in prediction, each network provides reasonable estimates of failure, so predicted failures pre-dating 2006 should not be dismissed as being inaccurate. Rather, pipes with predicted failures pre-dating 2006 should become a candidate for inspection as historical records indicate other similar pipes have already failed. These simulations and the knowledge learned from earlier discussions about the positive impact of CML and CP on reducing annual failures in Scarborough indicate CML and CP installation should be carried out for all unprotected candidate pipes to prevent escalation of future failure rates.
Figure 8.7: Predicted number of yearly failures for cast-iron pipes installed post-1946 without CML or CP.
While the neural networks are capable of providing a realistic estimate of the time to failure for individual pipes, each prediction is an expected value that should not be used as a stand-alone number for rehabilitation decision making. Rather, the predictions are more suited to supplementing information learned from existing records of failure and other failure-related predictions (i.e. predicted failure rates, condition scores or total number of failures) so that a more complete understanding of individualized pipe failure behavior is made available for inspection and rehabilitation decision making. While predictions obtained through statistics or data mining approaches cannot be used to identify where along the length of the pipe failure will occur, each candidate for rehabilitation can be inspected using state-of-the-art, free-swimming acoustic sensors that are now widely available. Capable of inspecting live pipelines up to 25 km long, this technology can be used to identify the location of existing leaks before they become catastrophic failures (PureTechnology, 2008). Inspection results from these in-line surveys can be used to determine if rehabilitation (i.e. cured-in-place pipe installation) or replacement actions need to be carried out for those water mains predicted to fail.

8.6.2 The Influence of Pipe-specific Attributes on ANN Prediction

The modeling software was used to carry out an independent variable importance analysis to evaluate the influence of each attribute on the output of the trained models (Table 8.4). The number of previous failures had the greatest influence on prediction for all three models (consistent with Goulter and Kazemi (1988), Sundahl (1997), and Asnaashari et al. (2011) where failure history strongly influ-
enced failure behavior over time). Predictions provided by the asbestos cement model were largely influenced by the number of previous failures, the year of construction and pipe diameter. While variations in CML and CP protection strongly influenced the predicted time to failure in the cast iron and ductile iron models, the introduction of soil type as an attribute within the model was shown to be largely irrelevant. The available division into either sand/gravel or silt/clay appears to be insufficient and specific information on the resistivity and corrosivity of the soils should be collected before the influence of soil on failure behavior can be fully evaluated.

Table 8.4: Independent variable importance analysis.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Asbestos Cement ANN</th>
<th>Cast Iron ANN</th>
<th>Ductile Iron ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Import. Normalized.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil</td>
<td>0.051 11.7</td>
<td>0.025 7.9</td>
<td>0.024 7.2</td>
</tr>
<tr>
<td>Length</td>
<td>0.051 11.8</td>
<td>0.071 22.5</td>
<td>0.049 14.8</td>
</tr>
<tr>
<td>Diameter</td>
<td>0.149 34.4</td>
<td>0.132 41.8</td>
<td>0.074 22.2</td>
</tr>
<tr>
<td>Const. year</td>
<td>0.316 73.0</td>
<td>0.136 43.1</td>
<td>0.103 30.8</td>
</tr>
<tr>
<td>CML</td>
<td>– –</td>
<td>0.186 58.7</td>
<td>0.282 84.6</td>
</tr>
<tr>
<td>CP</td>
<td>– –</td>
<td>0.133 42.1</td>
<td>0.135 40.4</td>
</tr>
<tr>
<td>Prev. failures</td>
<td>0.433 100.0</td>
<td>0.316 100.0</td>
<td>0.333 100.0</td>
</tr>
</tbody>
</table>

The developed neural networks rely on a set of basic attributes when predicting the time to failure for an individual water main and it is likely predictive performance could be improved if additional data were made available for model development. Future work in predicting failure should aim to include information on pipe location within the regional area if it is available as water infrastructure is now
commonly spatially integrated with maps of other municipal infrastructure assets (i.e. roadways and sewer lines). The sensitivity of modeling accuracy to variations in available failure history (i.e. failure records dating back 10 years vs. 20 years) is an additional area of research that merits investigation. Others looking to carry-out time to failure modeling should aim to include pipe wall thickness and water pressure information data if available as both are known to be important factors controlling leakage and break potential (Skipworth, 2002).

As indicated in this research, pipe failures in Scarborough are highly seasonal and failure in winter is likely attributed to the generation of circumferential pressure that develops as temperatures drop below 0 degrees C causing the soil around the pipe to freeze and expand. The penetration of the frost layer deep into the soil during the coldest months and the formation of ice lenses may cause additional frost heave pressure. Stress caused by the binding action of ice at the boundary between pipes and freezing soil may induce additional longitudinal strain in the pipe (Kim et al., 2005). Pipe failures during the early-stages of the spring thaw (March) are likely the result of reduced load support caused by a temporary saturation of the soil surrounding the pipe that occurs as the frost layer melts. Specific burial depth information for each pipe in Scarborough was not available for analysis at the time of this research and an evaluation of the influence of burial depth on failure is merited if depth information is made available in the future.

Incorporation of temperature data (winter severity) may also improve predictive accuracy of the neural networks. The influence of temperature-based covariates on predicted time to failure may prove to be pipe material specific as temperature has previously been found to have more of an impact on the total number of
pipe failures for subsets of cast iron water mains than ductile iron water mains in Scarborough (Rajani et al., 2012). Temperature covariates were less related to the total number of ductile iron pipe failures as this material fails more from the impact of perforations in the pipe wall induced by corrosion than fracture induced by temperature change.

8.7 Conclusions

An analysis of historical water main failures in Scarborough, Ontario, Canada illustrates the merits of cement mortar lining and cathodic protection installation practices for controlling deterioration in aging water distribution pipes. The installation of CML and CP reduced the number of annual water main failures from more than 300 in the mid-1980s (before installation) to less than 150 in 2005, which is approximately equal to the number of annual failures in the mid-1960s when the distribution system was half of its current size. CML and CP reduced the number of days in a year with multiple failures occurring at different locations in the distribution system from approximately 100 in the mid-1980s to 30 in 2005.

Water main failures in Scarborough were found to be strongly influenced by seasonal changes, with approximately 60% of all failures occurring from November to February. Winter pipe failures are likely influenced by the frost layer as it penetrates deep into the soil during colder months and an investigation into the influence of temperature (winter severity) on time to failure appears to be warranted.

Neural network models developed in this research can be used to predict the time to failure for individual watermains using commonly recorded pipe attributes. Net-
works were trained using the online back-propagation with momentum algorithm and a double-loop technique that determined optimum ANN architectures from the wide variety of possible network configurations. Unique models were developed for asbestos cement, cast iron and ductile iron water mains. Performance of the networks for instances reserved for evaluation indicate acceptable prediction capabilities (correlation coefficient of 0.70 for 37 asbestos cement instances, 0.70 for 1,169 cast iron instances, and 0.81 for 218 ductile iron instances). The predicted time to failure can be used to determine the year of failure and this can in turn be used to help guide inspection-related decisions for individual pipes in the distribution system. Simulations of future failure for pipes left unprotected by CML and CP indicate a return to the elevated failure rates of the mid-1980s if pipe protection practices are not extended to include all candidate pipes in the distribution system. The installation of CML and CP would extend the lifespan of aging water mains in the regional area and continued implementation would help maintain distribution system reliability. Adequate investment into this aging infrastructure would ensure reliable access to safe-drinking water, create jobs, recapture revenue lost from leaking water, and enhance North America’s economic competitiveness for years to come.
Chapter 9

Conclusions and Recommendations

The deterioration of aging water and wastewater pipes poses a significant environmental and economic threat. Aging sanitary sewer pipes inevitably deteriorate to a point where raw, untreated wastewater will leak out of the pipe and into the surrounding soil and nearby sources of groundwater. Structurally damaged stormwater pipes located in close proximity to broken sanitary sewer pipes provide leaking sewage direct access to surface water such as lakes and rivers. Water main breaks waste enormous quantities of drinking water, allow the ingress of contaminants into the distribution system, and cause flooding and extensive damage.

In this era of infrastructure replacement, a major change in the manner in which municipalities manage their water and wastewater assets needs to occur. The \textit{out-of-sight} and \textit{out-of-mind} approach to infrastructure management needs to be abandoned if municipalities are to have any hope of remaining economically viable.
as infrastructure inherited from previous generations continues to break down. This thesis examined the benefits of introducing data mining techniques to the field of infrastructure management with an intention to improve predictions of condition at the individual pipe level. This section summarizes the findings of the research and provides a list of recommendations for future work.

9.1 The Role of GIS and Inspection Technology

9.1.1 Inspection Technology

The majority of structural investigations of sewer pipe condition are carried out using closed-circuit television (CCTV) - a technique that provides a visual record of the internal condition of a pipe. Typically a scoring methodology (e.g. the WRc MSCC and WRc SRM) is used to standardized the results of the inspection and establish a condition grade to each inspected pipe. CCTV inspection technology is not without fault as it is expensive, time-consuming, and incapable of detecting defects below the water line. A suite of alternative inspection tools is increasingly being made available to municipalities for pipe inspection purposes (i.e. laser profiling, sonar, and electrical methods for leak detection).

9.1.2 The Role of GIS

Geographical information systems (GIS) play an increasingly important role in both the management and visualization of pipe inspection data.
9.1.3 Future Work Related to Pipe Inspection

1. Further investigation into the suitability of acoustic monitoring technology for pressurized sewer leak detection.

2. Comprehensive field study to quantify exfiltration rates in Canadian municipalities. Although sewage exfiltration into the environment is a recognized threat, data on the current rate of leakage is extremely limited. A variety of potential methods for detecting leaks exist - as identified in Chapter 2.

3. Develop methods of conveying condition information and rehabilitation-related work to the general public. Public perception of how municipalities are managing infrastructure using tax dollars is of critical importance. GIS may serve as an effective consensus-building tool - e.g. intelligent maps and reports created in GIS might foster better communication and cooperation among various stakeholders of a construction or improvement project and play a crucial role in maintaining transparency and accountability. As accounting practices change and environmental regulations become more widespread within the industry, cities will become increasingly reliant upon GIS to adequately account for and report their capital asset inventory in a complete, accurate, and detailed manner. When the time for remedial work is due people, although sensing the necessity for keeping the infrastructure functioning, are not altogether happy about the impact this work has on daily life. The implementation of any plans for rehabilitation are dependent on the municipalities ability to dissuade key stakeholders in the project (the general public, local government, etc.) through use of intelligent maps. This is the
information-age, and our tech-savvy society expects the municipality to be an excellent communicator that can keep all stakeholders informed.

9.2 Predicting Sanitary Sewer Pipe Condition

9.2.1 A Summary of Findings

The first investigations into the utility of data mining for managing linear water and wastewater infrastructure (presented as Chapter 3) focused on the novel application of decision trees to predict the condition of individual sanitary sewer pipes. Prior to this research, decision trees had never been used to predict pipe condition using data that would only be available prior to inspection. Furthermore, very few published accounts of the validity of predictions made at the individual-pipe level were available. Decision trees were demonstrated as a means of extracting information from limited inspection records, allowing sewer pipe condition to be predicted for pipes that have not yet been inspected. The challenge posed by class imbalance common in inspection datasets (a challenge that had posed problems when many other authors had attempted to predict individual pipe condition with any level of success) was formally recognized. Transformation to a binary classification problem was proposed (where pipes would be considered as being in either good or poor condition based on the condition grades). The model was capable of providing insight into a pipe condition dataset obtained after inspecting a portion of the sanitary sewers in Guelph, Ontario, Canada. Unlike the majority of existing models of pipe condition, the model was developed with minimal data
pre-processing effort and illustrated the influence of pipe-specific attributes (e.g. year of construction, diameter and length) on pipe condition in a format that can be easily shared with those unfamiliar with the data mining process. The predictive capability of the classification tree was validated using a stratified test set representative of the distribution of pipe condition existing in the sewer system.

CCTV inspection datasets are often imbalanced - with significantly more pipes in one condition class than another. This is problematic as data mining algorithms tend to be most effective when observations available for model development are balanced across classes. The optimally tuned classification tree predicted binary pipe condition (good vs. poor condition) with an overall accuracy of 76% (282 out of 364 instances of pipe condition correctly predicted in the stratified test set). The model achieved an acceptable test set area under the receiver operating characteristic (ROC) curve of 0.77 and can effectively identify individual pipes for future rounds of inspection. The data mining approach was capable of unlocking information contained within inspection records and enhances existing management practices used in the wastewater industry. The novel approach of screening pipes for inspection based on a predicted likelihood of being in poor condition and a close proximity to structurally poor condition stormwater pipes was also proposed. An indication was provided that utility of the model for inspection planning purposes might improve by attempting to classify pipes as being in either good or bad condition (addressed in Chapter 4) or adoption of other data mining algorithms (addressed in Chapter 5).

Further investigations into the predictive capabilities of data mining systems for individual sanitary sewer pipes were carried out in Chapter 4. The utility of decision
trees and support vector machines were compared using the Guelph, Ontario case study. An analysis of the detailed CCTV inspection records indicated the majority of all structural defects occur within 30 m of the nearest manhole. This finding has major implications for future inspection-related work, as it indicates the potential to incorporate zoom-camera technology to expedite future investigations into pipe condition carried out by the municipality. The modeling algorithms were implemented using open-source software and were tuned to counteract the negative impact on predictive performance resulting from class imbalance common within pipe inspection datasets (note - the class-imbalance problem was now more severe after switching to a good vs. bad classification task). The decision tree algorithm was found to be better than SVMs when modeling individual sewer pipe condition for a variety of reasons, including:

- The decision tree outperformed support vector machines for the classification task - achieving an acceptable area under the ROC curve of 0.77 and an overall accuracy of 76% on a stratified test set.

- The decision tree was easier to implement as the algorithm implicitly conducts feature selection (the predictive performance of SVMs hinges on presenting an appropriate subset of features to the algorithm and implementation of a complex feature selection algorithm is required prior to SVM development).

- There are less data preparation steps involved in the construction of decision trees as no attribute transformations or manipulation are required (data presented to SVMs needs to be transformed to a standardized format before model development to overcome scale differences).
• The decision tree algorithm is more intuitive and easier to explain than the complex mathematical manipulations involved in the construction of SVM classifiers.

• The decision tree model would be of greater benefit to municipalities seeking to learn more about why sewer pipes are breaking down as it reveals important relationships between pipe attributes and structural condition that remain hidden within the inspection dataset whenever SVMs are constructed.

The first split in the decision tree indicates structural condition is influenced by time - with 24% of pipes more than 50 years old in bad condition compared to only 5% of those less than 50 years old. The additional splits of the tree indicate municipalities cannot rely on a purely age-based approach to infrastructure management as a tendency towards bad structural condition is also the product of many other factors, including:

• Shallow burial depths: older pipes buried at depths less than 1.9 m have a 42% chance of being in bad condition compared to a 21% chance when they were installed at a depth greater than 1.9 m.

• The failure of aging water infrastructure: older sewer pipes placed at shallow burial depths have an 80% chance of being in bad condition if there has been at least one nearby water main failure. There is only a 33% chance of a similar sewer pipe being in bad condition when there has not been a water main failure in close proximity.

• Longer pipe lengths: older pipes with deep burial depths have a 24% of being in bad condition if they are longer than 33 m and only a 5% chance if they
are shorter than 33 m.

- Smaller pipe diameters: older pipes, with deep burial depths and longer lengths have a 26% chance of being in bad condition if they have a diameter less than 238 mm and only a 12% chance if they have larger diameters.

Chapter 5 presented the novel application of the random forests algorithm for predicting individual sewer pipe condition in Guelph, Ontario. An example of the inevitably poor predictive performance achieved when attempting to classify pipes into one of five condition grades using the available asset attributes was provided - indicating the necessity of adopting class imbalanced learning strategies when attempting to predict individual pipe condition. The algorithm was found to be better suited to tackling the class imbalance problem than decision trees - achieving a stratified test set false negative rate of 18%, false positive rate of 27% and an excellent area under the ROC curve of 0.81 in the Guelph, Ontario case study. The novel inclusion of condition information of pipes attached at either the upstream or downstream manholes of an individual pipe enhanced the predictive power for the bad condition pipes representing the minority class of interest (reducing the false negative rate to 11%, reducing the false positive rate to 25% and increasing the area under the ROC curve to 0.85).

Chapter 6 compared class-imbalance learning strategies (under-sampling during random forests construction and threshold adjustment) to evaluate any potential increase to predictive accuracy for the minority class of interest in the case study. An analysis of the inspection data indicated diminished structural integrity was most often the result of cracks, fractures and defective joints - representing 42%, 38% and 11% of all defects, respectively. Threshold adjustment was found to pro-
vide an optimal level of performance for the classification problem with the trained predictive random forests model achieving a false negative rate of 18%, false positive rate of 29% and an excellent area under the receiver operating characteristic (ROC) curve of 0.81 (considered to be excellent given the nature of the inspection dataset). An analysis of the predictive capabilities of the random forests algorithm trained using a dataset one-third the size of the one originally available to the case study municipality indicates the algorithm has utility for municipalities outside of the case study area. Network spatial analytics were introduced to visualize clusters of bad condition pipes in the sewer system. Visualization of pipe condition in this manner was found to be another effective tool for screening candidate pipes for inspection. A sensitivity analysis was carried out that indicates models developed using a significantly smaller dataset than the one available to the case study municipality may not have as high a level of performance, but will still be of significant utility when planning inspections.

The random forests algorithm achieved a lower classification error than both decision trees and support vector machines. It shares many of the previously discussed benefits of decision trees (minimal data preprocessing, built-in feature selection, and conceptually easier to understand than SVMs), making it an ideal candidate for municipalities seeking to achieve the highest level of predictive performance when determining individual pipe condition. Decision tree classifiers are, however, of greater utility for knowledge exploration purposes as the ensemble nature of random forests prevents learning how pipe-specific attributes are related to structural condition. As such, it is advisable that a municipality seeking to carry out similar data mining work should first implement a decision tree algorithm to extract information
from within their inspection dataset and then evaluate any potential gains in predictive performance that might be achieved through subsequent modeling using the random forests algorithm.

Although Guelph, Ontario was used as a case study area, the described data mining methodology was developed with an intention to improve the way all municipalities proactively manage their existing sewer infrastructure. As such, both decision trees and random forests capable of predicting individual pipe condition can have significant utility for other municipalities planning an upcoming round of condition inspection work for a variety of reasons, including:

- Implemented using free, open-source software (beneficial to those municipalities currently experiencing an infrastructure deficit).
- Efficiently implemented - with a minimal amount of data pre-processing.
- Leverages information contained within existing datasets of pipe condition that would remain hidden using many other modeling approaches.
- The decision tree model can be shared and easily interpreted by those unfamiliar with data mining (unlike other existing approaches to predicting pipe condition). The learned relationships between pipe attributes and condition may play a role in future rehabilitation and construction-related work.
- The predictions of condition provided by the models can help a municipality identify candidate pipes for inspection that pose a threat to the environment and public health.
- The novel application of gain chart analysis indicates and spatial analytics (network hotspot density analysis, proximity to defective stormwater infras-
structure, etc.) provides an opportunity to capture more information on bad condition pipes with significantly fewer inspections.

9.2.2 Future Work Related to Sanitary Sewers

1. Further develop the concepts of using risk of failure to screen pipes for condition inspection. Novel concepts have been introduced in this work - the use of gain chart analysis, clustered density of existing deteriorated pipe hotspots, proximity to other deteriorated infrastructure, etc. These have been presented individually, and there is merit in investigating the possibility to combine these aspects into a more complete risk methodology.

2. Carry-out field work in the municipality to investigate success associated with using model predictions to pick pipes for future inspection.

3. Creation of a user-friendly pipe-screening tool incorporated within the GIS environment would collate the results of data mining condition prediction, network spatial analytics and other pipe criticality information.

4. Develop models capable of predicting operational performance grade for individual sanitary sewer pipes.

5. Transfer knowledge to predict pipe condition in other municipalities. Alternatively, develop new location-specific models using inspection data previously collected by any new participating municipality.
9.3 Predicting Stormwater Pipe Condition

9.3.1 A Summary of Findings

Chapter 7 indicated the capability of using decision trees to extract asset condition information related to stormwater pipe integrity. Guelph, Ontario was again used as a case study. The classification tree illustrated the influence of construction year, diameter, length and slope on pipe condition in an easily interpretable format. An overall success rate of 71% (301/425 instances of pipe condition correctly classified in a stratified test set) indicates the utility of the model when predicting the condition of the remaining two thirds of the pipes in the stormwater system that have not yet been inspected.

9.3.2 Future Work Related to Stormwater Pipes

1. Improve predictions of stormwater pipe condition using random forests.

2. Carry-out field investigations to establish severity of sewage contamination of stormwater pipes in Canadian municipalities.

3. Use spatial analytics to identify clusters of bad condition stormwater pipes - areas of a municipality that may pose significant flood-risk.

4. Examine the potential to develop models based on data-mining to predict stormwater pipe condition for other Canadian municipalities outside of the case study area.
9.4 Predicting Water Main Failure

9.4.1 A Summary of Findings

Chapter 8 presented the final area of study - where effective management of aging water distribution infrastructure is shown to be essential for preserving the economic vitality of North American municipalities. This work examined historical water main failures within Scarborough, Ontario, Canada revealing a seasonal pattern to water main failure with the majority of failures occurring during the very cold winter months. Extensive installation of cement mortar lining and cathodic protection have extended the lifespan of aging water mains and reduced escalating failure rates. Artificial neural networks are found capable of predicting the time to failure for individual pipes using a range of pipe-specific attributes including diameter, length, soil type, construction year, and the number of previous failures. The developed models have correlation coefficients ranging from 0.70 - 0.82 on instances reserved for evaluating predictive performance and have utility on an asset-by-asset basis when planning water main inspection, maintenance and rehabilitation. Simulated failure scenarios indicate a return to high failure rates if cement mortar lining and cathodic protection are not extended to all candidate pipes in the distribution system.

9.4.2 Future Work Related to Water Main Failure

1. Combine predictions of time to failure with network spatial analytics to improve management of deteriorating water distribution infrastructure.
2. Use decision tree classifiers to gain an understanding of various pipe-specific attributes on water main failure. The neural networks proposed in Chapter 8 are essentially black-boxes, and decision trees may provide an opportunity to identify the influence and importance of water main length, diameter, age, etc. on the likelihood of failure and time to failure.

3. Evaluate the influence of burial depth on pipe failure (if that information can be made available by any participating municipality).

4. Investigation into the influence of temperature (winter severity) on water main rate of failure.

9.5 Concluding Remarks

It may be somewhat surprising, but there are striking similarities between the fall of one of the world’s first great empires and the situation facing many modern-day municipalities. Just as the Romans built aqueducts to bring water to their cities they also built drains and sewers to take away the waste water. Romans even were afforded the luxury of having a large number of public latrines connected to the sewerage system. About 2000 years ago, economic inflation in Rome reached 13% and lasted for 100 years. During that time the price of corn rose 300,000 times and it became impossible to levy taxes and public works (like sewer maintenance) were neglected. The main outfall of their great sewer was blocked, the Pontine Marshes were flooded, mosquitoes bred throughout the city and there was an outbreak of malaria which left the population of Rome decimated. The barbarians invaded and the dark ages supervened - in other words, a neglected sewer was a major factor in
the fall of the Roman empire (Read, 1997).

Municipalities face two choices: incur the cost of living with failing infrastructure (widespread environmental contamination, severe flooding, disruption of service, increased taxes, an erosion of public faith in government expenditure of tax dollars, etc.) or carefully prioritize and undertake infrastructure renewal investment to ensure the protection of the environment, public health and economic vitality.
Appendices
Appendix A

Permission to include published journal articles in thesis
Title: Predicting the Timing of Water Main Failure Using Artificial Neural Networks
Author: Richard Harvey, Edward A. McBean, Bahram Gharabaghi
Publication: Journal of Water Resources Planning and Management
Publisher: American Society of Civil Engineers
Date: 04/01/2014
Copyright © 2014, ASCE. All rights reserved.

Permissions Request
As an ASCE author, you are permitted to reuse you own content for another ASCE or non-ASCE publication.

Please add the full credit line "With permission from ASCE" to your source citation. Please print this page for your records.

Type of use: Dissertation/Thesis
Portion: full article
Format: print and electronic

Use of this content will make up more than 25% of the new work: no
Author of this ASCE work or ASCE will publish the new work: yes
Dear Richard Harvey,

Permissions request relating to material published in *Journal of Hydroinformatics*:

In response to your request for copyright clearance to reproduce the following material:

doi:10.2166/hydro.2014.007

Into a chapter of a PhD Thesis to be completed at the University of Guelph, Ontario, Canada; we are very happy to grant you permission to reproduce the material specified above without charge, provided that:

- the material to be used has appeared in our publication without credit or acknowledgement to another source;
- the version of the paper reproduced is the author’s proof and NOT the full published version;
- suitable acknowledgement to the source is given in accordance with standard editorial practice, e.g.,

  “Reproduced from Robert Richard Harvey and Edward Arthur McBean (2014) ‘Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure’ *Journal of Hydroinformatics* In Press, Uncorrected Proof, with permission from the copyright holders, IWA Publishing”;

- Reproduction of this material is confined to the purpose for which this permission is given.

I trust this permission will be satisfactory; if any point needs clarification or you have any further queries, please do not hesitate to contact us again.

Yours sincerely

[Signature]

Chloe Parker
Digital Publishing Executive
Appendix B

Useful References for Data Mining Theory

B.1 Overview

The following list of references would prove useful to those seeking to learn more about the data mining algorithms presented in this thesis.


- Chapter 1 - Introduction (containing information on the general goals of data mining with some key terminology)

- Chapter 13 - Nonlinear Classification Methods (with information on support vector machines)
• Chapter 14 - Classification Trees and Rule-based Models (with information on both CART and Random Forests)

• Chapter 16 - Remedies for Severe Class Imbalance (with information on adjusting classification thresholds)


• Chapter 3 - Decision tree learning

• Chapter 4 - Artificial neural networks (describing the backpropagation algorithm and the prevention of over-fitting using the early-stopping technique)


• Chapter 3 - Data Preprocessing (describing the importance of data cleaning)

• Chapter 8 - Basic Concepts of Classification (with information on decision tree induction and techniques to improve classification accuracy)

• Chapter 9 - Classification - Advanced methods (containing information on Support Vector Machines)
Appendix C

Comparing utility and cost of pipe inspection methods

C.1 Overview

There are a wide variety of inspection technologies available to a municipality in this current market, but there is limited information on the suitability of the technology for field work in the urban setting. The EPA recently carried out three weeks of field-testing in Kansas City, Missouri to determine the suitability of a variety of technologies (Table C.1) for assessing the condition of wastewater collection systems (Martel et al., 2011).
Table C.1: Inspection technology evaluated in Martel et al. (2011).

<table>
<thead>
<tr>
<th>Technology</th>
<th>Manufacturer</th>
<th>Pipe Material</th>
<th>Pipe Diameter</th>
<th>Flow Regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCTV</td>
<td>Manufacturer</td>
<td>Any</td>
<td>Not specified</td>
<td>Technology inspects dry pipe segments.</td>
</tr>
<tr>
<td>Digital</td>
<td>CUES - Optical Zoom II</td>
<td>Any</td>
<td>6 to 60 in.</td>
<td>Technology inspects dry pipe segments - line must be tested during periods of low flow.</td>
</tr>
<tr>
<td>Zoom camera</td>
<td>Cleanflow Multi-sensor (HD camera)</td>
<td>Any</td>
<td>&gt;6 in.</td>
<td>Technology inspects dry pipe segments - line must be tested during periods of low flow.</td>
</tr>
<tr>
<td>Electro-scanning</td>
<td>Envirosight QuickView</td>
<td>non-ferrous</td>
<td>3 to 60 in.</td>
<td>Pipe must be surcharged at face. Sliding plug system can be used to achieve surcharge.</td>
</tr>
<tr>
<td>Laser</td>
<td>Cleanflow Multi-sensor (laser unit)</td>
<td>Any</td>
<td>&gt;4 in.</td>
<td>Technology inspects dry pipe segments. Line must be tested during periods of low flow.</td>
</tr>
<tr>
<td>Sonar</td>
<td>Cleanflow Multi-sensor (sonar unit)</td>
<td>Any</td>
<td>&gt;12 in.</td>
<td>Technology inspects pipe below the water surface. A minimum depth is required to submerge the head of the sonar unit.</td>
</tr>
</tbody>
</table>

C.2 Experience with CCTV

Approximately 12,000 feet of pipe was inspected using CCTV. The inspection crews initially tried to inspect without first cleaning the pipes but debris and obstructions
prevented the self-propelled crawler from advancing through the pipe. At that point forward, a hydraulic jet truck equipped with a cutting head (to remove tree roots) was used to clean the pipes prior to CCTV inspection. It is not uncommon for there to be a large amount of debris in a large diameter pipe, so it is advisable to have the pipe cleaned before an inspection crew arrives to minimize inspection downtime. The inspection findings were consistent with historical records maintained by the utility - with the most common structural defects being cracks, fractures, and broken pipe. Debris in the bottom of the pipe made it impossible to effectively document the defects below the water line.

**C.3 Zoom Camera**

The field demonstration results illustrated a limitation of zoom camera inspection in pipes that were not cleaned - where objects in the pipe segment (*i.e.* spider webs, debris, and debris) caused the camera’s autofocus feature to focus on them rather than the pipe wall. The camera’s manual focus was inconsistent in its ability to sharpen the focus any further. A significant disparity between the surface temperature and the temperature at the bottom of the manhole (approximately 100 degrees F) caused a number of equipment issues - including overheating of an electrical connection at the control head. A notable factor limiting the camera’s sight distance was condensation inside the pipe (*headlights in fog effect*). A number of pipes were inspected both with CCTV and the Zoom Camera. Overall, the zoom camera identified 31 defects as compared to the 168 defects identified by CCTV for the same pipe segments. The poor results were attributed to the zoom camera’s
limited sight distance that resulted in partial inspection of each pipe length. This finding further stresses the need to consider zoom cameras solely as a pre-screening tool for follow-up inspection. The technology is useful for visualizing blockages, pipe fractures, and root intrusion but will not provide an accurate idea of the actual pipeline condition.

C.4 Electro-scanning

Electro-scanning does not require pipe cleaning before inspection but does require the pipe to be filled (a sliding plug can be used to fill small portions of the pipe at a time). Continuous pressure monitoring is required to maintain the pressure head below anticipated building invert elevations. At one location, water entered a basement through a floor drain due to the line being surcharged above the basement elevation. A comparison of electro-scanning to CCTV revealed the following:

- Pipes with a larger number of CCTV defects, especially defects associated with leakage (e.g., cracks, fractures, defective joints, faulty taps and root intrusion), generally have a larger number of electro-scanning anomalies.
- Clusters of CCTV defects often coincide with clusters of electro-scanning anomalies.
- Electro-scanning frequently registers more total leakage-related defects than CCTV, due primarily to the detection of more defective joints.
- Joint defects are identified by CCTV by the presence of roots and if a pipe is not in the vicinity of trees or has been cleaned prior to inspection, the ability
of CCTV to identify joint defects may be diminished. At this field site, there were abundant trees close to the pipes.

C.5 Multi-sensor Technology

The multi-sensor unit used in the EPA field-work was a combination of several inspection technologies mounted on a floating assembly (high resolution digital camera, laser scanner and a sonar head). The results from the CCTV and multi-sensor inspections performed on 12 segments of pipeline along the Line Creek Interceptor provided a basis for the comparison of defect detection. Unfortunately, the CCTV inspection did not identify any structural defects so a comparison of the technologies is limited to operational defects. Author’s note - it would have been of greater benefit to have a comparison of structural defects - particularly as sonar is thought to provide additional info on the potential of cracks hidden from the sight of CCTV to contribute to leakage. The high definition video recorded through digital scanning revealed minimal maintenance defects beyond those identified in the CCTV scan but did provide a detailed view of the pipe cross-section that is useful for condition assessment. The image quality of the digital scan video appeared to be superior to the video from the conventional CCTV based on a visual comparison. The analysis of laser and sonar data was not based on a standard defect coding system but relied on engineering judgment (i.e., knowledge of pipe wall construction) to assess the severity of defects and the need for subsequent maintenance. The laser data revealed and quantified corrosion above the water line that the conventional CCTV did not. The sonar scan results provided insight
into the depth and location of debris within the pipe that cannot be identified with CCTV.

Tables C.2 and C.3 illustrate the differences in complexity, cost and ease of operation between the inspection technologies as indicated by the EPA study. While zoom camera inspection is the cheapest option, it has limited sight distance and cannot be used to inspect the entire length of a pipe section. While multi-sensor scanning is quite fast in the field, the processing of the digital scan is labor intensive and requires specialized software. Analysis of the digital scan, sonar and laser results requires highly qualified personnel - as a result, the multi-scanner technology has the highest complexity rating of the group.
Table C.2: Complexity and ease of operation for inspection technologies tested in Martel et al. (2011).

<table>
<thead>
<tr>
<th>Contributing Factor</th>
<th>CCTV</th>
<th>Zoom Camera</th>
<th>Electro-scanning</th>
<th>Multi-sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Requirements</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>National Certification</td>
<td>PACP</td>
<td>PACP</td>
<td>None</td>
<td>PACP for digital scanning</td>
</tr>
<tr>
<td>Equipment Operation</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Pipe Preparation</td>
<td>Cleaning may be required</td>
<td>None required</td>
<td>None required</td>
<td>None required</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>Low to medium</td>
<td>Low to medium</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Overall Complexity Rating</td>
<td>Low to medium</td>
<td>Low to medium</td>
<td>Low to medium</td>
<td>Medium to high</td>
</tr>
</tbody>
</table>

Note - PACP is the Pipeline Assessment and Certification Program (PACP) developed by the National Association of Sewer Service Companies (NASSCO) in America. The PACP coding uses a simplified method of assigning severity values to the various defects and grades each pipeline similar to the internal condition grades used by the WRc SRM.
Table C.3: Cost comparison of inspection technologies tested in Martel et al. (2011).

<table>
<thead>
<tr>
<th>Cost Element</th>
<th>CCTV</th>
<th>Zoom Camera</th>
<th>Electro-scanning</th>
<th>Multi-sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning/Mobilization</td>
<td>-</td>
<td>$2257</td>
<td>$11047</td>
<td>$4000</td>
</tr>
<tr>
<td>Field Work</td>
<td>$34806</td>
<td>$7731</td>
<td>$11817</td>
<td>$13650</td>
</tr>
<tr>
<td>Data analysis and reporting</td>
<td>-</td>
<td>$15368</td>
<td>$6017</td>
<td>$12618</td>
</tr>
<tr>
<td>Total</td>
<td>$34806</td>
<td>$25356</td>
<td>$28881</td>
<td>$30268</td>
</tr>
</tbody>
</table>

Cost per foot inspected

| Total                        | $2.80 | $0.99       | $2.95            | $4.21        |

Daily cost

| Total                        | $5608 | $1222 - 6415 | $5776            | $15134       |

Cost as % of total inspection costs

| Planning                     | -     | 8.9         | 38.3             | 13.2         |
| Field Work                   | 100   | 30.5        | 40.9             | 45.1         |
| Data analysis and field work | -     | 60.6        | 20.8             | 41.7         |
| Total                        | 100   | 100         | 100              | 100          |
Appendix D

Sample code for developing predictive models

D.1 Overview

The following provides a general indication of the coding that can be used within the R environment to develop predictive models capable of predicting pipe condition. Comments in the coding are indicated by the # sign.
D.2 Code to Develop a Decision Tree Model

# Load the necessary packages for R.
library(AppliedPredictiveModeling)
library(caret)
library(e1071)
library(foreign)
library(DWD)

# The available modeling dataset is first edited in Microsoft Excel. Rows contain instances (inspected pipes) and columns contain the attributes (e.g. length, age, etc.). Numeric attributes are left as they are. Text descriptions are used for any categorical (factor) attributes. The Excel file is then converted to a CSV file so it can be opened in Weka and then converted to an ARFF file. ARFF files are useful as each attribute can be listed as being either numeric or a factor (categorical). Also, Weka can also be used to process the data (e.g. replace missing values), perform tasks like feature selection, and visualize the data.

# IMPORT THE DATA INTO R
TwoClass <- read.arff(Condition_Data) # the inspection data
str(TwoClass) # use this command to identify and confirm the structure of the file.

# PARTITION THE DATA USING THE CARET PACKAGE
# Use the technique of stratified random sampling (70% training, 10% evaluation and 20% testing). We first need to establish the training set:
set.seed(117) # a random number seed
split1 <- createDataPartition(TwoClass$Condition, p = .7)[[1]]
# where Condition represents the target class of interest (pipe in good or bad condition).
other <- TwoClass[-split1,]
TwoClassTrain <- TwoClass[split1,]
# Now create the evaluation and test sets:
set.seed(29)
split2 <- createDataPartition(other$Condition, p = 1/3)[[1]]
TwoClassEvaluation <- other[split2,]
TwoClassTest <- other[-split2,]

# IDENTIFY THE AVAILABLE PREDICTORS
# Establish the baseline set of predictors: Trunk, Material, Construction Year, Age, etc.

```r
predictors <- TwoClassTrain[c(1,2,3,4,5,6,7,8,9)]
str(predictors)
```

# Determine a predictor set without highly sparse and unbalanced distributions. We're using resampling methods, and a random sample of the training set may result in some predictors with more than one unique value to become a zero-variance predictor. These so-called "near zero-variance predictors" can cause numerical problems for some models. To identify these kind of predictors, two properties can be examined. It is suggested to remove predictors if: 1. The percentage of unique values is less than 20% and 2. The ratio of the most frequent to the second most frequent is greater than 20.

```r
isNZV <- nearZeroVar(predictors)
isNZV
```

# Which predictors should be removed (if any) will be indicated by a vector of integers

# To filter on correlations, we first get the correlation matrix for the predictor set. This only works with the numeric values. Make a new dataset for those. Assuming numeric predictors are contained in columns 1 - 4

```r
predictorsNumeric <- TwoClassTrain[cc(1,2,3,4)]
str(predictorsNumeric)
correlations <- cor(predictorsNumeric)
dim(correlations)
correlations[1:4, 1:4]
```

# To visually examine the correlation structure of the data, the corrplot package contains an excellent function of the same name. The function has many options including one that will reorder the variables in a way that reveals clusters of highly correlated predictors.

```r
library(corrplot)
corrplot(correlations, order = "hclust", tl.cex = .35)
```

# The size and color of the points are associated with the strength and correlation between predictor variables. To filter based on correlations, the findCorrelation function will apply the algorithm described in Section 3.5 of the Kuhn and Johnson (2013) text. For a given threshold of pairwise correlations, the function returns the column numbers denoting the predictors that are recommended for deletion.

```r
highCorr <- findCorrelation(correlations, .75)
length(highCorr)
head(highCorr)
filteredPredictors <- predictorsNumeric[cc, -highCorr]
```
str(filteredPredictors)

# CONSTRUCT A SINGLE TREE USING THE CART ALGORITHM

cvCtrl <- trainControl(method = "repeatedcv", repeats = 3, summaryFunction = two Class-
Summary, classProbs = TRUE)

# Establish the procedure used for cross-validation.

PredictorsTrain1 <- TwoClassTrain[c(1,2,3,4,5,6,7,8,)]

# Identify which predictors should be used as inputs into the model (i.e. use pipe age, length,
material of construction, etc.)

library(rpart)

set.seed(15)

CARTModel_Tuned <- train(x = PredictorsTrain1, y = TwoClassTrain$Condition, method = ":rpart", control = rpart.control(minsplit = 20, maxdepth = 5), tuneLength = 30, metric = ":ROC", trControl = cvCtrl)

CARTModel_Tuned

print(CARTModel_Tuned$finalModel)

# Select the top performing model over the series of cross-validations according to the highest
area under the ROC metric.

cat("\n")

fancyRpartPlot(CARTModel_Tuned$finalModel, main="Decision Tree With the Highest
Cross-Validation")

# Plot the resulting Decision Tree using the rpart.plot package.

# MODEL PREDICTIONS - TRAIN DATASET

# Use a data frame to house the predictions from different models.

trainResults <- data.frame(Condition = TwoClassTrain$Condition)

trainResults$CARTModel_Tuned ROC_1 <- predict(CARTModel_Tuned ROC_1 $final Model,
newdata = TwoClassTrain, type = "prob")[,1]

# Generate the class predictions

trainResultsClass <- data.frame (Condition = TwoClassTrain $Condition)

trainResultsClass$CARTModel_Tuned ROC_1 <- predict(CARTModel_Tuned ROC_1 $finalModel, TwoClassTrain, type = "class")

# Generate a training set confusion matrix

CARTModel_Tuned ROC_1_TrainCM <- confusionMatrix(trainResultsClass$CARTModel Tunned 
ROC_1, TwoClassTrain $Condition)
CARTModel_Tuned_ROC_1_TrainCM

# MODEL PREDICTIONS - EVALUATION DATASET
# Use a data frame to house the predictions from different models.
evaluationResults <- data.frame(Condition = TwoClassEvaluation$Condition)
evaluationResults$CARTModel_Tuned_ROC_1 <- predict(CARTModel_Tuned_ROC_1$finalModel, newdata = TwoClassEvaluation, type = "prob")[,1]
# Generate the class predictions
evaluationResultsClass <- data.frame(Condition = TwoClassEvaluation$Condition)
evaluationResultsClass$CARTModel_Tuned_ROC_1 <- predict(CARTModel_Tuned_ROC_1$finalModel, TwoClassEvaluation, type = "class")
# Generate an evaluation set confusion matrix
CARTModel_Tuned_ROC_1_EvaluationCM <- confusionMatrix(evaluationResultsClass$CARTModel_Tuned_ROC_1, TwoClassEvaluation$Condition)
CARTModel_Tuned_ROC_1_EvaluationCM

# MODEL PREDICTIONS - TEST DATASET
# Use a data frame to house the predictions from different models.
testResults <- data.frame(Condition = TwoClassTest$Condition)
testResults$CARTModel_Tuned_ROC_1 <- predict(CARTModel_Tuned_ROC_1$finalModel, newdata = TwoClassTest, type = "prob")[,1]
# Generate the class predictions
testResultsClass <- data.frame(Condition = TwoClassTest$Condition)
testResultsClass$CARTModel_Tuned_ROC_1 <- predict(CARTModel_Tuned_ROC_1$finalModel, TwoClassTest, type = "class")
# Generate a test set confusion matrix
CARTModel_Tuned_ROC_1_TestCM <- confusionMatrix(testResultsClass$CARTModel_Tuned_ROC_1, TwoClassTest$Condition)
CARTModel_Tuned_ROC_1_TestCM

# GENERATE ROC USING THE DATASETS
library(pROC)
CARTModel_Tuned_ROC_1_TrainROC <- roc(trainResults$Condition, trainResults$CARTModel_Tuned_ROC_1, levels = rev(levels(trainResults$Condition)))
CARTModel_Tuned_ROC_1_TrainROC

CARTModel_Tuned_ROC_1_EvaluationROC <- roc(evaluationResults$Condition, evaluationResults$CARTModel_Tuned_ROC_1, levels = rev(levels(evaluationResults$Condition)))

CARTModel_Tuned_ROC_1_EvaluationROC

CARTModel_Tuned_ROC_1_TestROC <- roc(testResults$Condition, testResults$CARTModel_Tuned_ROC_1, levels = rev(levels(testResults$Condition)))

CARTModel_Tuned_ROC_1_TestROC

# Plot the ROC curve for the evaluation set for analysis.
ROC_curve_Evaluation_Set <- plot(CARTModel_Tuned_ROC_1_EvaluationROC, print.thres = c(.5, CARTModel_Tuned_ROC_1_Thresh), type = "S", print.thres.pattern = "%.3f (Spec = %.2f, Sens = %.2f)", print.thres.cex = .8, legacy.axes = TRUE, ylab = "True Positive Rate (Sensitivity)", xlab = "False Positive Rate (1 - Specificity)"

# INVESTIGATE ALTERNATIVE THRESHOLDS
CARTModel_Tuned_ROC_1_Thresh <- coords(CARTModel_Tuned_ROC_1_EvaluationROC, x = "best", ret="threshold", best.method="closest.topleft")

CARTModel_Tuned_ROC_1_Thresh # the threshold closest to the top left of the evaluation set ROC curve.

CARTModel_Tuned_ROC_1_ThreshY <- coords(CARTModel_Tuned_ROC_1_EvaluationROC, x = "best", ret="threshold", best.method="youden")

CARTModel_Tuned_ROC_1_ThreshY # the threshold maximizing Youden's index.

testResultsClass$CARTModel_Tuned_ROC_1_AltTopLeft <- factor(ifelse(testResults$CARTModel_Tuned_ROC_1 > CARTModel_Tuned_ROC_1_Thresh,"Bad", "Good"),levels = levels(testResults$Condition))

CARTModel_Tuned_ROC_1_AltTopLeftTestCM <- confusionMatrix(testResultsClass$CARTModel_Tuned_ROC_1_AltTopLeft, TwoClassTest$Condition)

CARTModel_Tuned_ROC_1_AltTopLeftTestCM # Test set confusion matrix obtained using the threshold closest to the top left on the evaluations set ROC.

testResultsClass$CARTModel_Tuned_ROC_1_AltYouden <- factor(ifelse(testResults$CARTModel_Tuned_ROC_1 > CARTModel_Tuned_ROC_1_ThreshY,"Bad", "Good"),levels = levels(testResults$Condition))

CARTModel_Tuned_ROC_1_AltYoudenTestCM <- confusionMatrix(testResultsClass$CARTModel_Tuned_ROC_1_AltYouden, TwoClassTest$Condition)

CARTModel_Tuned_ROC_1_AltYoudenTestCM # Test set confusion matrix obtained using the Youden's threshold.

# DETERMINE THE LIFT PLOT
labs <- c(CARTModel_Tuned_ROC.1 = "CART Model Tuned ROC")
liftcurve <- lift(Condition CARTModel_Tuned_ROC.1, data = testResults, labels = labs)

library(lattice) # Add lattice options to produce a legend on top
# For info on xyplot see http://127.0.0.1:10769/library/lattice/html/xyplot.html

xyplot(liftcurve, ylab = "% of Bad Pipes Found", xlab = "% of Pipes Inspected", type = "l", lwd = c(3,7), lty = c(1,1), margins=c(3,3), scales = list(tck = -1, lwd = 2), par.settings = simpleTheme(col=c("grey50", "grey10")), auto.key = list(columns = 1,corner = c(0.95, 0.15),lines = TRUE, points = FALSE, lwd = c(3,5),lty = c(1,1), col=c("grey10", "grey10")))
References


Cardoso, A., Prigioebbe, V., Giulianielli, M., Baer, E., Benedittis, J.D., Coelho, S., 2005. Assessing the impact of infiltration and exfiltration in sewer systems using performance indicators - case studies of the APUSS project.


IBM, 2011. IBM SPSS modeler 14.2 - modeling user guide. IBM Corporation.


Kuhn, M., 2013. The caret package.

Kuo, S., Zhao, L., Mahgoub, H., Suarez, P., 2005. Investigation of ground penetratin}


Toronto, C.o., 2012. Watermains and watermain breaks - water supply fact sheets.


