ANALYSIS OF URBAN PIPE DETERIORATION USING COPULA METHOD

by

Farzana Atique

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering

Fall 2016

ANALYSIS OF URBAN PIPE

DETERIORATION USING COPULA METHOD

by

Farzana Atique

Approved: ____

Harry W. Shenton III, Ph.D.
Chair of the Department of Civil and Environmental Engineering

Approved: ____

Babatunde A. Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: ____

Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Nii O. Attoh-Okine, Ph.D., P.E.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Sue McNeil, Ph.D., P.E.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Javier Garcia-Frias, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Edgar P. Small, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisor Prof. Nii Attoh-Okine for the continuous support of my Ph.D. study and related research as well as for his patience, motivation, and immense knowledge. His guidance helped me throughout the research and writing of this dissertation. He not only has been a great academic advisor but also a great mentor and provided thoughtful advice for life in general.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Sue McNeil, Prof. Javier Garcia-Frias, and Dr. Edgar P. Small, for their insightful comments and encouragement as well as their thoughtful questions, which made me widen my research from various perspectives. I would also like to thank Mary Catherine Opila, who was gracious enough to share water pipe data that helped formulate the models.

To my family, friends, and colleagues, I would not be here without you. Thanks to my parents and in-laws for their encouragement. Special thanks to my three children Rashad, Raheel, and Raisah for their understanding and for letting their mom be a student. Lastly, I sincerely thank my husband, Mir Wahed, for his full support and encouragement throughout the whole process.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Aging water main systems are becoming a growing concern for maintenance. The structural deterioration of water mains is affected by different factors, such as pipe age, pipe material, soil conditions, pipe size, and climate conditions, among others. Since pipes are underground and obtaining data for pipes is difficult and expensive, various statistical modeling methods have been used to analyze the factors contributing to the pipe condition deterioration and to predict the failure of pipes. This research applies the copula method to urban pipe data analysis to generate data that can be used to determine remaining life. Copula modeling is an emerging method of modeling that has been widely used in financial sectors. It has recently been used in hydrology and pavement management sectors, but the method has not been applied to other civil engineering disciplines. This research uses copula modeling to determine dependency between several variables of pipe condition and to compare how it may be a better choice for determining correlation dependency for data that are non-normal and skewed. The copula method is very useful when marginal distributions of water pipe condition variables belong to different families of distributions. Copula modeling is used to generate large volumes of data. The large data sets generated can then be used for evaluating the current pipe condition models and the appropriateness of those models for determining the remaining life of a pipe or pipe condition. In this paper, multivariate vine copula modeling was applied to water pipe data. The Bayesian inference approach was applied for parameter estimation, and GIS was used for analyzing soil properties' effects on pipe condition assessment.

Chapter 1

INTRODUCTION

1.1 Introduction

The underground water infrastructure system in the United States was installed mainly in the periods of the 1800s, 1900-1945, and post-1945, the time periods coinciding with substantial population growth (Folkman 2012). The pipelines of these three eras fail decades due to age, corrosion, and improper installation. It has also been suggested that the life span of the materials used has become shorter with each new investment. Once every four years, American Society of Civil Engineers publishes a comprehensive assessment of the infrastructure system in the form of grades. The grades are assigned according to the following eight criteria: capacity, condition, funding, future need, operation and maintenance, public safety, resilience, and innovation. According to the ASCE report card published in 2013, the drinking water and wastewater infrastructure of the country received D grades (ASCE 2013), which suggests that most of the drinking water infrastructure are reaching the end of their useful life.

Nearly 170,000 public drinking water systems are located across the United States, of which 54,000 are community water systems. Failures in drinking water infrastructure can result in water disruptions, impediments to emergency responses, and damage to other types of infrastructure. Contamination of water due to broken pipes is also a health concern. Broken water pipes can damage roadways and structures and disrupt fire control measures. Unscheduled pipe repair work can disrupt transportation and commerce (ASCE 2013).

In 2012, the American Water Works Association (AWWA) estimated the aggregate replacement value for more than one million miles of pipes was approximately $2.1 trillion if all pipes were to be replaced at once. However, not all pipes need to be replaced immediately, and the most urgent investment could be spread over 25 years at an approximate cost of $1 trillion. In the United States, the Environmental Protection Agency (EPA) estimated that nearly 60% of total system costs are found in the distribution and transmission pipelines of water distribution systems. Management of this aging pipe network system has become an emerging issue in recent years.

Infrastructure asset management is an approach which can help maintain utility at a desired level of service at the lowest life-cycle cost. Asset management practices applied to underground infrastructure can help utility companies understand the timing and cost associated with rehabilitating, repairing, and replacing pipelines. Knowledge gained from these efforts also helps to develop pipe material selection criteria as part of the replacement strategy (Folkman 2012). A key component of the asset management practice is the condition assessment of each individual pipeline in order to identify failure-prone pipes and prioritize their renewal (Rogers and Grigg 2006).

Pipeline systems are critical infrastructure assets. These assets lose value over time as the system ages and deteriorates. This outcome can have a major impact on the overall performance of the system such that the costs of repair and rehabilitation may be very high. Major pipelines in the country have reached or passed their design life. It is very important that efficient and cost-effective maintenance and rehabilitation strategies are employed to prevent potential failure issues in the future. Knowing the condition of the pipelines and predicting the repair and maintenance needed is vital for proper asset management for pipe infrastructure.

Many factors affect the condition of the pipes. The structural deterioration of water mains and their eventual failure are affected by static factors, which include pipe material, size, age, and soil type, and dynamic factors, such as climate and pressure zone change factors. The physical mechanism that leads to pipe breakage is a complex process. There is incomplete knowledge on the pipe breakage process due to the fact that most pipes are buried underground, which makes it difficult to observe pipe behavior. Also, it is not uncommon for major utilities to have incomplete or limited data as the regions and districts that currently make up the utility jurisdiction may have different record keeping practices or may not have kept any records (Kleiner and Rajani 1999). Information on the current condition of the water main along with an understanding of the failure mode can help utility companies better manage their assets in a cost-effective manner. Data required for physical models are costly to obtain. High costs can only be justified for major transmission mains. For smaller diameter pipe networks, other methods such as statistical analysis models are more cost-effective.

1.2 Research Objectives

The main objectives of this research are to apply the copula method to water pipe data, to determine the correlations between factors influencing pipe deterioration and failure, and to develop multivariate models as a tool for asset management. The main objectives of this research will be achieved through the following sub-objectives:

1. To study the pipe asset management process and review the pipe deterioration process to identify the factors affecting pipe deterioration.
2. To study and develop copula models for determining the correlation between different factors that contribute to the pipe condition deterioration.
3. To compare the copula prediction model with widely used statistical models such as the regression prediction model.
4. To apply multivariate copula modeling and vine copula modeling to water pipe data.
5. To develop software codes based on the Bayesian inference process to estimate copula modeling parameters.
6. To use Geographic Information Systems (GIS) to study data on soil properties and find a correlation between soil properties and water pipe breakage.

These objectives signify the different applicability of using copula modeling method to pipe condition assessment. The approach of achieving these objectives are discussed and explained in the research overview sections and approach section. Each chapters of this research also narrates how the objectives are attained.

1.3 Research Overview

There are three modeling procedures used to predict pipe failures (Rostum 2000): descriptive analysis, physical/mechanical models, and statistical models. Descriptive analysis models summarize pipe failure data, which can be used to determine trends in pipe failures. These types of models can provide information on the breakage rate for groups of pipes but not information about individual pipes.

Physical models use the physical mechanism that leads to pipe failure to predict future pipe failure. The data required for physical models are hard to obtain. The high cost of obtaining data makes it appropriate for large lines but very costly for smaller pipe networks (Opila and Attoh-Okine 2011).

The third type, the statistical models, has been widely used for pipe failure modeling. Kleiner and Rajani (2001) have summarized and provided a list of the statistical models that have been proposed for water mains condition assessment. The statistical methods for predicting water main breaks use historical data on past failures to predict a future pipe breakage pattern. The statistical models are classified into two groups: deterministic and probabilistic models (Opila and Attoh-Okine 2011). The deterministic models predict breakage rates using two or three parameters. These models use the grouping criteria as covariates in the analysis while retaining a simple mathematical framework. Probabilistic single-variate models use probabilistic processes on grouped data to derive probabilities of pipe life expectancy, probability of breakage, and probabilistic analysis of break clustering phenomenon. The probabilistic multivariate models can consider most of the covariates in the analysis. Table 1.1 shows some of the statistical failure models.

Table 1.1 Sample statistical failure models (adapted from Opila and Attoh-Okine 2011)

Deterministic models	Probabilistic single variate models	Probabilistic multivariate models
Counting process Time-exponential models Time linear models Generalized linear models	Cohort-survival models Semi-Markov processes Break clustering Bayesian diagnostic	Accelerated lifetime models Time-dependent Poisson models Neural networks Proportional hazard model (PHM) -Bathtub Effect PHM -Poisson PHM -Weibull PHM

Copula modeling is an emerging method of statistical modeling. Even though the use of copulas has become dominant in actuarial science, the financial sector, and biology, there has been limited use of copula modeling in the civil engineering research field. There have been a few research papers on copula modeling in hydrology (Genest and Favre 2007), infrastructure dependence modeling (Attoh-Okine 2013), and modeling of vehicle axle weight (Srinivas et al. 2012).

In this research, I introduce copula dependence modeling to pipe condition data. The study of multivariate distributions has been dominated by the normal distribution (Frees and Valdes 1998). The choice of the multivariate normal distribution is appealing as the marginal distributions are also normal and the association between two random variables can be fully described by knowing a) their marginal distribution and b) the correlation coefficient. However, in engineering applications, non-normality can occur in the following ways: a) the marginal distribution of some of the variables may not be normal and b) in some cases, even though all the marginal distributions are normal, jointly these variables may not be multivariate normal (Yan 2006). Copulas are useful for generating joint distributions by combining given marginal distributions according to a specified form of a copula function. Copulas are appealing because they capture dependence more broadly than the standard multivariate normal framework.

Linear correlation is used as a measure of dependence between variables following a multivariate normal or elliptical distribution. The dependence measure most frequently used is the Pearson's correlation coefficient, which is related to linear dependence and the normal distribution (Accioly and Chiyoshi 2004). However, individual infrastructure conditions may have fat tails, skewness, and non-normal characteristics (Attoh-Okine 2013). This makes inference of the dependency based on correlation inaccurate for non-Gaussian data.

The copula approach uses Kendall's rank correlation and Spearman's rank correlation, which measure dependence across the entire distribution unlike Pearson's mean correlation, and are able to model tail dependencies in distributions. The advantages of copula functions have been summarized as follows (Srinivas et al. 2006):

1) This process is useful when multivariate simulation is to be carried out from dependent random variables from different classes of marginal distributions.
2) The copula approach provides a way to separate marginals from the dependence structure.

3) The limitation of using a linear correlation coefficient is avoided.
4) The invariance property of the copula dependence structure under transformations and its independence from the marginal distributions efficiently lead to simulations.

In summary, this research applies copula modeling to water pipe data. This research develops and applies copula modeling to water pipe data and by multivariate sumilation shows how the model is a better method for future predictions in comparison to standard regression statistical analysis.

1.4 **Approach**

The initial phase of the research consisted of an extensive literature review to determine the current state of the research area and gather water pipe data. One set of data was obtained from a utility company; it contains a pipe inventory which includes information on pipe age, size, material, length, etc. The agency also provided histories of pipe breakage. Initial analysis of the research focused on understanding the data and applicability of the copula method for analysis,

A literature review provided the state of the current research and potential areas where applying the copula modeling method is possible for pipeline data. Some pipe deterioration variables were found to be non-normal. Copula dependence modeling is appropriate for variables of non-normal marginals and where the dependence is non-linear and skewed. Bivariate copula modeling was used to generate large set of data points. Data points generated from the copula modeling was then used for regression analysis, and compared to regression analysis of the original data set. The literature review illustrated the scope of the applicability of multivariate vine copula modeling for multivariate condition modeling of pipe data. The research was broadened to include Bayesian analysis to estimate the parameters of copula modeling. The Bayesian software WinBUGS was used for analysis of the parameters of copula modeling, and a code was developed for the analysis.

To obtain various dependency factors for pipe leakage failure, research was conducted to include soil data. ArcGIS was used to obtain soil information and extract useful soil leakage factors to study the dependency modeling through copula modeling. Finally, areas were suggested where applicability of copula modeling for pipe data can further be evaluated. Figure 1.1 shows the main approach and outline of the dissertation.

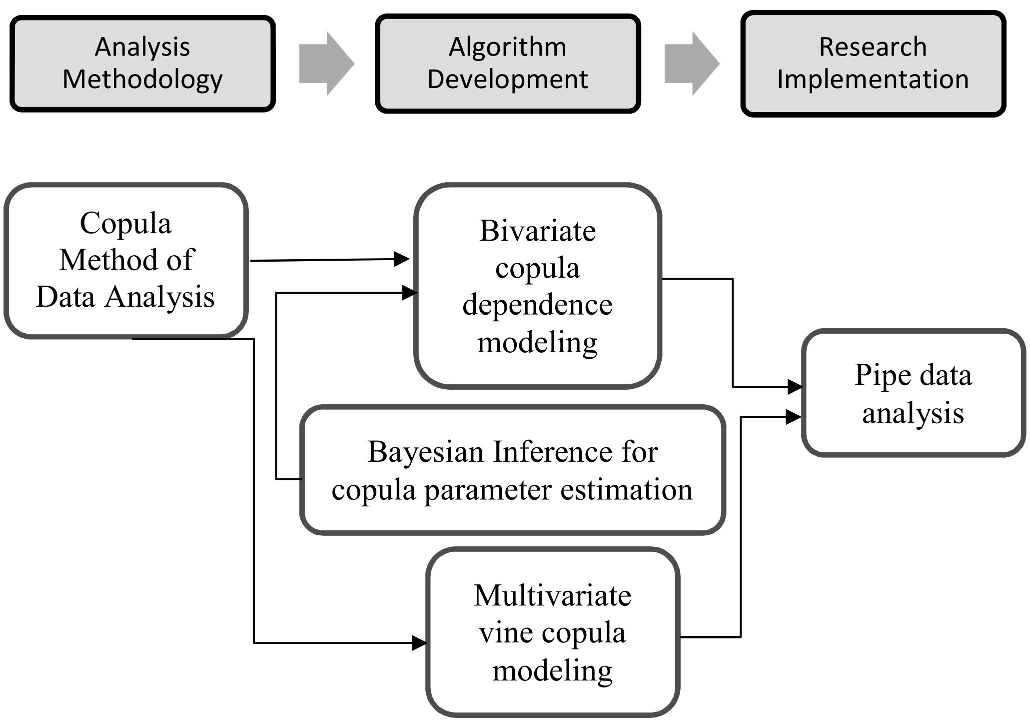


Figure 1.1   Dissertation outline and approach

1.5 Description of Chapters

The research results are presented in this dissertation, divided into the following chapters:

- Chapter 1: Summary and Overview—This chapter includes discussion about the importance of pipe condition monitoring and the need for maintenance work on water pipe infrastructure. It also gives the research overview, approach for the research, and applicability of applying the copula method to pipe data analysis.
- Chapter 2: Asset Management for Pipe Network—This chapter includes discussion about the asset management of water pipe systems. The pipe leak detection methods are discussed along with the steps that are practiced regarding the asset management programs.
- Chapter 3: Background Information on Pipe Failure—This chapter provides a literature review investigating pipe failure factors and pipe failure statistical models. The chapter presents an overview of the existing understanding of pipe failure and various pipe failure modeling methods.
- Chapter 4: Copula Modeling—The chapter presents an in depth overview of copula modeling. It gives details about the theory of copula modeling, discusses various types of copulas, types of measures of dependence, copula parameter estimation, and fitting of a copula model. The chapter also presents an example of a pipe leakage data set, which helps to show how data points simulated from pipe data can give a better prediction for pipe leakage.
- Chapter 5: Vine Copula Modeling—This chapter introduces different types of multivariate dependence modeling based on copulas. The vine copula method

is discussed in detail, and an application of the method with multivariate pipe data is shown as an example.

- Chapter 6: Bayesian Inference for Copula Parameter Estimation—This chapter gives details about Bayesian analysis and the method of application of Bayesian analysis for parameter estimation of copulas shown with an example of a pipe data set.
- Chapter 7: Hybrid Copula and GIS Analysis—Soil properties influence pipe deterioration. This chapter studies the effects of soil properties that influence the deterioration of pipelines. A step by step procedure is developed to obtain soil data from ArcGIS, and that information is used to study the correlation between soil properties and pipe deterioration using copula modeling.

Chapter 2

ASSET MANAGEMENT FOR PIPE NETWORKS

2.1 Introduction

As discussed in Section 1.1, the 2013 ASCE report card gave drinking water and wastewater infrastructure a grade of D. Asset management is a framework for improving the condition of infrastructure system recognizing the impacts of degraded system. There many adverse effects of water pipeline failure:

- Failures in drinking water infrastructure can result in water disruptions.
- Failures can cause disruption of transportation.
- Contamination of water due to broken pipes can cause health concern.
- Broken water pipes can damage roadways and structures and disrupt fire control measures.
- Water loss due to pipe failure can cause shortage of water supply.
- There is also monetary loss as the water that is lost has been treated and there are costs associated with the treatment of water.

The cost for repairing the exiting pipes is very expensive, by estimation of AWWA, it is valued as $2.1 trillion if all the pipes are to be replaced at once. Investment for drinking water infrastructure will cost $1 trillion over a span of 25 years based on assumption that pipes are replaced at the end of their service life and systems are expanded to meet the population growth (AWWA 2016). In the United States, the Environmental Protection Agency (EPA) estimated that nearly 60% of total system costs are related to the distribution and transmission pipelines of water distribution systems. Management of the aging pipe network system has become an emerging issue in recent years.

Asset management is the practice of managing infrastructural assets to minimize the total cost of owning and operating these assets while maintaining the desired level of service (EPA). It is an approach that can help maintain utility at a desired level of service at the lowest life-cycle cost. Asset management practices applied to underground infrastructure can help utility companies in many ways, such as

- Understanding the timing and cost associated with rehabilitating, repairing, and replacing pipelines.
- Knowledge gained from these efforts also helps to develop pipe material selection criteria as part of the replacement strategy (Folkman 2012).
- Doing condition assessment of each individual pipeline in order to identify failure-prone pipes and prioritize their renewal (Rogers and Grigg 2006).

Asset management is a process that utility companies can use to make sure that planned maintenance can be conducted and capital assets can be repaired, replaced or upgraded and there is enough funding to pay for it.

2.2 **Asset Management of Water Pipes**

Pipeline systems are critical infrastructure assets. With age these assets lose value over time. The cost for rehabilitation and maintenance for these aged pipes can be very expensive. Efficient and cost-effective maintenance and rehabilitation strategies needed to be employed to prevent potential pipe failure issues in the future.

There are five core questions of an asset management framework as suggested by EPA, which are as follows:

-What is the current state of the utility's assets?

-What is the utility's required sustained level of service?

-Which assets are critical to sustained performance?

-What are the utility's best "minimum life-cycle cost" strategies?

-What is the utility's

Knowing the condition of the pipelines and making predictions about the repair and maintenance work is vital for conducting a proper asset management for pipe infrastructure. Condition assessment is a process that helps to establish a record of the state of water pipelines. It is essential for cost-efficient repair and replacement programs. Condition assessment methods for pipes can generally be classified into direct or indirect methods as shown in Table 2.1.

Table 2.1 Condition assessment methods

Direct methods	Visual inspection, including closed circuit television (CCTV). Sampling programs (where sections of pipe or "coupons" are sent to a laboratory to have remaining wall thickness measured and a variety of material tests and analyses performed). Various non-destructive testing methods: acoustic emission, acoustic leak detection, remote field eddy current, magnetic flux leakage, ultrasonic pulse echo, and guided Lamb waves.
Indirect methods	Analysis of pipe failure history. Water audits and leak detection to determine leakage levels. Flow testing Measurement of soil resistivity to determine the risk of corrosion.

Even though there are methods available to detect failures in water pipes, each method has specific strengths and limitations. Careful selection and application of methods are needed to determine leak locations. Some of the difficulties in locating leaks are discussed below (Smith et al. 2000):

1. Sources of interference

   The acoustic-based leak location method requires the ability to detect and interpret the meaning of faint sounds; the interpretation is done by a skilled operator or a skilled operator supported by a computer correlator. The factors that alter or mask noise from the leak increase the difficulty of locating the leaks.

2. Pipe location

   Prior knowledge is required to locate leak location. Pipe locations shown on existing records may be of limited use, especially if the pipes belong to an old network.

3. Plastic pipe

   Leaks are difficult to detect in plastic pipes as the viscoelastic nature of plastic tends to dampen vibrations, so noise caused by leaks does not propagate as far as in metal pipe.

4. Multiple leaks

   Correlators are programmed to analyze and locate single leaks or breaks. Multiple leaks on the same line can be treated in different ways. If the leaks are close, the correlator treats them as a single leak. If the leaks are far apart, the two sounding points can be selected as individual leaks. However, when the points cannot be located to identify each leak individually and the leaks are not close together, correlators may not be effective.

It is thus imperative that careful consideration be given when selecting the method to detect pipe leaks and pipe failures. The following section describes the programs available for pipe asset management.

2.3 **Asset Management Program**

Water utilities require the development of a systematic, structured approach that allows for maintenance and renewal of assets at a manageable pace while at the same time maintaining an adequate level of performance. Many asset management programs and approaches are available to water utilities as they develop their own programs. Some are public while others are proprietary to consulting -firms, software providers, or consortia Water Research Foundation (WRF).

These frameworks include:

- International Infrastructure Management Manual (IPWEA 2011)
- Implementing Asset Management: A Practical Guide (AMWA et al. 2007)
- Sustainable Infrastructure Management Program Learning Environment (WERF 2008)

Water utilities should consider the following steps when developing or implementing an asset management program (Water Research Foundation, WRF, 2016):

1. Commence asset management activities by developing a plan (Cromwell and Speranza 2006, AMWA et al. 2007).
2. Establish an interdepartmental asset management team, including senior management.

3. Establish levels of service and key performance indicators for the water utility (Cromwell and Speranza 2006, AMWA et al. 2007, Damodaran et al. 2005, Thacher et al. 2011).
4. Create an inventory of assets throughout the utility (AMWA et al. 2007, Matichich et al. 2005).
5. Design a business risk assessment program, considering assets to be managed and how they might fail (Barnes et al. 2008, Gaewski and Blaha 2007).
6. Begin using data to establish the remaining life of water utility assets (Deb et al. 2002; Rajani, Kleiner, and Krys 2011; Thomson and Wang 2009).
7. Record all breaks and failures, including leaks (Friedman et al. 2010, Deb et al. 2002).
8. Consider condition assessment activities to gauge current condition of assets (Marlow et al. 2007, Thomson and Wang 2009).
9. Plan renewal activities based on the best possible evaluation of the water utility (Cromwell, Nestel, and Albani 2003, Grigg 2004, Deb et al. 2002).
10. Continuously improve asset management activities.

Pipe condition assessment is an important part of the pipe asset management program. Making decisions regarding the rehabilitation and repair of pipes depend on the proper assessment of the condition of the pipes and prediction of the failure of the pipes. Pipe data analysis using copula method is thus a critical part of overall pipe asset management process.

Chapter 3

BACKGROUND INFORMATION ON PIPE FAILURE

3.1 Introduction

Pipeline systems are critical infrastructure assets. There is mounting evidence that the integrity of the U.S. drinking infrastructure is at risk and needs a concerted effort to improve the management of key assets namely pipelines, treatment plants, and other facilities and a significant effort needed to maintain, rehabilitate, and replace these assets (Water Research Foundation, 2016). Pipe failure is a complex process involving the work of many factors. It is important to have an understanding of the different factors that cause the pipe leakage as well as the common mechanism of failure. Background of pipe information helps to formulate pipe deterioration models and helps in assessing pipe condition assessment.

3.2 Definitions

A water pipe is a pipe, frequently made of plastic or metal, that carries pressurized and treated fresh water to a building, as well as inside the building.

As a general definition, a pipe is considered to be in a failed condition if it does not provide an acceptable level of service. If a pipe fails, it requires such actions as rehabilitation, replacement, maintenance, or suspension of service. Failure can be from a small leak to a total pipe collapse in service. Water pipe failures can be classified into three types based on the amount of water leaked (Hamilton and Charalambous 2013):

- Weep/Small loss—a small amount of water that is less than 5 liters per minute leaving a pipe from a small failure
- Leak/Medium loss—an amount of water that leaves a pipe through orifice at an estimated flow of 90 liters per minute
- Burst/Large loss—an amount of water that leaves a pipe from an orifice at an estimated flow of 315 liters per minute
- Catastrophic failure—a complete rupture of the pipeline

Thus, the failures in the pipe can be from a very small leakage to a total failure. In this research the failure considered, consist of small, medium and large loss.

3.3 **Characteristics of Water Main Pipes**

The following sections describe some of the characteristics of water main pipes such as age, material, and size.

3.3.1 **Age of Water Pipes**

In Figure 3.1 (Morrison et al. 2013) shows the historical and projected age of water pipes in the US. As shown, the average pipe age in 2000 was about 38 years, but by the year 2050 the average age will be more than 50 years. The reason behind this is the boom in water pipeline installations after World War II. The rate of pipe installation from 1870 to 1945 was less than 20,000 miles of pipe per decade, but after 1945 the rate increased to 80,000 miles of pipe per decade.

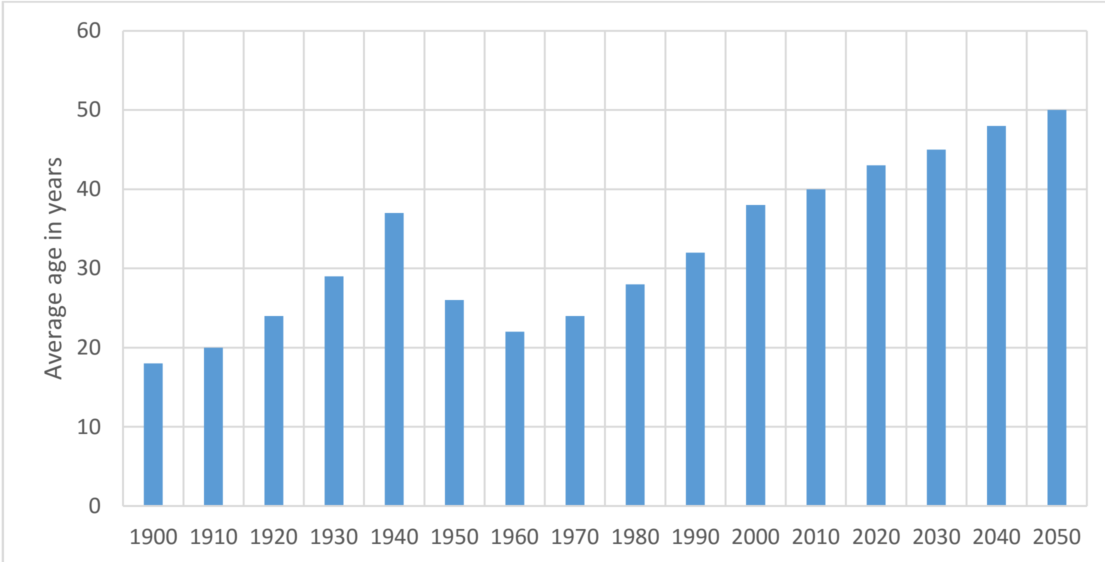


Figure 3.1 Historical and projected age of water pipes in the U.S.

Without proper rehabilitation, the overall condition of pipelines is deteriorating due to aging. Figure 3.2 and Figure 3.3 show the estimated pipe condition in 1980 and 2020, respectively. While in the 1980s 69% of pipes were in excellent condition, it is predicted that in 2020 only 33% will be in excellent condition. Very poor pipe condition will jump from 2% to 23%.

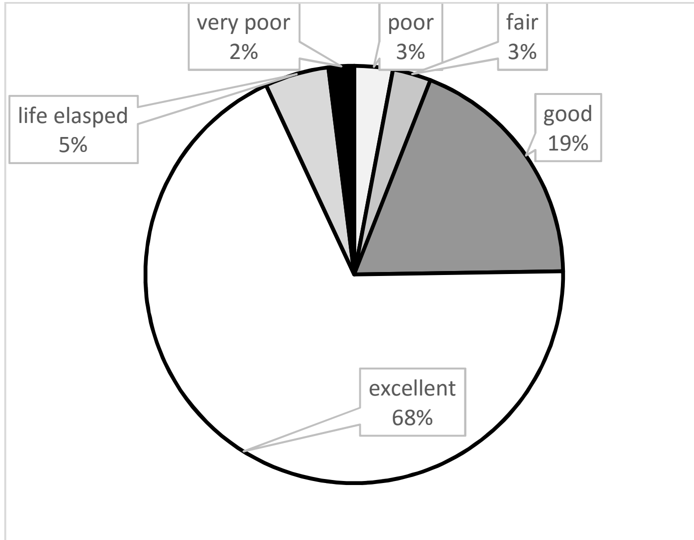


Figure 3.2 Pipe condition in 1980 (Morrison et al. 2013)

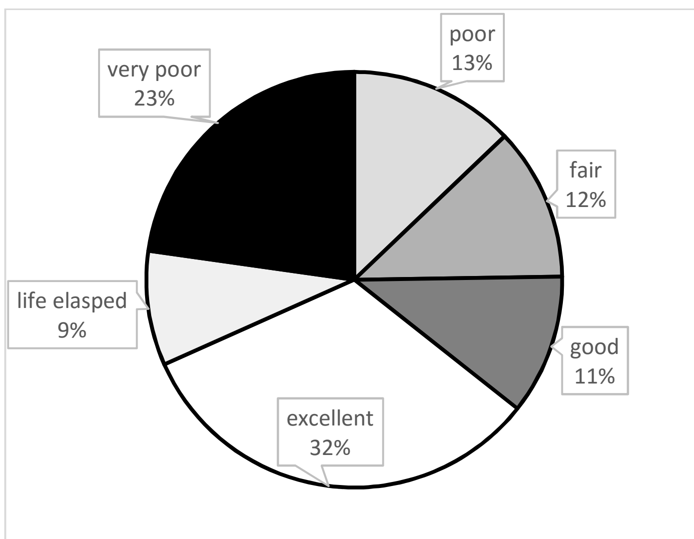


Figure 3.3 Projected pipe condition in 2020 (Morrison et al. 2013)

3.3.2 **Pipe Materials**

The materials currently used in water distribution systems are steel, ductile iron (DI), polyvinyl chloride (PVC), prestressed concrete cylinder pipe (PCCP), glass reinforced plastic (GRP), etc. There are also some materials that are no longer used for new installations but are present in the system, such as asbestos cement (AC) and cast iron (CI). Water mains can be divided into two categories: distribution piping (2 to 10 in.) or transmission mains (12 in. and greater) (EPA, 2009). Almost 73% of all water mains are distribution pipes based on length. Tables 3.1, 3.2, and 3.3 show the water distribution systems by material, diameter, and age, respectively.

Table 3.1 Water distribution systems by material (Morrison et al. 2013)

Material	Miles installed	% of total
CI (unlined, cement mortar-lined, and other)	341,715	39.6
DI (unlined, cement mortar-lined, and other)	189,115	21.9
AC	136,196	15.8
PVC	114,152	13.2
Steel	34,047	3.9
PCCP	23,584	2.7
PE (polyethylene)	3,349	0.4
GRP	665	0.1
Other/Not known	20,169	2.3
Total	863,000	100

Table 3.2 Water distribution system by diameter (Morrison et al. 2013)

Diameter range (in)	Miles installed	% of total
<6	107,200	12.4
6-10	523,200	60.6
12-16	138,600	16.1
18-24	29,700	3.4
30-48	5,700	6.7
>48	6,600	0.8
Total	863,000	100

Table 3.3 Water distribution system by age (Morrison et al. 2013)

Age (years)	Miles installed	% of total
0-10	245,000	28.4
10-25	325,500	37.6
25-50	156,500	18.1
>50	137,000	15.9
Total	863,000	100

3.4 **Pipe Failure Factors and Failure Modes**

Pipe condition is the cumulative effect of many factors acting on the pipe. Al Barqawi and Zayed (2006) classified these factors into three categories: physical, environmental, and operational, as shown in Table 3.4 and each factor is described in Table 3.5. This classification can be expressed as

Pipe deterioration=f (physical, environmental, operational).

The factors in the first two classes can be divided into static and dynamic. Static factors include pipe material, pipe geometry, and soil type, while dynamic factors include pipe age, climate, and seismic activity. Operational factors are inherently dynamic.

Table 3.4 Factors affecting pipe deterioration (Al Barqawi and Zayed 2006)

Physical factors	Environmental factors	Operational factors
Pipe age and material	Pipe bedding	Internal water pressure
Pipe wall thickness	Trench backfill	Leakage
Pipe vintage	Soil type	Water quality
Pipe diameter	Groundwater	Flow velocity
Type of joints	Climate	Back flow potential
Thrust restraint	Disturbances	Operation, maintenance
Pipe lining and coating	Stray electrical current	practices
Dissimilar metals	Seismic activity	
Pipe installation		
Pipe manufacture		

Table 3.5 Contributing factors to water system deterioration (Al-Barqawi and Zayed 2006)

Factor	Comments
Physical	
Pipe material	Pipes made from different materials fail differently.
Pipe wall thickness	Corrosion will penetrate thinner walled pipe more easily.
Pipe age	Effects of pipe degradation become more evident as time progresses.
Pipe vintage	Pipes made at a particular time and place may be more vulnerable to failure.
Pipe diameter	Smaller pipes are more susceptible to beam failure.
Type of joints	Some types of joints are susceptible to premature failure.
Thrust restraint	Inadequate restraint can increase longitudinal stresses
Pipe lining and coating	Lined and coated pipes are less susceptible to corrosion.
Dissimilar metals	Dissimilar metals are prone to galvanic corrosion.
Pipe installation	Poor installation procedures can damage pipes, making them vulnerable to failure.
Environmental	
Pipe bedding	Improper bedding may result in premature pipe failure.
Trench backfill	Some backfill materials are corrosive or frost susceptible.
Soil type	Soils can be corrosive, and some soils experience significant volume changes in response to moisture changes, resulting in pipe loading. Also, the presence of hydrocarbons and solvents in soil may result in pipe deterioration.
Groundwater	Certain groundwater is aggressive towards certain pipe material

Table 3.5 continued

Factors	Comments
Climate	Climate influences frost penetration and soil moisture and thereby enhances the chance of failure.
Pipe location	Migration of road salt into soil can increase the rate of corrosion.
Disturbances	Underground disturbances in the immediate vicinity of an existing pipe can lead to actual damage or load to loading structure on the pipe.
Stray electrical current	Stray currents cause electrolytic corrosion.
Seismic activity	Seismic activity can increase stress on pipe and cause pressure surges.
Operational	
Internal water pressure, transient pressure	Changes to internal water pressure can change stresses acting on the pipe.
Leakage	Leakage erodes pipe bedding and increases soil moisture in the pipe zone.
Water quality	Some water is aggressive, creating corrosion.
Flow velocity	Rate of internal corrosion is greater in unlined, dead-ended mains.
Backflow potential	Cross connections with systems not containing portable water can contaminate the water distribution system.
Operational and maintenance practices	Poor practices can compromise structural integrity of the pipes and water quality.

Many of the factors are difficult to measure or quantify. Also, the quantitative relationships between the factors and pipe failure are not often understood. So the use of two types of indicators has been established, namely distress indicators and inferential indicators for pipe condition assessment. Rajani et al. (2006) described distress indicators as the observable or measurable physical manifestations of the aging and deterioration process. The following tables provide the distress indicators for cast iron (CI) and ductile iron (DI) pipes, asbestos cement (AC), and PVC pipes. Unless otherwise noted, these tables are adapted from Liu et al. (2012).

Table 3.6 Distress indicators that influence pipe condition for cast and ductile iron pipes (Adapted from Liu et al. 2012)

Category	**Distress indicator**	**Comments**
External coating (poly wrap/tar/zinc)	Crack/tear	State of external coating will indicate how external corrosion will likely damage the pipe.
External pipe barrel/bell	Remaining wall thickness	Remaining pipe thickness can be obtained from NDE tests or from spot exhumations.
	Graphitization (pit) areal extent	Areal extent as a percentage of pipe diameter times unit length indicates the size of the affected area.
External pipe barrel/bell	Crack (pit) type	A crack is a mechanical response to stress. Circumferential cracks can indicate some type of longitudinal movement, loss of bedding support, or increase in vertical load.

Table 3.6 continued

External pipe barrel/bell	Crack (pit) width	This is another indicator of corrosion. A wide crack together with a deep pit is more detrimental to the pipe than a narrow but shallow crack.
Inner lining/surface	Cement lining (epoxy) spalling (blistering)	Inner lining deterioration is often due to incompatible water chemistry or abrasion due to the presence of high water velocities and sediments.
	Remaining wall thickness	Sometimes closed circuit television (CCTV) scans give estimates of the internal corrosion pit.
	Tuberculation	Heavy tuberculation can reduce water delivery.
Joint	Change in alignment	Changes in joint alignment indicate that pipe is susceptible to ground movement with large changes resulting in leakage.
	Joint displacement	Joints can displace without undergoing joint misalignment and may indicate other forces at play.

Table 3.7 Distress indicators that influence pipe condition for AC pipes (Adapted from Liu et al. 2012)

Category	Distress indicator	Comment
External coating (tar or bitumen)	Holiday	State of external coating will indicate how external soil properties encourage damage to the pipe.
External pipe barrel	Remaining wall thickness	Remaining pipe wall thickness (includes both external and internal walls) is usually obtained from spot test samples and performing a phenolphthalein test (to measure cement softening) or on-site measurements using the georadar technique.
	Corrosion areal extent	Areal extent as percentage of pipe diameter times pipe segment length indicates the size of the affected area. Severe corrosion may not always mean the pipe should have failed.
	Crack type	Circumferential cracks indicate bending or significant longitudinal movement has taken place. Longitudinal cracks occur due to exceedance of hoop resistance, occurrence of very high operational loads, or low remaining wall thickness as a result of sulfate attack.
	Crack width	Crack width is another indicator of corrosion. Wide cracks together with a deep softening of the asbestos cement matrix will be more detrimental to the pipe than a narrow but shallow crack.
Internal pipe surface	Remaining wall thickness	See above for external pipe barrel category.
	Corrosion areal extent	See above for external pipe barrel category.
Joint	Change in alignment	Changes in joint alignment (rotation) indicate that the pipe is susceptible to ground movement. Large changes can lead to leakage and eventually joint failure.

Table 3.8 Distress indicators for PVC pipes (Adapted from Liu et al. 2012)

Category	Distress indicator	Comment
External pipe barrel surface	Remaining wall thickness	Cavities or unfilled air bubbles introduced during manufacturing (and not detected upon installation) can be of significant size in PVC pipes.
	Scratch type	Longitudinal scratches are formed due to improper or rough handling. Circumferential scratches can form if lifted or handled using rough slings (e.g., chains). Also, sharp scratches have more detrimental effects than blunt scratches. Longitudinal scratches can eventually lead to longitudinal split failures.
	Scratch depth	Fatigue failure becomes an important consideration for deeper scratches, especially when they exceed 10% of pipe wall thickness.
Service connection	Split at tap	Inadequate tapping procedure or thin pipe walls can lead to a split in the PVC mains, usually on the inside of the pipe. This type of failure is commonly referred to as a fitting failure.
Joint	Change in alignment	Changes in joint alignment (rotation) indicate pipe is susceptible to ground movement. Large changes can lead to leakage.
	Joint displacement	Joints can displace without undergoing joint misalignment and hence are also an indicator of other forces at play.

Table 3.9 Inferential indicators for cast iron pipes (Adapted from Liu et al. 2012)

Category	**Agent**	**Comment**
Pipe vintage	Material type, historic standards, and installation practices	Pipes of specific vintages can experience a higher breakage rate.
Pipe joint	Joint type	Anecdotal reports indicate that leadite joints have performed poorly over the years.
Water quality	Water pH	Water with a low pH can leach the internal cement lining or pipe wall itself if lining is absent.
Water pressure	Operating pressure (OP)	High pressure subjects pipe to high stress and hence the pipe has a higher propensity to fail.
	Pressure change amplitude (% OP)	Large pressure changes (% of OP) can induce higher stresses than expected by design.
	Pressure change frequency	A slow or fast fatigue mechanism can induce early failure.
Location	Pipe embedment	Pipes exposed to wet/dry conditions have a higher failure rate than pipes totally below the water table or pipes totally exposed to the atmosphere.
	Surface loads - traffic type	Heavy surface loads will subject the pipe to high stresses and hence faster deterioration in the long term.
	Wet/dry cycle(s)	A changing environment can promote corrosion.
	Water table level	Water table position indicates if wet/dry cycle is likely to occur.
Soil	Soil type / backfill	Non-draining backfill leads to moisture retention and promotes corrosion.
	Soil resistivity	Low resistivity soil leads to higher corrosion rates. Soil chlorides (e.g., from de-icing salts) reduce soil resistivity.

Table 3.9 continued

Category	**Agent**	**Comments**
Soil	Soil chloride	Low chloride levels in high pH (> 11.5) environments can lead to corrosion.
	Soil sulfate	This accounts for microbial induced corrosion (MIC) and a possible food source for sulfate-reducing bacteria in anaerobic conditions under loose coatings.
	Soil sulfide	Sulfate-reducing bacteria give off sulfides that are excellent electrolytes.
	Soil pH	Low pH (< 4) means soil is acidic and likely to promote corrosion. High alkaline conditions (pH > 8) can also lead to
	Redox potential	High availability of oxygen promotes microbial-induced corrosion in the presence of sulfates and sulfides.

Table 3.10 Inferential indicators for AC pipes (Adapted from Liu et al. 2012)

Category	**Agent**	**Comment**
Pipe vintage	Material type, historic standards, and installation practices	Pipes of specific vintages have experienced a higher breakage rate.
Water quality	Water pH	Water with a low pH can leach the cement within the AC matrix.
	Water saturation index (SI)	Water with SI < 0.25 can leach the cement within the AC matrix.
Water pressure	Operating pressure (OP)	High pressure subjects pipe to high stress and hence a higher chance to fail.
	Pressure change amplitude (% OP)	Large pressure changes (% of OP) can induce higher stresses than expected by design.
	Pressure change frequency	The fatigue mechanism is not observed or documented for AC pipes.
Location	Surface loads—traffic type	Heavy surface loads subject the pipe to high stresses and thus to faster deterioration in the long term.
	Wet/dry cycle(s)	A changing environment promotes higher expansion of matrix than an unchanging environment.
	Water table level	The water table position will indicate if wet/dry cycle is likely to occur. Soil sulfate attack only occurs if sulfate is in solution.

Table 3.11 Inferential indicators for PVC pipes (Adapted from Liu et al. 2012)

Category	**Agent**	**Comment**
Pipe vintage	Material type, historic standards, and installation practices	Most PVC pipes used in North America are of the unplasticized PVC type. Newer modified PVC and oriented PVC have recently appeared on the market. Failures could be tied to certain manufacturing processes and standards or installation practices. Knowledge of the installer could also help to identify poor vs. adequate installation practices.
Water pressure	Operating pressure (OP)	High pressure subjects the pipe to high stress and hence a higher propensity to fail. Time to failure can be substantially reduced in PVC pipes under high pressure since PVC is a visco-elastic material.
	Pressure change amplitude (% OP)	Large pressure changes (% of OP) can induce higher stresses than expected by design.
	Pressure change frequency	The fatigue mechanism is the primary mechanism of PVC pipes if scratches or gouging are present.
Location	Surface loads—traffic type	Heavy surface loads will subject the pipe to high stresses and hence to faster deterioration in the long term, especially if PVC pipes have been previously scratched or gouged.
Soil	Hydrocarbons	PVC pipes are impervious to high-octane gasoline and gasoline-saturated water for periods of up to 2 years.
	Frost susceptibility (load)	PVC pipes are not designed for frost loads. If conditions exist to develop significant frost loads, then the pipe will be subjected to additional stresses (annual), leading to pipe failure if already significantly scratched.

3.5 **Pipe Failure Models**

Pipe failure in water networks happens when pipes are no longer able to meet water quantity and quality demands. In general, failure models help the utility companies to target pipes that require replacement and maintenance and rehabilitation of pipe work accordingly. Deterioration models of water mains are classified into different categories according to recent publications. Rostum (2000) categorizes pipe deterioration models into three groups: description analysis, physical/mechanical models, and statistical models. Liu et al. (2012) classified the models into two broad categories: physical/mechanistic models and statistical/empirical models.

As described in Section 1.3, each group of models do not provide individual pipe information. Physical models require expensive data and the data is expensive to obtain. While statistical models are a representation of the system and they can be used with various levels of input data.

Clair and Sinha (2012) performed extensive research on the available pipe deterioration models and divided the deterioration models into the following groups: deterministic, statistical, probabilistic, and advanced mathematical models, which consist of artificial neural networks (ANN), fuzzy logic, and heuristics. The following section discusses the models categorized by Clair and Sinha (2012), as shown in Figure 3.4.

3.5.1 **Deterministic model approach**

Deterministic models are mainly used in cases where the relationship between components is certain. Deterministic models can be empirical and mechanistic. In the empirical approach, failure rates are related to the attributes of the asset. This approach is applied only to cohorts of pipes. The mechanistic approach predicts the service lifetimes of individual assets. Following are some of the features of the deterministic model approach:

- The majority of the deterministic models are structural or functional performance models and primary response models.
- The model can predict an average single value of a dependent variable.
- Most of the existing prediction models have been developed through regression analysis, combined mechanistic-empirical analysis, and opinions from experienced engineers.
- The problem of the deterministic-prediction model is that the applicability of each individual model is restricted to a specific location.

In general, deterministic modeling is difficult to implement for an entire water pipe as sample provided would represent a certain section of pipe. Deterministic models are very site-specific and an entire system can only be generalized if the conditions remain the same throughout the site.

3.5.2 **Statistical Model Approach**

Statistical modeling is used for predicting the lifetime or failure time of infrastructure. Statistics are used to develop models from observed data. The statistical methodology is typically applied to asset cohorts that have recorded historical failure or condition data. Some of the features of the statistical model approach are listed below:

- Pipe infrastructure data are required to predict future condition.
- The condition of pipe ranking is required to predict future condition.

- This type of modeling is applied to a homogenous group of pipe infrastructure systems.
- Common features among the statistical models are that they are based on much long-term observed field data and processed through regression analysis.

A noted limitation of the regression-based approach is, it is not suitable for modeling the actual deterioration processes of a pipe since the sampling data frequently suffers from various limitations and selected examples (Clair and Sinha 2012).

3.5.3 **Probabilistic Model Approach**

Probabilistic modeling analyses the probability or relative frequency of an event occurring. The likelihood of these occurrences helps to describe the failure of an asset. Condition data and asset attribute information is needed in modeling the probability of failure.

Statistical and probabilistic pipe deterioration models are more popular due to lack of data availability for underground pipes. This is true for water pipe and sewer pipes. Scheidegger et al. (2011) stated that a major reason for problems of sewer pipe failure prediction models is the lack of complete and reliable datasets. In addition to insufficient data management, considerable errors in the derivation of data from CCTV can create substantial uncertainty in the available data. Followings are some of the benefits of simulation of data set for pipes Scheidegger et al. (2011):

- It allows scientists and utility managers to identify limitation of current deterioration models.

- It guides data management in terms of appropriate time-intervals as well as attributes that have to be collected for reliable model applications.

The same benefits can be applied to water pipe deterioration models with generation of data set. Copula method can achieve this.

3.5.4 **Advanced Mathematical Models**

Neural Network (NN) Approach

A neural network consists of interconnected processing elements often referred to as "neurons" that work together to provide a result. When properly trained, NN models can mimic the functioning of the human brain through pattern recognition and generalization capabilities. ANN models consist of the following characteristics:

- Nodes and the interconnections arranged within the layer of a given ANN determine its topology. The choice of the given topology depends on the type of problem.
- Increased levels of skill and training are required for these networks.
- Quality labeled data are required for supervised training and predicting the future condition.

Fuzzy Logic Model Approach

Fuzzy logic, a mathematical method, is used to deal with systems with uncertain information. Some of the characteristics of the fuzzy method are as follows:

- Fuzzy logic models are applied to some areas of infrastructure management, namely bridges, highways, oil and gas pipelines, and water pipe networks.

- Challenges are constructing fuzzy rule sets, selecting membership, and determining the defuzzification process.
- This technique implements the opinions of experts.
- These are used for systems that are subject to uncertainties, ambiguities, and contradictions.

Heuristic Model Approach

Heuristic models are for problems that are not well understood. This method illustrates failure risks with limited or no pipe data. The method is developed through subjective opinions from experienced field engineers and experts.

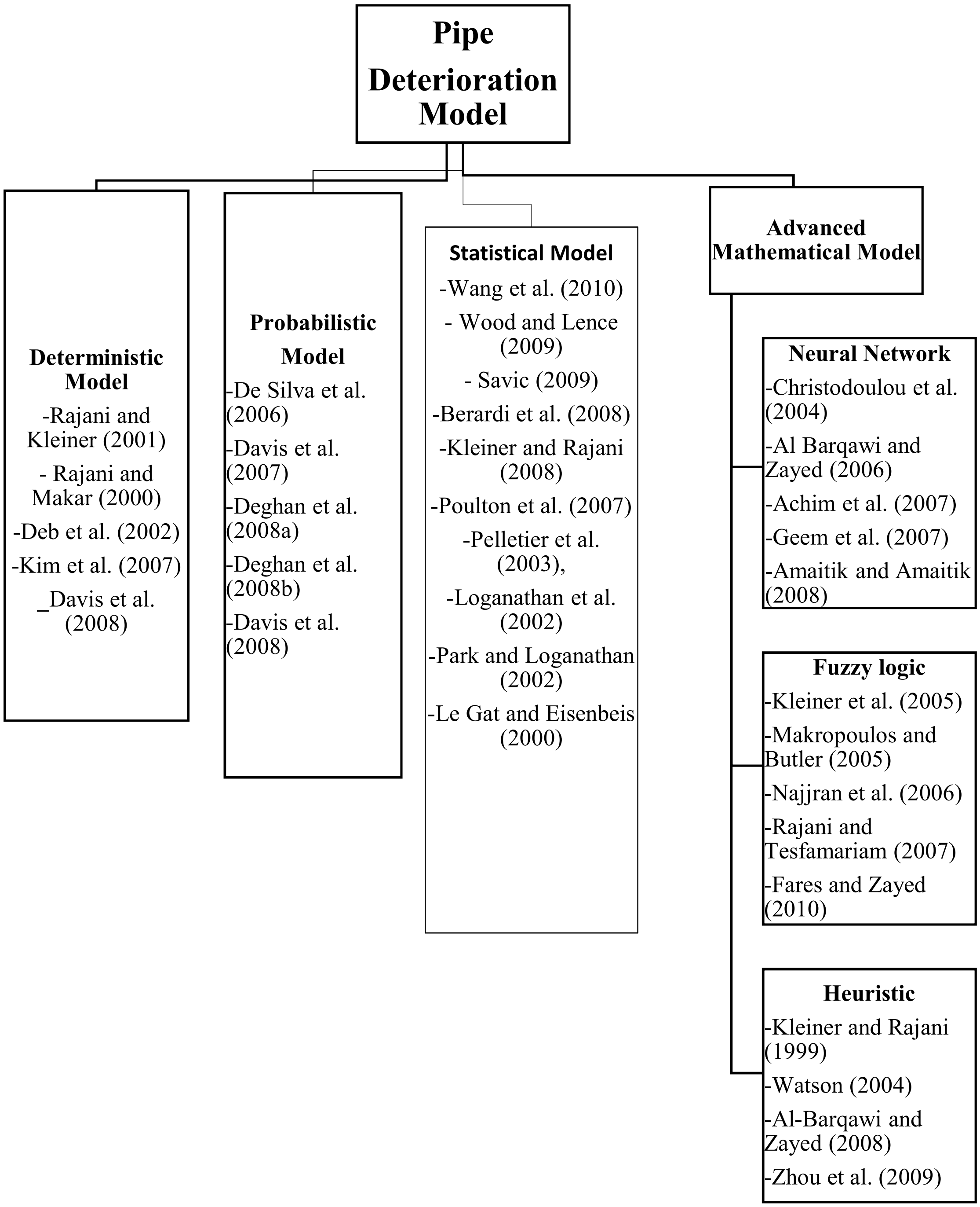


Figure 3.4 Classification of pipe deterioration model (Clair and Sinha 2012)

3.6 **Concluding Remarks**

Copula modeling falls into the probabilistic modeling category. Pipe deterioration process is often multidimensional and hence requires the joint modeling of several random variables. Traditionally, the pairwise dependence between variables has been described using classical families of bivariate distributions. Some of the most common models are the bivariate normal, lognormal, gamma, and extreme-value distributions. Bivariate normal being the most popular used bivariate distribution. The main limitation of this approach is that the same parametric family of univariate distributions must then characterize the individual behavior of the two variables. Copula method is free from this restriction. The main advantage of copula approach is that the selection of an appropriate model for the dependence between variables, represented by the copula, can proceed independently from the choice of the marginal distributions. Copula modeling also help to generate large data set. Multivariate simulation for copula modeling can be used to determine the effectiveness and comparison of the existing pipe deterioration models.

Chapter 4

COPULA MODELING

4.1 Introduction

The normal distribution has dominated the study of multivariate distributions (Frees and Valdes 1998). The choice of the multivariate normal distribution is appealing as the marginal distributions are also normal and the association between two random variables can be fully described by knowing a) their marginal distribution and b) the correlation coefficient. However, in engineering applications, non-normality can occur in such ways:

- The marginal distribution of some of the variables may not be normal.
- In some cases, even though all the marginal distributions are normal, jointly these variables may not be multivariate normal (Yan 2006).

Copulas are useful for generating joint distributions by combining given marginal distributions and a specified form of a copula function. Copulas are also appealing because they capture dependence more broadly than the standard multivariate normal framework.

Copulas were introduced in 1959 by Sklar. An n-dimensional copula is a multivariate distribution function defined on the unit cube $[1,0]^n$ with uniformly distributed marginal (Sklar 1959). The idea of dependence modeling with copula functions is based on Sklar's theorem. It states that if H is an n-dimensional cumulative distribution function with continuous marginal cumulative distributions F_1, $F_2, \ldots\ldots, F_n$ then there exists an n-dimensional copula C such that for all the real variables of $s = [s_1 s_2 \ldots\ldots s_n]^T$, i.e.,

$$H(s) = C(F_1(s_1), F_2(s_2), \dots., F_n(s_n)) \qquad \text{Eq. 4.1}$$

where C is unique for marginal distribution functions which are continuous.

Sklar's theorem also shows that for an n-dimensional joint distribution function, the n marginal distributions and the dependence structure can be separated. The copula function can completely describe the dependence between the n variables. By using the converse of Sklar's theorem, the multivariate distribution can be defined by linking the n univariate continuous marginal distributions of any type with the copula function. From equation (1), the multivariate probability density function h can be written as (Srinivas et al. 2006)

$$h(s) = \frac{\partial^n \left[C\left(F_1(s_1), \dots., F_n(s_n)\right)\right]}{\partial F_1(s_1), \dots \dots, \partial F_n(s_n)} \prod_{i=1}^{n} f_i(s_i) \qquad \text{Eq. 4.2}$$

$$h(s) = c(F_1(s_1), \dots \dots \dots \dots, F_n(s_n)) \prod_{i=1}^{n} f_i(s_i) \qquad \text{Eq. 4.3}$$

where $c(F_1(s_1), \dots \dots \dots \dots, F_n(s_n))$ = copula density associated with the copula function $C(F_1(s_1), \dots \dots \dots \dots, F_n(s_n))$ and f_i = marginal probability density function corresponding to F_i .

The basis for copula modeling can be described as shown in Figure 4.1.

There are many functions which exist satisfying the mathematical condition for being a copula function. The functions are known as families of copula functions, and there are two major types of copula: a) elliptical copulas and b) Archimedean copulas.

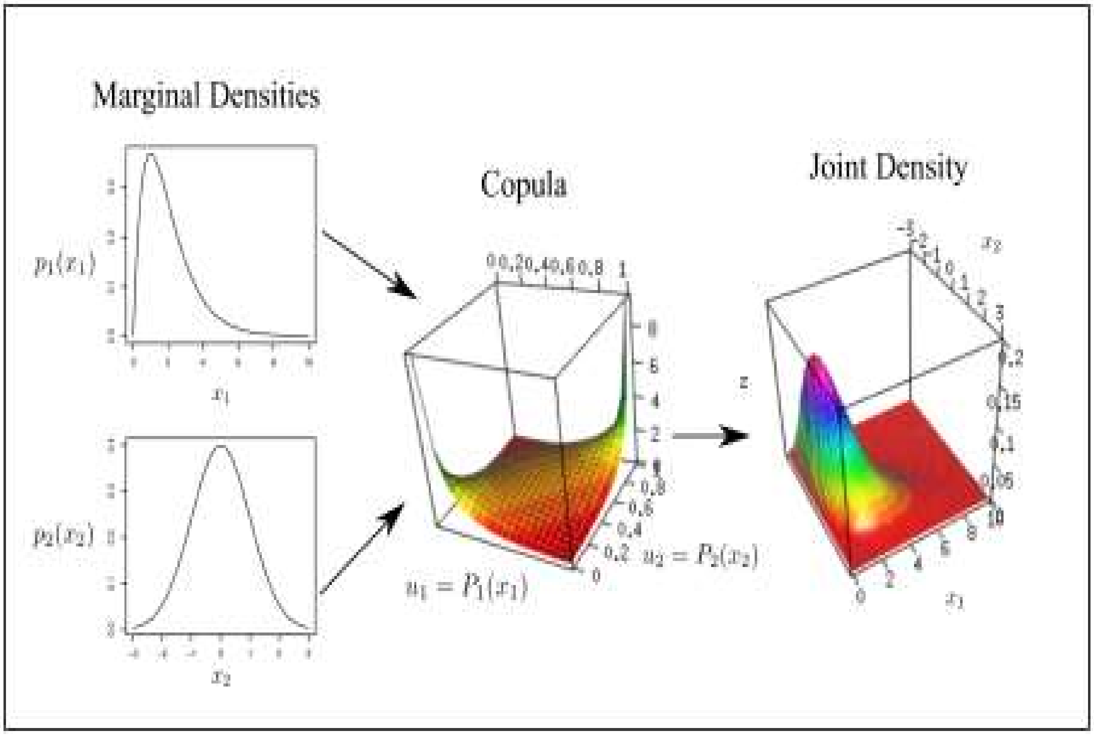


Figure 4.1 Copula modeling in a diagram (Hernandez-Lobato, 2016)

4.2 **Copula Family**

Generally there are two major types of copula: a) elliptical copulas and b) Archimedean copulas. These two types are discussed in the following sections.

4.2.1 **Elliptical Copulas**

Elliptical copulas are related to elliptical distributions. They have properties that are similar to multivariate normal distributions. The class of elliptical distributions provides a good source of multivariate distributions, which share many of the tractable properties of the multivariate normal distributions, and enables modeling of multivariate extremes and other forms of non-normal dependences (Embrechts et al. 2001). Two important copulas in this family are the Gaussian copula and the student's t copula. The student's t copula has a symmetrical dependence structure.

The elliptical copula can be defined by the following form:

$$C(q) = H\left(F_1^{-1}(q_{1,}F_2^{-1}(q_2), \dots \dots, F_2^{-1}(q_n)\right) \qquad \text{Eq. 4.4}$$

$$q = [q_1\, q_2 \dots \dots q_n] \in [0,1]^n \qquad \text{Eq. 4.5}$$

4.2.2 **Archimedean Copulas**

Archimedean copulas are a very popular copula family and have a wide range of applications for the following reasons (Bacigal 2006):

- They can be constructed easily.
- There are many comprehensive properties of the members of this family class.
- They can be applied whether the correlations between variables are positive or negative. The Archimedean bivariate copula is expressed as (Nelsen 1999)

$$C(u, v) = \varphi^{-1}\left(\varphi(u) + \varphi(v)\right), \quad 0 < u, v < 1 \qquad \text{Eq. 4.6}$$

Where φ, is a convex decreasing function from [0, 1] to [0, ∞] such that $\varphi(1) = 0$. Differentiating the generator twice, the copula density function is obtained:

$$c_{\varphi} = \frac{\partial^2 C_{\varphi}(u,v)}{\partial u \partial v} \qquad \text{Eq. 4.7}$$

By having different generators, there are forms of several important copulas. Some of the widely known Archimedean copulas are summarized in Table 4.1.

Table 4.1 Archimedean copulas (adapted from Attoh-Okine 2013)

Family of copula	Generator $\emptyset_t$	Bivariate copula $\boldsymbol{C_{\varphi}(u,v)}$
Independence	$-\ln t$	uv
Gumbel	$(-\ln t)^{\theta}$	$e^{-\left[(-\ln u)^{\theta}+(-\ln v)^{\theta}\right]^{-1/\theta}}$
Clayton	$t^{-\theta}-1$	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$
Frank	$-ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$\frac{1}{\theta} ln\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right)$

The Clayton copula tends to work well where there is strong left tail dependence; the Gumbel copula is good for positive tail dependence rather than negative tail dependence. Meanwhile, the Frank copula is a symmetric Archimedean copula where tail dependence is weak. Figure 4.2 and Figure 4.3 show the common copula family distributions.

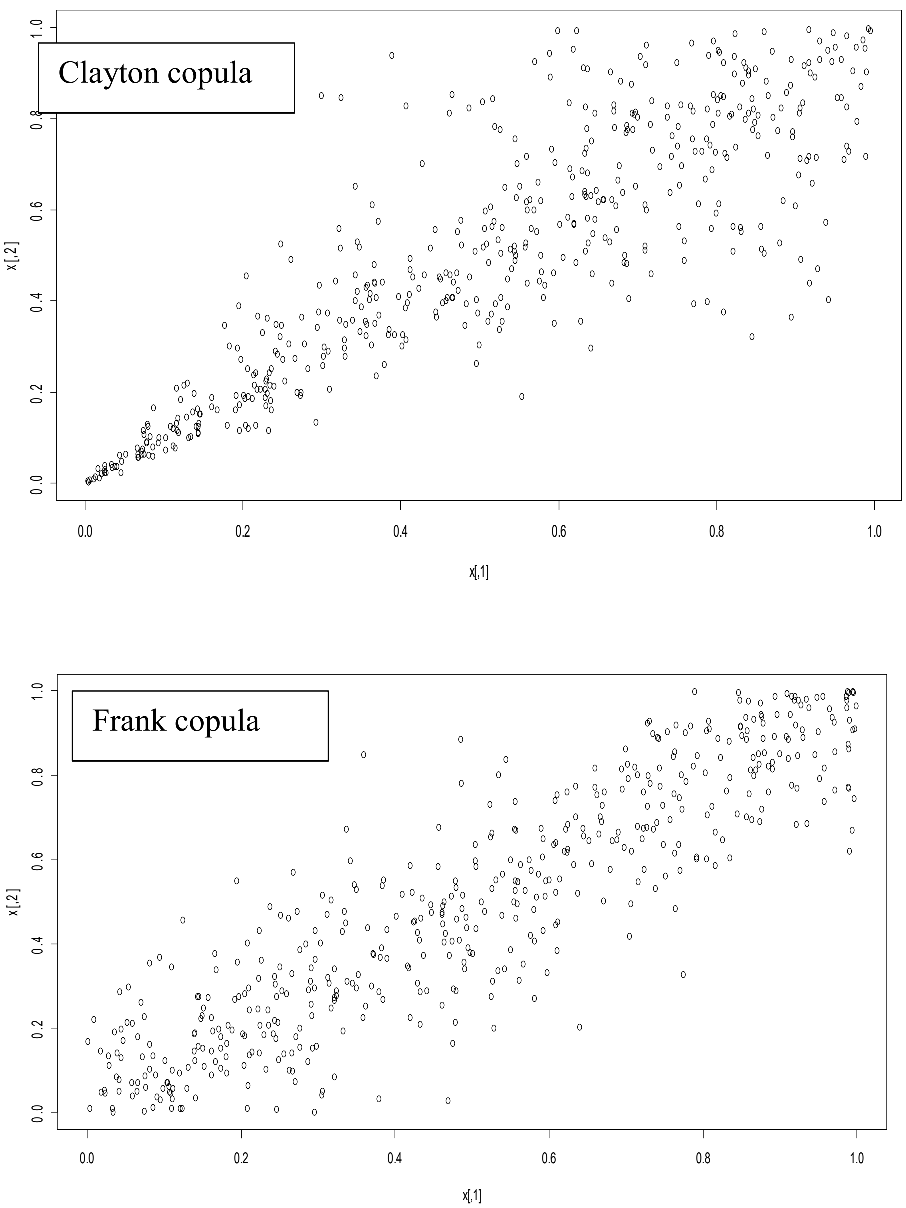


Figure 4.2 Common copula families (Clayton and Frank copula)

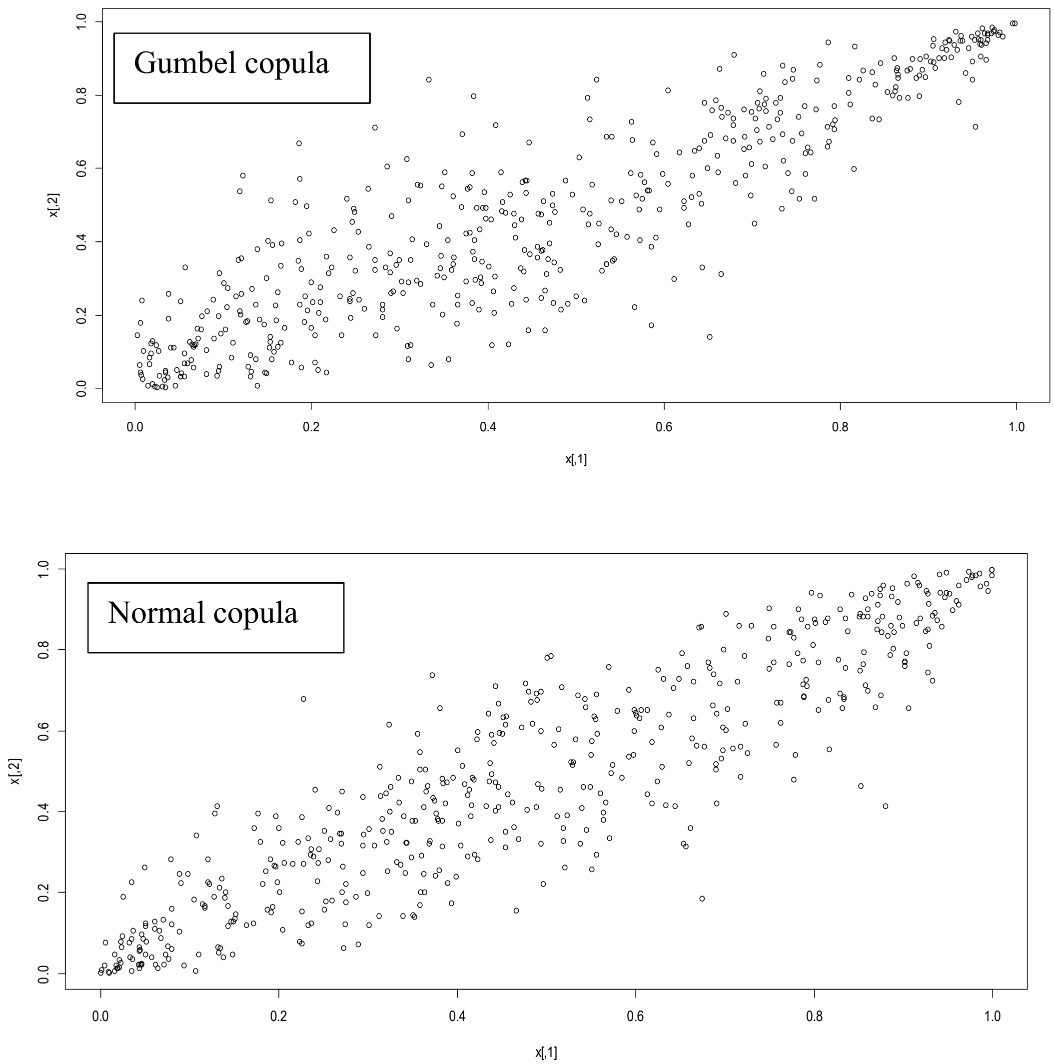


Figure 4.3 Common copula families (Gumbel and normal copula)

Table 4.2 The comparison between elliptical copulas and Archimedean copulas (adapted from Aas 2004)

	Elliptical copula	**Archimedean copula**
Definition	Elliptical copulas are of elliptically contoured distributions. Widely known elliptical distributions are: normal (Gaussian) and student's t copula.	Archimedean copulas are easily constructed and have attractive properties. Commonly used distributions are: Clayton, Frank, and Gumbel.
Advantages	The correlation between the marginals can be easily determined.	They are easily deduced.
Disadvantages	Absence of closed form expressions and impossible to have radial symmetry	The definition does not extend to a multivariate data set of *n* variables as there will be multiple values of tau.

4.3 Measure of Dependence

Common measures that are used for analyzing dependence are the 1) Pearson, b) Spearman, and 3) Kendall coefficients. For copula dependence measurement, Spearman's rho and Kendall's tau are used. The Kendall's tau (τ) of pair(x, y), distributed according to H, can be expressed as the difference between probabilities of concordance and discordance for two independent pairs (x_1, y_1) and (x_2, y_2), each with H distribution:

$$\tau(x,y) = \Pr\big((x_1 - x_2)(y_1 - y_2) > 0\big) - Pr\big((x_1 - x_2)(y_1 - y_2) < 0\big) \quad \text{Eq. 4.8}$$

where $(x_1 - x_2)(y_1 - y_2) > 0$ is concordant and $(x_1 - x_2)(y_1 - y_2) < 0$ is discordant. Kendall's tau can be expressed as

$$\tau(x,y) = \frac{\#Concordant\ pairs - \#Discordant\ pairs}{\#pairs} \qquad \text{Eq. 4.9}$$

Kendall's tau in terms of a copula can be expressed as

$$\tau(x,y) = \tau(C) = 4\int_0^1 \int_0^1 C(u,v)dC(u,v) - 1 \qquad \text{Eq. 4.10}$$

where *C(u,v)* is the copula of the bivariate distribution function of x and y. For the Gaussian and student's t-copulas and all other elliptical copulas, the relationship between the correlation coefficient and Kendall's tau is given by

$$cor(x,y) = \sin\left(\frac{\pi}{2}\tau\right) \qquad \text{Eq. 4.11}$$

where *cor* is the linear correlation coefficient.

Schweizer and Wolf (1981) established that for Archimedean copulas, Kendall's tau can be related to the dependence parameter; for the Clayton copula, it is

$$\tau(x,y) = \frac{\delta}{\delta+2} \qquad \text{Eq. 4.12}$$

For the Gumbel copula, it is

$$\tau(x,y) = 1 - \frac{1}{\delta} \qquad \text{Eq. 4.13}$$

Some of the properties listed for Kendall's tau include (Djehiche et al. 2004):

- Is insensitive to outliers
- Measures the average dependence between x and y
- Is invariant under strictly increasing transformations

Spearman's rho (ρ) can be expressed as follows:

$$\rho_C = 12 \int_0^1 \int_0^1 (C(u,v) - uv) dudv \qquad \text{Eq. 4.14}$$

Here *C* (*u, v*) is the copula of the bivariate distribution function of *X* and *Y*. Let *X* and *Y* have the distribution functions *F* and *G*, respectively. Then the following relationship between Spearman's rho and the linear correlation coefficient is obtained:

$$\rho_c = C(F(X), F(Y)) \qquad \text{Eq. 4.15}$$

For the Gaussian and student's t-copulas, the relationship between the linear correlation coefficient and Spearman's rho is

$$cor(X,Y) = 2\sin(\frac{\pi}{6}\rho_c) \qquad \text{Eq. 4.16}$$

Both $\tau_c$and $\rho_c$may be considered as measures of the degree of monotonic dependence between *X* and *Y*, whereas linear correlation measures the degree of linear dependence. Also, these measures are invariant under monotone transformations, while the linear correlation generally is not (Aas 2004).

4.4 Parameter Estimation

This section deals with the method of copula parameter estimation. Rank-based estimators are considered for estimation. The dependence structure captured by copulas does not have anything to do with the individual behavior of the variables, so any inference about the parameter indexing a family of copulas should rely on the rank of the observations (Genest and Favre 2007). Methods normally used for estimation are estimates based on Kendall's tau, Spearman's rho, and maximum pseudolikelihood.

4.4.1 Estimate Based on Kendall's Tau

The Kendall's tau estimator is based on the relationship between Kendall's tau and the copula parameter. More generally, if $\theta = g(\tau)$ for some smooth function g, then $\widetilde{\theta_n} = g(\tau_n)$, which may be referred to as the Kendall-based estimator of θ. It is implied that

$$\sqrt{n}\ \frac{\tau_n - \tau}{4S} \approx \mathcal{N}(0,1) \qquad \text{Eq. 4.17}$$

where

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(W_i + \widetilde{W_i} - 2\widetilde{W})^2 \qquad \text{Eq. 4.18}$$

and

$$\widetilde{W_i} = \frac{1}{n}\sum_{j=1}^{n} I_{ji} = \frac{1}{n}\#\{j: X_i \ll X_j, Y_i \ll Y_j\} \qquad \text{Eq. 4.19}$$

4.4.2 **Estimate Based on Spearman's Rho**

With the dependence θ parameter being real, an alternative rank-based estimator that remains in the spirit of the method of moments consist of taking

$$\widetilde{\theta_n} = h(\rho_n) \qquad \text{Eq. 4.20}$$

where $\theta = h(\rho)$ represents the relationship between the parameter and the population value of Spearman's rho. Following the convergence results, it is established that

$$\rho_n \approx N\left(\rho, \frac{\sigma^2}{n}\right) \qquad \text{Eq. 4.21}$$

4.4.3 **Estimate Based on Maximum Pseudolikelihood**

Here the maximum pseudolikelihood estimator is discussed in detail. The method which requires C_θ be continuous with density c_θ involves maximizing a rank-based log-likelihood of the form

$$l(\theta) = \sum_{i=1}^{n} log\left\{c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)\right\} \qquad \text{Eq. 4.22}$$

Where R_i stands for the rank of X_i among $X_1, \dots \dots, X_n$, and S_i stands for the rank of Y_i among $Y_1, \dots \dots, Y_n$. The equation is the same expression obtained when the unknown marginal distributions F and G in the classical log-likelihood

$$l(\theta) = \sum_{i=1}^{n} \log[c_\theta\{F(X_i), G(Y_i)\}] \qquad \text{Eq. 4.23}$$

are replaced by rescaled versions of their empirical counterparts

$$F_n(x) = \frac{1}{n+1}\sum_{i=1}^{n} 1(X_i \ \leq x) \qquad \text{Eq. 4.24}$$

and

$$G_n(y) = \frac{1}{n+1}\sum_{i=1}^{n} 1(Y_i \ \leq y) \qquad \text{Eq. 4.25}$$

This substitution yields a formula as it is realized that $F_n(X_i) = R_i/(n+1)$ and $G_n(Y_i) = S_i/(n+1)$ for all $i \in \{1, \ldots, n\}$.

Compared to Kendall's tau and Spearman's rho, this method may seem less attractive as it requires numerical work and requires the existence of a density c_θ. However, it is much more applicable than the other methods since it does not require the dependence parameter to be real. Letting $\dot{c_\theta}(u, v) = \partial c_\theta(u, v)/\partial\theta$, Genest et al. (1995) showed that

$$l(\dot{\theta}) = \frac{\partial}{\partial\theta} l(\theta) = \sum_{i=1}^{n} \frac{c_\theta(\frac{R_i}{n+1}, \frac{S_i}{n+1})}{\dot{c_\theta}\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)} = 0 \qquad \text{Eq. 4.26}$$

is unique. Furthermore,

$$\hat{\theta}_n = N(\theta, \frac{\nu^2}{n}) \qquad \text{Eq. 4.27}$$

where ν^2 depends exclusively on the underlying copula C_θ. The estimate of ν^2 is given by

$$\hat{v}_n^2 = \hat{\sigma}_n^2 / \hat{\beta}_n^2 \qquad \text{Eq. 4.28}$$

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(M_i - \bar{M})^2 \qquad \text{Eq. 4.29}$$

and

$$\hat{\beta}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(N_i - \bar{N})^2 \qquad \text{Eq. 4.30}$$

are sample variances computed from two sets of pseudo-observations with means $\bar{M} = (M_1 + \ldots + M_n)/n$ and $\bar{N} = (N_1 + \ldots + N_n)/n$.

4.4.4 **Other Estimation Methods**

Most common estimators are based on the maximization of the pseudolikelihood and on the inversion of either Kendall's tau or Spearman's rho. Among other estimator methods is a parametric two step procedure referred to as the "inference from margins" or the IMF method. The estimate of θ is obtained through the maximization of a function form:

$$l(\theta) = \sum_{i=1}^{n} log\left[c_\theta\{\hat{F}(X_i), \hat{G}(Y_i)\}\right] \qquad \text{Eq. 4.31}$$

The rank-based method takes $\hat{F} = F_n$ and $\hat{G} = G_n$ where (F_δ) and (G_δ) = suitable parametric families for the margins, and δ_n and η_n= standard maximum likelihood estimates of their parameters, derived from observed values of X and Y, respectively.

4.5 **Graphical Diagnostics**

For bivariate data, checking the adequacy of a copula model is to compare a scatter plot of the pair with an artificial data set of the same size generated from C_{θ_n}. For the bivariate case, a good method of generating a pair (U, V) from a copula C proceeds as follows:

- Step 1: Generate U from a uniform distribution on the interval (0, 1).
- Step 2: Given $U=u$, generate V from the conditional distribution:

$$Q_u(v) = P(V \leq v | U = u) = \frac{\partial}{\partial u} C(u, v) \qquad \text{Eq. } 4.32$$

by setting $V = Q_u^{-1}(U^*)$ where U^*= another observation from uniform distribution on the interval (0, 1). When an explicit formula does not exist for Q_u^{-1} the value $v = Q_u^{-1}(u^*)$ can be determined by trial and error.

Other options include chi-plots and k-plots applications. Chi-plots are based on chi-square statistics for independence in a two-way table. A chi plot is used as a graphical tool for detecting dependence.

For chi-plots, Genest and Favre (2007) introduced

$$H_i = \frac{1}{n-1} \#\{j \neq i : X_j \leq X_i, Y_j \leq Y_i\} = \frac{nW_i - 1}{n-1} \qquad \text{Eq. } 4.33$$

$$F_i = \frac{1}{n-1} \#\{j \neq i, X_j \leq X_i\} \qquad \text{Eq. } 4.34$$

and

$$G_i = \frac{1}{n-1} \#\{j \neq i, Y_j \leq Y_i\} \qquad \text{Eq. } 4.35$$

The k-plot is inspired by the notion of the Q-Q plot. The technique consists of plotting the pairs $\left(W_{i:n},\ H_{(i)}\right)$ for $i \in \{1, \dots, n\}$ where

$$H_{(1)} < \cdots < H_{(n)} \qquad \text{Eq. } 4.36$$

$W_{i:n}$ is the expected values of the *ith* statistic from a random sample of size *n* from a random variable $W = C(U, V) = H(X, Y)$ and under the null hypothesis of independence between U And V, where

$$W_{i:n} = n\begin{pmatrix} n & - & 1 \\ i & - & 1 \end{pmatrix}\int_0^1 wk_0(w)\{K_0(w)\}^{i-1}\{1 - K_0(w)\}^{n-i}dw \qquad \text{Eq. } 4.37$$

and

$$K_0(w) = P(UV \leq w) = \int_0^1 P\left(U \leq \frac{w}{v}\right)dv \qquad \text{Eq. } 4.38$$

$$= \int_0^w 1dv + \int_w^1 \frac{w}{v}\,dv = w - wlog(w) \qquad \text{Eq. } 4.39$$

and k_0 = corresponding density.

When K-plots are used for comparison of copula models, the process consists of comparing the empirical distribution K_n of the variables $W_{1,\dots\dots,}W_n$ with K_{θ_n}, the theoretical distribution of $W = C_{\theta_n}(U, V)$, where *(U, V)* is drawn from C_{θ_n}. One option is to plot K_n and K_{θ_n} on the same graph to see how well they agree. A Q-Q plot can be derived from the order statistics $W_{(1)} \leq \cdots \leq W_{(n)}$ by plotting the pairs

$(W_{i:n}, W_{(i)})$ for $i \in \{1, \ldots, n\}$. In this case, $W_{i:n}$ is the expected value of the *ith* order statistics from a random sample of size *n* from K_{θ_n} rather than from K_0

where,

$$W_{i:n} = n\binom{n-1}{i-1}\int_0^1 wk_{\theta_n}(w)\{K_{\theta_n}(w)\}^{i-1}\{1 - K_{\theta_n}(w)\}^{n-i}dw \qquad \text{Eq. 4.40}$$

where,

$K_{\theta_n}(w) = P\{C_{\theta_n}(U,V) \le w\}$ and $k_{\theta_n} = d\, K_{\theta_n}(w)/dw$

4.6 Goodness of Fit Formal Tests

The formal methodology for goodness of fit testing of Copula is in the early stage of application. Genest et al. (2006) proposed the following statistic form of

$$S_n = \int_0^1 |\mathbb{K}_n(w)|^2 k_{\theta_n}(w)dw \qquad \text{Eq. 4.41}$$

and

$$\tau_n = \frac{sup}{0 \le w \le 1} \ |\mathbb{K}_n(w)| \qquad \text{Eq. 4.42}$$

Where $\mathbb{K}_n(w) = \sqrt{n}\{K_n(w) - K_{\theta_n}(w)\}$, Genest et al. (2006) showed the following method of computing:

$$S_n = \frac{n}{3} + n\sum_{j=1}^{n-1} K_n^2\left(\frac{j}{n}\right)\left\{K_{n_\theta}\left(\frac{j+1}{n}\right) - K_{\theta_n}\left(\frac{j}{n}\right)\right\} \qquad \text{Eq. 4.43}$$

The p-values associated with these statistics are easy to obtain by bootstrapping.

4.7 **Pipe Data Analysis**

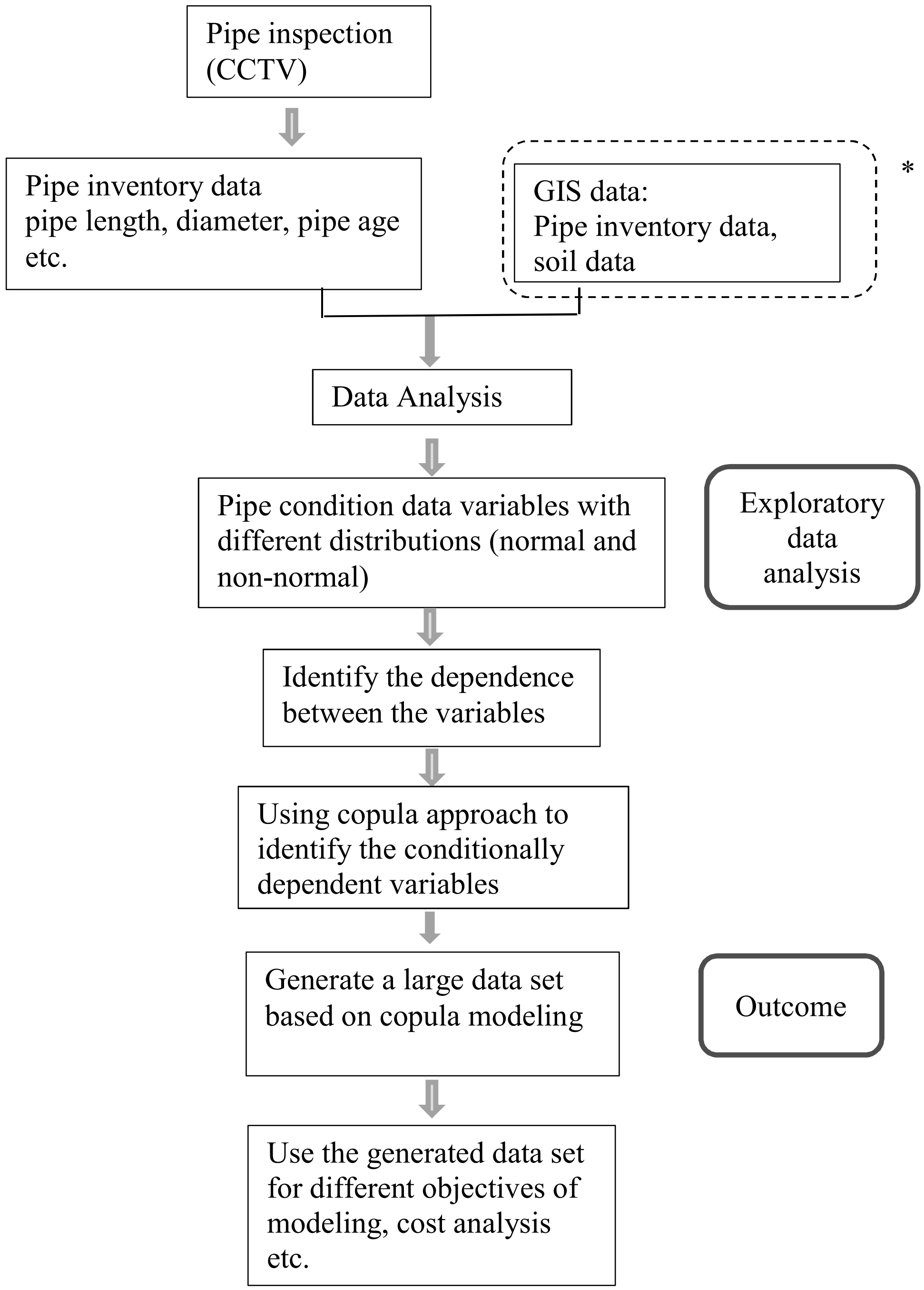
Two separate sets of pipe data were used for analysis copula modeling in this chapter. The details of the data sets used for analysis can be found in the Appendices as Data Set #1 and Data Set #2. Figure 4.4 shows a general flow chart for pipe data analysis using copula method. In the beginning pipe inspection data obtained from methods described in Section 2.2 are combined with pipe inventory data such as pipe length, pipe age etc. If GIS data are available, that can also be used for the pipe data analysis. In the data analysis phase, copula modeling can be used to determine the dependence between variables. In pipe deterioration process there is not only a large number of processes that can lead to the degradation of pipes, but also a correspondingly large number of exploratory variables (Scheidegger et al. 2011). Pipe construction has evolved through many different construction processes and use of different construction materials. In addition, the site conditions, i.e. soil conditions, water table location, water quality etc. can all affect the rate of deterioration. It is important to find, determine and chose exploratory variables for useful prediction models. For cases where the dependence is non linear and skewed, Spearman's rho and Kendall's tau can be better methods for dependence analysis. Dependence analysis can help to determine the relevancy of pipe deterioration variables. Generated large data sets can be used for different objectives. Based on bivariate copula dependence modeling, generated data points are used for regression analysis. Multivariate simulation for copula modeling can also be used to determine the effectiveness and comparison of the existing pipe deterioration models.

4.7.1 **Dependence Analysis of Pipe Data**

A set of water pipe data was obtained from a utility company in the desert west region. The data set (Data Set #1 in Appendix A) included pipes ranging from 4” diameter to 66” diameter. The oldest pipe was installed in 1943, and the newest one was installed in 2009. The data notation used for the variables is shown in Table 4.3. For analysis, 840 observations were used. These pipes had repairs done to them and were in active status. From Material variable M indicates material type indicated by code. The numbering of the material and corresponding material types are shown in Table 4.4. Distribution class varieties include distribution, fire hydrant lateral, service lateral, and transmission. The scatter plots of the variables are shown in Figure 4.5. For an example, the second column and top row gives the dependency figure for pipe material and pipe age. The x-axis shows the material code from 10 to 60 whereas the pipe age varies from 1 to 40,000. Since the age of the pipe can be a maximum of 24,000 days based on installation day, 40,000 value was an error. The data point is a PVC pipe that had an arbitrary installation date of 1/1/1900. The scatter plot provides an initial revelation that the variables do not provide joint normal distribution. Marginals of some of the variables are also found to be non-normal in further analysis. This provides a good opportunity to study copula modeling for this set of variables.

Table 4.3 Notation used for data analysis

Variable	Unit	Meaning
D	inch x 100	Diameter of pipe
M	Code number by material	Pipe material
L	feet	Length of pipe
PA	days	Pipe age
R		Ratio of pipe age/repair age



* In case available

Figure 4.4 Flow chart of pipe asset management with copula modeling

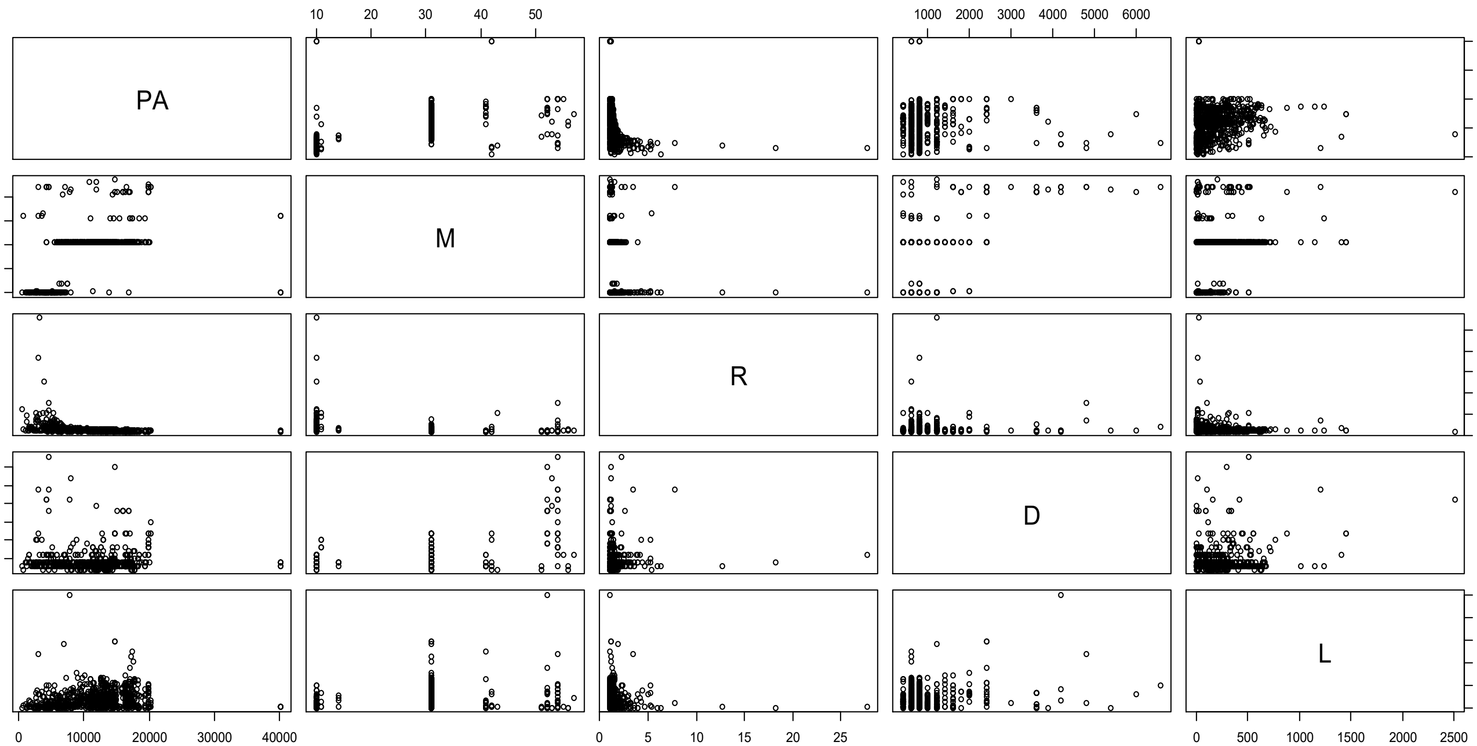


Figure 4.5   Scatter plot of pipe variables

The variable R is the ratio of pipe age to repair age. A higher value of R for a particular pipe indicates that, compared to a pipe with a lower value of R, repair was done at an earlier part of its lifespan. The Kendall's tau and Spearman's rho values for the pairs are shown in Table 4.4.

Table 4.4 Values of Kendall's tau and Spearman's rho of the five variables

	PA (Tau)	PA (Rho)	M (Tau)	M (Rho)	R (Tau)	R (Rho)	D (Tau)	D (Rho)	L (Tau)	L (Rho)
PA	1.00	1.00	0.44	0.53	-0.37	-0.51	-0.04	0.05	0.16	0.23
M	0.44	0.53	1.00	1.00	-0.28	-0.35	0.06	0.07	0.19	0.24
R	-0.37	-0.51	-0.28	-0.35	1.00	1.00	-0.04	0.06	-0.02	-0.03
D	-0.04	0.05	0.06	0.07	0.04	0.06	1.00	1.00	0.04	0.05
L	0.16	0.23	0.19	0.24	-0.02	-0.03	0.04	0.05	1.00	1.00

From the table, it is observed that pipe age has a strong positive dependence with type of pipe material and a somewhat positive dependence with pipe length. Cast iron and asbestos cement pipes were installed earlier than new material pipes such as steel or PVC pipes. So the strong positive correlation follows the notion that pipe age increases depending on the type of pipe material. Longer lengths of pipes are observed in transmission lines of larger diameter pipes, and the results show that pipe age tends to increase if the pipe length is higher. However, for this set of data there was no significant correlation between pipe age and pipe diameter. It was also observed that pipe age is negatively correlated with the ratio of pipe age to repair age; pipe age tends to be high for pipes where repair was done at a later part of their life spans. Comparing the Spearman's rho value to the Pearson correlation coefficient as shown in Table 4.4, is it apparent that for certain pairs of variables there is some difference between the two types of coefficients. For example, the Spearman's rho coefficient for pipe age (PA) and ratio R is -0.51, but the Pearson's coefficient is -0.32. Looking at the scatter plot for these two variables in Figure 4.5, it is apparent that there is a negative curvature and there are a few high value outliers for R. With those outliers removed, the Pearson's coefficient becomes -0.48. From this comparison, it is evident that using a ranking order correlation coefficient such as Spearman's rho helps to present a better dependence model than the Pearson's coefficient, where the variables are not in a straight line relationship. Another dependence number that stands out is the Pearson's coefficient for pipe diameter and pipe length; it shows a somewhat positive correlation of 0.19, whereas for Spearman's rho or Kendall's tau, the values are practically 0. That result indicates that there is no dependence between the two variables for this particular data set. The scatter plot shown in Figure 4.6 supports that notion.

Table 4.5 Pearson's correlation coefficient for five variables

	PA	M	R	D	L
PA	1.00	0.52	-0. 32	-0.01	0.15
M	0.52	1.00	-0.27	0.37	0.23
R	-0.32	-0.27	1.00	-0.32	-0.08
D	-0.01	0.37	0.07	1.00	0.19
L	0.15	0.23	-0.08	0.19	1.00

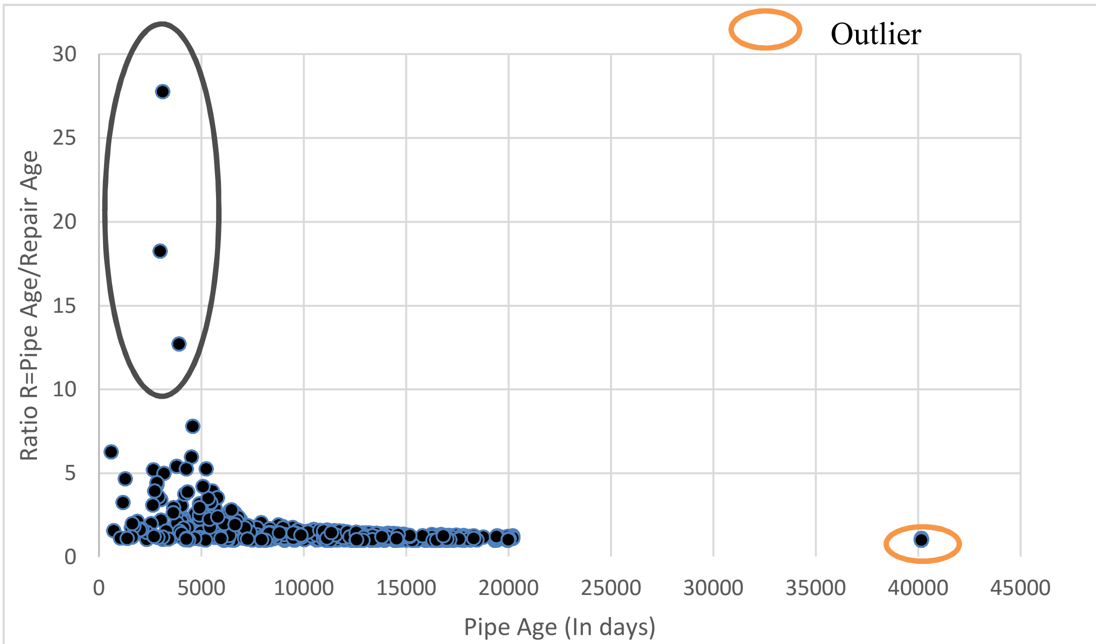


Figure 4.5 Scatter plot of ratio R and pipe age

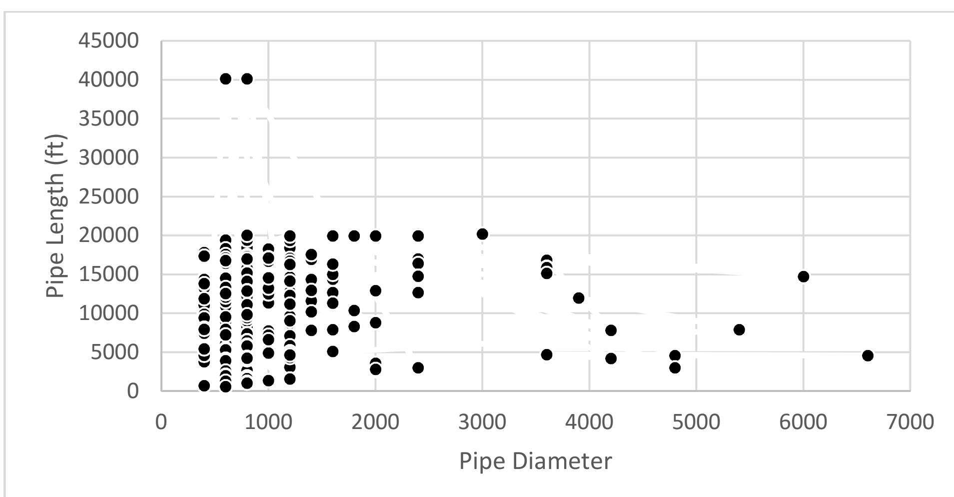


Figure 4.6 Scatter plot of pipe length and pipe diameter

4.7.2 **Choosing Copula Modeling for Pipe Data**

In this section, the whole process of copula modeling is demonstrated with the help of a set of pipe data. Continuing with the data set presented in the previous section, based on Table 4.4, pipe age and material variables were taken into consideration. For pipe material, it is assumed that lower numbered materials are of PVC material, which are relatively new pipe material compared to other pipe materials such as cast iron or asbestos cement pipes. The dependence between these two parameters showed a strong positive correlation with the Spearman's rho value of 0.53 and the Pearson's coefficient of 0.52. Figure 4.7 shows the k-plot, and Figure 4.8 shows the chi-plot for the data, both indicating that there is a good positive correlation between the variables. This follows the notion that pipes made of cast iron or steel, which are assigned a higher number, are of longer age in comparison to pipes made of PVC material. The list of pipe materials and their corresponding codes is given in Table 4.6.

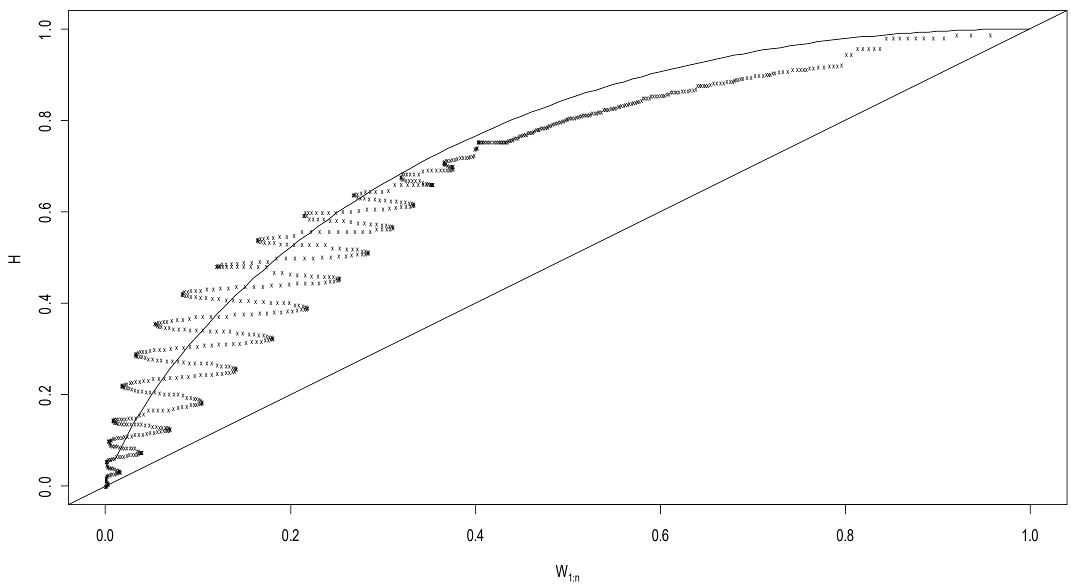


Figure 4.7 K-plot of the pipe data

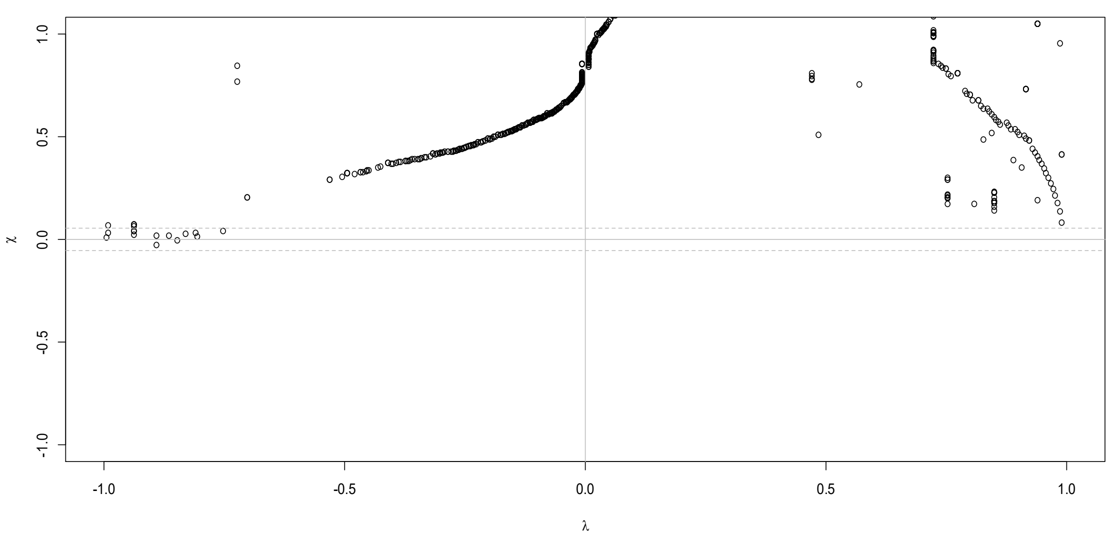


Figure 4.8 Chi-plot of the pipe data.

Table 4.6 Code number corresponding to pipe material

Code	Pipe Material
10	Polyvinyl chloride (C-900) 4"-12"
11	Polyvinyl chloride (C-905) 16"-36"
14	Polyvinyl chloride (general use)
31	Asbestos cement
41	Cast iron
43	Ductile iron pipe w/o baggie
51	Steel
52	Steel cylinder concrete - pretension
53	Steel cylinder concrete - prestressed
54	Steel mortar-lined mortar-coated
56	Steel concrete-mortar-lined mortar-coated

4.7.2.1 Marginal Distribution

The scatter plot of the two variables, pipe age and pipe material, is shown in Figure 4.10. The pipe age data was best fitted with a general extreme value distribution as shown in Figure 4.9. The parameters for the distribution were k=-0.37451, σ= 5126.9, and μ=10485, yielding a p-value of 0.0071. For pipe material, the distribution being a discrete distribution yielded a geometric distribution with parameter probability p=0.03327.

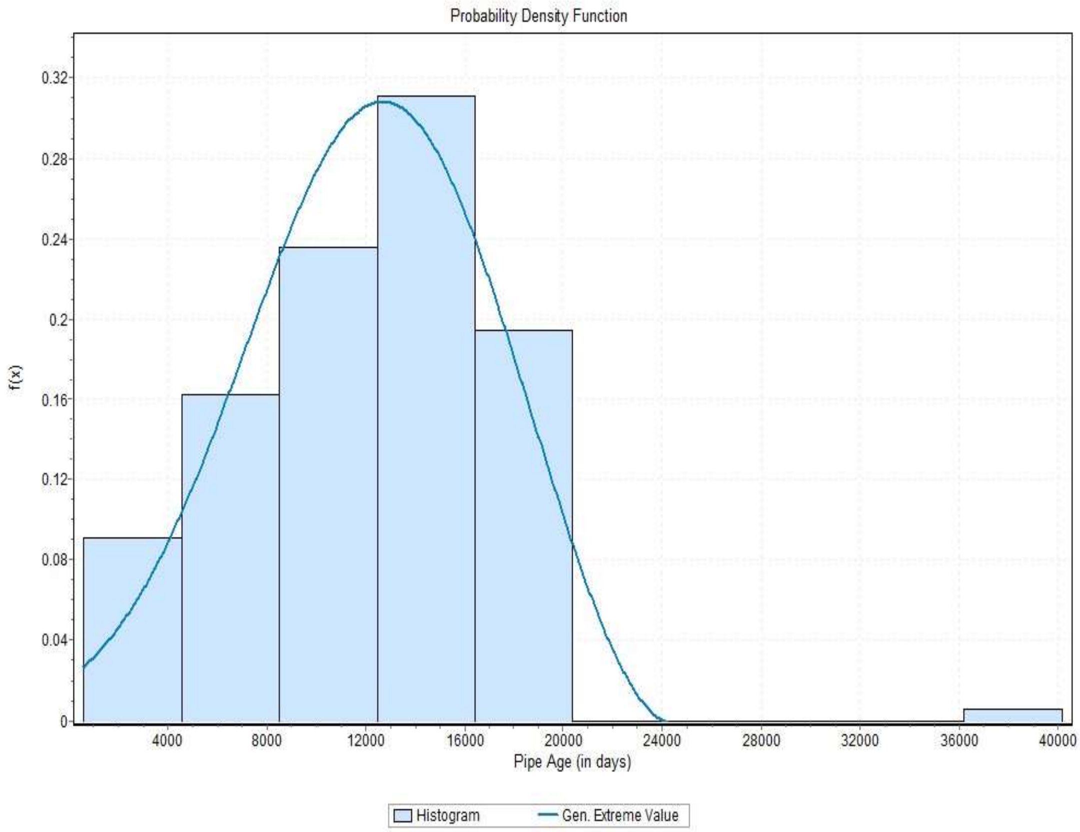


Figure 4.9 Marginal distribution of pipe age

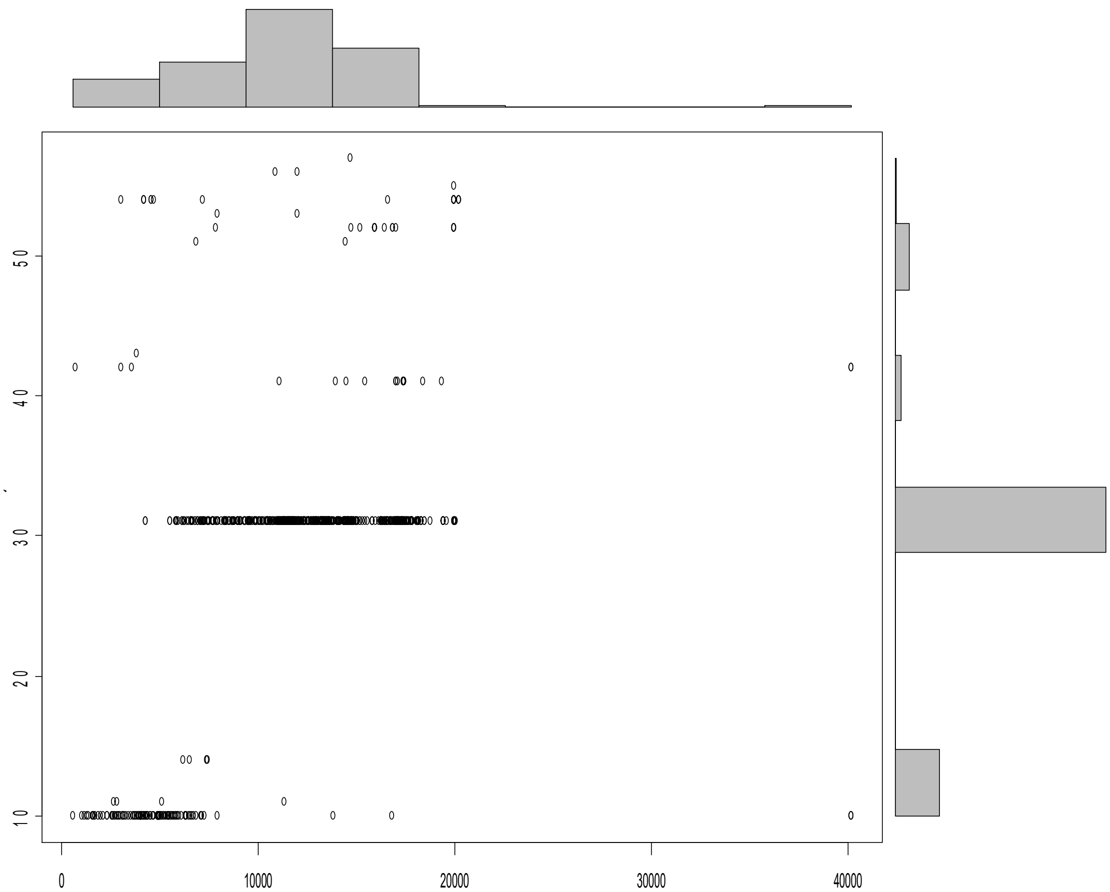


Figure 4.10 Scatter plot with histogram for pipe age and pipe material data

4.7.2.2 Parameter Estimation

To estimate the parameters of the copula family, three methods were applied, namely:

1) Inversion of tau
2) Inversion of rho
3) Maximum pseudolikelihood

For each copula family, these three methods of parameter estimation were applied. The results of the analysis can be found in Table 4.7. For the Clayton, Frank, and Gumbel copulas, the maximum pseudolikelihood method gave the copula parameter with least standard error. So for the Frank copula, the parameter chosen was 5.16. The parameter chosen for the Clayton copula was 1.47 and 1.6 for the Gumbel copula. These parameters along with the marginals for pipe age and pipe material were then used to generate 500 data points for the Frank, Clayton, and Gumbel copula families.

Table 4.7 Parameter estimation

Copula family	Method	Parameter	Standard error
Frank copula	Inversion of tau	4.67	0.48
	Inversion of rho	3.74	0.28
	Maximized pseudo likelihood	**5.16**	**0.24**
Clayton copula	Inversion of tau	1.55	0.20
	Inversion of rho	1.20	0.11
	Maximized pseudo likelihood	**1.47**	**0.07**
Gumbel copula	Inversion of tau	1.78	0.10
	Inversion of rho	1.60	0.06
	Maximized pseudo likelihood	**1.60**	**0.03**

4.7.2.3 Goodness of Fit of Copula Modeling

Copula models were checked by the goodness of fit graphical and formal test methods. In the following two sections, both of the processes are discussed.

4.7.2.3.1 Graphical Method of Goodness of Fit

Using the parameters obtained from maximized pseudolikelihood, 500 data points were generated using the Frank copula, Clayton copula, and Gumbel copula families. Graphs were created using the generated data points and the original points. Figure 4.11 shows the data points generated from the Frank copula family, Figure 4.12 shows data points generated from the Clayton copula family, and Figure 4.13 shows data points generated from the Gumbel copula family. The original pipe age and pipe material data are also shown in all three graphs. All three figures indicate that the generated points generally match the pattern of the original data.

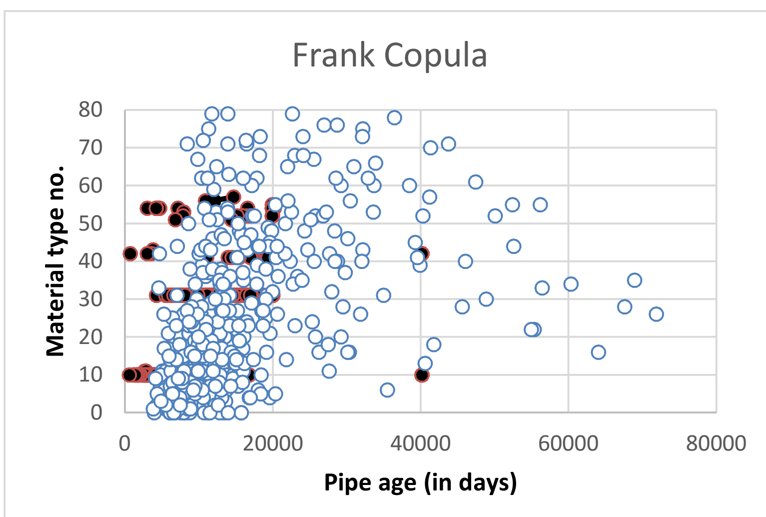


Figure 4.11 Data points from Frank copula modeling (hollow points) and original data points (solid points)

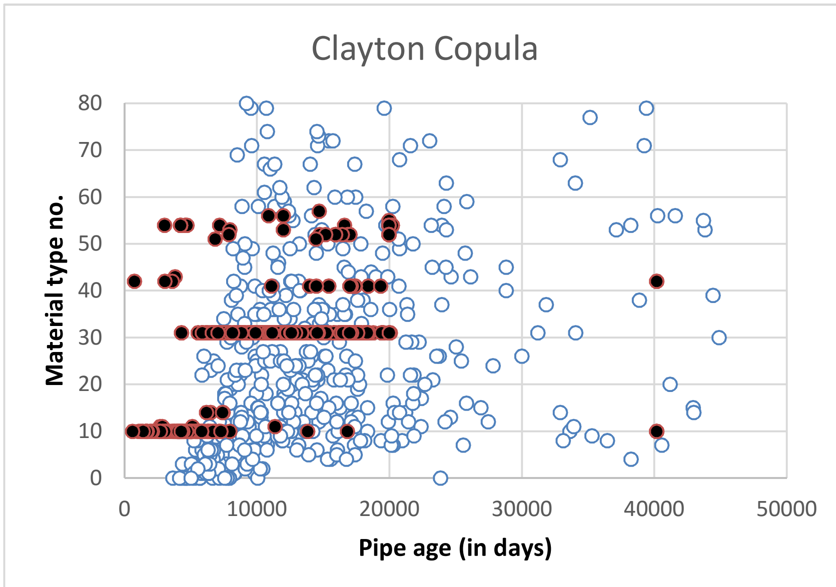


Figure 4.12 Data points from Clayton copula modeling (hollow points) and original data points (solid points)

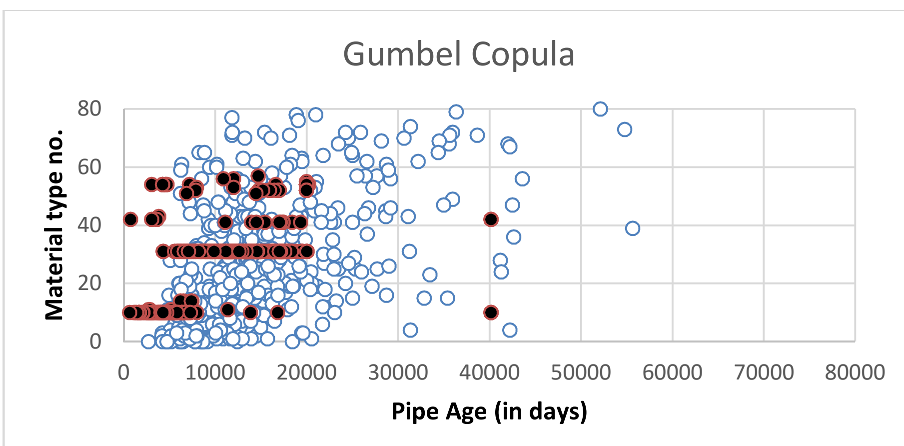


Figure 4.13 Data points from Gumbel copula modeling (hollow points) and original data points (solid points)

4.7.2.3.2 Formal Method of Goodness of Fit

Goodness of fit statistics, Cramer-von Mises functional S_n (Genest et al. 2009) was applied to the models, and the corresponding *p-values* are shown in Table 4.8. For a formal goodness-of-fit test, two goodness-of-fit statistics are used S_n and T_n. As discussed in Section 4.6, these statistics are based on the Kendall process and should be as small as possible (Genest et al. 2006). It can be observed empirically that the model for which the statistic is smallest generally has the largest *p value*. The lower value of statistic gives an indicator for choosing a family. The values were generated with N=1000 for bootstrapping. The results indicated that that at 5% significant level all three dependence families are statistically insignificant. This may indicate that a big enough N value was not chosen for bootstrapping methods for the models to give statistically significant results.

Table 4.8 Cramer-von Mises functional S_n and *p-values* for copula models.

Copula family	S_n	*p-value*
Frank	**20.99**	0.0005
Clayton	23.73	0.0005
Gumbel	24.49	0.0005

4.7.3 Using Copula Modeling to Develop Future Leakage Prediction Regression Models

The second data set (Data Set #2 in Appendix B) is the water pipe break data points obtained from the Washington Suburban Sanitary Commission (WSSC). The WSSC is among the largest water and wastewater utilities in the nation, with a network of nearly 5,600 miles of fresh water pipeline and over 400 miles of sewer pipeline. Its service area spans nearly 1,000 square miles in the Prince George and Montgomery counties in the Washington D.C. suburban area, and it serves 1.8 million residents.

The WSSC website has pipe data leak graphs relating leaks/damages per day to the temperature of water in the Potomac River. It is stated in their website that when the temperature drops below 40°F, there is a higher chance of pipe breakage. This study used the data for the month of January 2014, as the record indicates that temperature goes down below 40°F often during this month of the year for this specific region. The temperature range for the 31 days of the data set is from 37°F to 44°F.

The marginal for the breakage per day follows a geometric distribution, whereas the marginal for the water temperature follows a Weibull distribution, so for this pair in data set both the marginals are non-normal, and this serves as a good example to study the copula modeling for analysis.

The following parameters were used for analysis:

The coefficient of dependence based on different method yielded:

Pearson's coefficient =-0.625

Spearman's rho=-0.61

Kendall's tau=-0.41

Since the Frank copula can be used for negative dependency, this copula was used for analysis. The next step was to measure the parameter of the copula. This was done with the maximum pseudolikelihood estimator as discussed in the earlier part of this paper.

The method yielded a parameter of -4.32. The value was then checked for goodness of fit by graphical observation and a formal goodness of fit test. 500 random sample points were generated using a Weibull distribution marginal, a geometric distribution marginal, and Frank copula dependence. These data points were then plotted (Figure 4.14) with the original data, showing that the generated data points follow the original data well. The formal goodness of fit method described by Genest et al. (2009) was applied and gives a *p-value* of 0.39, which gives a statistically significant result.

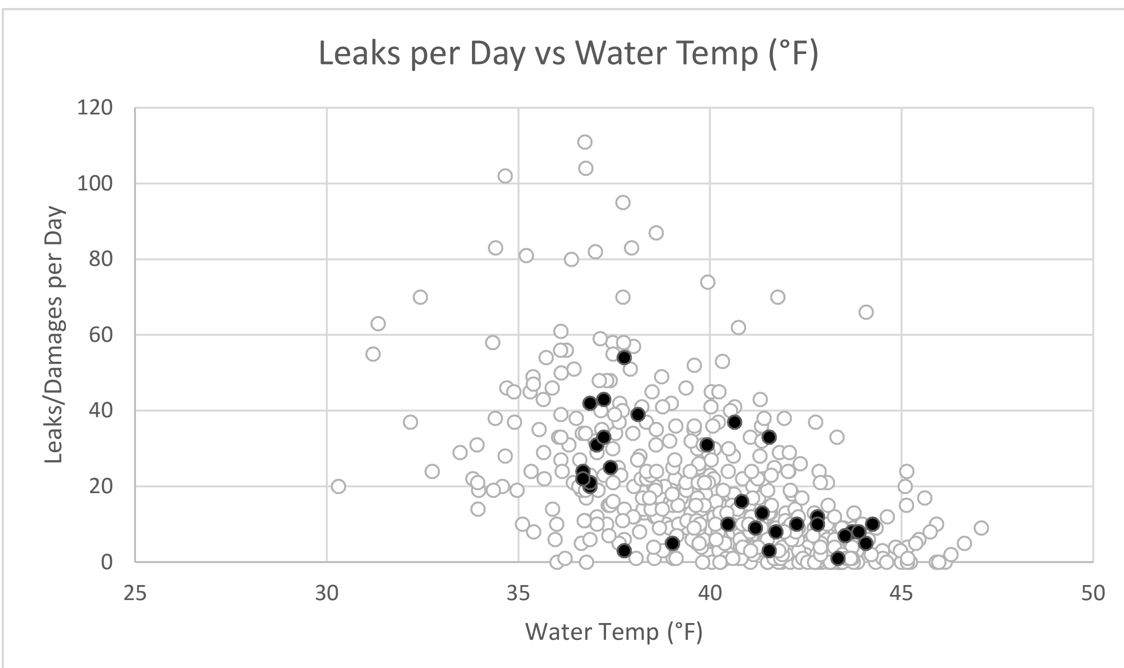


Figure 4.14 Observed values (black circles) and generated data using the Frank copula (hollow circles)

The next step, regression analysis, was done both on the original data and generated data points using copula modeling to generate prediction models for pipe breakage per day with the variation of water temperature. The regression equation for the original data set was Leaks/Damages per day=152.2-3.32*Water Temp in degrees Fahrenheit. However, in this case, the range for the lower 95% and the upper 95% difference was very wide: almost 102. Next, regression analysis was done on the points generated from the copula modeling. The regression equation for these data points was Leaks/Damages per day= 159.6-3.39*Water Temp in degrees Fahrenheit. The difference between the lower 95% and upper 95% was 39, a much lower value than the original regression data analysis. The statistical results for the two regression analyses are shown in Table 4.9 & Table 4.10. Comparing the two tables, it can be seen that for the regression model with copula-generated points, the standard error value is lower than that of the original data regression model. Also, the *p-values* are significantly low for both the intercept and the coefficient for water temperature, indicating that both the variables are statistically significant.

Table 4.9 Statistical results for regression analysis based on original data

	Coefficients	*Standard error*	*t stat*	*p-value*	*Lower 95%*	*Upper 95%*
Intercept	152.26	30.89	4.93	3.09E-05	89.09	215.44
Temp. of water (degrees Fahrenheit)	-3.31	0.77	-4.31	0.000167	-4.88	-1.75

Table 4.10 Statistical results for regression analysis based on copula-model-generated data

	Coefficients	Standard error	t stat	p-value	Lower 95%	Upper 95%
Intercept	159.65	9.93	16.07	4.06E-47	140.13	179.17
Temp. of water (degrees Fahrenheit)	-3.53	0.25	-14.31	3.81E-39	-4.01	-3.04

4.8 **Remarks**

The use of copulas is an emerging method of modeling and is useful when the marginals belong to different families of distributions. Copula modeling is also useful for generating large numbers of data points when it is difficult to obtain a data set, as is the case for pipe condition assessment. Underground pipe data information is expensive to obtain, and data sets may have random variables belonging to non-Gaussian family distributions. The large data sets generated can then be used for evaluating the current pipe condition models and the appropriateness of those models for determining the condition or remaining life of a pipe.

It was observed in the data set for leaks per day and temperature that the marginals were not following a normal distribution; the marginal for leaks per day was a geometric distribution, whereas that of the temperature data was a Weibull distribution. Using these two types of marginals and the Frank copula, it was possible to derive a large number of data points matching the original data distribution. The generated points were then used for generating regression analysis. It was shown that having a greater number of data points enabled reduction of the standard error value and significantly reduced the 95% confidence level difference compared to the original data set.

Identifying attributes associated with pipes helps to identify those factors that appear most susceptible to failure prediction. Through model variable correlation analysis, attributes are chosen to develop pipe deterioration models, which help to determine rehabilitation and pipe replacement strategies (Yan et al. 2013). In that way, obtaining variable dependency modeling helps utility companies in the process of developing asset management plans. These condition assessments help to develop the necessary steps and procedures to maintain water distribution systems running at desired levels of service in a cost-effective way. In the desert west region pipe data set, it was observed that among the variables there is a strong dependency between pipe age and pipe material and ratio of pipe age/repair age, whereas not much dependency was observed between pipe age and pipe size. The rank method of dependency Spearman's rho gave a better indication of the correlation if the data set had skewness and was non-normal. An example of two variables, pipe age and ratio of pipe age/repair age, was used to showcase how Spearman's rho gives a better indication of dependency than Pearson's coefficient for the two variables, because it considered the skewness and considered the whole distribution for measure for dependence

Chapter 5

VINE COPULA MODELING

5.1 Introduction

Copulas can be used in multivariate dependence modeling. Standard multivariate copulas have the following problems (Kramer and Schepsmeier 2011):

They can be inflexible in high dimensions.

They do not allow different dependency structures between pairs of variables.

The vine copula, as explained in detail by Aas et al. (2009), is a flexible graphical model for describing multivariate copulas built up using a cascade of bivariate copulas, where each pair-copula can be chosen independently from the others. Such bivariate copula construction decomposes a multivariate probability density into bivariate copulas, allowing each pair-copulas to be dealt with separately (Brechmann and Schepsmeier 2013).

5.2 Pair Copula Decomposition

A copula is a multivariate distribution, C, with a uniformly distributed marginal U (0, 1) on [0,1]. According to Sklar's theorem, multivariate distribution F with marginals $F_1(x_1)$ and $F_2(x_2)$

$$F(x_{1,}x_2) = C(F_1(x_1), F_2(x_2)) \qquad \text{Eq. 5.1}$$

The copula density for 2 dimensional distributions is

$$c_{12}(u_1, u_2) = \frac{\partial^2 C_{12}(u_{1,}u_2)}{\partial u_1 \partial u_2} \qquad \text{Eq. 5.2}$$

This implies joint density as

$$f(x_1, x_2) = c_{12}\big(F_1(x_1), F_2(x_2)\big).\, f_1(x_1).\, f_2(x_2) \qquad \text{Eq. 5.3}$$

The conditional density is given by the following equation:

$$f(x_2\,|x_1) = c_{12}\big(F_1(x_1), F_2(x_2)\big).\, f_2(x_2) \qquad \text{Eq. 5.4}$$

For d=3 dimensions, a possible decomposition is

$$f(x_3|x_1, x_2\,) = \frac{f(x_2, x_3|x_1)}{f(x_2|x_1)} \qquad \text{Eq. 5.5}$$

$$f(x_3|x_1, x_2\,) = \frac{c_{23|1}\big(F(x_2|x_1)F(x_3|x_1)f(x_2|x_1)\big)f(x_3|x_1)}{f(x_2|x_1)} \qquad \text{Eq. 5.6}$$

$$f(x_3|x_1, x_2\,) = c_{23|1}\big(F(x_2|x_1), F(x_3|x_1)\big)\, f(x_3|x_1) \qquad \text{Eq. 5.7}$$

$$f(x_3|x_1, x_2\,) = c_{23|1}\big(F(x_2|x_1), F(x_3|x_1)\big)c_{13}\big(F_1(x_1), F_3(x_3)\big).\, f_3(x_3) \qquad \text{Eq. 5.8}$$

The general equation can be expressed as follows:

$$f(x|v) = cxvj|v_{-j}(F\big(x\big|v_{-j}\big), F\big(vj\big|v_{-j}\big).\, f(x|v_{-j}) \qquad \text{Eq. 5.9}$$

for d dimensional vector v, and v_{-j} denotes the vector v, excluding the j^{th} component. So under appropriate regularity conditions, a multivariate density can be expressed as a product of pair-copulas, acting on several different conditional probability distributions.

For a high dimensional distribution, Bedford and Cooke (2001) proposed the use of vines for a large number of pair-copula constructions. The two types of vines suggested are a) C-canonical vines and b) D-vines. The n-dimensional density $(f(x_1, \ldots., x_n))$ corresponding to C-vines can be written as follows:

$$\prod_{k=1}^{n} f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,j+i|1\ldots\ldots,j-1} \left(F\left(x_j \middle| x_1, \ldots. x_{j-1}\right), F\left(x_{j+i} \middle| x_{i,} \ldots.. x_{j-1}\right)\right) \quad \text{Eq. 5.10}$$

The D-vine can be expressed as

$$\prod_{k=1}^{n} f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,,i+j-1} \left(F\left(x_i \middle| x_{i+1}, . x_{i+j-1}\right), F\left(x_{i+j} \middle| x_{i+1,} .. x_{i+j-1}\right)\right. \quad \text{Eq. 5.11}$$

The following examples illustrate vine copulas (Kramer and Schepsmeier 2011):

C vines: each tree has a unique node that connects all other nodes

$$f_{1234} = f_1 . f_2 . f_3 . f_4 . c_{12} . c_{13} . c_{14} . c_{23|1} . c_{24|1} . c_{34|12} \qquad \text{Eq. 5.12}$$

where

$$f_1 . f_2 . f_3 . f_4 = nodes\ in\ T_1$$

$$c_{12} . c_{13} . c_{14} = edges\ in\ T_1, nodes\ in\ T_2$$

$$c_{23|1} . c_{24|1} = edges\ in\ T_2, nodes\ in\ T_3$$

$$c_{34|12} = edge\ in\ T_3$$

D-vines: where each tree is a path

$$f_{1234} = f_1 . f_2 . f_3 . f_4 . c_{12} . c_{23} . c_{34} . c_{13|2} . c_{24|3} . c_{14|23} \qquad \text{Eq. 5.13}$$

$$f_1 . f_2 . f_3 . f_4 = nodes\ in\ T_1$$

$$c_{12} . c_{23} . c_{34} = edges\ in\ T_1, nodes\ in\ T_2$$

$$c_{13|2} . c_{24|3} = edges\ in\ T_2, nodes\ in\ T_3$$

$$c_{14|23} = edge\ in\ T_3$$

Figure 5.1 and Figure 5.2 show a four-variable C-vine and a four-variable D-vine, respectively.

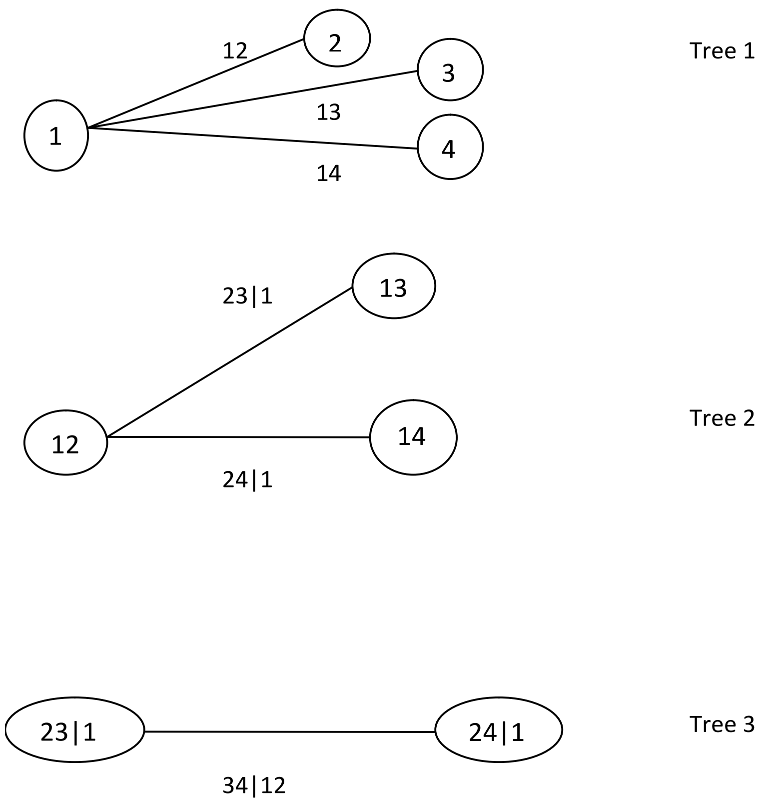


Figure 5.1   C-Vine for four variables

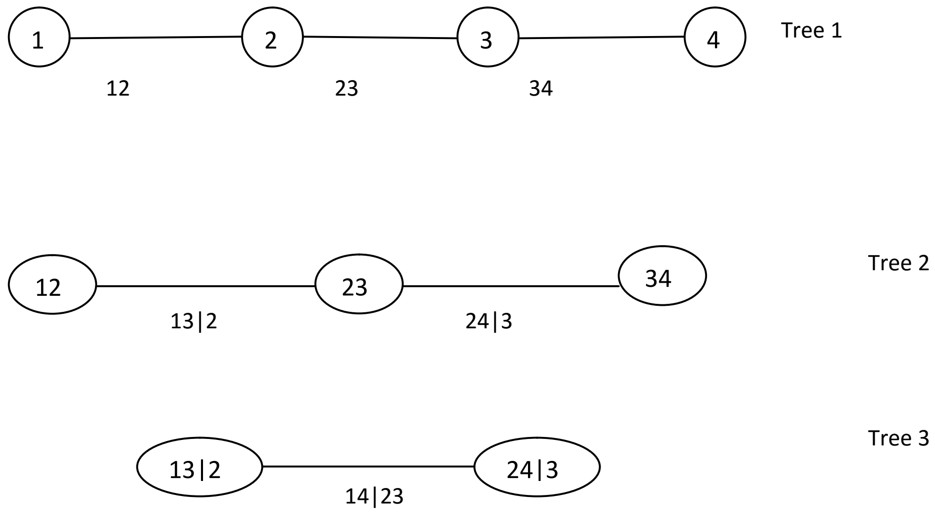


Figure 5.2 D-vine for four variables

Aas et al. (2009) stated that fitting a canonical vine might be advantageous when a particular variable is known to be a key variable that governs the interactions in the data set. In such a case, the variable can be placed at the root of the canonical vine, as variable 1 is placed in Figure 5.1. The notation of D-vines resembles independence graphs more than those of canonical vines.

Conditional independence can reduce the number of levels of the pair-copula decomposition and thereby simplify the construction. For a three-dimensional case, the general expression for both C-vines and D-vines is

$$f(x_1, x_2, x_3) = f_1(x_1).f_2(x_2).f_3(x_3).c_{12}\{F_1(x_1), F_2(x_2)\}.c_{23}\{F_2(x_2), F_3(x_3)\}.c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\}$$

Eq. 5.14

By assuming conditional independence, it reduces the number of levels of the pair-copula decomposition and thereby simplifies the construction. In Eq. 5.14, if X_1 and X_3 are independent given X_2, this results in $c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\}$= 1; thereby the pair-copula decomposition becomes

$f(x_1, x_2, x_3) = f_1(x_1). f_2(x_2). f_3(x_3). c_{12}\{F_1(x_1), F_2(x_2)\}. c_{23}\{F_2(x_2), F_3(x_3)\}$ Eq. 5.15

5.3 **Vine Copula Analysis**

Brechmann and Schepsmeier (2013) have described the following steps shown in Figure 5.3 for the vine copula method. Fitting a vine copula model involves four steps: first, an appropriate vine tree structure has to be identified. Such a structure may either be given by the data itself or has to be selected manually or through expert knowledge. For a given vine structure, copulas have to be selected. In the next step, the copulas have to be estimated. Finally, models are to be evaluated and compared to alternatives.

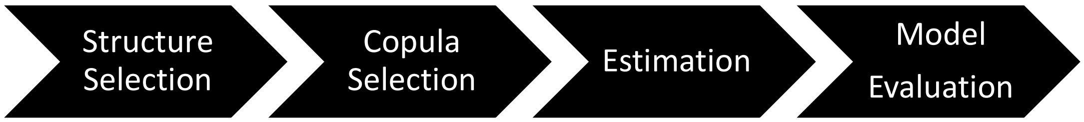


Figure 5.3 Vine copula process

The order of the variables had to be selected when specifying C- and D-vine copulas. For the D-vine, the order of the variables in the first tree had to be chosen, and for the C-vine, the root nodes for each tree needed to be determined. In a C-vine, the entries of this vector correspond to the following pairs and associated pair-copula terms:

(1, 2), (1, 3), (1, 4) ,….., (1, d), (Tree 1)

(2, 3|1, (2, 4|1) , ….., (2, d|1), (Tree 2)

(3, 4|1, 2), (3, 5|1, 2) , ….. , (3, d|1, 2), (Tree 3)

………

(d -1, d|1, … , d - 2) (Tree d-1)

Similarly, the pairs of a D-vine are specified in the following order:

(1, 2), (2, 3), (3, 4), … , (d – 1, d), (Tree 1)

(1, 3|2), (2, 4|3),……, (d – 2, d|d - 1), (Tree 2)

(1, 4|2, 3), (2, 5|3, 4), … , (d – 3, d|d – 2, d - 1) (Tree 3)

………

(1, d|2, …, d - 1) (Tree d-1)

Having decided the structure of the C- or D-vine to be used, pair-copula families for each (conditional) pair of variables were selected. Parameter estimation was done by inversion of the Kendall's tau method and the maximum likelihood estimation (MLE) method. Having fitted different vine copula models to a given data set, the best model was determined in terms of one or more criteria. The criteria used were the classical Akaike information criterion (AIC), Bayesian information criterion (BIC) and the Vuong and the Clarke tests.

The five variables used for D-vine copula generation were V1= PA=Pipe age, V2= M=Pipe material type (different code numbers used for different materials), V3= R=Pipe age/repair age, V4=D=Diameter of pipe, V5=L=Length of the pipe. Figure 5.5 and Figure 5.6 show C-Vine and D-Vine models with family and empirical tau values in each tree. They show the 4 trees for 5 variables, where F=Frank Copula, t=student's t copula, C=Clayton copula, G=Gaussian copula, and the empirical tau value is shown on the links with the copula family. The AIC value for the C-vine model was -2173, whereas for the D-vine model it was -2270. Since the D-vine model had the least value of AIC, the D-Vine model was used to generate 1000 values, which are shown in Figure 5.7. By vine copula modeling it was possible to generate multivariate pipe data.

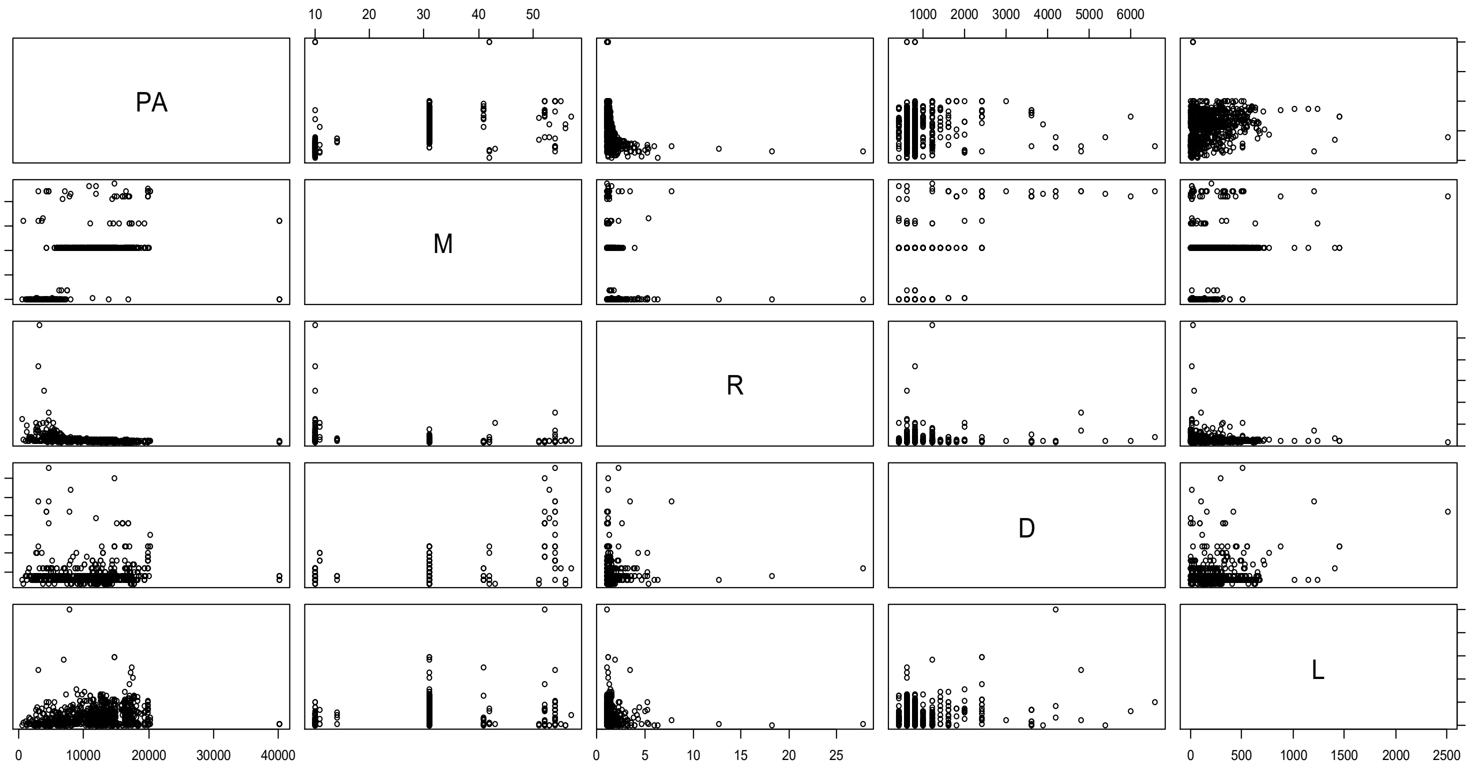


Figure 5.4 Scatter plot of pipe variables

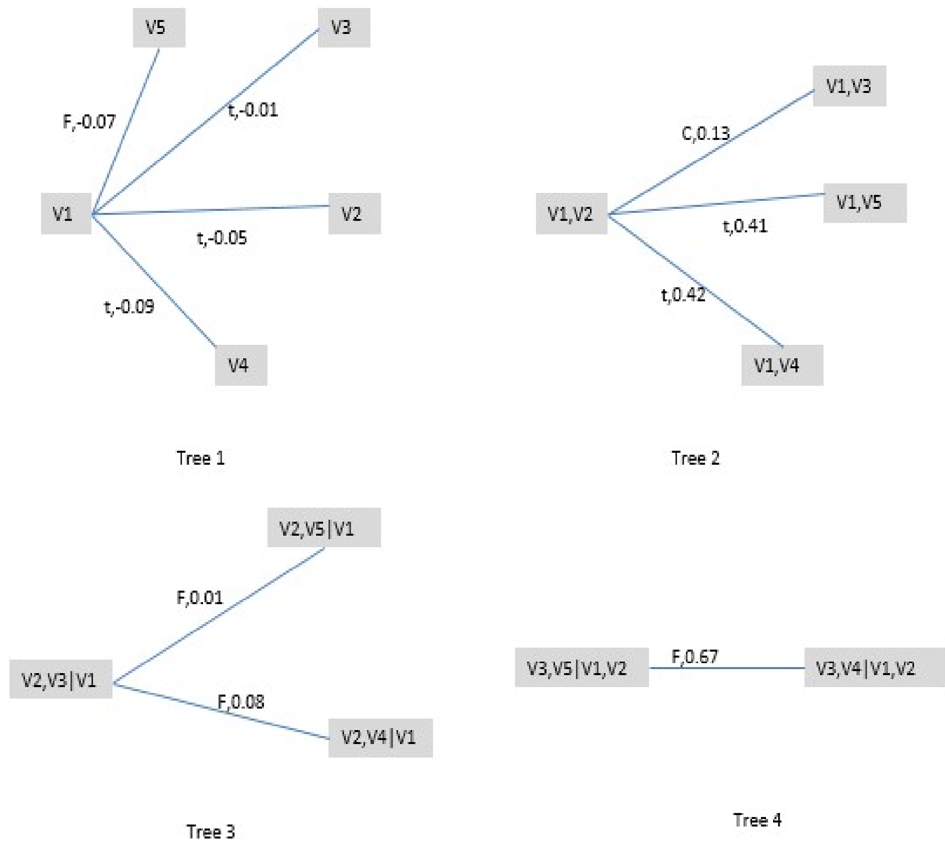


Figure 5.5 C-Vine of five pipe variables

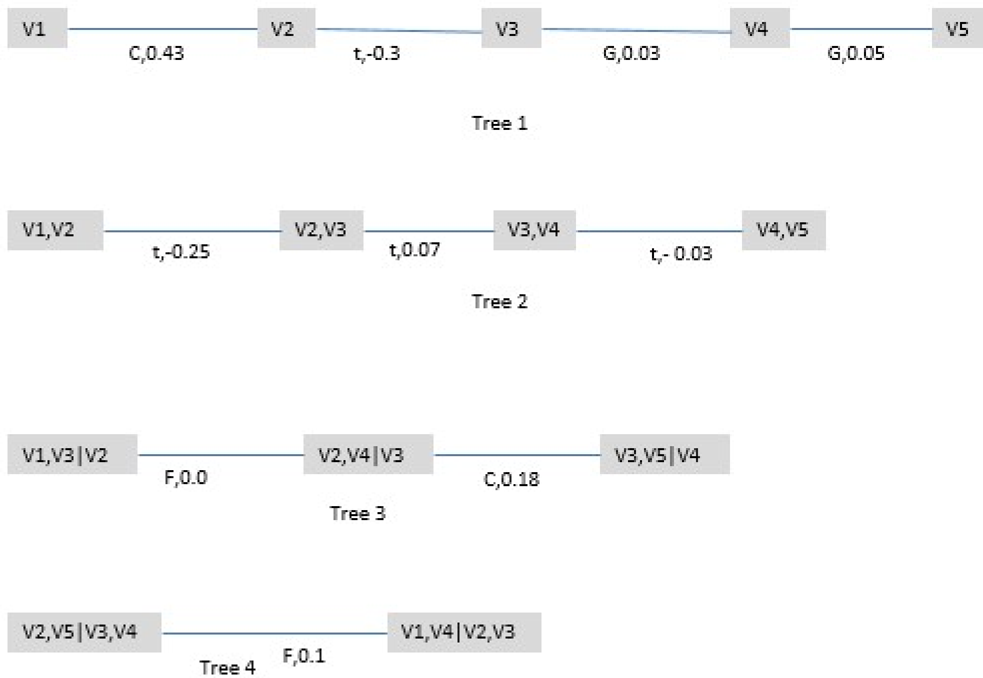


Figure 5.6 D-Vine of Five Pipe Variables

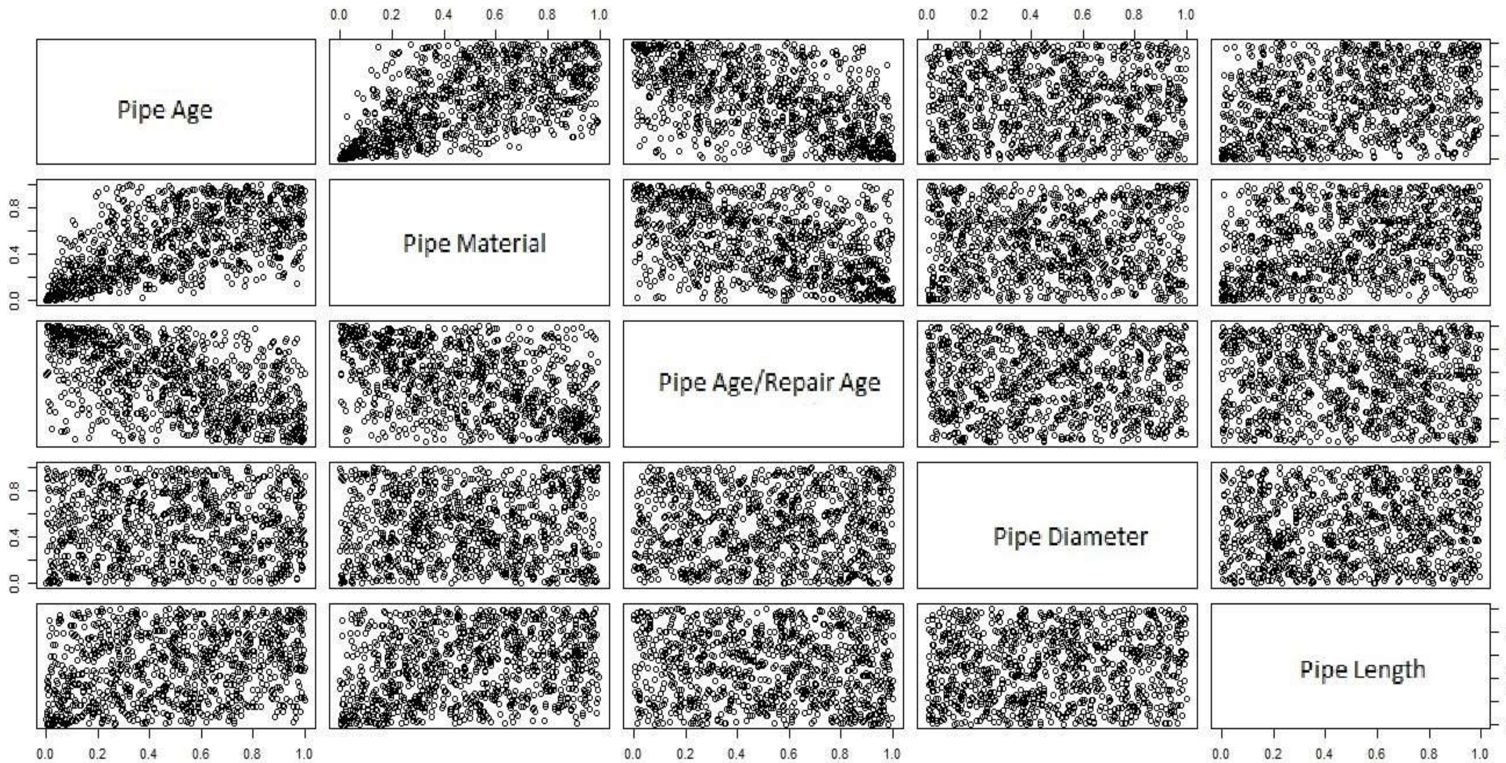


Figure 5.7 Generated data points from D-vine copula

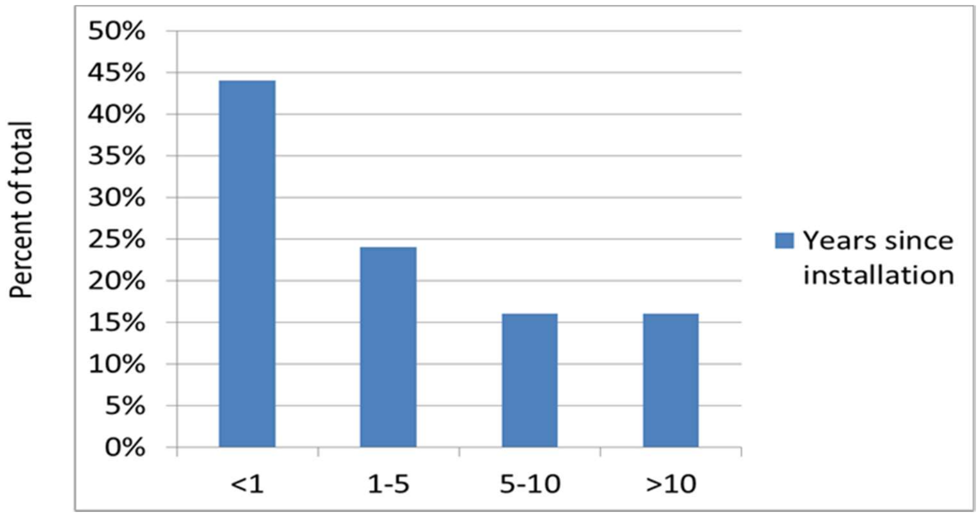


Figure 5.8 PVC failures as a function of time (adapted from Folkman 2012)

5.4 **Remarks**

When D-Vine copula modeling is applied to the pipe data set, the results indicate that pipe age and the ratio of pipe age to repair age are conditionally dependent given the pipe material. The result also helps to establish findings from other studies that, due to installation errors, repair work is needed in the early stages of PVC pipes after installation.

Chapter 6

BAYESIAN INFERENCE FOR COPULA PARAMETER ESTIMATION

6.1 Bayesian Inference

Bayesian statistical inference concerns unknown parameters that describe certain population characteristics, such as the true mean efficacy of a particular treatment. Inferences are made using data and a statistical model that links the data to the parameters. In frequentist statistics, parameters are fixed quantities, whereas in Bayesian statistics, the true value of a parameter can be thought of as being a random variable which is assigned a probability distribution, known as prior information. Bayesian models are suitable for complex cases of analysis. Bayesian inference can be performed on models as well as model parameters.

6.2 Bayes' Theorem

Bayes' theorem is based on the conditional probability as denoted by

$$p(y|x) = p(y,x)/p(x) \qquad \text{Eq. 6.1}$$

In words, the definition implies that the probability of y given x is the probability that they happen together relative to the probability that x happens at all. Multiplying both sides of Eq. 6.1 by p(x) results in $p(y|x)p(x) = p(y,x)$. Similarly, starting with $p(x|y) = p(y,x)/p(y)$ results in $p(x|y)p(y) = p(y,x)$. With two similar expressions for $p(y,x)$, one finds $p(y|x)p(x) = p(x|y)p(y)$, and dividing both sides by $p(x)$ results in

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \qquad \text{Eq. 6.2}$$

p(x) can be expanded as $p(x) = \sum_y p(x, y) = \sum_y p(x|y)p(y)$; substituting this into Eq. 6.2 yields

$$p(y|x) = \frac{p(x|y)p(y)}{\sum_y p(x|y)p(y)} \qquad \text{Eq. 6.3}$$

Eq. 6.2 and Eq. 6.3 are called Bayes' rule, and this simple relationship lies at the core of Bayesian inference. For a continuous-variable, Eq. 6.3 becomes

$$p(y|x) = \frac{p(x|y)p(y)}{\int dy\, p(x|y)p(y)} \qquad \text{Eq. 6.4}$$

Bayes' rule is applied to models and data. It can be used to know how strongly we should believe the model, given the data:

p (parameter values and model structure | data values)

For a model with data value D and parameter value θ, Bayes' rule can be expressed as

$$p\ (\theta|D) = p(D|\theta) \quad x \quad p(\theta) \ \ /\ p(D) \qquad \text{Eq. 6.5}$$

$$\boxed{\text{Posterior}} = \boxed{\text{likelihood}} \text{ x } \boxed{\text{prior}} \ /\ \boxed{\text{evidence}}$$

where

$$p(D) = \int d\theta\, p(D|\theta)p(\theta) \qquad \text{Eq. 6.6}$$

The prior $p(\theta)$ is the strength of the belief in θ without the data D. The posterior $p\ (\theta|D)$ is the strength of belief in θ when the data D have been taken into account. The likelihood $p(D|\theta)$ is the probability that the data could be generated with parameter value θ.

When there is observed data, Bayes' rule can be used to determine beliefs across competing parameter values in a model and to determine beliefs across competing models. The evidence $p(D)$ is the probability of the data according to the model, determined by summing across all possible parameter values weighted by the strength of the belief in those parameter values.

6.3 **Goals of Bayesian Inference**

The three goals of Bayesian inference are estimation of parameter values, prediction of data values, and model comparison (Kruschke 2011).Estimation of parameter values means determining the extent to which one can believe in each possible parameter value. From Eq. 6.5, the posterior distribution over the parameter values θ gives the estimate of those values. If the posterior distribution is narrow, with most of the probability within a small range of θ, then that predicts a fair amount of certainty about the possible values of θ.

Predicting the future data value can be obtained by averaging the predicted data probabilities across all possible parameter values, weighted by the belief in the parameter values:

$$p(y) = \int d\theta p(y|\theta)p(\theta) \qquad \text{Eq. 6.7}$$

Bayes' rule can also be used for model comparison. Let there be two models named M1 and M2. By Bayes' rule $p(M1|D) = p(D|M1)p(M1)/p(D)$ and $p(M2|D) = p(D|M2)p(M2)/p(D)$ where $p(D) = \sum_i p(D|M_i)p(M_i)$. The ratio of these two equations results in

$$\frac{p(M1|D)}{P(M2|D)} = \frac{p(D|M1)}{P(D|M2)} \frac{p(M1)}{p(M2)} \qquad \text{Eq. 6.8}$$

Eq. 6.8 reveals that the ratio of the posterior beliefs is the ratio of the evidence times the ratio of the prior beliefs.

6.4 Difficulties of Bayesian Inference

In achieving goals for Bayesian inference, one faces the difficulty of computing a difficult integral. The traditional way is to use likelihood functions with "conjugate" prior functions.

In Bayesian theory, if the posterior distribution $p(\theta|x)$ is in the same family as the prior distribution $p(\theta)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function. For example, the Gaussian family is conjugate to itself with respect to a Gaussian likelihood function; that is, if the likelihood function is Gaussian, choosing a Gaussian prior over the mean will ensure that the posterior distribution is also Gaussian. Table 6.1 shows some conjugate prior distributions.

Table 6.1 Conjugate prior distributions (Swiler 2006)

Sampling Distribution	Conjugate Prior Distribution
Binomial	Success probability is beta
Negative binomial	Success probability is beta
Poisson	Mean is gamma
Exponential with mean ($1/\lambda$)	λ is gamma

Normal with known variance and unknown mean	Mean is normal
Normal with unknown variance but known mean	Variance is an inverted gamma

In some cases, no reliable prior information about θ may exist; in such a case, a non-informative prior distribution can be used that contains no information about θ. When this happens, inferences made from the posterior distribution are regarded as objective rather than subjective.

6.5 Markov Chain Monte Carlo Method

The denominator function may not be analytically available. It may involve complex integration. For low-dimensional cases, there are specific methods to approximate Bayesian integrals. To calculate the posterior distribution of higher dimensions, the Monte Carlo method is often used to generate samples over which the integrand is calculated. Markov Chain Monte Carlo (MCMC) methods are simulation techniques through which posterior distributions can be obtained accurately by specifying the prior and likelihood distributions. MCMC is an iterative process that is based on the construction of a Markov chain that eventually "converges" to a stationary, posterior distribution. Unlike direct simulation methods, the MCMC output is a dependent sample generated from a Markov chain (Ntzoufras 2009). It is a stochastic process $\{\theta^{(1)},\ \theta^{(2)}, \dots \dots, \theta^{(t)}\ \}$ such that

$$f\left(\theta^{(t+1)} \middle| \theta^{(t)}, \dots \dots .., \theta^{(1)}\right) = f(\ \theta^{(t+1)}|\theta^{(t)}) \qquad \text{Eq. 6.9}$$

This implies that the distribution of θ at sequence $t+1$, given all the preceding θ values, depends only on the value $\theta^{(t)}$ on the previous sequence t. In the MCMC method, the purpose is to stimulate realizations from a Markov chain which has a stationary distribution, such as $f(U)$. Given a vector random variable $U = (U_1, \dots \dots U_k)$ with a joint distribution $(U_1, \dots \dots U_k)$, the expected value of some intractable function $h(U)$ can be approximated by obtaining independent random draws $U^{(t)}$, t=1,…,n from the distribution $f(U)$. The desired expectation can be approximated by

$$E(h(U) \approx \frac{1}{n} \sum_{t=1}^{n} h\left(U^{(t)}\right) \text{ as } n \rightarrow \infty \qquad \text{Eq. 6.10}$$

In Bayesian statistics, there are generally two MCMC algorithms that are used: the Gibbs sampler and the Metropolis-Hastings algorithm.

6.5.1 **Gibbs Sampling**

Gibbs sampling is a method used in the MCMC algorithm. It is a technique where random variables can be generated indirectly from a marginal distribution, sampling each variable from a conditional distribution where all other variables are considered known and random numbers can be easily simulated using standard functions in statistical and computing software. The Gibbs sampling process is described as follows (Hong and Prozzi 2006):

For a set of random variables $U_1, U_2, \dots \dots, U_3$, the joint distribution is denoted as $f(U_1, U_2, \dots \dots, U_m)$. For given arbitrary starting values of U_s's, say $U_1^{(0)}, U_2^{(0)}, \dots \dots, U_m^{(0)}$, the first iteration of random draws of U_s's obtained is

$U_1^{(1)}$ from $f(U_1 | U_2^{(0)}, U_3^{(0)}, \dots \dots \dots, U_m^{(0)})$

$U_2^{(1)}$ from $f(U_2|U_1^{(1)}, U_3^{(0)}, \dots \dots .., U_m^{(0)})$

$U_m^{(1)}$ from $f(U_m|U_1^{(1)}, U_2^{(0)}, \dots \dots .., U_{m-1}^{(1)})$

In a similar way, the second set of random draws of U_s 's is obtained through the update process. After r iterations, the series of U_s 's is obtained as $\left(U_1^{(r)}, U_2^{(r)}, \dots .., U_k^{(r)}\right)$. It is shown under mild conditions for each variable $U_s^{(r)} \rightarrow U_s^{(r)} \sim f(U_s)$ as $r \rightarrow \infty$ (Geman and Geman 1984). After enough iterations r, $U_s^{(r)}$, can be regarded as a random draw from the distribution $f(U_s)$.

Since Gibbs sampling is a convenient simulation technique, this research employed Gibbs sampling using WinBUGS software for obtaining parameters from the posterior distribution.

6.6 Applying Bayesian Inference Parameter Estimation Method on Pipe Data

When the data are continuous, the likelihood of an independent observation $y = \{y_1, \dots .. y_n\}$, each distributed as the equation $F(y_1, \dots \dots .., y_m) = C(F_1(y_1), \dots \dots, F_m(y_m))$
is $f(y|\Theta, \varphi) = \prod_{i=1}^{n} f(y_i|\Theta, \varphi)$, where $y_i = (y_{i1}, \dots \dots, y_{im})'$

and $f(y_i|\Theta, \varphi) = c(u_i;\ \phi) \prod_{j=1}^{m} f_i(y_{ij}; \theta_j)$

Where,$u_i = (u_{i1}, \dots \dots .., u_{im})'$ and $u_{ij} = F_j\left(y_{ij;}\theta_j\right), \Theta = \{_{\theta_1}, \dots \dots \theta_m\}$ are parameters of the marginal models and $f_j\left(y_{ij}; \theta_j\right) = \frac{\partial}{\partial_{yij}} F_j\left(y_{ij}; \theta_j\right)$ is the marginal density of $y_{ij;}$.

Most parametric copula functions have analytical expressions for the densities $c(u; \emptyset)$. Maximum likelihood estimation is often straightforward because of this. However there are cases where a Bayesian analysis can be preferable, according to Smith (2011):

- For more complex marginal models and/or copula functions, the likelihood can be hard to maximize directly. One of the solutions mentioned in the paper is to use a two stage estimator, where the marginal model parameters θ_j are estimated first and then the $\emptyset$ estimated conditional is on these. Another Bayesian alternative in this circumstance is to construct an inference from the joint posterior $f(\Theta, \emptyset|y)$ evaluated in a Monte Carlo manner, with Θ and $\emptyset$ generated separately in a Gibbs sampling scheme.
- Bayesian hierarchical modeling has been proven successful for the modeling of multivariate data.
- When estimating a copula model, the objective is often to construct inference on measures of dependence, quantiles, and/or functionals of random variable vector Y or parameters $(\Theta, \emptyset)$. The evaluation of the posterior distribution of these quantities is often easy using MCMC methods.

Data Set #1 in Appendix A was used for this part of analysis. Pipe age and repair age were the two variables chosen for analysis using the copula method and applying Bayesian inference for parameter analysis. The pipe age and repair age were converted to years instead of days for simplicity. The refined data set gave a correlation value of Spearman's rho 0.936 for pipe age and repair age. Figure 6.1

shows the histogram for pipe age and shows that the general extreme value type distribution fits the data well.

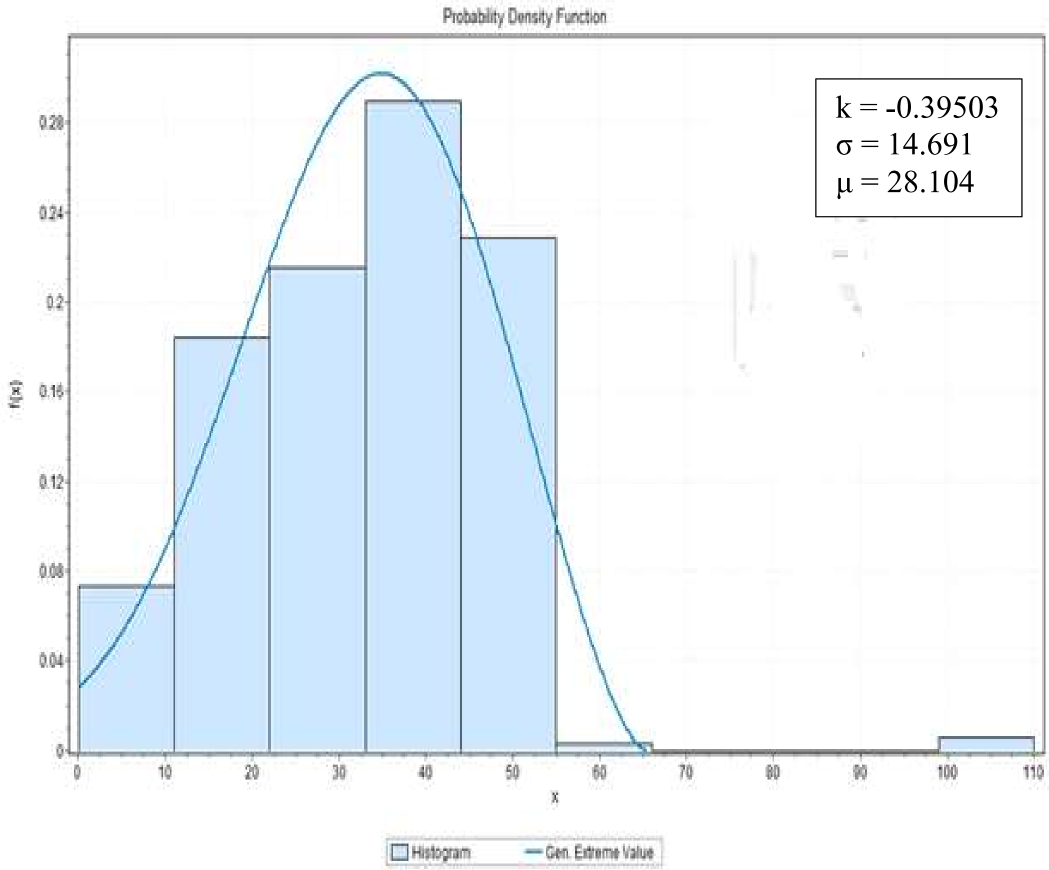


Figure 6.1 Pipe age data distribution

Figure 6.2 shows the general extreme value distribution fitting repair age data well.

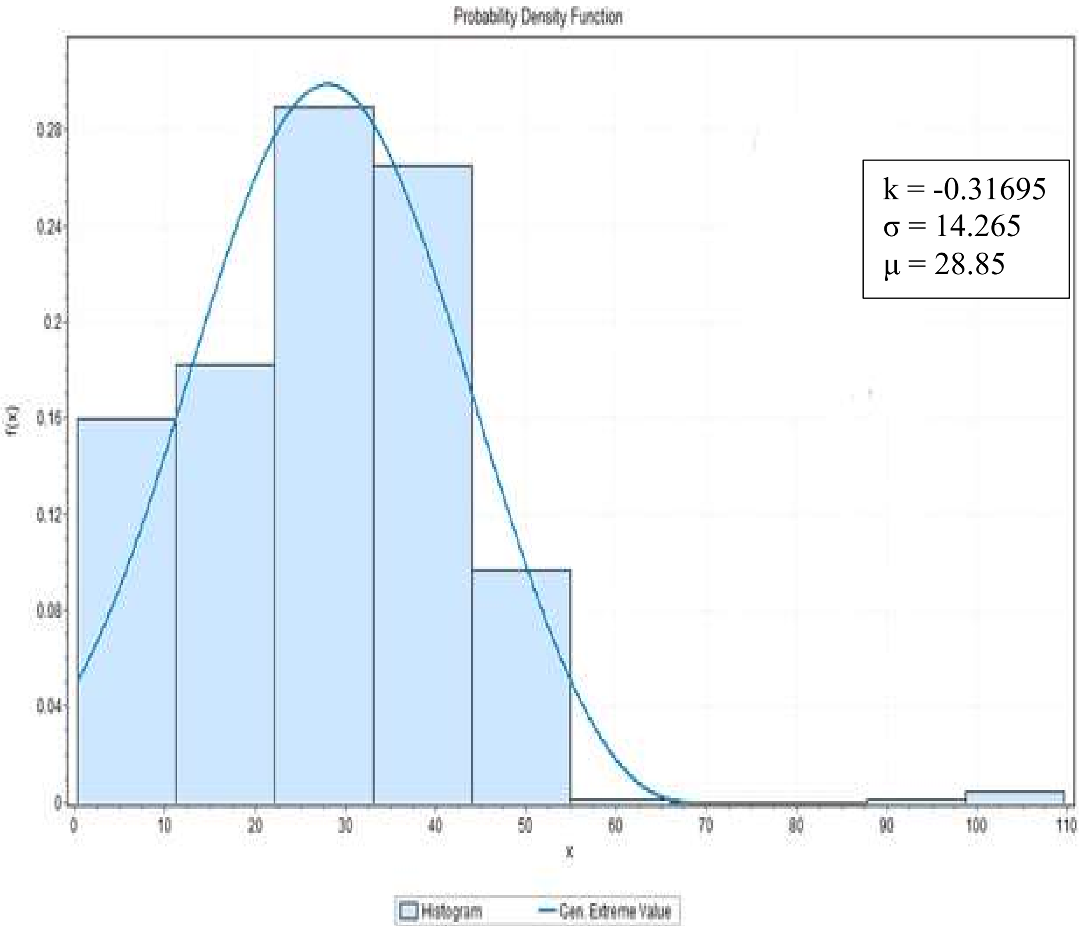


Figure 6.2 Repair age data distribution

Corresponding to the Spearman’s rho value of 0.936, a Gumbel copula parameter of 4.5 was obtained. Using the general extreme value marginal and the Gumbel copula 4.5, 1000 data points were generated.

Similarly, using a normal marginal, 1000 data points were generated. Both of these data sets were used for regression analysis, and the results are shown in Table

6.2. It is apparent that using general extreme value (GEV) as a marginal gives a better result for regression analysis, with an R squared value of 0.88 and a standard error of 4.89.

Application of Bayesian inference to the pipe data using the Gumbel copula and the Frank copula is discussed in the following sections.

Table 6.2 Comparison of regression analysis using copula modeling and original data

		Regression Statistics	
	Original data	Gumbel copula GEV marginal generated data points	Gumbel copula normal marginal generated data points
R squared	0.88	0.88	0.87
Standard error	4.92	4.89	5.06
Observations	902	1000	1000

6.7 **Gumbel Copula Parameter Estimation**

The likelihood function for Bayesian inference for a copula parameter can also be written as (Kelly 2007)

$$f(t_1, t_2) = f_1(t_1) f_2(t_2) c[F_1(t_1), F_2(t_2)] \qquad \text{Eq. 6.11}$$

The “zero tricks” as described in the WinBUGS manual is used to encode the likelihood function (Kelly 2007). In zero tricks, where x[i] observations with likelihood L[i], the Poisson (phi) observation of zero has likelihood exp(-phi), so if observed data is a set of 0's, and phi[i] is set to -log(L[i]), the correct likelihood is obtained.

A bivariate Gumbel copula is given by the following formula

$$C(u, v; \theta) = e^{-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{-1/\theta}} \qquad \text{Eq. 6.12}$$

The bivariate copula density is

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \qquad \text{Eq. 6.13}$$

$$c(u, v; \theta) = \frac{C(u, v; \theta)(log(u) \,.\, log(v))^{\theta-1}}{uv((-\,log(u))^{\theta} + (-\,log(v))^{\theta})^{2-\frac{1}{\theta}}} x(((-log(u))^{\theta} + (-\,log(v))^{\theta})^{\frac{1}{\theta}} + \theta - 1) \qquad \text{Eq. 6.14}$$

For a marginal with a normal distribution Eq. 6.11, can be expressed as

$$f_i(t_i) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau_i}{2}(t_i - \mu_i)^2} \qquad \text{Eq. 6.15}$$

where μ denotes the mean of the distribution and τ is known as precision, which is equal to $1/\sigma^2$.

where $u = F_1(t_1)$ and $v = F_2(t_2)$

The density function in Eq.6.11 was used as a likelihood function in Bayesian analysis, and a theta prior of uniform value of 1 to 100 was used for analysis. Using the 0.93 Spearman's rho value, the Gumbel copula value of 4.5 was chosen to generate 100 random data points. Using general extreme value distribution, the parameter estimated by the maximum pseudolikelihood method yielded a value of 4.95 with a standard error of 0.69. For the same data set of repair age and pipe age, using a normal marginal resulted in a parameter estimation of 4.81 with standard error of 0.67.

The data generated by using a normal marginal was used to obtain a parameter by Bayesian inference. The mean value of the parameter was 4.106. The range for the 2.5th percentile was 3.40 and for the 97.5th percentile was 4.912, as shown in Table 6.3. Comparing this theta value with those obtained from the maximized pseudolikelihood method showed that the value obtained from Bayesian inference was of a narrower range than the 95% confidence interval for the theta value obtained from the maximized pseudolikelihood method, as shown in Figure 6.3.

Table 6.3 Results for Gumbel copula parameter estimation

node	**mean**	**sd**	**MC error**	**2.5%**	**median**	**97.5%**	**start**	**sample**
mu1	31.56	0.8399	0.04827	29.81	31.58	33.15	4001	6000
mu2	26.77	0.7877	0.04559	25.2	26.75	28.29	4001	6000
tau1	0.004936	4.293E-4	1.128E-5	0.004128	0.004922	0.005831	4001	6000
tau2	0.005282	4.651E-4	1.26E-5	0.004428	0.00526	0.006252	4001	6000
theta	4.106	0.3857	0.007917	3.402	4.085	4.912	4001	6000

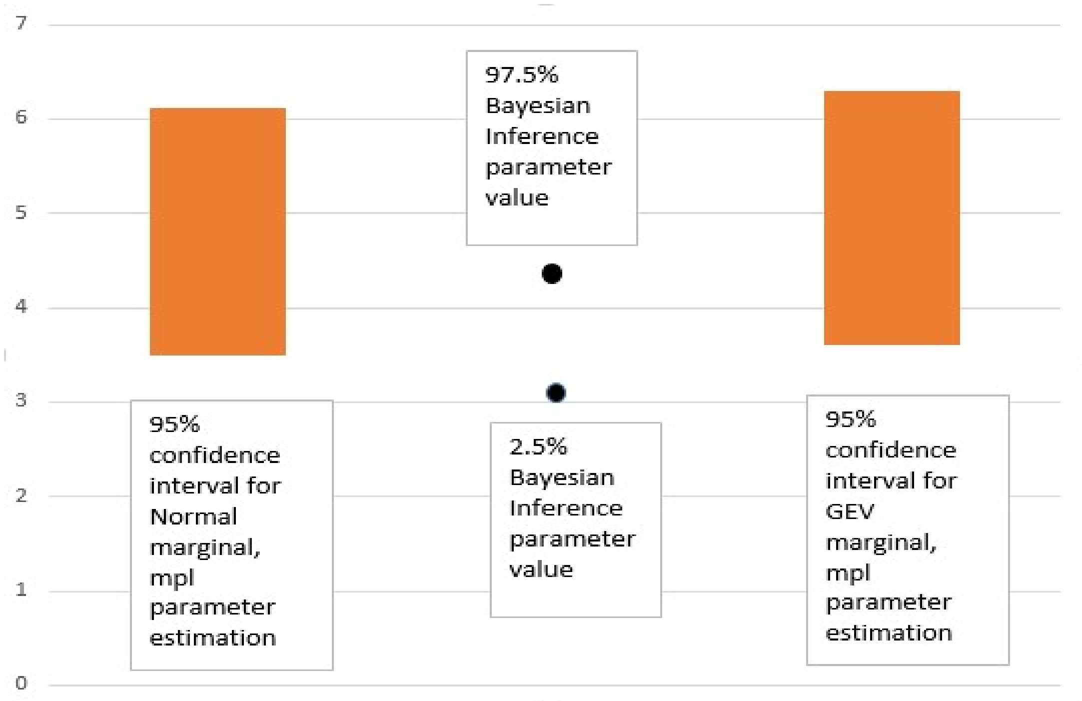


Figure 6.3 Comparison of parameter estimation based on Bayesian inference and the maximized likelihood method

The results from WinBUGS are shown in Figure 6.3, 6.4, and 6.5, where results for quantile, kernel density, and autocorrelation values for theta are shown, respectively.

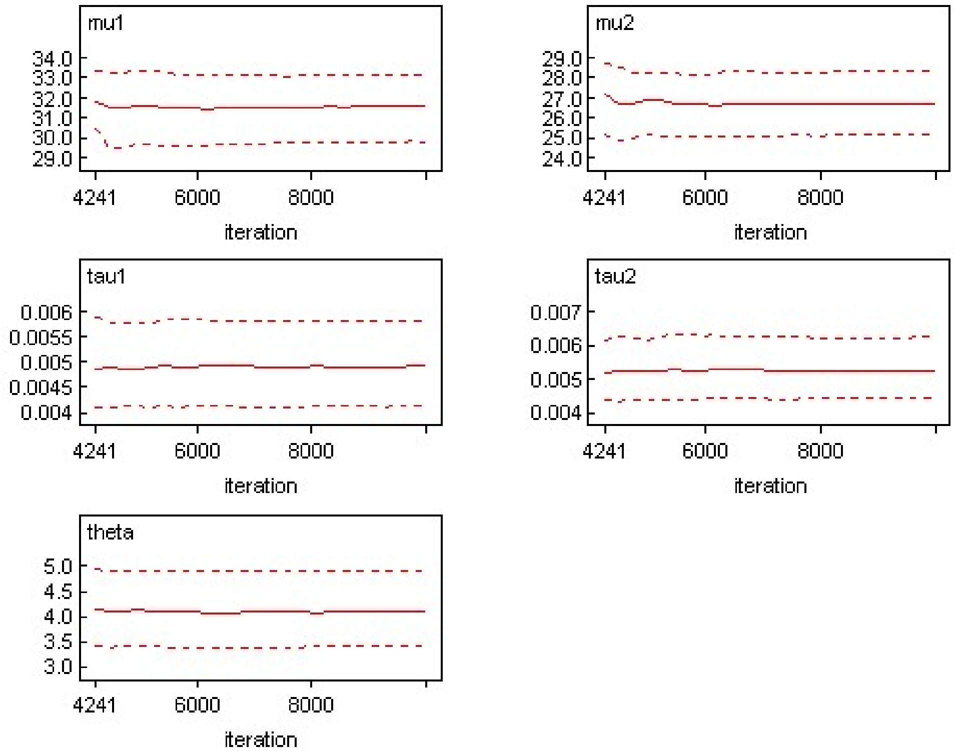


Figure 6.4 Quantile figures from WinBUGS analysis for Gumbel copula parameter estimation

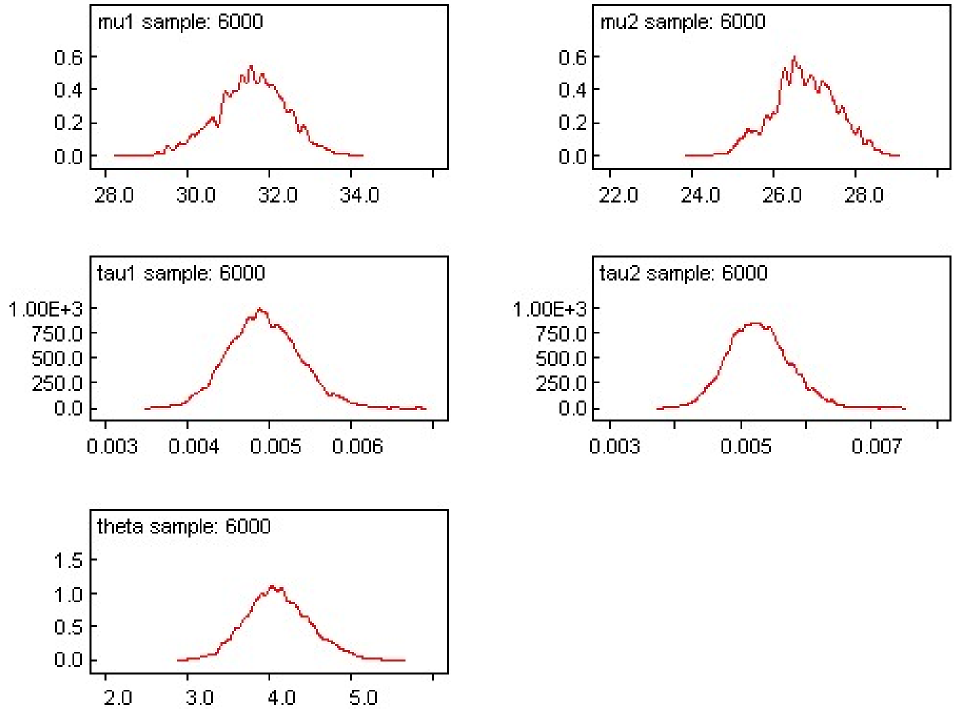


Figure 6.5 Density figures from WinBUGS analysis for Gumbel copula parameter estimation

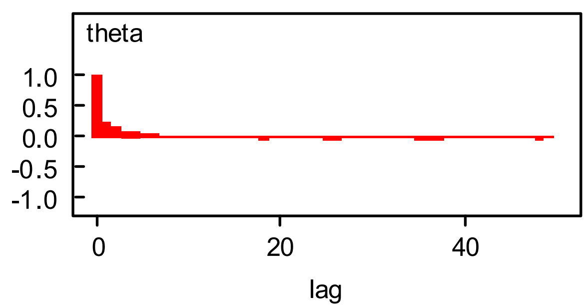


Figure 6.6 Convergence of simulation - autocorrelation plot

The trace plots for some of the parameters used in modeling are shown in Figure 6.4. The trace plots, which show the iterations versus generated values, are all within a parallel zone without strong periodicities; thus, convergence can be said to have been reached. The autocorrelation plot for the theta parameter is shown in Figure 6.6. For the autocorrelation plot, there is no obvious pattern with increasing lag, justifying the conclusion that convergence has been reached. An important parameter used as a check on the simulation is the Monte Carlo (MC) error. It measures the variability of the estimate due to the simulation. A low MC error is required to calculate the parameter of interest with increased precision (Mills and Attoh-Okine 2014). For all the parameter calculations from Table 6.3 and Table 6.4, the MC error value is significantly low.

6.7.1 **Frank Copula Parameter Estimation**

The bivariate Frank copula is expressed as

$$C(u,v;\theta) = \frac{1}{\theta} ln\left(+ \frac{\left(e^{-\theta u} - 1\right)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \qquad \text{Eq. 6.16}$$

Since the Frank copula is continuous, the bivariate copula density can be written as

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} \qquad \text{Eq. 6.17}$$

$$= \frac{\theta\left(1 - e^{-\theta}\right)e^{-\theta(u+v)}}{[1 - e^{-\theta} - (1 - e^{-\theta u})(1 - e^{-\theta v})]^2}$$

For the Frank copula, a value of 50 for θ corresponds to a Spearman's rho of 0.99, which is almost equal to 1. Therefore, a uniform distribution between 0 and 50 was used as a prior for theta. A normal distribution was chosen for the mean parameter, and a gamma distribution was chosen for the precision parameter tau.

Using two normal marginal distributions and Frank copula 14, based on the Spearman's rho dependency, the joint distribution was obtained. Fifty random samples were generated. These sample data were then used to obtain the parameter θ by the maximized pseudolikelihood method for the Frank copula, which was 12.07 with a standard error of 2.086. The 95% confidence level interval for the copula parameter yielded the lower limit of 7.98 and the upper limit of 16.16. The same set of 50 points was then used to test the script in the WinBUGS software to conduct a Bayesian analysis to obtain parameter θ. The result yielded a mean value of theta 11.25 with the 2.5th and 97.5th percentile values as 8.129 and 14.7, respectively. The Bayesian analysis result fit within the 95% confidence level from the likelihood method with a narrower variation. The summary for the result is shown in Table 6.4. Mu1 and tau1 correspond to the values of mean and tau=1/std.dev^2 of the pipe age marginal.

Table 6.4 Summary of results for Frank copula parameter

node	**mean**	**sd**	**MC error**	**2.5%**	**median**	**97.5%**	**start**	**sample**
mu1	34.07	1.011	0.0289	32.08	34.08	36.07	4001	17000
mu2	27.66	0.9524	0.02786	25.77	27.64	29.56	4001	17000
tau1	0.007146	9.604E-4	8.536E-6	0.0054	0.007097	0.009143	4001	17000
tau2	0.008054	0.001074	9.752E-6	0.00609	0.007995	0.01029	4001	17000
theta	11.25	1.679	0.01584	8.129	11.19	14.7	4001	17000

From the above analysis, it is shown that Bayesian analysis gives a parameter close to the original parameter of the chosen Frank copula. The values of mean and standard deviation are also close to the original values of the marginals.

6.8 Remarks

The parameter estimation of the copula can be achieved through different methods, such as maximized likelihood, inversion of Spearman's rho, inversion of Kendall's tau, and maximized pseudolikelihood, among which the maximized pseudo likelihood approach is based on ranked data. These parameter estimation methods are newly developed methods. Bayesian inference is an alternative method that can be used to obtain the parameter of the copulas. The results indicated that Bayesian inference can provide a good estimate of the copula parameter. Two types of copula, the Frank and Gumbel copulas, were tested with a normal marginal, and both gave a good estimation. In comparison to the maximum pseudolikelihood method of parameter estimation, the Bayesian result provided a narrower range for the parameter value. For more complex marginal models and/or copula functions, the likelihood can be hard to maximize directly; in such cases, Bayesian Inference using MCMC is a better alternative. For further research, the analysis can be extended to include methods applying Bayesian inference to choose a model of different copulas to fit the pipe condition data based on the Akaike information criterion (AIC).

Chapter 7

HYBRID COPULAS AND GIS ANALYSIS

7.1 Introduction

As discussed in section 3.5, soil type, soil characteristics, and water table position affect pipe deterioration condition. Some of the factors related to soil condition that influence pipe deterioration for cast iron pipes are the following:

- Non-draining backfill leads to moisture retention and promotes corrosion.
- Low resistivity soil leads to higher corrosion rates. Soil chlorides (e.g., from de-icing salts) reduce soil resistivity.
- Low pH (< 4) means soil is acidic and likely to promote corrosion. High alkaline conditions (pH > 8) can also lead to high corrosion.
- High availability of oxygen promotes microbial-induced corrosion in the presence of sulfates and sulfides.
- Low chloride levels in high pH (> 11.5) environments can lead to corrosion.
- Soil sulphate accounts for MIC and possible food source for sulfate-reducing bacteria in anaerobic conditions under loose coatings.
- Sulfate-reducing bacteria give off sulfides that are excellent electrolytes.

For cast iron pipes, the effects of the groundwater table are as follows:

- Pipes exposed to wet/dry conditions have higher failure rates than pipes totally below the water table or pipes totally exposed to the atmosphere.
- The water table position indicates if the wet/dry cycle is likely to occur.

Similarly, for asbestos cement pipes, the soil and groundwater effects are as follows:

- A changing environment promotes a higher expansion of the matrix than an unchanging environment.
- The water table position will indicate if the wet/dry cycle is likely to occur. Soil sulfate attack only occurs if sulfate is in solution.
- Non-draining backfill leads to moisture retention, promoting external corrosion.
- Low pH (< 5) means the soil is acidic and likely to promote corrosion.
- Soils with high sulfate (> 1000 ppm) can attack AC pipes with high free lime (type I AC pipes).
- AC pipes are not designed for frost loads. Significant frost load can cause substantial stress.

Soil and ground conditions do not have significant effects on pipe condition for PVC pipes.

The effects of soil on PVC pipes are as follows:

- PVC pipes are impervious to high-octane gasoline and gasoline-saturated water for periods of up to 2 years.
- PVC pipes are not designed for frost loads. If conditions exist that develop significant frost loads, then the pipe will be subjected to additional stresses (annual) and may fail if already significantly scratched.

Extensive amounts of soil data information are available through geographical informational systems (GIS). Obtaining pipe data through GIS is difficult, as access to the information is often restricted due to security reasons.

7.2 Uses of GIS in Water Pipe Asset Management

Geographical information systems (GIS) applications are used for collecting, manipulating, analyzing, and visualizing graphical information from water supply systems.

In this chapter, limited pipe data obtained from GIS is studied and copula modeling is applied to soil data and pipe data. An advantage of GIS is that these systems can store and organize large quantities of data and create graphical representations of different combinations of these data. GIS can provide valuable information for system modeling and can also integrate data sets as layers of maps for analysis. Table 7.1 shows examples of water utility data sets for a GIS.

Table 7.1 Examples of water utility data sets for a GIS (Smith et al. 2000)

Data Category	Example of Graphical Layers
Base data	Control information Planimetric feature Hydrology features
Facilities and distribution	Water piping Water valves and utility holes Fire hydrants Service areas Water plant facilities

Table 7.1 continued

Data Category	Example of Graphical Layers
Land records data	Property boundaries Easements Right of ways
Natural resources data	Groundwater data Drainage data Soil data Floodplain boundaries Topographic features Vegetation information
Transportation network data	Roads Intersections Bridges

7.3 **Analysis of Pipe Soil Data using Copula Modeling**

GIS is an important tool for storing data on pipe networks; many counties and local communities maintain GIS pipe network data; however, underground infrastructure such as water distribution networks to not have GIS readily available for public use. After extensive searching, water main GIS data was obtained for North Carolina. Even though the North Carolina state gives a statewide water pipe network, for this analysis, repair age was needed. Only 342 out of 118996 total pipes had information on repair age. The data from those 342 pipes was used for pipe variable analysis. Most of those 342 pipes were located in four counties: Duplin, Union, Perquimans, and Hyde.

7.3.1 **Source of GIS Data**

Water pipe data source:

Title: Water Distribution Pipes

Online link: www.nconemap.com

Date: 2000 (published)

Soil data source:

Soil information data was obtained for counties in North Carolina. For Duplin county:

Title: Soil Survey Geographic (SSURGO) database for Duplin County, North Carolina

Online link: http://websoilsurvey.nrcs.usda.gov

Date: 2007 (Published)

Figure 7.1 shows the map location of the four counties in North Carolina for which pipe information was obtained.

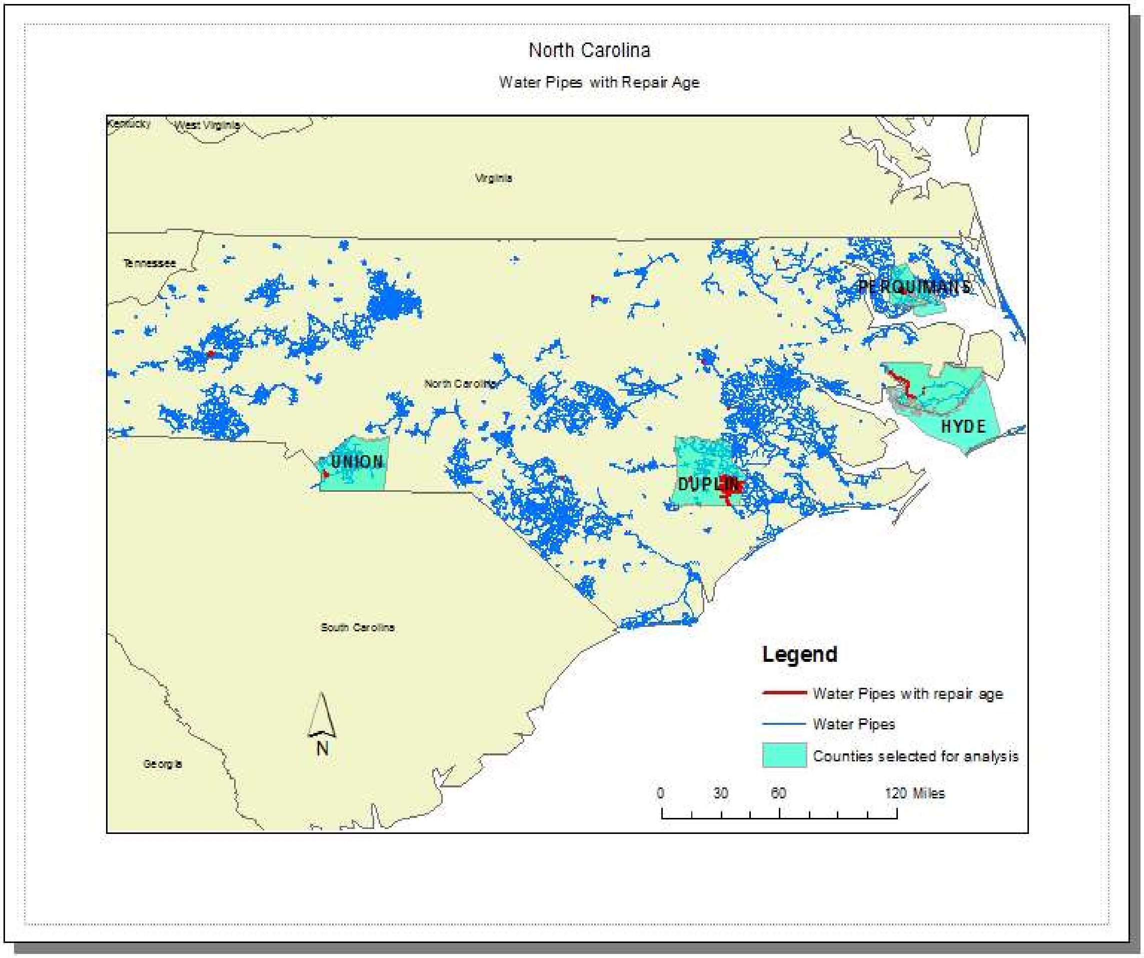


Figure 7.1 Water pipe network of North Carolina

7.3.2 **Flow Diagram of GIS Data Pipe Analysis**

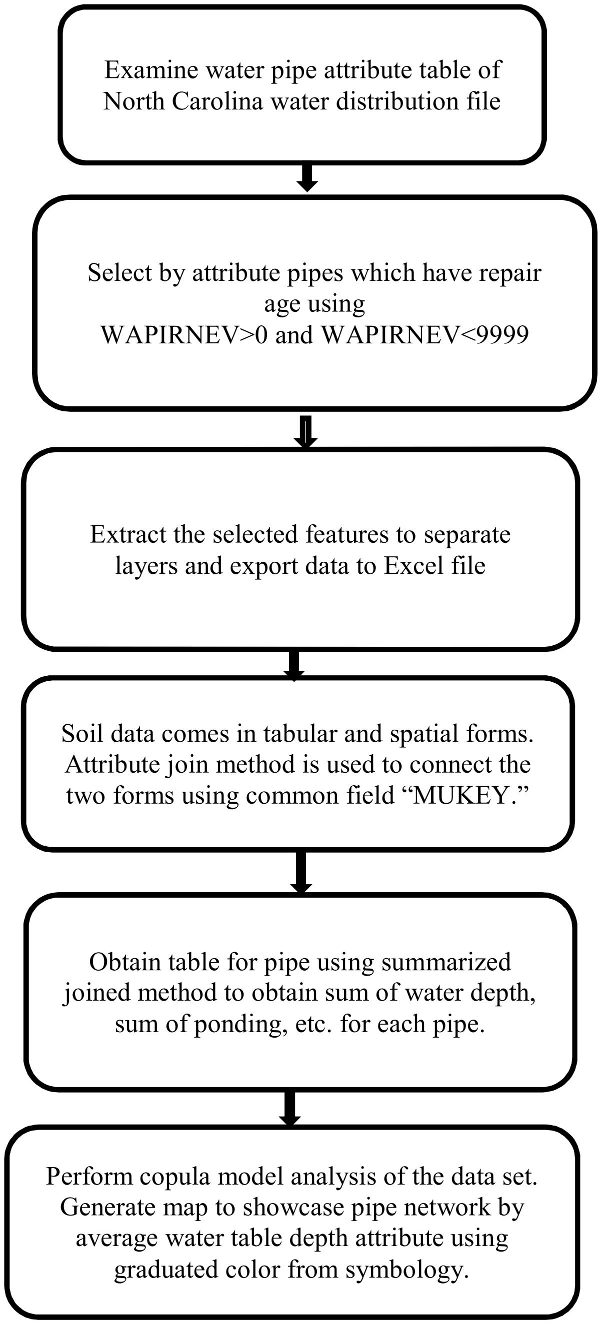


Figure 7.2 Flow diagram for pipe data analysis using GIS

7.3.3 **Pipe Data Analysis Using GIS**

Obtaining the excel file from the pipe network GIS map gave the following variables of the water pipe network for analysis:

- Pipe age = 2000 (year the map is revised) -pipe construction year
- Repair age, RA = Pipe renovation year-pipe construction year, pipe diameter and pipe length (LEN)

There was no significant variance for pipe diameter, so that variable was not used. From the GIS soil data, WTT = water table depth per length of pipe and average pH for pipe were used for analysis. The values of Spearman's rho for these four variables are shown in Table 7.2.

Table 7.2 Spearman's rho values for pipe variables

	Repair Age	pH	Water Table/Pipe Length	Pipe Length
Repair Age	1.00	-0.43	0.04	-0.40
pH	-0.43	1.00	-0.20	0.58
Water Table/Pipe Length	0.04	-0.20	1.00	-0.66
Pipe Length	-0.40	0.58	-0.66	1.00

The scatter plots of the variables are shown in Figure 7.3. From the scatter plots and the dependence values, it is apparent that pipe's repair age has a somewhat negative relationship with the pH value and length of the pipe. All the pipes considered for this analysis were made of PVC material. PVC pipe is known not to be electrically conductive and is unaffected by excessively hard or soft water and changes in pH. The values from Spearman's rho indicate that PVC pipes tend to have a higher repair age if the conditions are more acidic and the pipes are of shorter length. The five variables were then analyzed with the C-vine. The C-vine was chosen over the D-vine as repair age was placed at the first node for the analysis, and the other variables, pH, average water table depth, and average pH, were studied. The three trees from the analysis are shown in Figure 7.4, Figure 7.5, and Figure 7.6.

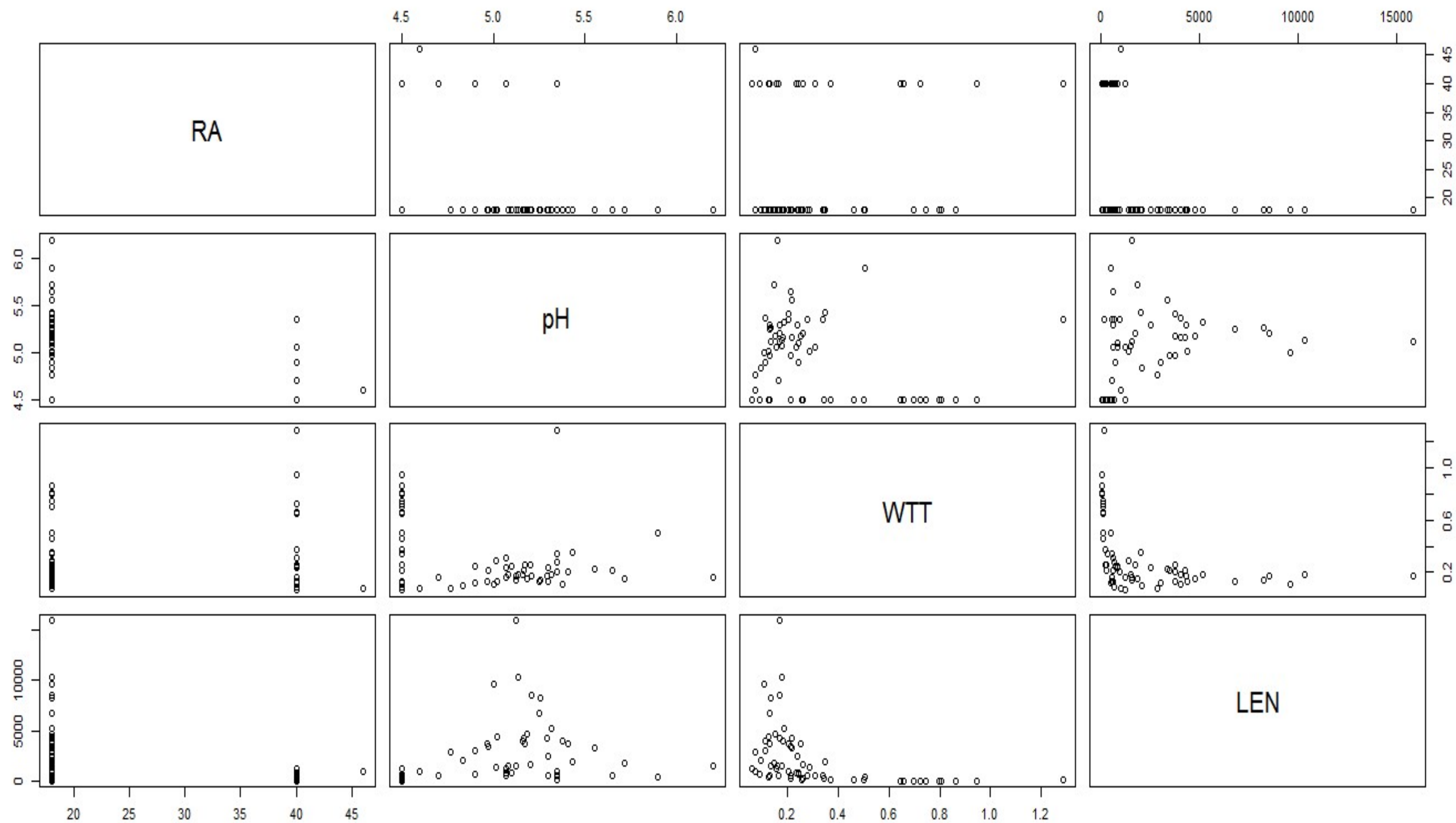


Figure 7.3 Scatter plots of pipe variables

TREE 1

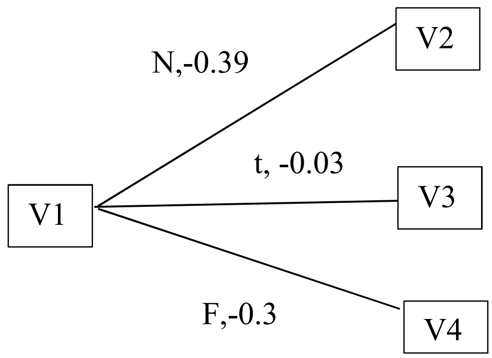


Figure 7.4 First tree of C-vine analysis, V1= repair age, V2=pH, V3= water table depth,V4=pipe length. Type of copula family and empirical Kendall’s tau value are noted on the edges of the tree.

TREE 2

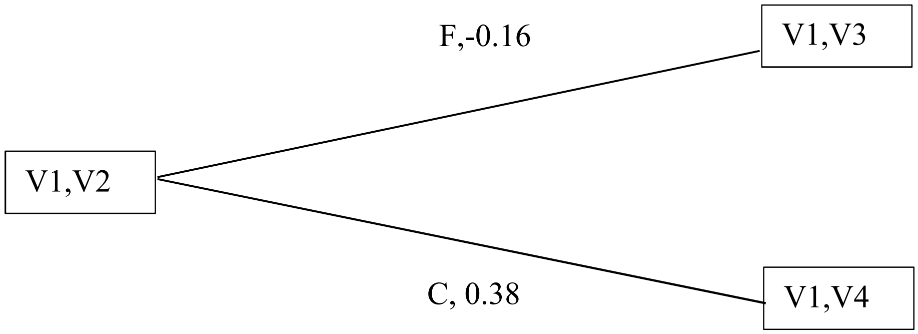


Figure 7.5 Second tree of C-vine analysis, V1= repair age, V2=pH, V3= water table depth, V4=pipe length. Type of copula family and empirical Kendall's tau value are noted on the edges of the tree.

TREE 3

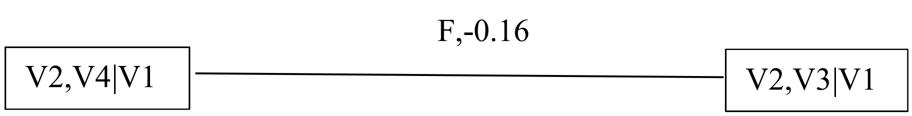


Figure 7.6 Third tree of C-vine analysis, V1= repair age, V2=pH, V3= water table depth, V4=pipe length. Type of copula family and empirical Kendall's tau value are noted on the edges of the tree.

GIS serves as a significant tool to provide graphical presentations to indicate pipes with greater water table depth. As shown in Figure 7.7, graduated color symbology indicates pipes with higher water table depth with darker colors. This type of graphical presentation can provide a general visual representation indicating water pipes with higher pH values or pipes with higher water tables, etc.

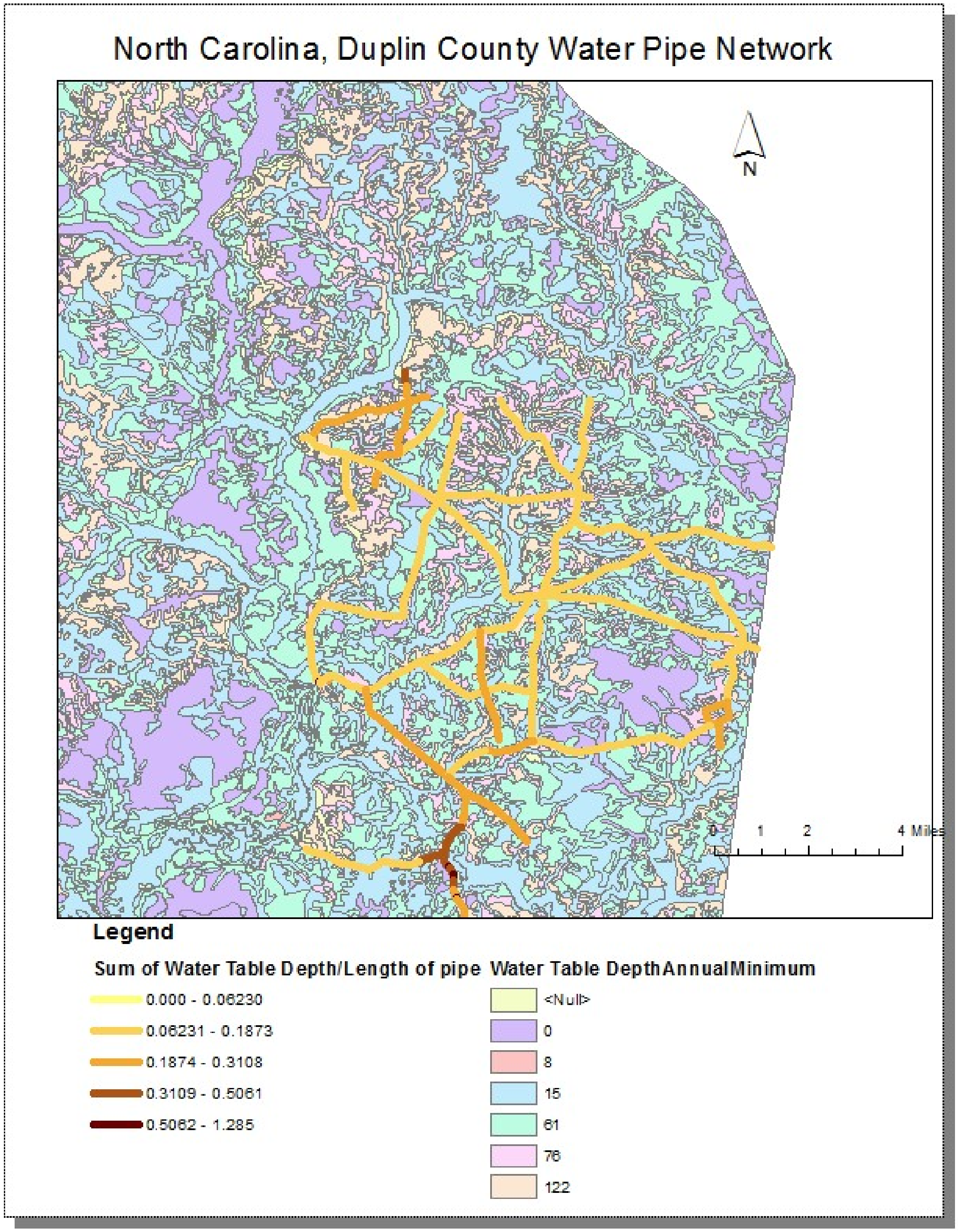


Figure 7.7 Pipe network at Duplin County in North Carolina showing graduated color for water table depth

7.4 Remarks

GIS is a powerful tool for analyzing factors for the deterioration of water pipes. GIS is capable of storing a large amount of data which can be used for analyzing models. This research was limited to data that was publically available, which limited the scope of the modeling. Not enough pipe data was available to draw conclusions regarding pipe age and/or repair age with soil or drainage information. GIS information coupled with MicroStation and InRoads software can produce variables such as whether pipes are subjected to wet/dry conditions or if the pipes are submerged in water or exposed to air. Pipes subjected to wet/dry conditions are more prone to deterioration. Survey information can be used to draw profiles of roadways, showcasing the water pipes in InRoads software. Information about the water depth can then be plotted on the profiles to show what level the water is with respect to the pipes. Figure 7.8 is an example of a water pipe with the depth of water shown.

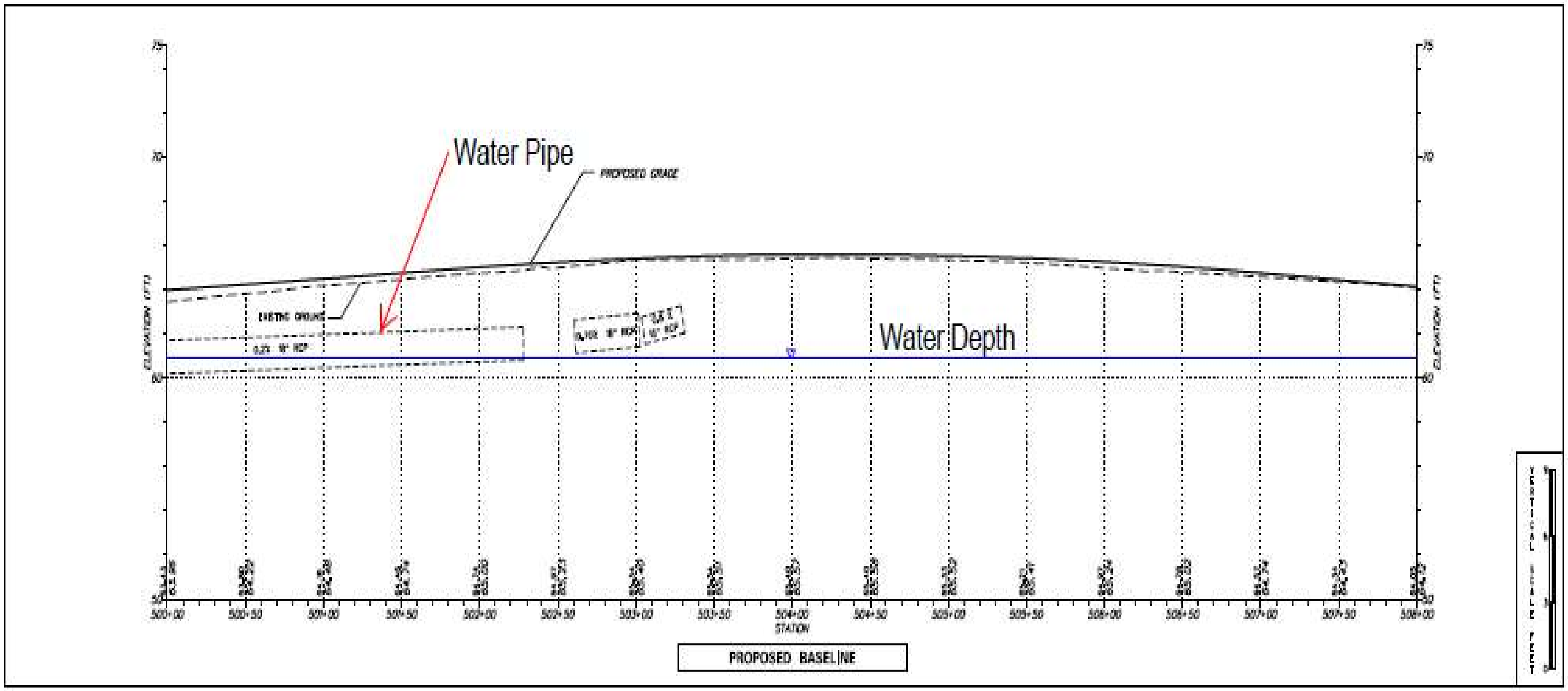


Figure 7.8 Profile of a water pipe with water depth shown

Chapter 8

SUMMARY AND CONCLUSIONS

8.1 Summary

Pipeline systems are critical infrastructure assets. As the pipes deteriorate, timely repair and rehabilitation is required otherwise the cost of repair can increase substantially. Major pipelines in the US have reached or passed their design life. It is therefore very important that efficient and cost-effective maintenance and rehabilitation strategies are employed to prevent potential failure issues in the future. Infrastructure asset management is an approach that can help maintain utility at a desired level of service at the lowest life-cycle cost. Asset management practices applied to underground infrastructure can help utility companies understand the timing and cost associated with rehabilitating, repairing, and replacing pipelines. Knowledge gained from these efforts also helps to develop pipe material selection criteria. Pipe failure is a complex process involving the work of many factors. It is important to have an understanding of the different factors that cause pipe leakage and to understand the common mechanisms of failure. Different pipe deterioration models are applied to predict the condition of the water pipes.

In this research paper, copula modeling was applied to pipeline engineering. Copula modeling is an emerging method of modeling and is useful in the case where the marginals belong to different families of distributions. Copula modeling is also useful for generating large numbers of data points when it is difficult to obtain data sets. This is the case for pipe condition assessment, where underground pipe data information is expensive to obtain and where data sets have random variables belonging to non-

Gaussian family distributions. The large data sets generated can then be used for evaluating the current pipe condition models and the appropriateness of those models for determining the remaining life of a pipe or its condition.

The conclusions in the following section are drawn from the study and research.

8.2 **Conclusions**

Following are the conclusions drawn from the research work.

1. A thorough review of pipe asset management and the pipe deterioration process helped to identify factors affecting pipe deterioration. Identifying factors for pipe deterioration is an important step in the pipe asset management process. Through model dependency analysis, attributes are chosen to develop a pipe deterioration model, which helps to determine rehabilitation and pipe replacement strategies (Yan et al. 2013). In that way, obtaining variable dependency modeling helps utility companies in the process of developing asset management plans. These condition assessments help to develop necessary steps and procedures to maintain water distribution systems running at desired levels of service in a cost-effective way.
2. It was observed in the pipe data set from the desert west region that among the variables there was a strong dependency between pipe age and pipe material and the ratio of pipe age/repair age. Whereas, not much dependency was observed between pipe age and pipe size. Even though it is known that smaller size pipes tend to have more failure and tend to have shorter life span. The

rank method of dependency Spearman's rho gave a better indication of correlation if the data set had skewness and was non-normal. An example of two variables, pipe age and ratio of pipe age/repair age, was used to showcase how Spearman's rho gives a better indication of dependency than Pearson's coefficient by considering the skewness and considering the whole distribution for the measure of dependence. Copula modeling dependence analysis use of spearman's rho and Kendall's tau is much better for taking consideration of outliers.

3. It was observed in the data set for leaks per day and temperature that the marginals were not following a normal distribution. The marginal for leaks per day was a geometric distribution, whereas that of the temperature data was a Weibull distribution. Using these two types of marginal and the Frank copula, it was possible to derive a large number of data points matching the original data distribution. The generated points were then used for regression analysis. It was shown that having a greater number of data points enabled reduction of the standard error value and significantly reduced the 95% confidence level difference compared to the original data set.
4. When D-Vine copula modeling was applied to the data set, the results indicated that pipe age and the ratio of pipe age to repair age were conditionally dependent given the pipe material. This result helped to support findings from other studies that, due to installation errors, repair work is needed in the early lives of PVC pipes after installation. Copula modeling

helped to simulate a multivariate modeling where it was evident that not all the marginals were normal and the dependency between variables differed. The dependency between variables were established by the use of different copula families, which did not depend upon the marginal distribution of the individual variables.

5. Parameter estimation of a copula can be achieved through different methods, such as maximized likelihood, inversion of Spearman's rho, inversion of Kendall's tau, and maximized pseudolikelihood, among which the maximized pseudolikelihood approach is based on ranked data. These parameter estimation methods are newly developed methods. Bayesian inference is an alternative method that can be used to obtain the parameters of the copulas. The results from the study indicated that Bayesian inference can provide a good estimate of copula parameters. Two types of copula, the Frank and Gumbel copulas, were tested with a normal marginal, and both gave good estimates. In comparison to the pseudolikelihood method of parameter estimation, the Bayesian result provided a narrower range for the parameter value.
6. This research also showcased how GIS can serve as an important tool to obtain data, including soil data. Many soil properties affect pipe deterioration. This data set can then be used for copula modeling. GIS data can be incorporated with other software, namely Microstation and Inroads, to identify pipes which are subjected to wet/dry conditions.

The research was able to draw conclusions based on the objectives set at the beginning of the work. Still more work can be done in future on copula modeling which are discussed in the following section.

8.3 **Future Research**

There is a wide scope for future research work to be done on copula modeling for water pipes and civil engineering infrastructure in general. Some of the opportunities for future work are mentioned below:

- Copula modeling is very useful when the marginals of variables are not of a normal distribution. Only a few variables affecting pipe deterioration were studied in this research; further work can be done with other variables, such as pipe lining and coating, pipe bedding, types of joints, flow velocity, etc.
- Pipe failure rate, such as pipe leaks per length per year, data availability will help to establish the correlation of the failure rate with different pipe deterioration variables using copula vine modeling.
- For more complex marginal models and/or copula functions, the likelihood can be hard to directly maximize. For such cases Bayesian Inference using MCMC is way to derive the copula family parameter. For future research, the analysis can be extended to include methods applying Bayesian inference to choose a model of different copulas to fit the pipe condition data based on the Akaike information criterion (AIC).
- Due to the restrictive availability of GIS public data, only a few soil data variables were applied for copula modeling with a limited amount

of pipe data. More robust studies can be done on different variables of soil data such as dry/wet condition of soil, chloride level, sulfate level of soil, etc. and their effects on pipe deterioration.

To summarize, copula modeling is a new modeling method that is starting to be used in different engineering disciplines. This research work can be extended to further analysis of different set of pipe data as well as other civil engineering data analysis.

REFERENCES

Aas, K. (2004). Modeling the dependence structure of financial assets: A survey of four copulas, Norwegian Computing Center, Oslo, Norway.

Aas, K., Czado, C., Fregessi, A., and Bakken, H. (2009). "Pair-copula construction of multiple dependence." *Insurance: Mathematics and Economics*, 44(2), 182-198.

Accioly, R., and Chiyoshi, F.Y. (2004). "Modeling dependence with copulas, a useful tool for field development decision process." *Journal of Petroleum Science Engineering*, 44, 83-91.

Achim, D., Ghotb, F., and McManus, K.J. (2007). "Prediction of water pipe asset life using neural networks." *Journal of Infrastructure Systems*, 13(1), 26–30.

Al-Barqawi, H., and Zayed, T. (2006). "Condition rating model for underground infrastructure sustainable water mains*." Journal of Performance Constructed Facilities*, 20(2), 126-135.

Association of Metropolitan Water Agencies (AMWA), National Association of Clean Water Agencies (NACWA), and Water Environment Federation (WEF) (2007). *Implementing asset management: A practical guide*, Association of Metropolitan Water Agencies, Washington D.C.

Attoh-Okine, N. (2013). "Pair-copulas in infrastructure multivariate dependence modeling." *Construction and Building Materials,* 49, 903-911.

American Water Works Association (2016). *Buried no longer: confronting America's water infrastructure challenge.* <http://www.awwa.org/Portals/0/files/legreg/documents/BuriedNoLonger.pdf>, Sept 2016.

Bacigal, T. (2006) "Fitting Archimedean copulas to bivariate geodetic observations*."5th International Conference APLIMAT,* Bratislava, Slovakia.

Barnes, Z.A., UKWIR (United Kingdom Water Industry Research), GWRC (Global Water Research Coalition), WaterRF (Water Research Foundation), WERF (Water Environment Research Foundation), and WSAA (Water Services Association of Australia). (2008). *Tool for risk management of water utility assets*, United Kingdom Water Industry Research, London, U.K.

Bedford, T., and Cooke, R. M. (2001). "Probability density decomposition for conditionally dependent random variables modeled by vines." *Annals of Math and Artificial Intelligence*, 32, 245-268.

Berg, D. (2008). "Using copulas: An Introduction to practitioners*"*, Norwegian ASTIN Society, Oslo, Norway.

Burn, S. (2006). *Long-term performance prediction for PVC pipes,* IWA Publishing Organization, London.

Breachman, E.C., and Schepsmeier, U. (2013). "Modeling dependence with C- and D-vine copulas: The R package CDVine." *Journal of Statistical Software*, 52(3), 1-27.

Christodoulou, S., Aslani, P., and Vanreterghem, A. (2004). "*Proceedings of the World Water and Environmental Resources Congress and Related Symposia"*, ASCE, Philadelphia, PA, 1-9.

Clair, M. A. and Sinha, S. (2012), "State-of-the-technology review on water ipe condition, deterioration and failure rate prediction models." *Urban Water Journal*, 9(2), 85-112.

Cromwell, J.E., Nestel, G., and Albani, R. (2003). *Financial and economic optimization of water main replacement programs*, AWWA Research Foundation and American Water Works Association, Denver, Colorado.

Cromwell, J.E., and Speranza, E. (2006). *Water infrastructure at a turning point: The road to sustainable asset management*, American Water Works Association, Denver, Colorado.

Deb, A.K., Grablutz, F.M., Hasit, Y.J., and Snyder, J.K. (2002). *Prioritizing water main replacement and rehabilitation*, AWWA Research Foundation and American Water Works Association, Denver, Colorado.

Damodaran, N., Pratt, J., Cromwell, J., Lazo, J., David, E., Raucher, R., Herrick, C., Rambo, E., Deb, A., and Snyder, J. (2005). *Customer acceptance of water main structural reliability,* AWWA Research Foundation, Denver, Colorado.

Davis, P., Burn, S., Moglia, M., and Gould, S. (2007). "A physical probabilistic model to predict failure rates in buried PVC pipelines." *Reliability Engineering and System Safety*, 92(9), 1258–1266.

Davis, P., Burn, S., and Gould, S., (2008). “Fracture prediction in tough polyethylene pipes using measured craze strength.” *Polymer Engineering & Science*, 48 (5), 843–852.

Davis, P., De Silva, D., Marlow, D., Moglia, M., Gould, S.J., and Burn, S. (2008). “Failure prediction and optimal scheduling of replacements in asbestos cement water pipes.” *Journal of Water Supply: Research and Technology-Aqua*, 57(4), 239–252.

Davis, P., and Marlow, D. (2008). “Quantifying economic lifetime for asset management of large diameter pipelines.” *American Water Works Association*, 100(7), 110–119.

De Silva, D., Moglia, M., Davis, P., and Burn, S. (2006). “Condition assessment and probabilistic analysis to estimate failure rates in buried metallic pipelines.” *Journal of Water Supply: Research and Technology-Aqua*, 55(3), 179–191.

Deb, A.K., Grablutz, F.M., Hasit, Y.J., Snyder, J.K., Loganathan,G.V., and Agbenowski, N. (2002). *Prioritizing water main replacement and rehabilitation*. Denver, CO: American Water Works Association Research Foundation.

Dehghan, A., McManus, K.J., and Gad, E.F. (2008a). “Probabilistic failure prediction for deteriorating pipelines: Nonparametric approach.” *Journal of Performance of Constructed Facilities*, 22(1), 45–53.

Dehghan, A., McManus, K.J., and Gad, E.F. (2008b). “Statistical analysis of structural failures of water pipes.” *Water Management*, 161(4), 207–214.

Djechiche, B., Liv, S., and Hulf, H. (2004). *An introduction to copulas with applications*, Svenka Akturiforeningen, Stockholm.

Embrechts, P., Lindkog, F., and McNeil, A. (2001). Risk Lab, ETH Zurich, Zurich, Switzerland.

Frees, E.W., and Valdez, E.A. (1998). “Understanding relationships using copulas.” *North American Actuarial J.*, 2(1), 1-25.

Folkman, S. (2012). *Water main break rates in the USA and Canada: A comprehensive study*, Utah State University Buried Structures Laboratory, Logan, Utah.

Gaewski, P.E., and Blaha, F.J. (2007). *Analysis of total cost of large diameter pipe failures*, AWWA Research Foundation, Denver, Colorado..

Geem, Z.W., Tseng, C., Kim, J., and Bae, C. (2007). “Trenchless water pipe condition assessment using artificial neural network.” *Proceedings of the Pipelines: Advances and Experiences with Trenchless Pipeline Projects*, ASCE, Boston, Massachusetts.

Genest, C., J. F. Quessy, and B. Rémillard (2006)*, “*Goodness-of-fit procedures for copula models based on the probability integral transformation*,” Scand. J. Stat., 33(2), 337–366.*

Genest, C., and Favre, A. (2007). “Everything you always wanted to know about copula modeling but were afraid to ask.” *Journal of Hydrologic Engineering*, 12, 347-368.

Genest, C., Ghoudi, K., and Rivest, L.P. (1995). “A semiparametric estimation procedure of dependence parameters in multivariate families of distribution.” *Biometrika*, 82(3), 545-552.

Genest, C., Remillard, B., and Beaudoin, D. (2009). “Goodness of fit for copula: A review and power study.” *Insurance: Mathematics and Economics*, 44, 199–213.

Genest, C., Quessy, J.F., and Remillard, B. (2006). “Goodness-of-fit procedures for copula models based on the probability integral transformation.” *Scand. J. Stat*., 33(2), 337-366.

Graler, B., Van den Berg, M.J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N.E.C. (2013). “Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation.” *Hydro Earth Syst. Sci.*, 14, 1281-1296.

Grigg, N.S. (2004). *Assessment and renewal of* water *distribution systems*, AWWA Research Foundation, Denver, Colorado.

Hamilton, S., and Charalambousm, B. (2013). *Leak detection: Technology and implementation*, IWA Publishing, London, UK.

Hernandez-Lobato, J.M. (2016). “Research” <http://jmhl.org/research> (Sept 7, 2016)

Hong, F., and Prozzi, J. (2006). “Estimation of pavement performance deterioration using Bayesian approach.” *Journal of Infrastructure Systems*, 12(2), 77–86.

Kelly, D.L. (2007). “Using copulas to model dependence in simulation risk assessment.” *2007 ASME International Mechanical Engineering Congress and Exposition*, American Society of Mechanical Engineers, Seattle, Washington, 81-89.

Kim, J., Bae, C., Woo, H., Kim, J., and Hong, S. (2007) “Assessment of residual tensile strength on cast iron pipes.” *Proceedings of the Pipelines:Advances and Experiences with Trenchless Pipeline Projects*, 1–7.

Kleiner, Y., and Rajani, B. (1999). “Using limited data to assess future needs.” *Journal American Water Works Association*, 91(7), 47-61.

Kleiner Y., and Rajani, B. (2001) “Comprehensive review of structural deterioration of water mains: Statistical methods.” *Urban Water*, 3, 131-150.

Kleiner Y., and Rajani, B. (2008). “Prioritizing individual water mains for renewal.” *Proceedings of the ASCE/EWRI World Environmental and Water Resources*, National Research Council Canada-CNRC-NRC, Honolulu, Hawaii, 1-10.

Kramer, N., and Schepsmeier, U. (2011). *Introduction to vine copulas*, NIPS Workshop, Granada, Spain.

Kruschke, J.K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*, Elsevier, Amsterdam.

Lam, P. (2016). “MCMC methods: Gibbs sampling and the Metropolis-Hastings algorithm.” < http://pareto.uab.es/mcreel/IDEA2015/MCMC/mcmc.pdf> Sept, 2016.

Le Gat, Y., and Eisenbeis, P. (2000). “Using maintenance records to forecast failures in water network.” *Urban Water*, 2(3), 173-181.

Loganathan, G.V., Park, S., and Sherali, H.D. (2002). “Threshold break rate for pipeline replacement in water distribution systems.” Journal of Water Resources Planning and Management, 128(4), 271-279.

Liu, Z., Kleiner, Y., Rajani, B., Wang, L., and Condit, W. (2012). *Condition assessment technologies for water transmission and distribution system.* Office of Research and development, U.S. Environmental Protection Agency, Cincinnati, Ohio.

Matichich, M., Booth, R., Rogers, J., Rothstein, E., Speranza, E., Stanger, C., Wagner, E., and Gruenwald, P. (2005). *Asset management planning and reporting options for utilities,* AWWA Research Foundation, Denver, Colorado.

Mills, L.O., and Attoh-Okine, N. (2014). "Analysis of ground penetrating radar data using hierarchical Markov Chain Monte Carlo simulation." *Canadian Journal of Civil Engineering*, 41(1), 9-16.

Morrison, R., Matthews, J., Sinha, S., and Sterling, R. (2013). *State of technology for rehabilitation of water distribution*, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio.

Moser, A.P., and Kellog, K. (1994). *Evaluation of polyvinyl chloride (PVC) pipe performance*, AWWA Research Foundation, Denver, Colorado.

Nelsen, R.B. (1999). *An introduction to copulas*, Springer, New York.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*, John Wiley & Sons, Inc., New York.

Opila, M.C., and Attoh-Okine, N. (2011). "Novel approach in pipe condition scoring." *Journal of Pipeline Systems Engineering and Practice*, 2(3), 82-90.

Park, S.W., and Longanathan, G.V. (2002). "Methodology for economically optimal replacement of pipes in water distribution systems: 2. Applications." *KSCE Journal of Civil Engineering*, 6(4), 545-550.

Poulton, M., Le Gat, Y., and Bemond, B. (2007). "The impact of pipe segment length on break predictions in water distribution systems." *Proceedings of LESAM-2nd Leading Edge Conference on Strategic Asset Management*, International Water Association and Leading-Edge Asset Management, Lisbon, Portugal, 1-11.

Rajani, B. and Kleiner, Y., 2001. "Comprehensive review of structural deterioration of water mains: physically based models." *Urban Water*, 3 (3), 151–164.

Rajani, B., Kleiner, Y., and Krys, D. (2011). *Long-term performance of ductile iron pipe,* Water Research Foundation, National Research Council of Canada, Commonwealth Scientific and Industrial Research Organisation, and UK Water Industry Research, Denver, Colorado.

Rajani, B. and Makar, J., 2000. "A methodology to estimate remaining service life of grey cast iron water mains." *Canadian Journal of Civil Engineering*, 27: 1259–1272.

Rogers, P., and Grigg, N. (2008). “Failure assessment model to prioritize pipe replacement in water utility asset management.” *Water Distribution Systems Analysis Symposium 2006*, American Society of Civil Engineers, Cincinnati, Ohio, 1-17.

Romeo, J.S, Tanaka, T.I., and Pedroso-de-Lima, A.C. (2006). “Bivariate survival modelling: A Bayesian approach based in copulas.” *Lifetime Data Analysis*, 12, 205-222.

Rostum, J. (2000). “Statistical modelling of pipe failures in water networks.” Doctoral dissertation), Norwegian University of Science and Technology, Trondheim, Norway.

Savic, D.A. (2009). “The use of data-driven methodologies for prediction of water and wastewater asset failures.” *Risk management of water supply and sanitation system,* Petr Hlavinek, Cvetanka Popovska, Jiri Marsalek, Ivana Mahrikova, and Tamara Kukharchyk, eds. Springer, Dordrecht, the Netherlands, 181-190.

Scheidegger, A., Hug, T., Rieckermann, J., and Maurer, M. (2011). “Network condition simulator for benchmarking sewer deterioration models.” *Water Research*, 45(16), 4983-4994.

Sklar, A., 1959. “Fonction de re´partition a` dimensions et leurs marges*.” Publ. Inst. Statist. Univ. Paris*, 8, 229-231.

Silva, R.D.S., and Lopes, H.F. (2008). “Copula marginal distributions and model selection: A Bayesian note.” *Statistical Computation,* 18, 313-320.

Sklar, A. (1959). “Fonctions de répartition à n dimensions et leurs marges,” Publications de l'Institut de Statistique de L'Université de Paris 8, 229-231

Smith, L.A., Fields, K.A., Chen, A.S.C., and Tafuri, A.N. (2000). *Options for leak and break detection and repair of drinking water systems*, Battelle Press, Columbus Ohio.

Smith, M.S. (2011). “Bayesian approaches to copula modeling.” In *Bayesian theory and applications*, Paul Damien, Petros Dellaportas, Nicholas G. Polson, and David A. Stephens, eds. Oxford University Press, Oxford, United Kingdom, 336.

Srinivas, S., Menon, D., and Prasad, A.M. (2006). “Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas.” *Journal of Transportation Engineering*, 132(12), 945-955.

Swiler, L.P. (2006). *Bayesian methods in engineering design problems*, Sandia National Laboratories, Albuquerque, New Mexico.

Thacher, J., Marsee, M., Pitts, H., Hansen, J., Chermak, J., and Thomson, B. (2011). *Assessing customer preferences and willingness to pay: A handbook for water utilities*, Water Research Foundation, Denver, Colorado.

Thomson, J., and Wang, L. (2009). *Condition assessment of ferrous water transmission and distribution systems, state of technology review report*, EPA Water Supply and Water Resources Division, National Risk Management Research Laboratory, Edison, New Jersey.

Vanrenterghmen-Raven, A. (2007). "Risk factors of structural degradation of an urban water distribution system." *Journal of Infrastructure System*, 13(1), 55-64.

Wang, C., Niu, Z., Jia, H., and Zhang, H. (2010). "An assessment model of water pipe condition using Bayesian inference." *Journal of Zheijang University-Science*, 11(7), 495-504.

Wang, Y., Zayed, T., and Moselhi, O. (2009). "Prediction models for annual break rates of water mains." *Journal of Performance of Constructed Facilities*, 23(1), 47-54.

Water Research Foundation (2016), "Asset management: elements and background." < http://www.waterrf.org/knowledge/asset-management/FactSheets/AssetMgt-ElementsBackground-FactSheet.pdf>, Sept 2016.

Washington Suburban Sanitary Commission. (2016). "Home." *WSSCwater.com,* <https://www.wsscwater.com/home.html> Jan 2015.

Wood, A., and Lence, B.J. (2009). "Using water main break data to improve asset management for small and medium utilities, District of Maple Ridge, B.C." Journal of Infrastructure Systems, 15(2), 11-119.

Yan, J. (2006). "Multivariate modeling with copulas and engineering applications." *Springer Handbook of Engineering Statistics*, Springer, London, 973-990.

Yan, J. (2006). "Multivariate modeling with copulas and engineering applications." In *Springer handbook of engineering statistics*, Hoang Pham, ed. Springer, London, 973-990.

Appendix A

DATA SET 1

A.1 Data Overview

The first data set was obtained from a utility company in the desert west which gave information about pipe attributes and pipe breaks. For the first data set, the discussion divides the available data into two groups: pipe attributes and pipe breaks.

A.1.1 Pipe Attributes

In Data Set #1, the pipe attribute data was stored in a database. The pipes in the data set were identified by unique IDFEATURE. The same IDFEATURE was also present in the pipe break database. Using this feature, the pipe attribute information was related to the pipe breakage information. The attributes in Data Set #1 were shown in Table A 1.1. In the following sections, the class, diameter, and material type are discussed in detail.

Table A.1 Pipe attributes

IDFEATURE	A unique number corresponding to each pipe

INSTALL DATE	Installation date in day/month/year
TERMINATION DATE	Termination date if a pipe is abandoned
STATUS	Status of the pipe in code number
STATUS DESCRIPTION	Explanation of status code, such as 1=existing active
CLASS	Class code
CLASS DESCRIPTION	Description of class, 31=Distribution pipe
DIAMETER	Size of pipe in numerical value
DIAMETER DESCRIPTION	Description of pipe diameter, such as 600 = 6" pipe
MATERIAL	Material type
LENGTH	Length of pipe

A.1.2 Pipe Class

After matching the pipe data set and pipe breakage data set based on IDFEATURE, a total of 978 pipes were used for analysis. The class was identified with a code number, and each number corresponded to a type of pipe. Table A.2 shows the code number for pipe class, the corresponding pipe type, and the percentage of pipes in this class. All the pipes have an assigned pipe class, so the sum of the percentages yields 100%.

Table A.2 Pipe class attributes

Code	Type	Percentage
31	Distribution	76.59
40	Fire Hydrant	10.53
11	Transmission	8.90
42	Service Lateral	2.76
43	AV/AR Lateral	0.82
44	Blow off Lateral	0.41
	Total	100.00

A.1.3 Pipe Diameter Attributes

The pipes considered for analysis varied in diameter from 4 inches to 66 inches. This is a good variation of pipe size, taking into account small to large pipe sizes. The highest percentage of pipes had diameters of 6 and 8 inches.

Table A.3 Pipe attributes—diameter

Code	Diameter in inches	Percentage
400	4"	6.95
600	6"	48.57
800	8"	24.74
1000	10"	3.27
1200	12"	7.77
1400	14"	1.43
1600	16"	2.04
1800	18"	0.72
2000	20"	0.82
2400	24"	1.64
3000	30"	0.10
3600	36"	1.02
3900	39"	0.10
4200	42"	0.31
4800	48"	0.20
5400	54”	0.10
6000	60”	0.10
6600	66”	0.10
		100.00

A.1.4 Pipe Material

The pipe’s material attribute is denoted by a code number specifying each type of pipe material. The list of pipe materials and their corresponding codes is shown in Table A.4. The top three materials listed in the data set are polyvinyl chloride (PVC), asbestos cement and cast iron. Additionally, 1.64% of pipes did not have any known material attached to them. These pipes were excluded from the analysis.

Table A.4 Pipe attributes—pipe material

Material Type	Percentage
10 Polyvinyl chloride (C-900)	17.38
11 Polyvinyl chloride (C-905)	0.41
14 Polyvinyl chloride (general use)	0.41
31-Asbestos cement	71.68
41-Cast iron	3.07
42 & 43-Ductile iron	0.82
51-Steel	0.20
52-Steel cyl. concrete-pretension	1.74
54-Steel mortar-lined mortar-coated	1.94
Unknown	1.64
Others	0.71
Total	100.00

Appendix B

DATA SET # 2

B.1 Data Overview

Data Set #2 used in this research was obtained from the Washington State Sanitary Commission (WSSC). The characteristics of the water mains of the WSSC are listed below:

- Serving a 1,000-square-mile area, the WSSC maintains more than 5,500 miles of water mains.
- The water mains range in size from 1 inch to 96 inches (8 feet) in diameter.
- Since 1977, the WSSC has used ductile iron pipes.
- Ductile iron pipes are stronger than cast iron, are pre-lined with cement mortar, and are not brittle.
- The most common size for a water main is 8 inches, followed by 6, 12, 10, and 16 inches.

B.2 Effect of water temperature on water pipes

A sudden temperature drop provides a kind of shock to the pipes. Most of the WSSC water comes from the Potomac River, which feeds the WSSC Potomac Water Filtration Plant. As air temperatures drop, the temperature of the water in the Potomac River drops as well. It takes a day or two, but an increase in breaks and leaks soon follows. Even a 10-degree Fahrenheit change in the air or water temperature can dramatically increase stress on pipes, including underground pipes. Water temperatures below 40 degrees F can cause pipes to become more brittle, and aboveground pipes can freeze when the air temperature hits freezing or below. That leads to increased external stress. The Patuxent River and the two reservoirs formed behind the Brighton and T. Howard Duckett dams provide about 30% of the water for WSSC customers via the Patuxent Water Filtration Plant. However, the reservoirs are deeper than the Potomac River, and their temperatures do not change as fast. Therefore, there are fewer breaks and leaks in areas served by the Patuxent Plant. Other factors known to contribute to pipe breaks are:

- **Material**: Half of the distribution pipes, approximately 2,900 miles, are cast iron pipe (16-inch diameter and smaller), which were used from 1916 to through 1976. As cast iron is a brittle material, these pipes are prone to breakage and are very sensitive to external pressure.
- **Soil Erosion**: A previous pipeline break, excavation, or nearby construction activity often erodes soil around water mains, which can lead to pipe breaks.
- **Corrosion**: Older pipes that are not cement lined can corrode inside and outside, increasing the chances of a break.
- **Pipe Diameter**: The smaller the pipe's diameter, the greater the risk of breakage.
- **Age**: The break rate for pipes usually increases after 60 years. However, age cannot always be used as an indicator of failure. It should be noted that some pipes installed in the early 1900s have never broken.

Figure B.1 shows the data set obtained from the WSSC website that was used for analysis in this research. January month data was used as that gave the more pipe breakage incidents as temp reached below 40 degree Fahrenheit.

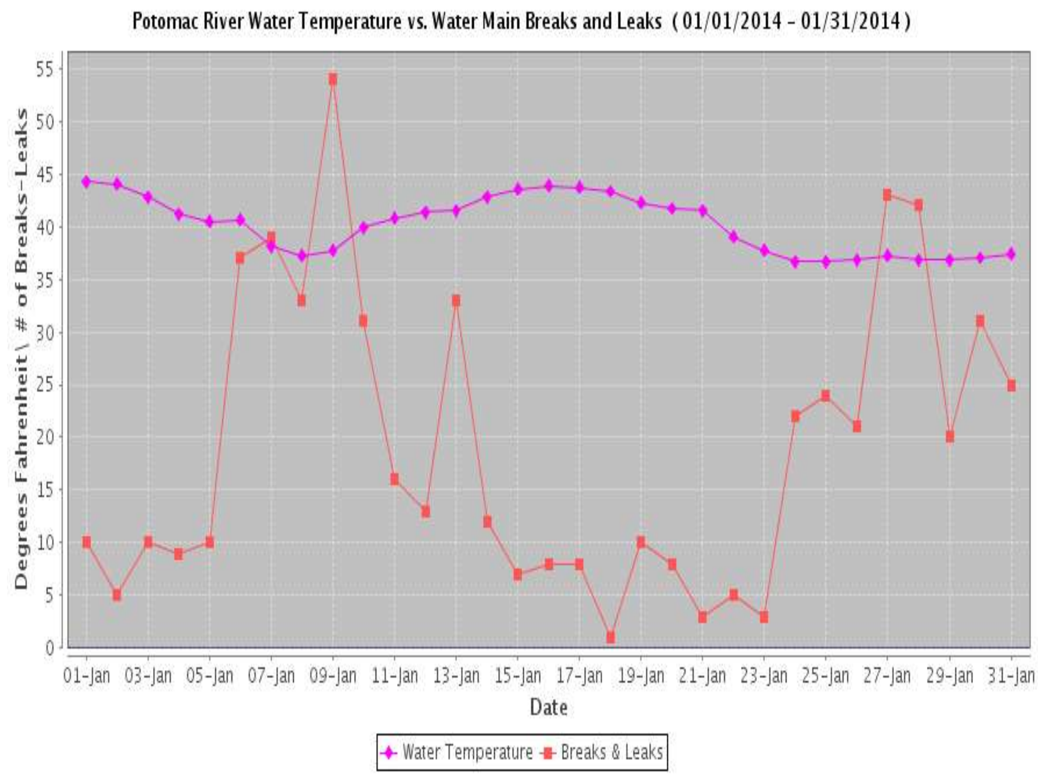


Figure B.1 Data Set #2 in graphical form (WSSC)

Appendix C

COPY RIGHT PERMISSION

C.1 Copy Right Permission from Published Journal

Figure C.1 shows the copy right permission showing the license number obtained from Construction and Building Journal. The published paper covers some materials of this dissertation.

12/1/2016 RightsLink Printable License

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 01, 2016

This Agreement between Farzana Atique ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4000340255694
License date	Dec 01, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Construction and Building Materials
Licensed Content Title	Using copula method for pipe data analysis
Licensed Content Author	Farzana Atique,Nii Attoh-Okine
Licensed Content Date	1 March 2016
Licensed Content Volume Number	106
Licensed Content Issue Number	n/a
Licensed Content Pages	9
Start Page	140
End Page	148
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	ANALYSIS OF URBAN PIPE DETERIORATION USING COPULA METHOD
Expected completion date	Jan 2017
Estimated size (number of pages)	151

Figure C.1 License for copy right