Chapter 1.
Introduction

1.1 Motivation

Because of its importance to the society and industrial growth, the need for adequate water supply systems draws attention among federal and state agencies and water utilities. Water utilities are concerned about large number of main breaks and the resulting direct costs and intangible costs, such as consumer productivity and economic loss, damage to nearby property due to flooding, disruptions to traffic, and possible contamination through cracked or broken pipes. Pipes in an aging water distribution system are unable to carry intended flows at required pressure heads due to tuberculation. They also lose strength due to corrosion. To make matters worse, ever expanding cities continually place growing demands on these pipes. These problems directly contribute to frequent breaks and resulting repair and rehabilitation costs. Among various remedies for water main deterioration problem, replacement represents large capital investment. Therefore, it is important that a decision regarding a pipe replacement be made in the most economical manner.

Decisions regarding continual repair or replacement involve assessing the failure mechanism and the associated contributing factors. The analysis should be focused at an individual pipe level. Assessing the failure mechanism of a pipe requires data related to each break event. The degree of deterioration of a pipe can be assessed most accurately by actual visual inspection and a structural integrity test in the field while the pipe is still in operation. However, this kind of exercise is simply impossible, especially if we are interested in every pipe in a system. In this case historical records and information related to each break event become the best resort.

The optimal replacement analysis of pipes is dependent on pipe break prediction models. Historic data of pipe breaks is used in obtaining parameters of the prediction model, and future break rate is predicted. If an estimator that can pinpoint the replacement time after which it is no longer economical to repair is developed, such an estimator will be of great use to the water utilities. Furthermore, the usual practice in pipe replacement, which is mostly by experience, would transform to more economical one and, correspondingly, communities will experience more reliable service.

The optimal replacement time is obtained as the time at which the pipe break rate equals or exceeds the threshold (economical) break rate. The pipe break rate is a time dependent phenomenon. New pipes clearly have smaller break rates. However, if the pipes are in a corrosive environment they deteriorate much faster. Similarly, loading conditions and pipe material control the deterioration rate. Unfortunately, the modeling of the effects of environmental factors such as loads, corrosion, and pipe material on pipe deterioration has remained elusive. In this research an approach that makes use of the directly available pipe failure records will be pursued.

Consider a buried pipe. It has a natural deterioration rate due to normal loads and material decay. This deterioration is considered as a time dependent background process. This process is accelerated by environmental factors such as soil corrosivity, thermal regimes, and increased loads.

Considering the pipe deterioration process as the product of a time dependent break arrival process and deterioration by environmental factors, it is possible to build a model that takes advantage of the directly available break data and accounts for the accentuating factors. In this regard, two models namely, the proportional hazards model and the proportional intensity model will be considered.

1.2 Research Goals

The specific objectives of the research are as follows:

a. Develop an economical, threshold break rate criterion beyond which it is no longer economical to repair a pipe in a water distribution system

b. Develop pipe break prediction models that make use of readily available pipe break record as well as account for the effects of governing environmental factors.

c. Develop a computer program to solve large-scale pipe replacement problems using the developed methodologies in this research and illustrate its application to a real world problem.

1.3 Organization of the Dissertation

The dissertation is organized as follows. Chapter 2 provides a comprehensive literature survey on pipe replacement analysis. Chapter 3 presents the derivation of the threshold break rate equation. The utility of the threshold break rate is illustrated with the aid of a number of examples. Also, design aids in the form of a table and curves are given in Chapter 3. Chapter 4 concentrates on the theoretical discussions of the proportional hazards and Weibull intensity models to pipe break data as forecasting tools for future breaks. Chapter 5 describes failure characteristics of the pipes in the study area and the database processing to transform the raw database into a usable input for the optimal replacement analyses developed in this research. Chapter 6 presents the applications of the proportional hazards and the Weibull intensity models to the case study area water distribution system. Application of the general break regression model is presented in Chapter 7. Chapter 8 contains the major findings of this study as well as recommendations for future research. Much of the mathematical development is provided in the form of Appendices (A through C) and the computer programs are given in Appendix D.

Chapter 2.
Literature Review

Aging, deteriorating water distribution systems are a crucial infrastructure problem. Pipes are heavily tuberculated, weakened by corrosion, too small to meet increasing demands, and unable to provide required pressure heads. These problems directly contribute to the increased replacement/repair costs, storage tank costs, and pumping costs. Because of the high costs of correcting these problems, it is necessary to identify and prioritize the failure prone pipes for replacement. There is a need for methods that can help in progressive system repair and replacement subject to budgetary constraints. Water utilities have found from experience that "do nothing until a system component fails" is not the best approach since it increases costs and leads to customer dissatisfaction and potential environmental problems. If "delaying" is considered good economic strategy, the "timing" of preventive maintenance becomes crucial for system maintenance and operation. For example, many municipalities coordinate the replacement of aged pipe with the resurfacing of pavement. For any given pipe, at a given time, there are three possible alternatives: do nothing, repair, or replace. The major factor in the decision process is the condition of a pipe in its current situation.

Unfortunately, there is not yet a comprehensive method for prioritizing failure prone pipes for replacement or repair. The available methods can be put in the following categories: (1) Deterioration Point Assignment (DPA) schemes and Depreciation/Break Even Analyses, (2) Regression and Failure Probability Methods, and (4) Mechanistic Methods.

2.1 Deterioration Point Assignment Schemes

In the deterioration point assignment schemes (DPA), a set of factors involved in pipe failures is identified. These may include age of pipe, pipe material, pipe size, type of soil, location, water pressure, discoloration and odor problems and history of previous breaks. Ordinal descriptions of these factors are associated with numerical failure score. For any pipe a total failure score is obtained by adding the failure scores of the factors for that pipe. If the total failure score exceeds a threshold value the pipe is considered a candidate for replacement/repair (Weston, Inc., 1997). The discriminatory power of the scheme is clearly limited and becomes an issue if there are other pipes competing for limited funding. Also, it is a here and now assessment and lacks the predictive power which is crucial for future course of action.

2.2 Depreciation/Break Even Analyses

The depreciation and break even analyses are cost based methods. In the depreciation analysis, the pipe is considered an asset facing depreciation. The repairs and maintenance constitute the depreciation charges. When the book value of the pipe is exhausted, the pipe is considered to have reached obsolescence and should be replaced. Clearly, this procedure does not account for proactive replacement strategies to minimize total cost and improved service. Another approach is to deposit certain sum at an interest rate and its compounded value should equal to the future repair and replacement costs.

2.3 Regression and Failure Probability Methods

The regression and failure probability methods are related to the DPA scheme in that they build on the same deterioration factors but bring in a predictive capability by assessing the probability of survival. Comprehensive reviews are given in O'Day et al. (1982) and Mays (2000).

Shamir and Howard (1979) applied regression analysis to obtain a relationship for the breakage rate of a pipe as a function of time. This relationship was used to find the optimal timing of pipe replacement to minimize the total cost of repair and replacement. It is clear that any error in the predictive model will alter the replacement time significantly. The proposed threshold break rate estimator (Brk_{th}) of this research assesses the break rate, solely based on current cost data; no prediction involved. It merely answers the question what is the most economically sustainable break rate to keep a pipe in the system at the present day cost levels. Then, we search for all pipes that exceed that sustainable rate as candidates for replacement. An error in Shamir and Howard's (1979) derivation is rectified in Appendix C.

Walski and Pelliccia (1982) subscribed to the idea of the threshold break rate. They adopted Shamir and Howard's (1979) model for predicting break rates. They derived an optimal replacement time estimator by setting the total repair costs to be equal to the replacement cost. It is not clear why such a criterion for replacement is valid.

Male et al (1990) described a procedure in which an arbitrary threshold break rate is fixed. The analysis involved consideration of five alternatives: (1) replace after one or more breaks, (2) replace after two or more breaks, (3) replace after three or more breaks, (4) replace after four or more breaks, and (5) do nothing approach. Alternative 2 turned out to be the most aggressive policy. Male et al also indicated that the choice of alternative is sensitive to the discount rate used in the calculation. A higher rate leads to a less aggressive policy and vise versa. The proposed optimal replacement threshold break rate removes the arbitrariness embedded in the choice of Male et al's break rate. It also clearly shows the trade-off between the repair and replacement costs and the role of the discount rate. Male et al. drew their conclusions from a number of simulation runs whereas the present work yields a closed form analytical model which clearly illustrates the role of the various factors.

Clark et al. (1982) suggested a model that combines two equations, one to predict the time elapsed until the first break occurred and the second to predict the number of subsequent breaks which were assumed to grow exponentially over time in an attempt to account for the relative impacts of various external agents. These equations had coefficients of determination (R^2) of 0.23 and 0.47, respectively. While Clark et al.'s (1982) procedure is a significant improvement in predicting pipe breaks, it does raise some concerns because of the low values of the coefficients of determination. Clark et al (1982) have made the following observations: only a subset of pipes have recurrent repairs; the time to first repair is quite long, typically about fifteen years; the time between repairs becomes shorter as pipes get older; large diameter pipes tend to have fewer problems; and industrial development in general results in more repairs.

Kettler and Goulter (1985) estimated simple regressions for the number of breaks versus diameter and time for cast iron and for asbestos-concrete mains in Winnipeg, Canada. Their estimates showed strong inverse linear correlations between failures and diameter (0.0625 less annual failures/km of main with each cm of larger pipe diameter, for diameters between 10 and 30 cm). The correlation was 0.96 ($R^2 = 0.92$). Comparisons with regressions on New York, Philadelphia and St. Catherines, Canada showed about 1/3 of decrease in failure rates per cm of increase in diameter in these three cities in which failures were found to increase linearly with time.

Mavin (1996) provided a review of the failure models in the literature. Mavin also pointed out the need to filter the data before constructing a failure model. It was suggested not to include breaks that occurred within three years of installation and six months of a previous break. Based on the filtered data, a set of regression equations was constructed for number of failures over a time period and time interval between breaks. Marks et al. (1985) used multiple regression techniques to establish that the variables affecting the pipe breakage rate were pipe diameter, length of pipe section, age, pressure, type, soil corrosivity, intensity of land development, number of previous breaks, time to the second break, and period of installation. Prioritizing of pipe rehabilitation considerations was subsequently discussed in the literature.

Andreou et al. (1987a, b) applied the proportional hazards model to predict failure probabilities of pipes in the early stages of deterioration and a Poisson-type model for the later stages of pipe deterioration. The basic idea of this model is to estimate a survivor function for each individual pipe, that will provide the probability for that pipe's survival beyond a future time period, given a set of risk factors. The model provides the hazard function as a product of the baseline hazard function dependent only on time and a scaling factor made up of external variables such as pipe diameter, length, soil type, and land use. It is because of the scaling factor that the proportional hazards model has a wide appeal.

Deb et al., (1997) discussed a probabilistic model called KANEW to estimate miles of pipes to be replaced on an annual basis. The primary objective of KANEW is to provide water utilities with a tool to develop their long-range pipe rehabilitation and replacement strategies. Based on the historical inventory of the water main and the estimated life-span data, KANEW predicts miles of different categories of pipes to be repaired and replaced on an annual basis. It is, however, not intended to provide location-specific repair and replacement information. The model uses the actual water main inventory, with the pipes categorized according to their age, material, diameter, and bedding quality. For each category 100^{th}, 50^{th} and 25^{th} percentile ages are obtained either by expert opinion or by an analysis. These percentiles are utilized to obtain the three parameters of the Herz probability density function from which the survival probabilities are obtained. These survival probabilities are used to obtain the expected survivors or its complement of non-survivors per year, which are to be renewed.

2.4 Optimization Methods

Kleiner et al (1998) considered repair costs in terms of improving the carrying capacity of an aging pipe by relining. An aging Hazen-Williams "C_{HW}" coefficient will increase the head loss and reduce pressure. Kleiner et al. incorporated mass balance, energy, and head-loss with time varying $C_{HW,}$ and minimum pressure constraints in a rolling time horizon to find optimal replacement times as relining costs become non-optimal. Their objective is to minimize total costs of relining and associated replacement cycles in an infinite time frame.

Karaa et al. (1987) described an appropriate technique for time phased replacement of failure prone pipes. In this procedure pipes that are to be rehabilitated, replaced, or constructed, are grouped into so-called bundles, based upon similarities in characteristics and criticality measures. A failure model may be used to delineate such bundles, based upon their role in determining the reliability of the system. Portions of bundles to be replaced subject to annual budget constraints can be obtained through a linear program. Sensitivity analyses can be performed to assess the variations in numbers of pipes to be replaced/repaired each year against budget changes. Such analyses also point to the budgetary requirements for various anticipated levels of upgrading the system. This information is crucial for proper planning, and assists in understanding the gravity of the problem. For example, to ensure a reliable network by the year 2000, the city of Boston should rehabilitate or replace 11 miles of pipe annually. In the same vein PPK consultants (1993) provided a comprehensive assessment procedure considering the water distribution system as a critical resource.

Su et al. (1987) and Wagner et al. (1988) addressed an alternative form of a reliability constraint based on the probability of satisfying nodal demands and pressure head requirements under various network failure configurations. Duan et al. (1990) considered the optimization and reliability of pumping systems in a network (also see Lansey and Awumah, 1994; Park and Liebman, 1993; and Pezeshk and Helweg, 1996). Although these models do not directly address the causal mechanisms of failures in pipes, they provide a framework for further economic analysis. Issues and methods related to pipe network optimization are fully covered in Loganathan et al (1990), Loganathan et al (1995) and Sherali et al (1998).

In addition to the three methods described so far, a number of researchers developed mechanistic methods to model pipe failure phenomena. The mechanistic methods model the corrosion (Romanoff (1957), Rossum (1969), Kumar et al (1984, 1986, 1987); Basalo, (1992)), temperature induced stresses (US Pipe and Foundry Co., (1962), Wedge, (1990), and Habibian, (1994)), frost load (Cohen and Fielding, (1979), Fielding and Cohen (1988)) and other failure processes using the underlying physical principles. While these methods help to make the failure processes more understandable, the predictive capability has to be brought in either through a correlation analysis or through a probabilistic analysis by considering the parameters/variables to be random.

To summarize, the basic foundations of pipe break prediction are established by Shamir and Howard (1979) through a simple regression equation and certain fundamental characteristics of pipe failures are given by Clark et al. (1982) and Kettler and Goulter (1985). Andreou (1987 a, b) and Deb et al. (1997) have provided methods for estimating the survival probabilities.

Chapter 3.
Economically Sustainable Threshold Break Rate

In this chapter, a new methodology for optimal pipeline replacement is presented. The methodology provides the threshold break rate equation that gives the critical break rate for optimal replacement of a pipe. In addition, the methodology shows how the threshold break rate is related to failure intensity, also known as ROCOF (rate of occurrence of failure), and hazard rate functions. Based on the relationships, equations to obtain future optimal pipe replacement times are developed.

To illustrate the methodology, four examples are presented by using arbitrarily assumed pipe break times. The assumed break times are generated based on the general pipe failure characteristic that times between breaks decrease as a pipe ages. Figures and Tables are developed for easy application of the proposed methodology by the end users. Suitable examples are also provided.

3.1 Development of the Threshold Break Rate Equation

In this section, a criterion for replacing a pipe is developed. At the time of the nth break, a decision has to be made whether to replace the pipe at a cost of F_n or to repair it at a cost of C_n. The scenario also implies that for the previous (n - 1) breaks, only repairs have been performed. If we assume that the pipe will be replaced at the time of nth break, t_n, we can write the present worth of the total cost of the pipe as

$$T_n = \sum_{i=1}^{n} \frac{C_i (1+i)^{t_i}}{(1+R)^{t_i}} + \frac{F_n (1+i)^{t_n}}{(1+R)^{t_n}} \tag{3.1.1}$$

in which
R = annual interest rate (1/year)
i = annual inflation rate (1/year)
t_i = time of ith break measured from the installation year (year)
t_n = time of nth break measured from the installation year (year)
C_i = repair cost of ith break (\$)
F_n = replacement cost at time, t_n (\$)
T_n = total cost at time ‘0’ (present worth) (\$)

When a pipe is new, it experiences very few breaks. An old pipe experiences more breaks under the same trench and load conditions. Therefore, the combination of time interval between breaks, relatively smaller repair cost, and a generally large replacement cost leads to the existence of a “U” shaped present worth of the total cost curve over time. The analysis that leads to the derivation of the threshold break rate seeks to find the time of the minimum present worth total cost.

For the total cost T_n at time t_n to be a minimum, assuming a unimodal function, it must satisfy the condition,

$$T_{n-1} > T_n < T_{n+1} \tag{3.1.2}$$

For $T_n < T_{n-1}$ consider

$$T_{n-1} = \sum_{i=1}^{n-1} \frac{C_i (1+i)^{t_i}}{(1+R)^{t_i}} + \frac{F_{n-1} (1+i)^{t_{n-1}}}{(1+R)^{t_{n-1}}} \tag{3.1.3}$$

From Eqs. (3.1.1) and (3.1.3) we obtain

$$T_{n-1} - T_n = \frac{F_{n-1}(1+i)^{t_{n-1}}}{(1+R)^{t_{n-1}}} - \frac{C_n(1+i)^{t_n}}{(1+R)^{t_n}} - \frac{F_n(1+i)^{t_n}}{(1+R)^{t_n}} \tag{3.1.3}$$

For $T_{n-1} - T_n > 0$ we have

$$\frac{F_{n-1}(1+i)^{t_{n-1}}}{(1+R)^{t_{n-1}}} > \frac{C_n(1+i)^{t_n}}{(1+R)^{t_n}} + \frac{F_n(1+i)^{t_n}}{(1+R)^{t_n}} \tag{3.1.5}$$

which is written as

$$F_{n-1} > \frac{C_n + F_n(1+i)^{t_n - t_{n-1}}}{(1+R)^{t_n - t_{n-1}}} \tag{3.1.6}$$

which becomes

$$\frac{(1+R)^{t_n - t_{n-1}}}{(1+i)^{t_n - t_{n-1}}} > \frac{C_n + F_n}{F_{n-1}} \tag{3.1.7}$$

and upon simplification

$$t_n - t_{n-1} > \frac{\ln\left(\frac{C_n}{F_{n-1}} + \frac{F_n}{F_{n-1}}\right)}{\ln\left(\frac{1+R}{1+i}\right)} \tag{3.1.8}$$

Recognizing $t_n - t_{n-1}$ is the time between (n – 1)th and nth breaks or time interval for the occurrence of one break at time t_n we obtain the threshold break rate, $Brk_{th,1}$, as the inverse of Δt_n where $\Delta t_n = t_n - t_{n-1}$. That is the threshold break rate is defined as

$$Brk_{th,1} = \text{break rate between subsequent breaks} = \frac{1}{t_n - t_{n-1}} = \frac{1}{\Delta t_{n-1}} \tag{3.1.9}$$

Therefore, the threshold break rate is expressed as

$$Brk_{th,1} < \frac{\ln\left(\frac{1+R}{1+i}\right)}{\ln\left(\frac{C_n}{F_{n-1}} + \frac{F_n}{F_{n-1}}\right)} \tag{3.1.10}$$

By considering $T_{n+1} > T_n$ for a minimum we obtain

$$Brk_{th,2} > \frac{\ln\left(\frac{1+R}{1+i}\right)}{\ln\left(\frac{C_{n+1}}{F_n} + \frac{F_{n+1}}{F_n}\right)} \tag{3.1.11}$$

We define the optimal threshold break rate to be Brk_{th} given by

$$Brk_{th} = \text{break rate between subsequent breaks} = \frac{1}{t_{n+1} - t_n} = \frac{1}{\Delta t_n} \tag{3.1.12}$$

and

$$Brk_{th} = \frac{\ln\left(\frac{1+R}{1+i}\right)}{\ln\left(\frac{C_{n+1}}{F_n} + \frac{F_{n+1}}{F_n}\right)} \tag{3.1.13}$$

If inflation rate can be ignored Eq. (3.1.13) is expressed as

$$Brk_{th} = \frac{\ln(1+R)}{\ln\left(\frac{C_{n+1}}{F_n} + \frac{F_{n+1}}{F_n}\right)} \tag{3.1.14}$$

Now from observed data for any given pipe we can derive a current break rate. Whenever the current break rate, Brk_{cur} equals or exceeds Brk_{th}, the pipe should be replaced. In other words, the condition for a pipe replacement at current time is expressed as

$$Brk_{cur} \geq Brk_{th} \tag{3.1.15}$$

Eq. (3.1.13) is chosen instead of Eq. (3.1.10) for the threshold break rate because, if Eq. (3.1.10) is used, current break rate would pass another replacement criterion which is Eq. (3.1.13). Therefore, one should consider replacing a pipe of interest whenever the current break rate of the pipe equals or exceeds the threshold break rate which is Eq. (3.1.13) or Eq. (3.1.14) if inflation is ignored.

The importance of the terms 'at current time' should be emphasized. If the current break rate is less than the threshold break rate and one wants to know when in the future the pipe needs to be replaced, he/she must be able to predict future break rates of the pipe and compare the predicted future break rates with the threshold break rate.

In the following sections a new methodology is presented for predicting pipe break rate as a function of time. A few related fundamental details are given first. They establish the empirical rate of occurrence of failure (ROCOF) and hazard function concepts and provide the equivalence between them and the threshold break rate given in Eq. (3.1.14). Since the repair and replacement costs consist mainly of pipe material and labor costs, the repair and replacement costs of pipes would not greatly be affected by inflation. Therefore, It is assumed that inflation can be ignored and Eq. (3.1.14) is used in the optimal replacement analyses developed in this dissertation.

3.2 Relationship between the Threshold Break Rate, Intensity and Hazard Functions

A break rate function is defined generally as

$$r(t) = \lim_{\Delta t \to 0} \frac{\text{No. of breaks in } (t, t+\Delta t]}{\Delta t} \tag{3.2.1}$$

Eq. (3.2.1) also can be expressed as the derivative of the cumulative number of breaks function $N_c(t)$, that is

$$r(t) = \frac{dN_c(t)}{dt} \tag{3.2.2}$$

Eq. (3.2.2) is in general called the rate of occurrence of failure (ROCOF) function. From Eq. (3.2.1) an empirical break rate in $(t_n, t_n+\Delta t_n]$, where t_n is the nth break time, and $\Delta t_n = t_{n+1} - t_n$, can be obtained as follows.

$$r(t_n) = \frac{N_c(t_{n+1}) - N_c(t_n)}{\Delta t_n} \tag{3.2.3}$$

Because t_{n+1} and t_n are successive failure times, we have $N_c(t_{n+1}) - N_c(t_n) = 1$ and the empirical break rate becomes

$$r(t_n) = \frac{1}{\Delta t_n} \tag{3.2.4}$$

which is the same as the threshold break rate given in (3.1.12) and

$$r(t_n) = Brk_{th} \tag{3.2.5}$$

The hazard function called as the hazard rate or the probability of instantaneous failure rate is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr[t < T_f \le t + \Delta t \mid T_f > t]}{\Delta t} \tag{3.2.6}$$

where T is a failure time random variable. The hazard function expresses the propensity to fail in the next small interval of time, given survival to time t. That is, for small Δt,

$$h(t) \cdot \Delta t \approx \Pr[t < T_f \le t + \Delta t \mid T_f > t] \tag{3.2.7}$$

The hazard function originally applies to non-repairable systems in which a failure implies the death of a system and is allowed only once in its lifetime. However, to apply the hazard function, we assume here that a system gains a new life after each repair. Similar to the case of the break rate function, the estimate of the hazard function at time t_n is expressed as

$$h_e(t_n) = \frac{N_c(t_{n+1}) - N_c(t_n)}{N_s(t_n)\Delta t_n} \tag{3.2.8}$$

where the numerator is interpreted as the number of deaths in Δt_n and $N_s(t_n)$ is the number of survivors at time t_n. Now consider the situation in which we are continuously monitoring a pipe for every break. In such a case

$N_c(t_{n+1}) - N_c(t_n) = 1$

and

$N_s(t_n) = 1$.

Therefore, the estimate of the hazard function at time t_n is

$$h_e(t_n) = \frac{N_c(t_{n+1}) - N_c(t_n)}{N_s(t_n)\Delta t_n} = \frac{1}{\Delta t_n} \tag{3.2.9}$$

which is the hazard rate of the lone survivor. Eq. (3.2.9) has the same definition as the threshold break rate shown in Eq. (3.1.12). Therefore, the optimal replacement time of a pipe can be obtained by solving the equation

$$h(t) = Brk_{th} \tag{3.2.10}$$

for time, t.

The application of the break rate concept along with the threshold break rate idea is shown in the next two sections. First, the break rate is assumed to be exponential and then a linear case is considered.

3.3 Exponential Break Rate Model

As an example, consider the equation (Shamir and Howard, 1979)

$$N(t) = N(t_0)e^{A(t-t_0)} \tag{3.3.1}$$

where

N(t) = number of breaks per 1000 ft length of pipe in year t

t = time in years

t_0 = base year for the analysis(pipe installation year, or the first year for which data are available)

A = growth rate coefficient (1/year).
In our notation N(t) is the break rate (that is, number of breaks/year at year t). Therefore, by setting N(t) = Brk_{th} in Eq. (3.3.1) we obtain

$$N(t) = N(t_0)e^{A(t-t_0)} = \frac{\ln(1+R)}{\ln\left(\frac{C_{n+1}}{F_n} + \frac{F_{n+1}}{F_n}\right)} \qquad (3.3.2)$$

Assuming $F_{n+1} \approx F_n$ at a rapidly deteriorating stage in which breaks occur in quick succession, we have

$$N(t_0)e^{A(t-t_0)} = \frac{\ln(1+R)}{\ln\left(1 + \frac{C_{n+1}}{F_n}\right)} \qquad (3.3.3)$$

Further for small x, putting
$\ln(1+x) = x$
we obtain

$$N(t_0)e^{A(t-t_0)} = \frac{\ln(1+R)F_n}{C_{n+1}} \qquad (3.3.4)$$

from which

$$t^* = t_0 + \frac{1}{A}\ln\left[\frac{\ln(1+R)F_n}{N(t_0)C_{n+1}}\right] \qquad (3.3.5)$$

which is an algebraic solution with respect to t for Eq. (3.3.4). A rigorous proof that avoids the assumptions built in Eqs. (3.3.3) and (3.3.4) and corrects Shamir and Howard's (1979) differentiation based approach is given in Appendix C.

3.4 Linear Break Rate Model

The linear model has following form:

$$N(t) = B + A(t - t_0) \qquad (3.4.1)$$

where
N(t) = number of breaks along the length of a defined pipe in year t
t = time in years
t_0 = base year for the analysis(pipe installation year, or the first year for which data are available)
A = growth rate coefficient (1/year)
B = coefficient of intercept
By setting the current break rate and the threshold break rate to be equal we obtain the optimal replacement time. Therefore, for

$$N(t) = Brk_{th} \qquad (3.4.2)$$

we have

$$B + A(t - t_0) = \frac{\ln(1+R)}{\ln\left(\frac{C_{n+1}}{F_n} + \frac{F_{n+1}}{F_n}\right)} \qquad (3.4.3)$$

from which

$$t^* = t_0 + \frac{1}{A}\left(\frac{\ln(1+R)}{\ln\left(\frac{C_{n+1}}{F_n} + \frac{F_{n+1}}{F_n}\right)} - B\right). \tag{3.4.4}$$

which is an algebraic solution with respect to t for Eq. (3.4.3).

3.5 Poisson Rate of Occurrence of Failure (ROCOF)

A ROCOF constant over time is used in homogeneous Poisson process. One of the properties of the homogeneous Poisson process is that the times between successive failures are independent random variables, each having an exponential density with parameter l, so that

$$\Pr\{\text{Failure Time} > x\} = e^{-lx}, \quad 0 < x < \infty. \tag{3.5.1}$$

Therefore, the homogeneous Poisson is called a stochastic point process with constant failure rate (l).

The non-constant ROCOF over time is used in nonhomogeneous Poisson process (NHPP). Consider an NHPP with time-dependent ROCOF $l(t)$, then the number of failures in the time interval $(t_1, t_2]$ has a Poisson distribution with mean

$$\int_{t_1}^{t_2} l(t)dt. \tag{3.5.2}$$

Thus the probability of no failures in $(t_1, t_2]$ is

$$\exp\left\{-\int_{t_1}^{t_2} \lambda(t)\,dt\right\} \tag{3.5.3}$$

Taking $l(t) = l$ for all t just gives the homogeneous Poisson process with constant ROCOF l. By choosing a suitable parametric form for $l(t)$, one obtains a flexible model for the failures of a repairable system.

Suppose we observe a system for the time interval $(0, t_0]$ with failures occurring at $t_1, t_2, \ldots, t_n$. The joint probability of observing no failures in $(0, t_1)$, one failure in $(t_1, t_1 + \Delta t_1)$, no failures in $(t_1 + \Delta t_1, t_2)$, one failure in $(t_2, t_2 + \Delta t_2)$ and so on up to no failures in $(t_n + \Delta t_n, t_0)$ is for small $\Delta t_1, \Delta t_2, \ldots, \Delta t_n$

$$\left\{\exp\left(-\int_0^{t_1} l(t)dt\right)\right\} l(t_1)\Delta t_1 \left\{\exp\left(-\int_{t_1+\Delta t_1}^{t_2} l(t)dt\right)\right\} l(t_2)\Delta t_2 \ldots \left\{\exp\left(-\int_{t_n+\Delta t_n}^{t_0} l(t)dt\right)\right\} \tag{3.5.4}$$

.

Dividing through by $\Delta t_1 \Delta t_2 \ldots \Delta t_n$ and letting $\Delta t_i \to 0$ (i = 1, 2, …, n) gives the likelihood function.

$$L = \left\{\prod_{i=1}^{n} l(t_i)\right\}\left\{\exp\left(-\int_0^{t_0} l(t)\,dt\right)\right\} \tag{3.5.5}$$

and the log-likelihood is thus

$$l = \sum_{i=1}^{n} \log \lambda(t_i) - \int_0^{t_0} \lambda(t)dt \tag{3.5.6}$$

In this research we use Weibull form of ROCOF function due to its wide use in repairable systems modeling. Weibull intensity function can be expressed as

$$l(t)=g \cdot d \cdot t^{d-1} \tag{3.5.7}$$

Therefore, for Weibull intensity case

$$l = n \log g + n \log d + (d-1)\sum_{i=1}^{n} \log t_i - g t_0^{\,d} .$$

Therefore, maximum likelihood estimates for g and d are obtained as

$$\hat{d} = \frac{n}{n \log t_0 - \sum_{i=1}^{n} \log t_i} . \tag{3.5.8}$$

and

$$\hat{g} = \frac{n}{t_0^{\,\hat{d}}} . \tag{3.5.9}$$

Here, the parameter g is identical to $1/h^b$ if the form of Weibull density is used as below

$$l(t) = \frac{b}{h}\left(\frac{t}{h}\right)^{b-1}, t > 0, \quad b > 0, \quad h > 0 . \tag{3.5.10}$$

Once the parameters of the model are obtained, we can use the threshold break rate equation to obtain optimal replacement time of each pipeline by using estimated Weibull intensity function. Optimal replacement time can be obtained when the intensity reaches threshold break rate. Thus,

$$\frac{b}{h}\left(\frac{t}{h}\right)^{b-1} = \frac{\ln(1+R)}{\ln\left(1+\frac{C}{F*L}\right)} . \tag{3.5.11}$$

in which $b = d$ and $h = g^{-1/b}$ in Eq. (3.5.7).
Therefore, optimal replacement time is obtained as

$$t^* = \left(\frac{\ln(1+R)}{\ln\left(1+\frac{C}{F*L}\right)}\frac{h^b}{b}\right)^{\frac{1}{b-1}} = \left(\frac{\ln(1+R)}{gd\,\ln\left(1+\frac{C}{F*L}\right)}\right)^{\frac{1}{d-1}} . \tag{3.5.12}$$

3.6 Application to Controlled Data

In this section four examples that use the same controlled break times are given. Example 1 relates the threshold break rate to empirical ROCOF function to obtain the optimal replacement time. In example 2, the time varying discrete break rate, $l(t) = 1/E(\Delta t)$, is considered. In example 3, a Weibull ROCOF function is fitted to a subset of the data given in Table 3.1. It is done this way because of the lack of flexibility of the Weibull ROCOF curve to cope with the changing nature of $l(t)$. Example 4 shows a table from which a practitioner can easily obtain the threshold break rate by using the cost ratio and the length of a pipe.

Table 3.1 Controlled Break Time Data
(n = break number, T_n = time at nth break (years))

n	T_n
1	10.0000
2	18.0000
3	24.4000
3	29.5200
5	33.6160
6	36.8928
7	39.5142
8	41.6114
9	43.2891
10	44.6313
11	45.7050
12	46.5640
13	47.2512
14	47.8010
15	48.2408
16	48.5926
17	48.8741
18	49.0993
19	49.2794

Example 1: Given the average times of failure for the nth break in Table 3.1, an annual interest rate of 7.5%, a constant repair cost of \$3,000 and a constant replacement cost of \$100,000, compute the optimal replacement time.

Solution:
The break times satisfy the equation $T(n) = 50(1-0.8^n)$. From this equation we obtain the break rate as $dn/dt = 0.08963/0.8^n$ which depends on the break number. From our analysis we know that the break rate $dn/dt = 0.08963/0.8^n = Brk_{th} = \dfrac{\ln(1+R)}{\ln\left(\dfrac{C_{n+1}}{F_n} + \dfrac{F_{n+1}}{F_n}\right)} = 1.844$

which gives the optimal break number as $n^* = 14$. For this particular data, the replacement time is 47.8 or 47 years.

Example 2. Given the average times of failure for the nth break in Table 3.1, an annual interest rate of 7.5%, a constant repair cost of \$3,000 and a constant replacement cost of \$100,000, compute the optimal replacement time in terms of the hazard function.

Solution

Table 3.2 shows the times between breaks obtained from the successive differences of the break times. Assuming the inter-break times follow exponential distributions, we obtain the exponential probability density parameter given in column 4 which is the reciprocal of the expected inter-break time. It is also true that the hazard function for exponential probability distribution is its parameter *Alpha*. As we have already established that the hazard function of a lone survivor equals the threshold break rate, we obtain $Brk_{th} = \dfrac{\ln(1+R)}{\ln\left(\dfrac{C_{n+1}}{F_n} + \dfrac{F_{n+1}}{F_n}\right)}$ = 1.844. The optimal replacement time from Table 3.2 is around 47.8 which belongs to the hazard 1.819 at the break number 14. Figure 3.1 shows a plot of the empirical break rate between breaks and empirical hazard rate between breaks (*Alpha*(n)).

Table 3.2 Break Times and Corresponding Hazard
($E(T_n)$ = expected time between breaks (year), *Alpha*(n) = 1/$E(T_n)$ (1/year))

n	T_n	$E(T_n)$	Alpha(n)
1	10.0000	10.0000	0.1000
2	18.0000	8.0000	0.1250
3	24.4000	6.4000	0.1563
3	29.5200	5.1200	0.1953
5	33.6160	3.0960	0.2441
6	36.8928	3.2768	0.3052
7	39.5142	2.6214	0.3815
8	41.6114	2.0972	0.4768
9	43.2891	1.6777	0.5960
10	44.6313	1.3422	0.7451
11	45.7050	1.0737	0.9313
12	46.5640	0.8590	1.1642
13	47.2512	0.6872	1.4552
14	47.8010	0.5498	1.8190
15	48.2408	0.4398	2.2737
16	48.5926	0.3518	2.8422
17	48.8741	0.2815	3.5527
18	49.0993	0.2252	3.4409
19	49.2794	0.1801	5.5511

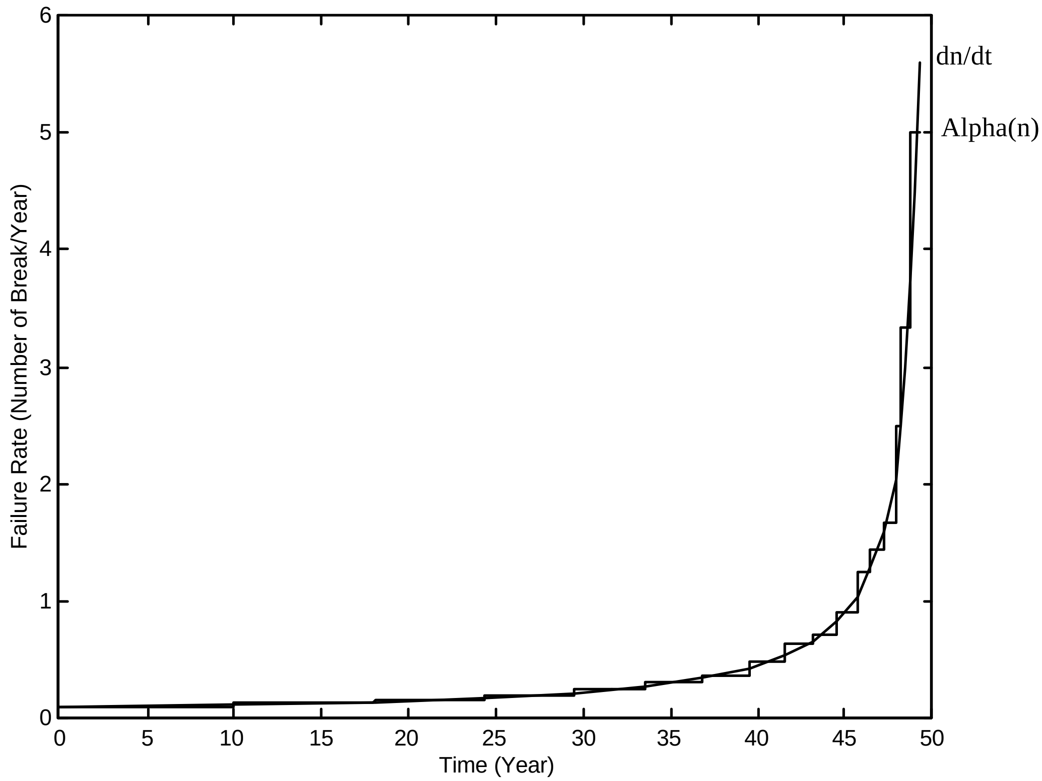


Figure 3.1 Plot of dn/dt and *Alpha*(n) for the Controlled Data.

Table 3.3 Controlled Break Times for Weibull ROCOF

Time	Time – 41.6114	$E(T_n)$	Alpha(n)
43.2891	1.6777	1.6777	0.5960
44.6313	3.0199	1.3422	0.7451
45.7050	4.0936	1.0737	0.9313
46.5640	4.9526	0.8590	1.1642
47.2512	5.6398	0.6872	1.4552
47.8010	6.1896	0.5498	1.8190
48.2408	6.6294	0.4398	2.2737

Example 3: Given the average times of failure for the nth break in Table 3.3, an annual interest rate of 7.5%, a constant repair cost of $3,000 and a constant replacement cost of $100,000, compute the optimal replacement time using Weibull ROCOF function.

Solution

First of all in Table 3.3, the break times are considered after the occurrence of the eighth break at 41.6114. The second column in Table 3.3 gives the break times since the eighth break. Using the break times after the sixth break, we have d = 2.212 from Eq. (3.5.8) and g = 0.1066 from Eq. (3.5.9) for n = 7 and t_0 = 6.6294. Therefore, the optimal replacement time from Eq. (3.5.12) is 5.455 years, which is close to the results of

Examples 1 and 2, since the 6^{th} break or 41.6114 + 5.455 = 47.076 years from the beginning. Figure 3.2 shows the plot of Weibull ROCOF function and *Alpha*(n) that is $1/E(T_n)$.

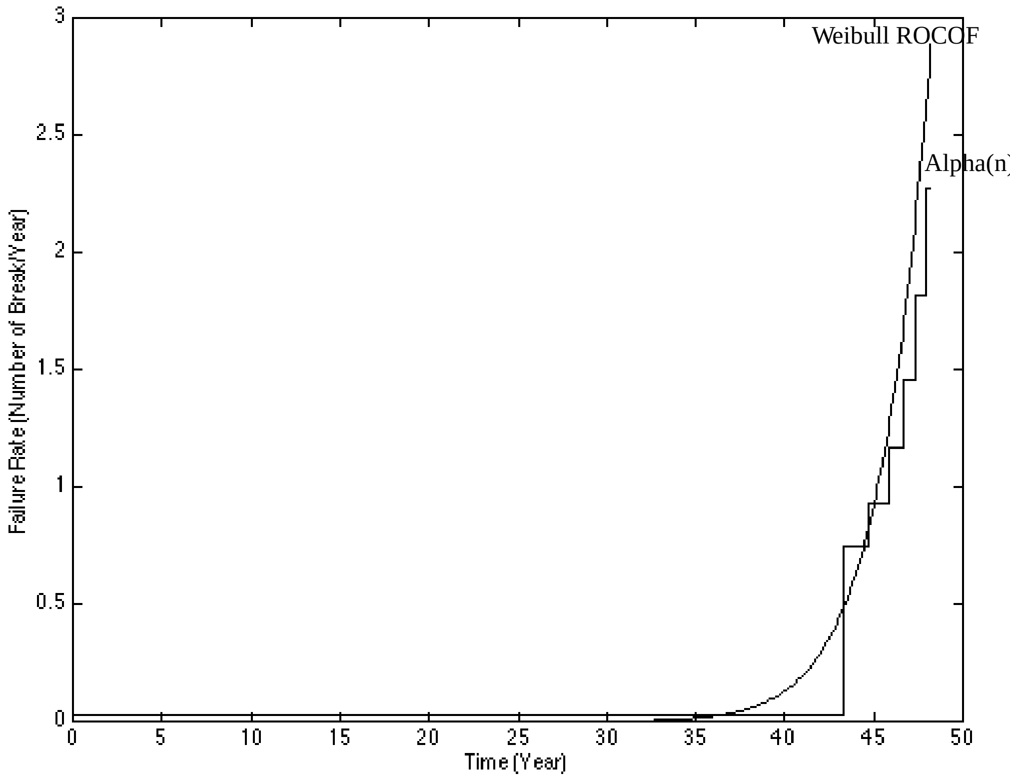


Figure 3.2 Weibull ROCOF function and *Alpha*(n).

Only the later part of the original data (from time 43.2891) is used in fitting the Weibull model. Weibull model gives a poor fit if entire data is fitted due to the lack of flexibility of coping with the trend in the controlled data. However, this approach of fitting only later part of the data would still work because the critical parameter in fitting ROCOF or intensity function is failure rate per unit time. A plot of Weibull fit to the whole data is presented in Figure 3.3.

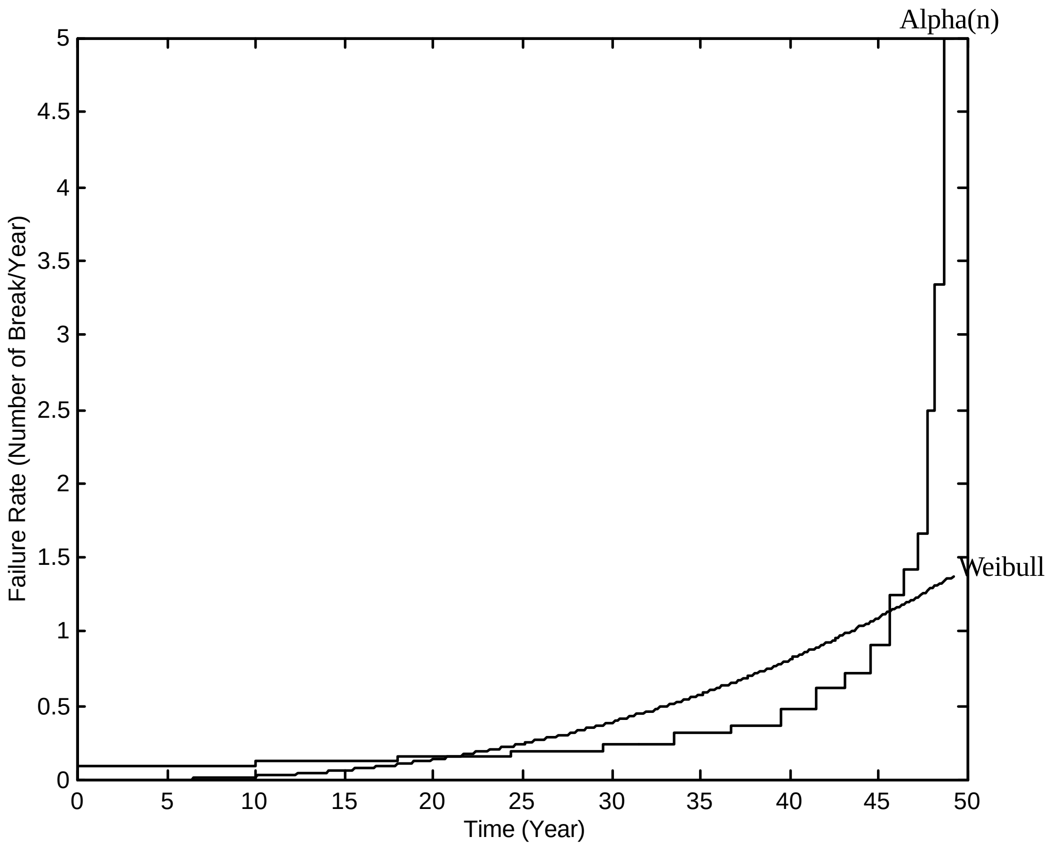


Figure 3.3 Plot of Weibull Fit for the Controlled Data.

Example 4 Given an annual interest rate of 7.5%, a constant repair cost of $3,000 and a constant replacement cost of $100 for a 1000 ft. pipe, compute the threshold break rate.

Solution

We can use the threshold break rate equation to obtain the solution. However, a series of tables can be generated to eliminate any kind of computation. Table 3.4 shows an example of many possible tables in which the discount rate is fixed and the cost ratio and the length of pipe are varied to obtain the threshold break rate. In this example cost ratio is C/F = 4000 ($) / 100 ($/ft) = 40 (ft) and the length is 1000 ft. Also, the discount rate is given as 7.5 %. Therefore, from Table 3.4 the threshold break rate for this pipe is given as 1.844.

Table 3.4 Example Threshold Break Rate Table (R = 0.075)

Length(ft) / Cost Ratio (C/F)	1000	2000	3000
30	2.447	3.857	7.268
31	2.369	3.702	7.035
32	2.296	3.556	6.816
33	2.228	3.419	6.611
34	2.163	3.290	6.417
35	2.102	3.169	6.235
36	2.045	3.054	6.063
37	1.991	3.945	5.900
38	1.939	3.842	5.746
39	1.890	3.745	5.599
40	1.844	3.652	5.460
41	1.800	3.564	5.328
42	1.758	3.480	5.202
43	1.718	3.400	5.082
44	1.680	3.323	3.967
45	1.643	3.250	3.857
46	1.608	3.180	3.753
47	1.575	3.113	3.652
48	1.543	3.049	3.556
49	1.512	2.988	3.464
50	1.482	2.929	3.375

3.7 Practical Usage of the Threshold Break Rate

By using the threshold break rate equation a series of graphs and tables can be generated for practical use. The threshold break rate is expressed as:

$$Brk_{th} = \frac{\ln(1+R)}{\ln\left(1+\frac{C}{F*L}\right)} \tag{3.7.1}$$

where F is replacement cost per unit length of a pipe ($/ft) and L is the length of a pipe (ft). Considering that pipes are easily categorized by size Eq. (3.7.1) is modified to represent the threshold break rate for different sizes of pipes. If a linear relationship is assumed between diameter and the cost ratio (C/F)

C/F = A*D + B

where, A and B are regression coefficients and D is diameter of pipe.

Then, Eq. (3.7.1) can be expressed as

$$Brk_{th,2} = \frac{\ln(1+R)}{\ln\left(1+\frac{A*D+B}{L}\right)} \tag{3.7.2}$$

where, L is the length of pipe.

Table 3.5 shows replacement cost per unit length (ft) and repair cost per break incident used in this study. Using this information, Table 3.6 shows the ratio of repair cost to replacement cost per unit length (foot) of the water distribution system used in this study.

Table 3.5 Cost Table by Pipe Size

Size(inch)	Replacement Cost ($/ft)	Repair Cost($)
4	90.00	2362.00
6	92.77	2814.00
8	96.95	3985.00
10	106.50	5869.00
12	116.05	7753.00
14	121.85	8140.65
16	127.94	8547.68
18	134.34	8975.06

Table 3.6 Cost Ratio Table by Pipe Size

Size(inch)	Repair Cost($)/Replacement Cost($/ft) (C/F)
6	30.333
8	41.104
10	55.108
12	66.807

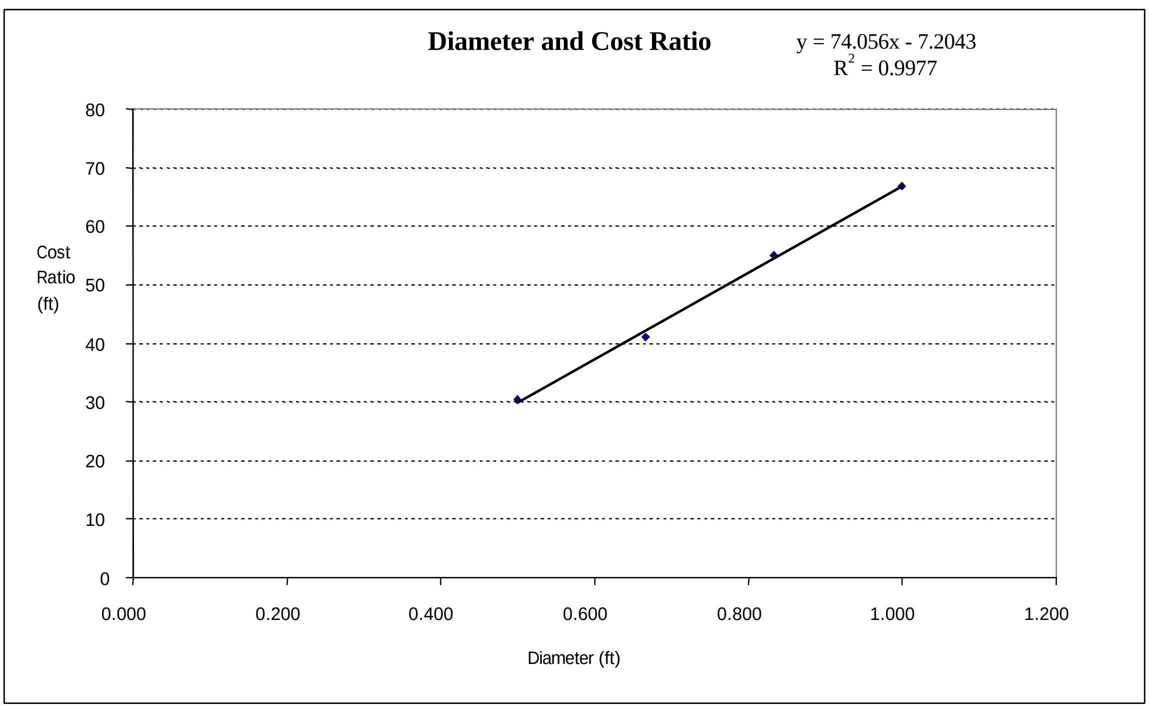


Figure 3.4 Diameter and Cost Ratio.

Figure 3.4 shows the linear relationship using actual repair and replacement costs. The linear Equation is expressed as R = 74.056 D – 7.204.
where, D is the diameter (ft) of pipe and R is the cost ratio (repair cost/replacement cost per foot) Therefore, Eq. (3.7.2) is expressed as

$$Brk_{th} = \frac{\ln(1+R)}{\ln\left(1+\frac{74.056*D-7.204}{L}\right)} \quad (3.7.3)$$

By using Eq. (3.7.3) a series of graphs can be generated to determine the threshold break rates for different sizes of pipes for given lengths and discount rates. Figure 3.5 and 3.6 show examples of such graphs.

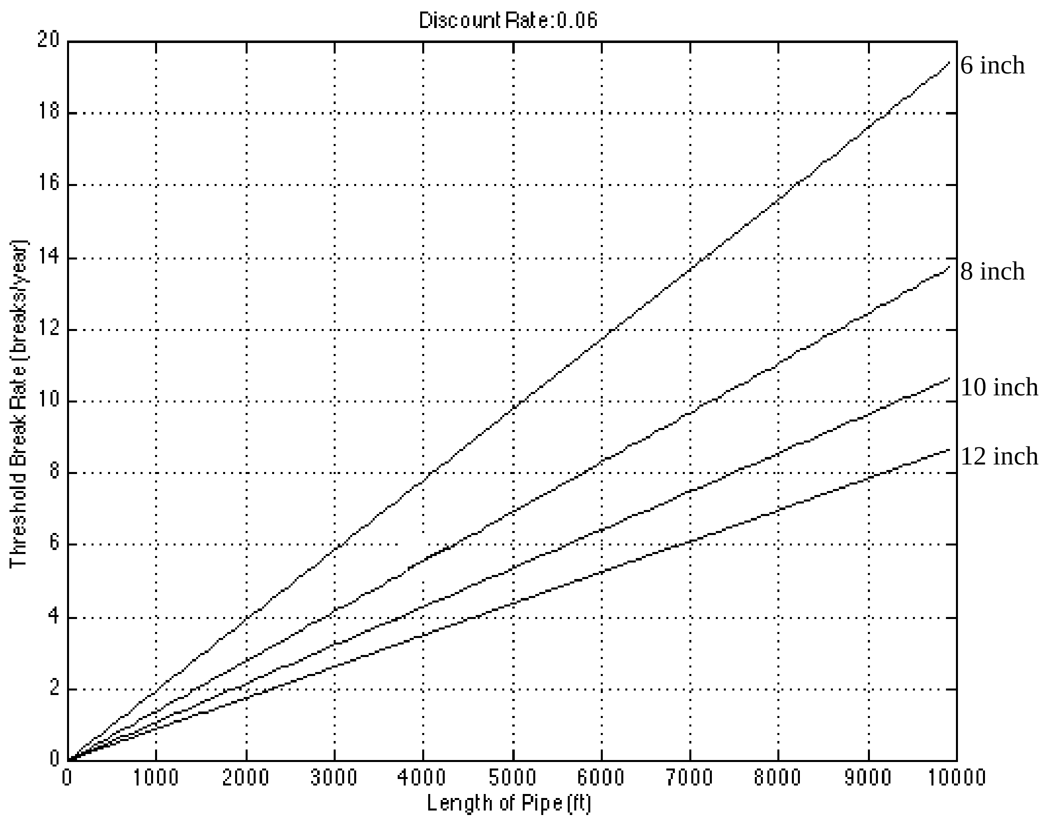


Figure 3.5 Threshold Break Rate Graphs for Pipe Size and Length (Discount Rate = 0.06)

According to Figure 3.5, 10 inch pipe should be replaced when the break rate (breaks/year) reaches 5 for a given length of 4000 ft and the discount rate of 0.06. On the other hand the threshold break rate of 8 inch pipe is shown to be about 5.5 given the same length and the discount rate. However, this result does not imply that bigger size pipes should be replaced more frequently than smaller size pipes since it takes a longer time for a bigger pipe to reach a certain threshold break rate than a smaller one. One should not confuse threshold break rate with optimal replacement time.

Another meaningful comparison can be made regarding the same threshold break rate and length of pipe. Let us consider 10 inch and 12 inch pipes with a threshold break rate of 5. Figure 3.6 shows that while 4000 ft of 10 inch pipe should be replaced when it reaches threshold break rate 5, 5000 ft of 12 inch pipe should be replaced when it reaches the same break rate. It says that because larger pipes fail so infrequently that we need to consider longer lengths of pipes to have the same break rate as a smaller pipe. Another interpretation is that for the same given length larger diameter pipes will have smaller break rates.

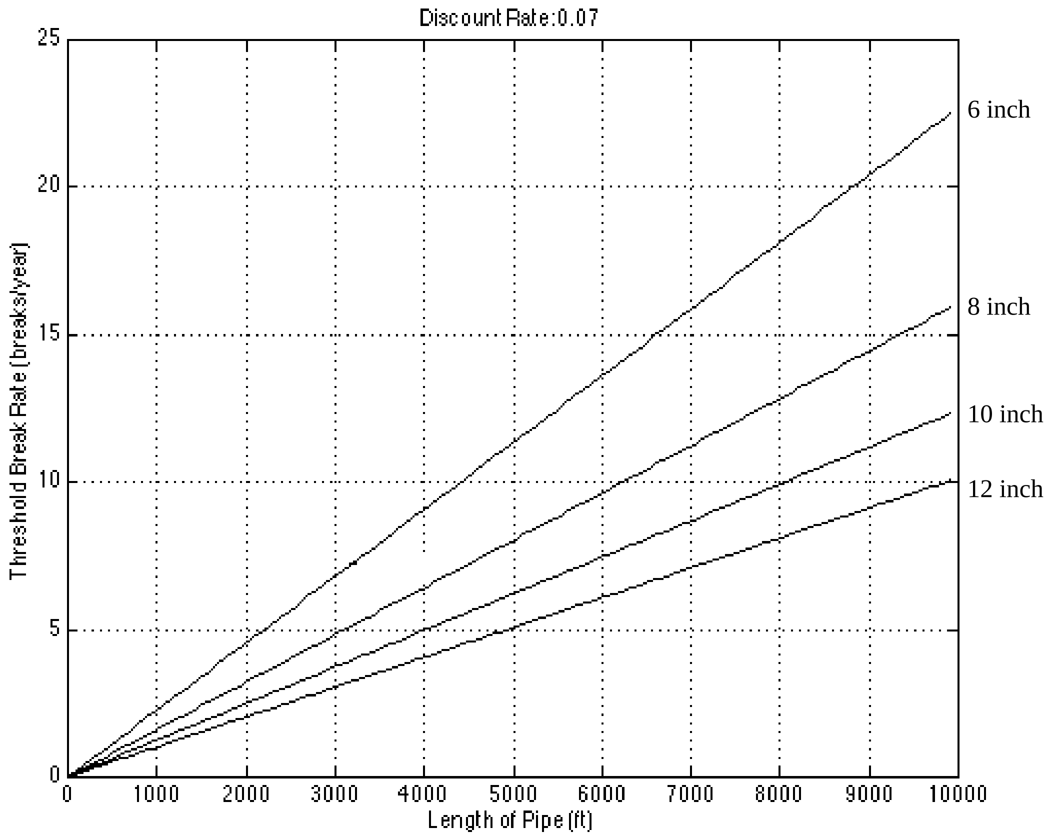


Figure 3.6 Threshold Break Rate Graphs for Pipe Size and Length (Discount Rate = 0.07)

Chapter 4.
Probabilistic Break Rate Models

4.1 Introduction

In this chapter three mathematical models to predict break rates of pipe are considered. These models are coupled with the optimal threshold break rate estimator to obtain the optimal replacement time. First is the proportional intensity model (PIM) which contains the break arrivals as a nonhomogeneous Poisson process. The Poisson parameter, l , is taken as a function of time. In addition, the factors dependent on environmental conditions of pipe further modifies the Poisson parameter, l , to account for the acceleration of pipe breaks due to the aggression of the environmental factors. Second is the proportional hazards model (PHM) in which a hazard function is taken to be a function of time (background process) multiplied by an environmental factor dependent constant (accelerating process). Third is a regression model that extends the trend of the break rate in a more obvious (say, eye fit) manner, than trying to decipher certain hidden structures in the PIM or the PHM. While PIM and PHM can yield failure probabilities of a pipe, such information cannot be elicited from the straightforward curve fit regression analysis. In the past the pipe break rate growth had taken to be exponential [Shamir and Howard (1979) and Walski and Pelliccia (1982)]. However, the presently available data at our disposal seems to follow a trend anywhere between an exponential and linear. Therefore, a general regression model that spans the linear to exponential trend is formulated. These models are applied to real life data described in Chapter 5 and the results are shown in Chapter 7.

4.2 Proportional Intensity Model (PIM)

Given that the failure characteristics of a pipe which experiences multiple failures in its lifetime, use of intensity model is appropriate. A fully parametric approach specifies the baseline intensity function $l_0(t)$ of the non-homogeneous Poisson process (NHPP) with proportional intensity function defined by

$$l(t) = l_0(t)exp(\mathbf{g}^T\mathbf{z}) \qquad (4.2.1)$$

where $\lambda_0(t)$ denotes the baseline intensity, assumed here to be continuous, $\mathbf{z}^\mathbf{T} = [z_1, z_2, \dots z_M]$ denotes the covariate vector with dimensionality M and $\boldsymbol{\gamma}^T = [g_1, g_2, \dots, g_M]$ denotes the covariate parameter vector.

Time dependent covariates can be used when the recorded environmental conditions are not time homogeneous. In such fragmented covariate structure situation the PIM is expressed as

$$l_j(t) = l_0(t)exp(\mathbf{g}^T\mathbf{z_j}) \qquad (4.2.2)$$

where $\mathbf{z_j}$ represents the recorded covariate values at the time of the jth failure,

$$t_{j-1} \le t \le t_j$$

where j = 1, 2 , … , n

$\lambda_0^*(t)$ can be obtained by assuming an appropriate distribution for the base-line failure rate. The Weibull distribution is known to yield a reasonable fit for many failure processes. Hence, the Weibull baseline intensity function is used for this application and it has the following form:

$$l_0(t) = \frac{b}{h}\left(\frac{t}{h}\right)^{b-1}, t > 0, \quad b > 0, \quad h > 0 \tag{4.2.3}$$

The estimates of the parameters can be obtained by using the method of maximum likelihood. The likelihood in failure data analysis in general is defined by:

*L = [Joint probability of failure at each failure time] * [Probability of having no failure from the last failure time until censoring time]*

By referring to Eq. (3.5.4) the likelihood function can be written as

$$\mathrm{L} = \prod_{j=1}^{n} \lambda(t_j) \exp\left[-\sum_{j=1}^{n+1} \int_{t_{j-1}}^{t_j} \lambda(u) du \right] \tag{4.2.4}$$

Note that the discretization (t_j, t_{j+1}) lends itself for substituting appropriate z_j values in $\lambda_j(t) = \lambda_0(t)\exp(\gamma^T z_j)$ from Eq. (4.2.2). We replace each $l(t_j)$ in Eq. (4.2.4) in terms of $\lambda(t_j) = \frac{\beta}{\eta}\left(\frac{t_j}{\eta}\right)^{\beta-1} \exp(\gamma^T z)$ from Eq. (4.2.1) and (4.2.3). The resulting likelihood function involves the variables b, h, and g^T, and their optimal values can be obtained. The entire derivation is given in Appendix A.

4.3. Optimal Replacement Time Analysis Using the Weibull Proportional Intensity Model

As described in Chapter 3 the threshold break threshold can be used with a break rate prediction model to obtain optimal replacement time of a pipe. Optimal replacement time using PIM is expressed as

$$l(t) = Brk_{th} \tag{4.3.1}$$

where $l(t)$ is proportional intensity function and Brk_{th} is threshold break rate. Thus, whenever intensity (rate) of break reaches threshold break rate, the pipe should be replaced. By applying the proportional intensity model and the threshold break rate equation, Eq. (4.3.1) is expressed as

$$\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{\gamma z} = \frac{\ln(1+\mathrm{R})\mathrm{F}*\mathrm{L}}{\mathrm{C}} \tag{4.3.2}$$

where time invariant repair and replacement cost is assumed and all other notations are the same as previous. Also, F is cost of replacing a pipe per unit length.

Therefore, the optimal replacement time is

$$t^* = \eta\left(\frac{\eta \mathrm{F}*\mathrm{L}\ln(1+\mathrm{R})e^{-\gamma z}}{\beta \mathrm{C}}\right)^{\frac{1}{\beta-1}} \tag{4.3.3}$$

Using x = ln(1+x) for small x, the optimal replacement time is expressed as

$$t^{*} = \eta\left(\frac{\eta \ln(1+R)e^{-\gamma z}}{\beta \ln\left(1+\frac{C}{F*L}\right)}\right)^{\frac{1}{\beta-1}} \quad (4.3.4)$$

in which C/(F*L) is taken to be small.
When three-parameter Weibull baseline intensity is assumed:

$$\frac{\beta}{\eta}\left(\frac{t-\theta}{\eta}\right)^{\beta-1} e^{\gamma z} = \frac{\ln(1+R)F*L}{C} \quad (4.3.5)$$

Therefore, the optimal replacement time is

$$t^{*} = \theta + \eta\left(\frac{\eta F*L\ln(1+R)e^{-\gamma z}}{\beta C}\right)^{\frac{1}{\beta-1}} \quad (4.3.6)$$

Again implementing C/(F*L) = ln(1 + C/(F*L)), the optimal replacement time is expressed as

$$t^{*} = \theta + \eta\left(\frac{\eta \ln(1+R)e^{-\gamma z}}{\beta \ln\left(1+\frac{C}{F*L}\right)}\right)^{\frac{1}{\beta-1}}. \quad (4.3.7)$$

4.4 Proportional Hazards Model (PHM)

Let T denote the time to failure of a pipe under consideration, with density function *f(t)* and reliability or survival function *S(t)*. The hazard function is defined as [Ascher and Feingold (1984)]:

$$h(t) = \frac{f(t)}{S(t)} \quad (4.4.1)$$

where: *h(t)dt* is the conditional probability of failure in the time interval *(t, t + dt)* given survival to time t. However, the hazard function of a pipe in general is influenced not only by the time, but also by the covariates under which it operates. The PHM is used to estimate their effects and to predict failure behavior of pipes in water distribution systems. The model has the form:

$$h(t;z) = h_0(t)\psi(\underline{z};\underline{\beta}) \quad (4.4.2)$$

where: $h_0(t)$ is an arbitrary and unspecified baseline hazard function, $y(\underline{z};\underline{b})$ is a scaling term. Also, $\underline{z}$ is a row vector consisting of the covariates, and $\underline{b}$ is a column vector consisting of the regression parameters. It is generally assumed that the functional form of $y(\underline{z};\underline{b})$ can be obtained. Some of the forms are: the exponential form, *exp*($\underline{b}\underline{z}$), the logistic form, *log(1+ exp*($\underline{b}\underline{z}$)), the inverse linear form, *1/(1+*$\underline{b}\underline{z}$*);* and the linear form, *(1+*$\underline{b}\underline{z}$*)*. Among these forms the exponential form for $y(\underline{z};\underline{b})$ is the most widely used because of its simplicity. In this case the hazard function can be written as

$$h(t;z) = h_0(t)\exp(\underline{\beta z}) = h_0(t)\exp(\sum_{j=1}^{q}\beta_j z_j) \qquad (4.4.3)$$

where z_j, j = 1, 2, ..., q, are the covariates associated with the system and b_j, j = 1, 2, ..., q, are the unknown parameters of the model, defining the effects of each one of the q covariates. The related survival functions are given by

$$S(t;z) = \left[S_0(t)\exp(\sum_{j=1}^{q}\beta_j z_j) \right] \qquad (4.4.4)$$

where

$$S_0(t) = \exp\left[-\int_0^t h_0(x)dx \right] = \exp[-H_0(t)] \qquad (4.4.5)$$

and $S_0(t)$ is the baseline survival function dependent only on time, and $H_0(t)$ is the cumulative baseline hazard function.

To construct the likelihood function, consider the probability that one fails among the survivors in the risk set, R(t), which is defined by

$$R(t) = \{k : t_k \geq t, \text{ where} : t_k = \text{failure time for item } k\}.$$

$$P[\text{one failure at time t, in R(t)}] = \sum_j P[t \leq t_j \leq t + \Delta t \mid t_j \geq t]$$

and

$$\sum_{j\in R(t)} P[t \leq t_j \leq t + \Delta t \mid t_j \geq t] = \sum_{j\in R(t)} h_j(t)\Delta t$$

(Recall the definition of $h_j(t) = P[t \leq T_j \leq t + \Delta t \mid T_j \geq t]/\Delta t$)

The probability that item 'i' fails at time t, given that an item fails in the risk set, R(t), is therefore written as

$$\frac{P[t \leq t_i \leq t + \Delta t \mid t_i \geq t]}{\sum_{j\in R(t)} P[t \leq t_j \leq t + \Delta t \mid t_j \geq t]} = \frac{h_i(t)}{\sum_{j\in R(t)} h_j(t)\Delta t} = \frac{\exp(\beta z_i)}{\sum_{j\in R(t)} \exp(\beta z_j)} \qquad (4.4.6)$$

in which $b = \{b_1, b_2, \ldots, b_q\}$, vector of regression coefficients; z_j = vector of component values observed for the jth item in the risk set. Therefore, the partial likelihood function, L, is obtained as the joint failure of the k observed items given by

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp(\beta z_i)}{\sum_{j\in R(t)} \exp(\beta z_j)} \qquad (4.4.7)$$

By maximizing the partial likelihood function we obtain the optimal b value. The details on estimating b appear in Appendix B.

The baseline hazard rate represents the hazard rate when b coefficient vector is zero. The maximum likelihood estimator of this baseline hazard rate at every failure time is given by [Breslow (1974)].

$$h_0(t_j) = \frac{d_j}{\left(t_j - t_{j-1}\right) \sum_{k \in R(t_j)} \exp(\beta z_j)} \tag{4.4.8}$$

in which d_j = number of failures at time t_j.
Note that in the above equation when b = 0, it yields the usual estimate of the ratio of failures to survivors per unit time. The estimate of the cumulative baseline hazard

$$H_0(t) = \int_0^t h(u)du$$

is written as [Breslow (1974)]

$$H_0(t) = \sum_{t_j \le t} \frac{d_j}{\sum_{k \in R(t_j)} \exp(\beta z_j)} \tag{4.4.9}$$

$H_0(t)$ in general will plot as discrete points for each observed failure time t_j. After fitting a smooth curve, its derivative is obtained as the baseline hazard rate $h_0(t)$. The hazard rate h(t) can, then, be written as

$$h(t; z) = h_0(t) \exp(\beta z) \tag{4.4.10}$$

4.5 Survival Analysis Using the Proportional Hazards Model

The methodology discussed above possesses flexibility for adapting to particular system characteristics, can directly provide the failure probabilities of individual pipes as a function of time and past history, and allows for more realistic and unrestricted assumptions concerning the underlying process. The effect of the causal factors leading to the breaks is of considerable interest. Such factors that affect failure time include:

a. Previous break history

b. External variables describing environmental characteristics like soil properties and land use activities in the vicinity of the pipes.

Usefulness of the PHM can be well examined by analyzing the structure of the model itself. A model of the type described by Equation (4.4.2) can be interpreted as follows. The function $h_0(t)$ represents an “aging” process that goes on independently of stress. For the case of deteriorating pipelines the aging process could represent a corrosion induced deterioration that makes the pipe weaker as time goes on. Assuming further that the conditional probability of failure at any time is the product of an instantaneous time-dependent term related to the “aging” process and a stress-dependent term described by $\exp(\underline{b}\, z)$, we obtain the underlying structure of the PHM given by Equation (4.4.2). The stress factors included in the exponent $\underline{b}\, z$ are assumed to act multiplicatively on the hazard rate, which in the case of failing pipes appears to be intuitively appealing. This is so, because it seems very likely that the interaction among various “high risk” factors would significantly increase the chances for a break to occur. The above observations tend to reinforce the argument that the structure of the PHM would be appropriate for describing the failure process of individual pipes.

It must be noted that in the theoretical development, it is assumed that the item dies at the failure time. The ordered sequence of failure times governs the parameter estimation process through the risk sets determined at each failure time. If we assume that the pipe is simply repaired and gets back into the population again, then the risk sets become ill defined. Therefore, for repairable systems the hazard function concept involving the risk sets is not amenable.

Let us consider the scenario in which after the nth break the pipe gets replaced. In such a scenario after the (n-1)th break the hazard concept seems to apply. It is in this vein that the proportional hazards model is applied in this dissertation. Therefore, the hazard model should be built on time between two consecutive breaks for different pipes which form the population. Therefore, there will be different hazard rates for the same pipe for installation to first break, first to second break, second to third break and so on. The following section describes the process in detail.

4.6 Hazard Function between Breaks

The usual PHM assumes only one failure for a piece of equipment. By applying the definition of the risk set at first failure, second failure, and so on, hazard function between breaks can be obtained. A risk set contains a list of survivors at risk at time, τ, denoted by $R(\tau)$. Consider a typical water main break example where multiple breaks occur. Figure 4.1 shows the history of breaks for selected pipes. Figure 4.2 and 4.3 illustrate the associated break times for 0 – 1 break and 1 –2 break.

In Figure 4.1 there are 4 pipes that have experienced multiple breaks. A line for each pipe represents that the pipe is under operation (alive) and breaks are denoted by '$x_{i,j}$' where i is pipe number and j is the order of the breaks. We want to develop the hazard function associated with 0 to 1 failure denoted by $h_{0,1}(t)$. The associated risk sets are given by

$R(t_1)$ = {pipe 1, pipe 2, pipe 3, pipe 4}
$R(t_2)$ = {pipe 2, pipe 3, pipe 4}
$R(t_3)$ = {pipe 3, pipe 4}
$R(t_4)$ = {pipe 4}
(For this case $t_1 = x_{11}$, $t_2 = x_{21}$, $t_3 = x_{31}$, $t_4 = x_{41}$.)

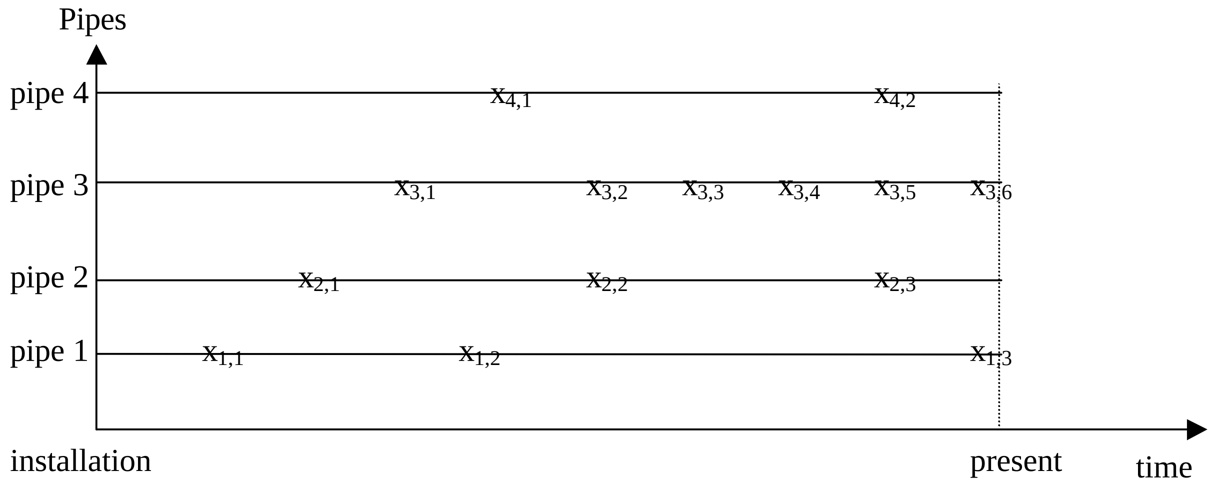


Figure 4.1 Water Main Break History

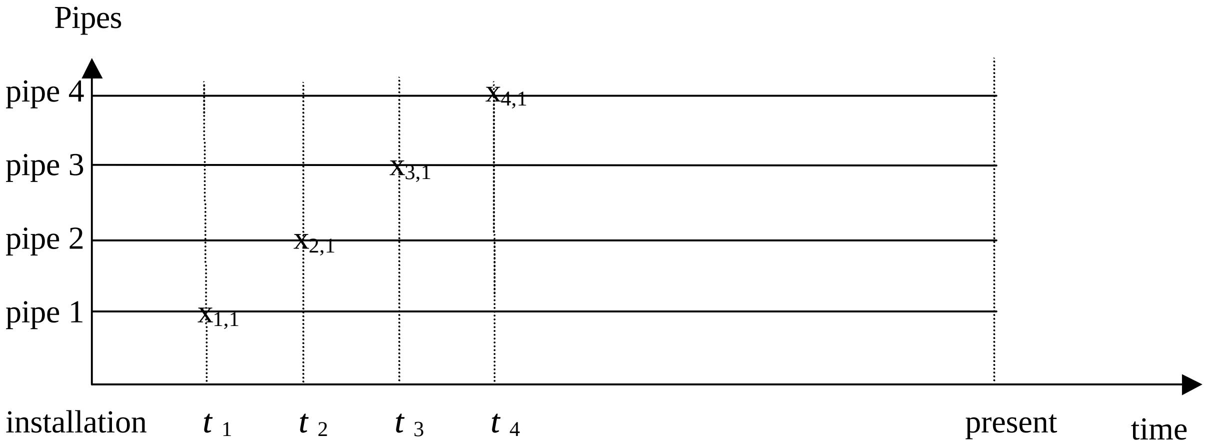


Figure 4.2 Water Main Break History (installation to the first break)

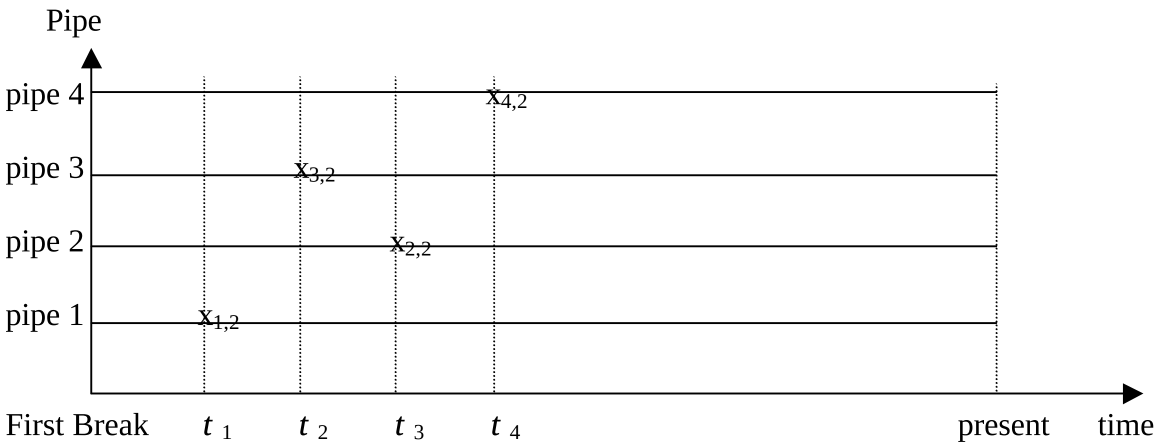


Figure 4.3 Water Main Break History (from the first break to the second break)

Now we want to obtain the hazard function associated with the time between the 1^{st} and 2^{nd} breaks denoted by $h_{1,2}(t)$. Here, given that one break has occurred, we want to determine the hazard pattern towards the second break. For example, given that the pipe has not experienced any breaks for two years since the first break, we want to determine the probability that the second break will occur during the next year.

Let us denote $R_2(t_j)$ as a risk set just before the jth ordered failure given that all jth ordered failures are second failures for each pipe. The risk set, $R_2(t_j)$ can be interpreted as "the set of pipes that has not failed since first break until just prior to the jth break time" and mathematically as

$$R_2(\tau_j) = \{k: t_{k,2} \geq \tau_j\}$$

where,
k = identity of pipes (pipe numbers)
$t_{k,2}$ = the second break time of pipe k
The corresponding risk sets for each pipe in the example are
$R_2(t_1)$ = {pipe 1, pipe 2, pipe 3, pipe 4}
$R_2(t_2)$ = {pipe 2, pipe 3, pipe 4}

$R_2(t_3)$ = {pipe 2, pipe 4}
$R_2(t_4)$ = {pipe 4}

Once the risk sets, $R_2(\tau_j)$ are defined, we can apply the maximum likelihood procedure to estimate the hazard function between the first and second breaks. The likelihood function for this case is

$$L(\underline{\beta}) = \prod_{j=1}^{n} \frac{h_{0_{1,2}}(\tau_j)\exp(\underline{\beta}\underline{z}_i)}{\sum_{k\in R_2(\tau_j)} h_{0_{1,2}}(\tau_j)\exp(\underline{\beta}\underline{z}_k)}$$

$$= \prod_{j=1}^{n} \frac{\exp(\underline{\beta}\underline{z}_i)}{\sum_{k\in R_2(\tau_j)} \exp(\underline{\beta}\underline{z}_k)} \qquad (4.6.1)$$

where, $h_{0_{1,2}}(t_j)$ denotes the baseline hazard function between the first and second breaks at t_j.

Even though the baseline hazard function cancels out in the likelihood function, it may have a different form from the one with 0 – 1 break case. In Eq. (4.5.1) the risk set, $R_2(t_j)$, which is a risk set for the 1 - 2, is used. The first and second derivative of log of Eq. (4.5.1) can be determined and maximum likelihood estimates of $\underline{b}$ can be obtained by an iterative method such as the Newton-Raphson method as shown in Appendix B.

As a result, the proportional hazards functions for each pipe between the first and second break time periods, $h_{1,2}(t)$, can be determined. In a similar manner the proportional hazard functions for each pipe between subsequent break time periods, $h_{2,3}(t)$, $h_{3,4}(t)$, and so on, can be determined.

Thus far, we have established a methodology that enables the use of the PHM for repairable systems, pipe systems for our interests. Although the PHM can be established for every break interval, the model of interest is the PHM in the last break interval of a pipe. Replacement decision is always required either at current time or in the future. Therefore, one only needs to check whether the current hazard rate exceeds the threshold break rate or not to obtain the future optimal replacement time. Thus, the criteria for replacement decision can be expressed as

(a) If Current Hazard Rate $\geq$ Threshold Break Rate then, replace at current time (year) and, if not,
(b) Obtain Optimal Replacement Time by Solving $h_L(t) = Brk_{th}$ for time t, where $h_L(t)$ is the PHM for the last break interval of a pipe.

4.7 General Break Regression Model

The accuracy of the fitted model plays a major role in the regression approach for the prediction of future pipe breaks and subsequent timing of replacement of water mains. Poorly fitted models produce either unrealistically high or low number of breaks in future years resulting in too early or late replacement. In this section a new model to better fit water main break data is developed. The exponential model which has been used by Shamir and Howard (1979) and other authors has been found often to over predict the number of future breaks of water mains. The following section shows how the combination of the exponential model and linear models, or the general break regression model, improves the pipe break prediction.

In the general model described in the section, a weighting factor is utilized to moderate the dominance of either exponential or linear model in the prediction of breaks for each pipe. This resulting general model offers two advantages over the exponential model: (1) it provides a better fit for the break history and it better predicts the future number of breaks for a pipe; (2) it is a break prediction model based on a cumulative number of breaks. Shamir and Howard (1979) use “number of breaks in a year per 1000ft length of pipe.” Considering the break patterns of a real pipeline underground, this approach may not generate satisfactory results because some pipelines simply don’t fail frequently even though the length is great (more than 1000 ft). Therefore, the break prediction model presented here incorporates cumulative number of breaks since installation.

The exponential part of the general model has the following form:

$$N_c(t) = B\_exp \cdot e^{A\_exp(t - t_0)} \quad (4.7.1)$$

where

$N_c(t)$ = cumulative number of breaks along the length of a defined pipe in year t

t = time in years

t_0 = base year for the analysis(pipe installation year, or the first year for which data are available)

A_exp = growth rate coefficient (1/year)

B_exp = coefficient of regression

The linear part of the general model has following form:

$$N_c(t) = B\_lin + A\_lin(t - t_0) \quad (4.7.2)$$

where

$N_c(t)$ = cumulative number of breaks along the length of a pipe in year t

t = time in years

t_0 = base year for the analysis(pipe installation year, or the first year for which data are available)

A_lin = growth rate coefficient (1/year)

B_lin = coefficient of regression

The general break regression model has the following form:

$$N_c(t) = (1 - wf)(B\_lin + A\_lin(t - t_0)) + wf \cdot B\_exp \cdot e^{A\_exp(t - t_0)} \quad (4.7.3)$$

where

$N_c(t)$ = cumulative number of breaks along the length of a pipe in year t

t = time in years

t_0 = base year for the analysis(pipe installation year, or the first year for which data are available)

WF = weighting factor to determine the best model for the data given, 0 <= WF <= 1

A_lin and B_lin = coefficient of linear model

A_exp and B_exp = coefficient of exponential model

The coefficients A_lin, B_lin, A_exp, and B_exp, are determined from a regression analysis. The weighting factor is determined between 0 and 1 that would result in the least sum of squared relative errors. The algorithm implemented for determining the weighting factor is shown in Figure 4.11. In addition to finding the best weighting factor, Figure 4.11 shows a method of enhancing the predictive capability of the general model. Following the steps in Figure 4.11, the weighting factor is determined as follows:

First, the value of the weighting factor is set to 0, which is the linear model, and the sum of the errors is computed for the later 1/3 portion of the data. For example, if there are 9 historic breaks, sums of the chi-square values are computed and stored for 7th, 8th, and 9th breaks. Then, the weighting factor is increased by some increment, ε and the sum of the errors is obtained. This process continues until the weighting factor is 1, which is the exponential model. The optimal weighting factor is the one with least sum of squared errors. The procedure can be mathematically represented as follows:

$$\min \; SSE = \sum_{j=1}^{n} \left(O_j - C_j\right)^2 \qquad (4.7.4)$$

Subject to:

$C_j = (1 - w_i)L(t) + w_i E(t)$

$w_i = i \cdot e$

where

$i = \{0, 1, \ldots, 1/e\}$

$1/e$ is an integer

$0 < e < 1$

where SSE = sum of squared errors for each i

O_j = each observed break time

C_j = computed value from the general model for each j

ω_i = weighting factor for each i

E(t) = the exponential model (Eq. (4.7.1))

L(t) = the linear model (Eq. (4.7.2))

n = number of breaks

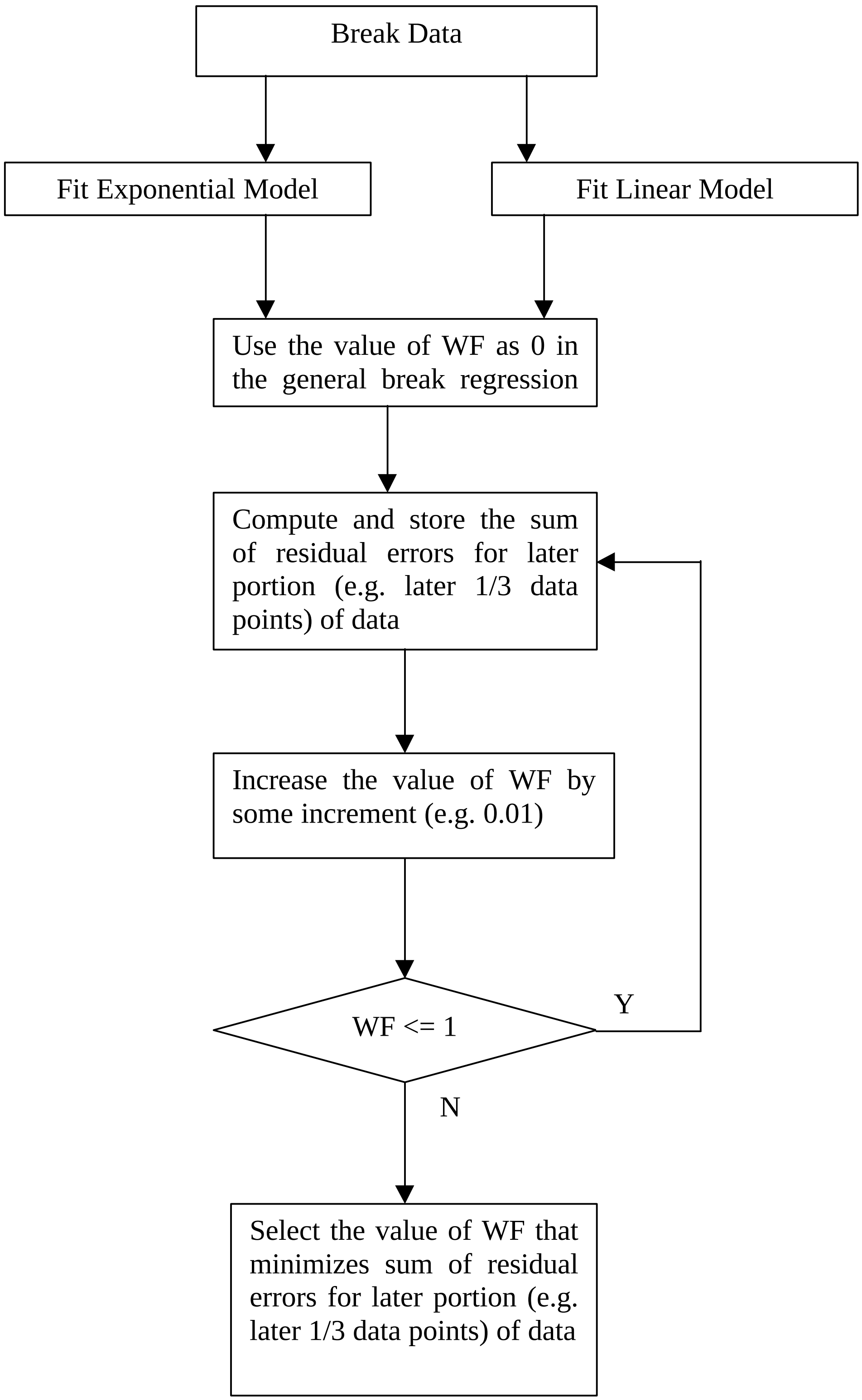


Figure 4.4 The General Break Regression Model Building Process.

4.8 Optimal Replacement Analysis with the General Break Regression Model

There are two ways to obtain optimal replacement time of a pipe by using the general break regression model. One is using the general break regression model with the threshold break rate. Another one is using a search method. Since Eq. (4.7.3) represents cumulative number of breaks at time t, it needs to be converted to a break rate function by taking a derivative with respect to time t, that is

$$\frac{d}{dt}N_c(t) = \frac{d}{dt}\left((1 - wf)(B\_lin + A\_lin(t - t_0)) + wf \cdot B\_exp \cdot e^{A\_exp(t - t_0)}\right)$$

$$= (1 - wf)A\_lin + wf \cdot A\_exp \cdot B\_exp \cdot e^{-A\_exp \cdot t_0} e^{A\_exp \cdot t}$$

By using

$$\frac{dN_c(t)}{dt} = Brk_{th} \tag{4.8.1}$$

the optimal replacement time, t*, is obtained as:

$$t^* = \frac{1}{A\_exp} \ln\left(\frac{Brk_{th} - (1 - wf) \cdot A\_lin}{wf \cdot A\_exp \cdot B\_exp \cdot e^{-A\_exp \cdot t_0}}\right) \tag{4.8.2}$$

In the search method the objective is to minimize the present value of the total cost of repairing and replacing a pipe occurring in a future year. The cost of repairing a pipe in year t is

$$C_m(t) = C_b N_c(t) = C_b\{N_c(t_p + i) - N_c(t_p + i-1)\}, \; i = 0,1,2,\ldots \tag{4.8.3}$$

where

C_b = cost of repairing a break($)

t_p = present year

By using the discount rate to obtain the cost of repair in future years the present value of repair cost is

$$P_m(t_r) = \sum_{t=t_p}^{t_r} \frac{C_m(t)}{(1+R)^{t-t_p}} = \sum_{i=0}^{x} \frac{C_m(t)}{(1+R)^{t_p+i-t_p}} = \sum_{i=0}^{x} \frac{C_m(t)}{(1+R)^i} \tag{4.8.4}$$

where

R = discount rate used in present year

t_p = present year

t_r = replacement year = $t_p + x$

Similarly, the present value of replacement cost is

$$P_r(t_r) = \frac{C_r}{(1+R)^{t_r-t_p}} = \frac{C_r}{(1+R)^x} \tag{4.8.5}$$

where C_r = cost of replacement($)

Hence, the economically optimal time for replacement is the year at which the total cost becomes minimal.

$$P_T(t_r) = P_m(t_r) + P_r(t_r) = \sum_{i=0}^{x} \frac{C_m(t)}{(1+R)^i} + \frac{C_r}{(1+R)^x}$$

Simplifying the equation above the present value of total cost becomes

$$P_r(t_r) = C_r(1+R)^{-x} + \frac{C_b(A\_lin(E-R)(D^x - 1 - R)(wf - 1) + B\_exp * e^{A\_exp(t_p - t_0 - 1)} E \cdot R(e^{A\_exp(1+x)} D^x - 1 - R)wf}{(E-R)R} \quad (4.8.6)$$

where
$E = e^{A\_exp} - 1$
$D = 1/(1+R)$
The optimal value of x that minimizes the present value of the total cost can be obtained by differentiating Eq. (4.8.6) with respect to x and find the value of x that make the differentiated equation as zero. The differentiated version of Eq.(4.8.6) with respect to x is
$d(P_T(t_r))/dx =$

$$C_b\left(J\frac{B\_exp \cdot H(F-K)}{G} + J\left(-\frac{A\_lin(1+R)}{R} + \frac{A\_lin(1+R)wf}{R} + \frac{B\_exp \cdot F \cdot I(1-e^{A\_exp})}{G} \right) \ln\frac{1}{1+R} \right) - C_r(1+R)^{-x}\ln(1+R) \quad (4.8.7)$$

where
$F = e^{A\_exp(tp - to)}$
$G = 1 - e^{A\_exp} + R$
$H = e^{A\_exp\, x}(1+R)\, wf \ln(e^{A\_exp})$
$I = e^{A\_exp\, x}(1+R)wf$
$J = D^{1+x}$
$K = e^{A\_exp(1+tp - to)}$

However, Eq. (4.8.7) does not have an algebraic solution for x that makes the equation zero. In this case the search method can be used to find the optimal replacement time. The search method essentially stores the present values of the total cost from the present year to some future years (say, 100 years from the present year) calculated by Eq. (4.8.6) and finds the year at which the present value of the total cost becomes minimal. Both of the methods yield identical results leaving the search procedure to be inefficient.

Chapter 5.
Description of the Case Study Water Distribution System

This chapter is part of a report submitted to the sponsoring utility. The author would like to thank Mr. Frank Grablutz of the Roy F. Weston, Inc. for permitting the use of the material as part of this dissertation.

5.1 Distribution System Overview

The selection and prioritization of pipes to be replaced require data on the installation, maintenance history, and current condition of the pipes. Many water utilities have not maintained these data adequately, or have only recently begun to computerize their data. The utility had the foresight to maintain a detailed, computerized water main break database since 1983. The availability of these 15 year's worth of data by 1998 is critical to this research. In addition, the utility's initiatives in developing Work Management Systems and related Geographic Information Systems (GIS) indicate that still more detailed data will be available for future applications of the models.This section describes the data sources within the utility that were used in applying the proposed methodologies. Key data sources are described and an overview of the existing pipe inventory and main break history of the case study area is provided.

5.2 Data Sources

Several sources of information related to the installation, maintenance history (i.e., breaks), and condition of pipe in the distribution system were identified. The utility also provided costs associated with main break repair and installation. These were used in conjunction with the forecasts of main breaks to conduct the optimal replacement analyses developed in this research. Before proceeding, it is important to define one data feature that was integral to all evaluations and modeling. The utility maintains much of the distribution system pipe information in terms of "Work Order." The Work Order for a pipe refers to the original project under which the water main was installed. Work Orders vary considerably in length, from less than 100 feet to more than 10,000 feet. The following key data sources were used in the evaluation of water main renewal needs and the development of the associated computer models.

5.2.1 Maintenance Department Break Database

The utility has maintained an electronic database of water main breaks since 1983. The database is maintained using the dBase IV software. These detailed historic records provide the foundation for the modeling. Key fields used in the modeling are described in Table 5.1. Actual dBase IV field names are provided in parentheses. The Break database used for this research was current through 1997, and included 32,242 break incident records. However, not all of these records were applicable to the needs of this research. Some of the records in the database related to hydrant leaks, breaks caused by third parties, and valve leaks. Others involved galvanized iron, concrete, asbestos cement, or plastic pipes that were not the focus of this study since they account for only a minor percentage of the pipe in the system.

Table 5.1 Break Database Key Fields.

Field	Description
Work Order (WORKORD)	The work order identifies the research under which the pipe was installed. This identifier was used to distinguish individual "pipes."
Footage (FOOTAGE)	The length of the work order, in feet.
Replaced (REPLACED)	A logical field that indicates if a pipe, or a portion thereof, was replaced as a result of a main break. This field is updated for older records when a replacement occurs on a portion of pipe that has experienced previous breaks.
Year Acquired (YEARACQ)	The year in which the pipe was installed (or in some cases acquired from another system).
Date (LKDATE)	The date on which the break was reported.
Subgrid (GRID)	The location of the break defined as grid symbol.
Diameter (SIZM)	The diameter of the pipe, in inches.
Material (TYPM)	The material of the pipe.
Break Type (LEAKCODE)	A descriptive code assigned by the Maintenance Supervisor after reviewing the failure. It is used to identify the type of the break (circumferential, longitudinal, broken hydrant, etc.).

5.2.2 Paper Break Records

Between 1965 and 1982 the utility maintained paper records of main breaks. Approximately 8,500 main failures occurred during this 18-year period, or an average of approximately 472 main breaks per year. This rate is considerably lower than the 32,242 recorded failures since 1983, or about 2,150 per year. Unfortunately, the paper records cannot be accessed as easily as the computerized records. However, the utility did investigate specific Work Orders in an attempt to produce a more complete break history for selected pipes. These data were compiled in the same format as the electronic database. A total of 318 additional main break records, called *the complete break database*, was added to the electronic data set in this fashion.

5.2.3 Depreciation Database

This database is maintained for accounting purposes within the utility. It is used to develop depreciation rates for pipes according to material and diameter. A portion of the database is shown in Table 5.2. Essentially, the database tracks the yearly history of pipe installations and retirements within the system. Applicable information from this database includes the dollar value of pipe installed in a given year (within the noted category), and the year and dollar value of any subsequent retirement of that pipe.

It should be noted that all dollar values are given in terms of the installation year. Examining Table 5.2, for example, $18,891.40 worth of 16-inch cast iron main was installed in 1959. In 1968 $144.21 worth (in 1959 dollars) of this pipe was retired, and another $3,512.10 worth (in 1959 dollars) was retired in 1970. Thus, a total of $3,656.31 worth of 1959 vintage 16-inch cast iron pipe was retired from the system. Expressed in another way, approximately 19% ($3,656.31 divided by $18,891.40) of the 16-inch cast iron pipe installed in 1959 is no longer in service.

Table 5.2 Example of Depreciation (Analysis) Database.

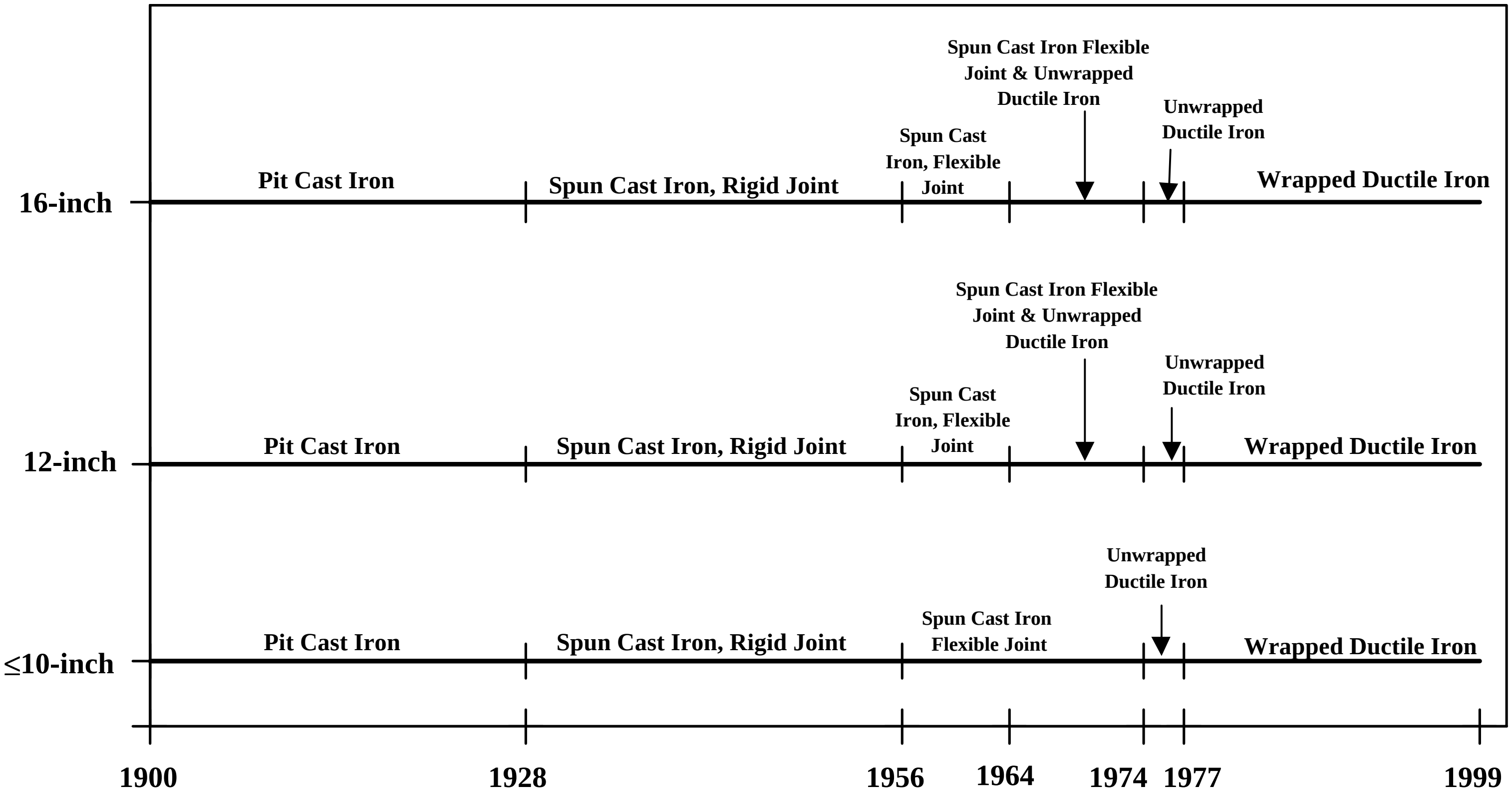
	Additions	Retirements	Sum of Retirements	Effective Year	Retirement Units	Cost of Removal	Cost of Salvage
343-11C16	T&D-Trans Mains-16" Cast Iron						
1959	18,891.40			1959			
		144.21		1968	2	55.16	
		3,512.10	3,656.31	1970	2	156.48	
1960	51,613.40			1960			
		1,036.86		1994	1		
		3,186.81	4,223.67	1996	2		
1961	115,790.68			1961			
		63.47		1984	2	101.2	
		2,743.27		1994	1		
		8,113.44	10,920.18	1996	6	4.12	
1962	283,653.44			1962			
		50.75	50.75	1996	1		
1963	5,217.66			1963			
		81.73	81.73	1974	2	5.86	
1964	92,425.23			1964			
		4,354.05		1969	2	16.25	315.9
		755.04	5,109.09	1985	2	26.25	
1965	11,570.14			1965			

These data are important because they provide the only information on the vintages of pipe remaining in the system. Dollar values for different years can be adjusted to a common year, and relative percentages calculated.

5.2.4 Pipe Material History

Various pipe materials have been used at different times over the history of the system. The utility provided a general timeline for when various materials were in use. Figure 5.1 provides an overview of this history. In general, pit cast iron pipes are the oldest in the system. Because of the manufacturing process used, this pipe had very thick walls compared to the spun cast pipe that followed. Spun cast iron pipe was introduced in 1970's, but two distinctive types of joints were used over that time. Initially rigid joints were employed that prevented the pipe from "flexing" under stress. These rigid joints, in combination with the thinner walls compared to pit cast iron pipe, are believed to contribute to greater break frequencies for this type of pipe. Later, flexible joints were introduced that allowed the pipe to flex under stress. Most recently, ductile iron pipe has been introduced to the system, and is now used almost exclusively by the utility. Wrapping of the ductile iron pipe with a polyethylene sheet helps to protect the pipe from potentially corrosive soil conditions. This practice became standard within the utility in 1977.

Figure 5.1
The Water Company
Timeline For Pipe Installation



5.3 System Characteristics

The data were reviewed to establish the existing conditions within the distribution system. This included both the characteristics of the pipe (material, diameter, vintage, etc.) and the condition of the pipe in terms of main break history.

5.3.1 Pipe Inventory

Figures 5.2 through 5.4 present overviews of the current composition of the distribution system in terms of material, diameter, and vintage. As shown in Figure 5.2, the current system is predominantly spun cast iron (66.5%), including pipe with rigid joints and flexible joints. Combined with the older pit cast iron pipe and the newer ductile iron pipe these four categories account for over 95% of the pipe in the system. Figure 5.3 breaks down the pipe inventory by diameter, and shows that 79% of the system (approximately 3,160 miles) is either 6- or 8-inch. Pipe that is 4-inch or smaller accounts for a little more than 2% of the total. This means that there is over 80 miles of this generally small pipe in the system. Large diameter transmission mains (16-inch) account for 9.4% of the system, or about 376 miles. Figure 5.4 examines the system in terms of pipe vintage. It shows that the majority of the system (63%) is less than 40 years old. On the other hand, 17% or 680 miles is more than 60 years old.

5.3.2 Main Break History

As the distribution system has expanded over time, it is illustrative to examine break rates in terms of number of breaks per 100 miles of pipe. Since 1965 the total length of pipe in the distribution system has almost doubled, growing from 2,422 miles to almost 4,000 miles in 1997. Figure 5.6 presents the break rates per hundred miles for the utility since 1983, and shows that break rates have generally increased since that time.

The Break database contains 32,242 records of incidents that occurred between January 1983 and December 1997. However, not all of these incidents are applicable to this research. Hydrant leaks, flush valve leaks, etc. are not relevant to the analysis of water main renewal needs. Considering only those records where the LEAKCODE was CB* (various types of circumferential breaks), LB* (various types of longitudinal breaks), or CH*(various corrosion holes) yields a total of approximately 24,000 applicable records

Figure 5.2
Pipe System Composition by Material

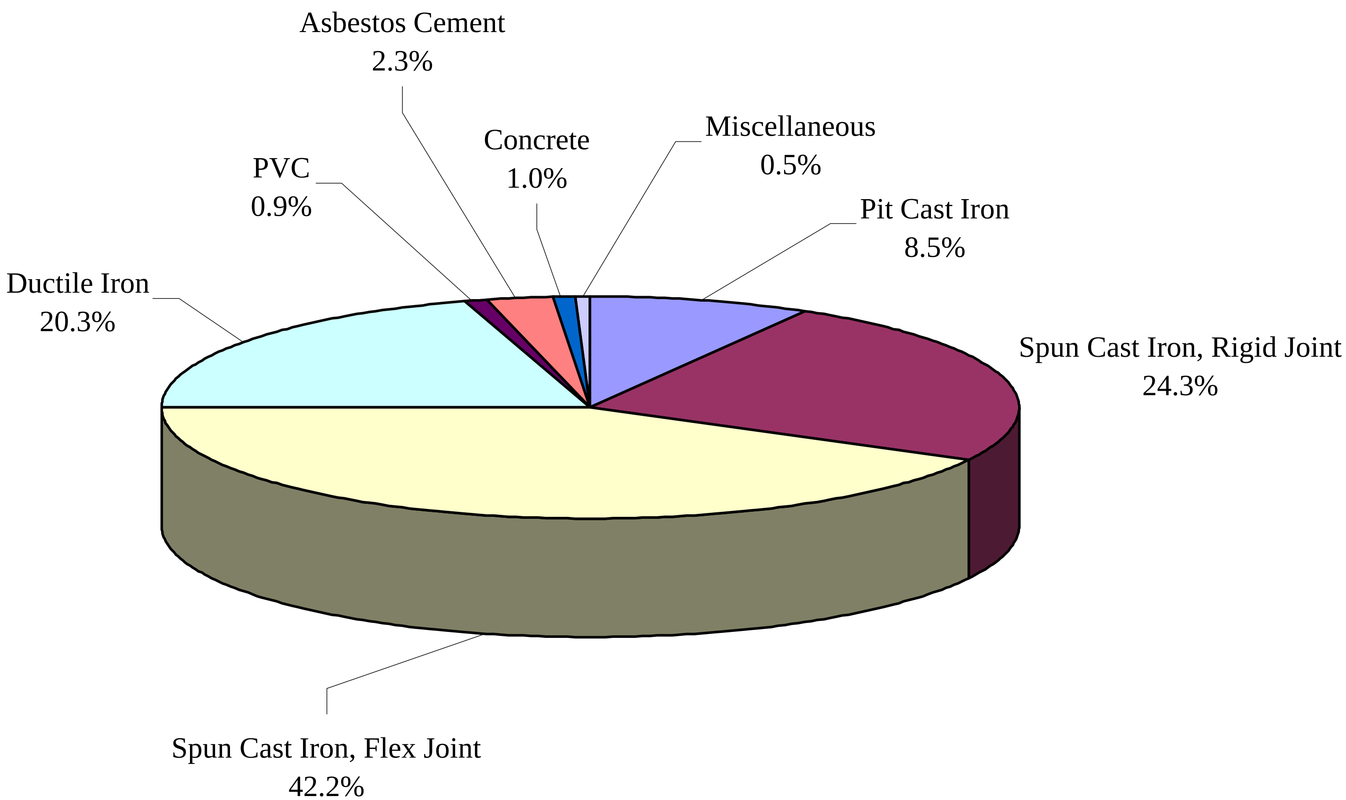


Figure 5.3
Pipe System Composition by Diameter

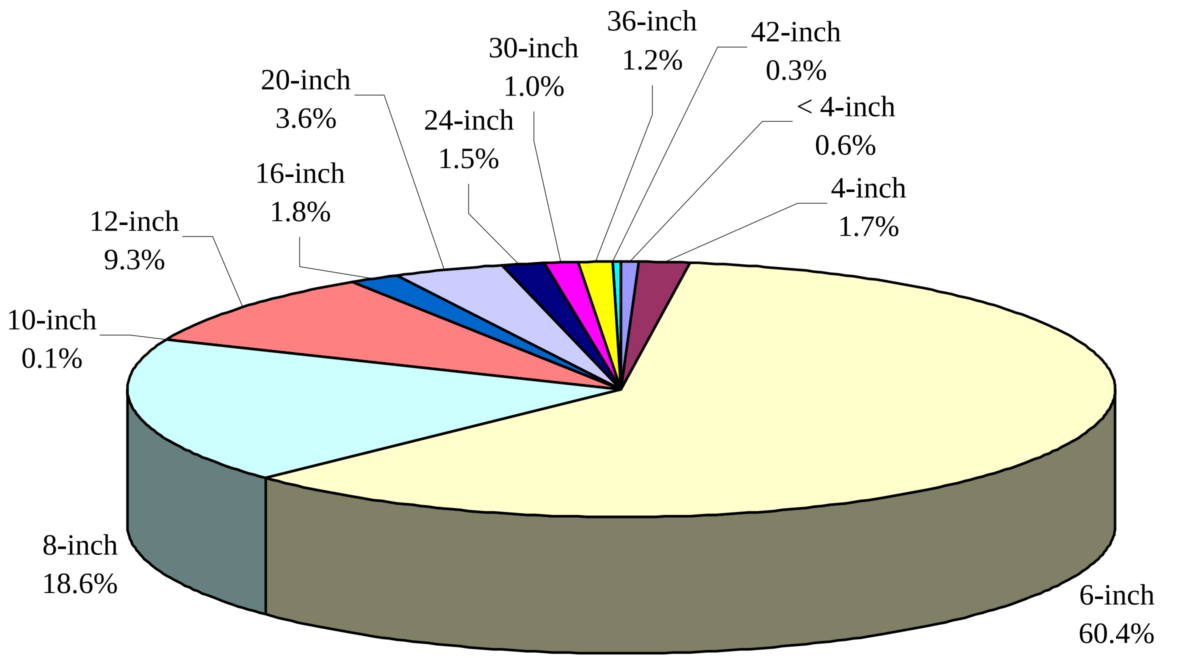


Figure 5.4
Pipe System Composition by Vintage

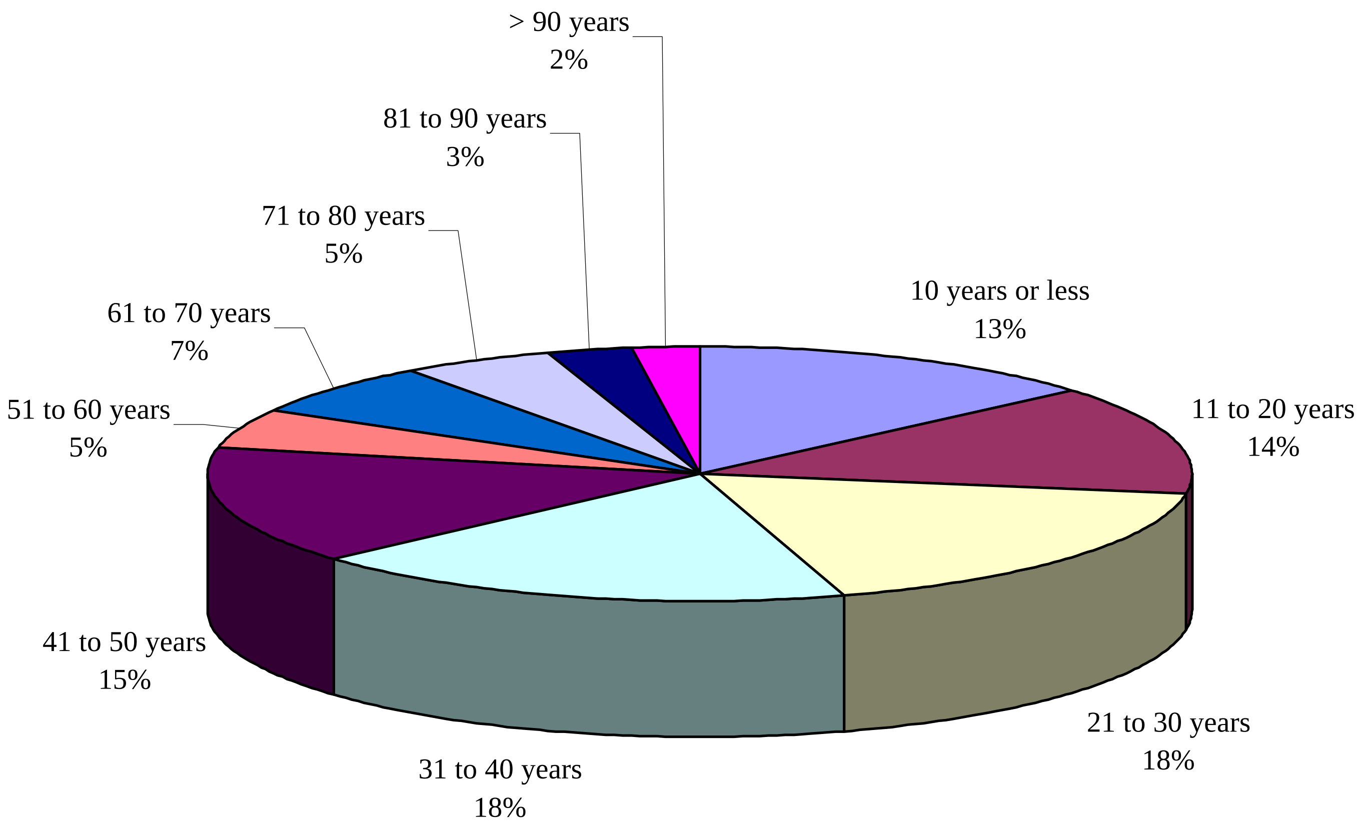


Figure 5.5
Main Breaks/100 mile since 1983

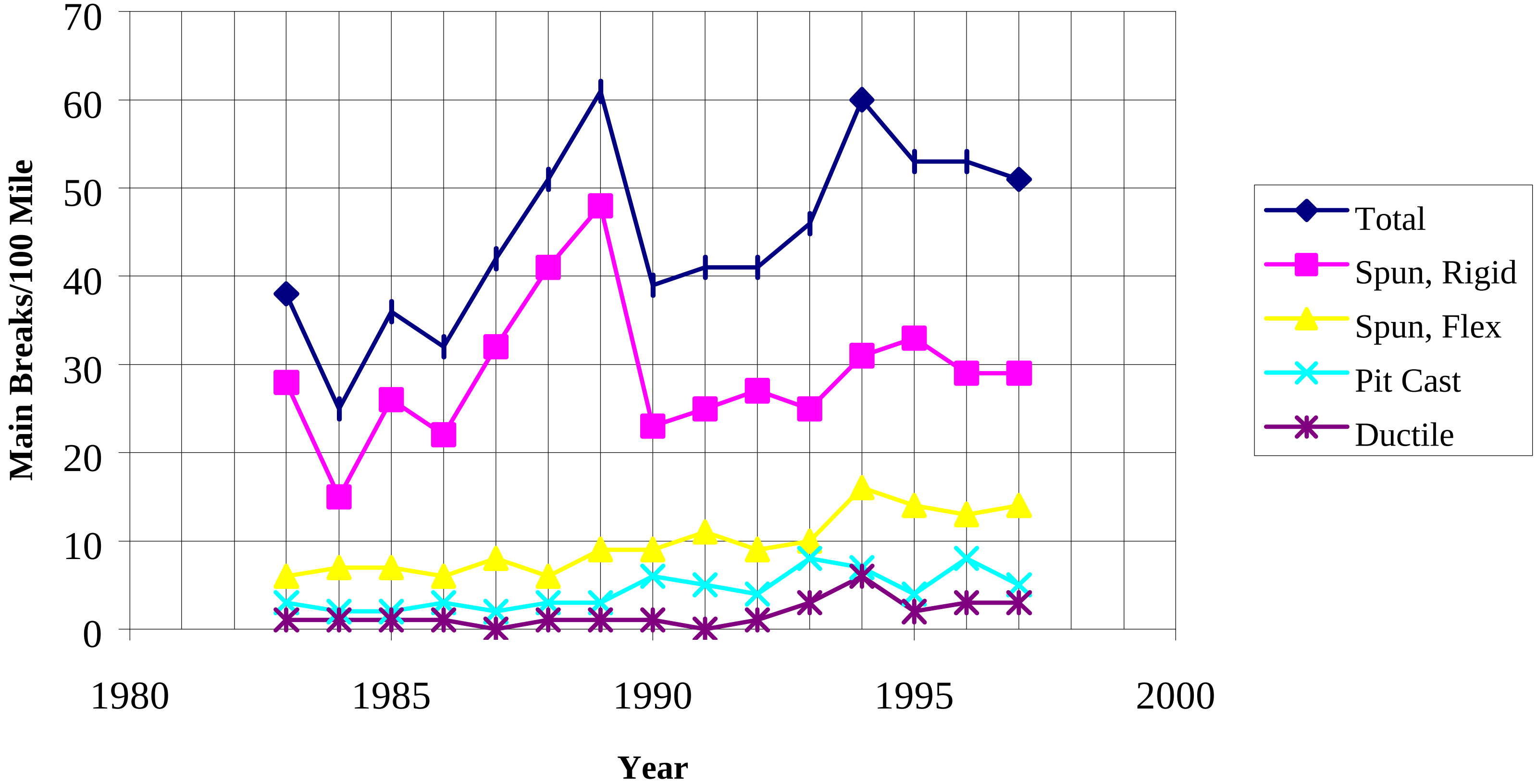


Figure 5.6 presents a breakdown of incidents from the Break database by pipe material. By far, most of the breaks occurred on spun cast iron, rigid joint pipes. Spun cast iron, flexible joint pipes accounted for the second largest number of breaks. Despite being the oldest pipe, only 6% of the incidents occurred on pit cast pipe.

A comparison of break incidents to the amount of a particular type of pipe in the system is presented in Figure 5.7. The figure shows that although spun cast iron, rigid joint pipe accounts for only 24% of the system, 60% of the main break incidents since 1983 involved this type of pipe. On the other hand, spun cast iron, flexible joint and ductile iron pipe had considerably lower percentages of breaks compared to the amount of pipe in the system.

Annual break incidents for each pipe material were also examined to identify trends over time. Figure 5.8 presents the number of incidents over time for each material, and shows that the number of incidents is generally increasing for each material. Since 1983 the spun cast iron, rigid joint pipe has annually accounted for the most incidents.

Figure 5.6
Percentage Main Breaks by Material

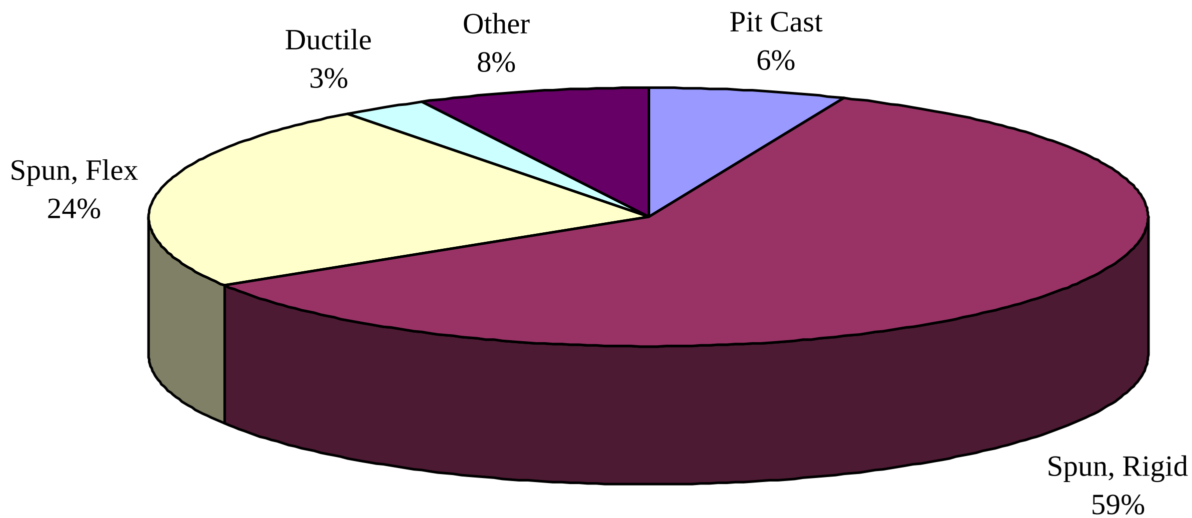


Figure 5.7
Percentage Main Breaks Since 1983 Compared to Type of Main

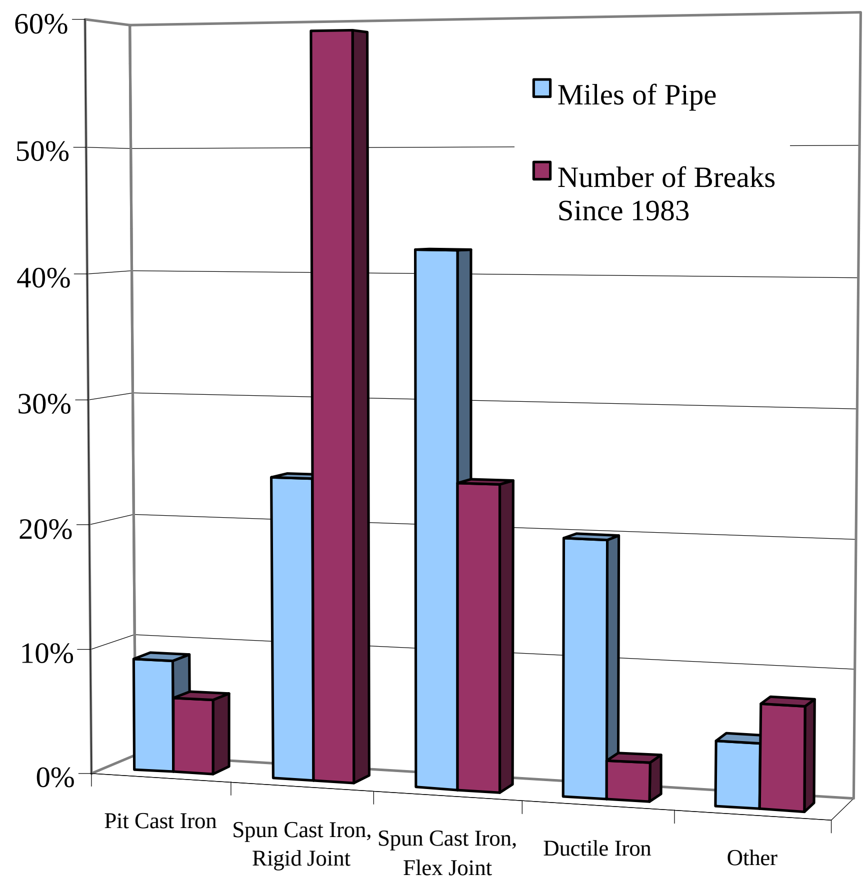
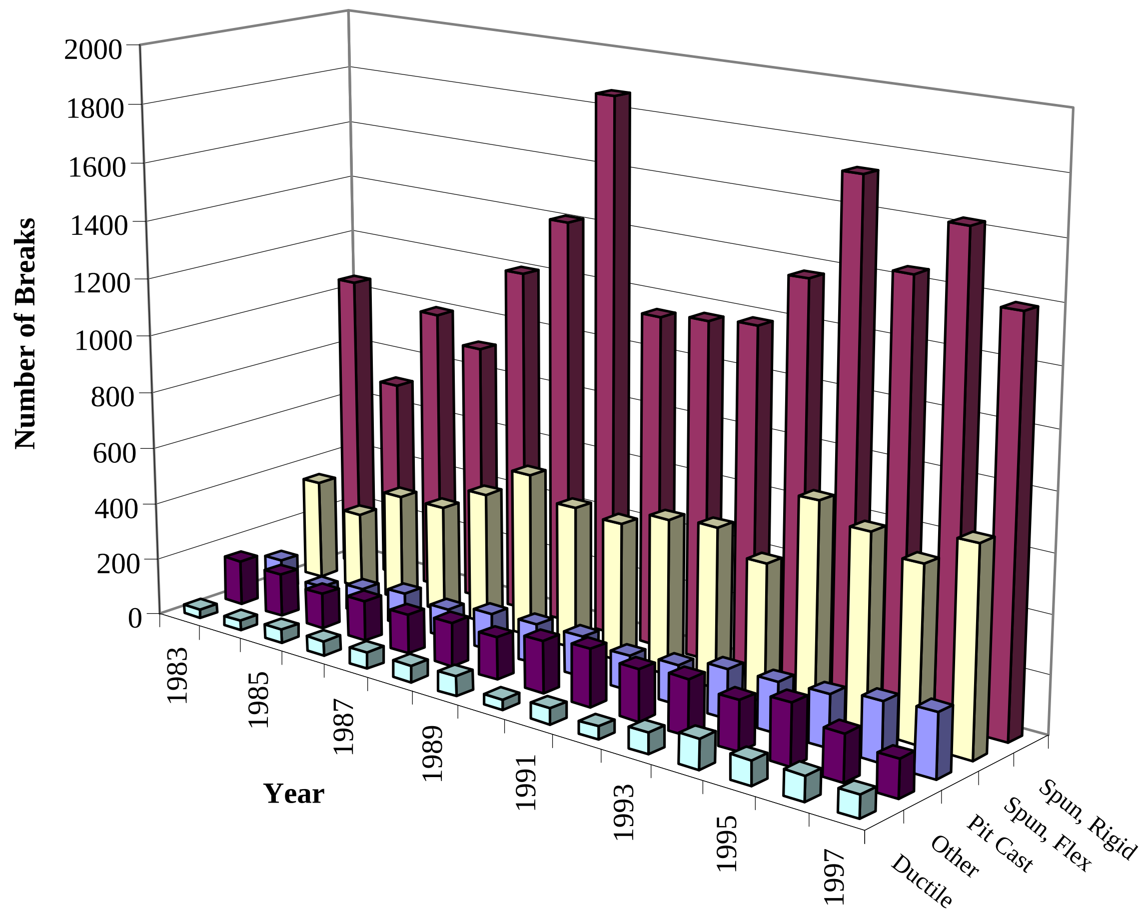


Figure 5.8
Main Breaks Over Time by Pipe Material



5.4 Modeling Basics

Analysis of pipe break data has been focused on long sections of pipe, e.g. 1000 ft or 1 mile etc, regardless of their characteristics such as size, installation time, and material. Another way to approach the analysis is to use a small or large section of pipe that is unique and identifiable. Unique and identifiable pipes must satisfy "measure of pipe identity," such as same installation date, or at least same installation year, same size, and material. Since most water utilities fail to keep accurate records for their pipe inventory and break history, a general and consistent way of defining an individual pipe has to be developed. In this research, a terminology to deal with this problem has been developed. This newly devised term is 'PIPE ID' which consists of work order, installation year, size, and material (type). Work order represents a way of identifying a series of pipes, which has been ordered to be installed for the utility in this research. This is a way of identifying an individual section of pipe in a system. Therefore, the concept of 'PIPE ID' will serve as a proper means to identify individual section of pipes for any kind of poorly maintained pipe break history database. Since the analysis based on 'PIPE ID' is able to pinpoint problematic stretches of pipes, only the specific problematic main sections can be replaced or repaired.

5.4.1 Pipe Identification

The utility does not presently maintain a pipe by pipe inventory of its distribution system. Historically, pipe installations and retirements were assigned unique Work Orders. In order to prioritize pipes for replacement, individual pipes must be defined. The utility maintains the information related to pipe breaks in terms of 'Work Order'. The Break database provides a rudimentary inventory of pipes (Work Orders) that have failed since 1983. The main break database consists of over 30,000 records and provides information on all water main breaks occurring in the system since 1983. Unfortunately, keypunch errors, data entry discrepancies, and difficulties in interpreting old paper records have led to inconsistencies in the database. An example of a typical inconsistency is shown below:

Work Order	**Installation Year**	**Material**	**Diameter**	**Length**
JAN 1970	1970	ST*	8	34285
ACQ 1970	1970	ST*	8	34285

*material type steel

In this case, both records obviously refer to the same pipe segment, but the Work Order was entered in two different ways. An initial effort was exercised to "clean up" the database by identifying records where the same Work Order had been entered in different ways, or where lengths of pipe were obviously wrong. It must be noted that the "cleaned up" database is still not perfect, and some inconsistencies remain. Since the database's inception, the utility has implemented some improvements in this area, and has plans for further refinement of the data entry process.

As the modeling progressed, it became apparent that Work Order alone was insufficient for identifying individual pipes in the database. To compensate for this problem, the concept of PIPE ID introduced earlier is used to systematically organize the

database. PIPE ID was generated using the Work Order (WORKORD), Year Acquired (YEARACQ), Material (TYPM), and Diameter (SIZM) fields from the Break database.
The use of PIPE ID (as an extension of Work Order) to identify individual pipe segments is a critical assumption in the modeling. It is unlikely that actual pipe replacements will follow the PIPE ID exactly. Rather, replacement programs will use the prioritized PIPE ID replacements as a general guide. Specific construction projects must address issues such as:

- Geographic grouping – Prioritizing pipes for replacement on the basis of economics is likely to result in a replacement schedule having pipes scattered throughout the distribution system. It is more efficient and less disruptive to organize replacement projects in limited areas.

- Coordination with other construction – Replacement projects should be coordinated with other roadway or utility construction to minimize disruption to residents.

- Previous replacements – Some or all of a pipe appearing in the Break database may have already been replaced. The lack of a "real-time" inventory of pipe in the system prevents the models from identifying how much of a pipe has been replaced. The utility's ongoing GIS development and associated asset management tools will provide the opportunity to better address this issue. The design of specific construction projects should include an investigation of prior replacements.

- Complete or partial replacement – The modeling assumed that the entire pipe shared the same characteristics. While this may be true for such basic features as diameter and material, it may not be true for break patterns. A long pipe may have many failures concentrated in a short section of pipe, and therefore, it may not be necessary to replace the entire pipe. Again, the design of specific construction projects should consider this.

5.4.2 Cost Information

Inflation and discount (interest) rates are important considerations in the optimal replacement analyses since the focus is on future events. For this modeling no inflation of repair or replacement costs is considered. An interest rate of 7% has been used to compute present worth of both repair and replacement costs. However, the models were constructed to allow a user to easily change these inputs and re-run the model.

The utility provided actual costs of water main installation and main break repairs. These costs were used in the modeling presented in this research, but can be updated whenever more representative data become available. For example, the utility is active in investigating innovative techniques for rehabilitating and replacing pipe. As these technologies become more common, the cost of pipe replacement may decrease. Costs vary by pipe diameter as summarized in Table 5.3. The threshold break rate shown in table 5.3 is obtained by assuming 1000 ft of pipe.

Table 5.3 Direct Pipe Replacement and Repair Costs

Diameter	Main Break Cost	Replacement Cost (per foot)	Threshold Break Rate
4-inch	$2,163	$90	3.12
6-inch	$3,120	$93	2.25
8-inch	$4,111	$97	1.80
12-inch	$6,597	$116	1.34