PREDICTIVE MODELS FOR PIPE BREAK FAILURES AND THEIR IMPLICATIONS ON MAINTENANCE PLANNING STRATEGIES FOR DETERIORATING WATER DISTRIBUTION SYSTEMS

by

STEFANOS ALEXANDER ANDREOU

Diploma, National Technical University of Athens

(1980)

S.M., Massachusetts Institute of Technology

(1982)

S.M., Massachusetts Institute of Technology

(1985)

SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS OF THE

DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February, 1986

Signature of Author_

Department of Civil Engineering

February, 1986

Certified by____________

David H. Marks

Thesis Supervisor

Accepted by____________

Francois M. Morel

Department Committee on Graduate Students

PREDICTIVE MODELS FOR PIPE BREAK FAILURES AND THEIR IMPLICATIONS ON MAINTENANCE PLANNING STRATEGIES FOR DETERIORATING WATER DISTRIBUTION SYSTEMS

by

STEFANOS ALEXANDER ANDREOU

ABSTRACT

Urban water distribution maintenance decisions present a very difficult multicriteria and multiobjective problem. Deterioration of those systems results primarily in breaks and leaks in pipes and also to reduction in carrying capacity from tuberculation of the interior pipe wall. Large investments are generally required for replacement and rehabilitation of water mains and it thus becomes critical to assess the current and expected future condition of the system for making maintenance decisions.

This thesis focuses on the derivation of predictive models for pipe break failures at the individual pipe level with main emphasis on the analysis of large diameter ($\geq$ 8 inches) cast-iron pipes. The applied methodologies can capture the high variability in break rates that exists among individual pipes of a given system and among various systems as a whole. The early phases of pipe deterioration, when only few infrequent breaks occur, were described by a proportional hazards type model, where the probability of failure changes as a function of time and depends on the number of previous breaks. (Non-homogeneous Markov Process.) When a pipe entered a stage of multiple and frequent breaks, the failure process was better characterized by a constant break rate estimated through an exponential type regression model and future breaks were represented as Poisson arrivals. A proportional hazards model is also applied to determine the probability for entering into the stage of multiple and frequent breaks. This detailed focus on the various phases of deterioration that a pipe can go through during its useful life consists a significant improvement over currently applied modelling techniques and provides a very useful quantitative tool of analysis that could assist water utility managers for making maintenance decisions.

The methodologies applied in this work have clearly shown that currently applied models and rules of thumb can in making cases be oversimplified and lead to very suboptimal repair, replacement and rehabilitation strategies. The integration of the proposed models in the decision making process is examined in detail and two case studies using actual break records from two water distribution systems in the U.S. are presented.

Thesis Supervisor: David H. Marks

Title: Professor and Head, Department of Civil Engineering

ACKNOWLEDGMENTS

This thesis is part of a research project sponsored by the Environmental Protection Agency (Grant No. CR810558-01-0) entitled "Maintenance Issues in Water Supply Networks." Dr. Robert Clark and Mr. Jeff Adams, EPA Officials, provided significant insights and comments for the analysis of the data sets.

Professor David H. Marks, my thesis supervisor, gave me ample encouragement and inspiration throughout this project. I feel fortunate to have been his student and I thank him for his constant interest, support, and guidance. I have also greatly benefitted from the advice of Professors J. Sussman and Y. Sheffi, who served on my Committee and their assistance is gratefully acknowledged.

An especially important contributor to this thesis was Professor W. DuMouchel, of the MIT Statistics Center. Without his valuable insights on the application of the statistical methodologies used in this work and his very useful suggestions, completion of it would have been impossible.

The assistance of C. Park and A. Zaslavsky, doctoral students at the MIT Statistics Center, has helped me overcome several difficult problems that I encountered during this research. I am very grateful to them for their ideas and constant effort while working with me on this project.

Special thanks go to Jackie Winton, and Janet Mahoney, for their excellent typing.

I would like to thank my friends at the Parsons Laboratory for making my stay here enjoyable. Those who have shared an office with me -- Aristotelis Mantoglou, Aris Georgakakos, Fadi Karaa, Jorge Ramirez, Carlos Puente, Alvaro Aldama and Mike Bruno deserve special mention.

Finally, I am indebted to my parents, Alexander and Helen, for their unfailing love, support and enccuragement. To them this thesis is dedicated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

CHAPTER 1

1.1 Introduction

Increasing public awareness about the decay of the nation's infrastructure caused by accumulated evidence of progressing deterioration of the physical condition of many urban systems (e.g., transportation, water supply, sewers), creates the need for more careful scientific studies to understand the failure patterns and propose methodologies to deal with the associated problems. Among the various systems that constitute the infrastructure of urban centers, water distribution systems play an essential role for the every day functioning of a modern city. The condition of water mains and the reliability of services they provide are strongly connected with community public health standards and potential for future growth and economic development. Two different types of problems are causing great concerns today in many water utilities; reduction in carrying capacity of pipes due to tuberculation, and breaks and leaks occurring at increasing rates along the pipe sections. Those problems are directly related to increasing future repair and pumping costs, unreliable services, potential for damages caused by breaks (flooding, disruptions in traffic, and operations of other utilities), and water quality problems caused from bacteria trapped in the tuberculous formed in the interior pipe wall. Remedies for the above problems consist of replacing or rehabilitating (i.e., cleaning and lining) sections of pipes in the system and represent large and important capital investments made by the water utilities. In order to make repair, replacement, and/or rehabilitation decisions for deteriorating pipelines, it is essential to develop insights about the failure mechanism and the interactions of the various factors contributing to breaks, so that predictions on the evolution of break trends with time could be made under various maintenance scenaria. However, great complexities arise when one attempts to analyze and predict the future behavior of individual pipes in a system, because of the high variability in failure patterns among different systems and among the various pipes of a given system. Nevertheless, such analysis at the individual pipe level is what is clearly needed for making maintenance decisions, under particular economic and reliability criteria. Although on site observation of the physical condition of water mains could reveal some information about their structural integrity, such an approach would be of primary use only when very severe deterioration is present. Otherwise, the localized nature of the various factors contributing to breaks, their interactions and the effect of the aging of the system cannot really be captured by inspections of particular points along the pipe length. Thus, there is a need for analyzing historical records of pipe breaks, and make use of all the available information on pipe characteristics and external environmental conditions, in order to develop a basic understanding of the existing patterns. Assessing with reasonable accuracy the expected number of future break events for each individual pipe can serve the following important purposes: a) determine the budgetary needs for future repairs under various replacement and rehabilitation strategies, b) perform economic evaluation to determine an optimum replacement time for breaking mains,

c) obtain estimates of the reliability of individual segments, which would be of great importance for certain pipes in the system, depending on their position in the network and the potential damages associated with their failures.

Although no clear definition of pipe breaks exists in the available data sets, the threshold between breaks and leaks usually considers that breaks, as opposed to leaks, result in a clear disruption of service and require some kind of repair action. Pipe leaks are associated with the problem of "unaccounted for water" and represent a distinct category of failures, different from that of pipe breaks. Nevertheless, it is believed that leaks are associated with breaks, since they can undermine the bedding material underneath the pipes and create localized concentration of stresses. Also, a clear distinction in failure patterns exists between small (less than or equal to 6 inches) and larger diameter pipes. Smaller pipes break more frequently, they do experience primarily circumferential breaks (ring cracks) and are clearly affected by weather induced temperature changes in the soil (frost penetration). The predominant failure mode of larger diameter pipes is the occurrence of longitudinal breaks and no seasonality patterns (associated with weather conditions) are observed.

Breaks occurring on larger diameter pipes can have major consequences on the quality of services that the water utility provides to its customers and can result in traffic disruptions, street and basement flooding, interference with the operation of other utilities (gas pipelines, sewers, cables), and disturbance in subway operations. Large diameter pipes can also become candidates for rehabilitation, since for small diameter pipes the rehabilitation alternative is not usually considered as economically efficient. Thus, being able to predict the expected repairs for such pipes, would be an important factor in the economic analysis for deciding whether they should be rehabilitated or replaced. In certain systems, where an advanced degree of general deterioration is observed, a small number of large diameter pipes also start experiencing very frequent breaks (i.e., a pattern similar to the one observed in smaller size pipes). It thus becomes critical to study the conditions under which the potential for entering such stage increases, since decisions for replacing or rehabilitating them consist of large capital investments.

The focus of this work is primarily on large diameter pipes (8 inches and above), although the methodologies proposed and insights generated could have extensions on decisions affecting smaller pipes.

1.2 Problems and Deficiencies of Currently Applied Repair/Replacement/ Rehabilitation Strategies and Proposed Methodologies for Analysis

Current knowledge about the failure mechanism in water mains lacks a deep understanding of the interactions of the various factors contributing to breaks. It is indeed extremely difficult to model the synergistic effects of corrosion (external and internal), improper bedding conditions, stresses from external loads, stresses from high internal pressure and properties of particular pipe materials that have changed through the years. Thus, although several factors associated with breaks can be identified, there is no coherent model describing the changes in the physical condition and structural integrity of a main with time, under the presence of those factors. This situation has laid to the development of several rules of thumb for assessing the future performance of deteriorating pipes and making replacement/rehabilitation decisions. Statistical techniques based on historical break records have also started to be applied, in order to develop greater insight about the failure patterns and quantify with greater accuracy existing trends. Many of the currently used rules of thumb recommend for example replacements based solely on pipe age and number of previous breaks, while clearly more scientific work is needed in order to examine whether reliance on such measures is justifiable.

Due to the complexity of the pipe breaking mechanism and the high variability in break rates existing among different pipes and water systems, statistical studies and attempts to obtain predictive models for future breaks, usually have failed to capture the detailed failure patterns on individual pipes and very often have transmitted confusing signals about the effect of pipe aging on the break rate.

Incomplete information on pipe break records, environmental conditions, pipe characteristics, and operating and maintenance practices, generates additional uncertainties about how the available data should be handled. Questions can also be raised concerning the reliability of existing data and the extent to which they can be used for making inferences about real trends in break patterns.

The two basic types of predictive models that have been proposed in the literature (Shamir and Howard, 1979; Clark et al, 1982) have oversimplified the characteristics of the failure process to an extent that leads to serious discrepancies between assumed in the model, and actual breaking behavior of individual mains. The underlying assumption in those models considers an increase of breaks with time, usually of an exponential form. No distinction is made in the basic model structure as to whether a pipe experiences very few unfrequent breaks or is in a frequently breaking mode. Also, often pipes of small and large diameter have been mixed together although they correspond to very distinct failure patterns. A more detailed review and critique of the methods of analysis proposed in previous studies is presented in Chapter 2.

The great variability in break rates among various cities in the U.S. (Table 1.1) suggests the need for developing a methodology that both relies on the basic dynamics of the failure process and possesses the flexibility to be adapted to the particular system characteristics. As far as large diameter pipes are concerned, because of the high consequences of breaks associated with them and the very often infrequent occurrence of such events, a probabilistic predictive model would appear

City	System Mileage	Number of Breaks (In Latest Year)	Breaks Per 1000 Miles	Type of Break: Pipe and Joint	Type of Break: Pipe Alone	Type of Break: Uncertain*	Average Age of Broken Pipe (Years)	Respondent's Judgment of Major Probable Causes of Breaks
1. Houston	3,988	5,146 (1973)	1,290	---	X	---	70	Shifting of soil, corrosion
2. New Orleans	1,426	730 ("average")	512	---	X	---	NA	Soil subsidence, corrosion
3. Detroit	3,422	1,309 (1973FY)	382	---	---	X	NA	----
4. Milwaukee	1,800	421 (1973)	234	---	X	---	37 to 100	Corrosion, stress, substandard material, temperature
5. Philadelphia	3,242	699 (1973-74)	216	---		X	74	Corrosion, temperature
6. Kansas City	1,924	403 (1973-74)	209	X	---	---	NA	NA
7. Baltimore	1,533	261 (1973)	170	---	---	X	NA	NA
8. Denver	1,793	280 (1973)	161	---	X	---	25	Corrosion, temperature
9. District of Columbia	1,350	171 (1973)	127	---	X	---	50	Temperature
10. San Francisco	1,176	125 (1973)	107	---	X	---	58	Construction activities, corrosion, installation, temperature
11. New York City	6,145	494 (1973)	80	---	X	---	83	Temperature, impact, age, utility interference
12. St. Louis	1,373	106 (1973)	77	---	X	---	NA	Temperature
13. Chicago	4,148	223 (1973)	54	---	X	---	71	Corrosion, settlement
14. Boston	1,077	54 (1973)	50	---	---	X	3 to 120	Pipe resting on solid object
15. Los Angeles	6,800	290 (1973-74)	43	---	X	---	NA	Corrosion, temperature
16. Seattle	1,693	21 (1973)	12	---	---	X	50	Temperature, pipe resting on solid object, traffic, excavations
Total - 16 Cities	42,890	10,731	250**					
Total Less Houston	38,902	5,587	144**					

NA = Not Available

* Not clear if data include pipe only or pipe plus joint failures

** Average value

Source: Report on Select Committee on Water Main Breaks in the City of New York (November 1977)

Table 1.1

WATER MAIN BREAKS IN 16 MAJOR CITIES

intuitively to be very useful. Knowing with some reasonable accuracy the failure probabilities of individual links in the network would provide water utility managers with the necessary quantitative measure for assessing the reliability of the provided services. Since in many cities with high economic activity (industrial, commercial, transportation), breaks in larger pipes can be associated with very serious consequences, the ability to estimate probabilities of failure becomes critical. Table 1.2 shows the watermain breaks damage claims in the city of New York. Given an average annual claim settlement cost of $500,000 for damages caused by breaks, it really becomes important to incorporate in the planning process for future capital investments, probabilistic estimates of the expected evolution in breaks under different maintenance scenaria.

1.3 Hypotheses Tested in this Work

The basic hypothesis developed and tested in this study is that in order to analyze and model historical break records of pipelines in water distribution systems a methodology based on proportional hazards models for analyzing failure time data would perform better than currently applied techniques. The techniques used in the past have proved inadequate in capturing the combined effect of the various factors contributing to breaks and their evolution with time at the individual pipe level. Thus, a methodology that possesses the theoretical properties that are appropriate for describing the underlying failure process and also avoids the major deficiencies of past methods of analysis, is being applied. The proposed methodology has its theoretical foundation

WATER MAIN BREAKS DAMAGE CLAIMS
1963-1964 - 1978-1979
(Five Boroughs)

Fiscal Year	Claims Filed: Number	Claims Filed: Amount Claimed $	Claims Settled: Number	Claims Settled: Amount of Claim $	Claims Settled: Amount of Settlement $	Percent of Claim
1963-64	127	1,188,516	163	1,370,370	514,078	37.6
1964-65	192	1,330,986	97	814,593	196,826	24.2
1965-66	325	4,427,823	120	1,723,736	619,514	35.9
1966-67	91	553,967	174	1,306,078	223,159	17.1
1967-68	170	1,661,851	83	537,454	152,482	28.4
1968-69	197	1,208,492	107	1,112,623	354,896	31.9
1969-70	127	1,146,227	103	642,613	172,487	26.8
1970-71	172	5,586,451	87	1,365,562	433,168	31.7
1971-72	183	5,518,397	142	956,745	301,075	31.6
1972-73	246	3,559,655	114	1,287,890	482,547	37.5
1973-74	149	2,511,419	155	4,639,145	714,127	15.4
1974-75	304	9,323,406	129	2,333,190	586,416	25.1
1975-76	212	3,150,488	111	706,983	503,951	71.3
Total	3,078	49,366,692	1,872	21,662,070	6,315,506	
Average	205	3,291,112	125	1,444,138	421,034	29.1

Table 1.2

WATER MAIN BREAKS DAMAGE CLAIMS,
CITY OF NEW YORK, NY

on statistical techniques developed for the analysis of failure time data and is believed to be able to reflect in much greater accuracy the various stages that a deteriorating water main is going through during its lifetime.

More specifically, the following problems and implications of the proposed technique are investigated by this study:

a. The failure mechanism in water pipes is very complex and cannot be accurately described by the rigid model structure of the methodologies currently proposed in the literature. In order to capture the high variability in break trends existing among different systems and among different pipes of an existing system, the focus of research should shift in modeling in greater detail the various phases of failure that an individual pipe can go through and it is believed that a methodology based on Cox's (1972) proportional hazards model would be more appropriate than currently applied techniques for achieving that goal. This methodology possesses high 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, than logistic type models or discriminant analysis. The basic hypothesis developed by other investigators that the break rate increases exponentially with time is challenged, at least for the larger diameter pipes. The hypothesis developed here states instead that a deteriorating pipe goes through various stages (or modes of failure) during its lifetime. Although a close to exponential relation between breaks and time could exist in the early stages of failure, it usually reaches a steady-state of multiple failures beyond a certain time period. Of course, a very small percentage of pipes in the whole system are likely to reach that state and it thus becomes very critical to identify pipes with such potential.

b. Missing break records in many data sets before a certain time period in the past pose serious questions about the validity of the derived predictive models based on such records. The question of using break record observations truncated from the left (left censored observations) in order to estimate failure probabilities in pipes is investigated by this study.

c. Decisions for repair/replacement/rehabilitation and reliability assessments could be highly affected by the proposed failure modelling methodologies. A problem that is investigated is how currently applied rules of thumb for pipe maintenance could lead to uneconomic, inefficient and unreliable strategies. Also the proposed methodologies would result in different replacement scenaria than the currently recommended based on the existing models. The economic and reliability implications of those changing replacement scenaria are thus being investigated and the integration of the proposed predictive models with the decision making process of a water utility is being examined.

d. The important question regarding the applicability of the developed methodologies in water distribution systems where break records are poor to perform any kind of statistical analysis, is examined. The detailed analysis performed in this study of two quite different systems in terms of failure patterns (one with much less frequent breaks than the other), is expected to set the stage for categorizing existing systems in terms of the severity of deterioration and indicate the appropriate methodologies for studying each particular system.

e. Besides any implications on maintenance practices, system operation and construction practices could also be affected by the results of this investigation. Depending on the explanatory power of the derived predictive models, covariates reflecting for example internal pipe pressure, or pipe material, could suggest changes in the way particular sections of the system need to be operated and also make inferences on design standards applied today and in the past periods.

The following pages are missing from the original

#26

2.2 Descriptive Statistical Studies for Deteriorating Water Mains

Descriptive statistical studies have been performed on several occasions for analyzing pipe breaks. Although such studies have revealed useful trends in the behavior of deteriorating pipes, many questions concerning the complex failure phenomenon of individual pipes have remained unanswered. Particularly, issues concerning the effect of aging on the break rate and the ability to provide reliable quantitative measures for assessing the condition of individual mains given their past break history and other pipe and environmental characteristics, have not been adequately emphasized and clarified in most of those studies. In general, because such studies have usually not focused on the analysis of failures at the individual pipe level, but rather in identifying trends in the overall system, they tend to obscure the high variability in failure patterns that exists among different pipes of a given system. The results of the most representative studies of this type are briefly presented and discussed in this section.

An extensive statistical analysis of main break failures in the city of Philadelphia has been performed by O'Day (1983). The most important findings related to break trends can be summarized as follows:

- Break rates were increasing at 1.8 percent per year since 1930.

- Large diameter ($\geq$ 16") main break rates have remained relatively constant since 1910.
- The rate of water main breaks increases significantly in the winter months, but also non-winter breaks had increased sharply in the past 20 years, at an annual rate of 2.5 percent.
- Break rates in small diameter mains are much greater than for larger diameter mains. Also, in terms of break type, circumferential breaks represented 71 percent of the breaks in the 6" diameter pipes, dropping to 34 and 31 percent for the 12" and 16" mains, respectively. Longitudinal breaks varied from 18 percent in 6" pipes to 47 percent for 10" pipes. No information is given about the relation between pipe size and breaks attributed to "holes" in pipes.
- An unusually high break rate has been observed for pipes installed during the 1940-1960 period. An important factor, believed to have contributed to such a high break rate, is the use, during that period, of a leadite joint material, which causes a joint to become rigid, instead of a lead material that was used in the past and which allowed the joint to be more flexible. Thus, a leadite joint material could restraint the pipe from thermal contraction or extraction and cause stresses that result in mechanical failure of the pipe.

- A strong correlation was found between break rate and residential development. Also, a great geographical diversity in break incidences existed across the city.

- The investigation of the relationship between main age and break rate did not lead to any satisfactory conclusion in terms of the effect of age on pipe failures. Although, older mains, on average, were found to experience higher break rates, it is argued that age should not be used as a criterion for replacement.

- Internal corrosion of pipes was found to have significantly reduced the wall thickness of most of the pipes examined in a sample test and also it was generally found deeper than external corrosion. It has thus been established that internal corrosion not only leads to the formation of tubercules which will block off the flow in the main, but it also contributes significantly to the undermining of the structural integrity of a pipe.

Another statistical analysis on main break patterns was performed for the city of Buffalo, New York, by the U.S. Army Corps of Engineers (1981). Its major findings revealed again a very seasonal pattern in main breaks with the majority of breaks also being in the smaller diameter pipes. The only difference in seasonality patterns found between that study and the one performed by O'Day for the City of Philadelphia, is that in Buffalo, NY, the break rate increased not only during the winter months but also during the summer months of high demand (June, July). The reason for that phenomenon happening could be found in different operating practices implemented at those two systems. For example, increased stresses in pipes during the summer months in the Buffalo system could be related to increases in operating pressures in the system in order to meet demands during that period. An attempt in the Buffalo study to relate high numbers of breaks with high annual average daily traffic volume did not lead to any satisfactory conclusion. Also, no information is given on the effect of other factors, such as main age, on the break rate. Statistical analysis of water main breaks has also been performed for the City of Manhattan, New York. The study was conducted by K. O'Day and G. Huguet (1980) for the New York District of the Corps of Engineers. The break records analyzed were those that occurred during the past 25 years. A detailed computer data base was created that included the break records and information on pipe location, diameter, length, date laid and number of hydrants. Discriminant analysis has been applied to identify environmental and pipe characteristics that could contribute to breaks. The study concluded that no consistent pattern of increasing breaks as pipes get older existed. Also, there was no evidence to show that the pipe material was deteriorating with age. Smaller diameter pipes ($\leq$ 6 inches) had the highest break rate and were primarily experiencing ring cracks. Although no data existed to confirm this, the study team concluded that leaks could be the most important factor causing main breaks, since, depending on the soil condition, the bedding material could be washed away when leaks occur. Their analysis pointed out that high break areas were usually related to high levels of other activities (heavy traffic, major reconstruction, subways, other underground utilities). The results of the Manhattan study concerning the application of discriminant analysis as a predictive tool for pipe breaks are discussed in Section 2.3.3.

In general, most of the statistical studies performed on various cities (New York City, U.S. Army Engineer District (1980); Cincinnati, Ohio, Clark et al. (1980); Binghamton, NY, Walski et al. (1982); Philadelphia, O'Day (1983); Buffalo, NY, U.S. Army Corps of Engineers (1981)), have observed a high break rate during the winter months. External forces on pipes induced by frost penetration and restraints caused by thermal contraction, are considered to be the primary causes of this trend. Clearly, those forces will depend on factors such as depth of cover of the pipe, soil type, joint and pipe material. There has not been any model thus far that can quantify the stresses caused under all those factors acting together on a water main.

A general controversy, although, exists in the literature about the effect of age on pipe failures. As previously mentioned, the

Manhattan study for the City of New York (1980) and the study by O'Day for the City of Philadelphia (1983), basically argue that the age of the pipe must be only a minor consideration in the main replacement program. The Manhattan study found that the mains were not wearing out with age. O'Day found a very weak correlation between break rate and age of mains when he looked at 6" diameter mains and he also found a higher number of breaks on average for older mains when he aggregated all pipe segments together. Clark et al. (1980), on the other hand, in their attempt to derive a predictive model for pipe break failures found that for the City of Cincinnati, Ohio, the age of metallic (cast iron and steel) pipes was an important factor in predicting future breaks for a main segment.

The probable reason for this controversy lies in the fact that the relationship between age of the pipe and break failures is a very complex one and cannot be captured by simplified statistical analysis techniques. Although intuitively it seems that age related factors, such as pipe materials used, design and construction practices at the time the pipe was laid, and deterioration due to corrosion, could result in a relationship between age and break rate, its form has not been established in the literature. What also appears to be happening is that certain pipes will experience breaks very early after installation due to defects in construction materials and/or particular localized unfavorable conditions causing excessive stresses in the pipe. The effect of aging should be expected to contribute to the failure rate much later in the life of the pipe. Thus, the age-break rate relationship is expected to be a non-linear one and possibly not a monotonic function.

Further investigation of this relation and the way it can be determined using the methodologies proposed in this thesis, will be examined in detail in later chapters.

The descriptive statistical studies for deteriorating water distribution systems have provided us with great insights about the failure patterns and possible break causing factors of water mains But such analyses also have major weaknesses which can be summarized as follows:

a. They do not provide useful information for the failure behavior of individual pipe segments.
b. The complex interaction among the various break causing factors cannot be captured.
c. Non-linearities in relations between the break rate and other factors (e.g., age) are difficult to reveal.
d. They very often provide an overwhelming amount of statistical information which is not clear on how it can be used for repair/replacement/rehabilitation decisions at the individual pipe level.

2.3 Predictive Models for Water Main Failures

The need to obtain a quantitative decision tool for making repair versus replacement decisions for deteriorating water pipes led to the derivation of predictive models for pipe break failures. There are three basic categories of such models developed thus far: Aggregate type models, where the expected number of breaks is a function of time t since a reference time period and certain constant model parameters; Regression type models, where the expected number of breaks, or the expected time to the next break, are predicted as a function of certain independent variables reflecting environmental conditions and pipe characteristics; Probabilistic or choice models, where discriminant analysis is applied on the data, and failure time models, where a survivor function is estimated for each individual pipe, which provides the probability that a pipe will survive without breaks beyond time t, as a function again of a number of independent covariates related to environmental conditions and other pipe characteristics.

2.3.1 Aggregate Type Models

Representative of this category are the models proposed by Shamir and Howard (1979). Two equations were used (one linear and one exponential) to describe the break rate as a function of time:

$$N(t) = N(t_o)e^{A(t-t_o)} \tag{2.1}$$

and

$$N(t) = N(t_o) + A(t-t_o) \tag{2.2}$$

where N is the expected number of breaks in year t per unit length, t_o is the base year time, A is a growth rate coefficient.

In this type of model, pipes with "similar" external and internal characteristics are pooled together and a function of the type (2.1) or (2.2) is fitted to them. An exponential growth model (Equation 2.1) is the most commonly applied.

The values for A proposed by Shamir and Howard (1979) are in the range of 0.01~0.15. Clark (1982) found a value of 0.086 and Walski (1982) reported values of 0.021 and 0.014 for pit cast iron and sandspun cast iron pipes respectively, when a similar modelling approach was applied by them on other data sets.

The Shamir and Howard models consisted of the first attempt to statistically analyze break records and apply the results for making better maintenance decisions. Their major advantage is their simplicity but they also have major drawbacks which can be summarized as follows:

a. They do not differentiate for various pipe characteristics, environmental factors and previous break history of individual pipe segments. Thus, they do not help in developing insights about the break causing mechanism and the key factors contributing to main failures.

b. As a general rule, studies where those models are reported, do not give any information about the goodness of fit tests and the statistical significance of the coefficients of those models. The great variability at the break rate of individual pipes, that exists in many of the studied systems, makes us believe that, such models, where all the break causing factors and failure dynamics are lumped in a few parameters, would give very unreliable estimates at the individual pipe level.

Thus, it is fair to believe that application of those models for making repair versus replacement decisions of water mains could lead to suboptimal replacement strategies, since their focus is not at the individual pipe level. The only advantage of such models is their simplicity and the smaller amount of data required as compared to other regression models. It is also plausible that they could provide better accuracy when applied to small diameter pipes (6" or less), where breaks are much more frequent than in larger diameter pipes and thus a growth rate for failures could possibly be estimated with greater accuracy. Nevertheless, the drawbacks previously mentioned will still hold, and any replacement decisions that are based on such models could be very misleading. It must be pointed out though that many investigators have proposed and used such models for pipe replacement decisions in the past (Walski, 1982).

2.3.2 Multiple Regression Type Models

The only model of this type reported in the literature is the one developed by Clark et al (1982), where the following two equations have been proposed:

First Event Equation

$$NY = 4.13 + 0.338D - 0.022P - 0.265I - 0.0983\ RES - 0.003\ LH + 13.28T \quad (R^2=0.23) \tag{2.3}$$

where:

NY = number of years from installation to first repair

D = diameter of pipe in inches

P = absolute pressure within a pipe, in pounds per square inch

I = percent of pipe overlain by industrial development in a census tract

RES = percent of pipe overlain by residential development in a census tract

LH = length of pipe in highly corrosive soil

T = pipe type (1 = metallic, 0 = reinforced concrete)

Accumulated Event Equation

$$REP = (0.1721)(e^{0.7197})^{T}(e^{0.0044})^{PRD}(e^{0.0865})^{A}(e^{0.0121})^{DEV} \cdot (SL)^{0.014} \cdot (SH)^{0.069} \quad (R^2=0.47) \tag{2.4}$$

where:

REP = number of repairs

T = pipe type

PRD = pressure differential, in pounds per square inch

A = age of pipe from first break

DEN = percent of land over pipe, which is developed

SL = surface area of pipe in low corrosive soil

SH = surface area of pipe in highly corrosive soil

It appears that in the above models the goodness of fit test (as indicated by the estimated values of R^2) shows a rather unsatisfactory fit of the models, particularly for model (2.3), and considering the fact that we do not know how statistically significant the estimated coefficients are. Also only the partial correlations of the independent variables with the dependent ones were given in the original study. It can thus be seen which variables have a stronger impact in the regression equation, but there is no way to evaluate the statistical significance of each variable appearing in the equation. Thus, the calculated R^2 could be artificially high.

Clearly, this type of model can generate significantly greater insights about the key factors related to failures in water mains and can also be much more helpful for replacement decisions than the models proposed by Shamir and Howard (1979). Of course, problems associated with the availability of appropriate data can limit their application in real world problems.

Some additional problems associated with the development of this kind of model can be:

a. A search is required for model identification.
   That is, a decision has to be made about the appropriate model structure, that will capture the dynamics of the system and the interaction among the various factors affecting the break rate. It becomes obvious from the

discussion about the models suggested by Clark et al. (1982) that serious questions exist as to whether the proposed model structure is a satisfactory one.

b. The effect of aging in the pipe break mechanism is not captured satisfactorily since the function between time and break events has to be specified apriori in a rather arbitrary way. Since there is little evidence about the proper shape of this function, the above fact provides a limitation of this type of model.

c. Information about pipes which have not broken (right-censored data) is not used effectively in the prediction equations. That is, the fact that the data can be censored from the right is not accommodated in this analysis. It is well known though, that a pipe may not have experienced a break since installation and still provide useful information of the probability of failure of other pipes with similar characteristics.

d. Pipes with small and large diameters along with pipes with very few and very many breaks, and pipes of different materials (metal and concrete) were mixed together in the developed regression equations. This can very well obscure the variability actually existing in failure patterns among individual pipes, since the model structure does not differentiate among the various failure modes that can exist in the pipes of a given system (e.g., less deteriorated vs. severely deteriorated pipes).

e. No distinction is made in the model of differences in pipe lengths corresponding to each particular observation. In other words, it is implicitly assumed that the risk factors (break causing factors) were uniformly distributed along the pipe length, no matter how highly varying this variable could be. It appears that such implicit assumptions could very well be contradicted by reality and thus a more sophisticated treatment of length would be required in the regression analysis.

2.3.3 Probabilistic Predictive Models

The focus in such models shifts from trying to predict a future break rate, into trying to estimate the probability that a break will occur at some future time and/or the probability of an item (a pipe in our case) to enter a particular state (e.g., severe deterioration with multiple failures).

Although several techniques for deriving probabilistic predictive models have been developed in the statistical literature, only discriminant analysis has been applied in the past to pipe failure data (Manhattan Study, 1980). The major findings of this study were described in Section 2.2. Regarding the application of discriminant analysis by that study the following observations can be made:

a. The analysis has been limited by classifying pipes in only two groups, those that have broken and those that have not. It thus fails to capture the evolution of failure patterns on individual mains, which could, for example, start from very few infrequent breaks and later advance into a multiple failure stage, characterized by frequent breaks. Thus, no use of the previous history of breaks is made, for making projections about future expected failures, conditional on the past performance of a pipe.

b. No useful quantitative descriptions of the effect of aging in pipes were possible to be made, and thus the predictive power of the derived models would be limited for longer time periods.

c. Since discriminate analysis was performed separately for each pipe diameter and the variables found to be statistically significant were highly varying for each category, general conclusions about the particular factors contributing to breaks were difficult to make. As it has also been argued in that study, the accuracy of the linear discriminant function depended upon the pipe diameter and in no case is expected to be accurate enough to be used along as a predictive tool.

Another type of probabilistic predictive models that appear to have very attractive features for applications in deteriorating pipes are the failure-time type models used for the analysis of survival data. Intrinsic to the notion of these models are the concepts of survivor and hazard rate functions. A formal definition of those terms is given in Chapter 3. The basic idea is to try to estimate a survivor function for each individual item, that will provide the probability for that item surviving beyond a future time period, given a set of explanatory variables. On the other hand an estimated hazard rate function will provide us with the instantaneous probability of failure for the particular item at any specified time.

Compared to the survival analysis models, discriminant analysis appears to be more limited. It usually involves concentration on a single point of the survival curve and by itself provides little insight into the way explanatory variables affect survival (Cox and Oakes, 1984). It could be more useful when the survival of each item can easily be classified as either very short or very long. It would also be a serious error to include the actual failure time as an explanatory variable, since it is part of the response and not part of the factors influencing the response (such discrepancy could for example be observed when logistic models are used for analyzing pipe break records).

The theoretical properties of the failure time models are examined in Chapter 3. It is one of the basic hypotheses of this work, that the analysis of pipe failures at the individual pipe level should, to a great extent, rely on the application of a particular type of failure model, which is Cox's proportional hazards model. The basic difference of this model from those previously described in this chapter, is that it can provide us with the probability of failure of a pipe segment at any future time period.

Two basic assumptions are associated with the structure of this model. The first implies a multiplicative effect of the break causing factors on the break rate. This assumption is physically justifiable, since it is reasonable to believe that the synergy of many factors acting together results in the final deterioration of the pipe wall and the occurrence of a break. The second assumption implies a log-linear relation between the break rate and the break-causing factors.

Of course, the fact that such a model structure is the appropriate one to describe failures in a deteriorating water distribution system cannot be proven apriori, unless the model is tested in a real world application. This task has been performed in this study for the New Haven, Connecticut and Cincinnati, Ohio water distribution systems, where the first real world application of Cox's Regression methodology on deteriorating water pipes has been completed. The results are described in detail in Chapter 5. In general, it was found that the model worked very satisfactorily and could provide reliable predictions of future pipe break failures and generate useful insights about failure patterns.

The following characteristics distinguish Cox's Regression methodology form the other techniques described previously in this chapter:

a. Cox's Proportional Hazards model provides us with the probability of break of a particular pipe rather than an expected number of breaks. For large diameter pipes ($\geq$ 8"), where breaks occur less frequently than in smaller diameter pipes, and very few or no breaks could happen during very long time periods, knowing the probability of a break can intuitively be interpreted much easier than a small fractional number of expected breaks that would have been obtained by the currently used regression type models. This probability will be changing in time as the pipe ages. It will also change if the pipe suddenly experiences a new break. In other words, the proportional hazards model is dynamic in nature, since the time which has elapsed after the last break and the previous history of breaks are incorporated in the analysis so that it becomes possible to update the failure probability of the pipe at any time. An additional benefit of a probabilistic model is its direct applicability in reliability analysis of the water system.

b. The proportional hazards model handles and uses effectively the information obtained from "censored" data. Censored data refer to pipes that have not experienced a break until the day that the study ended, or pipes for which breaks have not been recorded beyond a certain time period in the past. This is a very common phenomenon that appears in original data sets and as it has been pointed out previously it cannot be handled effectively by the methodologies proposed by Shamir and Howard (1979) and Clark et al. (1982). Further discussion of the censoring mechanism and its implications on parameter estimation is given in Chapter 3.

c. The proportional hazards model is very flexible in grouping (stratifying) the data according to various environmental and pipe characteristics and thus differences in failure trends among various pipe groups can be identified more easily.

d. The effect of the time elapsed since installation or since a previous break, on the future probability of failure does not need to be prespecified in any functional form (as for example in the Clark et al. (1982) model). This feature of the proportional hazards model is expected to provide great flexibility in investigating the problem of pipe aging on the failure rate.

A detailed analysis of the properties of the proportional hazards model is presented in Chapter 3, where the issues of data stratification, censoring and parameter estimation, are also extensively discussed. Overall, it appears that the advantages of the proposed methodologies for analysis of pipe break records over currently applied techniques will contribute to generating better insights about the break causing factors and lead to improved repair, replacement, and rehabilitation decisions.

CHAPTER 3

THE THEORY ON PROPORTIONAL HAZARDS MODELS AND THE JUSTIFICATION OF THEIR USE IN MODELLING FAILURES OF DETERIORATING PIPELINES

3.1 Introduction

The porportional hazards type models proposed by Cox (1972) belong in a broad family of models that have been widely used in the analysis of failure time data and are best described by the hazard function. Central to the structure of those models is the notion of failure time, defined as the length of time elapsed from a particular time origin up to a point event, often called failure. Assuming that observations on failure times of a group of individuals or items under consideration are available, those models attempt to assess the dependence of failure time on a set of explanatory variables that exist for each individual in the group. Numerous applications of these methodologies exist in medical studies and industrial life testing.

This Chapter examines the basic theoretical properties of the proportional hazards model and its major differences from other types of techniques available for the analysis of failure time problems. The appropriate estimation methods for this type of models are presented and the difficulties involved in the presence of censored observations are discussed. Finally a discussion is included about the justification for applying such methodologies to the analysis of deteriorating pipelines. The circumstances under which these methodologies are expected to provide better results than currently applied techniques are identified and arguments concerning their appropriate use during the phases that a deteriorating pipe goes through are presented.

3.2 Distribution of Failure Time and the Hazard Function

Assuming T to be a non-negative random variable that represents the failure time of an individual from a homogeneous population, then the survivor function $S_T(t)$ is defined as the probability that T is at least as great as a value t:

$$S_T(t) = pr(T \geq t) \quad , \quad T \varepsilon [0,\infty) \tag{3.1}$$

If T is continuous, which is mostly the case, then the probability density function $f_T(t)$ of the failure time is defined as follows:

$$f_T(t) = \lim_{\Delta t \to 0^+} \frac{pr(t \leq T < t + \Delta t)}{\Delta T} = - \frac{dS_T(t)}{dt} \tag{3.2}$$

and

$$S_T(t) = \int_t^\infty f_T(u)du \tag{3.3}$$

The hazard function $h(t)$ specifies the instantaneous rate of failure at time $T=t$ conditional upon survival to time t and is defined as follows:

$$h(t) = \lim_{\Delta t \to 0^+} \frac{pr(t \leq T < t+\Delta t / T \geq t)}{\Delta t} \tag{3.4}$$

$$= \frac{f_T(t)}{S_T(t)}$$

It can easily be seen that the functions $S_T(t)$, $f_T(t)$ and $h(t)$ provide mathematically equivalent ways of specifying the distribution of a continuous non-negative random variable. As Kalbfleisch and Prentice (1980) show, the hazard function can be written as:

$$h(t) = \frac{-d \log S_T(t)}{dt} \tag{3.5}$$

Given that $S_T(0) = 1$, we have:

$$S_T(t) = \exp\left(-\int_o^t h(u)\,du\right) \tag{3.6}$$

and the p.d.f. of T can be written as:

$$f_T(t) = h(t)\exp\left(-\int_o^t h(u)\,du\right) \tag{3.7}$$

Cox and Oakes (1984) provide several reasons why considering the hazard functions could be useful in many applications. The most important of them are:

a. Physical insights about the failure mechanism could be generated by considering the immediate "risk" of an individual known to have survived at age t.

b. Comparisons between groups of individuals are sometimes most incisively made through the use of the hazard function.

c. Hazard-based models are very often convenient to use when there is censoring of when several types of failure can exist.

d. Given the obtained functional form of the hazard rate, easy comparisons can be made about whether an exponential distribution would adequately describe the phenomenon.

3.3 Failure Time Models with Explanatory Variables - The Proportional Hazards Model

It is very often the case, as in the example of breaking pipes, that the effect on failure time of explanatory variables is of great interest. Such explanatory variables might include features thought to affect failure time, such as:

a. Previous break and maintenance history

b. Intrinsic properties of individual pipes, line material, internal pressure, size.

c. Exogenous variables describing environmental characteristics like soil properties and land activities in the vicinity of the pipes.

Explanatory variables can also be classified as constants with respect to time or time-dependent, and this distinction can have several implications in the estimation techniques discussed later in this chapter. Binary variables can also be used to distinguish between different standard sets of conditions.

Several types of regression models that try to determine the relationship between survival time t and a set of covariates $\underset{\sim}{z}$ have been proposed in the literature. Important factors in choosing which is appropriate to apply in a particular case, are the underlying assumptions about the nature of the failure mechanism and the degree of computational complexity involved in the estimation of parameters and the treatment of problems such as that of censoring.

The most commonly used regression models in practical applications are:

a. The Exponential and Weibull Models

For the exponential model, the hazard at time t for an individual with covariates z is given by:

$$h(t;z) = h.e^{zb} , \tag{3.8}$$

where h = constant

and $b = (b_1,...b_n)$, vector of regression parameters

The Weibull model assumes the hazard to take the following functional form:

$$h(t;z) = p(ht)^{p-1} e^{zb} \tag{3.9}$$

where p = constant

and b,z are defined as in equation (3.8)

Both models imply that the covariates act multiplicatively on the hazard function or in other words additively on the log of the survival time.

b. The Accelerated Life Model

The accelerated life model assumes that the effect of covariates is multiplicative on t rather than on the hazard function. A baseline hazard function $h_o(.)$ is assumed to exist and the effect of the regression variables is to change the rate at which an individual proceeds along the time axis, or equivalently the covariate vector z acts to accelerate or decelerate the time to failure. Thus, the accelerated life model takes the form:

$$h(t;z) = h_o(te^{zb})e^{zb} \tag{3.10}$$

where z and b are defined as in equations (3.8) and (3.9).

c. The Proportional Hazards Model

The proportional hazards model proposed by Cox (1972) takes the form:

$$h(t;z) = h_o(t)e^{zb} \tag{3.11}$$

where $h_o(t)$ is an arbitrary unspecified baseline hazard function and the covariates described by z act multiplicatively on the hazard function.

The exponential and Weibull models are special cases of the proportional hazards model, since $h_o(t)$ could be set equal to h or $hp(ht)^{p-1}$ respectively. Cox and Oakes (1984) have also shown that under constant explanatory variables, the accelerated life and proportional hazards models coincide only for the case where the hazard function follows the Weibull distribution.

The conditional on z density function of the survival time T is given by:

$$f_T(t;z) = h_o(t)e^{zb} \exp\left[-e^{zb}\int_o^t h_o(u)du\right] \tag{3.12}$$

and the corresponding survivor function becomes:

$$S_T(t;z) = \left[S_o(t)\right]^{\exp(zb)} \tag{3.13}$$

where $$S_o(t) = \exp\left[-\int_o^t h_o(u)du\right] \tag{3.14}$$

The fact that $h_o(t)$ is arbitrary provides the model with great flexibility for many applications, where it is not desirable or possible to assume a particular functional form describing the effect of time.

If some explanatory variables do not appear to have a multiplicative effect on the hazard function, the data can be divided into strata with separate baseline hazard functions for each stratum j and have only the remaining explanatory variables acting multiplicatively on the hazard rate. That is, the hazard function for the j^{th} stratum could be written:

$$h_j(t;z) = h_{oj}(t) \exp(zb) \tag{3.15}$$

Attention though should be paid on the simultaneous reduction of the sample size.

3.4 Physical Interpretation and Usefulness of the Proportional Hazards Model for Analyzing Break Records in Pipes

A model of the type described by equation (3.11) can be interpreted as follows: The function $h_o(t)$ represents an "ageing" process that goes on independently of stress. For the case of deteriorating pipelines the ageing process could represent a corrosion induced deterioration that makes several spots along the pipe length 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 "ageing" process and a stress-dependent term described by e^{bz}, we obtain the underlying structure of the proportional hazards model given by equation (3.11). The stress factors included in the exponential coefficient bz 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. On the other hand the model is non-cumulative in nature, since those stress factors are assumed to act independently from the on-going ageing process. The above observations tend to reinforce the argument that the structure of the proportional hazards model would be appropriate for describing the failure process of individual pipes at the phase where they have not yet reached the critical limit of deterioration, where they start completely falling apart. For this latter phase, other modelling techniques will be more appropriate since the main interest shifts from estimating the time to next failure of a rather structurally sound overall pipe, to calculating the number of expected failures that would occur in a very short time period once the pipe, or a particular section of it, has severely deteriorated.

It is important to emphasize that another attractive advantage of the proportional hazards model is the fact that the baseline hazard function does not need to be prespecified. This is expected to be of great help for the analysis of pipe breaks, since the way time affects the probability of failure remains controversial in the literature and thus deciding apriori on the shape of the time dependent function becomes very risky. As it has been argued by Kalbfleish and Prentice (1980), a remarkable feature of the proportional hazards model is that computationally efficient methods of inference of the regression coefficients b exist, without imposing any restriction on $h_o(t)$. Cox (1972) also suggests that it seems plausible that the loss of information about b arising from leaving $h_o(t)$ arbitrary is usually slight.

The last important feature of the proportional hazards model is the fact that censoring is relatively easily accommodated within its structure. The issue of censoring that may exist in many data sets and which is also present in the case of observations from pipe failures, is examined in detail in the next section.

3.5 The Problem of Censoring

It is often the case in the analysis of survival data that several individuals are not observed for the full time to failure. This, for instance, can happen in water pipes if upon termination of the study no failure has yet occurred. This type of truncated observation is called right censoring and, like failure, censoring is a point event, for which the period of observation must be recorded. A different type of censoring, also common in pipe break records, can occur when failure observations are missing beyond a time period in the past. This type of censoring is called left-censoring and could raise serious questions on the validity of statistical models derived under its presence. This section will primarily deal with the theoretical aspects concerning the issue of right-censoring. Appropriate ways for treating effectively left-censored data are described in detail in Chapter 6 , where an actual implementation of the proposed methodologies on pipe break records is carried out.

Censoring usually introduces great complications in the distribution theory for estimators used in survival models. Thus, at its presence, asymptotic methods for estimation and testing need to be developed. The various existing censoring mechanisms are presented in detail by Kalbfleish and Prentice (1980). For the case of pipe break records the censoring time for each pipe is fixed in advance (date of study termination). This type of censoring is considered a special case of random censorship, where the censoring time for each individual is a random variable.

3.6 Estimation Techniques

Among the proposed methods for estimating the parameters b, those of marginal and partial likelihood are the most commonly used in practice. Since the partial likelihood method developed by Cox (1972) is considered to be the most general of the existing estimation techniques and has also been used in the data analysis involved in the present study, further details on its derivation are given in this section.

If $\tau_1<\tau_2<\ldots<\tau_n$ denote the ordered failure times of n individuals and $\phi_i = i$, if and only if, individual i fails at $t_i = \tau_j$, then a risk set $R(t_j)$ is defined as $R(\tau_j) = \{i:t_i \geq \tau_j\}$.

The conditional probability that item i with covariate vector z_i fails at τ_j given that an item from the risk set $R(\tau_j)$ fails at τ_j is given by:

$$\frac{h_i(\tau_j)}{\Sigma_{k\varepsilon R(\tau_j)} h_k(t_j)} = \frac{\exp(\beta z_i)}{\Sigma_{k\varepsilon R(\tau_j)} \exp(\beta z_k)} \tag{3.16}$$

The baseline hazard function $h_o(\tau_j)$ cancels in (3.16) because of the multiplicative assumption of the proportional hazards model. Also, the conditional probability given by (3.16) is independent of $\tau_1, \tau_2, \ldots \tau_j$ (assuming no censoring occurs in (τ_{j-1}, τ_j), which is indeed the case for pipe break records). Thus, equation (3.16) equals the conditional distribution of ϕ_i given only

$\phi_1=i_1$, $\phi_2=i_2,\ldots\phi_{j-1}=i_{j-1}$, which is denoted by $p_j(i/i_1,i_2,\ldots,i_{j-1})$.
The joint distribution $p(i_1,i_2,\ldots,i_n)$ is obtained by applying the chain rule for conditional probabilities. Thus we have:

$$p(i_1,i_2,\ldots i_n) = \prod_{j=1}^{n} p_j(i_j/i_1,i_2,\ldots,i_{j-1})$$

$$= \prod_{i=j}^{n} \frac{\exp(bz_i)}{\sum_{k\varepsilon R(\tau_j)} \exp(bz_k)} \tag{3.17}$$

Given the fact that $h_o(t)$ is completely unspecified, no information about b is obtained from the observation that no failure occurs in the interval τ_{j-1},τ_j $(j=1,\ldots k)$.

The partial likelihood for b, L(b) is simply given by equation (3.17). Cox (1972) has shown that the above method for constructing the partial likelihood, although it does not provide a likelihood function in the usual sense (i.e., a function proportional to the conditional probability of an observed event), it gives maximum partial likelihood estimates that are consistent and asymptotically normally distributed (also see Efron (1977), Oakes (1977), Bailey (1979) and Tsiatis (1981)).

Several methods have also been proposed for dealing with the problem of ties in survival times (Cox and Oakes (1984), Kalbfleish and Prentice (1980)). The approximation used for the analysis of the pipe break records, in the case of ties, was that proposed by Breslow (1974), where the partial likelihood function is written as:

$$L(B) = \prod_{i=1}^{K} \left\{ \exp(bs_i) / \left[\sum_{j \varepsilon R(\tau_j)} \exp(bz_j) \right]^{m_i} \right\} \tag{3.18}$$

where m_1 is the number of failures at t_i and s_i is the vector sum of the covariates of the m_i items.

A variety of estimators for the baseline hazard function have been proposed. According to Kalbfleish and Prentice (1982), these estimators are in reasonable agreement and usually it does not matter much which is used. The one provided by Cox and Oakes (1984) is described as follows:

If $H_o(t)$ denotes the cumulative hazard obtained by:

$$H_o(t) = \int_o^t h_o(u)du \tag{3.19}$$

then a non-parametric estimator of $H_o(t)$ can be constructed as follows:

Let D(t) denote the total number of failures in (0,t), that is:

$$D(t) = \sum_{\tau_j < t} d_j \tag{3.20}$$

Then D(t) has the same expectation as the total cumulative hazard up to time t for all items:

$$H(t) = \sum_{i=1}^{n} H_i(y_i) \tag{3.21}$$

where $y_i = t$ if $t_i > t$ or otherwise y_i equals the failure or censoring time.

Thus, for the log-linear proportional hazard model we would have:

$$H(t) = \int_o^t \sum_{j \varepsilon R(u)} \exp(bz) h_o(u)\, du \tag{3.22}$$

and the estimator becomes:

$$\hat{H}_o(t) = \sum_{\tau_j < t} \frac{d_j}{\sum_{k \varepsilon R(T_j)} \exp(\hat{b}z)} \tag{3.23}$$

where the coefficient vector b is set equal to $\hat{b}$, obtained from the partial likelihood estimator.

The baseline survivor function $S_o(t)$ is then estimated by:

$$\hat{S}_o(t) = \exp\left[-\hat{H}_o(t)\right] \tag{3.24}$$

and the survivor function $S_T(t)$ is estimated by:

$$\hat{S}_{T_i}(t) = \left[\hat{S}_o(t)\right]^{\exp(\hat{b}z_i)}, \text{ for item } i \tag{3.25}$$

The above estimators could be used in assessing goodness of fit of the derived model. The residual for the i-th item is defined by:

$$R_i = - \ell_n \hat{S}_{T_i}(t) = H_i(t) \tag{3.26}$$

If the model fits the data, the residuals R_i should behave as a random sample of censored unit exponential variates. Thus, as an overall check of the model, the estimated values of $\hat{H}_i(t)$ can be plotted against their expected values for the unit exponential distribution. Nevertheless, since considerable investigation is still needed in order to examine the properties of those residuals, the above described goodness-of-fit test should be used with caution (Kalbfleish and Prentice (1980)). In other words, it has not yet been established in the statistical literature what kinds of deviations one would expect to observe in the residuals if the model does not fit well with the data or to what extent agreement with the anticipated line should be expected.

The conditional log likelihood obtained from equation (3.17) is given by:

$$L(b) = \sum_{i=1}^{k} z_i b - \sum_{i=1}^{k} \log\left[\sum_{\ell \varepsilon R(\tau_i)} \exp(z_\ell b)\right] = \sum_{i=1}^{k} \ell_i \tag{3.27}$$

where ℓ_i is the contribution to the likelihood from the i^{th} failure

In order to obtain maximum likelihood estimates of β the first and second derivatives of ℓ_i are first calculated. If z_{ir} denotes the value of the r^{th} component of the explanatory variable z on the i^{th} subject, then we have:

$$\frac{\partial \ell_i}{\partial \ell_r} = z_{ir} - \frac{\sum_{k \epsilon R(\tau_i)} z_{kr} e^{\beta^T z_k}}{\sum_{k \epsilon R(\tau_i)} e^{\beta^T z_k}} = z_{ir} - A_{ir}(\beta) \qquad (3.28)$$

Also:

$$\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s} = - \frac{\sum_{k \epsilon R(\tau_i)} z_{kr} e^{\beta^T z_k}}{\sum_{k \epsilon R(\tau_i)} e^{\beta^T z_k}} + A_{ir}(\beta) A_{is}(\beta) = -C_{irs}(\beta)$$

(3.29)

By summing over all risk sets $i=1,\ldots k$, we obtain from equation (3.28) the score function $U_r(\beta)$ for the r^{th} component:

$$U_r(\beta) = \sum_{i=1}^{k} \frac{\partial \ell_i}{\partial \beta_r} = \sum_{i=1}^{k} (z_{ir} - A_{ir}(\beta)) \qquad (3.30)$$

The information matrix $I(\beta)$ of negative second derivatives has elements:

$$I_{rs}(\beta) = \sum_{i=1}^{k} C_{irs}(\beta) \qquad (3.31)$$

Maximum likelihood estimates of b can be obtained by iterative use of equations (3.30) and (3.31). The Newton-Raphson iterative algorithm is applied in order to obtain those estimates from the analysis of the pipe failure data. Appendix C describes in detail the particular technique and applied covergence criteria.

If the censoring mechanism is independent of the explanatory variables, which is indeed the case with the pipe break records, then an exact test of the null hypothesis b=0 can be performed (Cox and Oakes (1984)). The appropriate test statistic is given by:

$$\{U(0)\}^T\{I(0)\}^{-1}\{U(0)\} \qquad (3.32)$$

The statistic described by equation (3.32) must, under the the hypothesis, have an asymptotic X^2 distribution with degrees of freedom equal to the number of coefficients set to zero.

3.7 Stratification of the Data Set and Testing of the Proportionaily Assumption

In Section 3.3 the notion of stratification of the data set has been introduced, in order to accommodate the cases where the ratio of hazard rates for different levels of an independent variable is not constant. This case may very well arise in the analysis of pipe failure data for variables reflecting for example different installation periods or pipe and soil characteristics.

That is, it is not known apriori whether the proportionality assumption of Cox's model would hold for various subgroups of the data. Thus, when application of Cox's model on this type of data is attempted, it would be desirable to first classify the data in groups that might experience different failure patterns among them. Changes in construction materials and practices along with different types of stresses imposed on various groups of pipes, will become suggestive of the appropriate ways for preliminary stratification of the data.

An easy way to test if the proportionality assumption holds for the various strata, is to plot the log of the minus log survival function $\ell n\left[-\ell n\hat{S}_T(t,\bar{z})\right]$ for each stratum, by setting the covariate vector z equal to the mean of the covariates of the corresponding stratum $\bar{z}$. If the proportionality assumption holds, the difference between the plotted curves for each stratum must be constant (Kalbfleish and Prentice, 1980). Another method for testing the proportionality assumption considers the use of time dependent covariates in the regression model, where a prespecified function of time is associated with the covariates suspected of having a nonproportional effect on the hazard rate function. In such a case, if the proportionality assumption holds, then the estimated coefficient for the time dependent covariate must be close to zero.

CHAPTER 4

GENERAL PROBLEMS ASSOCIATED WITH DATA SETS OF DETERIORATING WATER DISTRIBUTION SYSTEMS

4.1 Introduction

One of the most important problems to be considered when dealing with the derivation of a predictive model for pipe break failures is the availability of adequate and reliable data.

There is a great variability among the records kept in various cities in the U.S. about repair events for water mains, environmental characteristics, operating and maintenance practices, pipe materials, etc.. An underlying factor in most cases is that records were not collected and organized for the reasons they are expected to be used today. Although problems with the data sets will always exist, there will be cases where the information believed to be obtained by analyzing a data set, despite all the problems, could be useful in generating insights and helpful in making repair and replacement decisions. Also, there will be cases where the data could be considered very poor and unreliable for use in a more sophisticated analysis.

It must also be kept in mind that analyzing an "imperfect" data set could also generate insights about the proper collection of future data, as it will be pointed out in several occasions throughout this study.

Another important question raised by the inadequacy of data in many cities is the degree to which a derived predictive model for a certain city could be satisfactorily applied at another system with similar characteristics. Of course, there can never be a proof that a derived model at a particular city with adequate data is the right one to use for another case where the data are very poor. But, by comparing the derived models in different cities and also by trying to establish good physical reasons for the effect of the important covariates of those models on pipe failure, it is possible to start understanding whether such attempts would be justifiable or not in the future. This issue is discussed further when comparisons between the models obtained from the New Haven and Cincinnati systems are made.

4.2 Specific Problems Associated with Data Sets

Breaks Versus Leaks

It is usually hard to define a "threshold" between what kind of failure we characterize as a break and what kind we consider to be a leak. Clearly, that "threshold" should be related to the amount of discharge from the damaged main. Leaks from pipes are, as expected, much more difficult to detect than breaks which usually have an immediate impact at the system. Since the recorded break events very often represent repair events, a distinction should be made as to whether the repair was for a break or for a leak. This information though could be missing in many cases.

Break Type

Information on break type is important, since different break types will, in general, relate to different break causing factors. There are four basic break types in water mains: circumferential breaks (ring cracks), longitudinal breaks (split pipes), holes in pipes and joint breaks. It is indeed very often the case that information on break type is not provided in existing data sets. Break type had not been recorded for both of the analyzed data sets in this study. Only for a very limited part of the New Haven data such information has been available and an analysis has been performed (presented in Chapter 8), concerning the relation between break type, seasonality, patterns in breaks, and other pipe characteristics.

Break Location

Exact break location along a specified pipe length can be very useful, when available, since it makes it possible to investigate whether subsequent breaks occur in the same location or not and also it might provide information on particularly unfavorable local conditions. Very often though, information on break location is missing from the data sets, which makes the distinction between localized and more extended pipe deterioration extremely difficult.

Length of Pipe

The issue of defining an appropriate pipe length in the data set is directly related to the question of how a "pipe" must be defined. For the purposes of deriving predictive statistical models, a pipe should correspond to an entity with uniform geometric and operating characteristics (diameter, pressure) along its length and on which uniform environmental conditions apply (traffic loads, soil type, residential and industrial development, etc.). In general, by avoiding having high variability of pipe lengths and by choosing relatively short lengths to define a pipe, those goals about uniformity of conditions can be achieved. However, seldom are such considerations taken into account, when pipes were coded in the data files.

Previous Maintenance History

Previous maintenance practices (including anti-corrosion measures) are important, as long as they may have affected the structural behavior of several sections in the system. The utility's policy associated with leak detection can also be relevant since a close relationship between leaks and breaks is expected to exist. Often pipes that have been replaced due to bad performance still exist in the data sets and this could distort the results of a statistical analysis. Information on maintenance practices is rarely recorded on data sets, although it may significantly help in explaining many of the observed failure patterns.

Pipe Materials and Construction Practices

Pipe materials and construction practices (bedding standards, etc.) have significantly changed along the years, nevertheless such information is very often missing from the data. It is thus important that the data reflect this kind of information, since pipes are expected to show differences on break patterns, depending on those factors.

Cast-iron manufacturing has gone through several stages of evolution during the years. Starting from horizontal casting, for the pre-1850 pipes, it changed into "pit casting" after 1850, which resulted in more uniform wall thickness and inclusions of less foreign materials in the pipe wall. In the early 1930's, centrifugal casting became commercially common, which resulted in even more uniform wall thickness and complete uniformity of material. In the 1960's ductile iron has been introduced, which has better structural properties but its corrosion resistance is not expected to be any better than grey iron used in the older mains. (Corrosion Handbook, 1983) Also, the increase in iron strength that occurred with the years has been partially counterbalanced by reduced wall thickness in the most recent mains (Table 4.1, O'Day (1983)).

Table 4.1

STANDARD MAIN WALL THICKNESSES - INCHES
(1908-1980)

Year	Material	Nominal Diameter 6"	8"	12"	24"
1908	Pit Cast (11/31)	.48		.62	.89
1939	Centrifugal (18/40)	.38		.62	.89
1957	Centrifugal (21/45)	.35		.44	.68
1965	Ductile (42/72)	.25	.43	.49	.56

(11/31) bursting hoop strength/modulus
of rupture - 1,000's psi

Along with pipe material, joint material has been changing over the years. Properly designed joints are expected to provide the pipe with the ability to expand and contract under changing temperatures, so that expansion and contraction stresses along the pipe length are avoided. Flexible lead joints have been used for many years in most systems. Rubber gaskets have started replacing lead in the 1960's. Nevertheless, there are occasions, primarily because of cost and lead shortage considerations, that leadite (a lead substitute) was used in the late 1940's and early 1950's by several water utilities. Leadite, as opposed to lead, makes a rigid rather than a flexibile joint, with important negative implications on stresses caused along the pipe (O'Day, 1983).

As construction technologies and guidelines have also been modifying along the years, they must be related to a certain degree to observed failure patterns of different pipe categories.

Intertemporal Variation of Existing Conditions

Variables describing types of land development and pipe pressure, which often appear in the data sets, are expected to have changed with time. Usually, the data do not reflect such intertemporal changes and provide us with one unique value for those variables commonly obtained from a census tract or information available during the pipe installation time.

Related to this issue is the fact that the pipe's internal diameter could also change with time, depending on the material of the pipe wall, the flow velocity and the quality of water (pH etc.), through the formation of tuberculus in the interior of the pipe. This could result in changes of the operating characteristic of the system (pressure, carrying capacity). It is not clear how this process could affect the breaking mechanism in pipes, but it is believed that information on the pipe's condition, which is very often missing in the data, would be very useful for a statistical analysis.

Left Censoring of the Records

There are cases (e.g., Cincinnati data), where records on pipe breaks have been kept after a certain time period from installation. Thus, in analyzing such data, special problems occur for dealing with pipes which might have broken in the past, but for which no such records are available. Appropriate ways for dealing with the problem of left censoring in the context of the proportional hazards model proposed in this study are examined in Chapter 6.

Additional Problems

It is very often the case that information on depth of cover for pipes is missing and also information on the exact timing of breaks (i.e., day and month as opposed to only year). That information could have been useful for judging the effect of frost penetration on pipe breaks and the exact stresses caused by external loads.

Information on flow velocities would also have been useful, since flow velocity is directly related to dynamic pressures developed in a conduit from water hammer and conduit bends. But again, this information is commonly non-available.

4.3 Conclusions

Given that almost all of the problems mentioned in this chapter will be present in currently available data sets of various water distribution systems in the U.S., it must a priori be expected that they will tend to obscure the results of any statistical analysis performed. Particularly when an attempt is made to derive predictive models for future pipe failures, the explanatory power of certain variables used in regression analyses, should not be overestimated. For example, while a variable reflecting the type of land development covering a pipe might be used as a surrogate of external loads transmitted to the pipe, the exact relationship between such variables and the stresses caused at a pipe underneath, will not be clearly understood. Related to this is also the effect that a variable describing the degree of soil corrosivity might be interpreted of having at the break rate. Clearly such variable is associated only to one type of corrosion that could occur at a pipe. The presence of other types of corrosion (see Chapter 7) and primarily internal corrosion is not reflected by such variable and thus it should not be expected to fully explain the relationship between corrosion and pipe failures.

Under the above observations, the scope of a model building attempt becomes twofold: a) capture to the maximum possible degree the existing failure patterns in a system and generate insights about their causes and the factors associated with breaks, under the "imperfect" information available from existing data and, b) propose the necessary improvements in data collection practices, so that future models become more reliable and a deeper understanding of the failure mechanism is developed.

CHAPTER 5

ANALYSIS OF THE NEW HAVEN SYSTEM

5.1 General System Characteristics

The data set obtained from the New Haven water distribution system consisted of 1391 pipes installed during the 1900 - 1980 period with a total length of approximately 501 miles. Information had been coded for each individual pipe segment regarding its length, number and year of break events, pressure, diameter, degree of soil corrosivity, percentage of high, low, and medium land development covering the pipe, year of installation, type of pipe (metal or concrete), soil stability and presence of swamp surrounding the pipe.

Pipe lengths varied from 100 ft. to 14,000 ft. The criterion for defining length has been solely based on the fact of no intermediate connections and constant diameter size between nodes. As argued in Chapter 3, such configuration of pipe length is not the best for performing statistical analysis.

Pipe diameters ranged from 6 inches to 48 inches with most of the pipes having diameter greater than 8 inches (only 1.2% of the pipes were 6 inch). That is, the data set consisted of pipes that are, in general, considered large and believed to have quite different failure patterns than smaller size pipes (4 and 6 inches).

About 46% of the pipes were installed in the 1930-35 period. The rest of the pipes were installed in smaller clusters at several times during the 1980 - 1984 period. Approximately 96% of the pipes in the data set were made from cast iron and there existed only 20 concrete pipes.

Internal pressure showed a great variability ranging from 4 to 173 psi, but no information has been provided reflecting the exact period that these measurements were taken. As pointed out in Chapter 3, intertemporal variations in the internal pressure of various mains is very likely to have occurred, and that is expected to limit up to a certain degree the explanatory power of such variables.

Approximately 69% of the pipes were installed in non-corrosive soil while the rest were laid in corrosive soil. Soil corrosivity was defined as a 0,1 dummy variable and no further information was provided as to what exactly it represented, but most likely it was related to soil type. Since the issue of highly corrosive conditions is believed to be very important as far as pipe breaks are concerned, a detailed discussion regarding the explanatory power that such variables can have is provided in Chapter 7, where the physical interpretation of the results from this analysis is presented.

About 53% of the pipes were installed in unstable soil and 23% in moderately stable soil. Only 1% of the pipes was completely covered by swamp, while 95% had no contact at all with swamp areas.

About 511 pipes were covered 100% by minimal land development while 115 were covered 100% by maximum land development, with the rest of the pipes lying somewhere in between. No exact information was provided on what types of urban activities the land development values referred to. It is believed that they represent an average of industrial, transportation, residential, and commercial land uses, obtained from a 1972 census tract.

Only 292 pipes (21% of the total) had experienced one or more breaks. Among those, 202 had only one break since their installation and only about 3.7% of all pipes had 2 or more break events. Break data were sparce beyond the third break (only 24 pipes were broken more than three times). Since the data consisted primarily of larger diameter pipes, multiple breaks on individual mains are generally expected to be quite few as compared to smaller size pipes. The observed trend of total number of breaks per year in the New Haven system is shown in Figure 5.1. A general increase in the number of breaks is observed after the 1950 period.

Judging from the total number of breaks occurred in the large diameter pipes that the data set consisted of, it appears that the New Haven system is in good overall condition relatively to other systems in the U.S., where higher break rates are observed.

The average time to the first break was about 23 years, while then it dropped dramatically as we moved to the second and third

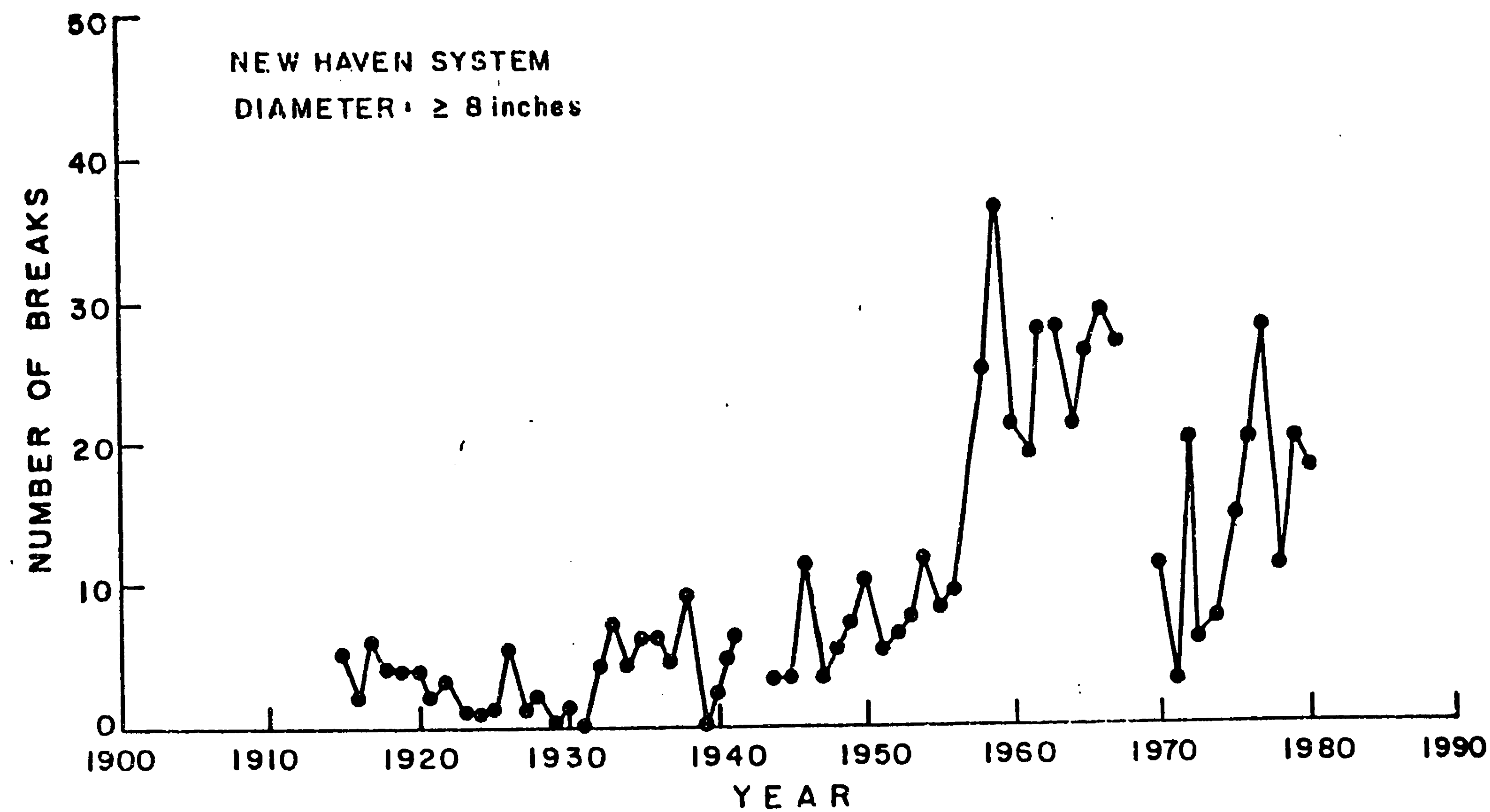


Figure 5.1

CUMULATIVE NUMBER OF BREAKS BY YEAR
(NEW HAVEN SYSTEM, ONLY FOR DIAMETER $\geq$ 8 INCHES)

break. Beyond the third break, data were scarce to obtain any statistically meaningful results but in general the time to next break became very short (up to 2 years on the average) with no further decreasing trend.

Descriptive statistics regarding all the previously discussed variables are shown in Appendix A.

5.2 Preliminary Bivariate Analysis

The reason for performing bivariate correlation analyses between pairs of variables in the data set was to identify any possible trends in break patterns and relations among variables that could help in the model building process that would follow this preliminary analysis.

Three types of bivariate correlation analyses were performed considering the whole set of pipes:

a. Parametric correlation analysis

   The calculated correlation coefficients between the number of breaks and time to first break with the rest of the variables believed to affect the break rate, are shown in Table A1 of Appendix A. The values of the correlation coefficients were very low for all the variables that could have been used as potential predictors of failure (calculated values did not exceed the range of 0.20 for such variables).

b. <u>Non-parametric correlation analysis</u>

The fact that many of the variables had values clustered together in various ranges and also showed very skewed distributions indicated that non-parametric correlation analyses might be appropriate for this type of data. Thus, the Spearman, Kendall, Tau and Pearsons correlation coefficients were also calculated for the various pairs of variables. Again, though, no statistically significant correlations were found. Representative results of the obtained correlation coefficients are shown in Tables A2 and A3 of Appendix A.

Another attempt was also made in which pipes were first categorized in clusters according to various installation periods and correlation coefficients were then calculated separately for each category. Again, in the majority of cases, no significant correlation among the potential failure predictors and the number of breaks, or the time to first break, can be found.

The general point that can be made, after this analysis has been performed for the New Haven data set, is that bivariate correlation analysis can very easily fail to give us satisfactory results and useful insights, when the process under study and the interactions among variables are fairly complex. Similar problems with those that appeared in this analysis were observed, as mentioned in Chapter 2, by O'Day (1983), when statistical analysis of break records was performed for the City of Philadelphia.

5.3 Application of the Proportional Hazards Model

The performance of the proportional hazards model was tested for the New Haven system. In order to accommodate the fact that a number of pipes had experienced several breaks, the following transformation was applied on the original data set; every interval time between two subsequent breaks was considered as a separate observation. A dummy variable indicating the number of previous breaks was used as a covariate together with the rest of independent variables describing pipe and environmental characteristics, in cases where Cox's model was applied on the entire data set. Several experiments were also performed by applying the model on separate groups of pipes independently, where groups were classified by the number of previous breaks or other distinguishing characteristics.

By observing the failure times of individual pipes (also see Figure A1, Appendix A), it becomes clear that beyond the third breaks no trend exists for the subsequent failure times. Also observations become very scarce for any meaningful statistical analysis, while most of the pipes in that stage experience multiple breaks within very short periods of time. For these reasons, it did not appear appropriate to extend the application of the model beyond the third break. If a larger number of observations existed however beyond that point, it would have made sense to apply other statistical techniques for estimating the break-rate once pipes were in that stage. That is, since many more frequent breaks are usually expected to occur after the third break, it would become important to estimate a break-rate rather than the probability of the next failure to occur.

Many experiments were performed by applying the model under different specifications for the independent variables used and possible interactions among them. A process of thorough stratification of the data in subgroups that could possibly demonstrate different failure patterns was followed. Such subgroups were defined in various ways reflecting differences in installation periods, pipe sizes, soil types, etc. The findings from intermediate experimental regression runs often became suggestive of further modification that would be appropriate in the model's structure. Some representative intermediate phases of the work that led to the final model configuration are presented in Section 5.4.

Stepwise regression has been applied in many of the preliminary runs, since no information was available about the importance of many of the independent variables used in the analysis. The maximum partial likelihood ratio test (MPLR) was applied in order to estimate the significance probabilities of the variables that were candidates to enter or be removed from the regression equation. This test computes the significance probabilities on the basis of a large sample partial likelihood ratio test using the chi-square value calculated from the log of the ratio of the maximized partial likelihood functions:

$$MPLE = 2/\log(L(b_{current})/L(b_{candidate})) \tag{5.1}$$

where b is the vector of coefficients.

By applying stepwise regression several of the variables included initially in the set of independent covariates were dropped out during further steps of analysis, because they were found to be statistically insignificant. For other variables appropriate transformations were applied. Such transformations were of two types; the first included functional transformations on a particular variable and for all observations in the data set. For example, the use of the logarithm of length, instead of length or the logarithm of age. The motivation for such transformations has always been the possibility that these variables could act on the hazard rate in the way specified through the transformation. The second type of transformations used in the analysis was motivated by the fact that certain covariates were found to affect the hazard rate only for certain subsets of the total observations. In such cases, they were transformed so that this effect would be captured. Stratification of the data in various ways that were believed to reveal differences in failure patterns has greatly contributed in identifying such transformations and also specifying necessary dummy variables for the analysis. Several stratification experiments that played an important role in leading to the final model configuration are presented in the following section.

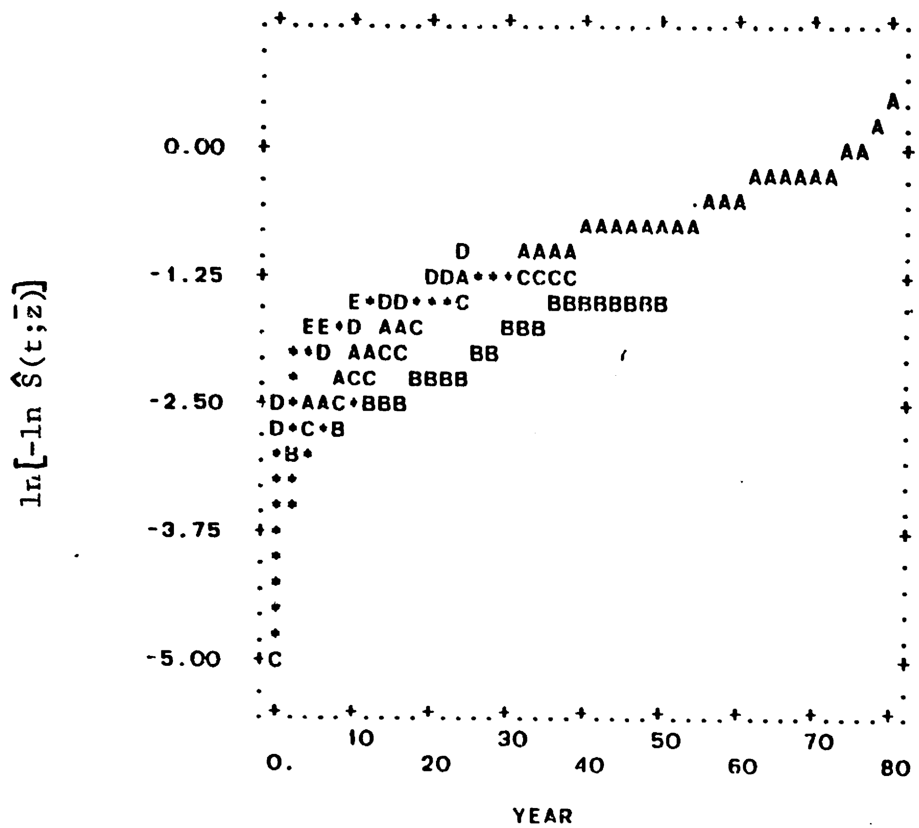
5.4 Stratification of the Data

The reason for stratifying the data is twofold: a) identify subsets of the whole data population that experience different failure patterns and, b) test the proportionality assumption in the proportional hazards model. This test is performed by plotting the log minus log survival function for each stratum, that is, the function $\ell_n[-\ell_n \hat{S}_T(t;\bar{z})]$, where $\bar{z}$ is the mean of the covariates. In cases where those curves are considered to be parallel, the same baseline hazard function could be used for each stratum and dummy variables could be introduced for representing the different strata. On the other hand, whenever this is not true, different baseline hazard functions should be estimated for each stratum. It is also important to notice that the plotted log minus log survival curves are best estimates of the "true" functions. Thus, they are subject to a random error, which should be taken into account, when judgments are made about the relative differences of various strata.

Three representative stratification experiments are described in this section.

a. Stratification by construction periods

Figure 5.2 shows the five different log minus log survival curves obtained by stratifying the observations according to construction periods. The symbols A, B, C, D and E refer to pipes installed during the periods 1900 - 1930, 1930 - 1935, 1935 - 1950, 1950 - 1965, and 1965 - present respectively. The results clearly



Period

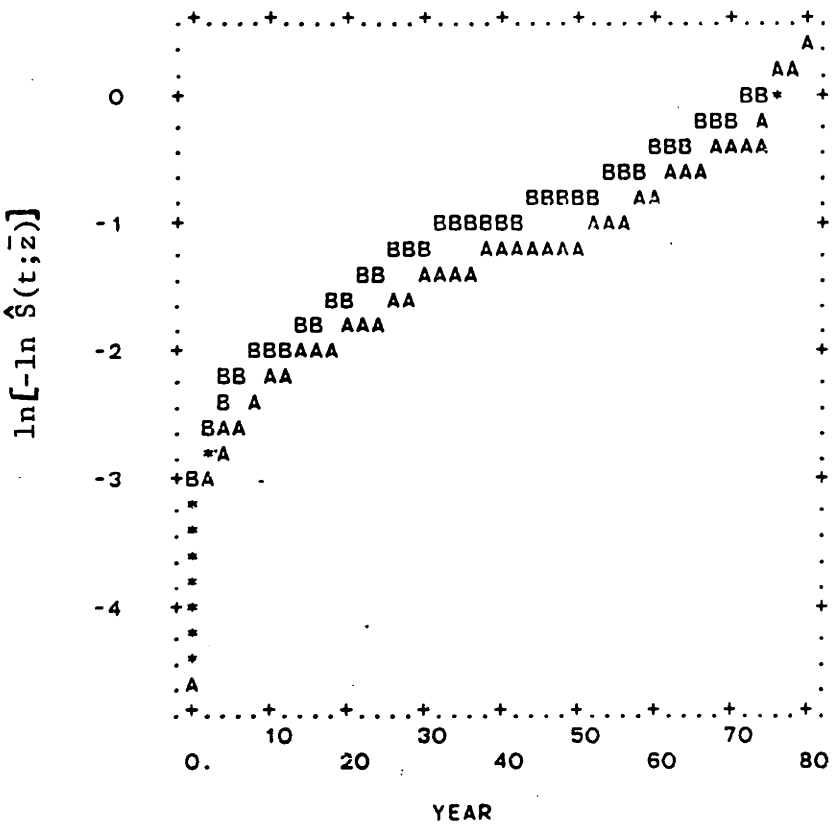
A	1900 - 1930
B	1930 - 1935
C	1935 - 1950
D	1950 - 1960
E	1960 - present

Figure 5.2

LOG OF MINUS LOG SURVIVAL FUNCTIONS - STRATIFICATION BY INSTALLATION PERIODS, NEW HAVEN indicate that pipes installed in the period 1930-35 have less chances for failure, while pipes installed after 1950 have the higher chances for breaking. The estimated survival curves become suggestive that pipes installed before 1930 and pipes installed during the period 1935-50 could be merged into one group. Also, pipes installed during 1950-65 and during 1965-present could be considered as one group. The basic reason for merging several groups together is that the estimated survival curves for each group do not show any statistically significant difference. According to the functions shown in Figure 5.2, it appears that the proportionality assumption is indeed a good approximation for the chosen strata. Thus, the same baseline hazard function is used and dummy variables reflecting the three subgroups, previously defined, are introduced. That is, the three broad pipe classifications to be used in further analysis are: Pipes installed during the 1930-35 period, pipes installed after 1950 and pipes that do not belong in the above two categories. As is has been shown here, these three pipe categories show distinctively different survival patterns.

b. Stratification according to soil corrosivity

Since it is intuitive to believe that soil corrosivity contributes to the acceleration of breaks, a comparison between the survival curves of pipes under corrosive and less-corrosive soil was performed. The plotted log minus log survival curves are shown in Figure 5.3, where the symbol A indicates pipes in less corrosive soil and the symbol B indicates pipes in corrosive soil.



A: Non-Corrosive Soil

B: Corrosive Soil

Figure 5.3

LOG OF MINUS LOG SURVIVAL FUNCTIONS - STRATIFICATION BY SOIL CORROSIVITY, NEW HAVEN

The difference between the two curves is two small to be statistically significant. Also soil corrosiveness did not show up as important covariate, when used in the regression equations. Thus, there is no evidence to support that higher soil corrosivity was associated with also higher break rates in New Haven. Since, however, the measure of soil corrosivity available in the data is quite ambiguous and several other factors related to pipe corrosion are not captured by that variable, such findings should not be considered surprising. A detailed discussion on the issue of corrosion and its effect on pipe failure is presented in Chapter 7, when the overall results of the statistical analysis are interpreted.

c. Stratification by number of previous breaks

After the number of covariates to be used in the regression equations have been reduced through the help of stepwise regressions and the previously described stratification experiments, three separate regressions were performed, where the observations were classified according to the number of previous breaks. This type of analysis assisted in identifying covariates that had an impact on pipe failures only for a certain category of observations. The results are shown in Table 5.1. The variables used in those final regressions are defined as follows:

	(1391 OBS) PREVBRK=0	(292 OBS) PREVBRK=1	(90 OBS) PREVBRK=2
GLOBAL X^2	134.4(5D.F.) p=0.0	14.3(6D.F.) p=0.0263	16.3(6D.F.) p=0.0124
	COEFF(St.Error)	COEFF(St.Error)	COEFF(St.Error)
LNLENGTH	0.581 (0.079)	0.345 (0.141)	0.289 (0.273)
PRESSURE	0.089 (0.028)	-0.048 (0.054)	-0.217 (0.098)
LOW	-0.58 (0.14)	-0.46 (0.29)	-0.42 (0.45)
C35	-0.688 (0.146)	-0.563 (0.253)	-0.802 (0.491)
C50	0.491 (0.173)	-0.144 (0.342)	-0.316 (0.517)
AGE	--	-0.006 (0.008)	-0.028 (0.013)

Table 5.1

ESTIMATED REGRESSION COEFFICIENTS AND STANDARD ERRORS - DATA GROUPED BY NUMBER OF PREVIOUS BREAKS, NEW HAVEN

LNLENGTH = The natural logarithm of length, in ft.

PRESSURE = (Pipe pressure, in psi/10)

LOW = (Percentage of low land development covering the pipe/100)

C35 = Dummy variable = 1, if pipe installed during the period 1930-35; 0, otherwise

C50 = Dummy variable = 1, if pipe installed after 1950; 0, otherwise

AGE = Age of the pipe, in years at the time of each break

We observe that there are differences in the coefficients for different numbers of previous breaks, especially for those of PRESSURE and AGE. Thus, if PREVBRK = 0, then the coefficient of PRESSURE is positive, while for PREVBRK = 1 or 2 it is negative, but also less statistically significant. For the coefficient of AGE we observe that when PREVBRK = 0, or 1, it is not statistically significant and has a negative sign. The impact on the hazard rate of ln(length), low land development and C35 is definitely important in all three strata. The increase in the standard error observed in the estimated coefficients of those three variables as the number of previous breaks changes from 0 to 1 and then to 2, is statistically justifiable if the reduction in the number of observations is considered. Pipes installed after 1950 (variable C50) have higher chances for a first break, while after they have already broken, no difference in the hazard rate exists between them and the rest of the pipes that do not belong in the 1930-35 category. The descriptive statistics of the variables included in those final regressions are shown in Table 5.2, where no significant differences are observed among the variables characterizing each group.

Figure 5.4 shows the log minus log survival functions corresponding to each stratum (Previous breaks equals to 0, 1, or 2). It is clear that pipes with zero previous breaks have the higher survival probabilities. It also appears that the two groups of pipes with 1 or 2 previous breaks could be merged into one group, since the differences in their survival curves are not statistically significant. Figure 5.4 clearly indicates that the estimated $\ell_n[-\ell_n\hat{S}_T(t;\bar{z})]$ functions are practically parallel, thus, the same baseline hazard function could be used for all strata, and a dummy variable would indicate whether there were any previous breaks or not.

5.5 Final Model Configuration

According to the results described in Section 5.4, the following variables were chosen for use in the final model:

1. LNLENGTH = ln[LENGTH], where length is the length of pipe in ft.
2. PRESBRK = (PRESSURE/10), if the number of previous breaks is zero

   0, otherwise
3. LOW = percentage of low land development covering the pipe, measured from 0 to 1.

PREVBRK = 0 (N=1391)

	MIN	MAX	MEAN	ST.DEV.
LNLENGTH	4.61	9.55	7.23	0.81
PRESSURE	0.4	17.3	7.93	2.02
LOW	0	1	0.40	0.48
C35	0	1	0.46	0.50
C50	0	1	0.30	0.46

PREVBRK = 1 (N=292)

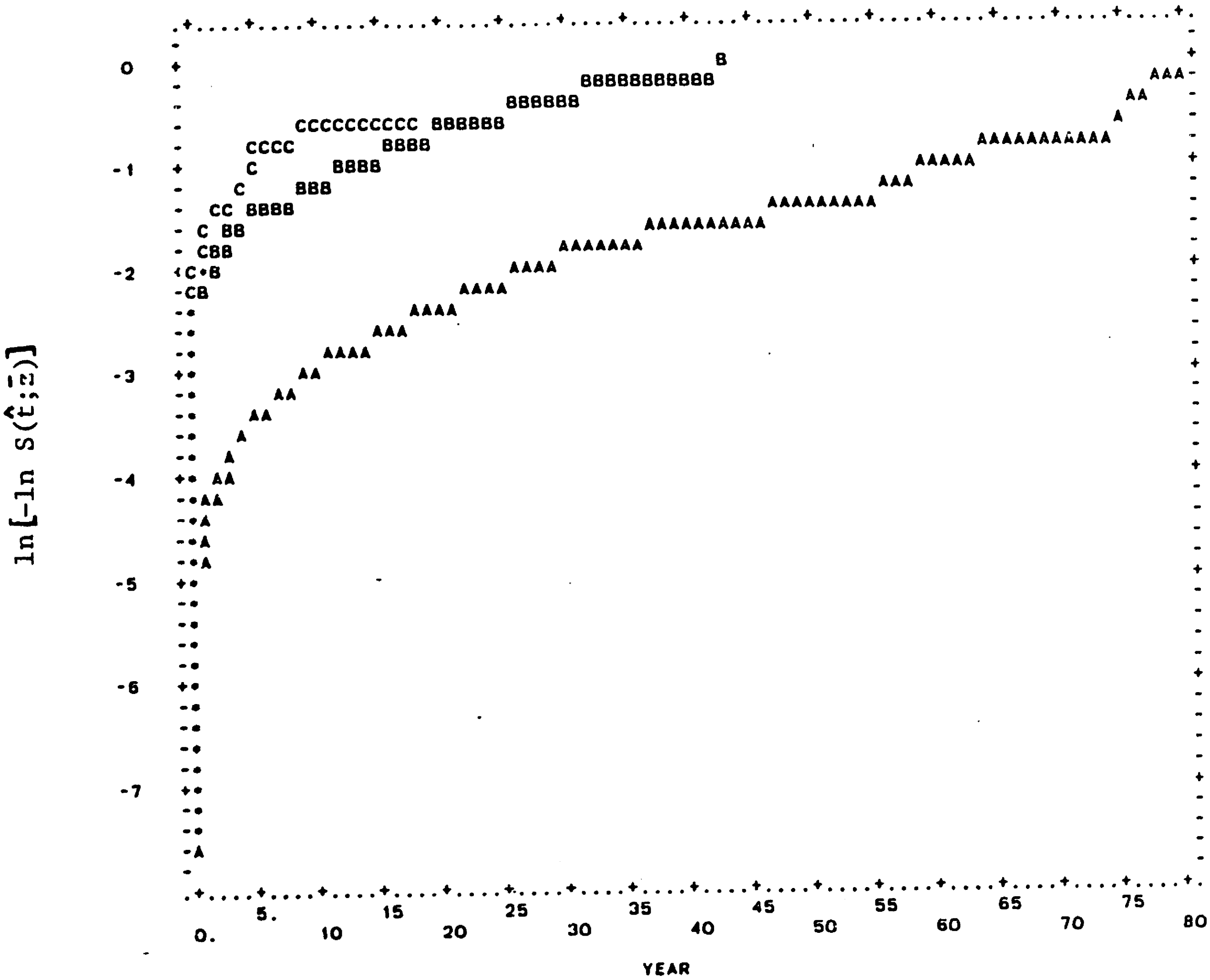
	MIN	MAX	MEAN	ST.DEV.
LNLENGTH	5.52	9.55	7.55	0.82
PRESSURE	3.0	17.3	8.40	2.07
LOW	0.0	1	0.33	0.45
C35	0	1	0.34	0.47
C50	0	1	0.24	0.43
AGE	0.0	80.0	22.79	17.94

PREVBRK = 2 (N=90)

	MIN	MAX	MEAN	ST.DEV.
LNLENGTH	6.21	9.11	7.68	0.74
PRESSURE	3.0	15.5	8.19	1.95
LOW	0.0	1.0	0.26	0.42
C35	0	1	0.28	0.45
C50	0	1	0.19	0.39
AGE	2.0	77.0	28.66	17.85

Table 5.2

DESCRIPTIVE STATISTICS OF MODEL COVARIATES
FOR PREVIOUS BREAKS EQUAL TO 0,1,2,
NEW HAVEN



A: PREVIOUS BREAKS = 0

B: PREVIOUS BREAKS = 1

C: PREVIOUS BREAKS = 2

Figure 5.4

LOG OF MINUS LOG SURVIVAL FUNCTIONS - STRATIFICATION BY NUMBER OF PREVIOUS BREAKS, NEW HAVEN

4. C35 (Dummy variable) = 1, if pipe installed during the period 1930-35

0, otherwise

5. C50 (Dummy variable) = 1, if pipe installed after 1950

0, otherwise

6. AGEBRK = Break rate = $\frac{\text{Number of breaks}(=2)}{\text{Age at time of break}}$, only if previous breaks equals to two

0, otherwise

7. P12 (Dummy variable) = 1, if previous breaks equals 1 or 2

0, otherwise

The estimated coefficients and the log minus log survival function for the average covariate vector $\bar{z}$ of the final model are shown in Table 5.3 and Figure 5.5 respectively.

We observe that all variables that appear in the final model are strongly statistically significant, with the only exception of C50, which can be considered only as marginally statistically significant. It has been included in the final model, in order to show the constrast in behavior between pipes installed during the 1930-35 period and those installed after 1950.

Log Likelihood = -2802.8503
Global Chi-square = 304.70 D.F. = 7 P-Value = 0.000

Variable	Coefficient	Standard Error	Coeff./S.E.	Exp(Coeff.)
LNLENGTH	0.5299	0.0666	7.9504	1.6987
PRESBRK	0.0931	0.0276	3.3777	1.0976
LOW	-0.5404	0.1222	-4.4216	0.5825
C35	-0.6459	0.1258	-5.1340	0.5242
C50	0.2631	0.1365	1.9274	1.3009
AGEBRK	1.7839	0.5831	3.3149	5.9531
P12	1.5726	0.2626	5.9895	4.8193

Table 5.3

ESTIMATED COEFFICIENTS OF THE FINAL
PROPORTIONAL HAZARDS MODEL,
NEW HAVEN

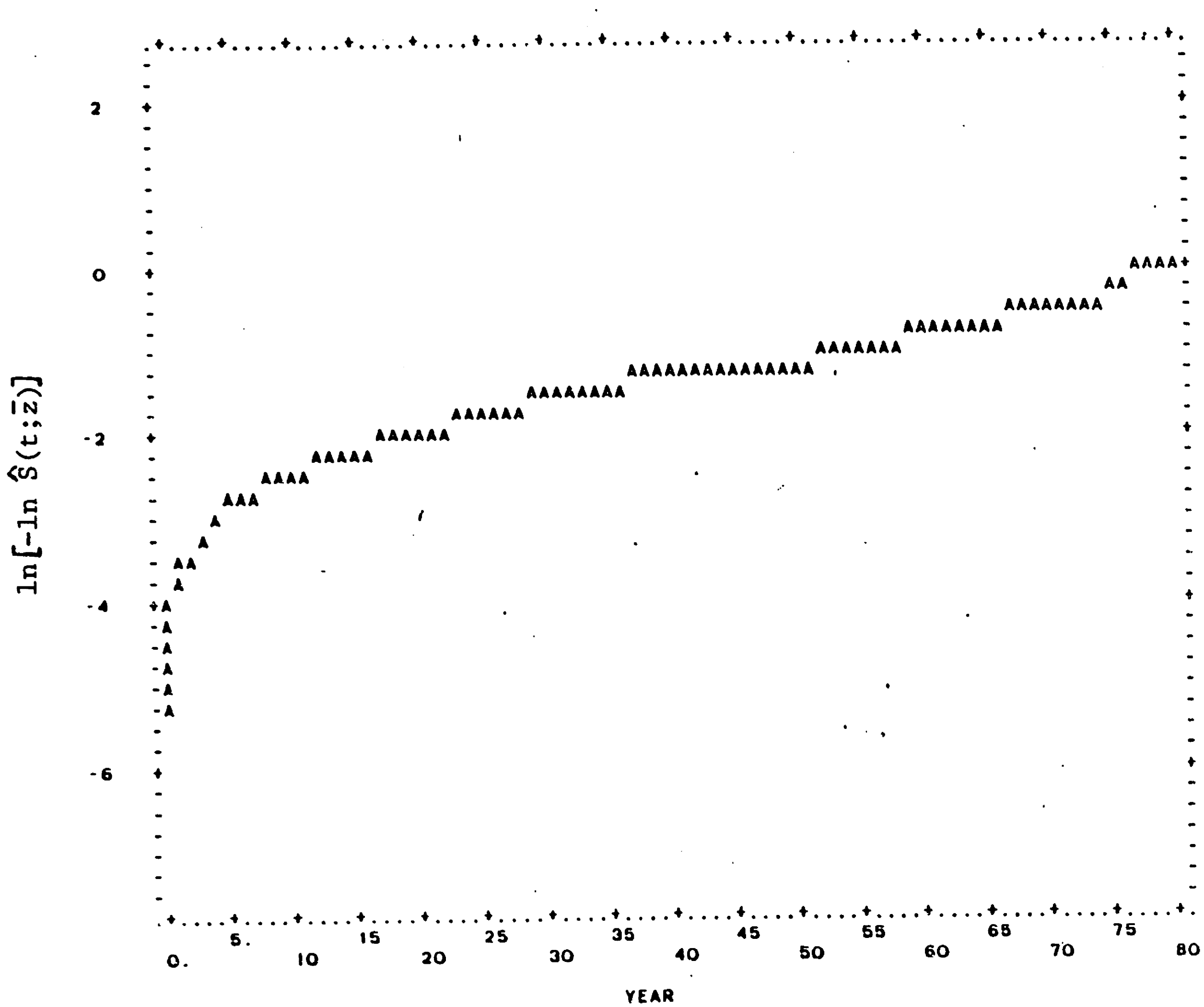


Figure 5.5

ESTIMATED LOG OF MINUS LOG SURVIVAL
FUNCTION FOR THE AVERAGE COVARIATE
VECTOR - FINAL MODEL, NEW HAVEN

5.6 Estimation of the Baseline Hazard Function - Effect of Aging on Pipe Failures

In order to evaluate the effect of aging in the pipes, the baseline hazard function $h_o(t)$ must be estimated, since the effect of time in the survival process is captured by that function. The baseline hazard function estimated by the model was approximated by a second degree polynomial, in order to derive a formula easily used in practical applications. The BMDP program plots a nonparametric estimate of the baseline hazard rate. If we approximate this by simply fitting a 3-parameter curve to the plot, the resulting smoothed estimate of $h_o(t)$ is:

$$h_o(t) = 0.000210476 - 0.000011171t + 0.000000199t^2 \tag{5.2}$$

where t is the survival time since installation, if breaks have not occurred or the time since the last break, if breaks have occurred.

A hazard function of this type indicates that the hazard rate initially decreases as the pipe does not experience any breaks in the early years of installation or early after a previous break. The effects of the aging process appear after approximately 28 years have passed and the hazard rate starts increasing after that time.

The fact that the probability of break initially decreases with time until the 28^{th} year of the pipe and then starts increasing can be shown if we set:

$$\frac{dh_o(t)}{dt} = 0 \tag{5.3}$$

from which we obtain a time $t^* = 28$ years, where the hazard rate function has a minimum.

A "bath tub" shape baseline hazard function is intuitively appealing because it is very reasonable to believe that the fact that the pipe had no break early after installation indicates the presence of favorable conditions. That is, the pipe is not defective and/or the break causing factors are not so severe for that pipe as compared to the others. Also, it is very reasonable to believe that age contributes to pipe failure much later after installation when internal corrosion has developed to a greater extent. Thus, in general, the probability of a break initially decreases as the pipe does not experience early breaks, and starts increasing again when the effect of age becomes important.

The estimated baseline hazard function also indicates that if a pipe does not break again very soon after a previous break, its chances of survival would initially increase up to a certain time period. This could be explained, if we assume that the fact that no break occurs soon after another one, implies that better remedial actions were taken for repairing the older breaks and thus it becomes less likely for a break to occur in the vicinity of the old break. Thus, for a pipe not experiencing an early next break, the chances of survival can be higher, exactly because this is a proof that appropriate remedial actions were taken for the previous break. Of course, such remedial actions are most likely to happen only in the vicinity of the break, while the remaining of the pipe length would be exposed to the various other break causing factors. But of course, those other factors are captured by the rest of the independent variables used in the model.

The above findings can be considered important since they provide a satisfactory answer to the controversy that exists in the literature about the effect of aging on deteriorating pipelines. It became clear from the discussion in Chapter 2 that other modelling techniques used in the past, could not capture this rather complex relation, which Cox's Regression allowed us to estimate through the baseline hazard function.

CHAPTER 6

ANALYSIS OF THE CINCINNATI SYSTEM

6.1 General System Characteristics

The part of the Cincinnati water distribution system analyzed in this work consisted of 452 pipes (both metal and concrete), which is about one third of the total number of pipes analyzed in the New Haven System. Similar information with that provided for New Haven was recorded for each individual pipe. (Table B1, Appendix B.) Certain differences though existed from New Haven in the way some variables were defined. Types of land uses were more accurately described by specifying the exact percent of pipe under industrial, commercial, residential and transportation land use. Instead of specifying the degree of soil corrosivity covering the whole pipe as it was done in the New Haven system, three variables were defined describing the percentage of pipe length under lowly, moderately and highly corrosive soil. In addition to the variables that were similar to those provided for New Haven, the following additional variables were defined: Pressure differential, which described the difference in pressure between the upstream and ownstream nodes of the pipe; average population density covering the pipe; a dummy variable indicating whether the pipe was cleaned and lined and a variable providing the date of cleaning and lining.

Pipe lengths in the Cincinnati system varied from 250 ft. up to 22,250 ft. About 50% of the pipes had lengths shorter than 2,000 ft. and about 15% longer than 6,000 ft.

Pipe diameters ranged from 6 to 96 inches, with only three 6 inch pipes. Thus, the data set consisted primarily of large diameter pipes (8 inches and above).

The oldest pipe was installed in 1855 and the most recent one in 1975. Approximately half of the system was installed before 1945 and about 10% of the system was installed before 1900.

Internal pressure values ranged from 15 psi up to 226 psi with a mean of 89 psi. About 50% of the pipes had average pressure higher than 77 psi. The average value for pressure differentials was 21 psi, but about 20% of the pipes were recorded with zero pressure differential, which makes the data regarding pressure differential rather unreliable.

About half of the pipes were covered by more than 50% by highly corrosive soil. Although the soil corrosivity variables were defined quite differently in the New Haven and Cincinnati systems, it appears that highly corrosive soil is more prevalent in Cincinnati.

There existed 81% metal pipes (primarily cast-iron) while the rest were made from concrete. The quite large percentage of concrete pipes points out to the need of analyzing separately this part of the data, since they are expected to follow a very different failure pattern tha.. the metal pipes.

The following general observations can be made regarding the types of land uses covering the pipes: About 60% of the pipes were under zero industrial land use. About 50% of the pipes were covered by soil with more than 40% of residential land and also less than 5% of commercial land use. Also less than 5% of transportation land use existed above the 50% of the pipes included in the data.

Only about 12% of the pipes in the system were cleaned and lined. Cleaning operations were recorded after year 1940 and the number of cleaned pipes per year increased steadily since then.

It is important to notice that break records were available only after 1940. Thus, the total number of failures recorded represents an underestimation of the actual number of failures. This issue of left-censoring is thoroughly investigated in Section 6.3.

The Cincinnati system had experienced considerably higher number of breaks than the New Haven system. About 41% of all pipes had broken one or more times, while in New Haven the number of broken pipes was only 21%. Also considerably higher number of pipes exerpienced multiple breaks (more than three) in the Cincinnati system. The time to consequent breaks reduced dramatically during the first few breaks (up to the third break), with no apparent trend beyond that point. More than 50% of all pipes with three previous breaks broke again within the next two years.

Descriptive statistics of the most important variables included in the analysis of the Cincinnati data are shown in Appendix B (Tables).

A descriptive comparison of some key variables included in the New Haven and Cincinnati systems is shown in Table 6.1. Besides the higher number of breaks, it is observed that pipes in the Cincinnati system operate on the average under higher pressure both in terms of mean and maximum values. Also larger diameter pipes represent a higher percentage of the system as compared with New Haven.

6.2 Bivariate Correlation Analysis

Pearson and Spearman correlation coefficients were calculated for pairs of the relevant variables. The results are shown in Tables and of Appendix B. No strong correlation between any of the variables appears to exist. Nevertheless, the variables representing percentages of industrial, residential, commercial and transportation land uses covering the pipes are moderately correlated ($r \approx 0.50$). Thus, although that would indicate that not all of them might be needed in the regression equations as potential predictors of failures, analysis by principle components failed to reveal any combination of them that would be more appropriate to use.

6.3 The left-censoring problem

Left-censored pipe break data represent a rather common problem

Major differences between the New Haven and Cincinnati data sets

Variable	Cincinnati	New Haven
Number of pipes	452	1391
Number of breaks	185 (41%)	292 (21%)
Diameter (inches)		
mean	26	14.5
median	24	12
minimum	6	6
maximum	96	48
Length (feet)		
mean	3,245	1,901
median	2,100	1,500
minimum	20	100
maximum	22,240	14,000
Pressure (p.s.i.)		
mean	89	79
median	78	75
minimum	15	4
maximum	226	173
Precentage pipe in highly corrosive soil (mean)	42%	*
Pipe type (% cast iron/steel)	81	96
Installation periods		
installation date, oldest	1856	1900
median installation date	1942	1935
installed before 1900	10%	--

*Soil corrosivity was defined differently in New Haven (Table A1, Appendix A). Instead of indicating the percentage of pipe in corrosive soil, it defined the average soil corrosivity corresponding to the total pipe length. In that respect, the average soil corrosivity for the whole New Haven system was 31%.

Table 6.1

COMPARISON OF GENERAL CHARACTERISTICS OF THE CINCINNATI AND NEW HAVEN SYSTEMS existing in many of the available records. The more the time that has elapsed between the year that a pipe has been installed and the year that recording breaks has started, the more information is lost concerning its "true" failure pattern. Thus, two important questions are immediately raised regarding that issue: a. To what degree the loss of information about the pipe's behavior beyond a certain period in the past is expected to limit the validity of derived predictive models for pipe failures and obscure the obtained results and b. What is the most appropriate method for treating the presence of left-censoring in the data.

Although it is very difficult to give a complete answer to the above questions, it appears that significant insights could be gained from uncensored from the left data sets, on which "artificial" left-censoring experiments are performed. That is, experiments where predictive models are derived assuming truncated from the left break record and comparisons are made between them and the model obtained from the original data. It must also be pointed out that the effect of left-censoring should not be expected to be that important for pipes that have already experienced multiple breaks during the period under observation. This is so because a more reliable trend can be established for those pipes regarding their failure pattern, due to the higher number of existing observations. It can thus be argued that left-censoring would be less of a problem in deriving predictive models for systems with many pipes under severe deterioration and thus many breaks already observed. On the contrary left-censoring must be expected to obscure relatively more the results in systems characterized by a small number of infrequent breaks.

Several experiments were performed on the New Haven set, where the available observations were transformed to reflect various left-censoring configurations. Since the Cincinnati data were truncated before 1940, that year was chosen as the starting point for the presence of left-censoring. Preliminary results obtained through the regression indicated that by setting the survival time of pipes installed before 1940 as equal to the time elapsed between the year of installation and the year of the first break after 1940, a rather severe bias was introduced in the final results as compared to those obtained by the original model. This bias basically consisted of penalizing excessively the most recent pipes, since the older ones were assumed not to have broken until 1940.

It appeared instead that the following configuration was in better agreement with the results of the original model: Break rates were assumed missing before the year 1940 and the survival time for pipes installed previous to that year was assumed to start at exactly that year. Thus, the first break of those pipes was considered to be the first one occurring immediately after 1940. The data were again stratified in order to reflect the various installation periods and Cox's regression analysis was performed on those transformed data. The results on the obtained coefficients are shown in Table 6.2 and are compared with the coefficient estimates obtained with

Predictive Model using the censored from the left data

Log Likelihood = -2423.2251
Global Chi-square = 278.02 D.F. = 7 P-Value = 0.0000

Variable	Coefficient	Standard Error	Coeff./S.E.	Exp(Coeff.)
LNLENGTH	0.5530	0.0731	7.5681	1.7385
PRESBRK	0.1171	0.0293	3.9929	1.1242
LOW	-0.4463	0.1267	-3.5230	0.6400
D1	-0.7247	0.1358	-5.3363	0.4845
D3	0.1806	0.1395	1.2944	1.1980
AGEBRK	1.7412	0.5701	3.0544	5.7043
P12	1.7785	0.2853	6.2348	5.9208

Predictive model using the original data set

Log Likelihood = -2802.8503
Global Chi-Square = 304.70 D.F. = 7 P-Value = 0.0000

Variable	Coefficient	Standard Error	Coeff./S.E.	Exp(Coeff.)
LNLENGTH	0.5299	0.0666	7.9504	1.6987
PRESBRK	0.0931	0.0276	3.3777	1.0976
LOW	-0.5404	0.1222	-4.4216	0.5825
D1	-0.6459	0.1258	-5.1340	0.5242
D3	0.2631	0.1365	1.9274	1.3009
AGEBRK	1.7839	0.5381	3.3149	5.9531
P12	1.5726	0.2626	5.9895	4.8193

Table 6.2

LEFT CENSORING EXPERIMENT, NEW HAVEN

uncensored from the left data. It is observed that no significant differences exist between the two models, given also the magnitude of the standard error of those estimates.

Those results become suggestive that a similar treatment of the left-censoring problem at the Cincinnati data where break records are missing before 1940, would be effective. It must also be expected that since for older pipes (i.e. pipes installed before 1940) the "true" number of previous breaks would be greater than or equal to the assumed value, pipes for which this would be indeed the case could be in a different failure mode than more recent pipes with the same number of previous breaks. Thus, in such cases, describing the interaction between installation periods and other explanatory variables in the regression, could become more necessary than otherwise, in order to reflect the underlying differences in failure patterns between "older" and "younger" pipes that are obscured by the presence of left-censoring.

6.4 Derivation of predictive models for pipe break failures in the Cincinnati water distribution system

The large number of pipes with many breaks in the Cincinnati system (more than 3 and as many as 35) motivated certain extensions of the methodologies applied in New Haven. Preliminary analysis indicated that pipes with multiple breaks should not be mixed in the same model with those that had experienced only few breaks, since they were clearly exhibiting different failure patterns. The threshold between pipes in the slow or fast-breaking mode was defined on the average by the occurrence of the third break. For pipes that had experienced up to three breaks, the average time between breaks clearly decreased as the number of previous breaks was increasing. Also most of these pipes could be characterized by rather infrequent break occurrences (slow-breaking stage). That is, a very similar failure pattern to that observed in the New Haven system existed. Thus, for that category of pipes the proportional hazards model appeared appropriate for their statistical analysis. Pipes that had experienced more than three breaks were usually entering a stage of multiple and frequent breaks at a rate though highly varying among indididual pipes at that stage. No trend also appeared to exist toward increasing or decreasing break-rate as the number of breaks increased. Therefore, it seemed appropriate to assume that breaks occurred at a constant rate for pipes at that stage. Since many breaks are expected to occur at that "fast-breaking" stage predictions of the yearly break-rate become more appropriate than the estimation of the yearly probability of failure. The great number of pipes in that stage made possible the derivation of predictive models for the break-rate, which are described in detail in Section 6.4.3. It must be pointed out that no such model was developed for New Haven because only a very small number of pipes were in that stage to obtain any statistically meaningful results.

Another trend observed in the Cincinnati data was that several pipes entered into the fast-breaking stage at a certain time-period after having experienced their first or second break. It thus becomes desirable to estimate the probability for entering that stage at any given time period and identify the factors that will enhance that probability. A proportional hazards model, where the time to failure was modified as being the time for entering the fast-breaking stage, would again be appropriate for modelling this process. This model is described in detail in Section 6.4.2.

6.4.1 Application of the proportional Hazards Model for pipes in the slow-breaking stage

According to the discussion about left-censoring presented in Section 6.3, the survival time for pipes installed before 1940 was measured from that year, while for the other pipes it was measured since the date of installation. Only metal pipes were included in the analysis, since concrete pipes had experienced very few breaks and a statistical analysis for this pipe category would not be meaningful.

Initially pipes were divided into three groups by period of installation: before 1920 (Period 1), 1920 to 1939 (Period 2) and

1940 to 1980 (Period 3). The year 1920 was chosen as a break point considering that it roughly corresponded to a time when pipe manufacture methods changed. The reason for defining period 3 as a separate group was that it included all pipes with observations not censored from the left.

A great number of exploratory regressions (both step-wise and ordinary) including various ways of data stratification were performed. The following variables were selected for the final regression models:

1. Logarithm of pipe length (LGLENGTH)
2. Absolute internal pipe pressure (ABSPRES)
3. Percentage of pipe in highly corrosive soil (PCTHIGH)
4. Pipe diameter (SIZEIN)
5. Land use variables:

   Percentage of residential land development covering the pipe (AVRESPCT)

   Percentage of industrial land development covering the pipe (AVINDPCT)

   Percentage of commercial land development covering the pipe (AVCOMPCT)

   Percentage of transportation land development covering the pipe (AVTRFCT)

   Population density (AVDENTY).

6. Period of installation: the following dummy variables were defined:

   Installed before 1920 (PERIOD 1)

   Installed during the period 1940-1980 (PERIOD 3)

7. Interaction variables: Several interaction variables were included in the analysis because in exploratory regressions the effects of certain covariates were found to be different in the groups of older and younger pipes respectively. All variables included in the interactions were centered by their means (corresponding to the particular age group) before the interaction levels were calculated. The following interactions were defined:

   Log length by period (LENP=LGLENGTH*PERIOD3)

   Pressure by period (ASBPRESP=ASBPRES*PERIOD3)

   Percentage of highly corrosive soil by period (PCTHIGHP=PCTHIGH*PERIOD3)

   Size by period (SIZEINP=SIZEIN*PERIOD3)

   Residential, industrial and commercial percentages by period (RESP, INDP, COMP respectively)

Since 57% of the pipes were installed before 1940 and 43% after 1940, the centered variable PERIOD3 was equal to -0.43 for the pre-1940 pipes and +0.57 for the after-1940 pipes. In cases where the main and interaction effects would cancel out in a certain period, the effect of the corresponding variable was considered zero for that period. The following variables belonged in this category:

SIZEIN3 (equal to SIZEIN for pipes installed after 1940 and 0 otherwise), AVRES3 (equal to AVRES for pipes installed after 1940 and 0 otherwise, PCTHIGH3 (equal to PCTHIGH for pipes installed after 1940 and 0 otherwise), ASBPRES2 (equal to ASBPRES for pipes laid before 1940 and 0 otherwise).

8. Previous breaking rate: the break-rate at the last break was used as a covariate in various regressions denoted by OLDRATE.

As mentioned previously the number of previous breaks affected the break-rate only before the occurrence of the third break. Similar were the findings in the New Haven system but as opposed to the trend appearing to exist there, the number of previous breaks did not have a proportional effect on the hazard rate for the Cincinnati system. This was discovered by plotting the log of minus log survival functions of groups of pipes stratified according to the number of previous breaks. Such finding indicated that different baseline hazard functions needed to be estimated conditional on the number of previous breaks.

a. Proportional hazards model conditional on one break

The analysis of the Cincinnati data concentrated primarily in deriving models that describe the time from the first break to a later break event, such as the second break, third break or entry into the "fast-breaking" stage and also the time from second to third break. Models that predict the hazard rate for pipes with no previous breaks were also derived and they are described later in this section. But given the fact that too many pipes have already broken in Cincinnati and a great number of them are already in severe deterioration, models predicting the failure probabilities of pipes with no breaks are expected to be only of secondary importance for maintenance decisions and their use would be primarily explanatory about the break causing factors.

The final proportional hazards model for time to the second break conditional on one previous break is described in Table 6.3. According to the variables found to be statistically significant, higher pressure increases the break rate of pipes installed in the post-1940 period. Also on the average pipes from the most recent period have a higher break rate than older ones.

Table 6.4 shows the derived model for time from first to third break. Such model can be useful for reliability assessments in order to know whether an early third break is likely to occur. It is observed that in addition to the variables found to be important for time from first to second break, the size of the pipe is also important for pipes installed after 1940, with the larger diameter pipes having a higher break rate. A higher percentage of commercial land development is associated with a higher

Variable	Coefficient	Standard Error	Coefficient/ Standard Error
LGLENGTH	0.261	0.111	2.35
ASBPRES2	0.0093	0.003	3.20
PCTHIGH3	0.0085	0.003	2.44
PERIOD3	0.607	0.206	2.94
SIZEIN3	0.024	0.014	1.71

Model Chi-square = 28.74 with 5 D.F, p = 0.0000

174 Observations

124 Uncensored Observations

Table 6.3

PROPORTIONAL HAZARDS MODEL FOR TIME TO SECOND BREAK CONDITIONAL ON ONE PREVIOUS BREAK, CINCINNATI

Variable	Coefficient	Standard Error	Coefficient/ Standard Error
LGLENGTH	0.320	0.056	5.74
ASBPRES2	0.0127	0.003	4.28
PCTHIGH3	0.0135	0.004	3.04
SIZEIN3	0.050	0.016	3.06
PERIOD3	0.573	0.243	2.36
AVCOMPCT	-0.004	0.029	0.14
COMP	0.043	0.019	2.21

Model Chi-square = 41.31 with 7 D.F, p = 0.0000

174 Observations

91 Uncensored Observations

Table 6.4

PROPORTIONAL HAZARDS MODEL FOR TIME TO THIRD BREAK CONDITIONAL ON ONE PREVIOUS BREAK, CINCINNATI probability of a sooner third break for the post-1940 pipes, while the opposite holds for the older pipes.

b. Proportional hazards model conditional on two breaks

Table 6.5 shows the derived model for time from second to third break. It is observed that fewer variables now affect the break-rate than in the previous models. The break-rate at the second break (i.e. the inverse of the time between second and first break) is a very significant predictor of failure. Higher pressure increases the hazard rate for all pipes and larger diameter pipes have higher chances for breaking. It is interesting to notice that length was not statistically significant in this model, while in the previous ones the logarithm of length always came in the regression as an important covariate. No full explanation can be provided for this phenomenon, although it could be argued that the covariate representing the break-rate at the second-break, which was found to be very significant in this model, could explain most of the observed variance in the hazard rates of the different pipes at that stage and thus a variable such as length ended up contributing too little in explaining this variation.

c. Proportional hazards model for pipes with zero previous breaks

A predictive model that provides the hazard rate for

Variable	Coefficient	Standard Error	Coefficient/ Standard Error
OLDRATE	0.960	0.318	3.02
ASBPRES	0.0067	0.003	2.55
SIZEIN	0.0195	0.010	1.99

Model Chi-square = 15.43 with 3 D.F, p = 0.0015

124 Observations

91 Uncensored Observations

Table 6.5

PROPORTIONAL HAZARDS MODEL FOR TIME TO
THIRD BREAK CONDITIONAL ON THE
SECOND BREAK, CINCINNATI

pipes which have not broken yet will not be so useful for maintenance decisions, particularly when many pipes in the system are already in serious deterioration and main attention is directed on them. Such model could though serve for explanatory purposes, since it is possible that within a selected set of pipes (e.g. pipes already broken), a factor may appear unimportant, while actually it was important in causing the pipe to be among those at risk of breaking.

The results obtained by applying the proportional hazards model for the time to first break are shown in Table 6.6. Pressure and length are strong predictors of failure for the older pipes but have no significant effect in the post-1940 pipes. Pipe size has a negative effect on breaks for the new pipes but no significant effect for the older pipes. Higher residential and industrial land development is also associated with higher hazard rates.

6.4.2 Application of the Porportional Hazards Model for Predicting the Probability of Entering to the Fast-Breaking Stage

Entry into the "fast-breaking" stage was defined by the occurrence of three breaks within a six-year time period and the time to fast-break was set equal with the time to the first of those three breaks. Although this measure seems somewhat arbitrary it gives a sense of the time that severe deterioration is very likely to have started. Table 6.7 shows the results obtained by that model.

(Pipes with no previous breaks or pipes censored from the left)

Variable	Coefficient	Standard Error	Coefficient/ Standard Error
LGLENGTH	0.407	0.084	4.85
LENP	-0.529	0.168	-3.15
ASBPRES2	0.010	0.002	5.00
SIZEIN3	-0.032	0.012	-2.67
AVRESPCT	0.015	0.004	3.75
AVINDPCT	0.062	0.017	3.65
PERIOD3	0.173	0.185	0.94

Model Chi-square = 83.94 with 7 D.F, p = 0.0

367 Observations

174 Uncensored Observations

Table 6.6

PROPORTIONAL HAZARDS MODEL FOR TIME TO FIRST BREAK, CINCINNATI

Variable	Coefficient	Standard Error	Coefficient/ Standard Error
LGLENGTH	0.230	0.147	1.56
ASBPRES2	0.0122	0.0034	3.48
PCTHIGH3	0.0085	0.0044	1.91
PERIOD3	0.493	0.276	1.79
AVCOMPCT	-0.007	0.011	0.59
COMP	0.057	0.021	2.64

Model Chi-square = 23.80 with 7 D.F, p = 0.0006

174 Observations

73 Uncensored Observations

Table 6.7

PROPORTIONAL HAZARDS MODEL FOR ENTERING
INTO FAST-BREAKING STAGE CONDITIONAL
ON ONE BREAK, CINCINNATI

Pressure is the most important predictor of higher break rates for the older pipes, while higher soil corrosivity increases the probability for entering into the fast-breaking stage for the younger pipes. Also, the post-1940 pipes have higher chances for entering that stage. A high percentage of commercial land development over the most recent pipes increases their probability for entering that stage, while the opposite was found to happen for the older pipes.

When a model was attempted to be derived that would predict the probability for entering into the fast-breaking stage conditional on the occurrence of two previous breaks, only the variable OLDRATE was found statistically significant with a coefficient of 0.937 and a standard error of 0.333.

6.4.3 Exponential Models for break rates after the third and sixth break

After testing the hypothesis of whether there was a trend in break times between subsequent breaks for pipes that had broken more than three times, it has been revealed that no such trend existed and the assumption of constant break-rate beyond the third break would be a very reasonable one. The third and sixth breaks were chosen as milestones for estimating the break-rate of pipes that had reached those stages respectively. Thus the assumptions of the model were: a. constant break-rate for the period under consideration, b. independent break events for each individual pipe, and c. the covariates had a multiplicative effect on the break-rate (as in the Cox regression model) and an exponential model structure was assumed.

It must be pointed out that the pipes included in these models were a highly selected set of all pipes (i.e. those that most likely are already under severe deterioration). Thus, the derived models are predicting break rates in comparison to other pipes that have reached the same starting point. For this reason it should not be considered as surprising that the effect of certain covariates on the break-rate could now be reversed as compared to the previously derived models. If for example, high pressure has contributed in bringing certain pipes at the stage of the third break, it should not necessarily be expected to appear as a significant covariate in the regression for the break rate beyond that point since there would also be pipes brought into that stage due to other break causing conditions and besides that, pressure might already be high and less varying amont the pipes of that subset.

All variables, for which interaction terms were also included in the regressions, were centered by their means. If we denote by R the estimated yearly break rate, by z the covariate vector and by b the set of estimated coefficient, then the proposed model takes the general form:

$$R = \exp(\beta z)+e \tag{6.1}$$

where e = error term of the model

Under the assumption that break events of that stage are represented by a Poisson Process with parameter tR, two sources of error will exist in predicting the number of breaks: a. The error generated by the Poisson Process, which takes place independently. That is, we cannot predict exactly how many breaks will occur in a specific time period, since break times are random. b. The model error e, due to imperfect information about the factors contributing to breaks. The Poisson error cannot be reduced, since it is associated with the underlying process. Thus, we know apriori that only part of the observed variance in the break rate could be explained by the model. Given the fact that the variance and mean of the Poisson Process are equal, the following procedure has been proposed (Zaslavsky, 1985) in order to minimize the mean square error of the estimated rate (i.e. the variance of e):

If $Var[e] = S^2$ and N is the number of breaks occured in time t, then $Var[N] = tR + t^2S^2$ and the variance of the break rate would be: $Var[N/t] = R/t + S^2$. Thus, the shorter the period of observation of a particular pipe, the more it contributes to the variance of the break rate. If we then denote by W the weight that should be given to that pipe, it will be given by: $w = \frac{1}{R/t+S^2}$ The following procedure was used in order to calculate the appropriate weights: A least-squares model was fit for the break rate $R = N/t$, using all available covariates and the residual sum of squares (RSS) was obtained. The Poisson component of the error R/t was estimated by the observed ratio N/t^2. This was summed over all pipes and subtracted from RSS to estimate the sum of e^2. Finally this estimate was divided by the number of pipes to estimate S^2. In the subsequent regressions weight W was given to each observation calculated by: $1/W = R/t+S^2$.

The results of the exponential models are shown in Table 6.8. The estimated R^2 can be considered satisfactory, particularly for the model after the sixth break. Given the fact that we approximately can explain no more than 70% of the observed variance (the other 30% is roughly associated with the Poisson Process), the derived models enable us to explain more than half of it. The covariates found to be statistically significant were fewer than in the previous models. Longer pipes were associated with higher breaking rate and also the post-1940 pipes were breaking at a higher rate. Highly corrosive soils increased the break rate. Higher pressure predicts higher break rates for the most recent pipes, while the opposite is true for older pipes. Higher industrial development was associated with lower break rate after the third break. As it has been argued previously, this reversal in the effect of certain covariates is explainable by the shift in focus towards a highly selected subset of pipes.

6.4.4 The effects of cleaning and lining

The effect of pipe cleaning and lining was tested by including a dummy variable for cleaning and lining in the regression models for time from first to second break and for break rate after third and sixth break. The dummy variable was set to 1 if the pipe had been cleaned and lined before the beginning of the event under

Variable	After break #3 Coefficient	After break #3 Coeff./S.E.	After break #6 Coefficient	After break #6 Coeff./S.E.
LGLENGTH	0.459	3.43	0.563	3.31
PERIOD3	0.644	3.40	0.945	4.81
PERIOD1	-0.946	1.87		
ASBPRES	-0.0017	0.59	0.000	0.03
ASBPRESP	0.0113	2.21	0.0125	2.14
PCTHIGH	0.0058	2.42	0.0051	1.90
AVCOMPCT			-0.022	1.69
AVINDPCT	-0.049	2.11		
Constant of the regression	-1.011	6.87	-0.660	5.20

Analysis of Variance	After break #3	After break #6
Residual sum of squares (RSS)	88.7	49.4
Regression sum of squares	45.6	42.2
Corrected total SS (CSS)	134.3	91.6
Uncorrected SS (USS)	246.6	180.7
R^2	0.34	0.46

Table 6.8

PREDICTIVE MODELS FOR YEARLY BREAK RATE
AFTER THE THIRD AND SIXTH BREAK,
CINCINNATI

consideration, i.e. before the first, third, or sixth break respectively. (Nevertheless, most pipes had been cleaned and lined before their first recorded break.)

In every case the effect of cleaning and lining was not statistically significant in the regression equations. Thus, no increase or decrease in the break rate was found to be associated with the cleaning and lining of pipes. Although this result can be important for maintenance considerations, it should still be viewed with caution, since we do not know how the cleaned and lined pipes were preselected. If the criteria for selection were partially based on the fact that they were in better or worse condition in terms of structural integrity than the other pipes, then the effect of cleaning and lining could have been masked in the regression equations. Nevertheless, it appears unlikely that the above could be important selection criteria, as it seems more plausible that candidates for cleaning and lining were chosen based on carrying capacity requirements, increased pumping costs and water quality problems.

6.4.5 The estimation of the baseline hazard functions

The baseline hazard functions of the derived proportional hazards models were estimated either by fitting simple polynomials to the cumulative hazard function (model for time to second break conditional on previous break) or by calculating average numerical values in 5 year time intervals in cases where the functional form was more complicated (models for time to third break and fast-breaking stage). The following baseline hazard functions were estimated:

a. Baseline hazard function for time to second break conditional on one previous break -- $h_0^{12}(t)$

$$h_0^{12}(t) = 0.1297 - 0.0086.t + 0.000213.t^2 \tag{6.2}$$

b. Baseline hazard function for time to third break conditional on two previous breaks -- $h_0^{23}(t)$

The estimated baseline hazard function is given in Table 6.9.

c. Baseline hazard function for time to fast-breaking stage conditional on one previous break -- $h_0^{1f}(t)$

The estimated baseline hazard function is given in Table 6.9.

We observe that the estimated baseline hazards for time to second break has a "bath tub" shape, a finding similar with that in New Haven. The baseline hazard function for time from second to third break is initially of bath tub shape (up to the fifteenth year since the second break). Beyond the fifteenth year though after the second break the hazard declines again. A similar irregular pattern is also observed for the baseline hazard for time from first break to fast break stage. For this reason tabulated values of those functions were used rather than polynomial approximations. The general point though that must be made is that pipes in the slow breaking stage will usually have higher chances for breaking again soon after a break and, at least in the next few years, the chances of breaking are reduced the more the pipe survives after the last break.

Years from start	Second to third break	First break to entry to "fast-break state"
	h	h
0	.139	.006
5	.037	.021
10	.127	.007
15	.040	.009
20	.028	.014
25		.009

Table 6.9

BASELINE HAZARD FUNCTION – $h_0^{23}(t)$,

BASELINE HAZARD FUNCTION – $h_0^{1f}(t)$
CINCINNATI

By comparing these functions with those derived for New Haven, it can easily be shown that the average pipe in the Cincinnati system has about three times greater chance for breaking than the average pipe in New Haven. This observation reinforces the argument that the Cincinnati system is under much more severe deterioration than the New Haven system.

If the derivative of the function $h_0^{12}(t)$ shown in Equation (6.2) is set equal to zero we can obtain the value of time t, for which the hazard rate starts increasing after the first break. This time is found to be equal to 20.18 years. The corresponding value obtained for the New Haven System was 28 years.

CHAPTER 7

INTERPRETATION OF RESULTS OBTAINED BY THE PREDICTIVE MODELS

7.1 Introduction

The results obtained from the analysis of the New Haven and Cincinnati systems have generated interesting insights about the various phases of pipe failure that can exist in deteriorating water distribution systems. They clearly point out the high variability in break rates among individual pipes of a given system and underline both the common and different characteristics in failure patterns that can exist among the various systems.

The analysis made clear that a pipe can go through different phases (or modes) of failure during its lifetime. Such phases would primarily be characterized by either few infrequent breaks occurring during rather long time spans, or by frequent multiple breaks occurring during short time periods. If the majority of pipes in the system experience the former type of failures, then the system is in a relatively "good" overall condition (New Haven), while if a high percentage of pipes experience the latter failure pattern then the system is in relatively "severe" deterioration (Cincinnati). The proportional hazards model proved to be useful both in describing the probability of failure for pipes in the slow breaking phase and the probability for entering the multiple (fast) breaking stage. After a pipe has already entered the stage where it experiences many frequent breaks (fast breaking state) then predictive models that provide the expected break rate rather than the probability of failure become more useful.

A detailed physical interpretation of the results obtained by the developed models is given in Section 7.3. The most important of the findings can be summarized as follows:

a. The proportional hazards model can describe very well the early phase of deterioration of a pipe, where the multiple breaking stage has not been reached. A variation of the proportional hazards model could also work very effectively in providing us with the probability that a pipe will enter that multiple breaking stage at a given time period.

b. The baseline hazard function turned out to have a "bath tub" shape during the period from zero to first and then to second break. That is, the probability of failure initially decreased with time and later started to increase. Such effect of time has not been captured by previous methodologies, because they do not have the flexibility to allow for an unspecified baseline hazard function (non-parametric method).

c. The number of previous breaks was found to affect the probability of failure only during the first two breaks. Beyond that point the majority of pipes entered a stage of very frequent breaks, where the occurrence of previous breaks was immaterial for future predictions. At that stage a rather constant break rate was established and the breaks could be modelled as Poisson arrivals.

d. Variables indicating internal pipe pressure, soil corrosivity and land development were generally found to have an impact on the probability of failure and the break rate. The effect of each one of them though varied depending on the stage of deterioration and other circumstances affecting individual pipes.

e. In both the New Haven and Cincinnati systems pipes installed in different periods experienced different failure patterns, with the most recent pipes (after 1940) performing on the average worse than the older ones.

f. The probability of failure did not increase proportionally with length, indicating a non-uniform distribution of break causing factors along the pipe and a localized nature of breaks. Such finding is important because in previous studies that proportionality was assumed to hold, and this also had implications on the way pipe lengths were coded in the data.

7.2 The issue of high variability in break rates among individual pipes

In order to interpret the results of our analysis, it is helpful to define a "reference pipe," describe its chance of breaking as a function of age, and then the risk of other pipes breaking can be described relative to that of the "reference" pipe. We arbitrarily define a "reference" pipe in the New Haven system as one in which:

Length = 1500 ft. (about average in our sample)

Pressure = 80 PSI

Low = 38%

Prevbrk = 0

Date of installation before 1930.

For such a pipe, the rate of breaks at age t is approximately

$$h(t) = h_o(t).e^{4.19}$$

where $h_o(t) = 0.00021[illegible]476 - 0.000011171\,t + 0.000000199\ t^2$ and $\beta z = 4.19$.

Thus, we see that yearly risk of a break for the reference pipe changes with age from about 0.00021 at age 0 to 0.00005 at age 30, to 0.00059 at age 80.

Figures 7.1 and 7.2 present the histograms of the covariates involved in the model and two different types of hazard rates. RH is the estimated relative hazard of each pipe as a whole, while RHPF is the relative hazard per foot of length of each pipe where $RH = \exp\ \{\beta^T(x-x_0)\}$ and $RHPF = RH \div (\text{length}/1500)$

$\underline{\beta}$ is the estimated coefficient

$\underline{x}$ is the covariate vector of any particular pipe in our data set

$\underline{x}_o$ is the reference covariate vector corresponding to LNLENGTH = 7.31, PRESBRK = 80, LOW = 0.38, C35 = 0, C50 = 0, AGEBRK = 0, and PREVBRK = 0.

The histograms are displayed on a log scale of relative risk and represent N = 1771 observations.

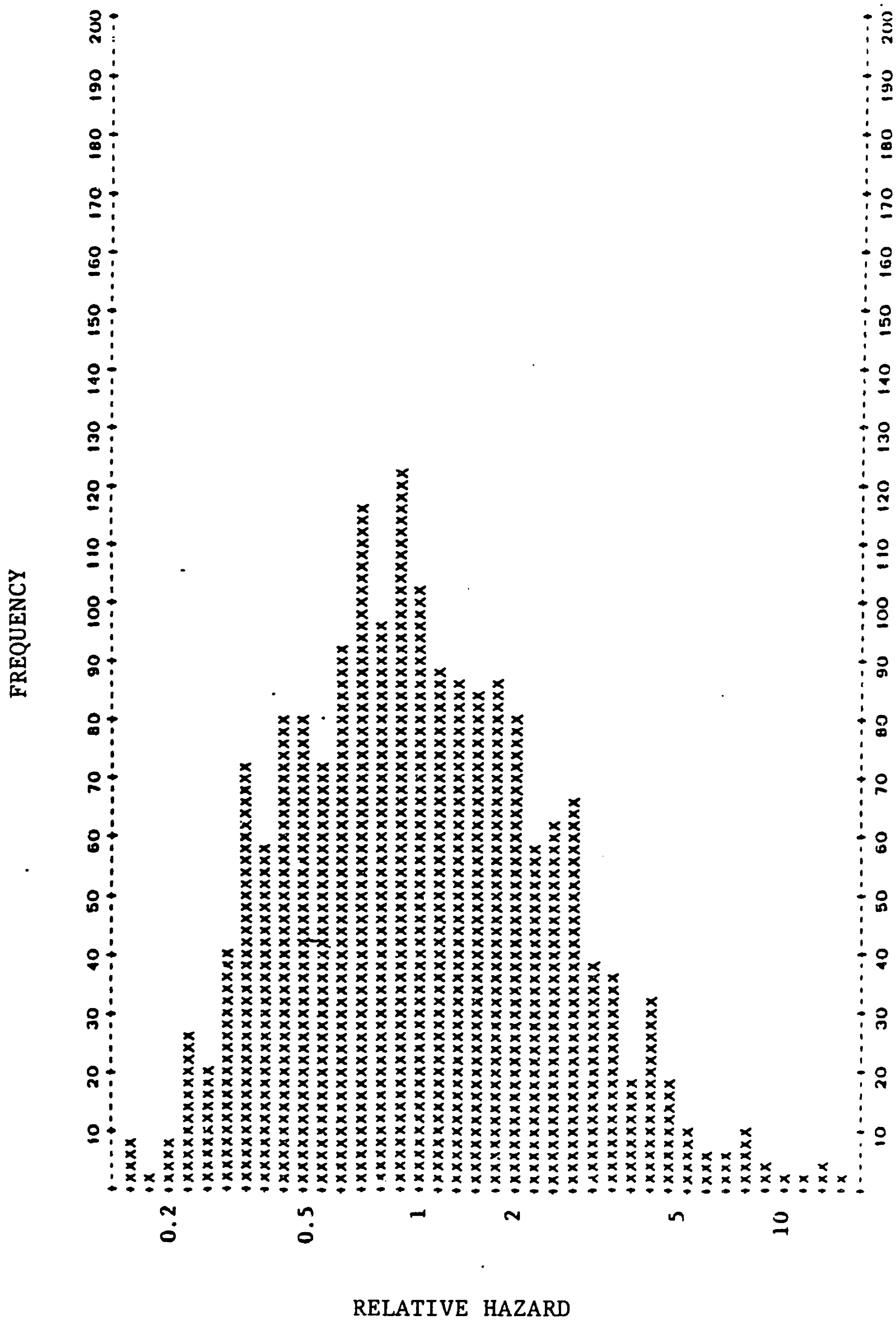


Figure 7.1

HISTOGRAM OF RELATIVE HAZARD RATE (RH),
NEW HAVEN

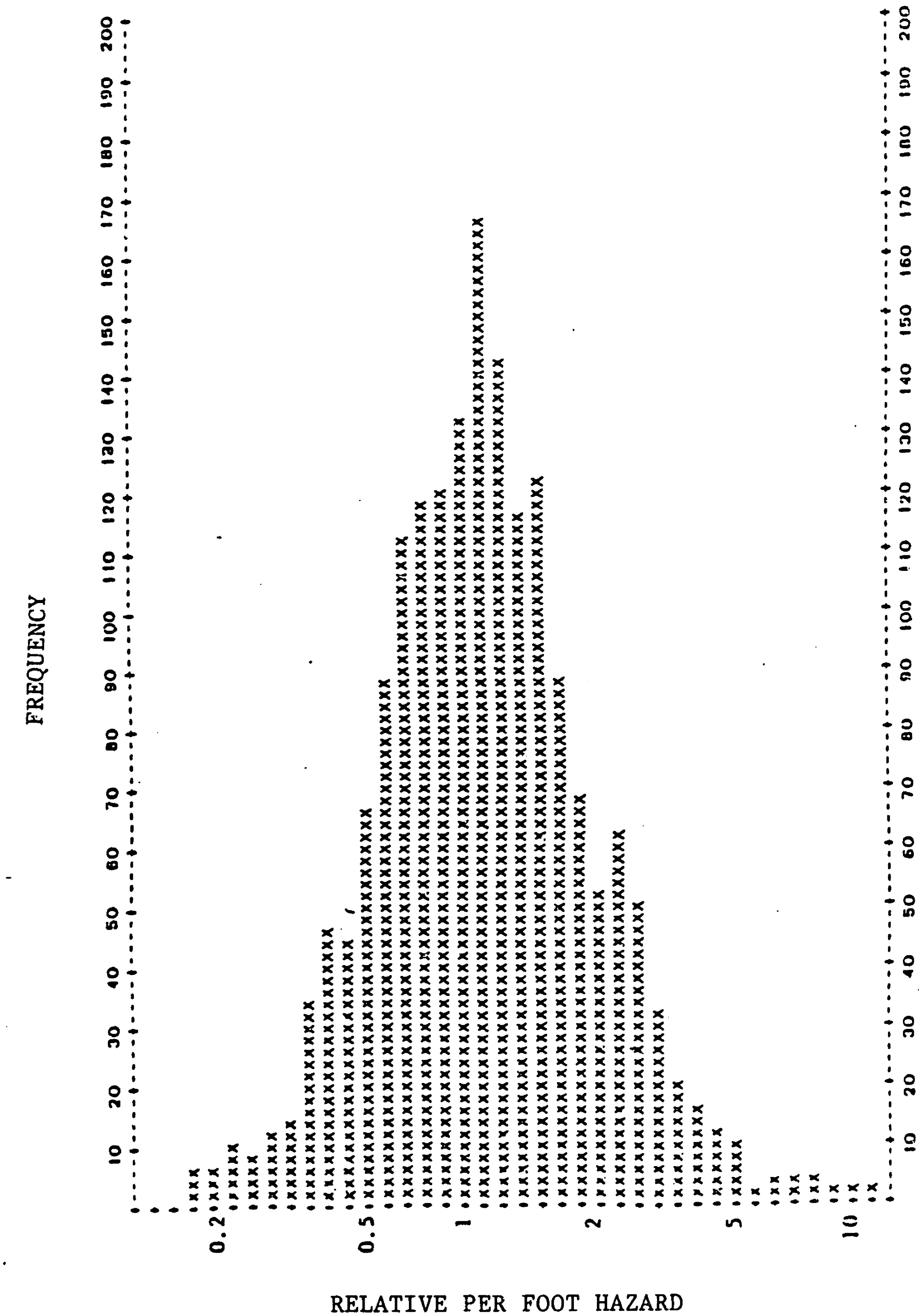


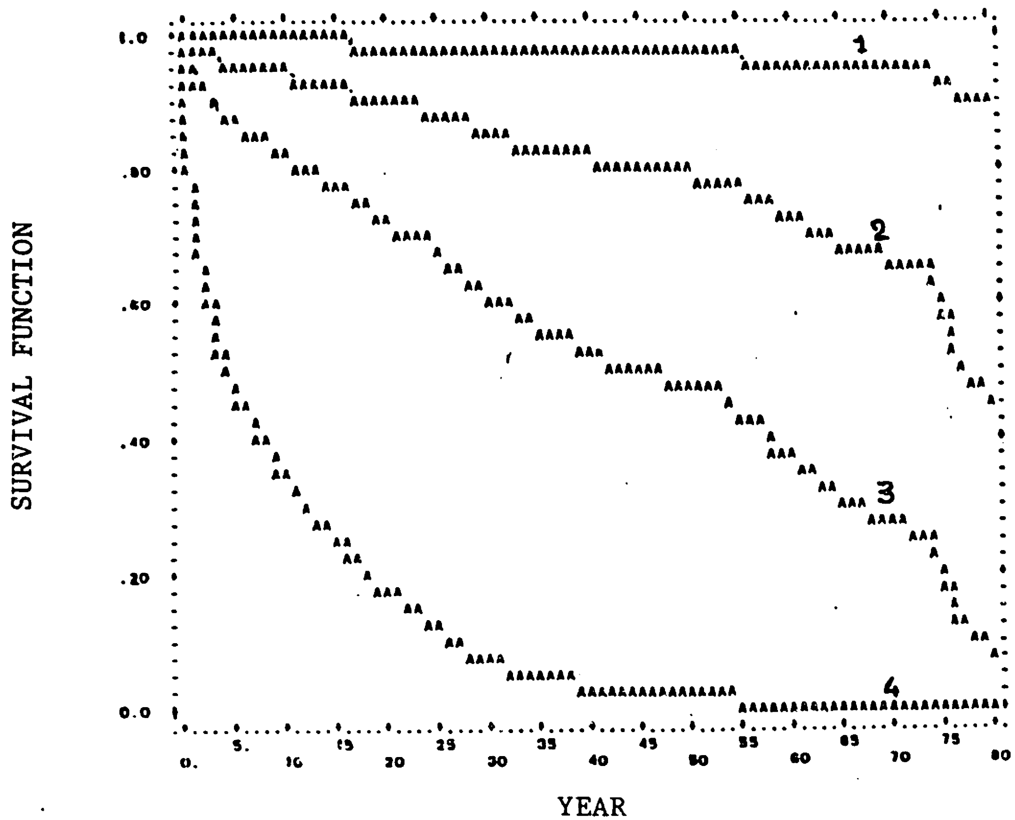
Figure 7.2

HISTOGRAM OF RELATIVE PER FOOT HAZARD
RATE (RHPF), NEW HAVEN

Note that both the hazard rate per pipe and the hazard rate per foot of pipe show great variation. Many pipes are 10 or 20 times as likely to break as are the other pipes. This display shows that there is potentially a large advantage to be gained by identifying pipes with high hazard rates and designing maintenance policies accordingly.

Examples of survivor functions of representative pipes in the New Haven system are demonstrated in Figure 7.3. The plotted functions demonstrate the high variability in hazards rates that can exist among individual pipes with different characteristics and previous break history.

Pattern 1 shows the estimated survival function of a 100 ft. pipe installed during 1930-35, covered totally with minimum land development which has already experienced 2 breaks with the last break occurring after 77 years from time to installation. Pattern 2 shows the survival function of a 100 ft. pipe with no previous breaks installed after 1950, covered with maximum land development and subject to very high internal pressure of 173 psi. The slope of the survival curve indicates the importance of internal pressure in decreasing the probability of survival for a pipe with no previous breaks. Pattern 4 shows the dramatic drop of the probability of survival for a "worst case" pipe, 14,000 ft. long, with two previous breaks and which had the second break only 4 years after installation. Pattern 3 shows the survival curve of a 100 ft. pipe with 2 previous



	PATTERN 1	2	3	4
LN[LENGTH]	4.6	4.6	4.6	9.5
PRESBRK	0	17.3	0	0
LOW	1	0	0	0
C35	1	0	0	1
C50	0	1	1	0
AGEBRK	2/77	0	2/4	2/4
PREVBRK	2	0	2	2

Figure 7.3

ESTIMATED SURVIVOR FUNCTIONS OF REPRESENTATIVE PIPES, NEW HAVEN breaks, installed after 1950, covered with maximum land development and which experienced the second break 4 years after installation. Again a steep decrease of the survival function is observed.

The ability to capture this high variability by the developed model will have important implications on reliability analysis and economic evaluation for repair, replacement and rehabilitation strategies. Those issues are discussed in Chapter 9.

7.3 Assessment of the predictive power of the models

The predictive power of the derived models can be tested by examining the spread of the predicted probabilities of failure. A model that can distinguish between pipes that are very likely to break and others that are not will be very useful, while one which gives about the same failure probability for every pipe will not be very useful. Figures 7.1, 7.2 and 7.3 have already demonstrated this capability of the model derived for New Haven.

A measure that can also provide a good sense of the predicted variation in failure probabilities is that of relative risk. Relative Risk (RR) is defined as the change that would occur at the hazard rate by a given change in one of the model covariates. Such measure has been calculated for all of the derived models. The changes in the covariates applied for the New Haven model are as follows:

Doubling the length, increasing pressure by 10 psi, changing the number of previous breaks from zero to one, assuming a pipe installed after 1950, assuming a pipe installed during the period 1930-1935, increasing the percentage of low land development from zero to 100%, assuming a change at the break rate from 0.05 to 0.50 breaks per year. The calculated values of relative risk, denoted by RR are shown in Table 7.1. The results of this table can be interpreted as follows: if for example length has a relative risk of 1.45 that would imply that by doubling the length the probability of failure would increase by 45%.

Variable	RR (Relative Risk)
LNLENGTH	1.44
PRESBRK	1.10
LOW	0.58
C35	0.52
C50	1.30
AGEBRK	2.23
P12	4.81

Table 7.1

RELATIVE RISK OF INDIVIDUAL PIPES,
NEW HAVEN

A similar measure has been calculated for the Cincinnati models. The following covariate changes were considered: Doubling of pipe length, changing diameter by one standard deviation, changing pressure by 10 psi, changing dummy variables from false to true, changing the percentage of highly corrosive soil from zero to 100%, changing the percentage of residential and commercial land development by one standard deviation, similarly changing the percentage of land development interaction terms by one standard deviation, changing the break rate to the second break from 1/3 to 1/20 breaks per year. The calculated values of relative risk are shown in Tables 7.2 and 7.3. It is interesting to observe that if a pipe is covered fully by highly corrosive soil, its chances of breaking again after the first break are higher by 134% from a pipe that is totally not covered by highly corrosive soil. Also, if a pipe is installed after 1940 its chances for having a second break are 84% higher than if it is not. Such pipe would also have 64% greater chance for entering into the fast-breaking stage. Among the pipes which have already broken three times those installed after 1940 have 90% higher break-rate than the older pipes.

If all pipes are ranked by their predicted hazard rates and we denote by PX the Xth percentile of the hazard rate (e.g. P90 = the 90th percentile) then a different measure of the predictive power of the model can be obtained if we consider the ratios of various percentiles. For example if we calculate the ratio P90/P50 it will provide us with a measure of how much worse on average all pipes in

CINCINNATI SYSTEM, Proportional hazards model for time from first break to second break and entry to "fast-break" stage - Relative Risk Calculations.

	To second break	To "fast-breaking" stage
Variable	RR	RR
LGLENGTH	1.20	1.17
ASBPRES2	1.10	1.13
PCTHIGH3	2.34	2.34
SIZEIN3	1.27	--
PERIOD3	1.83	1.64
AVCOMPCT	--	1.07
COMP	--	1.33

CINCINNATI SYSTEM, Proportional hazards model for time from second break to third break and entry to "fast-break" stage - Relative Risk Calculations

	To Third break	To "fast-breaking" stage
Variable	RR	RR
SIZEIN	1.22	--
ASBPRES	1.07	--
OLDRATE	1.21	1.21

Table 7.2

RELATIVE RISK OF INDIVIDUAL PIPES IN SLOW-BREAKING STAGE, CINCINNATI

	After third break	After sixth break
Variable	RR	RR
LGLENGTH	1.37	1.47
PERIOD3	1.90	2.57
PERIOD1	2.58	1.00
ASBPRESP	1.02	1.06
PCTHIGH	1.06	1.65
AVCOMPCT	1.79	1.25
AVINDPCT	1.28	—

Table 7.3

RELATIVE RISK OF INDIVIDUAL PIPES IN
FAST-BREAKING STAGE, CINCINNATI

the worst 10% will do compared with the median pipe. The ratios P90/50, P50/10, P90/P10 and P75/P25 were calculated for the various models and are shown in Table 7.4. We observe that as the number of previous breaks increases the predicted differences among the various pipe categories decrease. Such finding should be expected since as the number of previous breaks increase we focus more and more on the most "bad" pipes.

Model	P90/P50	P50/P10	P90/P10	P75/P25
Start to first break	2.4	3.0	7.4	2.4
First to second break	2.1	1.8	3.7	2.2
First to fast break rate	2.2	2.1	4.7	2.4
Second to third break	1.8	1.7	3.0	1.9
Second to fast break	1.6	1.1	1.8	1.1

Table 7.4

RELATIVE PREDICTED FAILURE RATES BY
DERIVED MODES, CINCINNATI

7.4 Physical Interpretation of Variables Found to be Significant Predictors of Failure at the Various Stages of Pipe Deterioration

It has been pointed out previously, that one of the benefits expected to be obtained by a rigorous statistical analysis, are the insights gained about the break causing mechanism, which could affect future construction and maintenance practices. A satisfactory physical interpretation of the variables found to be important by the statistical analysis will also help for making judgments about optimal repair, replacement and rehabilitation policies in citites, where it is impossible to derive a good predictive model, due to poor or unavailable data.

The following interpretation can be given for variables found to be significant from this analysis:

Internal Pressure

Pressure was found to play a role in pipe failures observed both in the New Haven and Cincinnati systems. Its effect though was not quite the same in the two systems. While higher pressure increased the probability of failure for a pipe with no previous breaks in the New Haven system, it was not found to be an important predictor of future failures once a pipe had already broken. On the contrary, in the Cincinnati system, higher pressure also increased the probability of future failures for pipes already broken once and which were installed before 1940. It also increased the probability for any pipe already broken, to enter a state of multiple and frequent failures. Although it appears that high pressure does not cause a loss of pipe wall thickness by itself, with time, it is reasonable to argue that the stresses imposed on the pipe wall due to high pressure contribute to the occurrence of a break if the wall is already eroded due to corrosion. Due to the considerably higher number of breaks per unit length in the Cincinnati system, it seems plausible that deterioration due to corrosion could have progressed much more than in the New Haven system. Thus, if this is indeed true, it is explainable why pressure turned out to be more significant than other factors in the Cincinnati system as compared to New Haven. It is also important to point out the significant effect of pressure on pipes with multiple breaks. As it was found from the analysis of the Cincinnati data, high pressure significantly increased the chances of a pipe to experience multiple and frequent breaks in the future. Such findings could have important implications on operating practices of a water utility and system design characteristics.

In order to explain better the effect that pressure can have on breaks, the forces caused by pressure should be examined. Internal pressure, including water hammer, creates transverse stress or hoop tension at a pipe. The hoop tension per linear inch of length is given by: $P/2 = pd/2$ where P is the total force [in ℓb], p is the intensity of pressure [in psi] and d is the diameter of the conduit.

The resulting transverse stress is then given by:

$s = pd/2t$ where t is the wall thickness in inches.

It can thus be seen that as the pipe wall thickness t decreases, the transverse stress s can rapidly increase. As it is known, cast iron pipes, experience localized (pitting) corrosion. Thus, it should be expected that local failure, under the presence of a corroded pipe wall, might be caused by high pressure. It would also have been useful if information on flow velocities had been available, since water hammer pressure and unbalanced dynamic pressures at bends are a function of it (water hammer pressure is a linear function of flow velocity, while dynamic pressures at bends change with the square of flow velocity). Such information nevertheless had not been provided in the data set.

Land Development

Variables describing the type and percentage of land development covering a pipe could be used as surrogates for external loads transmitted on pipes (e.g., traffic), presence of other structures (buildings and other underground utility lines), activities of other underground utilities (e.g., higher land development could indicate higher chances for excavations related to sewer and gas pipelines). The effect of frost penetration also becomes much more important at the presence of high land development because dynamic loads from traffic transmitted on pipes can then become very high.

A problem though associated with such variables that describe land development is the fact that no direct link can be made between them and the exact cause of a break along the pipe length. In other words, it becomes very hard to infer for which one of the above mentioned forces they represent a surrogate. The inter-temporal variation of such variables also complicates the attempt of explaining their exact impact on breaks. Direct comparisons of the results related to those variables, obtained from the New Haven and Cincinnati analyses, are impossible to be made, because in New Haven only one variable representing land development was available (an average of industrial, transportation, commercial and residential land uses), while in Cincinnati four different variables corresponding to the above types of land development were available. High land development was found to be an important failure predictor in the New Haven system. In the Cincinnati system, pressure was more important for the older pipes, while highly corrosive soil and land development (more precisely residential and commercial land use) were important only for the more recent pipes (installed after 1940). This finding could reinforce the argument that excessive corrosion was much more predominant in the Cincinnati system, and thus other factors related to corrosion induced failure (such as pressure and highly corrosive soil) could now explain much more about failures than in the New Haven system. Nevertheless, it is important to notice that high commercial land development became a very important predictor for younger pipes entering the multiple and frequent breaking stage in the Cincinnati system.

In order to understand better the linkage between the forces acting on a pipe and the presence of high land development, the following observations can be made regarding the stresses developed on underground mains:

External loads (soil cover, traffic loads, etc.), foundation reactions (manner of pipe support), and the weight of the conduit produce flexural stress at the pipe. The magnitude of external loads imposed on water mains depend on the presence of many factors, the most important of which are: a) the type of the covering and bedding material, b) the method of supporting the conduit, c) the width and depth of the trench, d) the method of backfilling, and e) the rigidity of the conduit. Exact information for several of the above factors is not provided in many cases and details about pipe installation procedures are usually missing from the data sets. Some information concerning those factors could, however, be revealed by general installation practices applicable during past time periods. As a general observation, forces acting on pipes from soil cover increase as we move from cohesionless granular soil (minimum force), to saturated clays(maximum force). The flexural stress caused on pipes by external loads is proportional to the amount of the external load and approximately inversely proportional to the square root of the depth of cover of the conduit. Those loads might, however, cause significantly higher than normal stresses if a) some uneveness of the surface on a street causes them to become dynamic, b) bedding material is not uniform under the whole pipe length (e.g., a pipe might be resting partially on a rock), thus causing bending moments on the pipe, and c) interaction exists between frost penetration and those kind of loads.

The presence of high land development is certainly an indication that those forces might be more severe than otherwise, and the fact that it has been found to be statistically significant in the regression analysis for both the New Haven and Cincinnati systems, reinforces that belief. Much more detailed records and additional research will be needed though in order to understand the exact relationship between any of the land development describing variables and the pipe failure mechanism.

Age at the Time of Second Break

The statistical analysis of both data sets has clearly shown that the age of the pipe at the time of the second break is a very important predictor of the hazard rate. The less the age of the pipe at the second break, or the higher the break rate up to the second break, the higher the chances for the pipe to break in the future. A possible physical explanation for this can be the fact that a pipe experiencing a second break early in its life is somehow "more defective" than a pipe which had a second break late in its life. Or, it might be the case that some unknown break causing factors are present and are certainly more severe in the case of a pipe that broke twice early in its life. (That is, an early second break could indicate a strong correlation between first and second break.)

The fact that age of the pipe at the first break did not come up as an important predictor is not fully explanable, although it could be definitely argued that a first break is certainly not as good proof of unfavorable conditions or defects as an early second break is. For the severely deteriorated system of Cincinnati, the stage between second and third break was a rather "grey" area, which would determine whether a pipe enters the fast breaking state or not. Thus, we observe that more in that system rather than in the New Haven system, the effect of other variables on the hazard rate after the second break, is masked by the effect of the variable describing the break rate up to that point. In any case though (i.e., "good" or "bad" overall condition of a water distribution system), an early second break would significantly increase the probability of having an early third break.

Installation Periods

Both studies (from the New Haven and Cincinnati systems) have clearly indicated that the most recent pipes (i.e., those installed in the 40's, 50's, and beyond) performed worse than the older ones. Particularly for the Cincinnati system, the most recent pipes not only had higher chances for experiencing breaks overall, but also were breaking at a much higher rate once they entered into the fast breaking state than the older ones.

The fact that pipes installed during the period 1930-35 in the New Haven system performed much better and pipes installed after 1950 performed worse and the very similar finding for Cincinnati, where the after 1940 pipes performed much worse, could be explained by changes in construction practices and materials used in various time periods. Similar findings are reported by O'Day (1984), where it is pointed out that technology changes have not always improved water main reliability.

In the city of Philadelphia, it was found that mains laid in the 1948-1952 period experienced a high rate of failure due to leadite joint related problems. This lead substitute material causes split bell failures even though the main wall is in relatively good condition. It is also known that a decrease in cast iron strength (lower safety factors, thinner walls) has been introduced for pipes installed after the 1935 period (pit cast versus sand spun, for the older and more recent pipes respectively).

Thus, the significantly better performance of the pipes installed in the period roughly before 1940 can satisfactorily be explained by the above argument and also by the fact that installation procedures could have been more careful and thorough during that period. This observation about the better performance of those pipes has also been confirmed by the water utility managers of the City of New Haven Although the structural causes (materials, joint-types, etc.) for explaining this finding are quite strong, another reason for such observation could be an existing pre-selection of pipes in the data set. That is, the very bad old pipes could have already been removed and not reported in the data. Of course, there is no evidence indicating to what extent this might be true. Whatever the case though, the fact that older pipes in no way were found to perform worse than the most recent ones, can have significant implications on currently applied replacement policies.

Number of Previous Breaks

The number of previous breaks was found to significantly affect the hazard rate of pipes. Since the models developed for predicting the next break in the New Haven system were merged into one model using as covariate a dummy variable indicating the number of previous breaks, the impact of the number of previous breaks on the hazard rate was reflected at the coefficient of that covariate. In the Cincinnati system, where no merging of the models was theoretically justifiable because of differences in the baseline hazard functions, the effect of previous breaks on the hazard rate is reflected in the change at the coefficients of the estimated baseline hazard function for each particular case. The negative effect of previous breaks on the hazard rate has also been established in numerous studies and observations of deteriorating water distribution systems (Clark, 1982). Since the break causing mechanisms and their interactions are not yet well understood, occurrence of breaks in a pipe simply reveals the fact that those break mechanisms are present in such a pipe and will thus increase the hazard rate and accelerate failures for that pipe in the future relatively more than that of pipes which have experienced less or no breaks in the past. Thus, previous number of breaks could be a surrogate for factors that were not available in the data, as for example, bedding conditions, soil disturbances caused by work from sewer contractors, pipe defects, and other localized unfavorable situations. It is also important to point out that the number of previous breaks affects the hazard rate only during the first few break occurrences. The analysis of the Cincinnati data clearly indicated that once a pipe enters into a "fast breaking" state then the number of previous breaks or the past break rate become immaterial for predicting future breaks. This finding clearly challenges many of the proposed model structures that have appeared in the literature.

Corrosive Soil

Highly corrosive soil was found to have an impact in increasing the hazard rate of the most recent pipes (installed after 1940) of the Cincinnati system. It also contributed to a higher break-rate for all pipes, which entered into the "fast breaking" state. Corrosive soil on the other hand was not found to affect the probability of failure in the New Haven system. A possible explanation for this happening could be the fact that highly corrosive soil had indeed contributed significantly in making the Cincinnati system much more failure prone than the New Haven system. That is, it could be the case that soil conditions were much more corrosive in Cincinnati, while in New Haven soil corrosivity (and thus external errosion of pipe wall) did not pose severe problems. If such argument is true, then it could also very well explain why so many more pipes were in a "fast breaking" state in the Cincinnati system. The above phenomenon could also be related to differences in anticorrosion measures taken by the two utilities. The fact that highly corrosive soil was found to be contributing to increased hazard rates during the first few breaks of only the most recent pipes in Cincinnati could be related to changes in pipe standards during the years (e.g., thinner pipe walls after 1940). As it can also be seen from the descriptive statistics obtained by period of installation for the Cincinnati system among the pipes installed before 1940, those which were also breaking more frequently were not on the average laid in more corrosive soil than those installed before 1940 and which did not break too much. Thus, even though soil corrosivity could very well have played a role in this category of pipes, its effect does not show up in the regression, since higher pressure for the older pipes was possible to explain more about the increased failure rate. On the other hand, among the most recent pipes, those that experienced the most of the breaks were also installed in highly corrosive soil, while those with fewer breaks were not.

It is critical at this point to mention that the variable describing the level of soil corrosivity only partially must be expected to represent the effect of corrosion in pipes (as also has been argued in Chapter 4). Corrosion in metal pipes can occur under many different ways, and most of its progression with time is expected to be captured by the baseline hazard functions used in the derived predictive models. It is essential to notice that all pipes are expected to corrode with time and the question becomes at what rate this will occur given the specific environmental conditions applied on each pipe. As it has been reported by a study performed by the National Bureau of Standards at more than 150 sites nationwide for more than 50 years (Husock, 1982), ferrous metals, including cast iron and ductile iron, corrode at the same rate underground under similar environmental conditions. Thus, one should not expect the most recent ductile iron pipes to corrode at a slower rate than the older pipes, as far as external corrosion is concerned. On the other hand, cement lining of the interior of pipes installed in the 1940's and beyond, should have dramatically reduced internal corrosion for these pipes. It must also be pointed out that the apparent corrosion resistance of cast iron often mentioned in the literature, could be attributed to the fact that graphitized cast iron can retain its appearance as a pipe even though much of the iron is gone (Husock, 1982). There are three broad classifications of corrosion types: uniform corrosion, localized corrosion and cracking (Newman, 1979). When uniform corrosion is present, the metal thins uniformly and in most cases the expected life of equipment can be estimated with reasonable accuracy. This type of corrosion usually occurs in copper pipes and thus has no relation with the primary focus of this research, which is on cast-iron pipes. Localized corrosion refers to metal deterioration only at particular locations along the pipe length and this is the type of external corrosion present on cast iron pipes. The three most important types of localized corrosion observed in cast iron pipes are:

a. Pitting

Localized corrosion is generally classified as pitting when the diameter of a cavity at the metal surface is the same, or less, than the depth (Newman, 1979). It can be caused by variations in the metal, such as surface defects, emerging dislocations, or incomplete surface films and coatings. Pits usually grow in the direction of gravity (i.e., downward on horizontal surfaces) and once pitting begins, the environment within a pit starts to change. That is, conditions within a pit become more aggressive and the rate of penetration increases with time.

b. Electrolytic Corrosion

Electrolytic corrosion results from direct current coming from outside sources. The locations along the pipe length where the current is picked up are not usually affected, while the locations where the direct current leaves the pipe to enter the soil become anodic and start to corrode. This type of corrosion is also referred as stray-current and its rate is affected by soil type and the level of soil moisture present in the vicinity of the pipe. The most common cause of this type of corrosion, are stray currents generated by rail transit systems in urban areas. If such sources are identified, appropriate protection measures can be taken. For that reason, it becomes important, as mentioned in other sections, to have information on past maintenance practices followed by the water utility.

c. Galvanic Corrosion

Galvanic corrosion can be present much more frequently in cast-iron pipes than electrolytic corrosion. On the other hand, it does not result in so rapid deterioration as electrolytic corrosion does. There are two predominant causes of galvanic corrosion: dissimilar metals and dissimilar environments. Dissimilar metals can for example be present in water pipes when a cast iron or ductile iron pipe is connected with copper water service lines, or when it is in contact with other underground pipelines of different metals (gas, etc.). Coating of pipes with various materials to protect against corrosion could also result in contact of dissimilar metals and thus accelerate corrosion instead. What basically happens is that when two dissimilar metals are in contact with each other at the presence of a conductive environment, a potential is developed between them and a current flows. The less resistant metal becomes anodic and the more resistant cathodic, with corrosion increasing on the less resistant metal. Dissimilar environments can also cause galvanic corrosion, when a single pipe is in contact with soils of varying composition. For example, if one section of the pipe is in heavy clay soil and another is in well aerated sandy soil, the section in the clay will corrode with respect to the section in the sand. For this reason, corrosion leaks are often found on the bottom of a pipe, even when soils are fairly uniform, because the bottom of the pipe is less accessible to oxygen than the remainder of the pipe (Husock, 1982).

The third category of corrosion induced failure is cracking corrosion, which results from the simultaneous action of corrosion and cyclic or static stresses imposed on the pipe. This mode of failure must be definitely associated with pipe breaks, as the predictive models have also revealed by indicating the significance of variables related to stresses in pipes (e.g., pressure and land development). That is, while corrosion of the previously mentioned types can be present at various degrees among different pipes, the presence of external stresses can accelerate failure by resulting in "corrosion fatigue". And as it is well known, the combined effect of these two factors is much greater than the effect of either one alone.

Given the above discussion on corrosion types and associated failures it becomes clear why a variable such as soil corrosivity can only give us limited information concerning the effect of corrosion on pipe failures. In addition to that, such variable does not reflect at all the degree of internal corrosion. Internal corrosion is also a very important factor in pipe deterioration. Its action will depend on water properties such as pH, dissolved oxygen, carbon dioxide, dissolved metals, temperature and flow velocities (since higher velocities allow oxygen or CO_2 to interact more easily with the surface of the conduit, remove protective films and cause increased corrosion rates). Given the fact that most recent pipes (installed after 1940) are believed to be much more resistant to internal corrosion than older pipes (because of cement lining), the results derived by the predictive models indicating that in some cases (pipes in slow-breaking stage) highly corrosive soil was not related to failures of the older pipes, could be explained by assuming that the most failure prone among the older pipes have much more severe internal deterioration than the most recent pipes and thus the presence of highly corrosive soil in those pipes does not have any significant explanatory power regarding their failure.

Pipe Size

In general, pipe size was not found to be related with the break rate in the New Haven system, while in the Cincinnati system larger diameter pipes were experiencing a faster third break if they had broken once or twice. Although a full explanation for this phenomenon is not possible, there are some plausible reasons why this would have happened. If, as argued earlier, the Cincinnati system is in a much worse overall condition than the New Haven system, then it is plausible that a case similar to that observed in the Manhattan, NY system (O'Day, et al, 1980) may exist. In that study it has been found that while for the broken 6 and 12 inch mains, less than 6% had below standard wall thicknesses, 22% of the 20 inch mains were below standard, and for the 36 inch and 48 inch mains almost all were below standard.

The fact that size did not turn out to be a factor for predicting the probability of a pipe entering a "fast breaking" state in the Cincinnati system, could be explained if we make the assumption that the very large mains are replaced after the third break (that is, before they enter the fast breaking state) because of the high consequences in the water network caused by their failure. Also, even if they are not totally replaced, it might very well be the case that much more careful remedial actions are taken for those pipes once they break than for the smaller ones. Thus, they do not appear in the set of pipes in "fast-breaking" state.

The relation between pipe diameter and stresses in pipes could also be explained by developed unbalanced static pressure at bends. Such pressure can cause a longitudinal force P proportional to the square root of the diameter. That is:

$P = \frac{1}{2} rd^2 \ p\sin \frac{1}{2}a$, where d = pipe diameter, p = internal pressure, and a = angle of bend.

Also, the unbalanced dynamic pressure at bends can cause a force P given by:

$$P = \frac{1}{2} r(d^2/144)(W/g)V^2 \sin \frac{1}{2} a,$$

where W = the weight of unit volume of water and V = the flow velocity.

There has not been reported, however, any investigation in the literature which could confirm the exact impact of those forces on the break rate, and thus additional research and more detailed data will be needed (e.g., information on flow velocity and number of bends along the pipe length), in order to understand more clearly such a relation.

It is also plausible that the higher break rates observed in failing large diameter pipes could be attributed partially to the higher likelihood of different environmental conditions surrounding those pipes exactly because of their larger diameter, which could cause some type of galvanic corrosion action.

The effect of pipe length

As previously noted pipe length was highly varying in the analyzed data sets. In the estimated regression equations length was entered as its natural logarithm. Since the derived models are exponential with respect to the covariates used, the effect of length on the hazard rate will vary as the b power of length, where b would be the estimated coefficient from the regression. In most previous analyses of pipe break records the hazard (or breaking) rate was assumed to vary proportionally with length, although such assumption has not been formally tested. In the models developed by this study the coefficient for length was, on the average, around 0.5, or in other words the hazard rate was approximately proportional to the square root of length.

Although such results might initially appear as paradoxical there can be few explanations about why this is happening. The following two reasons seem to be most plausible for justifying such effect:

a. Length might, up to a certain degree, be a surrogate for land development activities. That is, longer lengths might be associated with sections of the pipe being in more remote areas in the system with less land development. If high land development, as previously argued in this section, is associated with higher concentration of break causing factors, then, according to the above argument, longer pipes would be expected to break less frequently.

b. Break-causing factors could very well be not uniformly distributed along the pipe length. This would especially become more likely for the very long pipes. That is, although certain conditions that would contribute to a higher break rate might apply uniformly along the pipe length (e.g. internal corrosion), other secondary forces could be very localized. Such forces could be caused for example by non-uniformity of the bedding and differential soil subsidence, restraints by bends, branches, valves, etc., thermal and moisture changes in the bedding material, tree root growth pressures, vibrations due to heavy traffic and machinery. Thus, a previous pipe break would indicate a localized concentration of those forces in the vicinity of the break and not necessarily the presence of "high risk" facrors along the whole pipe length. If this is indeed so, then it should not be expected that the break rate in the future would vary proportionally to the length of the pipe.

Both of the above described reasons are certainly very plausible in order to explain the effects of pipe length in the various regression equations estimated by this study. Nevertheless, they basically consist of speculations about possible explanations, since no exact information is available to check their validity. As emphasized throughout this work a better method for coding pipe length is required in the data sets. A pipe segment defined in the data should not only reflect, as much as possible, uniformity of applying external conditions (soil, land-development, etc.) and operating characteristics, but it should also be small enough to allow better concentration on the failure pattern of various pipe sections. In order to achieve those goals, pipe lengths should be as short as possible and not highly varying. For practical reasons of coding the data, lengths in the order of 500 to 1000 ft. seem reasonable. Such change in the way lengths are coded would also contribute to more efficient replacement strategies, since the focus would be on shorter segments for which more precise information would be available.

The fact that pipe lengths were defined in the data without taking into account any of the above arguments, consists of a clear defficiency of those data. Nevertheless, since the derived models help in identifying the "high risk" pipes, more detailed focus could be directed on them later and more information could then become available about the distribution of the break causing factors for those particular pipes. Also since pipe units were defined arbitrarily, length is not an explanatory variable in the sense that, for example, pressure and soil corrosivity are explanatory for breaks. Thus pipe length could be considered as a nuisance or scaling variable in the analysis.

CHAPTER 8

BREAK-TYPE ANALYSIS AND SEASONALITY PATTERNS IN PIPE FAILURES

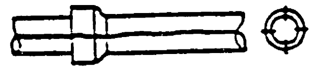
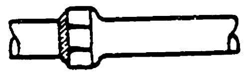
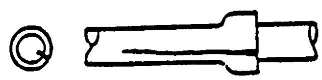
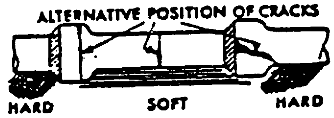
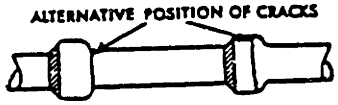
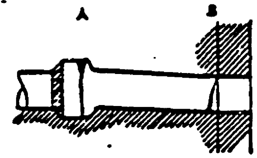
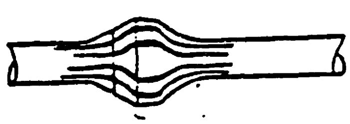
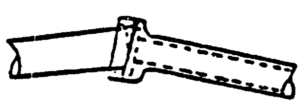
8.1 Introduction

The analysis performed on the New Haven and Cincinnati data did not make any distinction about break-type, because such information has not been available. It is true though that a thorough investigation of break-type and seasonality patterns in water main failures could certainly lead to a better understanding of the breaking mechanism, since it would help to focus our attention on the particular risk factors and stresses associated with each type of failure. Thus, it is expected that when such information is available, the results obtained by applying the methodologies proposed in this work would provide a considerably more accurate picture of the behavior of the deteriorating system. Observations and insights, however, developed by other independent studies on that issue, could also help in improving the interpretation of the results obtained from New Haven and Cincinnati and in better understanding their validity.

There is an almost general consensus in the literature (Walski, 1982) that pipe breaks increase during the winter months. In most instances, there is no clarification as to whether such seasonality pattern was observed to exist for all pipe sizes. This issue is of particular interest in this study, since only large diameter (≥8 inch) pipes were primarily examined, and thus it becomes important to know whether a significant part of the failures was weather induced.

On the other hand, results on the investigation of break-type patterns have been very scarce in the literature. A detailed description of possible break-types with their associated potential causes is shown in Figure 8.1 (after Clarke, 1968). The only extensive statistical studies that, to our knowledge, have investigated the problems of break-types and seasonality patterns are those performed by O'Day et al., for the cities of Philadelphia (1982) and Manhattan (1980), which showed quite similar results. In the Philadelphia study, the following four break categories have been investigated: Circumferential breaks (or otherwise called ring cracks), longitudinal breaks (split pipes), hole breaks and split bell breaks. The findings from this analysis can be summarized as follows (O'Day, 1982):

- Circular breaks almost always account for the largest portion of the breaks by decade.
- Hole breaks are very high in the 1931-1940 mains.
- Split bell breaks are very high in the 1941-1970 mains.
- Circular breaks represent 71 percent of six-inch breaks, dropping to 34 and 31 percent for 12 and 16-inch mains, respectively.

Type of failure	Appearance	Cause	Preventive measures
(*a*) Overload fracture		Excessive vertical load or inadequate bedding	Higher bedding class, or stronger pipe, or concrete surround
(*b*) Burst socket		Differential thermal or moisture expansion of jointing mortar	Resilient jointing material which does not cause excessive radial pressure on the socket
(*c*) Distortion fracture		Differential heating or cooling or moisture content	Protection of uncovered pipes against sun or cold night (or drying wind, with concrete pipes)
(*d*) Beam fractures		Uneven resistance of foundation, or soil movement, or differential settlement	Flexible joints and uniform hardness of foundation
(*e*) Pull fractures		Thermal or drying shrinkage of pipe or site concrete, drying shrinkage of clay soil	Flexible-telescopic joints and gaps in site concrete at pipe joints
(*f*) Shear fractures		Differential settlement of wall relative to pipe or/vice versa	Flexible joints at least at A and B and making AB not more than 3 ft.
(*g*) Bearing fracture		Hard spots in pipe bed	Elimination of hard spots
(*h*) Thrust fracture		Restrained thermal or moisture expansion of pipe or compression due to subsidence	Flexible-telescopic joints which do not cause excessive radial pressure on the sockets. Spigot end not hard up in socket
(*k*) Leverage fracture		Excessive angular displacement	Avoidance of excessive slew when laying

Types of failure *b*, *d*, *e* and *f* may occur with rigid (e.g. cement mortar) joints.
Type *k* may occur with flexible (e.g. rubber ring) joints.
The other failures are uninfluenced by type of joint.

Figure 8.1

TYPES AND CAUSES OF RIGID PIPE FRACTURES
(AFTER CLARKE, 1968)

- Longitudinal breaks represent 21 percent of all breaks, varying from 18 percent for six-inch to 47 percent for 10-inch mains.

- Split bell breaks represent only two percent of 6-inch breaks, 14 and 17 percent of 10 and 12-inch mains respectively.

O'Day (1982) provides the following structural causes for each break type:

Break Type	Structural Causes
Circumferential	Thermal contraction, beam failure
Longitudinal	Excessive ring load
Hole	Internal Pressure
Bell Crack	Thermal expansion and contraction

As it is pointed out in that study, the structural causes of failure only indicate the ultimate break cause. It is clear that the condition of the main (e.g., degree of internal or external corrosion), could have been a significant underlying factor. While it can often be the case that break-type is provided by the data sets, measurements on the progression of corrosion on the pipe wall are very scarce. In the O'Day study there has not been any attempt to examine the statistical correlation between break type and the structural break causing factors, because of lack of adequate data to describe those factors. Only for the split-bell breaks that occurred during the 1941-1950 period, it is strongly indicated that they are related to the leadite joint material used during the same period.

Concerning the seasonality patterns in main breaks, it has been well established in the literature that the break rate increases during the winter months (O'Day, 1980 (Manhattan), 1982 (Philadelphia); Walski, 1982; Buffalo, New York Study, 1981). For example, in the city of Philadelphia, the three months of December, January, and February, accounted for 51 percent of the main breaks during the 1961-1982 period. In the study performed at the City of Buffalo, NY, for the U.S. Army Corps of Engineers, the average break rate during the months of January, February, and March is about three times higher than the break rate during the fall months and about twice as high as the break rate during the summer months. It is also observed by the same study that a high break rate also occurs during the month of July, which correlates with the high water demand during that same period. It could thus be attributed to specific operating characteristics of the system (e.g., increases in internal pressure to meet higher demands). Only in the Manhattan Study (O'Day, 1980) there has been an attempt to investigate the relation between winter breaks and pipe size. The conclusion has been that only pipes smaller or equal to 12 inches in diameter were experiencing increased winter breaks. But no quantitative results are given concerning the exact degree of observed seasonalities in failure patterns for each particular pipe diameter below 12 inches.

As a general observation, it appears that an analysis of the relationship between break-type and seasonality patterns is currently missing from the literature. Additional data obtained from New Haven on break-type, for breaks which occurred after the 1972 period, made such an analysis possible. Unfortunately, no correspondence with the original data set existed, in order to make any direct inferences. Analysis of observed seasonality patterns, without differentiation for break-types has also been performed on both the New Haven and Cincinnati systems.

8.2 Seasonality Pattern in Pipe Failures of the New Haven and Cincinnati Systems

Very few 6 inch pipes (about 1%) existed in both the New Haven and Cincinnati data. Thus, the statistical analysis is primarily focusing on the larger diameter pipes. The results indicate that from a total of about 512 breaks recorded for New Haven since 1900, the distribution of breaks by month is as shown in Table 8.1. The same distribution of breaks by month is also shown for Cincinnati where break records were reported after 1940.

It is observed that a very slightly higher break rate during the winter months exists in the New Haven system, while a higher break rate during the summer months appears to exist in Cincinnati. Thus, it appears that overall no significant increases in breaks during winter months have occurred. The slight increase in winter breaks observed in New Haven is attributed, as further statistical analysis indicated, to the larger percentage of 8 inch pipes that were included in the data of New Haven as compared to those of Cincinnati. Thus, these results

NEW HAVEN SYSTEM

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Percentage of Breaks	11%	8%	6%	6%	8%	8%	9%	8%	10%	7%	9%	9%

SEASON	Percentage of Breaks %
Winter (Dec., Jan., Feb.)	28%
Spring (Mar., Apr., May)	21%
Summer (June, July, Aug.)	25%
Fall (Sept., Oct., Nov.)	26%

CINCINNATI SYSTEM

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept.	Oct	Nov	Dec
Percentage of Breaks	4.6%	7.2%	5.5%	6.5%	9.8%	11.3%	11.7%	8.8%	7.4%	9.1%	8.8%	9.2%

SEASON	Percentage of Breaks %
Winter	21%
Spring	22%
Summer	32%
Fall	25%

Table 8.1

SEASONALITY PATTERNS IN BREAKS

become suggestive of the fact that winter breaks are in most cases associated with the 8 inch and below pipes. The fact that in the great majority of other studies all pipes were pooled together in order to perform the seasonality analysis, can explain why the dampening of seasonal breaks that apparently occurs by larger diameters, did not show up. A possible physical explanation for this phenomenon could be given if we consider the fact that larger diameter pipes are structurally stronger (thicker pipe walls) and thus they can withstand more effectively the stresses caused by frost penetration and/or thermal contraction, during the winter months.

8.3 Break-Type Analysis of the New Haven System

A break-type analysis has been performed on a data set obtained from New Haven, which included information covering the period 1972 - 1984 about break type, break date, location of main, size and date of installation. No correspondence with the original data set existed because the link numbers, as defined in the original data set, were not available and also the 6" pipes had been included in the new data set. Nevertheless, useful results have been obtained concerning the seasonality of breaks corresponding to each break-type. Four major categories of main failures were defined: circumferential breaks (ring-cracks), longitudinal breaks (split pipes), holes in pipes and joint leaks.

A clear seasonal pattern for "ring-cracks" can be observed from

Table 8.2, where a considerably higher break rate occurs during the "colder" months of the year. Such a pattern does not exist for the other three break-types.

Table 8.3 shows the number of breaks associated with each break type and diameter size. We observe that about 56% of the breaks in the 6" and 8" pipes are of the ring-crack type and about 25% are of the split type. "Holes" in pipes are much more frequent in small diameters. "Ring-cracks" and "holes" in pipes were practically absent in 16" diameter and above. Larger diameter pipes (16" and above) basically experience splits and joint leaks. Thus, only split pipe failures and joint leaks can be associated with every pipe size. A split in a pipe can be caused by external loads transferred on the pipe (traffic, frost, soil cover, etc.), improper bedding conditions and improper handling of the pipe during installation. Clearly, such conditions can be present under any pipe size. "Joint leaks" are also observed in any pipe size, since joint materials are similar for every diameter and forces similar to those that can cause a split pipe can also cause excess stresses in joints. The fact that "holes" in the pipe wall were found to be associated only with smaller diameter pipes (12" and below) could be explained by the fact that the pipe wall is also thinner for smaller size pipes. Since corrosion is the clear underlying cause of those holes, we could argue that smaller pipes will be affected faster by corrosion than larger pipes. It has also been observed by other investigators that tuberculous produced from corrosion of the interior

NUMBER OF BREAKS (1972-84)
(6" pipes included)

Break Type \ Month	JAN.	FEB.	MAR.	APR.	MAY	JUNE	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.
RING CRACK	65	13	12	10	6	8	7	8	8	24	28	38
SPLIT PIPE (all types of split)	14	18	11	13	12	12	14	18	19	12	14	9
HOLE IN PIPE	7	5	4	4	3	5	7	7	1	4	8	3
JOINT LEAK	10	10	3	0	9	7	3	7	3	7	4	3

Table 8.2

SEASONALITY PATTERN IN BREAKS BY
BREAK-TYPE, NEW HAVEN

NUMBER OF BREAKS (1972-84)

Type of Break \ Diameter	6"	8"	10"	12"	16"	20"	24"	30"	36"	48"
<u>RING CRACK</u>	95	123	6	9	–	–	–	–	–	–
<u>SPLIT PIPE</u> (all types of split)	41	57	7	33	7	3	1	1	–	1
<u>HOLE IN PIPE</u>	26	14	–	4	–	–	–	–	–	–
<u>JOINT LEAK</u>	8	26	2	14	9	–	1	2	1	1

Table 8.3

BREAK-TYPE BY PIPE SIZE, NEW HAVEN

of the pipe wall are more frequent in smaller diameter pipes, thus indicating a faster progression of corrosion in those pipes. Flow velocity could be one of the reasons for this phenomenon. "Ring-cracks" are associated only with small diameter pipes and are clearly related to weather conditions. It can thus be argued that winter temperatures will basically affect pipes of 12" diameter and below, and will most likely lead to "ring-crack" failure.

Several utility managers have also argued (personal communication) that ring-cracks are associated with soil disturbances caused by previous works for other underground utilities such as sewer contractors It could be the case that those disturbances caused by nearby excavations create favorable conditions (e.g., stresses from pipe bending in the horizontal direction due to unbalanced soil pressures) for a weather induced structural failure.

8.4 Conclusions

The above analysis revealed that both in systems which are under rather severe deterioration (Cincinnati) and in systems which are in rather good overall condition (New Haven), there is practically no increase in pipe failures during the winter months for the larger diameter pipes. Only about half the pipes with diameters 8 inches and below will, according to our results, experience more breaks during the winter months. It is also interesting to notice that in the large diameter pipes (>8 inch) of the Cincinnati system, more frequent breaks are on the average observed during the summer months. Such findings could be related to pressure changes during that period in order to meet higher demand, or by possible thermal expansion of pipes due to higher temperature (for similar findings see also Buffalo study (1981), Section 8.1).

For the first time in the literature, to our knowledge, the break-type analysis performed on the New Haven system revealed that only circumferential breaks have a seasonal pattern with increased breaks during the winter months. Since about half of the 6 and 8 inch pipes were experiencing such breaks, it must be expected that a similar percentage of them would be related to increased breaks during the colder months of the year.

Given the above results, it appears that the majority of pipe failures analyzed for the New Haven and Cincinnati systems did not have any direct strong relation with stresses induced by cold weather conditions, and they were thus related to other modes of failure. In order to examine whether the smaller diameter pipes were experiencing a different failure pattern than the larger ones, the regression model was stratified in two groups. One included pipes with diameters less or equal to 8 inches. The estimated log of minus log survival functions for the average covariate vector are shown in Figure 8.2. No statistically significant difference in failure patterns between the two pipe categories is observed.

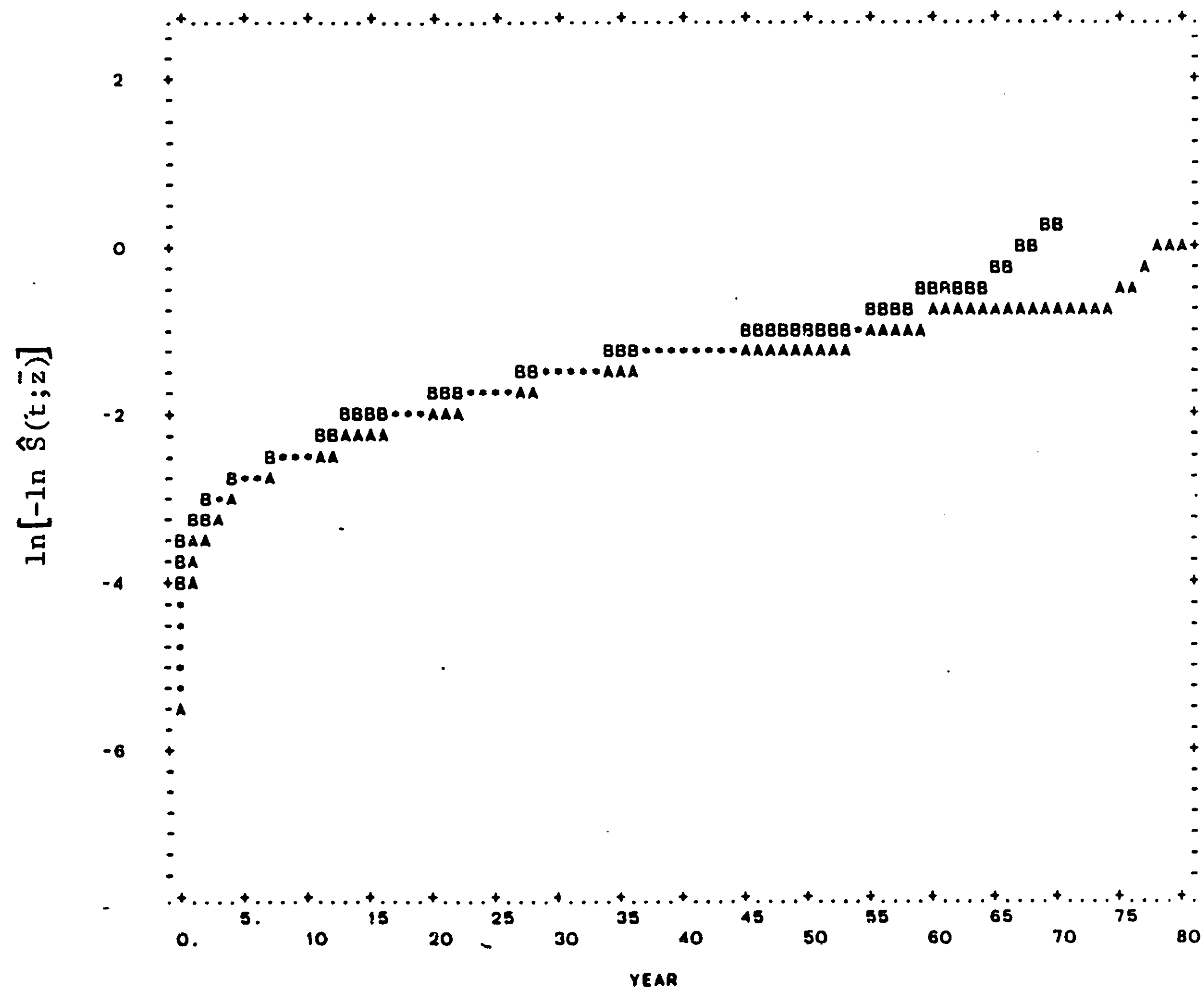


Figure 8.2

LOG OF MINUS LOG SURVIVAL
FUNCTIONS - STRATIFICATION BY DIAMETER
(B: DIAMETER $\leq$ 8 INCHES, A: DIAMETER > 8 INCHES)

The results from the seasonality and break-type analyses could also become suggestive of improvements in future engineering practices related both to pipe materials and appropriate installation methods (bedding materials, trench width, depth of cover, joint materials, etc.). For example, since smaller diameter pipes are basically those that experience ring-cracks, more flexible joints (which are recommended for relieving the kind of stresses associated with ring-cracks) would be highly recommendable for those pipes.

CHAPTER 9

REPAIR, REPLACEMENT AND REHABILITATION DECISIONS

9.1 Introduction

The current state of deterioration of many water distribution systems raises serious questions regarding the optimal repair, replacement and rehabilitation strategies that need to be implemented. The criteria used in making such maintenance decisions are classified in the following two broad categories; economic criteria and reliability of services criteria. The economic criteria involved in the repair/ replacement decisions take into account the costs of break repairs and the costs for replacement of pipe segments. One important objective becomes the minimization of expected future repair and replacement costs of a deteriorating pipeline. This can be achieved by deriving an optimal replacement time for each particular pipe based on predictions about the evolution of its failure pattern. The problem of deciding which pipes in the system to replace and when, based on economic efficiency criteria, becomes more complex when economies of scale can be realized depending on the size of the contracts chosen for replacement. Also many water utilities view as more efficient to perform replacement of pipes in bundles, corresponding to blocks defined by streets in the city. In such case, when pipes in a particular area are replaced at the same time, few of them are replaced at the "optimum" replacement time derived from the analysis at the individual pipe level. The reliability criteria involved in the repair/replacement decision usually include the following considerations: inconvenience to customers caused by interruptions in service due to breaks, potential difficulties for meeting fire flow demands in various parts of the network, potential damages caused by breaks due to flooding of streets, basements, subway stations and disruptions in traffic and other underground pipelines. Economic considerations can be involved here too, because of the potential liability of the utility for the associated damages.

The rehabilitation decision also involves both economic and reliability criteria. The reduction of the internal pipe diameter because of the build-up of tuburculous in the interior pipe wall is associated with increased pumping costs in order to meet the required flow demands and with water quality problems because of the bacteria that often develop within the tuberculous and the "red water" effects. Since rehabilitation of the pipe can be performed at about half the cost of replacement (particularly for the larger diameter pipes), it can be economically more efficient to rehabilitate rather than replace pipes with the above problems. Nevertheless, it has been unclear thus far how rehabilitation affects the future behavior of a pipe regarding its breaking pattern. It could thus be both uneconomical and unreliable to rehabilitate a pipe, which is expected to break quite frequently in the future and would rather be better to replace it in order to avoid incurring both the rehabilitation cost today and the replacement cost at some point soon in the future. Issues related to bundling are also clearly involved in the rehabilitation decision, since it is believed to result in greater efficiences than selective rehabilitation.

It must be pointed out that replacement of a pipe could be solely recommended based on increased demand requirements. It will also be the case that the water utility will face a limited budget each year for replacement and rehabilitation programs. Projecting future repair costs under a given maintenance strategy will also be important for budgeting purposes. It thus becomes apparent that the repair, replacement and rehabilitation decisions represent a very complex multicriteria and multiobjective problem.

The question examined in this chapter is how the derived models and the methodologies proposed in this work can be integrated in the decision making process. It will be shown that the detailed focus on the various phases of deterioration that a pipe can go through provided by the models of this study, has clear advantages against models and "rules of thumb" currently applied. A comparison between the current state of the art predictive models for pipe failures described in Chapter 2 and the models proposed in Chapters 5 and 6 of this study indicate that repair, replacement and rehabilitation decisions can be very different depending on which technique is applied.

Section 9.2 describes the application of the derived models on the economic evaluation for determining an optimal replacement time for individual pipes. Appropriate quantitative techniques applied respectively for pipes in the slow and fast breaking state are derived and implemented on real cases. Section 9.3 describes similar quantitative approaches that are appropriate for making the rehabilitation decision. Section 9.4 discusses in detail why currently applied models and "rules of thumb" can lead to very uneconomical replacement and rehabilitation decisions and why they can also result in future levels of system reliability that are not indeed desirable. Comparisons with results obtained by the proposed techniques are presented and discussed. Section 9.5 presents the application of the models for making better bundling decisions in contracting for replacement and rehabilitation and demonstrates how the obtained probabilities of failure can be used for identifying high risk pipes and high risk areas in the network.

9.2 The repair versus replacement decision

When a pipe starts experiencing breaks the question becomes whether it is more economical to let it break and incur the repair cost or replace it at some point in time and thus make the chances for breaking in the near future practically negligible. The replacement action should not necessarily involve the whole pipe length, but sections of the pipe, depending on the available information about the distribution of the break causing factors and the structural integrity of the pipe. The cost of repairing a break is usually small by itself if compared to the replacement costs. Thus, if only these types of costs are involved in an economic analysis, it must be expected that only the severely deteriorated pipes with frequent breaks will become candidates for replacement within a reasonable (10-15 years) time horizon. But since there would also be "social" costs associated with breaks and a potential liability for the water utility, replacement of a pipe could very well in many cases be based solely on reliability criteria. No matter though what the chosen criterion or criteria might be, the central piece of information needed for making such decisions is the expected evolution of failures associated with each individual pipe.

The following analysis is based on the results obtained from the New Haven and Cincinnati Systems, where primarily larger diameter pipes ($\geq$ 8 inches) were involved. It would thus be applicable only to such larger diameter pipes while possible extensions to smaller diameter pipes must be an issue of further research.

The analysis is based on the idea that two distinct stages can exist in a pipe failure pattern. The first stage is characterized by few infrequent breaks and according to the results presented in Chapters 5 and 6 it will usually last no further than the time that the pipe experiences its third break. The second stage is characterized by multiple and frequent breaks and it is indicative of severe deterioration of the pipe and/or very high concentration of break causing factors. For reasons of simplicity the first stage will be called "slow-breaking" stage, while the second stage will be called "fast-breaking" stage. Depending on the overall condition of a given system the break rates of pipes that are in either one of the two stages are expected to differ, as it has been observed in the case of New Haven and Cincinnati. If the general condition of the system in terms of breaks can be considered good (e.g. New Haven), then very few pipes will be observed in the fast-breaking stage. The contrary will happen in a system which, relatively speaking, is not in good overall condition (e.g. Cincinnati), where many pipes will be in a fast-breaking stage and, for those that are not, a quite high finite probability might exist for entering that stage.

The characterization of a system as being in a bad or good overall condition is basically subjective and an in depth analysis of available data will always be needed in order to make accurate inferences. As a first-order approximation though, the number of pipes in fast-breaking stage could be considered as a guideline. It can be argued that this number clearly shows for example the big difference in the level of deterioration between the New Haven and Cincinnati systems. Only 2.7% of the pipes in New Haven had more than two breaks, while the same was true for 21% of the pipes in Cincinnati (three previous breaks is basically the milestone for entering the fast breaking stage). As the detailed regression analyses revealed, this high difference in the percentage of pipes with more than three breaks, reflected in much higher on average probabilities of failure for the Cincinnati system. (Approximately three times higher than in New Haven, for the average pipe).

By treating New Haven and Cincinnati as two opposing cases in terms of degree of deterioration, the following analysis is developed for estimating the evolution of breaks with time for individual pipes:

a. Systems in "good" overall condition - Pipes in "slow-breaking" stage

If a system is in good overall condition and a pipe in that system is in the slow-breaking stage (implying that it has experienced zero, one or two previous breaks) it is safe to assume that the likelihood of two breaks occurring the same year is negligible.

Thus, no more than one break will be expected each year and the problem becomes to determine the probability for that break occurrence. As it can be easily seen from the model developed for New Haven, which applies to this case, the probability of failure varies with time and also is a function of the number of previous breaks. Thus, break events can be characterized at this stage as a non-homogeneous Markov process. If the number of previous breaks determines the states that a pipe can go through then the transition probabilities $P_{ij}^{(m)}(n)$ are given by the equation:

$$P_{ij}^{(m)}(n) = \sum_{k=1}^{r} P_{ik}^{(m-1)}(n)P_{kj}(n+m) \tag{9.1}$$

which is interpreted as follows: the probability that a pipe will be in state j at time n+m given that it was in state i at time n is the sum over all states, k=1,2,...r of the probabilities that the pipe goes to state k in m-1 steps and then passes to state j on the m^{th} step. The marginal probabilities of being at state j at time n, $q_j(n)$ are given by:

$$q_j(n) = \sum_{i=1}^{r} q_i(0)P_{ij}^{(m)}(0) \tag{9.2}$$

where $q_i(0)$ are the initial probabilities at time n=0.

The following notation is then used for calculating the expected number of breaks each year (or equivalently the probability of one break):

E_i = expected number of breaks in year i

$P_{ik\ell...}$ = probability of having i breaks in year 1, k breaks in year 2, ℓ breaks in year 3, etc., where i,k,ℓ will be either 0 or 1.

$P_{m/ik\ell...n}$ = probability of having m breaks in year n conditional on i breaks in year 1, k breaks in year 2, ...n breaks in year n-1, where m,i,k,ℓ,...n equal to 0 or 1.

$S(t)$ = survivor function at time t $\equiv S_T(t;z)$, where $S_T(t;z)$ is the survivor function defined in Chapter 3 for a pipe with covariate vector z. S(t) is the survivor function conditional on no intermediate breaks between the reference time and time t.

$S_i^j(t)$ = survivor function at time t, conditional on j additional breaks having occurred since the reference time and with the second break in year i.

According to the derived model (Chapter 5) the function $S_i^j(t)$ will vary depending on the values of i and j.

If the reference time chosen is zero, usually corresponding to the current year then the calculation for determining the expected number of breaks in future years can proceed as follows:

$$E_1 = P_1 \tag{9.3}$$

$$E_2 = P_{1/0} \cdot P_0 + P_{1/1} \cdot P_1 \tag{9.4}$$

$$E_3 = P_{1/00} \cdot P_{00} + P_{1/10} \cdot P_{10} + P_{1/11} \cdot P_{11} + P_{1/01} \cdot P_{01} \tag{9.5}$$

$$E_4 = P_{1/000} \cdot P_{000} + P_{1/100} \cdot P_{100} + P_{1/010} \cdot P_{010} + P_{1/001} \cdot P_{001}$$

$$+ P_{1/011} \cdot P_{011} + P_{1/101} \cdot P_{101} + P_{1/110} \cdot P_{110} \tag{9.6}$$

where the probability of having four breaks in four consecutive years was assumed negligible, given that pipes are in a slow-breaking stage and the system is in good overall condition.

It is observed that further simplifications could be made in equations (9.3)-(9.6) if probabilities of having two or three consecutive breaks are also assumed negligible. Depending on the overall condition of the system and the observed failure patterns, such assumptions could be very reasonable.

Substituting the expressions for the failure probabilities in equations (9.3) to (9.6) we have:

$$P_1 = 1 - \frac{S(t+1)}{S(t)} \tag{9.7}$$

$$P_o = \frac{S(t+1)}{S(t)} \tag{9.8}$$

$$P_{1/0} = \frac{S(t+1)-S(t+2)}{S(t+1)} \tag{9.9}$$

$$P_{1/1} = 1 - \frac{S_1^1(1)}{S(0)} = 1 - S_1^1(1) \tag{9.10}$$

$$P_{1/00} = \frac{S(t+2)-S(t+3)}{S(t+2)} \tag{9.11}$$

$$P_{00} = \frac{S(t+2)}{S(t)} \tag{9.12}$$

$$P_{1/10} = 1 - S_2^1(1) \tag{9.13}$$

$$P_{1/11} = 1 - S_2^2(1) \tag{9.14}$$

$$P_{1/01} = \frac{S_1^1(1)-S_1^1(2)}{S_1^1(1)} \tag{9.15}$$

$$P_{01} = \left[\frac{S(t)-S(t+1)}{S(t)}\right] \cdot S_1^1(1) \tag{9.16}$$

$$P_{1/000} = \frac{S(t+3)-S(t+4)}{S(t+3)} \tag{9.17}$$

$$P_{000} = \frac{S(t+3)}{S(t)} \tag{9.18}$$

$$P_{1/100} = 1 - S_3^1(1) \tag{9.19}$$

$$P_{1/010} = \frac{S_2^1(1)-S_2^1(2)}{S_2^1(1)} \tag{9.20}$$

$$P_{010} = \frac{S(t+1)}{S(t)} \cdot \left[\frac{S(t+1)-S(t+2)}{S(t+1)}\right] \cdot S_2^1(1)$$

$$= \left[\frac{S(t+1)-S(t+2)}{S(t)}\right] \cdot S_2^1(1) \tag{9.21}$$

$$P_{1/001} = 1 - \frac{S_1^1(3)}{S_1^1(2)} \tag{9.22}$$

$$P_{001} = \left[\frac{S(t)-S(t+1)}{S(t)}\right] \cdot S_1^1(2) \tag{9.23}$$

$$P_{1/011} = 1 - \frac{S_2^2(2)}{S_2^2(1)} \tag{9.24}$$

$$P_{011} = \left[\frac{S(t)-S(t+1)}{S(t)}\right] \cdot \left[1 - S_1^1(1)\right] \cdot S_2^2(1) \tag{9.25}$$

$$P_{1/101} = 1 - S_3^2(1) \tag{9.26}$$

$$P_{101} = \left[\frac{S(t)-S(t+1)}{S(t)}\right] \cdot S_1^1(1) \cdot \left[\frac{S_1^1(1)-S_1^1(2)}{S_1^1(1)}\right] \tag{9.27}$$

$$P_{1/110} = 1-S_3^2(1) \tag{9.28}$$

$$P_{110} = \frac{S(t+1)}{S(t)} \cdot \left[\frac{S(t+1)-S(t+2)}{S(t+1)}\right] \cdot \left[1-S_2^1(1)\right] \tag{9.29}$$

Thus the calculation of the expected number of breaks in years $i=1,2,\ldots$ can proceed as follows by substituting in (9.3)-(9.6) the expressions in equations (9.7)-(9.29):

$$E_1 = \frac{S(t)-S(t+1)}{S(t)} \tag{9.30}$$

$$E_2 = 1 - \left[\frac{S(t+2)+S_1^1(1)S(t)-S_1^1(1)S(t+1)}{S(t)}\right] \tag{9.31}$$

$$E_3 = 1 - \left[\frac{S(t+3)+S(t)\left[S_2^2(1)-S_2^2(1)S_1^1(1)+S_1^1(2)\right]}{S(t)} - \underline{S(t+1)\left[S_2^2(1)-S_2^1(1)+S_1^1(2)-S_2^2(1)S_1^1(1)\right]} - \underline{S(t+2)S_2^1(1)} \right] \tag{9.32}$$

$$\begin{aligned} E_4 = \; & S_2^2(1)-S_1^1(3)-S_2^2(1)S_1^1(1)-S_2^2(2)+S_2^2(2)S_1^1(1) \\ & - S_3^2(1)S_1^1(1)+S_3^2(1)S_1^1(2)+S_1^1(1) \\ & + S(t+1)\left[\frac{-S_2^1(2)-S_2^2(1)+S_2^2(1)S_1^1(1)+S_2^2(2)}{S(t)} \right. \\ & - \underline{S_2^2(2)S_1^1(1)-S_1^1(1)+S_3^2(1)S_1^1(1)-S_3^2(1)S_1^1(2)} \\ & \left. + \underline{1-(S_3^2(1)+S_3^2(1)S_2^1(1)+S_1^1(3)} \right] \\ & + \frac{S(t+2)}{S(t)} \left[S_2^1(2)-S_3^1(1)+S_3^2(1)-S_3^2(1)S_2^1(1) \right] \end{aligned} \tag{9.33}$$

By applying equations (9.30)-(9.33) a growth rate for the expected number of breaks can be established for each individual pipe in the system. The above described procedure can provide the growth rate in breaks with much greater accuracy than previously developed techniques since it uses the survivor functions estimated by the models presented in Chapters 5 and 6 which represent in great detail the failure process in deteriorating pipes.

If an average growth rate in the evolution of breaks is calculated then the expected number of breaks in year t will be given by:

$$E_t = E_o (1+g)^t \tag{9.34}$$

where E_t = expected number of breaks in year t

g = average break growth rate

The above model considers that, for pipes in the slow-breaking stage, it is reasonable to assume a constant growth of breaks with time for short time horizons (5-10 years). This growth rate will be highly varying though among the individual pipes of a given system depending on their previous maintenance history and the operating and environmental characteristics. By observing the break trends for pipes in the slow-breaking stage both in the New Haven and Cincinnati systems, it appears that the above assumption is a reasonable approximation. If we denote by C_b the cost of repairing a break and by C_r the cost of replacing a pipe (or the cost of appropriate remedial action for eliminating the possibility of breaks in the near future, e.g. replacement of only the parts of the pipe that are considered "high risk," given the available information on the location of previous breaks), then the present value of the expected costs, if replacement takes place in year t, will be given by:

$$PV \begin{bmatrix} \text{expected} \\ \text{repair and} \\ \text{replacement costs} \end{bmatrix} = E_o C_b \sum_{t=0}^{n} \left(\frac{1+g}{1+r} \right)^t + \frac{C_r}{(1+r)^t} \tag{9.35}$$

where: E_0 = expected number of breaks in year t=0, or equivalently, for pipes in slow-breaking stage, the probability of break in the reference year t=0

C_b = cost of repairing a break (\$)

C_r = cost of pipe replacement (\$)

g = break growth rate, 1/years

r = real interest rate, assumed constant over time.

By taking the derivative of equation (9.35) with respect to time t and setting it equal to zero we obtain an optimal replacement time t*, given by:

$$t^* = \ln\left[\frac{C_r}{C_b \cdot E_o} \ln\ (1+r)\right] / \ln(1+g) \tag{9.36}$$

where it was assumed that for small values of r, $(1+r)^t \approx e^{rt}$.

It must be pointed out that an optimal replacement time derived from equation (9.36) is only based on economic costs for repair and replacement. That is, no other criteria based on reliability or bundling considerations were included. Thus, in reality such measure should be only one of the factors that will result in the final decision of when to replace a pipe. Particularly for pipes in the slow-breaking stage, for which formula (9.36) will apply, the calculated time t* is expected to be quite far in the future for the majority of pipes, since E_0 and g will be very small. The value of t* will also critically depend on the value of C_r, which could highly vary depending on the cost of appropriate remedial action for restoring the structural integrity of the pipe.

Thus for systems in a good overall condition with the high majority of pipes in the slow-breaking stage, replacement decisions for most of the pipes should be expected to be based rather on reliability and bundling criteria as opposed to economic costs of repairs versus replacement. And for such criteria the probability of failure and its evolution with time, obtained by the derived models, provides a central piece of information.

In order to demonstrate the evolution of expected breaks in the New Haven system equations (9.30)-(9.33) were applied by substituting the values of the estimated survivor functions for the individual pipes. In the few cases where the required survivor function was not estimated by the model because of the small number of observations beyond the third break, the probability of survival was estimated just by the frequency of pipes that broke while being at that same stage.

Table 9.1 shows the expected breaks calculated for the next four years, starting from year 1984 as the reference year. The pipes are classified by installation period, where DATE indicates the year of installation (0→1900). For each year of installation the total number of expected breaks for the pipes installed during that year is shown

	E1	E2	E3	E4	LENGTH
	SUM	SUM	SUM	SUM	SUM
DATE					
0	1.12	1.20	1.27	1.31	38300.00
1	0.05	0.06	0.07	0.07	4700.00
6	0.19	0.19	0.19	0.19	8950.00
8	0.44	0.47	0.50	0.53	40150.00
10	0.09	0.10	0.11	0.12	3700.00
12	0.11	0.12	0.13	0.13	10700.00
13	0.08	0.08	0.08	0.08	4350.00
14	0.09	0.09	0.09	0.10	9700.00
16	0.19	0.22	0.24	0.26	7100.00
17	0.03	0.02	0.02	0.02	1250.00
18	0.07	0.07	0.07	0.08	3350.00
19	0.02	0.02	0.02	0.01	500.00
21	0.41	0.45	0.49	0.53	26450.00
22	0.07	0.08	0.09	0 10	4750.00
24	0.01	0.01	0.01	0.01	1500.00
25	0.28	0.31	0.35	0.38	20200.00
26	0.11	0.12	0.14	0.15	9950.00
28	0.33	0.36	0.39	0.42	20150.00
29	0.03	0.04	0.05	0.05	3250.00
30	2.54	2.78	3.02	3.24	492950.00
31	0.87	0.95	1.02	1.10	119650.00
32	0.33	0.36	0.39	0.43	60850.00
33	0.58	0.64	0.70	0.75	100300.00
34	0.35	0.38	0.41	0.44	97050.00

(CONTINUED)

Table 9.1

EXPECTED NUMBER OF BREAKS CLASSIFIED BY
PIPE INSTALLATION PERIODS, NEW HAVEN

	E1	E2	E3	E4	LENGTH
	SUM	SUM	SUM	SUM	SUM
DATE					
35	0.47	0.51	0.65	0.59	137750.00
36	0.02	0.02	0.02	0.02	4250.00
37	0.00	0.00	0.00	0.00	400.00
38	0.07	0.07	0.08	0.08	8250.00
39	0.06	0.08	0.09	0.10	11200.00
40	1.34	1.62	1.70	1.86	240550.00
41	0.10	0.12	0.14	0.15	10350.00
42	0.02	0.03	0.03	0.03	3550.00
44	0.03	0.03	0.04	0.04	5000.00
46	0.03	0.03	0.03	0.03	4600.00
47	0.02	0.02	0.03	0.03	6850.00
48	0.00	0.01	0.01	0.01	750.00
49	0.03	0.03	0.03	0.03	5100.00
50	0.46	0.52	0.57	0.63	106950.00
51	0.01	0.01	0.01	0.01	1000.00
52	0.06	0.06	0.07	0.08	18500.00
53	0.05	0.05	0.05	0.06	12000.00
54	0.03	0.03	0.03	0.03	4400.00
55	0.19	0.20	0.21	0.23	43950.00
56	0.17	0.20	0.22	0.23	36200.00
57	0.30	0.32	0.34	0.36	54850.00
58	0.20	0.21	0.23	0.24	43750.00
59	0.08	0.08	0.08	0.09	13800.00
60	0.28	0.31	0.34	0.36	72000.00

(CONTINUED)

	E1	E2	E3	E4	LENGTH
	SUM	SUM	SUM	SUM	SUM
DATE					
61	0.03	0.03	0.03	0.03	10400.00
62	0.08	0.08	0.08	0.08	8500.00
63	0.44	0.46	0.48	0.49	62650.00
64	0.22	0.24	0.25	0.26	79900.00
65	0.47	0.49	0.52	0.54	101850.00
66	0.06	0.06	0.07	0.07	8450.00
67	0.13	0.13	0.13	0.14	14100.00
68	0.12	0.13	0.13	0.14	22400.00
69	0.06	0.06	0.05	0.05	12250.00
70	0.50	0.54	0.57	0.59	102200.00
71	0.55	0.57	0.60	0.60	43600.00
72	0.26	0.29	0.31	0.32	43050.00
73	0.10	0.10	0.10	0.11	20250.00
74	0.03	0.03	0.03	0.03	6700.00
75	0.03	0.03	0.04	0.04	3850.00
76	0.09	0.09	0.09	0.09	5100.00
77	0.09	0.09	0.09	0.09	10600.00
78	0.03	0.03	0.03	0.03	6000.00
79	0.12	0.13	0.13	0.14	9650.00
80	0.04	0.04	0.04	0.04	4800.00
81	0.11	0.12	0.12	0.12	10400.00

under columns E1,E2,E3 and E4 corresponding to years 1984, 1985, 1986, 1987 respectively. Also the total length (in ft.) corresponding to each year of installation is shown under the column denoted by LENGTH. Given the fact that those breaks would represent rather major failures, since they correspond to the larger diameter pipes (99% of the pipes have diameters greater than or equal to 8 inches), the results presented in Table 9.1 are very useful for assessing the overall condition of the system and particularly obtain quantitative measures of how the pipes, classified by period of installation, are expected to perform.

Since the severity of the consequences of a break can be directly related to the size of the pipe, Table 9.2 shows the cumulative number of expected breaks for years 1984-1987, for each pipe diameter (diameter is denoted by the variable DIA and varies from 6 to 48 inches. Table 9.2 is very useful for assessing the overall reliability of the system.

From Tables 9.1 and 9.2 we can obtain an average value for the break growth rate in the whole system, which is equal to 7.3% per year. The total number of expected breaks in the reference year is equal to 15.98. If we assume that such average growth rate is indeed realized for the next 10 years then the total number of expected breaks per year is estimated to double by that time. The implications that various replacement scenaria can have on the total number of expected breaks are examined in Section 9.3.

It must be pointed out that the results demonstrated in Tables 9.1 and 9.2 correspond to aggregate measures for the overall system.

	E1	E2	E3	E4	LENGTH
	SUM	SUM	SUM	SUM	SUM
DIA					
6	0.23	0.25	0.27	0.28	19800.00
8	2.79	3.08	3.36	3.58	486500.00
10	2.03	2.22	2.41	2.57	260550.00
12	5.65	6.12	6.59	6.99	879575.00
16	2.59	2.80	3.00	3.18	424500.00
20	0.79	0.84	0.89	0.94	162150.00
24	0.88	0.94	0.99	1.04	104000.00
30	0.39	0.41	0.44	0.47	60750.00
36	0.34	0.37	0.41	0.43	55925.00
42	0.04	0.04	0.05	0.05	6250.00
48	0.25	0.25	0.26	0.27	62500.00

Table 9.2

EXPECTED NUMBER OF BREAKS CLASSIFIED
BY PIPE DIAMETER

For the repair vs. replacement decision at the individual pipe level the growth rate will be estimated from equations (9.30)-(9.33) for each pipe separately. This growth rate is expected to be highly varying among different pipes, as shows in Chapter 7 and the usefulness of the derived models lies very much on the fact that they can capture this high variability as opposed to previously developed techniques.

b. Systems in relatively severe deterioration - Pipes in slow-breaking stage

When a system appears to be under rather severe deterioration, as the Cincinnati system happened to be, two modes of pipe failure are expected to exist (Chapter 6). The first mode will be described by pipes in the slow-breaking stage (i.e. pipes with 0,1 or 2 previous breaks) and the second mode by pipes in the fast-breaking stage. Although pipes in the slow-breaking stage will have in general higher probabilities of failure than the pipes of a system in good overall condition (as was found to hold for the New Haven and Cincinnati system) the basic formulas for calculating the expected number of future breaks, given that they are in that stage, will again by given by equations (9.30)-(9.33). Of course, the estimated survivor functions will depend on the models derived for the particular system. The fact though that the system is not in a good overall condition would imply that the probability for entering into the fast-breaking stage at any given time period is not negligible any more but has a finite value calculated by a model such that developed for the Cincinnati system and described in Table 6.6 (Chapter 6). If this model is then used in the analysis, it will provide the probability P_{f_t} for entering into the fast-breaking stage at time t. The expected number of breaks

E_{b_t} in year t will then be given by:

$$E_{b_t} = P_{f_t} \cdot b_{f_t} + (1-P_{f_t})E_t \tag{9.37}$$

where E_t = expected number of breaks in year t if pipe is in slow-breaking stage (equations (9.30)-(9.33)).

P_{f_t} = probability for entering into the fast-breaking stage in year t.

b_{f_t} = break rate (breaks/year) if pipe enters into the fast-breaking stage.

According to the models derived for the city of Cincinnati the break-rate at the fast-breaking stage can be assumed to remain constant with time. Depending on the way the fast-breaking stage has been defined in the derived models the break rate b_{f_t} can be approximated as follows: a. Set roughly equal to the break-rate at which a pipe was considered of being in fast-breaking stage (e.g. set equal to 0.5 breaks/year, if the Cincinnati model is used). b. Set equal to a a fast-breaking rate estimated through a regression model derived for pipes that already are in that stage (e.g. models similar to those derived for Cincinnati to predict the break rate beyond the third break). In the latter case, a more accurate estimation would be possible, since the particular pipe characteristics will be explicitly taken into account.

For pipes in this category (i.e. pipes in slow-breaking stage but within a system undergoing severe deterioration) the estimated optimal replacement time based on break repair and replacement costs will again be calculated by the same procedure as in the case of pipes belonging to a good overall system. The only difference would be that the break-growth rate and/or the yearly probability of failure will now be much higher. Thus, the estimated optimal replacement times are expected to be shorter. It must also be pointed out that reliability considerations would become more important here than in systems in good condition. Thus, the calculated by the model failure probabilities are expected to be very useful quantitative measures for making reliability assessments.

c. Systems in relatively severe deterioration - Pipes in fast-breaking stage

The analysis of the Cincinnati data set revealed that when longer diameter pipes ($\geq$ 8 inches) enter into the fast-breaking stage, their break-rate does not show any trend of increasing or decreasing with time and can thus be approximated with reasonable accuracy by a constant rate. The models derived from Cincinnati for pipes that were in that state indicate that this break-rate can be highly varying depending on the particular operating and environmental characteristics of each pipe. After this break-rate is estimated for each pipe by the models described in Chapter 6, the assumption can be made that breaks are Poisson arrivals at that rate. Thus, the expected number of breaks each year will be equal to the estimated yearly break-rate and also equal to the variance of the yearly number of breaks.

For pipes at this stage both economic and reliability criteria are expected to become important factors in the repair versus replacement decision. The estimation of an optimal replacement time though based on repair and replacement costs is significantly simplified. Since the expected number of breaks will remain constant each year then it is easily shown that replacement would always be more economical if and only if:

$$C_r < \frac{E_o C_b}{r} \tag{9.38}$$

where: C_r = cost of replacement (in \$)

C_b = cost of repair (in \$), per break

E_o = expected number of breaks per year, or equivalently the estimated yearly break rate R

r = real interest rate.

The right hand side of the inequality (9.38) simply represents a perpetuity equal to $E_o C_b$. Of course it must be expected that for the majority of pipes in fast-breaking stage reliability considerations will have a great weight in the decision for replacement besides the economic costs.

The discussion presented in this section involved the use of repair and replacement costs for water mains. Such costs are not though the same for each pipe and they vary primarily with pipe diameter and break-type. Thus, when any of the derived formulas for economic analysis is applied the appropriate costs for each particular case must be substituted.

Shamir and Howard (1979) assumed a repair cost per break equal to $1,000, with a range of $500 to $2,000. No detailed discussion is though made of how they arrived at those figures. In the Manhattan, NY study (O'Day, 1980) the direct costs to the Water Supply Bureau were assumed equal to $7,323 (including 11 man-days of Water Bureau staff time per break). According to that study the average damage settlement was roughly $1,000 per break (1980 dollars). Stafford et al. (1981) reported that the average cost for break repairs ranged from $1,170 to $1,760 per break for the Cincinnati (Ohio) Water Works (in 1975 dollars). Walski and Pellicia (1981) developed synthetic cost function for breaks by calculating the costs of individual items involved in the break repair (labor and material costs). They included crew costs, equipment and sleeve costs, repaving and overhead costs. Table 9.3 shows the estimated total costs by pipe diameter for the cities of Binghamton, NY and Buffalo, NY respectively. It is argued in that study that the Buffalo costs are believed to be more representative of typical repair costs in an urban area.

Breaks are also a function of break-type, which apparently has not been taken into account in the previously mentioned studies. This is so because circumferential breaks can be repaired with a clamp while split bell and longitudinal breaks require that part of the pipe be cut out and replaced. In the Water Supply Infrastructure Study for the City of Philadelphia (King 1984a) these costs were estimated from

Pipe Diameter in.	Binghampton Costs $	Buffalo Costs $
4	718	1,455
6	786	1,558
8	839	1,679
10	896	1,780
12	920	1,872
16	1,266	2,315
18	1,305	--
20	1,415	2,434
24	1,770	2,755
30	---	3,289
36	--	3,485
48	--	4,107

Table 9.3

PIPE BREAK REPAIR COSTS (1983 DOLLARS)

416 breaks that occurred during the period 1975 to 1981. The results are shown in Table 9.4.

Replacement of pipes with new ones in urban areas (also called "relaying" a pipe) represents significantly higher costs than laying new pipes for the first time in undeveloped areas. Pipe replacement costs involved all costs paid to the project contractors. They include (Walski, 1975) the following items: excavation, abandoning existing pipe, laying new pipe, reconnecting services, pressure testing, disinfection, backfilling, repaving, contractor's overhead and profit. Typical replacement costs for the City of Philadelphia are provided by Walski (1985) (Table 9.5). Cost information was available as actual price per foot of pipe averaged over all projects in a single year and not on a project by project basis. In cases where both the water main and the sewer were replaced the allocated water main costs are only shown. Cost information has also been provided only for pipes 8 and 12 inches in diameter.

Costs of pipe relaying in New York City are provided by O'Day [1982]. They include the following items: protection and maintenance of traffic, removal of pavement, excavation, sheating and skoring, removal of existing main, dewatering maintenance and protection of existing structures, furnishing and replacing new main, backfilling using material from excavation removal and replacement of hydrants and valves, and temporary and permanent restoration of pavements (Table 9.6). Pipes smaller than 12 inches in diameter were not included because it is

Pipe Diameter in.	Cost for Indicated Type of Break, $		
	Circumferential	Split Bell	Longitudinal
6	930	975	1,058
8	895	1,202	1,053
10	1,149	1,380	1,611
12	1,362	1,087	2,516
16-48	2,237	3,904	5,620

Table 9.4

PIPE BREAK REPAIR COSTS BY TYPE OF BREAK
(1983 DOLLARS)

Pipe Diameter in.	Type of Project	Mean 1982 $/ft	Standard Deviation 1982 $/ft	Number of Years	Notes
8	Relay	96.6	11.8	10	
12	Relay	115.5	32.4	10	
8	R/R*	114.4	21.4	10	
12	R/R	149.0	38.3	9	
8	Relay	91.0	8.4	6	w/streets contract
8	R/R	65.9	20.2	5	w/streets contract
18	Reconstruction	152.4	25.6	6	sewer only
18	R/R	150.1	23.3	6	sewer only

*R/R = relay/reconstruction (water main portion of water main relay and sewer reconstruction project).

Table 9.5

COSTS FOR PIPE RELAY AND RELAY/RECONSTRUCTION

Pipe Diameter in.	Cost for Indicated Pipe, $/ft		
	Ductile Iron	Reinforced Concrete	Steel
12	106	---	342
20	132	---	---
24	144	---	---
30	354	319	342
36	472	461	496
42	589	561	608
48	685	662	721

Table 9.6

COSTS FOR PIPE RELAYING IN
NEW YORK CITY (1982 $)

Pipe Diameter in.	Cost $/ft
4	68
6	72
8	77
10	84
12	89
16	115
20	139
24	166
30	205
36	262
48	384

Table 9.7

COST FOR PIPE RELAYING IN BUFFALO, NY
(1982 $)

the current policy of New York not to install new mains with diameters of less than 12 inches.

Costs of pipe replacement were also estimated from the City of Buffalo, NY (U.S. Army Corps of Engineers Study), and they are shown in Table 9.7. Although no accurate explanation can be provided for the differences between the New York and Buffalo estimates they could be attributed to lower labor and excavation costs in Buffalo.

It must be pointed out that replacement costs will also vary depending on the size of the contracts. This issue is discussed in greater detail in Section 9.5 where the bundling decision is also considered. It might also be more economical if replacement of pipes is done "in house" by the water utility rather than using outside contractors (e.g. city of New Haven, CT). No accurate data though are available in order to support this argument.

Since many of the replacement decisions, particularly for the severely deteriorated pipes, will involve a high weight on reliability considerations, the following categories of failure impact can be identified in addition to the break repair costs:

a. Service disruption

Disruption of service will be of varying importance depending on the type of land use covering the pipe (residential, industrial, commercial or mixture of the above). If a failure occurs in a distribution main of great importance in the hydraulic network, then the disruption in service may extend well beyond the local point of failure. For this reason results obtained by the derived models are particularly useful, since after such pipes with high probability of failure are identified in the network, hydraulic analysis could be performed by assuming some of them to be out of service. This way, the impact of a break in remote areas outside the break location could be evaluated. It is well known a priori that water distribution systems include many redundancies in order to avoid extensive impacts in other parts of the network from systems failures. Such redundancies are also a function of the number of water supplies available and they certainly decrease as the size of pipes becomes larger. Since the focus of this study has been on larger diameter pipes, it becomes clear why probabilities of failure are particularly useful for such pipes.

b. Urban disruption

Urban disruption includes impacts on various activities. Water main failures could result in traffic disruptions because of street flooding, street damage and excavation within the street during repair. The level of impact will depend on traffic volume and also the expected duration of traffic disruption. Table 9.8 shows the estimated time to repair breaks by pipe diameter and break-type (King, 1984).

Other types of urban disruption may include flooding of basements and subway stations. Other underground utilities, particularly power and natural gas distribution, could also be threatened by water main failure.

Pipe Diameter in.	Philadelphia			
	Circular	Split Bell	Longitudinal	Binghamton
6	8.7	6.9	10.0	11
8	7.7	10.6	9.5	12
10	10.2	13.2	13.2	12
12	12.2	9.4	20.6	13
16-48	21.9	29.7	47.1	—
16	—	—	—	14
20	—	—	—	15
24	—	—	—	16

Table 9.8

TIME TO REPAIR BREAKS (IN HOURS)

9.3 The pipe rehabilitation decision

As water flows in the interior of cast-iron unlined mains electrochemical breakdown of the pipe wall can occur resulting in the formation of tuberculous. The rate at which iron is pulled out of the pipe wall depends on the properties of water and operating characteristics of the pipe. More specifically the factors that accelerate internal corrosion are: low pH, dissolved oxygen, insufficient alkalinity, high concentration of minerals and dissolved solids, elevated temperature and high flow velocities. It is also the case that when water is supersaturated with calcium or magnesium (high pH) scale may form in the interior of the pipe. The build-up of main sediments and tuberculation reduces the carrying capacity of a pipe by reducing its effective diameter and increasing frictional losses. The Hazen-Williams C coefficients can be reduced even more than three times from its initial value. A significant reduction in fire flow capacity and increase pumping costs would be some of the important consequences. Tuberculation can also have an impact on water quality. Water may become discolored and develop bad taste or odor. The most serious though problem associated with tuberculation is the growth of bacteria and viruses in the rough surface of tuburculous. The great difficulty in identifying the exact location of such bacteria along the pipe length makes this problem even harder to attack.

Cleaning and lining (or rehabilitating) a water main is considered by the water industry as the most effective method for resolving the problem of tuberculation. The cleaning process consists of mechanically or hydraulically scraping the inside of the pipe to remove all corrosion products. After cleaning is completed a thin lining of cement mortar is usually applied to the interior pipe wall.

The cost of pipe rehabilitation is on average about half to one third of the pipe replacement cost and does not vary as much as replacement cost with pipe diameter. Walski (1982) provides the following values for rehabilitation cost of water mains (Table 9.9):

Table 9.9

UNIT COST FOR REHABILITATION OF WATER MAINS

Diameter (inches)	Unit Cost ($/foot)
4	25.00
6	25.00
8	25.00
12	25.00
16	23.10
20	23.10
24	23.10
30	32.20
36	34.50
48	42.20

A more detailed evaluation of costs involved in pipe rehabilitation is provided by Walski (1985) and by a study performed for the U.S. Department of Housing and Urban Development by Brown and Caldwell (1984).

The decision for rehabilitating a pipe will be based primarily on economic considerations for increased pumping costs due to the reduction in carrying capacity. Also on reliability of services criteria related to reduction in fire flow capacity and water quality problems. Thus, the need for pipe rehabilitation is unrelated to the previous break history of the pipe. The decision though of rehabilitating it can very well be affected by the projected failure pattern of the pipe. There has been no conclusive argument in the literature as to whether cleaning and lining of the pipe accelerates or decelerates the break-rate. Apparently, rehabilitation is not expected to eliminate the existense of other break causing factors except internal corrosion. It will also not restore the structural integrity of the deteriorated pipe wall and there could be cases where corrosion might be enhanced by air trapped between the coating material and the interior wall. In the statistical analysis performed for the City of Cincinnati (Chapter 6) it has been found that cleaning and lining had no effect at the break-rate of pipes. Thus, given that further research and tests will be needed in order to understand the effect of cleaning and lining of pipes on the future evolution of breaks, a reasonable first-order approximation would be to assume that the break-rate remains unaffected by rehabilitation.

Under such assumption the derived probabilistic predictive models could provide a very useful piece of information for making the rehabilitation decision. This is so, because it might be uneconomical to rehabilitate a pipe that is expected to experience frequent breaks (and thus associated repair costs) and it would rather be better to replace it. If we denote by C_{rehab} the rehabilitation cost of a pipe and by C_r and C_b the replacement and repair cost per break respectively, then the following economic comparison can be made:

$$C_r \lesseqgtr C_{rehab} + E_o \cdot C_b \sum_{t=0}^{n} \left(\frac{1+g}{1+r} \right)^t \qquad (9.39)$$

where E_o = expected number of breaks in current year

g = average break growth rate

n = time horizon, usually set equal to 10-15 years

The values of E_o and g could be estimated according to the techniques presented in Section 9.2. If it turns out that the cost of replacing the pipe is less than the present value of rehabiliting it plus the expected repair costs, then it is more economical for the pipe to be replaced. Given the predictions about future pipe breaks, a rehabilitation decision could eventually be substituted by a replacement decision, solely based on reliability criteria. In any case, it appears that mostly pipes in the fast-breaking stage and very few pipes in the slow-breaking stage will exhibit break-rates such that a replacement decision would be made rather than a rehabilitation decision based either on economic or reliability considerations.

The procedure described in this section assumed that pipes that need rehabilitation would have been preselected according to the criteria chosen by the water utility and ranked into priority groups. The next step then would be to decide which of them will need replacement rather than rehabilitation.

9.4 Comparisons between replacement/rehabilitation scenaria based on the derived models and currently applied techniques

Water utilities use various criteria and methodologies for making replacement and rehabilitation decisions. Past studies about deteriorating systems have also proposed guidelines for maintenance decisions. A representative sample of those techniques, classified by system where they have been applied or by study where they have been proposed, is presented below. A discussion of how the developed strategies relate to the findings of this study is also included in each case.

A. The New Haven, CT System

The pipe replacement decision in the New Haven System was based only on reliability considerations and no economic evaluation at the individual pipe level was performed concerning future repair and replacement costs. The three most important factors on which replacement decisions are based were:

a. On site inspection to determine pipe wall thickness for problematic pipes. Pipes with severe wall deterioration were replaced.

b. Liability to the water utility from a potential break. Judgment on a case by case basis was required in order to make that assessment.

c. Inconvenience to customers caused by frequent future breaks. This assessment was also based on personal judgment and was highly influenced by population density.

Under the light of the results obtained in this study by analyzing the New Haven system it appears that several of their strategies are justifiable, while several of their decisions could have been assisted by the models developed here. More specifically, it appears reasonable not to base their replacement decision on economic costs of repairs versus replacement. This is so because according to the developed predictive models, the failure probabilities of the great majority of the pipes in their system are not high enough to result in future repair costs that would exceed replacement costs. Assuming for example a break repair cost of $1,400 per break, a replacement cost of $100/ft. and a pipe 1,500 ft. in length, it can be shown by applying equation (9.36) that in order for the optimal replacement time to be in less than 20 years, the yearly probability of failure should be greater than 0.77. (The real interest rate was assumed equal to 5% and the yearly break growth rate was set equal to the unrealistically high value of 10%.) Such high yearly failure probabilities are practically unobservable in the New Haven system. Only pipes that are in the fast breaking stage could have such high failure probabilities, but those are too few in that system. Nevertheless, if attention is focused on those pipes and by assuming a constant break rate at the fast-breaking stage, it can be shown by applying equation (9.38) under the rest of the assumptions of the previous example, that replacement would be economical only if the break rate is higher than 5.35 breaks/year. Thus, the above examples clearly justify the utility managers not to base replacement decisions on economic costs of this type.

Given the above discussion it becomes clear that the failure probabilities obtained by the models derived for the New Haven system would be very useful for reliability assessments in that system. In other words, if the impact of a break (e.g. inconvenience to customers or liability to the utility) that has been used as a criterion for replacement in that system is associated with an estimated failure probability, a better judgment is likely to occur than solely relying on personal experience. Of course this does not imply that the models should be used in any occasion as a substitute for subjective judgments, but they rather be used as complementary quantitative tools that would assist in making a final decision. Although the number of expected major breaks in the New Haven system is very small as compared to other systems, it would be useful to know that according to the results of the models in this study, they are expected to double in the next 10 years. Such observation might not only affect future maintenance practices but could also have implications for budgeting allocations related to expected repair costs.

B. The Manhattan, NY System

Although we do not know what the detailed maintenance policy has been in that system we can focus on the recommendations made by the New York City Water Supply Infrastructure Study (New York District Corps of Engineers, 1980). Currently applied by the city replacement priorities are not based on pipe age but rather on capacity requirements, previous break history and major street and sewer reconstruction projects in parts of the city. Mains on which two breaks have occurred in one block and mains that are believed to have a higher than average break rate are top candidates for replacement. Also if streets in the city are to be completely reconstructed mains in these streets are also replaced based on the above criteria or on whether they were laid prior to 1930 (because the pipes installed before 1930 were not cement-lined).

The decision not to base replacement on pipe age but rather on previous break history seems appropriate to a first order approximation. But as the derived predictive models for both the cities of New Haven and Cincinnati indicated, two pipes with the same previous maintenance history are not necessarily at the same risk for breaking. For example, previous break rate would be a factor predicting future rates, but according to the results of this study, this is very likely not to hold if a pipe is in fast breaking stage. Pipe age will also contribute to the failure probability up to the degree reflected by the baseline hazard function and thus it should also be taken into account. Other characteristics like land development, internal pressure, and installation period could result in variable break rates for pipes with the same break history. Given that not all "high risk" pipes can be replaced at once, capturing the variability in break rates among the set of pipes that are already at high risk is very important for prioritizing pipe replacements. This stage is where the developed models would be of help.

The methodologies applied by the Manhattan Study (1980) (primarily bivariate and discriminant analysis) were discussed in Chapter 2. The recommendations of the study regarding the main's replacement program as far as pipe breaks are concerned basically consist of replacing all pipes with multiple breaks in a city block within the next 10 years. The threshold though for multiple breaks has been defined as one previous break during the past 25 years. Clearly such recommendations would be effective if the budgets are available for implementing such program. But since it has been established that high variability in break rates among pipes can exist beyond the first break, such policy will not be optimal for a city where capital rationing exists for replacement. That is, when the economic resources for replacement programs are tight, then a more refined categorization of pipes in risk categories will be needed. Besides that, it is not clear, why should pipes be recommended for replacement after their first break even in cases where only reliability considerations are assumed. That is, there would certainly be pipes with low failure probabilities even after their first break.

C. Louisville, KY System

The water utility in the city of Louisville, KY is using the following criteria for main replacement decisions based on a point system (Table 9.10):

CRITERIA	POSSIBLE POINTS
1. Central Business District (50/5%):	
A candidate in the KIPDA Central Business District (CBD), Zone 1.	50
2. Redevelopment Zone (200/20%):	
An approved redevelopment zone candidate installed prior to 1937 in which street and/or sewer replacement is proposed.	
3. Main Size (0-200/20%):	
Four inch and smaller unlined pipe (CI, GA).	
Select four inch and smaller lined pipe (CU, CL, DI, PL).	50
4. KIPDA Roadway Classification (0-40, 4%):	
Interstate (4)	40
Major Arterial (3)	30
Minor Arterial (2)	20
Collector (1)	10
Local (0)	0
5. Main Break Data (0-50/5%):	
MB = ____, Main Break Frequency (MBF) points = MBF ÷ 10.	0-50
6. Joint Leak Data (0-50/5%):	
JL = ____, Joint Leak Frequency (JLF) points = JLF ÷ 10.	0-50
7. Field Pipe Samples (0-50/5%):	
Internal Corrosion (Tuberculation).	25
External Corrosion (Graphitization or Pitting). (light)	15
(heavy)	25
8. High Maintenance Priority (0-100/10%):	
An appropriate point value for special high maintenance candidates which do not fall into any specific category.	0-100

Table 9.10

CRITERIA FOR MAIN REPLACEMENT PROGRAM – LOUISVILLE, KY

Table 9.10 (continued)

9. Fire Flow Availability (0-50/5%):

Flow Rate at 20 psi = _______________ gpm.	
No hydrants within 750 ft.	25
Flow less than 500 gpm at 20 psi.	25
Flow 500 to 999 gpm at 20 psi.	15
Flow 1000 to 1500 gpm at 20 psi.	5
Flow greater than 1500 gpm at 20 psi.	0
Continuous red water during flow test.	25

10. Red Water Data (0-50/5%):

RW = ___, Red Water Frequency (RWF) points = RWF ÷ 10.	0-50

11. Documented Water Quality Data (0-100/10%):

An appropriate point value for specific water quality problems determined by field testing.	0-100

12. Corrosive Soil Zone (0/0%):

Data base too small at present time to evaluate.	0

13. Dead End Main (20/2%):

Dead end main excess of 500 feet.	20

14. Age/Type of Pipe (0-20/2%):

1856-1865 -	Sand cast iron - 9' and 12' lengths.	20
1866-1926 -	Sand cast iron - 12' lengths and unlined Delavaud - 18' lengths.	10
1927-1930 -	Unlined Delavaud - 18' lengths.	20
1931-1936 -	Unlined Delavaud - 18' lengths.	10
1937-1970 -	Lined Delavaud - 18' lengths.	0
1971-Present	Lined Ductile Iron - 18' lengths.	0
1930-1966 -	Asbestos - Cement.	0

15. Type of Joint (0-20/2%):

1862-1922 -	Lead.	10
1923-1955 -	Leadite (Sulphur).	20
1956-1960 -	Mechanical.	0
1960-Present	Push-on.	0

Total: 1000

Typical Range for Replacement: 250-400

The criteria used clearly include many other factors besides the projected break events. Apparently, many utilities are expected to use some type of point system (implicitly or explicitly) to rank pipes according to maintenance proiorties. The evaluation system described in Table 9.10 could certainly be improved if a measure such as the failure probability of a main is included among the criteria. Thus, for example, as argued extensively in previous sections, criteria like main break frequency, soil corrosivity and pipe age could be lumped under an estimated failure probability measure. Of course, the relative weight put on such measure with respect to other criteria will be highly varying depending on the priorities of each utility.

D. Boston, MA System

The Boston Water Distribution System has a very good break record as compared to other systems in the U.S. (Sullivan, 1982). The priority scheduling for replacing or rehabilitating a pipe is based on maintenance records, age records, loss-of-head tests, fire-flow tests and implementation of other roadway reconstruction projects in the city. A goal has been set to replace or rehabilitate all water mains 100 or more years old.

As far as pipe rehabilitation needs are concerned it must very well be expected that older pipes are in greater need for restoration. But as far as replacement of mains because of breaking is concerned, the results of the New Haven and Cincinnati systems clearly pointed out that utilities must be aware that older pipes are not necessarily the worst. Of course, the particular characteristics of the Boston system are not considered here for making this argument. Nevertheless, water utilities need to be aware of that fact. Arguing by analogy from the results obtained from the New Haven system if for example all pipes installed before 1920 were replaced in that system, the expected number of breaks per year would have been reduced only by 25%.

In order to clarify certain of the points made in this section, some further examples associated with the New Haven system maintenance strategies in relation to the derived predictive model are presented. All pipes included in the following analysis had no more than two previous breaks since for those pipes the derived model would apply (as argued in Chapter 5, pipes with equal or greater than three breaks were too few in the New Haven system to obtain any reliable predictive model from this part of the data).

As it has been shown in Tables 9.1 and 9.2, the expected number of breaks in the New Haven system during the year 1984 is approximately equal to 16. Pipes with zero, one or two previous breaks have contributed to a given percentage in arriving at that number. Table 9.11 shows the number of breaks that are expected according to the model to occur only on pipes with zero previous breaks. In this and the following tables the number of expected breaks during the current year is denoted by E1 while E2, E3 and E4 denote the expected breaks 2, 3 and 4 years hence. It is calculated that about 62% of the total expected breaks would be experienced by pipes with zero previous breaks. Of course it must be pointed out that such result is based solely on the statistical model and if, for example, an unknown break causing factor was not included in the regression it could be the case that the number of expected breaks calculated by the model for the pipes with zero breaks is an overestimate. Figure 9.1 shows a plot of the yearly failure probabilities (or equivalently expected number of breaks) of all pipes in the system as a function of the

	E1	E2	E3	E4	LENGTH
	SUM	SUM	SUM	SUM	SUM
DIA					
6	0.09	0.10	0.11	0.12	14300.00
8	1.49	1.67	1.85	2.03	348600.00
10	1.19	1.32	1.44	1.56	168700.00
12	3.67	4.02	4.36	4.70	673725.00
16	1.67	1.83	1.99	2.14	337250.00
20	0.51	0.55	0.60	0.65	125600.00
24	0.59	0.63	0.67	0.71	75750.00
30	0.27	0.30	0.32	0.35	52800.00
36	0.29	0.31	0.34	0.36	48675.00
42	0.04	0.04	0.05	0.05	6250.00
48	0.09	0.09	0.09	0.10	46000.00

Table 9.11

EXPECTED NUMBER OF BREAKS FOR PIPES WITH ZERO PREVIOUS BREAKS CLASSIFIED BY DIAMETER SIZE, NEW HAVEN

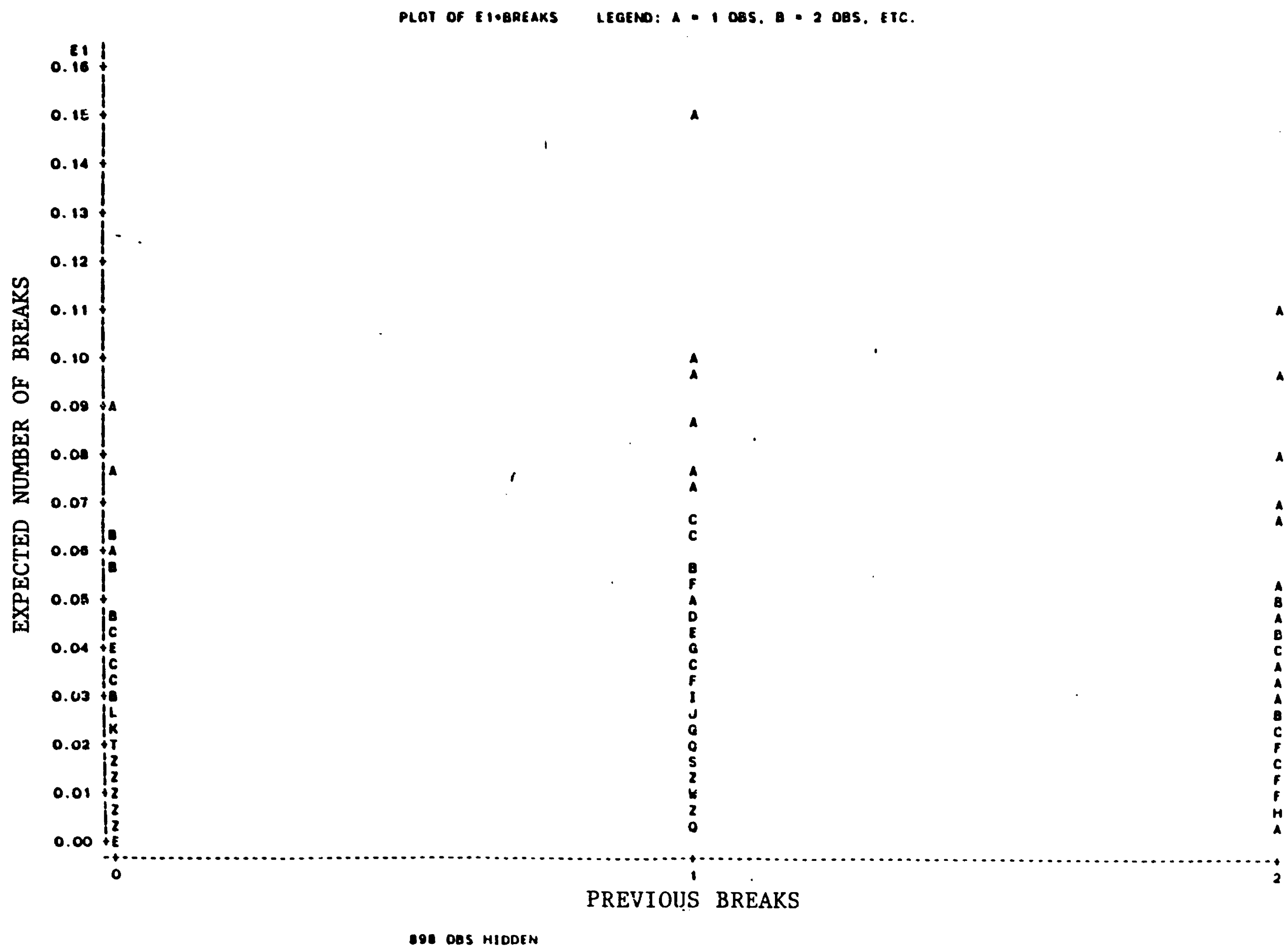


Figure 9.1

EXPECTED NUMBER OF BREAKS BY
NUMBER OF PREVIOUS BREAKS

number of previous breaks. The high number of pipes with zero breaks in the data (80%) has clearly contributed in making the percentage of expected breaks associated with them that high. Table 9.12 demonstrates the expected number of breaks associated with pipes with previous breaks greater than or equal to one and also with pipes that had two previous breaks. It is calculated that about 24% of the breaks expected from the whole set of pipes that have already broken one or more times would occur on pipes with two previous breaks. Such finding indicates that if a strategy focuses on replacing only pipes with two or more previous breaks it will not necessarily result in a significant reduction of the total number of expected breaks. Table 9.13 shows the number of expected breaks associated with pipes with yearly probability of failure greater than 5%. Those pipes contribute 15% of the total number of expected breaks (i.e. when also pipes with zero previous breaks are included), while they contribute 40% of the expected breaks of the set of pipes already broken. Those high risk pipes (failure probability > 0.05) consist on the other hand of only 23% of the total length of broken pipes.

Thus, it appears that a strategy of replacing all pipes among those that have broken, which have probability of failure greater than 5% (about 35 pipes) could result on the average in a reduction by 40% of the total number of expected breaks while by replacing only pipes with two previous breaks will result in a similar reduction by 24%. Given the fact that in the New Haven system pipes with two previous breaks have only about 10% less total length than the pipes with failure probability greater than 5%, it appears that a replacement strategy based on failure

PREVIOUS BREAKS ≥ 1

DIA	E1 SUM	E2 SUM	E3 SUM	E4 SUM	LENGTH SUM
6	0.14	0.15	0.16	0.16	5500.00
8	1.30	1.41	1.51	1.55	137900.00
10	0.83	0.90	0.97	1.00	91850.00
12	1.98	2.10	2.22	2.28	205850.00
16	0.92	0.97	1.02	1.04	87250.00
20	0.28	0.29	0.29	0.30	36550.00
24	0.29	0.30	0.32	0.33	28250.00
30	0.12	0.12	0.12	0.12	7950.00
36	0.06	0.06	0.07	0.07	7250.00
48	0.16	0.16	0.17	0.17	16500.00

PREVIOUS BREAKS = 2

DIA	E1 SUM	E2 SUM	E3 SUM	E4 SUM	LENGTH SUM
6	0.03	0.03	0.04	0.04	2000.00
8	0.46	0.58	0.67	0.70	34650.00
10	0.22	0.29	0.34	0.37	17550.00
12	0.46	0.59	0.70	0.74	55950.00
16	0.19	0.24	0.28	0.30	10700.00
24	0.06	0.07	0.09	0.09	4750.00
36	0.01	0.01	0.01	0.01	1000.00

Table 9.12

EXPECTED NUMBER OF BREAKS FOR PIPES THAT HAVE ALREADY BROKEN, NEW HAVEN

DATE	E1 SUM	E2 SUM	E3 SUM	E4 SUM	LENGTH SUM
0	0.67	0.70	0.74	0.76	16450.00
6	0.15	0.15	0.15	0.15	6200.00
8	0.21	0.23	0.25	0.27	18350.00
10	0.06	0.06	0.06	0.06	800.00
12	0.07	0.07	0.07	0.07	5700.00
16	0.08	0.09	0.10	0.11	1600.00
21	0.05	0.05	0.05	0.05	3000.00
25	0.06	0.07	0.08	0.09	5500.00
40	0.12	0.14	0.16	0.16	8500.00
41	0.05	0.06	0.07	0.07	2750.00
50	0.05	0.05	0.05	0.05	7000.00
56	0.05	0.07	0.08	0.08	5000.00
57	0.05	0.05	0.05	0.05	4500.00
58	0.06	0.06	0.06	0.06	6500.00
63	0.24	0.26	0.28	0.29	19750.00
65	0.07	0.07	0.07	0.08	7000.00
67	0.11	0.11	0.11	0.11	10100.00
71	0.17	0.20	0.21	0.21	6000.00
72	0.67	0.09	0.10	0.10	3500.00
76	0.07	0.07	0.07	0.07	3200.00

Table 9.13

EXPECTED NUMBER OF BREAKS FOR HIGH RISK PIPES
(PROBABILITY OF FAILURE > 0.05 CLASSIFIED)
BY DATE OF INSTALLATION, NEW HAVEN

probabilities would dominate one where replacement is only based on the number of previous breaks. It must again be pointed out though that failure probabilities are proposed to be used only as a tool for assisting the replacement decision and not as a substitute for judgment of the water utility managers.

In order to demonstrate the effect that the results of the derived models can have on pipe replacement strategies as far as replacement based on age is concerned, Figures 9.2 and 9.3 were plotted. Figure 9.2 shows the number of expected breaks for each pipe during the year 1984 as a function of the date of installation. Figure 9.3 shows the number of expected breaks as a function of the variable C35, where C35 = 1 if a pipe was installed during the period 1930-35 and zero otherwise. It becomes clear from those two figures that most recent pipes can also have high failure probabilities, while pipes of the 1930-35 period are expected to perform the best. This finding, as pointed out previously, indicates that focusing replacement only on the older pipes in the network can very well be a suboptimal strategy.

9.5 The "bundling" decision

It has been argued by several utilities that efficiencies can exist if replacement or rehabilitation of pipes takes place in "bundles" rather than selectively for the whole network. In other words, it appears that in many cases it is preferable to perform restoration of the system by areas where a large number of pipes

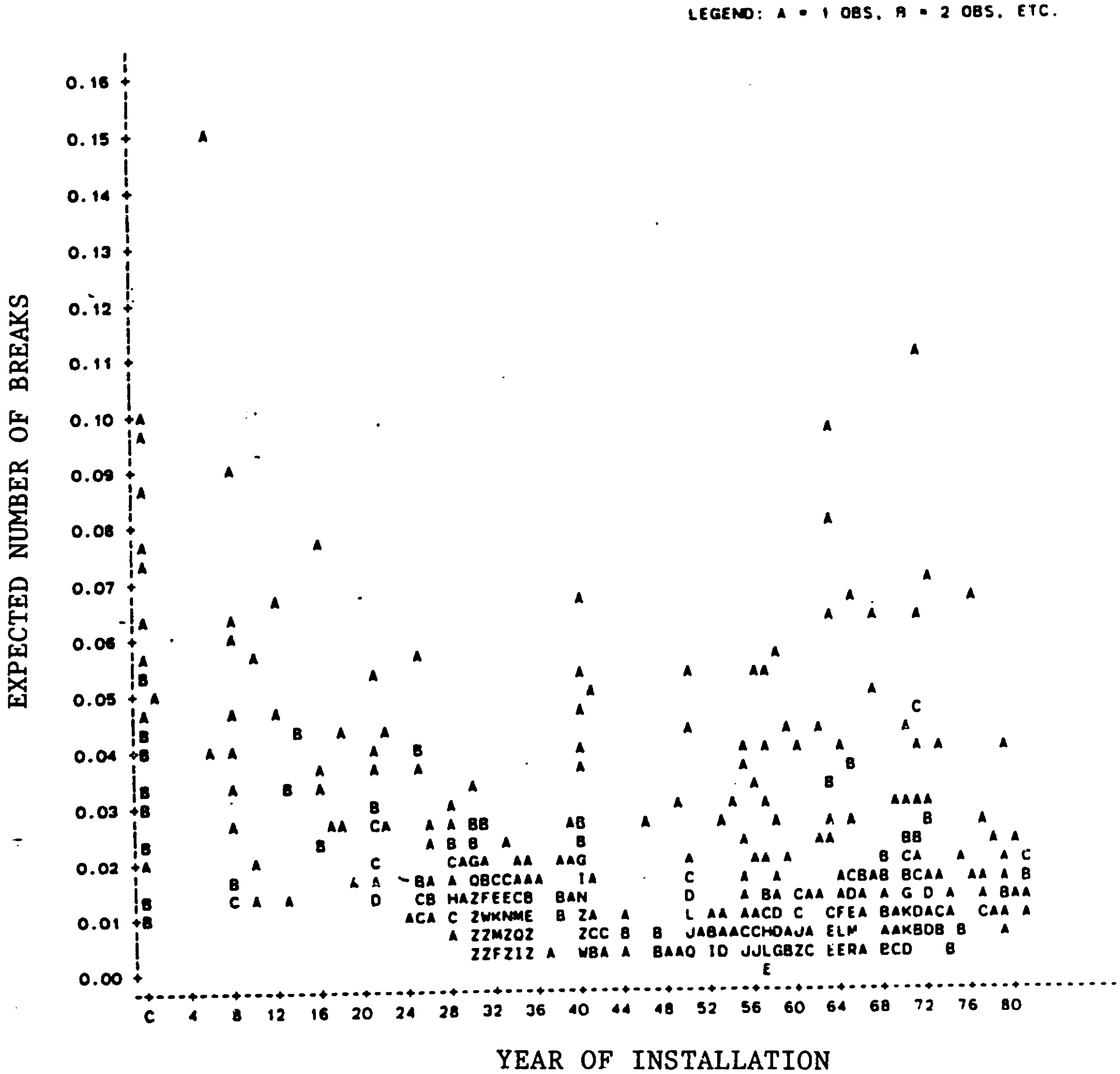


Figure 9.2

EXPECTED NUMBER OF BREAKS BY YEAR
OF PIPE INSTALLATION

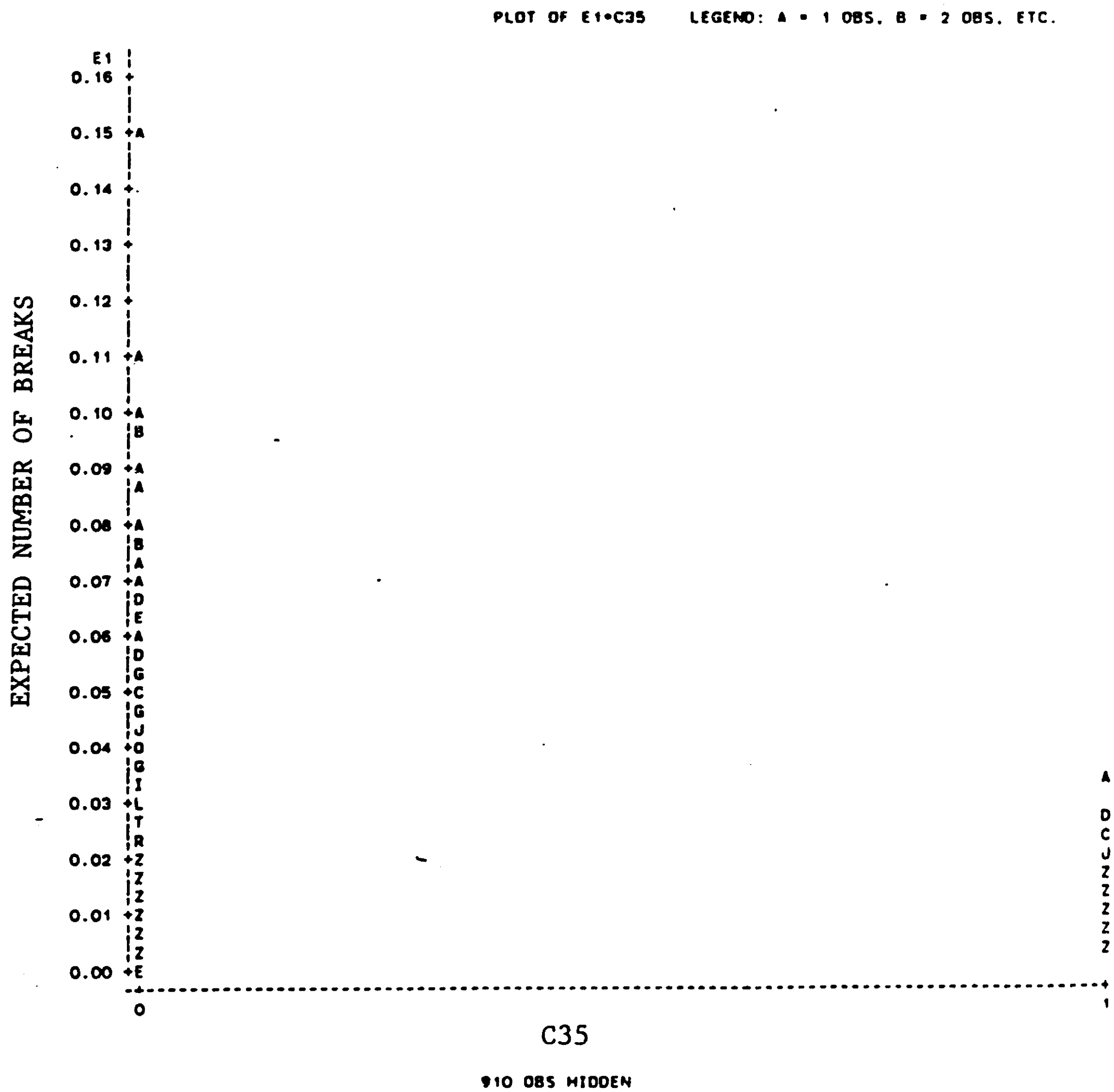


Figure 9.3

EXPECTED NUMBER OF BREAKS FOR
C35 = 1 and C35 = 0

in a given block or neighborhood is replaced or rehabilitated at once. Such policy is perceived, for example, as more efficient by the New Haven Regional Water Authority. A detailed description of the underlying reasons is presented by Ahrens et al [1983]. The most important of the efficiencies gained by "bundling" pipes together can be summarized as follows:

a. Economies of scale in contracting for new projects
b. Water quality improvements will benefit all area customers
c. Efficiencies in total time of operations for cleaning and lining
d. The chlorination operation is quicker and easier
e. Economies of scale in excavation and paving costs
f. Customers' neighborhoods are disturbed by construction activity only once rather than several times.

The "bundling" process in the New Haven system is performed according to the Grid System Method. It involves a systematic restoration of sections of the distribution system including street laterals. That is, not only the main streets but also the interior city street laterals are included in a given restoration project. Similar methods are followed by other utilities as well.

There are two ways that the derived probability of failure of individual pipe segments can be of help in the bundling decision. One is related with the identification of "high risk pipes" and consequently "high risk areas" in the water distribution network. Such high risk areas could become the focus of pipe bundles appropriate for replacement. The other is related to the selection of pipes that would need replacement within the area where rehabilitation has been decided to be performed. That is, if a grid consisting of main streets and laterals has been selected for rehabilitation, there would be pipes in that grid for which, according to the discussion in Section 9.4, it would be more economical and/or reliable to replace rather than rehabilitate them. Thus, estimated failure probabilities and break rates could serve exactly that purpose.

High risk pipes were identified for the system of New Haven by applying the derived model. The threshold for classifying a pipe as high risk was set arbitrarily equal to a yearly failure probability greater than 5%. Thirty-five pipes were selected as belonging in the high risk category. Their characteristics are shown in Table 9.11, where E1 denotes the expected number of breaks during the current year (1984) or equivalently the yearly probability of failure. Only pipes with up to two previous breaks were included, since the model applied only to this part of the data. The few pipes which have experienced three or more breaks will also belong in the high risk category unless they have already been replaced. By using maps of the New Haven water distribution system the high risk areas have been identified, where subsets of high risk pipes were grouped together. The results are shown in Table 9.12 where information is also provided about the link number for each pipe in the group, the geographical location of each particular area

OBS	LINK	DIA	LENGTH	BREAKS	DATE	E1
1	8	8	3500	2	72	0.070610
2	23	48	2500	1	63	0.063055
3	28	8	2750	2	41	0.050340
4	47	10	2000	2	0	0.097779
5	133	24	1500	0	0	0.057830
6	136	20	1500	1	0	0.077657
7	140	16	2000	1	0	0.086127
8	236	12	600	1	0	0.054992
9	290	12	1250	1	0	0.073798
10	438	12	3500	1	71	0.063268
11	468	24	6500	1	58	0.057726
12	492	16	3500	1	40	0.054932
13	601	8	5000	2	40	0.066729
14	693	6	3000	1	0	0.101223
15	698	12	3000	1	21	0.053655
16	706	16	3000	1	0	0.053655
17	708	8	2500	1	67	0.050211
18	781	24	1600	0	0	0.063101
19	861	12	7600	1	67	0.062554
20	922	10	800	1	10	0.056287
21	929	8	5700	1	12	0.066691
22	970	8	3200	1	76	0.067686
23	971	10	7000	1	65	0.067389
24	1021	12	6200	1	6	0.151299
25	1044	48	14000	1	63	0.097690
26	1069	16	3250	2	63	0.080973
27	1123	12	5500	0	25	0.057011
28	1162	12	4500	1	57	0.054464
29	1187	8	2500	2	71	0.111620
30	1274	12	5000	2	56	0.053073
31	1332	12	8600	0	8	0.090914
32	1339	12	4000	0	8	0.064968
33	1372	10	5750	0	8	0.058672
34	1373	8	7000	1	50	0.053510
35	1391	10	1600	0	16	0.076851

Table 9.14

SET OF HIGH RISK PIPES, NEW HAVEN

and the total number of breaks experienced by the pipes of each group. Since no exact information on pipe location in the network had been provided in the original data set and the derivation of Table 9.12 was based on inferences from street names, it should only be considered as a first order approximation in the attempt to identify the high risk areas in the system. By focusing the attention on those areas, additional break causing factors could be identified related to above ground activities, soil characteristics, etc. It is interesting, for example, to notice that all pipes in group 1 of Table 9.12, which is the highest risk group, are located close to the shoreline of the Quinnipiak River.

Another important point related to the application of the proposed methodologies for assisting in making the bundling decision is associated with the way pipes are coded in data sets. Since the focus for replacement or rehabilitation would be in pipes belonging in a particular system grid, the model could be used much more efficiently if predictions could be made corresponding to the pipe length included in each grid. Thus, for one more reason in addition to those mentioned in Chapter 4, pipe lengths should be rather short, possibly corresponding to a few street blocks, as grids for restoration would be defined in a similar scale.

Table 9.12

NEW HAVEN, High Risk Areas

Group Number	Link Number of Pipes in the Group	Information about the Geographical Location	Total Number of Previous Breaks for Pipes in that Group
1	1069, 1026, 1078, 1080, 1087, 1023, 21, 1107, 1100, 1045, 1109, 38, 63	Close to the Shoreline of the Quinnipiak River	85
2	644, 468	Orange	4
3	1020, 1021	North Haven, Sackett Point Rd.	4
4	1156, 1162	East Haven	4
5	107, 108, 290	New Haven (Longwharf, Brewery, Spring, Union)	9
6	865, 861	Hamden	5
7	723, 729, 726	New Haven Litchfield, Ave., Whallay Ave.	21
8	683, 684	West Haven (Forest St.)	6
9	1132, 1133	New Haven (Cosey Beach Ave.)	24
10	136, 140, 191	New Haven (Whitney, Church, Elm, Prospect, Canal)	5

CHAPTER 10

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

10.1 Introduction

The contribution of this work consists of the successful development of modelling methodologies for the analysis of the deteriorating water mains at the individual pipe level with clear implications of future maintenance practices. Given the fact that techniques developed in the past have failed to capture with any reasonable accuracy failure patterns at the individual pipe level, the successful implementation of the results of this work is expected to significantly influence the way current repair, replacement, and rehabilitation decisions are made. This investigation has clearly shown that the deteriorating process of larger diameter pipes should be modeled in a very different way than that used in current practice. Because of the great flexibility in the mathematical structure of the proposed modelling methodologies, significant insights have been generated about the various stages of deterioration that a water distribution system as a whole or individual mains, are likely to go through during their useful life. Issues pertaining to the effect of pipe aging on failure rates, which have resulted in many controversies appearing in the literature, have been clarified. The issue of missing records of pipe breaks beyond a certain time period in the past, that raised serious questions on the validity of models derived through such data sets, has been investigated and techniques for dealing effectively with the left censoring problem were tested under various conditions. For the

For the first time in the literature clear inferences and quantitative estimates could be obtained about the relation between the different pipe installation periods and the rates of deterioration of existing systems. The implications of the new methodologies on the economic evaluation of capital investments for pipe repair, replacement and rehabilitation are also important. The results clearly indicate that currently applied replacement rules and analytical techniques for economic evaluation, are oversimplified and could lead to great inefficiencies in management practices, both in terms of economic and reliability considerations. The approach followed in this study provided the necessary tools for categorizing existing systems according to their current and future degree of deterioration, so that the appropriate method(s) of analysis, as proposed by this work, could be readily applied. Suggestions on future data collection needs and implications of current operating practices in water utilities on future pipe failures are also made and new knowledge has been created concerning those issues.

The following section presents in greater detail the conclusions and contributions of this work.

10.2 Conclusions of the research

1. The study of the New Haven, CT and Cincinnati, OH water distribution systems clearly revealed the high variability in break rates than can exist among different mains of a particular system and among water distribution systems as a whole. Due to differences in maintenance practices, construction materials, soil and pipe characteristics, weather conditions and water properties, a broad range of failure patterns can be developed.

2. Two distinct types of failure patterns were identified: Slow breaks, where a pipe experiences infrequent breaks at an increasing break rate with time and fast breaks, where the pipe enters a fast breaking state with very frequent breaks (probability of failure greater than 50% each year). The above observation was found to hold for large diameter pipes (greater than or equal to 8 inches in diameter), while no data were available to perform any investigation on smaller diameter pipes. The fast breaking pattern was clearly much more pronounced in the Cincinnati system, where the overall condition of pipes (according to the failure records) appears to be much worse than that of the New Haven System.

Cox's proportional hazards model can work successfully in predicting breaks at the slow breaking stage. For pipes in that stage, the renewal assumption works very effectively in providing us with an appropriate failure time model. For pipes entering the fast breaking stage, a variation of Cox's model has been successfully applied, where the time to failure is considered to be equal to the time for entering the fast breaking state. The latter model estimates the probability for a pipe that has already experienced one or two previous breaks to enter a fast breaking state.

3. The idea, most commonly reported in the literature, that the number of breaks increases exponentially with time, once a pipe starts breaking, has been challenged by the findings of this research. Although the time to next break could be considered to decrease exponentially with time as the first one or two breaks occur, that indeed appears not to be the case, as multiple breaks start occurring. It has been found that, when a pipe enters the fast breaking stage, the break rate remains approximately constant. It has also been found that, although it is possible to predict an average value of the break rate for a pipe at that stage, very high variability of this rate does exist among the pipes entering that state, which only partially can be explained by the derived models.

4. The research strongly suggested that the problem of left censoring of the data sets can be effectively accommodated in the proposed Cox's regression methodology by setting the survival time of pipes with missing past break observations as starting at the time that records have become available and using the installation period of those pipes as a covariate. This finding is of great practical use, since many of the existing data sets have missing break records beyond a time period in the past.

5. The proposed methodologies for analyzing deteriorating water mains have avoided many of the deficiencies of currently applied techniques and modelling approaches that appear in the literature. Thus, problems associated with mixing pipes of different characteristics and breaking patterns together in the derivation of general predictive models for pipe failures, have been treated with theoretically sound methods. Also arbitrary model structures (e.g., exponential increase of break rate with time) have been avoided through the use of Cox's model, where the baseline hazard function does not need to be prespecified.

6. The covariates found to be statistically significant in the derived regression models can serve a dual purpose: a) Increase our understanding of the physical phenomenon of breaks in pipes and the interactions of the various factors contributing to breaks and b) be used as potential predictors of failure for forecasting future breaks. For pipes in a slow breaking state, the analysis revealed the following independent variables to be affecting the break rate:

    a. Internal pressure; although it is believed that pressure is not a direct cause for breaks, high internal pressure can clearly be associated with the acceleration of break occurrences once other stress factors are present (e.g., excessive corrosion of pipe wall, improper bedding conditions, etc).

b. Land development covering the pipe; this variable usually represents an average estimate of the percentages of industrial, commercial, transportation, and residential land use covering the pipe and could thus be considered as a surrogate of external loads transmitted at the pipe (in addition to loads imposed by soil cover and improper bedding conditions). The percentage of land development can also be associated with the type and number of other activities affecting water mains. For example, high land development could indicate greater chances for disruptions caused by excavations in the vicinity of the pipe, more frequent contact of water pipes with other structures believed to be contributing to breaks and greater number of pipe connections associated with potentially higher nunber of leaks that could undermine pipe's bedding material.

c. Early age of the second break; apparently, an early second break clearly points to certain deficiencies of pipe material or high concentration of break causing factors, that are expected to cause accelerated future breaks, if appropriate remedial measures are not taken promptly.

d. Number of previous breaks; since the occurrence of previous breaks reveals some information about the presence of unfavorable conditions for a particular pipe, it is intuitive to believe that it could be used as a potential failure predictor. This indeed has been the case for pipes in the slow breaking stage. Nevertheless, the number of previous breaks turned out not to be an important predictor of future failures for pipes in the fast breaking stage. This finding along with the detailed study of the various phases of failure indicated that the early stages of deterioration can be modelled as a non-homogeneous Markov Process, where the probability of a break varies both in time and with the number of previous breaks. When a pipe though enters into the stage of severe deterioration (fast breaking stage) then breaks can be modelled as Poisson arrivals at a constant rate.

e. Pipe length; as opposed to previously developed models, pipe length has been used in the regression models as a covariate. The omission to include length in the predictive models developed by other investigators implicitly assumes that the break causing factors are uniformly distributed along the pipe length. Such hypothesis though is very easily contradicted by reality. The inclusion of length in the predictive equations derived in this work indicated that the probability of failure is in most cases proportional to the suqare root of length. Such finding could be attributed to the following causes: i) longer pipes might lie on less developed areas of the city (since less connections are present), thus they are indirectly associated with more infrequent risk factors (excavations, traffic loads, etc.); ii) breaks are associated with unfavorable conditions developed only at certain points along the pipe length. Thus, if two pipes with different lengths have already broken, the longer pipe has higher chances for breaking in the future, since more length is exposed to risk factors, but the probability of failure does not increase linearly with length, because the distribution of potential break factors would be concentrated in the vicinity of the previous break.

f. Pipe installation period; both the New Haven and Cincinnati studies indicated that younger pipes (i.e., pipes installed in the 50's and 60's) were less reliable than older pipes. This finding could be explained by differences in pipe and joint materials used and also differences in construction practices. It is clearly suggestive of future needs for modifying pipe standards and applied construction methodologies.

g. For pipe entering a fast breaking state it has been found that besides length, which is used as a scaling factor, internal pipe pressure was the most important predictor of future pipe failures. This finding indicates that in a severely deteriorated pipe the stresses caused by internal pressure are critical for accelerating future failure events. This observation and the fact that the effect of pressure on the probability of failure can be quantified through the model are very important for utility operating practices and future standard's design. It could be possible, for example, under certain circumstances, to confine the number of future breaks at a desirable level by reducing pipe pressure. Also, pipe wall thickness standards could be affected by this finding.

h. Soil corrosivity; although higher soil corrosivity was not found to affect the break rate of pipes in the New Haven system, it had an important negative effect for the Cincinnati pipes. As it has been argued in previous chapters, soil corrosivity only partially captures the potential effects of corrosion on pipes and thus it should not be considered as the only factor associated with it. Nevertheless, the much more severe deterioration observed for pipes in the Cincinnati system as compared to those in New Haven could possibly be related to more corrosive soil properties in the former system.

7. Cox's Regression provided us with an estimate of the shape of the baseline hazard function, which gives an accurate representation of how time affects the break rate. The "bath tub" shape obtained in many cases for that function is an intuitively appealing result and provides a satisfying answer at the current controversy that exists in the literature about the effect of aging pipes. More irregular shapes of the baseline hazard functions obtained in certain cases also point out that the issue of the effect of time on breaks is not a trivial one.

8. The probabilities of failure and expected break rates by the model(s) of the individual pipe segments can be used directly

for assessing the reliability of a particular water distribution system. Two questions are raised concerning the reliability problem. The first is related with the assessment of the reliability of isolated links in the network. This apparently can be easily done by using the appropriate predictive model among those developed in this study. The second question relates to the more complicated issue of assessing the reliability of the system for delivering water at particular points in the network, given the failure of an individual link. What makes this problem complicated is the fact that hydraulic analysis at the network level is needed in order to evaluate the impact in areas of the system at some distance from the break. On the other hand, most of the existing systems have many redundancies built into them that minimize the effect of a break in points quite remote from it. The redundance of a system is associated both with the number of loops in the network and the number of sources (reservoirs) for water supply. Of course, it is expected that the redundancies in the system will be less as failures occur in larger diameter mains. For this reason, it is believed that water utilities could use the results of this research (primarily probabilities of failure and break rates of individual links) as inputs to the hydraulic models that usually they have set in place, to determine the impact (reduction in flow and pressure) of major breaks in various parts of the network.

9. The findings of the research have the following implications for repair/replacement/rehabilitation strategies:

a. In order to obtain estimates of the number of expected breaks and thus evaluate the future expected repair costs, an assessment of the general condition of the system (e.g. in terms of number of breaks per ft. each year) is necessary. This will give a preliminary idea about which of the proposed methodologies for analysis would be more applicable for that system. The mathematical computations for estimating the expected number of future breaks become more complex the more in "good" condition a system is, since our results suggest that in this case, where most pipes are in the slow breaking stage, the breaking mechanism is better approximated by a Markov Process, where the probability of failure varies with time and is conditional on the number of previous breaks. As we move to more severely deteriorated systems, many pipes have entered into the fast breaking stage, where breaks tend to become more frequent and more uniformly distributed along the pipe length (as opposed to being a localized phenomenon occurring in pipes that are structurally more sound). In such severely deteriorating systems several pipes will again be in a slow breaking stage. For the rest of pipes though which are in the fast breaking stage, the number of previous breaks is not any more an important predictor of future failures. In this latter case, making projections about future breaks is not as complex computationally as in the case of the more reliable overall pipes and the process can be modelled as Poisson.

b. Even in systems that are experiencing a relatively small number of breaks yearly (such as the New Haven system) and are thus in a generally good condition, the number of breaks in large diameter pipes is expected to double in about 10 years if proper remedial measures are not taken. The methodology proposed by this research provides utility managers with a quantitative tool for evaluating various replacement scenario and also for assessing future budgeting needs for pipe repairs.

c. The models developed in this work for describing the failure patterns in pipes result in quite different formulas for calculating an optimal replacement time, than the over-simplified formula currently applied by several investigators, which is based on the assumption of exponential increase of breaks with time.

d. The proposed methodology proved very helpful in identifying high risk areas in the network today and also determining areas with the potential for developing into "high risk" at some future time. It thus provides a very useful input for the bundling decision of pipe repairs.

e. Among the developed models, the one that provides an estimate of the probability for a pipe to enter into a fast breaking state, is a particularly useful input in the pipe rehabilitation decision. Since it is not economical to rehabilitate a pipe that is likely to frequently break in the future and it is instead better to replace it, the estimate of the probability for entering into a fast breaking state at some time in the future is critical for comparing the economic costs of the two alternatives.

10. The results indicate that currently used rules of thumb for pipe replacement by many utilities could lead to very suboptimal decisions. Thus, decisions for replacement based solely on criteria such as pipe age and number of previous breaks are miscalculating the real needs and ignore significant factors contributing to future breaks.

11. Given the fact that internal pipe pressure was found to be an important variable for determining the probability that a pipe would enter the fast breaking state once it starts breaking, some changes in system operating practices could occur by reducing pipe pressure, where it is possible, and thus slowing down the number of future breaks. Also, since both high pressure and most recent (after 1950) pipe materials were found to be related to acceleration of the failure rate, certain suggestions for future pipe design standards could be made.

12. This study clearly pointed out many of the deficiencies of existing data sets and indicated which of them would be most important to be corrected in the future. The problem of left censored data proved to be accommodated satisfactorily by the proposed methodologies. The following basic suggestions were then made concerning the collection of future data, that will

lead to more reliable models:

a. Pipe length does not need to be fixed in the data sets but a reduction of the high variability of its values should be attempted and the uniformity of external environmental conditions should become an important factor for its definition.

b. Information on the exact pipe break location and break type should be included in the future data sets, since this could resolve much of the ambiguity about how the various risk factors are distributed along the pipe length.

c. The intertemporal variation of certain variables (e.g., percentage of land development, pressure, pipe materials) is needed for a more accurate system representation. Also, information on maintenance practices and measures taken for corrosion protection would be useful for performing a more complete analysis.

10.2 Suggestions for Future Research

The following directions for future research are proposed within the framework of analyzing historical pipe break records:

1. Application of the proposed methodologies of analysis on data sets where improvements have been made according to the discussion presented in Chapter 4. Apparently many of the suggestions about better coding of break records and other relevant variables would apply only to future data collections, since for historical records much of the necessary additional information is expected to be missing. Nevertheless, it seems feasible that some of the improvements could be made even on past records. For example, information about the location of a pipe in the network could be useful in stratifying the data in a way that would reflect particular areal characteristics within the system. Also information on break-type and break location could possibly exist in historical records but not being yet properly coded, as it has been the case for the systems analyzed in this study. An improvement in that direction could even be made if pipe segments are redefined so that they correspond to shorter lengths.

2. The derived failure probabilities and break rates for individual pipes could be integrated with a numerical hydraulic model of the water distribution system, in order to examine the consequences of a break in terms of flow and pressure reductions in areas at various distances

from the break. Although water distribution systems have many redundancies, breaks in large diameter pipes could have an impact on locations far from the actual break.

3. The projections about expected breaks and the implied repair, replacement and rehabilitation strategies could be integrated with financing decisions, since now a more detailed assessment of future needs would be possible.

4. Extension of the proposed methodologies for analyzing failure records on smaller diameter pipes, sewers and gas pipelines, depending on the availability of data, could also generate useful insights about the behavior of such systems.

5. Statistical analysis of a water distribution system from a drier section of the country (e.g. California) would also be useful, since a comparison could be made with trends existing in cities of Eastern US, as it has been the case in this study.

REFERENCES

Ahrens, H., P. Gaewski and S. Turbak, "Justification for Cleaning and Cement Mortar Lining Cast Iron Water Mains," South Central Connecticut Regional Water Authority, October 1983.

BMDP Statistical Software, University of California Press, 1983, Adv. Appl. Prob., 10, pp. 431-451, 1978.

Brooks, B.J., "On the loss of Information through censoring," Biometrika, 69,1, 1982.

Gardner, D.R. and M.E. Taylor, "Modelling the effects of traffic loading on small-diameter rigid pipelines," Pipe and Pipelines International, March-April, 1983.

Clark, R.M., C.L. Stafford, J.A. Goodrich, "Water Distribution Systems: A Spatial and Cost Evaluation," Journal of the ASCE, WR3, October 1982.

Corrosion and Corrosion Protection Handbook, Edited by P.S. Schweitzer, Marcel Dekker, Inc, NY, NY, 1983.

Clarke, N.W.B., "Buried Pipelines," Published by Maclaren and Sons Ltd., London, 1968.

Cox, D.R., "Regression Models and Life-Tables," Journal of the Royal Statistical Society, 34, Series B, pp. 187-220, 1972.

Cox, D.R., "Partial Likelihood," Biometrika, 62,2, p. 269, 1975.

Cox D.R. and D. Oakes, "Analysis of Survival Data," Published by Chapman and Hall, New York, NY, 1984.

Dugovic, G.R., "There is Gold in Those Records - Sequencing the Replacement of Cast Iron Gas Main Sections," Rochester Gas and Electric Corporation, Report 80-D-39.

Effron, B., "The efficiency of Cox's likelihood function for censored data," J. Am. Statist. Assoc., 72, pp. 557-65, 1977.

Gaewski, P., "Water Main Restoration and Replacement Decisions," South Central Connecticut Regional Water Authority, August, 1984.

Heitzman, G. and J.L. Huber, "Louisville Water Company - Transmission-Distribution System Replacement and Rehabilitation Needs," Louisville Water Company, Kentucky, August, 1984.

Huber, J.L., "Water Main Replacement and Rehabilitation," AWWA Annual Conference Proceedings, pp. 253-242, 1980.

Hussock, B., "Causes of Underground Corrosion," Plant Engineering, 67-68, April 15, 1982.

Kalbfleisch, J.D. and R.L. Prentice, "The Statistical Analysis of Failure Time Data," John Wiley and Sons, New York, 1980.

Karaa, F., "A Decision Support Model for the Investment Planning of the Reconstruction and Rehabilitation of Mature Water Distribution Systems, Ph.D. thesis, Department of Civil Engineering, Massachusetts Institute of Technology, May, 1984.

Kay, R., "Proportional Hazard Regression Models and the Analysis of Censored Survival Data," Applied Statistics, 26, No. 3, pp. 227-237, 1977.

King, T., "Break Repair Costs Analysis," Water Department, City of Philadelphia, 1984a.

King, T., "Water Main Construction Cost Analysis," Water Department, City of Philadelphia, 1984b.

Kirmeyer, G.J. and G.S. Logsdon, "Principles of internal corrosion and corrosion monitoring," Journal of AWWA, Research and Technology, February, 1983.

Miller, R., and J. Halpen, "Regression with censored data," Biometrika, 69,3, pp. 521-31, 1982.

Moyer, E.E., J.W. Male, I.G. Moore and J.G. Hock, "The economics of leak detection and repair - a case study, Journal of AWWA, Management and Operations, January, 1983.

Newton, T.R., "Classifications of Corrosion Failures," reprinted from "Classification of Corrosion Failures witn Photomicrographs, Nalco Chemical Company, Oak Brook, IL, 1979.

Oakes, D., "The asymptotic information in censored survival data," Biometrika, 64, pp. 441-8, 1977.

O'Day, D.K. and L.A. Staeheli, "Information Systems for Facility Maintenance Decisions," Report No. 2, Study conducted for the Urban Infrastructure Network, The Urban Institute, contract No. 3264, October, 1983.

Peters, J.W., "Improve Maintenance of Cast Iron Mains," Pipe Line Industry, October, 1983.

Shamir, U. and C.P.D. Howard, "An Analytic Approach to Pipe Replacement," Journal of AWWA, 71,5, pp. 248-258, May 1979.

Singley, J.E., and T.Y. Lee, "Pipe Loop System Augments Corrosion Studies," Journal of the AWWA, Research and Technology, August, 1981.

Singley, J.E., and T.Y. Lee, "Determining Internal Corrosion Potential in Water Supply Systems," Journal of the AWWA, Research and Technology, August, 1984.

Stafford, C.G., "Maintaining Water Utility Infrastructure: A case study," Master Thesis, Institute of Environmental Sciences, Miami University, Oxford, Ohio, 1983.

Sullivan, P. Jr., "Maintaining aging systems - Boston's approach," Journal AWWA, Management and Operations, November, 1982.

Tsiatis, A.A., "A large sample study of Cox's regression model," Ann. Statist. 9, pp. 93-108, 1981.

U.S. Army Corps of Engineers, "Urban Water Study - Buffalo, New York, Volume I, Final Report, August 1981.

U.S. Army Corps of Engineers, New York District, "New York City Water Supply Infrastructure Study," Volume I, Manhattan, May 1980.

U.S. Department of Housing and Urban Development, "Utility Infrastructure Rehabilitation," Office of Policy Development and Research, November, 1984.

Walski, T.M., "Cost of Water Distribution System Infrastructure Rehabilitation, Repair and Replacement," Department of the Army, U.S. Army Corps of Engineers, Washington, DC, Technical Report EL-85-5, March 1985.

Walski, T.M., "Analysis of Water Distribution Systems," Published by Van Nostrand Reinhold Company Inc., New York, NY 1984.

Walski, T.M. and A. Pelliccia, "Economic analysis of water main breaks," Journal of the AWWA, Research and Technology, March, 1982.

Walski, T.M., "Selecting an Economical Water Main Rehab.," Civil Engineering, ASCE, September, 1984.

Walski, T.M., "Economic Analysis of Rehabilitation of Water Mains," Journal of the Water Resources Planning and Management Division, Proceedings of the American Society of Civil Engineers, Vol. 108, No. WR3, October, 1982.

Weiss, R., D.K. O'Day, D. Blair and P. Cairo, "Philadelphia Water Supply Infrastructure Study," paper presented at the Conference of the ASCE on Urban Water, Baltimore, MD, May 29, 1984.

Williams, J.S. and S.W. Lagakos, "Models for censored survival analysis: Constant-sum and Variable-sum Models," Biometrika, 64,2, pp. 215-24, 1977.

Zaslavsky, A., Doctoral Student, MIT, Statistics Center, Personal Communication, May 1985.

APPENDIX A

VARIABLE	EXPLANATION
1. LINK	Numeric identifier for each pipe segment in model. Total = 1391.
2. USNODE	Upstream Node: Numeric identifier of upstream node for each link.
3. DSNODE	Downstream Node: Numeric identifier of downstream node for each link.
4. DIA	Diameter: Diameter of each pipe (link) in inches. Range: 6-48 in.
5. LENGTH	Length of each pipe in model in feet. Range: 100-14,000 ft.
6. BREAKS	Total number of breaks on each pipe. Range: 0-14.
7. DATE	Year of pipe installation.
8. TYPE1	Dummy variable for pipe type (material). 1 = concrete.
9. TYPE 2	Dummy variable for pipe type. 1 = Cast Iron.
10. ORGCOST	Original cost of each pipe when installed. Range: $216.00-$820,400.00.
11. CORR	Corrosivity: Dummy variable indicator of corrosive or non-corrosive soil type.
12. S1	Dummy variable for soil stability. S1 = Unstable Soil.
13. S2	Dummy variable for soil stability. S2 = Moderately Stable Soil. (Classifications for vars. 11-13 taken from Soil Conservation Survey for New Haven).
14. PRESSURE	Absolute pressure of water traveling through each pipe in pounds per square inch. (PSI)
15. LOW	Percent of each pipe length covered by minimal land development.
16. MED%	Percent of each pipe covered by moderate land development.
17. HIGH%	Percent of each pipe covered by maximum land development.

Table A1

LIST OF VARIABLES PROVIDED FOR THE ANALYSIS OF THE NEW HAVEN SYSTEM

18.	SWAMP	Percent of each pipe covered by swamp. (Classifications for vars. 15-18 taken from U.S.G.S. Topographic Maps.)
20.	FIRST	Date of first repair for each pipe.
21. 33.	SECOND - FOURTEEN	Dates of second through fourteenth repairs if pipe experienced those repairs.
34.	YR1RPRS	Year to first repair: Number of years from date of installation to date of first repair. (var. 20 minus var. 7).
36.	YR2RPR - YR14RPR	Similar to var. 34: Number of years from previous repair to subsequent repairs.
49.	AGE83	Age of each pipe in 1983. (1983 minus var. 7).

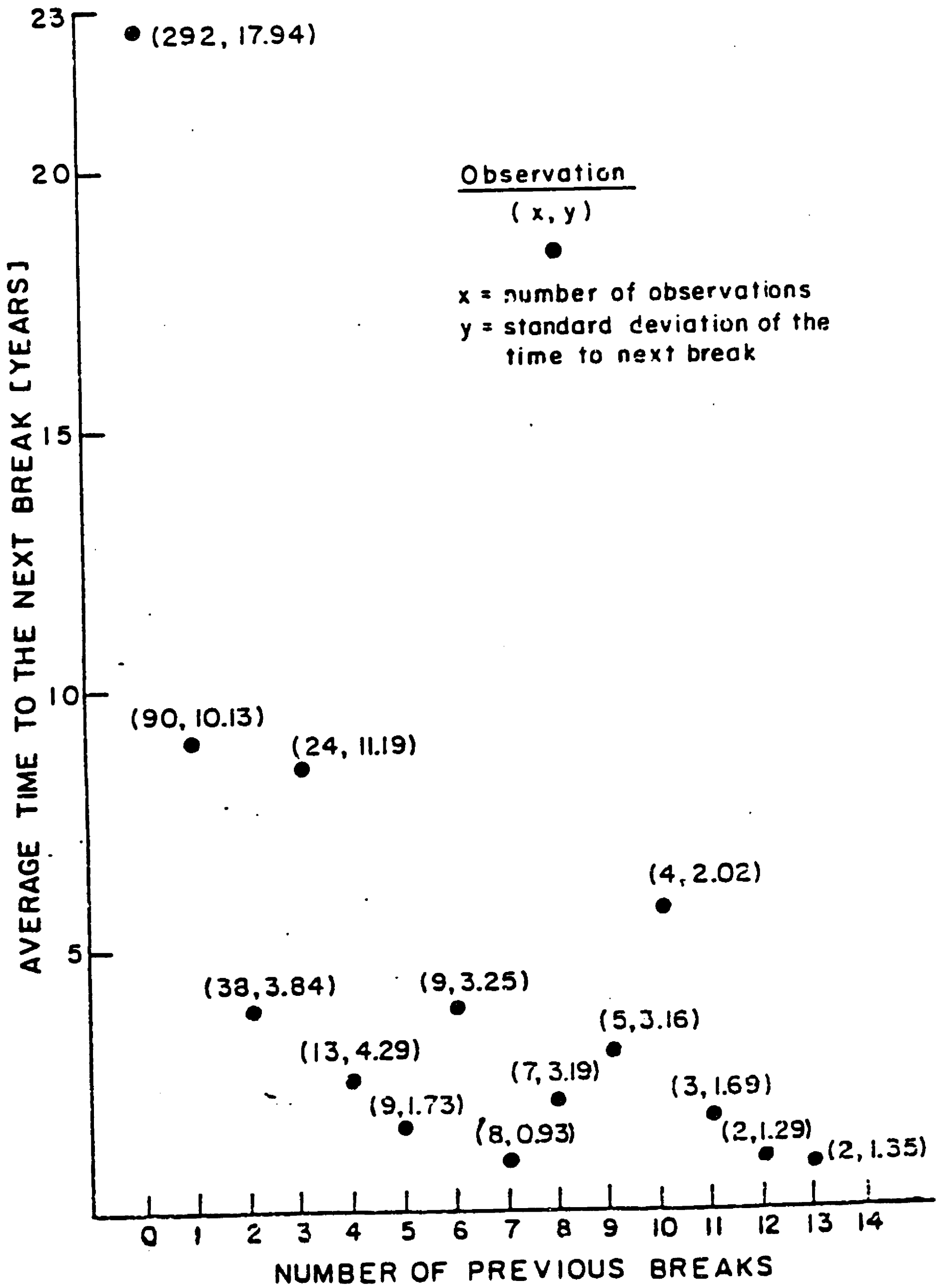


Figure A1

AVERAGE TIME TO NEXT BREAK BY NUMBER OF PREVIOUS BREAKS, NEW HAVEN

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
LENGTH	1391	1901.54	1602.94	2645050	100	14000
BREAKS	1391	0.36	1.09	506	0	14
CORR	1391	0.31	0.46	432	0	1
AGE83	1391	41.79	17.09	58138	2	83
YR1RPRS	292	22.78	17.94	6652.24	0.08	80
YR2RPR	90	9.04	10.13	813.63	0.08	43
YR3RPR	38	3.78	3.84	144.01	0.01	18
YR4RPR	24	8.85	11.19	212.55	0.01	40
YR5RPR	13	2.47	4.29	32.23	0.01	15
YR6RPR	9	1.61	1.73	14.53	0.01	5.24
YR7RPR	9	3.92	3.25	35.29	0.01	9
YR8RPR	8	0.93	0.93	7.49	0.01	3
YR9RPR	7	2.12	3.19	14.90	0.07	9
YR10RPR	5	2.81	3.16	14.09	0.17	7
YR11RPR	4	5.89	2.02	23.59	3.00	7
YR12RPR	3	1.80	1.69	5.41	0.66	3
YR13RPR	2	1.08	1.29	2.17	0.17	2
YR14RPR	2	1.04	1.35	2.08	0.08	2

Table A2

DESCRIPTIVE STATISTICS OF VARIABLES IN THE
NEW HAVEN SYSTEM

SPEARMAN BIVARIATE ANALYSIS

VARIABLE	N	MEAN	STD DEV	MEDIAN	MINIMUM	MAXIMUM
LENGTH	1391	1901.54565061	1602.94309508	1500.00000000	100.00000000	14000.00000000
BREAKS	1391	0.36376707	1.09279322	0	0	14.00000000
YR1RPRS	292	22.78164384	17.94002536	20.00000000	0.08000000	80.00000000
YR2RPR	90	9.04033333	10.13074800	4.00000000	0.08000000	43.00000000
YR3RPR	38	3.78973684	3.84543147	2.37499905	0.01000000	18.00000000
YR4RPR	24	8.85625000	11.19440264	3.51999950	0.01000000	40.00000000
AGE83	1391	41.79583034	17.09188069	49.00000000	2.00000000	83.00000000

SPEARMAN CORRELATION COEFFICIENTS / PROB > |R| UNDER H0:RHO=0 / NUMBER OF OBSERVATIONS

	LENGTH	BREAKS	YR1RPRS	YR2RPR	YR3RPR	YR4RPR	AGE83
LENGTH	1.00000	0.20046	-0.04375	0.09363	0.20387	0.10074	-0.13973
	0.0000	0.0001	0.4564	0.3801	0.2196	0.6395	0.0001
	1391	1391	292	90	38	24	1391
BREAKS	0.20046	1.00000	-0.13256	-0.07554	-0.06275	-0.42033	0.13873
	0.0001	0.0000	0.0235	0.4791	0.7082	0.0408	0.0001
	1391	1391	292	90	38	24	1391
YR1RPRS	-0.04375	-0.13256	1.00000	-0.13272	-0.10560	0.03296	0.57437
	0.4564	0.0235	0.0000	0.2124	0.5281	0.8785	0.0001
	292	292	292	90	38	24	292
YR2RPR	0.09363	-0.07554	-0.13272	1.00000	0.27700	0.27016	0.33322
	0.3801	0.4791	0.2124	0.0000	0.0922	0.2017	0.0013
	90	90	90	90	38	24	90
YR3RPR	0.20387	-0.06275	-0.10560	0.27700	1.00000	0.16030	0.11046
	0.2196	0.7082	0.5281	0.0922	0.0000	0.4543	0.5091
	38	38	38	38	38	24	38
YR4RPR	0.10074	-0.42033	0.03296	0.27016	0.16030	1.00000	0.40073
	0.6395	0.0408	0.8785	0.2017	0.4543	0.0000	0.0523
	24	24	24	24	24	24	24
AGE83	-0.13973	0.13873	0.57437	0.33322	0.11046	0.40073	1.00000
	0.0001	0.0001	0.0001	0.0013	0.5091	0.0523	0.0000
	1391	1391	292	90	38	24	1391

Table A3

SPEARMAN BIVARIATE ANALYSIS, NEW HAVEN

BIVARIATE ANALYSIS

VARIABLE	N	MEAN	STD DEV	SUM	MINIMUM	MAXIMUM
DIA	1391	14.50179727	7.51773661	20172.0000000	6.00000000	48.00000000
LENGTH	1391	1901.54565061	1602.94309508	2645050.0000000	100.00000000	14000.00000000
BREAKS	1391	0.36376707	1.09279322	506.0000000	0	14.00000000
CORR	1391	0.31056794	0.46289257	432.0000000	0	1.00000000
PRESSURE	1391	79.35011503	20.23135705	110376.0100000	4.33000000	173.24000000
AGE83	1391	41.79583034	17.09188069	58138.0000000	2.00000000	83.00000000

CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / N = 1391

	DIA	LENGTH	BREAKS	CORR	PRESSURE	AGE83
DIA	1.00000 0.0000	0.03149 0.2405	0.01840 0.4930	-0.00223 0.9338	-0.15306 0.0001	-0.07870 0.0033
LENGTH	0.03149 0.2405	1.00000 0.0000	0.16733 0.0001	0.08434 0.0016	0.05933 0.0269	-0.11669 0.0001
BREAKS	0.01840 0.4930	0.16733 0.0001	1.00000 0.0000	-0.05710 0.0332	0.03441 0.1996	0.16040 0.0001
CORR	-0.00223 0.9338	0.08434 0.0016	-0.05710 0.0332	1.00000 0.0000	-0.13397 0.0001	-0.16184 0.0001
PRESSURE	-0.15306 0.0001	0.05933 0.0269	0.03441 0.1996	-0.13397 0.0001	1.00000 0.0000	-0.04371 0.1032
AGE83	-0.07870 0.0033	-0.11669 0.0001	0.16040 0.0001	-0.16184 0.0001	-0.04371 0.1032	1.00000 0.0000

Table A4

BIVARIATE CORRELATION ANALYSIS, NEW HAVEN

CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / NUMBER OF OBSERVATIONS

	LENGTH	BREAKS	CORR	AGE83	YR1RPRS	YR2RPR	YR3RPR	YR4RPR	YR5RPR	YR6RPR	YR7RPR	YR8RPR	YR9RPR
LENGTH	1.00000 0.0000 1391	0.16733 0.0001 1391	0.08434 0.0016 1391	-0.11669 0.0001 1391	-0.02079 0.7235 292	0.04081 0.7025 90	0.09412 0.5741 38	-0.08167 0.7044 24	-0.00663 0.9829 13	-0.44816 0.2264 9	0.25478 0.5082 9	-0.32556 0.4313 8	0.30699 0.5030 7
BREAKS	0.16733 0.0001 1391	1.00000 0.0000 1391	-0.05710 0.0332 1391	0.16040 0.0001 1391	-0.11163 0.0567 292	-0.06647 0.5336 90	-0.17276 0.2996 38	-0.37807 0.0685 24	-0.22728 0.4552 13	-0.25911 0.5008 9	0.22613 0.5585 9	-0.58145 0.1306 8	-0.36076 0.4266 7
CORR	0.08434 0.0016 1391	-0.05710 0.0332 1391	1.00000 0.0000 1391	-0.16184 0.0001 1391	-0.06925 0.2381 292	-0.01161 0.9135 90	0.10097 0.5464 38	-0.23348 0.2722 24	0.14889 0.6273 13	0.32960 0.3864 9	-0.45126 0.2228 9	-0.39895 0.3276 8	-0.28449 0.5363 7
AGE83	-0.11669 0.0001 1391	0.16040 0.0001 1391	-0.16184 0.0001 1391	1.00000 0.0000 1391	0.61043 0.0001 292	0.37287 0.0003 90	0.12488 0.4551 38	0.41616 0.0431 24	0.25929 0.3923 13	0.30290 0.4282 9	0.37180 0.3245 9	0.35792 0.3840 8	0.40936 0.3618 7
YR1RPRS	-0.02079 0.7235 292	-0.11163 0.0567 292	-0.06925 0.2381 292	0.61043 0.0001 292	1.00000 0.0000 292	-0.15821 0.1364 90	-0.16024 0.3365 38	-0.02053 0.9241 24	-0.20845 0.4943 13	0.22993 0.5517 9	0.14347 0.7127 9	0.63083 0.0935 8	0.52399 0.2274 7
YR2RPR	0.04081 0.7025 90	-0.06647 0.5336 90	-0.01161 0.9135 90	0.37287 0.0003 90	-0.15821 0.1364 90	1.00000 0.0000 90	0.28113 0.0873 38	0.17656 0.4092 24	0.29113 0.3345 13	0.14736 0.7052 9	-0.21530 0.5780 9	-0.31881 0.4415 8	0.59986 0.1545 7
YR3RPR	0.09412 0.5741 38	-0.17276 0.2996 38	0.10097 0.5464 38	0.12488 0.4551 38	-0.16024 0.3365 38	0.28113 0.0873 38	1.00000 0.0000 38	-0.00663 0.9755 24	-0.15827 0.5827 13	0.28270 0.4611 9	0.40821 0.2754 9	-0.04778 0.9105 8	0.75276 0.0508 7
YR4RPR	-0.08167 0.7044 24	-0.37807 0.0685 24	-0.23348 0.2722 24	0.41616 0.0431 24	-0.02053 0.9241 24	0.17656 0.4092 24	-0.00663 0.9755 24	1.00000 0.0000 24	0.77069 0.0020 13	0.08262 0.8326 9	0.21083 0.5861 9	-0.12939 0.7601 8	-0.25342 0.5835 7
YR5RPR	-0.00663 0.9829 13	-0.22728 0.4552 13	0.14889 0.6273 13	0.25929 0.3923 13	-0.20845 0.4943 13	0.29113 0.3345 13	-0.15827 0.5827 13	0.77069 0.0020 13	1.00000 0.0000 13	-0.25434 0.5090 9	-0.11796 0.7625 9	-0.45188 0.2610 8	-0.26894 0.5598 7
YR6RPR	-0.44816 0.2264 9	-0.25911 0.5008 9	0.32960 0.3864 9	0.30290 0.4282 9	0.22993 0.5517 9	0.14736 0.7052 9	0.28270 0.4611 9	0.08262 0.8326 9	-0.25434 0.5090 9	1.00000 0.0000 9	-0.04193 0.9147 9	0.02811 0.9473 8	-0.02030 0.9656 7
YR7RPR	0.25478 0.5082 9	0.22613 0.5585 9	-0.45126 0.2228 9	0.37180 0.3245 9	0.14347 0.7127 9	-0.21530 0.5780 9	0.40821 0.2754 9	0.21083 0.5861 9	-0.11796 0.7625 9	-0.04193 0.9147 9	1.00000 0.0000 9	0.62541 0.0973 8	0.13666 0.7702 7
YR8RPR	-0.32556 0.4313 8	-0.58145 0.1306 8	-0.39895 0.3276 8	0.35792 0.3840 8	0.63083 0.0935 8	-0.31881 0.4415 8	-0.04778 0.9105 8	-0.12939 0.7601 8	-0.45188 0.2610 8	0.02811 0.9473 8	0.62541 0.0973 8	1.00000 0.0000 8	0.52442 0.2269 7
YR9RPR	0.30699 0.5030 7	-0.36076 0.4266 7	-0.28449 0.5363 7	0.40936 0.3618 7	0.52399 0.2274 7	0.59986 0.1545 7	0.75276 0.0508 7	-0.25342 0.5835 7	-0.26894 0.5598 7	-0.02030 0.9656 7	0.13666 0.7702 7	0.52442 0.2269 7	1.00000 0.0000 7

CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / NUMBER OF OBSERVATIONS

	LENGTH	BREAKS	CORR	AGE83	YR1RPRS	YR2RPR	YR3RPR	YR4RPR	YR5RPR	YR6RPR	YR7RPR	YR8RPR	YR9RPR
YR10RPR	-0.25722 0.6761 5	-0.39349 0.5122 5	0.46026 0.4354 5	0.51571 0.3738 5	0.75347 0.1414 5	0.55981 0.3264 5	-0.50800 0.3822 5	0.67245 0.2136 5	0.46085 0.4123 5	0.13092 0.8338 5	-0.70809 0.1808 5	-0.80304 0.1018 5	-0.47443 0.4194 5
YR11RPR	0.31316 0.6868 4	-0.31251 0.6875 4	0.03369 0.9663 4	0.98427 0.0157 4	0.68280 0.3172 4	0.59718 0.4028 4	0.22931 0.7707 4	0.81264 0.1874 4	0.23845 0.7616 4	0.76194 0.2381 4	-0.04981 0.9502 4	0.15373 0.8463 4	0.53544 0.4646 4
YR12RPR	-0.21328 0.8632 3	0.58437 0.6027 3	0.00000 1.0000 3	-0.99495 0.0640 3	-0.81780 0.3904 3	-0.83021 0.3764 3	-0.14343 0.9084 3	-0.99495 0.0640 3	-0.80979 0.3992 3	-0.95592 0.1897 3	0.18965 0.8785 3	-0.05345 0.9660 3	-0.48203 0.6798 3
YR13RPR	1.00000 2	0.00000 2	0.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2
YR14RPR	1.00000 2	0.00000 2	0.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2	1.00000 2

	YR10RPR	YR11RPR	YR12RPR	YR13RPR	YR14RPR
LENGTH	-0.25722 0.6761 5	0.31316 0.6868 4	-0.21328 0.8632 3	1.00000 2	1.00000 2
BREAKS	-0.39349 0.5122 5	-0.31251 0.6875 4	0.58437 0.6027 3	0.00000 2	0.00000 2
CORR	0.46026 0.4354 5	0.03369 0.9663 4	0.00000 1.0000 3	0.00000 2	0.00000 2
AGE83	0.51571 0.3738 5	0.98427 0.0157 4	-0.99495 0.0640 3	1.00000 2	1.00000 2
YR1RPRS	0.75347 0.1414 5	0.68280 0.3172 4	-0.81780 0.3904 3	1.00000 2	1.00000 2
YR2RPR	0.55981 0.3264 5	0.59718 0.4028 4	-0.83021 0.3764 3	1.00000 2	1.00000 2
YR3RPR	-0.50800 0.3822 5	0.22931 0.7707 4	-0.14343 0.9084 3	1.00000 2	1.00000 2
YR4RPR	0.67245 0.2136 5	0.81264 0.1874 4	-0.99495 0.0640 3	1.00000 2	1.00000 2

BREAKS	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	1099	1099	79.008	79.008
1	202	1301	14.522	93.530
2	52	1353	3.738	97.268
3	14	1367	1.006	98.275
4	11	1378	0.791	99.065
5	4	1382	0.288	99.353
7	1	1383	0.072	99.425
8	1	1384	0.072	99.497
9	2	1386	0.144	99.641
10	1	1387	0.072	99.712
11	1	1388	0.072	99.784
12	1	1389	0.072	39.856
14	2	1391	0.144	100.000

Table A5

FREQUENCY OF BREAK NUMBERS

APPENDIX B

Documentation of the data variables for Cincinnati area pipe file (SAS data set: CINCY.DATA[CWW])

1. PIPECODE: Arbitary numbering of all pipes staring with 1 - 801 with destinctions for the different sections of the city.

1 - 9:	Brecon
50 - 55:	California
100 - 185:	Clifton
200 - 288:	Central Service
300 - 347:	Eastern Hills
400 - 468:	Greenhills
500 - 525:	Mt. Washington
600 - 689:	Western Hills
700 - 731:	Winton Road
800 - 801:	Gravity Tunnels
>801 :	Covington and Kenton County

2. NCENT: number of the cenus tract that the pipe crosses

3. SIZEIN: size of pipe in inches

4. DATELAY: Date of pipe installation

5. PCOST78: Total cost of installation in 1978 dollars (update this value using Construction Cost Index ENR)

6. LENGTH: Length of pipe in feet

7. PRESDIF: Pressure differential between highest and lowest pressure values on a pipe

8. CL: Dummy variable for cleaning and lining 0 = NO
1 = Yes

9. CLDATE: Date of cleaning and lining if CL = 1
CLDATE = 0 If CL = 0

10. ASBPRES: Average pressure for pipe

11. YRTORP40: for Cincinnati data (CATITIL = 1)
YRTORP40 = number of years from installation date to the first repair occurring on a pipe in or after the year 1940. Repairs may have been made prior to 1940 but the data is missing. An exception is PIPE #217; for which entire repair history is known, i.e., install date = 1870, year of first break = 1886 and YRTORP40 = 16

for Covington and Kenton County (CATITIL = 2 or 3)
YRTORP40 = number of years from installation to the first repair occurring in or after the year 1968. Break data prior to 1968 is missing.

Table B1

LIST OF VARIABLES PROVIDED FOR THE ANALYSIS OF THE CINCINNATI SYSTEM

12. SUMREP40: SUM of repairs on a pipe assuming the conditions of YRTORP40

13. AGEFREP: number of years from first repair (assuming conditions of YRTORP40) to the year 1980.

14. YROFRE: year of first repair assuming conditions of YRTORP40

15. AGE: number of years from date of installation to the year 1980

16. AVINDPCT: percent of pipe under industrial land uses

17. AVCOMPCT: percent of pipe under commercial land uses

18. AVRESPCT: percent of pipe under residential land uses

19. AVTRPRCT: percent of pipe under transportation land uses

20. LENLOW: length of pipe in lowly corrosive soil

21. LENMOD: length of pipe in moderately corrosive soil

22. LEMGIGH: length of pipe in highly corrosive soil

23. AVDENTY: average population density over pipe

24. PIPETYPE: dummy variable 0 = concrete
1 = metal (cast iron, steel)

25. CATITIL: Cincinnati city data coded as 1 = CATITIL

Covington as 2

Kenton County as 3

26. CENT1-CENT8: census tracts that pipe crosses

27. CENTP1-CENTP8: percent of pipe that crosses associated census tract

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
SIZEIN	367	24.79	11.53	24.00	6.00	60.00
LGLENGTH	367	7.61	0.99	7.64	2.99	9.68
ASBPRES	367	90.27	38.05	79.00	15.00	226.00
AVRESPCT	367	37.26	19.77	36.00	0.00	78.00
AVINDPCT	367	3.47	4.76	1.00	0.00	28.00
AVCOMPCT	367	10.22	10.80	7.00	0.00	68.00
AVTRPRCT	367	18.09	11.52	15.00	4.00	57.00
AVTRPRCT	367	18.08	11.52	15.00	4.00	57.00
PCTHIGH	367	42.39	44.18	27.00	0.00	100.00
PCTLOW	367	11.17	27.22	0.00	0.00	100.00
AVDENTY	367	9.79	8.33	8.00	1.00	79.00
DATELAY	367	1930.91	27.42	1934.00	1856.00	1972.00

Table B2

DESCRIPTIVE STATISTICS OF VARIABLES IN THE CINCINNATI SYSTEM (ONLY FOR METAL PIPES)

Date Cleaned/Lined

ClDate	Frequency	Cum Freq	Percent	Cum Percent
0	395	395	87.389	87.389
1940	8	403	1.770	89.159
1945	7	410	1.549	90.708
1950	12	422	2.655	93.364
1960	12	434	2.655	96.018
1965	18	452	3.982	100.000

DESCRIPTIVE FREQUENCIES FOR PIPE DATA BY EVENT
(One record for each renewal of left-censoring)

Table of Clbefore by Prevbrks

Clbefore 1 if cleaned/lined before break PREVBRKS

Frequency Row Pct Clm Pct	0	1	2	3	4	5	6	7	8	9	Total
0	415 39.45 91.81	160 15.22 86.49	110 10.47 85.94	79 7.52 84.04	71 6.76 84.52	60 5.71 84.51	50 4.76 84.75	39 3.71 81.25	34 3.24 77.27	33 3.14 78.57	1051
1	37 23.72 8.19	25 16.03 13.51	18 11.54 14.06	15 9.62 15.96	13 8.33 15.48	11 7.05 15.49	9 5.77 15.25	9 577 18.75	10 6.41 22.73	9 5.77 21.43	156
Total	452	185	128	94	84	71	59	48	44	42	1207

Table B3

STATISTICS FOR PIPE CLEANING AND LINING

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
			Decade Pipe Laid=1850			
SIZEIN	Pipe Size (inches)	2	20.00	0.00	20.00	20.00
ASBPRES	Average Pressure	2	65.00	0.00	65.00	65.00
PCTHIGH	% in High Corrosive Soil	2	0.00	0.00	0.00	0.00
LGLENGTH	Log (Pipe Length)	2	7.98	0.64	7.52	8.43
AVRESPCT	% Under Residential Land	2	31 00	4.24	28.00	34.00
AVINDPCT	% Under Industrial Land	2	14.00	2.82	12.00	16.00
AVCOMPCT	% Under Commercial Land	2	14.00	2.82	12.00	16.00
AVTRPRCT	% Under Transportation Land	2	38.00	5.65	34.00	42.00
PERIOD1	Before 1920	2	1.00	0.00	1.00	1.00
PERIOD3	After 1940	2	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	2	28.00	5.65	24.00	32.00
DATELAY	Date of Pipe Installation	2	1856.00	0.00	1856.00	1856.00
			Decade Pipe Laid=1860			
SIZEIN	Pipe (inches)	5	24.00	13.56	16.00	48.00
ASBPRES	Average Pressure	5	63.00	10.77	48.00	76.00
PCTHIGH	% in High Corrosive Soil	5	6.59	14.75	0.00	32.98
LGLENGTH	Log (Pipe Length)	5	7.38	0.62	6.52	8.23
AVRESPCT	% Under Residential Land	5	12.40	8.20	4.00	24.00
AVINDPCT	% Under Industrial Land	5	8.00	3.80	3.00	13.00
AVCOMPCT	% Under Commercial Land	5	22.60	9.93	6.00	32.00
AVTRPRCT	% Under Transportation Land	5	45.20	7.79	37.00	54.00
PERIOD1	Before 1920	5	1.00	0.00	1.00	1.00
PERIOD3	After 1940	5	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	5	13.00	3.39	10.00	18.00
DATELAY	Date of Pipe Installation	5	1863.00	2.48	1862.00	1868.00

Table B4

DESCRIPTIVE STATISTICS OF METAL PIPES IN
THE CINCINNATI SYSTEM -
(PIPES GROUPED BY PERIOD OF INSTALLATION)

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Decade Pipe Laid=1870						
SIZEIN	Pipe Size (inches)	17	22.41	8.43	10.00	36.00
ASBPRES	Average Pressure	17	82.41	28.59	47.00	131.00
PCTHIGH	% in High Corrosive Soil	17	27.35	37.39	0.00	100.00
LGLENGTH	Log (Pipe Length)	17	7.09	1.27	2.99	8.76
AVRESPCT	% Under Residential Land	17	22.82	19.17	5.00	74.00
AVINDPCT	% Under Industrial Land	17	5.17	5.46	0.00	14.00
AVCOMPCT	% Under Commercial Land	17	10.17	9.25	1.00	29.00
AVTRPRCT	% Under Transportation Land	17	26.94	14.86	10.00	50.00
PERIOD1	Before 1920	17	1.00	0.00	1.00	1.00
PERIOD3	After 1940	17	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	17	11.52	4.97	4.00	20.00
DATELAY	Date of Pipe Installation	17	1876.58	3.29	1870.00	1879.00
Decade Pipe Laid=1880						
SIZEIN	Pipe Size (inches)	20	33.50	8.70	16.00	48.00
ASBPRES	Average Pressure	20	66.10	20.69	38.00	106.00
PCTHIGH	% in High Corrosive Soil	20	74.66	37.94	0.00	100.00
LGLENGTH	Log (Pipe Length)	20	7.27	0.69	6.32	9.05
AVRESPCT	% Under Residential Land	20	40.15	20.25	6.00	72.00
AVINDPCT	% Under Industrial Land	20	2.50	2.85	0.00	8.00
AVCOMPCT	% Under Commercial Land	20	11.80	9.42	2.00	37.00
AVTRPRCT	% Under Transportation Land	20	23.70	11.27	11.00	54.00
PERIOD1	Before 1920	20	1.00	0.00	1.00	1.00
PERIOD3	After 1940	20	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	20	15.55	6.00	3.00	29.00
DATELAY	Date of Pipe Installation	20	1880.45	1.09	1880.00	1883.00

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
	Decade Pipe Laid=1890					
SIZEIN	Pipe Size (inches)	3	30.66	9.23	20.00	36.00
ASBPRES	Average Pressure	3	74.00	33.15	51.00	112.00
PCTHIGH	% in High Corrosive Soil	3	50.65	42.82	3.01	85.97
LGLENGTH	Log (Pipe Length)	3	7.73	0.30	7.51	8.08
AVRESPCT	% Under Residential Land	3	44.00	9.16	36.00	54.00
AVINDPCT	% Under Industrial Land	3	5.00	1.73	3.00	6.00
AVCOMPCT	% Under Commercial Land	3	15.00	3.46	11.00	17.00
AVTRPRCT	% Under Transportation Land	3	29.00	4.00	25.00	33.00
PERIOD1	Before 1920	3	1.00	0.00	1.00	1.00
PERIOD3	After 1940	3	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	3	22.66	6.65	17.00	30.00
DATELAY	Date of Pipe Installation	3	1894.66	0.57	1894.00	1895.00
	Date Pipe Laid=1900					
SIZEIN	Pipe Size (inches)	28	35.92	14.02	16.00	60.00
ASBPRES	Average Pressure	28	80.80	19.23	54.00	133.00
PCTHIGH	% in High Corrosive Soil	28	37.98	47.01	0.00	100.00
LGLENGTH	Log (Pipe Length)	28	7.09	0.96	4.12	8.88
AVRESPCT	% Under Residential Land	28	33.96	21.02	4.00	48.00
AVINDPCT	% Under Industrial Land	28	3.10	3.30	0.00	14.00
AVCOMPCT	% Under Commercial Land	28	16.92	13.93	1.00	42.00
AVTRPRCT	% Under Transportation Land	28	23.89	12.17	7.00	57.00
PERIOD1	Before 1920	28	1.00	0.00	1.00	1.00
PERIOD3	After 1940	28	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	28	19.39	19.52	3.00	79.00
DATELAY	Date of Pipe Installation	28	1905.32	1.67	1902.00	1909.00

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
		Decade Pipe Laid=1910				
SIZEIN	Pipe Size (inches)	38	27.26	7.60	16.00	36.00
ASBPRES	Average Pressure	38	91.78	36.39	45.00	170.00
PCTHIGH	% in High Corrosive Soil	38	20.10	36.18	0.00	100.00
LGLENGTH	Log (Pipe Length)	38	7.70	1.01	4.78	9.37
AVRESPCT	% Under Residential Land	38	35.28	21.86	1.00	76.00
AVINDPCT	% Under Industrial Land	38	6.39	6.94	0.00	28.00
AVCOMPCT	% Under Commercial Land	38	12.36	11.71	1.00	68.00
AVTRPRCT	% Under Transportation Land	38	20.80	13.16	4.00	56.00
PERIOD1	Before 1920	38	1.00	0.00	1.00	1.00
PERIOD3	After 1940	38	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	38	9.65	4.73	2.00	20.00
DATELAY	Date of Pipe Installation	38	1913.18	2.65	1910.00	1918.00
		Decade Pipe Laid=1920				
SIZEIN	Pipe Size (inches)	28	18.14	4.77	12.00	30.00
ASBPRES	Average Pressure	28	117.85	51.05	36.00	196.00
PCTHIGH	% in High Corrosive Soil	28	45.67	41.47	0.00	100.00
LGLENGTH	Log (Pipe Length)	28	7.88	0.85	5.48	9.27
AVRESPCT	% Under Residential Land	28	41.75	19.56	5.00	67.00
AVINDPCT	% Under Industrial Land	28	4.03	4.61	0.00	14.00
AVCOMPCT	% Under Commercial Land	28	10.46	13.31	1.00	68.00
AVTRPRCT	% Under Transportation Land	28	17.14	10.46	4.00	48.00
PERIOD1	Before 1920	28	0.00	0.00	0.00	0.00
PERIOD3	After 1940	28	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	28	8.71	5.16	1.00	21.00
DATELAY	Date of Pipe Installation	28	1926.21	1.57	1924.00	1929.00

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Decade Pipe Laid=1930						
SIZEIN	Pipe Size (inches)	68	25.88	11.43	8.00	60.00
ASBPRES	Average Pressure	68	105.60	38.50	40.00	226.00
PCTHIGH	% in High Corrosive Soil	68	42.21	42.66	0.00	100.00
LGLENGTH	Log (Pipe Length)	68	7.68	0.96	5.48	9.46
AVRESPCT	% Under Residential Land	68	41.35	18.73	1.00	75.00
AVINDPCT	% Under Industrial Land	68	3.16	4.17	0.00	18.00
AVCOMPCT	% Under Commercial Land	68	8.39	8.59	0.00	45.00
AVTRPRCT	% Under Transportation Land	68	16.07	8.68	5.00	49.00
PERIOD1	Before 1920	68	0.00	0.00	0.00	0.00
PERIOD3	After 1940	68	0.00	0.00	0.00	0.00
AVDENTY	Average Pop. Density	68	9.55	5.55	2.00	27.00
DATELAY	Date of Pipe Installation	68	1933.44	2.76	1930.00	1939.00
Decade Pipe Laid=1940						
SIZEIN	Pipe Size (inches)	28	25.35	8.38	12.00	48.00
ASBPRES	Average Pressure	28	95.17	39.03	50.00	178.00
PCTHIGH	% in High Corrosive Soil	28	29.88	42.97	0.00	100.00
LGLENGTH	Log (Pipe Length)	28	7.35	1.06	4.78	9.11
AVRESPCT	% Under Resenditial Land	28	35.85	15.59	10.00	74.00
AVINDPCT	% Under Industrial Land	28	2.10	3.22	0.00	11.00
AVCOMPCT	% Under Commercial Land	28	15.00	17.07	0.00	68.00
AVTRPRCT	% Under Transportation Land	28	22.03	10.30	4.00	37.00
PERIOD1	Before 1920	28	0.00	0.00	0.00	0.00
PERIOD3	After 1940	28	1.00	0.00	1.00	1.00
ADVENTY	Average Pop. Density	28	7.60	5.29	1.00	18.00
DATELAY	Date of Pipe Installation	28	1944.57	3.09	1940.00	1949.00

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Decade Pipe Laid=1950						
SIZEIN	Pipe Size (inches)	84	23.14	12.28	6.00	48.00
ASBPRES	Average Pressure	84	81.58	36.84	15.00	190.00
PCTHIGH	% in High Corrosive Soil	84	50.52	46.39	0.00	100.00
LGLENGTH	Log (Pipe Length)	84	7.71	0.91	5.07	9.31
AVRESPCT	% Under Residential Land	84	43.45	18.99	7.00	78.00
AVINDPCT	% Under Industrial Land	84	1.90	3.69	0.00	20.00
AVCOMPCT	% Under Commercial Land	84	7.04	7.74	1.00	37.00
AVTRPRCT	% Under Transportation Land	84	12.26	5.03	4.00	30.00
PERIOD1	Before 1920	84	0.00	0.00	0.00	0.00
PERIOD3	After 1940	84	1.00	0.00	1.00	1.00
AVDNETY	Average Pop. Density	84	7.15	3.89	1.00	17.00
DATELAY	Date of Pipe Installation	84	1954.00	2.64	1950.00	1959.00
Decade Pipe Laid=1960						
SIZEIN	Pipe Size (inches)	41	18.24	10.99	8.00	48.00
ASBPRES	Average Pressure	41	84.09	34.35	41.00	164.00
PCTHIGH	% in High Corrosive Soil	41	48.87	44.29	0.00	100.00
LGLENGTH	Log (Pipe Length)	41	7.96	1.02	4.82	9.68
AVRESPCT	% Under Residential Land	41	27.12	17.09	0.00	72.00
AVINDPCT	% Under Industrial Land	41	4.04	5.73	0.00	18.00
AVCOMPCT	% Under Commercial Land	41	7.68	6.66	1.00	26.00
AVTRPRCT	% Under Transportation Land	41	13.63	12.04	4.00	49.00
PERIOD1	Before 1920	41	0.00	0.00	0.00	0.00
PERIOD3	After 1940	41	1.00	0.00	1.00	1.00
AVDNETY	Average Pop. Density	41	5.97	7.13	1.00	36.00
DATELAY	Date of Pipe Installation	41	1963.87	2.89	1960.00	1969.00

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
	Decade Pipe Laid=1970					
SIZEIN	Pipe Size (inches)	5	16.80	1.78	16.00	20.00
ASBPRES	Average Pressure	5	108.40	43.85	65.00	164.00
PCTHIGH	% in High Corrosive Soil	5	70.71	39.84	16.49	100.00
LGLENGTH	Log (Pipe Length)	5	7.20	1.36	5.07	8.60
AVRESPCT	% Under Residential Land	5	37.80	13.51	24.00	52.00
AVINDPCT	% Under Industrial Land	5	2.60	2.96	0.00	7.00
AVCOMPCT	% Under Commercial Land	5	4.60	3.57	1.00	9.00
AVTRPRCT	% Under Transportation Land	5	14.80	12.73	5.00	37.00
PERIOD1	Before 1920	5	0.00	0.00	0.00	0.00
PERIOD3	After 1940	5	1.00	0.00	1.00	1.00
AVDENTY	Average Pop. Density	5	8.80	9.03	1.00	23.00
DATELAY	Date of Pipe Installation	5	1971.40	0.89	1970.00	1972.00

SPEARMAN CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / N = 367

	SIZEIN	LGLENGTH	ASBPRES	AVRESPCT	AVINDPCT	AVCOMPCT	AVTRFRCT	AVTRFRCT	PCTHIGH	PCTLOW	AVDENTY	DATELAY
SIZEIN	1.00000	-0.23356	-0.06062	-0.01746	0.13600	0.27054	0.37107	0.37107	-0.06939	0.05483	0.32966	-0.30995
PIPE SIZE (INCHES)	0.0000	0.0001	0.2467	0.7388	0.0091	0.0001	0.0001	0.0001	0.1847	0.2948	0.0001	0.0001
LGLENGTH	-0.23356	1.00000	0.21550	-0.02060	0.02289	-0.14576	-0.30601	-0.30601	-0.00989	0.19787	-0.32075	0.20384
LOG(PIPE LENGTH)	0.0001	0.0000	0.0001	0.6940	0.6620	0.0051	0.0001	0.0001	0.8502	0.0001	0.0001	0.0001
ASBPRES	-0.06062	0.21550	1.00000	-0.03143	0.25555	-0.03481	-0.05338	-0.05338	0.03201	0.28353	-0.06116	-0.00140
AVERAGE PRESSURE	0.2467	0.0001	0.0000	0.5483	0.0001	0.5062	0.3078	0.3078	0.5410	0.0001	0.2425	0.9787
AVRESPCT	-0.01746	-0.02060	-0.03143	1.00000	-0.56661	-0.16658	-0.12789	-0.12789	0.24820	-0.27389	0.44802	0.04935
% UNDER RESIDENTIAL LAND	0.7388	0.6940	0.5483	0.0000	0.0001	0.0014	0.0142	0.0142	0.0001	0.0001	0.0001	0.3458
AVINDPCT	0.13600	0.02289	0.25555	-0.56661	1.00000	0.45956	0.33277	0.33277	-0.14280	0.33108	-0.02189	-0.22694
% UNDER INDUSTRIAL LAND	0.0091	0.6620	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001	0.0061	0.0001	0.6760	0.0001
AVCOMPCT	0.29054	-0.14576	-0.03481	-0.16658	0.45956	1.00000	0.43523	0.43523	-0.22110	0.10764	0.42833	-0.26245
% UNDER COMMERCIAL LAND	0.0001	0.0051	0.5062	0.0014	0.0001	0.0000	0.0001	0.0001	0.0001	0.0393	0.0001	0.0001
AVTRFRCT	0.37107	-0.30601	-0.05338	-0.12789	0.33277	0.43523	1.00000	1.00000	-0.17232	0.01445	0.53220	-0.44966
% UNDER TRANSPORTATION LAND	0.0001	0.0001	0.3078	0.0142	0.0001	0.0001	0.0000	0.0000	0.0009	0.7827	0.0001	0.0001
AVTRFRCT	0.37107	-0.30601	-0.05338	-0.12789	0.33277	0.43523	1.00000	1.00000	-0.17232	0.01445	0.53220	-0.44966
% UNDER TRANSPORTATION LAND	0.0001	0.0001	0.3078	0.0142	0.0001	0.0001	0.0000	0.0000	0.0009	0.7827	0.0001	0.0001
PCTHIGH	-0.06939	-0.00989	0.03201	0.24820	-0.14280	-0.22110	-0.17232	-0.17232	1.00000	-0.23510	0.09944	0.09655
% IN HIGH CORROSIVE SOIL	0.1847	0.8502	0.5410	0.0001	0.0061	0.0001	0.0009	0.0009	0.0000	0.0001	0.0570	0.0647
PCTLOW	0.05483	0.19787	0.28353	-0.27389	0.33108	0.10764	0.01445	0.01445	-0.23510	1.00000	-0.23524	-0.09371
% IN LOW CORROSIVE SOIL	0.2948	0.0001	0.0001	0.0001	0.0001	0.0393	0.7827	0.7827	0.0001	0.0000	0.0001	0.0730
AVDENTY	0.32966	-0.32075	-0.06116	0.44802	-0.02189	0.42833	0.53220	0.53220	0.09944	-0.23524	1.00000	-0.4500[illegible]
AVERAGE POP. DENSITY	0.0001	0.0001	0.2425	0.0001	0.6760	0.0001	0.0001	0.0001	0.0570	0.0001	0.0000	0.0001
DATELAY	-0.30995	0.20384	-0.00140	0.04935	-0.22694	-0.26245	-0.44966	-0.44966	0.09655	-0.09371	-0.4500[illegible]	1.00000
DATE OF PIPE INSTALLATION	0.0001	0.0001	0.9787	0.3458	0.0001	0.0001	0.0001	0.0001	0.0647	0.0730	0.0001	0.0000

Table B5

SPEARMAN BIVARIATE ANALYSIS, CINCINNATI

PEARSON CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / N = 367

	SIZEIN	LGLENGTH	ASBPRES	AVRESPCT	AVINDPCT	AVCOMPCT	AVTRPRCT	AVTRPRCT	PCTHIGH	PCTLOW	AVDENTY	DATELAY
SIZEIN	1.00000	-0.20322	-0.06464	-0.07346	0.06900	0.23448	0.27703	0.27703	-0.08542	0.07280	0.28149	-0.24813
PIPE SIZE (INCHES)	0.0000	0.0001	0.2167	0.1602	0.1872	0.0001	0.0001	0.0001	0.1023	0.1640	0.0001	0.0001
LGLENGTH	-0.20322	1.00000	0.19540	-0.03495	0.08378	-0.08990	-0.20986	-0.20986	-0.04866	0.09795	-0.22212	0.17663
LOG(PIPE LENGTH)	0.0001	0.0000	0.0002	0.5045	0.1091	0.0855	0.0001	0.0001	0.3526	0.0609	0.0001	0.0007
ASBPRES	-0.06464	0.19540	1.00000	-0.03205	0.28235	0.09468	-0.13770	-0.13770	-0.01634	0.15241	-0.10150	0.08672
AVERAGE PRESSURE	0.2167	0.0002	0.0000	0.5405	0.0001	0.0700	0.0083	0.0083	0.7551	0.0034	0.0520	0.0972
AVRESPCT	-0.07346	-0.03495	-0.03205	1.00000	-0.49689	-0.22308	-0.30990	-0.30990	0.23573	-0.29603	0.30853	0.10187
% UNDER RESIDENTIAL LAND	0.1602	0.5045	0.5405	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0512
AVINDPCT	0.06900	0.08378	0.28235	-0.49689	1.00000	0.24292	0.30349	0.30349	-0.15709	0.30447	-0.08593	-0.18010
% UNDER INDUSTRIAL LAND	0.1872	0.1091	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001	0.0025	0.0001	0.1002	0.0005
AVCOMPCT	0.23448	-0.08990	0.09468	-0.22306	0.24292	1.00000	0.27967	0.27967	-0.21966	0.17438	0.31869	-0.21250
% UNDER COMMERCIAL LAND	0.0001	0.0855	0.0700	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001	0.0006	0.0001	0.0001
AVTRPCT	0.27703	-0.20986	-0.13770	-0.30990	0.30349	0.27967	1.00000	1.00000	-0.26400	0.09497	0.33275	-0.43156
% UNDER TRANSPORTATION LAND	0.0001	0.0001	0.0083	0.0001	0.0001	0.0001	0.0000	0.0000	0.0001	0.0692	0.0001	0.0001
PCTHIGH	-0.08542	-0.04866	-0.01634	0.23573	-0.15709	-0.21966	-0.26400	-0.26400	1.00000	-0.30731	0.02085	0.08236
% IN HIGH CORROSIVE SOIL	0.1023	0.3526	0.7551	0.0001	0.0025	0.0001	0.0001	0.0001	0.0000	0.0001	0.6906	0.1152
PCTLOW	0.07280	0.09795	0.15241	-0.29603	0.30447	0.17438	0.09497	0.09497	-0.30731	1.00000	-0.16460	-0.11133
% IN LOW CORROSIVE SOIL	0.1640	0.0609	0.0034	0.0001	0.0001	0.0006	0.0692	0.0692	0.0001	0.0000	0.0016	0.0330
AVDENTY	0.28149	-0.22212	-0.10150	0.30853	-0.08593	0.31869	0.33275	0.33275	0.02085	-0.16460	1.00000	-0.37361
AVERAGE POP. DENSITY	0.0001	0.0001	0.0520	0.0001	0.1002	0.0001	0.0001	0.0001	0.6906	0.0016	0.0000	0.0001
DATELAY	-0.24813	0.17663	0.08672	0.10187	-0.18010	-0.21250	-0.43156	-0.43156	0.08236	-0.11133	-0.37361	1.00000
DATE OF PIPE INSTALLATION	0.0001	0.0007	0.0972	0.0512	0.0005	0.0001	0.0001	0.0001	0.1152	0.0330	0.0001	0.0000

Table B6

PEARSON BIVARIATE ANALYSIS, CINCINNATI

APPENDIX C

The Newton-Raphson Algorithm

The control of the Newton-Raphson Algorithm is described in detail in the BMDP Statistical Software [1983]. The most basic steps in applying this alorgithm are as follows:

The initial values for the parameter estimates are assumed to be zero, indicating that the covariates have no effect on survival. The algorithm terminates when the relative improvement in the partial likelihood function is less than the value of the "convergence" criterion or when the maximum number of "iterations" has been reached. The number used as the convergence criterion of the partial likelihood function was 0.00001. The maximum number of iterations allowed to maximize the partial likelihood function was 15.

When there is no improvement in the partial likelihood function between successive iterations, the computed parameter correction vector is halved and the partial likelihood function is recomputed. The halving process goes on until the partial likelihood function is greater than in the previous iteration, or until the maximum number of "halvings" is reached. If the maximum number of halvings is reached before any increase in partial likelihood, the computations terminate. In such a case, there is strong indication that problems exist with the data or the chosen model structure. The maximum number of halvings used was 5.

The tolerance for inversion of the information matrix is controlled by the "tolerance" limit. No change at a parameter estimate is made if it fails the tolerance test. The value used as tolerance limit was 0.00001. Failure to pass the tolerance test could indicate the presence of a singular or nearly singular information matrix, implying that possibly a problem might exist in the chosen regression equation.