

Chapter 6.

Application of Probabilistic Pipe Failure Models

6.1 Introduction

This chapter presents the application of the proportional intensity and hazards models (PIM and PHM) to a real world data set. The data set used for the models appear in Appendix D.11. They include PIPE ID, installation year, length, replacement and repair cost, and times of break since installation. The necessary MATLAB programs and SAS procedures used in the model building processes are given in Appendix D. Both the proportional intensity and hazard models (PIM and PHM) assume that the corresponding break rate functions are a product of a time dependent function and a stress dependent environmental factors. The covariates (environmental factors) are determined from the soil survey data of the case study area. The PIM is superior to the PHM and applies at a pipe level. The PHM applies for a set of bundled pipes and its predictive power is low.

6.2 Covariate Data Collection

Since information on the covariates for each pipe in the break database are not provided, an indirect way of obtaining the covariate information is pursued. As a result, the soil survey by the Soil Conservation Service, U.S. Department of Agriculture (April 1982) is used to obtain information on the characteristics of the soil and the level of land development in the region. Covariates for pipe break rate modeling should contain various information on soil characteristics surrounding a pipe and other factors affecting pipe breaks, such as hydraulic conductivity, pH, susceptibility to frost, risk of corrosion and the level of land development.

However, accurate information on these covariates at a precise location surrounding each pipe in a water distribution system is hard to obtain due to monetary and weather considerations for water utilities. Moreover, if a pipe is very long, for example, 1 mile, one has to obtain several soil samples along the pipe and the characteristics of the soil might be different along the pipe. To alleviate this problem either the length of the pipe should be redefined or the covariate values should be somehow averaged to treat the 1-mile pipe as an individual pipe. The soil survey by the Soil Conservation Service provides only limited information on the characteristics of the soil. The survey categorizes the soil in the area into 7 different soil association types. Each pipe is assumed to have the averaged value of the covariates of the soil associations for that pipe. Details on how the averaging of the covariate values are described in Section 6.5.

As described in Chapter 5, the entire pipe break database contained about 300 pipes that have complete historic break records which means breaks from installation to current year (Dec. 1997 in the database). The rest of the pipes in the database have break records only from year 1983. In other words, they have incomplete break records missing breaks that had happened before year 1983. Since the PHM with the threshold break rate considers only the last break interval (refer to Section 4.6), the number of breaks before the last break interval of a pipe is important because pipes with different break history

will have different hazard characteristics. Therefore, only the pipe break database that has *complete break records* are used in PHM and Weibull PIM analyses.

6.3 Soil Survey

The soil survey of the study area by the Soil Conservation Service divided the soil in the area into 7 soil association types. Typically, a soil association consists of one or more soil series and some minor soils. It is named for the major soil series. Each series consists of soils that have similar horizons in their profile. A soil horizon represents a layer of soil, approximately parallel to the surface, having distinct characteristics produced by soil-forming processes. A soil profile is a vertical section of the soil extending through all its horizons and into the parent material which is the consolidated organic and mineral material in which soil forms.

Table 6.1 Soil Type Information from the Case Study Area Soil Survey

Soil Association Type	Soil Series
Blake-Eudora-Waldron association	Blake (sandy loam): 43% Eudora (silty clay): 23% Waldron (silt loam): 18% Minor: 16%
Wilbur-Haymond-Elash association	Wilbur (silt loam): 46% Haymond (silt loam): 27% Elash (loam): 19% Minor: 8%
Freeburg-Ashton-Weller association	Freeburg (silty clay loam): 47% Ashton (silt loam): 38% Weller (silty clay): 6% Minor: 9%
Menfro-Winfield-Urban land association	Menfro (silty clay loam): 64% Winfield (silty clay loam): 24% Urban land: 8% Minor: 4%
Nevin-Urban land association	Nevin (silt loam): 55% Urban land: 40% Minor: 5%
Urban land-Harvester-Fishpot association	Urban land: 64% Harvester (silty clay loam): 22% Fishpot (silt loam): 6% Minor: 8%
Goss-Gasconade-Menfro association	Goss (clay): 31% Gasconade (clay): 24% Menfro (silty clay loam): 17% Minor: 28%

The soil survey provides physical and chemical properties of each of the soil series in the area in tabular form. On the other hand the soil associations in the area usually consist of three soil series. The soil survey provides percentage of each major series in a soil association. Covariate values of each soil association are obtained by weighting the covariate values according to the percentage of a soil series in a soil association. Table 6.1 shows the percentage of soil series in each soil association and the soil type (e.g. sandy loam, silty clay, silty loam, etc.) of each series. The soil survey contained the soil type of each soil series for different depths. Since pipes are buried 3 to 4 feet under the ground, soil types of each soil series are chosen at this depth. The information on soil type for each soil series is used to obtain the hydraulic conductivity of each soil series.

6.4 Description of the Covariates

For each soil series type the following information has been directly obtained from the soil survey.

- shrink-swell potential
- pH value
- potential frost action
- risk of corrosion
- liquid limit
- moist bulk density

In addition to this information, the information on land development for each soil series is indirectly obtained by using the Soil Survey. The hydraulic conductivity, HC, is obtained by using U.S. Department of Agriculture soil textural classification (Das, 1990) and ‘The Range of Values of Hydraulic Conductivity’ presented by Freeze (1979).

Shrink-swell potential, SSP, is the potential for volume change in a soil with a loss or gain in moisture. Volume change occurs mainly because of the interaction of clay minerals with water and varies with the amount and type of clay minerals in the soil. The size of the load on the soil and magnitude of the change in soil moisture content influence the amount of swelling of soils in place. If the shrink-swell potential is rated moderate to very high, shrinking and swelling can cause damage to buildings, roads, pipes and other structures.

The pH value, pH, is a measure of acidity or alkalinity and is important in determining the risk of corrosion. The pH is important in determining the risk of corrosion.

Potential frost action, PFA, is the likelihood of upward or lateral expansion of the soil caused by the formation of segregated ice lenses (frost heave) and the subsequent collapse of the soil and loss of strength on thawing. Frost action occurs when moisture moves into the freezing zone of the soil. Temperature, texture, density, permeability, content of organic matter, and depth to the water table are the most important factors considered in evaluating the potential for frost action.

Risk of corrosion, ROC, pertains to potential soil-induced electrochemical or chemical action that dissolves or weakens uncoated steel or concrete. The rate of

corrosion of uncoated steel is related to such factors as soil moisture, particle-size distribution, acidity, and electrical conductivity of the soil.

Liquid limit, LL, is that water content above which the soil is not plastic under stress and behaves like liquid. It is measured as the percentage of water content at which a groove cut in a standard liquid limit device is closed.

Moist bulk density, MBD, is the weight of soil (oven dry) per unit volume. Volume is measured when the soil is at field moisture capacity. Weight is determined after drying the soil at 105 degrees C. In the soil survey, the estimated moist bulk density of each major soil is expressed in grams per cubic centimeter of soil material that is less than 2 millimeters. The moist bulk density of a soil indicates the pore space available for water.

Hydraulic conductivity, HC, refers to the ability of a soil to transmit water or air. The estimates in the soil survey indicate the rate of downward movement of water when the soil is saturated. They are based on soil characteristics observed in the field, particularly structure, porosity, and texture.

The level of land development, LD, is considered as an important factor that affects pipe breaks. It is reasoned that high level of land development would induce more traffic to the area and, as a result, the pipes would experience more stress from the traffic loads.

Table 6.2 shows how these information are presented in the soil survey with examples and its abbreviations used as covariates in the PHM and Weibull PIM application to the break database.

Table 6.2 Example Values of the Covariates Based on the Soil Survey.

Information	Covariate Name Used in the PHM and the Weibull PIM	Example Values from the Soil Survey
Shrink-Swell Potential	SSP	low, moderate or high
pH	pH	5.6-7.3
Potential Frost Action	PFA	low, moderate or high
Risk of Corrosion	ROC	low, moderate or high
Liquid Limit(%)	LL	35-50
Moist Bulk Density (g/cm ³)	MBD	1.3-1.5
Hydraulic Conductivity(cm/sec)	HC	indirectly obtained
Land Development	LD	indirectly obtained

6.5 Evaluation of the Covariate Values

When varying covariate values dependent on soil type are presented in the soil survey, an averaged value is taken as the covariate value for each soil series.

Ex.: From Table 6.2 the value of pH is taken as $(5.6 + 7.3)/2 = 6.45$.

When the covariates are presented qualitatively such as low, moderate, and high in the soil survey, a numeric value is assumed for each of the qualitative measure and is assigned to each soil series, that is: low = 0.1, moderate = 0.5, high = 1.

In this fashion 6 covariate values (SSP, pH, PFA, ROC, LL, and MBD) for each soil series are directly obtained from the soil survey and are presented in Table 6.3. The second column in Table 6.3 shows the fraction of each soil series in a soil association. The rows labeled as 'Final' indicate the final values taken as covariates by weighting the

covariate values of each series according to the fraction of each soil series in a soil association.

Ex.: The shrink-swell potential (SSP) for Freeburg-Ashton-Weller (FAW) association =
 $0.5(\text{SSP for Freeburg series}) * 0.47$ (fraction of Freeburg series in FAW)
 $+ 0.1(\text{SSP for Ashton series}) * 0.38$ (fraction of Ashton series in FAW)
 $+ 1(\text{SSP for Weller series}) * 0.06$ (fraction of Freeburg series in FAW)
 $+ 0.533(\text{SSP for Minor series}) * 0.09$ (fraction of Minor series in FAW)
= 0.381.

Since no covariate information on the soil series 'Minor' is provided in the soil survey, an averaged covariate value of the major soil series covariate values are taken as a covariate value for the minor series.

Ex. SSP for Minor series in Freeburg-Ashton-Weller Association = $(0.5(\text{SSP for Freeburg}) + 0.1(\text{SSP for Ashton}) + 1(\text{SSP for Weller}))/3 = 0.533$

The land development (LD) covariate is obtained in the following way. Since the soil survey reports the surface paved as 'Urban land', the covariate 'land development' is assumed to have a value of 1 if the whole area of a soil association consists of 'Urban land'. However, none of the soil associations consists entirely of 'Urban land'. Therefore, the fraction of the 'Urban land' in an association is taken as the value of the land development covariate.

Ex.: The land development covariate of Menfro-Winfield-Urban land association (0.08) = The fraction of 'Urban land' in Menfro-Winfield-Urban land association = 0.08.

When an association does not mention any 'Urban land' fraction in its composition, the land development covariate is assumed to take one-third of the minor series fraction to consider at least some portion of land development in an association. This is done because the soil survey indicates that the 'Minor' category does include some urban land.

Ex.: The land development covariate of Wilbur-Haymon-Elash association = 0.08 (The fraction of Minor in series in Wilbur-Haymon-Elash association) / 3 = 0.0267

The value of the hydraulic conductivity, HC, of each soil association is obtained in the following manner. First, by using the soil type of each soil series the fraction of sand, silt, and clay in each soil type is obtained from the U.S. Department of Agriculture textural classification (Das, 1990). The composition of sand, silt, and clay in each soil type is assumed to be located approximately in the middle of each soil type.

Ex. Composition of sand, silt, and clay in the soil type silty loam are 0.2, 0.7, 0.1 respectively.

Second, the values of the hydraulic conductivity of sand, silt, and clay are assumed to take approximately middle values in the 'Range of Values of Hydraulic Conductivity and Permeability' provided by Freeze (1979).

Third, since each soil type consists of sand, silt, and clay, the values of the hydraulic conductivity are weighted according to the fraction of sand, silt, and clay in each soil type.

Ex.: HC for 'Silt Loam' = 0.2 (Sand in Silt Loam) * 0.01 (HC for Sand) + 0.7 (Silt in Silt Loam) * 0.00001 (HC for Silt) + 0.1 (Clay in Silt Loam) * 0.00000001 (HC for Clay) = 0.0020

Table 6.3 Covariate Values for Soil Associations.

Blake-eudora-waldron	percentage	SSP	pH	PFA	ROC	LL	MBD	LD
Blake	0.43	0.1	7.9	1	1	0.35	1.325	
Eudora	0.23	0.1	7.5	1	0.1	0.175	1.425	
Waldron	0.18	1	7.9	1	1	0.525	1.525	
Minor	0.16	0.4	7.76667	1	0.7	0.35	1.425	
Final		0.31000	7.78667	1.00000	0.74500	0.34125	1.40000	0.05000
Wilbur-haymond-elash	percentage	SSP	pH	PFA	ROC	LL	MBD	LD
Wilbur	0.46	0.1	6.7	1	0.5	0.3	1.375	
Haymond	0.27	0.1	6.45	1	0.1	0.315	1.375	
Elash	0.19	0.1	6.45	0.5	0.1	0.25	1.6	
Minor	0.08	0.1	6.53333	0.83333	0.23333	0.28833	1.45	
Final		0.10000	6.57167	0.89167	0.29467	0.29362	1.42375	0.02667
Freeburg-ashton-weller	percentage	SSP	pH	PFA	ROC	LL	MBD	LD
Freeburg	0.47	0.5	5.55	1	0.5	0.375	1.45	
Ashton	0.38	0.1	6.45	0.5	0.1	0.3	1.4	
Weller	0.06	1	5.4	1	1	0.5	1.475	
Minor	0.09	0.53333	5.8	0.83333	0.53333	0.39167	1.44167	
Final		0.3810	5.9055	0.7950	0.3810	0.3555	1.4318	0.0300
Menfro-winfield-urban	percentage	SSP	pH	PFA	ROC	LL	MBD	LD
Menfro	0.64	0.4	6.3	1	0.1	0.4	1.425	
Winfield	0.24	0.4	6	1	0.5	0.35	1.4	
Urban	0.08	0.4	6	0.9	0.5	0.35	1.4	
Minor	0.04	0.4	6.1	0.96667	0.36667	0.36667	1.40833	
Final		0.4000	6.1960	0.9907	0.2387	0.3827	1.4163	0.0800
Nevin-urban land	percentage	SSP	pH	PFA	ROC	LL	MBD	LD
Nevin	0.55	0.5	6.95	1	1	0.45	1.425	
Urban	0.4	0.5	6.95	1	1	0.45	1.425	
Dummy	0	0	0	0	0	0	0	
Minor	0.05	0.5	6.95	1	1	0.45	1.425	
Final		0.5	6.95	1	1	0.45	1.425	0.4
Urban-harvester-fishpot	percentage	SSP	pH	PFA	ROC	LL	MBD	LD
Urban	0.64	0.5	6.45	0.9	0.1	0.4	1.425	
Harvester	0.22	0.5	6.45	1	0.1	0.4	1.425	
Fishpot	0.06	0.5	6.45	1	0.5	0.375	1.375	
Minor	0.08	0.5	6.45	0.96667	0.23333	0.39167	1.40833	
Final		0.5000	6.4500	0.9333	0.1347	0.3978	1.4207	0.6400
Goss-gasconade-menfro	percentage	SSP	pH	PFA	ROC	LL	MBD	LD
Goss	0.31	0.5	5.9	0.5	0.5	0.6	1.4	
Gasconade	0.24	0.5	6.95	0.5	1	0.6	1.575	
Menfro	0.17	0.4	6.3	1	0.1	0.4	1.425	
Minor	0.28	0.46667	6.38333	0.66667	0.53333	0.53333	1.46667	
Final		0.4737	6.3553	0.6317	0.5613	0.5473	1.4649	0.0933

The values of the hydraulic conductivity for minor soil series are assumed to take about the average of the other soil type's values. Table 6.4 shows the values of the hydraulic conductivity for each soil type obtained through these processes. Since each soil series corresponds to a soil type, the hydraulic conductivity for each soil association

are obtained by weighting the values of them according to the fractions of each soil series in an association. Table 6.5 shows the values of the hydraulic conductivity for each soil association obtained through the weighting calculations.

Table 6.4 Hydraulic Conductivity by Soil Type.

	HC(cm/sec)	Silt Loam	Sandy Loam	Clay	Loam	Silty Clay	Silty Clay Loam	Minor
sand	0.01	0.2	0.55	0.2	0.4	0.1	0.1	
silt	0.00001	0.7	0.4	0.2	0.4	0.45	0.6	
clay	0.00000001	0.1	0.05	0.6	0.2	0.45	0.3	
Final HC		0.0020	0.0055	0.0020	0.0040	0.0010	0.0010	0.0035

Table 6.5 Values of the Hydraulic Conductivity by Soil Association.

Association	Hydraulic Conductivity(cm/sec)
Blake-Eudora-Waldron	0.00352
Wilbur-Haymond-Elash	0.00251
Freeburg-Ashton-Weller	0.00161
Menfro-Winfield-Urban land	0.00111
Nevin-Urban land	0.00208
Urban land-Harvester-Fishpot	0.00213
Goss-Gasconade-Menfro	0.00225

By combining Table 6.3 and Table 6.5 we have a total of eight covariate values (HC, SSP, pH, PFA, ROC, LL, MBD, and LD) for each soil association.

6.6 Assignment of Covariate Values to Pipes

The pipe break database contains the location information of each pipe in the system in the form of 'GRID'. Table 6.6 shows a portion of the database with 'GRID' information. Table 6.7 explains the meaning of the fields in the database.

Figure 6.1 shows the area represented in terms of the GRID map. A grid in the map represents 1250 square feet area. To assign a soil association for each pipe in the database, soil map of the case study area and Figure 6.1 are superimposed and each GRID is matched with the soil association. If a pipe crosses multiple soil associations, it is assumed that the pipe takes covariate values for the soil association that takes the greatest portion along the whole stretch of the pipe. Once each GRID is assigned to a soil association a MATLAB program is used to assign the covariate values for each pipe in the database. The MATLAB program developed for these purposes is presented in Appendix D.3.

Table 6.6 Sample Pipe Break Database with GRID Information.

PIPEID	Inst._Yr (year)	Length (ft.)	GRID	icpf (\$/ft.)	rcpb (\$)	MTB(1)	MTB(2)
01/04/29-1928-CI-6	1928	1985	21K	92.77	2814	744	
01/04/29-1929-CI-6	1929	1985	21K	92.77	2814	637	696
01/21/28-1928-CI-4	1928	1546	12J	92.77	2362	469	724
01/21/28-1928-CI-6	1928	600	12J	92.77	2814	687	
01/21/28-1928-CI-8	1928	1246	13J	96.95	3985	799	800
02/11/26-1926-CI-4	1926	452	20K	92.77	2362	608	846
02/21/29-1929-CI-12	1929	5499	08E	116.05	7753	481	577
03/02/33-1933-CI-12	1933	250	21L	116.05	7753	449	
03/06/28-1928-CI-12	1928	3494	11F	116.05	7753	471	481
03/06/28-1928-CI-6	1928	588	11F	92.77	2814	762	
03/07/29-1929-CI-6	1929	1797	15J	92.77	2814	457	576
03/24/28-1928-CI-8	1928	1323	12E	96.95	3985	601	793
03/26/32-1932-CI-12	1932	326	26J	116.05	7753	380	
04/09/32-1932-CI-12	1932	73	14K	116.05	7753	392	
04/14/32-1932-CI-12	1932	487	14K	116.05	7753	480	768
04/30/29-1929-CI-6	1929	1242	18J	92.77	2814	542	554
05/01/29-1929-CI-12	1929	1609	08E	116.05	7753	577	674
05/01/29-1929-CI-6	1929	326	14J	92.77	2814	696	805
05/04/28-1924-CI-6	1924	440	25H	92.77	2814	866	
06/01/22-1922-CI-12	1922	34750	16J	116.05	7753	502	535
06/06/24-1924-CI-6	1924	1240	20J	92.77	2814	680	708
06/14/29-1929-CI-6	1929	816	17H	92.77	2814	427	427
06/17/27-1927-CI-6	1927	649	27H	92.77	2814	584	621
06/25/28-1928-CI-12	1928	936	20O	116.05	7753	547	624
07/08/29-1929-CI-6	1929	800	19K	92.77	2814	637	

Inst_Yr = installation year (year)
 icpf = installation cost per foot (\$/ft)

MTB = months to break from installation (months)
 rcpb = repair cost per break

Table 6.7 Explanation of the Fields in the Database.

Field	Meaning
inst_yr	Installation Year
rcpb	Repair Cost per Break(\$/break)
icpf	Installation Cost per Foot(\$/ft)
MTB	Months to Break from Installation

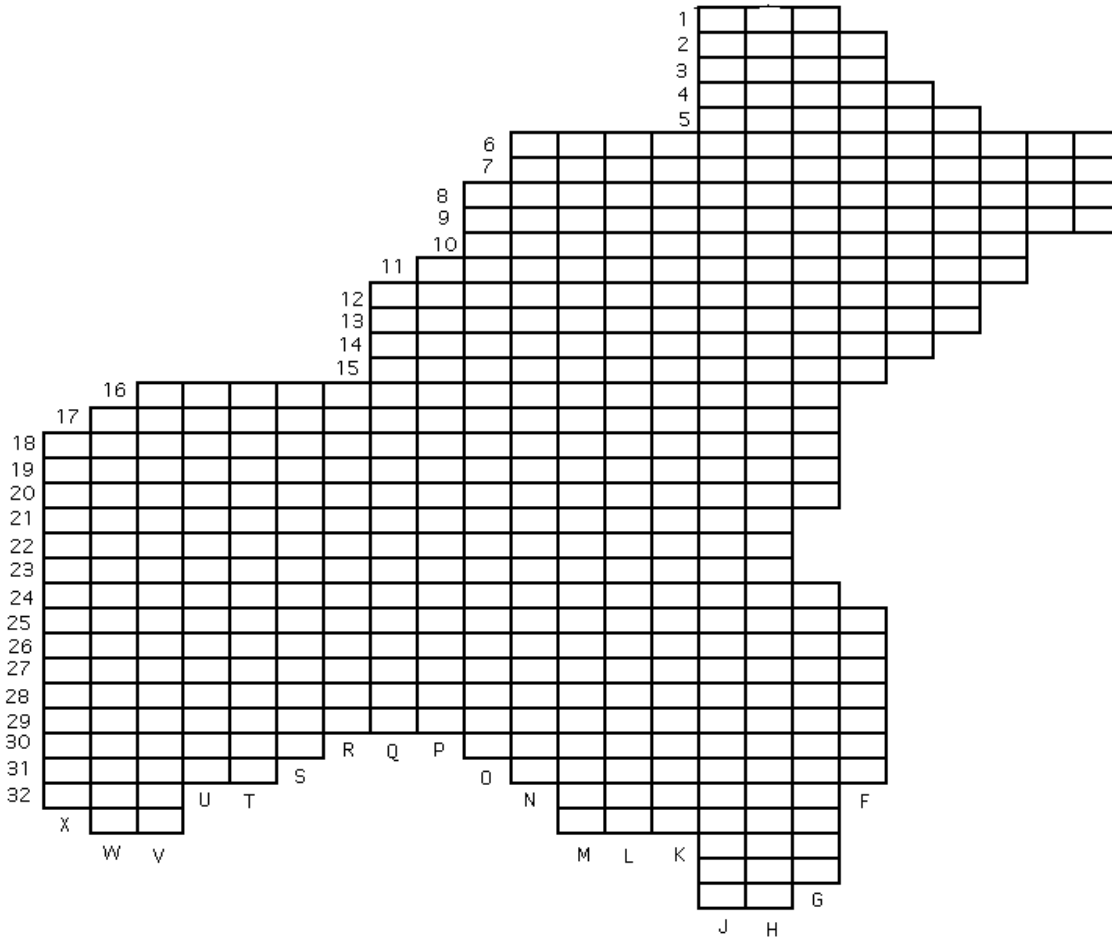


Figure 6.1 The GRID Map of the Study Area.

6.7 Application of the Weibull PIM

The Weibull PIM is constructed by the following steps.

Step 1. Obtain break times of each pipe in the system (the break database is given in Appendix D.11).

Step 2. Obtain the information on the covariates for each pipe in the system.

Step 3. Construct the Likelihood function for each pipe and obtain the values of the Weibull coefficients (\mathbf{b} and \mathbf{h}) and the covariate coefficients (\mathbf{a}) that maximize the Likelihood function:

$$l = n \log\left(\frac{\beta}{\eta}\right) + (\beta - 1) \sum_{j=1}^n \log\left(\frac{t_j}{\eta}\right) + n\alpha z - \exp(\alpha z) \left(\frac{t_0}{\eta}\right)^\beta$$

where t_0 is a censoring time, t_i is a recorded break time and n is the total number of breaks of a pipe. Censoring time is defined as the time when failure time recording is terminated.

Step 4. Construct the Weibull proportional intensity model for pipe i estimated from Step

$$3: \hat{\lambda}_i(t) = \hat{\lambda}_{0i}(t) \exp(\hat{\alpha}_i z_i) = \left(\frac{\hat{\beta}_i}{\hat{\eta}_i} \right) \left(\frac{t}{\hat{\eta}_i}\right)^{\hat{\beta}_i - 1} \exp(\hat{\alpha}_i z_i) \text{ where, } i = 1, 2, \dots, k \text{ and}$$

k is the number of pipes in the system.

Step 5. The optimal replacement time t^* is obtained by equating the Weibull PIM with the estimated coefficients and the threshold break rate as:

$$\hat{\lambda}_i(t) = \left(\frac{\hat{\beta}_i}{\hat{\eta}_i} \right) \left(\frac{t}{\hat{\eta}_i}\right)^{\hat{\beta}_i - 1} \exp(\hat{\alpha}_i z_i) = \text{BrkTh}_i.$$

Therefore, the optimal replacement time of pipe i , t_i^* , is obtained as:

$$t_i^* = \left(\text{BrkTh}_i * \exp(-\hat{\alpha}_i z_i) \frac{\hat{\eta}_i^{\hat{\beta}_i}}{\hat{\beta}_i} \right)^{\frac{1}{\hat{\beta}_i - 1}}. \quad (6.1)$$

The theory and much more detailed estimation methods of the Weibull PIM are described in Chapter 4 and Appendix A. Step 1 and Step 2 are done by the MATLAB program listed in Appendix D.3 by using the break database shown in Appendix D.11. A portion of the results of the program is given in Appendix D.12. Step 3, 4, and 5 are done by a MATLAB program listed in Appendix D.7 by using Appendix D.12 as input. The results of the program are given in Appendix D.8. The program constructs and estimates the likelihood function in Step 3. It also estimates optimal replacement time of a pipe by using Eq. (6.2). A numerical example is given at the end of this section. In addition the programs plot the empirical failure rates taken as the reciprocal of the inter break times, and the continuous failure rate which is the estimated Weibull PIM for each pipe.

Since at least three break records are required to build the Weibull PIM, the pipes with three or more breaks are used in the Weibull PIM building. As mentioned earlier, pipes with two or less breaks from their installation are not likely to be considered as

replacement candidates in near future. Therefore, considering the pipes with three or more breaks in the analysis would be sufficient for practical purposes. As a result, about 100 pipes are chosen for the analysis out of about 300 pipes in the database. Table 6.8 shows an example of the estimates of the coefficients and the corresponding optimal replacement times for all pipes. The complete results for the 100 pipes are listed in Appendix D.8. The plots of the Weibull PIM for the first five PIPE IDs in Table 6.8 are presented from Figure 6.2 to Figure 6.6.

Table 6.8 Sample Results of the Optimal Replacement Analysis Using the Weibull PIM.

PIPEID	beta	eta	HC	SSP	pH	PFA	LD	ROC	LL	MBD	Inst_YR	Length	icpf	rcpb	Brk_th	t*
05/01/29-1929-CI-12	54.10	8.02	1.37E-03	1.07E-03	-1.03E-04	4.85E-03	1.22E-03	6.86E-04	-1.39E-04	1.10E-03	1929	1609	116.05	7753	1.66	2005
08/12/27-1927-CI-12	50.58	7.79	1.27E-03	6.83E-04	-4.60E-05	8.17E-03	2.28E-03	2.20E-05	-7.75E-04	1.13E-03	1927	9614	116.05	7753	9.77	2020
10/09/26-1926-CI-4	57.85	6.34	1.65E-03	1.14E-03	-1.18E-03	8.57E-03	3.26E-03	1.03E-04	-7.50E-05	2.45E-03	1926	433	92.77	2362	1.18	2016
10094-1950-CI-6	29.18	2.77	1.20E-03	9.71E-04	-1.01E-03	1.35E-03	1.08E-03	1.01E-03	-1.26E-03	1.05E-03	1950	220	92.77	2814	0.52	2026
3027-1939-CI-6	39.35	5.81	1.55E-03	6.65E-04	-3.33E-04	9.30E-04	1.26E-03	5.50E-05	-7.31E-04	9.97E-04	1939	962	92.77	2814	2.18	2008
01/04/29-1929-CI-6	55.20	7.23	8.93E-04	9.90E-04	-2.85E-04	4.06E-04	1.34E-03	1.10E-03	-6.54E-04	2.88E-04	1929	1985	92.77	2814	4.46	2026
01/21/28-1928-CI-4	53.13	5.78	3.82E-04	6.30E-04	-1.10E-04	8.33E-03	8.08E-04	1.10E-03	-1.56E-03	3.10E-05	1928	1546	92.77	2362	4.14	2041
02/21/29-1929-CI-12	42.16	4.66	2.20E-05	7.61E-04	-7.58E-04	2.70E-03	8.97E-04	9.79E-04	-1.05E-03	1.52E-03	1929	5499	116.05	7753	5.60	2052
03/07/29-1929-CI-6	45.32	5.70	8.79E-04	1.71E-03	-1.61E-03	4.39E-04	9.01E-04	8.02E-04	-2.90E-05	-4.19E-04	1929	1797	92.77	2814	4.04	2024
04/30/29-1929-CI-6	41.42	5.41	3.07E-04	1.46E-03	-8.52E-04	1.79E-03	6.48E-04	1.32E-03	-1.10E-03	1.20E-03	1929	1242	92.77	2814	2.80	2012
06/01/22-1922-CI-12	40.74	2.57	8.22E-04	1.17E-03	-8.98E-04	2.08E-03	8.95E-04	9.85E-04	-1.04E-03	1.17E-03	1922	34750	116.05	7753	35.23	4223
06/06/24-1924-CI-6	59.65	8.30	1.11E-03	2.64E-03	-3.50E-03	1.98E-02	2.54E-03	6.83E-04	-3.92E-03	-1.43E-03	1924	1240	92.77	2814	2.80	2014
06/14/29-1929-CI-6	24.72	3.01	1.15E-03	1.39E-03	-9.77E-04	2.94E-03	1.19E-03	3.02E-04	-1.49E-03	1.29E-03	1929	816	92.77	2814	1.85	2025
06/17/27-1927-CI-6	46.56	4.59	1.16E-03	4.86E-04	-8.88E-04	3.14E-03	8.59E-04	7.25E-04	-6.68E-04	5.72E-04	1927	649	92.77	2814	1.48	2026
07/17/29-1929-CI-6	35.49	3.44	1.18E-03	3.57E-04	-1.83E-03	2.17E-03	1.12E-03	8.30E-04	-5.43E-04	1.33E-03	1929	800	92.77	2814	1.82	2046
08/18/33-1933-CI-6	45.35	4.48	2.15E-03	9.99E-04	-9.80E-05	1.03E-03	1.05E-03	4.44E-04	-2.00E-06	1.15E-03	1933	981	92.77	2814	2.22	2044
09/30/30-1930-CI-6	52.15	7.27	2.97E-04	2.24E-04	-2.49E-03	2.87E-03	7.15E-04	5.66E-04	-7.80E-04	7.94E-04	1930	1852	92.77	2814	4.16	2019
1034-1936-CI-6	44.97	3.47	6.44E-04	1.08E-03	-9.17E-04	1.60E-03	1.10E-03	1.01E-03	-9.08E-04	7.62E-04	1936	1752	92.77	2814	3.94	2157
1060-1936-CI-12	46.94	4.96	5.76E-04	1.16E-03	-6.91E-04	2.71E-03	1.02E-03	9.78E-04	-5.30E-04	4.61E-04	1936	1808	116.05	7753	1.86	2033
11/14/30-1930-CI-6	55.81	5.57	1.19E-03	6.75E-04	-3.34E-04	1.67E-03	1.72E-03	1.11E-03	-9.30E-05	4.42E-04	1930	550	92.77	2814	1.26	2027

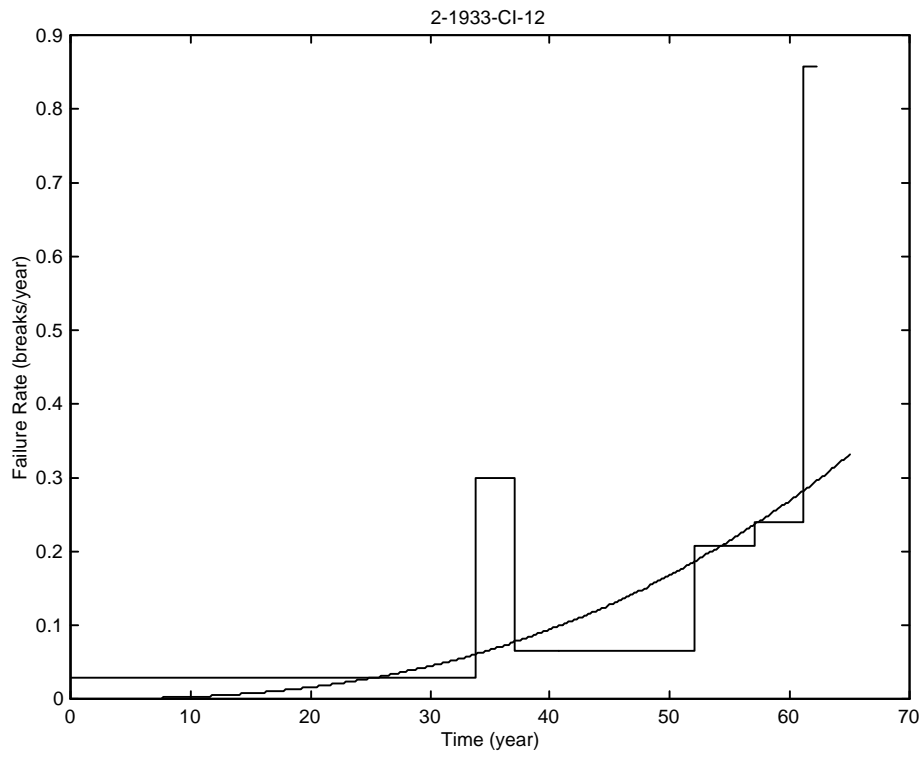


Figure 6.2 Failure Rate Plots of Pipe ID 2-1933-CI-12

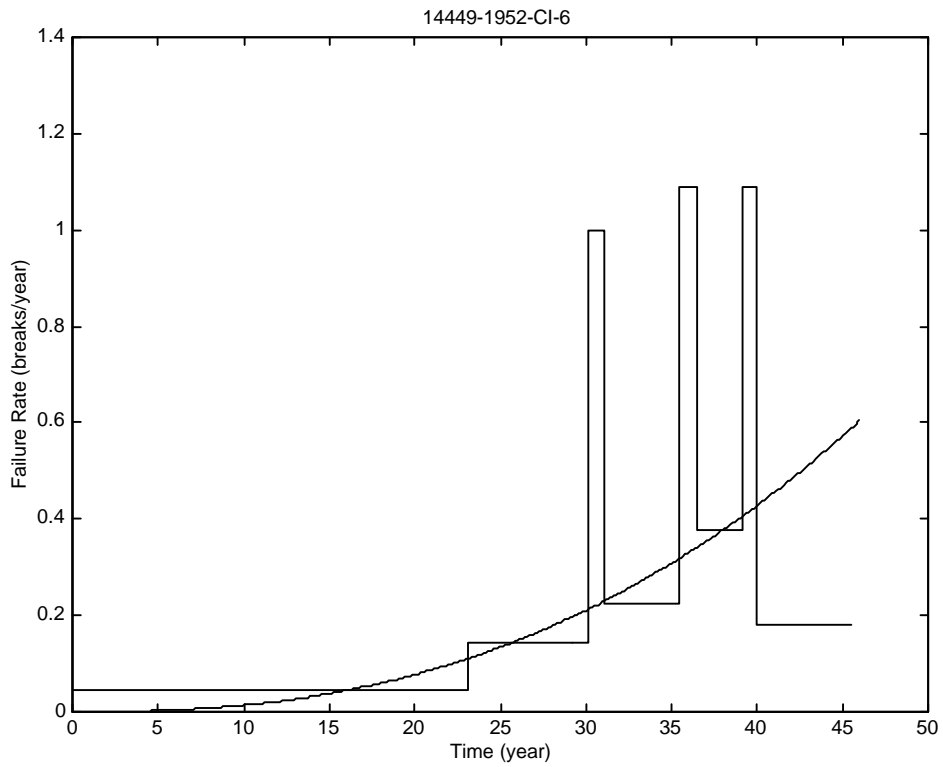


Figure 6.3 Failure Rate Plots of Pipe ID 14449-1952-CI-6

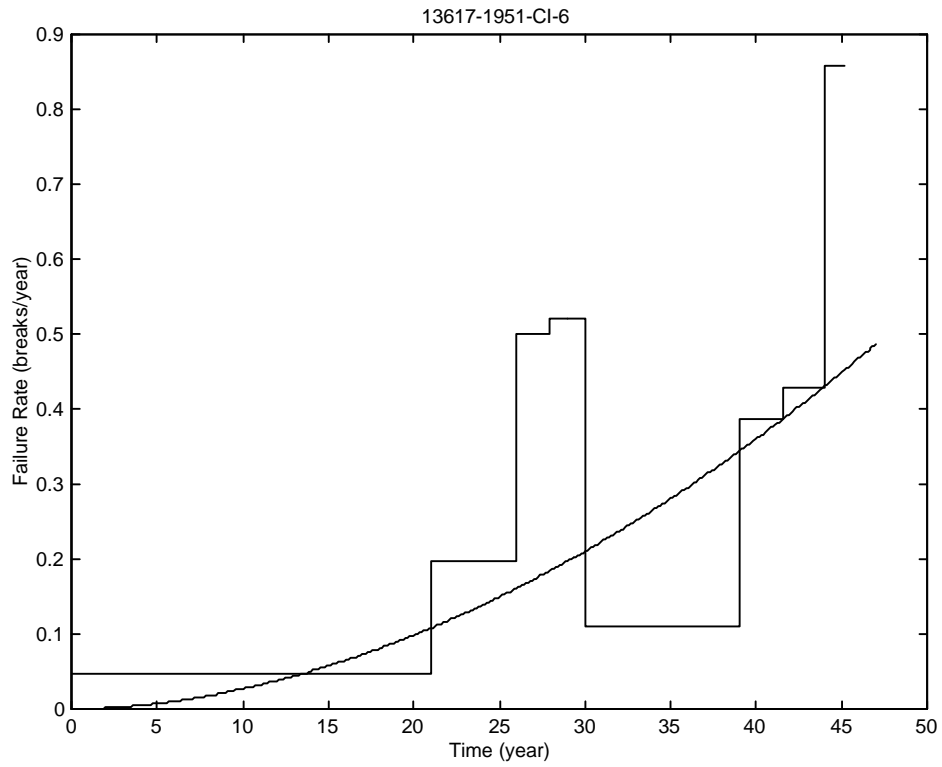


Figure 6.4 Failure Rate Plots of Pipe ID 13617-1951-CI-6

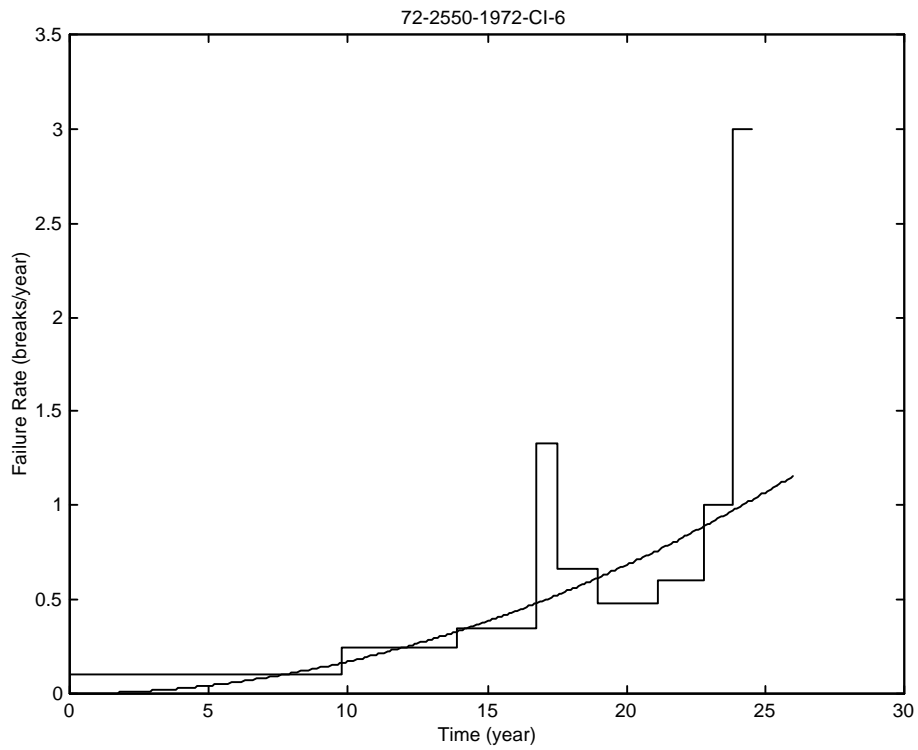


Figure 6.5 Failure Rate Plots of Pipe ID 72-2550-1972-CI-6

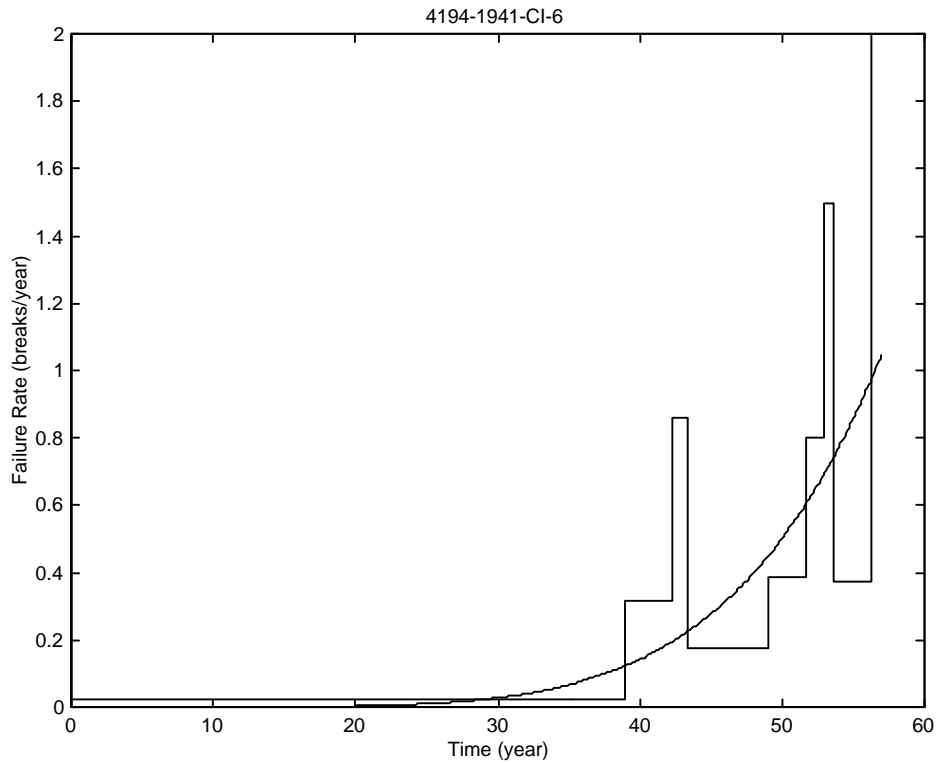


Figure 6.6 Failure Rate Plots of Pipe ID 4194-1941-CI-6

To illustrate how the Weibull PIM is established and the optimal replacement time is obtained for a pipe (PIPE ID), an example is shown below.

Example

By using the recorded historic break times and the covariate information obtain the Weibull PIM and the optimal replacement times of the pipe (PIPE ID) 14449-1952-CI-6. Use the interest rate as 7% per year.

Solution

The information on the properties of the pipe, 14449-1952-CI-6, obtained from the break database are as follows:

PIPE ID	Inst_Yr (year)	Length (ft)	icpf (\$/ft)	rcpb (\$)	Brk_th (breaks/year)
14449-1952-CI-6	1952	1363.5	92.77	2814	3.08

The recorded historic break times of the pipe, 14449-1952-CI-6 obtained from the break database are as follows;

MTB(1)	MTB(2)	MTB(3)	MTB(4)	MTB(5)	MTB(6)	MTB(7)	MTB(8)
277	361	373	426	437	469	480	546

The number in parentheses indicates the order of the break incidents. The covariate values of the pipe, 14449-1952-CI-6, are obtained by using the MATLAB program from Appendix D.3 and are given below.

PIPE ID	HC	SSP	pH	PFA	ROC	LL	MBD	LD
14449-1952-CI-6	0.0011	0.4	6.196	0.9907	0.2387	0.3827	1.4163	0.08

The meanings of the covariates are shown in the table below.

Covariate Name	Meaning
SSP	Shrink-Swell Potential (-)
pH	pH (-)
PFA	Potential Frost Action (-)
ROC	Risk of Corrosion (-)
LL	Liquid Limit (%)
MBD	Moist Bulk Density (g/cm ³)
HC	Hydraulic Conductivity (cm/sec)
LD	Land Development (-)

(Note: (-) stands for dimensionless quantity.)

By using the Weibull PIM structure (Eq. (4.2.2) and (4.2.3)) the Weibull PIM for this pipe is established as:

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \exp(\gamma_{HC} z_{HC} + \gamma_{SSP} z_{SSP} + \gamma_{pH} z_{pH} + \gamma_{PFA} z_{PFA} + \gamma_{ROC} z_{ROC} + \gamma_{LL} z_{LL} + \gamma_{MBD} z_{MBD} + \gamma_{LD} z_{LD})$$

To obtain the coefficients, the likelihood function, Eq. (4.2.4), is maximized by using the MATLAB program Appendix D.7. By using Eq. (4.2.2), (4.2.3), and (4.2.4) the likelihood function is expressed as

$$L = \prod_{j=1}^n \left[\frac{\beta}{\eta} \left(\frac{t_j}{\eta} \right)^{\beta-1} \exp(\gamma^T z) \right] \exp \left[- \sum_{j=1}^{n+1} \int_{t_{j-1}}^{t_j} \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \exp(\gamma^T z) du \right]$$

where

$$\gamma^T z = \gamma_{HC} z_{HC} + \gamma_{SSP} z_{SSP} + \gamma_{pH} z_{pH} + \gamma_{PFA} z_{PFA} + \gamma_{ROC} z_{ROC} + \gamma_{LL} z_{LL} + \gamma_{MBD} z_{MBD} + \gamma_{LD} z_{LD}$$

t_j = jth break time (MTB(j))

n = total number of breaks (8 for the pipe 14449-1952-CI-6).

The coefficients of the Weibull PIM for the pipe are obtained by using the MATLAB program in Appendix D.7 and are shown below.

PIPEID	b	h	HC	SSP	pH	PFA	LD	ROC	LL	MBD
14449-1952-CI-6	24.62	3.14	1.11E-03	7.71E-04	1.43E-03	7.74E-04	1.16E-03	7.75E-04	1.01E-03	1.06E-03

Therefore, the Weibull PIM for the pipe is expressed as

$$\lambda(t) = \frac{24.62}{3.14} \left(\frac{t}{3.14} \right)^{23.62} \exp(1.11E-3 \cdot z_{HC} + 7.71E-4 \cdot z_{SSP} + 1.43E-3 \cdot z_{pH} + 7.74E-4 \cdot z_{PFA} + 7.75E-4 \cdot z_{ROC} + 1.01E-3 \cdot z_{LL} + 1.06E-3 \cdot z_{MBD} + 1.16E-3 \cdot z_{LD})$$

The optimal replacement time of this pipe is obtained by using the covariate values, the cost and length information, and Eq. (4.3.4), that is

$$t_1^* = \eta \left(\frac{\eta \ln(1+R)e^{-\gamma^T z}}{\beta \ln\left(1 + \frac{C}{F^*L}\right)} \right)^{\frac{1}{\beta-1}}$$

$$= 3.14 \left(\frac{3.14 \ln(1+0.07)e^{-\gamma^T z}}{24.62 \ln\left(1 + \frac{2814}{92.77 * 1363.5}\right)} \right)^{\frac{1}{23.62}}$$

$$= 89$$

where

$$\gamma^T z = \gamma_{HC} z_{HC} + \gamma_{SSP} z_{SSP} + \gamma_{pH} z_{pH} + \gamma_{PFA} z_{PFA}$$

$$+ \gamma_{ROC} z_{ROC} + \gamma_{LL} z_{LL} + \gamma_{MBD} z_{MBD} + \gamma_{LD} z_{LD}$$

Since installation year is used as time '0' in this calculation, actual replacement year is expressed as

$$t^* = [\text{installation year}] + t_1^* = 1952 + 89 = 2041.$$

6.8 Application of the PHM

The procedure for building the PHM is shown below.

Step 1. Collect a set of pipes (n pipes) and obtain the last break interval of each pipe in the set (see the example at the end of this section and the break database is given in Appendix D.11).

Step 2. Obtain the information on the covariates for each pipe in the set.

Step 3. Construct the Likelihood function and obtain the values of the coefficients ($\hat{\beta}$) which maximize the Likelihood function:

$$L(\beta) = \prod_{i=1}^n \frac{\exp(bz_i)}{\sum_{k \in R(\tau_j)} \exp(bz_k)}$$

where,

$R(\tau_j)$: risk set at time t_j

n : total number of pipes considered

\underline{z}_j : vector of the values of the estimated covariates

Risk Set at time τ_j : the set of pipes which are not failed just prior to the j th failure time, τ_j

Step 4. For each recorded failure time t_j compute the cumulative baseline hazard function

$$\hat{H}_0(t) = \sum_{\tau_j < t} \frac{d_j}{\sum_{k \in R(\tau_j)} \exp(\hat{\underline{b}} \cdot \underline{z}_k)}$$

τ_j : observed and ordered failure times

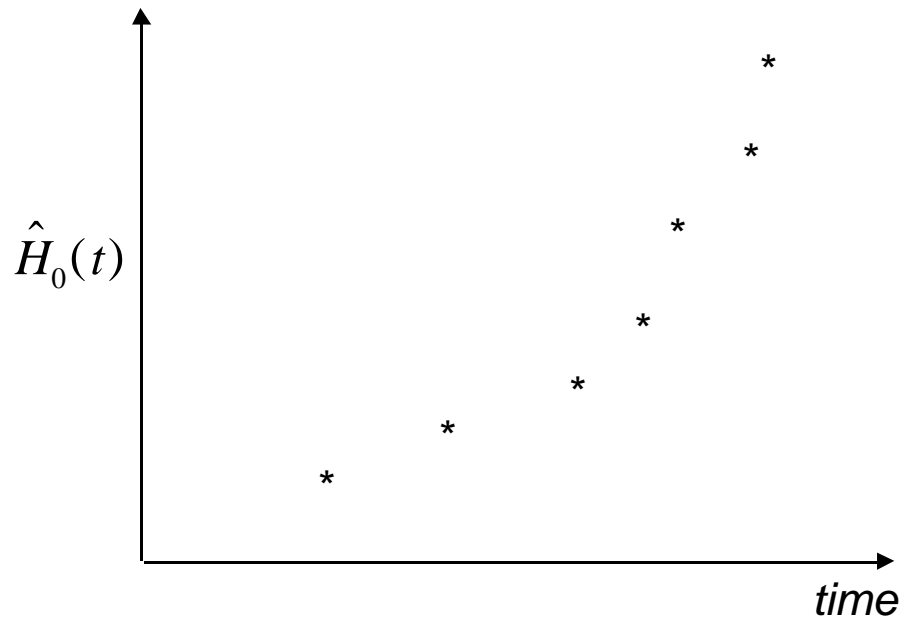
$\hat{\underline{b}}$: estimates of the coefficients of the covariates from the likelihood function

\underline{z}_k : values of the covariates for pipe k

d_j : observed number of failures at time τ_j

$R(\tau_j)$: risk set at time t_j

Step 5. Plot τ_j vs. $\hat{H}_0(t)$



Step 6. Fit an equation through the discrete data points for a continuous $\hat{H}_0(t)$

$$\hat{H}_0(t) = f(t)$$

Step 7. Take a derivative of $\hat{H}_0(t)$ with respect to t to obtain the baseline hazard function

$$\hat{h}_0(t) = \frac{d\hat{H}_0(t)}{dt}$$

The theory and much more detailed estimation methods of the PHM are described in Chapter 5 and Appendix B. From Step 3 and Step 7 the proportional hazards model for pipe i estimated from the data set is $\hat{h}_i(t) = \hat{h}_0(t) \exp(\underline{bz}_i)$ where, $i = 1, 2, \dots, k$ ($k =$ number of the pipes in the set). Therefore, the complete PHM for this application is expressed as:

$$h_i(t) = h_0(t) \exp(b_{HC} z_{HC_i} + b_{SSP} z_{SSP_i} + b_{pH} z_{pH_i} + b_{PFA} z_{PFA_i} + b_{ROC} z_{ROC_i} \\ + b_{LL} z_{LL_i} + b_{MBD} z_{MBD_i} + b_{LD} z_{LD_i})$$

in which the subscripts indicate each of the covariates as defined previously. A detailed numerical example is presented at the end of this section.

Steps 1 and 2 are accomplished by the MATLAB program listed in Appendix D.3 by using the break database shown in Appendix D.11. A portion of the results of the program is given in Appendix D.13. Step 3 to Step 5 are done by the SAS procedure 'phreg' by using Appendix D.13. The 'phreg' procedure in the SAS generates the log of baseline survival estimates expressed as 'LS' variable in outputs (refer to program listing D.5 in Appendix D). The 'MBLTB' (Months Between Last Two Breaks) variable, which is required in Step 1, is obtained from the break database. A partial listing is given as part of the numerical example.

By using $H_0(t) = -\log S_0(t)$, where $S_0(t)$ = the baseline survival function, the cumulative baseline hazard function ($H_0(t)$) is estimated and the baseline hazard function is obtained from Step 7. The baseline hazard function ($h_0(t)$) represents an aging process that goes on independently of stress or the covariates for a selected group of pipes. Since the baseline hazard function describes a common time dependent hazard for a group of the pipes, a group should consist of the pipes with similar failure characteristics. To obtain a meaningful baseline hazard function satisfying this condition groups of pipes should be categorized into one of the following categories: previous number of breaks, pipe material, or size.

Since the pipe material and the previous number of breaks are considered as the most important factors affecting the failure characteristics of the pipes, the pipes in the database are categorized into the groups with the same material and the same previous number of breaks. Here, previous number of break refers to the number of breaks previous to the most recent one. There are mainly two types of pipe materials in the database such as Cast Iron and Ductile Iron. A group of the pipes with zero number of previous breaks means that the group represents all of the pipes that have failed only once in their lifetime. If a pipe has three total breaks in its history the previous number of breaks is 2.

As mentioned earlier the pipe break database is categorized into groups with the same pipe material and the same previous number of breaks: According to this method, examples of the groups identified in the database are:

- a. Ductile Iron pipes with 0 previous number of breaks (DI-0),
- b. Cast Iron pipes with 0 previous number of breaks (CI-0),
- c. Ductile Iron pipes with 1 previous number of breaks (DI-1)
- d. Cast Iron pipes with 1 previous number of breaks (CI-1)
- e. Ductile Iron pipes with 2 previous number of breaks (DI-2)
- f. Cast Iron pipes with 2 previous number of breaks (CI-2) and so on.

The abbreviations in the parentheses will be used to denote the corresponding group for convenience.

Program listing D.4 in Appendix D shows the SAS program to obtain the coefficients of the covariates and the log of baseline survival (or cumulative baseline hazard) of a group of the pipes. Detailed estimation procedures are described in Chapter 4 and Appendix B. The contents of the program are common for all kinds of groups of the pipes and the only difference is the contents of the ‘data’ procedure. Therefore, the main program remains the same but the data consist of the last break interval represented in months and the covariates of each pipe in a group. Program listing D.4, which is for group DI-0, is shown as an example. The output of this program is listed in D.5 in Appendix D. The ‘baseline’ statement in ‘phreg’ procedure is used to obtain the estimates of the cumulative baseline hazard. By using the cumulative baseline hazard estimates, which are shown under the variable name, LS, in Appendix D.5.2., a plot for the cumulative baseline hazard for Group DI-0 is generated and is shown in Figure 6.7.

As can be seen in Figure 6.7 there is a clear hazard pattern in this group of the pipes. An exponential function is fit to the estimates and is obtained as $H_0(t) = 0.0015e^{0.021t}$ where $H_0 = -\log S_0(t) =$ cumulative baseline hazard function. Therefore, the baseline hazard function is obtained as:

$$h_0(t) = \frac{d}{dt} 0.0015e^{0.021t} = 3.15 \times 10^{-5} e^{0.021t} .$$

Likewise, plots of the estimates of the cumulative baseline hazard for other groups are shown from Figure 6.8 to Figure 6.12.

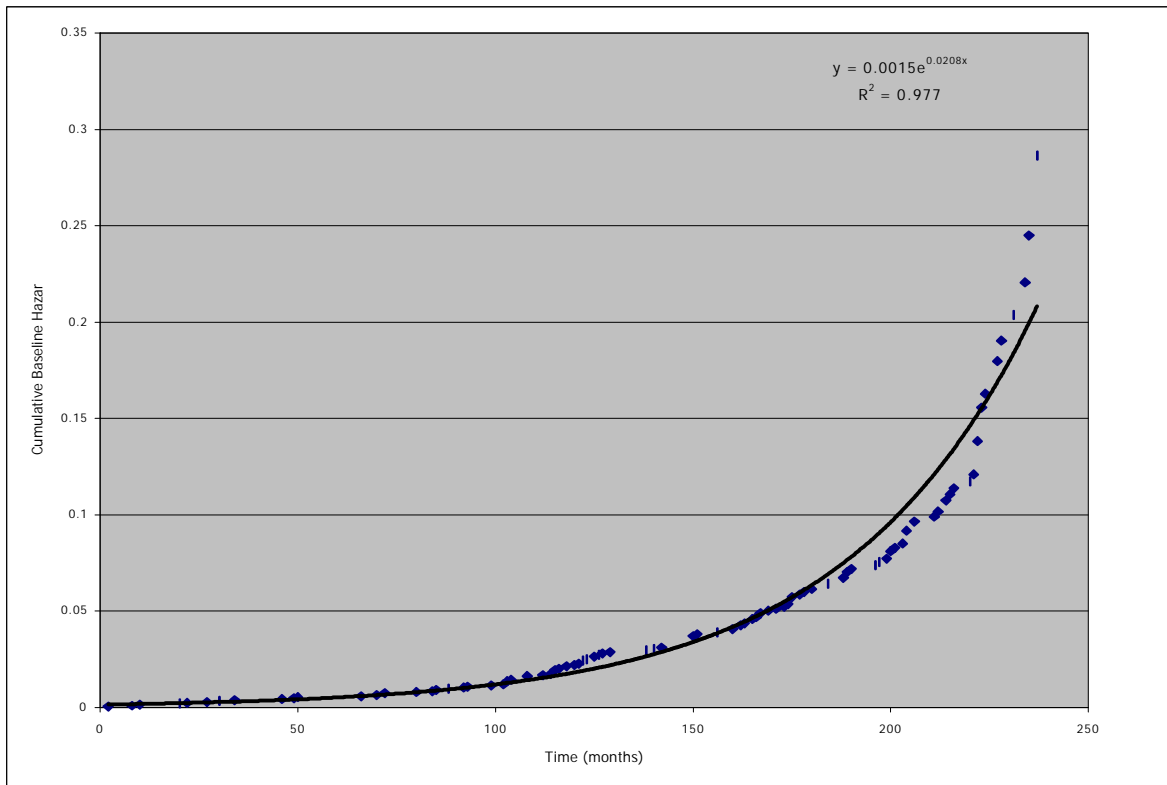


Figure 6.7 Cumulative Baseline Hazard for DI-0 group

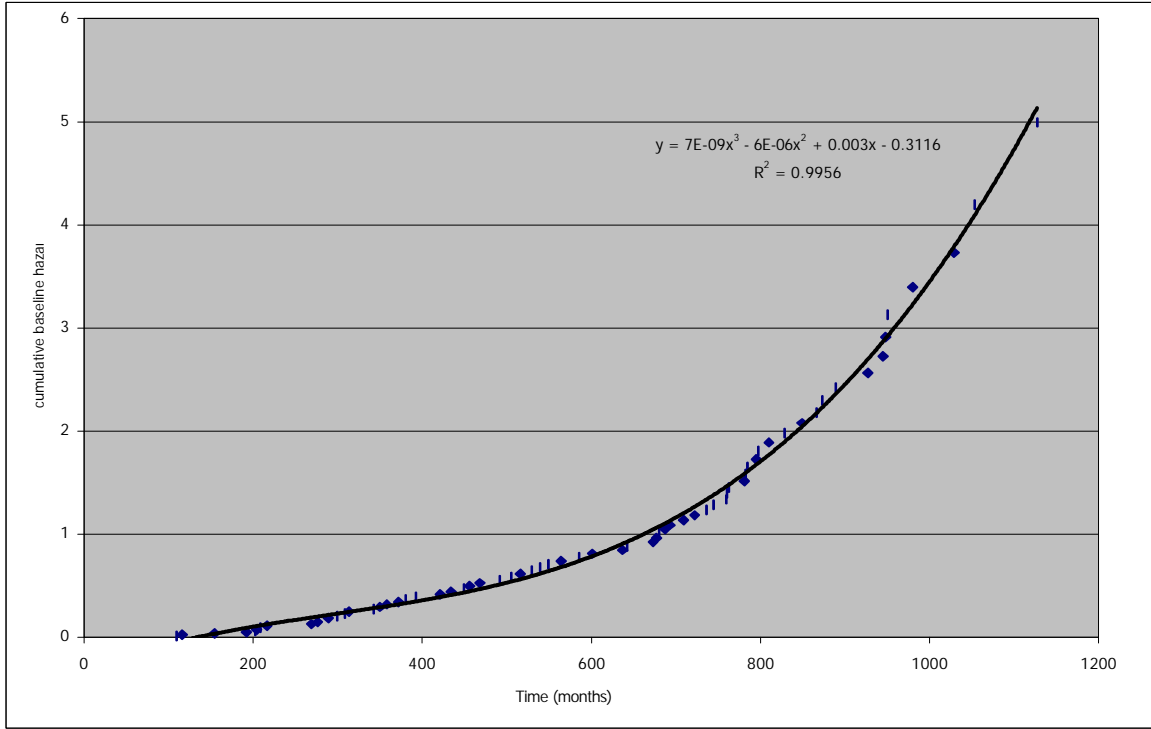


Figure 6.8 Cumulative Baseline Hazard for CI-0 group

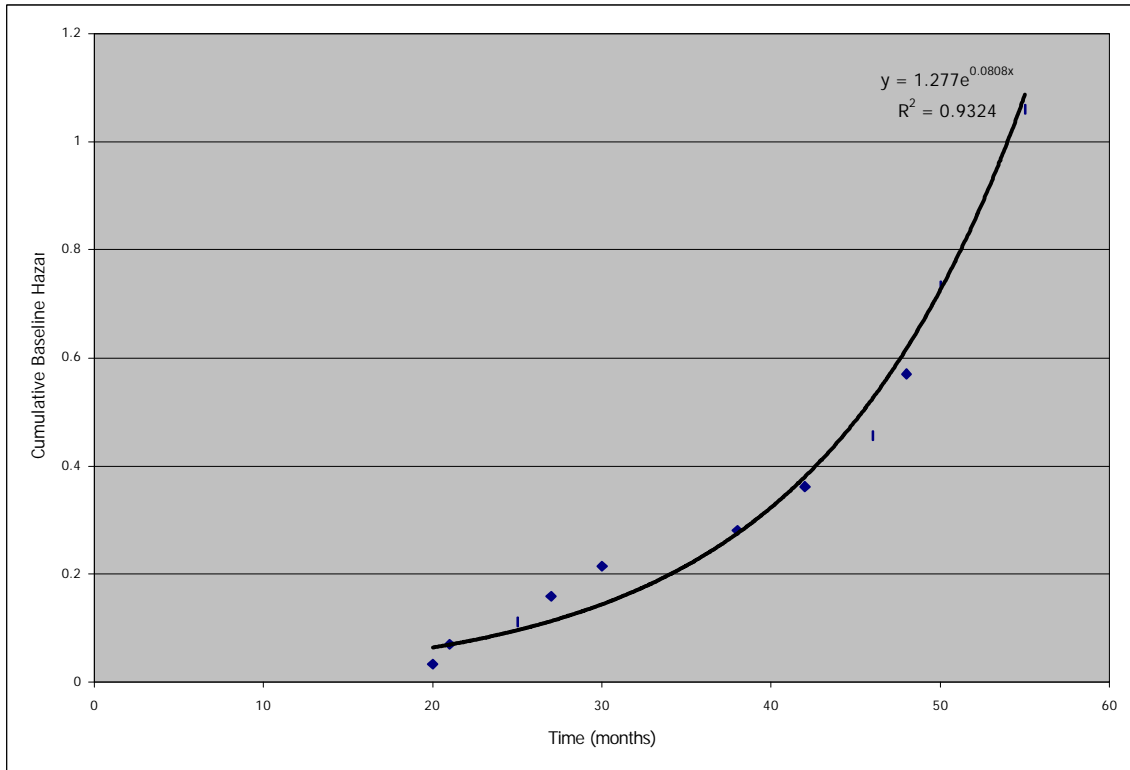


Figure 6.9 Cumulative Baseline Hazard for DI-1 group

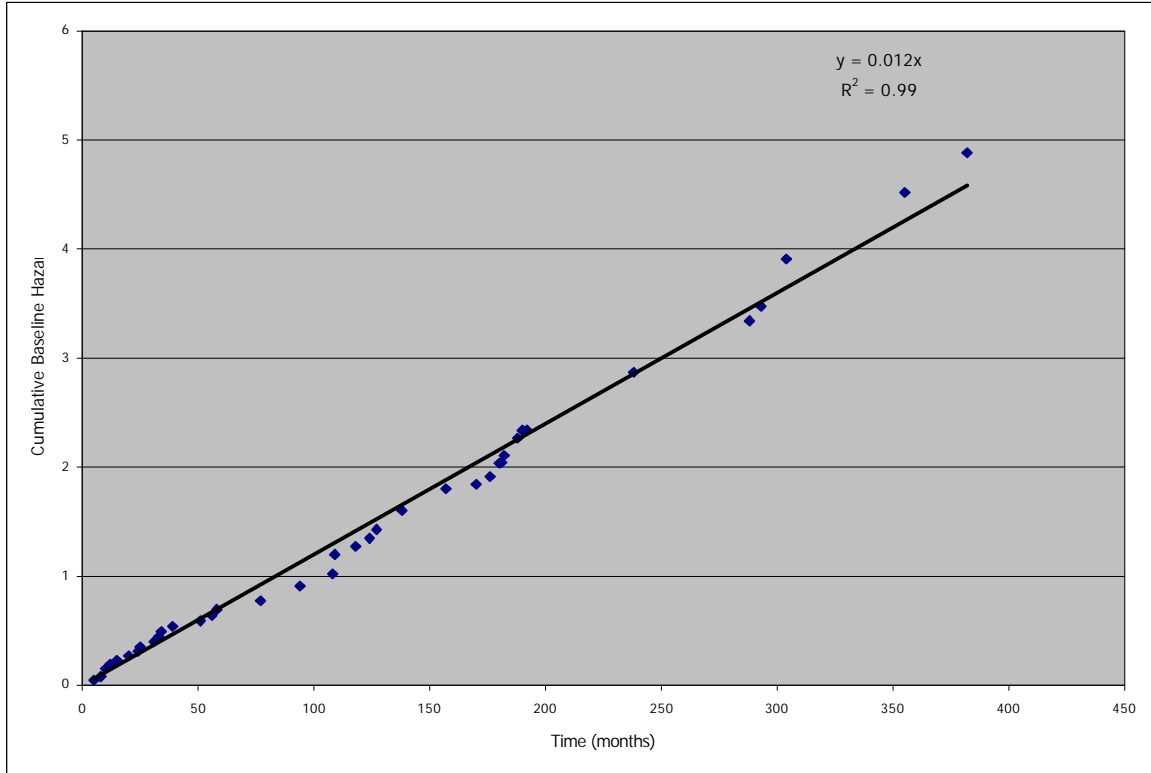


Figure 6.10 Cumulative Baseline Hazard for CI-1 group

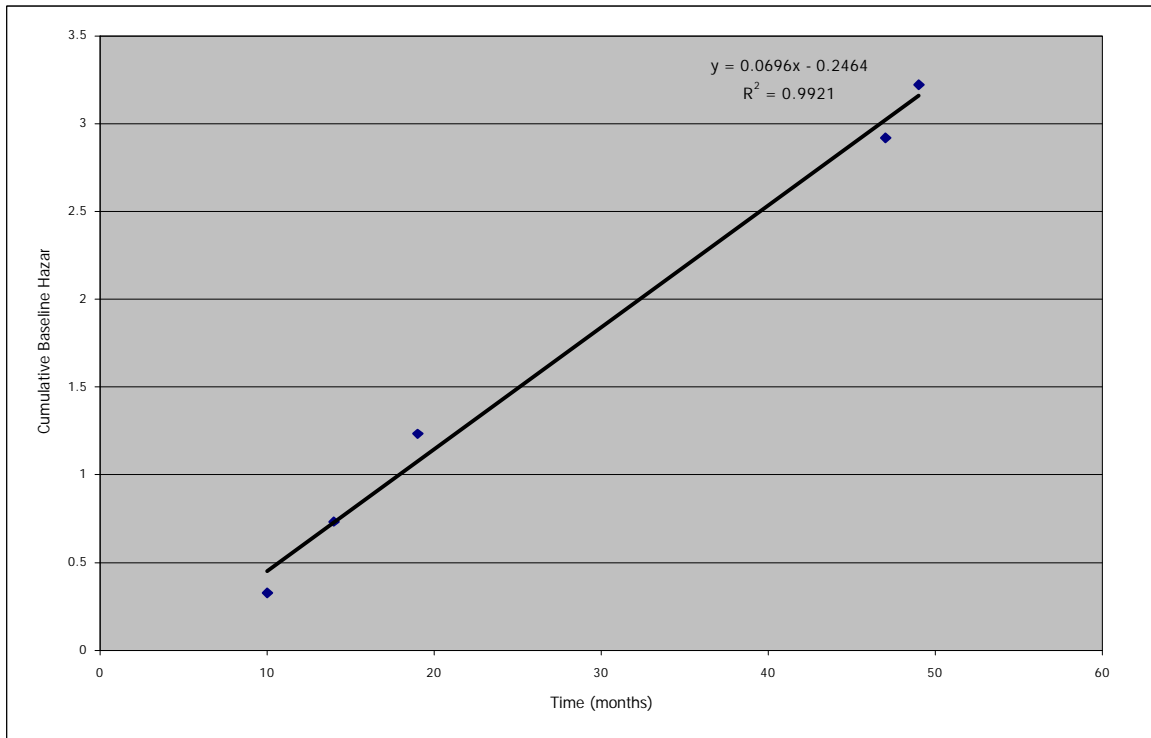


Figure 6.11 Cumulative Baseline Hazard for DI-2 group

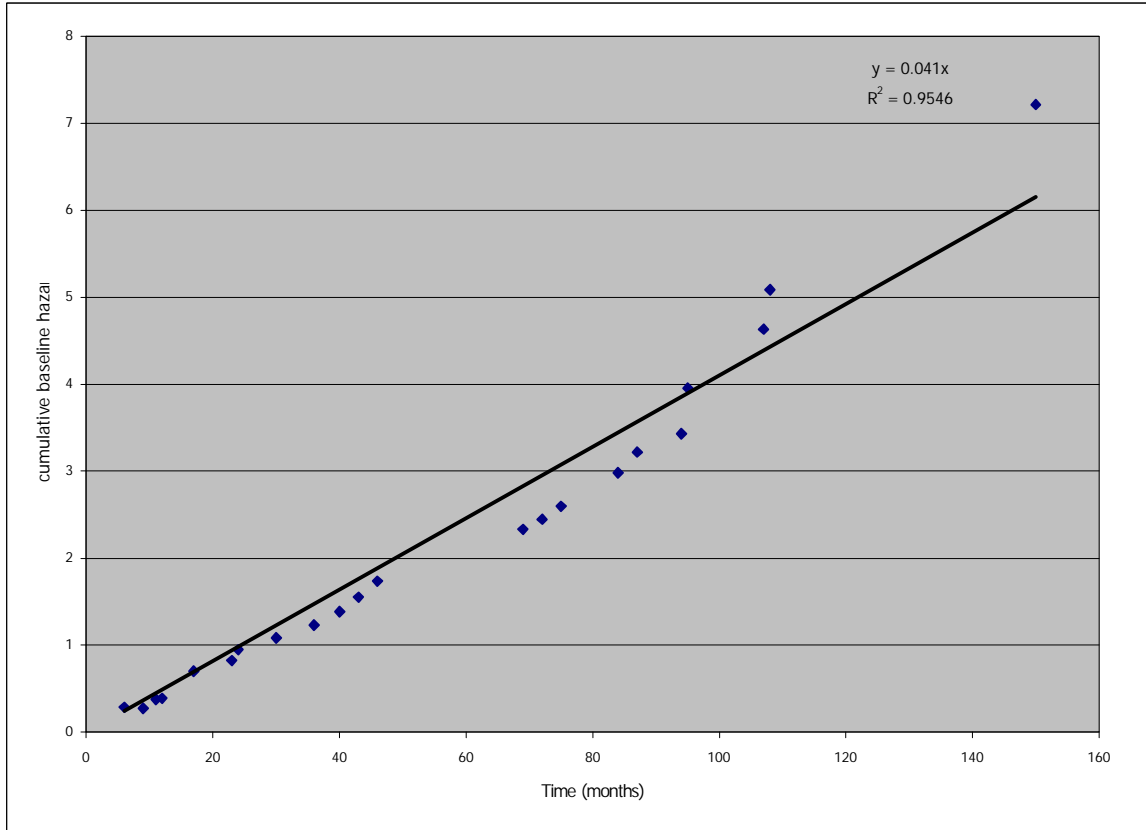


Figure 6.12 Cumulative Baseline Hazard for CI-2 group

The number of pipes with more than 2 previous number of breaks (or the total number of breaks is more than 3) has been found as too small. Therefore, different types of the pipes are lumped together to form a group of the same previous number of breaks such as group 3 and group 4 in which 3 and 4 denote previous number of breaks. Figure 6.13 and Figure 6.14 show the estimates of the cumulative baseline hazard for group 3 and group 4.

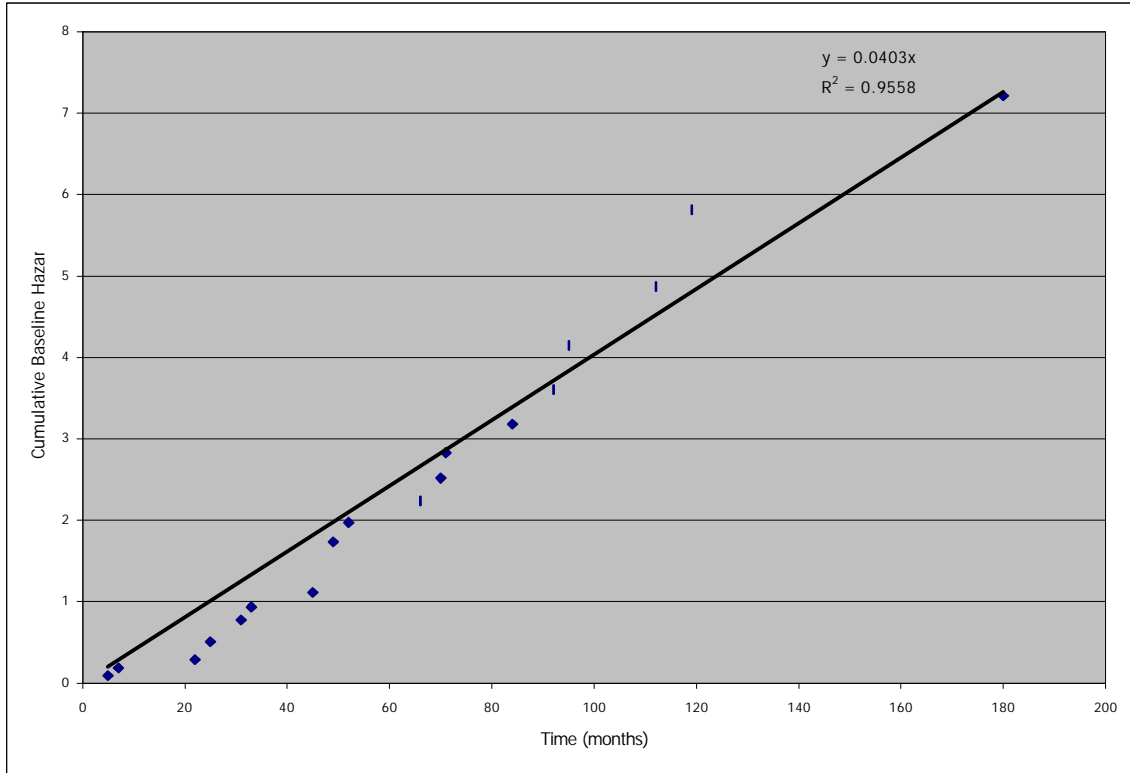


Figure 6.13 Cumulative Baseline Hazard for Group 3

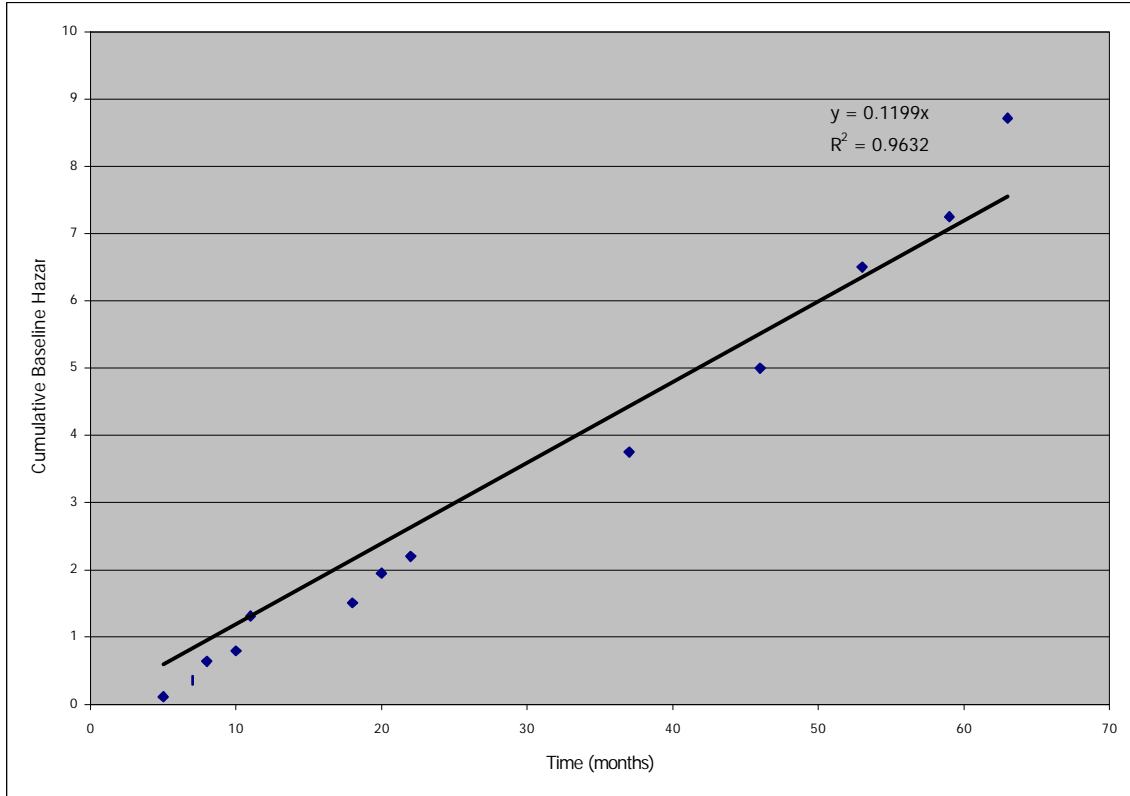


Figure 6.14 Cumulative Baseline Hazard for Group 4

As can be seen from Figure 6.10 to Figure 6.14 we find that the cumulative hazard ($H_0(t) = [-\log S_0(t)]$) function is almost linear after 2 breaks for the Ductile Iron pipes and after 1 break for the Cast Iron pipes, which leads to a constant hazard for the groups of the pipes. This finding implies that there will be no increase or decrease in hazards of breaks even if the groups of the pipes age after 2 breaks for the Ductile Iron and after 1 breaks for the Cast Iron. This conclusion drawn from a mathematical point of view is also reasonable when one considers the real pipe break phenomena. Pipes would be able to withstand all of the internal or external stresses up to some point in their lifetimes. When one of the pipes in a group breaks it would be due to either structural weakness or excessive amount of stresses (highly corrosive soil or extreme traffic load) at the precise location of the break. In the case of young (without many breaks) pipes they still have overall structural integrity and the structural integrity has a tendency to decrease over time. Some pipes in a certain location show a fast decline in their structural integrity due to the characteristics of the surrounding area and some have a slow decline in their structural integrity. This difference in the degree of the structural integrity loss over time makes the increasing hazard for a group of the pipes. When pipes become old and had some breaks in their lifetimes, overall structural integrity is used up and the internal and external stresses cause random breaks.

As a result, the probability of having a break in a group of pipes given that no break has happened at current time is the same over time leading towards a constant hazard. The critical point dividing an increasing hazard and a constant hazard is found to be 1 previous number of break for the Cast Iron pipes and 2 breaks for the Ductile Iron pipes for the site considered. The conclusion made here is based on the assumption that the Ductile Iron and the Cast Iron pipes with 3 or more breaks would also have a constant hazard, which could not be directly checked due to the data limitation. However, one can reason that they would have had the same trend considering the randomness of the pipe failure that were observed in the case of less than 3 breaks.

One should also note that a constant hazard does not mean a constant failure rate. A hazard is obtained by considering the probability of a failure for a sample of pipes given that no failures have occurred and a failure rate stands for the probability of a failure in a given period when multiple failures are occurring.

The optimal replacement analysis using the threshold break rate with a constant hazard leads to no replacement at current time and not even in the future, if the current hazard is less than the threshold break rate. Since a PHM with a constant baseline hazard such as in Eq. (6.2) does not have the time variable in its form, replacement is not needed even in the future.

$$h(t) = h_0(t) \exp(\underline{\alpha z}) = C \cdot \exp(\underline{\alpha z}) = \text{Brk}_{\text{th}} \quad (6.2)$$

where C is a constant.

Furthermore, if pipes have only one or two breaks, even though they show an increasing hazard over time (e.g. Group DI-0, DI-1, and CI-0), they are not likely to be considered as replacement candidates in near future, say in 30 years.

Therefore, although it is theoretically correct to use the PHM with the threshold break rate, it becomes almost infeasible in the actual application of the PHM with the threshold break rate to a real pipe break database due to the possibility of having a constant hazard. However, to illustrate the use of the PHM with the threshold break rate for the optimal replacement analysis, an example is shown below.

Example

By using the recorded historic break times and the covariate information obtain the proportional hazards model for Group DI-1 and the optimal replacement times of the pipes in the group. Use the interest rate as 7% per year.

Solution

The first step in the PHM building process is to obtain the last break interval of a pipe. Table below shows the break times of each pipe in Group DI-1, which are obtained from the break database.

PIPE ID	MTB(1)	MTB(2)	MBLTB
73-2342-1974-DI-6	150	175	25
78-1626-1978-DP-6	188	218	30
78-1686-1978-DP-8	183	229	46
78-1690-1978-DP-6	159	214	55
78-1803-1978-DP-6	213	234	21
78-2572-1978-DP-6	200	242	42
78-2842-1979-DI-6	186	224	38
79-1593-1979-DP-6	173	223	50
79-1726-1979-DP-6	137	185	48
79-2125-1979-DP-6	127	147	20
83-1394-1983-DP-6	90	117	27

MTB denotes months to a break from installation and MBLTB is months between last two breaks. Since DI-1 group has the pipes with total 2 breaks, MBLTB is obtained by subtracting the value of MTB(1) from MTB(2), which is shown in the third column of the table above. The last break interval and the covariate values of the pipes in Group DI-1 are obtained by using the MATLAB program Appendix D.3.

PIPE ID	MBLTB	HC	SSP	pH	PFA	ROC	LL	MBD	LD
73-2342-1974-DI-6	25	0.0021	0.5	6.45	0.9333	0.1347	0.3978	1.4207	0.64
78-1626-1978-DP-6	30	0.0016	0.381	5.9055	0.795	0.381	0.3555	1.4318	0.03
78-1686-1978-DP-8	46	0.0011	0.4	6.196	0.9907	0.2387	0.3827	1.4163	0.08
78-1690-1978-DP-6	55	0.0021	0.5	6.45	0.9333	0.1347	0.3978	1.4207	0.64
78-1803-1978-DP-6	21	0.0021	0.5	6.45	0.9333	0.1347	0.3978	1.4207	0.64
78-2572-1978-DP-6	42	0.0011	0.4	6.196	0.9907	0.2387	0.3827	1.4163	0.08
78-2842-1979-DI-6	38	0.0016	0.381	5.9055	0.795	0.381	0.3555	1.4318	0.03
79-1593-1979-DP-6	50	0.0016	0.381	5.9055	0.795	0.381	0.3555	1.4318	0.03
79-1726-1979-DP-6	48	0.0011	0.4	6.196	0.9907	0.2387	0.3827	1.4163	0.08
79-2125-1979-DP-6	20	0.0011	0.4	6.196	0.9907	0.2387	0.3827	1.4163	0.08
83-1394-1983-DP-6	27	0.0016	0.381	5.9055	0.795	0.381	0.3555	1.4318	0.03

Also, the information on the properties of the pipes in Group DI-1 are as follows:

PIPE ID	Inst_Yr (year)	Length (ft)	icpf (\$/ft)	rcpb (\$)	MTB (1) (months)	MTB (2) (months)	Brk_th (breaks/year)
73-2342-1974-DI-6	1974	260	92.77	2814	150	175	0.6131
78-1626-1978-DP-6	1978	365.5	92.77	2814	188	218	0.8486
78-1686-1978-DP-8	1978	1011	96.95	3985	183	229	1.6978
78-1690-1978-DP-6	1978	2971	92.77	2814	159	214	6.6607
78-1803-1978-DP-6	1978	837.5	92.77	2814	213	234	1.9017
78-2572-1978-DP-6	1978	713	92.77	2814	200	242	1.6240
78-2842-1979-DI-6	1979	658	92.77	2814	186	224	1.5013
79-1593-1979-DP-6	1979	758	92.77	2814	173	223	1.7243
79-1726-1979-DP-6	1979	765	92.77	2814	137	185	1.7400
79-2125-1979-DP-6	1979	893	92.77	2814	127	147	2.0255
83-1394-1983-DP-6	1983	363	92.77	2814	90	117	0.8431

Then, the coefficients of the covariates and the cumulative baseline hazard estimates are obtained by using the covariate and MBLTB values in the following SAS program.

```

data covals;
    input HC SSP pH      PFA      ROC      LL      MBD      LD;
cards;
0      0      0      0      0      0      0      0      0      0
run;

data PHM;
input MBLTB HC SSP pH      PFA      ROC      LL      MBD      LD status;
cards;
25      0.0021 0.5      6.45      0.9333 0.1347 0.3978 1.4207 0.64      1
30      0.0016 0.381    5.9055 0.795      0.381      0.3555 1.4318 0.03      1
46      0.0011 0.4      6.196      0.9907 0.2387 0.3827 1.4163 0.08      1
55      0.0021 0.5      6.45      0.9333 0.1347 0.3978 1.4207 0.64      1
21      0.0021 0.5      6.45      0.9333 0.1347 0.3978 1.4207 0.64      1
42      0.0011 0.4      6.196      0.9907 0.2387 0.3827 1.4163 0.08      1
38      0.0016 0.381    5.9055 0.795      0.381      0.3555 1.4318 0.03      1
50      0.0016 0.381    5.9055 0.795      0.381      0.3555 1.4318 0.03      1
48      0.0011 0.4      6.196      0.9907 0.2387 0.3827 1.4163 0.08      1
20      0.0011 0.4      6.196      0.9907 0.2387 0.3827 1.4163 0.08      1
27      0.0016 0.381    5.9055 0.795      0.381      0.3555 1.4318 0.03      1
run;

proc phreg data=PHM;
    model MBLTB*status(0)=HC SSP pH PFA      ROC      LL      MBD      LD;
baseline out=a covariates=covals logsurv=ls / nomean;
run;

proc print data=a;
run;

```

The 'phreg' procedure that is used in above SAS program uses the maximum partial likelihood method discussed in Chapter 4 and Appendix B to obtain the coefficients of the covariates. It also uses the 'baseline' statement to obtain the estimates of the cumulative baseline hazard. The results of the above program are obtained as follows.

Coefficients of the Covariates

The SAS System

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
HC	1	647.768397	839.22197	0.59578	0.1402
PH	1	-0.760531	1.58600	0.22995	0.1316
PFA	0	0	.	.	.
ROC	0	0	.	.	.
LL	0	0	.	.	.
MBD	0	0	.	.	.
LD	0	0	.	.	.
SSP	0	0	.	.	.

Estimates of the Cumulative Baseline Hazard

The SAS System

OBS	HC	PH	PFA	ROC	LL	MBD	LD	SSP	MBLTB	LS
1	0	0	0	0	0	0	0	0	0	0.00000
2	0	0	0	0	0	0	0	0	20	-0.03386
3	0	0	0	0	0	0	0	0	21	-0.07069
4	0	0	0	0	0	0	0	0	25	-0.11190
5	0	0	0	0	0	0	0	0	27	-0.15900
6	0	0	0	0	0	0	0	0	30	-0.21435
7	0	0	0	0	0	0	0	0	38	-0.28147
8	0	0	0	0	0	0	0	0	42	-0.36189
9	0	0	0	0	0	0	0	0	46	-0.45623
10	0	0	0	0	0	0	0	0	48	-0.57036
11	0	0	0	0	0	0	0	0	50	-0.73260
12	0	0	0	0	0	0	0	0	55	-1.06034

Therefore, by using the estimates of the cumulative baseline hazard at each break time (the values under the variable 'LS') Figure 6.9 has been generated and the cumulative baseline hazard is obtained as

$$H_0(t) = 1.277e^{0.0808t}$$

The baseline hazard function is obtained by taking the derivative of $H_0(t)$ with respect to time, t , which is expressed as

$$h_0(t) = d(1.277e^{0.0808t})/dt = 0.1032e^{0.0808t}$$

The coefficients of the covariates are obtained as $b_{HC} = 647.77$ and $b_{PH} = -0.7605$. All other coefficients are estimated as zeros. Therefore, the PHM for Group DI-1 is expressed as:

$$h_i(t) = 0.1032 \exp(0.0808t) \exp(647.77z_{HC_i} - 0.7605z_{PH_i}),$$

where $i = 1, 2, \dots, k$ ($k =$ number of the pipes in the group DI-1).

By equating the estimated hazard function with the threshold break rate of each pipe in group DI-1 we can obtain the optimal replacement time of each of the pipe. The optimal replacement time for pipe i in group DI-1, t_i^* , is expressed as:

$$t_i^* = \ln \left[\frac{\text{Brk}_{th_i} \exp(-(647.77z_{HC_i} - 0.7605z_{pH_i}))}{0.1032} \right] / 0.0808 \quad (6.3)$$

where Brk_{th_i} is the threshold break rate of pipe i .

For instance, the optimal replacement time of the PIPE ID 73-2342-1974-DI-6 in Group DI-1 is calculated by using the values of the threshold break rate and the covariate values of the pipe ($HC = 0.0021$ and $pH = 6.45$) in Eq. (6.3), that is

$$t_1^* = \ln \left[\frac{0.613 \cdot \exp(-(647.77 \times 0.0021 - 0.7605 \times 6.45))}{0.1032} \right] / 0.0808 \cong 66$$

Since time '0' in the PHM for the last break interval is the time of the start of the interval (time of (n-1)th break, where n is the total number of breaks of a pipe), actual replacement year for the PIPE ID is

$$t_{73-2342-1974-DI-6}^* = [\text{installation year}] + [\text{MTB}(1)]/12 + t_1^* = 1974 + 150/12 + 66 = \text{year 2052}$$

The optimal replacement times of other pipes in the group are obtained by using the procedure described above and a MATLAB program in Appendix D.6 is used to do the necessary calculations. The results are given in Table 6.9 as follows.

Table 6.9 Optimal Replacement Year for Group DI-1 by the PHM

PIPE ID	HC (cm/sec)	pH	Inst_Yr (year)	Length (ft)	icpf (\$/ft)	rcpb (\$)	MTB (1) (months)	MTB (2) (months)	Brk_th (breaks/year)	t*
73-2342-1974-DI-6	0.0021	6.45	1974	260	92.77	2814	150	175	0.6131	2052
78-1626-1978-DI-6	0.0016	5.9055	1978	365.5	92.77	2814	188	218	0.8486	2062
78-1686-1978-DI-8	0.0011	6.196	1978	1011	96.95	3985	183	229	1.6978	2077
78-1690-1978-DI-6	0.0021	6.45	1978	2971	92.77	2814	159	214	6.6607	2096
78-1803-1978-DI-6	0.0021	6.45	1978	837.5	92.77	2814	213	234	1.9017	2075
78-2572-1978-DI-6	0.0011	6.196	1978	713	92.77	2814	200	242	1.6240	2078
78-2842-1979-DI-6	0.0016	5.9055	1979	658	92.77	2814	186	224	1.5013	2070
79-1593-1979-DI-6	0.0016	5.9055	1979	758	92.77	2814	173	223	1.7243	2071
79-1726-1979-DI-6	0.0011	6.196	1979	765	92.77	2814	137	185	1.7400	2074
79-2125-1979-DI-6	0.0011	6.196	1979	893	92.77	2814	127	147	2.0255	2075
83-1394-1983-DI-6	0.0016	5.9055	1983	363	92.77	2814	90	117	0.8431	2059

Since all of the pipes in this group experienced only two breaks in about 20 years (from installation to the current time, Dec. 1997), the optimal replacement times are far in the future (about 50 to 100 years from year 1999).