ISSN: 2278-0181
Vol. 1 Issue 10, December- 2012

Prediction Of Water Main Failure Frequencies Using Statistical Modeling Techniques

Mrs.Reena D Popawala
Asso. Prof. in Civil Engineering Department
C.K.Pithawalla College of Engg. & Tech., Surat
Near Malvan mandir, via Magdalla port
Surat-Dumas road-395007
Surat, Gujarat,India

Dr. N C Shah
Prof. of CED & Section Head (TEP)
Sardar Vallabhbhai National Institute of Technology, Ichchanath
Surat, India

Abstract

The economic and social costs of pipe failures in water and wastewater systems are increasing, putting pressure on utility managers to develop annual replacement plans for critical pipes that balance investment with expected benefits in prediction based management context. In addition to the need for a strategy for solving such a multi-objective problem, analysts and water system managers need reliable and robust failure models for assessing network performance. An important concern for utility manager is the prediction of pipe failure frequencies of water mains. This paper presents analysis results of two models namely multiple and Poisson regression model. The model is developed using (SPSS statistic version 19) based upon 5 years historical data collected from Surat city in Gujarat state.

1. Introduction:

Each year, hundreds of kilometers of pipes worldwide are upgraded or replaced, in an attempt to mitigate the effects of pipe burst and water loss, and to maintain the uninterrupted transport of water. Existing water network are increasingly at risk due to numerous factors and the accidental or deterioration based breakage/leakage of water distribution system represents a large problems.
The driving force behind pipe replacement capital improvement projects have primarily been the mandate to safeguard the health of urban population, the need to increase the reliability of pipe networks and the service provided to people. As well as socio-economic factor in relation to the cost of operations and maintenance of piping network.

Sustainable water management system, though, should include not only new methods for monitoring, repairing or replacing aging infrastructure, but also expanded method for modeling deteriorating infrastructure conditions and proactive replace or repair strategies. The need for optimizing operating cost and network reliability is at the core of one of the most important dilemmas facing water distribution agencies: should an organization repair or replace aging and deteriorating water mains and in either case, what should the sequence of any such repairs be as part of long term network rehabilitation strategy?

2. State of Knowledge:

One of the major problems to be faced is the frequent pipe-breaks with unaccounted water leakages resulting in service disruption. Water service companies have begun to develop new leakage detection strategies in order to reduce leakages to an economical optimum level. The main objective is to propose reliable computational models to facilitate pipe replacement decisions in an effort to increase the overall reliability expected from the pipe network. An extensive amount of work on pipe rehabilitation and replacement has been published. The various algorithms developed have taken the form of non-linear, dynamic, heuristic and successive linear programming economic models, which assist decision-making based usually on historical statistics and cost information. In an early work Shamir and Howard (1979) proposed a model, which estimates the optimal time for pipe replacement based on pipe breakage history and the cost for repairing and replacing pipes. Kettler and Goulter (1985), identified a relationship between breakage rate and pipe diameter as well as a correlation between the

ISSN: 2278-0181
Vol. 1 Issue 10, December- 2012

number of pipe failures and pipe age. They proposed that improvements to pipe breakage or mechanical reliability may be achieved by selecting larger pipe diameters. Woodburn et al. (1987) presented a model for determining the minimum cost for rehabilitation, replacement or expansion of an existing network based on a combination of non-linear optimization and hydraulic simulation procedures. An explicit algorithm, implementing a graph theory approach, has been developed by Boulos and Altman (1991). The algorithm is capable of handling widespread applications, associated with future planning, expansion and improvement of fluid distribution networks. Arulraj and Rao (1995) proposed an optimality criterion called the significance index to rehabilitate existing networks. On many occasions when continuous quantities are selected as decision variables the results may be misleading. The use of statistical methods to discern patterns of historical breakage rates and use them to predict water main breaks has been widely documented. Kleiner and Rajani (2001) provided a comprehensive review of approaches and methods that had been developed prior to their review. Walski & Wade (1987) as well as Mavin (1996) also used exponential –based expression to model failure rates. However, instead of an exponential relationship between failure rate and age. Malandain et al (1999) applied a Poisson regression model to quantify the influence of the different variable namely diameter, material, and position of pipe on failure rate.

3. Multiple Linear Regression Analysis:

The preliminary analysis contains analysis of 788 random samples collected from south-west zone of surat city. The purpose of the analysis is to predict the pipe failure by using the Multiple Regression Analysis and Poission Regression Analysis. The table below shows the summery of variables included in the analysis. It contains the name of the variable, Type wheter continuous or categorical and measured scale.

Table 1: Summery of Variable

Sr. No	**Variable**	**Type**	**Measured Scale**	
			Minimum	**Maximum**
1	Number of Leakages	Continuous	0	24
2	Diameter – cm	Continuous	75	1500
3	Depth – Meter	Continuous	1	3.50
4	Type of Traffic	Categorical	1	3
5	Pipe Material	Categorical	1	3
6	Age – Year since installed	Continuous	5	31
7	Operational Pressure(kg/cm^2)	Continuous	1.5	3
8	C factor	Continuous	90	150
9	PipeThickness – mm	Continuous	6	18
10	Length of Pipe – m	Continuous	28	560

In order to predict the Number of Leakages, multiple linear regression analysis was performed. The regression analysis was carried out considering Number of Leakages as dependent variable and other variables as independent variables. In order to incorporate categorical variable in regression analysis, dummy coding is performed. There are two categorical variables, Traffic type and Material of pipe. Both variables are measured in 3 levels. The coding is done as below.

Table 2: Dummy Coding of Type of Traffic

Sr. No	Level of Categorical Variable	Dummy Coded Variable	
		X8	X9
1	Low Traffic	1	0
2	Moderate Traffic	0	1
3	High Traffic	0	0

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 1 Issue 10, December- 2012

Table 3: Dummy Coding of Type of Material

Sr. No	Level of Categorical Variable	Dummy Coded Variable	
		X10	X11
1	M.S Pipe	1	0
2	D.I Pipe	0	1
3	C.I Pipe	0	0

4. Regression Equation:

The model can be written as:

Y = a + b1*x1 + b2*x2 + b3*x3 + b4*x4 ++ b5*x5 + b6*x6+ b7*x7+ b8*x8 + b9*x9+b10*x810 + b11*x11 + e

Table 4 : Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.547[a]	.299	.289	2.712

The table below shows the summery of multiple regression analysis. The model suggests the R square for the regression was 0.299 and ANOVA (F= 29.99) was also significant (0.000) indicating the regression model is valid and the 11 independent variables are explaining 29.9 percent of variance in dependent variable Number of leakages.

The table below shows the summery of coefficients. It can be seen that majority of the coefficients are found significant at 5 percent level of significance. The Regression equation is written as below.

Number of Leakages = a - 0.005(Diameter) – 2.276 (Depth) + 0.034 (Age) + 3.641 (operational Pressure) + 0.001(C factor) + 0.685(Pipe Thickness) + 1.611(Log Length) -3.03(Low Traffic) - 2.478 (Medium Traffic) – 0.208 (M.S)-0.605 (D.I) + e

Table 5: ANOVA

Model	Sum of Squares	Df	Mean Square	F	Sig.
1 Regression	2426.574	11	220.598	29.991	.000[a]
Residual	5693.117	774	7.355		
Total	8119.691	785			

The chart below shows the scatter plot of predicted value and number of leakages. As R square value of 0.299 shows that the model is poor fit to data and predictability of the model is very low.

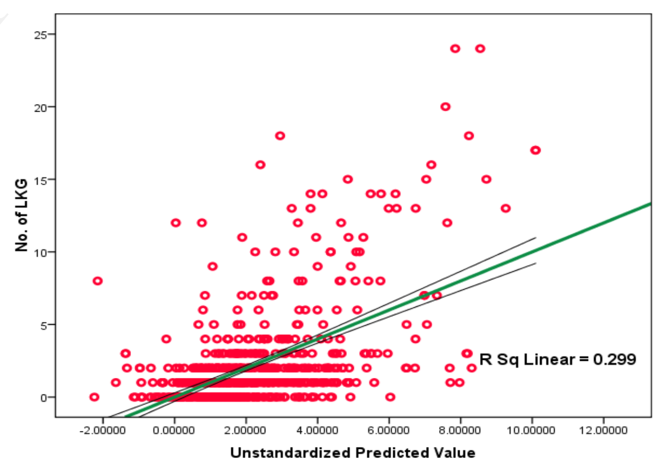

Figure 1: Scatter plot of Predicted value Vs Number of leakages

Table 6 : Summery of Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-6.506	3.102		-2.097	**.036**
Diameter	-.005	.001	-.330	-3.131	**.002**
Depth (m)	-2.276	.335	-.466	-6.803	**.000**
Age	.034	.012	.102	2.872	**.004**
Operational pressure	3.641	.304	.481	11.990	**.000**
C-factor	.001	.028	.007	.037	.971
Pipe thickness (mm)	.685	.127	.417	5.382	**.000**
Log Length	1.611	.561	.095	2.869	**.004**
Low Traffic	-3.030	.436	-.463	-6.945	**.000**
Medium Traffic	-2.478	.360	-.385	-6.889	**.000**
M.S	-.208	1.250	-.015	-.167	.868
D.I	-.605	1.299	-.089	-.466	.642

a.

5. Checking the Assumptions of Multiple Regression Analysis

5.1 Residual Analysis:

The analysis of regression residuals is an important tool for determining whether the assumptions of the multiple regression models are met. We will now discuss very important stage of checking the validity of the model assumptions in multiple regression analysis. Remember that under the assumptions of the regression model, the population errors *are* normally distributed with mean zero and standard deviation sigma. As a result, the errors divided by their standard deviation should follow the standard normal distribution: The chart below shows the histogram and P-P plot of Standardized Residuals. It can be clearly seen from the chart that the standardized residuals are not normally distributed violating the assumption of Multiple Regression Analysis.

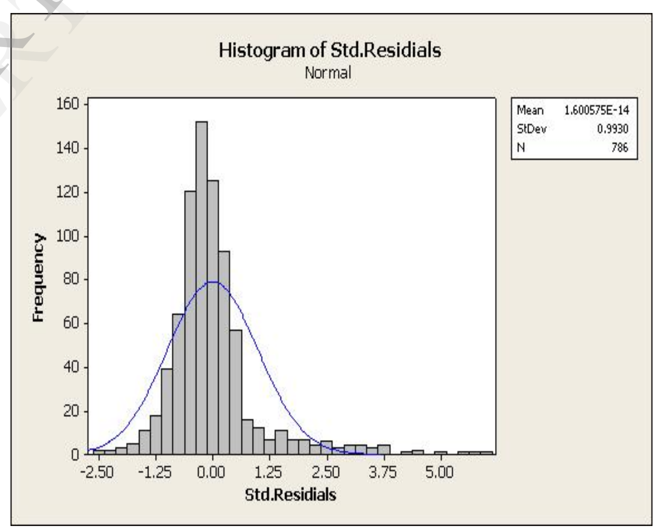

Figure 2: Histogram of standard residue

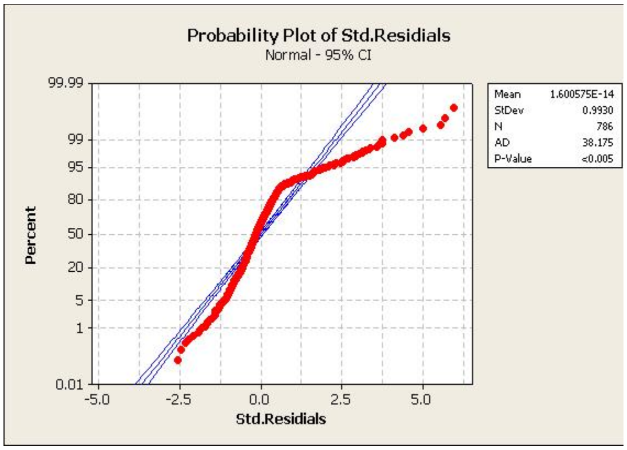

Figure 3: P-P Plot of Std. Residual

5.2 Error term has constant variance:

The second important assumption of Multiple Regression Analysis is Error term has constant variance for all levels of the predictor variables. To check this assumption, the scatter plot of Predicted value Vs Residuals is shown below. The graph clearly suggests that the error term do not have constant variance. The variance is increases as the number of predicted value increases.

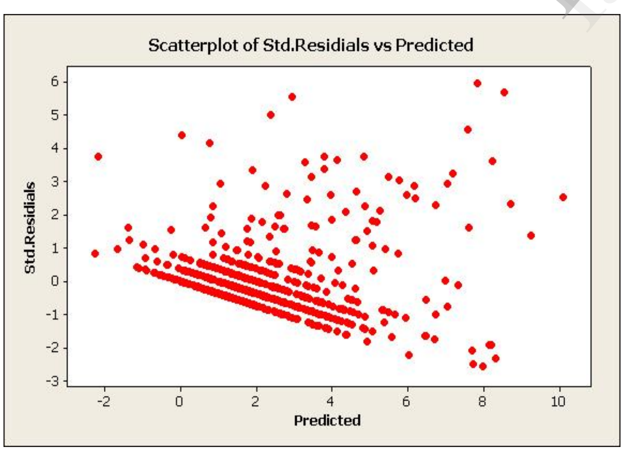

Figure 4: Scatter Plot of Std Residuals Vs Predicted Value

After checking the important assumptions of Multiple Regression Analysis, it can be concluded that the model is poor fit for data. Hence an alternative approach can be used to predict the number of leakages.

6. Poisson Regression:

Poisson regression assumes that data follows a Poisson distribution, a distribution that we frequently encounter when we are counting a number of events. Poisson distributions have three special problems that make traditional (i.e., least squares) regression problematic.

1. The Poisson distribution is skewed; traditional regression assumes a symmetric distribution of errors.
2. The Poisson distribution is non-negative; traditional regression might sometimes produce predicted values that are negative.
3. For the Poisson distribution, the variance increases as the mean increases; traditional regression assumes a constant variance.

In contrast, the Poisson regression model is not troubled by any of the above conditions. In particular, Poisson regression implicitly uses a log transformation which adjusts for the skewness and prevents the model from producing negative predicted values. As assumed for a Poisson model, our response variable (Number of Leakages) is a count variable, and each subject has the same length of observation time. The Poisson model, as compared to other count models (i.e., negative binomial or zero-inflated models), is assumed the appropriate model. In other words, we assume that the response variable is not over-dispersed and does not have an excessive number of zeros. The graph below shows the histogram and fitted Poisson curve. The fitted curve indicates that the distribution of number of leakages is more fitted to Poisson distribution as compared to normal distribution.

The model can be written as:

$\log e(Y)$ = a + b1*x1 + b2*x2 + b3*x3 + b4*x4 ++ b5*x5 + b6*x6+ b7*x7+ b8*x8 + b9*x9 + e

The Poisson Regression was performed by using IBM SPSS 19 and the result is shown below.

The equation can be written as

Number of leakages = exp (*-1.982 -0.002 (Diameter) – 0.694 (Depth) + 0.017 (Age) + 1.194 (operational Pressure) + 0.004 (C factor) + 0.223 (Pipe Thickness) + 0.002 (Length) -0.949 (Low Traffic) - 0.587 (Medium Traffic) – 0.045 (M.S)-0.517 (D.I) + e)*

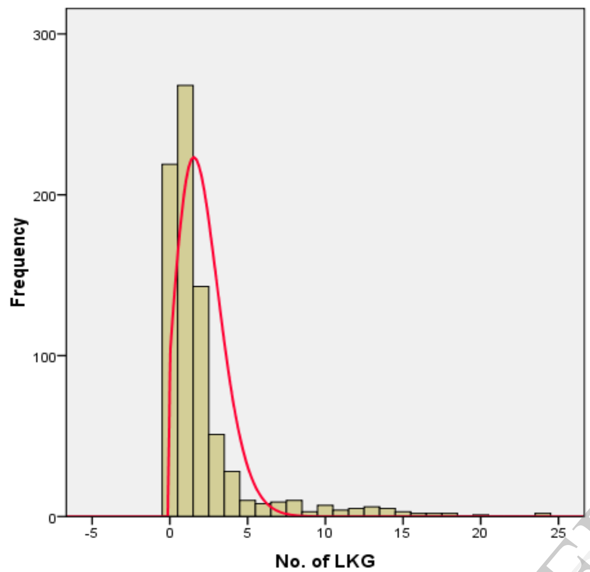

Figure 5: Histogram & Fitted Poisson distribution on Number of Leakages

REFERENCES:

Arulraj P, Suresh RH. Concept of significance index for maintenance and design of pipe networks. J Hydraul Engng 1995; 121(11):833–7.

Boulos P, Altman T. A graph-theoretic approach to exhibit nonlinear pipe network optimization. Appl Math Modeling 1991;15:459–66.

Constantine, A.G Darroch, J.N & Miller, R (1996) Predicting underground pipe failure, Australian water works association

Kettler, A. J. & Goulter, I. C. 1985 An analysis of pipe breakage in urban water distribution networks. Can. J. Civ. Eng. 12 (2), 286–293.

Malandain, J, Le Gauffire, P & Miramond, M. (1999) Modelling the aging of water infrastructure. Proceeding of the 13 th EJSW, 8-12 september, Germany, ISBN: 3-86005-238-1

Maurevin, K (1996) Predicting the failure performance of individual water mains. Urban water research association of Australia, Research report No. 114, Melbourne, Australia

Shamir, U. & Howard, C. D. D. 1979 An analytic approach to scheduling pipe replacement. J. AWWA 117 (5), 248–258.

Walski, T.M & wade, R. (1987) New York city water supply infrastructure study, Tech. Rep. EL-87-9 U.S Army Corps of engineers, New York.

Watson, T. 2005 A hierarchical Bayesian model and simulation software for water pipe networks. Civil Engineering. The University of Auckland, Auckland.

Woodburn J, Lansey K, Mays LW. Model for the optimal rehabilitation and replacement of water distribution system components.Proc Natl Conf Hydraul Engng, New York 1987;606–11.