Desalination and Water Treatment
www.deswater.com

1944-3994/1944-3986 © 2009 Desalination Publications. All rights reserved

2 (2009) 52–58

Probability model for risks of burst water pipes: a case study in Seoul

Kyoungpil Kim, Jayong Koo*, Minjeong Kim

Department of Environmental Engineering, University of Seoul, 90 Jeonnong-dong, Dongdaemun-gu, Seoul, Korea
Tel. +82 2 2210-2624; Fax. +82 2 2210-2245; email: jykoo@uos.ac.kr

Received 30 July 2007; Accepted 14 September 2007

ABSTRACT

Pipelines are a very important component of water supply systems. Specially, the pipe bursts and leaks are very useful indicators to show the condition of the network. To keep and improve the performance of the system, much accumulated know-how for inspections, operation, maintenance, and suitable rehabilitation to achieve the best performance is needed, as well as a logical method that can estimate the optimal time and range of replacement/rehabilitation work with an understanding of deterioration factors of pipe networks. Therefore, in this study, a statistical probability model for pipe burst risk was developed with various data from leak-repairing records and local characteristics of the circumstance on the real-scale distribution system in Seoul in order to utilize this method for management and operation of the water pipe network, including prioritization of pipe replacement/rehabilitation work.

Keywords: Pipe burst probability; Water distribution networks; Logistic regression

1. Introduction

The primary function of a water distribution system is to reliably deliver a sufficient quantity of good-quality water to its customers. The supply capacity of Korean waterworks is sufficient in most cities. However, Korean water pipelines are old, which gives rise to many problems including bursting of pipes, water contamination, leakage, and limited supply. Also, the demands of citizens who are pursuing improvement of life quality expand into waterworks. Consequently, much research is being conducted to solve these problems. Apparently, Korean waterworks have a period of maintenance from the period of expansion of facilities.

Water pipelines are very important components of a water supply system, much like blood vessels are to the human body. However, the diagnosis of pipe status is very difficult since its scale is huge, it has a very complicated composition, and it is buried underground. To keep and improve the performance of existing pipe systems, significant accumulated know-how on pipe inspection, operation, maintenance, and rehabilitation is necessary. In addition, the achievement of the best performance requires a logical method that can estimate the optimal time and range of replacement/rehabilitation work while understanding the factors in the deterioration of pipe networks. In some advanced countries, many studies have been conducted and are utilized for pipeline replacement/rehabilitation work as well as for the management/operation of pipe networks [1–3].

Pipe bursts and leaks are especially very useful indicators of the condition of a pipe network. In Korea, however, most past accidents were neither recorded well nor utilized scientifically because their importance was overlooked. Some existing studies forecasted pipe bursts with small numbers of impact factors such as the pipe age

---

*Corresponding author.

Presented at IWA Efficient 2007, May 20–23, 2007, Jeju, Korea.

and material. It is commonly accepted, however, that pipe bursts occur due to the complex action of various factors and not simple factors [4]. For more reliable results, good-quality and substantial data are necessary. Many water utilities, however, have little if any available data, since most of their pipes are buried underground and their records of pipe bursts and leak repairs are scanty, and the cost of constructing a database is very high through surveys or experiments. For these reasons, the execution of a physical model is difficult and has many limitations. Consequently, research is being conducted on the burst probability or the estimation of major transmission water mains that have significant damages and economic losses. In contrast, statistically derived models can be applied with various levels of input data and may thus be useful for minor water mains for which there are few data available or for which the low cost of failure does not justify expensive data acquisition campaigns [5].

In this study, therefore, a statistical probability model for burst risks was developed with various data from leak repair records and local characteristics of the circumstances. A logistic regression model was used for the burst probability. The logistic model was selected as it allows continuous variables such as pipe number density, and categorical variables such as soil type, to be combined [6].The logistic regression model can determine the burst probability of each pipe, whereas existing statistical models only estimate the burst probability of a group that has a similar feature. Consequently, there is a chance that the burst possibility of each pipe can be changed with the burst risk of each distribution block. Moreover, the results of this study can be used to formulate a plan for the management and operation of water distribution systems, including prioritization of pipe replacement/rehabilitation work. This prearranged plan is anticipated to reduce burst frequency.

2. Method

2.1. Study area

In this study, the availability of records of pipe bursts and leak repairs and of GIS was considered in the selection of the study area. The sufficiency of the volume of data for statistical analysis, including all ranges of pipe diameters in the system, was also considered. Consequently, one distribution block in Seoul was selected as the study area. Its map is shown in Fig. 1. The studied area was 47.1 km^2 and consisted of 99 small distribution blocks. The total pipe length was 614.7 km, the area population was 390,000, and the ratio of water supply service was 99%. More than a half of the total area was a green zone, and the north was a residential area that had a 90% population.

2.2. Investigation and analysis of pipe burst leak repair records using GIS

Pipe bursts and leakage accidents are not caused by a single factor but by a combination of many factors such as the material, the pipe diameter, the water quality, the groundwater level, the temperature, the laying year, the laying depth, the soil types, the traffic loading, and the road types. Therefore, in the analysis of burst features and the determination of the burst probability function, a multi-variate analysis that includes all the variables is necessary and not a separate analysis of each contributing factor.

To analyze a pipe burst feature and determine a pipe burst function, a database was constructed using leak repair records that have information on location, material, diameter, cause, burst type, and bursted part over a 3-year period (2001–2003). Then the leak repair records were combined with a digital map in GIS, and the cases were analyzed with respect to various aspects. The construction of the database in GIS is expected to lead to the effective management of pipe bursts, correlate pipe bursts with location information, and increase work efficiency and use of basic materials of areal management from existing lineal management.

2.3. Pipe burst probability function

To formulate a pipe burst probability function, a model that can include continuous variables and categorical variables is necessary. In this study, the logistic regression model was used to generate a probability function of pipe burst with considering the categorical factors extracted from operational and regional characteristic. When the case which occured a pipe burst is 1, and 0 otherwise, the burst probability (p) was defined as:

$$\text{Prob}=\frac{1}{1+e^{-Z}} \quad Z=b_0+b_1x_1+b_2x_2+\ldots+b_nx_n \tag{1}$$

in which $z = b_0 + b_1x_1 + b_2x_2 +\ldots+ b_nx_n$ are the independent variables, which are the contributing factors, and b_i refers to the coefficients of the model. The regression coefficients were selected by forward stepwise regression and the statistical significance for regression coefficients was tested by Wald statistics within the level of significance 0.05%. The logistic regression and the variable selection were carried out using the SPSS ver. 11.5 statistical package.

2.4. Cut-off value for determination of burst risks

The logistic regression model, a type of classification model, can be evaluated by comparing the real value with

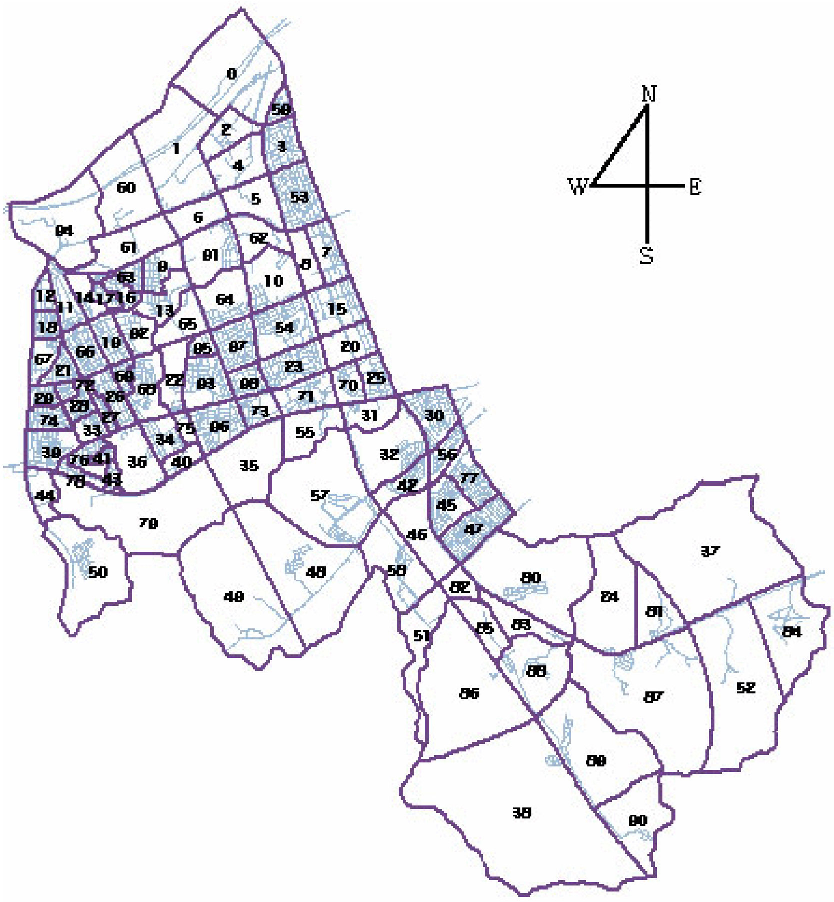

Fig. 1. Map of the studied area.

Table 1
Variables in statistical analysis

Variable	Type	Explanation
PROB	Binary	Dependent variable: 1, if pipe failure occurred; 0, otherwise
Laying year	Category	Laying year: categorize into seven groups from 1970
Diameter	Category	If diameter is ≥80 mm 1, 80 mm; 2, 100–150 mm; 3, 200–250 mm; 4, 300–350 mm; 5, 400–500 mm; 6, ≥600 mm
Material	Category	Pipe material: Galvanized steel (GS), polyvinyl chloride (PVC), copper (CO), stainless steel (SS), if pipe diameter ≤50 mm; Ductile cast iron (DCI), polyvinyl chloride (PVC), polyethylene (PE), cast iron (CI), coated and wrapped steel, if pipe diameter ≥80 mm
Average laying depth	Category	1, if average depth is lower than 1.0 m; 2, 1.0–1.5 m; 3, otherwise
Road width	Category	Road width: categorize into four groups: 1, not laid under the road ≤12 m; 2, laid under the road between 12–25 m; 3, laid under the road between 25–40 m; 4, laid under the road ≥40 m
Area	Category	Categories into three groups: 1, not busy; 2 residential; 3, mixed (residential + commercial)
Subway	Binary	1, if subway exists nearby; 2, otherwise

the predicted value classified at a cut-off value. Therefore, the accuracy of the estimation depends on the setting of the cut-off value. In this study, a type I error was one in which a broken pipe was not considered a burst pipe, and a type II error was when an unbroken pipe was considered a burst pipe. To decide on the optimal cut-off value, a cost loss was approximately calculated for each of the two types of error.

The cost loss of type I errors was evaluated as a water leakage that occurred while the burst pipe was let alone. The cost loss of type I errors in each pipe is defined in Eq. (2):

$$\mathrm{Cost}_{\mathrm{I}} = \frac{L \times P_1}{N_1} \tag{2}$$

in which L is assumed to be the water leakage that would have occurred had the leak not been repaired, and was calculated as the increase in the water volume through the repair of the burst pipes. P_1 is the cost of production, and N_2 is the number of repaired pipes.

The cost loss of type II errors was evaluated as the cost of the unnecessary detection and investigation of the water leakage, and is defined in Eq. (3):

$$\mathrm{Cost}_{\mathrm{II}} = \frac{P_2}{N_2} \tag{3}$$

in which P_2 is the annual cost of detection and investigation of the water leakage, which was calculated based on the annual staff labor cost, and N_2 is the number of detected water leakages.

2.5. Probability of pipe bursts in each distribution block

The distribution block system is recognized as a reasonable method of effectively managing a large water supply network that has a very complicated composition. It is widely used or planned to be used in many advanced countries. Also, a shift to areal management is necessary from the existing lineal management for effective pipeline replacement/ rehabilitation. This study proposes an areal management technique by computing the probability of a pipe burst in each distribution block using the burst risk in each pipe, which was calculated via the logistic regression model. The risks are defined in Eq. (4):

$$\text{If } p_i \geq p_0$$

$$R_k = \frac{\sum_{i=1}^{n} \left(p_i \times l_i \right)}{\sum_{i=1}^{n} l_i} \quad \left(i = 1,2,3,\ldots n,\ k = 0,1,2,\ldots,98 \right) \tag{4}$$

in which R_k is the risk of pipe bursts in the distribution block k; P_i is the probability, which was calculated using the burst probability function in pipe i; l_i is the length of pipe i; and P_0 is the cut-off value that was determined by static analysis. The flowchart of this study is shown in Fig. 2.

3. Results and discussion

3.1. Characteristics of pipe bursts in the study area

The investigation of records regarding pipe bursts and leak repairs and the understanding of their features are fundamental and very important because they are used as the basis for decision-making on the prioritization of pipeline replacement/rehabilitation work. The total number of pipe burst accidents was 2,215, and some accidents of other construction firms were excluded in the analysis of pipe burst features. The causes of pipe bursts were deterioration (1,241 cases), cracks (90), shock power (32), relaxation (23), electric corrosion (21), and others (349). In the records of these cases pertaining to pipe bursts and leak repairs, the detailed causes are not discussed and the record contents are formal and vague. Therefore, systematic and clear recordkeeping and management are necessary.

Seventy-six percent of total pipe bursts occurred in pipes with diameters of less than 50 mm, and the bursting of the pipe body was generally seen, but the bursting of the joint and the valve were seen slightly more in the case of DCI pipes. The numbers and probabilities of pipe bursts are shown in Table 2, in which the PVC and GS are more or less than 50 mm. The PVC and PE were more in the pipes with diameters of more than 80 mm. The numbers of bursts are as follows, in descending order: GS > CO > SS > DCI > PE > CI > PVC > SP. However, the number of bursts per unit length are as follows, in descending order: GS > PVC > CO > PE > SS > CI > DCI > SP. It can thus be seen that the number of pipe bursts per unit length of PVC and PE is high in comparison to the overall number of pipe bursts. In truth, PVC and PE pipes are continuously replaced because they are known to be improper as pipe materials for drinking water.

3.2. Pipe burst probability function in the case study area

The static analysis showed that in the pipes that were less than 50 mm, the road width, area characteristic, and laying year were significantly affected by the pipe burst, as were the diameter, material, road width, laying year, and subway approach in the over-80 mm pipes. In the over-80 mm pipes, the constant was excluded from the equation because the p-value was more than 0.05.

In the over-80 mm pipes, the pipe feature in itself was selected as a significant factor; and in the case of the less than 50 mm pipes, the regional feature was also selected and was more significantly affected by the regional

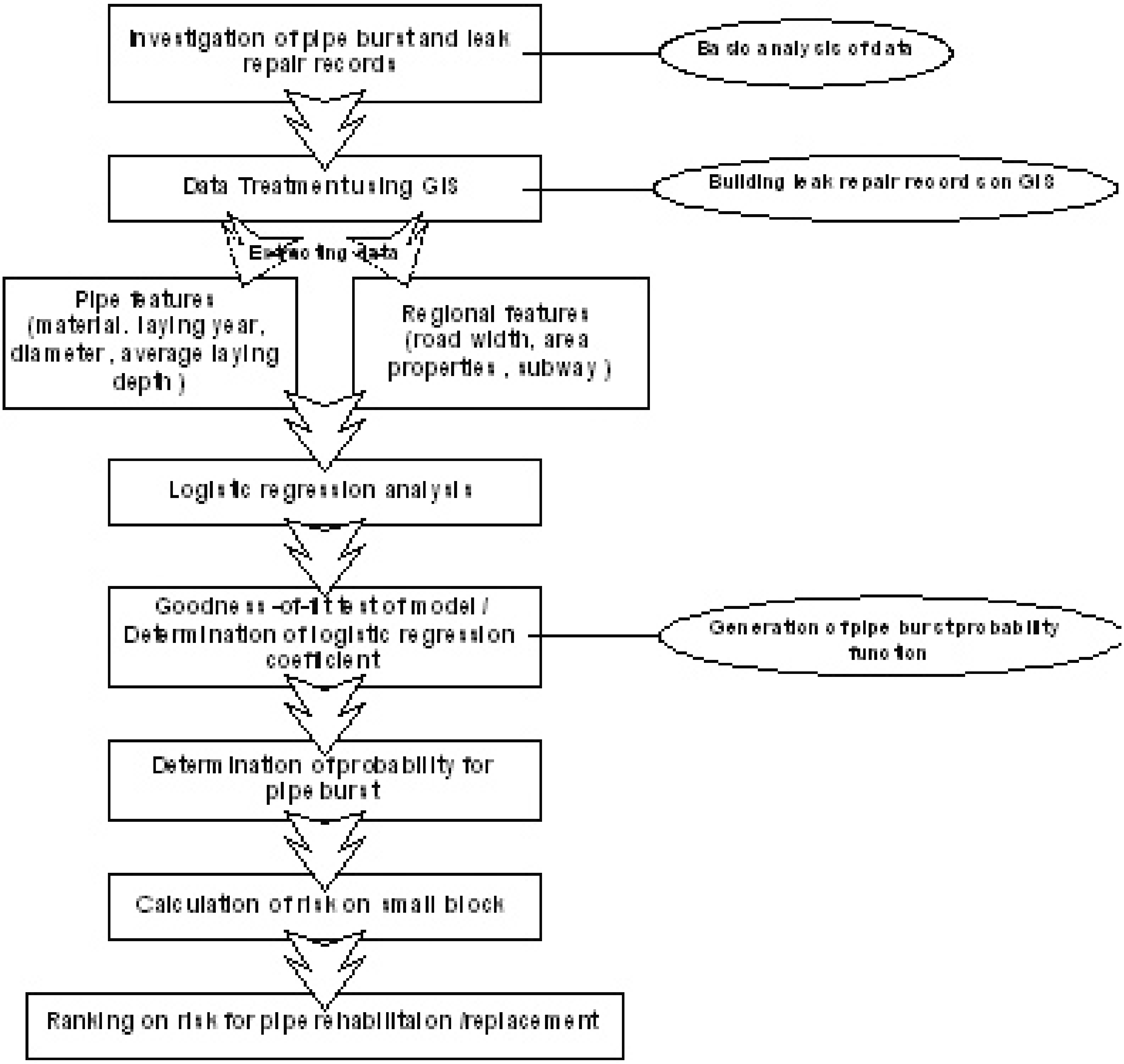

Fig. 2. Modeling a probability function of pipe bursts.

feature. Effects were seen not only on the diameter but also on the material.

In the less than 50 mm pipes, the pipe material, road width, area characteristic, and pipe age appeared to significantly affect the pipe burst. In of the over-80 mm pipes, the pipe diameter, material, road width, pipe age, and closeness to the subway appeared to affect the pipe burst. The probability function given in Eq. (1) was tested against several statistical values as well as using the Hosmer–Lemeshow goodness-of-fit test. The results of the function are shown as Eq. (5):

- where smaller than 50 mm:

$$\begin{aligned} Z = &-2.867 + (0.086A_1 - 0.341A_2) + (2.187M_1 + 5.373M_2 - 1.117M_3) \\ &+(-1.066R_1 - 0.453R_2 - 0.197R_3 - 0.544R_4) \\ &+(0.535L_1 + 1.057L_2 + 0.810L_3 + 0.629L_4 + 0.413L_5 + 0.546L_6) \end{aligned}$$

- where larger than 50 mm

$$\begin{aligned} Z = &-3.557 + (0.526S_1) + (-0.411R_1 + 0.196R_2 - 0.179R_3 - 1.000R_4) \\ &+(-2.142M_1 + 0.008M_2 + 0.989M_3 - 1.816M_4) \\ &+(3.238D_1 + 2.757D_2 + 3.264D_3 + 3.341D_4 + 1.45D_4) \\ &+(-1.929L_1 - 0.830L_2 - 0.935L_3 - 0.281L_4 + 0.086L_5 + 0.730L_6) \end{aligned} \tag{5}$$

Table 2
Features of pipe bursts in the study area according to material

	Total	GS	SS	CO	PVC	PE	CI	DCI	SP
Pipe length, km	614.674	14.108	73.052	40.258	2.469	10.541	53.910	362.936	57.400
Percentage	100	2.3	11.9	6.5	0.4	1.7	8.8	59.0	9.3
No. of pipe breaks	1,814	749	252	356	37	82	65	240	33
Percentage	100	41.3	13.9	19.6	2.0	4.5	3.6	13.2	1.8
No. of pipe breaks rate, No./km/y	0.98	17.70	1.15	2.95	5.00	2.59	0.40	0.22	0.19

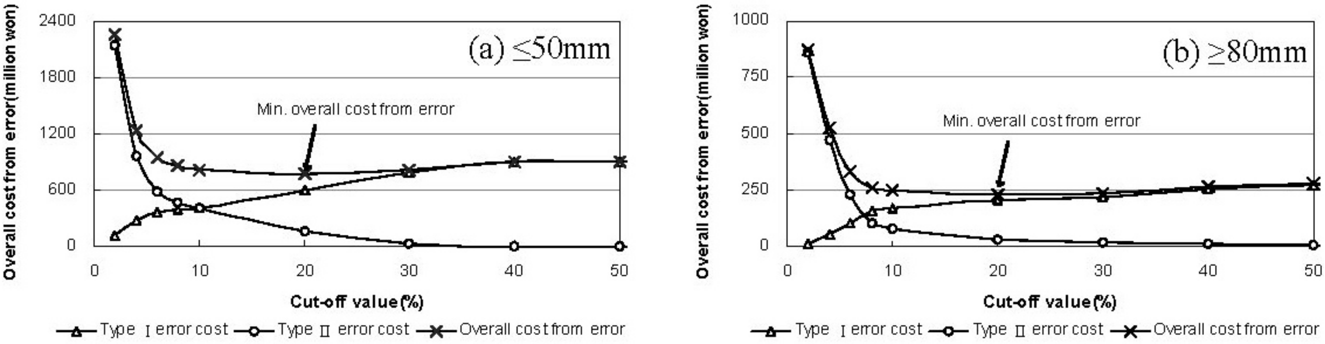

Fig. 3. Minimum cost by errors on cut-off value.

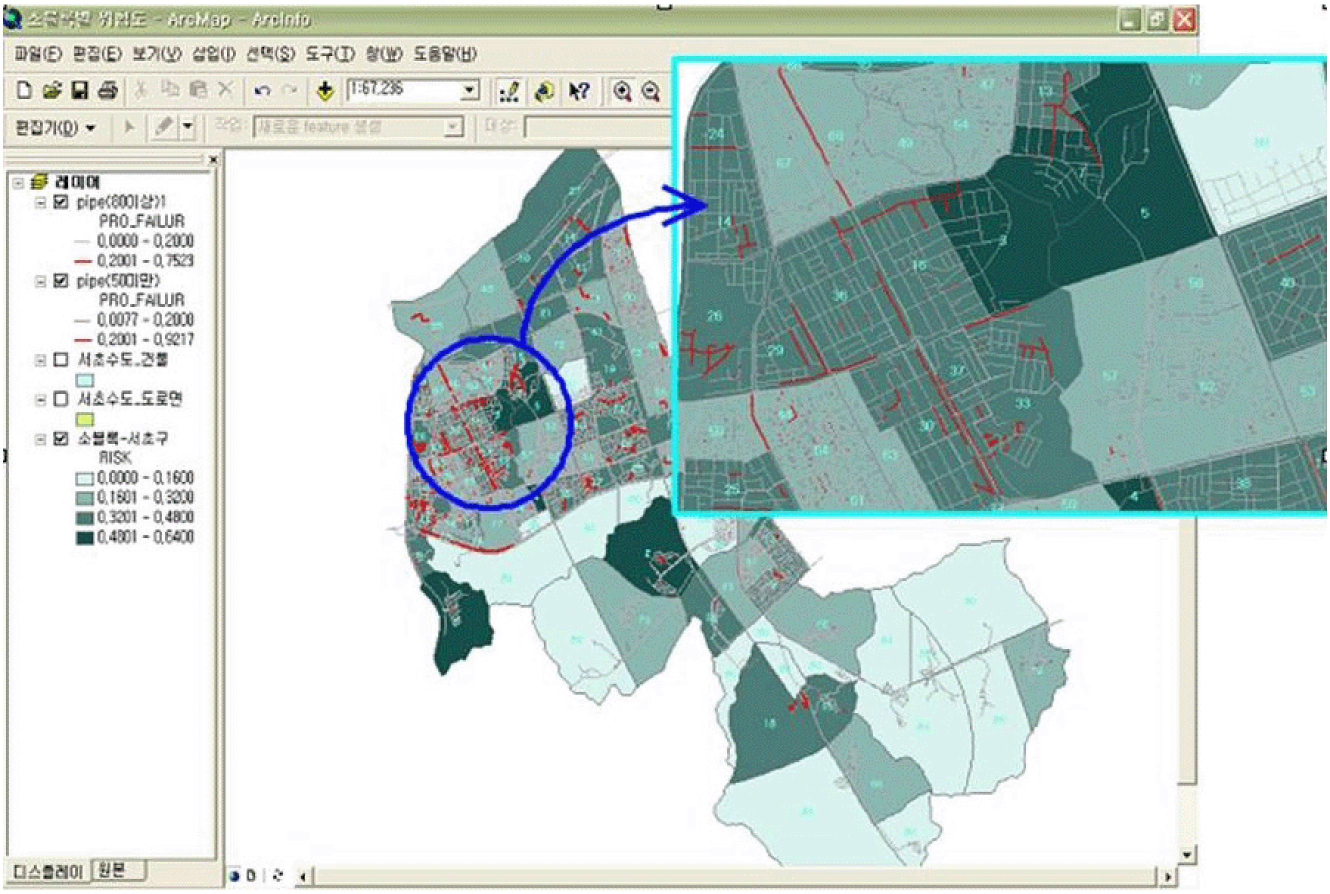

Fig. 4. Risk levels displayed in GIS interface.

where A is the area characteristics, M the pipe material, R the road width, D the pipe diameter, L the pipe age, R the road width and S is proximity to the subway.

3.3. Classification criteria for determination of the burst risk

The classification criteria were determined to minimize the total cost of the classification errors by 20% of the cut-off value, as shown in Fig. 3. The cost loss per pipe burst was calculated by product with the number of errors. However, in this study, the cost loss for errors is very rough. If the calculation method for cost loss is developed with more real data and research, the cut-off value will change more accurately.

3.4. Risk of burst probability in each distribution system

The study area was composed of 99 distribution blocks. First, the pipes that had burst risks because they were more than 20% of the burst probability in each distribution block were extracted and then calculated. In the case of pipes that had no burst risks; the risk of blockage was set at 0, and the risk of each distribution block was charged a value by rank. The greatest risk of danger was 0.6383, and the risk of the smallest block was 0.2095 when it was exempted in the case in which there was a 0 risk.

For practical use of the pipe burst probability function, the risk level of each distribution block was calculated and the order of the risk level was determined. The risk levels were classified into four groups and displayed using GIS, as shown in Fig. 4. This classified risk level can be used to prioritize pipeline replacement/rehabilitation work.

Since the calculated risk is shown as the burst probability per unit length, it can be used to prioritize pipeline replacement/rehabilitation because the cost is generally computed per unit length in Korea. However, the risk of each distribution based on the pipe burst probability was part of the planning for decision-making, and to get a better decision from the study with respect to the water balance in each block, the pressure during the minimum night flow, the cost loss due to pipe burst accidents, and the cost of replacement/rehabilitation of each material and diameter were necessary.

4. Conclusions

A probability model for burst risks was developed with various data from leak-repairing records and local characteristics of the circumstance on the real-scale distribution system in Seoul as a case study. A statistically significant probability function was composed and the risk levels of the distribution blocks were calculated. The risk levels may be used to scientifically prioritize pipeline replacement/rehabilitation work. However, this burst probability is not a concept predicting future pipe burst but showing a risk of the present pipe conditions based on past accident records. Therefore, the reliability of the result may change according to the accuracy of data used in statistic analysis.

If this approach is used with well-constructed data on leak repair records and a broader number of regional characteristics within a longer period, systematic and economic management and operation of water distribution systems would be possible.

Acknowledgement

This research was supported by a grant (4-2-3) from the Sustainable Water Resources Research Center of the 21st Century Frontier Research Program.

References

[1] R.K. Herz, Exploring rehabilitation needs and strategies for water distribution networks, J Water SRT-Aqua, 47(6) (1998) 275–283.
[2] Y. Kleiner and B. Rajani, Using limited data to assess future needs, J. AWWA, 91(7) (1999) 47–61.
[3] N.R. Cooper, G. Blakey, C. Sherwin, T. Ta, J.T. Whiter and C.A. Woodward, The use of GIS to develop a probability-based trunk mains burst risk model, Urban Water, 2 (2000) 97–103.
[4] D.K. O'Day, Organizing and analyzing leak and break data for making main replacement decisions, J. AWWA, 74(11) (1982) 589–594.
[5] Y. Kleiner and B. Rajani, Comprehensive review of structural deterioration of water mains: statistical models, Urban Water, 3(3) (2001) 151–164.
[6] C.T. Ta, A probability model for burst risk studies of water mains, Water Sci. Technol.: Water Supply, 2(4) (2002) 29–35.