Statistical Approach for Corrosion Prediction Under Fuzzy Soil Environment

Mincheol Kim[1], Toyono Inakazu[1], Akira Koizumi[1], Jayong Koo[2†]

[1]Department of Civil and Environmental Engineering, Tokyo Metropolitan University, Tokyo 192-0397, Japan
[2]Department of Environmental Engineering, University of Seoul, Seoul 130-743, Korea

Abstract

Water distribution pipes installed underground have potential risks of pipe failure and burst. After years of use, pipe walls tend to be corroded due to aggressive soil environments where they are located. The present study aims to assess the degree of external corrosion of a distribution pipe network. *In situ* data obtained through test pit excavation and direct sampling are carefully collated and assessed. A statistical approach is useful to predict severity of pipe corrosion at present and in future. First, criteria functions defined by discriminant function analysis are formulated to judge whether the pipes are seriously corroded. Data utilized in the analyses are those related to soil property, i.e., soil resistivity, pH, water content, and chloride ion. Secondly, corrosion factors that significantly affect pipe wall pitting (vertical) and spread (horizontal) on the pipe surface are identified with a view to quantifying a degree of the pipe corrosion. Finally, a most reliable model represented in the form of a multiple regression equation is developed for this purpose. From these analyses, it can be concluded that our proposed model is effective to predict the severity and rate of pipe corrosion utilizing selected factors that reflect the fuzzy soil environment.

Keywords: Discriminant function, Distribution pipe, External corrosion, Regression analysis, Replacement plan, Soil properties

1. Introduction

The water system plays an important role as a lifeline for our existence. It is through pipelines that water is distributed to customers for steadiness and safety. Water pipes, however, tend to be corroded as time goes by. Pipes laid underground are hardly examined without excavation. Some deteriorated pipes may be exposed to leakage or damages, resulting in the decline or interruption of water supply. In order to maintain customer service at a desirable level, replacement of such old pipelines is considered inevitable. As the replacement of pipelines requires huge costs and time, there is a pressing need to provide information on how to evaluate the pipe condition in an effective way.

Deterioration or corrosion of pipes may cause cracks and bursts, resulting in water leakage, pipe repair, and poses even the so-called water quality problem 'red water'. Occurrence of the corrosion relates to many factors: pipe materials, pipe age, surrounding soil conditions, water quality, pipe maintenance and management [1, 2]. It is, therefore, difficult to examine where the corrosion is taking place and the extent of it. Corrosion of the outer surface of the metallic pipes occurs mainly due to electrochemical reactions under the heterogeneous soil condition. The reaction rate is influenced by soil resistivity, permeability and mineral ions existing around the installed

pipes [3, 4].

Most of the previous researches dealt with causual analysis of pipe corrosion and prediction of future corrosion for water distribution pipes. Katano et al. [5] found the log-normal distribution fit their pit data the best and analyzed environmental factors using regression analysis. The environmental factors (soil type, pH, resistivity, redox potential, and sulfate ion) were found to be significant in determining pit depth. Kolovich and Kiefner [6] developed the Monte Carlo method for determining the corrosion rate distribution in buried pipelines that uses the probability distributions of corrosion depth and initiation time. Restrepo et al. [7] employed statistical techniques like cluster analysis to establish the sampling design and later for data analysis and obtaining a mathematical expression for external corrosion depth as a function of several experimental variables [7]. In addition, various studies have reported different methodologies used to be able to predict the future trend of corrosion for water pipeline.

The present study has strengths that although data is insufficient, effective evaluation of the current condition and prediction model can obtain a clear quantitative model by statistical approach and application is easy in other areas because of high reproducibility. In order to obtain a quantitative model for measuring pipe corrosion, we propose to apply an evaluation index to fuzzy environmental soil conditions around water pipes.

The present study aims at proposing an approach and method to predict the intensiveness and probable points of pipe corrosion in the pipe network. Base data utilized for the prediction are those sampled at the site. Objectives of the study are referred to: 1) obtain evaluation indexes for external corrosion reflecting the fuzzy soil nature of installation sites; 2) develop a multiple regression model to measure the degree of pipe corrosion under the fuzzy soil conditions; and 3) evaluate future risk and timing for decision of pipeline replacement.

2. Materials and Methods

The study focuses on the existing distribution pipe network in S city, South Korea during 2009–2010 [8]. Field data on pipe corrosions and soil conditions around pipes were obtained in the previous research project including those of the test pit excavation at 60 random locations along the water distribution pipelines (Fig. 1). The study area extends to 121.05 km² with a population of approximately 1.11 million [9].

Table 1 shows relevant data for this analysis with their mean values and standard deviation. The external corrosion is measured by two indexes: external corrosion depth given as localized corrosion index (Y_d), and corrosion spread on the pipe's surface given as general corrosion index (Y_a). Influential factors are grouped into two: factors related to 'characteristic of water pipe' and 'soil condition'. The former contains 3 items, all of which are obtained from geographic information system (GIS) database of the distribution pipe network. The latter are 4 items as measured from soil sample testing.

The previous studies suggest synergistic effect of the external corrosion depth (Y_d) and the corrosion spread (Y_a) on pipe breakage [10]. Following the spread of general corrosion, the localized corrosion may occur at a specific point on pipes coincidentally. So we have analyzed these two types of corrosion in parallel.

Prior to the main analysis, corrosion conditions are presented in histograms to show cumulative frequency vs. corrosion level (Figs. 2 and 3). No pipe corrosion is found at 47% of randomly selected sample points. Despite the pipe age of 12 to 52 years, some newer pipes (under 20 years) have deeper and/or wider spread of corrosion than the average. To the contrary, there are some uncorroded samples found in older (over 40 years) pipes. This tendency is considered to be closely related to characteristics of the surrounding soil conditions and the pipe conditions which are not protected by a polyethylene sleeve. Next, histograms are created for each of the soil properties in order to make it easy to understand a distribution of samples

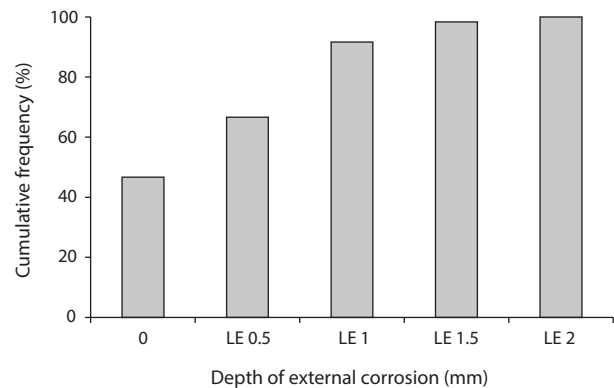

Fig. 1. Sampling points in S city.

Fig. 2. Histogram of the depth of external corrosion (Y_d).

(Figs. 4–7). For soil resistivity, it was identified by the American National Standards Institute (ANSI) method that only four samples had measurement values below 700 to be under the very highly corrosive soil condition, and around 72% of samples indicated values above 2,000 to be sampled from the extremely low corrosive soil condition. As pH's histogram shows, also, only one sampling point had acidity of less than 4, whereas 7 sampling

Table 1. Items for analysis

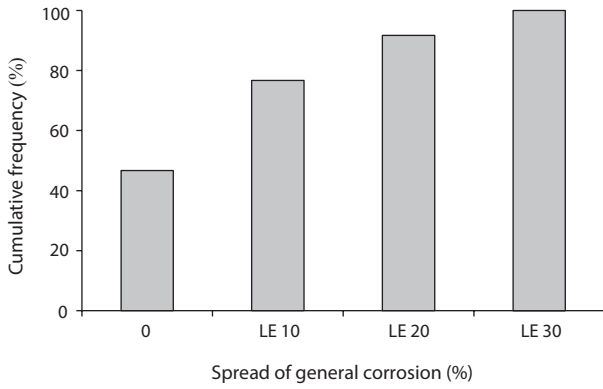|	Item	Variable	Mean	SD
External corrosion index	Index of external corrosion depth (mm)	Y_d	0.41	0.47
	Spread corrosion index (%)	Y_a	6.67	8.23
Characteristic of pipe	Diameter (mm)	X_1	182.00	102.11
	Water pressure (kg/cm²)	X_2	3.77	0.70
	Pipe age (yr)	X_3	23.10	8.50
Soil property	Soil resistivity (Ω-cm)	X_4	5,258	3,563
	Soil pH (-)	X_5	6.46	2.49
	Water content (%)	X_6	9.76	5.78
	Cl⁻ (mg/kg)	X_7	27.92	32.14

Fig. 3. Histogram of the spread of general corrosion (Y_a).
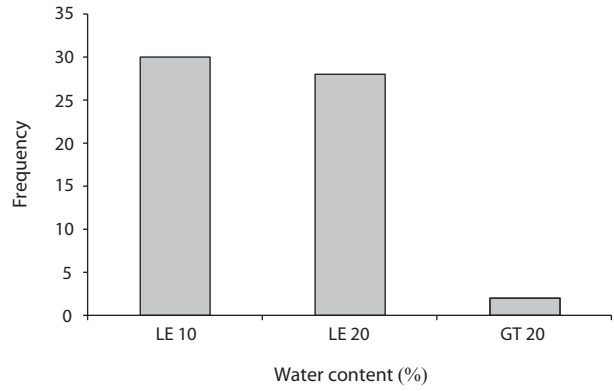

Fig. 6. Histogram of water content.
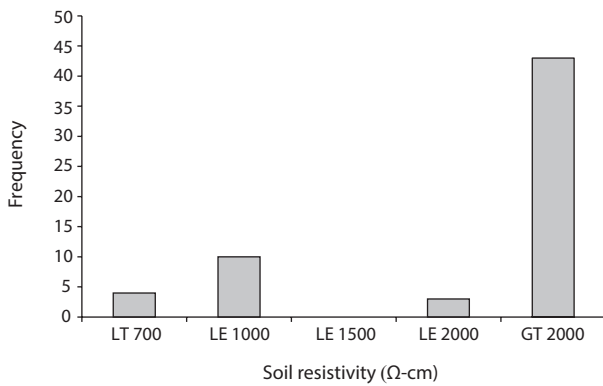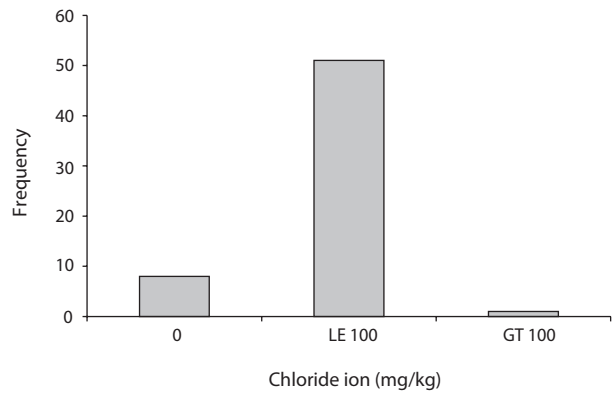

Fig. 4. Histogram of soil resistivity.

Fig. 7. Histogram of chloride ion.

Fig. 5. Histogram of pH.

methods for evaluation of soil corrosivity. As the results, nearly 47% of soil samples were evaluated as the corrosive condition (middle and strong corrosivity) by the modified ANSI method. On the other hand, around 88% of soil samples were found to be evaluated as weak corrosivity and 12% soil samples as middle corrosivity using the DVGW method. Also, the relation between the corrosive condition of each soil sample was evaluated by two methods and the actually measured external corrosion is shown in Figs. 8–11. As Figs. 8 and 9 illustrate, there are samples that actually measured external corrosion depth and spread of general corrosion have low values with the high score of modified ANSI and also have high values with the low score of modified ANSI. Similarly, between actually measured values of external corrosion and absolute values of DVGW score don't have definite relationship as shown in Figs. 10 and 11. These results mean that there are significant differences between the actually measured external corrosion and the corrosive condition evaluated by the two methods above. Thus, it is considered that a new evaluation method of corrosivity regarding soil samples should be developed.

In this study, in order to obtain a more reliable model to predict the degree of pipe corrosion, the relationship between external corrosion and soil property is further investigated by applying discriminant function analysis (DFA). From examining the discriminant function, we quantify the soil environment corrosivity using some combination of variables in which the
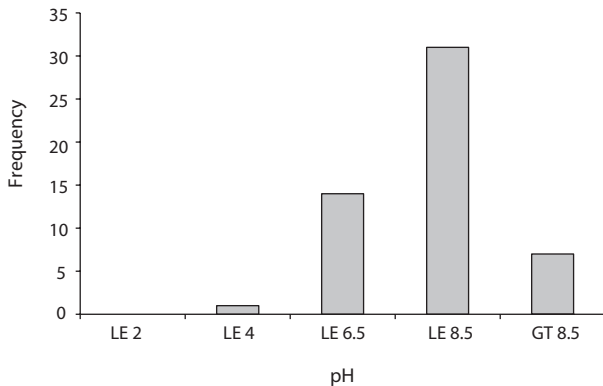
points were found to have alkalinity of over 8.5. Next, in relation to water content, exactly half the samples is obtained from dry contition (water content is less than 10%). Lastly, chloride ion was detected from 87% of soil samples. These figures illustrate that the sampling points covered a wide range of soil conditions.

Then, the corrosive condition of soil samples was evaluated using the modified ANSI method and the German Waterworks Association (DVGW) method which are the most representative
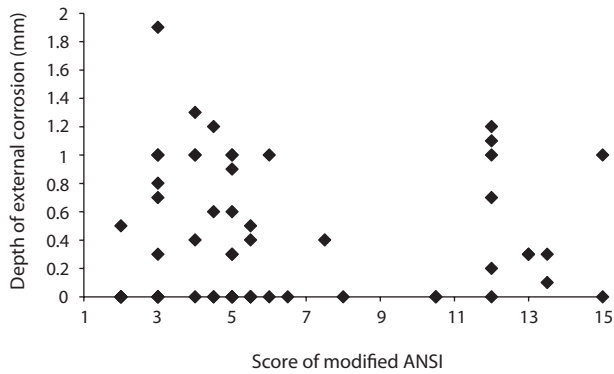
Fig. 8. Relation between score of modified ANSI and external corrosion depth. ANSI: the American National Standards Institute.
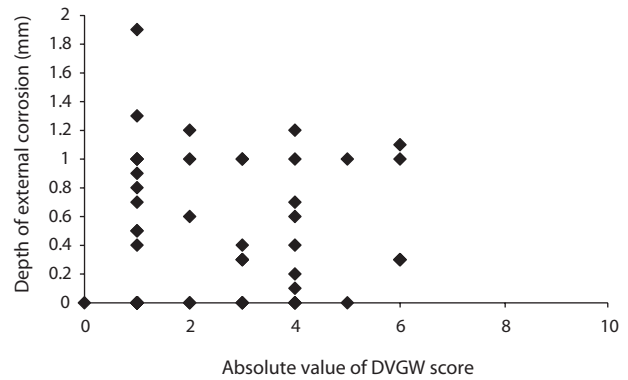

Fig. 10. Relation between absolute value of DVGW score and external corrosion depth. DVGW: the German Waterworks Association.
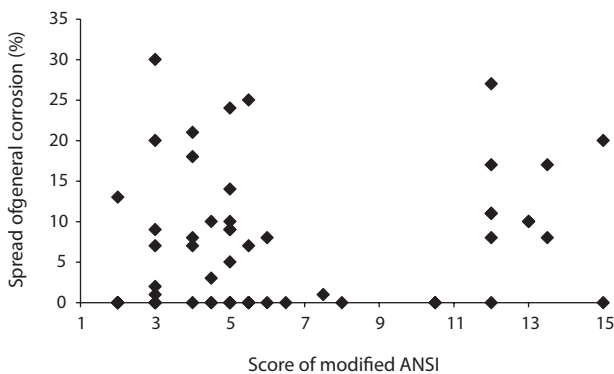

Fig. 9. Relation between score of modified ANSI and spread of general corrosion. ANSI: the American National Standards Institute.
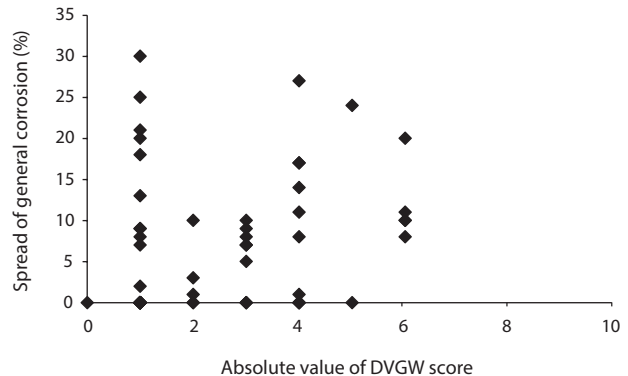

Fig. 11. Relation between absolute value of DVGW score and spread of general corrosion. DVGW: the German Waterworks Association.

fuzzy soil property is reflected. The multiple regression model (MRM) is to measure the degree of pipe corrosion. In developing MRM, some of the pipe characteristics are considered effective as explanatory variables together with the selected variables in the DFA. By developing the MRM, the degree of pipe corrosion at some points is predicted for evaluation of present and future risk.

3. Results and Discussion

3.1. Discriminant Function Analysis

In order to distinguish the corrosive soil environment in all soil samples, DFA is applied to this analysis. Common purpose of the DFA is to predict the group membership based on a linear combination of variables using a measure of generalized square distance assuming that each group has a multivariate normal distribution. Another purpose of DFA is to acquire insight into the relationship between the group membership and variables in the prediction model which is given as a discriminant function.

At the beginning of DFA, observations for each group are made including both high-ranking and low-ranking samples reflecting a characteristic of the universe (the population of all samples). In this study, two groups, namely, a corrosive group (named BAD group) and a non-corrosive group (named GOOD group) are considered for each discriminant function corre-

sponding to each corrosion indexes (Y_d and Y_a). The modeling process with the discriminant function is proposed in the following order as shown in Fig.12.

Step 1. Set sample condition for BAD and GOOD group

It is a normal procedure that DFA focuses merely on data ranked high and low (namely, the first quarter and the last quarter of the sorted samples). In this study, 16 samples (almost the last quarter) which had over 0.6 mm of corrosion depth are regarded as the BAD group corresponding to Y_d, and 12 samples with over 10% area corrosion are categorized into the BAD group corresponding to Y_a. On the other hand, the GOOD group consists of no external corrosion samples both for Y_d and Y_a. As the numbers of effective data differ between these groups, a measure to balance the sample data is required.

Step 2. Balance sample size between BAD and GOOD group

To ensure stability of DFA, two groups are preferably equal in data size. As the number of samples in the GOOD group is greater than sample size of the BAD group, random sampling methods are adopted to select samples for the GOOD group. Several sets of samples are arranged so that the Good group may have same sample size to the BAD group, and then the corresponding models for classification are examined as below.

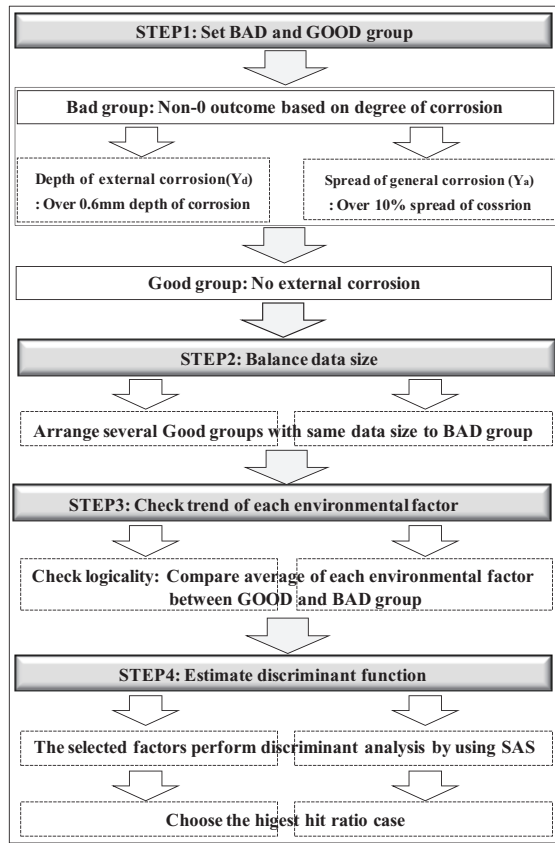Step 3. Examine whether each environmental factor is logically acceptable

Fig. 12. Development process of discriminant function analysis model.

$$Z_d = 0.601W_4 - 0.442W_6 - 0.081 \qquad (1)$$
$$W_4 = (X_4 - m_4) / s_4 \qquad (2)$$
$$W_6 = (X_6 - m_6) / s_6 \qquad (3)$$

On the other hand, as for the spread of general corrosion (Y_a), the soil resistivity (X_4) is found as an influential factor. Eq. (4) stands for a linear discriminant function obtained for Y_a. A standardized variable, W_4 in Eq. (4), is similar of the soil resistivity (X_4) as presented in the Eq. (2) above. If Z_a is greater than 0, then the sample is classified to the GOOD group. This implies that the spread of general corrosion is estimated as less than 10% of pipe surface.

$$Z_a = 0.859W_4 - 0.031 \qquad (4)$$

3.2. Modeling of External Corrosion

Evaluation indexes of fuzzy soil properties can be obtained from the linear discriminant functions. From these indexes, we can assess corrosiveness of the soil properties. It is found that low resistivity and/or high water content of the soil properties accelerate the rate of external corrosion depth. In the case of the general horizontal corrosion, however, the low resistivity of the soil is considered dominant in affecting an extent of the spread area and its corrosion rate. As to external corrosion prediction, regression analysis is applied to find influential factors related to pipe characteristics. Among 60 sampled data, nearly half of the sampled data didn't find corrosion on pipe surface. To minimize the effects on the regression model, these data are omitted in the succeeding analyses. The number of data utilized is twenty and nineteen samples for external corrosion depth and general horizontal corrosion, respectively.

It is believed that pipe characteristics and soil properties have linear and non-linear effects on pipe corrosion [11]. To incorporate these effects into the analyses on a same basis, this non-linear effect is first expressed in the form of power regression equations with a variable of pipe characteristics. Then, multiple regression equations for the two types of external corrosion $(Y_d$ and $Y_a)$ are formulated as a function of the selected factors expressed in linear and/or non-linear forms.

Prior to regression analysis, the correlation coefficient for the logarithmic data of pipe characteristic factors with corrosion indexes are estimated as shown in Table 2. Significant levels of reliability are confirmed as a result of the t-distribution test corresponding to sample size 'n' and statistically significant level ''. When an absolute value of the correlation coefficient is greater than its significance level, the factors are assessed as statistically significant. In Table 2, the depth of external corrosion (Y_d) has a non-linear relation with pipe age (X_3), and spread of general corrosion (Y_a) with diameter (X_1). These results suggest the effectiveness of adopting each factor as one of the explanatory variables in the corrosion prediction equations.

Step 1. Power regression model for external corrosion depth

Firstly, the regression model for between depth for external corrosion (Y_d) and pipe age (X_3) was made in the form of power function. According to Eq. (5), the estimated power coefficient (1.670) is greater than 1.0. This implies that the pipe age would affect pipe deterioration with acceleration as seen in Fig. 13.

$$Y_d = 0.00440X_3^{1.670} \qquad (R=0.439) \qquad (5)$$

After adjusting the sample size of both groups, the average value of each environmental factor for each group is compared. When the average value of an environmental factor in the BAD group is higher than that of the GOOD group, such a factor is judged to have logicality, excluding soil resistivity (X_4), which has the opposite tendency.

Step 4. Estimate discriminant function

The DFA model is developed with the aid of SAS ver. 6.03 (SAS Institute Inc., Cary, NC, USA) as a function of the selected factors, logically reasonable. For depth of external corrosion (Y_d) and spread of general corrosion (Y_a), 5 randomly sampled cases are analyzed respectively. The most reliable DFA models are selected for Y_d and Y_a which have the highest hit ratio, respectively.

As for external corrosion depth (Y_d), equations obtained from the analyses are assessed as highly reliable. Eq. (1) represents a linear discriminant function. If a classification index (Z_d) estimated from this equation is greater than zero, it is considered that the sample falls into the GOOD category. This implies the external corrosion depth within the range of less than 0.6 mm. Explanatory factors $(X_4$ and $X_6)$ are standardized as expressed in Eqs. (2) and (3), respectively. Table 1 shows their mean values (m_4 and m_6) and standard deviations (s_4 and s_6). From coefficients in the Eq. (1), W_4 has a larger absolute value than that of W_6. This indicates that the soil resistivity (X_4) is more influential than the water content (X_6) on the external corrosion depth.

Step 2. MRM for external corrosion depth

Following DFA and regression analysis in the form of power function, multiple regression analysis (MRA) is applied for further analysis. As the explanatory variables, the MRA utilizes pipe age expressed as 1.67^{th} power, soil resistivity (X_4) and water content (X_6) as previously verified effective by the discriminant function. This regression model expressed in Eq.
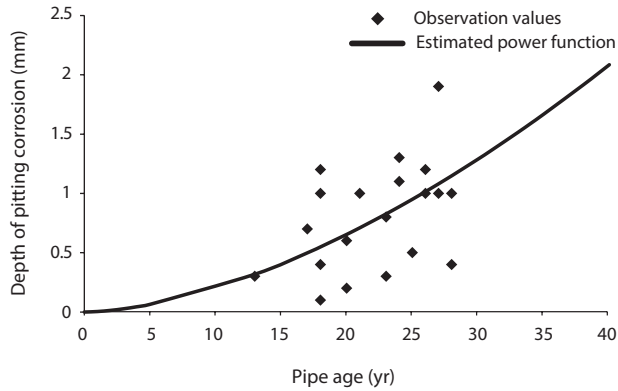
(6) is assessed statistically significant from its multiple correlation coefficient (R) obtained through the t-distribution test. The relation between the observed and the estimated values of external corrosion depth is shown in Fig. 14, which indicates sufficient accuracy of our model with a mean absolute error, 0.33 mm.
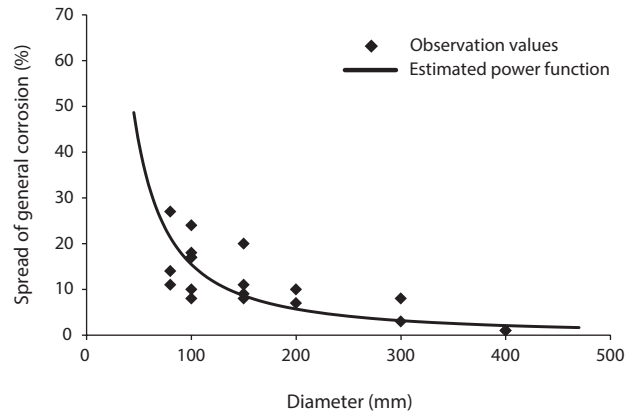

Fig. 13. Estimated trend of depth of external corrosion according to pipe age.

Fig. 15. Estimated trend of spread of general corrosion according to diameter.
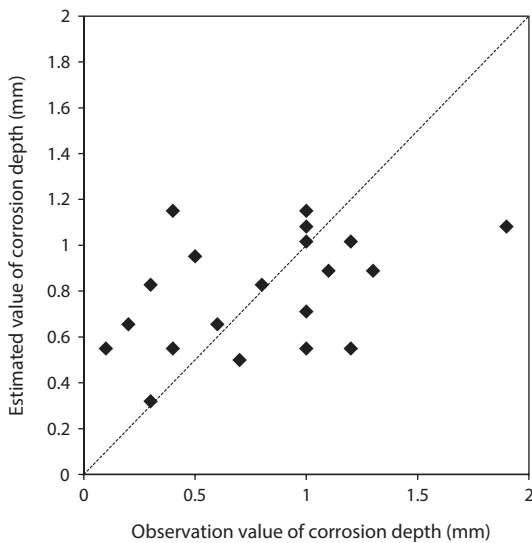

Fig. 14. Estimated values of external corrosion depth.

Fig. 16. Location where corrosion depth will exceed 2 mm by 2030.

Table 2. Logarithmic correlation between items of external corrosion and pipe characteristics

External corrosion index	Diameter (mm)	Water pressure (kg/cm²)	Pipe age (yr)
	log(X_1)	log(X_2)	log(X_3)
log(Y_d) [r (n=20, α=0.05) = 0.437]	0.080	-0.277	0.439[**]
log(Y_a) [r (n=19, α=0.05) = 0.456]	-0.836[**]	0.140	-0.436

r: standard value of correlation coefficient at α-significant level.
[**] The correlation coefficient is significant at 0.05 (bilateral).

$$Y_d = 0.00271X_4 + 0.00826X_6 + 0.00322X_3^{1.670} \quad (R = 0.482) \quad (6)$$

Step 3. Power regression model for spread of general corrosion

In the same way as stated above, a prediction model for the spread of general corrosion (Y_a) is expressed as a factor of the pipe diameter (X_1). The negative power (i.e., -1.435) of X_1 implies that value Y_a (the corrosion spread on pipe surface in percentage) tends to progressively decrease against an increase in value X_1 (the pipe diameter). This tendency is clearly seen in Fig. 15.

$$Y_a = 11481.536X_1^{-1.435} \quad (R = 0.836) \quad (7)$$

Step 4. MRM for spread of general corrosion

The main culprit behind the general corrosion is soil resistivity (X_4) among soil properties. According to the findings of regression analysis, pipe diameter (X_1) is considered to have a negative relation with the general corrosion (Y_a). In a multiple regression equation developed here, there are two independent variables as shown in Eq. (8). A correlation coefficient of the equation is estimated at 0.696. As examined by the t-distribution test, this value is able to ensure reliability of the equation.

Severity of the general corrosion for any pipes in the target area can be simply estimated from the equation. Data required for prediction are merely pipe diameter and soil resistivity in the targeted area. The equation is not expressed as the function of pipe age. This is due to the fact that the general corrosion rate per year is rather slower than that of the external corrosion depth. Continuous efforts are required to collect data on pipe corrosion. It may be possible to develop a more reliable prediction model based on the process as mentioned above.

$$Y_a = -0.0413X_4 + 13524.126X_1^{-1.435} \quad (R = 0.696) \quad (8)$$

3.3. The Corrosion Model Evaluation

In the previous paragraph, a prediction model for the external corrosion depth (Y_d) represented by pipe age (X_3), soil resistivity (X_4), and water content (X_6) was proposed.

Another method is a prediction which considers the soil environment around the pipes. Corrosion depth of 60 samples in the future is tentatively forecast in the Eq. (6), assuming that the pipes are left without any maintenance for 20 years. In this study a depth of external corrosion over 2 mm is considered serious and may result in pipe damage or leakage. Analyzing the result of forecast, a total of 27 out of 60 samples are expected to have serious external corrosion with a depth exceeding 2 mm on their pipe surface. Locations of those samples are given in Fig. 16. Out of total 60 soil samples, 39 consist of soft clay and 17 of red clay, while 4 samples are of sands. Soft and red clay, accounting for nearly half of the total are causing a rapid growth of corrosion, but the remnants are rather slow in corrosion, as they are predicted not to reach 2 mm. These results indicate that future risk of corrosion closely relates not only to the soil factors but also to the pipe characteristics.

4. Conclusions

This study intended to predict the severity of pipe corrosion in the target area. The study was carried out through assessment of various data collected at survey points. Results of the analyses indicate that influential parameters on external corrosion depth (Y_d) are soil resistivity (X_4), water content (X_6), and pipe age (X_3). It is also confirmed that the soil resistivity (X_4) and the pipe diameter, among others, affect the spread of general corrosion (Y_a). From all the above, it can be concluded that the multiple regression equation obtained herein provides valuable information on the degree and rate of external pipe corrosion in the target area. The proposed process for future risk evaluation is also effective to decide pipeline replacement and planning. Continuing efforts for collecting and storing up field data are, however, considered important to improve the reliability of the proposed prediction model obtained through statistical analyses.

References

1. Decker JB, Rollins KM, Ellsworth JC. Corrosion rate evaluation and prediction for piles based on long-term field performance. *J. Geotech. Geoenviron. Eng.* 2008;134:341-351.
2. Japan Ductile Iron Pipe Association (JDPA). Technical report about the factors of corrosion and anticorrosion measures for buried pipes. Tokyo: JDPA; 2001. p. 2-7.
3. Romer AE, Bell GE. Causes of external corrosion on buried water mains. In: Castronovo JP, ed. Pipelines 2001: advances in pipelines engineering and construction. Reston: American Society of Civil Engineers; 2001. p. 1-9.
4. Sarin P, Snoeyink VL, Lytle DA, Kriven WM. Iron corrosion scales: model for scales growth, iron release, and colored water formation. *J. Environ. Eng.* 2004;130:364-373.
5. Katano Y, Miyata K, Shimizu H, Isogai T. Predictive model for pit growth on underground pipes. *Corrosion* 2003;59:155-161.
6. Kolovich K, Kiefner JF. Calculation of a corrosion rate using Monte Carlo simulation. In: Proceedings of NACE International Corrosion 2007 Conference and Expo; 2007 Mar 11–15; Nashville, TN. p. 176-182.
7. Restrepo A, Delgado J, Echeverria F. Evaluation of current condition and lifespan of drinking water pipelines. *J. Fail. Anal. Prev.* 2009;9:541-548.
8. Suwon Metropolitan Waterworks. The technical analysis of water supply system in Suwon. Suwon: Suwon Metropolitan Waterworks; 2010. p. 247-286.
9. Choi TH, Lee SW, Koizumi A, Koo JY. Reliability assessment of water distribution systems using management reliability index. In: Proceedings of of the 4th IWA Leading-Edge Strategic Asset Management (LESAM 2011); 2011 Sep 27-30; Mulheim, Germany. London: IWA; 2011. p. 1-9.
10. Arai Y, Koizumi A, Umano H, Ashida H, Ozaki M, Yoshida E. Statistical analysis of the corrosion of water distribution pipes under their environmental factors. *J. Jpn. Soc. Civ. Eng. Environ. Syst. Res.* 2009;37:9-17.
11. Cleyo F, Velazquez JC, Valor A, Hallen JM. Probability distribution of pitting corrosion depth and rate in underground pipelines: a Monte Carlo study. *Corros. Sci.* 2009;51:1925-1934.