Fuzzy expert system to assess corrosion of cast/ductile iron pipes from backfill properties

Najjaran, H.; Sadiq, R.; Rajani, B. B.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1111/j.1467-8667.2005.00417.x
Computer Aided Civil and Infrastructure Engineering, 21, January 1, pp. 67-77, 2006-01-01

National Research Council Canada | Conseil national de recherches Canada

NRC · CNRC

Fuzzy expert system to assess corrosion of cast/ductile iron pipes from backfill properties

Najjaran, H.; Sadiq, R.; Rajani, B.B.

NRCC-46912

A version of this document is published in / Une version de ce document se trouve dans :
Computer-Aided Civil and Infrastructure Engineering,
v. 21, no. 1, January 2006, pp. 67-77.
Doi: 10.1111/j.1467-8667.2005.00417.x

Fuzzy expert system to assess corrosivity of cast/ductile iron pipes from backfill properties

Homayoun Najjaran, Rehan Sadiq, and Balvant Rajani
Institute for Research in Construction, National Research Council Canada (NRC)
Ottawa, ON, K1A 0R6; Tel: (613) 993-3806, Fax: (613) 952-8102
Email: Homayoun.Najjaran@nrc-cnrc.gc.ca

Abstract – Several factors may contribute to the structural failure of cast and ductile iron water mains, the most important of which is considered to be corrosion. The ANSI/AWWA C105/A21.5–99 10-point scoring (10-P) method is commonly used to predict corrosivity potential of a given soil sample using certain soil properties. The 10-P and other scoring methods use binary logic to classify the soil either as *corrosive* or *non-corrosive*.

Fuzzy logic extends binary logic in this context as it recognizes the real world phenomena using a certain degree of membership between 0 and 1. This paper presents a fuzzy logic expert system capable of predicting the deterioration of cast and ductile iron water mains based on surrounding soil properties. The proposed model consists of two modules: a knowledge base and an inference mechanism. The knowledge base provides information for better decision-making and is developed in a two-tier fuzzy modeling process. First in *direct approach*, the expert knowledge generates a subjective model to describe the characteristics of the system using fuzzy linguistic variables. Later in *system identification*, the field data is used to develop an objective model, which is eventually used in conjunction with the subjective model to provide a more reliable knowledge base for the expert system. The inference mechanism uses fuzzy approximate reasoning methods to process the encoded information of the knowledge base.

Keywords: Corrosion, soil corrosivity, fuzzy modeling, expert system, pipe deterioration, and 10-P method.

1 Introduction

Several factors may contribute directly or indirectly to the structural failure of metallic water mains. Factors such as casting and manufacturing defects may have an impact on the structural resilience of a pipe, while specific local and environmental conditions may act to exacerbate or sometimes alleviate stresses. Approximately 700 water-main breaks are reported in North America everyday, which accounts for about $1 billion annually in costs incurred (Lary, 2000). A similar study shows that water and sewer main failures cost $200 million per annum in Australia (Davis *et al*., 2003). It is now widely accepted that most breaks do not occur only in old pipes, rather corrosion is also found to play a major role in the premature failure of water mains (Spickelmire, 2002).

Water utilities use different criteria to assess the structural deterioration of pipes, among which the principal ones are breakage frequency or the growth rate of corrosion pits. The predominant deterioration mechanism on the exterior of metallic pipes is electro-chemical corrosion with the damage occurring in the form of corrosion pits in ductile iron (DI) and graphitized zones in cast iron (CI). Graphitization is a term used to describe the network of graphite flakes that remain behind after the iron in the pipe has been leached away by corrosion. Either form of metal loss will with time lead to a pipe break or leakage. The physical environment in which the pipe is placed has a significant impact on the deterioration rate. Factors that accelerate corrosion of metallic pipes are stray electrical currents, soil properties such as moisture content, chemical and microbiological content, electrical resistivity, aeration, and redox potential.

Deterioration modeling is an essential element of the decision making process for rehabilitation or renewal programs of water mains. Different mathematical and statistical techniques have been developed to model pipe deterioration. Probabilistic models are widely used in infrastructure deterioration modeling. Among this class of models, significant efforts have been dedicated to Markov and Markov derivative based methods. Some researchers have proposed multivariate regression models (Lei, 1997) and the Bayesian diagnostic models (Kulkarni *et al.*, 1986). In recent years, increasing research efforts seemed to be dedicated on the use of soft computing methods (e.g., fuzzy logic) for deterioration modeling (Sadiq *et al.*, 2004), primarily because field data are usually unavailable or, if available, they are usually qualitative, vague, and associated with a great deal of uncertainty.

There is a great deal of literature describing past and ongoing work on decision making for repair/renew/replacement of water mains. Rajani and Kleiner (2001) and Kleiner and Rajani (2001) provided comprehensive reviews of the published work related to physical and statistical models, respectively. It appears that the vast majority of the work has focused on relatively small distribution mains using breakage frequency as a surrogate measure of deterioration. This approach is only suitable for relatively small pipes where a certain number of breaks can be tolerated prior to pipe renewal. Little work has been done on making decisions prior to pipe failure, which is desirable in large transmission mains where failure consequences can be severe.

In deterioration modeling, the identification of potentially corrosive environments is paramount. If done prior to pipe installation, water utilities can save significant future costs and avoid failures by installing externally coated pipes or providing appropriate corrosion protection. Also the identification of corrosive environment in an existing water distribution network can save resources by concentrating attention on the pipe sections that are at high risk (Seica *et al.*, 2000; Doyle *et al.*, 2003). Corrosion protection measures are usually required in backfills (a term synonymously used with soil(s) in this paper) with low resistivity, high concentration of anaerobic bacteria, differences in soil composition, and differential aeration around the pipe.

Several evaluation processes are currently used to assess conditions that are corrosive to underground piping. The 10-point scoring (10-P) method was introduced by CIPRA (Cast Iron Pipe Research Association, predecessor of DIPRA, Ductile Iron Pipe Research Association) in 1964 for cast iron pipes, which was subsequently extended to ductile iron pipes (ANSI/AWWA C105/A21.5–99). DIPRA (2000) reported that the method has been used to determine soil corrosivity in more than 100 million feet of pipe installations in North America.

In this research, an expert system is developed to estimate pipe deterioration (corrosion rates based on maximum pit depth and pipe age) using a fuzzy model that relates pipe external corrosion to the backfill (soil) properties. This paper presents two aspects of the research: 1) the fuzzy logic expert system, and 2) fuzzy modeling. In the remainder of this section, an overview of the scoring methods as well as the basics of fuzzy sets and soft computing are provided to understand the difference between two approaches.

1.1 Point-scoring methods

The 10-P method is based on five soil properties: resistivity, pH, redox potential, sulfides, and moisture content. A summary of the method is provided in Table 1. If the sum of the scores of all five contributing properties for a given soil sample exceeds 10, the soil is considered *corrosive* to the water mains, requiring corrosion protection measures usually in the form of polyethylene wraps. The method essentially classifies the soil as either *corrosive* or *non-corrosive*. The scoring methods are weighted-averaging procedures in which the weights are implicit in the specific range of scores assigned to each property (attribute or factor). The scoring methods cannot provide information on the intensity of corrosivity. For instance, if the score is 10, the soil is classified as *corrosive*; however, if it is only slightly less than 10, say 9.5, the soil is rated as *non-corrosive* whereas in reality the latter is not significantly different from the former.

Metalogic (1998) proposed another method that takes into account twelve factors including: soil type, soil resistivity, water content, pH, buffering capacity, sulfides, chloride and sulfate concentrations, groundwater level, horizontal and vertical soil homogeneities, and electrochemical potential. The intensity scale used in rating the soil corrosivity is different from the 10-P method i.e., soil corrosivity is divided into 4 categories. Cumulative scores that are less than -10 represent a *highly corrosive* environment whereas positive values (>0) represent *virtually not corrosive* conditions. The remaining two classes, *slightly corrosive* and *corrosive*, lie in the intermediate ranges (Table 2).

Spickelmire (2002) proposed a 25-point scoring method that in addition to soil properties considers pipe factors such as: pipe location and leak repair difficulty, pipe minimum design life, pipe maximum design surge pressure, and pipe size. In this method, the soil corrosivity potential is divided into 4 categories: *mild*, *moderate*, *appreciable*, and *severe* as shown in Table 2.

Although the last two methods address some of the deficiencies of the original 10-P method, they still have the intrinsic problem of weighted-averaging methods (with weights implicit in the various score ranges). Furthermore, these methods do not explicitly address cases where one or more values of soil properties or other required variables are unavailable. A comparison of fuzzy-based method and 10-P method for predicting pipe deterioration has been performed by Sadiq et al. (2004) using two criteria, breakage rate and corrosion pit depth. The results showed that the fuzzy-based method outclasses the predictions of 10-P method for both criteria. The present study offers a more systematic fuzzy-based approach including fuzzy modeling and inferencing in an expert system that is used to predict pipe deterioration based on soil properties.

1.2 Fuzzy Sets and Soft Computing

In recent years, fuzzy-based methods have increasingly been applied to civil and environmental engineering problems from evaluation of concrete structures to water quality (Bardossy *et al.*, 1995; Dou *et al.*, 1995; Guyonnet *et al.*, 2000; Kleiner *et al.*, 2004; Najjaran *et al.*, 2004; Provenzano, 2003 and Provenzano *et al.*, 2004). Fuzzy logic provides a language with syntax and semantics to translate qualitative knowledge into numerical reasoning. In many engineering problems, the available information about the probabilities of various risk items is vaguely known or assessed; and hence, the information in terms of either measured data or expert knowledge is too imprecise to justify the use of crisp numbers. Zadeh (1996) introduced the term *computing with words* (CWW) to explain the notions of reasoning linguistically rather than with numerical quantities. In other words, the main contribution of fuzzy logic to modeling process is a methodology for computing with words.

The decision makers such as water utility managers, regulators, and engineers usually evaluate and describe systems using imprecise terms that may be translated into linguistic variables (e.g., very high, high, very low, low). On the other hand, there is usually some numerical information available for input and output data, although incomplete and uncertain in nature. The strength of fuzzy logic is that it can integrate descriptive (linguistic) knowledge and imprecise numerical data into a fuzzy model and use approximate reasoning algorithms to propagate the uncertainties throughout the decision process. A fuzzy model, as described by Zadeh (1973), contains following three distinguished features:

- linguistic variables instead of, or in addition to numerical variables;
- simple relations between the variables in terms of IF-THEN rules; and
- an inference mechanism that uses approximate reasoning algorithms to formulate complex relationships.

A linguistic variable can be regarded as a variable whose value is a fuzzy number, but fuzzy numbers can also represent numerical variables without being firmly connected to linguistic terms. A fuzzy number is a normal and convex fuzzy set in a continuous universe of discourse U. Finally, a fuzzy set is a collection of ordered pairs $A = \{x, \mu(x)\}$ that describe the relationship between an uncertain quantity x and a membership function $\mu(x)$, where $\mu(x) \in [0, 1]$. An excellent introduction to the fuzzy set theory and fuzzy logic can be found in (Klir and Yuan, 1995; Lee, 1990a, b).

The fuzzy set theory is an extension of the traditional set theory (based on binary logic) in which x is either a member of set A with $\mu(x) = 1$ or not a member of A with $\mu(x) = 0$. Fuzzy logic helps to address the inherent deficiencies of binary logic to account for uncertainties. Fuzzy models formulate the information on an intensity scale.

For example, soil with a score of 9.5 in the 10-P method would be rated non-corrosive, but a fuzzy model might assign the soil as being 0.8 corrosive and 0.2 non-corrosive (depending on predefined qualitative scales of corrosivity). It is anticipated that corrosion protection measures can be selected more efficiently if the degree of soil corrosivity is considered. Further, the qualitative determination of corrosion (deterioration) rates can improve risk assessment.

2 Fuzzy Logic Expert System

A fuzzy expert system is proposed to estimate the deterioration rate of metallic pipes using backfill properties. The expert system consists of two modules: a fuzzy knowledge base and an inference mechanism. The former is essentially a fuzzy model, which has two sources of information: expert knowledge, and field data obtained during the inspection, repair, or renewal of pipelines. The inference mechanism uses the knowledge base to deduce an output that corresponds to observed inputs. The modularized design of the expert system enables it to maintain a generic processing structure that is capable of dealing with various systems in different application domains (e.g., physical, medical, financial) as long as the knowledge base is constructed in a compatible format described later in Section 4.2. Another advantage of the modular design is that the expert system can be updated simply by expanding the knowledge base using new information, as it becomes available.

2.1 Knowledge Base

A fuzzy model determines the relationships between the inputs and outputs of a system using linguistic *antecedent* and *consequent* propositions in a set of IF-THEN rules. The fuzzy model of a multi-input single-output (MISO) system may be formulated in a set of IF-THEN rules as follows:

$$R_i : \text{IF } x_1 \text{ is } A_{i1} \text{ AND } x_2 \text{ is } A_{i2} \text{ AND } \ldots\ x_j \text{ is } A_{ij} \text{ THEN } y \text{ is } B_i\ ,\quad i = 1,\ldots,n \tag{1}$$

where R_i represents the i^{th} rule, n is the total number of rules, x_j $(j = 1,\ldots,r)$ are the input variables, y is the only output variable, A_{ij} are input fuzzy sets defined in the input space specified by r universes of discourse $U = U_1 \times \cdots \times U_r$, and B_i is the output fuzzy set defined in the output universe of discourse V. Thus, every rule is a local fuzzy relationship in $U \times V$ that maps a part of the multidimensional input space U into a certain part of the output space V.

The rule base of a complex system usually requires a large number of rules to describe the behavior of a system for all possible values of the input variables, referred to as "completeness". Hence, the appropriate number of rules depends on the complexity of the system in which the number of fuzzy rules corresponds to the order of a conventional model. This characteristic is similar to the traditional modeling approaches, where optimal model minimizes both the error and the number of rules (Sugeno and Yasukawa, 1993). The aggregation of the rules of equation 1 forms a rule base that is valid over the entire application domain and is given by,

$$R = \bigcup_{i=1}^{n} R_i = \ R_1 \text{ ALSO } R_2 \text{ ALSO } \ldots \text{ ALSO } R_n \tag{2}$$

2.2 Inference Mechanism

Fuzzy inference consists of three connectives: aggregation of *antecedents* in each rule (AND connectives); aggregation of the rules (ALSO connectives); and an inference based on implication relation (i.e. IF-THEN connectives). The type of operators performing these three connectives distinguishes fuzzy inference methods. The AND and ALSO connectives are chosen from a family of *t-norm* and *t-conorm* operators, respectively. Comprehensive discussions on *t-norm* (e.g. *minimum* and *product* operators) and *t-conorm* (e.g. *maximum* and *sum* operators) can be found in (Lee, 1990a, b; Lin, 1994; Emami, 1998). The IF-THEN connectives also use *t-norm* operators, not necessarily identical to the ones used for the AND connectives.

An efficient method of reasoning involves first inferring from individual rules, and then aggregating the results, called first-infer-then-aggregate (FITA). Among all FITA fuzzy reasoning methods, two types of fuzzy reasoning methods are most common in fuzzy logic control and modeling applications. Two inferencing methods are used in the proposed expert system, namely Mamdani's approximation reasoning (Mamdani, 1977) and Larsen's product operation rule (see Appendix A).

Defuzzification is a process to obtain a crisp value y^* that is the best representative of the fuzzy output. The fuzzy output (a possibility distribution) is analogous to a probability distribution function under monotonicity and identity conditions (Filev and Yager, 1991). Numerous defuzzification techniques have been introduced in the literature, but a more practical and generic defuzzification technique, *height method*, is implemented in the proposed expert system. In this technique, the elements of the fuzzy output with a membership value of less than α are disregarded, and the defuzzified value is calculated using the center of area of the elements that have a membership grade of not less than α:

$$y^* = \frac{\int_a^b y\,\mu_{B'}(y)\,dy}{\int_a^b \mu_{B'}(y)\,dy} \tag{3}$$

The *center of area* (COA) and *middle of maximum* (MOM) defuzzification techniques become special cases of *Height method*, when $\alpha = 0$ and $\alpha = \mu_{\max}(y)$, respectively.

3 Fuzzy Logic Modeling

Two basic approaches namely direct approach and system identification are used for fuzzy logic modeling (Yager and Filev, 1994a). In the direct approach, the information extracted from experts' knowledge is used to:

- specify the input, state, and output variables;
- determine the partitions of input and output variables in their universes of discourse, and may label the partitions with appropriate linguistic terms;
- define a set of linguistic (IF-THEN) rules that represent the relationships between the system variables;
- select an appropriate reasoning method; and
- evaluate the model adequacy.

Although *direct approach* is simple, it has inherent limitations because quantitative observations of the performance of the system are not explicitly used in the determination of the structure or parameters of the model. Thus, the adequacy of the model is restricted to the boundaries of the expert knowledge; if the expert knowledge about the system is faulty, then so is the model.

The *system identification* approach involves the use of input-output data of systems to introduce new knowledge and increase the objectivity of fuzzy modeling (Zadeh, 1991). System identification is divided into two components: structure identification, and parameter identification (Sugeno and Yasukawa, 1993). Similar to the direct approach, structure identification involves the determination of input and output variables, partitions of input and output variables (fuzzy sets), relationships between the input and output variables (IF-THEN rules) through the number of rules. Parameter identification involves the adjustment of the input and output membership functions.

The input variables are selected from a number of input candidates that most likely affect the output. Typically, there is no systematic way to specify the input candidates, so a heuristic method based on experience or common sense is often recommended. Given a finite number of input candidates, the input variables are selected based on the regularity criterion in the group method of data handling (Ihara, 1980) using the system input-output data.

The most important step to establish the fuzzy model is to generate the rules. Clustering of the input-output data is an intuitive approach to objective rule generation. The idea of clustering is to divide the output data into a certain number of fuzzy partitions. The appropriate number of clusters is determined so that the sum of the *Euclidian distance* of the output data from the center of the clusters is minimized. The determination of the number of rules is another important step for fuzzy modeling. A large number of rules, similar to a high order of a model, will bias the model towards specific data that are usually imprecise and subject to noise. On the other hand, less number of rules will likely increase the output error that is essentially equivalent to disregarding some valuable information. Thus, the optimal number of rules, n, is obtained so that both the output error and n are minimized.

4 Pipe Deterioration Knowledge Base

Figure 1 shows the structure of the expert system used to determine the deterioration rate of ductile and cast iron water mains using backfill soil properties. The inference mechanism is the generic inference tool that was discussed earlier. A two-tier fuzzy modeling process that involves both approaches of fuzzy modeling, namely, *direct approach* and *system identification* introduced in the previous sections, provides the knowledge base of the expert system.

Direct approach yields a subjective model based on expert knowledge. The subjective model adopts the qualitative aspect of fuzzy modeling. Input variables are selected from pertinent soil properties such as soil resistivity, pH, etc. The output is a soil corrosiveness criterion named corrosivity potential (*CoP*) that is a real number between 0 and 1, referring to *non-corrosive* and *most-corrosive* soil, respectively. The expert knowledge was established from published work on the condition assessment of water mains and then revised based on the results of a survey submitted to distinguished corrosion experts worldwide. The survey was conducted in the Internet where corrosion specialists were asked a variety of questions on how different soil properties influence corrosion of cast and ductile iron water mains.

The expert knowledge is used to select the input variables, determine the input partitions, and assign linguistic variables. It is also used to define the IF-THEN rules relating the soil properties to *CoP*. The subjectivity of the model is related to the descriptive basis of *CoP* and the fact that experts cannot always provide a quantitative relationship between the input and output variables in the model. Specifically, experts describe the severity of corrosion based on backfill soil properties but cannot quantify the deterioration of buried pipes, which is the result of an extremely random phenomenon, even if accurate measurements of the properties are available.

On the other hand, system identification provides an objective model that is exclusively based on field data, which may be obtained from nondestructive inspection of buried pipes or examination of exhumed pipes. The inputs to the objective model are identical to the subjective model, but the output cannot be the same because the model now requires a measurable quantity, such as breakage frequency or maximum pit depth. It is worth noting that traditional modeling techniques, such as linear regression, are inappropriate for this application because of high uncertainties associated with inaccurate field data. Above all, the data are scarce, which makes it too hard to establish an efficient outlier rejection procedure or use of statistical analysis. Fuzzy modeling provides a synergy between the subjective and objective models and augment the expert knowledge using the “imprecise” filed data, as well as simplify *system identification* that is usually an ill-defined process and not amenable to automated techniques.

Consequently, there is a need for a strategy to fuse the information of the subjective and objective models to generate the knowledge base. The difficulty in fusing the two models is that the outputs of the two models are not commensurate because experts cannot always explain the effect of soil properties on deterioration of pipes in a manner that is similar to that obtained from measurements (using instruments and sensors). Thus, it is proposed that the subjective model be used to aggregate multiple input parameters (i.e., soil properties) into a single entity, *CoP*. The latter is then used as an intermediate input parameter to generate the rules of the objective model, which is now reduced to a model of a single-input-single-output (SISO) system.

The fundamental idea of aggregating properties of fuzzy rules and their determination from input-output data were put forward by Zadeh (1971). Yager and Filev (1994b) provided background to these ideas to what they call template-based methods. This approach combines the expert knowledge and data. The expert knowledge provides templates for linguistic variables that are used to partition the input-output space, the fuzzy subsets (numbers) are given a priori. These template values are used to define the potential rules for the fuzzy system model. Input-output data are then used to generate weights or probabilities associated with the importance of the potential rules. Thus, the emphasis is on learning the weights (or *credibility*) of the rules.

The proposed expert system is developed in two parts. The first part focuses on the structure of the expert system and develops a subjective fuzzy model using the *direct approach* (i.e., based on expert knowledge). The application of the expert system for the determination of corrosivity potential (*CoP*) is illustrated via two examples in Subsection 4.1. The second part, the development of system identification to generate the objective model and subsequent fusion of the subjective and objective models will be described in subsequent publications.

4.1 Determination of Corrosivity Potential *(CoP)*

Two examples for the determination of *CoP* of cast iron mains using the proposed expert system are presented. Ductile iron pipes were not analyzed because sufficient field data were not available; nonetheless, the approach described here is equally applicable.

The first example demonstrates a knowledge base that consists of three input variables: soil resistivity, pH, and redox potential. Figure 2 shows the rule base that is based on expert knowledge. The input and output fuzzy sets, specified by the four edges of the trapezoids, are derived from the ranges used in the 10-P method. All input variables are partitioned by three fuzzy numbers low (L), medium (M), and high (H). The output has five fuzzy numbers very low (VL), low (L), medium (M), high (H), and very high (VH). Each fuzzy number is specified by four edges of a trapezoid. The rule base includes thirteen rules where the fuzzy numbers of the input variables in some of the antecedent propositions are concatenated to reduce the number of the rules. For example, the fuzzy number MH in soil pH shows that the rules 2, 12 and 13 correspond to both medium and high pH. As a result, the number of rules for the three 3-partition inputs is reduced to 13 (instead of 27 rules originally required). It is noted that the fuzzy set MH {3, 5, 12, 12} for pH is somewhat greater than the union of M {3, 5, 8, 10} and H {8, 10, 12, 12}; and hence, the concatenated rules may infer greater values in the areas between the two fuzzy sets (the shaded area of the pH fuzzy number in Figure 2). However, concatenation is justifiable under a reasonable assumption that a valid rule for both M and H is most likely valid for the values between them. The concatenation of the input fuzzy sets reduces the number of rules significantly, especially in rule bases with a large number of input variables. Finally, the linguistic variable LMH of an input variable in a rule implies that the input variable is ineffective; the output is specified regardless of the value of the input variable.

The universes of discourse of the input variables are determined based on the minimum and maximum values of the soil properties, induced by the definition (e.g., pH) or obtained from published data. Naturally, the maximum and minimum values of a parameter with an infinite range cast into a range in which the parameter has an effect on the output. For example, a soil resistivity of greater than 5000 Ω-cm is unlikely to have an effect on *CoP*, so the maximum value of resistivity is considered to be 5000 Ω-cm.

Although *CoP* is defined in the interval [0, 1], one cannot expect a defuzzified value for *CoP* greater than 0.774 (i.e., the center of area of the VH fuzzy number) and less than 0.033 (i.e., the center of area of the VL fuzzy number). It is noted that the defuzzified value is only a representative of the fuzzy output interval and this is not the limitation of the rules. More precisely, when the inferred fuzzy output is VH it means that *CoP* is in the interval [0.5, 1], and the best representative of this interval is 0.774.

Figure 3 shows the knowledge base of the second example that is based on five soil properties: soil resistivity, pH, % clay fines by weight (diameter < 0.002 mm), redox potential, and sulfide content. The rules are defined based on expert knowledge. The input and output fuzzy sets were derived from the ranges used in the 10-P method. All properties, except the percentage of clay fines, are similar to the parameters considered in the 10-P method. The percentage of clay fines replaces moisture content in the 10-P method in view of the fact that it reflects moisture retention capacity. Soil types are defined in Table 3 in broad categories ranging from gravel to clay in terms of percentage of clay fines present. Again, all input variables are defined using three fuzzy numbers (linguistic partitions), and the output, *CoP*, is defined using five fuzzy numbers. Concatenation reduces the number of rules to 45 (instead of 243 rules originally required for five 3-partition inputs). A comparison between the results obtained from the 3-input and 5-input knowledge bases is presented in the next section.

4.2 Deterioration Rate vs. Corrosivity Potential

The validation of the proposed expert system and knowledge base is carried out using the measurements of pipe age, soil properties, and maximum pit depth available from a previous study on cast iron mains (Rajani *et al.*, 2000). It is expected that a correlation between the *CoP* and corresponding deterioration rate will show that the hypothetical *CoP* calculated by the expert system is an appropriate indicator of the corrosion process. The soil properties and pit depth measurements were a snapshot of current conditions, and thus deterioration rates (*DR*) obtained from maximum pit depth and pipe age represent an average rather than instantaneous values. Table 4 shows the *CoP* values that are calculated by the expert system for a series of soil samples using the 3-input and 5-input knowledge base. Figure 4 shows the relationship between *DR* and *CoP* obtained using the 3-input and 5-input knowledge bases. The relationship shown in Figure 4 helps define the proposed strategy for the fusion of objective and subjective models. Figure 4 suggests that the deterioration rate is reasonably correlated with *CoP* obtained using the 5-input knowledge base; i.e., the higher the *CoP* the higher the deterioration rate. However, the *CoP* values obtained from the 3-input knowledge base do not follow any trend and show a poor correlation. Figure 4 shows that the *DR* versus *CoP* data points mostly fall in a region bounded between the two dashed lines, but the scatter occurs due to two reasons.

First, the subjective model of the fuzzy knowledge base is imprecise for a certain range of *CoP*. This may hint that either the number of fuzzy rules in the rule base are insufficient (i.e., the rule base does not satisfy the "completeness" condition), or the input and output partitions are not appropriately tuned in some range of their respective universe of discourse. Tuning up the model using field data, which will be addressed in future research, can alleviate these issues. Further, it could also mean that one or more input variables can become more dominant in certain ranges of *CoP*, but these variables have not been taken into account. The identification of the other input variables is challenging because measurements of a variety of input parameters must be available before selecting the most pertinent ones. This issue may be addressed to certain degree using the expert survey[1] to be more selective on possible inputs.

Second, obvious outliers among the field data exist that must be excluded from the database used in objective modeling. An accurate pipe deterioration rate is typically unavailable because the deterioration rate is determined under the debatable assumption of an average (constant) corrosion rate from the installation to exhumation of the pipe. Also, the corrosion rate is obtained by measuring the maximum pit depth in a few pipe sections that are randomly selected. Consequently, the choice of the pit as well as the measurement techniques imposes a great deal of uncertainty on the measurements. Issues such as manufacturing defects, changing water table, backfill chemistry (e.g., addition of salt during winter) and disturbance of backfill soil have an impact on the reliable determination of deterioration rates. For example, the first two rows of Table 4 refer to identical soil samples roughly, yet the corresponding deterioration rates are significantly different. These points are highlighted within Figure 4.

The uncertainty management in the proposed expert system relies on the fusion of the objective and subjective models. Research on the uncertainty management and its impacts on the assessment of *corrosivity potential* are ongoing, and therefore related issues will only be briefly addressed here.

The uncertainty of the deterioration model can be described by upper and lower values for deterioration rates with respect to *CoP*. The simplest example is a linear function between the deterioration rate, *DR*, and *CoP* that is given by:

$$DR = m \cdot CP + d \tag{4}$$

where m and d are the slope and intercept of the line, respectively. If the assumed expected values of the slope and intercept are $\overline{m} = E(m)$ and $\overline{d} = E(d)$ then the estimated values are as follows:

$$\begin{aligned} m &= \overline{m} + \varepsilon_m \\ d &= \overline{d} + \varepsilon_d \end{aligned} \tag{5}$$

where ε_m and ε_d are the slope and intercept disturbances, which are normally distributed with zero mean. Thus, for the linear relationship of equation 4, the uncertainty of the deterioration rate can be estimated by a linear combination of the normal distributions of ε_m and ε_d. Thus, it is possible to define a confidence level for the values obtained from the linear relationship between *DR* and *CoP*. The issues related to parameter identification and relationship between the deterioration rate and *CoP* will be examined in detail as more field data become available.

The expert system facilitates the uncertainty management process by providing not only a defuzzified crisp value for *CoP* but also an output fuzzy set that specifies memberships for all *CoP* values. Figure 5 presents the defuzzified value and *CoP* memberships of a soil sample corresponding to the marked point in Figure 4. The membership values may be represented by a membership function $\mu(CP)$, or converted to a probability distribution under the identity and monotonic conditions. Either the membership function or probability distribution can provide a reliability measure for *CoP*, which is now the independent variable. Recently, Davis *et al.* (2003) developed a relationship between the average failure rate of cast iron pipes and the point scores obtained from the 10-P method. The major shortcoming of their approach is that the 10-P method cannot provide an uncertainty measure for the soil score. Thus, it is impossible to determine the reliability of the soil scores required for uncertainty management. The other problem with their approach is that the average failure rate or breakage frequency can be affected by many other distressing factors besides soil properties. However, in the approach proposed here, only the influence of soil properties on deterioration rate as a result of external corrosion is considered.

It is anticipated that an enhanced fuzzy knowledge base of the proposed expert system will yield *corrosivity potentials* that are strongly correlated to deterioration rates. Hence, the predictions through this expert system can lead us to estimate the minimum remaining wall thickness of the pipes given the surrounding soil properties and predict the time of failure. Unlike the binary states of the corrosivity (*corrosive* vs. *non-corrosive*) obtained from the 10-P method, *corrosivity potential* can also be used to gauge the level of required corrosion protection. Specifically, the interval [0, 1] of the *corrosivity potential* can correspond to the six levels of corrosion protection measures recommended for ferrous pipe materials (Dechant and Smith, 2004). A more rigorous approach to match the *corrosivity potential* with a specific corrosion protection measure would require performing cost-benefit analysis.

5 Conclusions

A fuzzy expert system is proposed to determine the deterioration rate of cast and ductile iron water mains based on the backfill soil properties. The expert system consists of two modules: inference mechanism and knowledge base. The former is a generic inference tool based on the fuzzy set theory that can process the knowledge base of an arbitrary application as long as the encoded information is provided in an appropriate format. The knowledge base is developed in a twofold fuzzy modeling process. First, a subjective fuzzy model is developed using the direct approach of fuzzy modeling based on the information obtained from published literature and an online expert survey. Second, the *system identification* approach is used to develop an objective model based on the field data.

The subjective model provides a fuzzy relationship between a number of soil properties, perceived as the most significant contributors to the corrosion of cast and ductile iron pipes, and a proposed corrosiveness criterion (viz., corrosivity potential, *CoP*). It is shown that corrosivity potential is correlated with the deterioration rate, according to the field data. More precisely, the objective model refers to the deterioration rate as a function of *corrosivity potential* that is independently determined by the subjective model. As a result, the deterioration analysis is simplified significantly by considering only one parameter affecting the deterioration of the pipes. Further, *corrosivity potential* can be used to establish a cost-benefit analysis and determine the optimal level of corrosion protection required in municipal infrastructure based on the soil properties.

Future research will establish a relationship between the deterioration rate of cast/ductile iron pipes and *corrosivity potential* based on the field data. The relationship will link the objective model with the subjective model to enhance the knowledge base. The fusion of the two models will predict the deterioration rate more accurately and provide uncertainty measures. This will also help to explore the implicit weights and threshold values of 10-P method comprehensively, which are used as base line in the present research.

References

ANSI/AWWA C105/A21.5–99 (1999), American national standard for polyethylene encasement for ductile-iron pipe systems, *American Water Works Association*, Denver, CO.

Bardossy, A., Bronstert, A., and Merz, B. (1995), 1-, 2- and 3-Dimensional modeling of groundwater movements in the unsaturated soil matrix using a fuzzy approach, *Advances in Water Resources*, 18(4): 237-251.

Davis, P., Allan, I., Burn, S., and van de-Graaff, R. (2003), Identifying trends in cast iron pipe failure with GIS maps of soil environments, *Pipes Wagga 2003, Back to Basic: Design and Innovation*, Wagga City, Australia.

Dechant D. A., Smith G. (2004), Present levels of corrosion on ferrous water piping in municipal infrastructure: a manufacturer's perspective, *Materials Performance*, 43(1): 54-57.

DIPRA (2000), Polyethylene encasement – effective economical protection for ductile iron pipe in corrosive environments, Company Internet Website, Ductile Iron Pipe Research Association.

Dou, C., Woldt, W., Bogardi, I., and Dahab, M. (1995), Steady state groundwater flow simulation with imprecise parameters, *Water Resources Research*, 31(11): 2709-2719.

Doyle, G., Seica, M. V., and Grabinsky, M. W. F. (2003), The role of soil in the external corrosion of cast iron water mains in Toronto, Canada, *Canadian Journal of Civil Engineering*, 40: 225-236.

Emami, M. R., Turksen, I. B., and Goldenberg, A. A. (1998), Development of a systematic methodology of fuzzy logic modeling, *IEEE Transactions on Fuzzy Systems*, 6(3): 346-361.

Filev, D. P., and Yager, R. R. (1991), A generalized defuzzification method under BADD distribution, *International Journal of Intelligent Systems*, 6: 687-697.

Guyonnet, D. G., Come, B., Perrochet, P., and Parriaux, A. (2000), Comparing two methods for uncertainty in risk assessments, *ASCE Journal of Environmental Engineering*, 125(7): 660-666.

Ihara, J. (1980), Group method of data handling towards a modeling of complex systems-IV. *Systems and Control* (In Japanese), 24, 158-168.

Kleiner, Y., and Rajani, B. (2001), Comprehensive review of structural deterioration of water mains: statistical models, *Urban Water*, 3(3): 131-150.

Kleiner, Y., Sadiq, R., and Rajani, B. (2004), Modeling failure risk in buried pipes using fuzzy Markov deterioration process, *Proceedings of ASCE Pipelines2004*, San Diego, CA.

Klir, G. J., and Yuan, B. (1995), *Fuzzy sets and fuzzy logic – theory and applications*, Prentice-Hall Inc., Englewood Cliffs, NJ, US.

Kulkarni, R. B., Golabi, K., and Chuang, J. (1986), Analytical techniques for selection of repair-or-replace options for cast iron gas piping systems – Phase I, *Gas Research Institute*, PB87-114112, Chicago, I1.

Lary, J. (2000), Corrosion, not age, is to blame for most water main breaks, *Water World*.

Lee, C. C. (1990a), Fuzzy logic in control systems: Fuzzy logic controller-Part I, *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2): 404-418.

Lee, C. C. (1990b), Fuzzy logic in control systems: Fuzzy logic controller-Part II, *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2): 419-435.

Lei, J. (1997), Statistical approach for describing lifetimes of water mains – Case Trondheim Municipality, *SINTEF Civil and Environmental Engineering*, Report No. 22F007.28, Trondheim, Norway.

Lin, C. F. (1994), *Advanced control systems design*, Prentice Hall Inc. Englewood Cliffs, NJ, US.

Mamdani, E. H. (1977), Application of fuzzy logic to approximate reasoning using linguistic synthesis, *IEEE Transactions on Computers*, 26(12): 1182-1191.

Metalogic, N. V. (1998), Company Internet Website, <www.metalogic.be>, Metalogic N.V., Heverlee, Belgium.

Najjaran, H., Sadiq, R., and Rajani, B. (2004), Fuzzy Logic Expert System for Assessing Corrosion of Cast/Ductile Iron Water Mains using Soil Properties, *Proceedings of ASCE Pipelines2004*, San Diego, CA.

Provenzano, P. (2003), A fuzzy-neural network method for modeling uncertainties in soil structure interaction problems, *Computer Aided Civil and Infrastructure Engineering*, 18: 391-411.

Provenzano, P., Ferlisi, S., and Musso, A. (2004), Tentative interpretation of a model footing response through an adaptive neural fuzzy inference system, *Computer and Geotechnics*, 31: 251-266.

Rajani, B., and Kleiner, Y. (2001), Comprehensive review of structural deterioration of water mains: Physical based models, *Urban Water*, 3(3): 151-164.

Rajani, B., Makar, J., McDonald, S., Zhan, C., Kuraoka, S., Jen, C. K., and Veins, M. (2000), Investigation of grey cast iron water mains to develop a methodology for estimating service life, *American Water Works Association Research Foundation (AWWARF)*, Denver, CO.

Sadiq, R., Rajani, B., and Kleiner, Y. (2004), A fuzzy based method of soil corrosivity evaluation for predicting water main deterioration, *ASCE Journal of Infrastructure Systems*, 10(4): 149-156.

Seica, M. V., Packer J. A., Grabinsky M. W. F., Adams B. J., and Karney B. W., (2000) Evaluation and testing of cast iron and ductile iron water mains samples, *Final Report to City of Toronto,* Department of Civil Engineering, University of Toronto.

Spickelmire, B. (2002), Corrosion consideration for ductile iron pipe, *Materials Performance*, 41: 16-23.

Sugeno, M., and Yasukawa, T. (1993), A fuzzy-logic-based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems*, 1(1): 7-31.

Yager, R. R., and Filev, D. P. (1994a), *Essentials of fuzzy modeling and control*, John Willey & Sons Inc., NY, US.

Yager, R. R., and Filev, D. P. (1994b), Template based fuzzy systems modeling, *Journal of Intelligent Systems*, 2: 39-54.

Zadeh, L. A. (1971), Towards a theory of fuzzy systems, *Aspects of Network and System Theory*, R. Kalman, and N. DeClaris, Eds., Holt, Reinhart & Winston, New York 469-490.

Zadeh, L. A. (1973), Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Transactions on Systems, Man, and Cybernetics,* 3: 28-44.

Zadeh, L. A. (1991), *From circuit theory to system theory, in facets of systems science*, G. L. Klir Eds., New York, Plenum Press.

Zadeh, L. A. (1996), Fuzzy logic computing with words, *IEEE Transactions on Fuzzy Systems*, 4(2): 103-111.

Table 1 Scores of soil properties used in the 10-P scoring method

Soil	Values and characteristics	Points
Resistivity (Ω-cm)	< 1,500	10
	≥ 1,500 - 1,800	8
	> 1,800 - 2,100	5
	> 2,100 - 2,500	2
	> 2,500 - 3,000	1
	> 3,000	0
pH	0 - 2	5
	2 - 4	3
	4 - 6.5	0
	6.5 - 7.5	0
	7.5 - 8.5	0
	>8.5	3
Redox potential (mV)	> +100	0
	+50 - +100	3.5
	0 - +50	4
	< 0	5
Sulfides	Positive	3.5
	Trace	2
	Negative	0
Moisture	Poor drainage (continuously wet)	2
	Fair drainage (generally moist)	1
	Good drainage (generally dry)	0

Table 2 Other point scoring methods

Metalogic (1998)	
Soil Corrosivity	$\sum_{i}^{12} r_i$
Virtually not corrosive	> 0
Slightly corrosive	-1 to -4
Corrosive	-5 to -10
Highly corrosive	<-10

Spickelmire (2002)	
Soil Corrosivity	$\sum_{j}^{15} s_j$
Mild	0 to 14.5
Moderate	15 to 19.5
Appreciable	20 to 24.5
Severe	> 25

Table 3 Percentage clay fines for different soil types

Soil Type	% clay (soil particles < 0.002 mm) fines by weight
Granular material (gravel)	15
Sand	22
Silty sand	25
Silt	30
Silty clay	35
Clay	>40

Table 4 Field data and calculated corrosivity potential (*CoP*)

R •-cm	*pH*	Clay fines %	Redox potential mV	Sulfide presence	DR mm/yr	Corrosivity Potential (*CoP*) No of inputs	
						3	5
590	7.7	30	-29	1	0.025	0.467	0.774
580	7.7	30	-30	1	0.042	0.467	0.774
470	7.7	30	-39	1	0.022	0.467	0.774
1163	7.8	30.00	250	1	0.054	0.450	0.439
1575	5.8	22	309	-1	0.044	0.392	0.339
634	8.2	22	-103	-1	0.046	0.467	0.774
5417	7.4	42	-42	-1	0.090	0.133	0.133
…	…	…	…	…	…	…	…
3100	6.3	42	306	0	0.033	0.033	0.199
12929	4.6	22	268	0	0.027	0.033	0.185
14126	5.2	22	171	0	0.035	0.033	0.133
1640	5.7	42	177	0	0.048	0.381	0.321
1560	5.2	42	203	0	0.067	0.395	0.336
1300	7.6	22	-166	0	0.059	0.415	0.702
6700	5.5	22	-88	0	0.055	0.133	0.300

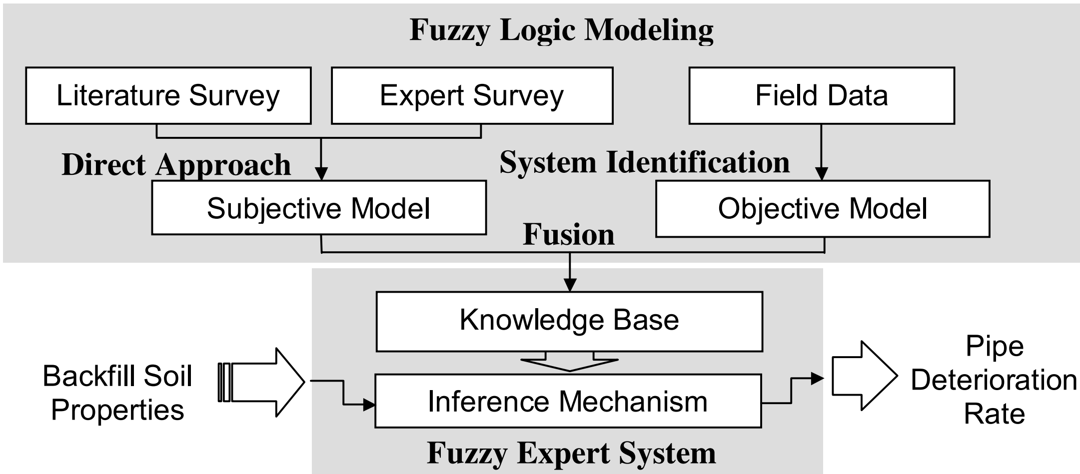
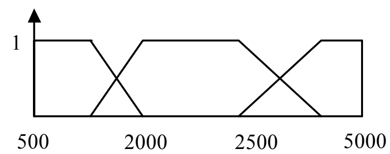
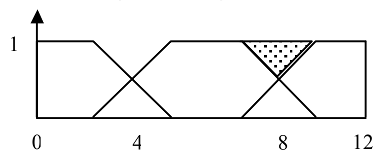


Figure 1 Structure of the fuzzy expert system

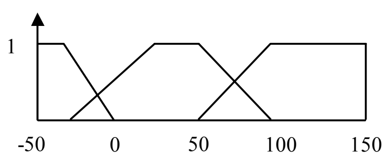
I1: Soil Resistivity (Ω-cm)				
L	500	500	1000	2500
M	1000	2000	2500	4000
H	2500	4000	5000	5000



I2: pH				
L	0	0	3	5
M	3	5	8	10
H	8	10	12	12



I3: Redox Potential (mV)				
L	-50	-50	-20	0
M	-20	25	50	100
H	50	100	150	150



Rule	Rule Base			
No	I1	I2	I3	O
1	L	L	LMH	VH
2	L	MH	LMH	H
3	M	L	LM	H
4	M	L	H	M
…	…	…	…	…
11	H	L	H	VL
12	H	MH	L	L
13	H	MH	MH	VL

O: Corrosivity Potential (*CoP*)				
VL	0	0	0	0.1
L	0	0.1	0.1	0.3
M	0.1	0.2	0.4	0.5
H	0.3	0.5	0.5	0.6
VH	0.5	0.6	1	1

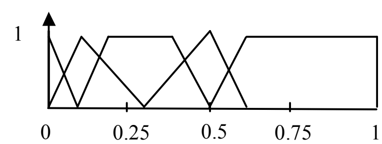
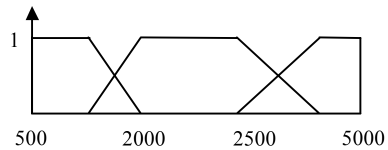
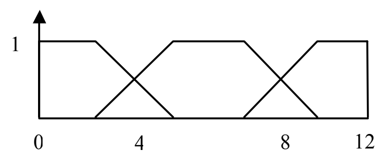


Figure 2 Fuzzy knowledge base with 3-input variables

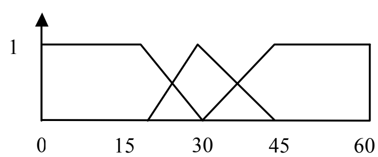
I1: Soil Resistivity (Ω-cm)				
L	500	500	1000	2500
M	1000	2000	2500	4000
H	2500	4000	5000	5000



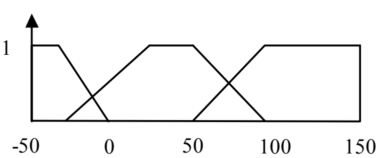
I2: pH				
L	0	0	3	5
M	3	5	8	10
H	8	10	12	12



I3: % Clay Fines				
L	0	0	20	30
M	20	30	30	45
H	30	45	60	60

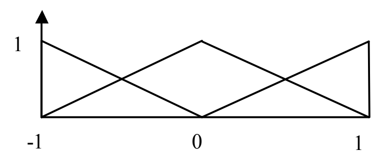


I4: Redox Potential (mV)				
L	-50	-50	-20	0
M	-20	25	50	100
H	50	100	150	150



Rule No	Rule Base I1	I2	I3	I4	I5	O
1	L	L	LMH	LM	LMH	VH
2	L	L	LMH	H	LMH	H
3	L	MH	LMH	LM	LMH	VH
4	L	MH	LMH	H	LMH	H
5	M	L	LM	L	LMH	VH
6	M	L	LM	MH	L	H
7	M	L	LM	MH	MH	VH
…	…	…	…	…	…	…
41	H	MH	L	MH	MH	L
42	H	MH	MH	L	L	L
43	H	MH	MH	L	MH	M
44	H	MH	MH	MH	L	VL
45	H	MH	MH	MH	MH	L

I5: Sulfide				
L	-1	-1	-1	0
M	-1	0	0	1
H	0	1	1	1



O: Corrosivity Potential (*CoP*)				
VL	0	0	0	0.1
L	0	0.1	0.1	0.3
M	0.1	0.2	0.4	0.5
H	0.3	0.5	0.5	0.6
VH	0.5	0.6	1	1

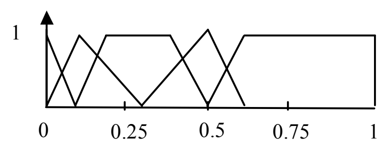


Figure 3 Fuzzy knowledge base with 5-input variables

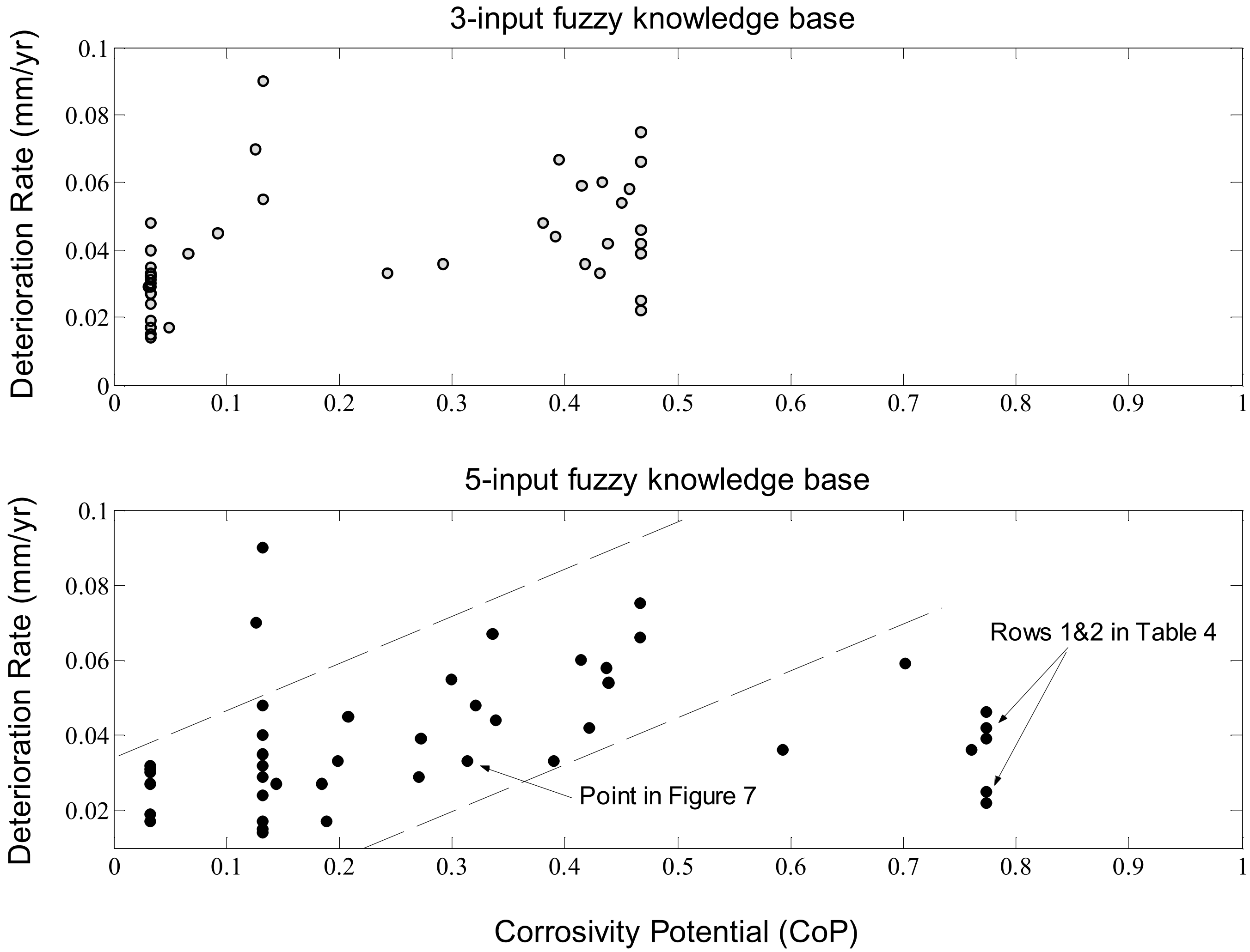


Figure 4 Correlation of deterioration rate with corrosivity potential

Example: 5-input …		
Resistivity	Ω-cm	2145
pH		7.9
Soil fines	%	30
Redox potential	mV	354
Sulfide		1
Corrosivity potential (*CoP*)		0.31

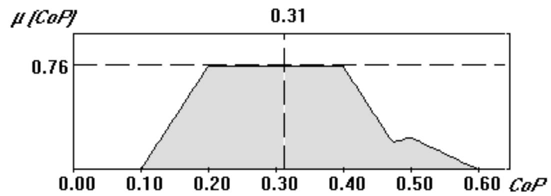
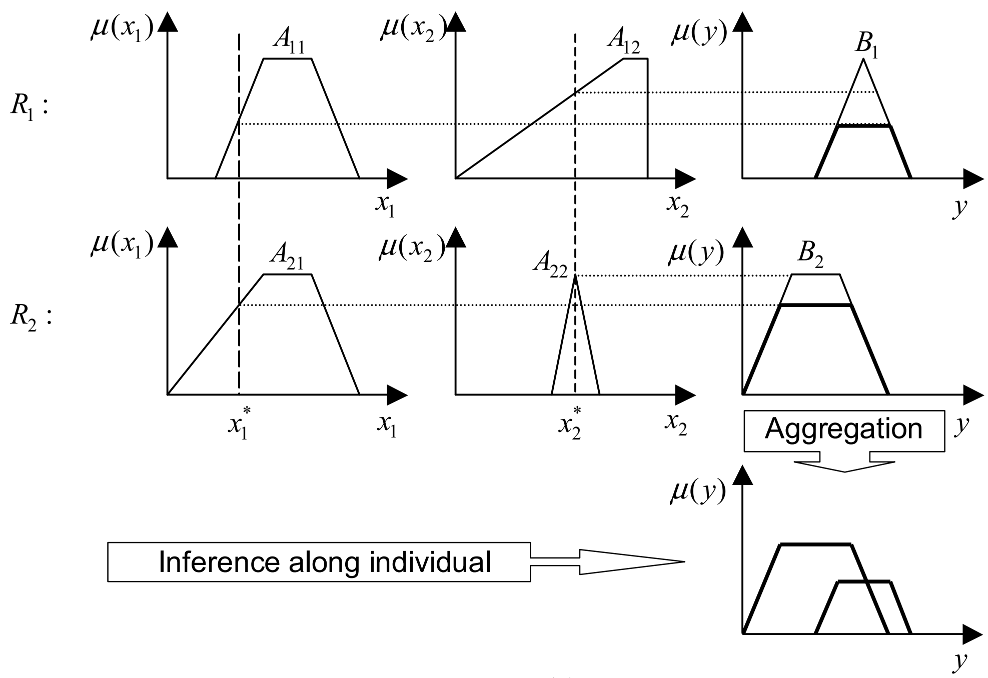


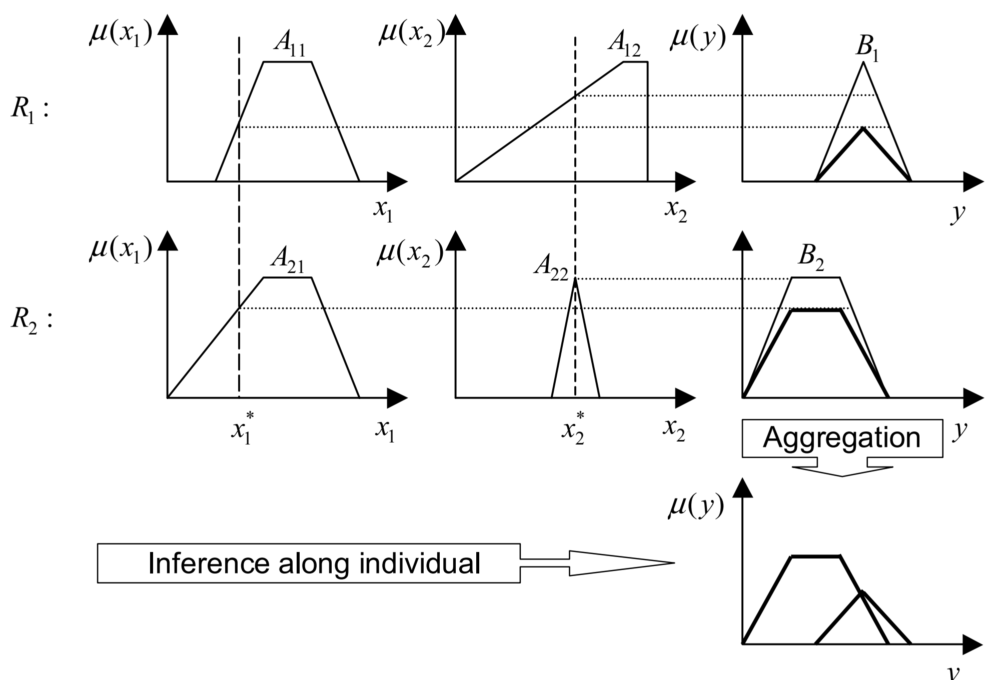
Figure 5 Defuzzified value and memberships of corrosivity potential (*CoP*) for a soil sample

Appendix A

FITA approximate reasoning using a) Mamdani's minimum operation rule and b) Larsen's product operation rule as the implication rule



(a)



(b)