Modelling the deterioration of buried infrastructure as a fuzzy Markov process

Yehuda Kleiner, Rehan Sadiq and Balvant Rajani

ABSTRACT

Dearth of data is the greatest acknowledged obstacle to the deterioration modeling of the buried infrastructure assets. In the last two decades numerous models have been proposed with a greater emphasis on Markovian Deterioration Processes (MDP). The MDP requires that the condition of the deteriorating system be encoded as an ordinal condition rating, based on numerous distress indicators obtained possibly from direct and indirect observations, as well as from non-destructive tests. The encoding of distress indicators into condition rating is inherently imprecise and involves subjective judgment. This imprecision is not considered, let alone propagated in the traditional application of the MDP.

In this paper a new approach is presented to model the deterioration of buried infrastructure assets using a fuzzy rule-based, non-homogeneous Markov process. This deterioration model yields the 'possibility' of failure at every time step along the life of the asset. The use of fuzzy sets and fuzzy techniques help to incorporate the inherent imprecision and subjectivity of the data, as well as to propagate these attributes throughout the model, yielding more realistic results.

This paper is the first of two companion papers that describe a complete method of managing failure risk of large buried infrastructure assets. The second companion paper describes how the condition ratings along the life of the asset are converted to risk values and how effective decisions can be made about the renewal and/or scheduling the next inspection of the asset.

Key words | condition rating, deterioration, failure risk, fuzzy Markov, infrastructure, large water mains

Yehuda Kleiner (corresponding author)
Senior Research Officer and Group Leader,
Institute for Research in Construction,
National Research Council Canada (NRC),
1200 Montreal Road, Building M-20, Ottawa ON,
Canada K1A 0R6
Phone: 1-613-993-3805
Fax: 1-613-954-5984
E-mail: *Yehuda.Kleiner@nrc-cnrc.gc.ca*

Rehan Sadiq
Research Officer, Institute for Research in Construction, National Research Council Canada (NRC),
1200 Montreal Road, Building M-20, Ottawa ON,
Canada K1A 0R6
Phone: 1-613-993-6282
Fax: 1-613-954-5984
E-mail: *Rehan.Sadiq@nrc-cnrc.gc.ca*

Balvant Rajani
Principal Research Officer, Institute for Research in Construction, National Research Council Canada (NRC),
1200 Montreal Road, Building M-20, Ottawa ON,
Canada K1A 0R6
Phone: 1-613-993-3810
Fax: 1-613-954-5984
E-mail: *Balvant.Rajani@nrc-cnrc.gc.ca*

NOTATION

A_i $(i = 1, 2, \ldots, 5)$	fuzzy triangular subsets (levels) in the fuzzy set A, which defines pipe age
A_t	fuzzy number representing the pipe fuzzy age at time step t
C_i $(i = 1, 2, \ldots, 7)$	fuzzy triangular subsets (states) in the fuzzy set C, which defines pipe condition
d_0	base deterioration unit
D'_i $(i = 1, 2, \ldots, 5)$	fuzzy triangular subsets (levels) in the fuzzy set D', which defines pipe deterioration rate
D'_t	fuzzy number representing the pipe fuzzy deterioration rate at time step t
$D_t^{i,j}$	deterioration rate at time step t from condition state i to condition state j
g_i	zero/one binary variables to compute membership 'flow' from state i to state j $(j = i + 1)$
G_i	threshold values which restrict the membership 'flow' from state i to state j $(j = i + 1)$
$h_x(u)$	height of a fuzzy set, or maximum membership in a fuzzy set
R_D	rule set governing the deterioration D' as a function of fuzzy age A and fuzzy condition C

doi: 10.2166/aqua.2006.074

$\mu_t^{D'_i}$	membership of a pipe to the fuzzy deterioration subset D'_i at time step t
$\mu_t^{A_i}$	membership of a pipe to the fuzzy age subset A_i at time step t
$\mu_t^{C_i}$	membership of a pipe to the fuzzy state (subset) C_i at time step t

INTRODUCTION

Large buried infrastructure assets, such as large-diameter trunk sewers and large water transmission mains, typically have low failure rates but when they fail the consequences can be disastrous. This low rate of failure, coupled with high cost of inspection and condition assessment, have contributed to the current situation where most municipalities lack the data necessary to model the deterioration rates of these assets and subsequently to make rational decisions regarding their renewal.

The condition assessment of a buried infrastructure asset is a costly process, and can be viewed as consisting of two distinct components. The first component involves the inspection of the asset using direct observation (visual, video) and/or non-destructive evaluation (NDE) techniques (radar, sonar, ultrasound, sound emissions, eddy currents, etc.). Inspection of an asset yields quantification and location(s) of distress, e.g., 2 mm wide crack at spring level and 19 broken wires, located 2 m from the bell end of a prestressed concrete cylinder pipe (PCCP). The second component of condition assessment involves the translation (Rajani *et al*. 2006) of these distress indicators to the condition rating of the asset. In this paper, the terms *distress indicator, observed distress indicator, and observation* are used interchangeably. The condition rating of an asset reflects the combined result of all observed distress indicators for one asset segment.

Managing the failure risk of infrastructure assets is improved with the use of a deterioration model to enable the forecast of the asset condition as well as its likelihood of failure in the future. The last two decades have witnessed a significant effort to model infrastructure deterioration. Various approaches have been suggested, but the use of Markovian Deterioration Process (MDP) has clearly dominated. For example, Madanat *et al*. (1997) used the probit model to show that the Markov deterioration process in bridge decks is state-dependent, i.e., non-homogeneous. Li *et al*. (1997) modelled deterioration as a non-homogeneous Markov process with ten condition states. Later they improved this method by incorporating a Bayesian updating process. Abraham and Wirahadikusumah (1999) and Wirahadikusumah *et al*. (2001) divided the life of sanitary sewers into four phases, whereby the deterioration in each phase is characterized by a homogeneous transition matrix. Jiang *et al*. (2000) proposed a “partially observable Markov process” as a basis for a decision approach for minimizing the life-cycle costs of structures. McKim *et al*. (2002) introduced risk ratios, defined as the percentage of pipes in a particular risk category (*low*, *medium* or *high*) out of all pipes in all risk categories. Mishalani and Madanat (2002) proposed a MDP, which is modified by exogenous effects such as material properties, environmental conditions, age, etc. Kleiner (2001) modelled the deterioration of buried infrastructure assets as a semi-Markov process. Probability distributions were used to derive the transition probability matrices for the asset to transit from one deteriorated state to the next.

Other types of statistical models have also been used, e.g., Lu and Madanat (1994) demonstrated the use of Bayesian updating with the aid of a two-parameter logistic function to model the deterioration of bridge decks; Cooper *et al*. (2000) introduced a GIS (geographical information system) framework to evaluate the failure risk of trunk water mains where the probability of failure is modelled as a logistic function of several covariates, etc.

In recent years, increased research effort has been dedicated to the application of *soft computing* methods to assess infrastructure deterioration. Soft computing methods include techniques such as artificial neural network (ANN), genetic algorithms (GA), probabilistic and evidential reasoning, and fuzzy techniques. Fuzzy techniques seem to be particularly suited to model the deterioration of infrastructure assets for which data are scarce and knowledge on cause(s)-effect(s) is imprecise and vague. Fuzzy techniques are a generalized form of interval analysis, which addresses uncertain and/or imprecise information (Zadeh 1965). Examples of the application of fuzzy techniques to infrastructure assets include Hajek and Hurdal (1993), who compared two artificial intelligence techniques ANNs and a rule-based system for maintenance of pavements; Chao and

Cheng (1998), who used a fuzzy-based model to diagnose cracks in reinforced concrete structures; Liang *et al*. (2001), who developed a multiple layer fuzzy evaluation method to monitor the health of concrete bridges; Sadiq *et al*. (2004), who employed a fuzzy-based method to determine the soil corrosivity as a surrogate measure for breakage/corrosion rate of cast iron pipes.

The proposed deterioration model was developed for large-diameter water mains. However, it could be trained on any system whose condition can be rated on an ordinal scale where condition states are enumerated either linguistically or numerically. Such a deterioration model would have useful application where access to assets is difficult and inspection is costly, hence the use of the term "buried infrastructure". The text refers to an "infrastructure asset" or simply an "asset" in a deliberate attempt to remain generic. The manner with which the model could be applied to a pipe segment, a pipe reach or an entire network requires further research, which was beyond the scope of this paper.

PROPOSED METHODOLOGY

Fuzzy sets

A fuzzy set describes the relationship between an uncertain quantity x and a membership function μ, which ranges between 0 and 1. A fuzzy set is an extension of the traditional set theory (in which x is either a member of set A or not) in that an x can be a member of set A with a certain degree of membership μ. Fuzzy techniques help address deficiencies inherent in binary logic and are useful in propagating uncertainties through models. A general definition of a fuzzy number is given by Dubois and Prade (1985): if x is a member of (sub)set A_i with a certain degree of membership $\mu_{A_i}(x)$, denoted as $A_i = \{(x, \mu_{A_i}(x))\}$, then A_i is a fuzzy number if x takes its value from the real numbers line and $\mu_{A_i}(x) \in [0, 1]$. The range over which x is defined is called the 'universe of discourse' of the fuzzy (sub)set A_i.

Any shape of a membership function is possible, but the selected shape should be justified by available information. Generally, triangular fuzzy numbers (TFN) or trapezoidal fuzzy numbers (ZFN) are used to represent linguistic variables (Lee 1996). A fuzzy set must be *normal*, *convex* and *bounded* (see Klir and Yuan (1995) for definitions of these terminologies) for it to qualify as a fuzzy number. Only TFNs are used in the model proposed in this paper, therefore, whenever the term 'fuzzy number' is used, it refers to TFN, unless noted otherwise. The fuzzy subsets A_i are triangular fuzzy numbers that can be defined by three points representing the three vertices of the respective triangle.

To illustrate these concepts, suppose that the age of the infrastructure asset is defined by five fuzzy subsets (or numbers), A_i, defined over universe of discourse where each subset represents an aging grade; $A_1 = new$, $A_2 = young$, $A_3 medium$, $A_4 = old$ and $A_5 = very\ old$, as illustrated in Figure 1. For example, the fuzzy subset $A_3 = medium$ has a membership function such that for age x below 20 years or above 60 years the membership to *medium* is zero, and for age between 20 and 60 years the membership follows straight lines that form a triangle.

Figure 1 shows the triangular membership values for an asset with five subsets as represented by the following set of equations,

$$\mu_{A_i}(x) = \begin{cases} 0, & 0 \le x \le q_{i,1} \\ \dfrac{x - q_{i,1}}{q_{i,2} - q_{i,1}}, & q_{i,1} \le x \le q_{i,2} \\ \dfrac{q_{i,3} - x}{q_{i,3} - q_{i,2}}, & q_{i,2} \le x \le q_{i,3} \\ 0, & x \ge q_{i,3} \end{cases} \quad (1)$$

Set	Qualitative scale	Triangular fuzzy number (TFN) representation
A_1	*new*	(*q1,1, q1,2, q1,3*) = (0, 0, 20)
A_2	*young*	(*q2,1, q2,2, q2,3*) = (0, 20, 30)
A_3	*medium*	(*q3,1, q3,2, q3,3*) = (20, 40, 60)
A_4	*old*	(*q4,1, q4,2, q4,3*) = (40, 70, 100)
A_5	*very old*	(*q5,1, q, q5,3*) = (70, 100, 100)

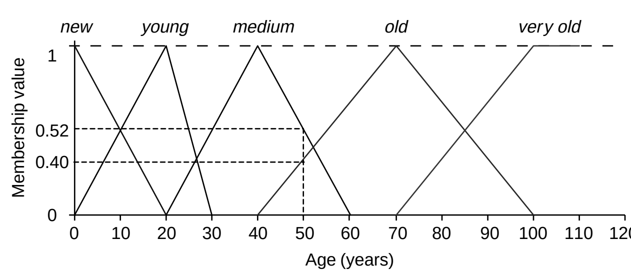


Figure 1 | Illustrative example of fuzzy sub-sets (numbers).

In this example, it can be seen (Figure 1) that for an asset of age $x = 50$ years the membership values are $\mu_{A_3}(x) = 0.52$, and $\mu_{A_4}(x) = 0.40$ and zeros for $\mu_{A_1}(x)$ $\mu_{A_2}(x)$ and $\mu_{A_5}(x)$. The fuzzy set representing the buried asset at age 50 can be written as the vector $A = (\mu_{A_1}(x), \mu_{A_2}(x), \mu_{A_3}(x), \mu_{A_4}(x), \mu_{A_5}(x)) = (0, 0, 0.52, 0.40, 0)$. Each element of the vector (tuple) depicts the asset membership value to the corresponding subset of aging grade (from *new* to *very old*). The term 'support' of a fuzzy set refers to all the tuples with non-zero memberships. For the cited example, the support of A is $S_A(X) = \{A_3, A_4\}$ or {*medium, old*}. The term 'height' of a fuzzy set refers to the highest membership value. In our example, the height of A is $h_A(X) = 0.52$. The term 'cardinality' of a fuzzy set refers to the sum of memberships of all its elements. The cardinality of A for the example is $|A| = 0.92$.

A range of definitions for arithmetic operations exists for triangular fuzzy numbers (operations with triangular shapes are simple in comparison with other shapes of fuzzy numbers). Details on these arithmetic manipulations are described by Klir and Yuan (1995). The term 'defuzzification' refers to a process to evaluate a crisp or point estimate of a fuzzy number. A defuzzified value is generally represented by centroid, often determined using the center of area method (Yager 1980).

Fuzzy rule-based algorithm

In fuzzy rule-based modelling, the relationships between variables are represented by means of fuzzy *if-then* rules of the form "*If* antecedent proposition *then* consequent proposition". The antecedent proposition is always a fuzzy proposition of the type "x is A" where x is a linguistic variable and A is a linguistic constant term. For example, a linguistic-based fuzzy rule might be: "if ambient temperature is 'quite cold' then 'people prefer to stay home'". The proposition's truth-value (a real number between zero and 1) depends on the degree of similarity between x and A. This linguistic model (Mamdani 1977) has the capacity to capture qualitative and highly uncertain knowledge in the form of *if-then* rules such as

$$\boldsymbol{R}_i : \textit{If } x \text{ is } A_j \textit{ then } y \text{ is } B_k; \quad i = 1, 2, \ldots, L; \; j = 1, 2, \ldots, M; \; k = 1, 2, \ldots, N \tag{2}$$

where x is the input (antecedent) linguistic variable and A_j is an antecedent linguistic constant. (one of M in set A) Similarly, y is the output (consequent) linguistic variable and B_k is a consequent linguistic constant (one of N in set B). The values of x and y, and A_j and B_k are fuzzy sets defined in the domains of their respective base variables. The linguistic terms A_j and B_k are selected from sets of predefined terms, such as *small*, *medium*, *large*. The rule set (comprising L rules) and the sets A and B constitute the knowledge base of the linguistic model. Each rule is regarded as a fuzzy relation: $\boldsymbol{R_i}$ $(X \times Y) \rightarrow [0, 1]$, which means that $\boldsymbol{R_i}$ is a function defined in the Cartesian space $X \times Y$, which takes values from the unit interval [0, 1]. This relation can be computed in two basic ways, either by using fuzzy implications or by using fuzzy conjunctions (Mamdani algorithm). In the proposed model, the Mamdani (1977) algorithm was used, in which fuzzy conjunction $A \wedge B$ is computed by a *minimum* operator (a *t*-norm):

$$\boldsymbol{R_i} = A_j \times B_k, \;\; i.e., \;\; \mu_{R_i}(x, y) = \mu_{A_j}(x) \wedge \mu_{B_k}(y) \tag{3}$$

where each possible pair x and y has a fuzzy relationship R_i. The strength (or the degree) of R_i is expressed as a membership function μ_{R_i} (x, y), which takes on the value of the *minimum* between μ_{A_j} (x) and μ_{B_k} (x). Note that the *minimum* operator is computed on the Cartesian product space of X and Y, i.e., for all possible pairs of x and y. The fuzzy relation $\boldsymbol{R}$ represents the entire model and is given by the *maximum* disjunction operator, i.e., *s*-norms, of L individual relations $\boldsymbol{R_i}$ $(i = 1, \ldots, L)$:

$$\boldsymbol{R} = \bigcup_{i=1}^{L} R_i, \quad i.e., \;\; \mu_R(x, y) = \max_{1 \le i \le L}\left(\mu_{A_j}(x) \wedge \mu_{B_k}(y)\right) \tag{4}$$

Now the entire knowledge base is encoded in the fuzzy relation $\boldsymbol{R}$ and the output of the linguistic model can be computed by the *max-min* composition operator, denoted in this paper by the symbol "o"

$$y = x \,\mathrm{o}\, \boldsymbol{R} \tag{5}$$

Suppose an input fuzzy value $x = A'$ has the output value B' given by the relational composition:

$$\mu_{B'}(y) = \max_X(\mu_{A'}(x) \wedge \mu_R(x, y)) \quad (6)$$

Substituting $\mu_{R_i}(x, y)$ from (3), the above expression can be rearranged as

$$\mu_{B'}(y) = \max_{1 \le i \le L}\left(\max_X\left[\mu_{A'}(x) \wedge \mu_{A_j}(x)\right] \wedge \mu_{B_k}(y)\right) \quad (7)$$

Defining $\beta_i = \max_X\left[\mu_{A'}(x) \wedge \mu_{A_j}(x)\right]$ as the degree of fulfillment of the ith rule's antecedent, the output fuzzy set of the linguistic model is:

$$\mu_{B'}(y) = \max_{1 \le i \le L}(\beta_i \wedge \mu_{B_k}(y)), \quad y \in Y \quad (8)$$

A fuzzy set can be 'defuzzified', i.e., assigned a representative crisp value. Several techniques for defuzzification have been proposed, but the one used here is the most widely accepted technique known as the centroidal method (Yager 1980).

The above algorithm is called the Mamdani inference. It is illustrated in equations (3) through (8) for single input and single output (SISO) model. It can be extended to include multiple inputs and single output (called MISO):

$$\boldsymbol{R_i}: \ \textit{If}\, x_1 \,\text{is}\, A_{1j} \,\textit{and}\, x_2 \,\text{is}\, A_{2j} \,\textit{and} \ldots \textit{and}\, x_p \,\text{is}\, A_{pj} \,\textit{then}\, y \,\text{is}\, B_k, \quad (9)$$

For example, a linguistic-based fuzzy rule might be: "if a buried pipe is 'fairly young' and its observed condition is 'quite poor' then its deterioration rate is 'very fast'". The algorithm can also be extended to multiple outputs (MIMO), which is formed by a set of MISO models. Other conjunction operators (t-norms) such as *product* can also be used instead of *minimum* operator Derivations and discussion on fuzzy rule-based modelling in this section were mainly adapted from Babuska (2003) and Yager and Filev (1994).

Fuzzy rule-based Markovian deterioration process (FR-MDP) model

Knowledge base

Figure 2 depicts the knowledge base for the proposed deterioration model. The age A of the asset is partitioned into 5 levels (from *new* to *very old*), represented by triangular fuzzy subsets $A_i\,(i = 1, 2, \ldots, 5)$, with underlying units of years. Similarly, the condition rating C of the asset is partitioned into 7 levels (condition states that range from *excellent* to *failed*) represented by triangular fuzzy subsets $C_i\,(i = 1, 2, \ldots, 7)$. C is mapped onto arbitrary non-dimensional relative scale in the interval [0, 1]. It should be noted that the *failed* state does not mean that collapse has already occurred (in which case the membership would be a clear unity), rather that it is imminent. The deterioration rate D', which is a measure of change in condition ratings in a unit time step, is partitioned into 5 levels (from *very slow* to *very fast*) represented by triangular fuzzy subsets $D'_i\,(i = 1, 2, \ldots, 5)$. D' is mapped onto a dynamic relative scale (Figure 2) that ranges from 0 to $4d_o$, where d_o is the base deterioration rate and has the underlying units of fractions of membership per year. The base deterioration rate d_o is a scaling parameter. It is required to account for possible large variations in deterioration rates of assets with different characteristics or subjected to different internal and external environments (e.g., a relatively slow deterioration rate of a cast iron pipe in corrosive environment may be considered relatively fast for a prestressed concrete cylinder pipe, PCCP, pipe in a benign environment). Typical range of the scale (from 0 to $4d_o$) for deterioration rates will usually be between zero and 0.2 membership per year (Figure 2).

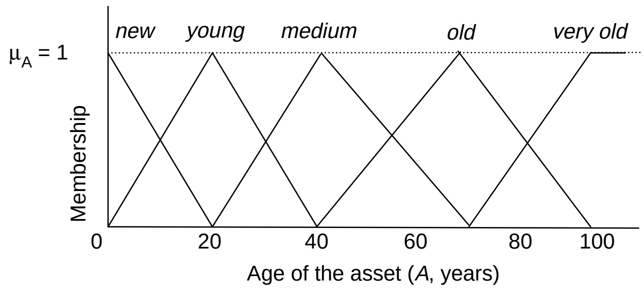
Numbers of partition levels (granularity) for age (A), condition rating (C) and deterioration rate (D') were established after considering the compromise between too many levels and the inability to practically judge the difference between any two contiguous levels. However, the model is general enough to accommodate other granularities. The table at the bottom of Figure 2 depicts the set of fuzzy rules R_D that governs this fuzzy rule-based Markovian deterioration process (FR-MDP) model. For example, if the asset age is $A =$ *young* and its condition is $C =$ *fair* then its deterioration rate is $D' =$ *fast*. The rule set R_D thus contains 35 fuzzy rules.

Deterioration process

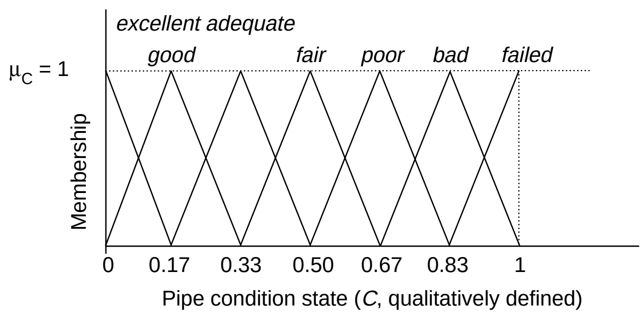
The proposed deterioration process is modelled essentially as a 'flow' of membership from one condition state to the

Knowledge-base

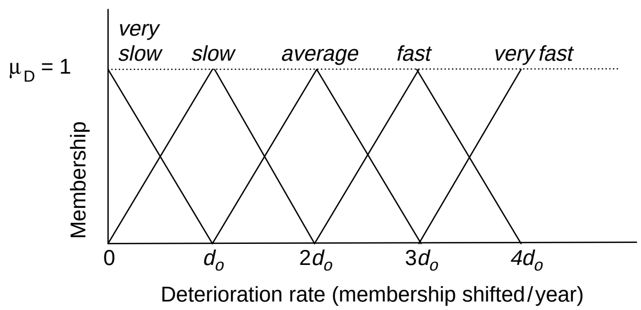
Age	*qi,1*	*qi,2*	*qi,3*
new	0	0	20
young	0	20	40
medium	20	40	70
old	40	70	100
very old	70	100	100



Condition state	*qi,1*	*qi,2*	*qi,3*
excellent	0	0	0.17
good	0	0.17	0.33
adequate	0.17	0.33	0.50
fair	0.33	0.50	0.67
poor	0.50	0.67	0.83
bad	0.67	0.83	1
failed	0.83	1	1



Deterioration rate	*qi,1*	*qi,2*	*qi,3*
very slow	0	0	d_O
slow	0	d_O	$2d_O$
average	d_O	$2d_O$	$3d_O$
fast	$2d_O$	$3d_O$	$4d_O$
very fast	$3d_O$	$4d_O$	$4d_O$



Fuzzy rule-set R_D

R_i = *If* asset age (A) is ***"A" and*** pipe condition state (C) is ***"C" then*** deterioration rate (D) is ***"D"*** (at time = *t*)

Pipe condition (C):	*excellent*	*good*	*adequate*	*fair*	*poor*	*bad*	*failed*
Age (A): ***new***	*slow*	*average*	*fast*	*very fast*	*very fast*	*very fast*	*very fast*
young	*slow*	*average*	*fast*	*fast*	*fast*	*very fast*	*very fast*
medium	*very slow*	*slow*	*average*	*average*	*fast*	*fast*	*very fast*
old	*very slow*	*very slow*	*slow*	*slow*	*average*	*average*	*fast*
very old	*very slow*	*very slow*	*very slow*	*slow*	*slow*	*average*	*average*

Figure 2 | Fuzzy sets for age, condition and deterioration rate and fuzzy rule-base for the Markovian deterioration model.

next lower condition state. The deterioration process at each time step comprises two stages. In the first stage, the asset age is fuzzified (mapped on A) to obtain the asset fuzzy age $A_t = \left(\mu_t^{A_1}, \mu_t^{A_2}, \ldots, \mu_t^{A_5}\right)$. The asset fuzzy condition rating at time step t (taken for convenience as a single year) is $C_t = \left(\mu_t^{C_1}, \mu_t^{C_2}, \ldots, \mu_t^{C_7}\right)$. The fuzzy deterioration rate at t, $D'_t = \left(\mu_t^{D'_1}, \mu_t^{D'_2}, \ldots, \mu_t^{D'_5}\right)$ is computed using the Mamdani (1977) algorithm detailed earlier for the MISO model – equation (9) where A_t and C_t are the inputs, D'_t is the output and R_D is the fuzzy rule-set from which the fuzzy inferences are established, i.e.,

$$D'_t = (A_t \text{ and } C_t) \circ R_D \qquad (10)$$

where "o" is a max-min composition operator explained earlier. D'_t, is a 5-tuple fuzzy set which is then defuzzified using the centroidal method. The defuzzified (crisp) value of the fuzzy deterioration D'_t is denoted by D_t. In the second stage, the condition rating of the asset in the next time step C_{t+1}, is calculated from its condition rating in the current time step, C_t, and the (defuzzified) deterioration rate D_t obtained by rule-base algorithm in the current time step as follows

$$C_{t+1} = C_t \otimes D_t \tag{11}$$

where $\otimes$ is an operator as explained below. Two assumptions are made with regards to the deterioration process. First, it is reasonably assumed that an asset cannot improve itself without any intervention. Thus if $D_t^{i,j}$ is the deterioration rate at time t from condition state i to condition state j, then $D_t^{i,j} = 0$ for all $i > j$. Additionally, it is assumed that the deterioration process is continuous and slow relative to the selected time step, therefore, an asset in state i can at the most deteriorate to state $i + 1$ within a single time step. This assumption has been made by others (e.g., Madanat *et al*. 1995; Kleiner 2001). Equation (11) can now be written in a matrix form

$$\left(\mu_{t+1}^{C_1}, \mu_{t+1}^{C_2}, \ldots, \mu_{t+1}^{C_7}\right) = \left(\mu_t^{C_1}, \mu_t^{C_2}, \ldots, \mu_t^{C_7}\right) \otimes \begin{bmatrix} D_t^{1,1} & D_t^{1,2} & 0 & 0 & \cdots & 0 \\ 0 & D_t^{2,2} & D_t^{2,3} & 0 & \cdots & 0 \\ \vdots & & & & & \\ 0 & 0 & \cdots & & D_t^{6,6} & D_t^{6,7} \\ 0 & 0 & \cdots & & & D_t^{7,7} \end{bmatrix} \tag{12}$$

where $\mu_{t+1}^{C_i}$ is the membership value to condition state i. The deterioration rate matrix in the RHS of (12) is analogous to a transition probability matrix in the traditional Markovian deterioration process. If, e.g., $D_t^{i,j} = 0.05$, it means that the asset loses 5% membership to state i in favour of state j $(j = i + 1)$ as it transits from time step t to $t + 1$. Consequently, (12) can be written

$$\left(\mu_{t+1}^{C_1}, \mu_{t+1}^{C_2}, \ldots, \mu_{t+1}^{C_7}\right) = \left(\mu_t^{C_1}, \mu_t^{C_2}, \ldots, \mu_t^{C_7}\right) \otimes \begin{bmatrix} 1 - D_t^{1,2} & D_t^{1,2} & 0 & 0 & \cdots & 0 \\ 0 & 1 - D_t^{2,3} & D_t^{2,3} & 0 & \cdots & 0 \\ \vdots & & & & & \\ 0 & 0 & \cdots & & 1 - D_t^{6,7} & D_t^{6,7} \\ 0 & 0 & \cdots & & 0 & 1 \end{bmatrix} \tag{13}$$

The operator $\otimes$ can now be viewed as a simple matrix multiplication, in which

$$\mu_{t+1}^{C_i} = \begin{cases} \mu_t^{C_1}(1 - D_t^{1,2}), & i = 1 \\ \mu_t^{C_{i-1}} D_t^{i-1,i} + \mu_t^{C_i}(1 - D_t^{i,i+1}), & i = \{2, 3, \ldots, 6\} \\ \mu_t^{C_7} + \mu_t^{C_6} D_t^{6,7}, & i = 7 \end{cases} \tag{14}$$

In practice, a fuzzy asset condition rating with substantial support (non-zero membership) for more than three contiguous tuples would be counterintuitive to expert opinion. One approach to incorporate this aspect of expert opinion is through the introduction of threshold parameters G_i, as follows.

$$\mu_{t+1}^{C_i} = \begin{cases} \mu_t^{C_1}(1 - D_t^{1,2}), & i = 1 \\ g_{i-1}\mu_t^{C_{i-1}} D_t^{i-1,i} + g_i \mu_t^{C_i}(1 - D_t^{i,i+1}), & i = \{2, 3, \ldots, 6\} \\ \mu_t^{C_7} + g_6 \mu_t^{C_6} D_t^{6,7}, & i = 7 \end{cases} \tag{15}$$

where

$$g_i = \begin{cases} 1 & if\ \max(\mu_\tau^{C_i}) \geq G_i\ for\ any\ \tau \leq t \\ 0 & otherwise \end{cases}$$

Equation 15 indicates that membership cannot start to 'flow' from state i to state j $(j = i + 1)$ at time t, unless the membership in state i exceeds a threshold value G_i at any time prior to t. These constraints ensure that the deterioration model avoids these counterintuitive situations in which an asset has significant membership values to more than two or three contiguous condition states (Figure 3). Figure 3 illustrates an example on the application of deterioration models with and without thresholds. At $t = 40$ years, e.g., the condition rating of the asset without threshold is

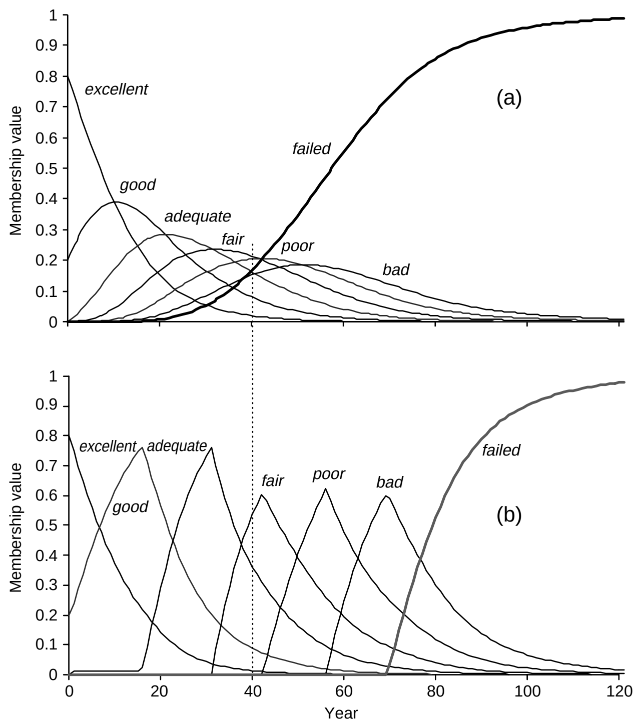


Figure 3 | Deterioration curves without (a) and with (b) membership threshold parameters.

approximately $C_{40} = \{0, 0.09, 0.16, 0.17, 0.20, 0.21, 0.17\}$, which means that the asset has relatively significant simultaneous membership values of $\mu^{C_2} = 0.09$ and $\mu^{C_7} = 0.17$ to state 2 (*good*) and to state 7 (*failed*), respectively. This result is of course unrealistic. In contrast, the model with thresholds yields a realistic condition rating of $C_{40} = \{0, 0.09, 0.36, 0.54, 0, 0, 0\}$. This, by the way, is one of the caveats of using the traditional Markov process to model the deterioration rate of a single asset (as oppose to an inventory of assets). Though the introduction of the artifice of threshold parameters in Equation 15 has no physical interpretation, it ensures a more realistic deterioration model and may better replicate realistic condition ratings of assets.

An instructive way to visualize the deterioration process is to consider a series of seven containers positioned one below the other, where water flows from a higher container to a lower one as illustrated in Figure 4. Each container represents a condition state (seven containers for seven condition states, only five are shown) and the water levels represent membership values. At the start of the process, all the water (membership) is in the top container (*excellent* condition state). As time goes by, water flows from the higher containers to the lower (more deteriorated states) ones, until gradually the bottom container (*failed* condition state) fills up. The effect of the threshold parameters can then be visualized as follows. In the deterioration process without the threshold parameters, the containers act as if they have a spout at the bottom, out of which water flows as soon as there is water in that container. As each container fills up, the flow into the next container increases due to increased head. Once the outflow becomes greater than the inflow the container begins to empty. In the deterioration process with the threshold parameters, the containers act as if they have an inverted siphon mechanism, instead of a spout at the bottom, where outflow does not start until the water level in the container reaches a specific threshold level.

It is worth noting that state 7, which is the *failed* state, is an 'absorbing' state. Furthermore, since the deterioration rate is re-evaluated according to the rule-set at each time-step, it follows that this deterioration process is analogous to a non-homogeneous Markov process. Membership to the *failed* state (state 7) at any given time t can be viewed as the possibility (not probability) of failure at that time. A notional difference between 'probability' and 'possibility' can be demonstrated by referring to "what will happen" (probability) versus "what can happen" (possibility). Possibility measure is therefore always more than the probability measure. While the area under the probability distribution function (pdf) is equal to 1, the area under the possibility distribution curve has no specific meaning, rather the 'height' of the possibility distribution is 1. Further discussion on this subject is given under the Discussion section.

Assumed or known condition rating of the asset immediately after installation, and at least one other at a later age t are required in order to train the deterioration model. The model is trained by using a non-linear regression procedure, where the sum of squared deviations between the observed and predicted membership values to condition states is minimized. The parameters that vary in the training process are those that control the scale and shape of the deterioration fuzzy set, namely, d_o (base deterioration rate) and G_i (threshold parameters). Note that not all G_i parameters need to be considered in the training process, rather only those that are significantly supported in the

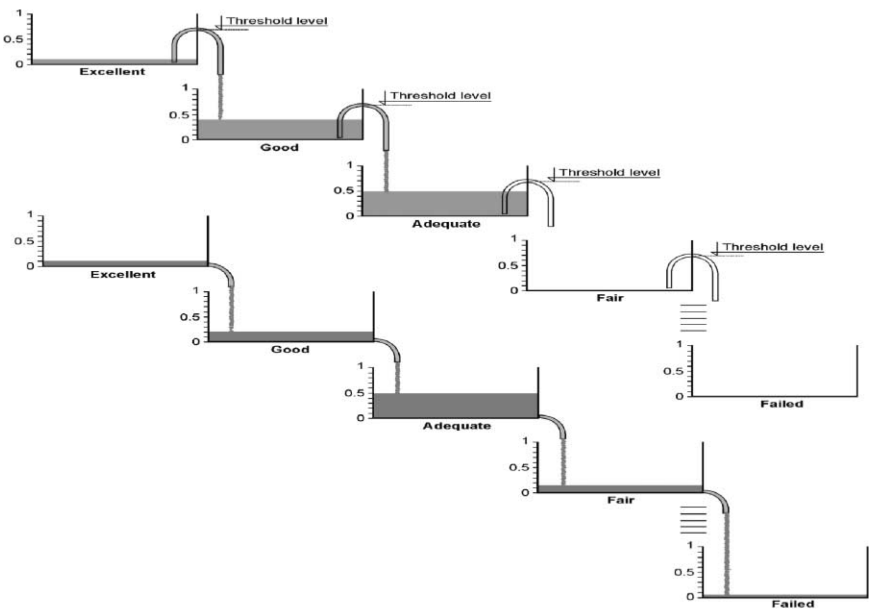


Figure 4 | Visualization of threshold effects: bottom reservoirs – analogous to traditional Markov process (no thresholds); top reservoirs – analogous to deterioration curves with thresholds.

observed condition rating. For example, if the observed condition rating at year t is $C_t = (0, 0.2, 0.7, 0.1, 0, 0, 0)$ only G_2, and G_3 need to be considered (although condition state 4 is supported, at a membership value of 0.1 the threshold parameter G_4 will not be triggered in the regression because typical G_i values are expected to be 0.5 to 0.9).

Figure 5 illustrates a numerical example in which the asset condition rating immediately after deterioration was assumed to be $C_{t=0} = (0.8, 0.2, 0, 0, 0, 0, 0)$, and at age 50 the observed condition rating was evaluated at $C_{50} = (0, 0.2, 0.7, 0.1, 0, 0, 0)$. The non-linear regression yielded a base deterioration rate parameter $d_o = 0.036$, and threshold values $G_2 = 0.754$ and $G_3 = 0.752$. Other threshold values understandably did not affect the results of the regression because the observed condition rating did not cause them to 'fire' (in Figure 5 they are arbitrarily depicted as $G_i = 0.6$ $(i = 4, 5, 6)$). The modelled condition rating was found to be $C'_{50} = (0.03, 0.18, 0.7, 0.09, 0, 0, 0)$ with a sum of squared deviations 0.00156.

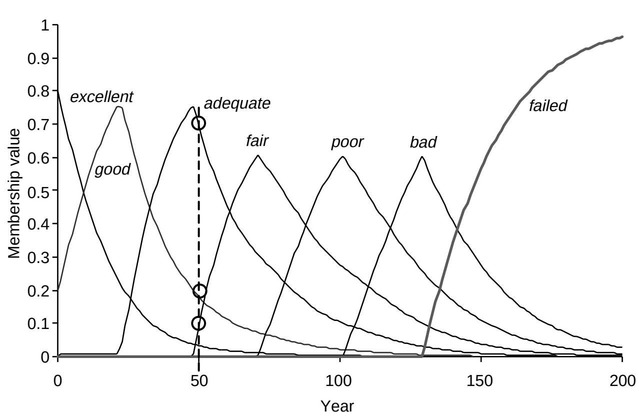


Figure 5 | Example model training on a single observation.

MODEL VALIDATION

The proposed fuzzy non-homogeneous Markov deterioration model is especially suited for assets for which data are scarce, and if available, are often imprecise and vague. Data scarcity results from the fact that most often these assets are difficult to access and expensive to inspect. In this paper, data on buried, large-diameter water transmission mains are used to conduct the validation exercise.

At least two consecutive observations of the asset condition are needed in order to validate the proposed model. Further, these observations need to be reasonably distant (in time) from each other, to avoid errors due to small inconsistencies due to the subjective nature of the condition assessment of an asset. The first observation is needed to train the model and predict future deterioration, whereas the second observation is required to validate the prediction. Unfortunately, available data are scarce and the data obtained for this research project were less than ideally suited for validation. Nonetheless, one case study is presented here to demonstrate the validation procedure.

Arizona Public Service Company (APS) submitted selected information on 69 prestressed concrete cylinder pipe (PCCP) segments of 96″ (2400 mm) in diameter and 98 PCCP segments of 114″ (2900 mm) in diameter, all installed in 1978. Following installation, some of the pipe segments have been inspected twice and others thrice, using visual inspection, hammer tapping (pulse echo) to detect concrete core condition, and remote field eddy current/transformer coupling (RFEC/TC) to detect the number of wire breaks (the most common failure mode in PCCP is one in which contiguous prestressing wires break due to corrosion). Observed distress indicators were translated into condition ratings as described by Rajani *et al*. (2006). Table 1 provides details for the specific pipe segment discussed here. The pipe condition rating upon installation was unknown and therefore assumed to be (0.9, 0.1, 0, 0, 0, 0, 0).

The pipe appeared to have deteriorated relatively slowly during the first 19 years, as well as during the subsequent five years. No detectable deterioration was observed between the second and third inspections, which is expected given the slow overall deterioration and the short period elapsed between inspections Figure 6a. Figure 6b illustrates the results obtained by training the deterioration model on the 1997 condition rating that was observed to be (0.09, 0.85, 0.06, 0, 0, 0, 0). The model yielded a condition rating (0.12, 0.87, 0.01, 0, 0, 0, 0) for 1997 (at age 19), with sum of squared deviations (SSD) of 0.004.

The model forecasted the condition rating for 1999 (age 21) to be (0.09, 0.9, 0.01, 0, 0, 0, 0), which is quite close to the observed (Table 1) condition rating (SSD = 0.01). For 2002 (age 24) the forecast was (0.06, 0.66, 0.28, 0, 0, 0, 0), which is still quite close to the observed condition rating (SSD = 0.07). It is noted that the threshold values for the future condition states (those that are not supported by the observed data) were arbitrarily set to 0.7, which is the mid-range of the expected values of all thresholds.

Figure 6b illustrates the results obtained by training the deterioration model on both 1997 and 1999 condition ratings. The fit was quite good (SSD = 0.023). With this calibration, the model forecasted the condition rating for 2002 (age 24) to be (0.07, 0.58, 0.35, 0, 0, 0, 0), which is moderately close, with SSD = 0.14.

It appears that the extra inspection data actually degrades the accuracy of the forecast rather than improve it. This degradation could possibly be attributed to the closeness of the inspections to each other in time as well as in observed condition ratings considering the inherent uncertainty in the inspection results. A more spread-out set of observations might have increased the robustness of the results but unfortunately such data were not available. Another source of poor model performance can be because of the lack of equivalency between any two non-commensurate measures of distress, such as visual inspection (1997 inspection) and RFEC (1999 and 2002 inspections). In the proposed framework, relative confidence in the inspection methods is expressed through the introduction of weights (credibility

Table 1 | Distress indicators and condition ratings for 2400 mm (96″) PCCP installed in 1978

Year	Inspection method	# wire breaks	Spalling	Cracks	Colouration	Pulse-echo sound	Observed condition rating
1978							(0.9, 0.1, 0, 0, 0, 0, 0)*
1997	Visual	–	None	No cracks	No stains	Very firm	(0.09, 0.85, 0.06, 0, 0, 0, 0)
1999	RFEC/TC	5	–	–	–	–	(0.06, 0.85, 0.09, 0, 0, 0, 0)
2002	RFEC/TC	5	–	–	–	–	(0.06, 0.85, 0.09, 0, 0, 0, 0)

*Assumed condition rating.

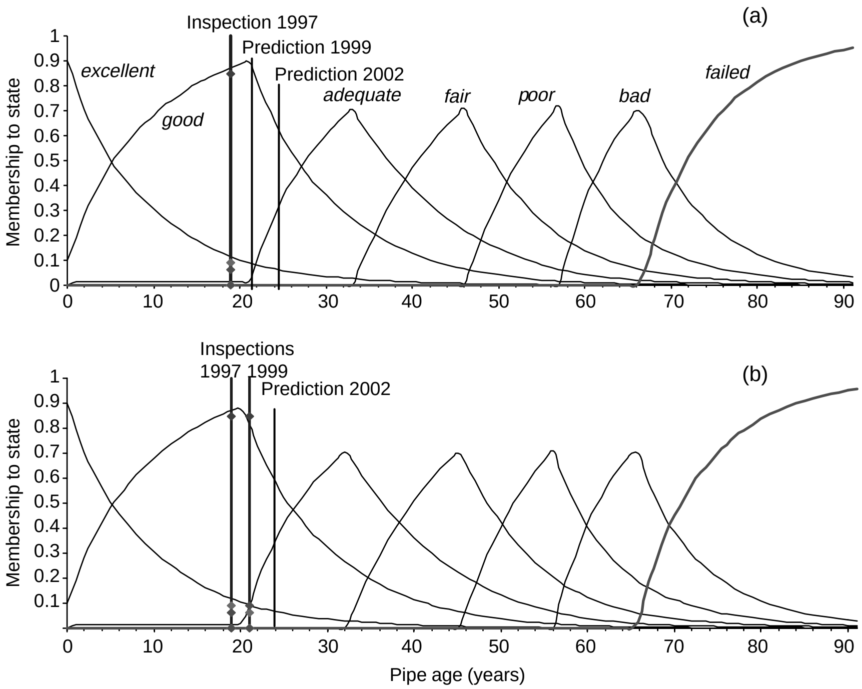


Figure 6 | Model validation with (a) one inspection in 1997 and (b) two inspections in 1997 and 1999.

factors) to the condition ratings (these weights are to be distinguished from weights introduced by Rajani *et al.* (2006), to assign relative 'importance' of distress indicators to determine condition rating). The issue of weights is discussed further in the next section.

α-CUTS AND FUZZY CONFIDENCE BAND

The concept of α-cuts as applied to fuzzy numbers (sets) is defined as follows. Assume A_i is a fuzzy number that corresponds to the definition described earlier. The α-cut of set A_i, denoted by $A_{i\alpha}$ is the subset of A_i that consists of all the elements in x for which $\mu_{A_i}(x) \geq \alpha$. Figure 7 illustrates the α-cut concept ($\alpha = 0.5$), using the fuzzy number $D'_{30} = (0.18,\ 0.25,\ 0.68,\ 0.14,\ 0)$, which represents the fuzzy deterioration rate at year 30. The dark solid outline in Figure 7 is a graphical representation of the 'mass' of the fuzzy number D'_{30}. The α-cut concept can be used to form a fuzzy confidence band (also could be referred to as a possibilistic confidence band), which is akin to (though not the same as) the probabilistic confidence interval used in traditional statistical methods. This possibilistic confidence band provides a plausible range (or an interval) for the best estimate. For the earlier example, a fuzzy confidence band $A_{i\alpha}$ for the best estimate of *failed* condition state is obtained by an α-cut at $[0.9max(\mu_t^{D'_1}, \mu_t^{D'_2}, \ldots, \mu_t^{D'_5})]$ which is illustrated in Figure 8. It is important to note that this confidence band should not be interpreted in the same manner as a probabilistic band.

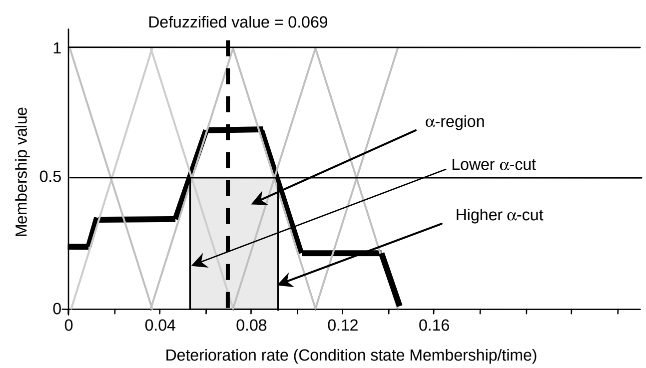


Figure 7 | Illustration of concept of an α-cut.

DISCUSSION

The example in the previous section illustrated how the deterioration model can be trained using one or two

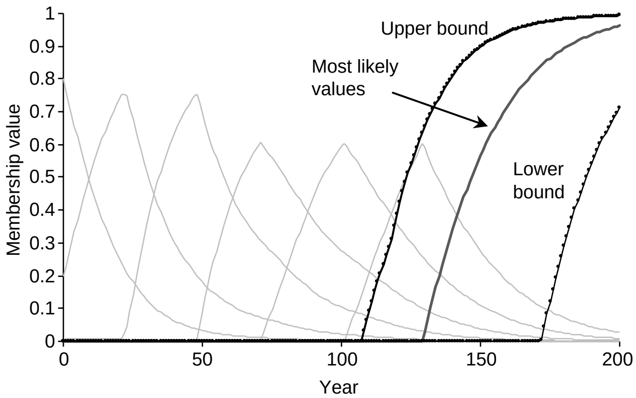


Figure 8 | Deterioration curves with fuzzy confidence band ($\alpha = [0.9\max(\mu_t^{D'_1}, \mu_t^{D'_2}, \ldots, \mu_t^{D'_5})]$) for the 'failed' state.

observations of the condition ratings of the asset. The training process could be extended to multiple observations throughout the life of the asset. Deterioration-causing conditions, such as climate, overburden, pressure regimes, can change because buried infrastructure assets have typically long lives, often measured in many decades. Consequently, when the model is trained on multiple observations that span a long period of time, with the purpose of forecasting the future condition ratings of the asset, the later observations could be perceived as having higher reliabilities with respect to their ability to predict future conditions. As a result, higher weights could be assigned to condition ratings that are closer (in time) to the present (e.g., see Yager (2004) for fusing evidences of different credibility using Dempster-Shafer rule of combination). As previously discussed, weights could also be assigned on the basis of relative credibility of the inspection techniques used.

In the proposed model, each condition rating adds seven degrees of freedom (corresponding to seven condition states) to the regression process. The deterioration base rate parameter (d_o) always reduces the degrees of freedom by one. Assuming that a typical observed condition rating yields a fuzzy set with support for two or three condition states (meaning that there are two or three membership values that are different than zero), the threshold parameters (G_i) will typically reduce the number of degrees of freedom by at least two. It follows that the maximum degrees of freedom for regression based on a single observed condition rating is four for a moderately deteriorated asset, and less for a more deteriorated one. Clearly, condition ratings based on multiple inspections will greatly improve the reliability of the regression, especially in the more deteriorated assets, provided that the inspection techniques are consistent.

Although in this paper a seven-state condition rating was used, the model is flexible enough to accommodate different numbers of states. The number of states is not anticipated to introduce a material change in the performance of the deterioration model. Further, it is possible that the fuzzy rules could be tweaked to improve the performance of the model. This has not been attempted in the current research.

It was assumed that the base deterioration rate d_o will persist in the process of predicting the future deterioration of the asset. However, the threshold parameters G_i cannot be ascertained for the condition states that are higher (more deteriorated) than the one observed in the latest condition assessment. Consequently, a need arises to assume reasonable values for the prediction of future deterioration. Reasonable values for the threshold parameters could be in the range of 0.6–0.8, and one could possibly consult threshold values obtained from lower states. Although threshold values of 0.7 were taken in the example described earlier, it is possible that they actually diminish as the pipe ages, causing an effective acceleration in pipe deterioration (faster 'flow' of membership from one condition state to the next). More research and more data would be required to investigate these premises.

There are two principal interpretations of probability, objective and subjective. According to the objective interpretation, probability is *the relative frequency of occurrence of an event*. The subjective (Bayesian) interpretation is that probability is *the degree of belief that an event will occur*. Subjective probabilists (also called *Bayesians*) maintain that probabilistic methods are also useful for problems where there are only a few observations and probability is based on one's experience or intuition. With these two definitions, probabilistic methods can model uncertainties of a stochastic nature as well as of a subjective nature, either when the available information is based on judgment or when it is based on measurements. Bayesian interpretation of probability is criticized for the additive axiom it has to abide and for the requirement that probabilities of all elementary events should be precisely expressed (Chen 2000).

Shackle (1961) first proposed the theory in which a decision process was modeled in term of "possibility". He stated that possibility of an event is equal to one minus a person's degree of surprise if the event occurs. He stated that possibility should be used instead of probability when the conditions under which we have to make a decision under uncertainty cannot be reproduced. Zadeh (1978) later formulated the contemporary possibility theory as an extension of fuzzy set theory, in which a possibility distribution is numerically equal to the membership function. Klir and Yuan (1995) compared the mathematical properties of possibility and probability theories. They found that possibility theory is well suited to represent imprecision or vagueness in the data, whereas probability is best suited to represent stochasticity (randomness).

SUMMARY

The scarcity of data about the deterioration rates of buried infrastructure assets, coupled with the imprecise and often subjective nature of assessment of pipe condition merits the usage of fuzzy techniques to model the deterioration of these assets. The deterioration process is modelled as a fuzzy rule-based non-homogeneous Markov process applied at each time step in two stages. In the first stage, the deterioration rate at the specific time step is inferred from the asset age and condition state using fuzzy rule-base algorithm. In the second stage, the condition rating of the asset in the next time step is calculated from the present condition rating and the deterioration rate. Essentially the deterioration process progresses as the asset gradually transits from higher memberships in good condition states to higher membership in worse states. Memberships thus 'flow' from higher to lower condition states. The process is formulated to imitate the reality in which a given asset at a given time cannot have significant membership values to more than two or three contiguous condition states.

The model is trained by using non-linear regression procedure where the sum of squared deviations between observed and predicted membership values of condition ratings is minimized. The model can be used to predict future deterioration rate of the asset, subject to some judgment-based assumptions. The model was partially validated using available data, however, a more rigorous validation would be beneficial, with data that have more historical depth as well as consistent inspection techniques. More research is required to investigate the assignment of weights to historical condition ratings. More research is also required to investigate the behaviour of threshold parameters as the asset ages and deteriorates.

ACKNOWLEDGEMENTS

This paper is based on a research project, which was co-sponsored by the American Water Works Association Research Foundation (AwwaRF), the National Research Council of Canada (NRC) and water utilities from the United States, Canada and Australia. The research report, accompanied by a prototype computer application is available from AwwaRF. The project report number is 91087 and its title is "Risk management of large-diameter water transmission mains".

REFERENCES

Abraham, D. M. & Wirahadikusumah, R. 1999 Development of prediction model for sewer deterioration. In: Lacasse, M. A. & Vanier, D. J. (eds) *Proceedings of the 8th Conference Durability of Building Materials and Components*. NRC, Vancouver, Canada, pp. 1257–1267.

Babuska, R. 2003 *Fuzzy systems, modelling and identification*, http://lcewww.et.tudelft.nl/ ~ babuska.

Chao, C.-J. & Cheng, F. P. 1998 Fuzzy pattern recognition model for diagnosing cracks in RC structures. *Journal of Computing in Civil Engineering, ASCE* **12**(2), 111–119.

Chen, Q. S. 2000 Comparing Probabilistic and Fuzzy Set Approaches for Design in the Presence of Uncertainty. Ph.D. Dissertation, Virginia Polytechnic Institute, Virginia, USA.

Cooper, N. R., Blakey, G., Sherwin, C., Ta, T., Whiter, J. T. & Woodward, C. A. 2000 The use of GIS to develop a probability-based trunk mains burst risk model. *Urban Water* **2**, 97–103.

Dubois, D. & Prade, H. 1985 Evidence measures based on fuzzy information. *Automatica* **21**(5), 547–562.

Hajek, J. J. & Hurdal, B. 1993 Comparison of rule-based and neural network solutions for a structured selection problem. *Transportation Research Record* **1399**, 1–7.

Jiang, M., Corotis, R. B. & Ellis, J. H. 2000 Optimal life-cycle costing with partial observability. *Journal of Infrastructure Systems, ASCE* **6**(2), 56–66.

Kleiner, Y. 2001 Scheduling inspection and renewal of large infrastructure assets. *Journal of Infrastructure Systems, ASCE* **7**(4), 136–143.

Klir, G. J. & Yuan, B. 1995 *Fuzzy sets and fuzzy logic- theory and applications*. Prentice- Hall, Inc., Englewood Cliffs, NJ, USA.

Lee, H.-M. 1996 Applying fuzzy set theory to evaluate the rate of aggregative risk in software development. *Fuzzy Sets and Systems* **79**, 323–336.

Li, N., Haas, L. R. & Xie, W.-C. 1997 Development of a new asphalt pavement performance prediction model. *Canadian Journal of Civil Engineering* **24**, 547–559.

Liang, M. T., Wu, J. H. & Liang, C. H. 2001 Multiple layer fuzzy evaluation for existing reinforced concrete bridges. *Journal of Infrastructure Systems, ASCE* **7**(4), 144–159.

Lu, Y. & Madanat, S. M. 1994 Bayesian updating of infrastructure deterioration models. *Transportation Research Record* **1442**, 110–114.

Madanat, S. M., Mishalani, R. & Wan Ibrahim, W. H. 1995 Estimation of infrastructure transition probabilities from condition rating data. *Journal of Infrastructure Systems, ASCE* **1**(2), 120–125.

Madanat, S. M., Karlaftis, M. G. & McCarthy, P. S. 1997 Probabilistic infrastructure deterioration models with panel data. *Journal of Infrastructure Systems, ASCE* **3**(1), 4–9.

Mamdani, E. H. 1977 Application of fuzzy logic to approximate reasoning using linguistic systems. *Fuzzy Sets and Systems* **26**, 1182–1191.

Mishalani, R. G. & Madanat, S. M. 2002 Computation of infrastructure transition probabilities using stochastic duration models. *Journal of Infrastructure Systems, ASCE* **8**(4), 139–148.

McKim, R., Kathula, V. S. & Nassar, R. 2002 The development of risk ratios for sewer prediction modeling. In: *Proceedings [CD-ROM] No-Dig Conference*, Montreal, Quebéc, April–May.

Rajani, B., Kleiner, Y. & Sadiq, R. 2006 Translation of pipe inspection results into condition rating using fuzzy techniques. *Journal of Water Supply: Research. & Technology-AQUA* **55**(1), 11–24.

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

Shackle, G. L. S. 1961 *Decision, Order and Time in Human Affairs*. Cambridge University Press, New York, USA.

Wirahadikusumah, R., Abraham, D. & Isely, T. 2001 Challenging issues in modeling deterioration of combined sewers. *Journal of Infrastructure Systems, ASCE* **7**(2), 77–84.

Yager, R. R. 1980 A general class of fuzzy connectives. *Fuzzy Sets and Systems* **4**, 235–242.

Yager, R. R. & Filev, D. P. 1994 *Essentials of fuzzy modeling and control*. John Wiley & Sons, Inc., NY, USA.

Yager, R. R. 2004 On the determination of strength of belief for decision support under uncertainty - Part II: fusing strengths of belief. *Fuzzy Sets and Systems* **142**, 129–142.

Zadeh, L. A. 1965 Fuzzy sets. *Information and Control* **8**(3), 338–353.

Zadeh, L. A. 1978 Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* **1**(1), 3–28.

First received 23 August 2005; accepted in revised form 8 December 2005