Water Resour Manage (2010) 24:139–156
DOI 10.1007/s11269-009-9441-2

A Neurofuzzy Decision Framework for the Management of Water Distribution Networks

Symeon Christodoulou · Alexandra Deligianni

Received: 23 June 2008 / Accepted: 13 April 2009 /
Published online: 12 May 2009

Abstract Among the most important components of sustainable management strategies for water distribution networks is the ability to integrate risk analysis and asset management decision-support systems (DSS), as well as the ability to incorporate in the analysis financial and socio-political parameters that are associated with the networks in study. Presented herein is a neurofuzzy decision-support system for the performance of multi-factored risk-of-failure analysis and pipe asset management, as applied to urban water distribution networks. The study is based on two datasets (one from New York City and the other from the city of Limassol, Cyprus), analytical and numerical methods, and artificial intelligence techniques (artificial neural networks and fuzzy logic) that capture the underlying knowledge and transform the patterns of the network's behaviour into a knowledge-repository and a DSS.

Keywords Water distribution networks · Decision support systems · Risk analysis

1 Introduction

Each year, hundreds of kilometres of pipes in urban areas are upgraded, or replaced, in an attempt to mitigate the effects of pipe bursts and water loss, and to maintain the uninterrupted transport of water. Given UN predictions that by 2025 the majority of earth's population (around 6 billion people) will live in urban areas then the continued, safe, and uninterrupted transport of clean water in urban centres becomes of paramount importance. Furthermore, the task of maintaining the underlying piping networks in good condition becomes an increasing challenge, both because of the predicted heightened demands and because of the additional risks incurred. In fact, the problem of aging infrastructure and the associated water losses in urban water distribution networks has been one of the biggest infrastructure problems

S. Christodoulou (✉) · A. Deligianni
University of Cyprus, Nicosia, Cyprus
e-mail: schristo@ucy.ac.cy

facing city and municipal authorities and a major task in their efforts to achieve efficient and sustainable management of water resources. Interestingly enough, even in developed countries "unaccounted-for" water is in the range of 20% to 30% (as reported by studies of the International Water Association), whereas in developing countries this percentage is even higher.

The driving forces behind pipe-replacement capital improvement projects have primarily been (1) the mandate to safeguard the health of urban populations, (2) the need to increase the reliability of the pipe networks and the service provided to people, and (3) the socioeconomic factors related to the cost of operations and maintenance of piping networks. Existing water distribution systems are increasingly at risk due to numerous factors (both internal and external to the distribution networks) and the accidental or deterioration-based breakage of water distribution systems represents a range of problems. Yet, advances in the development of technologies and tools for sustainable management of piping networks have been slow in pace (especially in terms of remote monitoring systems) and have yet to be fully embraced by the water utility industry.

2 Sustainable Management of Water Distribution Networks

Sustainable management of urban water distribution networks should include not only new methodologies for monitoring, repairing or replacing aging infrastructure, but more importantly it should facilitate the modelling of the state of each piping network and the predictive analysis of its behaviour over time. The goal should be the condition-assessment of deteriorating piping infrastructure and the utilization of historical incident data and risk-of-failure metrics for devising intelligent "replace or repair" strategies. The term "replace or repair dilemma" refers to the decision made by water-distribution network administrators at various points in time to replace (or not) a failing pipe based on their evaluation of the situation at hand. Exercising the replacement option would result in incurring replacement costs at the time of action, but it would lower the risk of future failures. Exercising the repair option would mean the retainage of the pipe in consideration and thus saving any replacement costs, but it would result in higher risk of future failures and thus increasing costs of repair, disruption of service and damages. The decision, thus, to repair or replace failing pipe segments boils down to the water agency's tolerance to risk, its appraisal of the situation and the associated risk level, and the socioeconomic implications from a possible pipe failure. A successful resolution of this strategic dilemma would help reduce water losses and life-cycle costs as well as increase the reliability of the network and the quality of service to citizens.

With that in mind, water distribution agencies are nowadays faced with the increasingly more complex task to intelligently and efficiently assess (or model) the condition of a pipe network and manage the network in ways that maximize its reliability and minimize its operational and management costs. Research to-date has helped identify a number of potential risk factors that contribute to pipe breaks, such as: the pipe diameter, material and length, the operating pressure in the network, a pipe's age, the number of previously observed breaks for each pipe segment, the soil conditions, and the external loads to the underground piping network. Some of the factors are time-invariant and some are time-dependent, but all have shown to be contributing factors to the overall risk-of-failure level.

2.1 Scope of Work

The scope of work for the research study reported herein was (a) to investigate possible risk-of-failure parameters and to develop a multi-criteria decision support system (DSS) for the modelling of the behaviour of water distribution networks, (b) to devise a methodology for arriving at an educated decision on whether to repair or replace a pipe-segment of interest, and (c) to devise expert rules for sequencing any such possible repairs as part of long-term network rehabilitation strategies.

The models and methodologies developed are based upon pipe-breakage incident data from New York City (U.S.A.) and Limassol (Cyprus) covering a combined period of about 10 years, and complement previous studies by adding to the analytical models a more general framework for failure and time-to-failure estimation by use of artificial intelligence techniques.

3 State of Knowledge

To-date, several studies on infrastructure assessment, deterioration modelling and network reliability have been reported upon. The intent has always been the provision of much-needed scientific insight on piping failures to operators of water distribution networks so that they can arrive at intelligent "repair-or-replace" decisions. The studies have most typically attempted to identify statistical relationships between break rates and influential risk factors.

For example, most studies in literature show a relationship between failure rates and time of failure (age of pipes), and some of them suggest a methodology to optimize the replacement time of pipes. Shamir and Howard (1979) reported an exponential relationship and Clark et al. (1982) developed a linear multivariate equation to characterize the time from pipe installation to the first break and a multivariate exponential equation to determine the breakage rate after the first break. A study by Andreou et al. (1987) suggested a probabilistic approach consisting of a proportional hazards model to predict failure at an early age, and a Poisson-type model for the later stages, and further asserted that stratification of data (based on specific parameters) would increase the accuracy of the model. A non-homogeneous Poisson distribution model was later proposed by Goulter and Kazemi (1988) to predict the probability of subsequent breaks given that at least one break had already occurred. Finally, Kleiner and Rajani (1999) developed a framework to assess future rehabilitation needs using limited and incomplete data on pipe conditions.

More recently, an inventory of water mains in New York City was utilized in a study by Vanrenterghem-Raven et al. (2004) for the development of analytical models in relation to structural degradation of urban water distribution networks. The models are based on survival analysis and proportional hazards. Additional work on the same case study was reported by Aslani (2003) and Christodoulou et al. (2003). The knowledge gained by the New York City case study was furthered and reported upon by Christodoulou et al. (2007) in a developed framework for integrated GIS-based management, risk assessment and prioritization of water leakage actions.

The problem of water losses was also at the center of the work by Tabesh et al. (2009), in their quest to develop an integrated model for evaluating water losses in water distribution networks. Their work parallels the work reported by Christodoulou et al. (2007) in terms of developing the means for evaluating and lowering non-revenue water by means of pressure reduction, water balance and minimum night-flow measures, and for linking mathematical models to a GIS application.

Other techniques found in literature include the work by Park et al. (2008), and Ekinci and Konak (2009). Park et al. (2008) presented and compared applications of a the log-linear ROCOF method and the power law process to model pipe failure rates and to estimate economically optimal replacement times, concluding that the log-linear ROCOF method performed better. Furthermore, they concluded that recording each failure time results in better modelling of the failure rate than observing failure numbers in some time intervals. Ekinci and Konak (2009) presented an optimization strategy based on the minimization of head losses, by use of a weighting approach in conjunction with the least-squares method.

4 Neurofuzzy Multicriteria Analysis

4.1 Traditional Methods for Multicriteria Analysis

A decision-support system (DSS) is a compilation of procedures that aim the provision of assistance to decision-makers for solving problems in a systematic manner and with a set of predefined steps. The term "multi-criteria analysis (MCA)" refers to any structured approach that is based on numerous decision variables and objectives. Multi-criteria analysis is a tool particularly useful in decision theory since in most cases decisions are based on the combined effects of more than one criterion. It is also highly useful in the case of hard-to-quantify parameters, non-numeric inputs and subjective appraisals. As Hajkowicz and Collins (2007) report, MCA has widespread and growing application in the field of water resource management, in topics ranging from water policy evaluation, strategic planning and infrastructure selection. In their report, 113 water management MCA studies from 34 countries were reviewed and categorized, concluding that fuzzy set analysis, paired comparison and outranking methods were among the most common methods.

Decision trees (branched models with a finite number of alternatives and a finite estimation of occurrence for each of these alternatives) are among the most common types of multi-criteria analysis tools. Also widely used is a process termed "analytical hierarchy process (AHP)", in which alternatives are compared between themselves in pairs and a ranking of their significance is obtained based on the pairwise comparisons. Expert systems, artificial neural networks, fuzzy logic and neurofuzzy systems are also widely used techniques in decision support systems.

4.2 Neurofuzzy Systems

Neurofuzzy systems are a relatively new method of analysis that is especially suitable for the representation and analysis of linguistic and patterned knowledge through artificial intelligence techniques. Such knowledge could be directly obtained from data analysis, or derived from previously acquired knowledge (past data and expert knowledge), or derived from implicit knowledge such as the experience of experts and/or heuristic rules.

Artificial neural networks (ANN) are the most common of artificial intelligence methods. ANN are ideal for the representation of non-linear parameter relationships and they are developed through pattern recognition algorithms and systematic knowledge acquisition processes that rely heavily on past data sets (training sets). In essence, most ANN applications utilize a supervised-learning process by which patterns in historical datasets are identified, understood and then represented in sets of inputs and outputs. The initial pattern recognition training phase is followed by a knowledge deduction phase in which the causal relations of the underlying data are deduced and a knowledge-base is formulated. This knowledge base is constantly and automatically updated based on the continuous inclusion of additional training datasets in the analysis. ANNs are among the most popular DSS tools, despite their "black box" nature (the inability of ANN to document and explain the computation steps and conclusions they reach).

Fuzzy logic (FL) is similar to ANN in many respects. FL differs, though, from ANN in the way it processes knowledge. FL deals with reasoning that is approximate rather than precisely deduced from classical predicated logic. Furthermore, unlike ANN FL is able to process non-numeric data such as linguistic terms of the form 'low risk', 'medium risk' and 'high risk'. Since its first introduction by Zadeh (1965) the concept of fuzzy logic has been applied to a variety of problems and has proven to be a valuable tool for the analysis of ill-defined and complex problems involving incomplete and imprecise information. The technique allows for set-membership values to range (inclusively) between 0 and 1, and in its linguistic form it allows for imprecise concepts (such as 'low', 'medium' and 'high') to be included in the analysis.

The combination of artificial neural networks with fuzzy logic results in a tool with a hybrid nature that is collectively more efficient than the two constituent parts (ANN and FL), for it combines pattern identification with linguistic data classification and rule deduction. Examples of neurofuzzy applications in water resource management can be found in the work by El-Shafie et al. (2007) and Lu et al. (2008). In the former, a neuro-fuzzy system is used for modelling inflow of the Nile river at the Aswan dam. In the latter, a fuzzy system is used for improving water resources management.

A neurofuzzy decision-support system usually consists of four levels of computation and analysis: an input level based on ANN, a fuzzification level with membership functions, an inference level with rules that process the input data and the fuzzy membership functions, and a defuzzification level that transforms the computational results to readable outputs. At the first level (the input layer) the input parameters $X_i (i = 1,2,3, ..., n)$ are read-in and an evaluation is made of the most important input factors. At the second level (the fuzzification layer) the input factors X_i are fuzzified according to predefined membership functions. Every input X_i has m degrees of membership $\mu_{A_i^j}(X_i)$, $(j = 1,2,3, ..., m)$, of the linguistic characteristics,

$$\mu_{A_i^j}(X_i) = f\left(a_i^j, b_i^j\right)$$

where $f\left(a_i^j, b_i^j\right)$ is the membership function used and a_i^j, b_i^j are the parameters of the membership functions. Membership functions are usually Gaussian, triangular or trapezoidal in shape, depending on the factor they are modelling and its assumed stochastic behaviour.

The third layer (the inference layer) is where the knowledge-processing and knowledge-extraction occurs based on the application of the predefined fuzzy rules. These are rules of the traditional IF-THEN form which are produced based on the collective experience and expertise of experts in the subject matter. The rules are, in essence, hypothetical scenarios of combinations of inputs and the assumed output (based on the experts' knowledge) and they attempt to capture the knowledge of the experts as they deal with the hypothetical input/output scenarios.

4.2.1 Fuzzy Membership Function, Fuzzification and Defuzzification

Membership functions (Fig. 1) represent the degree by which each element under consideration belongs to a set of interest (fuzzy theory permits the gradual assessment of the degree of membership of elements in a set). A fuzzy set is a pair (S,m) where S is a set and $m$is the mapping mechanism used to derive the degree of membership in the set (m: $A \to [0,1]$). For each value $x \in A$, the value $m(x)$ is the grade of membership of x in the set A. A value of "0" means that the element is not included in the set, while a value of "1" means that the element is fully included in the set. Values between "0" and "1" indicate partial membership to the set.

Should one define a_i as a fuzzy number such that $\forall a_i \in R$ (where R is the set of real numbers) then the value of a_i to be used in the fuzzy analysis depends on the membership function, $\mu(a_i)$, to be assumed (usually of triangular or trapezoidal shape) (Fig. 2). This number a_i ($i = 1,2,3, ..., m$) takes the form

$$a_i = \{x_1, x_2, x_3, x_4\}$$

in the case of a trapezoidal membership function, and the form

$$a_i = \{x_1, x_2, x_3\}$$

in the case of a triangular membership function, with $\{x_1, x_2, x_3, x_4\}$ being an ordered set ($x_1 < x_2 < x_3 < x_4$) and m being the number of parameters to be used in the analysis.

Fig. 1 Sample fuzzy-set membership function

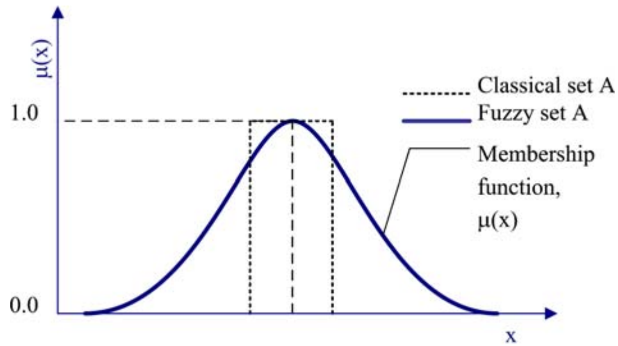
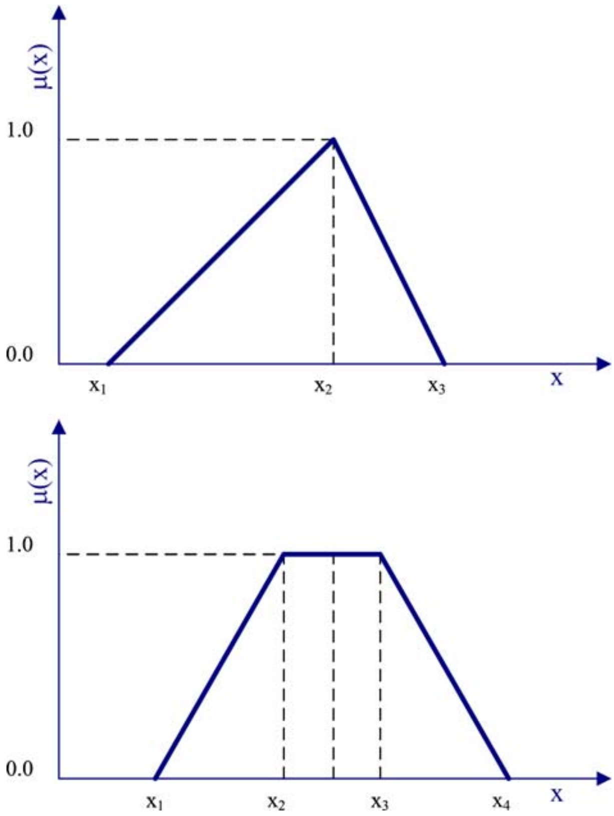


Fig. 2 Triangular and trapezoidal (typical) fuzzy-set membership functions



Defuzzification is the operation that reverts the fuzzy numbers to non-fuzzy (or crisp) values. The defuzzified value (d_i) produced by the aforementioned trapezoidal and triangular membership functions are

$$d_i = (x_1 + x_2 + x_3 + x_4)/4$$

and

$$d_i = (x_1 + 2x_2 + x_3)/4$$

respectively. This defuzzified value is used, in essence, for converting linguistic variables (such as low quality, high probability, high frequency) into well-defined numerical values with the intent of removing subjectivity and uncertainty from the decision-making process.

In typical multi-criteria decision-making problems the decision criteria are evaluated and scored based either on mathematical or numerical models (such as decision trees), or based on heuristic information (such as expert systems and AHP). The results of this analysis are then converted into rankings of the considered decision variables as related to their utility to the problem under investigation, which are then in turn converted into aggregated overall-performance values for each and every one of the alternatives in question. In comparison, in the case of fuzzy multi-criteria decision-making problems the ranking of alternatives takes into consideration the fuzzy scores and weights assigned to each decision criterion, and the fuzzified interaction between these criteria.

5 Case-Study Neurofuzzy System

5.1 Water Distribution Networks in Study (New York City, U.S.A. and Limassol, Cyprus)

Part of the current study is based on data collected from New York City's Department of Environmental Protection (NYCDEP). The study covers an urban area in the district of Queens (New York City) that includes sub-areas with high urban stress (dense network of highways and subways, elevated and underground) as well as quiet residential zones. The pipe network in the modeled area has a total length of about 365 km and consists of pipes of varying diameter, from 100 mm (~4 in.) to 1,800 mm (~72 in.). The majority of the pipes (51.4% of total length) have a diameter smaller than 200 mm and only a small percentage (5.5%) has a diameter over 900 mm. Almost all of these pipes are made of steel. In terms of material, 52.6% of pipes are made of cast iron, 15.8% made of lined cast iron, 15.8% made of ductile iron and 5% made of steel. It should also be noted that the majority of the pipes (52.6%) are more than 70 years old and only 17.2% are of less than 30-years of age. The data used in the study spans a 20-year period (1982–2002) and includes about 500 pipe breaks in a population of about 6,600 observed events (either related to true breakage events or simply false positives). The data used in the analysis included a number of presumed risk-of-failure factors, for which information was collected either at the time of the break event (such as age, diameter, etc.), or as an aftermath (such as vicinity to highways, subway system, residential and industrial areas, etc.).

The second water distribution network included in the study is from the city of Limassol (Cyprus). It is over 50 years of age and serves approximately 170,000 residents through approximately 64,000 consumer meters in an area of 70 km^2. The length of the piping network is about 795 km and the annual volume of potable water distributed through the piping network of pipes is about 13.7×10^6 m^3. The system infrastructure was developed in a well-organised fashion, utilizing pressure zones which are strictly governed by contours. Each pressure zone is subdivided into District Metered Areas (DMAs) of an average size of approximately 3,000 properties (Charalambous 2005). The diameters of the pipe distribution mains within the DMAs vary between 100 and 250 mm and where possible, interconnecting ring systems within the DMAs have been formed to minimize head loss at peak demands. The network owner (the Water Board of Limassol) maintains a proper water audit system (as per IWA's guidelines) and has over the years developed its infrastructure in such a way so as to be able to account efficiently and accurately for all water produced or "lost" (non-revenue water). The dataset used in the analysis spans a 5-year period (2002–2007) and about 2,000 incidents. The higher number of incidents in comparison to the New York City dataset is because of the wider city area covered in the study and because of the stricter waterloss-registering policies enforced in Limassol. These policies are greatly related to the fact that Cyprus is an arid country and water concervancy policies are of high national priority.

The datasets are unfortunately not publicly available (especially the New York City dataset) but the Limassol dataset could be provided by officially querying Limassol's Water Board.

5.2 Risk Factors and 'Repair or Replace' Decision Variables

As previously noted, a proper strategy for the operation and management of water distribution networks requires an assessment and quantification (if possible) of the condition of pipes, as well as an evaluation of the socioeconomic factors related to the piping network (such as the number of customers served, the proximity of pipes to residential or commercial areas, etc.). Direct examination of the pipes and evaluation of most of the risk-of-failure factors is time-consuming and costly. Thus, a means to infer their values through expert rules, data sampling and neurofuzzy systems appears to be an effective way to deduce conclusions on the comparative importance of the factors, their severity in time, and their possible outcomes in terms of the risk of failure. Neurofuzzy systems can process and link not only historical data and resulting data patterns (through the neural network component of them), but also the knowledge of experts in the particular field (through the fuzzy logic component of them).

The analysis is based on risk factors identified through prior studies (Vanrenterghem-Raven et al. 2004; Christodoulou et al. 2003, 2007). A difference between the aforementioned prior studies, though, and the current study can be found in the approach used for evaluating a network's vulnerability and the recommended management strategy: in prior studies a combination of statistical and survival analysis tools were used, whereas in this study neurofuzzy systems are employed.

For the New York City dataset, several presumed risk factors were originally considered but were later reduced to eight significant factors, based on statistical analysis of the dataset examined. The factors were: the number of observed previous breaks, the material, length, and diameter of a pipe, the traffic load in the viccinity of the pipe, the pipe's proximity to a highway, its proximity to an underground railway, and the pipe's proximity to a roadway/block intersection.

In the case of the Limassol dataset, only five risk factors were considered: the number of observed previous breaks, the material, length and diameter of a pipe, and the traffic load in the vicinity of each pipe. The list of "repair or replace" decision variables, though, was expanded to include socioeconomic factors such as the network's proximity to areas of high socioeconomic value, the type of neighbourhood (residential or industrial), coordination with other planned construction projects in the vicinity of the network, degree of inconvenience to the residents in the area, etc.

5.3 Establishing Weights for Decision Criteria

Even though the weights for the decision criteria are typically developed by use of expert systems or AHP, in the case of neurofuzzy systems these criteria can be obtained by use of the knowledge hidden in the historical data. This can be achieved through the ANN component of the neurofuzzy system, since ANN are very suitable for identifying patterns in the interactions of the underlying risk factors

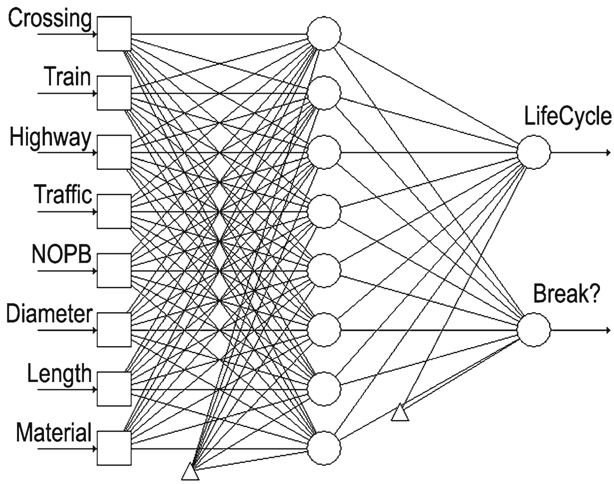


Fig. 3 Artificial neural network (ANN) implementation of the risk-of-failure analysis of water mains (New York City dataset)

(ANN inputs) and their possible contribution to a pipe's failure (the "*Break*" ANN output).

In the first stage of a three-stage analysis all historical data is processed through a three-layer back-propagation ANN consisting of one input, one hidden and one output layer (Fig. 3). The data used for training the ANN is the New York dataset with all eight presumed factors as ANN inputs, and with "LifeCycle (days)" and "Break?" used as the output neurons. It should be noted that both breakage and non-breakage incidents are included in the analysis.

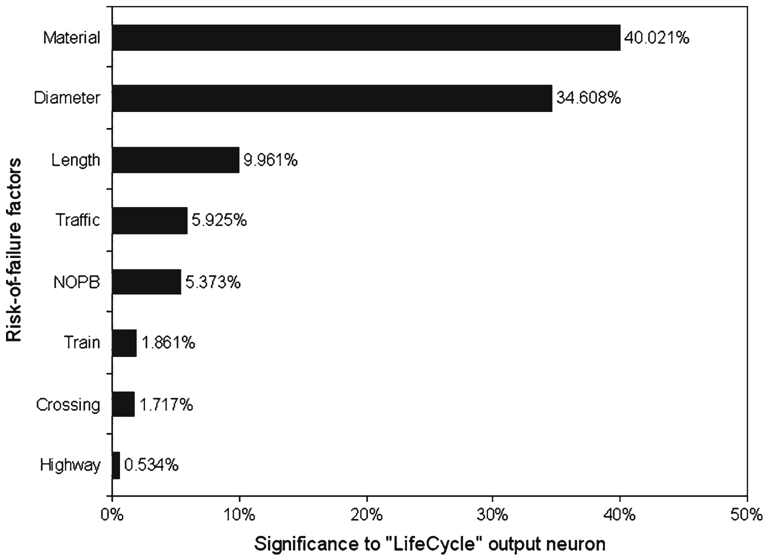


Fig. 4 Artificial neural network (ANN) risk-of-failure analysis and relative importance of the risk factors to the "LifeCycle" output neuron (New York City dataset)

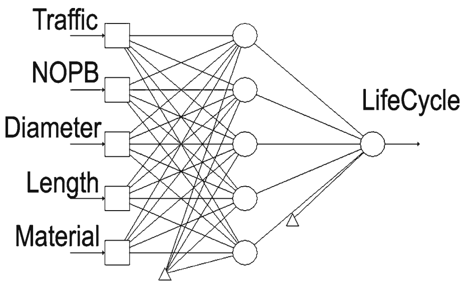


Fig. 5 Artificial neural network (ANN) implementation of the risk-of-failure analysis of water mains (Limassol dataset)

While ANN training is originally performed using all data in the set, once a 90% training accuracy is achieved the data is then filtered, non-breakage data is excluded from the analysis and the ANN training is amended. This reinforces the breakage-related ANN pattern recognition and helps identify the input factors most relevant to the "LifeCycle" output neuron. This information is provided by the ANN at the outset of the training phase and it is in the form of a relative-importance table that ranks the input factors based on the strength of the activation links between input and output neurons. As Fig. 4 depicts, for the dataset in study the ANN pattern classification identifies material type, diameter, length, traffic and number of observed previous breaks (NOPB) as the most important factors contributing to the risk for failure. The presence of a train, a road intersection or a highway seems to be of less risk-contributing importance than the former five factors.

This is followed by a second ANN stage in which the Limassol dataset is introduced to the analysis and tested against the previously trained network. The dataset consists of only the five most relevant risk factors, as identified in the previous stage, and the revised ANN consists of three layers, five input neurons and one output neuron (Fig. 5). It should be noted that, in this case the analysis contains only breakage events. As Fig. 6 shows, the NOPB risk factor now takes precedence over

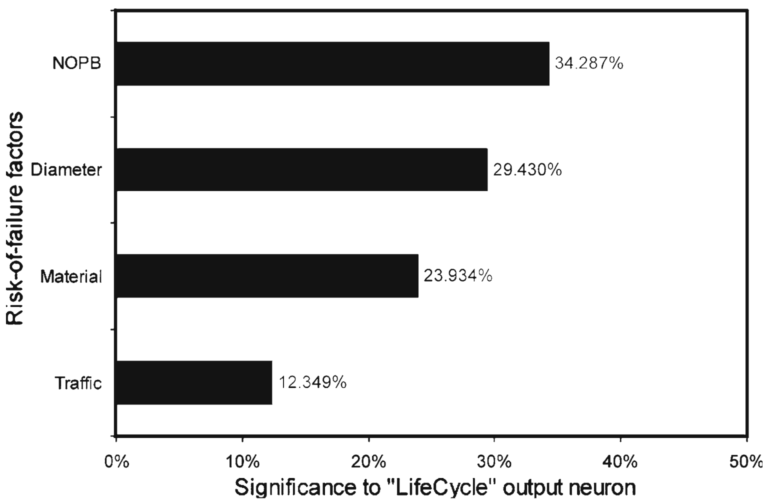


Fig. 6 Artificial neural network (ANN) risk-of-failure analysis and relative importance of the risk factors to the "LifeCycle" output neuron (Limassol dataset)

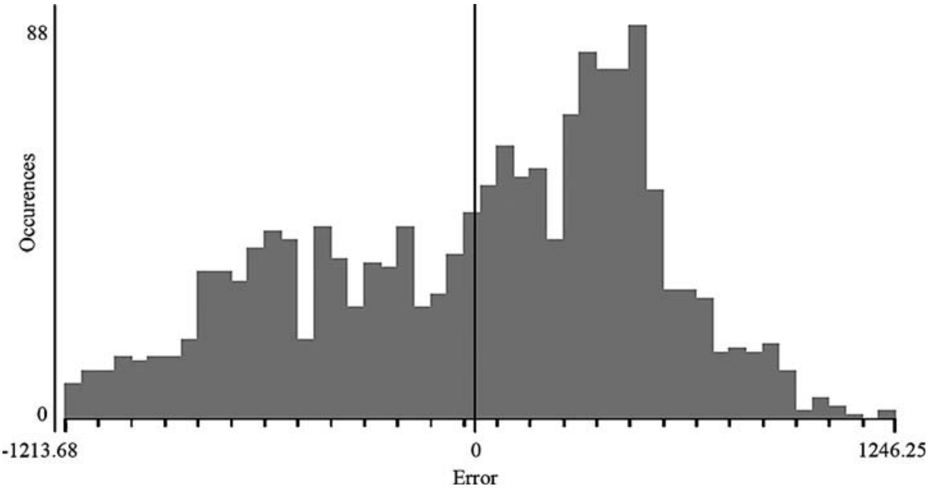


Fig. 7 Prediction error for the "LifeCycle" output neuron (Limassol dataset)

the other risk factors, followed by diameter, material and traffic. Furthermore, the error in pattern recognition and prediction of the "LifeCycle" neuron stays within [−1,213, 1,246] days, or else within [−3.3 years, 3.4 years] (Fig. 7).

The two ANN stages are followed by a third analysis stage in which the knowledge is transformed into rules by means of fuzzy logic. The deduced knowledge (fuzzy rules) is obtained by examining different combinations of inputs and recorded outputs (training vectors) and by combining the patterns in behaviour with expert rules. For example, the ANN data can be stratified by the number of previously observed breaks (NOPB) and the trained ANN rerun for several ranges of NOPB. The various input and output ANN combinations are used in devising FL rules of the form "IF NOPB is X, THEN LifeCycle is Y", or "IF NOPB is X AND Diameter is Y, THEN LifeCycle is Z", etc.

Table 1 shows the comparative results and the mean computational error of six such different applications of the neurofuzzy system. This indicates the degree of ANN training accuracy (the smaller the average error, the better the ANN pattern identification is). In the first application all eight presumed factors are taken into consideration (as inputs) and the neural network is made of two hidden layers. In the second application the eight inputs are maintained but only a single hidden layer is used. The rest of the applications are obtained by successively removing from the analysis one input at a time, based on the computed significance ranking of each input. In all cases, a simulation of 3,000 cycle iterations is used for training the network.

Table 1 Comparative results of ANN analysis

ANN structure (risk factors)	Average error in ANN training	
	2 hidden layers	1 hidden layer
8	0.079728	0.074911
7	0.076693	0.074370
6	0.080488	0.082210

Table 2 Fuzzy membership functions for risk-of-failure analysis' inputs and outputs

Risk factor	Membership function	Linguistic characterization and range of values			
Diameter	Triangular	Small: [4–30 in.]	Medium: [20–48 in.]	Large: [40–72 in.]	–
NOPB	Triangular	Small: [0–2]	Medium: [1–4]	Large: [3–9]	–
Train	Trapezoidal	0	1	–	–
Length	Triangular	Small: [0.25–5.5]	Medium: [4.5–14]	Large: [10–21]	–
Material	Trapezoidal	1	2	3	4
Traffic	Trapezoidal	0	1	2	–
Break?	Trapezoidal	0	1	–	–

The inputs and outputs of the system are then classified and assigned appropriate membership functions representing each factor's varying values within broader data sets, of linguistic nature. Examples of these linguistic values are characterizations of the form "low", "middle", "high" when describing and accounting for the contribution of the effect of a particular risk factor to the overall state of a pipe. In the case of the water distribution pipe network, the membership functions used were of triangular or trapezoidal form. The data boundaries of each factor and an overview of the fuzzy membership functions used in the analysis is tabulated in Table 2 and an example is shown in Fig. 8.

The defuzzification and decision-making process is introduced at the outset of the ANN, the fuzzification and the data-analysis processes (Figs. 2, 3, 4, 5, 6, 7). In this last stage the system builds on the knowledge acquired and transforms it into rules, while at the same time checking for the rules' validity and conformity with the observed input/output data sets.

For example, when the data set of [12, 1, 0, 6.5, 2, 1] is introduced to the neurofuzzy system as an input, the output produced for the "Break?" factor is 0, as expected. This was the case of a pipe with the characteristics: diameter of 12 units ($D = 12$), number of previously observed breaks equal to 1 (NOPB = 1), not in the vicinity of a subway (TRAIN = 0), with a length of 6.5 units ($L = 6.5$) and material of type 2 (MAT = 2), that it is subjected to heavy traffic (TRAF = 1). The combined effect of

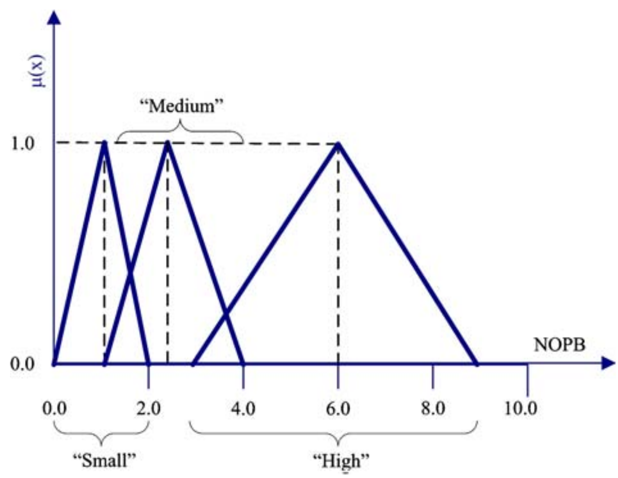


Fig. 8 Indicative graphical representation of fuzzy membership function for a linguistic risk-of-failure factor (the NOPB variable)

Table 3 Sample fuzzy rules for risk-of-failure analysis

Fuzzy rules					
If...					Then
Diameter	NOPB	Length	Material	Traffic	Break (1) OR not (0)
Small	Small	Small	2	1	0
Small	Small	Small	1	1	1
Large	Large	Small	4	2	1
Medium	Small	Small	4	2	0
Small	Small	Medium	–	1	0
Small	Small	Large	1	0	0
Medium	Medium	Small	4	2	1
Large	Large	–	2	–	1
–	Small	–	1	–	0

such acting risk factors is minimal on the risk-of-failure output factor ("Break?" = 0). Similarly, a sample dataset of [12, 0, 1, 2.25, 3, 0] results in a "Break?" output value of 0.5 meaning a possible failure, but in a not-so-crisp value (the pipe may possibly fail, or not). A dataset of [12, 3, 1, 6.5, 3, 1] results in a "Break?" output value of 0.7 indicating a heightened chance of failure.

Table 3 lists a subset of the derived rules and Fig. 9 shows a snapshot of the DSS rules for "NOPB = small". The presented in Table 3 top-level rules are only a subset of a broader set of rules devised by the neurofuzzy analysis, relating several input combinations to the "Break?" and "LifeCycle" outputs. Furthermore, the devised rules are dynamically updated by the neurofuzzy system through its ability for continuous training on newly acquired data. This is among the principle reasons for using such a sophisticated analysis tool. A variety of commercially available tools (such as MATLAB and Tiberius) were used for the analysis and several rules were derived based on the underlying data patterns.

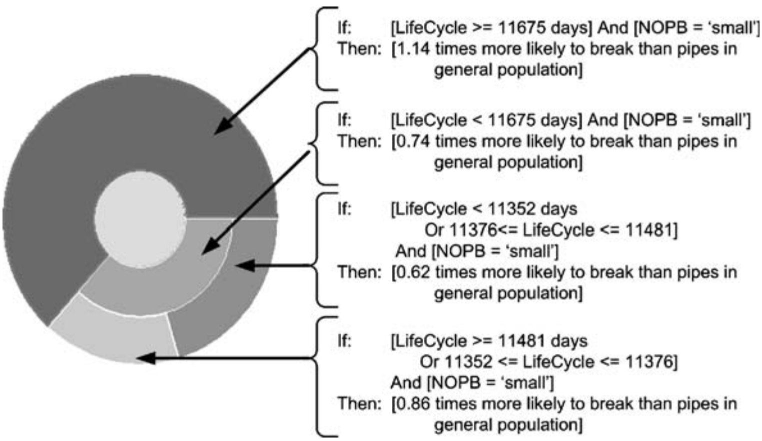


Fig. 9 Sample DSS rule derived from applying neurofuzzy classification

6 Fuzzy-Based Plan of Action

The present study, further to the development, training and validation of the neuro-fuzzy system, results in a management tool for sustainable and efficient management of water distribution networks. The management tool analyzes historical data, links it to neurofuzzy rules and through that to a prediction of the risk of failure over time for each pipe element in the network. The process used is outlined in Fig. 10.

A subset of the deduced top-level repair-or-replace rules deduced from the aforementioned neurofuzzy application and datasets is summarized below:

- Priority is given to areas in proximity of buildings of high public value (e.g. hospitals, schools).
- Priority is then given to areas combining residential and industrial use.
- Priority is then given to areas where other planned construction work is taking place (such as roadway rehabilitation) so as to maximize parallel work and minimize successive disruption.

Fig. 10 Flowchart of proposed neurofuzzy analysis

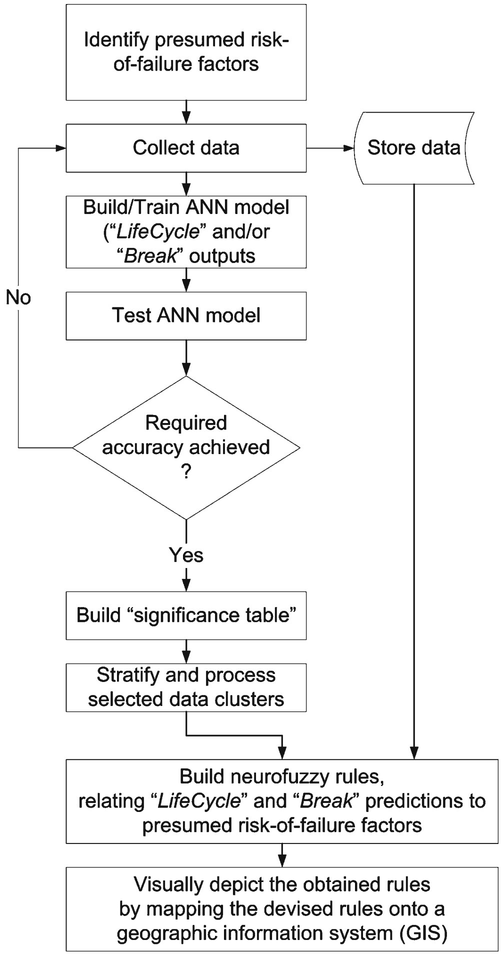


Fig. 11 Risk-of-failure analysis mapping at street level (colour variations indicate varying degrees for risk-of-failure)

- Priority is then given to pipes with a high number of observed previous breaks (NOPB).
- Then, pipes with large diameter take precedence.
- These are followed by pipes made out of cast-iron, followed in priority by steel pipes and plastic pipes.
- Finally, pipes subjected to heavy traffic loads take precedence.

One should note that the time and spatial priority of work (when and where to replace risk-prone pipes) is determined in close association with a risk-analysis study for a specific geographic area (as in the case of the applied neurofuzzy system and the datasets from New York City and Limassol) and can be better visualized by means of geographical information systems (GIS). The resulting integrated system can be deployed as a holistic decision support system for the management of water distribution networks, with the fusion of the deduced neurofuzzy rules mapped to a city street network as probability of failure metrics (Fig. 11) or risk contours.

7 Conclusions

The case-study presented (based on datasets from two locales: New York City, USA, and Limassol, Cyprus), aims at the development of knowledge related to water-main breaks and water loss in urban water distribution networks. The incomplete and time-sensitive multi-parameter nature of the data involved makes the employment of neurofuzzy systems a powerful tool for risk-of-failure analysis. The combination of artificial neural networks and fuzzy logic is extremely effective for the detection of patterns in the underlying data and then in the conversion of these patterns to knowledge and generic rules that can assist in risk-of-failure analysis and preventive maintenance of water distribution networks. The aforementioned pattern identification and acquired rule-based knowledge are dynamic in nature (the knowledge is updated with every new dataset added to the analysis) and easily adaptable to additional risk factors and decision criteria.

The work, which is still under development, is now piloted for implementation in two cities in Cyprus. The water utilities involved aspire, through this implementation, to reduce water losses in their water distribution networks and to improve on the reliability of their systems. The underlying knowledgebase and integrated decision support tools (statistics, ANN, fuzzy logic, GIS) aim to support these utilities in their endeavours and benefit their consumers the most.

Acknowledgements This report is based partially upon work supported by the National Science Foundation (under Grant No. CMS-0118376) and upon work supported by the Cyprus Research Promotion Foundation (under Grants No. AKGEN/1204/01 and PLYPH/0506/21). Gratitude is also extended to the Democritus University of Thrace for their participation as part of the EU-funded System Engineering and Management Program.

References

Andreou SA, Marks DH, Clark RM (1987) New methodology for modelling break failure patterns in deteriorating water distribution systems: applications. Adv Water Resour AWREDI 10:2–10

Aslani P (2003) Hazard rate modeling and risk analysis of water mains. MSc thesis, Polytechnic University

Charalambous B (2005) Experiences in DMA redesign at the water board of Lemesos, Cyprus. In: International water association's specialised conference on leakage. Halifax, Nova Scotia, Canada, pp 403–413

Christodoulou S, Aslani P, Vanrenterghem A (2003) A risk analysis framework for evaluating structural degradation of water mains in urban settings, using neurofuzzy systems and statistical modeling techniques. In: World water & environmental resources congress 2003 and related symposia, Pennsylvania, USA

Christodoulou S, Charalambous B, Adamou A (2007) Managing the 'repair or replace' dilemma on water leakages. In: International water association's (IWA) third specialty conference on water loss reduction (Water loss 2007), Bucharest, Romania

Clark RM, Stafford CL, Goodrich JA (1982) Water distribution systems: a spatial and cost evaluation. J Water Resour Plan Manage Div 108:243–256

Ekinci O, Konak H (2009) An optimization strategy for water distribution networks. Water Resour Manag 23:169–185

El-Shafie A, Taha MR, Noureldin A (2007) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. Water Resour Manag 21:533–556

Goulter IC, Kazemi A (1988) Spatial and temporal groupings of water main pipe breakage in Winnipeg. Can J Civil Eng 15:91–97

Hajkowicz S, Collins K (2007) A review of multiple criteria analysis for water resources planning and management. Water Resour Manag 21:1553–1566

Kleiner Y, Rajani B (1999) Using limited data to assess future needs. J Am Water Works Assoc 91:47–61

Lu HW, Huang G, Zeng GM, Maqsood I, He L (2008) An inexact two-stage fuzzy-stochastic programming model for water resources management. Water Resour Manag 22:991–1016

Park S, Kim BJ, Im GC (2008) Modeling of water main failure rates using the log-linear ROCOF and the power law process. Water Resour Manag 22:1311–1324

Shamir U, Howard C (1979) An analytical approach to scheduling pipe replacement. J Am Water Works Assoc 71:248–258
Tabesh M, Yekta AH, Burrows R (2009) An integrated model to evaluate losses in water distribution systems. Water Resour Manag 23:477–492
Vanrenterghem-Raven A, Eisenbeis P, Juran I, Christodoulou S (2004) Statistical modeling of the structural degradation of an urban water distribution system: case study of New York City. In: World water & environmental resources congress 2003 and related symposia, Pennsylvania, USA
Zadeh L (1965) Fuzzy sets. Inf Control 8:338–353