



# **NRC - CNRC**

## **Forensics of water quality failure in distribution systems – a conceptual framework**

**Sadiq, R.; Kleiner, Y.; Rajani, B.**

**NRCC-46742**

**A version of this document is published in / Une version de ce document se trouve dans :  
Journal of Indian Water Works Association, v. 35, no. 4, Oct-Dec. 2003, pp. 1-23**

<http://irc.nrc-cnrc.gc.ca/ircpubs>



# Forensics of water quality failure in distribution systems – a conceptual framework

Rehan Sadiq\*, Yehuda Kleiner and Balvant Rajani

Institute for Research in Construction, National Research Council Canada, Ottawa, Ontario K1A 0R6, Canada

Precise causes of water quality failures are often difficult to pinpoint. The complexity of the distribution system (many kilometers of pipes of different materials and ages), occurrences of physical/chemical/biological processes and the lack or absence of timely data make forensic analyses of water quality failure events very challenging. Water quality failure in the distribution system can occur through several pathways. These include intrusion of contaminants through failed or compromised pipes and cross-connections, regrowth of microbes in pipes and distribution storage tanks, leaching of chemicals or corrosion products from system components (pipes, tanks, liners), water treatment failure, deliberate contamination by terrorists and permeation of organic compounds through various plastic components of the system. Various indicators of water quality failure (symptoms) such as changes in the turbidity, odour, taste and colour, waterborne illnesses ranging from minor to serious, etc. can provide clues as to the causes of the failure in much the same way as symptoms of human health are used to diagnose causes and propose treatment.

A framework is proposed using fuzzy-based techniques and a knowledgebase on water quality. This framework is capable of backtracking symptoms to determine the candidate categories of contaminants, sources and/or pathways that cause a given water quality failure and rank them according to their respective likelihoods (or memberships). The usability of the framework is demonstrated with the help of a hypothetical case study.

***Key words:*** forensics, fuzzy sets, knowledgebase, water distribution, and water quality failure.

---

\*Corresponding author: [rehan.sadiq@nrc-cnrc.gc.ca](mailto:rehan.sadiq@nrc-cnrc.gc.ca)

# **1 INTRODUCTION**

## **1.1 WATER QUALITY FAILURE IN THE DISTRIBUTION SYSTEM**

A typical modern water supply system comprises the water source (aquifer or surface water source including the catchment basin), treatment plants, transmission mains, and the distribution system which includes pipes and distribution tanks. While water quality can be compromised at any component, failure at the distribution level can be extremely critical because it is closest to the point of delivery and, with the exception of a rare filtering device at the consumer level, there are virtually no safety barriers before consumption. Water quality failures, that compromise either the safety or the aesthetics of water in distribution networks, can generally be classified into the following major categories (Kleiner, 1998):

- Intrusion of contaminants into the distribution system through system components whose integrity was compromised or through misuse or cross-connection or intentional introduction of harmful substances in the water distribution system;
- Regrowth of microorganisms in the distribution network;
- Microbial (and/or chemical) breakthroughs and by-products and residual chemicals from the water treatment plant;
- Leaching of chemicals and corrosion products from system components into the water; and
- Permeation of organic compounds from the soil through system components into the water supplies.

Intrusion of contaminants into the water distribution system can occur through storage tanks (animals, dust carrying bacteria, infiltration) and pipes. Intrusion of contaminants through water mains may occur during maintenance and repair events, through broken pipes and gaskets, and cross-connections (Geldreich, 1990). Whenever the water pressure in a pipe is very low or negative, the risk of contamination through backflow or through leaky pipes increases. This can happen when the pipe is de-pressurized for repair or when the pipe is used for fire extinguishing or during transient pressures. Low or negative pressures, coupled with unprotected cross-

connections and/or contaminated soils, and/or leaky sewers in the pipe vicinity create a high risk of contaminant intrusion especially if the pipe is deteriorated with cracks and pinholes.

Biofilm is a deposit consisting of microorganisms, microbial products and detritus at the surface of pipes or tanks. Biological regrowth may occur when injured bacteria enter from the treatment plant into the distribution system. Under favorable conditions, such as nutrient supply (e.g., organic carbon) in the water and long residence time, these bacteria can attach themselves to surfaces, rejuvenate and grow in storage tanks and on rough inner surfaces of water mains. The regrowth of organisms in the distribution system results in an increased chlorine demand, which has two adverse effects: (a) a reduction in the level of free available chlorine may hinder the system's ability to contend with local occurrences of contamination (US EPA, 1999), and (b) an increased level of disinfection to satisfy the chlorine demand of biofilm results in higher concentrations of disinfection by-products (DBPs).

Disinfection is used to inactivate or kill pathogens. Chlorine has been highly successful in reducing the incidences of waterborne infections in human beings, but harmful DBPs are formed in the presence of natural organic matter (NOM) and bromide (from the source water) during chlorination. Other commonly used disinfectants are chloramines (combined chlorine), chlorine di-oxide and ozone. Ozone reacts with NOM and produces aldehydes, ketones and inorganic by-products. Ozone and chlorine di-oxide in the presence of bromide ion produce bromate and chlorate (and chlorite), respectively, which may have adverse effects on human health (US EPA, 1999).

Red water is one common water quality failure, which is a loss of aesthetics rather than a hazard to human health. Internal corrosion of metallic pipes and plumbing devices increases the concentration of metal compounds in the water. Different metals go through different corrosion processes, but in general low pH water, high dissolved oxygen, high temperature, and high levels of dissolved solids increase corrosion rates. Heavy metals such as lead and cadmium may leach into the water from pipes, causing significant health effects. Secondary metals such as copper (from home plumbing), iron (distribution pipes) and zinc (galvanized pipes) may leach into water and cause taste, odour and colour problems in addition to minor health related risks (Kleiner, 1998).

Permeation is a phenomenon in which contaminants (notably hydrocarbons) migrate through the pipe (plastic) wall. Three stages are observed in permeation phenomenon: (a) organic chemicals present in the soil partition between the soil and plastic wall, (b) the chemicals diffuse through the pipe wall, and (c) the chemicals partition between the pipe wall and the water inside the pipe (Kleiner, 1998). In general, the risk of contamination through permeation is relatively small as compared to other mechanisms. Figure 1 depicts a conceptual map of various factors that may play a role in the water quality arena. This map is by no means complete and comprehensive, but serves to demonstrate the complexity of the mechanisms.

Diagnostic or forensic analysis of contamination risk in water distribution systems is a difficult task due to the following:

- Water distribution system may comprise (depending on the size of the water utility) thousands of kilometers of pipes of different ages and materials;
- Operational and environmental conditions under which these pipes function, may vary significantly depending on the location of the pipes within the system;
- Since the pipes are not visible, it is relatively difficult and expensive to collect data on their performance and deterioration, and therefore indeed little field data are available;
- Some factors and processes affecting pipe performance are not completely understood; and
- It is often difficult to determine or validate an exact cause for water contamination or waterborne disease outbreak because such episodes are often investigated after the occurrence has ended. For these reasons, high uncertainties are inherent in any risk measure that may be assigned to the distribution system.

## **1.2 WATER QUALITY DEFINITIONS AND MONITORING**

Water quality is generally defined by a collection of upper and lower limits on selected possible contaminants in water (Maier, 1999). Water quality indicators (parameters or classes) can be classified into three broad categories: physical, chemical and biological contaminants. Within each class, a number of quality variables are considered. The acceptability of water quality for its intended use depends on the magnitude of these indicators (Swamee and Tyagi,

2000) and is often governed by regulations. A water quality failure (WQF) event is often defined as an exceedence of one or more water quality indicators from specific regulations, or in the absence of regulations, exceedence of guidelines or self-imposed, customer-driven limits.

As stated earlier, the physical, chemical and biological processes occurring in water distribution pipes are numerous and complex. A wealth of literature is available, describing the overall water quality represented by an aggregate index using various statistical and mathematical techniques. Swamee and Tyagi (2000) have discussed in detail the pros and cons of different techniques and approaches available for evaluating the overall index of water quality. Sinha *et al.* (1994) combined pH, chloride concentration, turbidity, residual chlorine, conductivity and MPN (Most probable number – a bacterial counting technique) into a single water quality index (0-100) through a weighting technique to represent an overall water quality at various nodes in the distribution system. Sadiq *et al.* (2003a) suggested a framework for the analysis of aggregative risk associated with water quality failure in the distribution system.

To monitor the quality of water in the distribution system, physical, chemical and biological parameters are recorded from routine grab sampling, followed by an analysis in the laboratory or using portable kits in the field. Sensor technology exists that enables capturing some parameters through online monitoring rather than grab samples. This technology is continually evolving to encompass more types of water quality indicators. Some common water quality indicators used for water distribution are alkalinity, dissolved oxygen, fluoride, nitrates, phosphate, residual disinfectant, specific organic compounds, temperature and turbidity (Hunsinger and Zioglio, 2002).

Regular monitoring programs help to identify a WQF if water quality indicators exceed regulatory regimes. Diagnostic/forensic techniques are used to relate the symptoms of WQF event to the causes (class of water quality indicators). The identification of such a breach is akin to identifying symptoms of a specific disease in the medical diagnosis. Pattern recognition techniques (an automatic recognition, description, classification and grouping of patterns) have been used extensively in variety of engineering and medical applications.

The fuzzy-based pattern recognition system typically involves data acquisition and pre-processing (knowledgebase), data representation and reporting, and decision-making

(identification of patterns). Fuzzy set theory (Zadeh, 1965) has been used for imprecise, incomplete and subjective knowledge for qualitative diagnostic modelling. Klir and Yuan (1995), Yao and Yao (2001) and Kiseliova and Peters (2001) used a decision-making technique for medical diagnosis using fuzzy sets and the compositional rule of inference. Chao and Cheng (1998) used fuzzy pattern recognition model for diagnosing cracks in reinforced concrete structures. Fuzzy-based pattern recognition techniques have been applied for character recognition (Chatterjee, 1982), texture classification (Hajnal and Koczy, 1982) and for earthquake modelling (Watada *et al.*, 1984).

## **2 WATER QUALITY FORENSICS**

### **2.1 FUZZY SETS AND FUZZY RELATIONS**

Fuzzy logic is the science of reasoning, thinking and inference that recognizes and uses the real world phenomenon that every thing is matter of degree. Fuzzy-based techniques are generalized form of interval analysis, which address uncertain and/or imprecise information. Zadeh (1965) introduced this concept, in which fuzzy sets are assigned to variables to represent uncertainties. A fuzzy set describes the relationship between an uncertain quantity  $x$  and a membership function  $\mu$ , 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  $\mu$  (Sadiq *et al.*, 2003b).

Fuzzy sets are distinguished from the probability concept in the type of uncertainty they describe. Probability is a way of handling and expressing randomness accounts for the likelihood that something is true or will happen. The concept centres around an all-or-none event, a binary concept- where the event can be defined as an entity having a particular characteristic, i.e., either it has that characteristic or not. Fuzzy sets on the other hand assess how much the entity is compatible with the definition of the characteristic.

The purpose of aggregation in fuzzy set theory is to meaningfully summarize and simplify a body of data. Familiar examples of aggregation techniques include arithmetic averages, geometric averages, harmonic averages, maximum and minimum values (Ayyub, 2001). Triangular norms ( $t$ -norm) and conorms ( $t$ -conorm) are generally used to define

aggregation operators. Ramik and Vlach (2001) and Sentz and Ferson (2002) provide a detailed description of the aggregation operators and fuzzy compositional rules used in this paper.

## 2.2 WATER QUALITY FAILURE INDICATORS (EFFECTS) AND CAUSES OF FAILURE – A RELATIONAL MATRIX

Let set  $P$  represent an array of water quality failure (WQF) indicators containing  $n$  elements (e.g., odour, colour, taste etc.)

$$P = [P_1 \ P_2 \dots P_i \dots P_n]_{1 \times n}; \quad i = 1, 2, \dots, n \quad (1)$$

Each WQF indicator  $P_i$  may be observed at various levels (symptoms), e.g., colour of water can be brownish, yellowish or greenish in case of a WQF failure, therefore each  $P_i$  can take on any of  $m_i$  elements

$$P_i = [P_{i1} \ P_{i2} \dots P_{ij} \dots P_{im_i}]_{1 \times m_i}; \quad j = 1, 2, \dots, m_i \quad (2)$$

A symptom  $P_{ij}$  may point to one or more WQF mechanisms or events. Let set  $C$  represents the classes of cause of a WQF event, such as physical (e.g., temperature), inorganic (e.g., metals), organic (e.g., trihalomethanes) or biological (e.g., *E-coli*). In total there are  $p$  classes in set  $C$  as follows:

$$C = [C_1 \ C_2 \dots C_k \dots C_p]_{1 \times p}; \quad k = 1, 2, \dots, p \quad (3)$$

A relational matrix  $R$  (to relate cause and effect) can be established between the classes of WQF and symptoms, based on hard (field) and soft (expert opinion) data. Let the matrix  $R_i$  represents the relationship between  $C_k$  (class of causes of WQF) and  $P_{ij}$  (symptom) as follows:

$$R_i = \begin{bmatrix} r_{i1,1} & r_{i1,2} & \dots & r_{i1,p} \\ r_{i2,1} & r_{i2,2} & \dots & r_{i2,p} \\ \vdots & & & \\ r_{im_i,1} & r_{im_i,2} & \dots & r_{im_i,p} \end{bmatrix}_{m_i \times p} \quad i = 1, 2, \dots, n \quad (4)$$

where  $r_{ij,p}$  represents a correlation (statistically derived or perceived qualitatively) between  $C_k$  and  $P_{ij}$ . Knowledge acquisition is required to explore and develop relationships between  $P_{ij}$  and  $C_k$ . Generally, knowledge acquisition consists of four distinct activities: preliminary analysis;



literature review; surveys/interviews and solicitations of opinions of an expert panel (McCauley-Bell and Badiru, 1996).

The preliminary analysis helps to obtain an overview of the problem and determine potential modular categories that would be useful in classifying various types of WQF risks. For water quality in distribution networks, this analysis could be carried out along the contamination pathways as illustrated in Figure 1. Literature review follows the preliminary analysis, which provides more comprehensive understanding of water quality failure. With comprehensive understanding, questionnaires and interview sessions can be designed to query the knowledge of utility personnel and other professionals working in the water industry. Finally, an expert panel can be assembled to discuss and organize the available information as well as to help fill identified knowledge gaps. The final data (correlation between WQF symptoms and the causes) may be qualitative, quantitative or a hybrid of both.

### 2.3 WATER QUALITY SAMPLING – A CONFIRMATIONAL MATRIX

Let  $N$  represent a set of  $l$  water samples, collected from the distribution system to monitor the WQF indicators.

$$N = [N^1 \ N^2 \ \dots \ N^h \ \dots \ N^l]_{1 \times j}; \quad h = 1, 2, \dots, l \quad (5)$$

where  $N^h$  represents a single sample. Each  $N^h$  encompasses measurements of  $i$  ( $i = 1, 2, \dots, n$ ) WQF indicators, and as stated in equation (2) can take a value of  $P_{im_i}$ . Let  $Q_i^h$  represents a water quality report about WQF indicator  $i$  in sample  $N^h$ :

$$Q_i^h = [q_{i1}^h \ q_{i2}^h \ \dots \ q_{ij}^h \ \dots \ q_{im_i}^h]_{1 \times m_i} \quad j = 1, 2, \dots, m_i \quad (6)$$

where  $q_{ij}^h$  represents a level of certainty that any water sample  $N^h$  may have certain symptom  $P_{ij}$ . For each sample  $N^h$ , the matrix  $Q_i^h$  contains  $m_i$  (number of) symptoms. If any symptom is not reported (or is below detection limit) that entry is considered as zero.  $Q_i^h$  is commonly referred to as confirmational matrix.

## 2.4 FUZZY COMPOSITION FOR WATER QUALITY FAILURE

Let  $T_i^h$  denote a fuzzy set with  $p$  members (or tuples), where each tuple  $t_{i,k}$  represents a certainty level that in sample  $N^h$  the WQF indicator  $i$  was caused by WQF class  $k$ .  $T_i^h$  can be estimated from the water sample report  $Q_i^h$  and the relational matrix  $R_i$

$$T_i^h = Q_i^h \circ R_i = [q_{i1}^h \quad q_{i2}^h \quad \cdots \quad q_{im_i}^h]_{1 \times m_i} \circ \begin{bmatrix} r_{i1,1} & r_{i1,2} & \cdots & r_{i1,p} \\ r_{i2,1} & r_{i2,2} & \cdots & r_{i2,p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{im_i,1} & r_{im_i,2} & \cdots & r_{im_i,p} \end{bmatrix}_{m_i \times p} \quad (7)$$

$$T_i^h = [t_{i,1} \quad t_{i,2} \quad \cdots \quad t_{i,p}]_{1 \times p}$$

where “o” represents composition operators defined with the help of  $t$ -norms and  $t$ -conorms. The extension of (7) to all WQF indicators  $i$  ( $i = 1, 2, \dots, n$ ) results in matrix  $T^h$  where each  $i$ th row corresponds to a specific  $T_i^h$

$$T^h = \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,p} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n,1} & t_{n,2} & \cdots & t_{n,p} \end{bmatrix}_{n \times p} \quad (8)$$

Let  $I$  denote a matrix representing the relative importance of the presence of a symptom  $P_{ij}$  as a testimonial to the occurrence of an event of WQF, and let  $B^h$  denote a non-weighted  $p$ -tuple fuzzy set, in which element  $b_k^h$  denotes the likelihood that a water sample  $h$  is evidence to the occurrence of a class  $k$  WQF:

$$I = [I_1 \quad I_2 \quad \cdots \quad I_i \quad \cdots \quad I_n]_{1 \times n} ; \quad i = 1, 2, \dots, n \quad (9)$$

$$B^h = I \circ T^h = [I_1 \quad I_2 \quad \cdots \quad I_i \quad \cdots \quad I_n]_{1 \times n} \circ \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,p} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n,1} & t_{n,2} & \cdots & t_{n,p} \end{bmatrix}_{n \times p} \quad (10)$$

$$B^h = [b_1^h \quad b_2^h \quad \cdots \quad b_k^h \quad \cdots \quad b_p^h]_{1 \times p}$$

where “o” represents composition operators defined with the help of  $t$ -norms and  $t$ -conorms. Equation (10) can be extended to  $l$  water samples as follows

$$B = \begin{bmatrix} b_1^1 & b_2^1 & \cdots & b_p^1 \\ b_1^2 & b_2^2 & \cdots & b_p^2 \\ \vdots & & & \\ b_1^l & b_2^l & \cdots & b_p^l \end{bmatrix}_{l \times p} \quad (11)$$

Water quality sampling can be carried out in various ways. Often, several samples are collected from the same sampling location to increase the reliability of the findings. Sometimes there is a reason to suspect that some samples are more (or less) reliable than others, in which case different weights can be assigned to samples according to their perceived reliabilities. Further, samples could also be weighted by the importance of their sampling locations, such as the relative importance of the water main sampled (e.g., based on size or flow rate), proximity to the water storage facility and the type of consumers (e.g., sensitive population i.e., close to school or hospital). Let  $W$  denote a sample weight matrix, in which each element  $W^h$  corresponds to the weight of sample  $h$

$$W = [W^1 \quad W^2 \quad \cdots \quad W^h \quad \cdots \quad W^l]_{1 \times l} ; \quad h = 1, 2, \dots, l \quad (12)$$

Importance and weight matrices ( $I$  and  $W$ ) can be established through the analytic hierarchical process (AHP). The AHP is a technique commonly used for multiple-criteria analysis. The AHP develops a linear additive model, which derives weights by performing pairwise comparisons between criteria or attributes (Ziara *et al.*, 2002). Once the  $W$  is established, the matrix  $B$  can be weighted

$$A = W \circ B = [W^1 \quad W^2 \quad \cdots \quad W^h \quad \cdots \quad W^l]_{1 \times l} \circ \begin{bmatrix} b_1^1 & b_2^1 & \cdots & b_p^1 \\ b_1^2 & b_2^2 & \cdots & b_p^2 \\ \vdots & & & \\ b_1^l & b_2^l & \cdots & b_p^l \end{bmatrix}_{l \times p} \quad (13)$$

$$A = [a_1 \quad a_2 \quad \cdots \quad a_k \quad \cdots \quad a_p]_{1 \times p}$$

where “o” represents composition operator defined with the help of  $t$ -norms and  $t$ -conorms as described before. The value of  $a_k$  represents the weighted aggregate likelihood, based on all  $l$

water samples, that a class  $k$  WQF event has occurred. The highest value of  $a_k$  is typically designated the most probable WQF class.

The relational matrices  $R_i$  provide a knowledgebase for the proposed framework. These relational matrices can be expanded to additional hierarchical levels, e.g. the water quality failure class  $C_k$  can be related to more specific classes of contaminants. These specific contaminants (classes of WQF) can be further linked to the basic sources and pathways of water quality failure.

### 3 HYPOTHETICAL CASE STUDY

Table 1 shows six major WQF indicators ( $P_i; i=1, 2, \dots, n; n=6$ ), where each indicator enumerates all possible symptoms  $P_{ij}$  ( $i=1, 2, \dots, n; j=1, 2, \dots, m_i$ ). For example, the WQF indicator “colour” can take one of three values— colour 1, colour 2 and colour 3, which may be symptoms of a WQF event.

Three classes  $C_k$  ( $k=1, 2, \dots, p; p=3$ ) of causes of WQF are considered in this example, namely, physical ( $C_1$ , e.g. temperature), chemical ( $C_2$ , e.g. presence of heavy metals, organics) and microbiological ( $C_3$ , e.g. *E-coli*, heterotrophic plate counts). Table 2 shows the relational matrices  $R_i$ , which express the correlations between symptoms ( $P_{ij}$ ) and the cause of classes,  $C_k$ , of WQF. Matrices  $R_i$  constitute the knowledgebase, which can be established from field, laboratory data and expert judgment or any combination thereof. Membership functions (Figure 2) can be used to define these correlations if available knowledge is vague or imprecise, or probabilities can be used if sufficient hard data are available. For example, if statistical data show that symptom  $P_{11}$  is caused 80% of the time by  $C_2$  and 20% of the time by  $C_3$ , the values 0.8 and 0.2 can be designated to  $C_2$  and  $C_3$ , respectively. If the data are not available, experts can assign these values using a linguistic scale similar to that shown in Figure 2 (Chao and Cheng, 1998).

Assume that the water quality is monitored at four sampling locations and that one sample  $N^h$  was collected from each location ( $h=1, 2, \dots, l; l=4$ ). Table 3 shows (from equation (6)) certainty levels,  $q_{ij}^h$  associated with water sample  $N^h$  for symptom  $P_{ij}$ . For example, a certainty level of 0.8 (*high to very high*) was assigned to symptom  $P_{11}$  (colour 1) in sample  $N^1$ ,

and a certainty level 0.7 (*medium to high*) for the same symptom in sample  $N^2$ , and so on. Note that in the sample  $N^1$  the observer is also about 0.3 (*low to medium*) certain that colour 2 was observed and 0.2 (*very low to low*) certain that it was colour 3. This illustrates the reality in which distinctions between observed phenomena are imprecise and often vague. Again the linguistic scale (Figure 2) can be used to define the certainty level of the presence of symptoms in samples. Matrices  $Q_i^h$  are compiled from Table 3.

Once both matrices  $Q_i^h$  and  $R_i$  are established, the matrix  $T_i^h$  can be calculated using equation (7) and subsequently  $T^h$  can be calculated using equation (8). Note that for comparison, equation (7) was applied using three different composition operators, *max-min* (*reflects low uncertainty range*), *sum-prod* (*reflects high uncertainty range*) and *mix* of both *max-min* and *sum-prod*. These three compositional operators express the various degrees of and-ness and or-ness in the application of fuzzy sets. Logical operators like *max-min* are more restrictive than, say, *sum-prod* and *max-prod*. Application of different operators leads to a range of outcomes that permits the analyst to assess the likely cause of a WQF event. Arbitrarily chosen importance factors ( $I$ ) for symptoms and weights for sampling locations ( $W$ ), required to establish matrices  $B^h$  and  $A$ , are shown in Table 4. The non-weighted likelihood that water sample  $h$  is due to class  $k$  WQF as per equation (10) is shown in the top part of Table 5. Dramatically different non-weighted likelihoods  $B^h$  can also serve as quality control measures for water sample tests. The weighted aggregate likelihood of all  $l$  water samples due to a class  $k$  WQF is shown in the last row of Table 5. Note those membership values  $b_k^h$  and  $a_k$  are normalized in Table 5, so that the sum of every row is unity. This facilitates the conversion of the results from membership values to probabilities (Yao and Yao, 2001). The *max-min* and *mixed* compositional operator indicate that, the highest likelihood (probability or a normalized membership) of the water quality failure in this example is due to chemical agents ( $C_2$ ), followed by microbial ( $C_3$ ) and physical agents ( $C_1$ ). However, the *max-min* compositional operator indicates that both chemical ( $C_2$ ) and microbial ( $C_3$ ) agents are equally likely to have caused the WQF event. This apparent discrepancy serves to point out how the outcome can differ when different compositional operators are used but this type of analyses offers leading clues of what issues should be addressed first.

#### 4 DISCUSSION AND CONCLUSIONS

Forensic analysis of a WQF event in a complex system inherently involves quantitative and qualitative assessments based on hard knowledge, supported by field data, as well as on soft imprecise knowledge acquired through anecdotal observations and vague notions of experts. The proposed framework using fuzzy-based approaches is uniquely suited to synthesize and make use of both types of knowledge. Fuzzy-based approaches allow the explicit consideration of uncertainties, which propagate through the analysis. It should however be noted that the framework presented in this paper is a simplified demonstration of the approach. A comprehensive structure would require a major effort, including the collaboration of several experts in the various related disciplines of knowledge.

The proposed method can be used to develop a system in which data (symptoms) obtained from sensors or field or lab tests are interpreted to help determine the likely source of a WQF and/or the pathway through which a contaminant was introduced into the system. The method is modular and scalable and thus able to accommodate new knowledge and information. It is easily programmable for computer applications and can be developed into a reliable water quality forensic tool for a water distribution system. It can also be linked to geographic information systems (GIS) and online monitoring systems to determine the likelihood of causes of water quality failures at various locations in a distribution system. Uncertainty in the diagnosis improves as the uncertainties in the various inputs and parameters are reduced. The proposed approach can help pinpoint those areas in which more data would yield the highest benefits in terms of accurate results.

Different compositional operators can be used at different hierarchical levels. The proposed method may be sensitive to the selection of compositional operators, as is illustrated in the variability of the results of the numerical example. A comprehensive analysis would require appropriate sensitivity tests. However, ultimately more research is needed to determine which compositional operators are most appropriate for various levels.

The proposed framework may also be sensitive to the weights that are assigned at various compositional levels of the analysis. There are numerous ways to derive weights, prominent among which is the analytic hierarchical process mentioned earlier. Regardless of the method

used for weight derivation, sensitivity analyses would have to be implemented in any credible analysis. It would also be possible to assign fuzzy rather than crisp weights, and then examine the resulting most-likely values and the tolerances (e.g., Deng, 1999; Triantaphyllou and Lin, 1996).

The proposed framework accommodates both qualitative and quantitative data as described before. Some data may be supported by rigorous observations (historical data based on probabilities), while other data may be based on loosely supported or anecdotal-based beliefs (expert opinion). These two types of data should have different weights in the aggregation process. The framework in its current form does not address this need to distinguish between data obtained from sources with different reliabilities.

Another issue that requires further research is the way in which the symptoms  $q_{ij}$  are reported in the water samples (confirmational matrix). In the presented approach, single values of memberships (called singletons in fuzzy set theory) are used to establish confirmational matrices. The analyses can be more flexible if fuzzy numbers were used instead. The same can also be said about the relational matrices for  $P_{ij}$ . Currently, all relations  $r_{ij,p}$  are represented by singletons, which carry the same weight. In reality, some of these relations would typically be supported by hard data, while others would be more loosely based on anecdotal information. A possible way of considering these differential reliabilities of knowledge is by expressing the relations using fuzzy rather than crisp numbers (memberships), where the less reliable relations would assume a fuzzy number with a wider range of uncertainty.

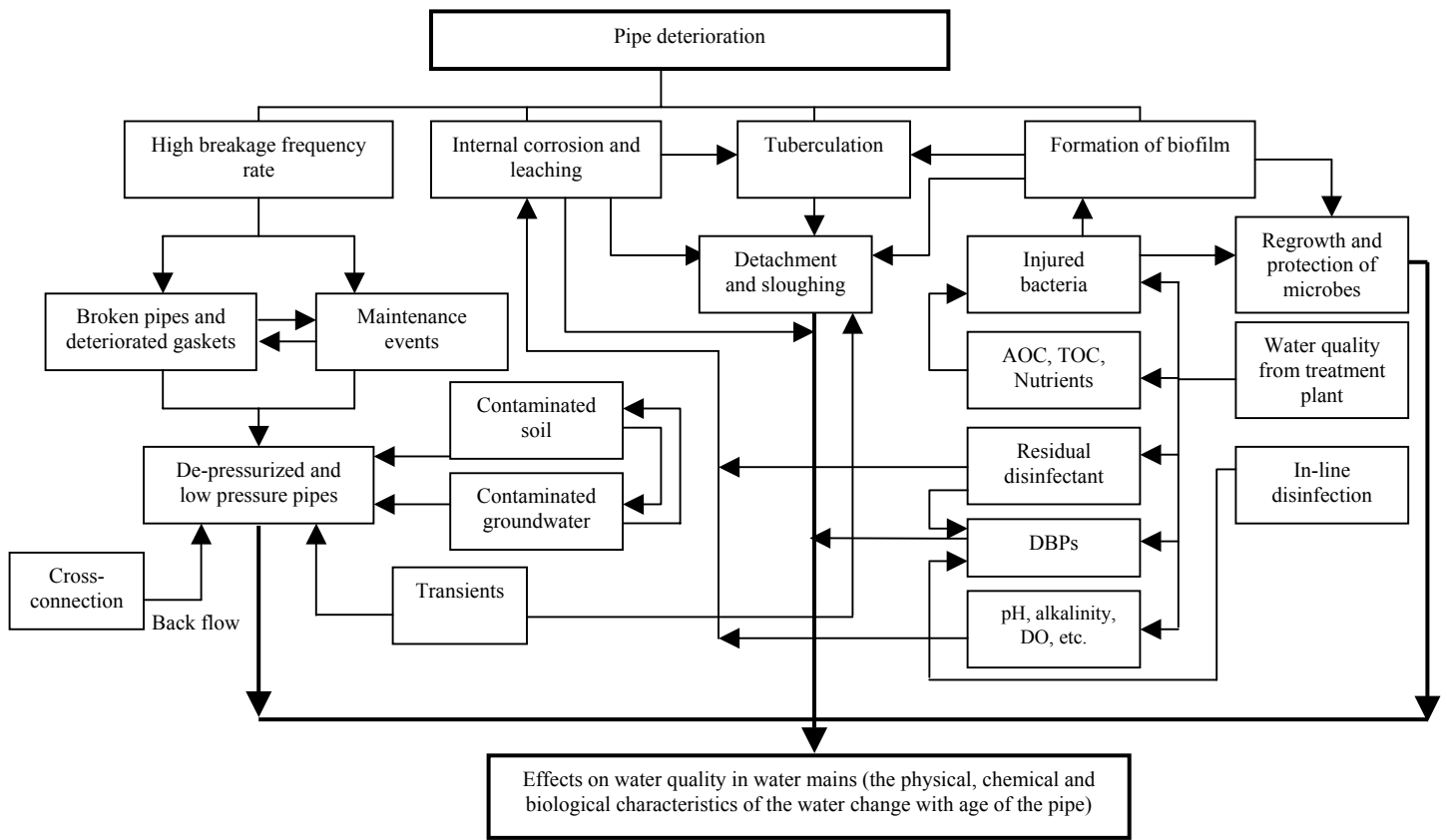
## 5 REFERENCES

- Ayyub, B.M. 2001. *Elicitation of expert opinions for uncertainty and risks*, CRC Press.
- Chao, C-J., and Cheng, F-P. 1998. Fuzzy pattern recognition model for diagnosing cracks in RC structures, *ASCE Journal of Computing in Civil Engineering*, 12(2): 111-119.
- Chatterjii, B.N. 1982. Character recognition using fuzzy similarity relations, In: *Fuzzy information and decision processes*, Ed. Gupta, M.M., and Sanchez, E., North-Holland Publishing Co, New York, 131-137.
- Deng, H. 1999. Multicriteria analysis with fuzzy pairwise comparison, *International Journal of Approximate Reasoning*, 21: 215-231.

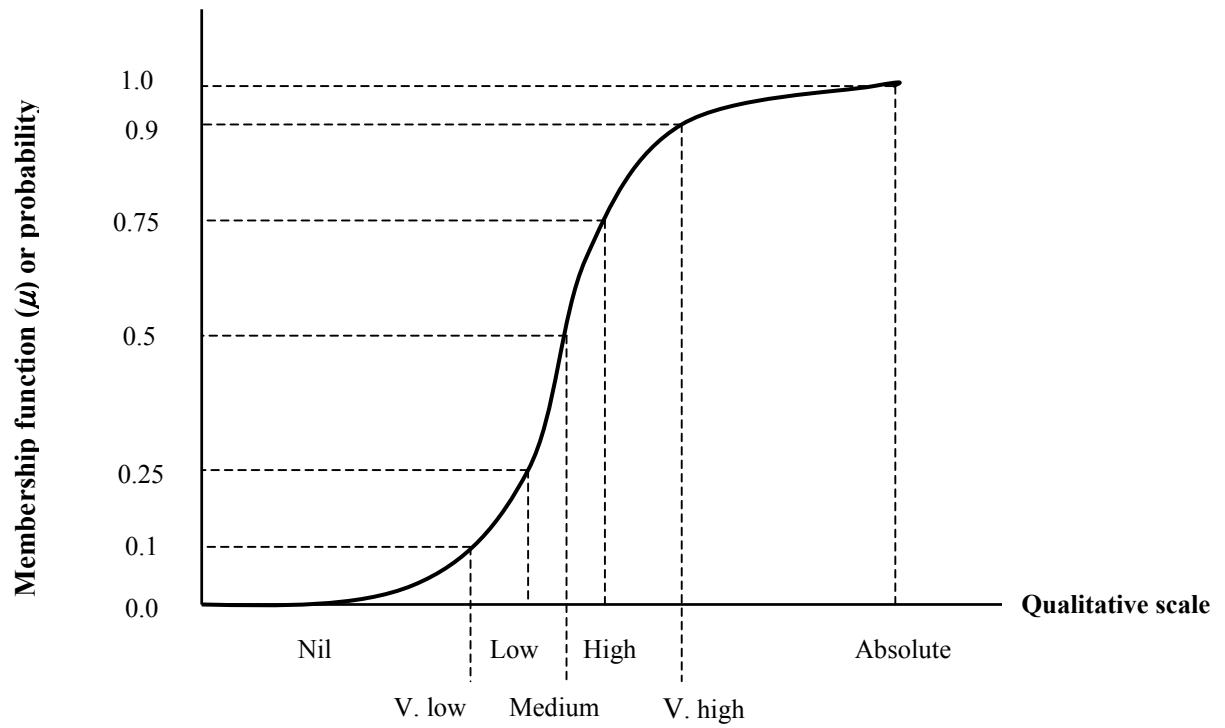
- Geldreich, E.E. 1990. Microbiological quality control in distribution systems, In: *Water quality and treatment*, Ed. Frederick, W.P., American Water Works Association, NY, McGraw-Hill Inc.
- Hajnal, M., and Koczy, L.T. 1982. Classification of textures by vectorial fuzzy sets, In: *Fuzzy information and decision processes*, Ed. Gupta, M.M., and Sanchez, E., North-Holland Publishing Co, New York, 157-164.
- Hunsinger, R.B., and Zioglio, G. 2002. Rationale for online monitoring, In: *Online monitoring for drinking water utilities co-operative research report*, Ed. Hargesheimer, E., Conio, O., and Popovicova, J., American Water Works Association Research Foundation, CO.
- Kiseliova, T., and Peters, H. 2001. Decision-making in diagnosis of oral Mucosa Lesions, In: *Fuzzy days 2001*, Ed. Reusch, B., LNCS 2206, pp. 33-39.
- Kleiner, Y. 1998. Risk factors in water distribution systems, *British Columbia Water and Waste Association 26<sup>th</sup> Annual Conference*, Whistler, B.C., Canada.
- Klir, G.J., and Yuan, B. 1995. *Fuzzy sets and fuzzy logic - theory and applications*, Prentice-Hall, Inc., Englewood Cliffs, NJ, USA.
- Maier, S.H. 1999. *Modeling water quality for water distribution systems*, Ph.D. thesis, Brunel University, Uxbridge.
- McCauley-Bell, P., and Badiru, A.B. 1996. Fuzzy modeling and analytic hierarchy processing to quantify risk levels associated with occupational injuries – Part I: the development of fuzzy linguistic risk levels, *IEEE Transactions on Fuzzy Systems*, 4(2): 124-131.
- Ramik, J., and Vlach, M. 2001. *Generalized concavity in fuzzy optimization and decision analysis*, Kluwer Academic Publishers, Boston.
- Sadiq, R., Kleiner, Y., and Rajani, B.B. 2003a. Aggregative risk analysis for water quality failure in distribution networks, Submitted to *AQUA - Journal of Water Supply: Research & Technology*.
- Sadiq, R., Rajani, B.B., and Kleiner, Y. 2003b. A fuzzy-based method to evaluate soil corrosivity for prediction of water main deterioration, Submitted to ASCE, *Journal of Infrastructure Systems*.
- Sentz, K. and Ferson, S. 2002. Combination of evidence in Dempster-Shafer theory, SAND 2002-0835.
- Sinha, R., Gupta, P., and Jain, P.K. 1994. Water quality modeling of a city water distribution system, *Indian Journal of Environmental Health*, 36(4): 258-262.
- Swamee, P.K., and Tyagi, A. 2000. Describing water quality with aggregate index, *ASCE Journal of Environmental Engineering*, 126(5): 451-455.



- Triantaphyllou, E., and Lin, C-T. 1996. Development and evaluation of five fuzzy multi-attribute decision-making methods, *International Journal of Approximate Reasoning*, 14: 281-310.
- US EPA 1999. Microbial and disinfection by-product rules – simultaneous compliance guidance manual, United States Environmental Protection Agency, EPA 815-R-99-015.
- Watada, J., Fu, K.S., and Yao, J.T.P. 1984. *Linguistic assessment of structural damage*, Rcp. CE-STR-84-30, Purdue University, West Lafayette, Indiana.
- Yao, J.F-F., and Yao, J-S. 2001. Fuzzy decision-making for medical diagnosis based on fuzzy number and compositional rule of inference, *Fuzzy Sets and Systems*, 120: 351-366.
- Zadeh, L.A. 1965. Fuzzy sets, *Information and Control*, 8: 338-353.
- Ziara, M., Nigim, K., Enshassi, A., and Ayyub, B.M. 2002. Strategic implementation of infrastructure priority projects: case study in Palestine, *ASCE Journal of Infrastructure Systems*, 8(1): 2-11.



**Figure 1.** A conceptual map of causes of water quality failure in water mains



**Figure 2.** *A qualitative or linguistic scale (modified after Chao and Cheng, 1998)*

**Table 1.** Symptoms for water quality failure in distribution system

<b>*WQF indicators</b>	$P_i$ $i = 1, 2, \dots, n; n = 6$	<b>** Symptoms</b>	$P_{ij}$ $j = 1, 2, \dots, m_i$
Colour	$P_1$	Colour 1	$P_{11}$
		Colour 2	$P_{12}$
		Colour 3	$P_{13}$
Odour	$P_2$	Odour 1	$P_{21}$
		Odour 2	$P_{22}$
Taste	$P_3$	Taste 1	$P_{31}$
		Taste 2	$P_{32}$
Turbidity	$P_4$	Level 1	$P_{41}$
		Level 2	$P_{42}$
		Level 3	$P_{43}$
Minor illness	$P_5$	Minor illness 1	$P_{51}$
		Minor illness 2	$P_{52}$
		Minor illness 3	$P_{53}$
Major illness	$P_6$	Major illness 1	$P_{61}$
		Major illness 2	$P_{62}$

\* Water quality failure indicators

\*\* Levels of water quality failure indicators

**Table 2.** Relational matrices  $R_i = P_{ij} \times C_k$  ( $k = 1, 2, \dots, p; p = 3$ )

Symptoms, $P_{ij}$ ( $j = 1, 2, \dots, m_i$ )	Physical $C_1$	Chemical $C_2$	Microbial $C_3$
$P_{11}$ (colour 1)	0.00	0.80	0.20
$P_{12}$ (colour 2)	0.00	0.70	0.30
$P_{13}$ (colour 3)	0.10	0.60	0.30
$P_{21}$ (odour 1)	0.20	0.40	0.40
$P_{22}$ (odour 2)	0.30	0.50	0.10
$P_{31}$ (taste 1)	0.10	0.60	0.30
$P_{32}$ (taste 2)	0.00	0.30	0.70
$P_{41}$ (turbidity level 1)	0.50	0.40	0.10
$P_{42}$ (turbidity level 2)	0.70	0.20	0.10
$P_{43}$ (turbidity level 3)	0.90	0.10	0.00
$P_{51}$ (minor illness 1)	0.00	0.30	0.70
$P_{52}$ (minor illness 2)	0.00	0.60	0.40
$P_{53}$ (minor illness 3)	0.10	0.10	0.80
$P_{61}$ (major illness 1)	0.00	0.80	0.20
$P_{62}$ (major illness 2)	0.10	0.50	0.40

$R_4$

$r_{42,2}$

**Table 3.** Water quality report (confirmational matrix) for 4 sampling locations  
 ( $h = 1, 2, \dots, l; l = 4$ )

$Q_i^h$	$q_{ij}^h$	Sample $N^h$			
		$N^1$	$N^2$	$N^3$	$N^4$
$Q_1^h$	$q_{11}^h$	0.80	0.70	0.90	1.00
	$q_{12}^h$	0.30	$q_{12}^2$ — 0.10	0.30	0.00
	$q_{13}^h$	0.20	0.00	0.20	0.00
$Q_2^h$	$q_{21}^h$	0.00	0.00	$Q_2^3$ — 0.00	0.00
	$q_{22}^h$	0.00	0.00	0.00	0.00
$Q_3^h$	$q_{31}^h$	0.00	0.00	0.10	0.00
	$q_{32}^h$	0.00	0.00	0.00	0.00
$Q_4^h$	$q_{41}^h$	0.00	0.00	0.00	0.00
	$q_{42}^h$	0.00	0.00	0.00	0.00
	$q_{43}^h$	0.00	0.00	0.10	0.10
$Q_5^h$	$q_{51}^h$	0.00	0.00	0.00	0.00
	$q_{52}^h$	0.00	0.00	0.00	0.00
	$q_{53}^h$	0.00	0.00	0.00	0.00
$Q_6^h$	$q_{61}^h$	0.00	0.00	0.00	0.00
	$q_{62}^h$	0.00	0.00	0.00	0.00

**Table 4.** *Weights assigned to symptoms and sampling locations*

$I_i$	Importance weights for symptoms	$W^h$	Weights for water samples
$I_1$	0.15	$W^1$	0.50
$I_2$	0.10	$W^2$	0.20
$I_3$	0.10	$W^3$	0.25
$I_4$	0.15	$W^4$	0.05
$I_5$	0.20		
$I_6$	0.30		

**Table 5.** Summary of the example forensic analyses

Matrices	Composition operator	*Likelihood to Physical $C_1$	Likelihood to Chemical $C_2$	Likelihood to Microbial $C_3$
$B^1$	<i>sum-prod</i>	0.02	0.75	0.23
	<i>max-min</i>	0.24	0.38	0.38
	<i>mixed</i>	0.08	0.67	0.25
$B^2$	<i>sum-prod</i>	$b_1^2$ — 0.00	0.79	0.21
	<i>max-min</i>	0.00	0.50	0.50
	<i>mixed</i>	0.00	0.78	0.22
$B^3$	<i>sum-prod</i>	0.08	0.69	0.23
	<i>max-min</i>	0.24	0.38	0.38
	<i>mixed</i>	0.08	0.67	0.25
$B^4$	<i>sum-prod</i>	0.08	0.74	0.18
	<i>max-min</i>	0.24	0.38	0.38
	<i>mixed</i>	0.09	0.73	0.18
$A$	<i>sum-prod</i>	0.04	$a_2$ — <b>0.73</b>	0.23
	<i>max-min</i>	0.24	<b>0.38</b>	<b>0.38</b>
	<i>mixed</i>	0.07	<b>0.69</b>	0.24

\* values represent the normalized membership values