NRC Publications Archive
Archives des publications du CNRC

Estimating risk of contaminant intrusion in distribution networks using Dempster-Shafer theory of evidence
Sadiq, R.; Kleiner, Y.; Rajani, B. B.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.
For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1080/10286600600789276
Civil Engineering and Environmental Systems, 23, September 3, pp. 129-141, 2006-09-01

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=ce3247fc-615c-40b6-b52f-4fcd08a6ef67
https://publications-cnrc.canada.ca/fra/voir/objet/?id=ce3247fc-615c-40b6-b52f-4fcd08a6ef67

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

NRC·CNRC CNRC·NRC
Institute for Research in Construction
Institut de recherche en construction
http://irc.nrc-cnrc.gc.ca

Estimating risk of contaminant intrusion in water distribution networks using Dempster-Shafer theory of evidence

NRCC-45408

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

A version of this document is published in / Une version de ce document se trouve dans: Civil Engineering and Environmental Systems, v. 23, no. 3, Sept. 2006, pp. 129-141. doi:
10.1080/10286600600789276

National Research Council Canada Conseil national de recherches Canada
Canada

Estimating risk of contaminant intrusion in water distribution networks using Dempster-Shafer theory of evidence

Rehan Sadiq[1]

Associate Research Officer

Urban Infrastructure Program, Institute for Research in Construction

National Research Council Canada, Ottawa, Ontario Canada K1A 0R6

Tel: (613)-993-6282; Fax: (613)-954-5984

E-mail: Rehan.sadiq@nrc-cnrc.gc.ca

Yehuda Kleiner

Senior Research Officer and Acting Director

Urban Infrastructure Program, Institute for Research in Construction

National Research Council Canada, Ottawa, Ontario Canada K1A 0R6

Tel: (613)-993-3805; Fax: (613)-954-5984

E-mail:Yehuda.kleiner@nrc-cnrc.gc.ca

and

Balvant Rajani

Principal Research Officer

Urban Infrastructure Program, Institute for Research in Construction

National Research Council Canada, Ottawa, Ontario Canada K1A 0R6

Tel: (613)-993-3810; Fax: (613)-954-5984

E-mail: Balvant.rajani@nrc-cnrc.gc.ca

---

1 **Corresponding author**

ABSTRACT

Intrusion of contaminants into water distribution networks requires the simultaneous presence of three elements; contamination source, pathway and driving force. The existence of each of these elements provides ‘partial’ evidence (typically incomplete and non-specific) to the occurrence of contaminant intrusion into distribution networks. Evidential reasoning, also called Dempster-Shafer (DS) theory, has proved useful to incorporate both *aleatory* and *epistemic* uncertainties in the inference mechanism. The application of evidential reasoning to assess risk of contaminant intrusion is demonstrated with the help of an example of a single pipe. The proposed approach can be extended to full-scale water distribution network to establish risk-contours of contaminant intrusion. Risk-contours using GIS may help utilities to identify sensitive locations in the water distribution network and prioritize control and preventive strategies.

Keywords: Contaminant intrusion, water distribution networks, evidence theory, Dempster-Shafer theory, risk, GIS, simplex plot, uncertainty

INTRODUCTION

Water quality in a distribution network can be described by specific microbiological, physico-chemical and aesthetic attributes of water. These attributes are generally maintained in a desirable range, predefined by upper and/or lower limits. Each water quality attribute encompasses a number of water quality indicators. The overall acceptability of water quality for its intended use depends on the magnitude of these indicators and is often governed by regulations and guidelines. A water quality failure is defined as a violation of regulations (or guidelines or self imposed limits) of one or more water quality indicators (Sadiq *et al.* 2004).

Five mechanisms/pathways can lead to water quality failures namely intrusion, regrowth, breakthrough, internal corrosion / leaching and permeation. Of the five, four mechanisms (breakthrough is the exception), are directly affected by pipes, either through pipe material type, size, structural condition, hydraulic / operational conditions and/or inner surface degradation. Table 1 shows a relative frequency (1971-1998) of deficiencies attributed to the outbreaks of illness that were traced to the distribution network (Lindley 2001, Lindley and Buchberger 2002). Environmental conditions such as the quality of the raw water, temperature and soil conditions around pipes can also have a direct or indirect impact on fluctuations of water quality in distribution networks (Kirmeyer *et al.* 2001).

The deterioration of pipe structural integrity can have a multi-faceted impact on water quality, especially in the domain of contaminant intrusion. Frequent pipe breaks increase the possibility of intrusion through the compromised sections in several ways. During repairs, intrusion can occur if flushing and local disinfection procedures are not appropriately followed. Furthermore, pipes are de-pressurized in the vicinity of a break during repair. This low pressure increases the potential of contaminant intrusion through unprotected cross connections. If the pipe has holes then de-pressurization will increase the *likelihood* of contaminant intrusion, which can be especially detrimental if the surrounding soil is contaminated by leaky sewers nearby, chemical spills, herbicides, pesticides, etc.

Data fusion refers to the scientific aggregation of information available in the form of observations and/or measurements. In some cases, different data sets give complementary information on various aspects of an event. In these cases there is motivation to collect additional information to increase the accuracy of the prediction. Information can also be redundant if it deals with the same aspect of the problem. Redundancy can improve the reliability of the prediction as measurement(s) / observation(s) are confirmed by the redundant one. Complementary and redundant information in data sets are the basis of data fusion applications in water quality modelling.

Quantitative aggregation of ‘incomplete’, ‘uncertain’ and ‘imprecise’ (vague) information / data warrants the use of *soft computing* methods, which are tolerant to imprecision, uncertainties and partial truths (Zadeh 1984). The term *soft computing* comprises an array of heuristic techniques such as fuzzy logic, evidential reasoning, neural networks, and genetic algorithms, which essentially provide rational solutions for complex real-world problems (Bonissone 1997). A traditional *soft computing* method for data fusion is the Bayesian (subjectivist) probability approach, which cannot differentiate between *aleatory* and *epistemic* uncertainties and is unable to handle non-specific, ambiguous and conflicting information without making strong assumptions. These limitations can be addressed by the application of Dempster-Shafer (DS) theory or the evidence theory. Evidential reasoning based on DS theory is named after Dempster (1967) and Shafer (1976) and is a generalization of the Bayesian theory. The DS theory was found to be flexible enough to combine the rigor of probability theory with the flexibility of rule-based systems. The DS theory applications in civil and environmental engineering range from slope stability (Binaghi *et al.* 1998), environmental decision-making (Attoh-Okine and Gibbons 2001), seismic analysis (Alim 1988), failure detection (Tanaka and Klir 1999), construction management (Sönmez *et al.* 2002), water quality (Sadiq and Rodriguez 2005) to climate change (Luo and Caselton 1997). Many more applications of DS theory can be seen in the detailed bibliography provided by Sentz and Ferson (2002).

This paper presents an innovative approach that uses evidential reasoning (DS theory) to quantify the risk of contaminant intrusion at a given location in a water distribution network. However, the proposed approach can be extended to full-scale water distribution network, which will help to establish risk-contours of contaminant intrusion using GIS. The risk-contours may help utilities to identify sensitive locations in the water distribution networks and prioritize their rehabilitation and control strategies.

The remaining paper is organised as follows: next section provides an introduction to the mechanism of contaminant intrusion in distribution water mains. The background and formulation of evidence theory is presented in the following section. It is then followed by an application of evidential reasoning to contaminant intrusion in the distribution network. Finally, a summary section concludes the paper.

CONTAMINANT INTRUSION

Intrusion of contaminants (hitherto referred to simply as "intrusion") into the water distribution network can occur through pipes and storage tanks (animals, dust-carrying bacteria, infiltration). Intrusion through deteriorated water mains can occur during maintenance and repair events, through broken pipes and gaskets, and cross-connections. Kirmeyer *et al.* (2001) ranked pathogen (contaminant) entry routes into the distribution network based on responses from an expert panel, the members of which were instructed to identify and rank the importance of routes of entry. Results are summarised in Table 2, indicating that intrusion was rated mostly "high". In addition to pathogens, intrusion can also introduce into the pipe chemicals, such as pesticides, herbicides, hydrocarbons (gasoline spills) as well as physical contaminants, such as plant debris and soil particles. Intrusion into water mains requires the simultaneous presence of three elements, a pathway, driving force (negative pressure differential between the pipe and its environment) and a contamination source. Brief description of these elements is provided in following sub-sections.

Pipe breakage / repair and cross-connection – a pathway

A water distribution network can never be completely water tight due to the existence of pipe cracks, holes, faulty gaskets and/or faulty appurtenances, which can serve as intrusion pathways. The driving force required for intrusion is usually a pressure differential. Therefore, it is *improbable* that intrusion will occur as long as the water pressure inside the network is greater than the pressure outside although movement of microbial or viral contaminants against the pressure gradient is *possible*. Pressure differential can occur during maintenance activities, such as during break repairs, flushing, etc., when parts of the distribution network are de-pressurized. Sources of contaminants include sewage water ex-filtrated from adjacent broken sewers, contaminated groundwater/soil and backflow through unprotected cross-connections.

Another direct pathway of intrusion is the actual exposure of a broken pipe to contaminated soil or water during repair. Contamination may occur if proper flushing and disinfection procedures are not implemented prior to re-commissioning. Clearly, the frequency of pipe breakage, the duration of repair jobs and the size of the network segment that can be isolated during maintenance are factors that have an impact on the risk of intrusion.

Cross-connection is a term used to describe a physical link through which it is possible for a non-potable liquid to enter into a potable water distribution network. Typically, when the pressure in the non-potable system is greater than that in the water distribution network, the existence of an unprotected cross-connection may result in the backflow of contaminants into the potable water supply system. One of the most severe cases of contamination attributed to cross-connection was recorded in Chicago in 1933 where 1,409 persons contracted *amoebic dysentery* of which 98 died (Anderson 1981). Since then, increased knowledge and awareness, extensive regulation and technical advances have reduced the risk of contamination through cross-connections but have not eliminated it completely.

Transient pressures – driving force

In addition to pressure differentials arising due to de-pressurisation of pipes, as discussed earlier, extreme transient pressures can also cause pressure differentials. Extreme transient pressures in a water supply system can occur as a result of power failure in a pumping station, fast closure of valves, fire flows, pipe rupture, etc. These transients can cause negative pressures in pipes, which sometimes may be exacerbated by peculiar topographical conditions. These negative pressures may provide a driving force for contaminants to intrude through compromised pipe walls and joint gaskets. Extreme transient pressures are more likely to occur in long transmission mains than in an urban distribution network in which users' faucets effectively serve as widely distributed pressure relief valves. An exception may be during fire flows or in the vicinity of a wet industrial facility. The volume of the inflow of the contaminated solute is typically quite small (less that 1% of the flow in the pipe) since the duration of transient pressures is quite short (Kirmeyer *et al.* 2001).

Sources of contamination

Contamination sources can be either chemical (pesticides, herbicides, petroleum products, fertilizers, solvents, detergents, pharmaceuticals, etc.) or microbiological (microbes, viruses, bacteria). Karim *et al.* (2003) reported concentrations of total coliform, fecal coliform, clostridium, bacillus, and viruses in soil and water samples taken around the water mains. Total and fecal coliforms were found in more than 60% and 40% of the samples, respectively. Bacillus was found in most of the samples as was expected because it is a natural soil organism. Enteroviruses, Norwalk and Hepatitis A viruses were also found around the pipe giving a strong indication of human and animal sources of contamination. Karim *et al.* (2003) also reported the range of concentrations for various organisms found in soil samples collected in the vicinity of the water mains.

Intrusion is a complex phenomenon, which depends on above three elements but generally the data on these elements are incomplete, non-specific and uncertain. Evidential reasoning provides a meaningful way to fuse / aggregate these data to make inferences on the risk of contaminant intrusion. The next section provides a background to DS theory and its possible application to assess the risk of contaminant intrusion in distribution networks.

EVIDENCE THEORY

Two major types of uncertainties, *aleatory* (natural heterogeneity and stochasticity) and *epistemic* (subjectivity, ignorance) are observed in natural systems. The traditional approach to handle *aleatory* uncertainty is through probabilistic analysis based on historical data (a *frequentist* approach). Traditionally, *epistemic* uncertainty has been addressed through Bayesian approach, however, the approach has limitations, since it requires a-*priori* assumptions (Sentz and Ferson 2002).

Consider a case of water quality deterioration in distribution network due to intrusion, where possible outcomes (*water quality condition states*) are *low*, *medium* and *high* denoted by {L}, {M}, and {H}, respectively. The traditional Bayesian approach treats these outcomes only as *disjoint* bodies of evidence, i.e., probabilities can be assigned to only *singletons* {L}, {M}, and {H}. Further, according to the basic axiom of probability, $p(L) + p(M) + p(H) = 1$. Consequently, the probability of the complement of {L}, i.e., $p(\neg L) = 1 - p(L) = p(M) + p(H)$. The inference about $p(\neg L)$ is based on a rather strong assumption, i.e., the *Principle of Insufficient Reason* (Sentz and Ferson 2002) that ignorance has to be distributed uniformly among the remaining singletons {M} and {H}. For example, if observation of water quality implies that $p(L) = 0.5$ and no further information is available, then $p(M) = p(H) = 0.25$ is assumed due to *Principle of Insufficient Reason.*

The DS theory can be interpreted as a generalization of the Bayesian approach, where probabilities are assigned to *subsets* and not only to mutually exclusive *singletons* (Sentz and Ferson 2002). For example, in the above case, in addition to singletons {L}, {M}, {H}, subsets of outcome (with less specificity) such as {L, M} (read: L or M), {M, H}, {L, H} and {L, M, H} are also considered as candidates for a *basic probability assignment* (*bpa*, this concept is detailed in the next section). The Bayesian approach could therefore be viewed as a special case of DS theory, where sufficient evidence exists to assign probability to singletons only (highly specific situations) and ignore less specific subsets. Therefore, the Bayesian approach is unable to differentiate both *aleatory* and *epistemic* uncertainties efficiently and cannot handle less specific and ambiguous evidences. The evidential reasoning (or theory of evidence) addresses these issues effectively. In the above example instead of assigning p(M) = p(H) = 0.25, the remaining probability will be acknowledged as ignorance (*epistemic* uncertainty) and will be dealt with using DS theory.

Basic concepts

In DS theory, the *frame of discernment* Θ is defined as a set of mutually exclusive alternatives, which allows the *power set* "*A*" to have a total of $2^{|\Theta|}$ subsets in the domain, where $|\Theta|$ is the cardinality of the *frame of discernment*. For example, if the *frame of discernment* Θ = {L, M, H}, its power set comprises 8 subsets (the cardinality is 3), due to *closed world* assumption over "union" (i.e., the possible outcomes are *exhaustive* and can not be outside the *frame of discernment*). This power set *A* contains the 8 subsets A_i (i = 1, 2, …, 8), i.e., ϕ (a null set), {L}, {M}, {H}, {L, M}, {M, H}, {L, H}, and {L, M, H}. Thus, depending on the evidence, probability mass (also referred to as *bpa*) can be assigned to *low*, *medium*, *high*, *low* or *medium*, *low* or *high, medium* or *high*, and *low* or *medium* or *high* (the last subset denotes a fully ignorant situation). Recall that this concept is different from the Bayesian approach in which there are only three possible outcomes on this *frame of discernment* Θ, which are {L}, {M} and {H}. Four important concepts, namely, *basic probability assignment* (*m* or *bpa*), *belief* (*bel*), *plausibility* (*pl*) and *pignistic probability* (*bet*) functions are used in DS theory.

The *basic probability assignment* (*bpa* or *m*) expresses the proportion of all available relevant evidence that supports the claim that a particular element of power set *A* belongs to the (sub)set A_i but to no particular subset of A_i (Klir 1995; Klir and Folger 1988). For a given $m(A_i)$, every subset A_i for which $m(A_i) \neq 0$ is called a *focal element*. The mass $m(A_i)$ is defined over the interval [0, 1], but it is different from the classical definition of probability. The *bpa* of the null subset $m(\phi)$ is zero and the sum of the *basic probability assignments* $m(A_i)$ in a given evidence set "<$m(A_i)$, A_i>" is "1". Thus,

$$m(A_i) \rightarrow [0,1]; \quad m(\phi) = 0; \quad \sum_{A_i \subseteq A} m(A_i) = 1 \tag{1}$$

Continuing on intrusion example described in the previous section, the *focal element* of a given evidence "<$m(A_i)$, A_i>" can be written as m(L) = 0.5, therefore it implies that m(Θ) = m(L, M, H) = 0.5. This is because {L, M, H} represents complete ignorance and the DS theory dictates that all missing evidence is always assigned to ignorance (as opposed to the Bayesian approach that distributes missing evidence over the remaining disjoint subsets).

The *basic probability assignment* is used to determine the two non-additive measures *belief* and *plausibility*, which represent the lower and upper bounds of a probability, respectively. The lower bound, *belief* (*bel*), for a set A_i is defined as the sum of all the *basic probability assignments* of the proper subsets A_k of the set of interest A_i, i.e., $A_k \subseteq A_i$. The general relation between *bpa* and *belief* can be written as

$$bel(A_i) = \sum_{A_k \subseteq A_i} m(A_k) \tag{2}$$

It can be shown that

$$bel(\phi) = 0; \quad bel(\Theta) = 1 \tag{3}$$

It should be noticed that *bel*(L, M) ≥ *bel*(L) + *bel*(M) because DS theory allows some mass to be assigned to less specific subset *m*(L, M), which was not permitted in Bayesian approach. Therefore, DS theory relaxes a strong *additivity* constraint of probability theory to more relaxed constraint of *monotonicity*.

The belief functions for above example are given by

bel(L) = *m*(L) = 0.5; *bel*(M) = *m*(M) = 0; *bel*(H) = *m*(H) = 0

bel(L, M) = *m*(L) + *m*(M) + *m*(L, M) = 0.5 + 0 + 0 = 0.5

bel(L, H) = 0.5; *bel*(M, H) = 0; *bel*(L, M, H) = *m*(L) + … + *m*(Θ) = 1

The upper probability bound, *plausibility*, is the summation of *basic probability assignment* of the sets A_k that intersect with the set of interest A_i, i.e., $A_k \cap A_i \neq \phi$, and therefore it can be written as

$$pl(A_i) = \sum_{A_k \cap A_i \neq \phi} m(A_k) \tag{4}$$

In addition, the following relationships for *belief* and *plausibility* functions hold true in all circumstances

$$pl(A_i) \geq bel(A_i); \quad pl(\phi) = 0; \quad pl(\Theta) = 1; \quad pl(\neg A_i) = 1 - bel(A_i) \tag{5}$$

In our example, the *plausibility functions* are given by

pl(L) = *m*(L) + *m*(L, M) + *m*(L, H) + *m*(Θ) = 1

pl(M) = 0.5; *pl*(H) = 0.5; *pl*(L, M) = 1;

pl(L, H) = 1; pl(M, H) = 0.5; $pl(\Theta)$ = 1

The *belief interval* (I) is an interval between *belief* and *plausibility* representing a range in which true probability may lie, therefore a narrow *belief interval* represents more precise probabilities, and it can be shown that the probability is uniquely determined if $bel(A_i) = pl(A_i)$ (note that probability theory is applicable only where all probabilities are unique and disjoint (Yager 1987)). If $I(A_i)$ has an interval [0, 1], it means that no information is available; conversely, if the interval is [1, 1], it means that A_i has been completely confirmed by $m(A_i)$.

The *belief intervals* for our example are

I(L) = [0.5, 1]; I(M) = [0, 0.5]; I(H) = [0, 0.5]

Beliefs manifest themselves at two levels - the *credal* level (from credibility) where *belief* is entertained, and the *pignistic* level where beliefs are used to make decisions. The term "pignistic" was proposed by Smets (2000) and originates from the word pignus, meaning '*bet*' in Latin. Pignistic probability is used for decision-making and uses *Principle of Insufficient Reason* to derive from *basic probability assignment*. It is a point (crisp) estimate in a *belief interval* and can be determined as

$$bet(A_i) = \sum_{A_i \subseteq A_k} \frac{m(A_k)}{|A_k|} \tag{6}$$

The denominator $|A_k|$ in the above equation represents the cardinality (number of elements) of the (sub)set A_k. The *pignistic probabilities* in our example are

$bet(\mathrm{L}) = m(\mathrm{L})/1 + m(\Theta)/3 = 0.5 + 0.17 \approx 0.67$

$bet(\mathrm{M}) = m(\mathrm{M})/1 + m(\Theta)/3 = 0 + 0.17 \approx 0.17$

$bet(\mathrm{H}) = m(\mathrm{H})/1 + m(\Theta)/3 = 0 + 0.17 \approx 0.17$

The sum of pignistic probabilities is always 1.

Dempster–Shafer (DS) rule of combination

The purpose of data fusion/ aggregation is to summarize and simplify information in a rational manner. The DS theory assumes that the sources of information are independent. Alim (1988) described that the 'combined' (or 'fused') belief not only represents the total belief of a set A_i and all of its subsets but also takes into account the contributions of different sources of evidence about A_i. The DS inference uses combination operators that compromise on precision but require less information than the Bayesian inference (Sentz and Ferson 2002).

The DS rule of combination strictly emphasizes agreement between multiple sources and ignores all conflicting evidence through *normalization*. A strict *conjunctive logic* through AND-type operator (product) is employed in the combination of evidence. So far, *bpa* was referred to as m for any body of evidence. Since two or more bodies of evidence are introduced in the subsequent discussion, the subscript j, i.e., m_j, is introduced in reference to body of evidence j. The DS rule of combination for j = 1, 2 determines the joint $m_{1\text{-}2}$ from the aggregation of two *basic probability assignments* m_1 and m_2 by (Klir and Folger 1988)

$$m_{1-2}(A_i) = \frac{\sum_{A_p \cap A_q = A_i} m_1(A_p) m_2(A_q)}{1-K} \qquad when \quad A_i \neq \phi; \qquad \text{and} \qquad m_{1\text{-}2}(\phi) = 0 \tag{7}$$

where $K = \sum_{A_p \cap A_q = \phi} m_1(A_p)\, m_2(A_q)$ is the *degree of conflict* in two sources of evidence and $m_1(A_p)$ and $m_2(A_q)$ are their corresponding masses. The denominator (1-K) is a normalization factor, which counterbalances the effect of conflicting evidence on aggregation. The above equations can be rewritten as

$$m_{1-2}(A_i) = \frac{\sum_{A_p \cap A_q = A_i} m_1(A_p) m_2(A_q)}{\sum_{A_p \cap A_q \neq \phi} m_1(A_p)\, m_2(A_q)} \tag{8}$$

Continuing the previous example, now assume another water quality observation is available with the following body of evidence $\langle m_2(A_q), A_q \rangle$, where

m_2(L) = 0.5; and m_2(M, H) = 0.5

Applying DS rule of combination as indicated by equation 7 (or 8), the final combined results are provided in Table 3. The *belief*, *plausibility* and *pignistic* probability are determined using equations 2, 4 and 6 respectively for the combined evidence.

Combining sources of varying credibility

Equations (7) and (8) above implicitly assume that all sources of information are equally credible, but this may not always be the case. For example, sampling locations for monitoring water quality may be representative of a particular part of the water distribution network, e.g., if one sample is collected from major distribution main and the other is collected from a minor main, the influence zones of the two samples are different. Similarly, if the samples are collected at the same point when two different flow conditions prevail, the evidence of water quality also needs to be adjusted based on the flow conditions. Also, if water utility staff with different levels of expertise collects water samples, the observations may need to be adjusted based on their experience.

Therefore, the bodies of evidence obtained from different sources of information need to be discounted using *credibility factor* (α) depending on its relative strength and/or reliability. The evidence can be discounted as

$$\left.\begin{aligned} m_{1-2}(A_i)_\alpha &= m_{1-2}(A_i)\cdot\alpha \\ m_{1-2}(\Theta)_\alpha &= m_{1-2}(\Theta)\cdot\alpha + (1-\alpha) \end{aligned}\right\} \tag{9}$$

The credibility factor is constrained by $0 \le \alpha \le 1$, where '0' represents 'fully incredible evidence', and '1' represents 'fully credible evidence'. The following section provides a simple hypothetical application of DS theory to determine the risk of contaminant intrusion in distribution network.

ESTIMATING RISK OF CONTAMINANT INTRUSION

The *frame of discernment* of risk of an intrusion can be described by a universal set Θ = {P, NP}, in which 'P' denotes 'possible' and NP denotes 'not-possible' intrusion. The power set of the risk of intrusion consists of two singletons {P} and, {NP}, a universal set {P, NP} and the empty set {ϕ}. As described earlier, the risk of intrusion of contaminants can be evaluated based on three bodies of evidence, a pathway (e_1), a driving force (e_2), and a contamination source (e_3).

In this example, the breakage rate (# of breaks/100 km/year) is taken as a surrogate measure for an intrusion pathway, transient pressure (psi) is taken as a surrogate for the existence of a driving force and the separation distance (meters) between a contaminant source and a water main as a surrogate measure for a source of contamination. We selected these surrogate measures due to simplicity in data collection. The *frames of discernment* for all three bodies of evidence are mapped to attain the *basic probability assignments* (i.e., m_1, m_2 and m_3), where each of them is assigned to the subsets {P}, {P, NP}, and {NP} of universal set Θ risk of intrusion. This multi-valued mapping is performed using a plot similar to fuzzy sets. But, the overlap of subsets does not refer to vagueness (for which fuzzy sets are used) rather this overlap is more conceptual and represents ambiguity (Beynon 2005). This multi-valued mapping makes sure that the sum of *bpa* (represented by *y*-axis) is "1" over the *frame of* *discernment*. It is not the basic requirement in case of fuzzy sets, which are represented by memberships on the *y*-axis.

Figure 1 refers to a scenario in which the following bodies of evidence are observed:

1) Pipe breakage rate is '10 breaks/100 km/year', from Figure 1a

   m_1 (P) = 0 m_1 (P, NP) = 0 m_1(NP) = 1,

2) The possibility of pressure drop to '0 psi' at the respective node, from Figure 1b

   m_2 (P) = 0 m_2 (P, NP) = 1 m_2 (NP) = 0, and

3) A leaky sewer is located at a distance of '3 m', from Figure 1c

   m_3 (P) = 1 m_3 (P, NP) = 0 m_3 (NP) = 0.

Suppose that three bodies of evidence m_1, m_2 and m_3 are assigned credibility factors (α) of 0.7, 0.9 and 0.6, respectively. Therefore, after *bpas*, are estimated (from mapping) they are adjusted by credibility factors (using equation 9). For example, the credibility factor for contaminant source α_3 is 0.6, therefore the adjusted basic probability assignment is $m_{\alpha 3}$ (P) = 0.6, $m_{\alpha 3}$ (P, NP) = 0.4, and $m_{\alpha 3}$ (NP) = 0.

Three bodies of evidence are combined using the DS rule of combination as described above. The *simplex plot* is used to illustrate three dimensions (subsets) of risk of intrusion, i.e., {P}, {P, NP}, and {NP} as shown in Figure 2. The *simplex plot* is an equilateral triangle in which, ant point is represented by three offsets measured from the axes opposite to vertices of a triangle. The sum of these perpendicular distances is always equal to 1 anywhere in the triangle. To see more details on simplex plots reader should refer to Marschak (1950), Walley (1991), Denoeux (2000) and Beynon (2005). For example, the vertex {P, NP} represents the ignorance; therefore any point closer to this vertex represents higher level of ignorance and ambiguity (because it has a maximum perpendicular distance from the opposite axis {NP}-{P, NP}). Points e_1, e_2 and e_3 represent three bodies of evidence which are fused together using the DS rule of combination to obtain combined evidence (e_c) of (0.31, 0.21, 0.48) as shown in Figure 2. The interpretation is that the *belief* of intrusion risk *bel*(P) is 31%, the corresponding *plausibility pl*(P) is 52%, and that there is 48% belief *bel*(NP) of no intrusion risk. The *belief interval* [0.31, 0.52] represents the lower and upper values of probabilities. The *belief interval* can be converted into a crisp or point estimate *bet*(P) (also termed 'expected utility') of risk of intrusion. The pignistic transformation of the imprecise probabilities yields *bet*(P) = 0.41.

Five additional scenarios for the example in Figure 1 are examined (their details are provided in the table at the bottom of Figure 2). The pressure and intrusion pathways measures in scenario 2 remain unchanged from those in scenario 1, but the contaminant source is assumed to be at a distance of 20 m rather than 3 m from the water main. The *belief* is now reduces to zero but the *plausibility* is 0.3, which is due to low pressure at that node. However, if transient pressure is increased to 50 psi for scenario 3 (a normal operating pressure), the *belief* remains zero but the *belief interval* drops to [0, 0.03]. In scenario 4, the breakage rate is increased to very high rate of 30 breaks/100 km/year, *belief* remains the same (at zero) but *plausibility* increases to 0.1, making it a more uncertain event. The pignistic probability increases to 0.17 from 0.05 for scenario 5, where the sewer is very close (at 3 m) to the water main, and the breakage rate is very high. Scenario 6 corresponds to an extreme case (negative pressures, contaminant source very near and very high breakage rate) in which all bodies of evidence hint to a 'certain' intrusion and the ignorance is almost negligible.

These scenarios illustrate how the risk of intrusion can vary with variations in three bodies of evidence considered here. The suggested approach enables the determination of risk of contaminant intrusion even when available information is incomplete, ambiguous or qualitative in nature. This approach can help utilities to establish risk-contours of contaminant intrusion for their distribution network using GIS. The evidential reasoning methodology can be implemented and be viewed as two layers in GIS. First layer may represent the risk of contaminant intrusion using *pignistic probabilities*, whereas the second layer may show the confidence over those values using *belief interval*. The pipe segments with higher *bet*(*P*) are points of concern, however, pipes with lower *bet*(*P*) but larger *belief intervals* are also of concern.

SUMMARY

The intrusion of contaminants into water distribution networks requires the simultaneous presence of three elements, a pathway, a driving force and a contamination source. Each of these elements provides an independent body of evidence of the risk of 'intrusion' in the distribution network.

In this paper, evidence theory was introduced as an innovative methodology that can be used to simplify and improve the understanding and interpretation of data generated through routine water quality monitoring in distribution networks. Evidential reasoning, also called Dempster-Shafer (DS) theory, has proved effective in dealing with this type of situation.

Bodies of evidence representing, intrusion pathway(s), driving force(s), and contamination source(s) are mapped over a *frame of discernment* of intrusion risk. Subsequently the DS rule of combination is applied to make an inference on the occurrence of intrusion. The implementation of this evidential reasoning method to assess risk of intrusion in distribution network is described with the help of a *simplex* plot where vertices of an equilateral triangle represent potential for intrusion, not-intrusion and ignorance. Six scenarios were generated to demonstrate the application of the proposed method under varying conditions.

The proposed method will help to quantify the risk of contaminant intrusion in a given pipe. However, the concept can be extended to water distribution network, which will help to establish risk-contours of contaminant intrusion. The risk-contours may help utilities to identify sensitive locations in the water distribution networks using GIS and prioritize control strategies.

ACKNOWLEDGEMENTS

This paper presents the results of a preliminary investigation of an on-going research project, which is co-sponsored by the American Water Works Association Research Foundation (AwwaRF) and National Research Council (NRC) of Canada.

REFERENCES

Alim, S. 1988. Application of Dempster-Shafer theory for interpretation of seismic parameters, *ASCE Journal of Structural Engineering*, 114(9): 2070-2084

Anderson, D.C. 1981. Cross-connections - their importance and control, Proceedings 1981 *AWWA WQTC*, Annual Conference

Attoh-Okine, N.O., and Gibbons, J. 2001. Use of belief function in brownfield infrastructure redevelopment decision making, *ASCE Journal of Urban Planning and Development*, 127(3): 126-143

Beynon M.J. 2005. A novel technique of object ranking and classification under ignorance: An application to the corporate failure risk problem, *European Journal of Operational Research,* 167(2): 493-517

Binaghi, E. Luzi, L., Madella, P., Pergalani, F., and Rampini, A. 1998. Slope instability zonation: a comparison between certainty factor and fuzzy Dempster–Shafer approaches, *Natural Hazards*, 17: 77–97

Bonissone, P.P. 1997. Soft computing: the convergence of emerging reasoning technologies. *Soft Computing*, 1: 6-18

Dempster, A. 1967. Upper and lower probabilities induced by a multi-valued mapping, *The Annals of Statistics*, 28: 325-339

Denoeux, T. 2000. Modeling vague beliefs using fuzzy belief structures, *Fuzzy Sets and Systems* 116: 167–199

Karim, M., Abbaszadegan, M., and LeChevallier, M.W. 2003. Potential for pathogen intrusion, *Journal of the American Water Works Association*, 95(5): 134-146

Kirmeyer, G.J., Friedman, M., Martel, K., and Howie, D. 2001. *Pathogen Intrusion into Distribution System*, AwwaRF, Denver, CO, USA

Klir, G.J. 1995. Principles of uncertainty: what are they? why do we need them?, *Fuzzy Sets and Systems*, 74: 15-31

Klir, G.J., and Folger, T. 1988. *Fuzzy sets, uncertainty,* and *information,* Englewood Cliffs, NJ, Prentice-Hall

Lindley, T.R. 2001. A framework to protect water distribution systems against potential intrusions, *MS thesis*, University of Cincinnati, Cincinnati

Lindley, T.R., and Buchberger, S.G. 2002. Assessing intrusion susceptibility in distribution systems, *Journal of the American Water Works Association*, 94(6): 66-79

Luo, W.B., and Caselton, B. 1997. Using Dempster-Shafer theory to represent climate change uncertainties, *Journal of Environmental Management*, 49(1): 73-93

Marschak, J., 1950. Rational behaviour, uncertain prospects, and measurable utility. *Econometrica*, 18(2): 111–141

Sadiq, R., and Rodriguez, M.J. 2005. Predicting water quality in the distribution system using evidential theory, *Chemosphere*, 59(2): 177-188.

Sadiq, R., Kleiner, Y. and Rajani, B. 2004. Aggregative risk analysis for water quality failure in distribution networks, *Journal of Water Supply: Research & Technology – AQUA*, 53(4): 241-261

Sentz, K. and Ferson, S. 2002. Combination of evidence in Dempster-Shafer theory, SAND 2002-0835

Shafer, G. 1976. *A mathematical theory of evidence*, Princeton University Press, Princeton, N.J., USA

Smets, Ph. 2000. Data Fusion in the transferable Belief Model, Proceedings of 3rd International Conference on Information Fusion, *Fusion 2000*, pp. PS21-PS33, Paris, France, July 10-13

Sönmez, M., Holt, G.D., Yang, J.B. and Graham, G. 2002. Applying evidential reasoning to prequalifying construction contractors, *ASCE Journal of Management in Engineering*, 18(3): 111-119

Tanaka, K. and Klir, G.J. 1999. Design condition for incorporating human judgement into monitoring systems, *Reliability Engineering and System Safety*, 65: 251-258

Walley, P. 1991. *Statistical reasoning with imprecise probabilities*, Chapman & Hall, London

Yager, R.R. 1987. On the Dempster-Shafer framework and new combination rules, *Information Sciences*, 41: 93-137

Zadeh, L.A. 1984. Review of books: A mathematical theory of evidence, *The AI Magazine*, 5(3): 81-83

Table 1. Deficiencies in distribution systems resulting in documented outbreaks of waterborne illness in USA from 1971-1998 (Lindley 2001)

Cited deficiency causing illness outbreak	# of events	% of total	Intrusion pathway	Adverse pressure	Contaminant source
Cross-connection and back siphonage	60	53.1	✔	✔	✔
Inadequate separation of water main and sewer	1	0.9	✔	✔	✔
Broken and leaky water mains	10	8.8	✔	✔	✔
Contamination in storage	15	13.3	✔		✔
Contamination during construction/repair	6	5.3	✔		✔
Contamination of household plumbing	8	7.1			✔
Metal corrosion and metal leaching	13	11.5			✔
Total	113	100			

Table 2. Microbial risk in the water distribution system - routes of entries (modified after Kirmeyer *et al.* 2001)

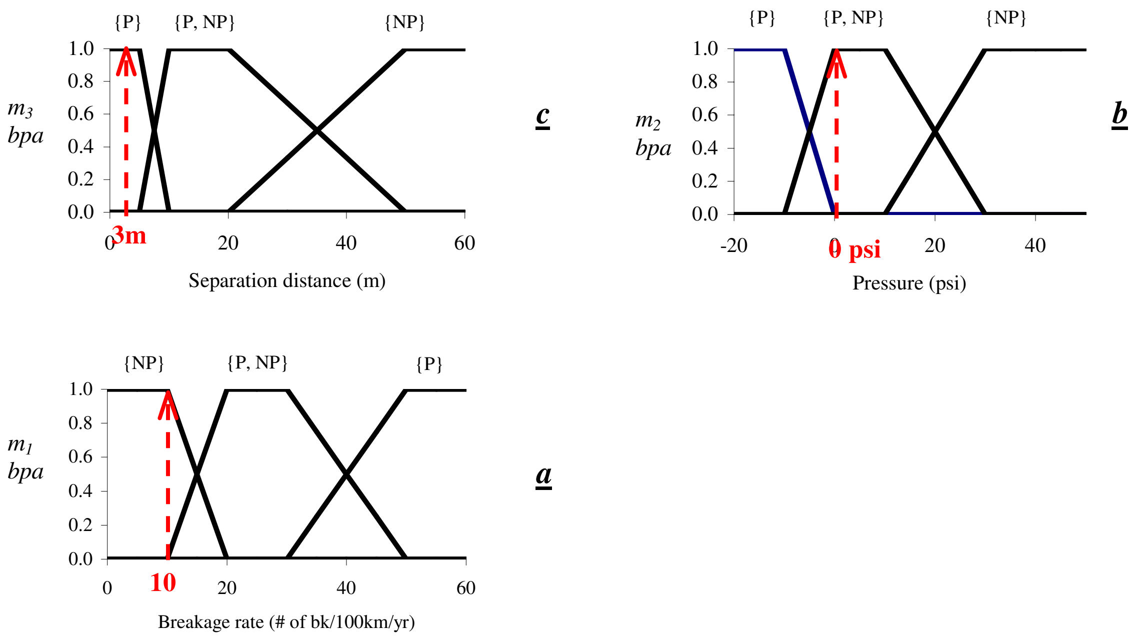
Route of entry	Priority/risk level
Water treatment breakthrough	High
Transitory contamination (intrusion)	High
Cross connection (intrusion)	High
Water main repair/break (intrusion)	High
Uncovered storage facilities (intrusion)	Medium-High
New main installations (intrusion)	Medium
Covered storage facilities (intrusion)	Medium
Growth/re-suspension	Low
*Purposeful contamination (intentional intrusion)	No

*After the recent terrorist activities, the purposeful contamination might be rated as a higher-level risk. Recently, AwwaRF has initiated a research project entitled "*Vulnerable points in the water distribution systems*" that addresses this issue.

Table 3. Results of DS rule of combination for two bodies of evidence

A_i	$m_1(A_p)$*	$m_2(A_q)$**	$m_{1\text{-}2}(A_i)^{\clubsuit}$	$bel_{1\text{-}2}(A_i)$	$pl_{1\text{-}2}(A_i)$	$bet_{1\text{-}2}(A_i)$
{L}	0.5	0.5	0.67	0.67	0.67	≈0.67
{M}	0	0	0	0	0.33	≈0.17
{H}	0	0	0	0	0.33	≈0.17
{L, M}	0	0	0	0.67	1	
{M, H}	0	0.5	0.33	0.33	0.33	
{L, H}	0	0	0	0.67	1	
Θ	0.5	0	0	1	1	
			Sum of {L}, {M} and {H} =	0.67 < 1	1.33 > 1	= 1

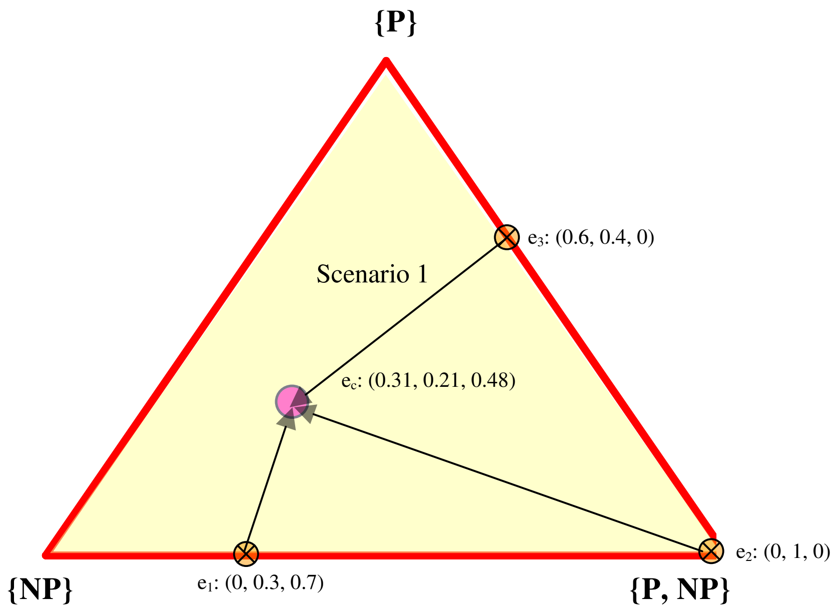
* First body of evidence
** Second body of evidence
♣ K = 0.25 (degree of conflict is 0.25) and the normalization factor is 0.75.



Evidence	Value	{P}	{P, NP}	{NP}	α
Pathway, e_1 (# breaks/100 km/year)	10§	m_1 (P) = 0	m_1 ({P, NP}) = 0	m_1(NP) = 1	$\alpha_1 = 0.7$
Pressure, e_2 (psi)	0	m_2 (P) = 0	m_2 ({P, NP}) = 1	m_2 (NP) = 0	$\alpha_2 = 0.9$
Contaminant source, e_3 (m)	3	m_3 (P) = 1	m_3 ({P, NP}) = 0	m_3 (NP) = 0	$\alpha_3 = 0.6$

§Statistical information obtained based on pipe diameter, age, material and surrounding soil conditions

Figure 1. Estimating the risk of contaminant intrusion using evidential reasoning



Scenario	Breakage rate e_1 (# breaks/100 km/year)	Pressure e_2 (psi)	Separation distance e_3 (m)	e_c ({P}, {P, NP}, {NP})	*bl*(P)	*pl*(P)	*bet*(P)
1	10	0	3	(0.31, 0.21, 0.48)	0.31	0.52	0.41
2	10	0	20	(0, 0.3, 0.7)	0	0.3	0.15
3	10	50	20	(0, 0.03, 0.97)	0	0.03	0.02
4	30	50	20	(0, 0.1, 0.9)	0	0.1	0.05
5	30	50	3	(0.13, 0.09, 0.78)	0.13	0.22	0.17
6	30	-20	3	(0.96, 0.04, 0)	0.96	1	0.98

Figure 2. *Simplex* plot representing individual and combined bodies of evidence