Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 70 (2014) 1688 – 1696

Procedia Engineering

www.elsevier.com/locate/procedia

12th International Conference on Computing and Control for the Water Industry, CCWI2013

Probabilistic backtracing of drinking water contamination events in a stochastic world

P. van Thienen[a,*], D. Vries[a,b], B. de Graaf[c], M. van de Roer[d], P. Schaap[e], E. Zaadstra[f]

[a] *KWR Watercycle Research Institute, Nieuwegein, The Netherlands*
[b] *Wetsus, Centre of Excellence for Sustainable Water Technology, Leeuwarden, The Netherlands*
[c] *Vitens, Zwolle, The Netherlands*
[d] *Dunea, Zoetermeer, The Netherlands*
[e] *PWN, Heemskerk, The Netherlands*
[f] *Brabant Water, 's Hertogenbosch, The Netherlands*

Abstract

In this paper, we investigate the relevance of the stochastic nature of water demand for backtracing of contaminations in drinking water distribution networks. We present an approach to deal with the uncertainty introduced by stochastic demand, which is applied to a full detail part (all pipes) of a hydraulic model of a distribution network in the Netherlands. It is demonstrated that stochastic water demand can introduce significant amounts of uncertainty for backtracing in some parts of tertiary (reticulation) networks in specific, looped configurations. In other parts, the additional uncertainty introduced by stochastic water demand can be limited.

Selection and peer-review under responsibility of the CCWI2013 Committee

Keywords: Backtracing; backtracking; contamination; probabilities; stochastic flow

1. Introduction

Along with the increased awareness of the vulnerability of drinking water distribution systems for contamination with a chemical or biological agent in the last decade, a range of approaches for their early detection and impact mitigation were developed (e.g. Berry et al., 2010). One important question to be answered when a utility is faced with a contaminant in its drinking water is what its source (location) is. Backtracing (or backtracking) methods have been developed to answer this question. These methods allow the user to track the transport of water and any substance contained therein in a hydraulic model in reverse time, i.e. back into the past towards the production location. Any substance found in the water which was not present at the production site must have been introduced into the water somewhere along the path(s) the water has followed from production location to sampling location. Knowledge of the contamination source helps to mitigate its effects by allowing effective isolation and accurate information to

* Corresponding author. Tel.: +31-30-6069602
E-mail address: peter.vanthienen@kwrwater.nl

1877-7058
Selection and peer-review under responsibility of the CCWI2013 Committee
doi:10.1016/j.proeng.2014.02.186

consumers. Several methodologies have been presented in the literature to perform backtracing calculations. The most straightforward, and widely applied, method implements water parcel backtracing in a Lagrangian framework (e.g. Shang et al., 2002), but only gives the paths which were followed by a parcel of water. More elaborate approaches allow, to some degree, to narrow down the location of origin. These include model optimization approaches (Laird et al., 2005, 2006, De Sanctis et al., 2010), a model tree approach (Preis and Ostfeld, 2006), and probabilistic inversion (Propato et al., 2010).

Central in any backtracing exercise is knowledge or a model of the (time varying) flow field in a distribution network. The accuracy with which this flow field is known for each location and each moment in time is generally an inverse function of the scale or detail level one is interested in. Because of the stochastic nature of water demand, a significant amount of uncertainty is introduced into the backtracing calculation from resulting uncertainties in the state and variation of the local flow field in the meshed parts of the tertiary distribution network (reticulation system, neighborhood level). Blokker (2011) compared the flow fields resulting from uniform domestic demand patterns with those resulting from stochastic domestic demand patterns and found that a significant part of the tertiary (reticulation) network studied shows regular flow reversals and/or stagnation which were not observed in the former, widely applied, approach. So far, this stochastic variability of the flow field has not been considered in backtracing studies.

The aim of this paper is to develop an approach to backtracing in a network with realistic, time-varying, stochastic demand and investigate the relevance of stochastic water demand for backtracing calculations. Our approach is based on a suite of stochastic network velocity models designed to accurately represent the variability of the flow field in the network throughout the day and between the days based on demand patterns generated by the stochastic end user demand simulator SIMDEUM (Blokker et al., 2010, Blokker, 2011). In essence, it is an expanded variation on the classic Lagrangian backtracing method.

The developed approach is applied to the drinking water distribution system of a Dutch city.

2. Methods

2.1. Flow field generation

SIMDEUM (Blokker et al., 2010, Blokker, 2011) is a water demand simulator that simulates individual types of water use and water using appliances for individual households. The amounts of water use and their timings and durations are sampled from statistical distributions which are based on time use surveys and technical data on appliances. Combined with demographic information, SIMDEUM generates realistic stochastic demand patterns for individual households or any number of households combined. Fig. 1 shows the combined demand pattern for a number of households ranging from 100 to 10,000 (flows scaled to same magnitude by dividing flow rates by the connection number ratio). The average demand and a two standard deviation margin are shown for an ensemble of 10 demand realisations (stochastic demand patterns for 10 days) for 5 minute time blocks. It is clear from this figure that when a sufficiently large number of connections is combined, stochastic variations average out, becoming smaller and smaller relative to the average. This effect is illustrated in Fig. 2, which shows the all day mean of the standard deviation of the demand patterns as a function of the number of demand pattern realisations and the number of connections. Fig. 2a demonstrates that with 5-10 stochastic realisations, a significant part of the stochastic variability is described. Fig. 2b shows an exponential decay of the varibility for increasing number of connections. However, (parts of) network models which have not been skeletonized may have small numbers of connections per network node. In these cases, individual demand patterns generated with SIMDEUM may differ significantly.

Hydraulic simulations (72 hours) have been performed for a network model (described below) for 5 different sets of stochastic demand patterns. Each of these five runs for a network model results in a time varying flow field, which we call a realisation of the flow field. Together, these 5 realisations are assumed to be representative of the stochastic variation of the flow field, assigning equal likelihood to each of the realisations. As is shown in Fig. 2a, 5 realisations are sufficient to capture most of the variability of the flow field in terms of flow magnitude. However, it is likely that relevant realisations of the flow field are missed when only 5 sets of samples from the demand distributions are taken. This should be kept in mind.

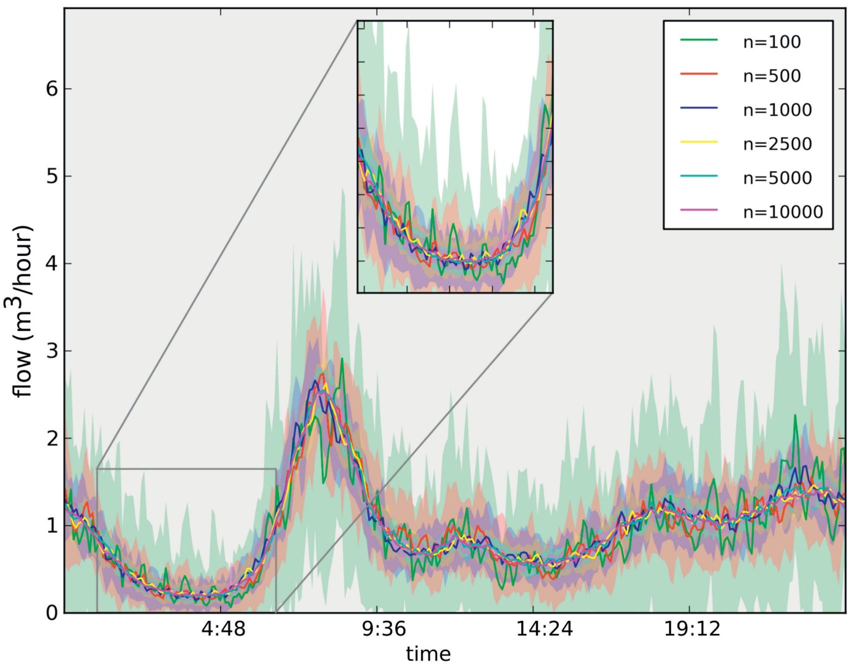


Fig. 1. Average demand patterns with 2 standard deviation margins for varying numbers of connections n (total demand scaled to be the same for all). Means and standard deviation values are computed from a set of 10 stochastic pattern combinations for 5 minute time intervals.

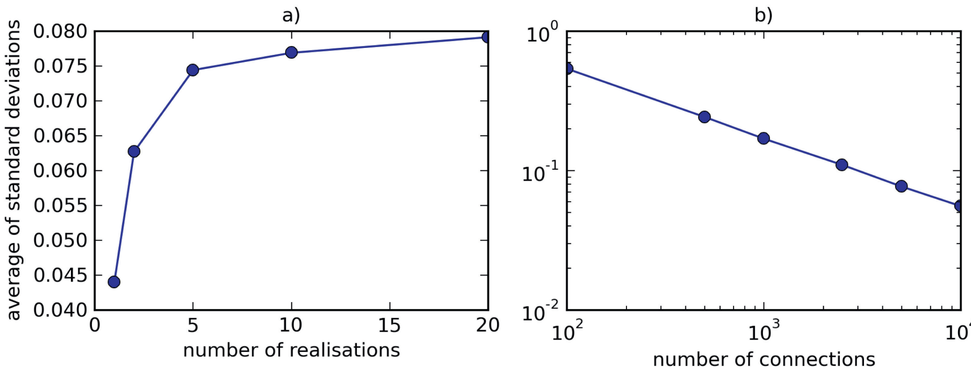


Fig. 2. Mean value of standard deviations (in m^3/minute) representing demand pattern variability as a function of the number of stochastic realisations (a, for 5000 connections) and the number of household connections (b, for 10 realisations, scaled to same total demand). The all day mean demand is about 1.25 m^3/minute for these patterns.

2.2. Backtracing by combination of Lagrangian forward traces

In contrast to the widely applied Lagrangian backtracing approaches discussed in the introduction, our method does not actually perform backtracing as such, but is based on forward tracing instead, using EPANET-MSX (Shang et al., 2008). For a single realisation of the flow field, at each node at each instant of (model) time, a tracer substance is released which is specific for that node, such that a fixed concentration at that node is obtained. All these tracer substances are transported through the network, with a simulation time longer than the longest residence time in the network. This requires some computation time and significant amounts of storage (custom software has been written to combine and compress EPANET-MSX output files), but when these calculations are done, backtracing is a simple exercise of retrieving data from the results files: for a network node *i*, the complete path of origin is contained in the local concentrations of all the tracer substances. Also, the full dynamic history of the flow field is included in the backtrace. All node specific tracer substances for origin nodes *j* which have a non-zero concentration indicate that the water being sampled at node *i* has passed these nodes *j*. The relative proportions of the concentrations correspond to the relative proportions of the water flow which passed nodes *j* to the flow of water passing node *i*.

This approach is expanded to a stochastic framework simply by repeating the analysis for a number of realisations of the flow field and combining the resulting traces. When a sufficiently high number of traces is combined, the ratio of the number of traces in which a node *j* is part of the traces to the total number of traces represents the likelihood that node *j* is part of the actual flow path.

2.3. Network model

A full 1:1 network model (all pipes) for a Dutch city of about 200,000 inhabitants was provided by its water company. The model was partially skeletonized: a single neighborhood in the network was left in full detail, the rest of the network was skeletonized (Fig. 3). Stochastic demand was assigned to the full detail neighborhood, whereas standard top-down patterns were applied to the skeletonized parts of the network.

3. Results

3.1. Backtraces

Fig. 4 shows flow traces for two selected sample nodes for selected individual flow field realisations (a,c and b,e, respectively), as well as origin probability maps combining five realisations. The maps for individual traces show

(a) (b)

Fig. 3. Network model: a) full model, for a large part skeletonized; b) full detail neighborbood.

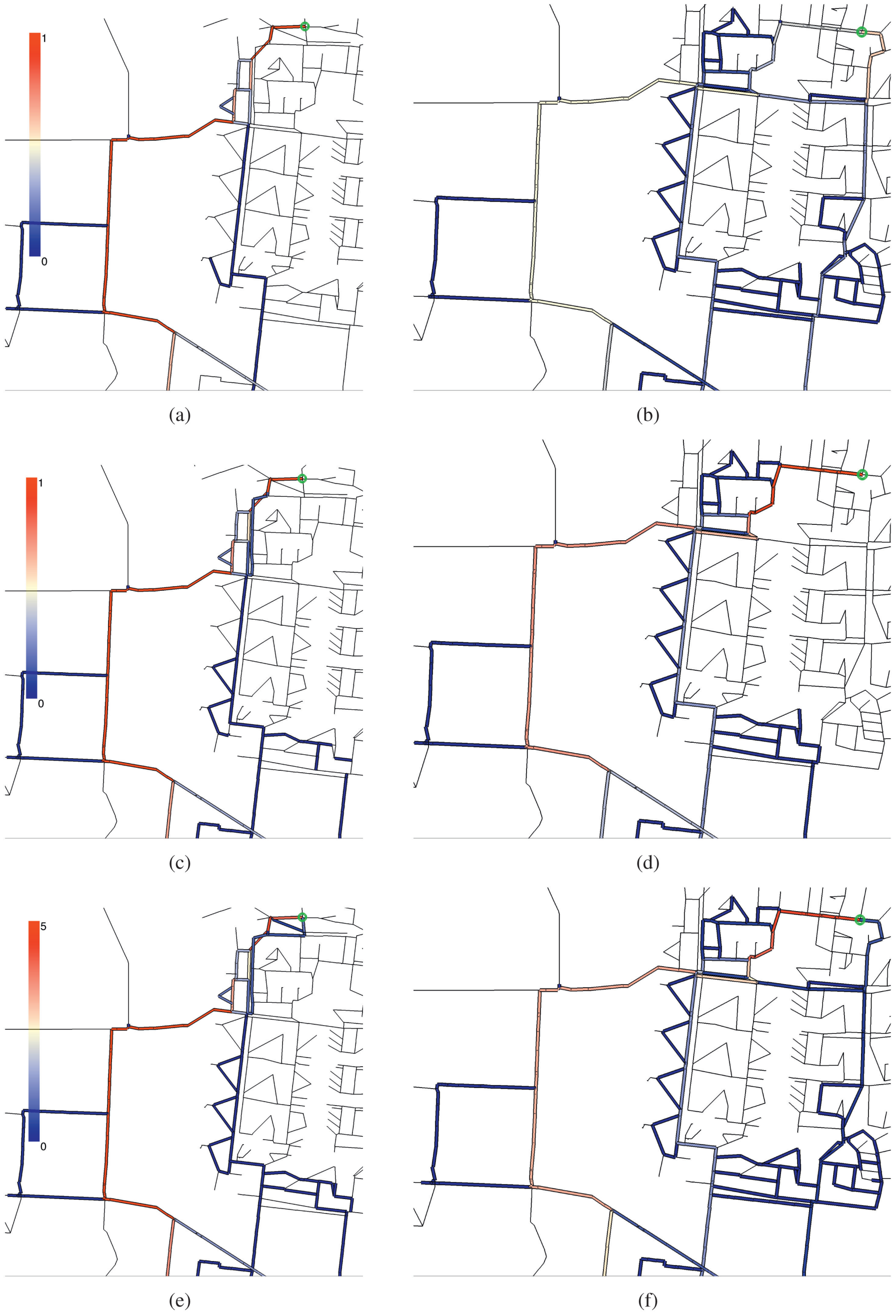


Fig. 4. Origin maps for water sampled at two marked locations (a,c,f and b,d,f) at time t=72 hours for two different realisations of the flow field (a,c and b,d, respectively) with combined maps for 5 realisations (e,f).

which fraction of the water being ”sampled” at the sample node has passed any section of the network, ranging from small fractions (blue) to large fractions (red). The combined map shows a simple stacking of these individual traces. When one assumes that all realisations considered are equally likely, the combined map represents relative likelihoods of a contamination being picked up at the sample point originating from the corresponding locations in the network. For the first of the sample nodes (Fig. 4a,c,e), there is relatively little (and exclusively low amplitude) variation between the flow field realisations and the combined map resembles the individual traces. For the second sample node (Fig. 4b,d,f), however, the main trace of origin may differ between realisations. Nevertheless, the combined map shows a strong preference for one of these paths, suggesting that the alternate path is only important in one of the five realisations.

3.2. Statistics

A slightly different approach to evaluating the effect of stochastic demand on traces is by considering the amount of overlap between traces for individual realisations of the flow field. Fig. 5a shows a selection of 12 nodes in the network model, for which backtraces have been computed for all 5 realisations of the flow field. The amount of overlap between the traces for these five realisations is visualized in Fig. 5c by means of pie charts. These pie charts show which fraction (in terms of numbers of network nodes, see Fig. 5b) of the combined trace is covered by one, two, three, four or five realisations. We only consider those parts of the traces which are inside the detailed part of the network model (see Fig. 5b). Note that this approach considers the backtraces in a binary way: a node is either part of the backtrace or it is not. Largely or completely blue (n=5) pie charts indicate that for these nodes, the backtrace is the same for all realisations of the flow field. Nodes for which the blue (n=5) fraction is smaller, however, show a considerable variability in the backtrace for this node for different realisations of the flow field. Backtracing from these nodes is therefore quite sensitive to the stochastic nature of water demand. In Fig. 5c, these sensitive nodes are generally found in the perifery of the tertiary (reticulation) network.

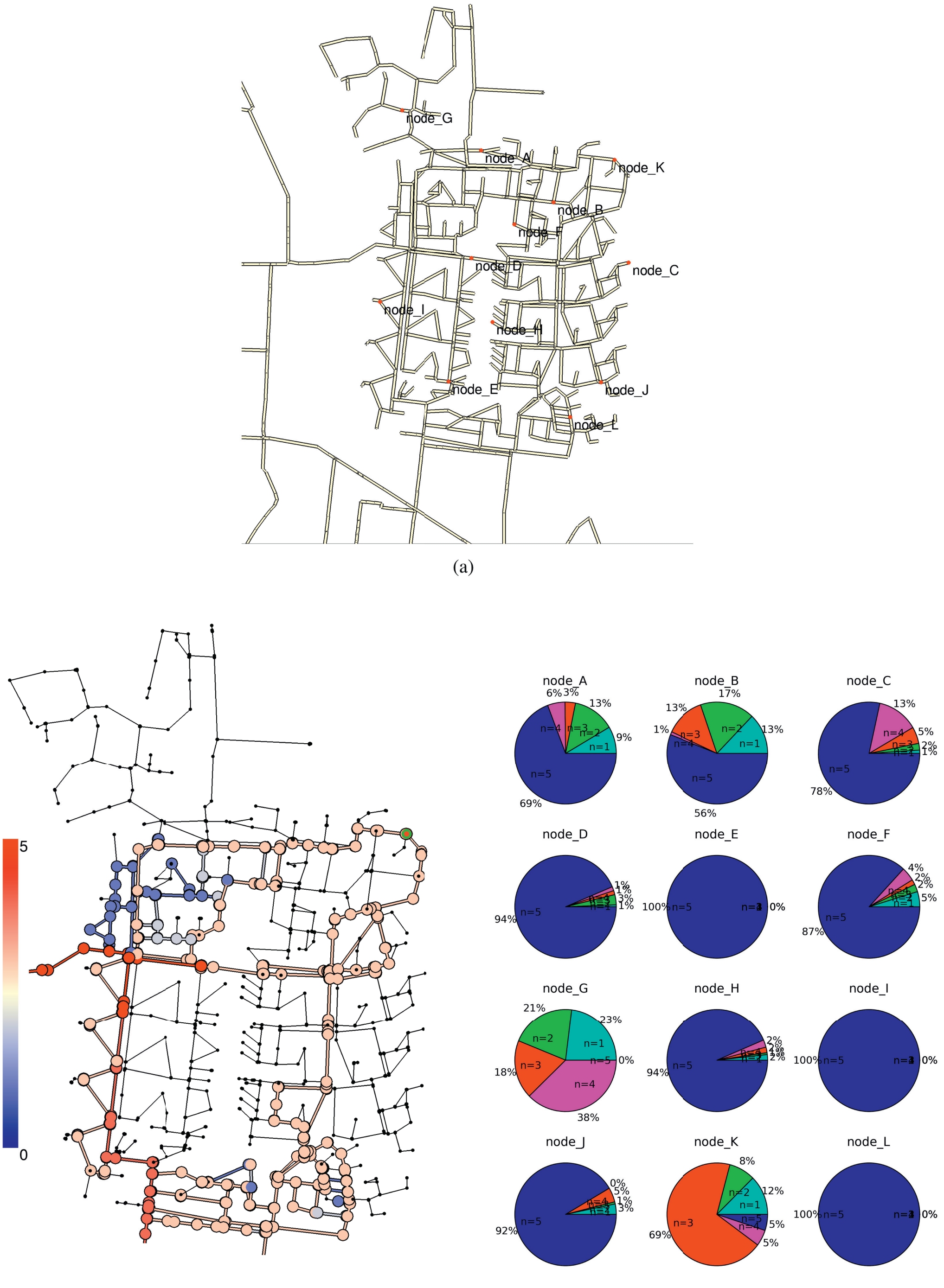
4. Discussion

Even though this paper only presents a number of example traces and is by no means an exhaustive analysis of the network model being studied, we have demonstrated that in some parts of the tertiaty (reticulation) network, the stochastic nature of water demand may result in water taking different routes at the same hour of different days. Therefore, backtracing calculations in full detail hydraulic models of the perifery of the distribution network should consider the stochastic nature of water demand. When doing so, the backtracing problem is no longer deterministic, but rather a probabilistic one with no single answer. If water quality sensors are placed in a network with the (secondary) aim of facilitating contamination source location, these nodes with varying source routes are the less suitable choice.

It must be kept in mind that a primary prerequisite for any kind of backtracing exercise is a network model which is accurate and up to date. For example, the status of valves in the distribution network may be changed in the course of time, e.g. during maintenance, without the changes being registered in the central data system of the water company. Any water company that wants to perform accurate backtracing calculations needs to ensure that network configuration information, including valve statuses, and customer/demand information is accurate and up to date. Even though this point is obvious, we illustrate the potential effect of a single unknown closed valve for backtracing in Fig. 6, because in practice, water companies often do not have accurate and up to date information about their valves.

5. Conclusions

The stochastic nature of water demand locally affects flow patterns in the perifery of the reticulation network. Therefore, backtracing calculations in these parts of the network require consideration of stochastic demand. However, an accurate and up to date network model, including correct valve statuses, is a prerequisite.



(a)

(b)

(c)

Fig. 5. Path node distribution for backtraces in 5 realisations of the flow field for twelve selected sampling nodes. a) Sampling node locations; b) stacked origin maps for node_K; c) pie charts for stacked origin maps. Pie charts indicate which fraction of all nodes encountered in traces for the 5 realisations is found in which number of traces. Note that only the parts of the traces which are within the detailed part of the model (b) are considered.

Fig. 6. The closing of a single valve (a) leads to a complete change in the origin map (b to c) for a node which is close to this valve.

References

Berry, J., Boman, E., Riesen, L., Hart, W.E., Phillips, C.A., Watson, J., Murray, R., 2010. Users Manual TEVA-SPOT Toolkit Version 2.4. Technical Report. Office of Research and Development, U.S. Environmental Protection Agency. EPA 600/R-08/041B.

Blokker, E., Vreeburg, J., van Dijk, J., 2010. Simulating residential water demand with a stochastic end-use model. Journal of Water Resources Planning and Management 136, 19–26.

Blokker, M., 2011. Stochastic water demand modelling: Hydraulics in water distribution networks. IWA publishing, London.

De Sanctis, A.E., Shang, F., Uber, J.G., 2010. Real-time identification of possible contamination sources using network backtracking methods. Journal of Water Resources Planning and Management 136, 444–453.

Laird, C., Biegler, L., van Bloemen Waanders, B., Bartlett, R., 2005. Contamination source determination for water networks. Journal of Water Resources Planning and Management 131, 125–134.

Laird, C., Biegler, L., van Bloemen Waanders, B., 2006. Mixed-integer approach for obtaining unique solutions in source inversion of water networks. Journal of Water Resources Planning and Management 132, 242–251.

Preis, A., Ostfeld, A., 2006. Contamination source identification in water systems: A hybrid model trees-linear programming scheme. Journal of Water Resources Planning and Management 132, 263–273.

Propato, M., Sarrazy, F., Tryby, M., 2010. Linear algebra and minimum relative entropy to investigate contamination events in drinking water systems. Journal of Water Resources Planning and Management 136, 483–492.

Shang, F., Uber, J.G., Polycarpou, M.M., 2002. Particle backtracking algorithm for water distribution system analysis. Journal of Environmental Engineering 128, 441–450.

Shang, F., Uber, J., Rossman, L., 2008. EPANET Multi-Species Extension Software. Technical Report. U.S. Environmental Protection Agency. Washington, DC. EPA/600/C-10/002.