Drinking Water
Engineering and Science

Open Access

Technical note: Efficient online source identification algorithm for integration within a contamination event management system

Jochen Deuerlein, Lea Meyer-Harries, and Nicolai Guth

3S Consult GmbH, 76137 Karlsruhe, Germany

Correspondence to: Jochen Deuerlein (deuerlein@3sconsult.de)

Abstract. Drinking water distribution networks are part of critical infrastructures and are exposed to a number of different risks. One of them is the risk of unintended or deliberate contamination of the drinking water within the pipe network. Over the past decade research has focused on the development of new sensors that are able to detect malicious substances in the network and early warning systems for contamination. In addition to the optimal placement of sensors, the automatic identification of the source of a contamination is an important component of an early warning and event management system for security enhancement of water supply networks. Many publications deal with the algorithmic development; however, only little information exists about the integration within a comprehensive real-time event detection and management system. In the following the analytical solution and the software implementation of a real-time source identification module and its integration within a web-based event management system are described. The development was part of the SAFEWATER project, which was funded under FP 7 of the European Commission.

1 Introduction

Drinking water distribution networks (WDNs) are an integral part of the technical infrastructure. Their task is to deliver a sufficient amount of drinking water of perfect quality at any place and any time. Since WDNs are exposed to a number of different risks, they belong to the critical infrastructures. One important risk is the (deliberate) contamination of drinking water with harmful substances. An adequate risk management system is required to manage the different threats in case of an event. Such a system normally consists of components for prevention, detection and response. Prevention measures include for example the protection of facilities such as water treatment plants, storage tanks and pumping stations or network sectorization for isolation of contaminants (Di Nardo et al., 2014). Among others, online monitoring of water quality within the distribution system is an essential part of a contamination warning or early detection system (Hall et al., 2007). As important, in case of a contamination event, is a well-prepared and efficient response

plan. For more than 1 decade a number of researchers have been working on methods for civil protection, real-time detection of contaminations and specific sensor development. Different software tools have been developed tackling problems such as optimal placement of sensors in the system (e.g. TEVA-SPOT, Sandia National Laboratories, 2017b; Chang et al., 2011) and detection algorithms (e.g. CANARY, Sandia National Laboratories, 2017a). Even though sensor-based online detection systems are capable of raising an alarm almost in real time if the quality of the water that reached the sensor deteriorated, the location of the source of this deterioration, in general, remains unclear. Since the efficiency of response actions is strongly dependent on the knowledge of the location of the source and the time when the contamination started, a tool is required that calculates the possible location of the contamination source based on the sensor information (SI: source identification). Then, in combination with the knowledge of the flow velocities in the pipes, the current spread of contamination and therewith the affected population can be estimated.

For solution of the source identification problem there are several categories of formulations and solving methods. Traditionally, SI was treated as an inverse parameter estimation problem. The parameters are the node time pairs of possible contaminations. Laird et al. (2005) formulated the SI problem as a non-linear, infinite-dimensional optimization problem subject to algebraic, ordinary differential, and partial differential constraints. Input parameters are the flow profiles calculated by hydraulic simulation and measured concentrations at sensor locations. As output, the time-dependent concentration along pipes and at junctions and the mass input at junctions as a function of time are sought. The method was tested for a network with 469 nodes. Even for this small model the number of unknowns in the non-linear programming (NLP) problem already reached 210 000. As a consequence, the method cannot be applied to real-time calculation of the large-size network models. Other authors use stochastic optimization methods for solution of the parameter estimation problem. Preis and Ostfeld (2008) combine the EPANET hydraulic and water quality simulation software with a genetic algorithm (GA). The objective function consists of a least squares function of measured and calculated concentrations. For the measurements, imperfect sensors are also taken into account. The calculation time for a test network with less than 1000 nodes was about 1 h. Liu et al. (2011) use an evolutionary algorithm (EA) for adaptive dynamic optimization (based on updated observations) and continually search for optimal solutions of a modified least squares function. Other authors (e.g. Propato et al., 2007) propose applying an input–output model for contaminant source identification. The model consists of a linear relationship between the possible source concentrations and the concentration measured at the sensors. The equation uses a linear transport matrix that is, in general, underdetermined, leading to non-unique solutions. Since for real-time application the execution time of the method is the critical performance indicator, optimization-based methods are not suited. Therefore, De Sanctis et al. (2010) use a modified particle backtracking method (PBA) for identification of all possible contaminant source locations. For alarm generation, binary sensor information is introduced. The authors claim that formulating the SI problem as an inverse water quality problem is difficult for three reasons: (1) ill-posedness (sparse sensor grid in contrast to a huge number of possible sources, e.g. hydrants, house connections), (2) problem size (number of possible sources times the number of time steps within the detection time), and (3) assumption of the existence of perfect quality sensors that are capable of measuring the concentration of all relevant substances that do not exist in reality. In addition to the available number of sensors and their reliability, the design of the sensor network is also essential for the effectiveness and accuracy of the source identification algorithm (Ung et al., 2017).

In the following a software tool is presented that is suitable for real-time application and integration within risk or event management systems. It was developed as part of EU-funded project SAFEWATER (contract number: 312764). The algorithmic development is based on a simplified transport algorithm that runs forward and backward in time. Whereas backtracking is used for source identification, forward simulation delivers the estimated current spread of contamination based on the source locations. As a possible response action, the tool also proposes valves to be closed for isolation of contaminants. The software is part of comprehensive event management software (EMS) that collects all information from the field and from different software components that are connected with the EMS, including a newly developed event detection system (EDS) as well as offline and online hydraulic and water quality simulators. The paper starts with an overview of the theoretical background, including a simplified transport model and source identification algorithm. Then, the integration within the EMS framework is outlined. Finally, the application of the tool is presented for the SAFEWATER test lab water network at Water Supply Zurich.

2 Theoretical background

2.1 Transport model

The theory of modelling reaction and transport of substances within a water distribution system is described in a number of books and articles. A good overview of the methods that are included in most of the commercially available simulation software tools can be found in Rossman and Boulos (1996). In the following a very brief presentation of the problem of contaminant injection with sharp quality fronts (discontinuity) is given. Reaction and diffusion are not considered here. In this case the transport is described by the so-called Riemann problem (p. 49 in Toro, 2009):

(PDE) $\quad c_t + vc_x = 0, \quad -\infty < x < \infty, \quad t > 0,$ (1a)

(IC) $\quad c(x,0) = c_0(x) = \begin{cases} c_L & \text{if} \quad x < 0, \\ c_R & \text{if} \quad x > 0. \end{cases}$ (1b)

The partial differential equation (1a, PDE) includes the partial derivatives of the concentration c after time (c_t) and location within the pipe (c_x); v denotes the flow velocity. The initial value problem (IVP) of Eq. (1) differs from conventional transport equations only in the initial condition (Eq. 1b, IC). Here, there are two constant water qualities c_L and c_R that are upstream and downstream of point $x = 0$. For $x = 0$ the initial condition has a discontinuity. For solution the method of characteristics (MOC) can be used. When time passes, the discontinuity is propagated along the characteristic lines with flow velocity v. Accordingly, the solution of this simplest

case of Riemann problems (Eq. 1) is

$$c(x,t) = c_0(x - vt) = \begin{cases} c_L & \text{if } x - vt < 0, \\ c_R & \text{if } x - vt > 0. \end{cases} \quad (2)$$

For the Riemann equation, the characteristic that passes through $x = 0$ is of special interest. It separates the (concentration) surface above the x–t-plane into one part where the concentration is c_L and one part where the concentration is c_R. This particular characteristic line is the only one across which the concentration changes. For implementation, the IVP (1) is solved for all pipes using a MOC. For later processing of the results, the traces of a particle that travels along the special characteristic line that starts from $x = 0$ is stored. In the easiest case the flow velocity is constant and only the slope has to be stored. However, in extended period simulations the flow velocity and even the flow direction may change from external time step to time step. Therefore, it proved to be convenient to store the [time, position, value] triples of all particles for every external time step. The term external time step refers to the time interval between updates of the boundary conditions of the underlying hydraulic simulation model. Since the flow velocities change only after such an external time step, the full information [value, time, position] can be re-established from this information for any time and location.

In order to move from the pipe level – with the initial assumption of infinite pipe length from above – to the network level, additional boundary conditions have to be considered for a finite pipe length. The IVP then is transferred to an IBVP (initial boundary value problem). In general, the combination of pipes at a junction requires the formulation of additional mixing conditions at network nodes. Since no critical concentration can be given for unknown substances, here, the binary information (contaminated yes/no) of the front is transferred unchanged to all pipes having outflows from the junction.

2.2   Source identification

The proposed source identification (SI) module is distinguished from existing solutions by its real-time capabilities, its integration within the EMS and the permanent calculation of the monitoring state of the sensor network even in the case with no alarm. The latter is useful especially when flow directions change due to network operations. The continuous backtracking of negative sensor alarms allows the visualization of the current monitoring state of the system. For any location in the network, the last time of observation is calculated. Given the limited coverage of the distribution network by the sensor network, a contamination may not be detected by the sensors because it can be outside of the covered area. It is important to be aware of this issue if other detections, like customer complaints, indicate an event contradicting the negative alarm states of the sensors.

In the context of SAFEWATER, one basic requirement was that the source identification algorithm must deliver results almost in real time. For that reason, all kinds of optimization-based approaches (see the Introduction) are not suitable due to the huge number of simulations and thus the long computing time that are required for stochastic optimization models. Therefore, a more direct approach was chosen for the solution of the inverse problem. It has some similarities to the particle backtracking algorithm that was first presented by Shang et al. (2002) and the approach presented by De Sanctis et al. (2010). Its core component is an event-driven simplified transport algorithm that does not consider reaction and diffusion mechanisms as described in the previous section. Another key performance indicator is the usage of memory. In the presented approach, a special format was developed that is used for storing the characteristic lines of the transport equation MOC. This information can then be used for particle tracking (forward and backward) as well as reconstruction of time curves (at a certain location) or the concentration along a pipe at a certain time. For implementation in SAFEWATER, an event-driven method is used that strongly simplifies the common water quality equations: instead of concentrations, only binary quality states are calculated (contaminated/not contaminated), neglecting reaction and diffusion terms and using simplified mixing equations. An event is triggered every time a change in the (binary) boundary conditions occurs (for example, start of intrusion in the forward case or release of a sensor alarm in the backward case). The algorithm sends a separator front through the system following the flow velocity of the water in the pipes (forward) or working against it (backtracking case). The moving front separates the network into regions that have distinct values for the binary quality state (contaminated/not contaminated). In the reverse case, a change in the sensor alarm state is considered a binary signal (no alarm $\rightarrow$ alarm). Based on this simplifying assumption, memory requirements and calculation time are minimized in comparison with common water quality solvers. The fact that in the case of a real event the substance and input concentration of a contamination are presumably not known may serve as a justification for the simplifying assumptions.

For solution of the source identification problem at each external time step (change in the flow vector), the alarm states of the sensors (positive, negative, unknown) are sent through the system in reverse time. If at least one of the sensors is positive, the nodes and pipes upstream of the sensor are identified, which are possible candidates for the location of the contamination source. If more sensors have positive alarm states, normally released at different times, the results of the backtracking calculations are combined. Necessary conditions for source candidate locations are that (1) signals of all positive sensors passed the location during the backtracking, and (2) no signal of a negative sensor was observed. Since the conditions are necessary but not sufficient, the backtracking, in general, cannot give unique source lo-

cations. Based on a worst case assumption a unique source location is selected that serves as input for a simplified look-ahead transport calculation that gives an estimation over the future spread of contamination. As a worst case, the node with the biggest outflow volume is chosen. This second step is necessary since the backtracking algorithm in general does not give unique results. The region of possible source locations can be quite large depending on the efficiency of the sensor network.

It is important to mention that negative sensor alarms also deliver useful information. The backtracking of negative alarm states (under the assumption of perfect sensors) identifies the nodes upstream of these sensors and the times when the signal passes the nodes. A contamination starting from such a node before the calculated arrival time of the negative separator front is not possible since, in this case, the sensor alarm state must have been positive as well. Consequently, the nodes can be withdrawn from the list of source candidates. Another important application of backtracking of negative sensor alarms consists in the continuously updated monitoring state of the system. The backtracking algorithm calculates a kind of reverse water age for each location under the protection of the sensor network. That means that for any location the time of the last observation is known. In addition, the nodes and pipes that are not covered by the sensor network can be visualized and updated in real time.

For estimation of the future spread of contamination and the identification of isolation valves, the simplified forward transport calculation is used. Here, the assumption is again that more complicated water quality calculations such as incomplete mixing, reaction and diffusion are of secondary importance for this particular problem. It is expected that at the beginning of a real contamination event the kind and severity of the substance are unknown. In the SAFEWATER project a second water quality solver was implemented that deals with all of these issues. Once more information about the chemical properties of the agent is available, the enhanced water quality solver can be used for more detailed calculation of concentrations.

3   Outline of the integration within the EMS

For integration of the different software components within a real-time environment and the SAFEWATER-EMS, the existing SirOPC software tool was extended by implementation of additional plugins. Originally, SirOPC was developed for connecting hydraulic simulation software with common OPC server software provided by the SCADA system for receiving and sending real-time operational data. The plugin technology allows the flexible extension of the software. For connection with external data, adapter plugins are used (blue boxes in Fig. 1). Application plugins allow the integration of specialized software tools (green boxes in Fig. 1) with the data managed by the core component. The plugins provide the in-

terfaces between the central data and native applications. For example, the hydraulic solver plugin includes the interface for updating boundary conditions of a hydraulic simulation model with SCADA system data that are delivered through SirOPC and the OPC data adapter. The SIRA plugin is used to connect the source identification solver with the core software and manages the update of the flows calculated by the hydraulic solver. For exchange of mass data, an additional database interface exists. For communication with the EMS, an ActiveMQ (2015) data adapter plugin has been developed in the project that is able to receive messages about changes in alarm states of the sensors that are detected by the EDS. Based on this information, the online variables are updated in the SirOPC core, from where the information is transferred to the SI algorithm. By implementation, the developed SI module runs in combined online mode together with the hydraulic solver. The boundary conditions of the hydraulic solver are updated in regular time intervals by receiving data from the OPC server. After each time step calculated by the hydraulic solver, the flow velocities in the source identification algorithm are updated and new backtracking calculations are carried out. Positive alarms are generated by the EMS as soon as a pending alarm is acknowledged by the operator. Alarms could trigger calculations automatically, but in order to avoid false alarms due to known events, at this stage of the development every alarm has to be manually acknowledged. After that, an ActiveMQ message is sent to SirOPC. The next SI calculation considers the positive alarm and calculates the possible locations for the source of contamination that are consistent with the alarm states of all sensors, including negative sensor alarms.

As a possible response action, the valves are identified that have to be closed for isolation of the contaminant. The results of the different algorithms are presented in a native GUI of the SI application as well as in the EMS map. The execution loop, which is in mutual feedback with the hydraulic simulator, is controlled by so-called custom commands, a kind of virtual state machine (Petri-Net). The results of the SI calculations are stored in a database that can be accessed by the EMS for further processing and visualization.

4   Source identification module example

4.1   Calculation of source candidates

In the following section the SI algorithm is demonstrated using the example of the SAFEWATER test network in Zurich (Fig. 2). The total pipe length of the network is about 160 m. The system consists of transparent PVC pipes (diameter 100 mm) and has three major outlets that allow the operation with adjustable demands. In addition to flow and pressure measurements at the system inlet there are four additional flow measurements installed within the pipe network. The measurement devices are connected to a central OPC server. The hydraulic online simulation is connected with
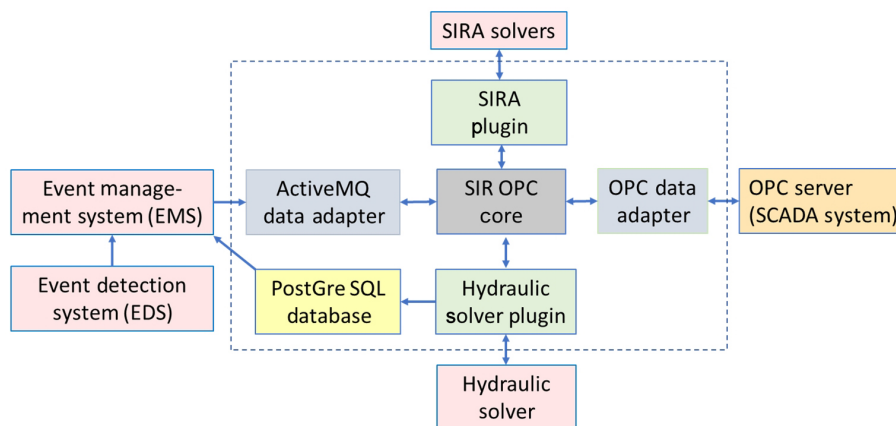
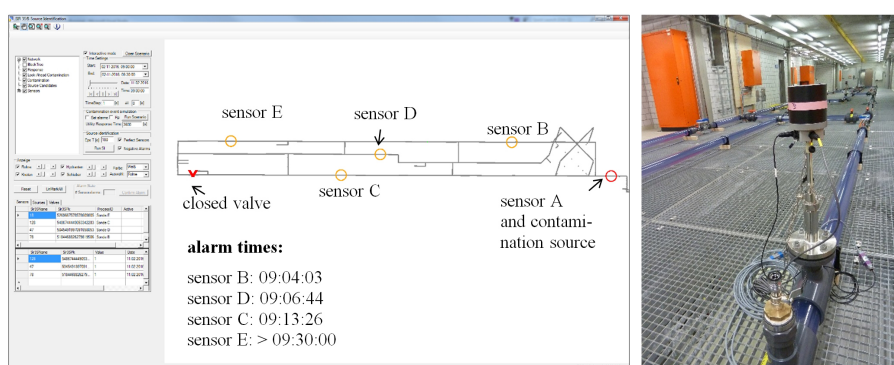**Figure 1.** System architecture of SI integration within the SAFEWATER event management system.

Figure 2. Test network with four sensors (orange circles) and contamination source (red circle).

the OPC server through a bidirectional interface provided by the SirOPC online OPC client. The hydraulic simulation runs every 5 s with updated boundary values from the OPC server. The calculation results, namely the flow velocities, are stored in a native binary data format from where they can be accessed by the SI plugin and transferred to the transport solvers.

Water quality sensors are installed at five locations. Four sensors are distributed over the network (orange circles) and one is located directly downstream from the contamination source (red circle) in order to control the entrance water quality for the injection scenarios. For demonstration purposes, the sensor alarm times were calculated by running a forward transport calculation that starts at 09:02. The SI software tool is able to record the earliest time when the contamination passed the sensors. During the tests, a valve on the lower left-hand side was closed. Therefore, there is almost no flow in direction towards sensor E and the alarm time is beyond the scenario end time (09:30). The subsequent alarm times are also shown in Fig. 2. From Fig. 3 it can be seen that the more information is available (through subsequent additional alarms), the smaller the area of possible source candidates (red marked pipes). In Fig. 3a all pipes upstream of sensor B

are possible locations. In Fig. 3b the combination of the upstream pipes of sensor B and of sensor D decreases the set of source candidates because the locations directly upstream of sensor B cannot explain the alarm at sensor location D. The area is further reduced after the third alarm (Fig. 3c). Possible source locations must also be upstream of sensor C. The orange and blue marked pipes show the estimated spread at the current time and in a predefined look-ahead time, respectively.

4.2 Impact of negative sensor alarms on source identification results and selection of a single source

The sensors do not only deliver valuable information in case of an alarm. Backtracking of negative sensor alarm states also helps to reduce the area of possible source candidates (see for comparison De Sanctis et. al., 2010). Assuming perfect sensors, the upstream locations of the negative sensors cannot be the contamination source since otherwise the contamination had to be detected by the sensor. Negative sensors can clear parts of the source candidate locations. A node (or pipe) is a possible candidate for the contamination source if the sensor signals of all sensors with positive alarms are observed
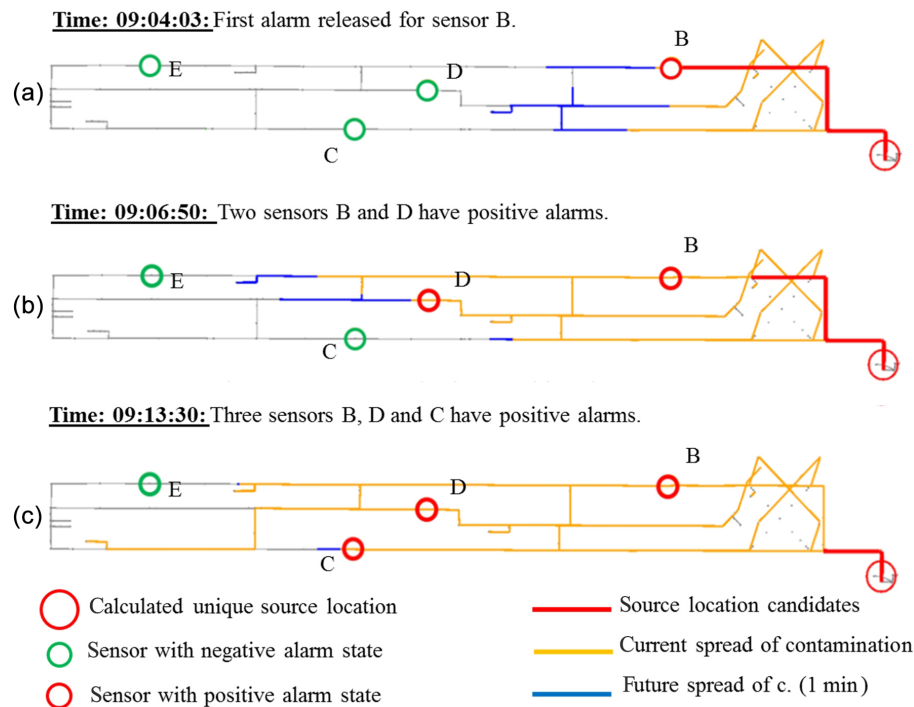
www.drink-water-eng-sci.net/10/53/2017/

Drink. Water Eng. Sci., 10, 53–59, 2017

Time: 09:04:03: First alarm released for sensor B.

(a)

Time: 09:06:50: Two sensors B and D have positive alarms.

(b)

Time: 09:13:30: Three sensors B, D and C have positive alarms.

(c)

○ Calculated unique source location   ▬ Source location candidates

○ Sensor with negative alarm state   ▬ Current spread of contamination

○ Sensor with positive alarm state   ▬ Future spread of c. (1 min)

Figure 3. Sequential release of sensor alarms and results of the SI algorithm, current spread of contaminant and look-ahead.

and no signal of negative sensors can be found for these locations. A particular weighting scheme has been implemented, considering also the duration and time delay between the different signals. The criterion for a node to be a source candidate location is that all signals that passed that node are positive signals. The set of source candidate locations consists of a more or less large area, depending on the design of the sensor network, the number of sensors and the topology of the network graph. For example, if forward calculations shall be carried out for the estimation of the current spread of contamination, a single source node has to be selected.

5 Conclusions

In case of a contamination event, finding fast, efficient and simple response actions after detection is the key issue for mitigating contamination of drinking water distribution networks. The online source identification module indicates the possible locations of the source of contaminants and shows the current monitoring state at any location and any time. The presented SI module is integrated in combination with hydraulic online simulation within the SAFEWATER event management system. On the SI input side the hydraulic online simulator delivers system-wide flow velocities needed for transport calculations based on actual process data that are received from a SCADA system using OPC technology. The continuous update of the model by online data aims at maintaining a sufficient match between the model parameters and the real situation in the field for changing operational conditions. This is a step forward and improves the situation compared to calculations based on offline models. However, it must not be forgotten that uncertainty in model parameters still exists, especially the demands, affecting also the accuracy of the results of the source identification algorithm. The number of unknown parameters normally exceeds the number of measurements by magnitude. The alarm states of sensors are sent by the EMS using the ActiveMQ communication channel. As output, the module generates current monitoring states of the network and, in the case of a positive sensor alarm, the source candidate locations. The output is visualized in the GIS map of the EMS. In addition to the tests for the network on a lab scale, the functionality and applicability of the development of the approaches were also proven within the project for three pilot zone systems of real existing networks (details cannot be presented for security reasons). Future work should focus on the enhancement of the calculation cycle. While the run time of the algorithms is considered to be sufficiently short also for large networks, the huge amount of data to be processed and transferred through the different modules is still a challenge for large real-world applications.

Advisory Group) project. The motivation for and the objectives and achievements of the SAFEWATER project can be viewed at https://www.youtube.com/watch?v=Bs5SljKUxgE.

Edited by: Edo Abraham
Reviewed by: two anonymous referees

References

ActiveMQ: available at: http://activemq.apache.org, last access: 23 May 2017.

Chang, N. B., Pongsanone, N. P., and Ernest, A.: Comparisons between a rule-based expert system and optimization models for sensor deployment in a small drinking water network, Expert Syst. Appl., 38, 10685–10695, 2011.

De Sanctis, A., Shang, F., and Uber, J.: Real-time identification of possible contamination sources using network backtracking methods, J. Water Res. Pl.-ASCE, 136, 444–453, 2010.

Di Nardo, A., Di Natale, M., Msumarra, D., Santonastaso, G. F., Tzatchkov, V., and Alcocer-Yamanaka, V. H.: Dual-use value of network partitioning for water system management and protection from malicious contamination, J. Hydroinform., 17, 361–376, https://doi.org/10.2166/hydro.2014.014, 2014.

Hall, J., Zaffiro, A. D., Marx, R. B., Kefauver, P. C., Krishnan, E. R., Haught, R. C., and Herrmann, J. G.:. On-line water quality parameters as indicators of distribution system contamination, J. Am. Water Works. Ass., 99, 66–77, 2007.

Laird, C. D., Biegler, L. T., Bartlett, R. A., and van Bloemen Waanders, B. G.: Contamination source determination for water networks, J. Water Res. Pl.-ASCE, 131, 125–134, 2005.

Liu, L., Ranjithan, S. R., and Mahinthakumar, G.: Contaminant source identification in water distribution systems using an adaptive dynamic optimization procedure, J. Water Res. Pl.-ASCE, 137, 183–192, 2011.

OPC Foundation: OPC, available at: https://opcfoundation.org, last access: 23 May 2017.

Preis, A. and Ostfeld, A.: Genetic algorithm for source characterization using imperfect sensors, Civ. Eng. Environ. Syst., 25, 29–39, 2008.

Propato, M., Tryby, M. E., and Piller, O.: Linear algebra analysis for contaminant source identification in water distribution systems, World Environmental Water, Tampa, Florida, USA, ASCE, 1–10, 2007.

Rossman, L. and Boulos, B.: Numerical methods for modeling water quality in distribution systems: a comparison, J. Water Res. Pl.-ASCE, 122, 137–146, 1996.

Sandia National Laboratories: CANARY, available at: https://software.sandia.gov/trac/canary, last access: 23 May 2017a.

Sandia National Laboratories: TEVA-SPOT, available at: https://software.sandia.gov/trac/spot/, last access: 23 May 2017b.

Shang, F., Uber, J. G., and Polycarpou, M. M.: Practical backtracking algorithm for water distribution systems, J. Environ. Eng., 128, 441–450, 2002.

Toro, E. F.: Riemann Solvers and Numerical Methods for FLuid Dynamics – A Practical Introduction, 3rd Edn., Springer, Berlin Heidelberg, Germany, 2009.

Ung, H., Piller, O., Gilbert, D., and Mortazavi, I.: Accurate and optimal sensor placement for source identification of water distribution networks, J. Water Res. Pl., 143, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000777, 2017.