SAND2017-0714J

Testing Contamination Source Identification Methods for Water Distribution Networks

Arpan Seth [1], Katherine A. Klise [2], John D. Siirola [3],
Terranna Haxton [4], and Carl D. Laird, A.M.ASCE [5]

ABSTRACT

In the event of contamination in a water distribution network (WDN), source identification (SI) methods that analyze sensor data can be used to identify the source location(s). Knowledge of the source location and characteristics are important to inform contamination control and cleanup operations. Various SI strategies that have been developed by researchers differ in their underlying assumptions and solution techniques. The following manuscript presents a systematic procedure for testing and evaluating SI methods. The performance of these SI methods is affected by various factors including: size of WDN model, measurement error, modeling error, time and number of contaminant injections, and time and number of measurements. This paper includes test cases that vary these factors and evaluate three SI methods on the basis of accuracy and specificity. The tests are used to review and compare these different SI methods, highlighting their strengths in handling various identification scenarios. These SI methods and a testing framework that includes the test cases and analysis tools presented in this paper have been integrated into EPA's Water Security Toolkit (WST), a suite of software tools to help researchers and others in the water industry evaluate and plan various response strategies in case of a contamination incident.

[1]Graduate Research Assistant, School of Chemical Engineering, Purdue University, 480 Stadium Mall, West Lafayette, IN 47907, USA.

[2]Senior Member of Technical Staff, Geoscience Research and Applications Group, Sandia National Laboratories, PO Box 5800 MS 0751, Albuquerque, NM 87185, USA.

[3]Principal Member of Technical Staff, Computing Research Center, Sandia National Laboratories, PO Box 5800 MS 1326, Albuquerque, NM 87185, USA.

[4]Environmental Engineer, National Homeland Security Research Center, U.S. EPA, E-mail: haxton.terra@epa.gov, 26 W. Martin Luther King Dr., Cincinnati, OH 45268, USA.

[5]Associate Professor, School of Chemical Engineering, Purdue University, 480 Stadium Mall, West Lafayette, IN 47907, USA (corresponding author). Tel: +1 765 494 0085; E-mail: lairdc@purdue.edu.

Finally, a set of recommendations are made for users to consider when working with different categories of SI methods.

Keywords: Drinking water distribution system; Source identification; Testing

INTRODUCTION

Municipal water distribution networks (WDNs) are large complex systems with many access points, leading to the potential for accidental or intentional contamination. Rapid response and mitigation of contamination incidents requires a three-part approach. First, improve security at WDN interface points (e.g., physical security at treatment plants and storage tanks, and backflow preventers at customer interfaces). Second, implement an event detection system (EDS) that includes contamination sensors to rapidly alert system operators to the presence of contamination. Third, develop response plans with a goal to rapidly contain, remove, or neutralize contamination using actions like closing isolation valves, flushing network pipes, or injecting decontaminant agents. Response actions can be more effective with an accurate understanding of the extent of the contaminant progress within the WDN; estimating the extent of contamination requires having an accurate real-time model of the network and knowledge of the contamination source. Accurate real-time WDN modeling is vital for both source identification (SI) and response planning and is being studied by various researchers (Shang et al., 2006; Davidson and Bouchart, 2006; Preis et al., 2009).

The SI problem is typically formulated as an inverse problem with the objective to find the source location of a contamination incident using the limited measurement data available from a sparse set of sensors. Several researchers have proposed different methods to solve this problem. Given the diversity of SI methods proposed in the literature, there is the need for a common set of tests to evaluate their performance. Thus, this paper presents a testing methodology for SI techniques. This methodology includes a comprehensive set of potential contamination scenarios designed to cover a wide variety of factors that impact the effectiveness of SI techniques. In addition, three basic techniques are evaluated for their effectiveness by comparing both accuracy and specificity. These SI methods and a testing framework that includes the test cases and analysis tools presented in this paper have been integrated into EPA's Water Security Toolkit (WST), a suite of software tools to help researchers and others in the water industry evaluate and plan various response strategies in case of a contamination incident (EPA, 2014). The testing framework uses network models available in the literature. The framework also provides flexibility for academics, consultants, and water utilities to test new and existing SI methods on their own WDN models.

BACKGROUND

Given the occurrence of a water quality incident, SI is a critical step needed to begin control and cleanup of the contaminant spread. Solving this problem is challenging due to limited sensor information, the uncertainty associated with potential source locations, the accuracy of the WDN model, the demand estimates, and the size and complexity of the distribution system itself. A significant body of research exists describing different approaches for this problem.

Early work assumed the availability of contaminant concentration measurements from water quality sensors. Shang et al. (2002) present a water quality modeling framework called the Particle Backtracking Algorithm (PBA) and suggest its application in the identification of unknown contamination sources in WDNs. Laird et al. (2005) present a least-squares formulation that seeks to find the contamination source profile that minimizes the sum of squares of the difference between calculated and measured contaminant concentrations observed at water quality sensors. The authors later extend this approach to identify multiple contamination sources (Laird et al., 2006) and to deal with the non-uniqueness inherent in the inverse problem (Laird et al., 2005). An alternate approach is introduced in Preis and Ostfeld (2006), where a large number of contamination simulations are performed using EPANET (Rossman, 2000b) to build an approximate model in the form of hybrid model trees. Once contamination has been detected, the SI algorithm involves "climbing backwards" in the model tree and solving a Linear Programming problem at each step. In addition, a number of simulation-optimization approaches use a water quality simulator (e.g., EPANET) as a black-box to perform SI. Pattern-search methods, or Genetic Algorithms (GAs), are common approaches for finding the solution to these black-box problems. Preis and Ostfeld (2007) demonstrate a method to perform SI, which links EPANET with a GA and extend their work by building an input-output relationship matrix through an offline simulation process to help better initialize the GA (Preis and Ostfeld, 2008). Alternatively, Guan et al. (2006) provide a simulation-optimization approach that solves a least squares optimization problem using EPANET as a black-box function with gradient information from a finite difference approximation.

The ability of an SI technique to correctly determine the true injection scenario is significantly limited by the accuracy, reliability, and placement of sensors in the contamination warning system (Tryby et al., 2010). Due to the lack of prior knowledge about the contaminant (chemical or biological), recent developments in contamination detection technology utilize fault detection approaches by monitoring standard water quality measures (e.g., pH, free chlorine, turbidity, conductivity) to provide a binary yes/no indication of the presence or absence of contamination in the network (EPA, 2010; Oliker and Ostfeld, 2014; Zhao et al., 2014). The aforementioned EPANET-GA based algorithm of Preis and Ostfeld (2008) does consider three types of measurements - concentration, fuzzy (low, medium, high), and binary (yes/no) - and concludes that finding unique solutions to the SI problem becomes more difficult going from complete concentration information to only binary information. Cristo and Leopardi (2008) provide an input-output model based SI technique in which the model is built by running a large number of EPANET water quality simulations (the use of PBA is also suggested). Although the authors assume the availability of concentration information, the adverse effect of measurement error is considered to verify the robustness of the overall algorithm. Liu et al. (2011) extend the work of Zechman and Ranjithan (2009) by a presenting an adaptive evolutionary strategy linked with EPANET that considers binary measurements in the form of detection thresholds. The Contaminant Status Algorithm (CSA) of De Sanctis et al. (2010) utilizes binary measurement data and the input-output model generated from PBA to identify possible candidate injection nodes as being safe, unsafe, or unknown. This work was later extended by De Sanctis et al. (2008) to provide a Bayesian probabilistic approach that explicitly considers false positives and false negatives. Using the same modeling technique, Yang and Boccelli (2014) propose an improved (in terms of selectivity) source probability updating scheme based on a Beta-Binomial approach.

Apart from the challenges associated with contaminant measurement, network modeling errors can also be introduced due to demand variability, inaccurate estimation of pipe friction factors, and contaminant reaction dynamics. In order to address these uncertainties, various researchers have proposed statistical approaches to this problem. Given a prior probability of a node being a contamination node, the Bayesian methodology developed by Propato et al. (2010) is designed to calculate the corresponding posterior probability by minimizing an entropy function that represent the amount of information available. Liu et al. (2011) propose a similar entropy minimization approach that builds a Logistic Regression Model by running a large number of contamination simulations and then use this model to calculate probability values required to evaluate the entropy function. Perelman and Ostfeld (2013) also propose an entropy minimization method that uses a network clustering mechanism to represent a WDN as an acyclic graph, which can then be used to run small number of simulations required to calculate probabilities for the entropy function. See Wagner et al. (2015), Wagner and Neupauer (2013), and Wang and Harrison (2014) for more recent advancements in probabilistic approaches for contamination SI. Other researchers have also proposed a data mining approach that requires building a large database containing historical or simulated contamination scenario characteristics (Huang and McBean, 2009; Shen and McBean, 2012). This database is used in a real-time situation with a likelihood maximization method to identify the contamination source(s).

The majority of the SI methodologies proposed in the literature assume measurement information coming from sensors placed at fixed locations around the network. Alternatively, using mobile sensors or manual sampling teams to dynamically choose measurement locations during a response to a contamination incident has shown promising results (Eliades and Polycarpou, 2012; Mann et al., 2012). Mann et al. (2012) present an algorithm based on two Mixed Integer Linear Programming (MILP) models, with the first performing SI and the second selecting nodes for subsequent manual sampling to refine the source identified by the first model.

PROBLEM DEFINITION

In this section, the formulation being considered in this paper is defined. Measurement data is assumed to be available from a fixed number of sensors located at specific nodes in a WDN. An EDS provides discrete yes/no measurements that indicate the presence or absence of contamination in the water. In addition, all sensors in the network are assumed to provide measurements at the same constant frequency. Note that although this assumption is used in this paper, it is easy to relax this assumption for any of the tests. For example, one probable response to the initial detection can be to obtain additional grab sample measurements to confirm the contamination. Although, the SI methods studied in this work can make use of the information provided from these grab sample measurements, the test scenarios discussed in this paper do not consider these kind of measurements. Therefore, using measurements from fixed continuous sensors, the goal of performing SI is to identify the candidate locations where contamination could have taken place. This inverse problem is solved considering a fixed historical time period called the *time horizon*, providing a measure of likeliness for all nodes. This measure is used to provide a ranking of all nodes where a higher value indicates a greater chance of being the contamination source. It is assumed that contaminant ingress can take place at any node (junctions, tanks, and reservoirs) in the entire network, and both single and multiple injections can occur.

PERFORMANCE METRICS

It is assumed that SI methods output a list of possible injection nodes along with their corresponding *measure of likeliness*. The likeliness measure is dependent on the SI method being used. For example, for a probability-based method that reports a probability of a node being the true injection node, this probability value is used as the measure of likeliness. The measure of likeliness used for each SI methods studied in this paper is described later in the section describing these methods. To measure the performance of a SI method, two important criteria must be considered. First and foremost, a method should be able to correctly identify the true injection location(s) as the most likely location(s). Secondly, a method should be able to distinguish the true injection location(s) from the rest of the candidate nodes as effectively as possible. To this end, Yang and Boccelli (2014) introduce two performance metrics - *Accuracy* and *Specificity* - and this paper uses modified versions of these metrics as shown in Equations 1 and 2.

$$\mathit{Accuracy}\,(\%) = \frac{\mathit{Likeliness\ measure\ of\ the\ true\ injection\ node}}{\mathit{Highest\ likeliness\ measure\ over\ all\ candidate\ nodes}} \times 100 \tag{1}$$

$$\mathit{Specificity}\,(\%) = \frac{\mathit{Number\ of\ nodes\ with\ lower\ likeliness\ than\ true\ injection\ node}}{\mathit{Total\ number\ of\ candidate\ nodes}} \times 100 \tag{2}$$

Here, a 100% accuracy indicates that the true injection node had the highest likeliness value , while a high value of specificity indicates a high rank of the true injection node among all the candidate nodes. For example, if the true injection node is given a likeliness value of 5, and the remaining nodes are 2, then the accuracy is 100%, and the specificity is as close to 100% as possible. However, it is possible to have a high accuracy with low specificity. For example, if a method returns a likeliness value of 2 for all nodes, the accuracy value is still 100%, but the specificity value is zero. Although Equations 1 and 2 are defined for scenarios with only a single true injection node, in the case of multiple injection nodes, multiple values of both metrics are calculated with respect to each true injection node.

TESTING METHODOLOGY

In this section, some of the challenges inherent for SI methods in a real-time response system are introduced. The following subsections describe a category of test cases designed to demonstrate the effectiveness or identify limitations of different SI methods. Most of the test cases presented in this paper are created using Net3 (an example network from EPANET (Rossman, 2000a)). Table 1 provides the standard specifications used in generating these test cases. Some test cases require varying these standard specifications, while most of the test cases require additional specifications discussed in the corresponding subsections.

Measurement Error

Various researchers have demonstrated the impact of measurement error on their SI method's performance. Water quality sensors can encounter measurement error in the form of both false positives and false negatives. In general, false positives can lead an SI calculation to identify an unnecessarily large number of candidate nodes, while false negatives could decrease the likelihood of identifying the true contamination node(s).

In order to test whether an SI method is reliable in the likely occurrence of both false positive and false negative measurements, a set of tests with a range of false positive and false negative rates is used. These tests are generated by simulating contamination scenarios in EPANET Net3 at five different nodes located in different parts of the WDN. The specifications provided in Table 1, with modified measurement start and end times, are used in designing this test set. For each scenario, the binary measurements from five sensors (optimally placed using TEVA-SPOT in WST) are collected over a 12-hour span starting 8 hours before the detection time and ending 4 hours after the detection time. Next, measurement error is artificially introduced by randomly selecting which measurements are in error based on the specified false positive rate (FPR) and false negative rate (FNR). All permutations of FPR and FNR values in the set - [0, 0.1, 0.2, 0.3, 0.4] are chosen in designing these tests. Finally, each combination of FPR and FNR is sampled 50 times to obtain statistics. To summarize, each test is identified by an injection node, an FPR value, an FNR value, and a sample number, giving a total of 6005 tests (FNR=FPR=0 does not require multiple samples).

Modeling Error

Due to the limited availability of data for model tuning and a lack of real-time demand information, error can be expected between a network model and the true flow fields in the distribution system. Therefore, it is important to include test cases that are designed to assess the performance of a SI method in the presence of network modeling errors.

Demand variability is a naturally stochastic phenomenon that is challenging to estimate especially with the lack of real-time data. Typically, SI methods incorporate a hydraulic model that uses estimated demand patterns to model the flow rates and directions inside each pipe in the WDN. Therefore, errors in demand estimates are propagated to errors in modeled flow rates and directions. Furthermore, errors in flow direction can have a drastic impact on the results by potentially switching the set of upstream nodes which could contain the true injection node(s). Inaccurate flow rates can also impact SI results by shifting the estimated time profile of the contaminant at a sensor node. To evaluate a method in the presence of modeling error, a set of test cases are generated with different levels of demand variability between the model used to simulate the contamination incident and the model used to perform SI. This produces error between the model used by the SI method and the measurement data. Details of the specifications used in designing this test set are provided in Table 1. In this set of tests, a base case model is used to generate the measurement data. To form the base case, the demands of all nodes of Net3 are reduced by 20%. This is done to avoid infeasibly high demand values when random error is added to the system. Base cases are formed from each of the five contamination scenarios described earlier, and the simulations are used to obtain the measurement data. Next, random demand error over a range of error percentages - [1, 2, 4, 8, 10, 20] - is added to this base case to form "erroneous" models given to the SI methods. This is achieved by generating model input files (EPANET INP format) containing error in the base demand values at each node. Finally, for every error range value, multiple samples (50) are taken to generate a number of different input files. Hence, each test in this set consists of measurements obtained for a particular injection scenario, a model input file with a particular error range, and a sample number, giving a total of 1500 test cases.

Injection Characteristics

Because of the limited number of sensor measurements, the SI problem is inherently non-unique (many possible locations and/or injection profiles). Most SI methods acknowledge these limitations and some methods impose additional constraints on the characteristics of a possible solution. Two such characteristics are the number of simultaneous injection locations considered during a contamination scenario, and the length of the contaminant injection at a node.

While an algorithm that assumes a single injection location might be able to more accurately identify the source in a single source test, it is important to know how it performs if multiple injections occur. The accuracy of methods that do not make this assumption should be evaluated. Therefore, this study includes test cases where the number of injection nodes in a scenario is varied from a single injection node to three injection nodes in Net3. Details of the specifications used in designing this test set are provided in Table 1. Three injection scenarios are generated - single injection at node 151, two simultaneous injections at nodes 111 and 151, and three simultaneous injections at nodes 111, 151, and 189. In the case of multiple injections, the time horizon is chosen in reference to the longest detection time of any of the individual injections.

This study also incorporates tests to investigate the capabilities of a SI method in identifying contamination scenarios of varying injection lengths. In the presence of measurement and modeling errors, longer injections are generally more distinguishable compared with shorter injections, which could be completely missed by periodic sampling of the sensors or produce a short pulse of positive measurements that can be more difficult to assess. The details of the first set of tests are provided in Table 1 with the exception of the injection length, which is varied over a range. For each injection node, this range contains the following injection lengths (in minutes) - [60, 120, 240, 480, 720]. In order to capture the difference in measurement profiles produced by the injections of different lengths, the measurement end time is increased to include 12 hours following the initial detection. Another set of tests that contain the same simulated injections (nodes and injection lengths) but with added measurement error (FPR=10%, FNR=10%) is also included in this test set. The test cases with measurement error are run 50 times with different random seeds to obtain performance statistics. Each case in both injection length test sets (with and without error) is identified by an injection length, an injection node, and (in case of measurement error) a sample number. Therefore, this category contains a total of 1275 tests.

Time Horizon

In order to keep the size of the SI problem reasonable, the calculations are typically performed by limiting the window of time under consideration. This is generally called the analysis time horizon or the time horizon. While reducing the time horizon can save computational expense, if it is too short, it can falsely limit the space of potential injection locations. Picking the right time horizon is not straightforward. Ideally, it should be at least as big as the longest flow path to any sensor in the WDN. However, for realistic large-scale networks this can limit the efficiency of real-time SI. Therefore, good methods need to be aware of the limitations imposed by a selected time horizon and indicate these limitations in the results they produce. This study includes tests that vary the time horizon used by a method. Details of this test set are provided in Table 1, with the exception of the time horizon that is varied over the set - [1, 2, 4, 8, 16, 24] (hours). Each case in this test set consists of an injection node and a time horizon used to perform SI for that injection, giving a total number of 30 tests.

Network Size

The size and complexity of the WDN can impact the performance of a SI method. One major factor is the non-uniqueness of the solution, which, for a fixed number of sensors, increases significantly with the network size. Not only can the quality of the solution to the SI problem be negatively impacted as the network size increases, but the computational effort in terms of solution time and memory requirement can also increase substantially. A set of tests are provided that contains WDNs of three different sizes - EPANET Net3 (97 Nodes), Network2 (3,358 Nodes) (Watson et al., 2009) and BWSN (Battle of the Water Sensor Networks) Network 2 (12,523 Nodes) (Ostfeld, A., et al., 2008). The system diagrams of these three networks are provided in the supplemental data (Figures S1, S2, and S3). To obtain average performance statistics for each of these networks, injection scenarios are simulated by injecting a contaminant at several nodes selected from different parts of that network. Note that for each of these scenarios, all the nodes in a network are considered as candidate injection nodes for the SI methods. The number of injection scenarios selected for each network is provided in Table 2 along with other specifications used in designing this test set.

Sensor Placement

The cost of buying, operating, and maintaining water quality sensors limits the number of sensors that can be installed in a WDN. Therefore, researchers have proposed optimal placement of fixed sensors to minimize the impact on the population or the network infrastructure due to a contamination incident (Berry et al., 2005; Ostfeld and Salomons, 2004; EPA, 2010). Although optimal sensor placement is a separate problem to SI, the location of these sensors can have a major impact on the performance of SI methods. Only a few papers have investigated sensor placement for better SI. Tryby et al. (2010) provide a sensor placement technique that is designed to reduce the ill-posedness of the SI problem. An approach of dynamically selecting manual grab sample locations to improve distinguishability between candidate source nodes has also been proposed in Mann et al. (2012). However, other objectives typical for sensor placement might not be appropriate for SI. For example, a typical maximum coverage objective would place a sensor that detects multiple scenarios, which would reduce a SI methods ability to distinguish between those scenarios.

The sensor density in a network is defined as the percentage of nodes where a water quality sensor is located. This corresponds to the amount of information available for SI. A good SI method should be able to accurately identify candidate contamination nodes with limited information. To investigate the impact of sensor density and layout, two different sets of tests are provided. The first test set varies the density of optimally placed sensors, while the second test set varies the density of sensors that are randomly placed at nodes around the network. Two networks of different sizes are used for this test set: EPANET Net3 (97 Nodes) and Network2 (3,358 Nodes). For Net3, apart from the sensor nodes, the rest of the specifications used in designing the optimal and random sensor placement test sets are provided in Table 1. The tests for Network2 use the same set of parameters with different injection nodes. The list of sensor placements are selected by varying the sensor density over the set [2%, 4%, 6%, 10%, 20%] for Net3 and [0.2%, 0.4%, 0.6%, 0.8%, 1%] for Network2.

For the test set with optimally placed sensors, sensor placement is performed using WST with the objective set to minimize population exposed. To summarize, each case in both test sets (optimal and random) is identified by a network, a sensor density and an injection node, for a total of 100 test cases.

SOURCE IDENTIFICATION METHODS USED IN COMPARATIVE STUDY

The testing methodology described in the previous section is used to compare the performance of three different SI methods. The following subsections describes a Bayesian probability-based method that is provided in WST, the Contaminant Status Algorithm (De Sanctis et al., 2010), and an optimization-based method (Mann et al., 2012). The later two methods are briefly overviewed and readers are referred to their respective publications for more details. The underlying assumptions of each method is highlighted to help explain the performance results presented later.

Bayesian Probability-Based Method

This method operates by simulating all candidate contaminant injections and then calculating the probability of each injection based on how well the simulated measurement profile matches the true measurements obtained from the sensors. These probability calculations are performed using Bayes theorem:

$$P(i|m) = \frac{P(m|i)P(i)}{P(m)} \qquad \forall\, i \in \mathbf{C} \tag{3}$$

Where $\mathbf{C}$ is a set of all possible contamination incidents. An injection can start at any node and any time step, and is assumed to continue for the complete simulation duration (i.e., continuous injections are assumed). $P(i|m)$ is the probability of an incident i given a vector of measurements m. $P(i)$ is the prior probability of contamination incident i. This formulation assumes that only a single injection incident is possible, and therefore $P(i)$ is set to a uniform prior that is the inverse of the cardinality of $\mathbf{C}$ given by $1/|\mathbf{C}|$. Since an estimate of $P(m)$ (the prior probability of the observed measurement) is generally not available, it is common to replace this calculation by normalizing the calculated values of $P(i|m)$ so that they sum up to one. Finally, $P(m|i)$ is the probability of measurement m given injection incident i. It is calculated using the following equation:

$$P(m|i) = (1 - p_f)^{match(i)} p_f^{num\_meas - match(i)} \tag{4}$$

Where, p_f is a user specified estimate of measurement failure probability (false positive or false negative), num_meas is the total number of available measurements, and $match(i)$ is the number of actual discrete measurements that match the discrete measurements obtained by simulating incident i.

The overall algorithm for this method is as follows. Following detection, hydraulic simulations are performed (using EPANET) for a specified time window preceding the detection time and the flow data is used to build a linear input-output water quality model using the origin-tracking algorithm introduced by Laird et al. (2005). Next, the set of candidate injections, $\mathbf{C}$, is populated by analyzing the input-output model and choosing injection node-time pairs that are hydraulically connected only to the positive measurements. Next, all the candidate injections are simulated using the linear input-output model to obtain simulated measurement profiles that are then compared to the actual measurement profile to get the number of matches. The posterior probability of each injection node-time pair is calculated using Equations 3 and 4. Finally, the probability of each node being the injection node is chosen as the highest posterior probability value over all injection node-time pairs containing that particular node. This posterior probability value for each node is used as the measure of likeliness for the performance metrics calculation.

Contaminant Status Algorithm

The Contaminant Status Algorithm (CSA), proposed by De Sanctis et al. (2010), performs SI by assigning status to each candidate node-time pair as either being safe (not an injection candidate), unsafe (possible injection candidate), or unknown. Since the performance metrics calculation requires the results of a SI method to be in the form of a list of candidate nodes with their corresponding measure of likeliness, the CSA was modified to assign a likeliness measure of 1 to a node if it is contained in the list of unsafe node-time pairs, while all other nodes are assigned a likeliness measure of 0. Essentially, CSA operates by iterating over all measurements and pruning out (marking as safe) upstream node-time pairs that are hydraulically connected to negative measurements. Consequently, CSA allows for multiple simultaneous injections, however, it assumes perfect measurements when marking candidate injections as safe.

Optimization-Based Method

The third method used for this comparative study is the MILP based technique from Mann et al. (2012). This method incorporates the linear input-output water quality model Laird et al. (2005); Mann et al. (2014) directly into an optimization formulation that seeks to find an injection source profile that minimizes the mismatch between yes/no measurements and those in the model. This formulation assumes that a sensor yields a positive measurement if the contaminant concentration is above a specified detection threshold concentration and a negative measurement otherwise. Therefore, if a sensor yields a positive measurement, any corresponding calculated concentration from the water quality model above the threshold is in agreement with this measurement data. Hence, while constructing an objective for estimation, only calculated concentrations below this threshold are penalized. Likewise, if a sensor yields a negative measurement, only the corresponding calculated concentration above the threshold is penalized. To identify a number of possible source candidates, the method repeatedly solves the MILP problem, each time adding integer cuts to remove previously found solutions until the objective value at the solution has deteriorated significantly. Note that for each candidate solution, the corresponding inverse of the objective value is used to represent the measure of likeliness of all nodes identified in that solution. Also note that this method allows for multiple simultaneous injections.

DISCUSSION OF TEST RESULTS

In this section, the performance of the three SI methods is compared on the test cases. Each of the following subsections provides the performance plots for the three methods using a particular test set.

Measurement Error

An increase in the false positive and false negative rate is expected to degrade the performance of all three SI methods. Figures 1 and 2 highlight how the three source identification methods behave differently in the presence of false positives versus false negatives.

Figures 1a and 2a show that the Bayesian probability-based method performs worse in the presence of false positive measurements as opposed to false negative measurements (both in terms of accuracy and specificity). This behavior can be attributed to the fact that the probability-based method starts off by selecting a list of initial candidate injections that contains upstream node-time pairs that are hydraulically connected to positive sensor measurements. These candidate injections are then simulated to obtain measurement profiles, which are then matched against the true measurements. Therefore, a higher number of false positives leads to more candidate injections and also increases the possibility of finding injections that match the measurement data better than the true injection, hence degrading both accuracy and specificity.

In Figure 2b, the CSA shows a dramatic decrease in specificity with increased false positive rate. This is because the CSA selects, as injection location candidates, all node-time pairs that are hydraulically connected to any positive measurements. Therefore, an increase in the number of positive measurements leads to an increase in the size of the candidate set. Figure 1b also shows that the CSA maintains 100% accuracy across all levels of negative and positive measurement error. The CSA removes node-time pairs from the candidate set (marks node-time as safe) only when a negative measurement confirms the absence of contamination. Within the framework of this study, the CSA was modified to aggregate over time, including a candidate node if it appears in at least one node-time pair. Therefore, while the CSA will remove a node-time pair in the face of a false negative measurement, as long as the true node is hydraulically connected to at least one positive measurement, the accuracy will be 100%.

The optimization-based method is generally more balanced in the trade-off between accuracy and specificity even in the presence of large amounts of measurement error. A trend to notice in Figure 1c is that an increase in the FNR has a higher impact on the accuracy of this method compared with a similar increase in the FPR. This is likely due to the fact that a typical test scenario has a fewer number of positive measurements (often from a single location of initial detection) compared with the number of negative measurements taken from all sensors over the complete 12 hour time horizon.

Modeling Error

Figure 3 shows a decrease in performance for all three methods as the amount of demand error is increased. For instance, the mean specificity of the optimization-based method at 0% demand error is 90%. This means that considering the average performance over the 5 simulated scenarios, 10 nodes will have to be investigated before finding the true injection node. However, when the demand error is increased to 10%, the specificity reduces to 80%, which means that 20 nodes will have to be investigated. Therefore, for the optimization-based method the percentage of candidate nodes to be investigated doubles with only a

10% increase in the demand error. Similar conclusions can be made about the other two SI methods. This highlights the need for accurate demand estimation when performing SI.

Injection Characteristics

In the first set of tests, multiple simultaneous injections are simulated and the performance of the three methods is measured based on their ability to identify each individual injection node. Tests are run with a single injection, two simultaneous injections, and three simultaneous injections.

The Bayesian probability-based method assumes only a single injection node when performing SI. This method uses a binomial distribution to calculate the prior source probabilities. For a large number of measurements, the binomial distribution has a sharp peak, which means that small changes in the number of matching measurements can lead to big differences in the probability values. This results in a big drop in probability values over the ranked list of candidate nodes. This means that even though the true injection node(s) can be high in rank (high specificity), it can still have low accuracy. Hence, while the test results show that this method has 100% accuracy for a single injection location, for two simultaneous injections the method can only identify one true injection node with 100% accuracy, while the other injection node has 1% accuracy and 80% specificity. For three simultaneous injections, the accuracy for all injection nodes drop to 1% while the specificities are 90%, 70%, and 60%. In contrast, the CSA allows for multiple injections and has 100% accuracy going from one to three simultaneous injections. However, the specificity deteriorates quickly from being 90% for one injection, to 75% for two injections, to 30% for three injections. The optimization-based method performs reasonably on these tests even though the maximum number of injections were set to one in the optimization formulation. The accuracy value(s) for one injection is 100%, for two simultaneous injections is 100% and 60%, and for three simultaneous injections is 100%, 70%, and 40%. The specificity values for the optimization-based method are very similar to the Bayesian probability-based method. These go from 90% for a single injection, to 100% and 60% for two simultaneous injections, to 95%, 80%, and 78% for three simultaneous injections. To further test the impact of modeling error on this analysis, a pairwise sensitivity study of the multiple injection tests that simultaneously considers hydraulic uncertainty has been provided in the supplemental data (Tables S1, S2, and S3).

SI techniques often make assumptions about the length of candidate injections. The probability-based method assumes continuous injections and therefore performs poorly for tests that involve short injection durations. For injection lengths below 4 hours, both mean accuracy and specificity values are under 50%. As expected, both metrics have high mean values ($\sim$100%) when the injection length is over 8 hours.

On the other hand, CSA does not make any assumption regarding the injection length and is therefore more capable of identifying short injections. The method shows 100% mean accuracy and close to 60% mean specificity for all injection lengths. The optimization-based method also shows 100% mean accuracy for all injection lengths, while the mean specificity goes from 60% for 1 hour injections to 80% for 12 hour injections. The optimization-based method does assume continuous injections, but it is still able to accurately identify short injections since an injection at the true node matches the measurement better than any other injection node, even if it is a poor match (many negative measurements do not match). However, short injections are more difficult to identify in the presence of measurement error. Therefore, as expected, adding 10% measurement error (FNR=10%, FPR=10%) leads to an extra 20% reduction in mean accuracy for both optimization-based method and Bayesian probability-based method on injection lengths less than 4 hours, while the CSA still shows 100% mean accuracy on these tests. Adding measurement error results in similar trends in mean specificity for all three methods over all injection lengths.

Time Horizon

As expected, the performance of all three methods improves as the time horizon is increased from 1 to 24 hours. For some cases, the true incident time is outside the time horizon. In those cases, the SI method can identify the true source node, but with an incorrect incident time. Since the node-time pairs were aggregated to only identify nodes in the metrics, in some cases, the correct node will be identified. Nevertheless, on average, small time horizons are expected to result in poor performance. The mean accuracy of Bayesian probability-based method ranges from 1% for 1 hour horizon to 100% for 24 hour horizon, while the mean specificity ranges from 35% to 90%. The mean accuracy for both CSA and optimization-based method ranges from 20% to 100%, while the mean specificity for CSA has a slightly lower range (15% to 80%) as compared to the optimization-based method (40% to 100%). A pairwise sensitivity study that simultaneously considers hydraulic uncertainty is provided in the supplemental data (Tables S4, S5, and S6).

Network Size

Figure 4 shows an increase in the specificity values as the size of the network increases, however this is difficult to compare since the size of the networks differ. Therefore, the figure also includes the numeric values above each specificity bar to indicate the mean number of nodes that need to be investigated before the true injection node is identified. For instance, using CSA on the BWSN2 network on average involves investigating 214 nodes before the true node is identified. On the other hand, for the same network the Bayesian probability-based method requires only 45 nodes to be investigated. This is primarily because the CSA allows for the possibility of multiple injections and also because it produces a relatively large list of all equally likely candidate incidents. All methods showed 100% accuracy on all tests.

Sensor Placement

As expected, for both optimal and random sensor placement, the specificity of all three SI methods (as shown in Figure 5 and Figure 6) improves with higher sensor density due to the increase in the amount of measurement information available from a larger number of locations around the network. All methods showed 100% accuracy on all test cases. It is interesting to see that the optimal sensor placement does not perform as well as the random sensor placement. This is due to the fact that optimal placement of sensors is typically done based on an objective (e.g., minimize population impact, maximum coverage) that is not designed for SI. Typical optimal sensor placement results in sensors being placed at locations that detect larger number of scenarios, which can have a negative impact on a SI method's ability to distinguish between possible injection scenarios. A pairwise sensitivity study that simultaneously considers hydraulic uncertainty is provided in the supplemental data (Figures S4 and S5).

CONCLUSIONS AND FUTURE WORK

In this manuscript a systematic testing methodology for contamination SI methods is presented. This methodology includes metrics and a set of test cases designed to show strengths and weaknesses of different methods. The methodology is then used to perform a comparative study of three different SI methods - a Bayesian probability-based method, the Contaminant Status Algorithm, and an optimization-based method. In general, the test cases presented in this paper were effective at illustrating key differences in the methods, and while a detailed comparison of these methods is not the primary goal of this paper, the following basic conclusions can be drawn. Note that these results are not necessarily indicative of the performance of all methods of a particular class (i.e., all Bayesian probability-based methods, all optimization-based methods, or all CSA type methods).

- The Bayesian probability-based method assumes only single candidate injections and therefore performs poorly (at least in terms of accuracy) in the presence of multiple simultaneous injections. This method does not explicitly consider hydraulic connections between the sensor and candidate nodes (unconnected nodes-time pairs can match negative measurements). Furthermore, this method has poor accuracy in the presence of a large amount of false positive measurements. However, with reasonably good information (low measurement error, low demand error) this method shows higher accuracy and specificity in identifying single injections compared with the other two methods.
- The Contaminant Status Algorithm has higher accuracy than the other two methods,

but typically shows lower specificity since it provides an exhaustive list of hydraulically connected node-time pairs with no negative measurement to mark them as safe. Unlike the other two methods, CSA does not make any assumptions about the length of candidate injections and therefore shows better performance in identifying short injection lengths. The specificity of this algorithm becomes worse as the number of positive measurements are increased, since more candidate injections are hydraulically connected to these measurements. Nevertheless, the fact that this method has good accuracy in the presence of large amount of measurement and modeling error can be used to shortlist the candidate set for further SI calculations. More recent work by De Sanctis et al. (2008) extends this method to a Bayesian probabilistic approach.

- The optimization-based method shows good performance in most test cases, especially in the presence of large amount of measurement error. However, this method has tuning parameters (e.g., detection threshold) that could affect performance in a real system, is more difficult to implement, and can be computationally intensive.

The results presented in this paper highlight the importance of analyzing various factors that impact the performance of SI methods. Random noise on individual measurement points was added to show that the performance of all three methods can start to degrade at high false negative and false positive rates. In the future, it will be interesting to study the impact of more systematic sensor failures. Using the test set that varied the amount of modeling error, the importance of accurate demand estimation on the performance of SI methods is highlighted. Real-time demand estimation, which is an area of active research (Shang et al., 2006; Davidson and Bouchart, 2006; Preis et al., 2009), can play a crucial role in improving the performance of SI methods. Furthermore, while higher sensor density typically leads to better performance of SI methods, typical criteria used for optimal sensor placement are not ideal for SI and can even be worse than a simple random sensor placement method. Identifying sensor placements to improve SI performance is an interesting topic for future study. The effectiveness of different SI methods with an increase in the network size was also studied. Given the short window of data considered in these tests, for larger networks, the number of injection nodes identified can sometimes become prohibitively large. For these cases, a manual sampling strategy to help distinguish between candidate nodes could be an extremely effective way to adaptively increase available information. In future work, the test bed will be extended to include cases for SI methods that incorporate measurements in the form of manual grab samples. The methods described in this paper do not make use of the specific identity of the contaminant (e.g., to model decay/reactions). In general, the SI methods should be effective immediately, before the contaminant may be identified through laboratory analysis. If the specific compound is known, then the water quality models could be modified to include kinetic models. This is a reasonable direction for future work in these methods. Furthermore, it is important to consider the online and offline computational performance of SI methods since they need to be useful for real-time response to a contamination incident. Network size and complexity will have a major impact on the computational performance of various SI methods and therefore a metric to measure computational efficiency (both speed and memory usage) will be included in the future. A testing framework that includes the set of tests proposed in this paper, while allowing users to generate their own set of tests, has been developed as a part of EPA's Water Security Toolkit (WST) and will be included as a module in future releases. Others are invited to use this tool for testing new SI methods and to contribute to the testing framework. With the simultaneous development of real-time data collection systems, real-time modeling tools, and real-time SI tools, there is a need to study and optimize their interactions, which opens up new challenges associated with monitoring and protecting drinking water networks.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support provided by Sandia National Laboratories and U.S. Environmental Protection Agency.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S.

Department of Energys National Nuclear Security Administration under contract DE-AC04 94AL85000. SAND Number SAND2014-20020 J.

The U.S. Environmental Protection Agency (EPA) through its Office of Research and Development funded and collaborated in the research described here under an Interagency Agreement (IA # DW8992192801) with the Department of Energys Sandia National Laboratories. It has been subject to an administrative review but does not necessarily reflect the views of the Agency. No official endorsement should be inferred. EPA does not endorse the purchase or sale of any commercial products or services.

SUPPLEMENTAL DATA

Figures S1-S5 and Tables S1-S6 are available online in the ASCE Library (ascelibrary.org).

REFERENCES

Berry, J., Fleischer, L., Hart, W., Phillips, C. A., and Watson, J.-P. (2005). "Sensor placement in municipal water networks." *Journal of Water Resources Planning and Management*, 131(3), 237–243.

Cristo, C. and Leopardi, A. (2008). "Pollution source identification of accidental contamination in water distribution networks." *Journal of Water Resources Planning and Management*, 134(2), 197–202.

Davidson, J. and Bouchart, F.-C. (2006). "Adjusting nodal demands in scada constrained real-time water distribution network models." *Journal of Hydraulic Engineering*, 132(1), 102–110.

De Sanctis, A., Shang, F., and Uber, J. (2010). "Real-time identification of possible contamination sources using network backtracking methods." *Journal of Water Resources Planning and Management*, 136(4), 444–453.

De Sanctis, A. E., Boccelli, D. L., Shang, F., and Uber, J. G. (2008). "Probabilistic approach to characterize contamination sources with imperfect sensors." *World Environmental and*

Water Resources Congress 2008, Reston, VA, American Society of Civil Engineers, 1–10 (May).

Eliades, D. and Polycarpou, M. (2012). "Water contamination impact evaluation and source-area isolation using decision trees." *Journal of Water Resources Planning and Management*, 138(5), 562–570.

EPA, U. S. (2010). "Water quality event detection systems for drinking water contamination warning systems: Development, testing, and application of CANARY." *Report No. EPA/600/R-10/036*, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

EPA, U. S. (2014). "Water security toolkit user manual: Version 1.1." *Report No. EPA/600/R-13/353*, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

Guan, J., Aral, M., Maslia, M., and Grayman, W. (2006). "Identification of contaminant sources in water distribution systems using simulation-optimization method: Case study." *Journal of Water Resources Planning and Management*, 132(4), 252–262.

Huang, J. and McBean, E. (2009). "Data mining to identify contaminant event locations in water distribution systems." *Journal of Water Resources Planning and Management*, 135(6), 466–474.

Laird, C., Biegler, L. T., Waanders, B. G. v. B., and Bartlett, R. A. (2005). "Contamination source determination for water networks." *Journal of Water Resources Planning and Management*, 131(2), 125–134.

Laird, C. D., Biegler, L. T., and van Bloemen Waanders, B. G. (2006). "Mixed-integer approach for obtaining unique solutions in source inversion of water networks." *Journal of Water Resources Planning and Management*, 132(4), 242–251.

Liu, L., Ranjithan, S. R., and Mahinthakumar, G. (2011). "Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure." *Journal of Water Resources Planning and Management*, 137(2), 183–192.

Mann, A. V., Hackebeil, G. a., and Laird, C. D. (2014). "Explicit water quality model generation and rapid multi-scenario simulation." *Journal of Water Resources Planning and Management*, 140(5), 666–677.

Mann, A. V., McKenna, S. a., Hart, W. E., and Laird, C. D. (2012). "Real-time inversion in large-scale water networks using discrete measurements." *Computers & Chemical Engineering*, 37, 143–151.

Oliker, N. and Ostfeld, A. (2014). "Minimum volume ellipsoid classification model for contamination event detection in water distribution systems." *Environmental Modelling & Software*, 57, 1–12.

Ostfeld, A. and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *Journal of Water Resources Planning and Management*, 130(5), 377–385.

Ostfeld, A., et al. (2008). "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms." *Journal of Water Resources Planning and Management*, 134(6), 556–568.

Perelman, L. and Ostfeld, A. (2013). "Bayesian networks for source intrusion detection." *Journal of Water Resources Planning and Management*, 139(4), 426–432.

Preis, A. and Ostfeld, A. (2006). "Contamination source identification in water systems: A hybrid model treeslinear programming scheme." *Journal of Water Resources Planning and Management*, 132(4), 263–273.

Preis, A. and Ostfeld, A. (2007). “A contamination source identification model for water distribution system security.” *Engineering Optimization*, 39(8), 941–947.

Preis, A. and Ostfeld, A. (2008). “Genetic algorithm for contaminant source characterization using imperfect sensors.” *Civil Engineering and Enviornmental Systems*, 25(1), 29–39.

Preis, A., Whittle, A., and Ostfeld, A. (2009). “Online hydraulic state prediction for water distribution systems.” *World Environmental and Water Resources Congress, American Society of Civil Engineers (ASCE).*

Propato, M., Sarrazy, F., and 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(4), 483–492.

Rossman, L. (2000a). “EPANET 2 (Software), U.S. Environmental Protection Agency, <http://www2.epa.gov/water-research/epanet>.

Rossman, L. (2000b). “EPANET 2:Users Manual. Cincinnati, OH: U.S. Environmental Protection Agency.

Shang, F., Uber, J. G., and Polycarpou, M. M. (2002). “Particle backtracking algorithm for water distribution system analysis.” *Journal of Environmental Engineering*, 128(5), 441–450.

Shang, F., Uber, J. G., van Bloemen Waanders, B. G., Boccelli, D., and Janke, R. (2006). “Real time water demand estimation in water distribution system.” *8th Annual Water Distribution Systems Analysis Symposium, Cincinnati, OH.*

Shen, H. and McBean, E. (2012). “False negative/positive issues in contaminant source identification for water-distribution systems.” *Journal of Water Resources Planning and Management*, 138(3), 230–236.

Tryby, M., Propato, M., and Ranjithan, S. (2010). "Monitoring design for source identification in water distribution systems." *Journal of Water Resources Planning and Management*, 136(6), 637–646.

Wagner, D. E. and Neupauer, R. M. (2013). "Probabilistic contaminant source identification in water distribution systems with incomplete mixing at pipe junctions." *World Environmental and Water Resources Congress 2013*, ASCE, 930–935.

Wagner, D. E., Neupauer, R. M., and Cichowitz, C. (2015). "Adjoint-based probabilistic source characterization in water-distribution systems with transient flows and imperfect sensors." *Journal of Water Resources Planning and Management*, 141(9), 04015003.

Wang, H. and Harrison, K. W. (2014). "Improving efficiency of the bayesian approach to water distribution contaminant source characterization with support vector regression." *Journal of Water Resources Planning and Management*, 140(1), 3–11.

Watson, J.-P., Murray, R., and Hart, W. (2009). "Formulation and optimization of robust sensor placement problems for drinking water contamination warning systems." *Journal of Infrastructure Systems*, 15(4), 330–339.

Yang, X. and Boccelli, D. L. (2014). "Bayesian approach for real-time probabilistic contamination source identification." *Journal of Water Resources Planning and Management*, 140(8), 04014019.

Zechman, E. M. and Ranjithan, S. R. (2009). "Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system." *Journal of Water Resources Planning and Management*, 135(5), 334–343.

Zhao, H., Hou, D., Huang, P., and Zhang, G. (2014). "Water quality event detection in drinking water network." *Water, Air, & Soil Pollution*, 225(11), 1–15.

List of Tables

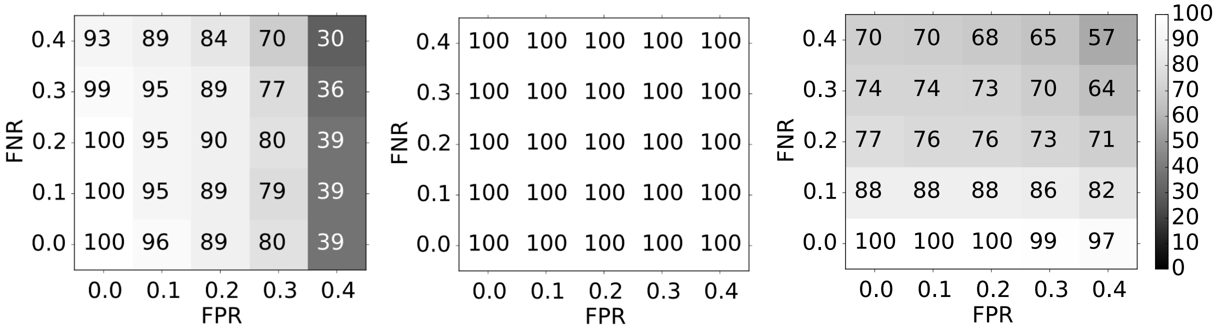
TABLE 1: Standard specifications used for generating test sets.

Specification Parameter	Value(s)
Network	EPANET example Net3
Injection Nodes	111, 151, 189, 183, 229
Sensor Locations	149, 117, 167, 213, 253
Injection Start Time	24 hours into simulation
Injection Length	10 hours
Measurement Start Time	10 hours before detection
Measurement End Time	2 hours after detection
Sensor Frequency	2 measurements per hour
Time Horizon	24 hours

TABLE 2: Specifications used for generating the network size test set.

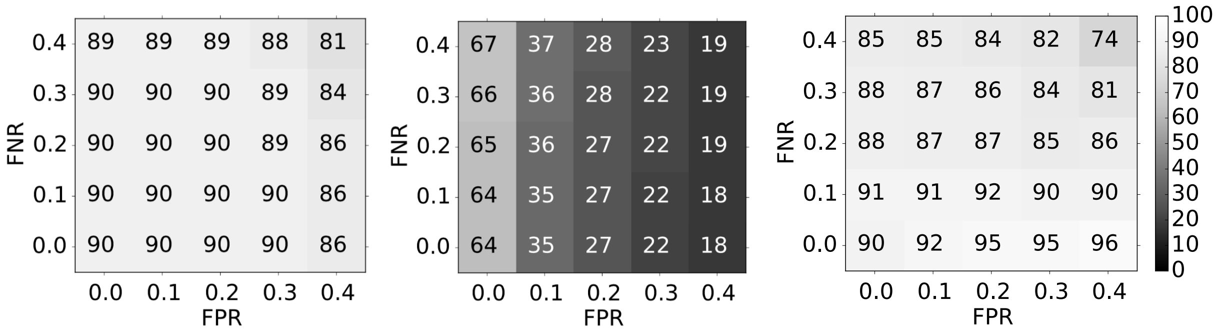
Specification Parameter	Value(s)
Number of Nodes	97, 3,358, 12,523
Number of Injection Scenarios	5, 30, 30
Number of Sensors	5, 30, 130
Injection Start Time	12 hours into simulation
Injection Length	10 hours
Measurement Start Time	10 hours before detection
Measurement End Time	2 hours after detection
Sensor Frequency	2 measurements per hour
Time Horizon	12 hours
Total Number of Test Cases	65

List of Figures



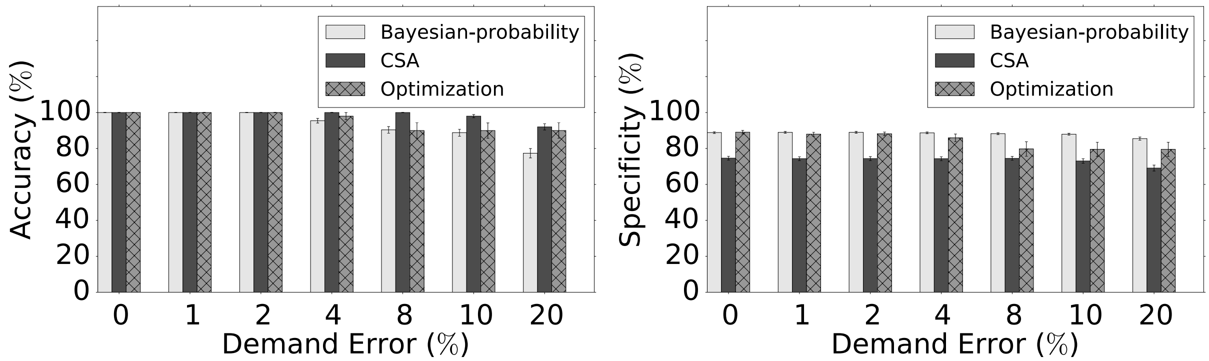
(a) Probability-based (b) CSA (c) Optimization-based

FIG. 1: Mean accuracy of the three SI methods as a function of FPR and FNR calculated over 5 injection nodes and 50 samples.



(a) Probability-based (b) CSA (c) Optimization-based

FIG. 2: Mean specificity of the three SI methods as a function of FPR and FNR calculated over 5 injection nodes and 50 samples.



(a) Accuracy

(b) Specificity

FIG. 3: Mean accuracy and specificity of the three SI methods as a function of demand error calculated over 5 injection nodes and 50 samples. The error bars represent ± standard deviation of the mean.

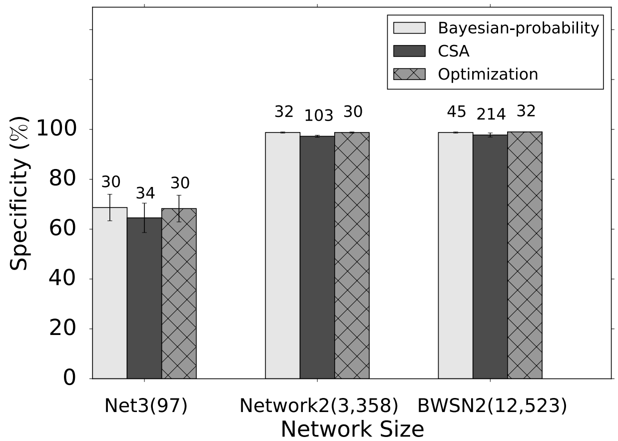


FIG. 4: The effect of network size on the performance of all three SI methods. Each bar represents a mean specificity over the number of injection locations provided in Table 2. Error bars represent ± standard deviation of the mean. The number above each bar represents the absolute specificity value (i.e., the number of nodes with higher or equal likeliness to the true injection node).

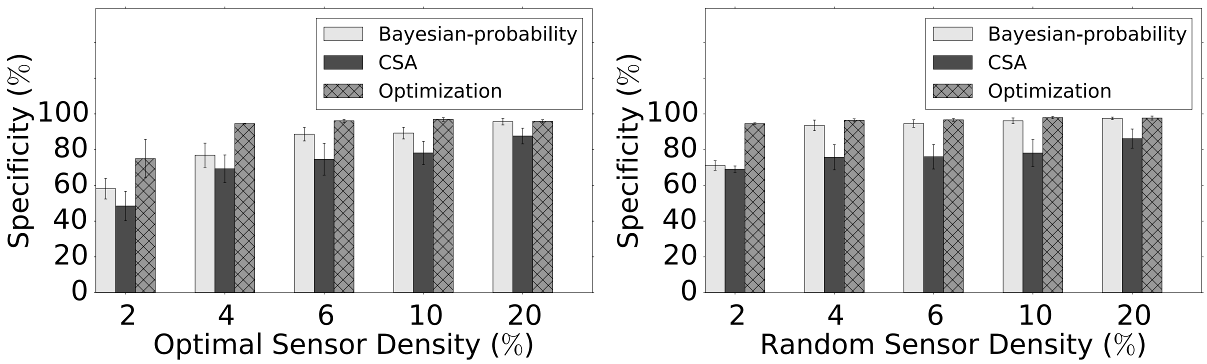


FIG. 5: The effect of sensor density and sensor placement on the specificity of all three SI methods using Net3 (97 nodes). Each bar represents the mean specificity over 5 different injection locations. Error bars represent ± standard deviation of the mean.

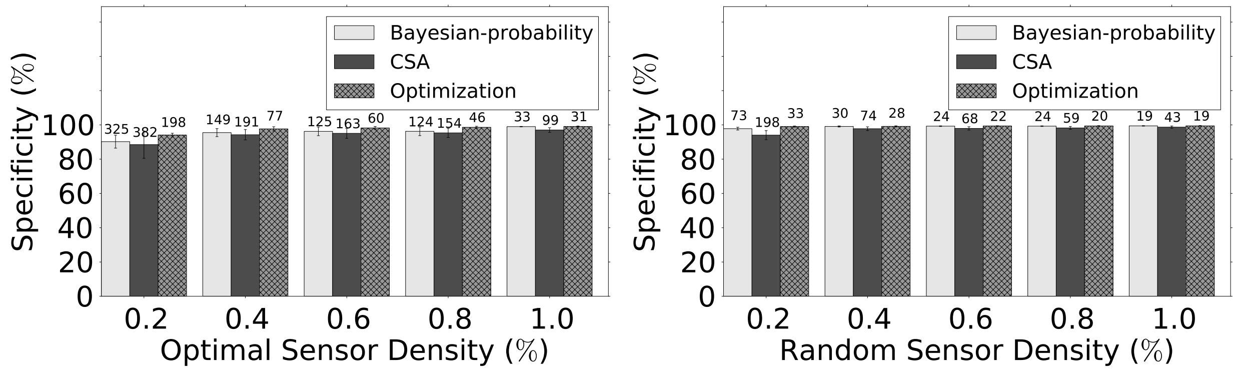


FIG. 6: The effect of sensor density and sensor placement on the specificity of all three SI methods using Network2 (3,358 nodes). Each bar represents the mean specificity over 5 different injection locations. Error bars represent ± standard deviation of the mean. The number above each bar represents the absolute specificity value (i.e., the number of nodes with higher or equal likeliness to the true injection node).