Validation and Assessment of Integer Programming Sensor Placement Models

Jonathan W. Berry* William E. Hart* Cynthia A. Phillips*
James G. Uber† Jean-Paul Watson*

Abstract

We consider the accuracy of predictions made by integer programming (IP) models of sensor placement for water security applications. We have recently shown that IP models can be used to find optimal sensor placements for a variety of different performance criteria (e.g. minimize health impacts and minimize time to detection). However, these models make a variety of simplifying assumptions that might bias the final solution. We show that our IP modeling assumptions are similar to models developed for other sensor placement methodologies, and thus IP models should give similar predictions. However, this discussion highlights that there are significant differences in how temporal effects are modeled for sensor placement. We describe how these modeling assumptions can impact sensor placements.

1 Introduction

Public water distribution systems are inherently vulnerable to accidental or intentional contamination because of their distributed geography. The use of on-line, real-time early warning systems (EWSs) is a promising strategy for mitigating these risks. The general goal of an EWS is to identify a low probability and high impact contamination incident while allowing sufficient time for an appropriate response that mitigates any adverse impacts. An EWS complements conventional routine monitoring by quickly providing information on unusual threats to a water supply. Although several European countries have deployed EWSs to monitor riverine water supplies (Drage et al., 1998; Schmitz et al., 1994; Stoks, 1994), relatively few systems have been deployed for U.S. water supplies.

A key element of the design of an effective EWS is the placement of sensors throughout the water network. A variety of technical approaches have been developed to formulate and solve sensor placement problems in water networks, including integer programming models (Berry et al., 2004, 2005; Lee et al., 1991;

*Discrete Algorithms and Math Dept, Sandia National Laboratories, Mail Stop 1110, P.O. Box 5800, Albuquerque, NM, 87185-1110; PH (505)844-2217, (505) 845-7296; {jberry,wehart,caphill,jwatson}@sandia.gov. Sandia is a multipurpose laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

†USEPA, Cincinnati, OH and Dept. of Civil Engineering, University of Cincinnati, Cincinnati, OH; PH (513)569-7974 uber.jim@epa.gov

Lee and Deininger, 1992; Watson et al., 2004), combinatorial heuristics (Kessler et al., 1998; Kumar et al., 1999; Ostfeld and Salomons, 2004), and general-purpose metaheuristics (e.g. Ostfeld and Salomons (2004)). Integer programs can often be solved to optimality in practice, thereby provide the ability to ensure that the best solution is found. However, formulations like integer programs generally rely on simplifying assumptions to limit the number of design parameters and to enable the model to be solved with a particular method.

In this paper, we reconsider two integer programming formulations that we have recently proposed for sensor placement in water networks. The *static* model proposed by Berry et al. (2005) simply considers whether an attack can reach a downstream population, while the *dynamic* model proposed by Berry et al. (2004) uses the temporal dynamics of contaminant flow to determine whether a downstream population is affected before the contaminant is detected. We critique these models by considering the type of simplifying assumptions that they make, and we contrast these sensor placement formulations with models used by other sensor placement techniques. Further, we analyze the types of solutions that will be generated by the static and dynamic formulations, and note practical trade-offs in the solution of these problem formulations. Empirical comparisons of these sensor placement formulations on real-world water networks confirm the types of differences that we predict.

The general conclusion that we reach with these results is that both of these integer programming formulations are worthy of further consideration. These sensor placement formulations make simplifying assumptions that are similar to models that have been optimized with heuristic search methods. Thus, we should generally expect similar solutions. We have observed that static IP formulations are often easier to solve than dynamic formulations, which suggests that these formulations can be more effectively applied to large-scale problems. Although static formulations can only approximately capture temporal aspects of contaminant flows, our experiments show that they can generate solutions whose expected performance is similar to solutions of the dynamic formulations.

2 Comparing Sensor Placement Formulations

Sensor placement problems can be naturally formulated as optimization problems. Although our focus is on detecting contaminant events within an EWS, methodologies for placing water quality monitoring stations are directly related to sensor placement problems for EWS design. Consequently, we include them in our comparison of modeling approaches, and for simplicity of presentation we refer to the placement of water quality monitoring stations as a sensor placement problem.

For EWS design, the goal of a sensor placement optimization formulation is simple: to place a limited number of sensors in a water distribution network such that the impact to public health of an accidental or intentional injection of contaminant is minimized. However, no specific, concrete formulation for sensor placement has emerged that is widely accepted by the water community. There are a wide range of alternative objectives that are also important when considering sensor placements, such as minimizing the cost of installing and maintaining the sensors, minimizing the response time to a contamination event, and minimizing the extent of contamination (which impacts the recovery costs). Additionally, it is difficult to quantify the health impact of a contamination event because human water usage is often poorly characterized, both in terms of water consumption patterns, as well as how the water consumption impacts health effects. Consequently, surrogate measures like the total volume of water consumed at all sites have been used to model health impacts; this measure assumes that human water consumption is proportional to water consumption at all junctions within the network.

One common feature of sensor placement formulations is the simplifying assumption that sensors can accurately measure water quality and/or the presence of contaminants. Although this may be reasonable for water quality measurements, it remains unclear how well this assumption will apply to EWS design activities. New sensor technologies are needed to detect contaminant threats, but these robustness and accuracy of these contaminant-specific sensors remains unclear.

Sensor placement formulations can also be categorized by the manner in which contaminant events are modeled. In a *dynamic* sensor placement formulation, the impact of a contamination event at a network junction is determined exactly, using a detailed water quality simulation to compute contaminant concentration time-series for each junction in the network. In a *static* sensor placement formulation, the impact of a contamination event is <u>estimated</u> by analyzing some combination of (1) flow directions and velocities obtained via hydraulic simulation, (2) pipe lengths, and (3) junction demands. This categorization reflects the fact that dynamic formulations capture temporal dynamics, while static formulations assume one or more patterns of water flow separately. Although water network models typically describe water flow with a set of flow patterns, static sensor placement formulations do not model transitions between these flow patterns.

2.1 Static Formulations

Almost all of the research on sensor placement for water networks has considered static sensor placement formulations. These sensor placement formulations are distinguished by (a) the objective used for optimization and (b) how network flows are modeled. An objective related to set covering was developed by Lee et al. (1991; 1992) and subsequently refined by several researchers (see Ostfeld and Salomons (2004) for a review). Sensor placement objectives are motivated by the observation that water quality measurements at a sensor reflect the quality of water at nearby points within the network. Specifically, network flow information is used to compute a matrix that determines what fraction of water flow passes through each junction, and this information is used to find sensors that maximize the coverage of water flow.

Kessler et al. (1998) and Ostfeld and Kessler (2001) have developed a static sensor placement formulation whose objective is to ensure that a pre-specified maximum volume of water consumed prior to detection can be guaranteed. This objective is solved using a set covering formulation, where sensors cover junctions for which detection can be guaranteed within the pre-specified "level of service". This is a static sensor placement formulation, so the calculation of contaminant consumption is an estimated quantity. This calculation is performed using an auxiliary network, which has directional edges that are determined via analysis of hydraulic simulation outputs. A directed edge is added in this network from junction v_i to v_j if there is flow from v_i to v_j at any point in the simulation. These directed edges are weighted by the *average* velocity from v_i to v_j over the course of simulation, which allows the use of this auxiliary graph (along with network pipe lengths) to estimate the shortest travel time between all pairs of vertices in the original water network.

Finally, Berry et al. (2003; 2005) and Watson et al. (2004) describe a variety of static sensor placement formulations that are formulated as integer programs (IPs). The objective of the IP described by Berry et al. (2003; 2005) is to minimize the expected fraction of the population exposed to a contamination event. Hydraulic simulation results are used to compute a fixed flow orientation for each pipe in the network over a series of p distinct non-overlapping time intervals, referred to as patterns. A population consuming water at node v_j is considered exposed to a contamination from vertex v_i there exists a flow path from v_i to v_j along which there is no sensor. Watson et al. (2004) generalize this formulation to consider a range of optimization objectives, some of which account for travel times by considering contaminant propagation within each flow pattern separately (instead of aggregating these into an auxiliary network as is done by Kessler et al. (1998)).

2.2 Dynamic Formulations

The previous static sensor placement formulations cannot effectively model time varying flow characteristics like contaminant dilution, concentration level, and transport interactions. Instead, these models simply track the presence or absence of contaminant at various network points. The rate of contaminant flow can be modeled, but these flow calculations are only approximate because they do not account for temporal variations (e.g. the effects of shifts between flow patterns).

By contrast, *dynamic* sensor placement formulations precisely characterize the impact of a contamination event on the rest of the network. Dynamic SPOP formulations use detailed water quality simulations to compute contaminant concentration time-series for each junction in a network. These time-series can be used to exactly determine the impact of any contamination event, including how contaminant impacts junctions through the network (e.g. how much contaminant is consumed at every junction).

Ostfeld and Salomons (2004) propose a dynamic sensor placement formulation whose objective is ensure that the expected impact of a contaminant event is within a pre-specified level of service. In their model, this level of service is a maximum volume of water consumed prior to detection that is above a given contaminant concentration level. Mirroring the earlier approach of Kessler et al., Ostfeld and Salamons solve this objective using a set covering formulation, where sensors cover junctions for which detection can be guaranteed within the pre-specified level of service. The calculation of contaminant consumption is computed using PiplineNet (Bahadur et al., 2003).

We have developed a similar dynamic sensor placement formulation for minimizing the expected volume of contaminated water consumed before detection (Berry et al., 2004). This formulation is expressed and solved as a mixed-integer program. Contaminant flow is modeled with a discrete event simulator in our analysis, though we have subsequently demonstrated the use of EPANET for more general water quality simulations.

3 Comparing Static and Dynamic Formulations

In the previous section, we described how dynamic sensor placement formulations can be used to more accurately model the impact of contamination events. Dynamic formulations have the added advantage that a full range of attack types and sensor characteristics can be modeled; the accuracy of the formulation is strictly limited by the accuracy of the water quality simulation.

However, the accuracy of dynamic sensor placement formulations comes with a price, specifically in the form of a very large number of expensive water quality simulations. For large-scale applications, the computation of these quality simulations is a clear bottleneck, and storage of the output of this simulation data can be excessive. For example, on a large data set (with 3000+ junctions and pipes), the size of our model input data is over 100MB. Furthermore, using water quality simulation data for dynamic sensor placement optimization may require the use of a high-end workstation. For example, on real-world sensor placement applications (with 10,000+ junctions and pipes), the linear programming relaxation of the dynamic sensor placement IP described by (Berry et al., 2004) requires more than the 4 gigabyte memory limitation of 32-bit workstations.

By contrast, static sensor placement formulations are based strictly on comparatively cheap hydraulic simulations. This contrast begs the question of whether the additional accuracy in the dynamic formulations generates qualitatively better sensor placement results. Although we have noted that there are many ways that a dynamic formulation can more accurately model contaminant impacts, the approximations made by a static formulation may be reasonable given the fidelity of data used for sensor placement. For example, in several large real-world datasets, we have noted that the direction of water flow never changes on a large percentage of pipes. This suggests the there are may not be many drastic changes in contaminant flow within some water networks.

Thus it is clear that a better understanding of the differences between static and dynamic problem formulations for sensor placement problems could be of significant practical utility. In the next two sections, we consider a specific comparison between static and dynamic sensor placement formulations for the objective of minimizing the number of junctions contaminated within a network. This

$$
\text{(Static)} \quad \text{minimize} \quad \frac{1}{nP} \sum_{i=1}^{n} \sum_{p=1}^{P} \sum_{j=1}^{n} c_{ipj}
$$

$$
\text{where} \quad \begin{cases} c_{ipi} = 1 & \forall i = 1 \dots n, p = 1 \dots P \\ c_{ipj} \geq c_{ipk} - s_j & \forall i = 1 \dots n \\ \sum_{i=1}^{n} s_i \leq S_{\max} & \\ s_i \in \{0, 1\} & \forall i = 1 \dots n \end{cases}
$$

Figure 1: The IP formulation of the static sensor placement formulation to minimize the number of junctions that are contaminated before detection.

$$
\text{(Dynamic)} \quad \text{minimize} \quad \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}}^{n} \sum_{i \in \mathcal{L}} w_{ai} b_{ai}
$$

$$
\text{where} \quad \begin{cases} \sum_{i \in \mathcal{L}} b_{ai} = 1 & \forall a \in \mathcal{A} \\ b_{ai} \leq s_i & \forall a \in \mathcal{A}, i \in \mathcal{L} \\ \sum_{i=1}^{n} s_i \leq S_{\max} & \\ s_i \in \{0, 1\} & \forall i \in \mathcal{L} \end{cases}
$$

Figure 2: The IP formulation of the dynamic sensor placement formulation to minimize the number of junctions that are contaminated before detection.

objective was chosen because it seems like an objective that would be minimally by temporal effects; specifically, the value of this objective does not directly involve the rate of contaminant flow (e.g. as does an objective like the volume of contaminant consumed).

3.1 IP Models

In this section we describe the static and dynamic formulations that we consider in our experiments. Figure 1 shows the static IP formulation for sensor placement to minimize the number of junctions that are contaminated before detection, and Figure 2 shows corresponding dynamic IP formulation. In both models, the decision variables $s_i = 1$ if we place a sensor on junction i, and 0 otherwise. A sensor at a junction detects contaminants moving in any direction through that junction. Both sensor placement formulations can use a bounded number of sensors, $S_{\max}$.

The static formulation in Figure 1 considers each of the P flow patterns separately. This formulation uses auxiliary variables c_{ipj} to indicate whether junction j has been contaminated from an attack at junction i in flow pattern p.

The dynamic formulation in Figure 2 considers entire the temporal flow pattern. Consequently, the times and locations $\mathcal{A}$ of contaminant injection events are explicitly defined. The auxiliary variables b_{ai} indicated whether a sensor at location i would raise an alarm for a contamination from a (note that a refers to both the time and location of a contamination event). The set $\mathcal{L}$ denotes the set of all possible sensor locations, in addition to a dummy location (which is used to capture the event that an attack is not detected). Finally, the values w_{ai} are the precomputed impact of a contamination event at a if the contaminant is first detected at i.

3.2 Experimental Comparisons

We apply and evaluate our dynamic and static formulations on two real-world datasets:

- **Dataset SNL-4**: A dataset adapted from a local area network. This network has approximately 450 nodes and 600 pipes, with a small number of pumps and tanks.
- **Dataset SNL-5**: A dataset adapted from a moderately large southwestern city. This network has approximately 3500 junctions and 3800 pipes, with tanks and wells spread throughout the water network.

For each of these datasets, EPANET 2.0 (Rossman, 1999) was used to calculate the flow directions for the attack scenarios. Furthermore, the impact of contamination events was performed to compute the w_{ai} data used by the dynamic model; no water quality data was used in these calculations. Four flow patterns were used in the static model, and for the dynamic model contamination events were set up at four uniformly spaced times throughout a 24 hour period. The optimization formulations are implicitly setup to consider contamination events at all junctions within these water networks, with uniform likelihood of all contamination events.

We used the AMPL modeling language (Fourer et al., 2002) to formulate the static and dynamic IPs, and these IPs were solved on a 64-bit Linux workstation using the CPLEX 9.0.2 IP solver. Table 1 summarizes the performance of the static and dynamic formulations on these two datasets. This table shows the final value of the sensor placement found for these two formulations. The value of solutions to the static formulation are estimates of their values in the dynamic formulation. Thus this table includes results that validate the value of these solutions in the dynamic model.

The results in Table 1 suggest that the static model may be able to identify near-optimal sensor placements to dynamic formulations. Except for the zero-sensor case for SNL-5, the static model's predictions are reasonably close to both the validated values, as well as the values of the optimal solutions for the dynamic formulation. Figures 3 and 4 provide additional information about the difference between the predict impact and the validation impact calculation (which considers the performance of the static formulation's solution within the dynamic model, for an arbitrary set of attack times). Note that these figures rescale the densities using a log-scale to enhance the contrast within these distributions, so in fact these distributions are even more biased towards the dark regions of these figures than the shading suggests.

Both Figure 3b and 4b indicate that sensors can be placed such that the predicted impact from the static formulation is close to the predicted impact from

SNL-4	Static		Dynamic
Num Sensors	Predicted Value	Validated Value	Optimal Value
0	39.37	30.34	30.34
20	7.70	3.95	2.79
SNL-5	Static		Dynamic
Num Sensors	Predicted Value	Validated Value	Optimal Value
0	308.11	529.56	528.06
100	8.57	7.97	6.57

Table 1: Optimal solutions for the static and dynamic sensor placement formulations on the SNL-4 and SNL-5 datasets.

the dynamic formulation. Although there is considerable spread in these distributions, the variance is small because the distributions are strongly biased towards the diagonal lines, where these predictions are equivalent. We can interpret these results to suggest that sensors can be placed such that temporal impacts can be largely ignored, because the sensors detect contaminant events sufficiently quickly.

The results in Figures 3a and 4a consider the case where there are no sensors, and as we might expect these results are not as consistent. The predictions of the static formulation in SNL-4 are close to the optimal value in the dynamic formulation, which we believe is due to the fact that there are few flow changes in this model. However, there as many as one quarter of all pipes within SNL-5 exhibit flow changes within a 24 hour period. The static formulation fails to adequately capture the impact of these flow changes, and thus its predictions are poor in this case. However, even though the *predicted* value is qualitatively different from the value in the dynamic formulation, the validated value of the static formulation's optimal solution is close to the value of the dynamic formulation's optimal solution. Thus this formulation may provide near-optimal solutions even when its predictions are skewed by temporal effects.

4 Conclusions

The initial motivation for this work was the need to more carefully compare and contrast the IP models developed by Berry et al. (2003; 2004; 2005) and Watson et al. (2004) with other sensor placement formulations. Our discussion in Section 2 highlights the fact that these IP formulations make the same type of simplifying assumptions as previous work. Further, this comparison highlighted the distinction between dynamic and static sensor placement formulations, which was the focus of our experimental studies.

Our empirical comparisons consider a “best case” scenario for the comparison of static and dynamic sensor placement formulations, since the number-of-junctions-contaminated metric does not depend on water quality values. Our experiments confirm what we have seen in practice: that values of static sensor

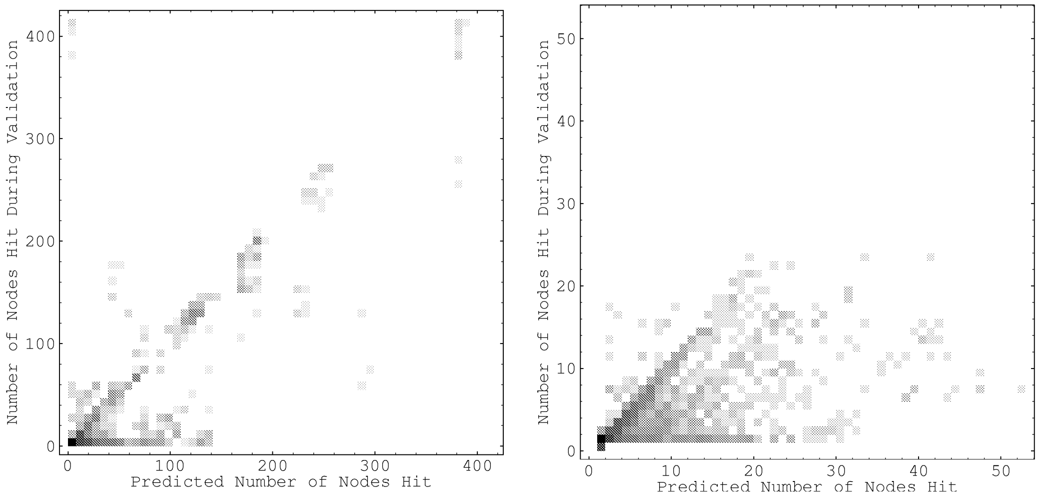

Figure 3: Analysis of the optimal solution for the static formulation on dataset SNL-4 with (a) zero sensors and (b) 20 sensors. For each contaminant event, these comparisons plot the predicted impact vs. a validated estimate of the true impact.

placement formulations appear to be near-optimal in dynamic sensor placement formulations. Our results clearly indicate that static formulations need to be applied with care, as temporal effects can play an important role in interpreting their value. However, it should be noted that the SNL-5 may be an extreme case for temporal effects: this model concerns the core water distribution network of a municipality with many water sources (wells and tanks) spread throughout the network. For water networks that are fed from a small number of localized sources (e.g. river or reservoir), static formulations may be quite reliable.

Finally, we note that our discussion and empirical comparisons do not address the more general question of whether the objectives considered in these IP formulations are well suited for practical applications. As we noted in Watson et al. (2004), many of the objectives of interest for sensor placement formulations are competing formulations. Thus, we should not in general expect that any one sensor placement objective would be strictly preferable to all others. However, our results also indicate that near-optimal trade-offs of different objectives can be achieved, so that many different objectives might be nearly optimal for a given sensor placement.

References

R. Bahadur, W. B. Samuels, W. Grayman, D. Amstutz, and J. Pickus. Pipelinenet: A model for monitoring introduced contaminants in a distribution system. In *Proc., World Water & Environmental Resources Congress 2003 and Related Symposia*. ASCE, 2003. CD-ROM.

J. Berry, L. Fleischer, W. E. Hart, and C. A. Phillips. Sensor placement in

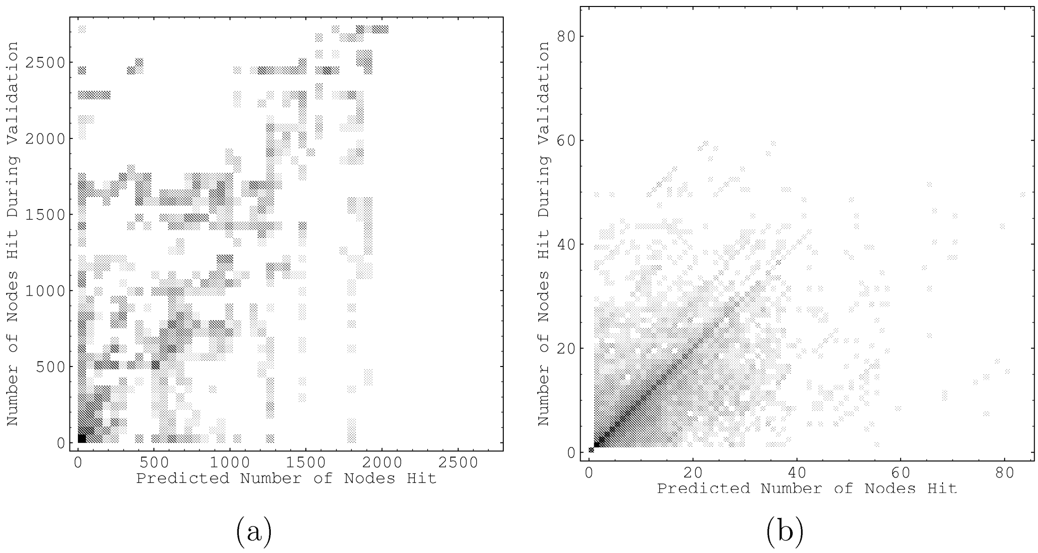

Figure 4: Analysis of the optimal solution for the static formulation on dataset SNL-5 with (a) zero sensors and (b) 100 sensors. For each contaminant event, these comparisons plot the predicted impact vs. a validated estimate of the true impact.

municipal water networks. In P. Bizier and P. DeBarry, editors, *Proceedings of the World Water and Environmental Resources Congress*. American Society of Civil Engineers, 2003.

J. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J.-P. Watson. Sensor placement in municipal water networks. *J. Water Planning and Resources Management*, 2005. (to appear).

J. Berry, W. E. Hart, C. A. Phillips, and J. Uber. A general integer-programming-based framework for sensor placement in municipal water networks. In *Proceedings of the World Water and Environment Resources Conference*, 2004.

B. E. Drage, J. E. Upton, and M. Purvis. On-line monitoring of micropollutants in the river trent (UK) with respect to drinking water abstraction. *Water Science and Technology*, 38(11):123–130, 1998.

R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming.* Brooks/Cole, Pacific Grove, CA, second edition, 2002.

A. Kessler, A. Ostfeld, and G. Sinai. Detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management*, 124(4):192–198, 1998.

A. Kumar, M. L. Kansal, and G. Arora. Discussion of 'detecting accidental contaminations in municipal water networks'. *Journal of Water Resources Planning and Management*, 125(4):308–310, 1999.

B. H. Lee and R. A. Deininger. Optimal locations of monitoring stations in water distribution system. *Journal of Environmental Engineering*, 118(1):4–16, 1992.

B. H. Lee, R. A. Deininger, and R. M. Clark. Locating monitoring stations in water distribution systems. *Journal, Am. Water Works Assoc.*, pages 60–66, 1991.

A. Ostfeld and A. Kessler. Protecting urban water distribution systems against accidental hazards intrusions. In *Proceedings IWA Second Conference*. IWA, 2001. CD-ROM.

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

L. A. Rossman. The EPANET programmer's toolkit for analysis of water distribution systems. In *Proceedings of the Annual Water Resources Planning and Management Conference*, 1999. Available at `http://www.epanet.gov/ORD/NRMRL/wswrd/epanet.html`.

P. Schmitz, F. Krebs, and U. Irmer. Development, testing and implementation of automated biotests for the monitoring of the river rhine, demonstrated by bacteria and algae tests. *Water Science and Technology*, 29:215–221, 1994.

P. G. Stoks. Water quality control in the production of drinking water from river water. In M. Adriaanse, J. van der Kraats, P.G. Stoks, and R.C. Ward, editors, *Proceedings: Monitoring Tailor-made*, RIZA, Lelystad, The Netherlands, 1994. (ISBN 9036945429).

J.-P. Watson, H. J. Greenberg, and W. E. Hart. A multiple-objective analysis of sensor placement optimization in water networks. In *Proceedings of the World Water and Environment Resources Conference*, 2004.