11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

ALTERNATIVE STRATEGIES FOR OPTIMAL WATER QUALITY SENSOR PLACEMENT IN DRINKING WATER DISTRIBUTION NETWORKS

PETER VAN THIENEN

Drinking water infrastructure group, KWR Watercycle Research Institute, Groningenhaven 7, 3433 PE Nieuwegein, Netherlands

In this paper, we demonstrate the application of different sensor location optimization strategies in drinking water distribution networks, with aims such as maximization of distribution network coverage with redundancy and optimization of contamination source identification. We present and compare results of these different approaches applied to hydraulic models of a real drinking water distribution network in the Netherlands. The selection of results presented in this paper illustrates that it is important for a water company to decide what its main objective is with regard to water quality sensors before installing them in the field. Different optimization criteria for their spatial configuration result in very different configurations, which may perform well for one purpose but much less so for another. However, some sets of objectives are compatible in the sense that a configuration optimized for one objective also performs well for the other.

INTRODUCTION

The most commonly applied strategies for optimal water quality sensor placement in drinking water distribution systems are based on the philosophy of contamination early warning systems [e.g. 1,2 and references therein]. These strategies aim to minimize the number of people affected in case of a deliberate contamination of drinking water in the distribution system, and provide a valuable tool. A number of factors which are usually not taken into account, including the response strategy to the identification of a contamination event, the fallibility of sensors and changes in network configuration (valve manipulation) and operation, may affect the results of these strategies. Since the quickness and effectiveness of a response is generally also a function of the location of the contamination event (both source and first detection), knowledge on the response strategy should also be part of the sensor placement optimization methodology.

Besides contamination early warning systems, there are several other reasons for placing water quality sensors in distribution network, including process control and monitoring, regulatory monitoring, etc. These may require a different approach to optimization of the sensor network in terms of sensor locations.

In this paper, we demonstrate the application of different sensor location optimization strategies in drinking water distribution networks, with aims such as maximization of distribution network coverage with redundancy and optimization of contamination source identification. We present and compare results of these different approaches applied to hydraulic models of a real drinking water distribution network in the Netherlands.

CONTAMINATION SCENARIOS, OPTIMIZATION CRITERIA AND APPROACH

Methodology

A hydraulic model (EPANET-MSX [3]) is used to calculate the transport of contaminants through a distribution network. Postprocessing of calculated transport from and to all nodes allows a complete consideration of the problems treated here.

Contamination scenarios

In the network model discussed below, a set of potential contaminant injection points is defined, which is a subset constructed by optimizing for an equal number of connections surrounding each potential injection location. An equal probability of contaminant injection is assumed at each potential injection point and at each hour. The same node set is used as the set of potential sensor locations.

Optimization objectives

Detection likelihood and time to first detection

Mean detection likelihood and mean time to first detection are commonly applied objectives for sensor placement optimization, which are included here for comparison. Ostfeld et al. [2] note that detection likelihood and time to first detection are criteria which oppose each other in the sense of sensor location optimization. However, this assertion is closely related to their choice to not include non-detected events in their analysis. When, for example, a large penalty time would be given to each non-detected event, a detection time optimized sensor configuration would move towards a configuration optimized for detection likelihood. A reasonable choice for this penalty time might be the estimated time for a contamination to surface by other means (e.g. customer complaints). However, not all events will surface in this way. A different approach is to take the maximum residence time of the water in the injection point at the time of injection as the penalty time. This approach is followed in this work.

Network Coverage and redundancy

Network coverage is defined here as the fraction of the network which from which water is sampled during a predefined observation window. It is closely related to detection likelihood, but allows the user to choose in what way the fraction of the network is expressed, e.g. in network length, number of connections, number of connected costumers, etc. Redundancy is introduced by demanding concurrent observation of a network segment by at least *n* sensors. Redundancy allows water companies to, for example, start preparatory actions at the first detection and escalate when a confirmation from a second sensor is obtained.

When using a (partially) skeletonized network model, as is the case here, it is important to express network coverage in terms of a parameter which is conserved in the skeletonization process, such as number of connections, rather than a parameter which is not, such as pipe length or volume.

Contamination Source Identifiability

The most important tool for determining the source area of a contaminant is an accurate hydraulic model of the distribution network, in which a contaminant can be traced back in time from its point of observation to all the parts of the network where it might have originated. Several approaches to this backtracing or backtracking have been presented in the literature [e.g. 4,5]. Our approach is based on combination of forward traces. Note that any alternative backtracing algorithm which also takes into account the dynamic flow field renders equally suitable backtraces for the following.

The backtrace of from a single node in the network forms the complete potential area of origin for a contaminant which was observed at this node. This backtrace contains all dynamics and variability of the flow field throughout the day. It does, however, not contain the stochastic variations in demand and resulting variations in the flow field. These may have a significant influence on backtracing in specific parts of the tertiary (reticulation) network (see [6]), but are expected not to be important outside these areas. When one assumes that multiple observations (in space and/or time) of an anomaly by sensors constitute the same contamination event, their combination can be used to narrow down the potential area of origin in the distribution network. This becomes a simple exercise when the backtraces are considered in a binary way. For a set of observations, the set of nodes which constitutes the potential contamination source area S is formed by the intersection of the backtraces of the individual observations T_i for the same and/or different sensor locations at the same and/or different observation times:

$$S = T_1 \cap T_2 \cap T_3 \ \ldots \quad (1)$$

The design objective for this optimization objective, *contamination source identifiability*, is the minimization of the mean minimum potential source area size (pipe length) which can be determined for the contamination scenario set. This means that for each scenario, the minimum non-zero potential source area size is determined (which can occur at any time after the start of the contamination event), and the mean for all scenarios is taken as the performance parameter. Note that this objective is computationally very expensive. Therefore, the results below have been computed for a single contamination time (midnight) at each of the potential source nodes.

Optimization approach

Optimization is performed using a genetic algorithm implemented in the *inspyred* library [7]. In order to verify the performance of the algorithm, several comparisons against brute force global optimizations have been performed for simple networks.

RESULTS

Distribution network

The approaches described above are applied to a drinking water distribution network model, see Figure 1. It is a skeletonized version (306 km of pipe) of a distribution network in The Netherlands (part of the Vitens Innovation Playground [8]) of in total 612 km of pipe.

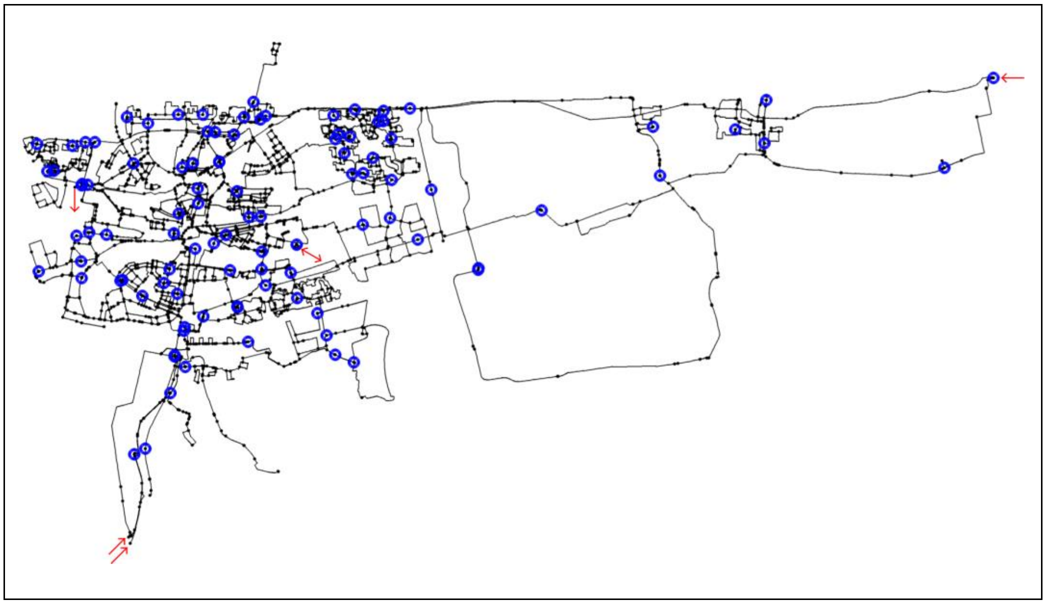

Figure 1: Skeletonized distribution system with in and outgoing water flows (red arrows) and 100 potential contamination/sensor locations (blue circles).

Sensor networks and performance

Overview

For all optimization criteria discussed above, optimal sensor configurations have been computed for the test network. For all of these optimal configurations, the performance of the sensor network with respect to each of these criteria has been determined. The results are presented in Table 1, and discussed in more detail in the following paragraphs. A uniformly distributed (over the connections) set of 15 sensor locations has been included for comparison.

Table 1: Overview of sensor network design performances for the objective they were optimized for and the other objectives. The uniform configuration is non-optimized. Higher performance values are better for the detection likelihood and coverage, lower is better for time to first detection and contamination source identifiability (CSI). Optimized configuration labels are: DL: detection likelihood, TFD: time to first detection, CR1-3: coverage with redundancy, CSI: contamination source identifiability.

		performance with respect to objective					
	scenario	*detection likelihood*	*time to first detection (hours)*	*coverage and redundancy (fraction of all connections)*			*CSI (km)*
				n=1	*n=2*	*n=3*	
applied optimal sensor configuration	DL	0.50	9.22	0.54	0.34	0.26	262
	TFD	0.45	9.13	0.52	0.40	0.22	272
	CR1	0.39	11.80	0.62	0.33	0.21	272
	CR2	0.40	11.11	0.55	0.48	0.29	263
	CR3	0.38	10.28	0.41	0.26	0.31	268
	CSI	0.43	9.67	0.47	0.30	0.16	238
	uniform	0.26	12.48	0.44	0.25	0.11	285

It is clear from the table that almost all optimized designs (the exception is design CR3 for objective *coverage n=1*) perform better on each of the individual objectives than the uniformly distributed configuration.

Detection likelihood and time to first detection

Optimal sensor locations for 15 sensors with respect to mean detection likelihood and two different approaches to the mean time to first detection are shown in Figure 2. When non-detections are ignored, i.e. when events that are not observed by any sensor do not contribute in a negative way to the performance indicator of the network (cyan circles in Figure 2), a configuration with many sensors close to the water sources and transport mains is found. When non-detections contribute a penalty detection time which is equal to the local maximum residence time (large magenta circles in Figure 2), a configuration which is much closer to that of the maximum detection likelihood configuration is found, both geometrically and in terms of performance for both objectives (see Table 1). This shows that these objectives do not necessarily oppose each other, as was noted in [2].

Network coverage and redundancy

The sensor configurations optimized for coverage in terms of numbers of connections with 15 sensors and single or multiple (n=2, n=3) redundancy for the test network and their performance are shown in Figure 3. Requiring (multiple) redundancy in the sensor network results in a very clear contraction of the configuration and the covered area.

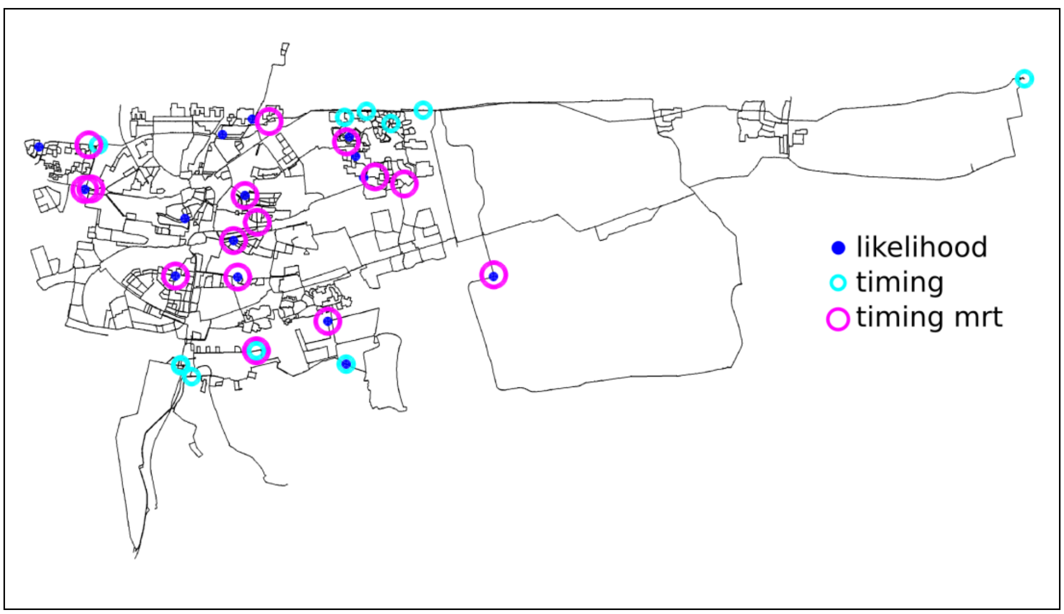

Figure 2: Optimal sensor configurations for 15 sensors with optimization for detection likelihood (*likelihood*, blue dots), time to first detection ignoring non-detected events (*timing*, small cyan circles), and time to first detection using the local maximum residence time at the contamination site as time penalty in case of non-detection (*timing mrt*, large magenta circles).

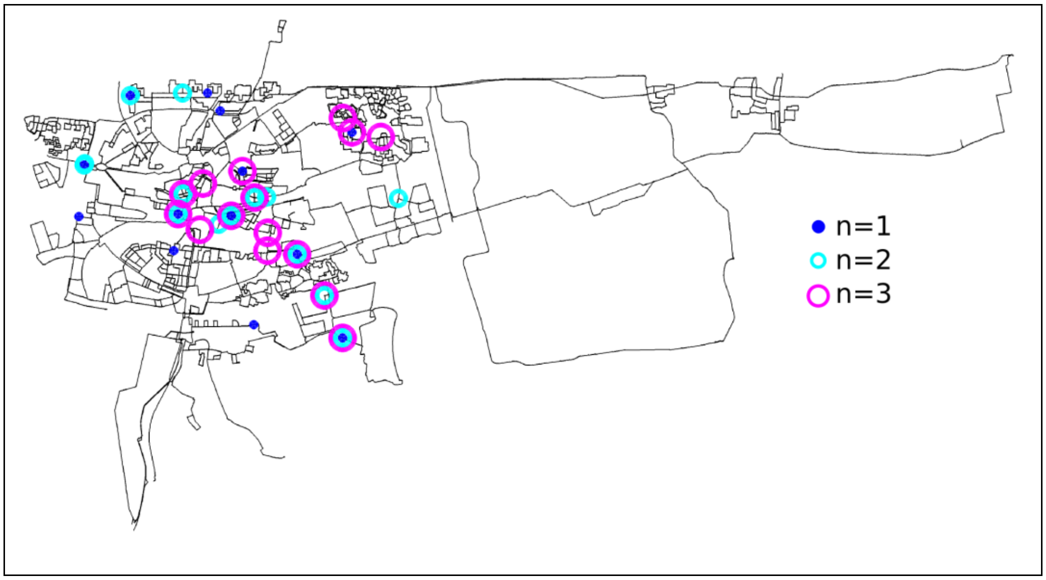

Figure 3: Optimal sensor locations for maximum network coverage in terms of numbers of connections without (*n=1*, blue dots) and with redundancy (*n=2*, cyan circles, *n=3*, large magenta circles).

Contamination source identification

The sensor configuration for 15 sensors optimized for contamination source identifiability is shown in Figure 4 (blue dots). For comparison, the optimal configuration for the detection likelihood objective is also shown (cyan circles). The two configurations are relatively close, both in terms of sensor locations and in terms of performance with respect to the different objectives.

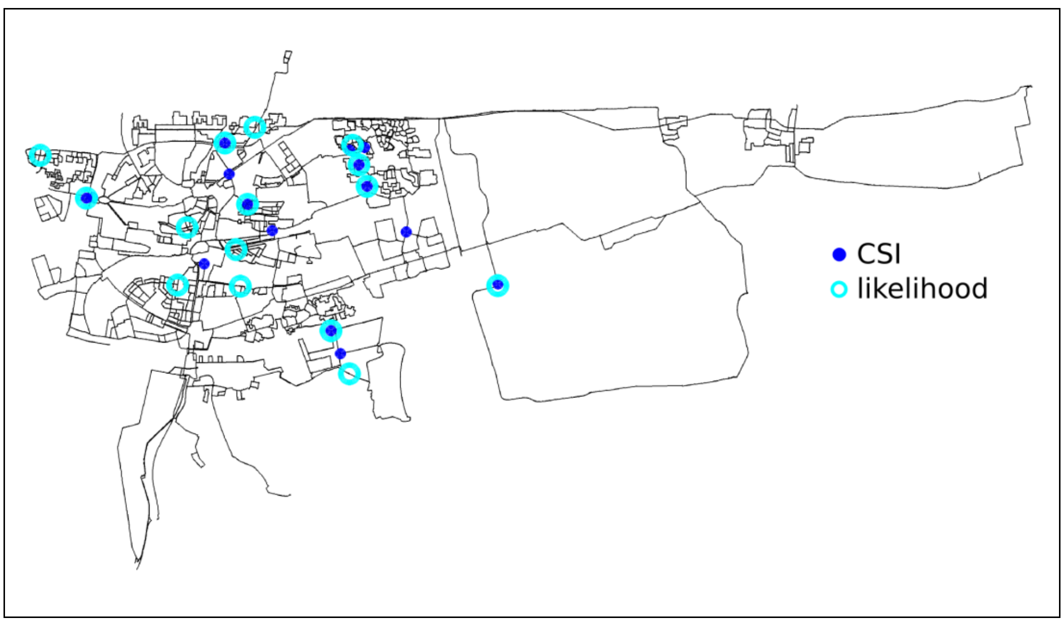

Figure 4: Optimal sensor locations for maximum contamination source identifiability (CSI) and detection likelihood.

The quantitative values of the CSI performance of the different designs listed in Table 1 do not appear to be very different. However, these scores are diluted by the assignation of the full network as the potential source area in case of non-detection.

DISCUSSION AND CONCLUSIONS

Detection likelihood and network coverage appear, at first glance, to be the same parameter. However, detection likelihood has a focus on the network, whereas network coverage can be defined to have a focus on the consumers. This results in a different configuration, with a reduced performance for the other objective. Requiring redundancy in the network coverage may have its clear uses for water companies (for example, it allows water companies phase their response to detected contaminants as a function of the number of detections), but obviously the number of connections which can be covered with the same number of sensors is much lower.

Minimizing the time to first detection is a very useful approach when one want to protect a population. Using the maximum residence time for an individual node as a penalty time when optimizing for this objective presents a more balanced approach than ignoring non-detections, as has been done in the past. The resulting sensor configuration is quite different, and much closer to that for the detection likelihood objective.

The optimization for contamination source identifiability is a novel criterion, which directly connects to mitigative and corrective measures taken in case of a contamination. This criterion results in a qualitatively similar sensor configuration compared to the detection likelihood objective for the case which was studied here.

This selection of results illustrates that it is important for a water company to decide what it wants to get out of water quality sensors before installing them in the field. Different optimization criteria for their spatial configuration result in very different configurations, which may perform well for one purpose but much less so for another, as is illustrated in Table 1. However, some sets of objectives are compatible in the sense that a configuration optimized for one objective also performs well for the other.

Hydraulic models generally play a central role in the optimization of sensor placement. The validity of their computations strongly depends upon accurate and up to date information on the network, which is often not fully available (e.g. unregistered valve status changes). This is a point of concern, which requires attention.

ACKNOWLEDGMENTS

This research was performed in the framework of the joint research program of the Dutch drinking water companies (BTO). Contributing ideas by and discussions with Mirjam Blokker, Bendert de Graaf, Maurice van de Roer, Peter Schaap and Veerle Sperber are gratefully acknowledged.

REFERENCES

[1] Berry J., Boman E., Riesen L.A., Hart W.E., Phillips C.A. , Watson J.P, and Murray R. *"User's Manual TEVA-SPOT Toolkit Version 2.4."* Office of Research and Development, U.S. Environmental Protection Agency, EPA 600/R-08/041B, (2010).

[2] Ostfeld, A. , Uber J.G, Salomons E., Berry J.W., Hart W.E, Phillips C.A., Watson J.P., Dorini G., Jonkergouw P., Kapelan Z., di Pierro F., Khu S.-T., Savic D., Eliades D., Polycarpou M., Ghimire S., Barkdoll B.D., Gueli R., Huang J.J., McBean E.A., James W., Krause A., Leskovec J., Isovitsch S., Xu J., Guestrin C., VanBriesen J., Small M.,

Fischbeck P., Preis A., Propato M., Piller O., Trachtman G.B., Wu Z.Y. and Walski T. "*The Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms*", J. Water Resour.Plann. Manage., 134(6), (2008), pp. 556-568.

[3] Shang F., Uber J. G. and Rossman L. "*EPANET Multi Species Extension Software.*" U.S. Environmental Protection Agency ,Washington D.C., EPA/600/C-10/002, (2008).

[4] Shang F., Uber J. G. and Polycarpou, M. M. "*Particle backtracking algorithm for water distribution system analysis".* Journal of Environmental Engineering, 128 (5), (2002), pp. 441-450.

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

[6] Van Thienen P., Vries D., De Graaf B., Van de Roer M., Schaap P. and Zaadstra E. *"Probabilistic backtracing of drinking water contamination events in a stochastic world"*, 12th International Conference on Computing and Control for the Water Industry,(2013).

[7] *"inspyred: Bio-inspired Algorithms in Python"*. http://inspyred.github.io, March 12, 2014.

[8] De Graaf B.R. Williamson F., Klein Koerkamp M., Verhoef J.W., Wuestman R., Bajema B., Trietsch E. and Van Delft W. *"Implementation of an innovative sensor technology for effective online water quality monitoring in the distribution network"*, Water Practice & Technology, vol. 7, no. 4, (2012).