Bounds on water quality sensor network performance from design choices and practical considerations

P. van Thienen*, B, de Graaf**, J. Hoogterp**, J. van Summeren*, and A. Vogelaar*

* KWR Watercycle Research Institute, Nieuwegein, The Netherlands
(E-mail: peter.van.thienen@kwrwater.nl; joost.van.summeren@kwrwater.nl; ad.vogelaar@kwrwater.nl)

** Vitens, Zwolle, The Netherlands
(E-mail: bendert.degraaf@vitens.nl; jan.hoogterp@vitens.nl)

Abstract
Because of limitations on traditional approaches to optimal water quality sensor placement in drinking water distribution networks, oriented towards obtaining information and mitigating effects, approaches optimizing the utility’s response to a contamination event merit wider study and application. In this contribution, we study and discuss the performance of these different approaches. We also show that practical considerations may impose significant limitations on the achievable performance of a water quality sensor network. Both aspects should be taken into account when performing a sensor placement optimization for a real drinking water distribution network.

Keywords
DWDS, water quality sensors, sensor networks, placement optimization, practical considerations

INTRODUCTION
Since the start of the 21st century, events in society and technological advances have pushed the development of techniques for online water quality monitoring in drinking water distribution systems, in order to protect customers from incidental and/or intentional contamination of drinking water (e.g. Kroll and King, 2010). Because of budgetary constraints, the number of sensors for online monitoring which can be placed in any system is always limited, and therefore methods have been developed to determine optimal sensor placement (Ostfeld et al., 2008, Berry et al., 2010) within a drinking water distribution network. Optimality is, however, a matter of definitions and requirements. Many objectives have been presented in the literature, and these can be roughly classified into 3 categories, aimed at obtaining information, facilitating utility response, and mitigating the effects of a contamination (Table 1). The literature has mostly focused on the first (Ostfeld et al., 2008) or the last (Berry et al., 2010).

Although many studies have been presented on the optimal placement of water quality sensors, very few water companies in the world have, to our knowledge, actually deployed a water quality sensor network in their distribution network. Vitens water company has done so in their Vitens Innovation Playground (De Graaf et al., 2012). This provides a unique opportunity to study sensor placement in practice.

In this contribution, we compare the three objective classes in the context of this live network, discussing in more detail the response oriented approach, and discuss theoretical and practical considerations concerning the application of optimal sensor placement strategies in a real distribution network.
Table 1. Classification of sensor placement optimization objectives.

<table>
<thead>
<tr>
<th>Objective class</th>
<th>Orientation of optimization strategy towards…</th>
<th>Information</th>
<th>Utility response</th>
<th>Effect mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples</td>
<td>Detection likelihood, time to first detection, network/customer coverage</td>
<td>Redundant detection, identifiability of contamination source</td>
<td>Population affected, ingested volume, number of people above dose threshold</td>
<td></td>
</tr>
<tr>
<td>(Dis) advantages</td>
<td>Simple, but several steps from information to actual customer protection</td>
<td>Close to operational practice</td>
<td>Objective matches final objective of utility, but this objective is complex to compute and the results show a strong dependence on utility response (time)</td>
<td></td>
</tr>
</tbody>
</table>

**APPROACH**

**Methodology and tools**
In order to optimize sensor placement and apply the strategies described in Table 1, we have created a custom tool name Contamination Source Toolkit (Van Thienen, 2014), which implements several information and utility response oriented strategies. The tool uses EPANET-MSX (Shang et al., 2008) to calculate contaminant transport and genetic algorithms implemented in the Inspyred library (Garrett, 2015) in order to perform optimizations. In addition to this, we have used TEVA-SPOT (Berry et al., 2010) for performing optimizations aimed towards effect mitigation.

**Optimization objectives**
In the following, we describe how the optimization objectives listed in Table 1 have been approached in this work (following Van Thienen, 2014, to which the reader is referred for more details).

*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. (2008) 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. In this work, we use the maximum residence time of the water in the injection point at the time of injection as the penalty time.

*Network coverage and redundancy*
Network coverage is defined here as the fraction of the network 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 (see below), 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. Shang et al., 2002, Laird et al., 2006). Our approach is based on combination of forward traces (see Van Thienen, 2014 for more details), but any alternative backtracing algorithm which also takes into account the dynamic flow field renders equally suitable results.

Testing area: Vitens Innovation Playground
Vitens has appointed a dedicated area of its distribution network, the Vitens Innovation Playground (VIP), to the testing of innovative added-value technologies, one aspect of which is application of sensors (sensor hardware, data transfer, algorithms etc.). The VIP is located in the north of the Netherlands and consists of both rural and urban areas. For the present numerical work discussed below, only a part of the VIP with about 600 km of pipes (1/4 of the VIP) was studied, with some skeletonization for performance purposes resulting in a model with about 300 km of pipes. The distribution network in the VIP can be supplied by either a single water source or from the three sources supplying the area. These degrees of freedom are an important tool that can be used in the understanding of sensor response and modelling the distribution network.

RESULTS AND DISCUSSION

Optimization approaches
Of the three classes of sensor placement optimization objectives, the information oriented ones are the simplest to compute (see Ostfeld et al., 2008), requiring only a network model (hydraulics and material transport). The more complex effect oriented approach has been implemented in TEVA-SPOT (Berry et al., 2010) and applied to a network model of the VIP. Many simulations were performed to study the relation between assumed and actual response time of the utility (after which consumption is assumed to have ceased) on the one hand and the performance of the sensor network (reduction of the number of people affected) on the other. Some results are shown in Figure 1. A number of observations can be made:

![Figure 1](image-url)

**Figure 1.** Reduction of the number of people affected by a contamination scenario relative to a situation without sensors as a function of the number of sensors placed (optimized configuration in each case), the assumed response time of the utility in the optimization and the actually realized response time.
sensors are useless for event detection if the response time of the utility is too long;
the law of diminishing returns plays up, resulting in every additional sensor contributing less
to the objective than the previous;
optimizing the sensor network for a long response time results in a deteriorated performance
if the actual response time is shorter.

These observations have also been made in previous studies. However, the clearest message from this
figure, which cannot be stressed often enough, is that the utility response time is a critical parameter
in both sensor placement (which varies for different assumed response times) and in the effect of a
sensor network on the reduction of population exposure. This means that on the one hand, this
optimization approach requires a realistic estimate of the response time, which may be difficult to
obtain and may vary significantly from case to case, and on the other hand that utilities need to realize
a short response time in order for a sensor network to have any practical use as an early warning
system for contaminations.

With a custom made optimization tool Contamination Source Toolkit (see Van Thienen, 2014), sensor
placement optimization calculations were also performed for information and response oriented
objectives. Two response oriented objectives were considered. The first is redundant coverage. This
is a response oriented objective in the sense that a utility will not (fully) act upon a single sensor
signal, but would like some confirmation first. Therefore, this objective maximizes the observability
of events by at least two sensors. The second is contamination source identifiability, which minimizes
the mean potential source area size which can be reconstructed from backtraces of contaminations
from multiple sensors (Van Thienen and Vertommen, 2016). Figure 2 shows the relative performance
of several optimized sensor networks in the VIP for these objectives and the presently installed
configuration of water quality sensors, based on experience and practical considerations. Different
sensor network configurations optimized for different objectives and the actual configuration at that
time are organized in rows. For each configuration, the performance is indicated for three information
oriented objectives and two response oriented objectives. It can be clearly seen that
1. the sensor networks perform best on the objectives they have been designed for, and
2. the original configuration based on insight into network hydraulics, experience, and practical
considerations performs relatively poorly in comparison to numerically optimized
configurations.

In this diagram, there is a positive correlation between optimization towards and performance on the
information oriented objectives detection likelihood and time to first detection, meaning that
configurations optimized for detection likelihood also perform well on time to first detection and vice
versa. This is contrary to earlier results reported in Ostfeld et al. (2008). This difference is the result
of the different treatment of non-detections in our approach.

On the other hand, the response oriented objectives perform relatively poorly on sensor networks
which have been optimized for information oriented objectives and vice versa. The main message
from Figure 2 is therefore that the objective of a sensor network needs to be well defined before
designing the sensor network (and of course that optimization pays off).

Practical considerations
Even though it is tempting to use all network model nodes as potential sensor locations, practical
considerations limit the number of actual suitable locations. Some examples are shown in Figure 3.
The problems encountered include accessibility, the availability of power, the possibility of
transmission of data, etc. In the example of Figure 3a, the optimal location is in a residential street.
Figure 2. Performance of information oriented (maximum detection likelihood, minimum mean time to first detection, maximum network coverage) vs. response oriented (redundancy, contamination source identifiability) sensor network designs, with a comparison with the present non optimized configuration.

Even though (some) other locations on the same pipe in the same street would probably be equally suitable, in any case either a manhole should be created in the street and fitted with electricity and a data connection, or a sensor should be installed in a home in the street, using the owner’s electricity and possibly his broadband connection as well. Also, in order to make sure that the sensor is exposed to a sufficiently large and continuous flow of water, a permanent flow must be created at this home, either by wasting a lot of water, or by bypassing the pipe running through the street. Similar considerations are valid for the second example (Figure 3b, similar situation) and the third (Figure 3c, no buildings available).

When the utility elects to consider only practically suitable locations as potential sensor locations, this may have a significant effect in the overall performance of the network, as is illustrated in Figure 4. This figure shows how optimizing a sensor network configuration based on practically available locations not only results in a performance loss compared to sensor networks of which the sensors can be located basically anywhere. A comparison is made with optimizations for different sets of uniformly distributed nodes as potential locations (300, and all 2700, respectively). It can be seen that for very small and very large numbers of sensors, the results vary significantly between the practical set and the limited set of 300 potential locations, but that for moderate numbers, the differences are small. Note that in some cases, the network based on the practical set of potential locations performs even better than a configuration based on a larger number of uniformly distributed nodes. The practical set may include, by coincidence, very suitable locations which may be absent from the uniformly distributed sets. However, when all 2700 network nodes are considered as potential locations, a significantly better performance can be achieved (grey curve in Figure 4).

Figure 4 also shows (again) that a numerical optimization of the sensor locations is expected to result in a better performing sensor network design than the current. Based on some of the results reported here and other calculations, a number of sensors have been relocated by Vitens since. Analysis of the effects of these translocations on the effectiveness of the sensor network are ongoing.
Figure 3. Some examples of theoretical optimal sensor locations and the associated practical situations.

Nevertheless, the suboptimally positioned sensor network (generic water quality) has been able to detect all known events (change in water hardness, failure of the discoloration at the treatment plant, turbidity increases associated with bursts) during its evaluation period. Thus it has been demonstrated that this sensor network works, but questions remain on optimal sensor density with respect to investment and returns (in terms of safety/security).
Figure 4. Performance comparison (network coverage, percentage of connections) of the non-optimized situation based on experience, an optimized configurations without practical limitations (300 and 2700 potential sensor locations) and a configuration with practical limitations.

CONCLUSIONS
Since the response time is such a critical parameter in the actual effect on the population of a contamination event, sensor placement strategies aimed at optimizing the response strategy, such as verification through redundancy or contamination source identifiability, have the potential of strongly contributing to the operational success of a water quality sensor network in a drinking water distribution system. Practical limitations, such as the availability of power and data connections (or the willingness of the utility to invest in these at every sensor location) determine to what degree an optimal sensor network can be realized. Nevertheless, the sensor network in the VIP has been shown to pick up many water quality events. Combining these practical considerations of sensor placement with numerical evaluation and optimization of the sensor network allows for realistic expectations towards these sensor networks before installation, and ultimately the best protection of the consumers.
REFERENCES


