

PAPER

Measuring imperfections of water quality sensors in water distribution networks

To cite this article: Casper de Winter *et al* 2019 *Meas. Sci. Technol.* **30** 095101

View the [article online](#) for updates and enhancements.

Measuring imperfections of water quality sensors in water distribution networks

Casper de Winter^{1,2}, Venkata Reddy Palleti^{2,5} , Daniel Worm³
and Robert Kooij^{3,4}

¹ Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

² iTrust Centre for Research in Cyber Security, Singapore University of Technology and Design, Singapore

³ Cyber Security and Robustness, TNO, The Hague, The Netherlands

⁴ Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

E-mail: venkata_palleti@sutd.edu.sg

Received 6 June 2018, revised 24 April 2019

Accepted for publication 2 May 2019

Published 19 July 2019



CrossMark

Abstract

Water distribution networks (WDNs) are critical to provide safe, clean drinking water around the globe. However, they are susceptible to accidental or deliberate contamination, potentially resulting in poisoned water, many fatalities and large economic consequences. In order to protect against such intrusions, an efficient sensor network should be placed in a WDN. Finding the optimal placement for water quality sensors is a challenging problem. Several sensor placement strategies have been proposed, but the vast majority of these strategies rely on the assumption that the sensors are perfect. In this paper we provide evidence for the imperfection of water quality sensors, by conducting measurements in an operational environment. We investigate the imperfection of four types of water quality sensors being employed in actual WDNs for the purpose of contamination detection. We describe experiments conducted at the WaDi testbed, a realistic water distribution facility at the Singapore University of Technology and Design. Through these experiments we study the imperfection, sensitivity and degradation of the water quality sensors, under normal conditions (water flow without contaminants present) as well as under attack conditions. It is shown that several aspects of sensor imperfection do occur, including missing values, inexplicable jumps and drifting.

Keywords: water distribution networks, water quality, sensors, contamination detection, imperfect sensors, experiments, testbed

(Some figures may appear in colour only in the online journal)

1. Introduction

Water distribution networks (WDNs) are a critical infrastructure, providing clean, safe drinking water to billions of people around the world. There are several threats to a WDN which can be divided into physical disruptions and chemical disruptions. Physical disruptions, such as leaking pipelines, failing pumps or intentional attacks on the network itself, will have a big economical impact but are not considered a serious risk

to human beings. The biggest threat to a population comes from intentional or accidental chemical contamination within the water network, see for instance [7]. In order to protect the public from such intrusions, it is necessary to monitor the quality of the drinking water in a WDN effectively and efficiently, by deploying appropriate sensors. Most sensors are based upon a small range of chemical and physical parameters, such as pH, ORP and turbidity. However, the use of sensors for these traditional water quality parameters, has several limitations. In particular, electrochemical sensors demand frequent maintenance and calibration, see [16]. One of the

⁵ Author to whom any correspondence should be addressed.

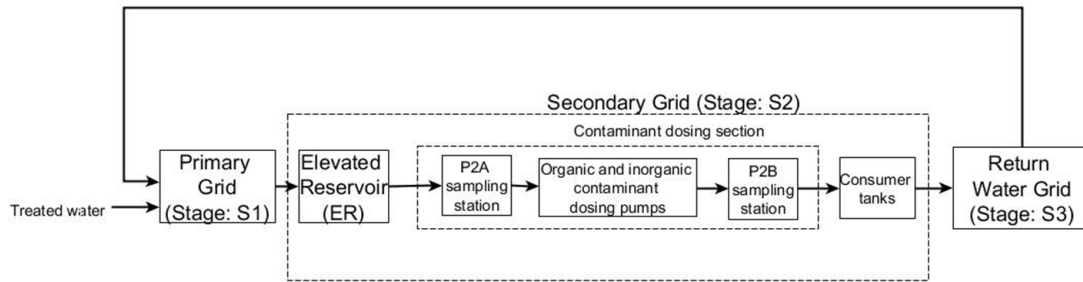


Figure 1. An overview of the WaDi testbed.

challenges with using water quality sensors is to find a proper placement for the sensors. Several sensor placement strategies have been proposed, see for instance [17] and [11]. However, most of these strategies rely on the assumption that the sensors are perfect, i.e. they detect a contamination with a probability of 1. The number of papers that take imperfection of sensors into account for the sensor placement problem in WDNs, is quite limited. For instance, a very recent survey paper, see [8], only mentions one paper, namely [4], that considers sensor imperfection, from a list of 18 papers. Likewise, the survey paper [15], only mentions three papers, [2, 19], and again [4], out of a list of 23 papers that deal with imperfect sensors. At the same time, it is known [2, 5], that the imperfection of sensors has an impact on the optimal sensor placement.

The aim of this paper is to provide evidence for the imperfection of water quality sensors, by conducting measurements in an operational environment. Therefore we have conducted a number of experiments at the WaDi (WaterDistribution) testbed, a realistic water distribution facility at iTrust, the Centre for Research in Cyber Security⁶ at the Singapore University of Technology and Design. WaDi, and the two other cyber-physical testbeds at iTrust, SWat (secure water treatment) and EPIC (electric power intelligent control), are scaled-down replicas of real-life cyber physical systems containing all the elements of real-life infrastructures. The data generated at these testbeds is made available on iTrusts website, which anyone can request for their research and education purposes. According to [1], the iTrusts datasets form the only openly shared data for research and education, in the context of cyber-physical systems.

Through our experiments conducted at WaDi we have studied the imperfection, sensitivity and degradation of the water quality sensors. Experiments were conducted under both normal conditions (water flow without contaminants present) and attack conditions. For the attack scenario, three different contaminants were considered, a sodium hypochlorite solution, an ammonia solution and a hydrogen chloride solution. The attack scenarios were performed for different values of water flow rate, dosing rate and dosing duration. Although we focus on sensor imperfection related to chemical properties in this paper, imperfection can also reflect the impact of cyber attacks. For instance, in 2015, the US States ICS-CERT (industrial control systems cyber emergency response team) received and responded to 295 incidents [9]. The Water Sector account for 8.5% of these incidents. Therefore sensor

imperfection could also be a result of a hacked sensor, which communicates spoofed data to the system.

The rest of this paper is organized as follows. In section 2 we describe the WaDi testbed in detail as well as the contaminants selected for the experiments. In section 3 we describe the three types of experiments we designed as well as their results. Finally we conclude in section 4.

2. The WaDi testbed

An overview of the WaDi testbed will be given in section 2.1. Next, more details on the contaminants and sensors used in our research will be given in section 2.2.

2.1. Overview of the testbed

The WaDi testbed is a small real-life water distribution network with tanks, pipes, valves, pumps, and sensors. As we are mostly interested in the sensors and their location in the network, the focus of this overview will be on those sensors. Figure 1 shows the overview of the WaDi testbed. The directions in this figure show the flow of water from one stage to another stage. The complete overview of the WaDi testbed is given in figure A1 in the appendix. WaDi contains a total of four water quality sensors. The same type of sensors are used by public utilities [6]. Each sensor consists of multiple probes measuring four different parameters, each of which gives an indication of the water quality. These parameters are conductivity, turbidity, pH and ORP. We did not measure the water temperature. According to [14] temperature is usually constant if measured in short periods of time, although it changes with the seasons. Figure 2 shows a picture of a part of the WaDi testbed.

The WaDi testbed consists of three stages, namely, primary grid (Stage: S1) consisting of two raw water tanks, secondary grid (Stage: S2) consisting of two elevated reservoir tanks and the return water grid (Stage: S3). Further, Stage S2 is divided into sections S2A and S2B. A more detailed overview of the testbed is given in the appendix, figure A1. Water flows in the representation of figure A1 in the appendix, from the top left corner in a clockwise direction back to the top left corner. The water takes the following route:

1. The two raw water tanks represent reservoirs in a city, and these tanks are fed by two different sources. One source is the Singaporean public water distribution network, providing tap water. The second source is the

⁶<https://itrust.sutd.edu.sg/>

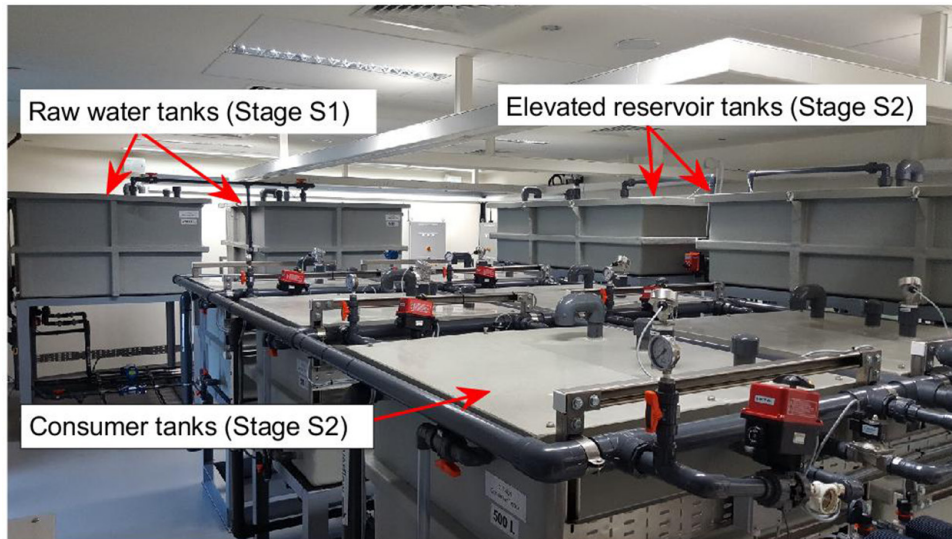


Figure 2. Picture of part of the WaDi testbed.

WaDi testbed itself, providing cleaned return water that has traveled through the WaDi pipelines before. From these sources, the water passes a first water quality sensor (P1) (see figure A1), upon entering the first raw water tank. The task of this sensor is to decide whether the quality of the water is high enough. If the quality does not meet a preset threshold, the raw water tanks will be drained directly.

2. After the raw water tanks, water is pumped to the two elevated reservoirs which represent elevated tanks in a real-world scenario. From these tanks, the water flows further through the system. The amount of water which exits these elevated reservoirs is equal to the required water demand by the consumer tanks described in step 4. After the reservoirs, the water first passes sensor P2A after which contaminants can be injected into the pipe, both from an organic and an inorganic tank.
3. From that point onwards, the water can take two routes to the consumer tanks, via a gravity feed or via the booster pump stations. It is possible to regulate which of these routes is used by closing certain valves. The gravity feed contains another water quality sensor (P2B). The contaminants which were inserted into the system at the end of step 2, can be detected by sensor P2B. Due to this, the booster station valve was always closed in our research. The gravity feed also includes a draining option after sensor P2B, for situations in which the toxic water should be drained immediately. More details on the part of the testbed from the injection of a contaminant to sensor P2B will be given in section 2.2.
4. After the gravity feed, or the booster pumps, the water will reach the consumer tanks. The amount of water going to each consumer tank, and thus the flow through the system, can be regulated beforehand. The required demand of each consumer tank can be changed every 5 min, emulating to some extent, dynamic demand.
5. Finally, the water coming through the consumer tanks will reach the return water system. In this last part, the

water can be cleaned in order for it to return to the start of the process. Here the final sensor P3 is located, which is used to check if the water is clean enough to return to the first part of the system.

2.2. Contaminants and sensors

Contaminants can be inserted into the WaDi testbed after the elevated reservoirs between sensors P2A and P2B. After a small section of pipes, these contaminants can be detected by sensor P2B. Sensors further downstream in the network, notably sensor P3, cannot easily detect those contaminants as they will be largely diluted after passing through the consumer tanks. In this section, it will be explained how contaminants can be inserted and in what way a sensor can detect contaminants.

When a liquid is placed in the contamination tanks, it can be pumped into the network with a chosen flow rate and for a chosen duration. The dosing pumps present at the WaDi testbed can pump a contaminant into the pipes with a maximum dosing rate of 0.59 liters per hour. Figure 3 shows a picture of the dosing tanks and pumps. The contaminated water will then flow through some pipes to sensor P2B.

In our research, we used three different solutions as contaminant: sodium hypochlorite, hydrogen chloride and ammonia. Their pH levels were approximately 12, 12.5 and 2 respectively. According to environmental public health regulations, the pH levels of these solutions are outside the healthy range of drinking water [13].

Information on the length and diameter of the pipes between the contamination point and the sensor P2B, is given in table 1. It is also good to note there are eleven 90 degrees bends present in this section of pipes. The effect of these bends are ignored for simplicity reasons. The total volume of the pipes between the point of injection and the sensor is $10.659 \text{ dm}^3 = 10.659 \text{ l}$. As the flow rate in these pipes is continuously measured, it can be calculated how long it should

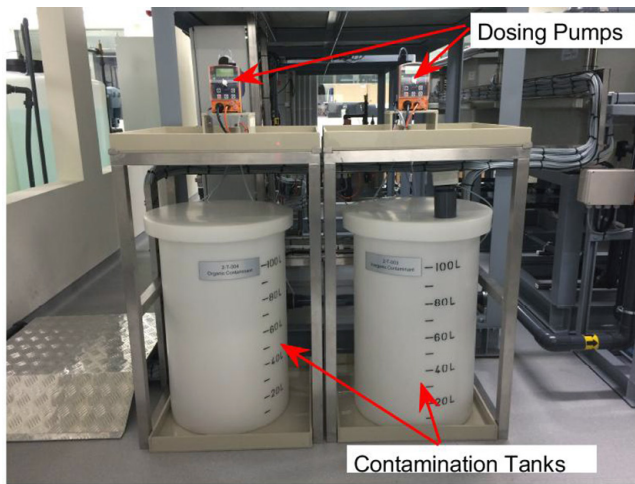


Figure 3. Contamination tanks with the dosing pumps on top.

Table 1. Measurements on the pipes between the injection point and sensor P2B which could detect the contaminant.

	From injection point to branching point	From branching point to sensor
Length (m)	5.24	1.08
Inner diameter (cm)	5.080	0.635
Volume (dm ³)	10.625	0.034

take for sensor P2B to detect the contamination after the injection of a contaminant. If the flow rate is for example 0.1 liter per second, it is expected that the contaminant reaches sensor P2B after 106.59 s.

For each water quality parameter, the healthy range for drinking water is shown in table 2. Conductivity sensors measure the ability of a liquid to conduct electricity. This value changes when more ions are present in the water. Turbidity readings are optical measurements of the diffraction of light in the water and relate to the particle load in the water. pH is an indication of how acidic or basic a liquid is. ORP (oxidation reduction potential) is used as a measure of disinfection potential which is related to how long bacteria can stay alive in that liquid. The units of conductivity, turbidity and ORP are mS/cm (milliSiemens per centimeter), NTU (nephelometric turbidity unit) and mV (millivolt) respectively. The measure pH is unitless. Table 2 also shows an estimation of the four parameters in normal tap water.

The sensors in the WaDi testbed are programmed in such a way that an alarm is raised when the measured value of the sensor is outside of this healthy range. Furthermore, as P2A and P2B are located close to each other, an alarm is also raised if the difference between the readings of P2A and P2B is too large.

The sensors used in the testbed have a lifetime of approximately one year. Maintenance, for example calibration of the sensors, should be done every couple of months [6]. In this calibration, it is checked for each of the four parameters with several standard solutions whether or not the performance of the sensors is still sufficiently good.

Table 2. Ranges of the sensor parameters for which the quality is considered to be sufficient, together with the normal reading of tap water.

	Healthy range	Normal tap water
Conductivity (mS cm ⁻¹)	0–0.32	0.15–0.20
Turbidity (NTU)	0–4	0–0.5
pH	6.5–9.5	7.5–8
ORP (mV)	–500 to 500	300–500

3. Experiments and results

We designed three different experiments to investigate the imperfection of sensors and the way contaminants are being detected. First, the WaDi testbed was put into operation for multiple days to check how sensors behaved in the long run and to get an initial idea on imperfection. Second, the performance of the sensors was evaluated by comparing known levels of standard solutions with the sensor readings several times. Finally, contaminants were added into the network to see how the sensor performed in detecting the contamination. The experiments and their results are presented in sections 3.1, 3.3 and 3.4, respectively. In section 3.2 we investigate the impact of the sensors sampling rate.

3.1. Long runs

The first experiment is a simple procedure in which the WaDi testbed is put into operation for a couple of days such that water is continuously flowing through the system and a large amount of sensor data is being collected. Our main focus is on the sensor data of P2A and P2B as these are very close to each other. The same water that passes P2A will reach P2B within a couple of minutes, depending on the flow rate. Therefore, it is expected that the long run sensor data for the two sensors is similar, or at least highly correlated. Another point of interest is finding indications of imperfection in the resulting time series.

Two different long runs are being considered. The first one consists of data of four days and the second one of almost three days. We will refer to them as LR1 and LR2 respectively. Both long runs were separated by several weeks. In figures 4(a) and (b), the conductivity and pH levels over time of LR1 are being shown. In each figure, the red line represents the time series of sensor P2A, while the time series of P2B is represented by a blue line.

Upon a visual inspection of figures 4(a) and (b) it is obvious that there is quite a large gap between the time series for the two sensors. This should not be the case as no contaminant was added for the described experiment. Such deviations could lead to a lot of false alarms, in case the threshold for the difference between the two time series is set too low. In the next subsection we give a possible explanation for the observed gap. Apart from the gap between the two time series, some other observations can be made. There is an inexplicable

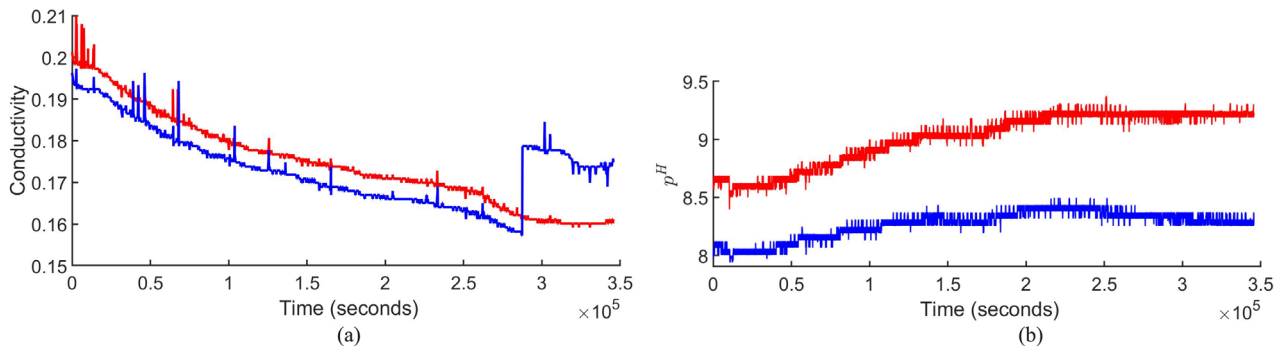


Figure 4. Two time series of sensor readings of P2A (red) and P2B (blue) in LR1. (a) Conductivity levels over time in LR1. (b) pH levels over time in LR1.

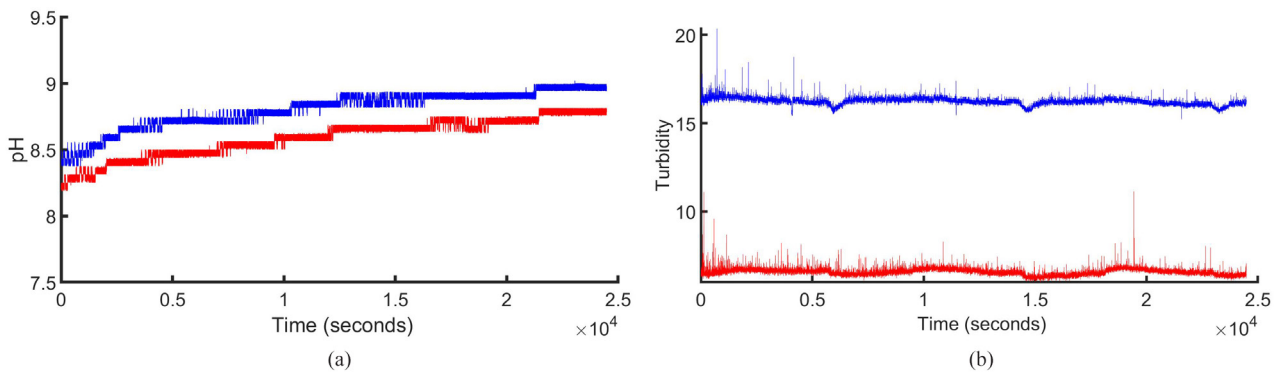


Figure 5. Two time series of sensor readings of P2A (red) and P2B (blue) in LR2. (a) pH levels over time in LR2. (b) Turbidity levels over time in LR2.

jump present in the P2B time series of conductivity, in which the conductivity level suddenly increased with 0.021. Also this jump can cause the sensor to raise an alarm in real WDNs even though no contaminant was present in the system. A missing value was present close to the end in the P2A time series of pH in figure 4(b). This missing value is probably due to measurement or communication errors. In total, there were two inexplicable jumps and four time periods with missing values in all collected data. The longest time period in which a sensor reading was missing, was a period of almost two minutes in the P2B time series of turbidity. A final observation is that while the conductivity levels easily fit within the healthy range, the pH level according to P2A is close to the upper bound (pH = 9.5).

In figures 5(a) and (b), two LR2 time series of, respectively, pH and turbidity, are given. If the sensors are working accurately, it is expected that the time series of sensors P2A and P2B should be approximately the same with a slight delay because the same water passes both sensors. Again, figure 5(b) exhibits a gap between the two time series, where the difference between the signals for turbidity is very large, and also outside the healthy range. We will apply Pearson's correlation coefficient, see [12], to quantify the correlation between the two time series. For this, the time it takes for water to travel from sensor P2A to P2B has to be taken into account. In LR1, the average flow rate was 313 liters per hour and in LR2 it

was 630 liters per hour. Using these values and the volume of the pipelines in between the two sensors, it can be calculated that the water should reach P2B in approximately two minutes and one minute in LR1 and LR2, respectively. Furthermore, as outliers can have a large impact on the correlation, the missing values were replaced with the last correctly measured value. The time series with a sudden large jump are split in two time series in order to evaluate the correlation before and after the permanent increase.

It is expected that the sensor readings of P2A and P2B show a high correlation. The highest correlation was found with the conductivity time series of LR1 before the jump which was shown in figure 4(a). After the jump, the correlation was much lower (0.995 versus 0.652). The correlation of the other conductivity time series was just below 0.9 and all pH and ORP time series were highly correlated with values higher than 0.9.

In contrast to these results of pH, ORP and conductivity, the turbidity time series hardly showed any correlation. The correlation between the time series of P2A and P2B were -0.0914 and 0.2931 for LR1 and LR2, respectively. This is a clear indication that the turbidity sensor is quite unstable. As can be seen in figure 5(b), the turbidity sensor will often raise a false alarm due to the large number of peaks. Therefore, the turbidity sensors will not be trusted by operators and might be ignored when a real contaminant passes that sensor.

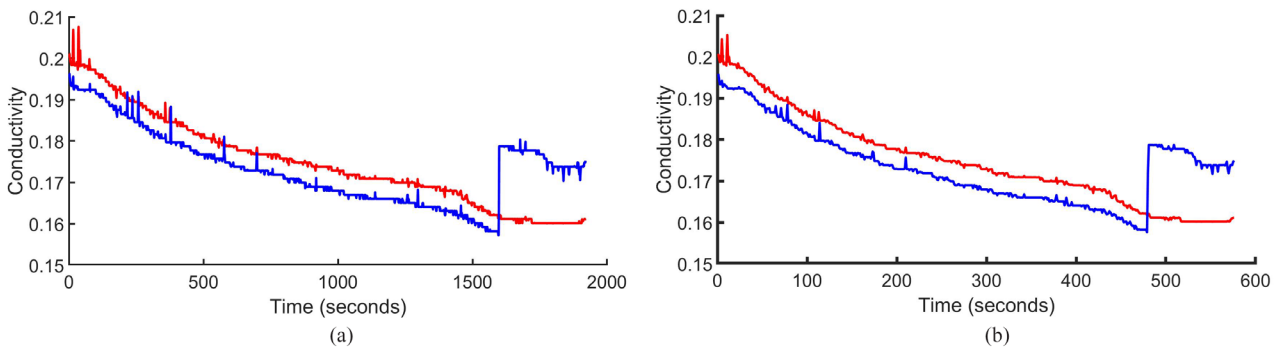


Figure 6. Conductivity in LR1 experiment for different sampling rates (P2A (red) and P2B (blue)). (a) Sampling rate-180 s. (b) Sampling rate-600 s.

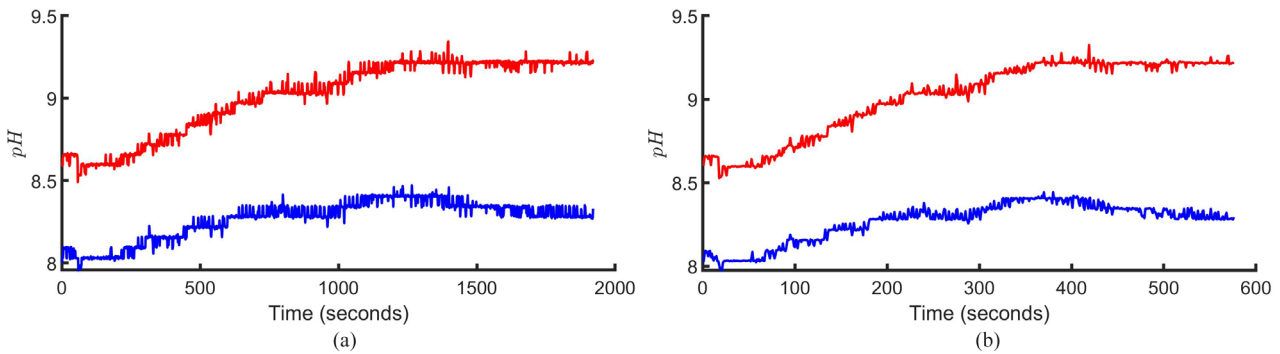


Figure 7. pH in LR1 experiment for different sampling rates (P2A (red) and P2B (blue)). (a) Sampling rate-180 s. (b) Sampling rate-600 s.

Table 3. Correlation coefficients for Conductivity and pH at different sampling rates.

Parameter	1	60	180	600
Conductivity	0.7272	0.7275	0.7281	0.7293
pH	0.9062	0.9208	0.9254	0.9342

So, using these long runs, the first evidence of sensor imperfection is found, through the occurrence of measurement or communication errors, inexplicable jumps and instability in some readings.

3.2. Impact of sampling rate

In our experiments the sensors have been sampled every second. This rate was chosen because the testbed was primarily designed to study the impact of cyber attacks on this critical infrastructure. Detection methods for cyber attacks typically operate at these sampling rates, or even higher. In this subsection we show that for our experiment the impact of lowering the sampling rate is small. Figures 6 and 7, show conductivity and pH, respectively, for LR1, when sampling every 180 and 600 s. As expected, lower sampling rate leads to smoother curves.

Table 3 shows the impact of the sampling rate on the correlations between sensors values at P2A and P2B, for conductivity and pH. It is clear that the sampling rate has little impact on the correlation. Finally, according to [3], typical sampling rates in operational water systems are once per minute and lower.

Table 4. The performance of the pH sensors over some time.

Sensor	Standard solution: pH = 10.01			
	P1	P2A	P2B	P3
Last checked	August 10	August 29	August 29	July 28
September 7	—	9.68	10.05	—
September 13	12.31	9.79	9.95	9.82
September 19	12.71	9.93	10.12	10.09
September 25	12.98	10.10	10.31	10.15

3.3. Standard solutions

In section 2.2, it was mentioned that the performance or accuracy of the sensors should be checked every couple of months using some standard solutions and that the sensors should be calibrated if necessary. This procedure can be illustrated with an example. Suppose we have a standard solution with a known pH level of 7. A sensor which is put into a sample of this liquid should measure a pH of 7. If this measured value is not 7 but 5 or 8.5, the sensor must be recalibrated such that it will give the accurate value. This calibration should be done every three months according to the calibration software. Most of the parameters measured at a sensor were calibrated using a two-point calibration and thus with two standard solutions.

This calibration was done at the WaDi testbed before and in between several other experiments. It became clear that a parameter for a sensor could become unreliable very fast. For example, the ORP value of sensor P2A was showing 530 instead of 430 within a week after a calibration. Therefore, experiments were done to formally show this inaccuracy or

Table 5. The sensor readings of sensor P2A over time for conductivity, ORP and turbidity.

P2A Std. solution value	Conductivity		ORP	Turbidity	
	<0.1	0.447	430	0.1	20
September 7	0.066	0.514	425	2.7	37.0
September 13	0.059	0.590	421	6.7	28.8
September 19	0.055	0.505	420	7.8	21.8
September 25	0.057	0.587	412	8.1	47.4

Table 6. A summary of all sensors and parameters which shows which sensor readings were still accurate (+), inaccurate (−) or indeterminable (±) after a month.

	Conductivity	ORP	pH	Turbidity
P1	±	−	−	−
P2A	±	+	+	−
P2B	±	+	+	±
P3	±	−	+	−

instability over time and to find evidence for sensor degradation or drifting. During these experiments, the sensors were not calibrated and the readings were tested every couple of days for all parameters and sensors to check how far each sensor reading was off. Before the start of the experiments, sensors P2A and P2B were checked and, if necessary, calibrated approximately a month after the other two sensors were last checked.

In table 4, the results for all four sensors for the standard solution with a pH level of 10.01 over time are shown. The first time, on September 7, P2A and P2B were the only sensors which were being checked. After September 7, the accuracy of all sensors was checked every six days.

In order to evaluate whether or not a sensor still gives accurate readings, we allow for an arbitrary, five percent margin. If the readings are within 5% of the known value of the standard solution, the performance of the sensor is still considered to be sufficient. For the standard solution with pH level of 10.01, the sensor reading should be roughly between 9.5 and 10.5. It can be seen in table 4 that all sensor readings are still within this range except for sensor P1. The readings of this sensor are between 12 and 13. For the other standard solution with a known pH level of 7, this sensor also shows readings that are too high. This means that the pH sensor of P1 is unreliable after just a month and needs to be recalibrated.

Even though the other three sensors are within the right range, they all show some upwards slope. These pH sensors are all slowly drifting away from the right value which can also cause problems in the long-term. When sensors are programmed to raise an alarm when a sensor reading is outside a certain range, drifting can lead to false positives or even to non-detections of attacks. Non-detections can for example occur when a sensor reading of pH drifts towards the maximum value within the healthy range when the quality of the tap water is in fact still in the middle of this healthy range. An acidic attack causes a large decrease in the pH value such

Table 7. Different values of flow rate, dosing rate and dosing duration used.

	Water flow rate	Dosing rate	Dosing duration (min)
High	±950 liters per hour	0.472 liters per hour	5
Medium	±600 liters per hour	0.354 liters per hour	—
Low	±300 liters per hour	0.236 liters per hour	2

that the sensor reading is close to the minimum threshold of the healthy range, while the actual pH value is below this threshold, being harmful for humans.

For sensor P2A, the readings over time for conductivity, ORP and turbidity are shown in table 5. ORP is the only parameter for which only one standard solution was available.

A downwards drifting pattern was found for the ORP reading. This was the case for all four sensors. The conductivity readings were more stable but in general a bit too high for the second standard solution compared to the value of 0.447. We suspect that this is mainly due to the fact that conductivity was checked every time, but the last time it was calibrated was more than half a year ago. The values around 0.5 were probably tolerated in combination with the good sensor readings of the other ‘< 0.1’ standard solution. The turbidity readings were, just as we concluded in section 3.1, very unstable.

Finally, table 6 shows for each sensor and parameter which readings were still correct (+) after a month and which were inaccurate (−). A plus-minus sign means that the sensor reading was still stable and within or close to the five percent range for most of the evaluations or that it was unable to form a conclusion with just three or four data points.

As all conductivity readings were somewhat too high, a ± sign was given for all of them as the last calibration happened more than six months ago and a formal conclusion can therefore not be drawn. Remember that according to the guidelines, the sensors should be calibrated every three months. After one month, only two ORP and three pH readings were still accurate and the turbidity readings were unstable and inaccurate for most sensors. As the considered sensors are almost a year old and thus close to the end of their lifetime, the results lead to clear conclusions. Maintenance or calibration should happen more often than currently recommended because the sensors are drifting quite fast. As we suspect that the performance of the sensors is better in the beginning of their lifetime, sensor degradation appears to have occurred. Thus, we have found more evidence for sensor imperfection.

3.4. Adding contaminants

Three different liquids were used in our research as a contaminant and added into the WaDi testbed: a sodium hypochlorite solution (Hypo), an ammonia solution (Ammonia) and a hydrogen chloride solution (HCl). Our main objective when adding these contaminants into the water network was to see which sensor parameters could detect the contaminations, the time needed to detect a contamination and how long

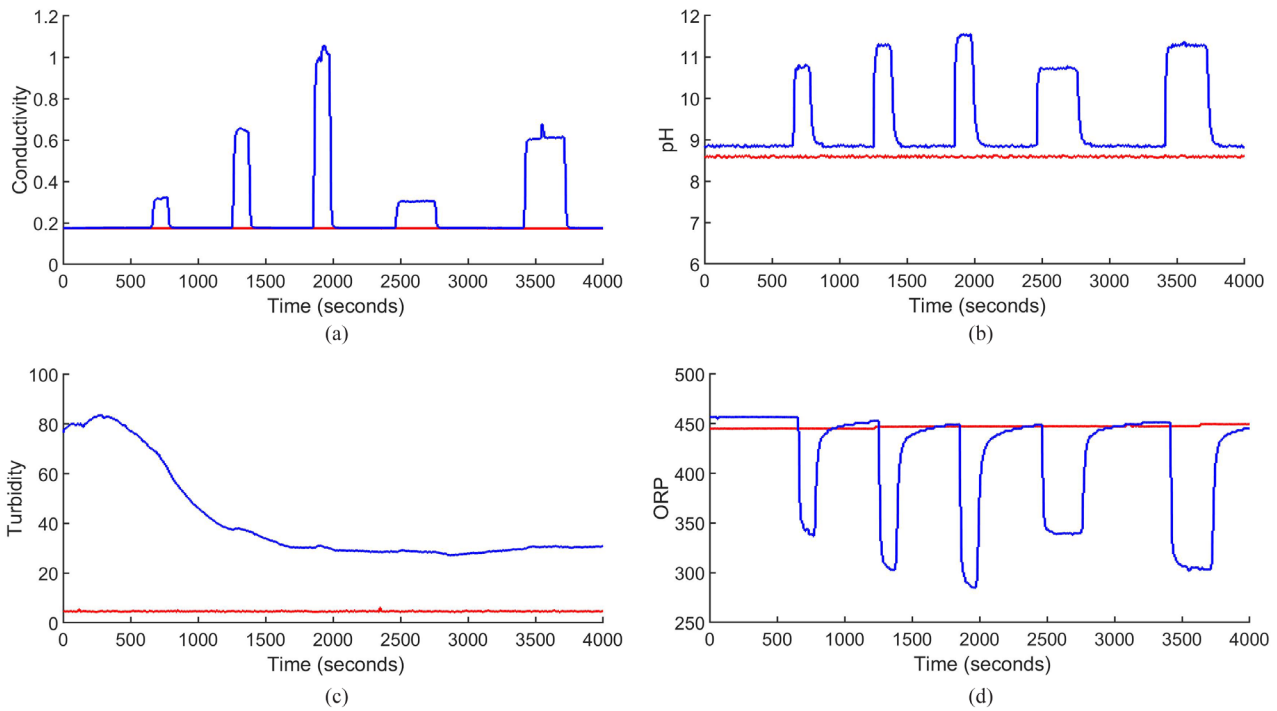


Figure 8. Time series of the four sensor readings of P2A (red) and P2B (blue) over time when Ammonia is added into the network, with different settings for dosing rate and duration. (a) Conductivity levels over time. (b) pH levels over time. (c) Turbidity levels over time. (d) ORP levels over time.

Table 8. Parameters that reacted (+) to each of the contaminants. The arrow shows if it is a positive peak or a negative peak. A minus sign means that the sensor parameter was unable to detect the contamination and the plus-minus sign shows that there were some peaks present, but not as clear as with the other parameters.

	Conductivity	Turbidity	pH	ORP
Hypo	+ (↑)	-	+ (↑)	± (↑)
Ammo	+ (↑)	-	+ (↑)	+ (↓)
HCl	+ (↑)	-	+ (↓)	+ (↑)

the contaminant stayed in the pipelines. Besides that, we focus on the effects of factors such as flow rate, dosing rate, dosing duration and the type of contaminant, on the detection.

Multiple experiments were performed for each contaminant. The liquid was added into the network in a different setting every time, i.e. a different combination of the water flow rate, the dosing rate and the dosing duration. The different settings for each experiment can be found in table 7. Recall that the maximum dosing rate for the contamination pumps is 0.59 liters per hour. The maximum flow rate for the water through the gravity feed of the WaDi testbed is 1000 liters per hour. Only two dosing durations were used as the supply of these contaminants was limited.

As a contaminant is inserted between sensors P2A and P2B, sensor P2B should be able to detect the contamination. Figure 8 shows the time series of the four sensor readings of P2B when Ammonia is added. The time series of sensor P2A is also added to the picture as a benchmark. Ammonia was added five times in these figures, each time in a different

Table 9. An overview of the detection times of contaminants in the WaDi testbed for different flow rates.

	Stable flow rate	Expected detection time (s)	Observed detection time (s)
High flow rate	960 liters per hour	40	49
Medium flow rate	610 liters per hour	63	67
Low flow rate	312 Liters per hour	123	111

setting. In each setting, the water flow rate was high. The combination (Dosing rate, dosing duration) was (low, low), (medium, low), (high, low), (low, high) and (medium, high) for settings one to five respectively.

The most remarkable result can be found in the turbidity levels in figure 8(c). All other three time series show clear peaks when the contaminant is inserted and the influence of the dosing rate and dosing duration is very clear. Take for example the conductivity readings of figure 8(a). The higher the dose, the higher the peak and the longer the duration, the wider the peak. These findings are the same for all three contaminants. In these figures, the flow rate was high each time. When less normal water was flowing through the pipelines per hour, the peaks were somewhat higher.

So, the turbidity sensors are not able to detect Ammonia. In table 8, a summary is presented for all three contaminants

showing which parameters can detect a certain contaminant and if the sensor reading increases or decreases.

As can be seen in table 8, turbidity is not able to detect any of the three contaminants. The turbidity sensor measures the diffraction of light in the water which means that a reading states something about the particle load in the water. As all three considered contaminants are transparent liquids, none of these make the contaminated water turbid. However, the resulting time series of these experiments showed again that the turbidity sensor readings are quite unstable as was also shown in the previous sections. Conductivity, pH and ORP all showed some peaks when the contaminant was added. The peaks in the ORP time series when Hypo was added were not as clear as the ones of conductivity and pH such that ORP alone is not enough to detect that contaminant.

Finally, we are interested in the time it takes for a contaminant to reach the sensor and how long the contaminant remains visible in the water. Using the calculated volume between the point of injection and sensor P2B in section 2.2 and the measured flow rate, the expected detection time in seconds can be calculated. This formula is shown in equation (1).

$$\text{Expected detection time in seconds} = \frac{\text{Volume of pipelines (dm}^3\text{)}}{\text{Flow rate (dm}^3 \text{ s}^{-1}\text{)}}. \quad (1)$$

The volume of the pipelines is 10.659 dm³ and three different flow rates were used in our research. In table 9, the average flow rate in the three settings is shown. With each flow rate, the expected detection time can be calculated. This value is compared with the average observed detection time of the added contaminant. This observed detection time is based on the first observation after contaminant was added for which the sensor readings differ more than 5% from the mean of the previous 100 observation points. For this, the well-known CUSUM method is used [10]. So, not on the range or difference between the sensor reading of P2A and P2B as it was described in section 2.2. This was chosen due to the instability and inaccuracy of the time series as observed in section 3.1.

The difference between the observed and expected detection time is quite interesting. When the flow rate is high, the expected time is too low and when the flow rate is low, the expected time is too high. The average observed detection time for all different flow rates is very stable as the variance is quite small. The detection of contaminants when the flow rate was high was for example always close to 50 s. There was hardly any difference in the detection times for the different contaminants or different dosing rates.

We are not able to give a solid explanation for this difference between the observed and expected detection time. It is possible that the measured flow rate is not correct as this flow rate

was only measured in one point of the pipelines. We leave clarification of this issue for future research. Furthermore, figure 8 also shows that after adding of contaminants was stopped, most sensor readings quickly return to their old values.

4. Conclusion

Through the experiments performed on the WaDi testbed we have found strong evidence for sensor imperfection. Failing to detect a contaminant can occur as missing values were observed in the data series. Besides that, the sensor readings are quite unstable as inexplicable peaks and jumps in the sensor readings were found and the sensors become inaccurate and unreliable within several weeks after calibration, due to drifting. Especially the sensor readings for turbidity are very unstable. Still, all added contaminants are detected by at least two sensor parameters. After the contaminant is added, the sensor readings quickly increase or decrease to a new static level. This means that when the sensor misses this initial change due to measurement errors, the contaminant will possibly not be immediately detected. Furthermore, low flow rates and high contaminant dosing rates result in higher peaks of the sensor readings, which may increase the probability of detecting a contaminant. This implies that the probability for a sensor to detect a contaminant depends on the effect of dilution. Sensors which are placed further away from the source of contamination, might only observe small peaks due to branching and dilution effects.

Finally, the data of the experiments on the WaDi testbed indicate that many false positives occurred. A sensor can measure a rapid change in a certain parameter and therefore raise an alarm when no contaminants are being added. Due to the large number of false positives and the fact that a water network will not be shut down for each positive sensor reading, it is possible that even when a contaminant is added and the sensor is detecting something, it can be considered as a false positive by the system or by employees. This can result in a non-detection even when it was detected. The occurrence of false positives should therefore also be taken into account in further research.

Our qualitative study has shown the imperfection of water quality sensors in an operational network. Methods for optimal sensor placement for water distribution, need quantitative values for this imperfection as input. For future research we will design and conduct experiments that allow us to quantify the imperfection of water quality sensors. For future experiments we will also use sensors which are not based upon electro-chemical principals, such as the refractive index, see [18].

Appendix

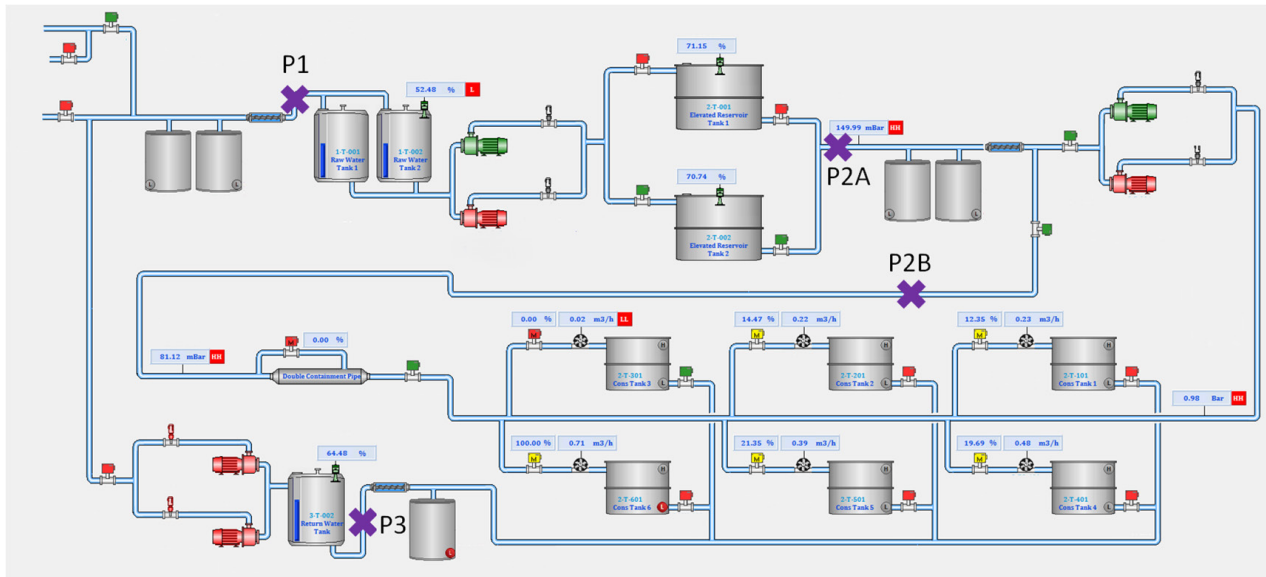


Figure A1. An overview of the WaDi testbed. The four sensors are presented with purple crosses.

Table A1. LR1 correlation coefficients.

	Conductivity (2A)	Turbidity (2A)	pH(2A)	ORP (2A)	Conductivity (2B)	Turbidity (2B)	pH(2B)	ORP (2B)
Conductivity (2A)	—	0.1925	-0.9452	0.9282	0.7272	0.2471	-0.8092	0.7797
Turbidity (2A)	—	—	-0.2093	0.2508	0.2988	-0.0914	-0.2058	0.2241
pH(2A)	—	—	—	-0.9599	-0.8216	-0.2312	0.9062	-0.8831
ORP (2A)	—	—	—	—	0.8219	0.2377	-0.8946	0.9228
Conductivity (2B)	—	—	—	—	—	0.0988	-0.8662	0.8773
Turbidity (2B)	—	—	—	—	—	—	-0.2222	0.2302
pH(2B)	—	—	—	—	—	—	—	-0.9585

Table A2. LR2 correlation coefficients.

	Conductivity (2A)	Turbidity (2A)	pH(2A)	ORP (2A)	Conductivity (2B)	Turbidity (2B)	pH(2B)	ORP (2B)
Conductivity (2A)	—	0.2650	-0.5870	0.5496	-0.6682	0.3527	-0.6503	0.6324
Turbidity (2A)	—	—	-0.2122	0.1583	0.0670	0.2931	-0.1885	0.2176
pH(2A)	—	—	—	-0.9838	0.6115	-0.4429	0.9534	-0.9694
ORP (2A)	—	—	—	—	-0.6093	0.4467	-0.9542	0.9754
Conductivity (2B)	—	—	—	—	—	-0.2626	0.6917	-0.6402
Turbidity (2B)	—	—	—	—	—	—	-0.4774	0.4855
pH(2B)	—	—	—	—	—	—	—	-0.9794

ORCID iDs

Venkata Reddy Palleti  <https://orcid.org/0000-0001-8799-7728>

References

- [1] Almgren M, Andersson P, Bjrkman G, Ekstedt M, Hallberg J, Nadjm-Tehrani S and Westring E 2018 Rics-el: building a national testbed for research and training on scada security *13th Int. Conf. on Critical Information Infrastructures Security*
- [2] Berry J, Carr R D, Hart W E, Leung V J, Phillips C A and Watson J-P 2009 Designing contamination warning systems for municipal water networks using imperfect sensors *J. Water Resour. Plan. Manage.* **135** 253–63
- [3] Chappell N A, Jones T D and Tych W 2017 Sampling frequency for water quality variables in streams: systems analysis to quantify minimum monitoring rates *Water Res.* **123** 49–57
- [4] Comboul M and Ghanem R 2013 Value of information in the design of resilient water distribution sensor networks *J. Water Resour. Plan. Manage.* **139** 449–55
- [5] de Winter C, Palleti V R, Worm D and Kooij R 2019 Optimal placement of imperfect water quality sensors in water distribution networks *Comput. Chem. Eng.* **121** 200–11
- [6] FlowTech Engineering Technician 2017 private communication
- [7] Gleick P H and Heberger M 2014 *Water Conflict Chronology (The Worlds Water)* (Berlin: Springer) pp 173–219
- [8] Hu C, Li M, Zeng D and Guo S 2018 A survey on sensor placement for contamination detection in water distribution systems *Wirel. Netw.* **24** 647–61
- [9] U.S. Department of Homeland Security Industrial Control Systems-Cyber Emergency Response Team 2016 *ICS-CERT 2016 NCCIC/ICS-CERT Year in Review: FY 2015* (Washington, DC: U.S. Department of Homeland Security Industrial Control Systems-Cyber Emergency Response Team)
- [10] Page E S 1954 Continuous inspection schemes *Biometrika* **41** 100–15
- [11] Palleti V R, Narasimhan S, Rengaswamy R, Teja R and Bhallamudi S M 2016 Sensor network design for contaminant detection and identification in water distribution networks *Comput. Chem. Eng.* **87** 246–56
- [12] Pearson K 1895 Note on regression and inheritance in the case of two parents *Proc. R. Soc.* **58** 240–2
- [13] PUB 2018 Singapore drinking water quality (www.pub.gov.sg/Documents/Singapore_Drinking_Water_Quality.pdf)
- [14] Raciti M, Cucurull J and Nadjm-Tehrani S 2012 *Anomaly Detection in Water Management Systems* (Berlin: Springer) pp 98–119
- [15] Rathi S and Gupta R 2014 Sensor placement methods for contamination detection in water distribution networks: a review *Proc. Eng.* **89** 181–8 (*16th Water Distribution System Analysis Conf.*)
- [16] van den Broeke J 2014 The benefits of using refractive index for water quality monitoring in distribution networks *Optiqua Technol. Whitepaper*
- [17] van Thienen P 2014 Alternative strategies for optimal water quality sensor placement in drinking water distribution networks *11th Int. Conf. on Hydroinformatics (New York)*
- [18] Williamson F, van den Broeke J, Koster T, Koerkamp M K, Verhoef J W, Hoogterp J, Trietsch E and de Graaf B R 2014 Online water quality monitoring in the distribution network *Water Pract. Technol.* **9** 575
- [19] Xu J, Small M, Fischbeck P and Van Briesen J 2010 Integrating location models with bayesian analysis to inform decision making *J. Water Resour. Plan. Manage.* **136** 209–16