Development of risk assessment tools for early detection of bio-contamination in water distribution systems

S. Tinelli, I. Juran and W. P. Cantos

ABSTRACT

Water distribution systems (WDSs) could be an easy target for accidental or intentional contamination in any city. The lack of reliable, real-time specific microbial and chemical contamination monitoring raises major public concerns as delayed detection accelerates the fast growth of the public health risk. In this context, the proposed research aims to develop, adapt and demonstrate the technical feasibility of smart Bio-Sensing and Information Management (Bio-SIM) that will effectively enable water operators to ensure smart water monitoring and quasi real-time quality control for early contamination detection and preemptive warning. Facing a lack of labeled data, numerical simulations with engineering models (EPANET-MSX) have been adopted to produce synthetic data for pattern recognition of the effect of bio-contamination on chemical and physical water parameters (e.g. pH, total organic carbon, turbidity and free chlorine). Statistical and stochastic models enabled non-specific bio-anomaly detection, based on pattern recognition of Chlorscan data and selected parameters that can be monitored simultaneously online using multi-parameter sensors (e.g. s::can). The results were compared with laboratory model tests and numerical simulations. This paper outlines the main results of the feasibility assessment of the Bio-SIM prototype system for early bio-contamination detection, which is expected to significantly contribute to alleviating consumers' bio/chemical contamination risks.

Key words | bio-contamination, chlorine decay, EPANET-MSX, risk assessment, water distribution system

S. Tinelli (corresponding author)
DICAr-Dipartimento di Ingegneria Civile e Architettura,
University of Pavia,
Via Ferrata 3, Pavia 27100,
Italy
E-mail: *silvia.tinelli01@universitadipavia.it*

I. Juran
W. P. Cantos
New York University Tandon School of Engineering,
15 MetroTech Center, Brooklyn, NY 11201,
USA

INTRODUCTION

The potential for contamination of a nation's water supply is very real because contamination could be either intentional or accidental and could involve a variety of classic and non-traditional chemical agents, toxic industrial chemicals and/or toxic industrial materials. For example, biological agents are most likely to be introduced in the water supply system because some of the microbes cause diseases and death even at low concentrations and they can remain in the water supply system for the long term causing dangerous risks of contamination.

First of all, assessing the potential health impacts from drinking water contamination events requires understanding the fate and transport of contaminants through the distribution system, and the exposure of consumers to contaminated water (Davis & Janke 2009). The estimation of the potential risk from short-term contamination events was initially deepened through a Quantitative Microbial Risk Assessment (QMRA) coupled with hydraulic modeling. This method has previously been used to evaluate the risk from microbial intrusion during negative pressure events that led to contamination of distribution systems (Teunis *et al.* 2010; Yang *et al.* 2011): the focus was on contamination through transients and the effect of chlorination.

A similar issue, the issue of timing, was previously examined through the use of a hydraulic model combined with numerous exposure models to estimate the impact on

doi: 10.2166/ws.2018.036

dose. However, only later (Blokker *et al.* 2013; Van Thienen 2014) was the combination of a hydraulic model with an exposure and risk model investigated to look at the impact of the number of consumption events on the probability of infection by a pathogen in a community. The results showed that the time of consumption is important, while the event location within the network and the amount of consumption are of smaller importance. Van Thienen (2014) also created a custom tool named the Contamination Source Toolkit which calculates contaminant transport and implements utility response.

For these reasons, being aware that not all the potentially injected substances can be considered, the first challenge of the current paper is to track the transport and fate of a specific pathogen, that is the *Escherichia coli* bacterium (*E. coli*), in chlorinated drinking water in order to evaluate the variations of the standard water quality, the time of consumption, and consequently the risks associated with that specific species. Therefore, an appropriate EPANET-MSX model (Shang *et al.* 2008) was defined, enabling the modeling of the interaction between the bulk species and pipe wall surface, and taking into account sets of differential-algebraic equations, along with all the required kinetic constants, and equilibrium equations for the simulation of the reaction between the *E. coli* and chlorine. In order to test the proper functioning and accuracy of the EPANET-MSX model before its application to a real distribution network, a pilot laboratory site was created at the University of Lille, able to test the behavior of the injected bacterium. In fact, the pilot laboratory allowed comparison with the numerical simulations for the model reproducing the real conditions of water distribution systems (WDSs), especially in terms of materials, velocity and pressure, and performing the same tests that had been modeled through EPANET-MSX (Abdallah 2015).

After evaluating the water contamination by the *E. coli* bacterium, the second challenge of the current paper is the creation of a smart detection system: the detection of a contamination event requires that the related variations in the values of the measured parameters have to be distinguished from the normal daily and/or seasonal fluctuations (the so-called background noise). In practice, an Early Detection System (EDS) is based on a continuous acquisition of the values of the measured parameters and their transmission to Supervisory Control and Data Acquisition (SCADA) integrated with early detection software that reads and interprets the acquired data by distinguishing the abnormal variations from the normal fluctuations. The abnormal variations then require prompt attention or intervention. To this purpose, CANARY software (open source), developed by Sandia National Laboratories in collaboration with US-EPA (Hart *et al.* 2007; US-EPA 2010, 2012; Hagar *et al.* 2013), provides both real-time and off-line analysis tools, giving particular emphasis to the following features: (i) use of a standard format for input and output of water quality and operations data streams; (ii) the ability to connect various detection algorithms, both in MATLAB and compiled library formats, for testing and evaluation by using a well-defined interface; (iii) an operations approach that simulates the utility operator mode; and (iv) comparison of tools for different evaluation metrics, including Receiver Operating Characteristic (ROC) curves, time to detect, and false-alarm rates. Traditionally, water utilities use set points (thresholds) to identify changes in water quality parameters: set points provide alarms when the actual values of the water parameters go above or below the set-point values. For example, free chlorine levels near zero need to be communicated immediately to an operator. Through the use of different detection algorithms used by CANARY, it is now possible to identify the water quality values that are significantly different from the background values whether or not they exceed the set-point limits. A comparative analysis of the performance of CANARY and other commercial early detection software (OptiDES, ana::tool, BlueBox, Event Monitor) has been published by US-EPA (2013). The Evaluation Center in Cincinnati, OH, together with the researchers in sensor industries such as the Hach Corporation in Loveland, CO, were also involved with water quality sensor testing, developing the Guardian Blue EWS to detect, alert, and classify a wide variety of threat contaminants in drinking WDSs (Kroll 2006). Allgeier *et al.* (2011) tested CANARY software, using the Cincinnati Pilot field and simulating several events (1,588). For the simulated events, their detection rate of the true positives was 40%, leaving 60% as false negatives. However, the authors showed that even if only 40% of the simulated contamination incidents were detected, those undetected scenarios had small consequences. In addition, investigating the opportunities to improve event detection, Vugrin *et al.* (2009) used historical water quality data from the utility to identify recurring patterns and saved those patterns in a library that can be accessed during online operation. This pattern-matching capability was implemented within CANARY in order to demonstrate a decrease in false alarms. Finally, a significant false-alarm decrease was noticed through the method proposed by Koch & McKenna (2011), according to which data can be combined from multiple stations, considering the location and time of individual detections.

Consequently, the present paper aims at: (i) analyzing *E. coli* behavior in a chlorinated drinking water system in order to detect the free and total chlorine variations during a contamination event; and (ii) developing an automated-based approach to detect bio-anomalies in generic WDSs, starting from a database which consists of chlorine time series. In particular, Chlorscan sensors, which are a three-electrode, amperometric sensor that detects free and total chlorine and chlorine dioxide, were used to collect data regarding the chlorine measurements.

Overall, it is implicit that this paper gives an important contribution in terms of *E. coli* analysis in countries where the use of a disinfectant, such as chlorine or chloramine, is required. These countries include the United States and the United Kingdom (Rosario-Ortiz *et al.* 2016).

STATE OF PRACTICE

Engineering model data analysis

Due to the lack of bio-contamination data, synthetic data have been produced starting from biological and chemical contaminations and through the use of engineering models, such as EPANET and EPANET-MSX. The research assumed a multi-species model. Following Dominic Boccelli's research, which has proven that the reactive species model will always represent chlorine decay better than a first-order model (Boccelli *et al.* 2003), *E. coli* inactivation due to chlorine and chlorine first-order decay based on the initial chlorine concentration are here added to the four-species model. Two input files, an output file and a user-defined program are required to run multi-species simulations using EPANET-MSX. The first input file consists of the EPANET file (.inp), and the second consists of the definitions and reactions of the various chemical and biological species that are to be simulated.

The file defines the eight parameters provided in the multi-species model (tryptic soy broth reaction rate coefficients, *E. coli*/chlorine inactivation coefficient), which are estimated using the Hooke–Jeeves optimization algorithm (Hooke & Jeeves 1961). The objective function is the sum of relative squares of difference between the experimental chlorine residuals and the predicted chlorine concentrations from the model. The output file has an extension of '.rpt', which is used to report the results of the multi-species simulations.

The multi-species model is used to simulate the chlorine decay, the transportation of contaminants in the network in terms of time and chemical reaction, and the interaction between the analyzed chlorine with the injection of *E. coli*. In particular, in the EPANET-MSX model, all the reactions between the species are expressed in the form of differential equations; the *E. coli* inactivation because of the chlorine is represented by the chlorine first-order decay from a predefined initial chlorine concentration (refer to Tinelli & Juran 2017 for the model's details).

Pilot site in Lille University

The lack of on-line sensors capable of directly detecting bio-contamination led to the establishment of a laboratory in the University of Lille. In fact, as part of the European project Smart Water for Europe (SW4EU), the pilot laboratory site was built at the Laboratory of Civil Engineering and Geo-Environment, located on the Campus of Lille University (a scientific city in northern France). As already mentioned, the laboratory was built to test chemical reactions and biological transformations that occur once the *E. coli*, as well as the chlorine, have been injected into the network. In addition, being connected to some sensors, the pilot laboratory can test the ability and the sensitivity of sensors available on the market for the detection of chemical or biological contamination in drinking water. To achieve this objective, the laboratory had to reproduce the real conditions of water distribution, especially in terms of materials, velocity and pressure, and it enabled the performance of tests according to different scenarios (planned injections). For this purpose the pilot site consisted of PVC pipes for the flow of water tanks, simulating a linear network, 61 m long and

16 mm in diameter, injections of chemical/biological products, pumps to allow the circulation of water, valves to control the flow direction, devices for pollutant injections and instrumentation for water pressure/velocity.

As part of the European project SmartWater4Europe, s::can, Optiqua and Intellisonde were selected for the monitoring of water quality. The water quality in a water distribution system may vary according to a number of conditions, including: the quality source of water, the treatment of process water, variations in the distribution system, accidental rupture of a pipeline and intentional actions; thus the pH, total organic carbon, conductivity, turbidity and chlorine are subject to variations. The research outcome of Abdallah (2015) demonstrated that for chemical contaminants, EventLab showed high reliability in detecting low concentrations of chemical contaminants while the s::can probes, especially the spectro::lyser, showed a good ability to detect chemical contaminants. Regarding biological contaminants, EventLab showed no ability to detect these types of contaminants while the s::can probes showed an ability to detect biological contaminants for bacterial concentrations above 10^6 CFU/mL.

Comparison between numerical modeling and experiments

The comparison between the experimental activity and the numerical model is illustrated in Figure 1: Figure 1(a) and 1(b) illustrate the chlorine decrease with an *E. coli* injection of 10^5 CFU/mL in the laboratory experiments and the numerical model, respectively. In detail, the system water demand was assumed to vary with hourly steps, and the simulations were run for 24 hours with the same features as the experimental activities, thus they considered:

- initial chlorine concentration of 0.3 mg/L, 0.5 mg/L and 1 mg/L;
- *E. coli* injection equal to 10^5 CFU/mL.

Figure 1(c) illustrates the comparison in terms of chlorine decrease, for an initial chlorine concentration of 0.3 mg/L (and an *E. coli* injection of 10^5 CFU/mL). It shows that chlorine decay is faster in the numerical model than in the experimental activity. In fact, according to the numerical model, chlorine reaches zero more rapidly. For this reason, the two red-circled points in Figure 1(c) differ from the bisector: they specifically represent the delay of the experiments in achieving zero, as compared with the numerical model (please refer to the online version of this paper to see Figure 1 in color: http://dx.doi.org/10.2166/ws.2018.036). These two points are also responsible for the reduction of the factor R^2, which is still satisfactory in water quality analyses.

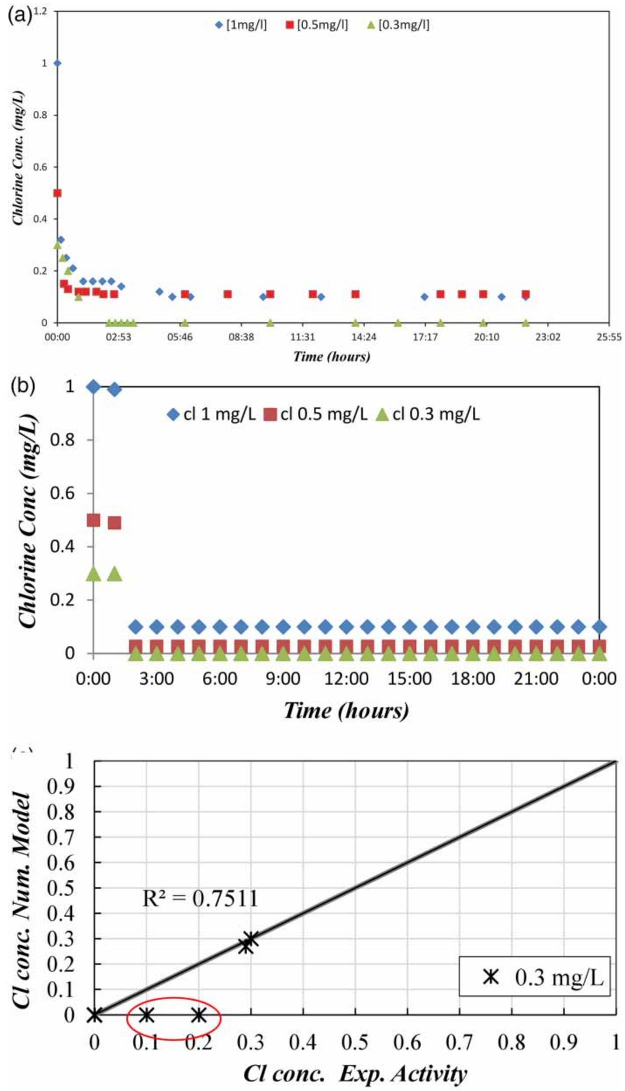

Figure 1 | (a) Laboratory results for the decay of different concentrations of chlorine with an injection of *E. coli* equal to 10^5 CFU/mL; (b) numerical simulation (EPANET-MSX) results for the decay of different concentrations of chlorine with an injection of *E. coli* equal to 10^5 CFU/mL; (c) comparison of chlorine concentrations between the experimental activity and the numerical model with an *E. coli* injection of 10^5 CFU/mL.

An improvement of the equations in the EPANET-MSX model and an enhancement in the model calibration could lead to a better matching of the two chlorine decreases. Approximately the same results were obtained comparing the two studies for initial chlorine concentrations of 0.5 and 1 mg/L with an *E. coli* injection of 10^5 CFU/mL.

This aspect identifies EPANET-MSX as an essential support for qualitative studies in WDS because, unlike the traditional software used in the literature (such as EPANET), it is able to describe the behavior of the most common chemical/physical water reactions during *E. coli* presence and identify variation patterns in the biological and chemical parameters (such as the chlorine).

Chlorscan data analysis

After simulating contamination events to detect biological contamination in drinking water systems, as explained above, a data processing procedure, whose functional architecture is composed of pre-processing/processing, training, validation, and forecasting has to be established in order to define a risk assessment model, together with a decision support system for an EDS definition. For this purpose typical Chlorscans data have been used: since they are sensors used for chlorine measurements, they are considered as relevant indicators of potential non-specific bio-contamination. Chlorscan data analysis requires a first pre-processing/ processing step which consists in cleaning data and removing spurious data points from the raw data. Taking as input the recorded Chlorscan measurements, which are aggregated into a continually updated data file, this phase produces a list of anomalies identifying:

- the data source (sensor or meter ID);
- the period over which the anomaly has been detected;
- the type of anomaly;
- a flag supporting the decision to achieve the detection based on the dataset.

Consequently, false alarms due to non-specific water quality indicators and their high temporal variability are filtered out. Chlorscan data analysis requires then a post-processing step, which supports the final alarm-triggering by crosschecking different sources of information. A multi-spots approach is used to compare the anomalies detected by Chlorscan at different locations over the water distribution network.

METHOD

Risk assessment analysis

This paper presents a statistics-based approach to detect bio-anomalies in a generic WDS. Starting from a database, which consists of Chlorscan time series, statistical tests are implemented to establish the first, second, and third standard deviations. Afterwards, the analysis requires the following steps:

- Chlorscan data are normalized to the average (F) to filter out 'noise'.
- Five threshold levels (insignificant, low, moderate, high, very high) of normalized Chlorscan data are defined to establish likelihood and risk severity levels corresponding to the amplitude of normalized concentration deviation from the average (ΔF).
- The likelihood is defined as a function of ΔF amplitude and the elapsed time period (ΔT in hours) of the detected anomaly.
- Using the selected thresholds of the state parameters (ΔF, ΔT) the likelihood matrix is established as shown in Figure 2(a).
- The risk severity matrix is obtained, using a similar process, considering respectively the normalized Chlorscan concentration deviation data (ΔF) and the exposure period (ΔT_{ex} in hours). Therefore, using the selected thresholds of the state parameters (ΔF, ΔT_{ex}) the severity matrix is established as shown in Figure 2(b).
- Using the likelihood scale and the severity scale the risk matrix is defined as shown in Figure 3.
- Finally, using the risk scale, the time series of the risk indicator is defined based on the appropriate state color of the time step.

This paper illustrates (i) the application of the risk analysis method to typical 'synthetic' Chlorscan data of a real case study, (ii) the results of the numerical simulations obtained using the EPANET-MSX contaminant transport model to

(a)

Likelihood Matrix							
	Hours						
ΔF%	1	2	3	4	>4	Likelihood Scale (0-100%)	
0-4%						0 – 10 %	*Insignificant*
4-10%						10- 30 %	*Low*
10-20%						30-60 %	*Moderate*
20-30%						60-90%	*High*
>30%						>90	*Very High*

(b)

Severity Matrix							
	Exposure Hours						
ΔF%	0-1	1-3	3-12	12-24	>24	Severity Scale (1-5)	
0-4%						1	*Insignificant*
4-10%						2	*Low*
10-20%						3	*Moderate*
20-30%						4	*High*
>30%						5	*Very High*

Figure 2 | (a) Likelihood scale and likelihood matrix; (b) severity scale and severity matrix.

Risk Assessment Matrix							
Likelihood	Severity Scale						
Scale	1	2	3	4	5	Risk Scale (0-1)	
0-10%						0-0.1	*Insignificant*
10-30%						0.1-0.2	*Low*
30-60%						0.3-0.6	*Moderate*
60-90%						0.6-0.9	*High*
>90%						0.9-1	*Very High*

Figure 3 | Risk scale and risk assessment matrix.

bio-contamination propagation, and (iii) the detection of *E. coli* on the same real network.

Case study

The research project was deployed at the SUNRISE Demonstration Site located in Lille University Campus (a scientific city in the north of France). The campus simulated a small city as its features include: (i) 23,000 student, faculty and staff population; (ii) 15 km-length of subsurface water network, buried under the roads or under vegetated lands, located in walkable technical galleries; (iii) 140 buildings; (iv) 49 fire hydrants; (v) 250 valves; (vi) 93 Automatic Meter Readers (AMR) measuring hourly water consumption; (vi) five pressure sensors; and (vii) two Virtual District Metering Areas (VDMA). Figure 4 illustrates the configuration and sectorization of the demonstration site. An EPANET hydraulic model was integrated on a Google Earth platform. In this case study the system water demand was assumed to vary with hourly steps. One month of data was averaged to form a 1-day (24 hr) pattern used for the demand multiplying coefficients and the simulations were run for 1 day.

Application of risk analysis to typical Chlorscan data

Data from specific Chlorscans have been investigated in the present paper for the assessment of bio-contamination detection.

In order to calibrate, validate and test the application of the proposed analysis, Chlorscan data were used, as available:

Figure 4 | Configuration and characterization of Lille University, scientific city, France, as the demonstration site (Juran 2016; Cantos *et al.* 2017).

- Chlorscan 1: September 16, 2011, from 0 am to 9 am;
- Chlorscan 2: September 4, 2011, from 2.30 pm to 10.40 pm.

Since each time series was regularly updated every 10 minutes for 8–9 hours and directly aggregated into a data file once collected, they were continually pre-processed. Therefore, spurious dots were not presented and the pre-processing phase was not required.

Table 1 illustrates Chlorscan 2 data that were used in the analysis here reported.

RESULTS AND DISCUSSION

The proposed statistical procedure is shown below through the Chlorscan 2 data.

Figures 5 and 6 illustrate the Chlorscan 2 data analysis, as follows.

In detail, Figure 5 shows the trend in percentage of the normalized data along with time: the sensor is able to detect an anomaly from 5.30 pm to 7.50 pm. In particular, the sensor detects an initially increasing anomaly, which once it reaches its peak, tends to be constant. The statistical tools of the first, second and third standard deviations were used as default values for identifying the thresholds for the likelihood and the severity levels, reaching a very high likelihood red color-coded on the likelihood scale with ΔF greater than 30%. Operators can input threshold levels based on their experience. Figure 6 illustrates the time series of the risk indicator with its scale, taking into account the likelihood and the severity scale. It is evident that the risk indicator indicates a high risk level coded orange and an alarm should be emitted for the water utility operators in order to support the decision-makers in their decision for the public community.

Numerical simulation of bio-contamination detection

Since the EPANET-MSX numeric model was validated as shown above in the section 'State of Practice', it was possible to apply it to the same real case study of Lille University. An EPANET hydraulic model was created in order to simulate the behavior of the network towards bio-contamination injections. Then, the EPANET-MSX five-species model was used for running the simulations on the Lille network for 24 hours. In the absence of a pumping system capable of injecting chlorine into the distribution network, an initial chlorine concentration of 1 mg/L was assumed.

Table 1 | Chlorscan 2 data used in the statistics-based analysis

Chlorscan 1

Day	N. meas.	Chlor. (mg/L)
2011-09-04 14:30:00	1	164
2011-09-04 14:40:00	2	164
2011-09-04 14:50:00	3	163
2011-09-04 15:00:00	4	163
2011-09-04 15:10:00	5	166
2011-09-04 15:20:00	6	166
2011-09-04 15:30:00	7	162
2011-09-04 15:40:00	8	161
2011-09-04 15:50:00	9	154
2011-09-04 16:00:00	10	157
2011-09-04 16:10:00	11	152
2011-09-04 16:20:00	12	151
2011-09-04 16:30:00	13	158
2011-09-04 16:40:00	14	154
2011-09-04 16:50:00	15	157
2011-09-04 17:00:00	16	153
2011-09-04 17:10:00	17	151
2011-09-04 17:20:00	18	145
2011-09-04 17:30:00	19	146
2011-09-04 17:40:00	20	125
2011-09-04 17:50:00	21	93
2011-09-04 18:00:00	22	80
2011-09-04 18:10:00	23	75
2011-09-04 18:20:00	24	72
2011-09-04 18:30:00	25	74
2011-09-04 18:40:00	26	78
2011-09-04 18:50:00	27	68
2011-09-04 19:00:00	28	72
2011-09-04 19:10:00	29	75
2011-09-04 19:20:00	30	78
2011-09-04 19:30:00	31	69
2011-09-04 19:40:00	32	156
2011-09-04 19:50:00	33	162
2011-09-04 20:00:00	34	163
2011-09-04 20:10:00	35	159
2011-09-04 20:20:00	36	159
2011-09-04 20:30:00	37	160
2011-09-04 20:40:00	38	153

(*continued*)

Table 1 | continued

Chlorscan 1

Day	N. meas.	Chlor. (mg/L)
2011-09-04 20:50:00	39	161
2011-09-04 21:00:00	40	161
2011-09-04 21:10:00	41	161
2011-09-04 21:20:00	42	162
2011-09-04 21:30:00	43	163
2011-09-04 21:40:00	44	162
2011-09-04 21:50:00	45	162
2011-09-04 22:00:00	46	160
2011-09-04 22:10:00	47	161
2011-09-04 22:20:00	48	161
2011-09-04 22:30:00	49	161
2011-09-04 22:40:00	50	160

The monitored nodes are highlighted in Figure 7, for which the chlorine decay process is shown in Figure 8. As expected (Hallam *et al.* 2002), after an initial drop due to chlorine reactions with organic materials, the tank wall and substances present in the water (e.g. metals), the chlorine increases. Thus, the remaining chlorine is the total chlorine divided into (i) the amount of chlorine that has reacted with nitrates and is unavailable for disinfection, which is called the combined chlorine, and (ii) the free chlorine, which is the chlorine available to inactivate disease-causing organisms, and is used as a measure to determine the potability of water. Some of the monitored nodes were analyzed before and after *E. coli* injections in order to (i) illustrate the chlorine trend without any contaminant injections and (ii) corroborate the sudden reaction between chlorine and *E. coli*. Lines J-1, J-14, J-23, J-60 show the chlorine trend for the respective nodes: as long as the distance from the tank increases, the trend of the chlorine decay process results in a time delay of the concentration increase in the nodes J-1, J-14 and J-23. Node J-60, which is the farthest from the tank among the monitored nodes, illustrates the stabilization of the chlorine concentration.

Figure 8 also shows the EPANET-MSX simulations of *E. coli* injections. The *E. coli*, a Gram-negative rod-shaped bacterium, is injected at the J-1 node (Figure 7) with different concentrations: 2, 10 and 100 CFU (as units diluted in

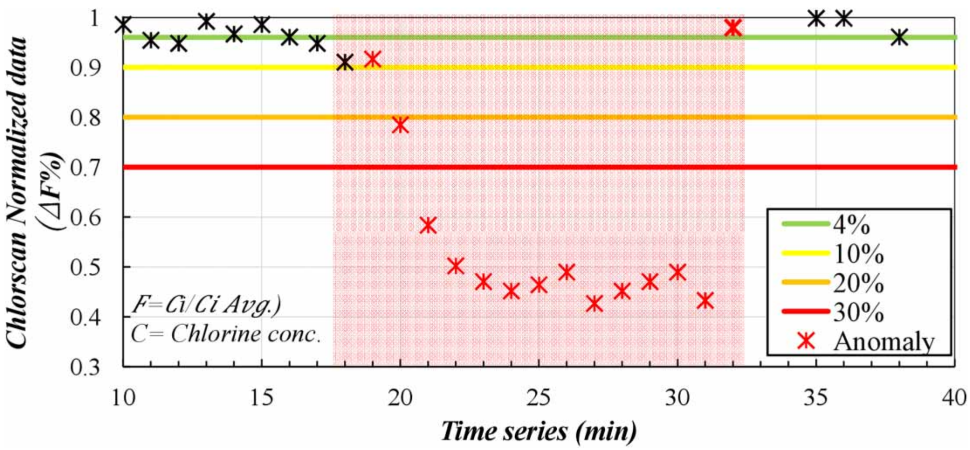

Figure 5 | Normalization of Chlorscan data to the average (*F*) and contamination likelihood assessment.

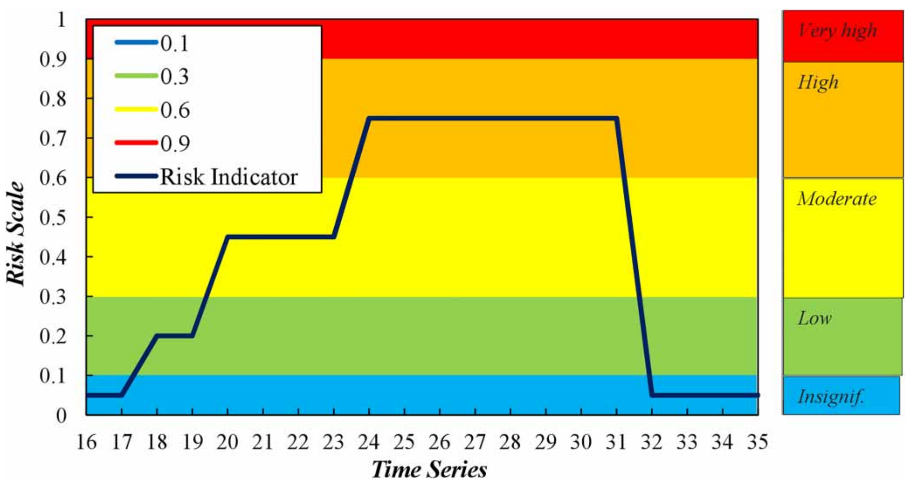

Figure 6 | Risk indicator with its scale for the Lille case study.

1 mL). The trend of the chlorine in the presence of the three *E. coli* concentrations is shown for the J-14 node with lines EC_2, EC_10, EC_100. It illustrates that the chlorine drops down to almost zero in the presence of the bacterium without reaching zero. The results are consistent with the laboratory experiments and their simulations, proving that (i) the chlorine, which is one of the most common disinfectants, is consumed when it reacts with pathogens and (ii) the chlorine drops down to zero only when the *E. coli* injections reach the order of magnitude of 10^8 CFU/mL (Tinelli & Juran 2017). Therefore, the analysis showed a clear possibility of detecting *E. coli* bacteria by analyzing the level of the chlorine in the network. In fact, the *E. coli* presents an immediate effect on the network: if the *E. coli* quantity increases, the level of total chlorine directly decreases.

CONCLUSIONS

Since contamination events can occur in a water distribution network for several reasons, risking many lives within minutes, this paper aims to underline the importance of gaining knowledge regarding the early detection of such events within a WDS.

Applying statistical concepts on one side and engineering models (EPANET-MSX) on the other side, the authors

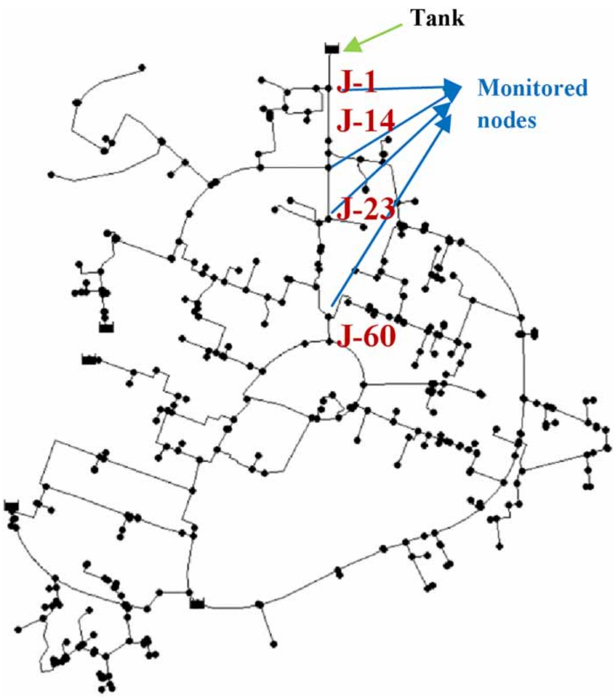

Figure 7 | Monitored nodes in the Lille network.

showed the behavior of total chlorine in a specific network and proposed a risk assessment analysis capable of producing a decision support system integrated with a color-based bio-anomaly detection protocol, filtering false alarms, in order to support water utility managers in their decisions for citizens' safety.

In this context, the paper identifies the software EPANET-MSX as an essential support for qualitative studies in WDSs because, unlike the traditional software used in the literature, it describes dynamic interactions among the contaminants, the water and the pipe/tank walls.

In fact, the multi-species numerical model used in the EPANET-MSX software to carry out the numerical simulations is able to show an initial drop in the chlorine due to its reactions with organic materials, tank walls and other substances present in the water. Then, the model reports residual chlorine stabilization, and it correctly corroborates the fact that the presence of natural or injected organic matter (like *E. coli*) in the WDS plays a vital role in the fate and transportation of chlorine.

Both numerical simulations and experimental model results demonstrate that injections of *E. coli* result in significant reduction of the free chlorine to a residual level which mostly depends on the chlorine concentrations.

The results of this research confirm the feasibility of early detection of bio-anomalies (such as *E. coli*) in the drinking WDS through the use of the chlorine trend: in particular, chlorine measurements were exploited to develop an automated prototype system for early anomaly detection, that efficiently enables water operators to ensure a real-time management of the procedures for water quality control, as well as a preemptive decision-making process.

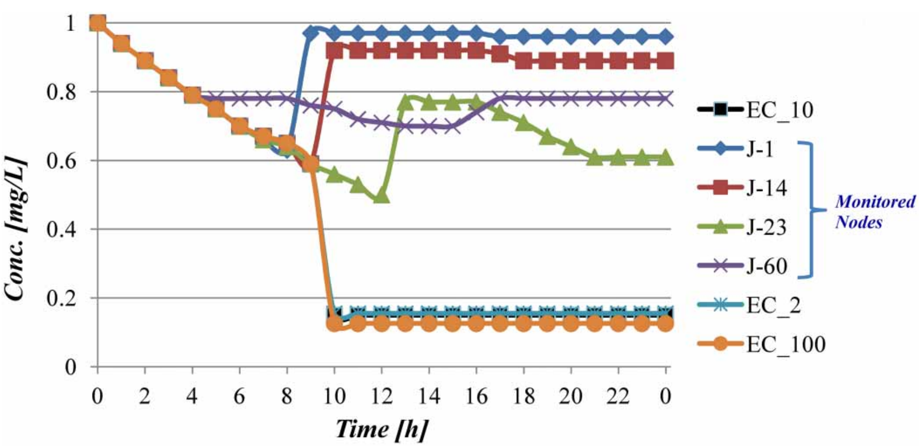

Figure 8 | Chlorine concentrations before *E. coli* injections for nodes J-1, J-14, J-23 and chlorine concentration for node J-14 after *E. coli* injections of 2, 10 and 100 CFU (as units diluted in 1 mL).

ACKNOWLEDGEMENTS

The authors acknowledge and thank WS4EU Consortium Partners for support and data sharing and in particular Prof. Isam Shahrour for access to the University of Lille demonstration site data, technical discussion on the laboratory pilot testing and advice on network simulation, as well as W-SMART for access to the operator experience of water utilities.

REFERENCES

Abdallah, A. 2015 *Reseaux d'eau potable: surveillance de la qualite de l'eau par des capteurs en ligne*. PhD thesis, University of Lille, France.

Allgeier, S. C., Umberg, K., Johnson, R. & Sahni, N. 2011 Detection of distribution system contamination incidents using online water quality monitoring. In: *Proceedings of the AWWA Water Quality Technology Conference, Phoenix*, 13–17 November. American Water Works Association, Denver, CO, USA.

Blokker, E. J. M., Vreeburg, J. & Speight, V. 2013 Residual chlorine in the extremities of the drinking water distribution system: the influence of stochastic water demands. In: *12th International Conference on Computing and Control for the Water Industry*, Perugia, Italy.

Boccelli, D. L., Tryby, M. E., Uber, J. G. & Summers, R. S. 2003 A reactive species model for chlorine decay and THM formation under rechlorination conditions. *Water Research* **37** (11), 2654–2666.

Cantos, W. P., Tinelli, S. & Juran, I. 2017 Risk assessment for early water leak detection. In: *ASCE International Conference on Sustainable Infrastructure* (L. Soibelman & F. Peña-Mora, eds), ASCE, Reston, VA, USA, pp. 306–318.

Davis, M. J. & Janke, R. 2009 Development of a probabilistic timing model for the ingestion of tap water. *Journal of Water Resources Planning and Management* **135** (5), 397–405.

Hagar, J., Murray, R., Haxton, T., Hall, J. & McKenna, S. 2013 *The Potential for Using the CANARY Event Detection Software to Enhance Security and Improve Water Quality*. Center for Homeland Defense and Security, Monterey, CA, USA.

Hallam, N. B., West, J. R., Forster, C. F., Powell, J. C. & Spencer, I. 2002 The decay of chlorine associated with the pipe wall in water distribution systems. *Water Research* **36**, 3479–3488.

Hart, D., McKenna, S. A., Klise, K., Cruz, V. & Wilson, M. 2007 CANARY: a water quality event detection algorithm development tool. In: *Proc. World Environmental and Water Resources Congress* (K. C. Kabbes, ed.), ASCE, Reston, VA, USA.

Hooke, R. & Jeeves, T. A. 1961 'Direct search' solution of numerical and statistical problems. *J. ACM* **8** (2), 212–229.

Juran, I. 2016 *EU-FP7 SW4EU Interim Progress Report*. W-SMART Association.

Koch, M. W. & McKenna, S. A. 2011 Distributed sensor fusion in water quality event detection. *J. Water Res. Pl. ASCE* **137** (1), 10–19.

Kroll, D. J. 2006 *Securing Our Water Supply: Protecting a Vulnerable Resource*. PennWell Corp., Tulsa, OK, USA.

Rosario-Ortiz, F., Rose, J., Speight, V., von Gunten, U. & Schnoor, J. 2016 How do you like your tap water? *Science* **351** (6276), 912–914.

Shang, F., Uber, J. G. & Rossman, L. 2008 *EPANET Multi-Species Extension (EPANET-MSX) Software and User's Manual*. US Environmental Protection Agency, Washington, DC, USA, EPA/600/C-10/002.

Teunis, P. F. M., Xu, M., Fleming, K. K., Yang, J., Moe, C. L. & LeChevallier, M. W. 2010 Enteric virus infection risk from intrusion of sewage into a drinking water distribution network. *Environmental Science & Technology* **44** (22), 8561–8566. doi:10.1021/es101266 k.

Tinelli, S. & Juran, I. 2017 Numerical modeling of early bio-contamination in a water distribution system and comparison with laboratory experiments. In: *ASCE International Conference on Sustainable Infrastructure* (L. Soibelman & F. Peña-Mora, eds), ASCE, Reston, VA, USA, pp. 258–269.

US-EPA 2010 *Water Quality Event Detection Systems for Drinking Water Contamination Warning Systems*. US Environmental Protection Agency, Washington, DC, USA, EPA 600/R-010/036.

US-EPA 2012 *CANARY Users Manual Version 4.3.2*. US Environmental Protection Agency, Washington, DC, USA, EPA 600/R-08/040B.

US-EPA 2013 *Water Quality Event Detection System Challenge: Methodology and Findings*. US Environmental Protection Agency, Office of Water, Washington, DC, USA, EPA 817-R-13-002.

Van Thienen, P. 2014 Alternative strategies for optimal water quality sensor placement in drinking water distribution networks. In: *11th Int. Conf. on Hydroinformatics, HIC2014*, New York, USA.

Vugrin, E., McKenna, S. A. & Hart, D. 2009 Trajectory clustering approach for reducing water quality event false alarms. In: *Proceedings, World Environmental and Water Resources Congress* (S. Starrett, ed.), ASCE, Reston, VA, USA.

Yang, J., LeChevallier, M. W., Teunis, P. F. & Xu, M. 2011 Managing risks from virus intrusion into water distribution systems due to pressure transients. *Journal of Water & Health* **9** (2), 291–305.

First received 7 September 2017; accepted in revised form 31 January 2018. Available online 15 February 2018