

# FIELD STUDY OF REAL-TIME WATER QUALITY CONTROL

CHRISTINE SAAB<sup>1,2</sup>, ELIAS FARAH<sup>1</sup>, ISAM SHAHROUR<sup>1</sup> & FADI HAGE CHEHADE<sup>2</sup>

<sup>1</sup>Laboratory of Civil Engineering and Geo-Environment, University of Lille, France

<sup>2</sup>Civil Engineering Department, Doctoral School of Sciences and Technology, Lebanese University, Lebanon

## ABSTRACT

The objective of a drinking water network is to provide a good quality of water to users. Accidental or intentional contamination can degrade the water quality and consequently threatens the consumer's health. Generally, the water quality is monitored using traditional methods, based on manual sampling, which can take several days. Early warning of water contamination can be achieved using smart technology. This paper presents a field study of the use of this technology in real-time monitoring of the water quality. The field study is conducted at the Campus of the University of Lille in the North of France within the European Project "SmartWater4Europe" (<http://www.sw4eu.com>). Two sensors are installed in the campus: S::CAN and EventLab which measure several water quality parameters such as TOC (Total Organic Carbon), turbidity, refractive index, etc. This paper presents analysis of these parameters as well as the influence of hydraulic parameters on the water quality. It presents also an event detection system, which is developed using CANARY software. A sensitivity study is presented to determine the appropriate parameters in order to reduce false alarms and to determine the probability of possible event.

*Keywords: water quality, monitoring, event detection, smart technology.*

## 1 INTRODUCTION

The main objective of the water supply is to ensure a safe drinking water for consumers. Water distribution system (WDS) can be subjected to accidental contamination or malicious attack which can affect the water quality. Although treatment works are closely monitored and controlled in the developed countries, this is not the case for all less developed countries. Many factors such as backflow, pipes break, physico-chemical and biological reactions and corrosion, can induce the deterioration of water quality.

In order to protect public health from contamination events, water utilities are concerned by a real-time monitoring of the water quality. In general, water quality control is based on taking manual samples in a periodic basis, and then doing several laboratory analyses: i) physico-chemical to determine organoleptic characteristics of tested water; and ii) microbiological to identify the presence of pathogenic microorganisms. These traditional methods can take from several hours to some days.

Contamination could occur in WDS and causes serious diseases. Recently, a contamination by industrial chemical (MCHM) occurred in the distribution networks of Charleston, West Virginia (USA) and affected more than 300,000 consumers [1].

In order to prevent contamination events earlier and to protect the water network from any intrusion, water companies could use smart technology, which allow a rapid detection of abnormal events. On-line water quality sensors can be used to enhance the monitoring of the water quality in real-time [2]. According to a recent survey by Walsby [3], the smart technology could lead to high savings in water industry (between 7.1 and 12.5 billion US\$ per year). The use of the smart technology in real-time monitoring of the water quality is recent. It requires yet both laboratory and field studies for its implementation at high scale. This paper concerns this issue, it presents the results of the use of the smart technology for



the real-time control of the water quality in the Scientific Campus of the University of Lille, which stands for a town of around 25,000 users.

## 2 MATERIALS AND METHODS

Within the European project SmartWater4Europe, we have used S::CAN and Optiqua EventLab for the on-line control of the water quality.

EventLab uses an optical probe that ensures an Early Warning System (EWS) for water distribution grid, without the need of reagents or any consumables. It can be installed and accessed for servicing without interrupting the main flow [4]. Based on Mach-Zehnder Interferometer (MZI) principle, this device measure continuously, each minute, the variation of refractive index (RI), which is an effective indicator of the status of water quality. Any substance dissolved in water affects the refractive index of the water matrix [5]. Basing on its benefits, the refractive index (RI) is used for early alert of water contamination. The use of refractive index (RI) has a number of advantages for water quality monitoring and the detection of water quality incidents: consistent sensitivity for all substances, response linear with concentration and high resistance to matrix interference [6]. The variation of refractive index (RI) is illustrated by the measurement of variation of phase to which it is directly proportional. The water temperature is also monitored. Data is accessible via server web. This sensor has proved a high sensitivity in the rapid detection of any type of chemical contaminant: organic or inorganic [4].

S::CAN micro::station allows an online monitoring of various water quality parameters. It contains three main elements: Spectro::lyser, S::CAN probes and controller assembled on one panel. With a 3G SIM card, the data are transmitted continuously to a web server that can be connected to a Supervisory Control and Data Acquisition (SCADA) system. A con::cube terminal with moni::tool software ensures data acquisition, data display and station control [7]. The main function of this device is to detect anomalies in the controlled parameters. A reference line for each parameter is obtained for a good quality of water. Then, any significant deviation from the reference is detected as possible events occurred in water, due to the potential presence of a contaminant. Four main probes ensure the measurement, every minute, of the water quality parameters:

- i::scan probe allows the control of various reliable indicator of quality, the absorbance UV254, the turbidity (EPA and ISO), organic substance by measuring TOC (Total Organic Carbon) and DOC (Dissolved Organic Carbon), the color and the temperature.
- Condu::lyser measure the conductivity, as indicator of water mineralization, and the temperature.
- pH::lyser ensure the measurement of pH and temperature.
- Chlo::lyser evaluate the quantity of free chlorine in water.

In addition to the deployment of quality sensors in the distribution network, an Event Detection System (EDS) constitutes an important part of the Contamination Warning System (CWS). An EDS can ensure an interface that analyses real-time data and detects any unexpected data. The open source software Canary is used in this work for analysis of water quality data collected in the Campus. The main purpose of Canary is to analyze, in near real-time, the data (water quality signals and operations data) from a SCADA System. Then, it identifies the probability of possible events in water. An event in the water signal is defined as an anomalous reading from sensors that differ from normal data. Water quality event detection algorithms within Canary automatically identify significant deviations from expected water quality values [8].



The detection methodology used in Canary software can be summarized in four steps [8]. Firstly, a future value is predicted based on historical data. This can be done by two main approaches: Linear Prediction Coefficient Filter (LPCF) and Multivariate Nearest Neighbor (MVNN). The LPCF method uses the weighted average of previous data to estimate the value at the next step. The MVNN approach compares the Euclidean distance between the current measured water quality and any water quality measurement within the recent past in the multivariate space [9]. Once the value at the next step is available from the sensors, it is compared with the estimated one by calculating the residual (difference between observed and predicted value) in units of standard deviation. The maximum residual is then compared with a user predefined threshold, and classified as an outlier when it exceeds the limit. The final step is Binomial Event Discriminator (BED), based on the binomial distribution, which calculates the probability of event as a function of the number of outliers in a defined window. It is important to note that Canary works in both offline (historical data) and online (via SCADA connection) modes.

### 3 WATER QUALITY CONTROL OF THE SCIENTIFIC CAMPUS OF THE UNIVERSITY OF LILLE

#### 3.1 Site description

The water quality control system was implemented at the water distribution network of the Scientific Campus of the University of Lille. This campus is located in Villeneuve d'Ascq, Lille in the north of France. It stands for a small town. It covers an area of around 110 hectares hosting around 25,000 users in 150 buildings (for administration, teaching and research, student's residence, restaurant). The campus includes about 100 km of urban networks (drinking water, electrical network, heating and sewage).

The water network length is about 15 km of grey cast iron pipes (diameter between 20 and 300 mm). It is equipped with 250 isolations valves, 49 hydrants for fire fighting, in addition of stabilizers and purges [10]. The campus is supplied at five sections located in the North, West and South of the campus.

#### 3.2 Instrumentations

In order to cover the majority of zones in the campus and to compare the water quality in two types of building's usage, two locations has been chosen for the installation of EventLab and S::CAN (Fig. 1). The first installation concerns the engineering school Polytech'Lille, while the second location is located at Barrois restaurant. The installation uses a main pipeline passing near the target location, a connection is taken from this pipe and then the sensors are installed on this connection. The water samples are taken continuously by sensors which measure each minute various parameters and then the water passes throughout a discharge.

### 4 ANALYSIS OF WATER QUALITY SIGNALS

#### 4.1 EventLab

Any change in the composition of the water matrix will induce a change in the combined refractive index (RI). These changes in refractive index (RI) are monitored by EventLab [6].

A linear relationship is maintained between refractive index (RI) and concentration of a contaminant [11]. Fig. 2(a) illustrates the concentration of Sodium Sulfate in pure water and the resulting change in refractive index (RI) [12]. Fig. 2(b) shows the response of EventLab after the injection of different concentrations of chemical product (Chloride of Mercury  $\text{HgCl}_2$ ) [13].



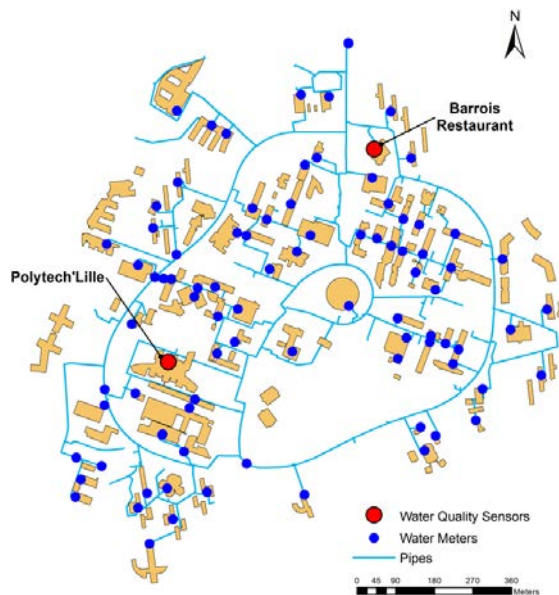


Figure 1: Water distribution network of the campus with sensors locations.

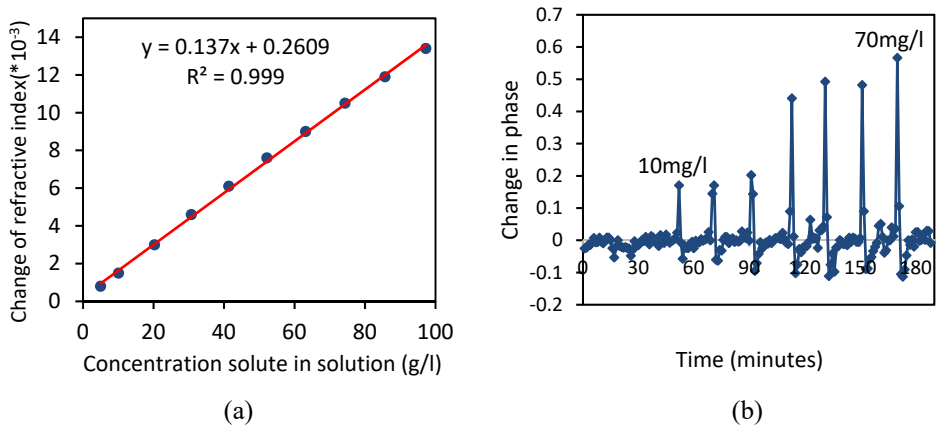


Figure 2: EventLab. (a) Variation of refractive index (RI) as function of the concentration of Sodium Sulfate (Source: Wolf, 1984.); (b) Response of EventLab after injection of Chloride of Mercury  $HgCL_2$  [13]. (Source: Abdallah [13].)

The main response of EventLab system concerns the phase measured each minute. In fact, the change in refractive index of the water  $\Delta n_{water}$  is given by [11]:

$$\Delta n_{water} = 4 \times 10^{-4} (\Delta \Phi m / 2\pi). \quad (1)$$

At each time step  $i$ , variation of phase,

$$\Delta \Phi (i) = \Phi (i + 1) - \Phi (i). \quad (2)$$



In normal drinking water, the refractive index is quasi-constant as well as the variation of phase. Any exceed from the normal variation ( $\pm 3\sigma$ ) is considered as an event [14], which could indicate a possible contamination. An example of EventLab response is illustrated in Fig. 3 for the period of October 3 and 4, 2016 for the location (Polytech'Lille). We observe that the majority of variation does not exceed the accepted limit, which indicates a safe drinking water. However, some events are detected (red circle in Fig. 3). In order to find an explanation for these events, the water consumption profile is displayed in Fig. 3. It shows that a sudden increase in the consumption can affect water signal and cause an outlier in the response. This result indicates that the water flow due to consumption could be responsible of the extraction of substances from the aging water pipes.

#### 4.2 S::CAN

Multiple reliable indicators of water quality are continuously measured by S::CAN, each parameter is represented by a signal. For safe drinking water, signals are constant and below the standard thresholds. Therefore, any anomaly in the water affects the stability of the baseline reference and could induce a significant deviation. Fig. 4 shows data collected by S::CAN in July, 2016 at Polytech'Lille. Although the signals are generally quasi constant, some events could be identified. In order to analyze the source of these anomalies, the hydraulic parameters (consumption and pressure) are displayed in Fig. 4. We observe that pressure does not affect the water quality, while a significant correlation is observed between consumption and peak of signals. An increase in the consumption is accompanied by an increase in the quality parameters: During the night, the consumption is low and all signals are stable, while the water consumption in the morning, induces a variation of the majority of water quality parameters, in particular UV and turbidity.

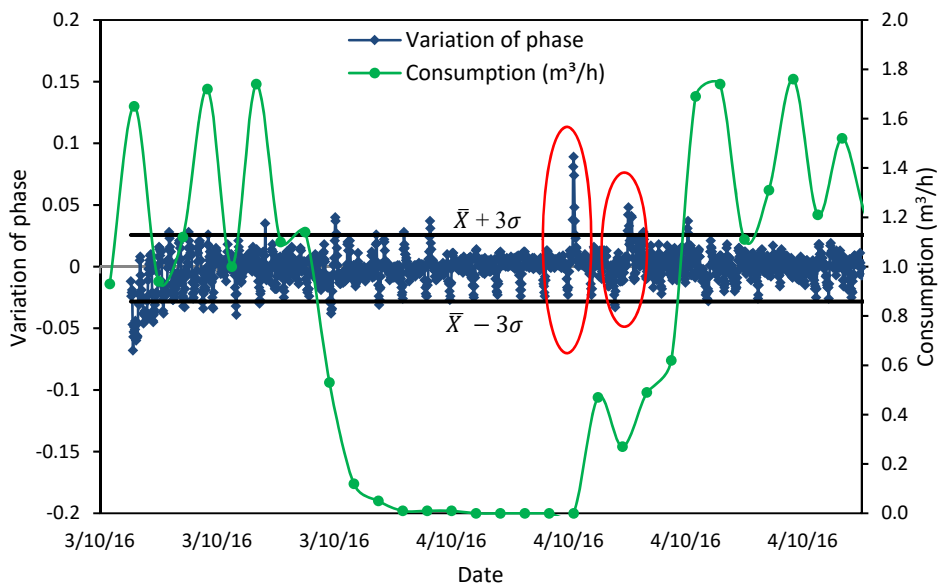


Figure 3: EventLab response with consumption profile for October 3 and 4, 2016.



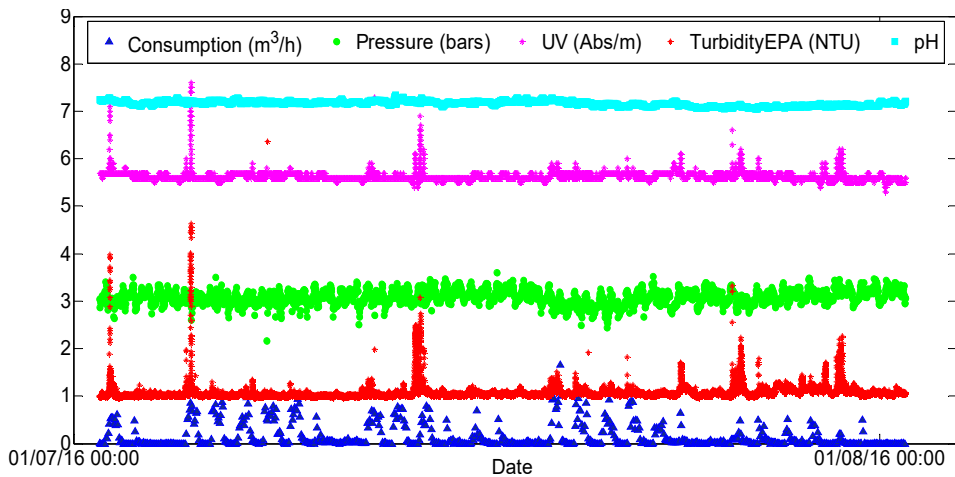


Figure 4: SCAN response with hydraulics parameters for July 2016.

## 5 SENSITIVITY ANALYSIS

### 5.1 Methodology

In order to test the ability of CANARY, as an EDS, to identify earlier any anomaly in water signal, a performance analysis is needed for a set of historical data. A water contamination event is identified when the real-time water quality data are outside the expected range of allowable water quality criteria, at which point an alert is issued [9]. The main objective is to configure the EDS in a way to decrease the number of missed detection, with the minimum number of false alarms. For this purpose, a configuration file, written in YML markup language, has to be adjusted to find the best parameters ensuring a high detection of events. Five principal parameters affect generally the detection capacity of the software:

- Event Detection algorithm (LPCF or MVNN): used in the first prediction step of future value of water parameter.
- History window: defines the number, in units of time steps, of previous data used to estimate the value at the next time step.
- Event threshold: determines the threshold probability of Event ( $P(\text{event})$ ) to generate an alarm.
- Outlier threshold: measured in unit of standard deviation, and indicates the limit to be exceeded to identify an outlier.
- BED window: determines the number of time steps for the binomial distribution.

The S::CAN sensor installed in Polytech'Lille is chosen to identify the appropriate parameters that adjust the configuration file. Ten parameters were measured: UV254, Turbidity ISO, Turbidity EPA, TOC, Color, Temperature1 (from i::scan), Conductivity, Temperature2 (from Condu::lyser), pH and free chlorine. The collected data cover the period from August 1, 2016 till October 10, 2016. During this period, about 29 events were detected in water signals. The time step selected for Canary running is 1 minute (same that measurement interval of S::CAN sensor). The aim of the sensitivity analysis conducted is to select adequate values for the five parameters cited above.

## 5.2 ROC curve analysis

Before explaining the concept of Receiver Operating Characteristics (ROC) curve, it is important to define the basic outcomes of this analysis, as follows:

- True Positive (TP): A real event is detected. The sensors detected an event and the method reported it.
- False Negative (FN): An actual event is not detected. The method does not identify a real event.
- False Positive (FP): Generation of false alarm, while there is no real event.
- True Negative (TN): There is no real event and the method does not generate an alarm.

The sensitivity and the specificity are calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \text{TPR (True Positive Rate)}. \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \text{TNR (True Negative Rate)}. \quad (4)$$

By definition, the ROC curve is a technique for visualizing, organizing and selecting classifiers based on their performance [15]. In fact, it is two-dimensional graphs where TPR is plotted in function of  $(1 - \text{TNR}) = \text{FPR}$  (False Positive Rate). The main objective is to ensure a correct detection with minimum of false alarms by making a tradeoff between increasing the sensitivity and decreasing the specificity. This can be represented by the point (0, 1) in the ROC curve where both sensitivity and specificity reach 100%. The best performance of the detection algorithm occurs when the result is closest to this point (0, 1) in the graph. The ROC curve analysis is used in this study to select the appropriate configuration file for the water quality monitoring.

## 5.3 Selection of event detection algorithm

For each of the two types of algorithms (LPCF and MVNN), the precision and the accuracy are calculated in terms of TP, FP, TN and FN as follows:

$$\text{Precision (PPV)} = \frac{TP}{TP + FP}. \quad (5)$$

$$\text{Accuracy (ACC)} = \frac{TP + TN}{TP + FN + FP + TN}. \quad (6)$$

In fact, for the two algorithms, we determine the precision and accuracy for five cases of history window from 0.5 days (720 time steps) till 3 days (4320 time steps). However, other parameters are fixed as follows: outlier threshold 0.85, event threshold 0.99 and BED window 200. Fig. 5(a) shows that a better precision is obtained for the case of LPCF algorithm. The precision in LPCF case is between 62% and 85%, while for the MVNN case, the precision is below 49%. In addition, Fig. 5(b) indicates different values of accuracy evaluated for the five history window for each algorithm. It is clear that the LPCF method is more accurate (75%), but the accuracy in MVNN case is between 40 and 60%. From this result, we choose obviously the LPCF algorithm.



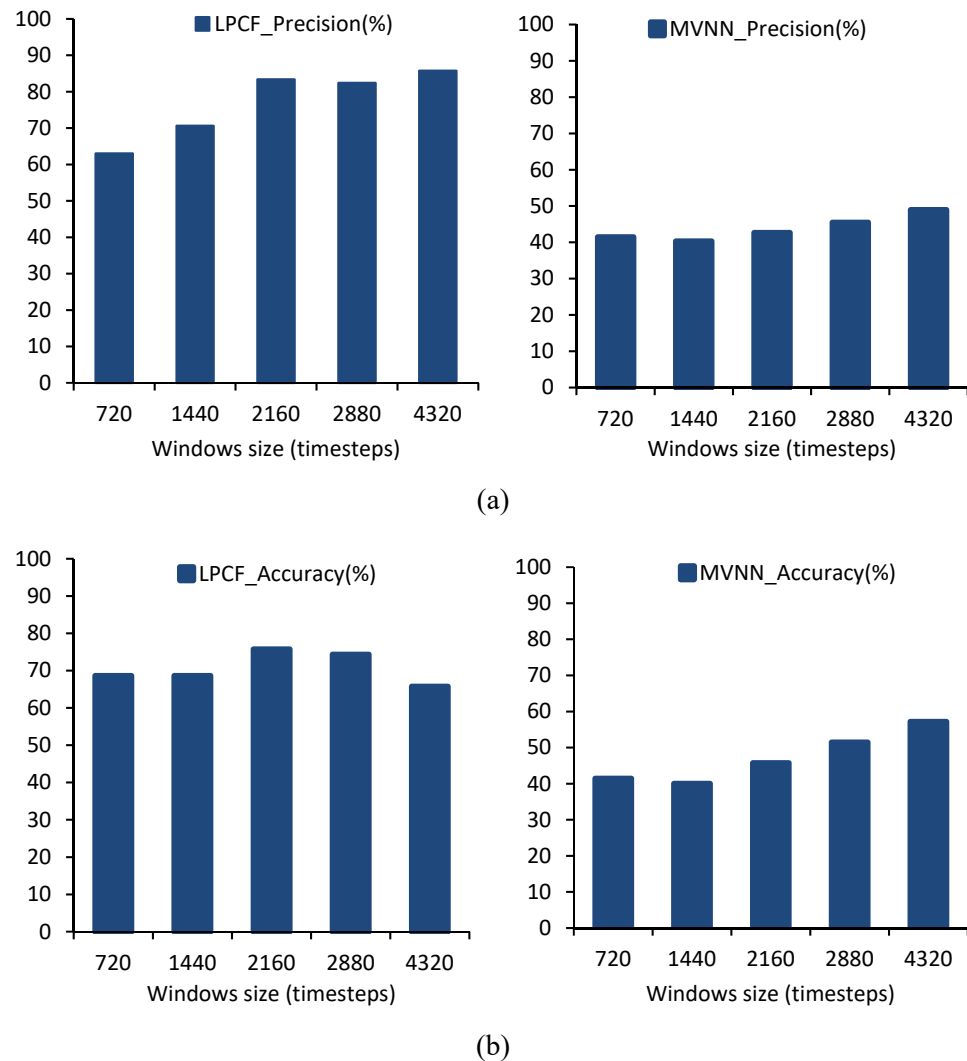


Figure 5: Comparison of detection algorithm. (a) Precision; (b) Accuracy.

#### 5.4 Selection of history window

The selection of the best history window is based on the calculation of the absolute average residuals and the corresponding standard deviations. The results of these calculations are illustrated in Fig. 6. The accurate predicted values are obtained in case of lowest average residual and standard deviation. Fig. 6 shows that from a window size of 2160 (1.5 days), the values of residual (Fig. 6(a)) and standard deviation (Fig. 6(b)) reach the minimum and start to converge. In a way to reduce the computational time, a choice of 1.5 days as history window is considered precise and useful. An exception is observed in Turbidity EPA after 1.5 days, this can be due to some fluctuations in the signal during the training phase.





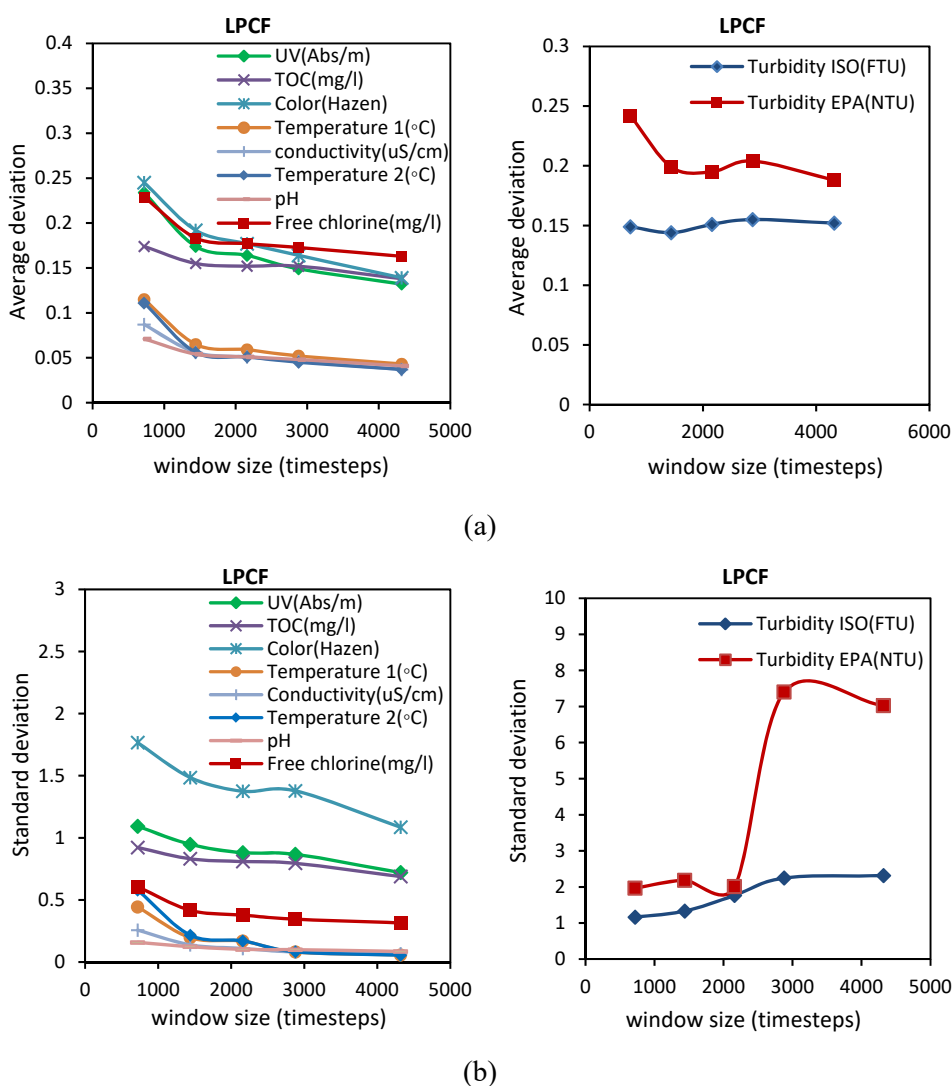


Figure 6: Comparison of history window. (a) Average deviation; (b) Standard deviation.

### 5.5 Determination of event threshold

This parameter indicates a maximum acceptable value of probability. When this limit is exceeded, an alarm of event is generated. A parametric study was conducted by varying the event threshold between 0.75, 0.85, 0.9 and 0.99. Other parameters are fixed (LPCF, history 1.5 day, outlier threshold 0.85, BED window 200). The sensitivity analysis shows close results between different event thresholds, so we choose the value 0.99 that minimize the FPR with the high precision and accuracy. The sensitivity analysis is detailed in Table 1.



Table 1: Sensitivity analysis for different event threshold.

Event threshold	TPR (%)	TNR (%)	FPR (%)	PPV (%)	ACC (%)
0.99	51.72	92.68	7.32	83.33	75.71
0.90	48.28	92.68	7.32	82.35	74.29
0.85	48.28	92.68	7.32	82.35	74.29
0.75	51.72	92.68	7.32	83.33	75.71

5.6 Parametric study for outlier threshold and BED window

The selection of outlier threshold and BED window is done by a parametric analysis, where five values for each parameter are taken into consideration. The threshold values, that give the better detection of obvious variations in water quality, were chosen. The BED window is selected in a way of making tradeoff between the rapid detection of event and the generation of false positive alarms. Fig. 7 shows the calculation of average prediction residual using LPCF algorithm with 1.5 days as history window and 0.99 as event threshold. The residual is minimal (0.108) for the combination of 0.75 as outlier threshold and 50 as BED window.

In order to verify the selected parameters, the sensitivity analysis using ROC curve was performed by fixing 0.75 for the threshold value and varying the BED window values. To increase the probability of detection (TPR) with low value of false alarm (FPR), a BED window of 100 is more appropriate.

In the same way, the ROC curve analysis is plotted for multiple values of outlier threshold for both BED window 50 and 100. The results are shown in Fig. 8. A high sensitivity with an accepted probability of false events is obtained with a combination of an outlier threshold of 0.75 and a BED window of 100.

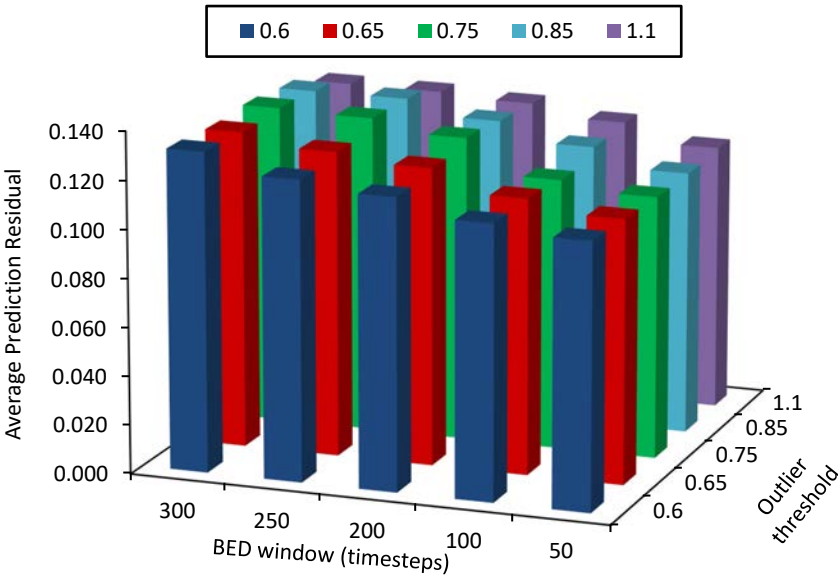


Figure 7: Average prediction residual in function of outlier threshold and BED window.

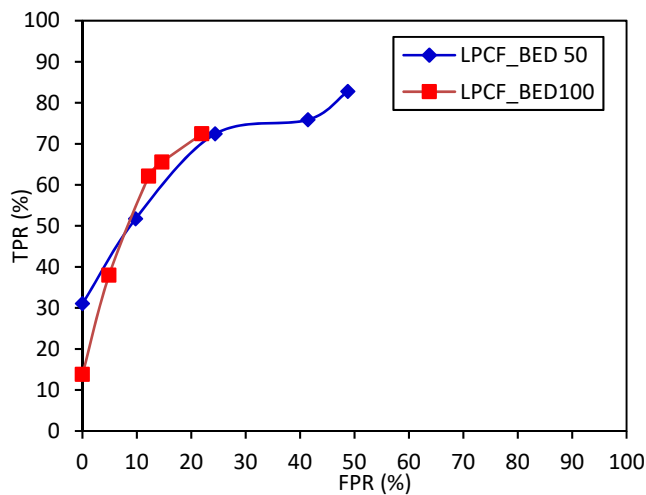


Figure 8: ROC curve for multiple outlier thresholds for a BED window of 50 and 100.

### 5.7 Application of selected parameters

The adjusted configuration file, obtained from the training phase, is applied to test the data from October 10, 2016 till November 1, 2016. During this period, ten events were observed in the water signals. Using Canary as event detection algorithm, with the selected parameters, 7 real events have been successfully detected, with 30% as false alarm rate. An example of results is given in Fig. 9, which shows the probability of event detection for the period from October 10, till October 17, 2016 according to LPCF.

## 6 CONCLUSION

The real-time monitoring of water quality constitutes a great challenge for a safe water supply to users. The use of smart technology allows water companies to conduct an on-line effective control of the water quality and to develop an early warning strategy for the safety of the water supply. This paper has presented the application of this technology in the Scientific Campus of the University of Lille. Two on-line water quality sensors (S::CAN and EventLab) were installed in two sections of the campus. The use of these sensors showed some events in water quality, which were correlated with water flow due to consumption, in particular in the morning.

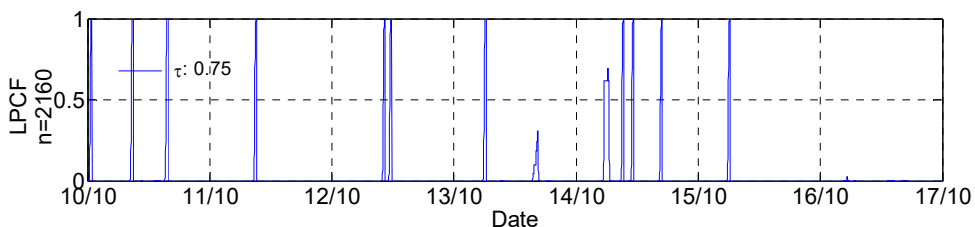


Figure 9: Probability of event for the week of October 10, 2016.

A sensitivity study of event detection algorithm was conducted using Canary software. Analysis led to an appropriate selection of parameters, which is based on increasing the probability of detection and reducing the generation of false alarms.

#### REFERENCES

- [1] Cooper, W.J., Responding to crisis: The West Virginia Chemical Spill. *Environmental Science and Technology*, **48**(6), pp. 3095–3095, 2014.
- [2] Eliades, D.G., Contamination Event Detection in Water Distribution Systems using a Model-Based Approach. *Proceedings of 16<sup>th</sup> Water Distribution System Analysis Conference*, pp. 1089–1096, 2014.
- [3] Walsby, C., The power of smart water networks. *Journal AWWA*, **105**(3), pp. 72–77, 2013.
- [4] Optiqua EventLab, Real time on-line water quality monitoring, [www.optiqua.com](http://www.optiqua.com)
- [5] van Wijlen, M.A.B., et al., Innovative sensor technology for effective online water quality monitoring. *Proceedings of the Singapore International Water Week*, 2011.
- [6] Broeke, J. van den., The benefits of using refractive index for water quality monitoring in distribution network, a comparison of Optiqua EventLab and traditional water quality parameters, Optiqua, Published by Optiqua Technologies, [www.optiqua.com](http://www.optiqua.com)
- [7] Micro::station, Monitoring Stations, [http://www.s-can.at/medialibrary/datasheets/microstation\\_dw\\_hp.pdf](http://www.s-can.at/medialibrary/datasheets/microstation_dw_hp.pdf)
- [8] Murray, R., et al., *Water Quality Event Detection Systems for Drinking Water Contamination Warning Systems, Development, Testing and Application of CANARY*, US Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Sec. 1, 2010.
- [9] Hou, D., et al., Detection of water-quality contamination events based on multi-sensor fusion using an extended Dempster-Shafer method. *Measurement Science and Technology*, **24**(5), 18 pp., 2013.
- [10] Farah, E., et al., SunRise: Large scale demonstrator of the Smart Water System. *International Journal of Sustainable Development and Planning*, **12**(1), pp. 112–121, 2017.
- [11] Tangena, B.H., et al., A novel approach for early warning of drinking water contamination events. *Water Contamination Emergencies: Monitoring, Understanding and Acting*, pp. 13–31, 2011.
- [12] Wolf, A.V., et al., Concentration properties of aqueous solutions. *Handbook of Chemistry and Physics* (65th ed.), CRS Press, 1984–1985.
- [13] Abdallah, A., *Réseaux d'eau intelligents: Surveillance de la qualité de l'eau par des capteurs en ligne*, Doctoral dissertation, Lille, 2015.
- [14] Saw, J.G., et al., Chebyshev inequality with estimated mean and variance. *The American Statistician*, **38**(2), pp. 130–132, 1984.
- [15] Fawcett, T., An introduction to ROC analysis. *Pattern Recognition letters*, **27**(8), pp. 861–874, 2006.

