Detection of water quality failure events at treatment works using a hybrid two-stage method with CUSUM and random forest algorithms

Gerald Riss, Michele Romano, Fayyaz Ali Memon and Zoran Kapelan

ABSTRACT

Near-real-time event detection is crucial for water utilities to be able to detect failure events in their water treatment works (WTW) quickly and efficiently. This paper presents a new method for an automated, near-real-time recognition of failure events at WTWs by the application of combined statistical process control and machine-learning techniques. The resulting novel hybrid CUSUM event recognition system (HC-ERS) uses two distinct detection methodologies: one for fault detection at the level of individual water quality signals and the second for the recognition of faulty processes at the WTW level. HC-ERS was tested and validated on historical failure events at a real-life UK WTW. The new methodology proved to be effective in the detection of failure events, achieving a high true-detection rate of 82% combined with a low false-alarm rate (average 0.3 false alarms per week), reaching a peak F_1 score of 84% as a measure of accuracy. The new method also demonstrated higher accuracy compared with the CANARY detection methodology. When applied to real-world data, the HC-ERS method showed the capability to detect faulty processes at WTW automatically and reliably, and hence potential for practical application in the water industry.

Key words | CUSUM, event recognition, online monitoring, random forest, water treatment works

HIGHLIGHTS

- The novel HC-ERS combines the conventional SPC-type method with RF advanced machine-learning technique to ultimately detect WTW-level failure events.
- When applied on unseen data HC-ERS proved to be capable of detecting failure events in WTW processes in near-real-time.
- HC-ERS outperformed threshold-based and CANARY event detection methods.
- HC-ERS showed potential for practical application in the water industry.

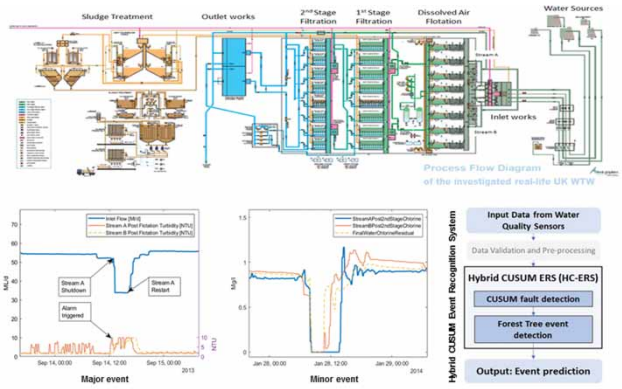
Gerald Riss (corresponding author)
Fayyaz Ali Memon
Centre for Water Systems,
University of Exeter,
Harrison Building, North Park Road, Exeter,
Devon,
UK
E-mail: *gr312@exeter.ac.uk*

Michele Romano
United Utilities Group PLC,
Lingley Mere Business Park,
Warrington WA5 3LP,
UK

Zoran Kapelan
Department of Water Management, Faculty of Civil Engineering and Geosciences,
Technical University of Delft,
Delft,
The Netherlands
and
College of Engineering, Mathematics and Physical Sciences,
University of Exeter,
Exeter,
UK

doi: 10.2166/ws.2021.062

GRAPHICAL ABSTRACT



Detection of water quality failure events at treatment works using a hybrid two-stage method with CUSUM and Random Forest algorithms

INTRODUCTION

Water utilities around the world face considerable challenges in ensuring that their WTWs produce water of the required quality and quantity. To operate at lowest expenditure, WTWs are already heavily monitored and automated using online sensors deployed at the different treatment stages. Near-real-time detection of faulty sensors and/or the WTW's processes is essential for efficient and effective plant operation. However, due to varying water demand, changing influent conditions, dynamics in water treatment processes and imperfect, missing or incorrect sensor data, this is a difficult task to achieve. In the UK, most WTWs use event recognition systems (ERS), which apply thresholds to generate alarms and detect abnormal behaviour in observed signals. Unfortunately, those threshold-based systems have the major drawback that they result in low true-detection and high false-positive rates (Riss *et al.* 2018).

Nevertheless, more sophisticated applications for event detection at WTWs have already been developed, such as CANARY (Hart *et al.* 2007) released by the United States Environmental Protection Agency (USEPA) (USEPA 2010) or GuardianBlue from Hach Lange (Hach Homeland Security Technologies 2007). However, this first generation of software packages still suffers from a number of shortcomings, such as insufficient real detection capability or too many false alarms (Bernard *et al.* 2015). To overcome the above shortcomings, new and more efficient technologies need to be developed focusing on innovative, cost-effective and, wherever possible, predictive near-real-time event detection systems.

In this paper, we investigate the application of the novel hybrid CUSUM event recognition system (HC-ERS) for the detection of failure events at WTWs and demonstrate improvements achieved by evaluating the detection performance of the HC-ERS for real sensor data and historical events. In addition, we compare HC-ERS's performance to the performance of (i) the threshold-based WTWs event detection system currently used by one of the largest water companies in the UK and (ii) the well-known CANARY event detection algorithms.

BACKGROUND

Online monitoring of water quality to control the treatment processes of WTWs has made considerable progress in recent years (Storey *et al.* 2011). A broad range of fault detection techniques has already been developed (Das *et al.* 2012). For complex systems such as treatment processes at WTWs, where the generation of analytical models is too difficult or not possible, the application of data-driven event detection

methods based on statistical analyses of process data is preferred (Verron *et al.* 2008).

Most common data-driven approaches apply conventional statistical techniques such as statistical process control (SPC) or statistical classifiers to identify deviations in the behaviour of observed process variables by comparing their actual behaviour with normal operating conditions. For example, Schraa *et al.* (2006) discussed the practical aspects of univariate Shewhart, cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) control charts and demonstrated that SPC charts are suitable control schemes for advanced fault detection at wastewater treatment works (WWTW), although they are difficult to apply due to autocorrelation, seasonality and non-constant variance of treatment plant measurements. Aguado & Rosen (2007) presented different multivariate statistical approaches for detection and diagnosis of treatment processes at WWTW using adaptive PCA with two complementary control charts (Hotelling's T^2 and squared prediction error) combined with fuzzy *c*-means clustering for fault diagnosis. This study demonstrated that faster PCA model adaption to changing process conditions results in higher detection speeds but also causes an increased number of false alarms. George *et al.* (2009) combined PCA with Hotelling's T^2 charts for fault detection at a multistage WTW. When applied to the time series data of 23 parameters collected from sensors deployed at a real-life WTW over a 14-day period, the method showed feasibility in detecting abnormal process conditions and was able to identify specific parameters which contributed to disturbances in the process. Although the model seems to perform well over a short period of time, its validation over a long-term period with changing process conditions was not demonstrated in this study. Inspired by the monEAU vision (Rieger & Vanrolleghem 2008), Alferes *et al.* (2013) presented a PCA-based method for real-time monitoring of water systems and detection of sensor faults aiming to achieve an advanced monitoring system with automatic data collection, evaluation, correction and alarm triggering. In their study, Alferes *et al.* used PCA in combination with T^2 and Q-statistics for sensor data validation. Unlike previous PCA models, this approach used sensor data pre-processing to remove outliers and perform auto-scaling (mean centring and variance scaling) before applying the PCA model.

Several event detection systems, such as CANARY, have been recently developed demonstrating substantial improvements in detection of failure events at WTWs. The CANARY open source software provides three algorithms for event detection: time series increment (INC), linear prediction coefficient filter (LPCF) and multivariate nearest-neighbour (MVNN) algorithms (Klise & McKenna 2006). Although LPCF and MVNN have proven to be the most effective detection algorithms (USEPA 2014), they still generate high false-alarm rates (USEPA 2013). Therefore, conventional detection algorithms, as used by CANARY, are often criticised for producing low true-positive and high false-positive rates (Liu *et al.* 2015).

Artificial intelligence (AI), especially machine-learning (ML) techniques, seem to be most promising to achieve further improvements in the field of event recognition at WTWs, because of their ability to extract useful information for operational decisions and to deal efficiently with imperfect sensor data collected by the supervisory control and data acquisition systems (SCADA) commonly used by water utilities (Romano *et al.* 2014). Although Lennox *et al.* (2001) presented the first studies on artificial neural networks (ANNs) for monitoring and controlling filtration processes at WTWs as early as 2001, most of the ML techniques used for fault detection at WTWs have only appeared in the last decade (e.g. Chen & Huang 2011; Padhee *et al.* 2012). These studies predominantly applied approaches based on one-class support vector machine (SVM) and ANNs. For example, an immune feed-forward neural network (IFNN) using an ANN for fault detection in water quality monitoring equipment was developed by Chen & Huang (2011). In their study Padhee *et al.* (2012) combined PCA for fault detection in a WWTW's processes with an ANN based on a back-propagation algorithm as classification technique, ascertaining normal or faulty conditions of a multistage WWTW. Page *et al.* (2017) introduced an adaptive technique to monitor changes in water quality based on multivariate pattern analysis using multivariate analysis and ANNs. Piciaccia *et al.* (2018) developed a data-driven approach for learning the optimal control parameters using an SVM algorithm to predict WWTWs' process behaviour in terms of future

plant states, estimation of optimal chemical dosage and identification of the most influential parameters. The study carried out by Dogo *et al.* (2019) provides an overview of work done in anomaly detection on drinking-water quality data focusing on recent AI and ML approaches applied to water distribution systems, but also presents a specific approach for detecting anomalies in WTWs (Inoue *et al.* 2017). Fehst *et al.* (2018) compared the effect of automatic feature learning using a long short-term memory (LSTM) recurrent neural network (RNN), with manual feature selection with feature subset selection for dimensionality reduction on drinking-water quality anomaly detection. LSTM showed far better performance, with an F_1 score (see section 'Detection Performance Assessment') of 0.8, than the statistical analysis of variance (ANOVA) model, with an F_1 score of 0.48. In their study, Muharemi *et al.* (2019) investigated the performance of SVM, ANN, LSTM, RNN, deep neural network (DNN) and linear regression models for the detection of water quality anomalies when applied to real-world data. SVM outperformed all other approaches with an F_1 score of 0.98, but all models that were tested showed vulnerability to unbalanced data and achieved much lower F_1 scores, e.g. 0.36 for SVM when demonstrated on unseen data. A probabilistic outlier detector implemented by a DNN for anomaly detection at WTWs was proposed by Inoue *et al.* (2017). In their work, Inoue *et al.* applied a DNN consisting of an LSTM layer followed by feed-forward layers of multiple inputs and outputs to time series data of a testbed treatment plant to predict engineered contamination events. Although the proposed method demonstrated promising results with a true-detection rate of 68% and F_1 score of 0.8, further improvements in detection performance, in particular higher true-positive rates, are required for use in engineering practice. More importantly, the AI-based methods presented above approach the detection problem in a sub-optimal way by developing detection methods that usually apply only a single multivariate classification technique for anomaly/event detection.

This paper presents an alternative, fundamentally different approach for near-real-time event detection capable of classifying faults detected on individual water quality signals into faulty/not faulty processes at WTWs. The HC-ERS combines, i.e. hybridises, the SPC (i.e. conventional) method-based fault detection at individual water quality signals with the random forest (RF – i.e. machine-learning) method to ultimately detect WTW-level failure events.

HYBRID CUSUM EVENT DETECTION METHODOLOGY

The proposed HC-ERS method works in two stages. As shown in Figure 1, at the first 'CUSUM fault detection' stage, the CUSUM method is used to identify abnormal deviations of individual water quality and other signals from their normal process conditions. The end result of this stage is a set of labelled individual deviations (i.e. faults) for all signals analysed. This output from the first stage is then used as input for the second stage of 'Forest Tree event detection' in which the RF classifier is trained to learn what combinations of individual signal faults result in failure events at the WTW. The output of the RF method is an estimated probability of the presence of a failure event at the WTW. An alarm is then raised when this probability reaches some pre-specified level.

CUSUM fault detection method

Assuming X failure events (classified into major and minor events, see section 'WTW Minor and Major Events') and Y

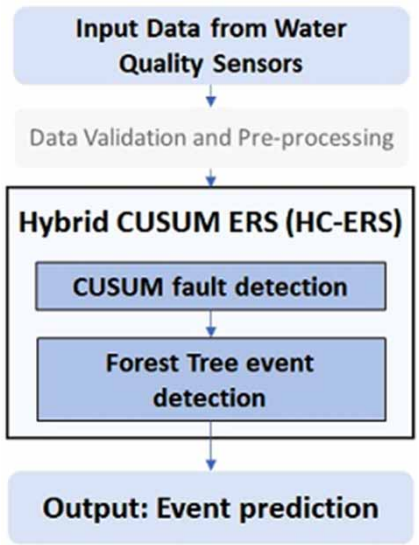


Figure 1 | Process scheme of the hybrid CUSUM event recognition system.

water quality sensors deployed at a WTW with N treatment processes, the first fault-detection stage of the HC-ERS aims to detect the presence of the X failure events at the WTW by identifying relevant deviations of the Y water quality parameters from normal process conditions. The detection method itself is applied to the continuous data of Y water quality signals utilising the data-driven SPC CUSUM control chart technique (Page 1954). CUSUM control charts have already proven their ability to perform well for the detection of small shifts in the process mean (Montgomery 2009). In general, this technique involves the monitoring of process variables or parameters derived from process data (e.g., mean, range, etc.) over a period of time (in the following also referred to as 'window size') by statistical control charts. The parameters of interest were charted over time and compared with control limits to determine whether the process is in 'normal state' (Schraa *et al.* 2006) or 'out of control'.

Typical control charts contain a centre-line that represents the baseline (e.g. average) of a statistical measure across all samples when the process is in control (Freund *et al.* 2010), and two other lines, the upper control limit (UCL) and the lower control limit (LCL) that represent thresholds within which the measure is allowed to vary when a process is in control. UCL and LCL are usually quantified in numbers (n) of standard deviations above and below the centre-line, where n is an integer equal to or greater than 1 ($1 \leq n \leq 6$ are typically used values). Any observation below the LCL or above the UCL indicates that the process is 'out of control'.

Following the approach of Yang *et al.* (2010), the median-based CUSUM control charting technique was used to make the system more robust against outliers. Using a sliding window technique, control charts were generated separately for each of the Y water quality signals by calculating upper and lower cumulative sums of allowed deviations from signals' target values (median of the respective signal). Figure 2 illustrates the CUSUM charting technique by showing an example control chart of a real-world water quality signal. The CUSUM chart monitors the cumulative sums of deviations of observed measures from a target value over time and localises statistically significant anomalies ('out of control' points or sequences) relative to the 'normal state pattern' for each of the water quality signals analysed. The anomalies identified this way, shown in the middle and on the right-hand side of Figure 2, are marked with squares and circles on the upper and lower cumulative sum respectively. Once an 'out of control' point is detected by the chart, the corresponding time step is labelled with the binary value '1'. In the case of the signal's

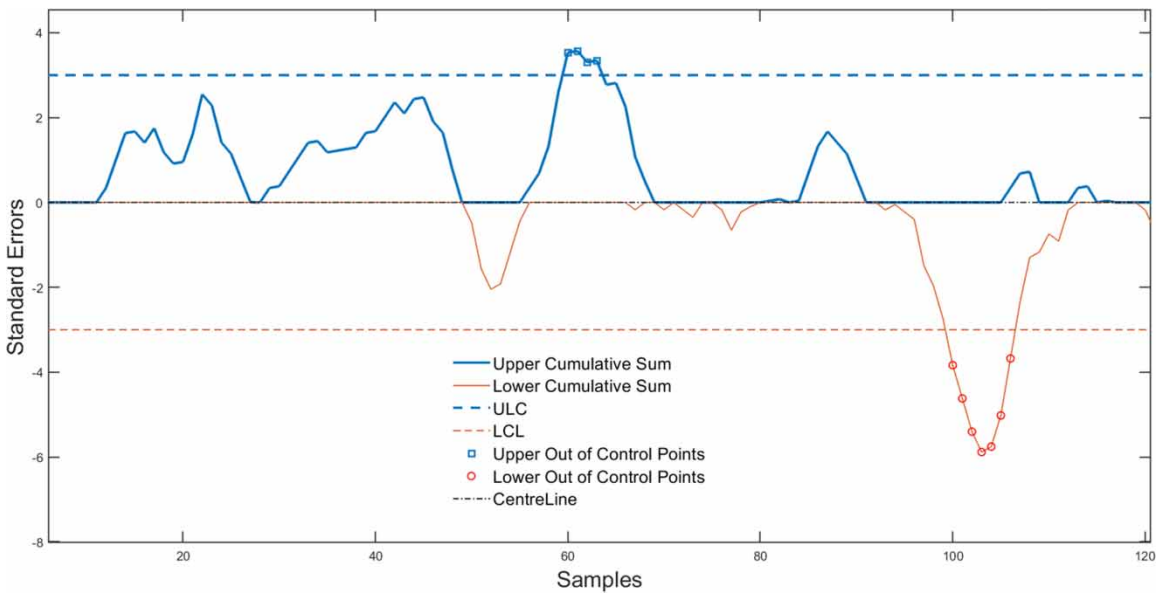


Figure 2 | An example CUSUM control chart.

normal condition, the respective time step is labelled with '0'. In this way, a vector containing ones or zeros at each observed data point was generated for each of the observed Y water quality signals as an output of the applied CUSUM fault detection methodology.

Even though CUSUM charts are largely automated, some parameters can be fine-tuned for an optimal adaptation to the specific fault detection application. In particular, CUSUM control charts require a precise definition of the reference value K, which is often chosen as halfway between the target value and the 'out of control' value of the mean. By changing the reference value, the sensitivity of the CUSUM method can be adjusted. The higher the K value, the less sensitive the CUSUM charting method becomes. Therefore, a fine-tuning of the system was conducted by adjusting the CUSUM parameters for each of the Y water quality signals individually to investigate the optimal control limits and K value combination, with the aim to explore the best possible CUSUM output to serve as input for the subsequent RF event detection method. To achieve this, a sensitivity analysis was performed by gradually changing the K values (from 1σ to 9σ in 0.5σ increments) for different control limits (1σ, 3σ, 6σ and 12σ) and time windows (one day and one week). This way, new CUSUM control charts were created for each signal, resulting in corresponding CUSUM output vectors labelled for each observation with binary values of '0' and '1' for normal and abnormal condition, respectively. For each sensor signal, the optimised new K value and control limits combination were then derived by selecting the specific combination that showed the maximum number of true detections (sum of true positives, TP, and true negatives, TN, see Figure 3).

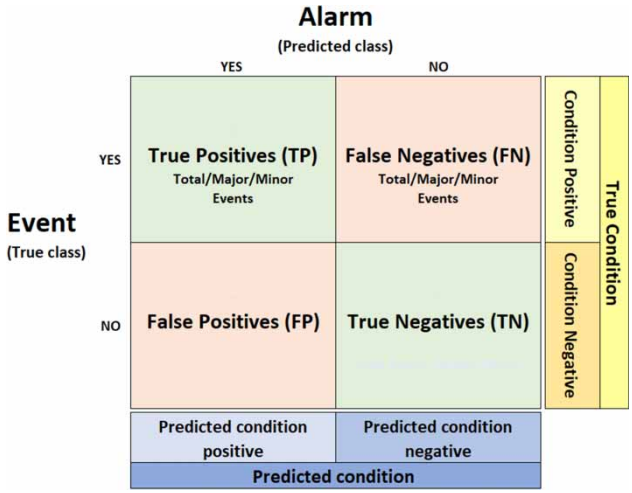


Figure 3 | Confusion matrix.

Random forest event detection method

The objective of the event detection methodology is to investigate possible improvements to the CUSUM fault detection performance by moving away from applying detection rules to individual water quality/other sensor signals only. Indeed, it is expected that moving away from treating individual signals independently (i.e. using a univariate detection method) towards a more sophisticated multivariate event recognition system will increase the true detection rate and, in particular, reduce the false-alarm rate. Once the binary output for each signal is generated as a result of the CUSUM fault detection process, a prediction about the likelihood of a WTW failure event occurring is made by a trained RF classifier (Breiman 2001). It has often been shown that the RF method outperforms other one-class classifier methods by a significant margin (Hempstalk *et al.* 2008), hence it was selected here.

The RF method applied in this study works by using a set of input variables (CUSUM method outputs), which are then passed onto each of the decision trees in the forest. RF classifiers implement randomness in the modelling process, by selecting at each node of the decision tree the variable for splitting as a randomly selected sample of the independent input variables. Each tree gives a prediction and the mean of these values is the prediction of the RF. In the event detection method used here, the RF classifier estimates the probability of the presence of a failure event at the WTW. Similar to CUSUM fault detection, the RF classification method is data-driven and learns relevant relations from the dataset of the observed Y water quality signals, that contains pre-labelled events, aiming to classify the condition of WTW processes into normal or faulty, respectively, to predict the presence of a failure event. For reliable predictions of process conditions, suitable relations between the candidate signals, i.e. across the Y water quality signals, needed to be analysed by the classifier. To achieve this, the fine-tuned CUSUM's binary output of the Y analysed signals

served as input data for the training of the classifier. RF classifiers that make use of 'bagging' in tandem with random feature selection growing a combined ensemble of decision trees to let them vote for the most popular class (Breiman 2001) were tested on the CUSUM output of Y water quality signals. The optimal number of trees used for the RF classifiers was explored by growing template trees and comparing the ratio of true-positive rate (TPR) and false-discovery rate (FDR) (see Figure 4) for the respective number of trees. For an effective implementation of the RF classifier it is assumed that the training database contains a sufficient number of identified historical process faults and events. Each tree utilised in an ensemble of decision trees was individually trained using the data from Y water quality signals to generate the decision rules, according to which each tree generated its vote for the estimated class (event or no event) for each observation. The proportion (non-weighted average) of votes from all trees in the ensemble in favour of a class represents the estimated probability of the class membership. Finally, an alarm is triggered if the estimated probability of an event is above a pre-specified threshold value (e.g. 0.5 used in the case study here). After the training, the classifier model was tested on unseen data and performance was evaluated by quantifying detection statistics on observed, historical data and events. This classification process results in triggering an alarm if a possible failure event is predicted.

Based on the system developed so far, additional improvements of the classifier model were investigated by optimising the feature selection procedure for the classification process aiming at removing redundant signals and those signals that possibly adversely affect the performance of a classifier. To achieve this, stepwise backward elimination using a wrapper method similar to the approach of Kohavi & John (1997) was used to identify and reject the signals that have been considered as insignificant or counterproductive for the model's performance. This optimisation process resulted in a final model that was assumed to perform best, i.e. to demonstrate the best ratio between TP and false positives (FP) (see Figure 3) by using only a subset of the original Y water quality signals.

Performance Metrics

True Detections				False Detections		
True Positive Rates (TPR)			Positive Predictive Value (PPV)	False Discovery Rate (FDR)	False Positives (FP)	False Negative Rate (FNR)
Total	Major	Minor				

Figure 4 | Performance statistics.

Detection performance assessment

The performance assessment of all detection methods (i.e. ERSs) used here was conducted by simulating the ERSs using historical time series data (5 min intervals) of Y water quality signals with X pre-labelled events contained in the datasets. All ERS methods were first calibrated using the data from the calibration time period. The performance of calibrated ERS methods was then assessed using unseen data of the validation time period. This was done by creating two-by-two confusion matrices with true/false positives/negatives, showing the distribution of possible outcomes for Y water quality signals (see Figure 3). Performance statistics were then calculated for each of the Y water quality signals as shown in Figure 4. The detection performance of the overall ERS is evaluated by averaging the detection rates and summation of FP over all Y observed water quality signals.

The derived performance statistics (see Figure 4) contain the true positive rate (TPR), also referred to as recall or sensitivity, calculated by $TPR = \sum True\,positive / \sum Condition\,positive = TP/TP + FN$ for total events (sum of major and minor events) on the one hand and for major and minor events separately on the other. Additionally, the positive predictive value (PPV), also known as precision, was derived by $PPV = \sum True\,positive / \sum Predicted\,condition\,positive = PPV/TP + FP$. Both TPR and PPV describe the true-detection capabilities of the system. Instead of the more common false-positive rate (FPR), the false-discovery rate (FDR) calculated by $FDR = \sum False\,positive / \sum Predicted\,condition\,positive = FP/TP + FP$ was used to indicate the rate of false alarms raised by the detection system. Additionally, the number of FP and the false-negative rate (FNR), which is derived by $FNR = \sum False\,negative / \sum Condition\,positive = FN/TP + FN$ and indicates the miss rate, are shown in the performance statistics. FDR, FP and FNR refer to false detections and are therefore suitable measures of performance for faulty predictions of the system. In addition to the above detection statistics, the

number of $False\,Alarms\,per\,week = \sum False\,positive / Number\,of\,weeks$ and the so-called F measure or F_1 score, often used in the literature (Inoue *et al.* 2017) to compare the detection performance of different models, was calculated by $F_1 = 2 * \text{precision} * \text{recall} / \text{precision} + \text{recall}$ to evaluate the detection performance of each ERS.

CASE STUDY

WTW description

The real WTW used for this study is situated in the north-west of England and supplies water to around 200,000 domestic and industrial customers with a 73.5 ML/d flow capacity. The process flow diagram of the WTW under scrutiny is shown in Figure 5. This WTW is heavily automated and controlled in near real-time by using a SCADA system. Multiple water quality parameters such as pH, turbidity, iron, chlorine, etc. are monitored by online sensors at the different treatment stages. As can be seen from Figure 5, raw water is abstracted from different water sources and enters the WTW at the inlet works, where it is mixed with supernatant recycled flow from dirty backwash water and afterwards split into two separate streams (stream A and B). After dosing for coagulation and pH adjustments, water from each stream is treated by dissolved air flotation (DAF), first-stage filtration and second-stage filtration processes. After filtration, treated water enters the water holding tanks at the outlet works where both streams are re-combined and presented for the final disinfection procedure.

WTW sensor data

The methods described in this work have been developed, tested and validated using real data from the aforementioned WTW. Historical data from 56 sensors over three and a half calendar years (from 01/01/2012 to 30/06/2015), at a five-minute resolution, were provided by the relevant UK water company. Initial data screening resulted in 28 water quality signals relevant for event detection. It is well known that the quality of sensor data utilised for event detection affects the performance of any detection system. Indeed, low data quality may lead, in the worst case, to wrong conclusions (Rieger *et al.* 2010). For this reason, the quality of the provided data streams was assessed on the basis of criteria such as data availability and data consistency by means of a missing data analysis and a statistical analysis, respectively. The aim of these analyses was to create a final dataset that only contains data of sufficiently high quality, crucial for robust event detection. As part of the first analysis, the data of individual signals were examined to identify large numbers of missing data over a significantly long time period (one month, used here). If data was missing for more than one month

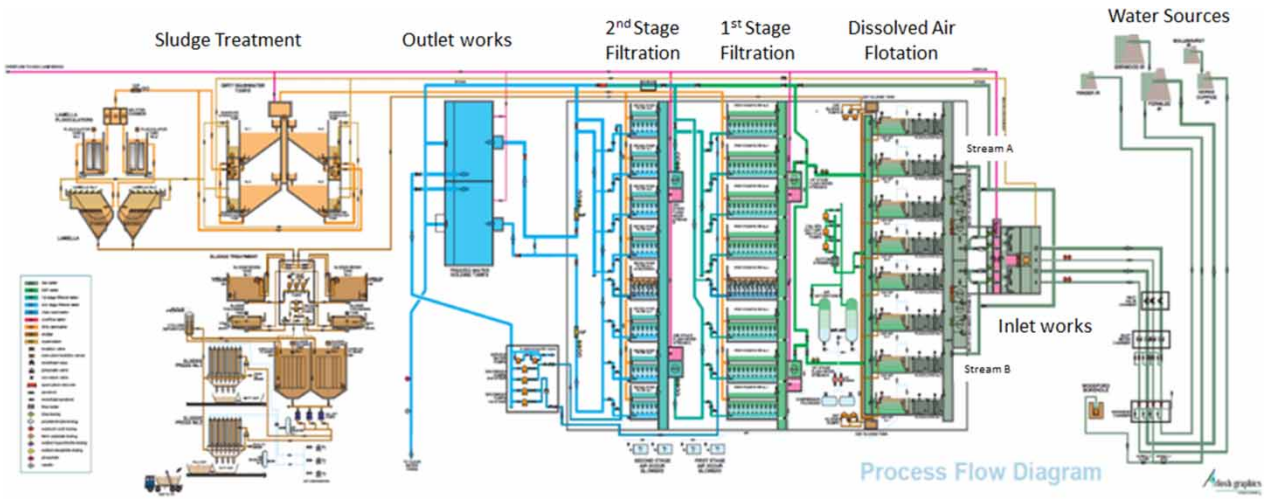


Figure 5 | Process flow diagram of the investigated real-life UK WTW.

continuously in some signal, that signal was considered unreliable. This way, six signals, i.e. first-stage iron (stream A and B), post-second-stage colour (stream A and B), outlet contact tank chlorine and outlet contact tank pH signals were identified as unreliable and hence omitted from further analysis due to poor data quality. As part of the second analysis, time periods were identified in which multiple signals simultaneously showed poor data quality, i.e. data inconsistencies such as frozen values (flat line faults). Figure 6 shows a range of pH and turbidity signals in the period from 02/2015 to 06/2015 where the data quality of the pictured signals continues to decrease with progressing time from 09/03/2015 on (i.e. increasing number of flat line faults marked with grey bars) until all graphed signals show frozen values on 27/05/2015.

The data validation analyses performed here resulted in the final dataset (i.e. used for further analysis) containing 22 signals (see Figure 7) and covering the time period from 01/01/2012 to 01/03/2015. These 22 signals can be mapped to their corresponding sensors and treatment stages for streams A and B as shown in Figure 7. Each of the 22 signals was then split into a dataset for the calibration of the detection models (time period from 01/01/2012 until 28/02/2014, i.e. ~70% of the total time period) and a dataset for the follow-on validation of the detection models on unseen data (time period from 01/03/2014 until 01/03/2015, i.e. ~30% of total time period).

WTW minor and major events

As part of this study, historical events were manually identified and classified into major and minor events to enable the computation of performance measures.

A total of five major events were reported by the water company. These are events that resulted in unplanned shutdowns (full or partial) of the WTW. Figure 8 shows an example of a major event causing a shutdown of the WTW's stream A at 12:40 on 14/09/2013. This was the result of an alarm triggered by stream A's post-flotation turbidity signal. The partial shutdown was followed by a drop of the inlet flow from around 55 ML/d to approximately 35 ML/d. The inlet flow recovered to normal state after the restart of stream A at 16:45 on 14/09/2013.

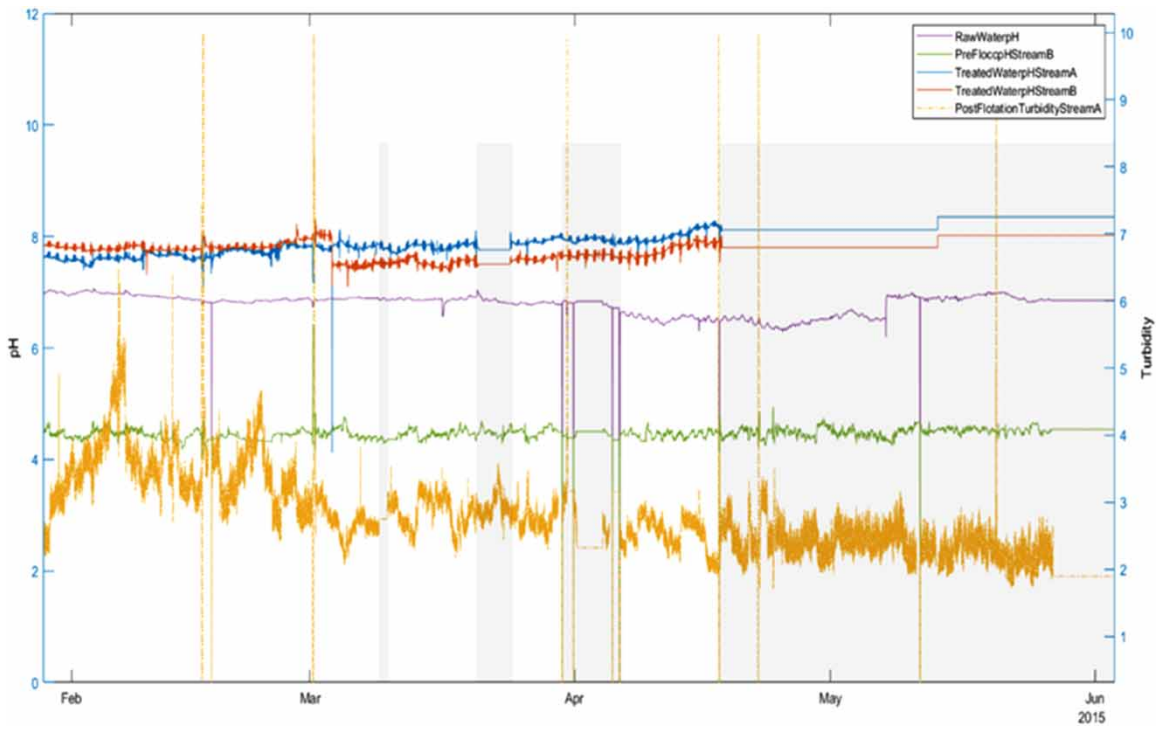
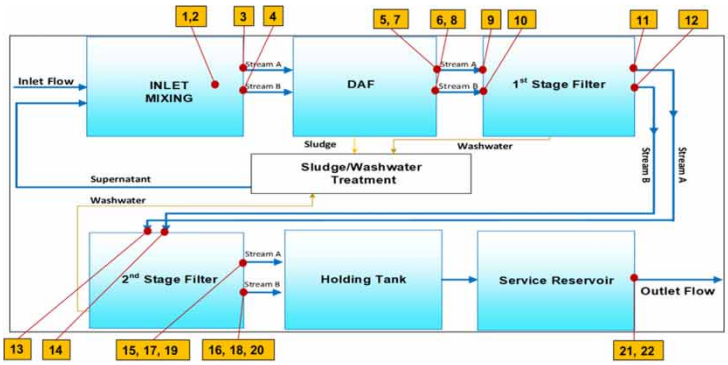


Figure 6 | Example pH and turbidity signals showing flat line faults during the time period from 03/2015 until 06/2015.



Signal (No.)	Water Quality Parameter	Stream	Treatment Stage
1	Raw Water Turbidity	Combined	Inlet Works
2	Raw Water pH	Combined	Inlet Works
3,4	Pre Flocculation pH	A,B	Flocculation & Flotation
5,6	Post Flotation Turbidity	A,B	Flocculation & Flotation
7,8	DAF Iron	A,B	Flocculation & Flotation
9,10	Post 1st Stage pH	A,B	1st Stage Filtration
11,12	Post 1st Stage Turbidity	A,B	1st Stage Filtration
13,14	Pre 2nd StagepH	A,B	2nd Stage Filtration
15,16	Post 2nd Stage Turbidity	A,B	2nd Stage Filtration
17,18	Post 2nd Stage Chlorine	A,B	2nd Stage Filtration
19,20	Treated Water pH	A,B	2nd Stage Filtration
21	Final Water pH	Combined	Outlet Works
22	Final Water Chlorine Residual	Combined	Outlet Works

Figure 7 | Basic schematic of mapped sensor locations.

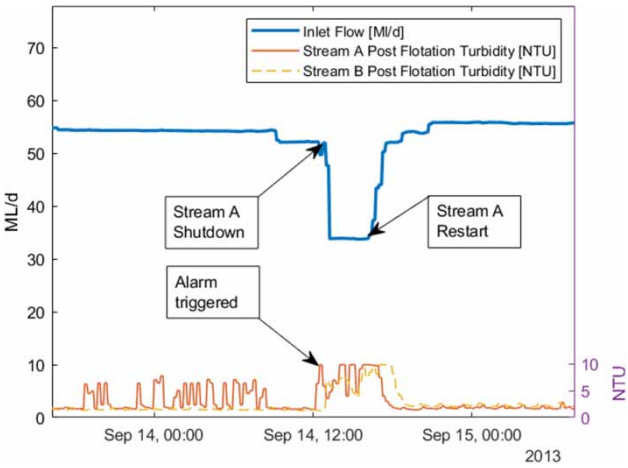


Figure 8 | Example major event – shutdown and restart of WTW's stream A.

The identification and classification of minor events, on the other hand, was carried out by visual inspection of the 22 signals across the entire time period of analysis. Minor events were identified by looking for simultaneous deviations of more than one signal from their normal operating process conditions (without causing a shutdown). Specifically, the WTW's normal operating conditions were established first based on common statistical indicators for minimum, maximum, mean and range. Bivariate correlations between parameters were then calculated using Spearman's correlation coefficient to derive possible related deviations of multiple signals from the corresponding normal values. Following all this, abnormal conditions were identified by visual inspection of the displayed deviations by plotting all the analysed signals below each other for the full time period analysed (01/01/2012 to 01/03/2015). In the case of simultaneous deviations of two or more signals, the presence of a minor event was assumed. An example of this situation is shown in Figure 9. It can be noticed that stream A's post-second-stage chlorine, stream B's post-second-stage chlorine and the final water chlorine residual dropped to zero almost simultaneously at 08:15 on 28/01/2014. Using this methodology, a total of 158 minor events were identified during the analysed time period. Bearing this in mind, it is important to stress that a limited number of the identified minor events were reviewed by an expert from the water company to confirm the validity of the method used against the water company's records.

Once the events were identified as per the above, major and minor events within the final dataset were labelled accordingly.

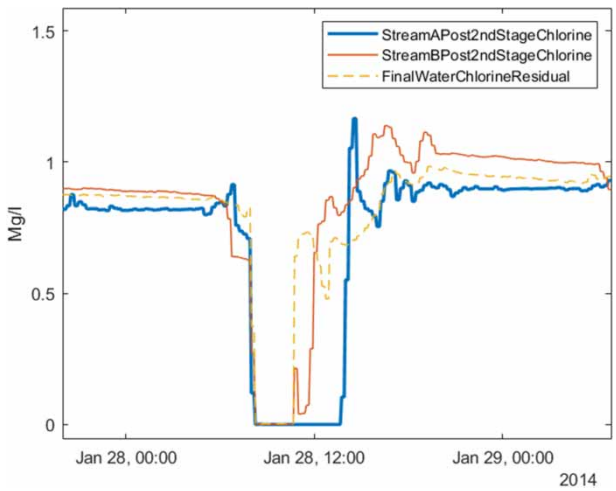


Figure 9 | Example of minor event.

Existing WTW detection system

The existing ERS (E-ERS) makes use of pre-defined threshold limits for each signal and carries out default actions (alarm/no alarm) in the case of limit violations. Every five minutes each signal is checked against the default low and/or high threshold(s). In addition to this, persistence times are used. Persistence defines the time a signal has to remain continuously above/below a threshold value before an alarm is triggered. If two threshold values are set on a single signal, i.e. low and high limit, the same persistence value is used for both limits. For example, the E-ERS applies low and high limits of 5.8 and 7.5 respectively for pre-first-stage pH signals, both with a default persistence of ten minutes (see Table 1). Therefore, an alarm is only raised if the pH value goes below 5.8 or above 7.5 for longer than ten minutes.

RESULTS AND DISCUSSION

The aims of the analysis conducted are to (a) evaluate the performance of the developed HC-ERS method in terms of its detection capabilities and (b) to compare the performance of the HC-ERS method with the performance of the E-ERS and the CANARY methods.

E-ERS performance assessment

The performance of the E-ERS is evaluated here by using the final dataset with labelled major and minor events. The results of this analysis serve as a baseline for the assessment of the improvements achieved by the HC-ERS. The E-ERS's threshold and persistence values used for this analysis are shown in Table 1.

For each signal, confusion matrices were generated and the corresponding detection statistics were calculated according to the formulae shown in Figure 4. The detection statistics for the overall E-ERS were calculated by averaging the detection rates and summation of false positives across all the signals. The E-ERS's detection statistics for the validation dataset are shown in Table 2. It should be noted that, in this study, constant (over the entire time period of analysis) limit and persistence values were used to assess the E-ERS's performance. However, in real life those values are frequently reviewed and, if necessary, adjusted. Therefore, as different values may have been used for different periods of time in real life, it is likely that fewer alarms may have been triggered by the actual E-ERS than those considered here to calculate the E-ERS's detection statistics.

Table 1 | E-ERS detection thresholds and persistence times

Signal	Unit	Low limit	High limit	Persistence [5 min]
Raw Water Turbidity	NTU	–	10.00	0
Raw Water pH	pH	5.50	7.90	1
Pre-Flocculation pH Stream A	pH	4.0	4.80	0
Pre-Flocculation pH Stream B	pH	4.0	4.80	0
Post-Flotation Turbidity Stream A	NTU	0.01	6.50	1
Post-Flotation Turbidity Stream B	NTU	0.01	6.50	1
DAF Iron Stream A	mg/l	–	2.50	6
DAF Iron Stream B	mg/l	–	2.50	6
Pre-First-Stage pH Stream A	pH	5.80	7.50	2
Pre-First-Stage pH Stream B	pH	5.80	7.50	2
Post-First-Stage Turbidity Stream A	NTU	–	0.50	2
Post-First-Stage Turbidity Stream B	NTU	–	0.50	1
Pre-Second-Stage pH Stream A	pH	6.80	8.60	1
Pre-Second-Stage pH Stream B	pH	6.80	8.60	2
Post-Second-Stage Turbidity Stream A	NTU	–	0.40	3
Post-Second-Stage Turbidity Stream B	NTU	–	0.25	3
Post-Second-Stage Chlorine Stream A	mg/l	0.60	1.40	1
Post-Second-Stage Chlorine Stream B	mg/l	0.60	1.40	1
Treated Water pH Stream A	pH	6.80	8.60	0
Treated Water pH Stream B	pH	6.80	8.60	0
Final Water pH	pH	7.00	9.00	1
Final Water Chlorine Residual	mg/l	0.60	1.35	0

Table 2 | Detection statistics of analysed event detection systems tested on the validation time period

	True-Positive Rate (TPR) *Total Events*	True-Positive Rate (TPR) *Major Events*	True-Positive Rate (TPR) *Minor Events*	Positive Predictive Value (PPV)	False-Discovery Rate (FDR)	False Positives (FP)	False-Negative Rate (FNR)	False alarms per week	F_1 score
HC-ERS	82%	100%	82%	86%	14%	13	18%	0.3	0.84
Canary (LPCF)	79%	100%	78%	69%	32%	33	21%	0.6	0.73
Canary (MVNN)	100%	100%	100%	48%	52%	145	0%	2.8	0.65
E-ERS	22%	64%	21%	62%	38%	354	78%	6.8	0.31

As can be seen from Table 2, the E-ERS is able to detect only 22% of total events, 64% of major and 21% of minor events respectively. The significant higher true-detection rate for major events was expected since these events are easier to detect than the minor ones. The E-ERS also generates a considerably high number of false alarms, as demonstrated by the FDR of 38% and the high number of FP events (i.e. 354) produced within the one-year validation time period. All this resulted in approximately 6.8 false alarms per week (derived as the ratio of 354 FP alarms and 52 weeks). This value is just below the critical value of seven false alarms per week after which an ERS can be considered of 'limited practical relevance' (USEPA 2013). Furthermore, the calculated F_1 score is only 0.31, further confirming a rather poor detection performance.

HC-ERS performance assessment

The performance of the HC-ERS is evaluated here in the same way as it was done for the E-ERS. After stepwise elimination of the redundant signals, the performance of the HC-ERS was evaluated using the 16 signals shown in Table 3.

The HC-ERS's detection statistics for the validation dataset are presented in Table 2. The HC-ERS performance, with a TPR of 82% and an FDR of 14%, demonstrated major improvements against the threshold-based E-ERS. Compared with E-ERS, the novel HC-ERS achieved 60% higher TPR and 24% lower FDR. The resulting F_1 score of 0.84 and 0.3 false alarms per week (in contrast to E-ERS' F_1 score of 0.31 and 6.8 false alarms per week) further evidence the HC-ERS's improved performance.

Table 3 | Signals identified as most important for the detection performance of HC-ERS

Signals used by HC-ERS after stepwise elimination of redundant signals	
Raw Water Turbidity	Pre-First-Stage pH Stream A
Raw Water pH	Post-First-Stage Turbidity Stream A
Pre-Flocculation pH Stream A	Pre-Second-Stage pH Stream A
Pre-Flocculation pH Stream B	Pre-Second-Stage pH Stream B
Post-Flotation Turbidity Stream A	Post-Second-Stage Turbidity Stream B
Post-Flotation Turbidity Stream B	Treated Water pH Stream A
DAF Iron Stream A	Final Water pH
DAF Iron Stream B	Final Water Chlorine Residual

CANARY performance assessment

The performance of the well-known CANARY method is evaluated here in the same way as it was done for the E-ERS and for the HC-ERS. As mentioned in the background section, CANARY makes use of three event-detection algorithms (i.e. INC, LPCF and MVNN). Since the LPCF and MVNN algorithms have proven to be the most effective (USEPA 2014), the INC algorithm is not used in this work. Both the LPCF and MVNN event detection algorithms require five key parameters to be defined: (a) the length of the history window, measured in time steps, used to calculate the baseline variability of signals, (b) the outlier threshold, measured in units of standard deviation, used for the detection of outliers, (c) the window size of the binominal event discriminator (BED), measured in time steps, used to provide the event probability for

comparison against (d) the user-defined number of outliers (NRO) required to determine an event, and (e) the event threshold, as a probability value, used to declare a group of outliers as an event. The LPCF and MVNN algorithms were tested using the USEPA-recommended configuration parameters listed in Table 4.

Since it was demonstrated that increasing the number of data points used in the history window results in fewer alarms, while lower values (less than 1.5 days) increase the number of alarms (USEPA 2010), a history window of 2,016 data points (i.e. data from seven days with a resolution of 5 min) was chosen for this analysis. Corresponding to the experiments conducted by USEPA, a window size of 12 time steps (1 hr) was selected for the BED window because, similar to above, shorter BED sizes increase the number of alarms, while events of short duration (shorter than the BED) will not be detected with larger BED window sizes. The NRO used for the analyses were calculated as $N_{\mathrm{RO}} = \sum_{i=0}^{2}(2/3BED + i)$. NRO can then be used to calculate the event thresholds. The event thresholds applied for the sensitivity analysis were defined as $Event\,Thresholds = \sum_{i=0}^{N_{\mathrm{RO}}}(BED!/i!(BED - i!))(1/2)^{BED}$, and can be calculated as $Event\,Thresholds = BINODMIST$ (N_{RO}, BED, 0.5, $True$).

Once the configuration parameters were defined, the sensitivity analysis was conducted by gradually increasing the outlier threshold in increments of 0.25 standard deviations from 1 to 3 and evaluating the test results for each event threshold value. This way sensitivity tests were carried out for both the LPCF and the MVNN detection algorithms resulting in corresponding detection statistics and F_1 scores. The optimal outlier and event threshold combination for LPCF and MVNN algorithms was then derived by selecting the combination with the maximum F_1 score. Detection statistics and F_1 scores of both methods using the optimised outlier and event thresholds are shown in Table 2.

Table 4 | Configuration parameter values used for the sensitivity analysis

Parameter	Initial configuration values
History window	2016 data points
Outlier threshold	0.5–3.0 standard deviations
BED window	12 data points
Number of outliers (N_{RO})	8, 9, 10
Event threshold	0.927, 0.981, 0.997
BED, binomial event discriminator	

Overall performance assessment

CANARY's LPCF algorithm, which employs an outlier threshold of 2.75 standard deviations combined with an event threshold of 0.981, demonstrated the best detection performance among all configurations of the two tested CANARY algorithms. Therefore, the detection statistics of CANARY's system applying the LPCF detection algorithm were used for comparison with the E-ERS and HC-ERS detection performances. For better comparison of all the tested systems, a summary of the detection statistics of each method supplemented by the number of false alarms per week and F_1 score is shown in Table 2 (sorted in order of highest to lowest F_1 scores).

From the above table it can be seen that the HC-ERS method outperforms the other ERSs for many of the key performance indicators. The good performance of HC-ERS is illustrated by a 3% higher TPR for total events and a more-than-halved FDR in contrast to the LPCF CANARY system. The system shows the highest F_1 score of 0.84 and, with 0.3 false alarms per week, by far the lowest rate among all tested event detection systems.

In addition to the above, HC-ERS is also computationally efficient. Indeed, HC-ERS is capable of processing approximately 300 observations per second, including the sensor data validation and pre-processing procedure, while CANARY processes around 100 observations per second. These results were obtained on a laptop with an i5 2.2 GHz processor having 12 GB RAM.

CONCLUSION AND FUTURE WORK

The work presented in this paper introduces a new methodology for near-real-time detection of failure events at WTWs. The novel HC-ERS makes use of CUSUM-based fault detection and RF event detection. The new method was tested, validated and demonstrated using data from a real WTW. The HC-ERS performance was compared with the E-ERS and the well-

known CANARY event detection methods. Based on the results obtained, the following key conclusions can be drawn:

1. The new HC-ERS detection methodology is capable of effectively and efficiently identifying the presence of failure events in WTW processes in near-real-time by processing signals coming from sensors deployed at a WTW. The effectiveness of the HC-ERS method can be seen in the obtained high true-positive detection rate of 82%, accompanied by a low false-alarm rate of only 0.3 false alarms per week, all on unseen data. This is due to the fact that, unlike other ERSs found in the literature, which usually deploy a single method for event detection, either statistical-, knowledge- or machine-learning-based, the HC-ERS follows a hybrid approach that uses two data-driven methods, namely the SPC-type method and the RF advanced machine-learning technique.
2. When compared with the well-known CANARY detection methods, the new HC-ERS method performed better on unseen data. With the true-positive detection rate of 82%, the F_1 score of 0.84 and the 14% false-alarm rate (equivalent to 0.3 false alarms per week), the HC-ERS method demonstrated improved performance over the CANARY method, which achieved a true-detection rate of 79%, an F_1 score of 0.73 and a false-alarm rate of 31% (0.6 false alarms per week).
3. The E-ERS, based on flat-line thresholds and persistence times that are pre-specified for the analysed signals, has demonstrated only moderate detection performance. The system achieved a modest F_1 score of 0.31 with a barely acceptable 6.8 false alarms generated per week. This demonstrates the clear limitations of threshold-based detection methods which, unfortunately, continue to be predominantly used in engineering practice.

Future work should involve further validation of the new HC-ERS method on additional real-world data collected at different WTW sites and should also consider shifts in the time series, since one event would have a signature at different points in time for different measured water quality parameters. In this work, testing and validation was done on a single WTW due to limitations in availability of real-world data. Tests on additional WTWs with potentially different sensors and failure events would not only enable a more thorough validation and demonstration of the proposed HC-ERS detection method, but, more importantly, would provide an opportunity to gain additional knowledge and hence further generalise the observations made in this paper. In general, to ensure reliable predictions of the HC-ERS, accurate sensor data from multiple water quality signals, i.e., data with a low level of missing data and frozen values, should be used. The test results presented show that using all available water quality signals with accurate data (16 sensors in total) worked out well in the case study shown here. This, of course, may not be true for other case studies and the selection of sensors to use needs to be identified on a case-by-case basis, via suitable preliminary analysis. Assuming good quality data, the selection of sensors will depend largely on the characteristics of events being detected and whether and how these events manifest themselves in different water quality signals. Regarding this, water quality signals containing complementary information (i.e. sensors of different type) are especially useful as this helps with the detection. Having said this, using redundant sensor information (i.e. multiple sensors of the same type) can be useful too, as it enables the detection of events with higher true-detection rates and lower false-alarm rates. Finally, when using the HC-ERS on data from other WTWs, it is important to ensure that a sufficiently large number of real failure events are collected and used for the training of RF classifiers. Again, the exact number of events and their characteristics needs to be decided on a case-by-case basis, depending on the nature and characteristics of events being detected.

The use of enhanced sensors that can provide the 'health status' of assets should also be investigated, to examine possible options for integrating this additional metadata (asset condition) into the detection process. Providing additional information could be beneficial for more reliable detection results and would likely improve the system's overall detection performance.

ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council in the UK via grant awarded for STREAM, the Industrial Doctorate Centre (IDC) for the water sector, which is gratefully acknowledged. This work was also kindly supported by United Utilities which

provided the data and industrial insights, which is equally gratefully acknowledged.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Aguado, D. & Rosen, C. 2007 Multivariate statistical monitoring of continuous wastewater treatment plants. *Engineering Applications of Artificial Intelligence* **21**, 1080–1091.

Alferes, J., Tik, S., Copp, J. & Vanrolleghem, P. A. 2013 Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection. *Water Science & Technology* **68** (5), 1022–1030.

Bernard, T., Mossgraber, J., Madar, A. E., Rosenberg, A., Deuerlein, J., Lucas, H., Boudergui, K., Ilver, D., Brill, E. & Ulitzur, N. 2015 SAFEWATER – innovative tools for the detection and mitigation of CBRN related contamination events of drinking water. In: *13th International Conference on Computing and Control for the Water Industry (CCWI2015)*, 2–4 September, Leicester, UK.

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.

Chen, X. & Huang, H. 2011 Immune feedforward neural network for fault detection. *Tsinghua Science and Technology* **16** (3), 272–277.

Das, A., Maiti, J. & Banerjee, R. N. 2012 Process monitoring and fault detection strategies: a review. *International Journal of Quality & Reliability Management* **29** (7), 720–752.

Dogo, E. M., Nwulu, N. I., Twala, B. & Aigbavboa, C. 2019 A survey of machine learning methods applied to anomaly detection on drinking-water quality data. *Urban Water Journal* **16** (3), 235–248.

Fehst, V., La, H. C., Nghiem, T.-D., Mayer, B. E., Englert, P. & Fiebig, K.-H. 2018 Automatic vs manual feature engineering for anomaly detection of drinking-water quality. In: *Genetic and Evolutionary Computation Conference Companion, GECCO '18*, Kyoto, Japan.

Freund, R. J., Wilson, W. J. & Mohr, D. L. 2010 Probability and sampling distibutions. In: *Statistical Methods*, 3rd edn. Academic Press, New York, USA, pp. 67–124.

George, J. P., Chen, Z. & Shaw, P. 2009 Fault detection of drinking water treatment process using PCA and Hotelling's T^2 chart. *International Journal of Computer and Information Engineering* **3** (2), 430–435.

Hach Homeland Security Technologies 2007 *GuardianBlue – Early Warning System*. Hach Company, Loveland, CO, USA.

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

Hempstalk, K., Frank, E. & Witten, I. H. 2008 One-class classification by combining density and class probability estimation. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases – Part I*, Antwerp, Belgium.

Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C. M. & Sun, J. 2017 Anomaly detection for a water treatment system using unsupervised machine learning. In: *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, New Orleans, LA, USA.

Klise, K. A. & McKenna, S. A. 2006 Water quality change detection: multivariate algorithms. In: *Defense and Security Symposium*, International Society for Optics and Photonics, Orlando, FL, USA.

Kohavi, R. & John, G. H. 1997 Wrappers for feature subset selection. *Artificial Intelligence* **97**, 273–324.

Lennox, B., Montague, G. A., Frith, A. M., Gent, C. & Bevan, V. 2001 Industrial application of neural networks – an investigation. *Journal of Process Control* **11**, 497–507.

Liu, S., Smith, K. & Che, H. 2015 A multivariate based event detection method and performance comparison with two baseline methods. *Water Research* **80**, 109–118.

Montgomery, D. C. 2009 The cumulative sum control chart. In: *Introduction to Statistical Quality Control*, 6th edn, Wiley, New York, USA, pp. 400–419.

Muharemi, F., Logofătu, D. & Leon, F. 2019 Machine learning approaches for anomaly detection of water quality on a real-world data set. *Journal of Information and Telecommunication* **3** (3), 294–307.

Padhee, S., Gupta, N. & Kaur, G. 2012 Data driven multivariate technique for fault detection of waste water treatment plant. *International Journal of Engineering and Advanced Technology* **1** (4), 45–50.

Page, E. S. 1954 Continuous inspection schemes. *Biometrika* **41**, 100–115.

Page, R. M., Waldmann, D. & Gahr, A. 2017 Online water-quality monitoring based on pattern analysis. In: *CCWI 2017 – Computing and Control for the Water Industry*, 5–7 September, Sheffield, UK.

Piciaccia, L., Croce, D., Basili, R., Pettersen, J. & Ryfors, P. 2018 A data-driven approach for optimal control parameters in WWTP: the VEAS Experience in Scandinavia. In: *HIC 2018 – 13th International Conference on Hydroinformatics*, vol. 3 (G. La Loggia, G. Freni, V. Puleo & M. De Marchis, eds), EasyChair, pp. 1648–1655.

Rieger, L. & Vanrolleghem, P. A. 2008 mon*EAU*: a platform for water quality monitoring networks. *Water Science & Technology* **57** (7), 1079–1086.

Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P. A. & Comeau, Y. 2010 Data reconciliation for wastewater treatment plant simulation studies – planning

for high-quality data and typical sources of errors. *Water Environment Research* **82** (5), 426–433.

Riss, G., Romano, M., Woodward, K., Kapelan, Z. & Memon, F. 2018 Improving detection of events at water treatment works: a UK case study. In: *HIC 2018 – 13th International Conference on Hydroinformatics*, vol. 3 (G. La Loggia, G. Freni, V. Puleo & M. De Marchis, eds), EasyChair, pp. 1766–1771.

Romano, M., Kapelan, Z. & Savić, D. A. 2014 Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management* **140** (4), 457–467.

Schraa, O., Tole, B. & Copp, J. B. 2006 Fault detection for control of wastewater treatment plants. *Water Science & Technology* **53** (4–5), 375–382.

Storey, M. V., van der Gaag, B. & Burns, B. P. 2011 Advances in on-line drinking water quality monitoring and early warning systems. *Water Research* **45** (2), 741–747.

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

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

USEPA 2014 *Configuring Online Monitoring Event Detection Systems*, EPA 600/R-14/254. US Environmental Protection Agency, Cincinnati, OH, USA.

Verron, S., Tiplica, T. & Kobi, A. 2008 Fault detection and identification with a new feature selection based on mutual information. *Journal of Process Control* **18** (5), 479–490.

Yang, L., Pai, S. & Wang, Y.-R. 2010 A novel CUSUM median control chart. In: *International Multiconference of Engineers and Computer Scientists, IMECS 2010*, 17–19 March, Hong Kong.

First received 16 November 2020; accepted in revised form 22 February 2021. Available online 10 March 2021