Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 309 – 318

Procedia Engineering

www.elsevier.com/locate/procedia

13th International Conference on Computing and Control for the Water Industry, CCWI2015

Improving root cause analysis of bacteriological water quality failures at water treatment works.

K. Ellis[a]*, S. R. Mounce[a], B. Ryan[b], C. A. Biggs[c], M. R. Templeton[d]

[a]Department of Civil and Structural Engineering, The University of Sheffield, Sheffield, S1 3JD, UK
[b]Research and Development, Severn Trent Water Ltd., Coventry, CV1 2LZ
[c]Department of Chemical and Biological Engineering, The University of Sheffield, Sheffield, S1 3JD, UK
[d]Department of Civil and Environmental Engineering, Imperial College London, London, SW7 2AZ, UK

Abstract

Variations in spot-sampled and continuously-monitored water quality data were assessed to determine whether they could be linked to regulatory coliform failures. Data were available from raw water to the final monitoring point at water treatment works (WTW)-B and included climate, physico-chemical and bacteriological data. These were analysed using cross-correlation and self-organising maps in MATLAB®. The results highlighted rainfall and upstream coliforms and turbidity as important factors in the coliform failures. Further examination showed that failures correlated with low turbidity and low coliform loading, but relatively high rainfall. This outcome could be used to improve bacteriological compliance at WTW-B and similar sites.

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: Bacteriological water quality; data analysis; monitoring data; proactive failure prevention

1. Introduction

Water companies conduct bacteriological quality monitoring to assure the safety of drinking water for consumers and to monitor the performance of treatment processes. Water samples are routinely collected from water treatment works (WTWs), service reservoirs and customers' taps. Due to the low numbers of bacteriological

* Corresponding author. Tel.: +44 114 222 5761.
E-mail address: kate.ellis@stream-idc.net

Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.890

pathogens in drinking water under normal circumstances, water samples are tested for indicator organisms. The principal bacteriological indicators are coliforms, *Escherichia coli*, Enterococci and *Clostridium perfringens* (Standing Committee of Analysts, 2002). All four parameters have prescribed values of 0 cells per 100 ml and any detections of these microorganisms is indicative of environmental or faecal contamination of treated water (Council of the European Communities, 1998). Larger sample volumes, for example 1 L, may be used for surveys.

Our previous work (Ellis et al., 2014) showed that cross-correlation and self-organising maps (SOMs) could be used to inform the root cause analysis of a coliform detection at a surface water WTW in the UK (WTW-A). Cross-correlation is a measure of the similarity of two variables as a function of a time lag between them (Bracewell, 1965). This tool could therefore give WTW operators a time period in which to amend treatment processes to prevent a bacteriological failure. SOMs enable the correlation of more than two parameters (with no specific time element) (Kangas and Kohonen, 1996) and were used to understand the broader water quality at the time of the coliform detections. The methods showed some promise for the improvement of root cause analysis, but a limitation of the case study was that data were only available from the final monitoring point. This meant that there was no practical time lag between changes in water quality and the detection of coliforms.

This work builds on the findings from WTW-A and focuses on WTW-B, which produces 160 ML d^{-1}. Both WTWs are owned and operated by Severn Trent Water Ltd. (STW), UK. It treats surface-water using the process outlined in Fig. 1. On the 14th March, 1st April and 12th April 2013 there were 1 L coliform failures from samples at WTW-B collected as part of a water quality survey. Despite extensive investigations, STW have been unable to determine the cause(s). This is the outcome for approximately two thirds of all bacteriological failure investigations (UK Water Industry Research, 2009; Ellis et al., 2013). Since no cause could be identified, these failures were selected for the data analysis in this work. The supply network for WTW-B is extensive and it is important to STW to determine the causes of these non-compliances so that they can protect their consumers. This paper analyses data from 1st January to 31st May (accounting for all three 1 L coliform detections) and for the week 9th to 15th April (focusing on the third 1 L coliform failure) to assess the utility of the analytical methods at both scales.

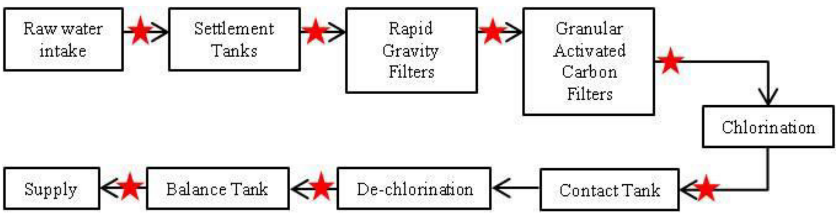

Fig. 1: Process flow diagram for WTW-B; ★ marks the location of the on-line monitors and spot-sampling points.

Data were available throughout WTW-B (Fig. 1), from raw water through to final water. The aim of this study was to see whether through-plant data could identify a time lag which enables operators to act to prevent future bacteriological failures.

Nomenclature

CFU	colony forming units
FTU	formazin turbidity units
GAC	granular activated carbon
NTU	nephelometric turbidity units
RGF	rapid gravity filter
SOM	self-organising map
STW	Severn Trent Water Ltd.
WTW	water treatment works

2. Methods

2.1. Data collection

Water quality and climate data were collected for the period 1st January to 31st May 2013 and a subset created for the 9th – 15th April 2013 (hereafter called "Week Three"). WTW-B's treatment process is outlined in Fig. 1. There are four blocks of settlement tanks and four blocks of rapid gravity filters (RGFs). RGF blocks A, B and C result in a combined spot sampling point; block D has a separate sampling point. RGF block D is fed solely by settlement tank block D. Data from the block D treatment stream were used in this study. These data were for coliforms, colony forming units (CFU) 100 ml^{-1} (excluding final water); 1 L coliforms, CFU 1 L^{-1} (final water only); turbidity, nephelometric turbidity units (NTU) (excluding RGF block D and granular activated carbon (GAC) filters); and free and total chlorine, mg l^{-1} (contact and balance tanks and final water). Raw water and through-plant analyses up to and including GAC filters were collected every two to four days and contact tank, balance tank and final water samples were collected daily. In addition, data for weekly rainfall, mm, were received from STW's Water Resources Strategy team.

For the same time period, the following archived on-line monitoring data were received from STW's Asset Creation Data team:

- Raw water temperature, °C, measured using an ABB AX400 (ABB Ltd., UK);
- Final free chlorine, mg l^{-1}, measured using a Capital Controls® TVU/CC1930 (Severn Trent Services, Philadelphia);
- Filter Block D turbidity, formazin turbidity units (FTU), measured using a Hach 1720E (Hach Lange, UK);
- GAC and final turbidity, NTU, measured using a Sigrist AquaScat WTMA (Sigrist, Germany).
- Temperature data were archived every 15 min and chlorine and turbidity data were archived every 1 min.

2.2. Routine sample collection and coliform analyses

STW samplers conducted spot sample analyses and collected bacteriological quality samples in accordance with standard protocols defined by the Standing Committee of Analysts (2010) and summarised in Ellis et al. (2014).

All samples were conveyed in refrigerated containers to the microbiological laboratory within 24 h. Coliforms were enumerated on membrane lactose glucoronide agar following the manufacturer's protocol (Oxoid, 2012), which conforms to Methods for the Examination of Water and Associated Materials (Standing Committee of Analysts, 2009). The number of colonies were counted and recorded as CFU 100 ml^{-1}/CFU 1 L^{-1}.

2.3. Data manipulation

The datasets were imported into MATLAB® R2012a (The MathWorks Inc., Massachusetts). The date fields were converted to date-number format during the import. The analyses required that all columns contained the same number of rows; to achieve this linear interpolation and zero-padding were used. For this purpose, a template dataset was created with time stamps at 1 min intervals. Linear interpolation was used for all parameters except contact tank, balance tank and final coliforms. Gaps in coliform data from the last three treatment stages were filled with zeros to ensure that when colonies were recorded, the results remained as integers. Full outer joins were used to create a single time-aligned dataset with 1 min time intervals.

2.4. Cross-Correlation

One hundred and seventy-nine cross-correlations were applied to the joined datasets, as detailed in Fig. 2, using the un-biased XCORR function in MATLAB®. The output from this process was a data table of time lags between peaks in data for the first factor (down the side of Fig. 2) and peaks in data for the second factor (across the top of Fig. 2).

A subset of this data table was created containing only the cross-correlations where the time lags were both positive and <24 h (conducted using the same protocol as in Ellis et al., 2014). Positive time lags mean that peaks in the first factor occurred before peaks in the second factor and could have affected them. A time lag of <24 h was selected because the final 1 L coliform sampling was conducted daily.

	Raw water temperature	Raw turbidity (spot)	Raw coliforms	Sett.Tank D turbidity (mon.)	Sett.Tank D coliforms	RGF D turbidity (mon.)	RGF D coliforms	GAC turbidity (mon.)	GAC coliforms	Cont.Tank turbidity (spot)	Cont.Tank free chlorine (spot)	Cont.Tank total chlorine	Cont.Tank coliforms	Bal.Tank turbidity (spot)	Bal.Tank free chlorine (spot)	Bal.Tank total chlorine	Bal.Tank coliforms	Final turbidity (mon.)	Final free chlorine (mon.)	Final free chlorine (spot)	Final total chlorine	Final coliforms
Rainfall	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Raw water temperature		x	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
Raw turbidity (spot)			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Raw coliforms					x		x		x				x				x					x
Sett.Tank D turbidity (mon.)					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Sett.Tank D coliforms							x		x				x				x					x
RGF D turbidity (mon.)								x	x	x			x	x			x	x				x
RGF D coliforms									x				x				x					x
GAC turbidity (mon.)									x	x			x	x			x	x				x
GAC coliforms													x				x					x
Cont.Tank turbidity (spot)											x	x	x	x	x	x	x	x	x	x	x	x
Cont.Tank free chlorine (spot)												x	x	x	x	x	x	x	x	x	x	x
Cont.Tank total chlorine													x	x	x	x	x	x	x	x	x	x
Cont.Tank coliforms																	x					x
Bal.Tank turbidity (spot)															x	x	x	x	x	x	x	x
Bal.Tank free chlorine (spot)																x	x	x	x	x	x	x
Bal.Tank total chlorine																	x	x	x	x	x	x
Bal.Tank coliforms																						x
Final turbidity (mon.)																			x	x	x	x
Final free chlorine (mon.)																				x	x	x
Final free chlorine (spot)																					x	x
Final total chlorine																						x

Fig. 2: Cross-correlations applied to joined datasets (spot = spot-sampled data; mon = monitor data).

2.5. Self-Organising Maps

The SOM analyses were carried out using the MATLAB® SOM Toolbox version 2.0 (Laboratory of Computer and Information Science, Finland). The analyses were conducted on parameters that showed a time lag greater than 0 h with bacteriological parameters across the treatment process as identified during the cross-correlation analyses. The SOM algorithm first normalises the datasets and conducts rough training on these to learn the global structure. After which, fine training is completed before producing the SOM plots. The default settings of linear initialisation and batch training were selected for both analyses. Each variable (such as raw water turbidity) is represented by a colour-coded rectangular plot called a component plane; the same point in one plot is related to that location in all corresponding plots enabling an understanding of how parameters change respective to one another.

For the Week Three 1 L coliform failure, the SOM analysed twelve parameters in the joined dataset based on the results of the cross-correlation: rainfall, air temperature, raw water coliforms, raw water turbidity, filter block D turbidity, GAC filter coliforms, contact tank turbidity, contact tank free chlorine, contact tank total chlorine, balance tank turbidity, final turbidity and final 1 L coliforms. For the five month dataset, the SOM analysed: raw water temperature, settlement tank D turbidity, settlement tank D coliforms, RGF D turbidity, RGF D coliforms, contact tank turbidity, balance tank turbidity, final turbidity, final total chlorine and final 1 L coliforms.

3. Results

3.1. Cross-correlation

For the Week Three 1 L coliform detection, of the 179 cross-correlations conducted, 131 yielded results that were both positive and between 0 and 24 h (Fig. 3). Where the results were 0 h, they indicated that the two parameters changed respective to one another and there was no time lag between them. This accounted for the majority of results for this dataset. Of interest are the cross-correlation results between the following parameters: rainfall and GAC coliforms, 7 h; RGF D turbidity and GAC coliforms, 8 h, final turbidity, 15 h and final 1 L coliforms, 18 h; GAC coliforms and final 1 L coliforms, 23 h; contact tank turbidity and final 1 L coliforms, 23 h; contact tank free chlorine and final 1 L coliforms, 23 h; contact tank total chlorine and final 1 L coliforms, 23 h; and balance tank turbidity and final 1 L coliforms, 23 h (Fig. 3).

Over the five month period, of the 179 cross-correlations conducted, 13 yielded results that were both positive and between 0 and 24 h (Fig. 4). The cross-correlations between rainfall and final turbidity, and settlement tank D turbidity and RGF D coliforms, were both 0 h, suggesting that these parameters changed respective to one another and that there were no time lags between them. Unlike the Week Three results, in the five month dataset the majority of qualifying results had time lags greater than 0 h. Of especial interest, are the time lags for the following parameters: settlement tank D turbidity and settlement tank D coliforms, 3 h; and, RGF D turbidity and RGF D coliforms, 4 h. These results suggest that turbidity has an impact on the bacteriological quality of water from these unit processes.

3.2. Self-Organising Maps

The SOMs for the Week Three 1 L coliform detection and for the five month period are shown in Fig. 5 and Fig. 6, respectively. There are two parts to the SOM output: the summary U-matrix and the component planes for the individual parameters. The U-matrix displays the overall cluster patterns in the input dataset after the model has been trained. The component planes are coloured in accordance with the underlying numerical values for the parameters as shown in the scale bars to the right of each plot. Blue shades represent low values and red shades correspond with high values. The ranges for the coliform results from the contact tank, balance tank and in the final water have been altered by the algorithm as a result of the zero-padding; the SOM output is blue where the result was 0 CFU 100 ml^{-1}/1 L^{-1} and green/red where coliforms were detected. The region of interest across all component planes has been highlighted in the Results for ease of interpretation.

During Week Three, data for the majority of monitored parameters were evenly spread over the component planes (Fig. 5). The ranges were: rainfall, 8.5 – 9.1 mm week^{-1}; air temperature, 6.6 – 8.6 °C; raw water coliforms, 2970 – 3980 CFU 100 ml^{-1}; raw water turbidity, 4.9 – 5.6 NTU; RGF D turbidity, 0.15 – 0.18 FTU; GAC coliforms, <1 – 2 CFU 100 ml^{-1}; contact tank turbidity, 0.1 – 0.2 NTU; contact tank free and total chlorines, 1.98 – 2.01 mg l^{-1} and 2.09 – 2.11 mg l^{-1}, respectively; balance tank turbidity, 0.1 – 0.2 NTU; and, final coliforms, 0 – 1 CFU 1 L^{-1}. The results for final turbidity were 0.04 – 0.29 NTU; the component plane for this parameter shows a small cluster of high values (red) in the top left corner. The coliform detection correlated with low rainfall (8.5 – 8.6 mm week^{-1}), high air temperature (7.9 – 8.6 °C), high raw water coliform counts (3640 – 3980 CFU 100 ml^{-1}), high raw water turbidity (5.3 – 5.6 NTU), low RGF D turbidity (0.15 – 0.16 FTU), low GAC coliform counts (1 CFU 100 ml^{-1}), low contact tank turbidity (0.14 – 0.15 NTU), low contact tank free and total chlorines (1.98 – 1.99 mg l^{-1} and 2.09 mg l^{-1}, respectively), low balance tank turbidity (0.13 – 0.16 NTU), and low final turbidity (0.04 – 0.12 NTU).

The five month dataset resulted in an uneven spread of data clusters in the component planes (Fig. 6). The ranges were: raw water temperature, 4.9 – 16.2 °C; settlement tank D turbidity, 0.50 – 1.23 NTU; settlement tank D coliforms, 37 – 647 CFU 100 ml^{-1}; RGF D turbidity, 0.05 – 0.86 FTU; RGF D coliforms, 56 – 164 CFU 100 ml^{-1}; contact tank turbidity, 0.10 – 0.33 NTU; final turbidity, 0.09 – 0.21 NTU; final total chlorine, 0.71 – 0.97 mg l^{-1}; and final 1 L coliforms, 0 – 1 CFU 1 L^{-1}. The 1 L coliform detection correlated with low-medium raw water temperature (4.9 – 12.4 °C), low-medium settlement tank D turbidity (0.50 – 0.99 NTU), low settlement tank D coliforms (37 – 240 CFU 100 ml^{-1}), low RGF D turbidity (0.05 – 0.32 FTU), low-medium RGF D coliforms (56 – 128 CFU 100 ml^{-1}), low-medium contact tank turbidity (0.10 – 0.30 NTU), low balance tank turbidity (0.09 – 0.25 NTU), and the full range of both final turbidity and final total chlorine residual.

	Raw water temperature	Raw turbidity (spot)	Raw coliforms	Sett.Tank D turbidity (mon.)	Sett.Tank D coliforms	RGF D turbidity (mon.)	RGF D coliforms	GAC turbidity (mon.)	GAC coliforms	Cont.Tank turbidity (spot)	Cont.Tank free chlorine (spot)	Cont.Tank total chlorine	Cont.Tank coliforms	Bal.Tank turbidity (spot)	Bal.Tank free chlorine (spot)	Bal.Tank total chlorine	Bal.Tank coliforms	Final turbidity (mon.)	Final free chlorine (mon.)	Final free chlorine (spot)	Final total chlorine	Final coliforms
Rainfall	0	0	0	0	0	0	0	0	7	0	0	0	-	0	0	0	-	0	0	0	0	-
Raw water temperature		0	0	0	0		0		0	0	0	0	-	0	0	0	-	0	0	0	0	-
Raw turbidity (spot)			0	0	0	0	0	0	0	0	0	0	-	0	0	0	-	0	0	0	0	-
Raw coliforms					0		0		0				-				-					-
Sett.Tank D turbidity (mon.)					0	0	0	0	0	0	0	0	-	0	0	0	-	0	0	0	0	-
Sett.Tank D coliforms							0		-				-				-					-
RGF D turbidity (mon.)								0	8	0			-	0			-	15				18
RGF D coliforms									0				-				-					-
GAC turbidity (mon.)									0	0			-	0			-	-				-
GAC coliforms													-				-					23
Cont.Tank turbidity (spot)											0	0	-	0	0	0	-	0	0	0	0	23
Cont.Tank free chlorine (spot)												0	-	0	0	0	-	0	0	0	-	23
Cont.Tank total chlorine													-	0	0	0	-	0	0	0	0	23
Cont.Tank coliforms																	-					-
Bal.Tank turbidity (spot)															0	0	-	0	0	0	0	23
Bal.Tank free chlorine (spot)																0	-	0	0	0	0	-
Bal.Tank total chlorine																	-	0	0	0	0	-
Bal.Tank coliforms																						-
Final turbidity (mon.)																			0	0	0	-
Final free chlorine (mon.)																				0	0	-
Final free chlorine (spot)																					0	-
Final total chlorine																						-

Fig. 3: Cross-correlation results for the Week Three 1 L coliform failure.

	Raw water temperature	Raw turbidity (spot)	Raw coliforms	Sett.Tank D turbidity (mon.)	Sett.Tank D coliforms	RGF D turbidity (mon.)	RGF D coliforms	GAC turbidity (mon.)	GAC coliforms	Cont.Tank turbidity (spot)	Cont.Tank free chlorine (spot)	Cont.Tank total chlorine	Cont.Tank coliforms	Bal.Tank turbidity (spot)	Bal.Tank free chlorine (spot)	Bal.Tank total chlorine	Bal.Tank coliforms	Final turbidity (mon.)	Final free chlorine (mon.)	Final free chlorine (spot)	Final total chlorine	Final coliforms
Rainfall	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	-
Raw water temperature		-	-	-	-		-		-	12	-	-	-	4	-	-	-	-	-	-	-	-
Raw turbidity (spot)			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Raw coliforms					-		-		-				-				-					-
Sett.Tank D turbidity (mon.)					3	7	0	-	-	14	-	-	-	14	-	-	-	-	-	-	-	-
Sett.Tank D coliforms							-		-				-				-					-
RGF D turbidity (mon.)							4	-	-	1			-	-			-	-				-
RGF D coliforms									-				-				-					-
GAC turbidity (mon.)									-	-			-	-			-	-				-
GAC coliforms													-				-					-
Cont.Tank turbidity (spot)											-	-	-	4	-	-	-	17	-	-	-	-
Cont.Tank free chlorine (spot)												-	-	-	-	-	-	-	-	-	-	-
Cont.Tank total chlorine													-	-	-	-	-	-	-	-	-	-
Cont.Tank coliforms																	-					-
Bal.Tank turbidity (spot)															-	-	-	-	-	-	-	-
Bal.Tank free chlorine (spot)																-	-	-	-	-	-	-
Bal.Tank total chlorine																	-	-	-	-	-	-
Bal.Tank coliforms																						-
Final turbidity (mon.)																			-	-	5	-
Final free chlorine (mon.)																				-	-	-
Final free chlorine (spot)																					-	-
Final total chlorine																						-

Fig. 4: Cross-correlation results for the five month period, during which time there were three 1 L coliform detections.

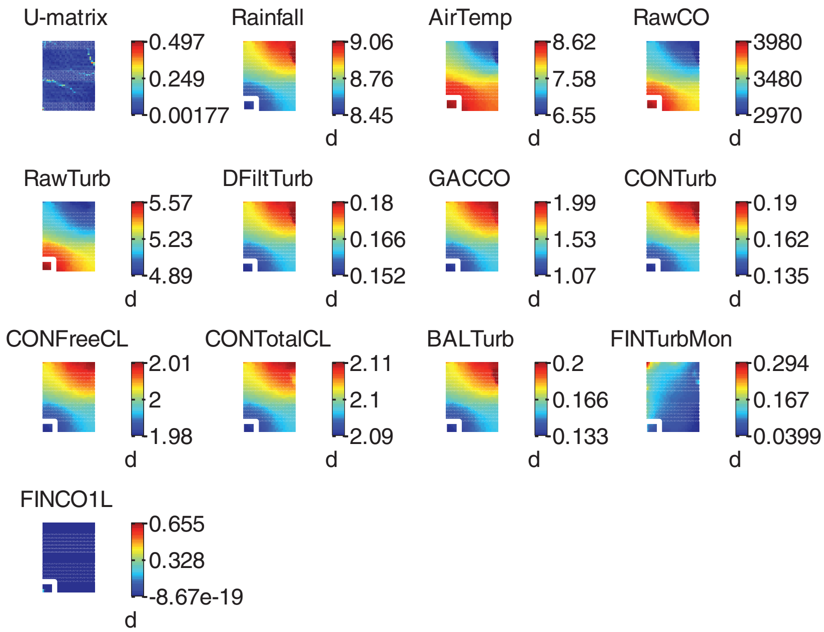

Fig. 5: Self-organising map for the Week Three 1 L coliform failure at WTW-B. Region of interest highlighted.

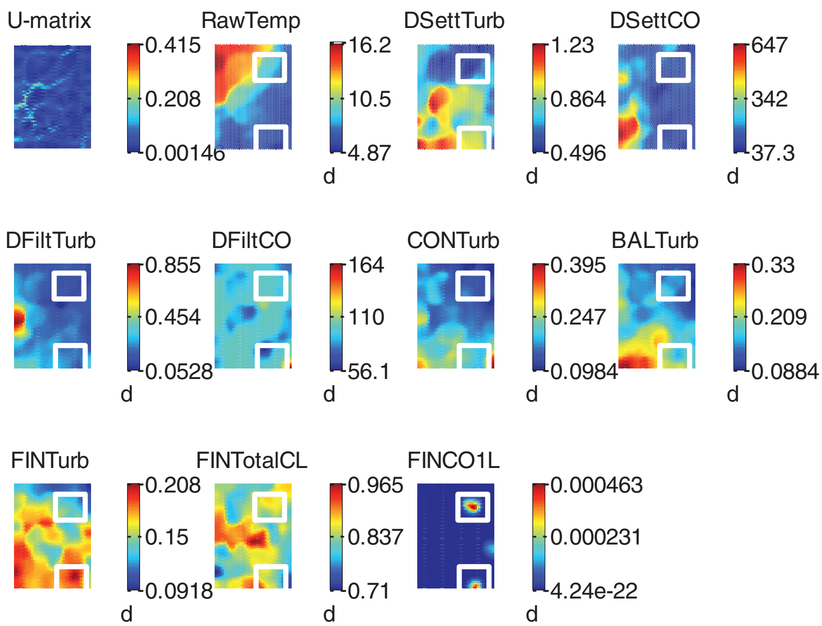

Fig. 6: Self-organising map for the five month period at WTW-B. Regions of interest highlighted.

3.3. Combined power of cross-correlation and SOMs

The results from the cross-correlation analysis at the week-scale showed that the majority of time lags met the selection criteria, but were for 0 h, which does not allow time for operators to act to improve the bacteriological quality of drinking water. In contrast, the majority of results for the five month period did not meet the selection criteria. Never-the-less, it was possible to use the applicable cross-correlation results in the selection of parameters for the SOM analyses. The results suggest that the following conditions, separately or combined, impact upon downstream bacteriological water quality at the week-scale with viable time-lags: low rainfall; low RGF D turbidity; low GAC coliforms; low contact tank turbidity; low free and total chlorine; and low balance tank turbidity. These parameters give operators between 7 and 23 h to act to improve treatment efficacy for final coliforms. At the five month-scale, low-medium turbidity from settlement tank D and low turbidity from RGF D were correlated with coliform results at their respective treatment stages, with time lags of 3 and 4 h, correspondingly.

3.4. Flow through WTW-B

Data provided by STW show that water passes through the selected pathway in 6.9 h at maximum output (and assuming 100% efficiency) and in 8.9 h at minimum output (calculated based on minimum flow-rate).

4. Discussion

4.1. Comparison of the results from cross-correlation and SOMs

Cross-correlation was used to assess the relationship between 179 pairs of parameters for data at the five month- and week-scale; from this, only the results with the strongest correlation were selected. This means that the results were greatly simplified and did not relate to the whole dataset for either the five month or Week Three datasets. Furthermore, only time lags that were both positive and <24 h were considered as they corresponded to the bacteriological sampling frequency. The SOMs incorporated all the data, but had no time element. These tools are powerful individually and when combined can provide valuable information for WTW operators.

The results of the cross-correlation analyses show that many of the parameter pairings appeared to change simultaneously; an observation that was also made in our previous paper (Ellis et al., 2014). It was more likely that a qualifying time lag would be identified at the week-scale. Under these conditions, there will be fewer peaks or troughs in the data trend to account for in arriving at a single time lag. The cross-correlation results showed that rainfall and upstream turbidity and coliform counts impacted the likelihood of detecting a coliform failure at WTW-B. When these parameters were analysed using SOMs, it was observed that low rainfall, low-medium upstream turbidity and low upstream coliform counts were predictors. Previous research has demonstrated a link between climatic factors and bacteriological water quality, including Curriero et al. (2001), Schets et al. (2005), Thomas et al. (2006), and Pitkänen et al. (2008). These studies found, conversely, that heavy rainfall was implicated in drinking water system failures, and not low rainfall. High turbidity and bacteriological loading of the disinfection process consumes disinfectant and reduces treatment efficacy (LeChevallier et al., 1988; Levi, 2004; Al-Jasser, 2007); previous research advises that operators seek low turbidity and low bacteriological counts by the time water reaches the disinfection stage, as was achieved at WTW-B. The range of rain falling during the period of the Week Three failure was 8.45 mm week^{-1} to 9.06 mm week^{-1}; whilst the 1 L coliform detection correlated with the low end of this range, 8.45 mm week^{-1} is more than one third the amount of rain that fell during April 2013 (21.8 mm month^{-1}) (MetOffice, 2013). This finding shows that it is important to compare climatic data with the annual trend to identify its significance with regard to bacteriological compliance. In this case, it seems likely that rainfall was a factor in the coliform detection at WTW-B.

The cross-correlation results suggested that operators at WTW-B would have between 3 and 23 h to act in order to prevent a bacteriological failure. Since water spends between 6.9 h and 8.9 h passing through the WTW before entering the distribution system, care must be taken in assigning operator value to time lags greater than 8.9 h. This therefore means that at WTW-B operators would be advised to take note of weather forecasts for rainfall and also to closely monitor the water quality at the earlier treatment processes (settlement tanks and RGFs).

Cross-correlation and SOMs provide different information to analysts. Cross-correlation provides operators with time lags, but not parameter states (low, medium or high), whilst SOMs provide the reverse. The two tools, when used together, offer useful information for the optimisation of WTW operation. The dataset analysed from WTW-B was very comprehensive, not just in the range of parameters available, but also the time-frame and resolution of the data. It was more time-efficient to select SOM parameters as a subset of those analysed with cross-correlation (compared with the approach used in Ellis, 2013 and Ellis et al., 2014); this may have resulted in excluding data that would develop a clearer picture of what was happening at the WTW. Only a single stream through WTW-B has been presented; the impacts of water passing through blocks A, B and C treatment streams (settlement tanks and RGFs) have been omitted. Bearing in mind these limitations, these results show that developing an understanding of water quality throughout the WTW is more powerful than focussing only on the final monitoring point. In the future, both SOMs and cross-correlation could be incorporated into on-line monitoring tools to give near-real-time information about water quality.

4.2. Bacteriological failures under nominally ideal operating conditions

These coliform detections at WTW-B were obtained under nominally ideal operating conditions: low turbidity, low coliform loadings, and adequate free and total chlorine concentrations. This concurs with the work that we did at WTW-A, where data were only available from the final monitoring point (Ellis et al., 2014). The reliance on monitoring for these water quality parameters as surrogates for microbiological safety may also result in the high proportion of bacteriological water quality failures ex-works having unknown causes (UK Water Industry Research, 2009; Ellis et al., 2013). These wider findings infer that parameters other than those for which data are readily available can serve to reduce bacteriological compliance. Future case studies would benefit from the inclusion of data for organic carbon (Dukan et al., 1996; American Water Works Association, 2003) and cell counts (Berry et al., 2006; Berney et al., 2008). At present, these data are not routinely collected by UK water companies. Adopting these newer monitoring tools, with on-line capability, would give water companies a head start in preventing failures which currently have no cause attributed to them and enable the development of more advanced site-specific action plans for maintaining water quality.

5. Conclusions

This project used cross-correlation and SOMs to determine whether on-line water quality data from throughout a WTW could be used to identify the cause of coliform detections at a WTW final monitoring point. Two time periods were analysed: one of five months (1st January to 31st May 2013) and incorporating three 1 L coliform detections and another relating just to the Week Three 1 L coliform failure.

Cross-correlation results that were considered for further analysis were both positive and <24 h. The Week Three dataset yielded 131 qualifying results. The majority of these were 0 h, which suggests that the parameters changed respective to one another. The five month dataset yielded 13 qualifying results, of which 11 were greater than 0 h. The reduced success of cross-correlation in obtaining a time lag at the five month-scale suggests that it is better applied to smaller datasets. Results that were greater than 0 h could be considered useful for operators seeking to make changes to the performance of the WTW in order to avoid a bacteriological failure. There cross-correlation results showed that operators had between 3 and 23 h to act to reduce the likelihood of a coliform non-compliance.

Both datasets highlighted cross-correlation results of upstream turbidity and coliform counts with downstream bacteriological water quality. When these relationships were analysed using SOMs it was found that the coliform detections occurred under conditions of low turbidity, low coliform loadings and adequate disinfectant residuals. While cross-correlation and SOMs were not able to provide a definitive cause for the coliform failures, they suggested that the Week Three failure was impacted by relatively high rainfall. This result is useful to operators at WTW-B, as they can focus their efforts on settlement and filtration efficacy during wet weather.

Improvements in water infrastructure and management practices have made bacteriological failures rare in the UK. Resolving failures with unknown causes will require investment in more advanced technologies, which can be linked with cross-correlation and SOMs to further improve drinking water quality compliance.

Acknowledgements

The authors wish to thank Darren Watson and Daniel Booth (both STW) for providing spot-sampling and on-line monitoring data, respectively. This research was funded under the STREAM Industrial Doctorate Centre and was collaborative among the Engineering and Physical Sciences Research Council, Severn Trent Water Ltd., The University of Sheffield and Imperial College London.

References

Al-Jasser, A. (2007). Chlorine decay in drinking-water transmission and distribution systems: Pipe service age effect. *Water Research, 41*, 387-396.

American Water Works Association. (2003). *Influence of distribution system infrastructure on bacterial regrowth.* Denver: AWWA Research Foundation.

Berney, M., Vital, M., Hülshoff, I., Weilenmann, H.-U., Egli, T., & Hammes, F. (2008). Rapid, cultivation-independent assessment of microbial viability in drinking water. *Water Research, 42*(14), 4010-4018.

Berry, D., Xi, C., & Raskin, L. (2006). Microbial ecology of drinking water distribution systems. *Current Opinion in Biotechnology, 17*, 297-302.

Bracewell, R. (1965). *Pentagram notation for cross-correlation: the Fourier transform and its application.* New York: McGraw-Hill.

Council of the European Communities. (1998). *Council Directive of 3 November 1998 on the quality of water intended for human consumption. 98/83/EC.* European Commission.

Curriero, F., Patz, J., Rose, J., & Lele, S. (2001). The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. *American Journal of Public Health, 91*(8), 1194-1199.

Dukan, S., Levi, Y., Piriou, P., Guyon, F., & Villon, P. (1996). Dynamic modelling of bacterial growth in drinking water networks. *Water Research, 30*(9), 1991-2002.

Ellis, K. (2013). *Improving root cause analysis of bacteriological water quality failures.* EngD thesis: University of Sheffield.

Ellis, K., Mounce, S., Ryan, B., Templeton, M., & Biggs, C. (2014). Use of on-line water quality monitoring data to predict bacteriological failures. *Procedia Engineering, 70*, 612-621.

Ellis, K., Ryan, B., Templeton, M., & Biggs, C. (2013). Bacteriological water quality compliance and root cause analysis: an industry case study. *Water Science and Technology: Water Supply, 13*(4), 1034-1045.

Kangas, J., & Kohonen, T. (1996). Developments and applications of the self-organizing map and related algorithms. *Mathematics and Computers in Simulation, 41*, 3-12.

LeChevallier, M., Cawthon, C., & Lee, R. (1988). Factors promoting the survival of bacteria in chlorinated water supplies. *Applied and Environmental Microbiology, 54*(3), 649-654.

Levi, Y. (2004). Minimizing potential for changes in microbial quality of treated water. In R. Ainsworth, *Safe Piped Water: managing microbial water quality in piped distribution systems* (pp. 19-37). London: IWA Publishing.

MetOffice. (2013, May). *Regional values - April 2013*. Retrieved May 5, 2015, from MetOffice: http://www.metoffice.gov.uk/climate/uk/summaries/2013/april/regional-values

Oxoid. (2012). *Membrane Lactose Glucoronide Agar*. Retrieved November 2012, from Oxoid: http://www.oxoid.com/UK/blue/prod_detail/prod_detail.asp?pr=CM1031&org=71&c=UK&lang

Pitkänen, T., Miettinen, I., Nakari, U.-M., Siitonen, A., Kuusi, M., Takkinen, J., et al. (2008). Faecal contamination of a municipal drinking water system in association with Campylobacter jejuni infections. *Journal of Water and Health, 6*(3), 365-376.

Schets, F., During, M., Italiaander, R., Heijen, L., Rutjes, S., van der Zwaluw, W., et al. (2005). Escherichia O157:H7 in drinking water from private supplies in the Netherlands. *Water Research, 39*, 4485-4493.

Standing Committee of Analysts. (2002). *The Microbiology of Water 2002: Part 1 - Water Quality and Public Health.* London: Her Majesty's Stationery Office.

Standing Committee of Analysts. (2009). *The Microbiology of Water 2009: Part 4 - Methods for the isolation and enumeration of coliform bacteria and Escherichia coli (including E. coli O157:H7).* London: Her Majesty's Stationery Office.

Standing Committee of Analysts. (2010). *The Microbiology of Water 2010: Part 2 - Practices and Procedures for Sampling.* London: Her Majesty's Stationery Office.

Thomas, M., Charron, D., Waltner-Teows, D., Schuster, C., Maarouf, A., & Holt, J. (2006). A role of high impact weather events in waterborne disease outbreaks in Canada, 1975-2011. *International Journal of Environmental Health Research, 16*(3), 167-180.

UK Water Industry Research. (2009). *Validating the cause of coliforms in drinking water.* London: UK Water Industry Research Limited.