



This is a repository copy of *A data quality assessment framework for drinking water distribution system water quality time series datasets*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/197122/>

Version: Published Version

---

**Article:**

Gleeson, K. [orcid.org/0000-0002-3767-3001](https://orcid.org/0000-0002-3767-3001), Husband, S., Gaffney, J. et al. (1 more author) (2023) A data quality assessment framework for drinking water distribution system water quality time series datasets. *Journal of Water Supply: Research and Technology-Aqua*, 72 (3). pp. 329-347. ISSN 0003-7214

<https://doi.org/10.2166/aqua.2023.228>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## A data quality assessment framework for drinking water distribution system water quality time series datasets

Killian Gleeson<sup>a,\*</sup>, Stewart Husband<sup>a</sup>, John Gaffney<sup>b</sup> and Joby Boxall<sup>a</sup>

<sup>a</sup> Department of Civil and Structural Engineering, The University of Sheffield, Sheffield S10 2TN, UK

<sup>b</sup> Siemens UK, Manchester M20 2UR, UK

\*Corresponding author. E-mail: kgleeson1@sheffield.ac.uk

 KG, 0000-0002-3767-3001

### ABSTRACT

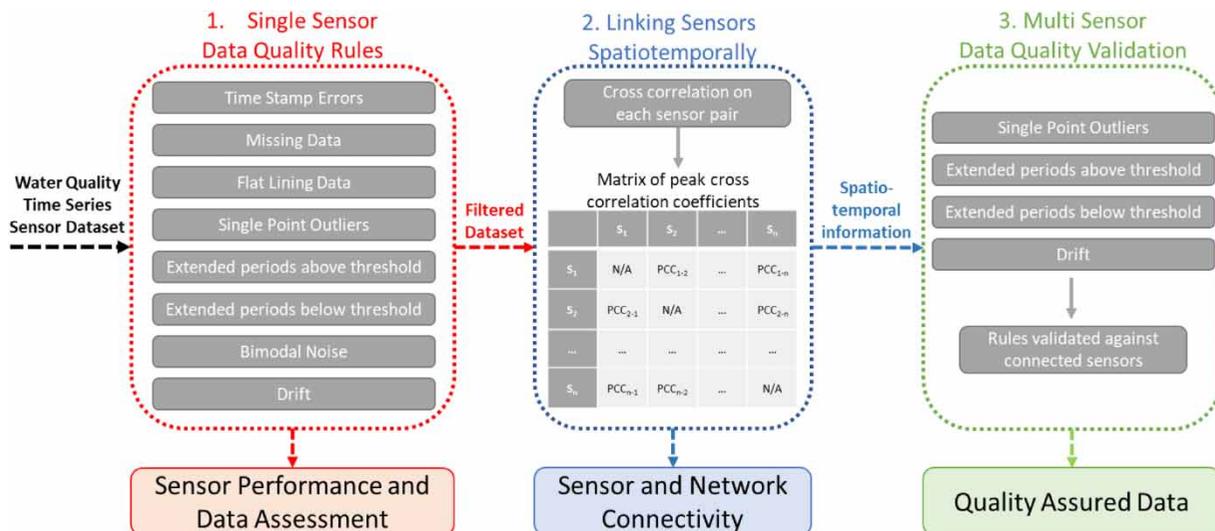
The derivation of information from monitoring drinking water quality at high spatiotemporal resolution as it passes through complex, ageing distribution systems is limited by the variable data quality from the sensitive scientific instruments necessary. A framework is developed to overcome this. Application to three extensive real-world datasets, consisting of 92 multi-parameter water quality time series of data taken from different hardware configurations, shows how the algorithms can provide quality-assured data and actionable insight. Focussing on turbidity and chlorine, the framework consists of three steps to bridge the gap between data and information; firstly, an automated rule-based data quality assessment is developed and applied to each water quality sensor, then, cross-correlation is used to determine spatio-temporal relationships and finally, spatiotemporal information enables multi-sensor data quality validation. The framework provides a method to achieve automated data quality assurance, applicable to both historic and online datasets, such that insight and actionable insight can be gained to help ensure the supply of safe, clean drinking water to protect public health.

**Key words:** data quality, drinking water distribution systems, drinking water quality, water quality analytics, water quality sensors

### HIGHLIGHTS

- Automated data quality rules assess turbidity and chlorine sensor performance while deployed within drinking water systems.
- Cross-correlation determines how sensor locations are linked spatiotemporally, enabling multi-sensor analytics.
- A multi-sensor data quality assessment framework uses spatiotemporal information to further validate and, therefore, assure data quality prior to the derivation of actionable insight.

### GRAPHICAL ABSTRACT



This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

## 1. INTRODUCTION

Monitoring of distributed drinking water quality typically consists of periodic discrete sampling that fulfills regulatory purposes but only offers a limited opportunity to understand the performance of these extensive and complex engineered environments. The sparse data from discrete sampling does not enable examination of water quality deterioration processes that are known to occur between treatment and tap, such as hydraulic-induced discolouration (Husband & Boxall 2011) and disinfection residual decay (Speight & Boxall 2015). Without more dense data, water utilities can only be reactive, informed of water quality incidents through customer contacts (Mounce 2020). This is becoming increasingly unacceptable, including an estimated 4–12 million cases of gastrointestinal illnesses attributable to public drinking water systems in the United States (Colford *et al.* 2006). High-frequency water quality monitors (generally considered as sampling every 15 min or less) suitable for deployment within drinking water distribution systems (DWDS) offer the potential to change this. Such instruments measure parameters such as turbidity and free chlorine, both of which can indicate pathogen presence. Turbidity has been linked to gastrointestinal illness (Mann *et al.* 2007) and is a proxy measurement for discolouration (Boxall & Saul 2005), which is primarily caused by hydraulic changes mobilising material accumulated on pipe walls, which can include pathogens from biofilms (Husband *et al.* 2016). While disinfection residuals are relied upon to provide protection against planktonic cells and limit microbial regrowth within DWDS (Thayanukul *et al.* 2013), meaning a drop in free chlorine may indicate increased pathogen risk. However, because these are relatively (to flow and pressure) sensitive scientific instruments, the quality of data obtained from their deployment can vary.

The potential for such instruments is clear and utilities are embracing these sensing technologies within DWDS, but the questionable data quality due to instrument sensitivity and issues connected to the often remote and harsh locations is currently a major barrier to the resulting data being used to inform network operations. Many turbidity sensors have optical lenses in contact with sample streams that can get fouled by accumulating material (Mounce *et al.* 2015). Online chlorine sensors commonly rely on membrane technology, which requires regular recalibration and servicing. Even with regular maintenance, the collected data may not be representative of the water quality being studied. This has resulted in water quality data often requiring extensive manual data quality assessment and cleaning to remove spurious signals before analysis is possible (Mounce *et al.* 2015).

There is a need to develop rapid and robust automated methods for checking water quality sensor performance and assessing data quality. Differentiating between sensor errors and real system events is difficult without the ability to cross-validate with other sensors in a network if applying the logic that system events will be seen in multiple sensors, unlike sensor faults (Krishnamachari & Iyengar 2004). Sensors deployed within a DWDS can be entirely unconnected to each other or separated by network features such as service reservoirs, valves, and pumps, which alter the water quality to such a degree that direct comparison may not be possible. It is also not considered practical to install two sensors at every location, so understanding how sensors at different locations are connected to each other is a key step in improving the effectiveness of data quality assessment.

### 1.1. Background

In general, sensor data quality describes how accurately the sensor data represents the system under observation. There are generally two routes to automatically assessing sensor data quality: define normality for the system being monitored and quantify the degree of conformity to this normality (Teh *et al.* 2020); or define data quality metrics or errors and quantify the degree these errors present in the sensor data (Kirchen *et al.* 2017). The selection of a data quality method depends both on the type of data available and the intended usage of the data. Normality can be modelled from past observations or taken from an assumed distribution, but this may not always be available or applicable. A systematic review of sensor data quality detection and correction by Teh *et al.* (2020), revealed that outliers were the most commonly studied sensor error, followed by missing data, bias, drift, repeated values, uncertainty, and ‘stuck-at-zero’. The fact that outliers are both indicative of sensor faults and real system events in sensor networks (Zhang *et al.* 2010) demonstrates the need to be able to validate these and other potentially erroneous occurrences, with other sensors. A rule-based approach, looking at features such as data spikes and missing data, was employed on river water quality sensors in a study from Australia in 2019 (Talagala *et al.* 2019), but this approach has not yet been applied to DWDS water quality sensor data. Though this work focuses on assessing sensor data quality, the subsequent analysis may require that any removed or missing data to be filled in. A review of missing data imputation techniques using DWDS demonstrated the range of potential methods, from simple statistical single imputation to model-based and machine learning multiple imputation algorithms (Osman *et al.* 2018). A recent

study compared such approaches on river water quality parameters and found that most will work well for short periods, but longer gaps require consideration of the temporal fluctuations present in water quality time series (Zhang & Thorburn 2022).

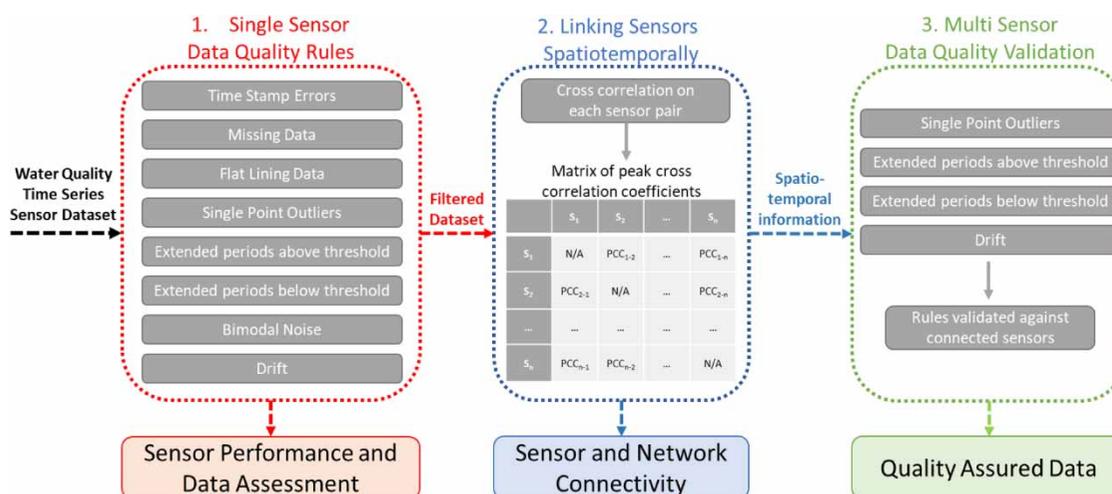
Understanding how simultaneously recorded time series are related to each other spatiotemporally has been studied in areas such as seismology (Vandecar & Crosson 1990), astronomy (Peterson *et al.* 1998), ultrasound imaging (Bonnetfous 1986), and psychology (Boker *et al.* 2002). A variety of similarity metrics have been used, ranging from simple Euclidean distances to dynamic time warping (DTW) and correlation coefficients (Kianimajd *et al.* 2017). Cross-correlation is the most commonly used method for determining the strength of the relationship and time lag between two time series signals (Benesty *et al.* 2004). This involves shifting one time series relative to another and calculating a correlation coefficient at each step, with the step giving the highest correlation taken as the time lag. Pearson's correlation coefficient (PCC) is a commonly used coefficient as it measures the linear relationship between two variables. Many variants on cross-correlation, such as detrended cross-correlation analysis, have been developed to deal with non-stationarity and the presence of unwanted periodicity (Horvatic *et al.* 2011). In DWDS, one recent study used cross-correlation analysis between flow and pressure sensors in a DWDS to detect leakages, indicated by sudden drops in cross-correlations (Gomes *et al.* 2021). DTW is another method for quantifying the similarity of two time series and can deal with different durations and sample rates (Keogh & Pazzani 2001). It has been shown to effectively determine transit times in sewers using temperature sensors (Dürrenmatt *et al.* 2013), though its similarity metric is not as easily interpretable as PCC, which provides a value between  $-1$  and  $1$  that informs about the strength of the relationship. Though cross-correlation has not been used previously to relate water quality sensors spatiotemporally, semblance correlations between turbidity and hydraulic data have been used to infer changes in the risks of asset deterioration (Mounce *et al.* 2015).

The aim of this work was to develop a data quality assessment framework suitable for water quality monitoring within DWDS. Specifically, this work aimed to establish and automate an appropriate method for accurate detection and quantification of anomalous data in high-frequency remote turbidity and chlorine sensors. A key element of the framework was to develop a method to understand the connectivity between water quality sensors at different locations, enabling data quality assessments to be cross-validated. A final stage would allow data quality assurance, providing confidence for further analysis.

## 2. METHODS

### 2.1. Multi-sensor data quality assessment framework

A rule-based data quality assessment approach was decided upon, as opposed to an approach involving defining system normality, which would be problematic due to the lack of labelled datasets combined with the fact that water quality data is neither stationary nor normally distributed. Therefore, a framework was developed, illustrated in Figure 1, for assessing data quality of water quality sensors deployed within DWDS. The framework consists of three steps that work sequentially to perform high confidence data quality assessments for water quality sensor networks deployed within DWDS. The first step



**Figure 1** | Multi-sensor data quality assessment framework.

involves a single sensor data quality assessment, using eight data quality rules that were developed to identify data quality issues. These are developed here specifically for turbidity and chlorine sensor data but will have wider applicability. Detecting and quantifying the prevalence of data flagged by the rules in each time series allows for the performance of the sensors to be ranked and compared within and across datasets. However, the removal and/or replacement of flagged data depends on the needs of any subsequent analysis. When used in the multi-sensor framework presented in this work, the first single sensor pass of the data quality rules involves filtering out any period of data that is flagged by a data quality rule. This is to enable the cross-correlation analysis to be performed on the remaining data, without anomalous features negatively impacting the correlation calculations. Next, cross-correlation analysis is performed to gain an understanding of how the sensors are related in time and space. This results in a peak Pearson's cross-correlation coefficient (PCC) for each sensor pair as well as an estimated transit time between the highly correlated sensors. This information can enhance other water quality analyses by allowing sensor information to be combined across locations. In terms of this data quality assessment framework, this spatiotemporal information is used to enhance the data quality assessment, by enabling cross-sensor validation to take place in Stage 3 on four of the eight rules identified in Stage 1. Methods for each framework stage were written in Python, primarily using the data science library Pandas (McKinney 2010).

## 2.2. Data quality rules

The data quality issues, which were identified as the basis of each rule, are outlined in Table 1, along with the corresponding method to detect their occurrence and a possible cause described. In many cases, the causes cannot be determined with confidence without other supporting information, such as the data from other related sensors (Stage 3). The reasoning behind each rule, along with the detection methods, are expanded upon in this section. This rule-based approach does not rely on predicting or modelling these highly complex non-stationary parameters but instead focuses on developing methods to detect the presence of the specific issues identified.

### 2.2.1. Timestamp errors

Timestamp errors refer to data points that have an unintended sampling interval compared with the previous data points. Imbalanced datasets as such suffer from bias, with the lack of consistent timestamps being problematic for time series analyses and the data may require interpolation, leading to information loss (Bors *et al.* 2017). It also could be indicative of malfunctioning instrumentation and/or human intervention. Perfect detection of timestamp errors requires knowledge of intended sampling rates, information not always available and can change strategically during monitor deployment. A more robust method, not requiring such information, is therefore used, involving calculating the sampling interval for each data point and detecting any instances of interval changes compared with the previously taken data point. In the datasets reviewed, target sample rates changed at most a couple of times a year, meaning that this method would result in a negligible percentage of timestamp errors with normal strategic sample rate changes and well-functioning data acquisition.

**Table 1** | Data quality rules: detection methods and possible causes

Rule	Detection by	Possible cause
Timestamp errors	Sample rate changes	Malfunctioning data acquisition
Missing data	Resample and compare to maximum possible data points at resampled rate	Battery or communications issue
Flat-lining data	Repeated values for minimum duration	Sensor or communications issue
Single-point outliers	<i>z</i> -score for window pre and post each data point	Interference with sensor measurement or location issue (if persistent)
Extended periods above threshold	Minimum rolling values for minimum duration	Sensor issue, such as fouling or real event
Extended periods below threshold	Maximum rolling values for minimum duration	Sensor issue, such as loss in sensitivity or real event
Bimodal noise	Minimum median non-zero delta in a window	Sensor issue related to power cycle or other electrical interference
Drift	Successive duration of weekly median increases	Lens/membrane fouling

### 2.2.2. Missing data

For remotely deployed water quality sensors, it is likely that at some point some data will be missing, often due to battery or communication issues. How to detect and handle missing data is a subject that is widely studied, with some form of imputation usually employed where a complete dataset is highly desirable (Allison 2000; van Buuren 2018). Detecting and quantifying the degree of missing data in a time series is once again helped by knowledge of the target sample rate. To avoid having to rely on this knowledge, an alternative approach involved first resampling the time series to the highest employed sample rate (generally 15 min). The maximum number of samples for this sample rate and timeframe was then calculated and compared with the samples in the resampled time series. This method also ensures that periods of oversampling do not interfere with the calculation of missing data, as these periods would be resampled first.

### 2.2.3. Flat-lining data

Flat-lining data occurs when sensors return the same value repeatedly. This would not be expected for sensitive water quality instruments in a dynamic environment and often (but not always) occurs at (close to) zero or at the maximum sensor value. To detect period of flat-lining data that are erroneous and a sign of a faulty sensor, it was decided to look at the total time that a sensor repeats the same value. This is intended to make detection less sensitive to the sampling rate as opposed to looking at the total number of data points with repeating values. For example, a sensor sampling every 10 s might return the same value 10 times in a row, but this is quite a different prospect from a sensor sampling every 15 min returning the same value 10 times in a row, as that would mean no water quality changes have been detected for 2.5 h. Nonetheless, sampling rate is unavoidably a significant influencing factor on a sensor's tendency to flatline, as is the resolution of the sensor.

### 2.2.4. Single-point outliers

Single-point outliers (SPOs) refer to values that are unrepresentative relative to the data before and after. These can occur in turbidity sensor data due to the presence of air bubbles or single, highly reflective particles occurring at the point of measurement. It may, however, also represent a genuine, if short-lived, event (again influenced by the sample rate). As potentially unrepresentative, these are flagged for further inspection before further analysis (Kazemi *et al.* 2018). A method was written to compare each individual data point to its surrounding data (Kazemi *et al.* 2018). The  $z$ -score, which is difference from the sample mean divided by the standard deviation, is a commonly used metric for single-point outliers in univariate signals (Grubbs 1969). This rule involves calculating the  $z$ -score for a window both 'pre' and 'post' the data point in question. The data point is considered a SPO if it exceeds a threshold, which was selected using sensitivity analysis. The window size to consider for each data point was also included in sensitivity studies.

### 2.2.5. Extended periods above (or below) a threshold

Extended periods above a threshold can indicate a sensor error or external interference but could also be a real event, so they are flagged for further inspection. Extended periods below a threshold can equally indicate a sensor error and are also flagged. The thresholds used here were designed to be generic across datasets but could also be tuned to specific network locations.

### 2.2.6. Bimodal noise

Bimodal noise was an issue identified as specific to the turbidity sensors during this trial when the sensor often fluctuated between two distinct data points. A detection method was developed, which involves calculating the median non-zero delta (with delta being the difference in amplitude from one data point to the next) over a period of time. This method uses the knowledge that turbidity sensors monitoring at or below every 15 min in DWDS are expected to record small changes in NTU and if the median delta is high, then that indicates the presence of bimodal noise. This issue was identified because of this work, with the rule developed and added to the data quality assessment, highlighting the ability to simply add or amend when employing a rule-based approach.

### 2.2.7. Drift

Drift can occur in turbidity sensors, historically linked to light source degradation but now more likely due to optical lens fouling from material accumulation, usually manifesting as a gradual baseline increase over several weeks. Drift can also occur in chlorine sensors due to deteriorating or fouling membranes, although chlorine sensor drift is often related to the sensitivity of the membrane and its ability to respond, meaning that it requires recalibration but may not exhibit gradual baseline drift behaviour. A drift detection method was developed that involved calculating the median weekly values and looking

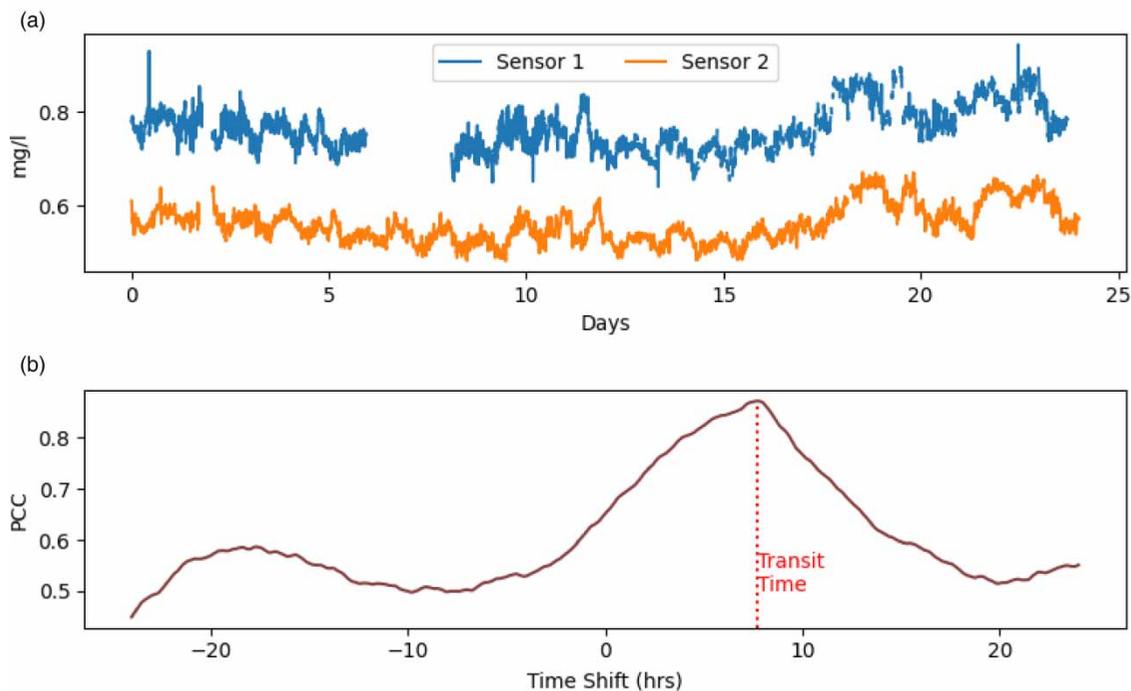
for periods that saw successive changes. As there is evidence that drift in turbidity sensors can be corrected (Gaffney & Boulton 2012), a drift correction method was developed that involves fitting the drift data using asymmetric least squares (Peng *et al.* 2010).

### 2.3. Linking sensors spatiotemporally

To understand the spatiotemporal relationships between water quality sensors deployed within a DWDS, a cross-correlation method previously developed (Gleeson *et al.* 2023) was used on both turbidity and chlorine, as well as some lesser measured parameters such as pH, temperature, and conductivity. As cross-correlation is particularly sensitive to the presence of erroneous data, it is important that this follow the Stage 1 rules. Cross-correlation is then applied to determine the strength of relationship and transit time between two water quality sensors, as illustrated in Figure 2. Transit time is defined as the average difference in hydraulic arrival times between two locations, which are not necessarily directly in line. The top plot in Figure 2 shows two chlorine sensor time series, over the course of 24 days. The bottom plot displays the cross-correlation curve, the peak of which is the time shift, which results in the strongest correlation. The maximum correlation coefficient was found to occur for a time shift of 7.7 h (indicated by the dotted red vertical line). As this example shows, this method can handle missing data (seen in Sensor 1), which is vital for use in this framework due to the first stage involving removing flagged data points. However, calculating meaningful cross-correlations from chlorine sensor data is not always as straightforward as this example may imply due to network and hydraulic complexities. The steps required before such calculations can be done are explored in the results and discussion.

### 2.4. Multi-sensor data quality validation

The final stage of the data quality framework combines the output from the single sensor data quality rules and the spatiotemporal information. Where the spatiotemporal information indicates that a sensor has one or more sensors with strong connectivity, the detected data quality rules can be reassessed with the additional context of these connected and therefore comparable sensor(s). The derived transit time between the sensor locations can also be helpful in synchronising the errors. For example, if Sensors A and B are connections and have data quality rules detected that are synchronised according to the derived transit times, these periods of data can be considered to be real and not containing sensor errors. Similarly, if a flagged



**Figure 2** | Two example chlorine time series (a) and corresponding cross-correlation curve (b). Please refer to the online version of this paper to see this figure in colour: <https://doi.org/10.2166/aqua.2023.228>.

rule is only seen on one sensor, it should be investigated as a sensor error. In reality, it is not possible to make absolute statements without physically inspecting the sensor or obtaining metadata regarding network operations, but this framework provides the tools to perform a cross-sensor validated data quality assessment purely based on turbidity and chlorine time series data.

## 2.5. Datasets

Three real-world water quality DWDS time series datasets from three different parts of the UK were used to develop and demonstrate the data quality assessment framework. The details of these datasets are given in Table 2. All water quality sensors listed monitor turbidity and chlorine, with some other water quality parameters also included less frequently, such as conductivity, pH, and temperature. Dataset B was completed before this work began, with Datasets A and C becoming available while monitoring was ongoing. This meant that the ability of the framework to assess sensor performance could be assessed both off-line with historic data and in near real-time via an API (application programming interface).

## 3. RESULTS

### 3.1. Data quality assessment rules

The rules were developed and refined using the 92 available multi-parameter sensors from the three independent DWDS datasets.

#### 3.1.1. User-definable input parameter values

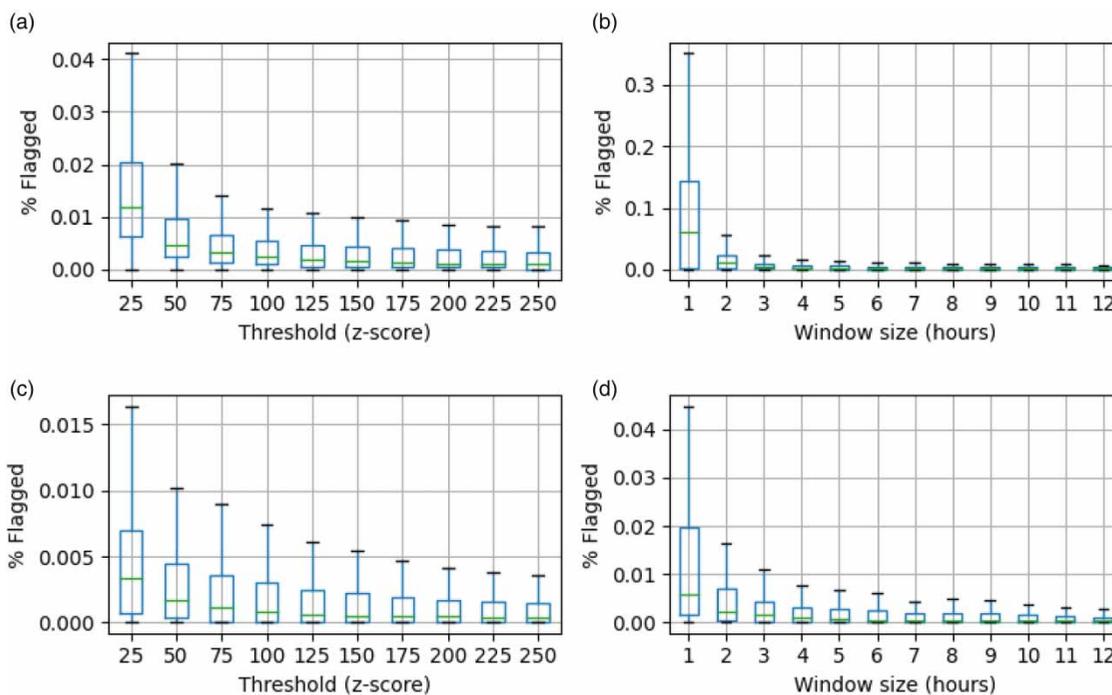
Sensitivity studies were conducted for rules where there were user-defined input parameters in order to understand the impacts of changing these values on the total amount of data points flagged. Figure 3 shows an example of the sensitivity data for single-point outliers, which informed the selection of a  $z$ -score threshold of 100 with a window of 6 h, ensuring that only significant instances of single-point outliers are detected. The values ultimately selected for each rule, determined with the aid of further sensitivity analysis, are listed in Table 3. There was no need to perform sensitivity studies on missing data or timestamp errors, as these do not have user-definable values. It is noted that input duration or window size was selected using all datasets so accommodates the different sampling rates encountered in this work, while thresholds were based on reasonable expected values. An advantage of the rule-based approach adopted here is that there is no need for labelling of sensor errors. However, if labelled sensor errors were available, it would be possible to investigate and fine-tune parameter selection.

#### 3.1.2. Timestamp errors and missing data

Figure 4 shows a plot of the sampling intervals for a single water quality sensor over the course of 18 months (blue dots), with the red line showing the accumulated timestamp errors, where each detected timestamp error equals 1. In this example, 13.7% of the data points were detected as timestamp errors. Using an assumed target sample interval of 15 min for the first 12 months and 1 min for the last 6 months, the actual intervals can be compared with this assumed target. Using this method, results in 12% of the data points being flagged as timestamp errors. This shows that the original method slightly over-estimated the prevalence of timestamp errors, explained by this example of fluctuating in and out of target sample intervals (the first instance of return to the target sample rate will be flagged as a timestamp error). To determine the quantity of missing data in this example, the time series resampled to 15-min intervals results in 44,264 nonempty samples, compared with a maximum of 52,323 for this timeframe, meaning that a total of 8,059 missing resampled data points were calculated (or 15% of the maximum potential data).

**Table 2** | Datasets used

Dataset	Number of sensors	Duration	Sampling interval
A	12	1 year	2 min
B	62	1.5 years	51 sensors at 15 min for 12 months, then 1 min for 6 months 11 sensors every 2 min
C	18 originally (later reduced to 11)	2.5 years (ongoing)	15 min for first 20 months, 2 min for 2 months, 5 min for last 8 months



**Figure 3** | Box plots showing sensitivity of single-point outlier rule for turbidity sensors, with (a) z-score threshold and (b) window size and for chlorine sensors, with (c) z-score threshold and (d) window size.

**Table 3** | Data quality rules: input parameter values

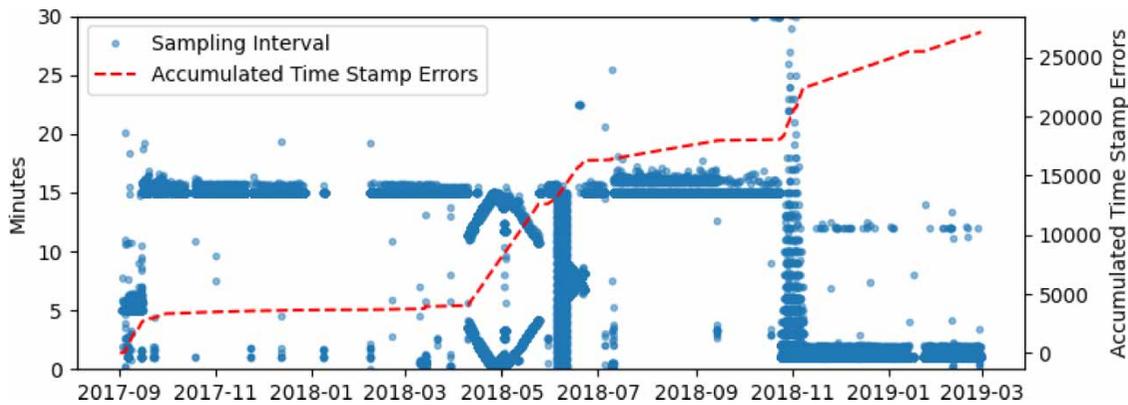
Rule	Detection by	User-defined input parameter values
Flat-lining data	Repeated values for minimum duration	Minimum duration = 8 h
Single-point outliers	z-score for window pre and post each data point	z-score threshold = 100 Window size = 6 h
Extended periods above threshold	Minimum rolling values for minimum duration	Minimum threshold = 1.5 NTU/1.5 mg/l Cl Minimum duration = 6 h
Extended periods below threshold	Maximum rolling values for minimum duration	Maximum threshold = 0.05 NTU/0.15 mg/l Cl Minimum duration = 6 h
Bimodal noise	Minimum median non-zero delta in a window	Minimum threshold = 0.1 Window size = 6 h
Drift	Successive duration of weekly median increases	Minimum duration = 4 weeks Minimum overall rise = 0.3 NTU/0.3 mg/l Cl

### 3.1.3. Flat-lining data

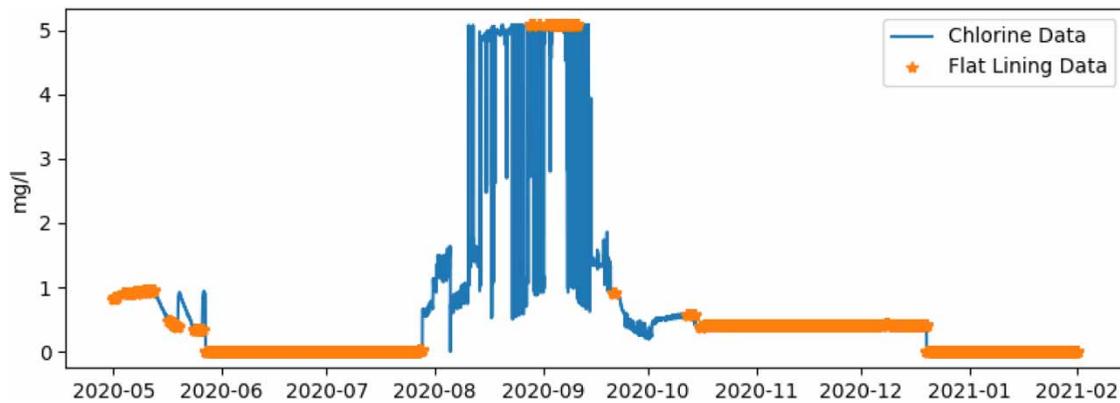
Figure 5 shows an example of a chlorine sensor with significant levels of flat-lining. The flat-lining duration threshold here was set to 8 h, with 66% of the data surpassing this level in the 8 months' worth of data shown here.

### 3.1.4. Single-point outliers

Figure 6 shows an example, using a z-score threshold of 100 and a window size of 6 h, of a detected single-point outlier on the left, and two spikes left undetected on the right. In the LHS example, a single-point outlier was detected as the z-score both 6 h pre and post scored above the threshold, while the RHS example illustrates how this method deals with two spikes occurring within the same window. For the first spike, only the pre z-score went above 50 while the post z-score was influenced by the presence of the next spike. The opposite occurred for the second spike.



**Figure 4** | Time series plots of sampling intervals (blue) and accumulated timestamp errors (red). Please refer to the online version of this paper to see this figure in colour: <https://doi.org/10.2166/aqua.2023.228>.



**Figure 5** | Chlorine time series with flatlines of at least 8 h highlighted.

### 3.1.5. Extended periods above (or below) a threshold

Figure 7 shows a chlorine time series where both extended periods above and below the set thresholds were detected. The upper limit used in this example was 1.5 mg/l, with the lower limit being 0.15 mg/l, and the minimum duration was 6 h. In this example, nearly 1 month of data was above 1.5 mg/l, followed by about a week at very low levels below 0.15 mg/l. The disinfection residuals for the UK system in which this was deployed were designed to stay above 0.2 mg/l and below 1 mg/l.

### 3.1.6. Bimodal noise

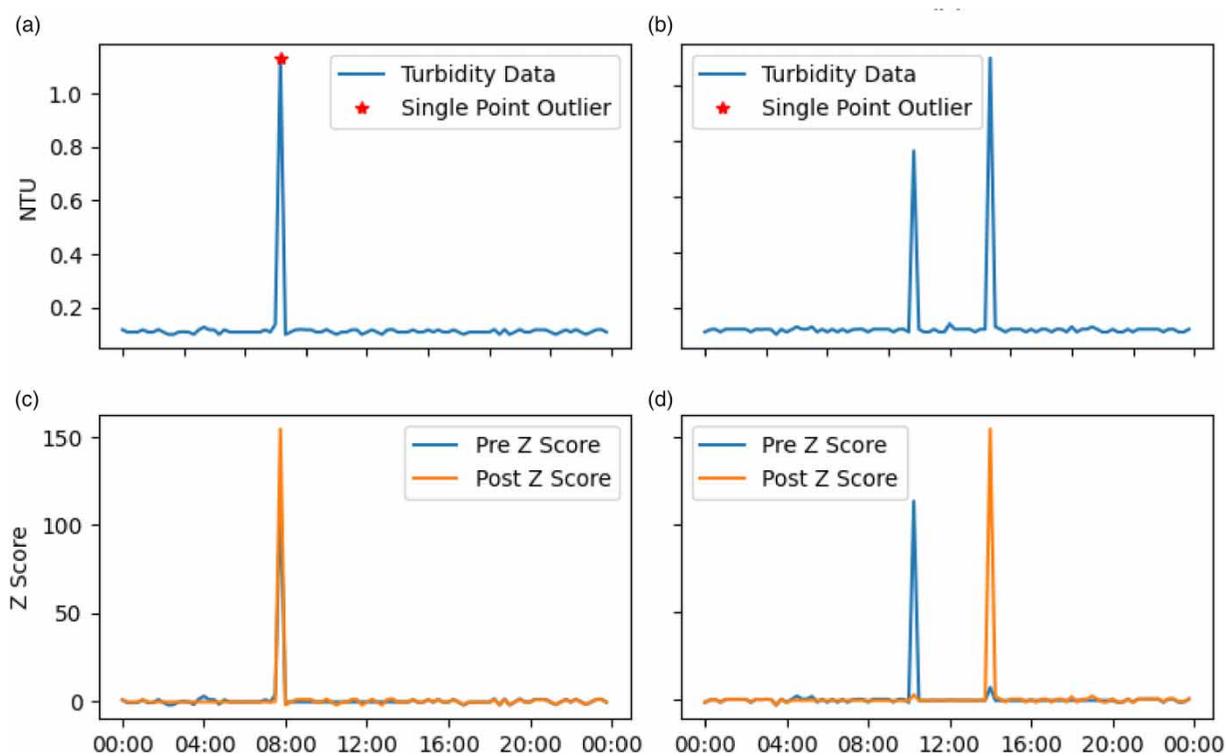
Bimodal noise is illustrated in Figure 8, where bimodal noise was detected to be occurring for around 78% of the 15-month period shown, using a threshold of 0.1 NTU and a window size of 6 h.

### 3.1.7. Drift

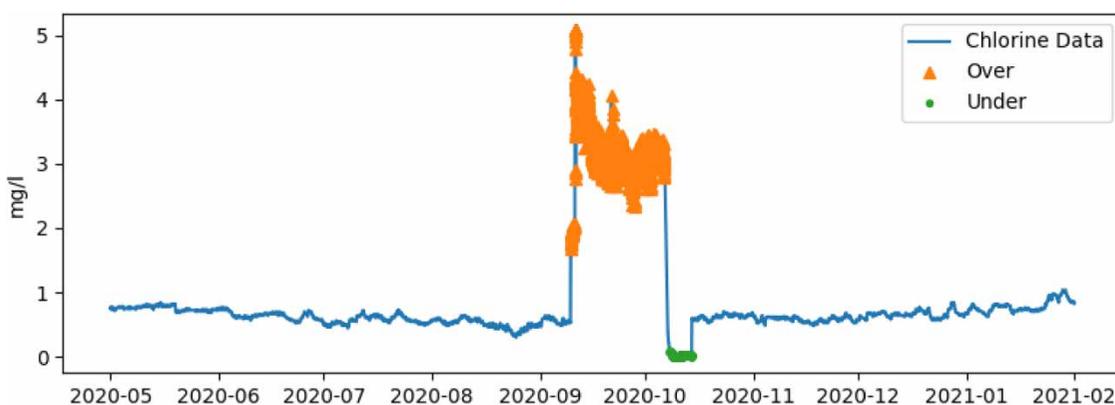
Figure 9 shows a turbidity sensor that was prone to drift. Over the 11-month period shown in this plot, over 90% of the data was calculated to be part of a drift period, using a minimum of 4 weeks of successive weekly median increases as a drift period and a minimum overall increase of 0.3 NTU. The bottom plot of Figure 9 shows the drift data corrected using asymmetric least squares.

### 3.1.8. Rules applied to datasets

The eight rules were applied to each of the three datasets listed in Table 2, enabling sensor performance to be assessed and ranked, as illustrated in stacked bar charts in Figure 10. These results highlight the multiple data quality issues seen in these



**Figure 6** | (a) Turbidity time series with single-point outlier detected, (b) turbidity time series with no single-point outlier detected, (c) pre and post z-scores corresponding to (a), and (d) pre and post z-scores corresponding to (b).



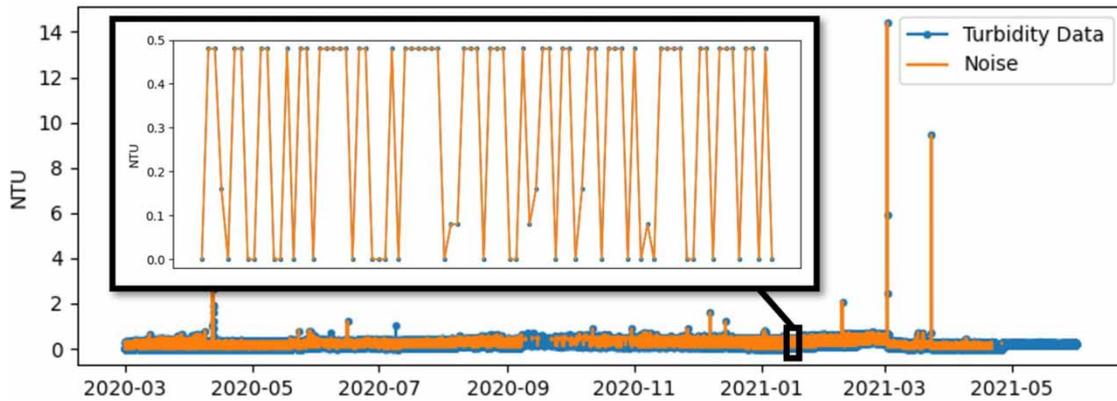
**Figure 7** | Chlorine time series with regions staying over 1.5 mg/l, or under 0.15 mg/l, for 6 h highlighted.

three datasets and show how prevalent each rule was for turbidity and chlorine sensors. Timestamp errors were only seen in Dataset B. Missing data was consistently seen in all datasets. Flat-lining data was seen most in Dataset C and was in both chlorine and turbidity sensors, often (but not always) at the same time. Single-point outliers and bimodal noise were more common in turbidity sensors. Drift was seen equally in turbidity and chlorine.

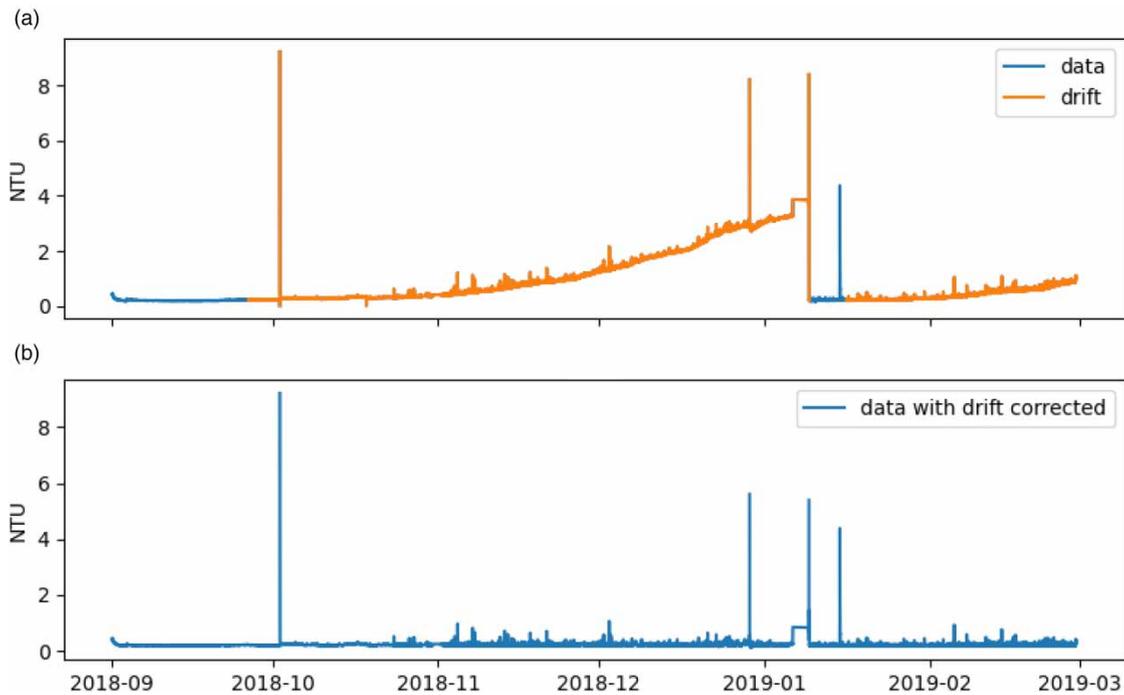
### 3.2. Linking sensors spatiotemporally

#### 3.2.1. Cross-correlating water quality sensors

Cross-correlation was tested on multiple water quality parameters to determine which were the most suited. Suitability was determined by manually examining DWDS schematics and by discussing with utilities. For the correlations to be valid sensors must provide sufficient good-quality data in common, set at 50% of the total window length for this work. This 50%



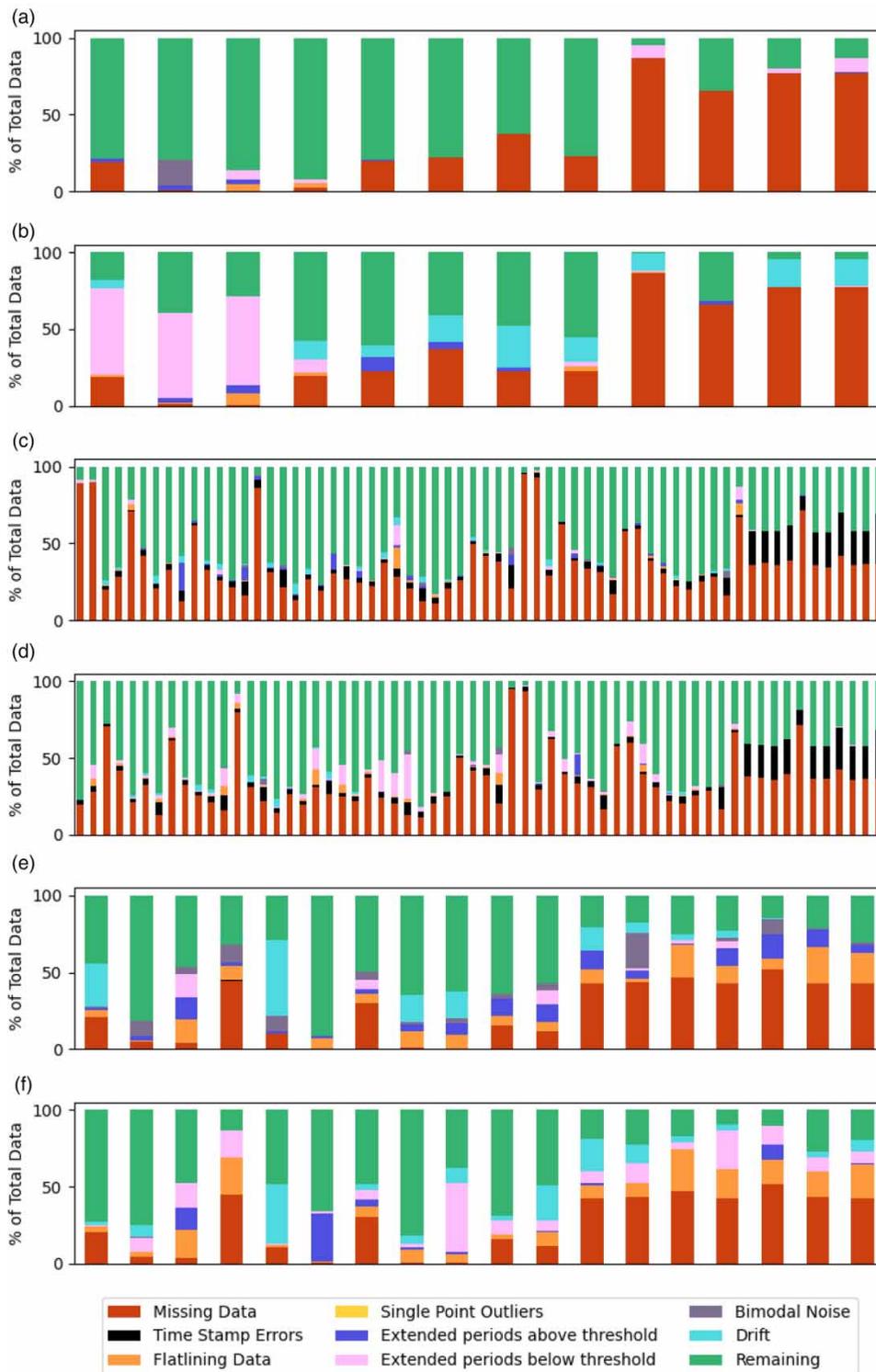
**Figure 8** | Turbidity time series with closeup of highlighted bimodal noise.



**Figure 9** | (a) Turbidity time series with two periods of drift highlighted and (b) drift corrected turbidity data.

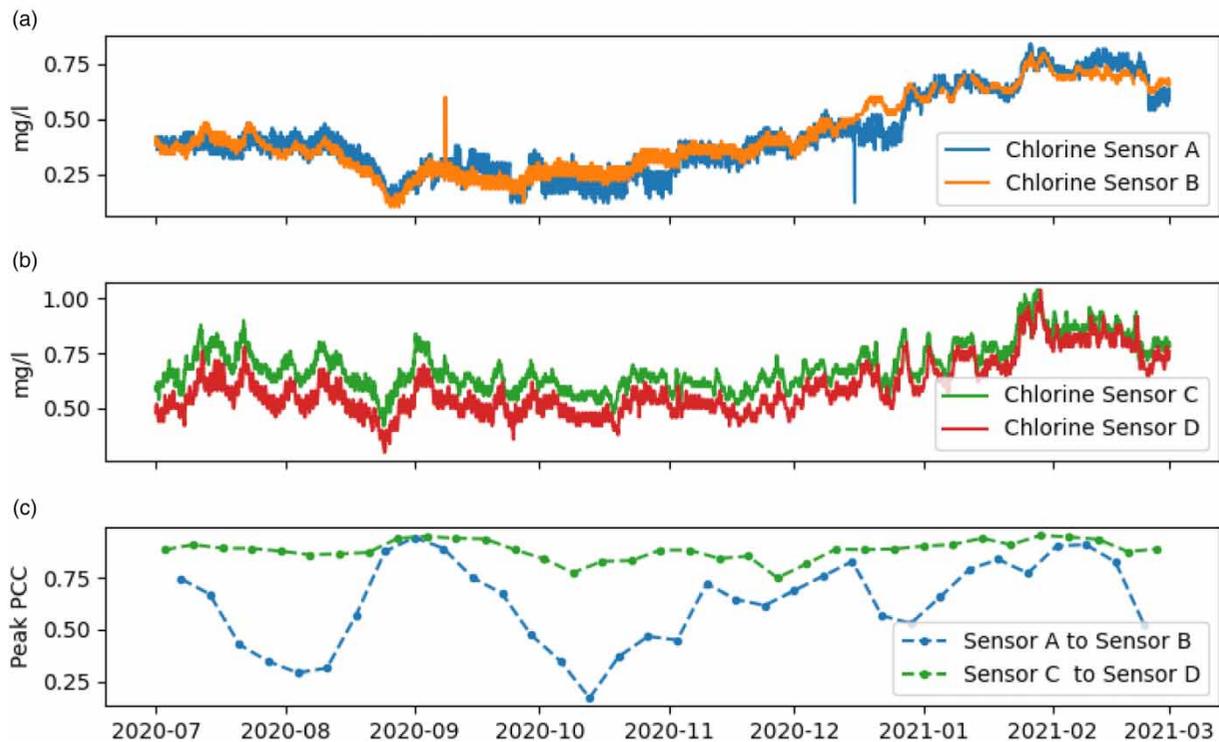
commonality limit was selected to ensure that correlations were meaningful while also allowing for the long periods of missing or low-quality data experienced. PCCs are calculated at different time shifts for each possible sensor pair, with the strength of connectivity represented by the highest correlation coefficient and with the temporal shift of this highest correlation also designating the transit time. The transit time is only valid if the maximum PCC is sufficiently high. For this work, a threshold of 0.7 was used, as any values above this are widely accepted to indicate a strong correlation (Schober *et al.* 2018). For longer time series up to a year in duration, the cross-correlations were done on shorter 4-week periods, with the median cross-correlation coefficient across the entire time series reported. This was done to avoid the correlations being dominated by seasonal trends shared by many unrelated locations, with shorter time frames more likely to support hydraulic connectivity.

The cross-correlations results presented were all calculated on chlorine time series data, a parameter that was well-suited for this method. Figure 11 illustrates why longer time series need to be split into smaller sections for the correlations to be meaningful. In this example, two chlorine sensor pairs (A and B; and C and D) were found to be highly correlated over the 8-month period shown, but upon inspection, A and B were only distantly related in the network but over a long timeframe



**Figure 10** | Data quality rules applied to Dataset A turbidity (a) and chlorine sensors (b), Dataset B turbidity (c) and chlorine sensors (d) and Dataset C turbidity (e), and chlorine sensors (f).

displayed similar seasonal chlorine trends, possibly due to their sharing of the same treatment works. When cross-correlations were performed using window sizes of 4 weeks, the median PCC for A–B was below the significance level of 0.7, while the median PCC for C–D was above. A 4-week window size was used, calculated once a week, with median coefficients presented



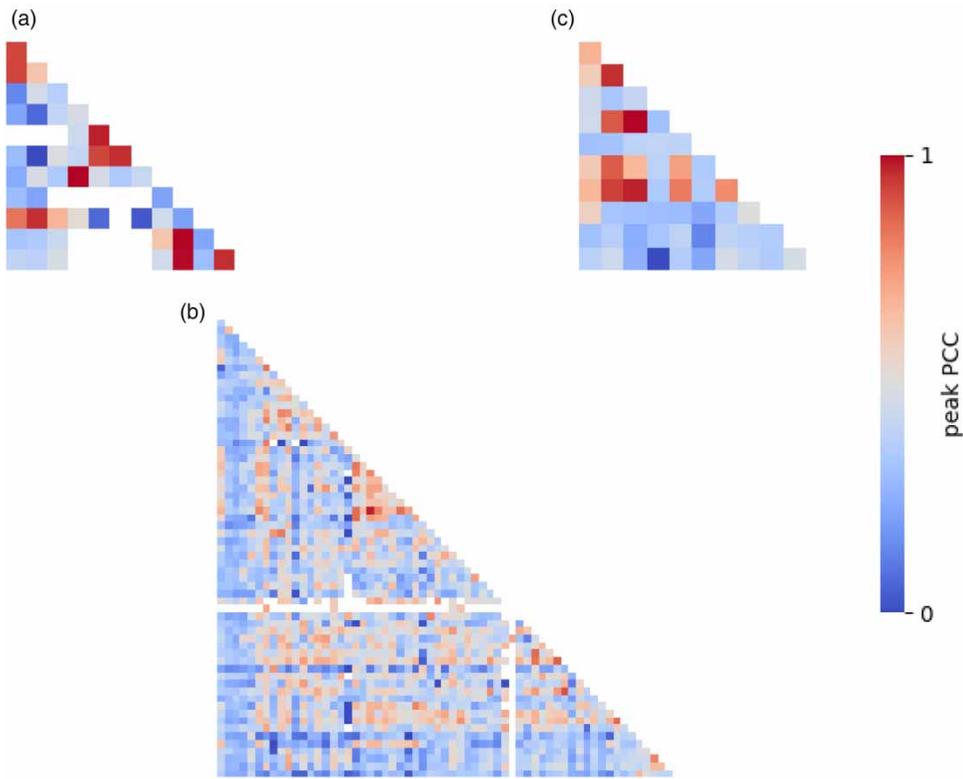
**Figure 11** | Chlorine time series profile for two pairs, shown in (a) and (b), and (c) the sliding cross-correlation coefficients calculated using overlapping 4-week windows every 7 days.

for each sensor pair in heatmaps in Figure 12(a)–12(c) for Datasets A, B, and C, respectively. The implied connectivity was verified using schematics showing sensor locations and through discussion with utility operators for Dataset C, indicating this method's suitability in implying sensor connectivity. The blank squares in these plots indicate sensor pairs with insufficient data in common, after removing flagged anomalies.

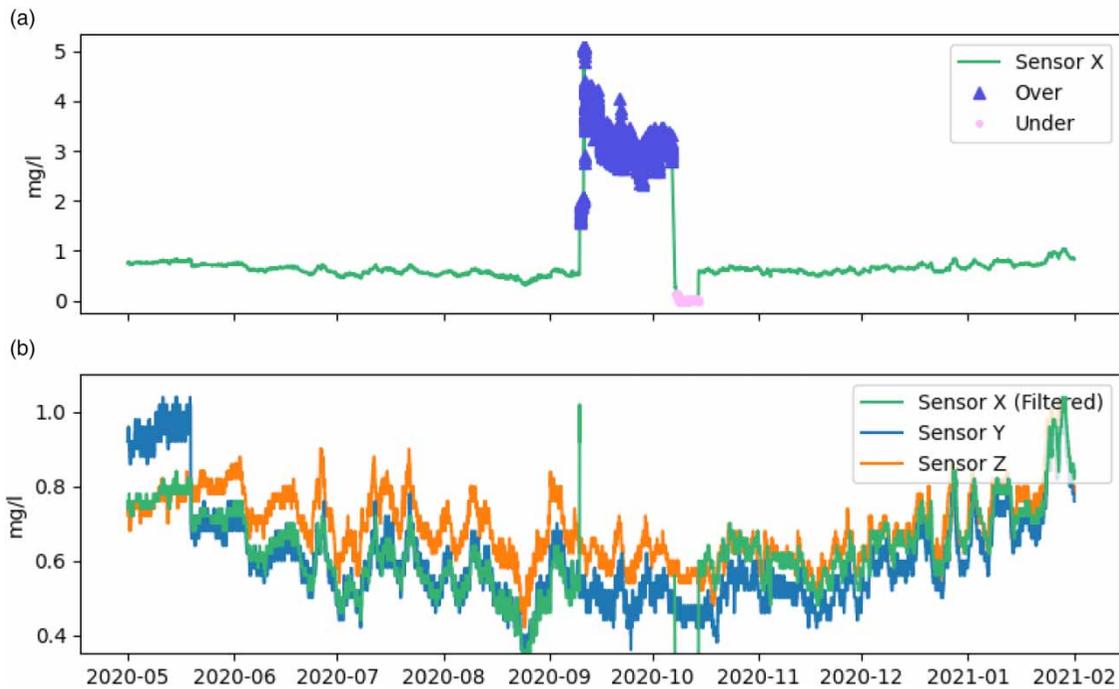
### 3.3. Multi-sensor data quality validation

The final step in the framework uses the spatiotemporal information, derived from the cross-correlation of chlorine sensors, to enable multi-sensor data validation. Figure 13(a) shows an example of chlorine Sensor X with two extended periods above and below selected thresholds. In this case, the cross-correlation results provided information that two other connected sensors, shown in Figure 13(b), along with chlorine Sensor X with anomalous periods filtered out, continued to record chlorine data similar to that seen in normal operation. This provides higher confidence that these anomalous periods are data quality issues, rather than real network events. Hence, rather than flagging this data, it can be more confidently removed.

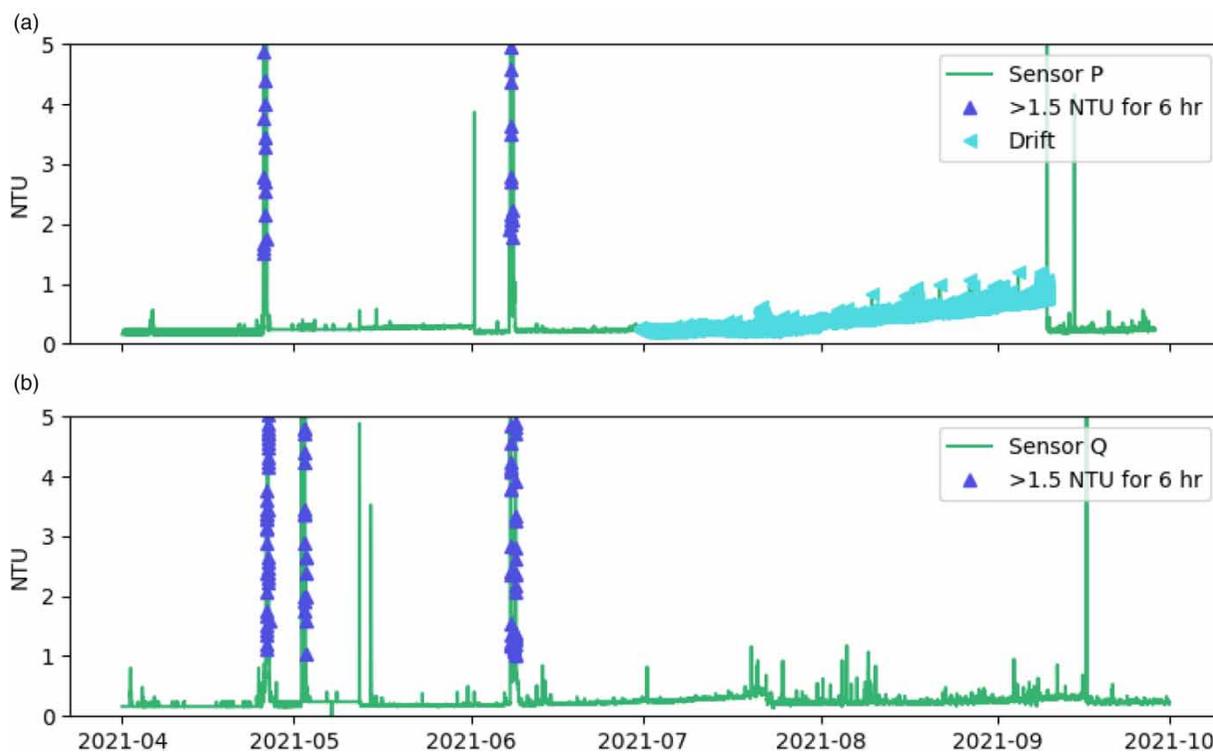
Figure 14 illustrates how the spatiotemporal information derived from cross-correlations done on chlorine sensors can be utilised for other parameters at the same locations. In this case, the chlorine-based cross-correlations provide information that Sensor P in Figure 14(a) and Sensor Q in Figure 14(b) are connected and can be compared. In this example, a period of drift is detected in Sensor P but not seen in Sensor Q, indicating that this drift is likely a sensor issue. Sensor P also had two detected instances of extended periods above 1.5 NTU and a detected period of drift. Sensor Q had similar corresponding extended periods above 1.5 NTU, indicating that these are real events passing through this network section. Sensor Q also had an additional period above 1.5 NTU in May 2021, not seen in Sensor P. As this third event cannot be validated as a real network event, this remains a potential sensor error. Of course, this could also be an event localised to Sensor Q, particularly as it has confirmed similar events before and after. This example highlights the complexity of these natural systems and underlines why this final stage of the framework currently requires a subjective evaluation using all the information at hand. The only way to be certain these are sensor errors is through a physical inspection of the sensor, though this framework provides tools to make informed, cross-validated decisions based purely on turbidity and chlorine data.



**Figure 12** | Heatmaps with peak PCC for each chlorine sensor pair in Datasets A (a), B (b), and C (c). Blank squares indicated insufficient data.



**Figure 13** | Multi-sensor data quality validation example, with (a) flagging of periods above 1.5 mg/l and below 0.15 mg/l for 6 h in Sensor A and (b) showing the absence of anomalous feature in two sensors calculated to be correlated to Sensor A.



**Figure 14** | Multi-sensor data quality validation example, with (a) showing instances of periods above 1.5 NTU for 6 h and a detected period of drift in Sensor A and (b) showing detected periods above 1.5 NTU for 6 h in Sensor B, which was calculated to be highly correlated to Sensor A.

## 4. DISCUSSION

This work provides a data quality assessment framework for water quality sensors within DWDS, developed and demonstrated on turbidity and chlorine. Of the three-stage framework, the first has been automated, meaning that a single sensor data quality assessment can be quickly provided for any dataset containing turbidity and chlorine time series data. The second stage has not been fully automated, as care must be taken when interpreting cross-correlation results to be indicative of network connectivity and it is still recommended to perform a visual check on the chlorine data. The final stage is the least automated and requires an analyst to evaluate rules detected by single sensors using the derived spatiotemporal information. However, the logic presented in this framework provides a platform from which an automated multi-sensor, multi-parameter data quality assessment system could be developed. The data sets used comprised different hardware and software and installation and maintenance practices with the quality assurance framework being agnostic to these. Another strength of the framework is that, as it is unsupervised and applicable across multiple parameters, it does not require labelled data sets, such as those previously developed for water quality sensors deployed in rivers (Talagala *et al.* 2019). Such labelling of outliers by experts, to inform and compare detection performance, is (if possible) time-consuming and is often parameter-specific and even data set-specific. A risk of the rule-based approach taken is the requirement for user-defined values, but the sensitivity studies and application of single values across the diverse datasets explored here give confidence that these were robust. The framework presented was developed on UK DWDS datasets and focused on turbidity and chlorine, two of the most measured water quality parameters. The datasets did not include sufficient data from other parameters, but the limited exploration possible showed them to be suitable for this approach. Two of the three datasets were analysed during sensor deployment, by accessing uploaded sensor data through an API, enabling sensor maintenance and deployment strategies to be informed by the latest sensor performance and demonstrating the near real-time potential of this framework.

### 4.1. Rules, framework Stage 1

The data quality rules, Stage 1, were developed to detect specific anomalous instances and have been shown to effectively achieve this, but each rule, and hence the resulting data flagged, is subject to user-defined input variables. The single-point

outlier rule considers that a data point unrepresentative of its surrounding data is an error, but this is not necessarily the case and care must be taken to avoid the removal of real, if irregular, data. If a sensor has repeated instances of single-point outliers, this could be indicative of a real but potentially undesired external factor, such as a nearby valve, which could cause short discolouration events due to small amounts of material building up and becoming dislodged. Higher frequency sampling (closer to 1 sample/min) would assist in determining whether single-point outliers are genuine or not. The ‘extended period above/below a threshold’ rule has perhaps the greatest potential for flagging of valid data. The dismissal of these periods as sensor errors without cross-sensor validation is not recommended. Hence, the framework revisits the appropriate rules in Step 3 following cross-correlation in Stage 2. This re-checking of data flagged by the rules following cross-correlation is also recommended for single-point outliers and drift. An additional advantage of a rule-based approach is that they can easily be added or changed, as was the case with the bimodal noise rule, which was added after being identified during monitoring.

The data quality assessment results shown in [Figure 10](#) provide a visual impression of the data quality seen in each dataset. The rule-based approach allows for the prevalence of each specific feature to be compared across sensor locations and datasets. Timestamp errors were only seen in Dataset B, which was the earliest of the datasets and used similar instrumentation, indicating that this data acquisition error may have since been fixed. However, due to the potential negative implications of this error, it is worth continuing to detect. Missing data was consistently seen in all datasets, but quantifying the missing data was limited by the lack of knowledge of intended sensor deployment timeframes. For example, a sensor in a dataset may have been intentionally taken out of service, but this analysis did not always have access to that kind of operational information. Dataset C had the most flat-lining data, the cause of which is unknown but could potentially be related to sensors being removed from deployment but continuing to take data as maintenance information was not always available. Single-point outliers and bimodal noise were more common in turbidity sensors, as expected as these rules were originally developed for errors seen in turbidity sensors. Drift was seen equally in turbidity and chlorine, but not with equal confidence. Chlorine ‘drift’ could be due to changes in chlorine dosing, such as in response to seasonal temperature change, with revisiting chlorine flagged as drifting following cross-correlation valuable.

#### 4.2. Cross-correlation, framework Stage 2

The cross-correlation analysis was found to work well for chlorine time series data, due to the way chlorine decays steadily as it passes through a DWDS, leaving connected sensors with similar chlorine time series profiles, with a decay and time lag that cross-correlation can determine. This method was not found to work as well for turbidity sensors due to the time series profiles tending to be flatter unless there were specific network events. Cross-correlation was explored for the limited data for other parameters available here. pH and conductivity showed some promise, as did temperature. Patterns in temperature data occur and propagate within DWDS due to the heating or cooling effects of the surrounding ground as a function of patterns of residence time. Hence, the parameters likely to be effective for cross-correlation are those with an expected time-dependent behaviour occurring within DWDS pipes. Where the chlorine time series is too flat, for example, immediately downstream of a well-controlled dosing point, this method will not work well. DWDS network features such as service reservoirs, valves, pumps, etc., also interfere with water quality, making sensors either side of such features difficult to correlate. As shown in the results, window size is a major factor to consider when doing this analysis and using too big a window can lead to the correlations being dominated by common seasonal trends. Performing the cross-correlation using the overlapping 4-week windows helped ensure that over the course of long-term datasets, flat periods or seasonal trends would not dominate the results. An issue with chlorine sensors is the need for regular recalibration to promote confidence in the baseline values. As correlations are not however, affected by absolute values, this method is unaffected by poorly calibrated chlorine sensors, resulting in an effective method for determining the spatiotemporal relationship between chlorine sensors.

The cross-correlation analysis provides an indication of connectivity and transit time. It should be noted that connectivity is not as simple as up and down stream rather than the two-sensor location experiences similar water with some time lag. This could, for example, be sensors on two legs of a branched system. Hence, the transit time is not simply the time to go from A to B, it can also be the difference in time for similar water to reach to different points in a network. This is still a valuable insight, but care must be taken in the interpretation of its meaning and its further use. The spatiotemporal information is used in this work to improve data quality assessment, but it can also be used to characterise network events. For example, an event could be described as local to a specific sensor or global and seen by multiple sensors. Knowing the connectivity and transit times is necessary to be sure about such conclusions. Global events that travel through the network can be assessed with knowledge

of hydraulic transit times, which could help in locating the source and destination of an event, allowing for event mitigation. Connectivity and transit time information can support and improve hydraulic models, particularly useful when adding water quality functionality as higher standards of calibration would be required (Boxall *et al.* 2004), and accuracy of otherwise not-straightforward disinfection residual modelling (Speight & Boxall 2015). The use of cross-correlation analysis in DWDS is not unique and has been used to detect leaks by looking for drops in cross-correlations (Gomes *et al.* 2021). A similar method could potentially be explored to detect anomalous data in chlorine sensors that are correlated, but data quality rules would be required to enable the correlations initially. A form of correlation, semblance analysis, was also previously used to associate daily cycles in turbidity and flow or pressure (Mounce *et al.* 2015). This analysis relied on time-consuming manual data quality checking, addressed here, but more importantly showing the value and deeper insight that can potentially be gained by further analysis of quality-assured data integrated across quantity and quality data.

### 4.3. Multi-sensor validation, framework Stage 3

Whether to remove, flag, or interpolate detected anomalies depends on the requirements of the subsequent analysis. For cross-correlation, it was desired to correlate the baseline performance of chlorine sensors. Therefore, the rules were applied with any detected instances removed. Even though this may have resulted in real network events being removed, this was desired in this case. For most other analytic needs rules such as timestamp errors, flat-lining data, and bimodal noise should rarely be left in. For other detections, a single sensor and parameter does not give enough information, which is why the cross-sensor validation is required. This idea is illustrated with the two case studies in Figures 13 and 14. Figure 13 shows a chlorine sensor with detected anomalous periods of both high and low chlorine. Comparisons to other connected sensors, known from cross-correlation results, show that this anomalous behaviour is likely the result of a fault specific to this sensor, rather than a real network event. Such a conclusion could not have been made with high confidence without the additional context provided by the connected sensor data. Figure 14 shows how the cross-correlation results from chlorine sensors can be applied to other parameters, in this case turbidity. In this example, both sensors experience periods of elevated turbidity at the same time, indicating a real network event. The transit time information given by the cross-correlation analysis allows for the direction of travel to be known and for the event to be assessed as it tracks through this network section. Other information, such as flow data, maintenance records and customer complaints, would be useful in determining the cause of flagged rules. The final multi-sensor validation stage would be challenging to fully automate as comparing data quality rules across connected sensors is somewhat subjective. However, it could be automated to the point that any flagged data quality rule would come with details on whether similar issues are seen in known connected sensors, enabling operators to make quick decisions. The additional value from linking sensors spatiotemporally has been demonstrated and has implications for sensor deployment strategies, with connected monitoring locations providing greater potential for network insights. However, the connectivity between locations cannot always be inferred from schematics and can often be unexpected. Therefore, a practical approach to obtaining a spatiotemporally connected and performance assured sensor network within a DWDS is by applying this framework and redeploying until desired connectivity is achieved.

## 5. CONCLUSIONS

- This work presents and demonstrates an effective multi-sensor data quality assessment framework that combines an automated, single sensor rule-based data quality assessment with spatiotemporal cross-correlation, facilitating data quality assurance for turbidity and chlorine sensors deployed within DWDS.
- The framework worked for different hardware configurations across three extensive real-world DWDS water quality datasets and was demonstrated to work both on historic data and in near real-time.
- The rule-based approach developed detected and quantified the presence of anomalous features, allowing sensor performance to be evaluated and possible causes to be proposed. The nature of the rules allows rapid and simple modification, but with standardised settings found (via sensitivity studies) and used here across three large, disparate datasets.
- Cross-correlation has been shown to work effectively on chlorine data, supporting data quality assessment and understanding of system connectivity, including transit time between sensors.
- By applying this multi-sensor data quality assessment framework, water utilities can extract added value from water quality sensors and provide high confidence data for further automated or manual analysis, helping bridge the gap between data and actionable information

## ACKNOWLEDGEMENTS

This research has been supported by an Engineering and Physical Sciences Research Council (EPSRC) studentship as part of the Centre for Doctoral Training in Water Infrastructure and Resilience (EP/S023666/1) with support from industrial sponsor Siemens UK, Anna Taliana and Paul Gaskin of Welsh Water, Jez Downs of Southern Water, Katrina Flavell of Yorkshire Water and Derek Leslie of ATi.

## DATA AVAILABILITY STATEMENT

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

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Allison, P. D. 2000 [Multiple imputation for missing data: a cautionary tale](#). *Sociological Methods and Research* **28** (3), 301–309. doi:10.1177/0049124100028003005.
- Benesty, J., Chen, J. & Huang, Y. 2004 [Time-delay estimation via linear interpolation and cross correlation](#). *IEEE Transactions on Speech and Audio Processing* **12** (5), 509–519. doi:10.1109/TSA.2004.833008.
- Boker, S. M., Rotondo, J. L., Xu, M. & King, K. 2002 [Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series](#). *Psychological Methods* **7** (3), 338–355. doi:10.1037/1082-989X.7.3.338.
- Bonnefous, O. 1986 [Time domain formulation of pulse-doppler ultrasound and blood velocity estimation by cross correlation](#). *Ultrasonic Imaging* **8** (2), 73–85. doi:10.1016/0161-7346(86)90001-5.
- Bors, C., Bögl, M., Gschwandtner, T. & Miksch, S. 2017 [Visual support for rastering of unequally spaced time series](#). In: *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction*. ACM, New York, NY, USA, pp. 53–57.
- Boxall, J. B. & Saul, A. J. 2005 [Modeling discoloration in potable water distribution systems](#). *Journal of Environmental Engineering* **131** (5), 716–725. doi:10.1061/(ASCE)0733-9372(2005)131:5(716).
- Boxall, J. B., Saul, A. J. & Skipworth, P. J. 2004 [Modeling for hydraulic capacity](#). *Journal – American Water Works Association* **96** (4), 161–169. doi:10.1002/j.1551-8833.2004.tb10607.x.
- Colford, J. M., Roy, S., Beach, M. J., Hightower, A., Shaw, S. E. & Wade, T. J. 2006 [A review of household drinking water intervention trials and an approach to the estimation of endemic waterborne gastroenteritis in the United States](#). *Journal of Water and Health* **4** (S2), 71–88. doi:10.2166/wh.2006.018.
- Dürrenmatt, D. J., Del Giudice, D. & Rieckermann, J. 2013 [Dynamic time warping improves sewer flow monitoring](#). *Water Research* **47** (11), 3803–3816. doi:10.1016/j.watres.2013.03.051.
- Gaffney, J. W. & Boulton, S. 2012 [Need for and use of high-resolution turbidity monitoring in managing discoloration in distribution](#). *Journal of Environmental Engineering* **138** (6), 637–644. doi:10.1061/(ASCE)EE.1943-7870.0000521.
- Gleeson, K., Husband, S., Gaffney, J. & Boxall, J. 2023 [Determining the spatio-temporal relationship between water quality monitors in drinking water distribution systems](#). *IOP Conference Series: Earth and Environmental Science* **1136** (1), 012046. doi:10.1088/1755-1315/1136/1/012046.
- Gomes, S. C., Vinga, S. & Henriques, R. 2021 [Spatiotemporal correlation feature spaces to support anomaly detection in water distribution networks](#). *Water* **13** (18), 2551. doi:10.3390/w13182551.
- Grubbs, F. E. 1969 [Procedures for detecting outlying observations in samples](#). *Technometrics* **11** (1), 1. doi:10.2307/1266761.
- Horvatic, D., Stanley, H. E. & Podobnik, B. 2011 [Detrended cross-correlation analysis for non-stationary time series with periodic trends](#). *EPL (Europhysics Letters)* **94** (1), 18007. doi:10.1209/0295-5075/94/18007.
- Husband, P. S. & Boxall, J. B. 2011 [Asset deterioration and discoloration in water distribution systems](#). *Water Research* **45** (1), 113–124. doi:10.1016/j.watres.2010.08.021.
- Husband, S., Fish, K. E., Douterelo, I. & Boxall, J. 2016 [Linking discoloration modelling and biofilm behaviour within drinking water distribution systems](#). *Water Supply* **16** (4), 942–950. doi:10.2166/ws.2016.045.
- Kazemi, E., Mounce, S., Husband, S. & Boxall, J. 2018 [Predicting turbidity in water distribution trunk mains using nonlinear autoregressive exogenous artificial neural networks](#). **3**, 1030–1019. doi:10.29007/9r3b.
- Keogh, E. J. & Pazzani, M. J. 2001 [Derivative dynamic time warping](#). In: *Proceedings of the 2001 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, Philadelphia, PA, pp. 1–11.
- Kianimajd, A., Ruano, M. G., Carvalho, P., Henriques, J., Rocha, T., Paredes, S. & Ruano, A. E. 2017 [Comparison of different methods of measuring similarity in physiologic time series](#). *IFAC-PapersOnLine* **50** (1), 11005–11010. doi:10.1016/j.ifacol.2017.08.2479.
- Kirchen, I., Schutz, D., Folmer, J. & Vogel-Heuser, B. 2017 [Metrics for the evaluation of data quality of signal data in industrial processes](#). In: *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE, pp. 819–826.

- Krishnamachari, B. & Iyengar, S. 2004 Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. *IEEE Transactions on Computers* **53** (3), 241–250. doi:10.1109/TC.2004.1261832.
- Mann, A. G., Tam, C. C., Higgins, C. D. & Rodrigues, L. C. 2007 The association between drinking water turbidity and gastrointestinal illness: a systematic review. *BMC Public Health* **7** (1), 256. doi:10.1186/1471-2458-7-256.
- McKinney, W. 2010 Data structures for statistical computing in python. In: *Proceedings of the Python in Science Conferences*. SciPy, Austin, Texas.
- Mounce, S. R. 2020 Data science trends and opportunities for smart water utilities. In: *ICT for Smart Water Systems: Measurement and Data Science*. Springer, pp. 1–26.
- Mounce, S. R., Gaffney, J. W., Boulton, S. & Boxall, J. B. 2015 Automated data-driven approaches to evaluating and interpreting water quality time series data from water distribution systems. *Journal of Water Resources Planning and Management* **141** (11), 37–39. doi:10.1061/(ASCE)WR.1943-5452.0000533.
- Osman, M. S., Abu-Mahfouz, A. M. & Page, P. R. 2018 A survey on data imputation techniques: water distribution system as a use case. *IEEE Access* **6**, 63279–63291. doi:10.1109/ACCESS.2018.2877269.
- Peng, J., Peng, S., Jiang, A., Wei, J., Li, C. & Tan, J. 2010 Asymmetric least squares for multiple spectra baseline correction. *Analytica Chimica Acta* **683** (1), 63–68. doi:10.1016/j.aca.2010.08.033.
- Peterson, B. M., Wanders, I., Horne, K., Collier, S., Alexander, T., Kaspi, S. & Maoz, D. 1998 On uncertainties in cross-correlation lags and the reality of wavelength-dependent continuum lags in active galactic nuclei. *Publications of the Astronomical Society of the Pacific* **110** (748), 660–670. doi:10.1086/316177.
- Schober, P., Boer, C. & Schwarte, L. A. 2018 Correlation coefficients. *Anesthesia & Analgesia* **126** (5), 1763–1768. doi:10.1213/ANE.0000000000002864.
- Speight, V. & Boxall, J. 2015 Current perspectives on disinfectant modelling. *Procedia Engineering* **119**, 434–441. doi:10.1016/j.proeng.2015.08.906.
- Talagala, P. D., Hyndman, R. J., Leigh, C., Mengersen, K. & Smith-Miles, K. 2019 A feature-based procedure for detecting technical outliers in water-quality data from in situ sensors. *Water Resources Research* **55** (11), 8547–8568. doi:10.1029/2019WR024906.
- Teh, H. Y., Kempa-Liehr, A. W. & Wang, K. I. K. 2020 Sensor data quality: a systematic review. *Journal of Big Data* **7** (1), 11. doi:10.1186/s40537-020-0285-1.
- Thayanukul, P., Kurisu, F., Kasuga, I. & Furumai, H. 2013 Evaluation of microbial regrowth potential by assimilable organic carbon in various reclaimed water and distribution systems. *Water Research* **47** (1), 225–232. doi:10.1016/j.watres.2012.09.051.
- van Buuren, S. 2018 *Flexible Imputation of Missing Data*, 2nd edn. CRC Press LLC, Utrecht.
- Vandecar, J. C. & Crosson, R. S. 1990 Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares. *Bulletin – Seismological Society of America* **80** (1), 150–169.
- Zhang, Y. & Thorburn, P. J. 2022 Handling missing data in near real-time environmental monitoring: a system and a review of selected methods. *Future Generation Computer Systems* **128**, 63–72. doi:10.1016/j.future.2021.09.033.
- Zhang, Y., Meratnia, N. & Havinga, P. 2010 Outlier detection techniques for wireless sensor networks: a survey. *IEEE Communications Surveys & Tutorials* **12** (2), 159–170. doi:10.1109/SURV.2010.021510.00088.

First received 14 December 2022; accepted in revised form 16 February 2023. Available online 1 March 2023