



University of
Sheffield



Decoding Distribution Data:

**Extracting Actionable Information from Drinking Water
Distribution System Water Quality Time Series**

Killian Gleeson

Thesis submitted for the degree of Doctor of Philosophy
(Submitted September 2023)

Department of Civil and Structural Engineering,
University of Sheffield

Academic Supervisors: Prof. Joby Boxall, Dr. Stewart Husband

Industrial Supervisor: Dr. John Gaffney

Declaration

I declare that no portion of the work contained in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or institute.

The work presented is my own except where indicated.

Abstract

Advancements in continuous in-situ water quality monitoring provide a unique opportunity to enhance our understanding of water distribution systems, ensuring safe and clean drinking water for all while mitigating contamination risks. However, this kind of monitoring is relatively new, with its true value yet to be fully demonstrated. This research developed novel analytical routines to extract actionable information from diverse real-world datasets provided by five different water service providers. A vital first step to maximising the value of these datasets was taken with the development of a data quality assessment framework, specifically tailored to address the challenge of water quality sensors' sensitivity to errors when deployed in drinking water distribution systems. Next, an investigation into the optimal analysis of in-network discolouration events using turbidity time series was conducted. This was informed by an innovative crowd-sourced event labelling exercise, integrating the perspectives of 48 domain experts and employing time series forecasting to devise a turbidity event scale. This scale effectively distinguishes between advisory (<2 NTU), alert (2-4 NTU), and alarm (>4 NTU) events, enabling reactive and proactive analysis of network events. An overarching finding of this research was the demonstration that the level of insight obtainable when moving from single parameter single sensor to multiple parameters and sensors increases in a multiplicative fashion, as evidenced by the application of developed approaches to multiple real-world examples. This research paves a clear path towards enhanced intelligent utilisation of water quality sensor networks to improve network management capabilities. This digitalisation-driven approach, in the face of increasing climate change related challenges, promises to provide the resilience required to safeguard these vital public health assets.

Acknowledgements

First of all, I would like to extend my gratitude to my two academic supervisors Prof. Joby Boxall and Dr. Stewart Husband. Without their knowledge, dedication, guidance and kindness this PhD would not have been possible nor as enjoyable as it has been. Thank you to my industrial supervisor Dr. John Gaffney for his support and valuable insights over the past 4 years. I'd also like to thank Neeraj Shah (formerly of Siemens) for his support, particularly for helping me setup a more efficient coding environment.

Many thanks to the various water service providers and sensor manufacturers who have not only provided the datasets that made this research possible, but have been willing to engage and collaborate with me. Particularly notable are James Burton and Derek Leslie, both of ATi. James for showing me the ins and outs of installing and maintaining sensors in the field and for his willingness to discuss sensor issues on the phone, and Derek for his enthusiasm in discussing and sharing his expertise in how to interpret and analyse water quality time series data. Thank you also to all participants in the event labelling exercises.

Thank you to all my fellow WIRE students, particularly Cohort 1 - I think we all have a shared bond from starting our PhD right before a global pandemic hit. A special mention for Clervie Genevois who has been a good friend and source of comic relief throughout. Thank you to my friends and family back in Dublin and in particular my parents, John and Zita, without their support I would not be here. Finally, I want to thank my wife Stephanie, for always being there for me and having faith in me even when I didn't.

Contents

Declaration	<i>i</i>
Abstract	<i>ii</i>
Acknowledgements	<i>iii</i>
Contents.....	<i>iv</i>
List of Abbreviations.....	<i>ix</i>
List of Figures	<i>xi</i>
List of Tables	<i>xvi</i>
Chapter 1: Introduction.....	<i>1</i>
1.1 Introduction to Project	<i>1</i>
1.2 Thesis Structure.....	<i>2</i>
1.3 Publications and Talks	<i>3</i>
1.3.1 Journal Papers.....	<i>3</i>
1.3.2 Conference Talks.....	<i>3</i>
Chapter 2: Background	<i>5</i>
2.1 Introduction	<i>5</i>
2.2 Water Quality in Drinking Water Distribution Systems	<i>6</i>
2.2.1 Drinking Water Quality	<i>6</i>
2.2.2 Drinking Water Distribution Systems.....	<i>7</i>
2.2.3 How Water Quality Changes in Drinking Water Distribution Systems	<i>7</i>
2.2.4 Managing Water Quality in Drinking Water Distribution Systems	<i>10</i>
2.3 Monitoring Water Quality in Drinking Water Distribution Systems.....	<i>12</i>
2.3.1 Current State of Monitoring	<i>12</i>
2.3.2 Water Quality Parameters	<i>12</i>
2.3.3 Temporal and Spatial Resolutions	<i>15</i>
2.3.4 Deploying and Maintaining Water Quality Sensors	<i>16</i>
2.4 Analysing Water Quality Time Series Data	<i>17</i>
2.4.1 Introduction to Time Series Data	<i>17</i>
2.4.2 Data Quality	<i>18</i>

2.4.2.1 Assessing Data Quality.....	19
2.4.2.2 Handling Sensor Errors	20
2.4.3 Combining Parameters and Sensor Locations	21
2.4.3.1 Understanding Relationships Between Variables.....	22
2.4.3.2 Utilising Relationships Between Variables	23
2.4.4 Time Series Forecasting	24
2.4.4.1 Traditional and Modern Approaches.....	25
2.4.4.2 Forecasting within Drinking Water Distribution Systems.....	26
2.4.5 Event Detection	27
2.4.5.1 Selecting a Detection Method	28
2.4.5.2 Event Detection in Drinking Water Distribution Systems.....	30
2.5 Summary.....	32
<i>Chapter 3: Aims and Objectives</i>	<i>34</i>
<i>Chapter 4: A Data Quality Assessment Framework for Drinking Water Distribution System Water Quality Time Series Datasets.....</i>	<i>35</i>
4.1 Abstract	35
4.2 Introduction	36
4.2.1 Background	37
4.3 Method	39
4.3.1 Multi-Sensor Data Quality Assessment Framework	39
4.3.2 Data Quality Rules.....	40
4.3.2.1 Time Stamp Errors	41
4.3.2.2 Missing Data	41
4.3.2.3 Flat Lining Data	42
4.3.2.4 Single Point Outliers	42
4.3.2.5 Extended periods above (or below) a threshold	43
4.3.2.6 Bimodal Noise.....	43
4.3.2.7 Drift.....	43
4.3.3 Linking Sensors Spatiotemporally	44
4.3.4 Multi Sensor Data Quality Validation.....	45
4.3.5 Datasets	46
4.4 Results	46
4.4.1 Data Quality Assessment Rules.....	46
4.4.1.1 User Defined Input Parameter Values	46

4.4.1.2 Time Stamp Errors and Missing Data	48
4.4.1.3 Flat Lining Data	49
4.4.1.4 Single Point Outliers	49
4.4.1.5 Extended periods above (or below) a threshold	50
4.4.1.6 Bimodal Noise.....	51
4.4.1.7 Drift.....	51
4.4.1.8 Rules applied to datasets.....	52
4.4.2 Linking Sensors Spatiotemporally	54
4.4.2.1 Cross Correlating Water Quality Sensors	54
4.4.3 Multi-sensor Data Quality Validation.....	56
4.5 Discussion	58
4.5.1 Rules, framework stage 1.....	58
4.5.2 Cross correlation, framework stage 2	60
4.5.3 Multi-sensor validation, framework stage 3.....	61
4.6 Conclusions	62
<i>Chapter 5: Algorithms to Mimic Human Interpretation of Turbidity Events from Drinking Water Distribution Systems</i>	<i>64</i>
5.1 Abstract	64
5.2 Introduction	65
5.2.1 Background	66
5.3 Method	69
5.3.1 Methodology.....	69
5.3.2 Event Labelling Exercise	69
5.3.3 Event Score Calculation.....	70
5.3.3.1 Forecasting Methods	71
5.3.3.2 Event Score and Comparison to Labels	73
5.4 Results	74
5.4.1 Labelled Results	74
5.4.2 Event Analysis Results.....	77
5.4.2.1 Flatlines.....	78
5.4.2.2 Calculating Event Scores	78
5.4.2.2.1 Forecasting Methods	80
5.4.2.2.2 Comparison to Labels.....	82
5.5 Discussion	84

5.5.1 Human Interpretation of Turbidity Events.....	85
5.5.2 Event Score Calculation.....	87
5.5.3 Proactive versus Reactive Events.....	89
5.6 Conclusions	90
<i>Chapter 6: Extracting Actionable Information from Water Quality Time Series Data: From Single-Parameter Single-Sensor to Multi-Parameter Multi-Sensor</i>	<i>91</i>
6.1 Introduction	91
6.2 Example 1 – Building from SPSS to MPSS to MPMS	93
6.2.1 SPSS.....	93
6.2.2 MPSS	96
6.2.3 SPMS	99
6.2.4 MPMS.....	101
6.3 Example 2 – Building from SPSS to SPMS to MPMS	106
6.3.1 SPSS.....	106
6.3.2 SPMS	108
6.3.3 MPMS.....	115
6.4 Example 3 - MPMS Discolouration and low chlorine at end of network	119
6.5 Example 4 - MPMS Turbidity event in a single long main	129
6.6 Example 5 – MPMS Dimensionality reduction and anomaly detection.....	134
6.7 Example 6 - MPMS Increase in daily turbidity profile.....	139
6.8 Example 7 - MPMS Coliform failures in service reservoirs.....	142
6.9 Summary.....	146
<i>Chapter 7: Discussion.....</i>	<i>149</i>
7.1 Multiplicative Value of Multiple Parameters and Multiple Sensors.....	149
7.2 From Reactive to Proactive Management of Network Contamination Events	152
7.2.1 Reactive Management	152
7.2.2 Proactive Management.....	153
7.3 Practical Considerations	155
7.3.1 Data Quality	156
7.3.2 Deployment Strategies	156

7.4 Future Research	158
7.4.1 From Source to Tap.....	158
7.4.2 Health Risk of Contamination Events.....	159
7.4.3 Role of Artificial Intelligence	160
Chapter 8: Conclusions.....	161
References.....	163

List of Abbreviations

ANN – artificial neural network

ARIMA – autoregressive integrated moving average

BED – binomial event discriminator

DWDS – drinking water distribution systems

EE – elliptical envelope

ETS – error trend seasonality

EWM – exponential weighted mean

IF – isolation forest

LOF – local outlier factor

LPCF – linear prediction correction filter

LSTM – long short-term memory

MPMS – multiple parameters multiple sensors

MPSS – multiple parameters single sensor

MVNN – multi-variate nearest neighbour

NTU – nephelometric turbidity unit

OCSVM – one-class support vector machine

ORP – oxidation reduction potential

PCA – principal component analysis

PCC – Pearson's correlation coefficient

RMSE – root mean squared error

SARIMA – seasonal autoregressive integrated moving average

SPMS – single parameter multiple sensors

SPO – single point outlier

SPSS – single parameter single sensor

SVM – support vector machine

WSP – water service providers

WTW – water treatment works

List of Figures

Figure 3.1. Analysis moving from single parameter single sensor (SPSS) to multi parameter multi sensor (MPMS)..... 34

Figure 4.1. Chapter 4 graphical abstract. 36

Figure 4.2. Multi-Sensor Data Quality Assessment Framework. 40

Figure 4.3. Two example chlorine time series (a) and corresponding cross-correlation curve (b)..... 45

Figure 4.4. Boxplots 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. 47

Figure 4.5. Time series plots of sampling intervals (blue) and accumulated time stamp errors (red). 49

Figure 4.6. Chlorine time series with flat lines of at least 8 hours highlighted. 49

Figure 4.7. (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), (d) pre and post z scores corresponding to (b)..... 50

Figure 4.8. Chlorine time series with regions staying over 1.5 mg/l, or under 0.15 mg/l, for 6 hours highlighted. 51

Figure 4.9. Turbidity time series with closeup of highlighted bimodal noise. 51

Figure 4.10. (a) turbidity time series with two periods of drift highlighted, (b) drift corrected turbidity data. 52

Figure 4.11. 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). 53

Figure 4.12. 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. 55

Figure 4.13. Heatmaps with peak PCC for each chlorine sensor pair in Datasets A (a), B (b), and C (c). Blank squares indicated insufficient data. 55

Figure 4.14. Multi-sensor data quality validation example, with (a) flagging of periods above 1.5 mg/l and below 0.15 mg/l for 6 hours in Sensor A and (B) showing absence of anomalous feature in two sensors calculated to be correlated to Sensor A. 57

Figure 4.15. Multi-sensor data quality validation example, with (a) showing instances of periods above 1.5 NTU for 6 hours and a detected period of drift in Sensor A and (b) showing detected periods above 1.5 NTU for 6 hours in Sensor B which was calculated to be highly correlated to Sensor A. 57

Figure 5.1. Chapter 5 graphical abstract. 65

Figure 5.2. Screenshot of turbidity example 4 in event labelling tool, with the theoretical event types highlighted..... 70

Figure 5.3. Boxplot of total percentage labelled data across all six example datasets per session. 75

Figure 5.4. Six turbidity time series examples, in plots (a) to (f) respectively, along with corresponding average label results (the x-axis and NTU y-axis ranges are scaled for each dataset). 76

Figure 5.5. Mean absolute turbidity value for each average label value; for Examples 1 & 2 (blue squares) and Examples 3-6 (purple stars). 77

Figure 5.6. Flatline thresholds differentiating Alert (2-4 NTU) and Alarm (>4 NTU) events in Example 1.....	78
Figure 5.7. Example 1 with expanding ARIMA forecast values (a) and Example 3 with expanding time-based average forecast values (b).....	79
Figure 5.8. Optimised sigmoid function for ARIMA tuned on all events and time-based (tb) average method tuned on advisory events (a), with corresponding outputs compared to Example 1 and 3 labels, (b) and (c) respectively.	79
Figure 5.9. Lowest RMSE for each type of forecasting methods, divided into tuning for all event types (darker shades) and advisory only events (lighter shades) and evaluated on all event types (a) and advisory events (b). Note that the y-axis is clipped from 0.1 for visual interpretation.	83
Figure 5.10. Example 2 turbidity data (clipped exclude the alarm event on day 11) and averaged-out labels versus sigmoid output using time-based averaging method, optimised for advisory events.....	84
Figure 5.11. Advisory event detection on Example 6 using a 0.2 NTU limit applied to the residual, and a 0.5 threshold using the sigmoid output (a), with equivalent average labels and sigmoid output (b).	90
Figure 6.1. Data quality rules applied to over 2 years of turbidity time series.	94
Figure 6.2. Turbidity time series containing identified event, with drift corrected version shown.	95
Figure 6.3. Turbidity event scale applied to example, with alert (>2 NTU)and alarm (>4 NTU) event identified in the turbidity time series (a), and an advisory time series shown in (b).	95
Figure 6.4. Close-up of alarm (> 4 NTU) turbidity event.	96
Figure 6.5. Data quality rules applied to turbidity (a) and chlorine (b) time series, with temperature (c), pH (d), pressure (e), and flow (f).....	97
Figure 6.6. Turbidity (a), chlorine (b), temperature (c), pH (d), pressure (e), and flow (f) before and after event.	98
Figure 6.7. Turbidity time series during alarm event (a), with pressure (b) and flow time series (c), and calculated material flux (d), where total material is labelled within the highlighted area.	99
Figure 6.8. Data quality rules applied to 11 chlorine time series, with 'ref' referring to the chlorine sensor at the location being investigated.....	100
Figure 6.9. Median correlation coefficients between reference chlorine data and 10 network deployed chlorine sensors.	101
Figure 6.10. Chlorine time series at reference and location 6 (a), sliding cross-correlation PCC (b) and sliding offset (c).....	101
Figure 6.11. Simplified schematic showing reference location (green R) and upstream location 6 (blue L6), with both downstream of a service reservoir and separated by some offshoots.	102
Figure 6.12. Turbidity time series at both locations during event on 5 th April, with the times of each peak labelled.	103
Figure 6.13. Turbidity event scale applied to the reference, with alert and alarm events shown in (a), and advisory event score in (b), and location 6, with alert and alarm events shown in (c), and advisory event score in (d).....	104
Figure 6.14. Turbidity (a), flow (b), and pressure (c) at both locations in lead up to April 5 th event.	104

Figure 6.15. Temperature (a), pH (b), and chlorine (c) at both locations, where the alarm event is indicated by the dotted red line.	105
Figure 6.16. Secondary Turbidity event responses at each location with turbidity (a), flow (b), and net material flux (c).	106
Figure 6.17. Data quality rules applied to turbidity time series.	107
Figure 6.18. Turbidity event scale applied to example turbidity, with alert and alarm events identified (a), and the advisory score time series (b).	108
Figure 6.19. Close-up of alarm turbidity event.	108
Figure 6.20. Data quality rules applied to 11 turbidity sensors.	109
Figure 6.21. Flagged data quality rules in 10 turbidity sensors from 10 th to 13 th April.	110
Figure 6.22. Data quality rules applied to turbidity data at location 10 from March to June.	111
Figure 6.23. Turbidity time series at reference and locations 2 and 7 during alarm event, with the times of each peak labelled.	111
Figure 6.24. Turbidity time series at reference and locations 2 and 7, with alarm event marked (a) and daily peak advisory scores at each location (b).	112
Figure 6.25. Data quality rules applied to 11 chlorine time series, with 'ref' referring to the chlorine at the location in question.	113
Figure 6.26. Median correlation coefficients between reference chlorine data and 10 other chlorine sensors.	113
Figure 6.27. Chlorine time series at reference and locations 2 and 7 (a), sliding cross-correlation PCC's (b) and sliding offsets (c).	114
Figure 6.28. Simplified schematic showing reference location (purple R) and upstream location 2 (green L2) and downstream location 7 (yellow L7), with all installed on different offshoots downstream of a service reservoir.	115
Figure 6.29. Mains peak daily flow from 2018 to 2020.	116
Figure 6.30. Turbidity at the three locations (a), upstream mains flow (b), flow at the three locations (c), and pressure at the three locations (d) in lead up to alarm event.	116
Figure 6.31. Temperature (a), pH (b), and chlorine (c) at all three locations, where the alarm event is indicated by the dotted red line.	117
Figure 6.32. Turbidity event at all three locations with the net turbidity in (a), flow (b), and net material flux (c).	118
Figure 6.33. Mains flow (a), turbidity at the three locations (b), and network daily discolouration contacts (c).	118
Figure 6.34. Turbidity during alarm event at three locations (a), material flux (b), and accumulated discolouration contacts per location and all (pink line and RHS y-axis) (c).	119
Figure 6.35. Simplified schematic showing sensor locations A, B, C and D and significant take-offs.	120
Figure 6.36. Data quality rules applied to turbidity and chlorine at locations A, B, C and D from February to December 2020.	120

Figure 6.37. Cross correlation heatmap between locations A, B, C and D, with median offsets shown for pairs with peak PCC above 0.7.	121
Figure 6.38. Turbidity (a), chlorine (b), temperature (c), pH (d), pressure (c) and flow (f) at locations A, B, C, and D.....	123
Figure 6.39. Turbidity at locations A, B, C and D, with alert and alarm events marked (a), and rolling daily peak advisory scores for the same four locations (b).	124
Figure 6.40. Peak daily temperature at locations A, B, C, D and air (a), and peak daily flow at A, B and C (b)...	124
Figure 6.41. Turbidity (a) and chlorine (b) at locations B, C and D, and flow at locations A, B and C (c) for April, May and June 2021.....	126
Figure 6.42. Turbidity at location B (a) and C (b) with alarm and alert events marked, and rolling daily peak advisory scores at the same locations (c).	127
Figure 6.43. Turbidity (a), flow (b), and material flux (c) at locations B and C, with total material estimated in (c) for the alarm event on April 25 th and 26 th 2021.	128
Figure 6.44. Turbidity (a), flow (b), and material flux (c) at locations B and C, with total material estimated in (c).....	129
Figure 6.45. Simplified schematic show locations 1-6 in a single straight mains.	130
Figure 6.46. Turbidity time series from July to end of September for locations 1 (a), 2 (b), 3 (c), 4(d), 5 (e), and 6 (f).	130
Figure 6.47. Median daily standard deviation for locations 1-6.....	131
Figure 6.48. Median peak daily advisory score (LHS y-axis and yellow bars) and number of alert and alarm events (orange and red bars, respectively and RHS y-axis)	131
Figure 6.49. Upstream mains flow (a) and turbidity time series at each location (b) between 15 th and 23 rd August.....	132
Figure 6.50. Turbidity (a), flow (b) and material flux (c) during event first wave at each location 1-6.	133
Figure 6.51. Turbidity (a), flow (b) and material flux (c) during event second wave at locations 5 and 6.....	133
Figure 6.52. Total material passing locations 2-6 during the event first and second waves.	134
Figure 6.53. Scatter plots of two principal components for location H (a) and I (b), with boundary lines shown for unsupervised anomaly detection methods isolation forest, elliptic envelope, local outlier factor and one-class SVM.	135
Figure 6.54. Two connected eight-parameter water quality time series from March to October s with turbidity (a), chlorine (b), temperature (c), pH (d), pressure (f), flow (g), conductivity (h), ORP (i) and two principal components in (e) and (j) respectively.	136
Figure 6.55. Detected anomaly at Location H shown at both locations for turbidity (a) and ORP (b).	137
Figure 6.56. Detected anomaly at location H shown in both locations for turbidity.	137
Figure 6.57. Detected anomaly at Location I shown at both locations for turbidity (a) and ORP (b).	138
Figure 6.58. Multiple detected anomalies at location I, plotted on parameters turbidity (a), pH (b), pressure (c) and ORP (d) with location H also plotted.....	138

Figure 6.59. Simplified schematic showing sensors V, W, X, Y and Z (line lengths are not representative of distances)..... 140

Figure 6.60. Mains peak daily flow from 2019 to 2021. 140

Figure 6.61. Turbidity at five locations for the first three days of August (a) and throughout August (b). The bottom plot shows the peak daily advisory scores at each location in August (c)..... 141

Figure 6.62. Net daily material (a) and total material passing each location (b) in August 2021. 142

Figure 6.63. Average weekly Log WQRS (a), total chlorine (b), temperature (c) and total coliforms (d) across over 300 service reservoirs for over 4 years..... 143

Figure 6.64. Scatter plot of total chlorine versus temperature, where blue circles represent all samples, orange squares are samples with a WQRS above 9 and black stars are samples with coliform failures (a), and a histogram showing the distribution of WQRS for samples with coliform failures (b)..... 144

Figure 6.65. Log WQRS (a), free chlorine (b), water and air temperature (c), rainfall (d), and total coliforms (e) at a single service reservoir for over 2 years..... 145

Figure 6.66. Log WQRS (a), free chlorine (b), water and air temperature (c), rainfall (d), and total coliforms (e) at a single service reservoir for over 2 years..... 145

List of Tables

<i>Table 2.1. Types and examples of water quality deteriorations within DWDS.</i>	9
<i>Table 4.1. Data Quality Rules: detection methods and possible causes.</i>	41
<i>Table 4.2. Datasets used.</i>	46
<i>Table 4.3. Data Quality Rules: Input Parameter Values.</i>	48
<i>Table 5.1. Forecasting methods and associated adjusted parameters.</i>	72
<i>Table 5.2. Event Labelling Sessions.</i>	74
<i>Table 5.3. Turbidity Event Scale.</i>	77
<i>Table 5.4. Optimal parameters found for averaging and time-based averaging.</i>	81
<i>Table 5.5. Optimal parameters found for ARIMA and SARIMA.</i>	81
<i>Table 5.6. Optimal parameters found for EWM and ETS.</i>	81
<i>Table 5.7. Optimal parameters found for Prophet.</i>	82
<i>Table 5.8. Optimal parameters for LPCF using sigmoid function and BED.</i>	82
<i>Table 6.1. Analysis possibilities moving from SPSS to MPMS.</i>	92
<i>Table 6.2. Summary of examples presented.</i>	147

Chapter 1: Introduction

1.1 Introduction to Project

Decoding Distribution Data is a University of Sheffield PhD project 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. This project has been a collaboration with multiple UK water service providers (WSP), who have provided time series datasets from water quality sensors continuously deployed along drinking water distribution systems (DWDS). It is relevant to note that these datasets were not collected bespoke for this project, they were from WSPs' own internal deployment projects, though insights obtained from applying the methods developed in this research influenced deployment strategy and improved the quality of data taken in some cases. The research in this project has focused on developing analytic routines to extract information from the datasets that could inform improved network management. Particular attention has been given to the two most commonly measured parameters in these datasets: turbidity, a measure of cloudiness in water, and chlorine, which is relied upon for prolonged disinfection. In exploring the use of continuously monitoring water quality sensors (generally taking measurements at intervals between 1 and 15 minutes), this research represents a move away from the status-quo when it comes to monitoring water quality post treatment works. This has traditionally comprised only of infrequent grab sampling, which is insufficient for such a dynamic complex environment that entire societies continuously rely on for their health and wellbeing. It also represents a move away from the reliance on customers to inform of drinking water quality incidents to a more proactive digitalised form of operating these vital assets and ensuring they perform as intended, ensuring the delivery of the high quality treated drinking water safely to customer taps. The existence of this collaborative project and the availability of the necessary datasets demonstrates how widespread such monitoring is becoming, but also suggests that WSP have yet to fully reap the rewards. This is partially due to such monitoring being a relatively new pursuit, meaning the necessary datasets for developing analytical routines have only been recently available. Therefore, for such monitoring to continue to expand in its use and become a more widespread practice for managing these assets, it is vital to develop

analytical methods to translate the raw data into useful actionable information capable of informing and improving DWDS management.

1.2 Thesis Structure

This thesis consists of 8 chapters, with the significant research contributions detailed in **Chapters 4-6**. **Chapter 2** provides the background and motivation behind water quality monitoring within DWDS. An overview of drinking water quality within DWDS is included, including how it can change on its journey from treatment to tap, and common approaches to managing these assets. The next section discusses the practice of deploying water quality sensors within DWDS, with a focus on what parameters are best suited and different deployment strategies. The final section discusses methods for analysing water quality time series data, with a focus on data quality assessment, combining parameters and sensors, time series forecasting, and event detection. **Chapter 3** presents the aims and objectives for this research, which were driven by the knowledge gaps identified in **Chapter 2**.

Chapters 4 and 5 are from manuscripts prepared for publication in peer review journals, that address the first two objectives listed in **Chapter 3**. **Chapter 4** describes a data quality assessment framework for turbidity and chlorine sensors, including a cross-correlation method that determines connectivity between sensor locations, with this spatiotemporal information finally used to cross-validate any flagged data quality issues. **Chapter 5** is focused entirely on turbidity events, and features a unique crowd-sourcing approach that asked domain experts their opinions on events in turbidity time series, before developing algorithms that were able to best mimic such human interpretation. **Chapter 6** demonstrates how actionable information can be derived from DWDS water quality time series datasets and explores the impact of different combinations of parameters and sensors. In order to perform a robust examination and review multiple examples, **Chapter 6** is a longer-form chapter than the previous manuscript-type chapters. **Chapter 7** is a discussion that connects and builds upon the previous 3 chapters, adding depth and significance to the overall work, reflecting on the contributions made with respect to prior research reviewed in **Chapter 2**, as well as providing recommendations for operators and future research. The conclusions of this research project are stated in **Chapter 8**.

1.3 Publications and Talks

This research has produced a number of publications and conference talks that are presented in this section.

1.3.1 Journal Papers

Gleeson, Husband, Gaffney, Boxall. “A data quality assessment framework for drinking water distribution system water quality time series datasets.” *IWA Aqua: Water Infrastructure, Ecosystems and Society* (2023). doi:10.2166/aqua.2023.228 (Chapter 4 is a reproduction of this paper)

Gleeson, Husband, Gaffney, Boxall. “Algorithms to Mimic Human Interpretation of Turbidity Events from Drinking Water Distribution Systems.” *IWA Journal of Hydroinformatics* (Accepted for publication, September 2023). (Chapter 5 is a reproduction of this paper)

1.3.2 Conference Talks

Gleeson, Husband, Gaffney, Boxall. “Automated Data Quality Assurance for Water Quality Sensors in Drinking Water Distribution Systems.” *AWWA Water Quality Technology Conference* (Washington, USA, 2021). (Based on work from Chapter 4)

Gleeson, Husband, Gaffney, Boxall. “Determining the spatiotemporal relationship between water quality monitors in drinking water systems.” *IWA Hydroinformatics Conference* (Bucharest, Romania, 2022). Resulting conference paper was published in *IOP Conference Series: Earth and Environmental Science* 1136 (1), 012046. doi:10.1088/1755-1315/1136/1/012046. (Based on work from Chapter 4)

Gleeson, Husband, Gaffney, Boxall. “Linking Water Quality Sensors in Distribution Systems.” *Water Wastewater and Environmental Monitoring* (Telford, UK, 2022). Finished second in the Sensor for Water Interest Group (SWIG) Early Career Innovation Poster Prize. (Based on work from Chapter 4)

Gleeson, Husband, Gaffney, Boxall. “Leveraging Water Quality Data to Support Intelligent Decision Making.” *WWT Smart Water* (Birmingham, UK, 2023). (Based on work from Chapter 5)

Gleeson K, Husband S, Gaffney J, Boxall J. “Root-cause Analysis of Discolouration Events in Drinking Water Distribution Systems using Time Series Data.” *Computing and Control in the Water Industry* (Leicester, UK, 2023). (Based on work from Chapters 5 & 6)

Chapter 2: Background

2.1 Introduction

DWDS are extensive buried engineered networks that function to provide a continuous source of high-quality drinking water to entire populations. Drinking water is generally treated to high standards; in the UK 99.97% of tested drinking water complied with the Water Supply (Water Quality) Regulations in 2021 (DWI 2022). However, most water quality monitoring is done as the water leaves the water treatment works (WTW), which is not representative of the water that eventually reaches consumers via DWDS. It would be optimistic to expect even the most pristine DWDS to continuously transport treated water from treatment to tap without any risk of deterioration occurring along the way. However, these assets are buried, often many decades old and consist of sections with burst or leaking pipes, which can allow contaminants to enter. There are therefore many complex and poorly understood chemical, biological, and physical processes that occur within DWDS that can lead to water quality deterioration and ultimately pose a risk to public health. Due to the vital importance of DWDS to public health and well-being, it is worth investigating the use of more sophisticated and widespread monitoring methods, to ensure the quality of transported drinking water.

This project is focused on deriving insight from continuously deployed water quality sensors along DWDS. In this context, the term continuous refers to the sensors taking measurements at regular intervals, generally between 1 and 15 minutes, as opposed to discrete grab sampling. In recent years WSP have started deploying continuous water quality monitors along DWDS, a practice driven by regulatory pressures under the Drinking Water Inspectorate (DWI 2023) to move from reactive to proactive asset management and to better understand and manage these complex systems. However, little work has been done to understand how actionable information can be derived from this newly available time series sensor data. Without knowing the potential value, WSP can't perform a cost-benefit analysis regarding increased investment in monitoring campaigns. This Chapter provides a background firstly on drinking water quality in DWDS (**Section 2.2**), including how it can change and how it can be managed by WSP. The subsequent **Section 2.3** focuses on monitoring water quality within DWDS, looking at both traditional grab sampling and newer continuous online sensing. The final **Section 2.4** reviews methods of analysing the resulting

water quality time series data, with focuses on sensor data quality, combinations of different parameters and sensors, time series forecasting and event detection. A summary of the main research gaps identified is provided in **Section 2.5**, which was used to inform the overall project aims and objectives in **Chapter 3**.

2.2 Water Quality in Drinking Water Distribution Systems

2.2.1 Drinking Water Quality

Johnson et al. (1997) defined water quality as:

“a measure of the condition of water relative to the requirements of one or more species and/or to any human need or purpose”.

Water quality is therefore a subjective term that depends on who is using it and for what purpose. Since this work is focused on drinking water, water quality refers to water’s suitability and safety for human consumption. As a matter of survival, humans have been acutely aware of the importance of drinking water quality throughout history, and have relied on senses of smell, vision and taste to determine its suitability for consumption (Angelakis and Mays 2014). However, water can contain many different pathogens and contaminants undetectable to human senses but potentially leading to serious illnesses or death.

Waterborne disease was not well understood until the 19th century. After John Snow linked a cholera outbreak to a water supply in England, Louis Pasteur published his “germ theory” which recognised the threat posed by bacteria in water (Pasteur 1861). These findings initiated efforts to pipe water from safer sources and eventually lead to the development of WTW, where the objective was to monitor and remove dangerous contaminants. In the middle of the 20th century, the long-term toxic effects of chemicals found in DWDS, such as lead, were also better understood (Gray 2008). This resulted in the development of drinking water quality standards in many parts of the world, dictating concentration limits for various chemicals and pathogens (EU 1998; WHO 2017; DWI 2018; USEPA 2018).

The palatability of water to human senses also remains an instinctive factor for consumers. Tap water could be entirely safe for human consumption but if it is unappetising it may be avoided. Since there are health benefits associated with drinking water as opposed to sugary soft drinks or alcohol, and environmental benefits from avoiding bottled water, the

palatability of water supplied by a DWDS is important, and is impacted by appearance, odour and taste.

2.2.2 Drinking Water Distribution Systems

Throughout human history, societies have depended on availability of drinking water, a critical factor in determining the location of cities. DWDS have evolved over several thousand years, with sophisticated pipe water distribution networks developed by ancient civilisations in places such as Pakistan, Greece (which was later built upon by the Roman Empire) and Mesopotamia as early as 2600 BC (Angelakis and Mays 2014). Most of the advances in hydrology, bacteriology, public hygiene and water treatment, that lead to the development of modern day treated and distributed drinking water, came in the past 2-3 centuries, after the Renaissance. Such advances have improved public health. Life expectancy in the United States rose from 47 to 63 between 1900 and 1940, with access to safe drinking water attributed for nearly half the overall mortality rate drop, and three-quarters of the infant mortality reduction (Cutler and Miller 2004).

Modern day DWDS deliver enormous volumes of water to industrial and private users through vast underground pressurised piped networks, broken up by service reservoirs, pumping stations and valves. Many are ageing and have deteriorated over many decades. They have not been proactively maintained and sections have been repaired or replaced upon failure, leading to huge variations in pipe material, age and dimensions. How DWDS will evolve in the future is difficult to predict due to potentially unforeseen political, social, economic and environmental changes. However, it can be expected they will come under increased strain due to increasing urban populations and the devastating impacts of climate change. These impacts are already being seen, with water quantity, water quality, and water-related extremes identified as three major push factors influencing several million migrants and similar conditions may affect more than half the global population by 2050 (UN-INWEH 2020). Therefore, it is crucial for DWDS to be resilient and responsive to threats with increased monitoring an essential part of achieving this.

2.2.3 How Water Quality Changes in Drinking Water Distribution Systems

The long journey treated drinking water makes from treatment to tap comes with a risk of contamination and/or deterioration. This can range from the introduction of deadly pathogens

to subtler degradations impacting the aesthetic quality of the final drinking water received by customers. Such changes can occur quickly and unpredictably and the buried nature of DWDS, combined with its ability to quickly spread contaminants to large populations, meaning the results can be catastrophic. In the UK, 36% of waterborne disease outbreaks from 1911 to 1995 have been attributed to issues within the DWDS (Ainsworth 2013). Since most DWDS were built many decades ago, many of the materials used in their construction are now thought to have negative health impacts, such as coal-tar linings, asbestos and lead (Gray 2008). Coal-tar linings, used to protect ductile iron, are known to add carcinogenic polycyclic aromatic hydrocarbons (PAHs) into drinking water. Asbestos has been found in drinking water in the UK and, though the effects of consumption through drinking water are unclear, there has been a sharp increase since the 90's of mesothelioma, a lung cancer known to be caused by inhaling asbestos, in people who have not inhaled asbestos (Gray 2008) and the WHO currently warns that latest evidence suggests ingesting asbestos through drinking water increases risk of getting cancer (WHO 2021). Aside from the presence of harmful materials embedded within DWDS infrastructure, the journey water takes from treatment to tap offers additional avenues for water quality to deteriorate.

Some common causes of water deterioration within DWDS are summarised in Table 2.1. These processes are all interlinked and all impacted by the state of the pipework, which is highly variable and often unknown, most obviously in the case of bursts and leaks allowing contaminants to enter the DWDS (Fox et al. 2016). It is difficult to determine how commonly this occurs though one study found a positive correlation between broken pipes and internet searches for gastrointestinal illnesses (Shortridge and Guikema 2014). The way water interacts with pipe surfaces plays a major role in water quality deteriorations. In the UK, the most common water quality related customer complaint is related to discolouration (DWI 2022), which is often caused by increased daily flow rates mobilising material from pipe walls. There are two main theories behind how this material accumulates: gravitational settling of larger particles during low flow rates or the formation of cohesive layers on pipe-walls (Boxall et al. 2023). Regardless of accumulation process, the subsequent mobilisation of pipe-wall material is driven by increased hydraulics. UK WSP face regulatory fines and/or monitoring instructions based on discolouration contacts (DWI 2023), making it a priority to reduce their occurrence. Discoloured water is often safe to consume but indicates an increased risk of the presence of suspended solids such as iron or manganese (Boxall and Saul 2005) and consumption of water with increased turbidity has been associated with an

increased risk of gastrointestinal illness (Mann et al. 2007). Even when not posing a health risk, discoloured water could be unappetising to customers and reduce trust in the utility. As well as having links to flow rate increases, discolouration is also more likely at higher air temperatures (van Summeren et al. 2015). Additionally, the microbiology of pipe wall biofilms have been shown to have a stronger impact on the bacteria present in tap water than water source and treatment (El-Chakhtoura et al. 2018), further demonstrating the strength of impact pipe-wall material has on final drinking water quality.

Table 2.1. Types and examples of water quality deteriorations within DWDS.

Type	Example
Interaction with pipe surface	Material accumulating along pipe wall and subsequently becoming mobilised through increased hydraulics (Husband et al. 2008).
Bulk water transformation	Excessive disinfectant residual decay, often overlooked due to simplistic modelling (Speight and Boxall 2015), leaving sections unprotected against contamination.
Infrastructure failure	Damaged and leaking pipes allow contaminants to enter DWDS from surrounding soil (LeChevallier et al. 2003).

Most countries rely on a disinfectant residual, most commonly chlorine, to provide lasting protection against contamination at a relatively low cost (McGuire 2006). Consuming excessive levels of chlorine is bad for human health and can cause dangerous disinfection by-products so careful dosing is required to ensure sufficient but not excessive residuals reaches the farthest points in the network, with additional dosing points common. Predicting chlorine decay is not straightforward, particularly due to complex interactions with pipe wall biofilms (Speight and Boxall 2015). Badly deteriorated pipework has a serious impact, with rusted pipes found to rapidly increase disinfection decay (Savane et al. 2019). Such an interaction with organic matter can lead to the formation of disinfection by-products (Sadiq and Rodriguez 2004), posing a chemical risk to human health and adding complexity to the decision of whether to rely on disinfection to protect against contamination. The risk of any form of deterioration increases for remote properties towards the end of DWDS, both due to more opportunities for contamination and the difficulty in maintaining a sufficient disinfection residual. In summary, many avenues exist that can lead to drinking water quality

worsening as it travels from treatment to tap but, without sufficient monitoring in place, the true scale is not well understood.

2.2.4 Managing Water Quality in Drinking Water Distribution Systems

How the system is managed plays a significant role in how likely water quality deteriorations are (Weston et al. 2022). A number of practices have been developed by water utilities to prevent and limit water quality deterioration within DWDS, including routine repairs, planned responses to deterioration events, online monitoring, primary and secondary chlorination dosing, and network cleaning operations to manage material accumulation along pipe walls (Speight et al. 2020). However, major changes to DWDS, such as switching water sources or changing disinfection strategy, can have disastrous results on water quality if the risks are not properly understood (Liu et al. 2017). Though online monitoring is sometimes used as a preventative method in itself, it is also an important part of any attempt to manage DWDS deterioration as it enables operators to examine the impacts of any intervention. Service reservoirs have commonly been assumed to be the source for many DWDS deterioration events, leading to many utilities installing monitors at their outlet. However, assuming any detected events could only be from the service reservoir ignored the entire upstream section of DWDS, meaning inlet and outlet monitoring at service reservoirs is required to paint a full picture (Doronina et al. 2020).

As already covered, DWDS tend to have sections that are badly ageing and in need of repair or replacement. Pipe replacement is however very disruptive particularly in dense urban environments. These repair activities can also be disruptive to water quality, as they risk external contaminants entering the system and without careful planning could cause major hydraulic disturbances leading to mobilisation of pipe wall material. How a utility responds to a water quality deterioration incident and how much of the risk posed to customers is minimised is a factor in how well-equipped utilities are for managing water quality in DWDS. Without online water quality monitoring within DWDS, the majority of decisions utilities take regarding issues such as discolouration are dealt with in a reactive and unscientific manner with utilities often learning of discoloured water solely through customer contacts (Cook et al. 2016; Boxall et al. 2023). The actions the utility takes in response to such an incident is vital, though they are hampered by being forced to respond to something that has already reached and impacted customers.

Once a network section is identified as requiring cleaning, a number of options are available. The most intensive, expensive and disruptive option is to shut down a section for mechanical cleaning, while the cheapest and least disruptive options is to perform flushing or conditioning exercises (Friedman et al. 2012). Traditionally, mechanical cleaning has been widely used though flushing has become more commonplace, as well as ice pigging and swabbing. Ice pigging involves forcing ice through a water main to remove bulk water sediment, while swabbing is similar but uses a sponge like object and effectively scrapes the pipe walls. However, such operations risk causing disturbances that increase the risk of discolouration in other network areas (Husband et al. 2010). Flushing or conditioning activities involve carefully planned increased hydraulics such that pipe wall material is mobilised in a controlled manner. Flushing was originally done in response to customer complaints but more recently has been employed on a more regular basis (Boxall et al. 2011), with the idea of reducing risk of a later sudden uncontrolled mobilisation of large quantities of pipe wall material which could potentially contain dangerous bacteria.

Though conditioning activities have been shown to reduce overall discolouration risk (Sunny et al. 2019), utilities are often nervous to attempt them due to the perceived potential risk to customers if done poorly. A Variable Condition Discolouration Model (VCDM) was developed to simulate turbidity responses to hydraulic interventions (Furnass et al. 2019), enabling utilities to plan their conditioning practices so that a certain level of turbidity is not exceeded and reducing any potential risk to customers. VCDM was developed using the concept of material accumulating on pipe walls in cohesive layers conditioned by the daily peak flow rates (Husband and Boxall 2016). Flow conditioning gives utilities a way to proactively manage discolouration, minimising more disruptive reactive approaches that involve taking sections of DWDS out of service. One of the main downsides is its excessive use of water, which may not always be appropriate particularly during droughts, which are becoming a more common factor. Monitoring both flow rate and turbidity is an essential part of any flow conditioning exercise, and having continuous monitors before and after allows for the impact to be determined.

Where disinfection is used, management of residual disinfection levels throughout DWDS is a factor. However, chlorine residual is unpredictable and challenging to accurately model due to the impact of local characteristics like pipe wall chlorine demand. This makes ensuring a sufficient residual at all sections very difficult and often secondary disinfection dosing points

will be installed. Such dosing points can also be reliant on a chlorine sensor to inform when dosing is required. The accuracy of this is therefore dependent on how well calibrated the chlorine sensor is. Though disinfection is relied upon in most countries, including the UK, recent research showed that in certain environments a disinfection residual can actually promote the colonisation of pathogens in pipe-wall biofilm as well as increased discolouration (Fish et al. 2020), suggesting more research is needed to better understand the overall impacts of disinfecting DWDS.

2.3 Monitoring Water Quality in Drinking Water Distribution Systems

2.3.1 Current State of Monitoring

Drinking water quality guidelines have been developed by many international governing bodies (EU 1998; WHO 2017; DWI 2018; USEPA 2018) but the majority of samples are taken as the water exits the WTW. There is a lack of international consistency with regards to equipment, parameters measured and sampling protocols, making any comparisons within and between countries challenging. Regulatory post WTW monitoring tends to consist of periodic grab sampling, with WSP often becoming aware of in-network deterioration through customer contacts meaning any interventions are by definition reactive (Mounce 2020). Periodic grab samples post-WTW only offer a snap shot into these 24/7 complex dynamic and unpredictable DWDS that provide such a vital public health and wellbeing service. With advances in remote data acquisition and communication technologies, it is now possible to monitor water quality at greater temporal and spatial resolutions, yet the benefits of such monitoring remains largely unknown. In recent years, such technology is being deployed and experimented with and it has been expected to become more commonplace in future DWDS (Boxall et al. 2011). However, this remains a new practice and work needs to be done to understand what parameters are most useful and how sensor networks can be deployed and maintain as to ensure high quality data is taken. This subsection will review strategic options for continuous deployment of water quality sensor networks within DWDS.

2.3.2 Water Quality Parameters

There are a variety of different parameters that can be used to describe drinking water quality, ranging from the direct detection of specific contaminants or pathogens, to measurements of indirect proxy parameters. Water quality parameters can loosely be categorised as microbial

and non-microbial. In an ideal world, many of these could be combined in order to build up a complete and thorough understanding of the water quality. However, not all parameters are equally suited for continuous remote measurement. Therefore, this work considers which parameters are both suited to DWDS deployment and can lead to the best understanding of water quality within DWDS.

Awareness and restriction of microbial pathogens in drinking water has reduced transmission of disease in the DWDS of many countries. Disease carrying pathogens in drinking water include protozoans (5-100 μm), bacteria (0.5-1 μm) and viruses (0.01-0.1 μm) (Gray 2008). Total coliform bacteria is a useful parameter for indicating the presence of faecal contamination without having to monitor each contaminant. Their ease of detection has led to them becoming standardised. However, many coliforms are not pathogens and they can form part of the pipe wall's biofilm (Camper et al. 1998), meaning lack of detection does not mean zero risk of future contamination while detection does not always infer a health risk. *E. coli* is a more reliable pathogenic indicator as it is found in human and animal faeces but, like coliforms, is more sensitive to disinfection than certain harmful pathogens (Payment et al. 2003). While a positive coliform or *e. coli* test requires immediate attention, negative tests provide a false sense of security. Dissolved oxygen (DO) is a useful parameter for indicating how suitable an environment the water is for certain living organisms. Oxidation reduction potential (ORP) offers a wider scale than DO as it can monitor anaerobic conditions as well as aerobic and can indicate the level of disinfection in the water. Heterotrophic plate counts are often used to count microbial cells but can only detect certain microorganisms (Boe-Hansen et al. 2002). Flow cytometry has emerged as a popular sampling method that produces a more complete microbial profile in just 15-20 minutes, compared to over a day for more traditional methods (Berney et al. 2008). These microbial indicator parameters tend to be popular due to their inexpensive testing but cannot be easily automated and require well-trained laboratory technicians though flow cytometry is more suited to online monitoring.

Physical and chemical parameters tend to more suited to automatic sampling but their link to water quality can be complicated. Turbidity, for example, is an optical measure of how clear the water is and has been shown to be well-suited for measuring discolouration material (Boxall and Saul 2005), making it an important parameter due to the pervasiveness of discolouration incidents. Where relied upon, monitoring of the disinfection residual within DWDS can inform about water quality deterioration. Chlorine decay rates are challenging to

accurately model (Speight and Boxall 2015), meaning monitoring is required to ensure sufficient but not excessive residuals are present. These decay rates can be suddenly and unexpectedly accelerated due to increased demand from sources such as bacteria, natural organic matter (NOM), and the pipe-wall biofilm. Therefore, monitoring the disinfection residual can inform about water quality changes and risk of contamination. Additionally, research looking into intentional contaminants in DWDS found that free chlorine responded to 10 of 15 contaminants tested (Murray and Haxton 2010), indicating its use as a proxy contamination indicator. Temperature is a useful parameter as any in-network reaction is influenced by water temperature. Water held at elevated temperatures increases the rate of disinfection decay and can lead to infection via deadly pathogens such as legionella, with the WHO (2017) recommending temperatures be kept below 25°C and ideally below 20°C. Water temperature within DWDS is determined more by the surrounding conditions and pipe diameter than source water temperature (Blokker and Pieterse-Quirijns 2013; Douterelo et al. 2019). DWDS in urban environments competes for space with heating pipes, underground infrastructure and electric cables, all of which can heat sections leading to localised hot spots that can pose a risk to end users.

Other useful water quality parameters are conductivity, which indicative the level of ionic salts, including toxic salts (Banna et al. 2014), and potential of hydrogen or pH, which is a measure of water acidity and usually remains within its ideal range of 6.5 and 8.5 within the DWDS (Payment et al. 2003) though the impact of smaller deviations are not well understood (Fish et al. 2016). DWDS pipe characteristics such as pipe material, size, surface roughness and level of deterioration can all have dramatic localised impact on water quality and should be considered where available. Water age refers to how long the water has spent inside the DWDS and is related to water quality due to the increased opportunities for contamination (Machell and Boxall 2014). However, water age is not a measurable parameter. Hydraulic parameters, such as flow and pressure, are also relevant to understanding water quality due to the tendency of hydraulic changes to initiate water quality deterioration events, as was highlighted in Section 2.2.3. Flow and pressure sensors are more widely deployed in DWDS than water quality parameters as part of efforts to tackle leakage rates, meaning they may be available to supplement any deployed water quality sensors. In general, it is not well understood what parameters are best suited to online monitoring, or indeed what combination of parameters can yield the greatest level of insight regarding the state of DWDS water quality.

2.3.3 Temporal and Spatial Resolutions

Selecting the temporal and spatial resolution of a sensor network is a trade-off between cost and the value of increased quantities and densities of data. WSP in the UK have traditionally used 15-minute sampling intervals for flow rate and pressure monitors (S. R. Mounce et al. 2012), and this practice has been adopted when deploying water quality sensors. However, there are clear benefits of increased sampling frequencies of up to 1 minute, particularly when you consider that these are instantaneous measurements of a remarkably dynamic and unpredictable subject matter (Gaffney and Boulton 2012). The clearest trade-off with an increased sampling rate is that these remote installations rely on a battery, which will run out quicker with increased sampling. However, dial out rate, the frequency with which the sensor uploads data to a server, is a more significant factor for battery life than sampling rate (ATi 2022). Renewable energy sources represent an alternative approach for powering water quality sensors and have been used in other remote environmental monitoring applications (Dewan et al. 2014).

When considering the optimal spatial resolution of a water quality sensor network, the number of suitable installation locations is of course a major factor as is the ultimate objective of the sensor deployment project. Research on sensor placement within DWDS has produced algorithms that are capable of optimising sensor placement strategies for a given objective, such as quickest detection time or protection of most members of the public (Krause et al. 2008). An open-source Python package called Chama was recently released that uses mixed-integer, stochastic programming methods to determine optimal sensor placement (Klise et al. 2017). However, much of this work requires the number of sensors to be already selected, and it is also focused on simple detection of specific contaminants. Sankary and Ostfeld (2018) showed that non-specific water quality parameters such as turbidity, chlorine and pH could be used to detect intentional contamination and sensor placement could be optimised based on simulated responses. Such research is interested in multiple parameters at each location and in combining information from multiple sensor locations, which distance between locations may dictate. Being able to understand how sensor locations are connected could be vital in enabling a move from single sensor analysis to analysing DWDS network sections as connected entities. Therefore, this research will explore sensor connectivity between water quality sensors installed within DWDS, including what spatial densities facilitates such an approach.

2.3.4 Deploying and Maintaining Water Quality Sensors

Once the quantity and spatial density of sensors is chosen, consideration is needed for ensuring the sensors perform to take representative data. Typically a maintenance plan is required, which involves visiting installation sites and validating sensor performance. How frequently this is needed depends on the desired data quality, the parameters monitored, the deployment length and the installation site itself. Water quality sensors monitoring parameters such as turbidity and free chlorine were traditionally designed for laboratory use and can become fouled and damaged when installed remotely in the harsh DWDS environment. Turbidity sensors are optical and their optics can get fouled or damaged during deployment, often resulting in data drift (Mounce et al. 2015). Turbidity drift is a common problem and, though there is evidence that drift corrected data is still useful, after a certain point the sensor's optical lens will need to be cleaned. The problem of drift in turbidity sensors has led to the use of the daily standard deviation as a metric for assessing discolouration risk that, unlike averaging, is unaffected by drift (Cook et al. 2016). Chlorine sensors tend to rely on a membrane electrolyte which loses sensitivity over time and therefore require regular re-calibration (Garcia et al. 2020).

Selecting a suitable installation location for water quality monitoring is difficult primarily due to the buried nature of DWDS. Additionally, much DWDS infrastructure was built many decades ago and the precise location and geometry of entire sections can be unknown. Installing directly into pipes would in theory lead to the most representative sampling but without custom built sensor chambers, there are very few installation locations that provide this opportunity. Sensors are commonly installed at hydrants (fire, wash out etc), which are more convenient to access than the buried pipework, though it is unclear how representative this water is of the water in the actual pipes. By the time the water reaches a hydrant, the water has slowed down considerably, with potential impacts on water quality. Some installations include a purge mechanism which flushes out any stagnant water before measurements are taken to account for this (Gaffney and Boulton 2012). The ease of access for an installation point not only dictates difficulty of install, but also serves to make subsequent sensor validation and maintenance operations more challenging and costly.

2.4 Analysing Water Quality Time Series Data

The main analysis done to water quality data from DWDS is related to meeting regulatory compliance using discrete sampling data (Boxall et al. 2011). Continuously monitored water quality data, resulting in time series datasets, are becoming more commonplace particularly in response to a particular network issue or event. However, potential analytical approaches will be entirely different for such higher-frequency time series datasets. Therefore, a first general introduction into time series data is provided, to highlight the value in continuous timestamped monitoring. As outlined already, remotely deployed water quality sensors can be prone to erroneous measurements, as opposed to regulatory grab sampling which are typically done in a controlled laboratory environment. Therefore, the next section reviews data quality assessment approaches, along with methods for handling sensor errors. Next, the benefits of understanding the relationships between time series from different water quality parameters and sensor locations is discussed. The final two sections cover two common time series analysis tasks: time series forecasting and event detection, both of which are highly desired in the context of DWDS.

2.4.1 Introduction to Time Series Data

Sensors are increasingly linking the physical and digital worlds through vast amounts of time series data and are driving a new industrial revolution focused on autonomous systems (Vaidya et al. 2018). Time series data refers to datasets ordered chronologically with each entry usually accompanied by a timestamp. This provides powerful temporal information that allows for parameters and cross parameter relationships to be analysed over time, which is vital for forecasting future values and for understanding interactions in complex systems and control engineering (Box et al. 2015). Time series data is found in increasing volumes in many applications such as economics, epidemiology, network monitoring, social science, medicine, and engineering (Shumway and Stoffer 2017) though the nature of the time series data can vary enormously depending on the data sources. Important attributes of any time series dataset include: the number of parameters, sample rates, availability of labels, availability of metadata, nature of the data, and data quality. These factors impact what analytic approach can be taken, which can range from simple statistical descriptors, applicable to univariate time series data, to supervised machine learning algorithms, applicable to multivariate labelled datasets.

Before sophisticated analytics can be done on any time series dataset, there is usually some primary reviewing of the dataset, in order to gain a better understanding of the nature of the data. This practice is called exploratory data analysis (EDA), a term first coined by John Tukey (Tukey 1977), and involves visualising the data in various different graphical representations and producing simple statistical descriptions, in order to understand the underlying data better. This stage will often include an assessment of the data quality, which would inform any data cleaning processes necessary to prepare the dataset for subsequent analysis. Data quality assessment is seen as a vital first step in analysing any continuously deployed water quality sensor data, both in terms of preparing the dataset for further analysis and assessing sensor performance while deployed. There exist many potential manipulations of raw datasets, a process often called feature engineering, such as normalisation, particularly important for certain machine learning algorithms, reformatting data, replacement of missing or removed data points, or generating custom variables (Reid Turner et al. 1999). Often it is desirable to understand how each variable is correlated, as such information could help inform feature engineering operations. In a sensor network, such correlations are important in differentiating local and global events. Once a time series dataset is prepared and validated, further analytics can be done with confidence. Two main topics of interest are forecasting future values, and the detection of unusual events, often termed outliers or anomalies. Understanding the variables in time series datasets is often an important goal of analytics, with other common goals focused on forecasting future values, and detecting. This section reviews data quality, relating time series variables, forecasting and event detection, with respect to water quality time series datasets.

2.4.2 Data Quality

Data quality refers to how well suited data is for its intended purpose (Wang and Strong 1996) and therefore its meaning changes according to application and context. Where sensor data is describing real-world systems, data quality can be considered to be how accurately it represents the system under observation. Therefore, its importance when it comes to generating accurate information from sensor data cannot be understated. As this accuracy can be inhibited by sensing errors or artefacts introduced environmentally, data quality can be determined by assessing the degree of sensing and environmental errors in the sensor data (Karkouch et al. 2016). Assessing degree of sensing errors has the benefit of assessing sensor performance. As with most technological equipment, failures can occur and sensors can

degrade or malfunction leading to spurious measurements, particularly for sensitive traditionally laboratory-based instruments like water quality sensors, when deployed in a harsh environment such as a DWDS. Detecting sensor errors generally requires knowledge both of the sensor and the application, so that unrealistic or untrustworthy measurements can be detected.

The water industry in general has a perception of being ‘data rich information poor’ (DRIP) (Mounce 2020), meaning there is a lot of willingness to gather and store data, yet the vast majority of collected data remains largely unused or used only within specific silos (Kyritsakas et al. 2023). While the quality will be better for data required for regulatory purposes, this does not translate to continuously monitored DWDS water quality data. This has the unfortunate result of requiring any data science project, seeking to extract information and knowledge from water industry datasets, to start with a time-consuming and challenging process of manually trawling through the data and determine which parts are suitable for subsequent analysis (Kyritsakas et al. 2023). Therefore, there is a need to introduce more robust data acquisition practices that focus more on ensuring sensor functionality.

2.4.2.1 Assessing Data Quality

There are many different approaches to assessing data quality, both in terms of methods of detection to approaches for handling erroneous data for subsequent analytic use, a process often referred to as data pre-processing or data cleaning. Selection of an appropriate approach for assessing sensor data quality requires an appreciation of the nature of the sensor technology, the availability of any reference sensors, and the difficulty of the measurement they are intended for. For some applications, data quality assessment will be based on a model of normal operating behaviour, with any deviations detected as erroneous (Teh et al. 2020). Such a model could be based on labelled past data or an assumed or theoretical distribution. Common model-based approaches are principal component analysis (PCA) and artificial neural networks (ANN) (Teh et al. 2020). PCA is a common dimensionality reduction technique and has been used to detect sensor faults in systems where sensors are highly correlated with each other (Dunia et al. 1996). ANN are supervised machine learning algorithms consisting of multiple neurons, that together aim to learn in a way that mimics the human brain. ANN can be used for detecting sensor faults by learning the normal behaviour and subsequently detecting deviations. However, a model-based approach is not always applicable in applications without a clearly definable set of normal operating conditions.

An alternative approach involves defining data quality metrics and writing rules to detect the presence of such metrics or errors (Karkouch et al. 2016; Kirchen et al. 2017). Defining data quality rules for any application would again be helped by labelled data, this time with the errors labelled. However, there are some errors that are commonly accepted to be undesirable for any sensor time series data. These include outliers, missing data, bias, drift, repeated values, and ‘stuck-at-zero’ (Teh et al. 2020). Though outliers are the most commonly studied sensor error (Teh et al. 2020), care needs to be taken to ensure real but abnormal events are not being incorrectly detected as errors and subsequently removed or imputed. Differentiating between sensor errors and real system events is difficult without the ability to cross-validate with other sensors in a network, using the logic that system events will be seen in multiple sensors unlike sensor faults (Krishnamachari and Iyengar 2004). García et al. (2017) developed a method for differentiating between sensor faults and real water quality events using known-to-be spatially correlated sensors, following the detection of an unusual data point. Time series data requires specific consideration for the consistency of timestamps (Pastorello et al. 2014; Gschwandtner and Erhart 2018), as many time series analytics require timestamps to be equally spaced. In the absence of labelled sensor errors, developing data quality rules requires knowledge of the sensor technology and the nature of the measurements.

2.4.2.2 Handling Sensor Errors

In most real-world applications, sensor data must be prepared and cleaned before any meaningful analysis can be undertaken. The extent to which this must be done depends both on the sensor data quality and the intended usage of its data. There are many different options to take with a poor quality time series dataset. In some cases, the data quality may be so poor that any analysis would not only be meaningless but could lead to erroneous interpretations. In these instances, the data should not be used and the source of the issue should be rectified for future monitoring. The two main issues that must be dealt with are erroneous data points and missing data. Erroneous data points are either removed or replaced, depending on the specific needs of the subsequent analysis. The extent to which erroneous data points must be dealt with depends on how frequent they are and what the required data formats are for the subsequent analytics. Gaps caused by missing data can greatly limit analytics where consistent timestamps are needed, commonly a requirement for time series analysis, but excessive imputation can lead to a dataset that no longer accurately represents the original

measurement. Nonetheless, missing data is inevitable to occur at some point, commonly due to battery issues or data corruption, and some analysis methods require them to be replaced. The process of replacing data points is known as imputation and common methods include smoothing, averaging or interpolation, usually based on a window of previous values, or comparing to a model describing normality (Palma 2016). Multiple imputation involves replacing missing data with several different plausible values, leading to a range of possible analysis solutions representing the uncertainty of the missing data (van Buuren 2018).

As mentioned in the previous section, methods for event detection and the detection of sensor errors can be similar in terms of methods and differentiating between these can be challenging. This can lead to a highly undesirable situation in which real and potentially alarming events are being automatically filtered by a data quality system. A famous example of this occurred when a British Antarctic Survey (Farman et al. 1985) detected a large drop in ozone concentrations. NASA had been monitoring the Antarctic ozone concentrations since the 1970's but their data quality control system had disregarded these previously unseen low measurements, which they then corrected in response to Farman et al. (Stolarski et al. 1986). This example demonstrates the importance of reporting and investigating anomalous data rather than simply excluding it. In a highly-correlated sensor network, differentiation of real events from sensor errors can be assisted by looking at whether the unusual feature occurred in more than one sensor. Li and Parker (2014) presented a method for replacing missing values using spatio-temporally correlated sensor data, a method that has greater accuracy and suitability for complex networks than other imputation techniques. This is similar to the approach used by Garcia et al. (2020), where chlorine loss of sensitivity incidents were differentiated from real low chlorine events using spatially correlated sensors, and confirmed by domain expert labels. However, the more complex a DWDS the harder it is to assume spatial connectivity between sensors, meaning without a method for assessing sensor connectivity, the above approach is reliant on having directly upstream sensors always available which could only be achieved with a very dense and expensive network of water quality sensors.

2.4.3 Combining Parameters and Sensor Locations

Multivariate time series datasets allow not only for each parameter to be examined over time, but for inter-variable relationships to be both discovered and examined. Taken at face value and without context, a multivariate time series dataset is a collection of independent

variables. However, understanding which variables are related to each other, and examining the nature of that relationship, can enhance overall understanding of the dataset. This can lead to informed feature engineering activities and a more enlightened analysis. In a sensor network with multiple-parameter monitors installed at a range of locations, it is valuable to understand not only relationships between different parameters in one location, but also between installation locations, in order to enable data fusion and for subsequent analytics to focus on multi-parameter multi-sensor data rather than being limited to analysing each parameter and sensor individually.

2.4.3.1 Understanding Relationships Between Variables

There are many situations where it is desirable to understand the relationship between time series variables, including to study known or assumed relationships in more detail and to discover previously unknown relationships. In some cases, the link between two variables may be obvious from the context of the monitoring, such as in input/output control processes (Box et al. 2015). This takes advantage of known connectivity between sensors in highly interconnected processes, and aims to develop a transfer function model that captures this relationship. Where sensor interconnectivity is less certain, it is often desirable to calculate the strength of relationship between multiple time series signals. Cross-correlation involves correlating two signals for various different time lags, and is the most common method for determining strength of relationship and time lag between time series signals (Benesty et al. 2004). This method can also be used where the strength of relationship is known but determining the time lag between variables is of importance, such as estimating tele seismic arrival times (Vandecar and Crosson 1990).

Dynamic time warping (DTW) also quantifies the similarity of two time series, and allows for a greater degree of fuzziness in terms of similarity of signal profiles than cross-correlation (Keogh and Pazzani 2001). Wavelet-based semblance analysis allows for the frequency content of two parameters to be compared over time in a 2d scalogram plot (Cooper and Cowan 2008). However, Pearson's correlation coefficient (PCC), commonly used in cross-correlation, has the advantage of providing an easily understood value between -1 and 1 that informs about strength of relationship. In some analytics, such as regression, the presence of correlations between independent variables can lead to unreliable results, a phenomenon known as multicollinearity (P. Vatcheva and Lee 2016). Knowledge of inter-variable relationships can lead to custom variables being generated during feature engineering, which

can lead to improved understanding and is particularly helpful for training machine learning models.

For water quality monitoring within DWDS, it is uncommon to have multiple water quality parameters monitored or to have multiple linked locations monitored. Where these are monitored, some of these parameters will have inherent relationships with each other that can be exploited in any subsequent analysis. One example is chlorine and temperature, where higher temperature water increases the rate of chlorine decay, as discussed in Section 2.3.2. Another is turbidity and flow rate, as sudden increased flow rate mobilises material leading to a period of increased turbidity, as discussed in Section 2.2.3. The relationship between hydraulics and turbidity in a UK DWDS, pre and post flushing operations, was examined using wavelet-based semblance (Mounce et al. 2015). The relationship between sensors at different locations within DWDS is generally more challenging to determine, with precise locations often unknown and hydraulic models are not accurate enough to determine transit times between sensor install locations. Even with flow rates being monitored, transit time is difficult to determine due to often lack of availability of pipe dimension. Therefore, estimating relationship and transit times between locations would enhance such datasets. DTW was used to estimate transit time between sensors in a sewer (Dürrenmatt et al. 2013), though the temperature sensors used were installed directly inline meaning determining the strength of connectivity was not necessary. Cross-correlation has been used to determine transit time between known-to-be connected locations in a straight DWDS using chlorine time series data (Bowden et al. 2006), though it has not been used to determine the strength of connectivity between sensor locations in more complex network layouts.

2.4.3.2 Utilising Relationships Between Variables

Once the relationship between two variables is understood, overall ability to examine the process or system under observation is enhanced. Returning to the example of modelling a transfer function between input and output processes, such a model will improve the ability to respond when the output time series changes, as the model will inform what necessary input signal changes can help to maintain and keep the process on track. Similarly, this relationship could be used to forecast future output values. With regards to DWDS water quality time series, knowledge of a strong relationship could be used to make estimates of expected values in correlated parameters. This could be valuable when detecting abnormal data, sensor errors, local and global events, and for replacing missing data. Relationships between variables also

refers to parameters that have a known association, such as turbidity and flow rate, and across and between different sensor locations.

Feature engineering can involve taking advantage of known or assumed links between variables to create a custom variable that can aid analysis (Geron 2019). Continuing the water quality example of turbidity and flow rate, knowledge of this link can be used to determine if the cause of turbidity increases are hydraulic or not. However when it comes to accurate analysis of discolouration events, turbidity alone does not tell the whole story, as it is effectively informing about the concentration of particulate matter in the water (Boxall and Saul 2005). Therefore, to understand the quantity of particulate matter, or discolouration material, passing a sensor location the turbidity signal needs to be multiplied by the flow rate signal, to generate a custom parameter often called material flux with units $\text{NTU}\cdot\text{m}^3/\text{hr}$ (Furnass 2015). Material flux can then be integrated to calculate total volume of discolouration material passing a location for a given time period, in units of $\text{NTU}\cdot\text{m}^3$. This can also be converted to total mass of particulate matter, if the relationship between particulate matter and turbidity is known (Gaffney and Boulton 2012). Less work has been done to combine data from more than one monitoring location in a water quality sensor network deployed in a DWDS, but this has the potential to enable a move from single sensor to network-wide analytics.

2.4.4 Time Series Forecasting

Forecasting is the practice of predicting the future, based on past and current observations. There are many fields in which time series forecasting plays a leading role, such as meteorology, stock broking, demand predictions, and epidemiology (Hyndman and Athanasopoulos 2021). Such forecasting applications range widely in their general importance, prediction accuracy, and how predictable the phenomenon of interest is. The accuracy of any forecast depends on several factors, such as our understanding of the underlying factors, data availability and quality, and how similar the future is to the past (Hyndman and Athanasopoulos 2021). The forecast horizon refers to how far into the future it is wished to predict, and of course this will impact forecast accuracy, as it is much easier to predict something right before it happens. For example, short-term weather forecasts are now very accurate due to improvements in weather modelling but predicting the weather with a longer forecast horizon, say a week, is much more challenging. Time series forecasting has been an active field of development since the 1980's with two main types of forecasting:

exponential smoothing and ARIMA (autoregressive integrated moving average), both with many variations to account for time series components such as seasonality and trend (De Gooijer and Hyndman 2006).

2.4.4.1 Traditional and Modern Approaches

Exponential smoothing as a form of time series forecasting has been studied since the 1950's (Holt 1957; Brown 1959), the original simple exponential smoothing (SES) approach has been expanded into around 15 different methods that account for different trend and seasonality (De Gooijer and Hyndman 2006) and are conveniently viewed through the ETS (error trend seasonality) framework (Hyndman and Athanasopoulos 2021). Exponential smoothing has been found to perform robustly for a variety of different univariate time series (Chatfield et al. 2001) and, unlike ARIMA-based approaches, model selection can be done automatically by minimising the Information Criterion (IC) (Hyndman and Athanasopoulos 2021). The use of ARIMA-based forecasting was formalised by Box and Jenkins (1970), who compiled previous research developing autoregressive (AR) and moving average (MA) approaches into an accessible framework. This book was very influential and there are many examples of research using ARIMA-based forecasting models throughout the 1980's and 1990's (De Gooijer and Hyndman 2006). Multivariate VARIMA (vector ARIMA) and ARIMAX (where X represents exogenous variables) were developed, though multivariate exponential smoothing is less common (De Gooijer and Hyndman 2006). In cases where there is a known relationship between two time series, a linear or non-linear regression model can be used to forecast one variable from the other (Hyndman and Athanasopoulos 2021), but of course this is dependent on having a suitable predictor variable.

The recent non-linear regression Prophet model (Taylor and Letham 2018) was developed to perform automatically on time series with significant and complex seasonality considerations. Significant research has also gone into adapting neural network approaches for time series forecasting, including neural network autoregression (NNAR) which uses lagged values (Hyndman and Athanasopoulos 2021) and LSTM (long short-term memory) (Hochreiter and Schmidhuber 1997). The increased interest in neural networks came following computing advances in the 1990's, making neural networks more practical (Geron 2019). However, it has been shown that classical time series forecasting tends to outperform machine learning in time series forecasting while also requiring far less computational power (Makridakis et al. 2018). Research has also shown that combining multiple forecasting approaches and taking

an average of their forecasts can lead to improvements compared to relying on any one forecasting approach (Bates and Granger 1969; Clemen 1989). This is clearly demonstrated by 12 of the 17 most accurate methods in the M4 time series forecasting competition being combinations (Makridakis et al. 2020).

2.4.4.2 Forecasting within Drinking Water Distribution Systems

Forecasting water quality parameters within DWDS is inherently challenging due to the dynamic unpredictable nature of these vast networks that are linked to human and weather patterns but also contain many inter-dependent and unknowable complex biological, chemical and physical reactions. Existing hydraulic models are not considered accurate enough to be expanded to water quality capabilities (Vreeburg 2007; Machell et al. 2014). Additionally, when you consider the amount of notice a water utility would require to be able to respond to a predicted water quality incident, this becomes an extremely difficult task. Significant research has gone into forecasting water demand, seen as important for strategic planning and useful in detecting leaks. As water quality is intrinsically linked to in-pipe hydraulics, water quality forecasting can learn from demand forecasting research. Longer term demand forecasting of several years horizon requires more consideration of anthropogenic factors, while shorter term forecasts of the scale of hours and days are improved by using meteorological data (Sebri 2016). Mu et al. (2020) found that LSTM produced better forecasts for water demand, at time resolutions between 15 minute to 24 hours, though ARIMA performed comparably for 24 hour resolutions. LSTM was shown to outperform other machine learning forecasting methods in predicting water demand in a network in China by Du et al. (2021), after pre-processing input data that included weather and holiday patterns using discrete wavelet transform (DWT) and PCA. A large-scale study using over 2 million domestic water meters compared seven different traditional and machine learning forecasting approaches and found that ARIMA performed the best, with the multi-layer perceptron (MLP) neural network approach the least reliable (Karamaziotis et al. 2020). Stańczyk et al. (2022) showed that modelling weekly seasonality can greatly improve short-term demand forecasting.

Less research has focused on DWDS forecasting water quality as opposed to quantity. Machine learning approaches have shown promise in predicting turbidity in DWDS several hours ahead. Random forests work by combining multiple decision trees, all trained on a subset of data and features, and Meyers, et al. (2017) found that they outperformed ANN and

SVM for classification-based predictions while Kazemi et al. (2018) used a nonlinear autoregressive exogenous ANN (NARX) to predict turbidity responses and demonstrated the value in calculating new input variables as it showed that the addition of a custom input parameter describing the time since last turbidity event improved model performance. However, these papers were based on very little data, the former using one turbidity time series and latter using two, meaning they are unlikely to generalise well. Additionally, both these examples involved training models using data containing events and utilised the clear relationship with flow rate to predict turbidity data during events, by including flow rate as a model input. However, training data containing turbidity events may not always be available and flow rate data is not always available so research is needed to forecast turbidity, and detect turbidity events, without requiring flow rates or previous events. At best, a turbidity forecasting method would consider any potential seasonality, though daily or sub-daily seasonality is difficult to capture (Hyndman and Athanasopoulos 2021), in order to estimate future turbidity values. Though, any unexpected water quality deteriorations, such as but not only hydraulic interventions, will most likely not be predictable without more information.

The challenges in maintaining a residual chlorine sufficient to prevent contamination has also lead to research forecasting chlorine residual levels in DWDS. Gibbs et al. (2006) found that a MLP outperformed other predictive models in forecasting chlorine concentration levels, using data from the WTWs and chlorine and temperature data from the two forecasted locations in an Australian DWDS. Bowden et al. (2006) developed a GRNN (general regression neural network) to forecast chlorine residuals at a single location in a DWDS up to 3 days in advance, with 24 hours the optimal forecast horizon, using 3 chlorine sensors within the DWDS. Turbidity, flow rate, pH and temperature were also used as model inputs. Maintaining adequate residual chlorine in drinking water was given increased importance due to the need for clean drinking water during the COVID-19 pandemic, with García-Ávila et al. (2021) developing a chlorine decay model with the aim to ensure a minimum concentration of 0.5 mg/l at end points in a DWDS in a town in Ecuador. However, this is challenging to perform on a wide scale as the wall decay factor will change depending on local pipe characteristics.

2.4.5 Event Detection

Algorithms detecting events using sensor data are found everywhere in modern life, from washing machines detecting an unbalanced load, to cars reporting system faults. Frequently a

human analyst will be able to visually identify events from looking at graphical representations of datasets. Likewise, a human expert may be able to quickly tell if there is a fault in a system. For example, a car mechanic will know sooner than someone unfamiliar with vehicles if there is a serious problem with the car they are driving. However, most people operating cars do not have this level of expertise and rely on dashboard alerts to tell them when something is going wrong. In many industries and applications, the quantity of data being measured is increasing exponentially and it is considered excessively expensive to rely on manual human interpretation for each event of interest (Shaukat et al. 2021). Additionally, there are applications where no human expert could feasibly handle the complexity of the data as efficiently as a computer can. Therefore, computing algorithms are necessary to provide adequate safety and efficiency in many processes.

The meaning of event detection varies enormously depending on application, in part because the word *event* is often not defined in this context and even when it is, definitions can be very different (Yu et al. 2020). This has led to many highly specialised approaches with limited cross-sector applicability. Where forecasting is about predicting the future, event detection is about accurate and timely recognition of *something of interest*. That the accurate and timely detection of things like natural hazards, pollution events, disease outbreaks, and network hacking (Aggarwal 2016; Yu et al. 2020), is still considered so vital speaks to our inability to accurately predict the future, despite advanced modelling existing in all these areas. For some event types, such as disease outbreaks or resource contamination locating the source spatially is vital in mitigating the effects. An obvious example is the recent COVID-19 outbreak, which is thought to have started in China around November 2019, yet was allowed to spread seemingly without notice for several weeks (Roberts et al. 2021). However, it is difficult to detect something you are not actively looking for and the same can be said of water quality deterioration events in DWDS: without taking high quality data, such events will only be learned about through customer contacts, as is currently frequently the case (Mounce 2020).

2.4.5.1 Selecting a Detection Method

Determining a suitable event detection approach requires consideration for both the type of event in question, and the nature of available data (Chandola, Banerjee, et al. 2009). Events can broadly be categorised as either unusual events, sometimes termed outliers or anomalies, and important but not necessarily unusual events. Many methods have been developed to detect anomalous or outlying events in data, with the availability of supporting data labels

influential in method choice. Supervised anomaly detection requires a training dataset with all data labelled either anomalous or not, while semi-supervised approaches require either normality or anomalies to be labelled (Chandola, Cheboli, et al. 2009). Unsupervised approaches have been heavily researched due to many real-world applications not having labelled data available, and tend to be based on assumptions that the normal portion of data will be represented by a particular distribution. Isolation forests (Liu et al. 2008), elliptical envelopes (Rousseeuw and Driessen 1999) and local outlier factor (LOF) (Breunig et al. 2000) are popular unsupervised anomaly detection methods that look to create a boundary between normal and abnormal datapoints. Isolation forests identify anomalous datapoints as opposed to profiling normal data, elliptical envelopes assumes the normal data follows a Gaussian distribution, and LOF works by measuring local deviations from each neighbouring point. One-class SVM (OCSVM) is a variant of the popular supervised classification method SVM that is unsupervised and splits data in two distinct classes. The suitability of any of these approaches is dependent on the nature of the data, but these are most commonly applied to multivariate datasets. PCA is a dimensionality reduction technique that was discussed in Section 2.4.2.1 as a data quality assessment method. It has been found to be an effective way to reduce such highly-correlated high-dimensional datasets into smaller uncorrelated datasets more suited to unsupervised anomaly detection (Dunia et al. 1996; Aggarwal 2016).

Once commonality between the many event detection approaches is some sort of attempt to model or capture normal patterns of behaviour (Aggarwal 2016), with a boundary line often made to distinguish normal from abnormal. However, in many cases this distinction is not black and white. For events that are not necessarily rare, a more specific approach is usually required, that appreciates the nature of the event in question and the sensing technology relied upon to detect it. Where the data is a time series, the temporal context for each datapoint is an important consideration, particularly where there is some seasonality involved. For example, a thermometer measuring 35 degrees in winter might be a sensor error but in summer would be believable. The former is often termed a contextual anomaly. Continuing this example, any suspected anomalous thermometer could be checked against other thermometers or weather stations to confirm where the measurement is real or not. Where this is not available, suspected anomalies should be checked against preceding values. The most common method for anomaly detection in time series involves comparing a value to its forecast equivalent, with a threshold above which an event is reported (Aggarwal 2016). Supervised approaches for anomaly detection in time series include ANN-based approaches, with the Long Short

Term Memory (LSTM) variant of recurrent neural networks (RNNs) proving to produce excellent results for time series where labelled datasets are available (Hochreiter and Schmidhuber 1997; Shaukat et al. 2021).

Events within time series data are often multiple continuous data points, or subsequences, rather than single points, and dynamic time warping (DTW) has been successfully implemented to find the most unusual subsequence within a time series (Keogh and Pazzani 2001), an important problem for analysing electrocardiogram (ECG) time series (Lin et al. 2005). Where a specific subsequence signal is the event in question, pattern matching can be done to look for its presence within a time series. A robust method was developed to achieve this based on a piecewise linear representation (Keogh 2012). These are also sometimes called collective anomalies (Shaukat et al. 2021). Change point or novelty detection algorithms look for the point at which there is a change in the overall state. This can be achieved by looking for points where the model best describing the underlying data changes, as proposed by Guralnik and Srivastava (1999). This paper also asked four human experts to label change points and the resulting labels showed significant disagreements and bias from including noise for two of the experts, highlighting the complexity of determining ground truth for what is and isn't an event. Often events in time series are required to be detected in real-time, which can be done by comparing some expected or forecast value to the latest detected value,

2.4.5.2 Event Detection in Drinking Water Distribution Systems

Detection of leaking or bursting pipes is the most common event detection task for DWDS, with algorithms developed tending to require either acoustic or pressure time series data (Romano et al. 2014; El-Zahab and Zayed 2019). Other approaches have also been investigated, such as the use of pattern matching to search for specific subsequences (Mounce et al. 2014). Though this only had limited success due to the difficulty in generating a collection of subsequences to cover all potential scenarios. However, the link between leaking pipes and acoustic and pressure data is strong, while the link between water quality contamination and commonly monitored water quality parameters can be more complex. Significant research, led by the US Environmental Protection Agency (EPA), developed methods to detect intentional contamination of DWDS by comparing detected values to values predicted by windowed statistical algorithms, with the addition of a Binomial Event Discriminator (BED) that outputs a probabilistic event score from counting detected outliers

(McKenna et al. 2007). The BED approach effectively treats each BED window as a Bernoulli process, which means each data point in a sequence is considered an independent trial carried out under the same conditions. However, this assumption might not fully hold for water quality time series data, which could somewhat restrict the interpretability of the BED probability score. Nonetheless, it was shown to effectively cluster groups of detected outliers and reduce the very high level of false positives (>0.9 for 4/9 thresholds tested in McKenna et al. (2007) from using the multivariate nearest neighbour (MVNN) approach on its own. This research led to the production of the open-source water quality event detection tool CANARY (Hart et al. 2007). CANARY's use of sliding window statistics, to effectively describe normality within that window, is designed to account for the ever-changing nature of water quality time series.

A review of recent research into water quality in DWDS by Weston et al. (2022) found that 13% of related publications since 2000 reference intentional contamination, often referring to the '9/11' terrorist attack in New York, though these have become less frequent over time. Perelman et al. (2012) proposed an ANN model for multivariate water quality and Bayesian analysis to identify outliers from the residuals. The ANN was trained on labelled normal data, with simulated intentional contamination events, with Gaussian distributions and 8 hour durations, introduced and subsequently detected. As mentioned previously in Section 2.3.2, chlorine has been shown to be useful in detecting some potential contaminants and Eliades et al. (2014) developed a contamination event detection system, by comparing modelled chlorine concentrations, based on chlorine input data, to actual chlorine sensors. Yu et al. (2017) proposed combining known-to-be hydraulically connected sensors to improve detection ability, once again using simulated intentional contamination events. Li et al. (2019) used a multivariate GAN (generative adversarial network) to train a LSTM model to detect simulated contamination events using data generated from a Water Distribution (WADI) testbed. Muharemi et al. (2019) won a water quality event detection competition, using a real-world multivariate dataset from a DWDS in Germany, with a LSTM approach. However, the dataset provided had artificial events inserted, as has become common practice for DWDS water quality event research. Simulated events were also used to develop CANARY (Murray and Haxton 2010), highlighting one of the central problems of water quality event detection in DWDS: the difficulty of obtaining real-world confirmed events.

Furthermore, intentional contamination events are significantly different and more rare than unintentional DWDS water quality deterioration events. As identified in Section 2.2.3, discolouration events are the most common type of water quality deterioration event and, though their frequency will vary depending on sensor location, millions of people in the UK and worldwide are forced to deal with discoloured water each year (DWI 2022). Therefore, utilities have, in theory, millions of examples of confirmed discolouration events. But the problem is the absence of water quality time series data from within relevant DWDS. The link between DWDS turbidity time series and discolouration events at customer taps is therefore difficult to study. The two turbidity forecasting research papers outlined previously in Section 2.4.4 included a focus on turbidity events, yet had very different definitions. Meyers et al. (2017) used classification metrics based on event thresholds of 1, 2 and 4 NTU, while Kazemi et al. (2018) did not clearly define a threshold though the events presented are all under 1 NTU. However, both of these papers attempted to forecast future turbidity values, events and all, as opposed to first calculating a residual. In further developing turbidity event detection methods, there is a need to first better understand what a turbidity event is. Although turbidity events are not necessarily rare, they can still be considered deviations from normal operating conditions meaning the approach of comparing forecasts may be useful.

2.5 Summary

Advancements in technology have empowered WSPs to implement continuous monitoring of water quality within DWDS. This evolution marks a departure from traditional reliance on sporadic grab sampling or customer interactions to ascertain network conditions and identify potential contamination incidents. Nevertheless, the deployment of water quality sensors within DWDS is relatively new, meaning both deployment strategies and analytic routines have not attracted much research attention. Water quality sensors have thus far been mainly used in laboratory environments and are prone to sensing errors when remotely deployed. Therefore, research is required to understand the nature of these errors and to develop effective methods for their detection and mitigation.

This review has identified discoloration events as being a particularly promising focal point for investigation. This is both due to the frequency of in-network discoloration events and the fact that turbidity sensors are proxy discoloration sensors that are suited to online deployment

within DWDS. Furthermore, little research has gone into water quality sensor deployment strategies. Additionally, more research is required to understand what parameters are most useful and how combinations of parameters and sensors can improve understanding and insight. In particular, the presumed increase in insight obtainable with more parameters and sensor locations requires better understanding, as this directly impacts the value of deployed sensors. These knowledge gaps have limited the progress of DWDS water quality sensor networks and require addressing to realise the benefits of more sophisticated digitalised DWDS water quality monitoring.

Chapter 3: Aims and Objectives

The aim of this research is to develop analytical routines that transform water quality time series data from DWDS into actionable insight in order to aide operational management. A central concept driving this research is that the value obtained increases moving from analysis based on single parameter single sensor to multiple parameter multiple sensors, as illustrated in Figure 3.1, but the nature of this increase is unknown.

The following objectives have been defined:

1. Develop methods to assess sensor performance, both to assist with analysis of historical datasets and to improve sensor management during deployment.
2. Investigate and understand how discolouration events can be identified and understood through water quality time series data.
3. Explore how different combinations of parameters and sensors impacts the level of insight that can be derived.

Objective 1 and 2 are addressed in **Chapters 4** and **5**, respectively. Objective 3 will be examined throughout but in particular through the longer form **Chapter 6**, which features examples ranging from single parameter single sensors (SPSS), to combining multiple sensors measuring the same parameter (SPMS), combining multiple parameter in a single location (MPSS) to finally explore the added value from having multiple parameters monitored at different locations (MPMS).

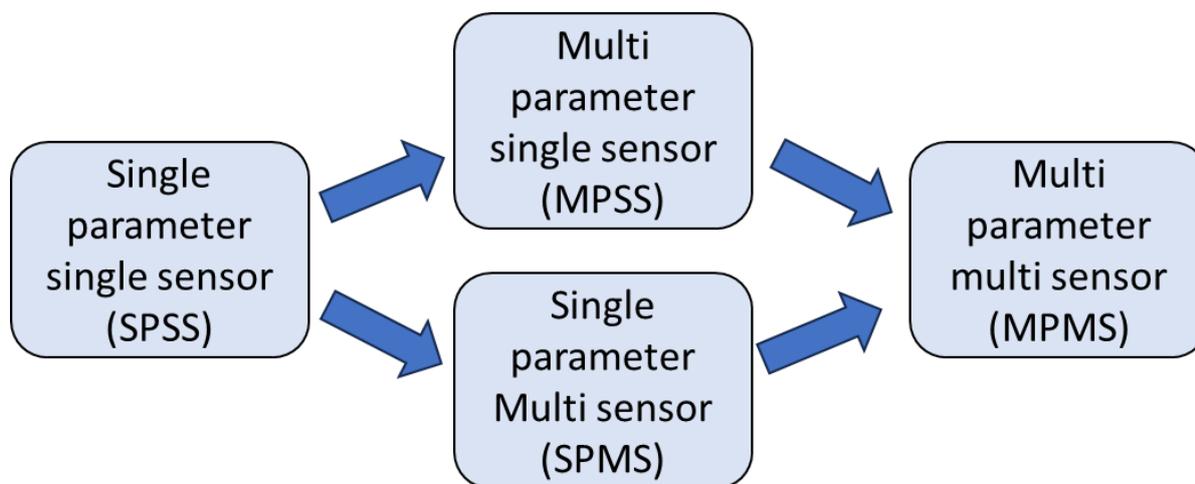


Figure 3.1. Analysis moving from single parameter single sensor (SPSS) to multi parameter multi sensor (MPMS).

Chapter 4: A Data Quality Assessment Framework for Drinking Water Distribution System Water Quality Time Series Datasets.

Reproduced from Gleeson K, Husband S, Gaffney J, Boxall J. A data quality assessment framework for drinking water distribution system water quality time series datasets. *IWA AQUA - Water Infrastructure, Ecosystems and Society* (2023) 72 (3): 329–347. *Submitted December 2022. Published March 2023.* <https://doi.org/10.2166/aqua.2023.228>

Declaration

Chapter 4 is a peer-review journal paper published in *IWA Aqua: Water Infrastructure, Ecosystems and Society*. The contribution of the authors are the following:

1. **Killian Gleeson** is the PhD candidate and first author of this chapter. Killian developed the framework, wrote all of the code involved in each step of the framework, and wrote the first draft of the chapter.
2. **Dr. Stewart Husband** and **Prof. Joby Boxall** are Killian’s academic supervisors, and **Dr. John Gaffney** is Killian’s industrial supervisor, and all three are co-authors of this chapter. All three provided critical feedback and input into the thinking behind the framework. Additionally, they provided constructive advice about the structuring of this chapter, including proposing changes to the narrative and presentation.

4.1 Abstract

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. Focusing on turbidity and chlorine, the framework consists of three steps to bridge the gap between data and information; firstly, an automated rules-based data quality assessment is developed and applied to each water quality sensor, then cross-correlation to determine spatiotemporal

relationships and finally spatiotemporal information enabled multi-sensor data quality validation. The framework provides a method to achieve automated data quality assurance, applicable to both historic or 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.

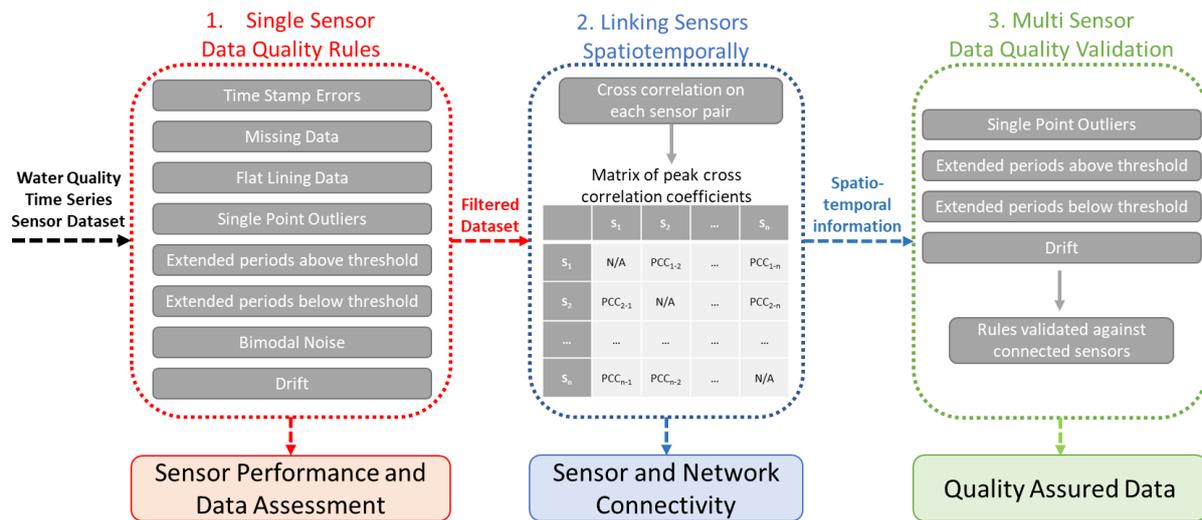


Figure 4.1. Chapter 4 graphical abstract.

4.2 Introduction

Monitoring of distributed drinking water quality typically consists of periodic discrete sampling that fulfils regulatory purposes but only offers 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 and Boxall 2011) and disinfection residual decay (Speight and 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 minutes 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 and 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, these are relatively (to flow rate and pressure) sensitive scientific instruments and as such the quality of data obtained from their deployment can be variable.

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 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 and 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.

4.2.1 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 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). 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 (Yang Zhang et al. 2010) demonstrates the need to be able to validate these, and other potentially erroneous occurrences, with other sensors. A rules-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, subsequent analysis may require 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, longer gaps require consideration of the temporal fluctuations present in water quality time series (Zhang and Thorburn 2022).

Understanding how simultaneously recorded time series are related to each other spatiotemporally has been studied in areas such as seismology (Vandecar and 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 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 rate 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 and 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 relationship. Though cross-correlation has not been used previously to relate water quality sensors spatiotemporally, semblance correlations between turbidity and hydraulic data have been to infer changes in 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 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 in further analysis.

4.3 Method

4.3.1 Multi-Sensor Data Quality Assessment Framework

A rules-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 water quality data neither being stationary nor normally distributed. Therefore, a framework was developed, illustrated in Figure 4.2, 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 assessment for water quality sensor networks deployed within DWDS. The first step 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, 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 for 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, 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).

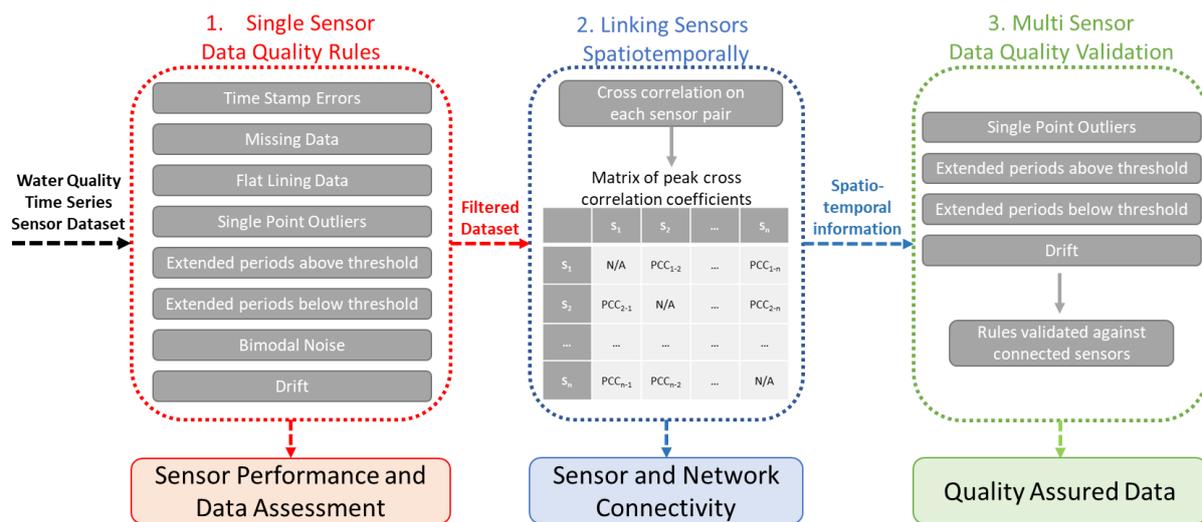


Figure 4.2. Multi-Sensor Data Quality Assessment Framework.

4.3.2 Data Quality Rules

The data quality issues, which were identified as the basis of each rule, are outlined in Table 4.1, 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

rules-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.

Table 4.1. Data Quality Rules: detection methods and possible causes.

Rule	Detection by	Possible Cause
Time Stamp 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	Z-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

4.3.2.1 Time Stamp Errors

Time stamp errors refer to data points that have an unintended sampling interval, compared to the previous data points. Imbalanced datasets as such suffer from bias, the lack of consistent time stamps 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 time stamp 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 to the previously taken data point. In the datasets reviewed, target sample rates changed at most a couple of times a year, meaning this method would result in a negligible percentage of time stamp errors with normal strategic sample rate changes and well-functioning data acquisition.

4.3.2.2 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 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 minutes). The maximum number of samples for this sample rate and timeframe was then calculated and compared to 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.

4.3.2.3 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 sampling rate, as opposed to looking at total number of data points with repeating values. For example, a sensor sampling every 10 seconds might return the same value 10 times in a row, but this is a quite a different prospect to a sensor sampling every 15 minutes returning the same value 10 times in a row, as that would mean no water quality changes have been detected for 2 and a half hours. Nonetheless, sampling rate is unavoidably a significant influencing factor on a sensor's tendency to flat line, as is the resolution of the sensor.

4.3.2.4 Single Point Outliers

Single point outliers (SPO) refer to values that are unrepresentative relative to 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 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, difference from 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.

4.3.2.5 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 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.

4.3.2.6 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 minutes 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 rules-based approach.

4.3.2.7 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 ability to respond, meaning 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 see successive changes. As there is evidence that drift in turbidity sensors can be corrected (Gaffney and Boulton 2012), a drift correction method was developed that involves fitting the drift data using asymmetric least squares (Peng et al. 2010).

4.3.3 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 followed 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 4.3. Transit time is defined as the average difference in hydraulic arrival times between two locations, which are not necessarily directly inline. The top plot in Figure 4.3 shows two chlorine sensor time series, over the course of twenty-four 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 hours (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.

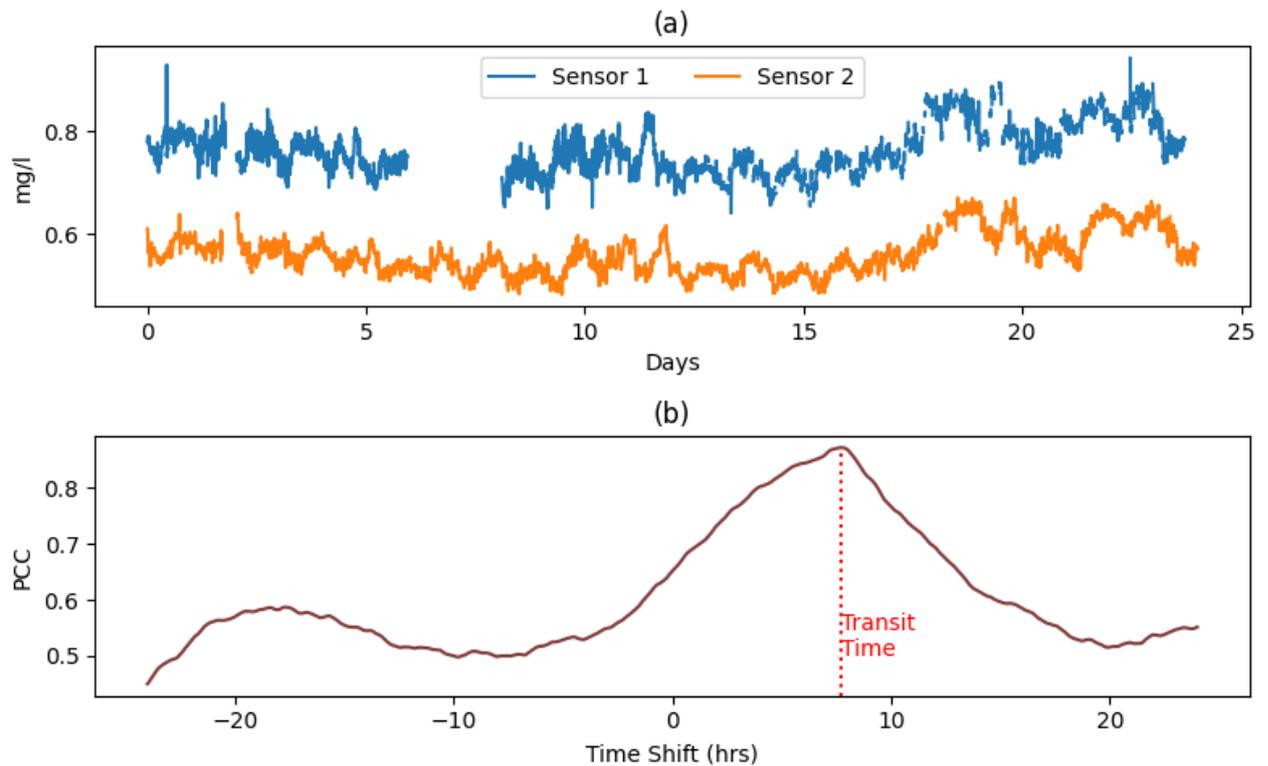


Figure 4.3. Two example chlorine time series (a) and corresponding cross-correlation curve (b).

4.3.4 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 connected, 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 rule is only seen in 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.

4.3.5 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 4.2. All water quality sensors listed monitored 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 A and C becoming available while monitoring was ongoing. This meant that 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).

Table 4.2. Datasets used.

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

4.4 Results

4.4.1 Data Quality Assessment Rules

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

4.4.1.1 User Defined 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 4.4 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 hours, 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 4.3. 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, whilst thresholds were based on reasonable expected values. An advantage of the rules-based approach adopted here is 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.

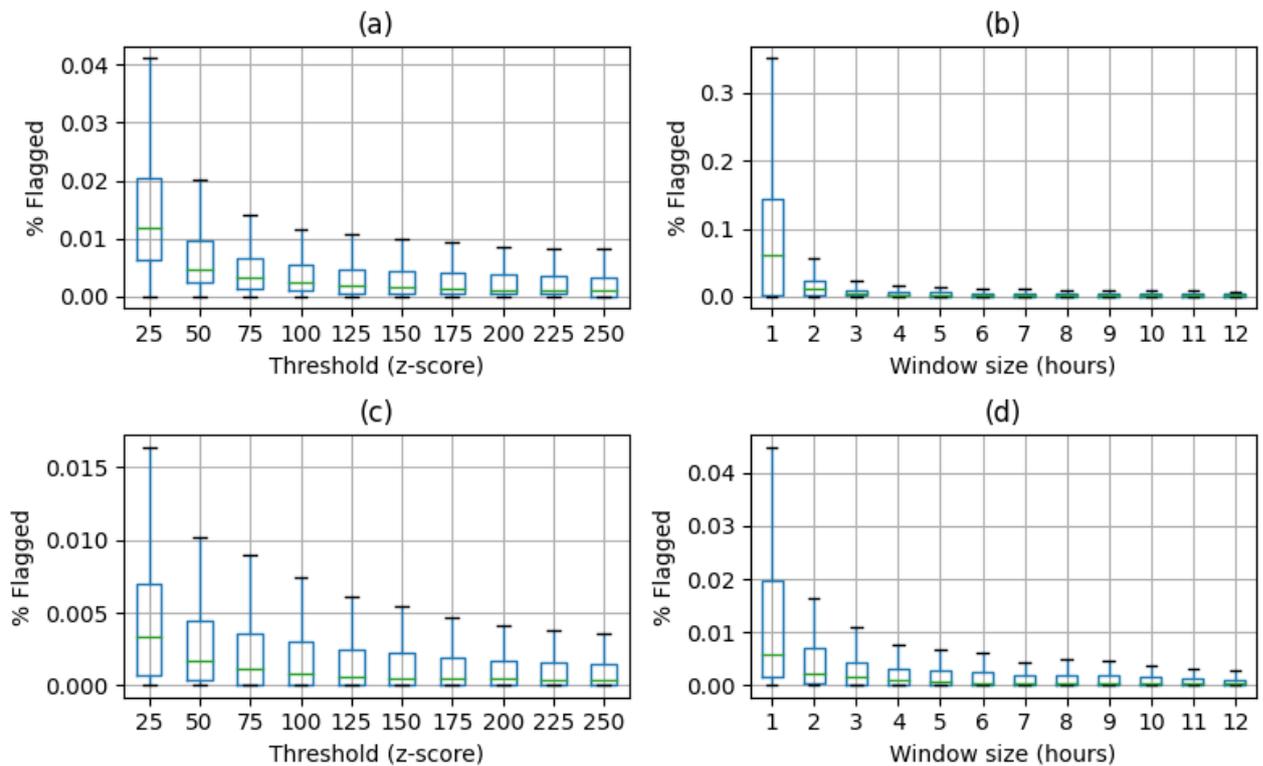


Figure 4.4. Boxplots showing sensitivity of single point outlier rule for turbidity sensors, with (a) z-score threshold and (b) window size, and for chlorine sensors, with (c) z-score threshold and (d) window size.

Table 4.3. 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 hours
Single Point Outliers	Z-score for window pre and post each data point	Z-score threshold = 100 Window size = 6 hours
Extended periods above threshold	Minimum rolling values for minimum duration	Minimum threshold = 1.5 NTU/ 1.5 mg/l Cl Minimum duration = 6 hours
Extended periods below threshold	Maximum rolling values for minimum duration	Maximum threshold = 0.05 NTU/0.15 mg/l Cl Minimum duration = 6 hours
Bimodal Noise	Minimum median non-zero delta in a window	Minimum threshold = 0.1 Window size = 6 hours
Drift	Successive duration of weekly median increases	Minimum duration = 4 weeks Minimum overall rise = 0.3 NTU/ 0.3 mg/l Cl

4.4.1.2 Time Stamp Errors and Missing Data

Figure 4.5 is a plot of the sampling intervals for a single water quality sensor over the course of 18 months (blue dots), with the redline showing the accumulated time stamp errors, where each detected time stamp error equals 1. In this example, 13.7% of the data points were detected as time stamp errors. Using an assumed target sample interval of 15 minutes for the first 12 months, and 1 minute for the last 6 months, the actual intervals can be compared to this assumed target. Using this method results in 12% of the data points being flagged as time stamp errors. This shows that the original method slightly overestimated the prevalence of time stamp errors, explained by this example fluctuating in and out of target sample intervals (the first instance of return to target sample rate will be flagged as a time stamp errors). To determine the quantity of missing data in this example, the time series resampled to 15-minute intervals results in 44,264 nonempty samples, compared to a maximum of 52,323 for this timeframe, meaning a total of 8,059 missing resampled data points were calculated (or 15% of maximum potential data).

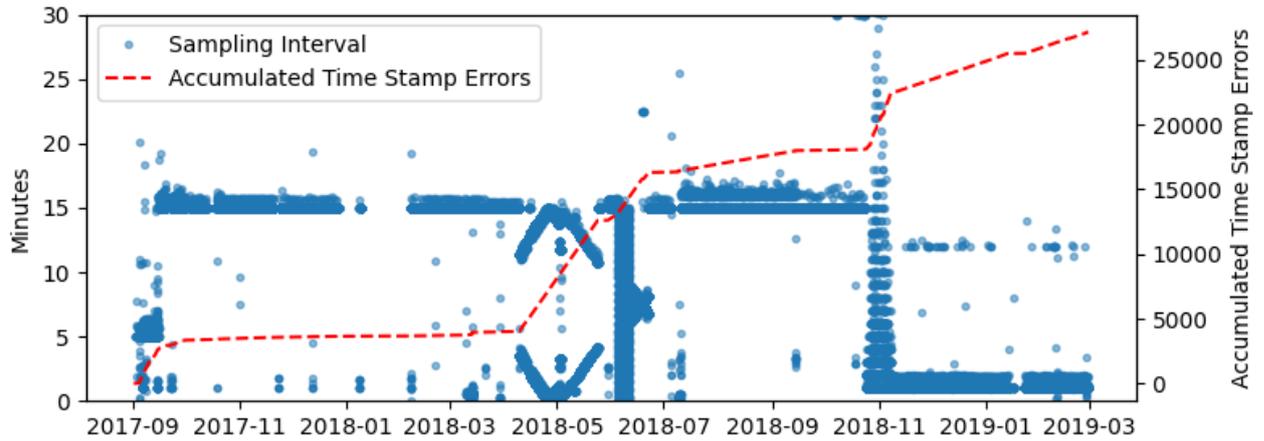


Figure 4.5. Time series plots of sampling intervals (blue) and accumulated time stamp errors (red).

4.4.1.3 Flat Lining Data

Figure 4.6 is an example of a chlorine sensor with significant levels of flat lining. The flat lining duration threshold here was set to 8 hours, with 66% of the data surpassing this level in the 8 months' worth of data shown here.

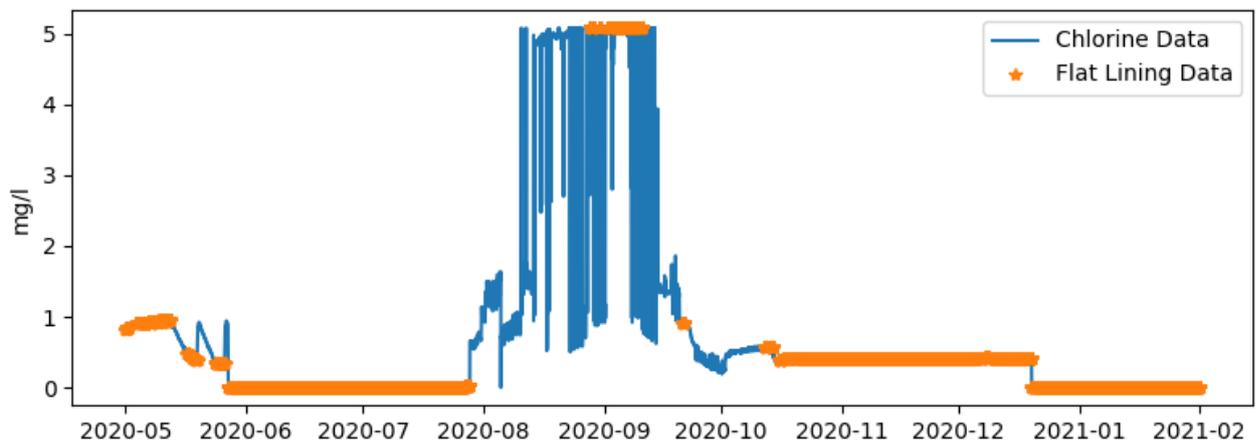


Figure 4.6. Chlorine time series with flat lines of at least 8 hours highlighted.

4.4.1.4 Single Point Outliers

Figure 4.7 shows an example, using a z-score threshold of 100 and window size of 6 hours, 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 hours 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.

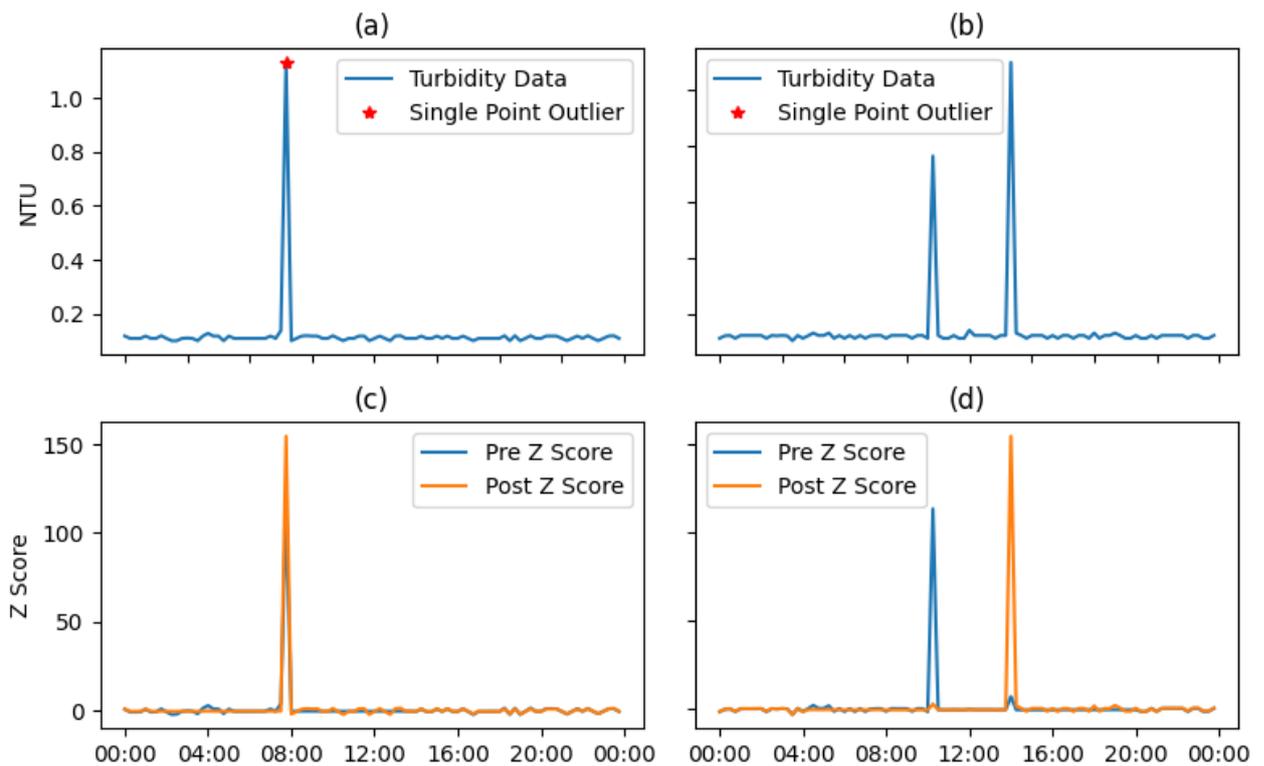


Figure 4.7. (a) Turbidity time series with single point outlier detected, (b) turbidity time series with no single point outlier detected, (c) pre and post z scores corresponding to (a), (d) pre and post z scores corresponding to (b).

4.4.1.5 Extended periods above (or below) a threshold

Figure 4.8 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 0.15 mg/l, and the minimum duration was 6 hours. 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. Disinfection residuals for the UK system in which this was deployed was designed to stay above 0.2 mg/l and below 1 mg/l.

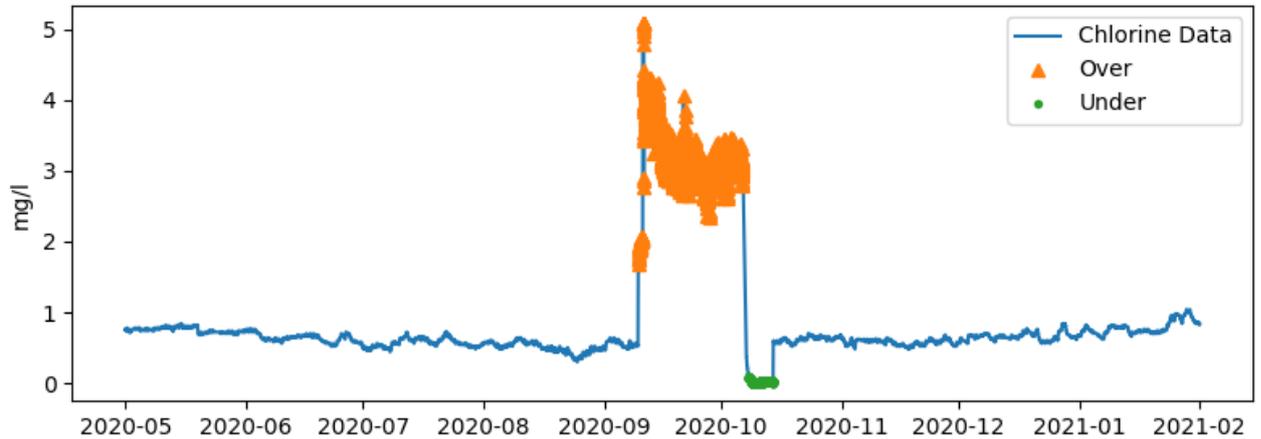


Figure 4.8. Chlorine time series with regions staying over 1.5 mg/l, or under 0.15 mg/l, for 6 hours highlighted.

4.4.1.6 Bimodal Noise

Bimodal noise is illustrated in Figure 4.9, 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 window size of 6 hours.

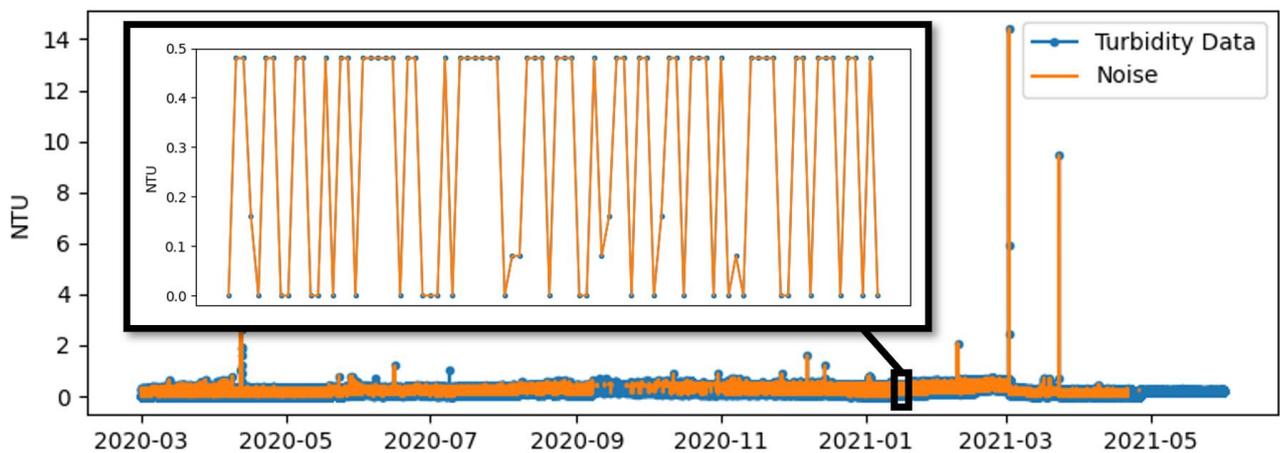


Figure 4.9. Turbidity time series with closeup of highlighted bimodal noise.

4.4.1.7 Drift

Figure 4.10 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 minimum overall increase of 0.3 NTU. The bottom plot of Figure 4.10 shows the drift data corrected using asymmetric least squares.

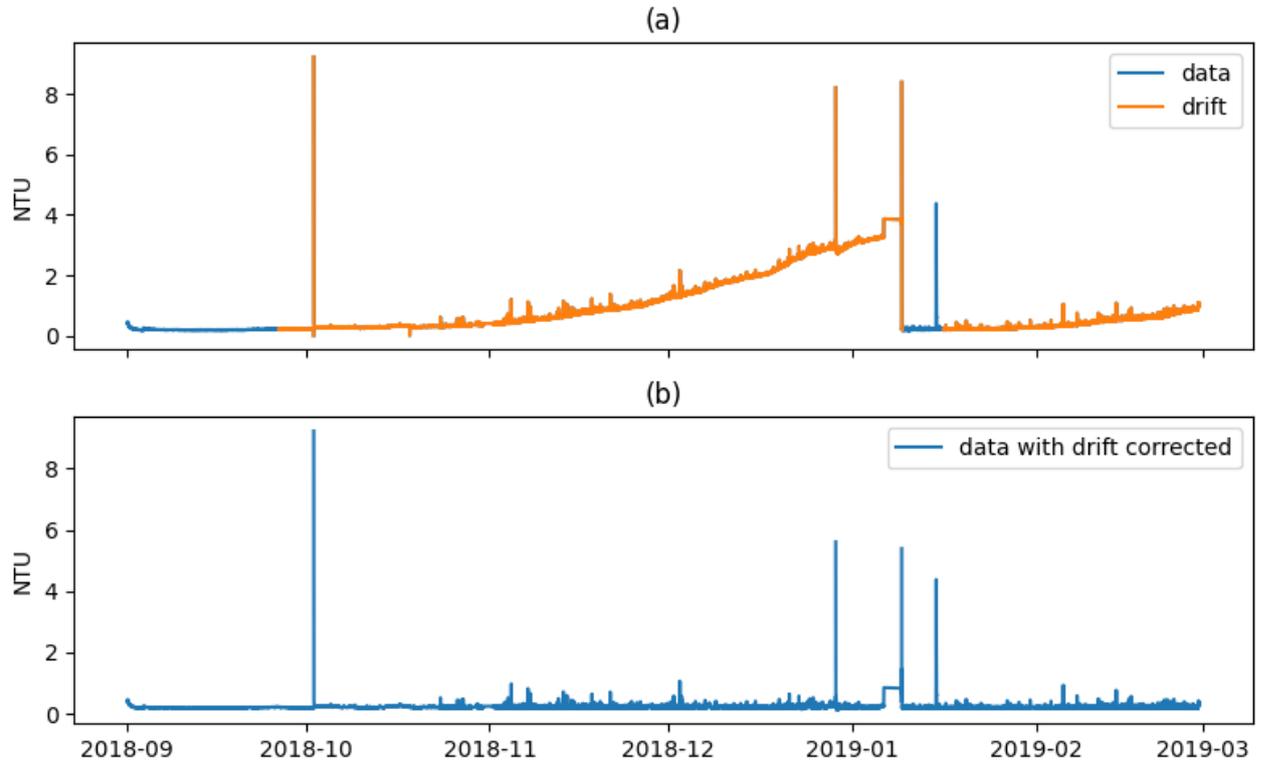


Figure 4.10. (a) turbidity time series with two periods of drift highlighted, (b) drift corrected turbidity data.

4.4.1.8 Rules applied to datasets

The eight rules were applied to each of the three datasets listed in Table 4.2, enabling sensor performance to be assessed and ranked, illustrated in stacked bar charts in Figure 4.11. These results highlight the multiple data quality issues seen in these three datasets and show how prevalent each rule was for turbidity and chlorine sensors. Time stamp 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.

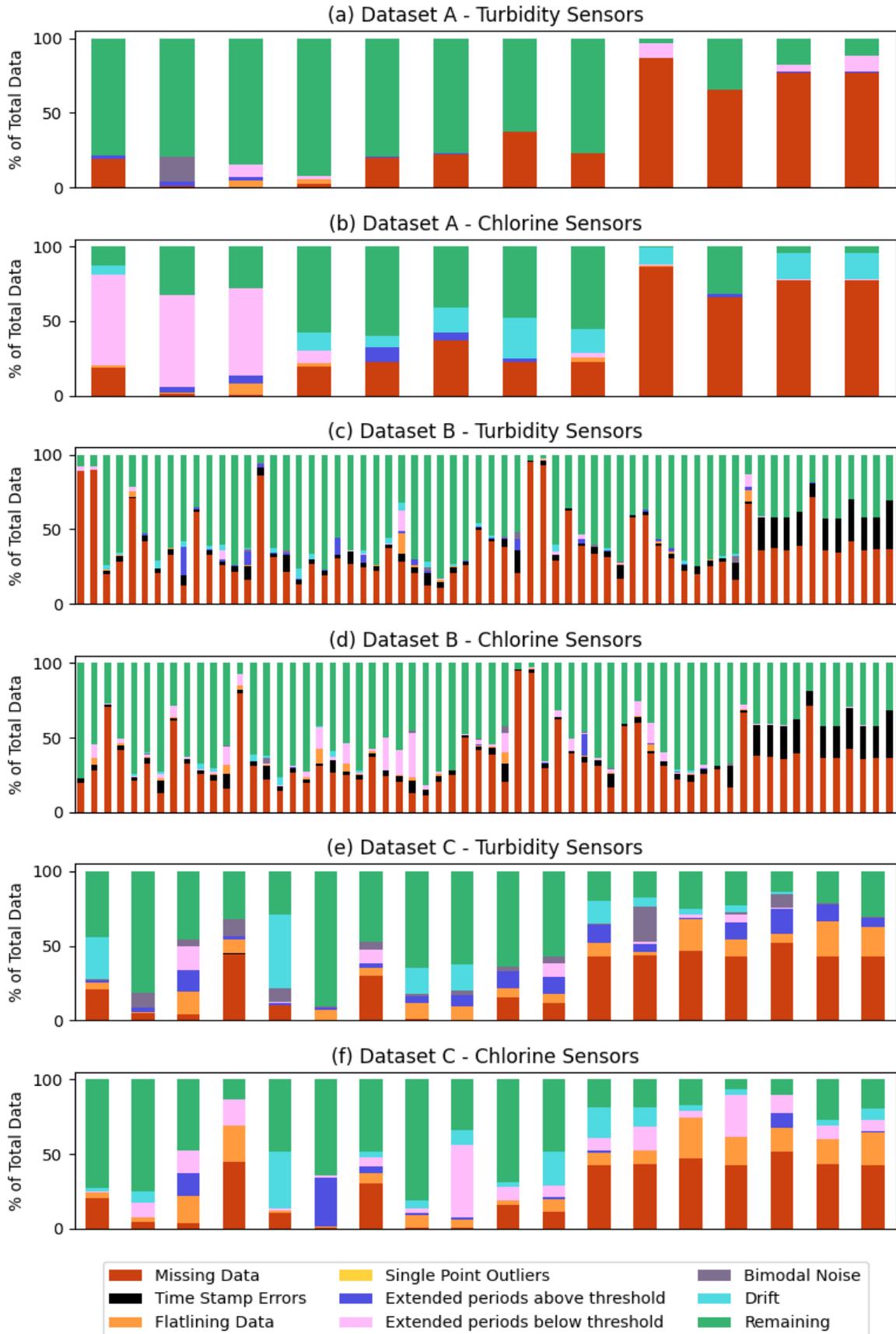


Figure 4.11. Data quality rules applied to Dataset A turbidity (a) and chlorine sensors (b), Dataset B turbidity (c) and chlorine sensors (d), and Dataset C turbidity (e) and chlorine sensors (f).

4.4.2 Linking Sensors Spatiotemporally

4.4.2.1 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 as 50% of total window length for this work. This 50% commonality limit was selected to ensure that correlations were meaningful while also allowing for the long periods of missing or low-quality data experienced. PCC's are calculated at different time shifts for each possible sensor pair, with 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 four-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.

Cross-correlations results presented were all calculated on chlorine time series data, a parameter that was well-suited for this method. Figure 4.12 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 displayed similar seasonal chlorine trends, possibly due to their sharing of the same treatment works. When cross-correlations were performed using window sizes of four 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 four-week window size was used, calculated once a week, with median coefficients presented for each sensor pair in heatmaps in Figure 4.13(a), (b) and (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.

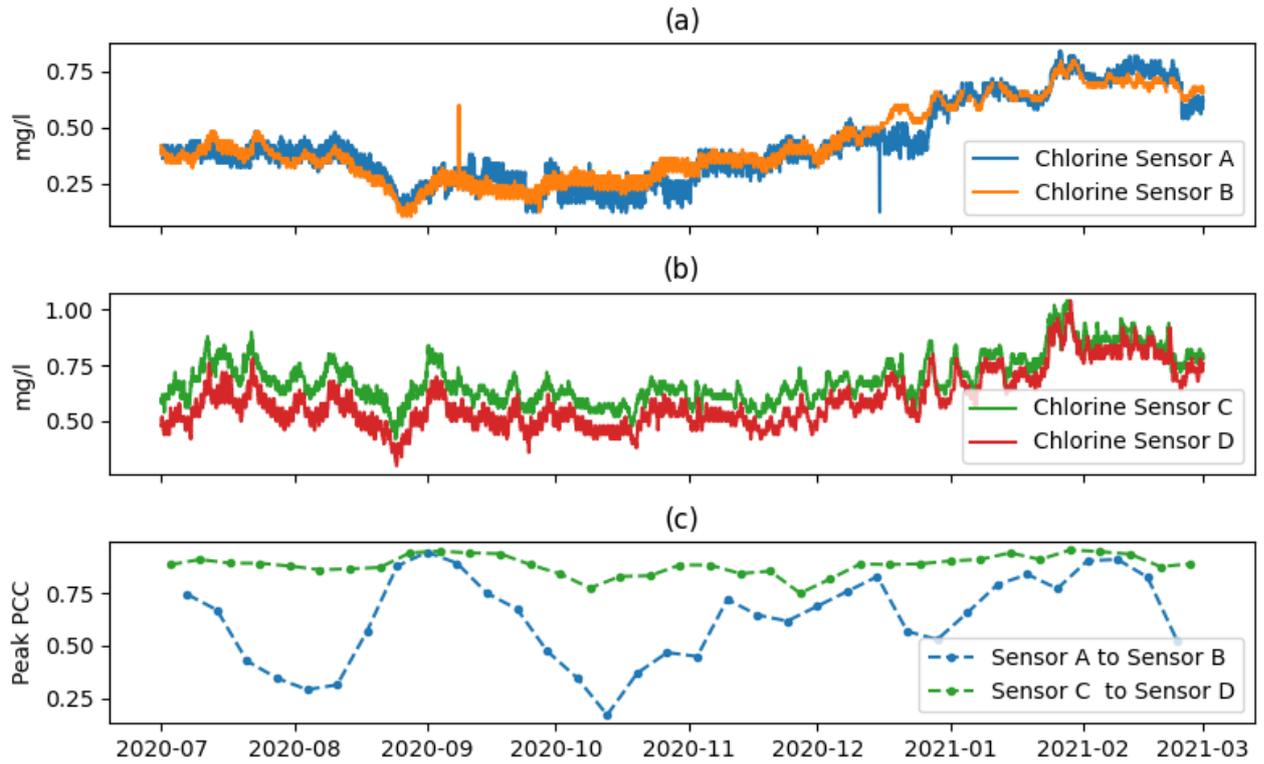


Figure 4.12. Chlorine time series profile for two pairs, shown in (a) and (b), and (c) the sliding cross-correlation coefficients calculated using overlapping 4-week windows every 7 days.

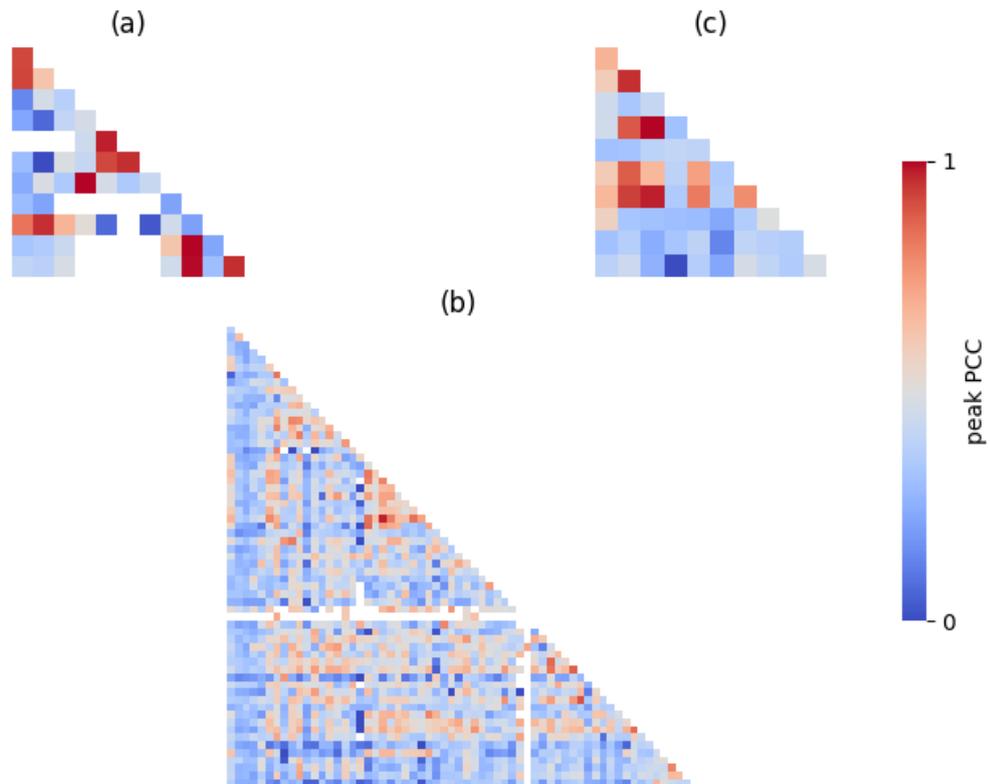


Figure 4.13. Heatmaps with peak PCC for each chlorine sensor pair in Datasets A (a), B (b), and C (c). Blank squares indicated insufficient data.

4.4.3 Multi-sensor Data Quality Validation

The final step in the framework uses the spatiotemporal information, derived from cross-correlation of chlorine sensors, to enable multi-sensor data validation. Figure 4.14(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 4.14(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 4.15 illustrates how the spatiotemporal information derived from cross-correlations done on chlorine sensors can be utilised for other parameters in the same locations. In this case, the chlorine based cross-correlations provide information that Sensor P in Figure 4.15(a) and Sensor Q in Figure 4.15(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.

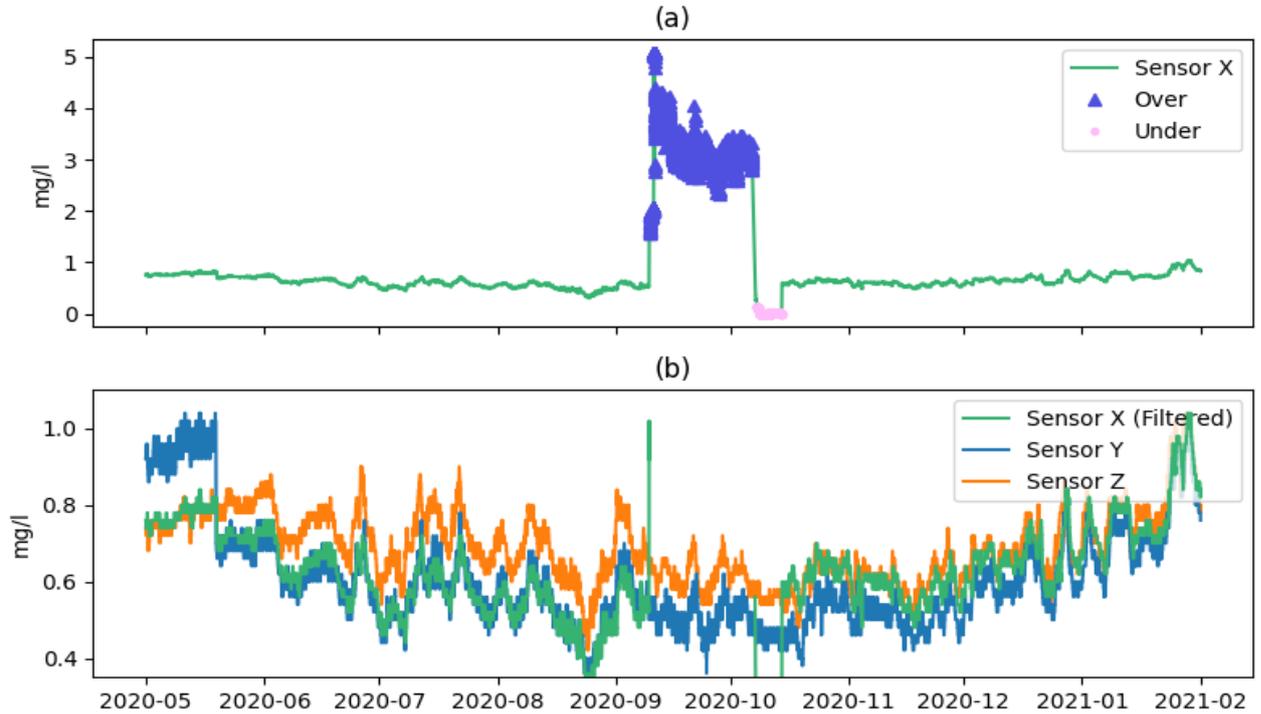


Figure 4.14. Multi-sensor data quality validation example, with (a) flagging of periods above 1.5 mg/l and below 0.15 mg/l for 6 hours in Sensor A and (B) showing absence of anomalous feature in two sensors calculated to be correlated to Sensor A.

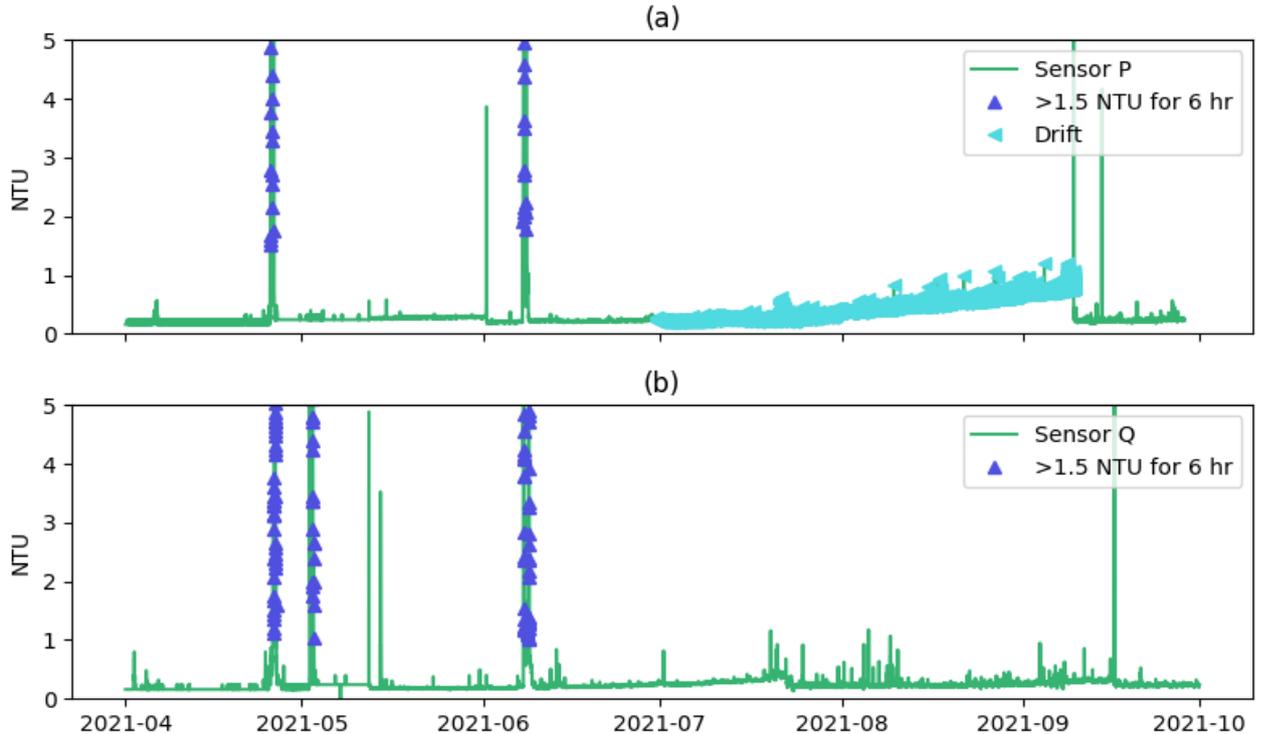


Figure 4.15. Multi-sensor data quality validation example, with (a) showing instances of periods above 1.5 NTU for 6 hours and a detected period of drift in Sensor A and (b) showing detected periods above 1.5 NTU for 6 hours in Sensor B which was calculated to be highly correlated to Sensor A.

4.5 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 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 in 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 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 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 rules-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 gives 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 to 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 latest sensor performance and demonstrating the near real time potential of this framework.

4.5.1 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 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 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/minute) 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. 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 rules-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 4.11 provide a visual impression of the data quality seen in each dataset. The rules-based approach allows for the prevalence of each specific feature to be compared across sensor locations and datasets. Time stamp 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 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.

4.5.2 Cross correlation, framework stage 2

The cross-correlation analysis was found to work well for chlorine time series data, due 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.

Conductivity and pH showed some promise, as did temperature. Patterns in temperature data occur and propagate within DWDS due to 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 in-common seasonal trends.

Performing the cross-correlation using the overlapping four-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 that the two-sensor location experience similar water at 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 valuable insight, but care must be taken in interpretation of meaning and 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 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 and 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 the 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 rate 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.

4.5.3 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 time stamp 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 cross-sensor validation is required. This idea is illustrated with the two case studies in Figure 4.14 and Figure 4.15. Figure 4.14 shows a chlorine sensor with detected anomalous periods of both high and low chlorine. Comparisons to connected sensors, known from cross-correlation results, shows that this anomalous behaviour is likely 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 4.15 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 rate 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.

4.6 Conclusions

- This work presents and demonstrates an effective multi-sensor data quality assessment framework that combines an automated single sensor rules-based data quality assessment and 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 near real-time.
- The rules-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.

Chapter 5: Algorithms to Mimic Human Interpretation of Turbidity Events from Drinking Water Distribution Systems

Reproduced from Gleeson K, Husband S, Gaffney J, Boxall J. Algorithms to Mimic Human Interpretation of Turbidity Events from Drinking Water Distribution Systems. *Submitted to IWA Journal of Hydroinformatics June 2023. Accepted September 2023.*

Declaration

Chapter 5 has completed peer-review for publication as a journal paper in IWA Journal of Hydroinformatics. The contribution of the authors are the following:

1. **Killian Gleeson** is the PhD candidate and first author of this chapter. Killian developed and led the labelling exercise, wrote the code for the automated algorithm, and wrote the first draft of the chapter.
2. **Dr. Stewart Husband** and **Prof. Joby Boxall** are Killian's academic supervisors, and **Dr. John Gaffney** is Killian's industrial supervisor, and all three are co-authors of this chapter. All three provided critical feedback and input into the thinking behind the labelling exercise and interpretation of the results.

5.1 Abstract

Deriving insight from the increasing volume of water quality time series data from drinking water distribution systems is complex and is usually situation and individual specific. This research used crowd-sourcing exercises involving groups of domain experts to identify features of interest within turbidity time series data from operational systems. The resulting labels provide insight and a novel benchmark against which algorithmic approaches to mimic the human interpretation could be evaluated. Reflection on the results of the labelling exercises resulted in the proposal of a turbidity event scale consisting of advisory <2 NTU, alert $2 < \text{NTU} < 4$ and alarm >4 NTU levels to inform utility response. Automation, for scale up, was designed to enable event detection within these categories, with the <2 NTU category being the most challenging. A time-based averaging approach, based on data at the same time of day, was found to be most effective for identifying these advisory events. Simple flat-line

event detection was sufficient to identify higher-level alert and alarm events. The automation of event detection and categorisation presented here provides the opportunity to gain actionable insight to safeguard drinking water quality from ageing infrastructure.

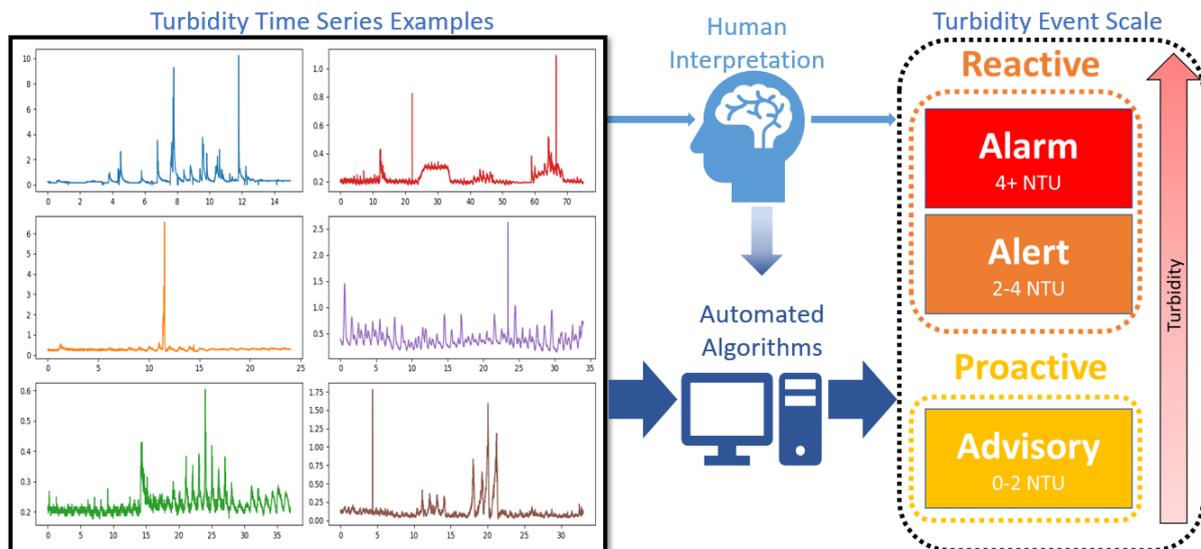


Figure 5.1. Chapter 5 graphical abstract.

5.2 Introduction

Continuous water quality monitoring within drinking water distribution systems (DWDS) enables network events to be captured and understood at a level of spatial and temporal detail that regulatory periodic discrete sampling cannot achieve. Causes of post-treatment DWDS water quality events range from hydraulic-induced mobilisation of pipe wall material (Husband et al. 2008), infrastructure failures allowing contaminant ingress (LeChevallier et al. 2003), to bulk water transformation such as excessive chlorine decay leaving no residual protection against contamination (Speight et al. 2019). A primary source of water quality-related customer contacts is discoloured water (DWI 2022) with turbidity sensors, using optics to measure the light scattering of water, considered a proxy measurement (Boxall and Saul 2005). Turbidity has also been shown to provide network specific correlation with iron and manganese (Cook et al. 2016), so also providing some insight into these parameters. Time series turbidity data taken from within DWDS is therefore of particular interest to operators who wish to understand and hence reduce the likelihood of discolouration events and customer contacts. Utilities are increasingly deploying turbidity sensors within DWDS, with the resulting datasets currently relying on manual interpretation that is reactive, subjective, situation specific and time consuming.

Data visualisation and interpretation is a powerful human skill due to the brains ability to subconsciously process visual information in as little as 13 milliseconds (Potter et al. 2014), significantly faster than text or numbers. An expert analyst can quickly identify and label periods of data of interest from interpreting graphical representations, yet the subjective nature limits the ability to cross compare. The sheer volume of data now being collected, along with the 24/7 nature of DWDS, makes reliance on such subjective manual assessment unviable, particularly as human brains can only accurately and quickly comprehend up to four variables at once (Halford et al. 2005). There is therefore a need to better understand the human process and to develop computing algorithms that can automate aspects of the interpretation and analysis of turbidity data to rapidly provide actionable information for operational decisions. The IWA's recent series of white papers on digital transformation (IWA 2022) stresses the need to move to more proactive infrastructure management and analysis of DWDS. Higher frequency turbidity time series data has the potential to enable this and improve our understanding of discolouration processes that will aid sustainable and safe delivery of high quality drinking water.

5.2.1 Background

Detection of interesting, undesired, or anomalous events in datasets is a widely studied and varied topic. The most common form is in detecting rare or unusual data points, often termed outliers or anomalies, by seeking deviations from assumed or modelled normality (Aggarwal 2016). Successful examples are found in network hacking, credit card fraud, and medical diagnostics (Aggarwal 2016). A review of anomaly detection techniques by Chandola et al. (2009) identified the nature of the available data and the type of event detection required as two key factors that dictate what methods are suitable. The availability of labelled data, an agreed designation where one or more labels identify properties, characteristics or classifications, opens an array of supervised machine learning approaches. These include support vector machines (SVM) and artificial neural networks (ANN) that can be more effective than unsupervised techniques as they use knowledge of known previous examples (Aggarwal 2016). Another important factor is the number of variables in a dataset, with significant research being done to detect anomalies in applications where high-dimensional datasets are the norm such as financial records and online interactions (Thudumu et al. 2020). When the data is in a time series, the temporal context of each dataset requires consideration, and detection methods rely either on a statistical or forecasted expected value, from which the

real values are compared and some sort of outlier score is determined (Gupta et al. 2014; Blázquez-García et al. 2020). The field of time series forecasting is of direct importance here, with ARIMA (autoregressive integrated moving average) and exponential smoothing two of the most popular approaches (Hyndman and Athanasopoulos 2021). Important considerations are the quantity of data used to determine a forecast and the forecast horizon. ARIMA models utilise the autocorrelations in a time series in order to make forecasts (Hillmer and Tiao 1982), while exponential smoothing gives greater importance to more recent data and has been adapted to account for trend and seasonality (Blázquez-García et al. 2020). Seasonality in time series data can refer to patterns occurring on a repeated periodic basis, such as yearly, monthly, weekly, or daily. Seasonality is relevant to DWDS time series due to the strong links to seasonal weather and human behaviour patterns. SARIMA (seasonal ARIMA) is a modification of ARIMA that is capable of accounting for seasonality while VAR (vector autoregression) and ARIMAX (X representing exogenous variables) models are adaptations that can consider additional variables. The ETS (error, trend, seasonal) framework describes nine exponential smoothing variations, based on how the error, trend and seasonal components are calculated and combined (Hyndman and Athanasopoulos 2021). Recent advances in time series forecasting include the use of neural networks, with LSTM (long short-term memory) particularly popular for supervised multi-variate time series forecasting (Hochreiter and Schmidhuber 1997), and Prophet, which can be applied automatically and considers holiday effects (Taylor and Letham 2018).

Research on detecting events in DWDS has been dominated by leakage detection methods, most commonly looking for unusual patterns in acoustic or pressure sensor data (El-Zahab and Zayed 2019). Detection of water quality events within DWDS has not attracted as much attention, but research has been done to detect intentional contamination of DWDS by the US Environmental Protection Agency (EPA), who produced an open source event detection software package called CANARY, which consists of various different statistical algorithms to detect outlier values based on rolling window statistics, from which an event probability is calculated for each window using a Binomial Event Discriminator (BED) (McKenna et al. 2007). The use of rolling windows is a common way to account for the temporal context in time series. CANARY has been applied to DWDS data in the UK, where it has shown promise in detecting multi-parameter events (S. Mounce et al. 2012). The difficulty of linking detected events to confirmed real world actions is highlighted by this research, where only 28% of detected events could be linked either to customer contacts or hydraulic disturbances.

Labelled data in DWDS are uncommon and the process of linking data to information from network operations or customer interactions is time-consuming. Additionally, deciding what constitutes a water quality event is not clear cut, meaning any labels cannot be considered ground truth. Crowd-sourced labels are commonly used in machine learning and research has been done to understand how to deal with inevitable human error (Ustalov et al. 2021) with strategies that include multiple labellers per example. However, the labels used are generally definitive, such as whether a picture contains a cat or a dog, and little research has been done to understand how labels can be combined in cases where the question posed is highly subjective.

When developing methods for analysing events in turbidity time series, it is important to first understand the nature of turbidity data and the desired events to detect and study. This is not a trivial challenge. Depending on turbidity event definition, these may occur frequently or as unique incidents and are linked to network and sensor installation location. In the UK, legislation dictates that the water at customers taps should not exceed 4 NTU, nor 1 NTU exiting treatment works (DWI 2018). Therefore, network turbidity sensors recording values more than 1 NTU are evidence of in-transit deterioration, and this may represent actionable information. In reality, turbidity levels leaving treatment works are generally much lower than 1 NTU, with less than 0.01 % of regulatory turbidity samples exiting treatment works exceeding 1 NTU in 2021 (DWI 2022). Therefore, even turbidity events occurring below 1 NTU may relate to variations in discolouration risk and also be worthy of identification and study. Yet analysis of DWDS turbidity time series data has tended to focus on reacting to larger events, meaning the information at lower turbidity levels has remained unused. Computing and modern analysis techniques however offer the potential to rapidly analyse lower-level turbidity data but require specific instructions which are currently not well understood.

The aim of this research was to explore and to improve understanding of what constitutes an event worthy of further consideration in turbidity time series data and then to develop and assess automated computing algorithms that can rapidly review and identify such events, mimicking human judgements and intuitive extrapolation to inform both reactive and proactive utility responses.

5.3 Method

5.3.1 Methodology

The difficulty and subjectivity of linking turbidity data with real-world evidence of water quality deterioration led to a crowd-sourcing approach being taken that involved a time series labelling exercise, with domain experts being tasked to label what they considered to be events of interest within turbidity time series examples. This approach takes advantage of human brain power, which computer algorithms can only approximate when given specific instructions. To overcome the problem of bias and subjectivity, the same time series examples were shown to different groups at different meetings, with each of the resulting Boolean labelled time series combined and averaged. This averaging of results for each turbidity datapoint returned an associated ‘label average’ score, between 0 and 1. This value could then be used as a benchmark to evaluate the suitability of algorithmic approaches such as flat-line detection and the calculation of event score time series of similar form to the averaged labels. The labelled data would also inform whether a single approach can handle different event types or whether a combination of approaches is more suitable.

5.3.2 Event Labelling Exercise

An interactive labelling exercise was compiled using the open-source browser-based time series labelling tool Trainset (Geocene 2020), to enable users to label six turbidity time series examples. The examples were selected after reviewing approximately 100 turbidity sensors across 4 different UK DWDS, which combine to a rough total of 150 years’ worth of turbidity time series data. As this research is interested in both reactive and proactive aspects of turbidity events, six turbidity examples were selected to represent a range of different event types and magnitudes that were observed in the wider datasets. Each turbidity example had been quality assured according to the procedures set out in Chapter 4. Different durations were used to represent realistic but manageable sensor deployment time frames, ranging from 2 to 10 weeks. Care was taken to ensure the participants were not aware of the reasoning behind the examples. Six time-series of 16 to 75 days in duration was considered to be a safe limit to ensure the human experts maintained a high level of focus and attention to detail. Additionally, limiting this exercise to just examples meant it took roughly 10 minutes to complete, which was considered a realistic expectation of participants. A screenshot of the event labelling software with example number 4 is shown in Figure 5.2 where the pink

highlighted data is an example of what user-labelled data looks like. Example 4 is unique out of the six time series turbidity datasets in that it was artificially constructed (by splicing and combining data) to represent some different theoretical types of turbidity event; (1) a hydraulic-induced material mobilisation event, (2) a single point event, (3) a baseline-change event, and (4) an increase in diurnal turbidity event. These event types are marked in Figure 5.2, where event (5) is a combination of the first four. The other examples are all unedited turbidity time series from different UK DWDS (displayed in Figure 5.4). To ensure consistency between data from multiple sources and reporting intervals, all examples were resampled to a 15-minute sampling interval.

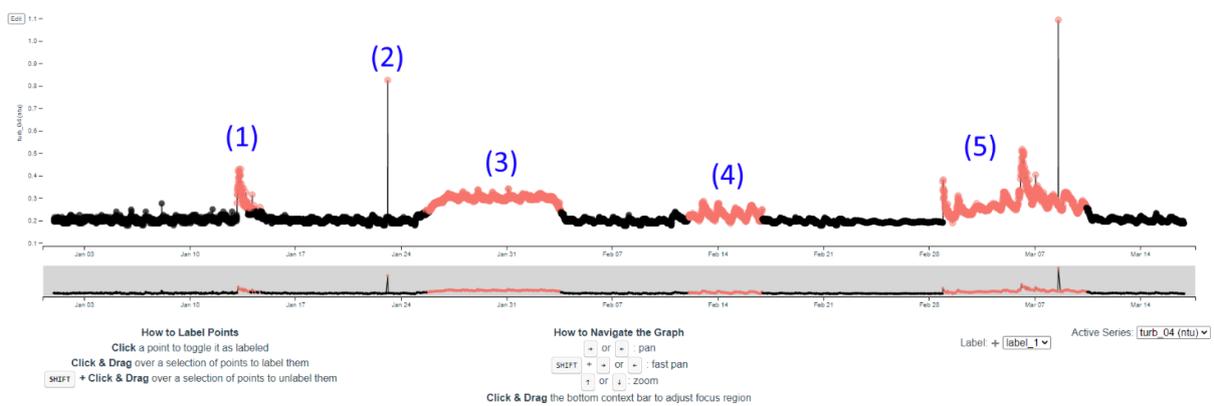


Figure 5.2. Screenshot of turbidity example 4 in event labelling tool, with the theoretical event types highlighted.

The labelling exercise was run across multiple sessions with anonymity retained and participant consent required to confirm they understood what they were taking part in before they could proceed to the labelling interface. Upon completing the labelling exercise, users were directed to an upload page on a dedicated webpage, which had an upload button that anonymously uploaded the labelled data to a dedicated server folder. For the labelling sessions of the exercise, users were simply given the instruction to ‘label events’ with the following provided as an event definition:

“An event is described as a noteworthy period of data to be flagged for further consideration”

5.3.3 Event Score Calculation

Each algorithmic approach involved first making a forecast (other than flat-line), which is then subtracted from the turbidity data to obtain a residual time series. The next step was to

transform the residual into an event score time series, which could be compared to the averaged-out labels. This transformation was achieved using a sigmoid function. To compare different time series forecasting methods, the sigmoid function was optimised to find the lowest error against the labels, for each residual calculated. Each approach involved adjustable parameters, which were investigated in a sensitivity analysis with the goal of determining the combination that most closely captured the information gained from the labelled datasets.

5.3.3.1 Forecasting Methods

The different methods used to make a forecast, from which a residual was calculated, are listed in Table 5.1. All approaches were employed for sliding windows of 24, 48 and 72 hours, as well as expanding windows for forecast horizons of single point and between 2-72 hours ahead. CANARY was an exception as it only produced next step forecasts and does not include expanding windows. However, the remaining approaches all share window and forecast horizon parameters, with all method-specific adjustable parameters listed in Table 5.1. Averaging methods were based on using data within the specified window, with different quantile levels examined, as well as mean values. The time-based average method represented a deviation from the typical sliding window approach. Instead of using a window directly preceding each datapoint, this method looks at previous data at the same time of day, accounting for the diurnal patterns often seen in DWDS data that is heavily linked to human behaviour. New adjustable parameters were introduced here, the size of window to include each day (e.g. for a datapoint at 8:30 am, a 2 hour window would mean any data between 7:30 am and 9:30 am would be included) and the averaging method used. The averaging and time-based averaging approaches were developed in Python using the Pandas (McKinney 2010) library.

ARIMA has three input parameters: the lag order (p), the degree of differencing (d), and the order of moving average (q) [16]. These parameters make up the order, often shown in the form: (p, d, q). SARIMA also has seasonal ordering parameters P, D, Q , and m , where m is the seasonal period. Wherever the seasonal period was a possible option, 96 represent the diurnal patterns that turbidity time series can exhibit as this is how many samples were in a day (at 15-minute sampling rates). Exponential smoothing methods were explored using the ETS framework that looked at the impacts of different error, trend, and seasonal component calculations. Each component can be either additive or multiplicative. An exponential

weighted mean (EWM) approach was also included, which requires the decay to be specified, either in terms of centre of mass, span, half-life, or as a smoothing factor. Other methods investigated were Prophet and CANARY. Prophet was run using both its automatic functionality, and for different growth methods and seasonality modes. The ARIMA, SARIMA, ETS, and Prophet approaches were developed using the machine learning for time series interface library sktime (Loning et al. 2019). The CANARY software was run and included in this analysis, using the linear prediction correction filter (LPCF) method and BED. The alternative multi-variate nearest neighbour (MVNN) method is not applicable to this univariate problem. LPCF uses the MATLAB filter and lpc functions to estimate the next value based on weighted filter applied to a window of normalised data proceeding each datapoint (Murray and Haxton 2010). The user needs to specify window size and the threshold, in standard deviations, above which is considered an outlier. Window sizes between 1 and 72 hours were included in the sensitivity analysis while standard deviation thresholds were looked at between 0.5 and 1.5. Parameters event timeout, the number of timesteps after an event is found before alarm is silenced automatically, and event window save, a parameter related to plotting identified events, were not adjusted as this research was more interested in the residual calculated. Table 5.1 lists the adjusted parameters for each forecasting approach, aside from the window and forecast horizons.

Table 5.1. Forecasting methods and associated adjusted parameters.

Approach	Variants (number of adjusted parameters)	Adjusted Parameters
Averaging	Mean (0), median (0), quantiles (1)	quantile value
Time-based Averaging	Mean (1), median (1), quantiles (2)	Window size (hours), averaging method, quantile value
ARIMA-based	ARIMA (3) SARIMA (6)	p, d, q (ARIMA) p, d, q, P, D, Q (SARIMA)
Exponential Smoothing	ETS (4) EWM (1)	Error, trend, seasonal, damped trend (ETS) Alpha (EWM)
Prophet	Auto and with settings (3)	Growth method, growth cap (if method is logistic), seasonality mode
CANARY	CANARY LPCF (1)	Outlier threshold

5.3.3.2 Event Score and Comparison to Labels

After residuals were calculated for each of the forecast methods, the sigmoid function was used to transform the residuals into event scores. The sigmoid function maps inputs to outputs between 0 and 1 using the following Equation (1).

$$y = \frac{1}{(1 + e^{-c*(x-b)})} \quad (1)$$

Where, x = input data, b = sigmoid centre point, c = sigmoidal width. The two sigmoid parameters were optimised to minimise the error against the labelled data using SciPy's optimisation function (Virtanen et al. 2020). The optimisation was done with all examples and using just 3-6, so without examples 1 and 2. Examples 1 and 2 contain significant large-scale events exceeding 4 NTU and were excluded from one optimisation to investigate what approaches work best for lower-level turbidity events, in this case with all data below 2 NTU. RMSE (root mean squared error) was used to evaluate each approach. The root mean squared error (RMSE) is a commonly used forecasting metric, and is represented by Equation (2):

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (x_j - y_j)^2}{n}} \quad (2)$$

The RMSE is known for being sensitive to outliers or large errors (Chai and Draxler 2014). As all these values are between 0 and 1, the largest possible error is when the labels are 1 and the event detection system is 0 (or vice versa). For this research, it is desirable to punish these outcomes, so the RMSE is a suitable performance metric. To include more forecasting methods, windows, and horizons, the first 3 days of each example was omitted when calculating the RMSE. Methods such as ETS require two full cycles of data to account for seasonality. Other approaches such as the time-based averaging require at least 1 day of data, while some methods worked best for forecast horizons of 24-48 hours. This also handles the 'cold start' problem many forecasting methods have, where it is very difficult to make predictions without any prior data.

For the CANARY LPCF, the BED approach was used in addition to the sigmoidal approach already outlined. Since BED requires a Boolean input, this could not be used for other residuals without adding an additional outlier threshold step, which risks losing complexity and adds an unnecessary additional input. BED uses probability theory to estimate event probability for each datapoint, based on the number of outliers present within a specified window. BED takes two input parameters, window size and outlier probability. Outlier

probability is a probability threshold above which events are counted. Since this research is only interested in the probability score, the probability threshold is not needed. The CANARY manual recommends using BED windows between 4 and 18 timesteps (between 1 and 4.5 hours for 15-minute data) so this was the range examined in the sensitivity analysis.

5.4 Results

5.4.1 Labelled Results

The turbidity time series labelling exercise was run four times during different academic and industry events, with a total of 48 participants returning complete labelled data. Session 1 took place during an online meeting by a university research group who focus on DWDS. Session 2 took place during a water utility-academia event, focusing on discolouration in DWDS, with 12 UK utilities represented. Session 3 was run independently by a water utility’s network modelling team. The final session was run during a separate water utility-supply chain-academic (industry dominated) event which focuses specifically on water quality within DWDS, with 16 different water utilities and at least 8 supply chain companies present. Aside from session 3, these exercises were run during academic and industry meetings that focussed on discolouration and water quality issues in distribution systems, meaning the participants were not selected specifically for the purpose of participating in this exercise. The labelling exercise sessions are summarised in Table 5.2. Figure 5.3 shows boxplots of the total percentage labelled data points from each session, illustrating the variety in responses within and across sessions. Session 3 stands out as having the lowest amount of labelled data. This session was run externally, without the authors of this research in attendance. A stricter definition of what constitutes an event was used, with attendees focussing on significant events, potentially of regulatory concern. This highlights the challenges in defining what constitutes an event of interest, or with respect to this research, what information is required to inform what decisions from the data and the impact of differing instructions and perspectives. The session 3 results were included in the subsequent analysis as the perspectives of session 3 participants were considered equally valid. This enabled the inclusion of a wide range of different expert opinion.

Table 5.2. Event Labelling Sessions.

Session	Event	Valid Labelled Datasets
1	University Research Group	8
2	Water Utility-Academia Event	9

3	UK Water Utility	12
4	Water Utility-Supply Chain-Academia Event	19

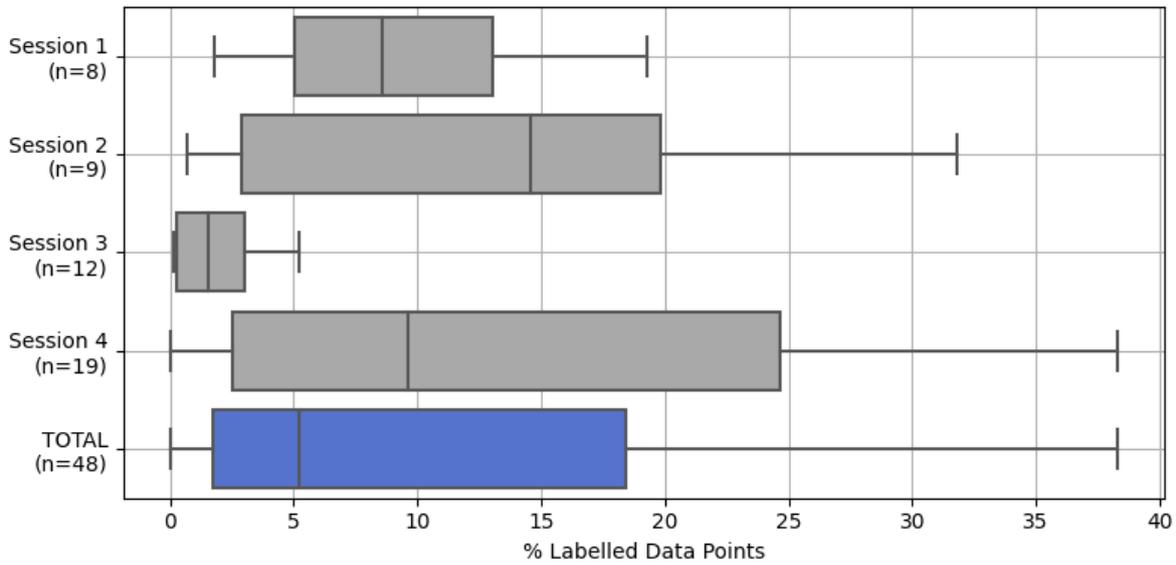


Figure 5.3. Boxplot of total percentage labelled data across all six example datasets per session.

Figure 5.4 plots each turbidity time series example along with the labelling results, averaged out for each datapoint. The value of each data-point indicates the fraction of participants that deemed it to be noteworthy. These event score time series provide a useful way to interpret the labelling results and a benchmark to evaluate algorithmic approaches against. One of the challenges of human analysis is inherent bias towards higher turbidity events. In these time-series, examples 3-6 had little data above 2 NTU, meaning that the lower-level events were more easily visualised than in examples 1 and 2. Figure 5.5 is a scatter plot showing the average absolute turbidity values for each averaged out label, divided into those from examples 1 and 2 with significant higher turbidity events, and those from examples 3-6. This illustrates the impact of bias on the presence of higher turbidity events and how the human participants then interpreted the lower-level turbidity data present in the same dataset. In order to examine automated analysis methods that would work well at analysis of lower-level turbidity events, the data from examples 3-6 were treated separately. To distinguish the low-level events, a threshold of 2 NTU was identified and analysis of events below this are termed ‘advisory’. At the same time events exceeding the regulatory value at customer taps of 4 NTU, and therefore typically requiring immediate attention, are considered ‘alarm’. Between these levels, events are considered ‘alert’, representing significant deterioration compared to the maximum permitted turbidity of 1 NTU leaving a treatment works. This turbidity event scale naming convention and the boundaries are summarised in Table 5.3.

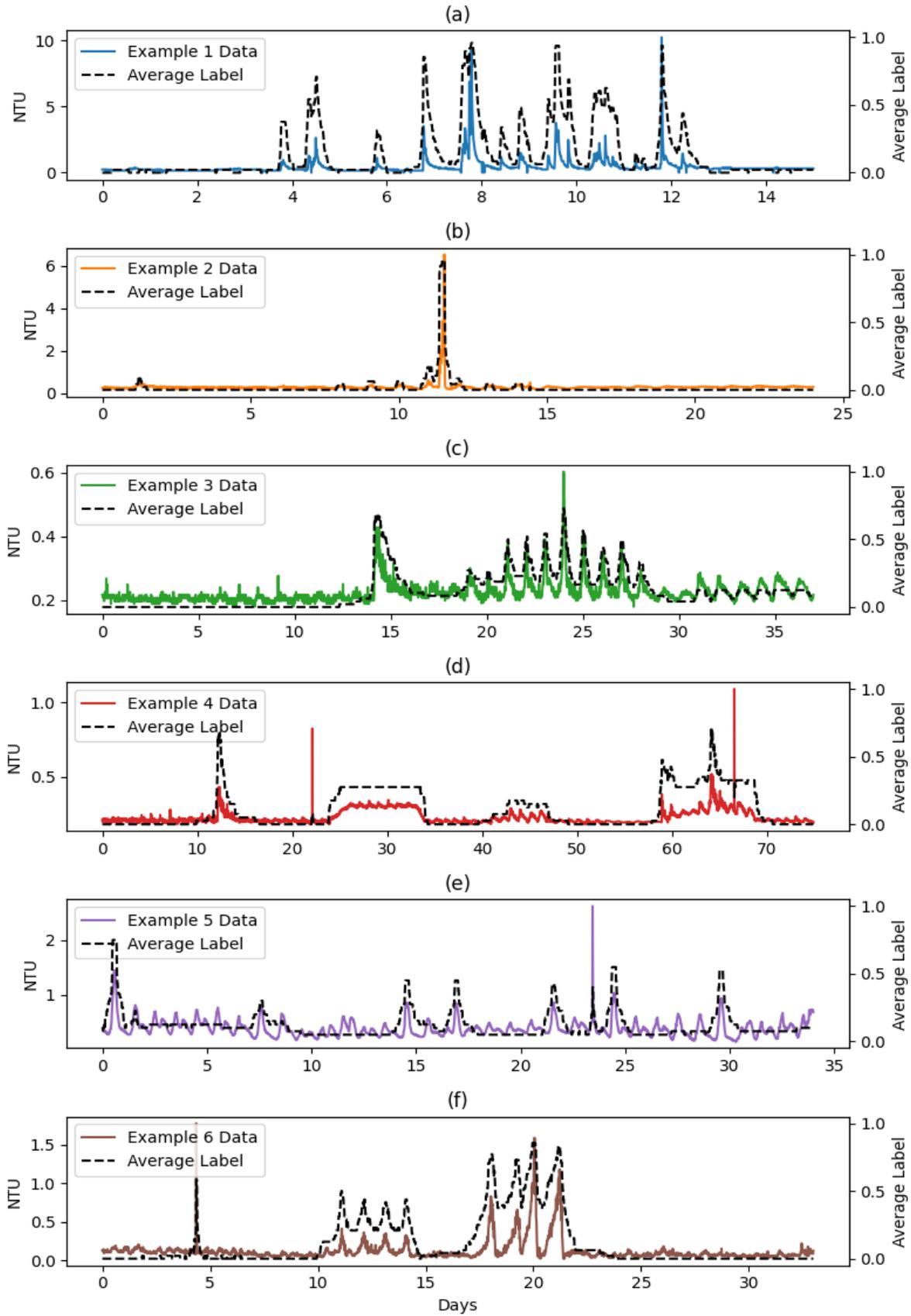


Figure 5.4. Six turbidity time series examples, in plots (a) to (f) respectively, along with corresponding average label results (the x-axis and NTU y-axis ranges are scaled for each dataset).

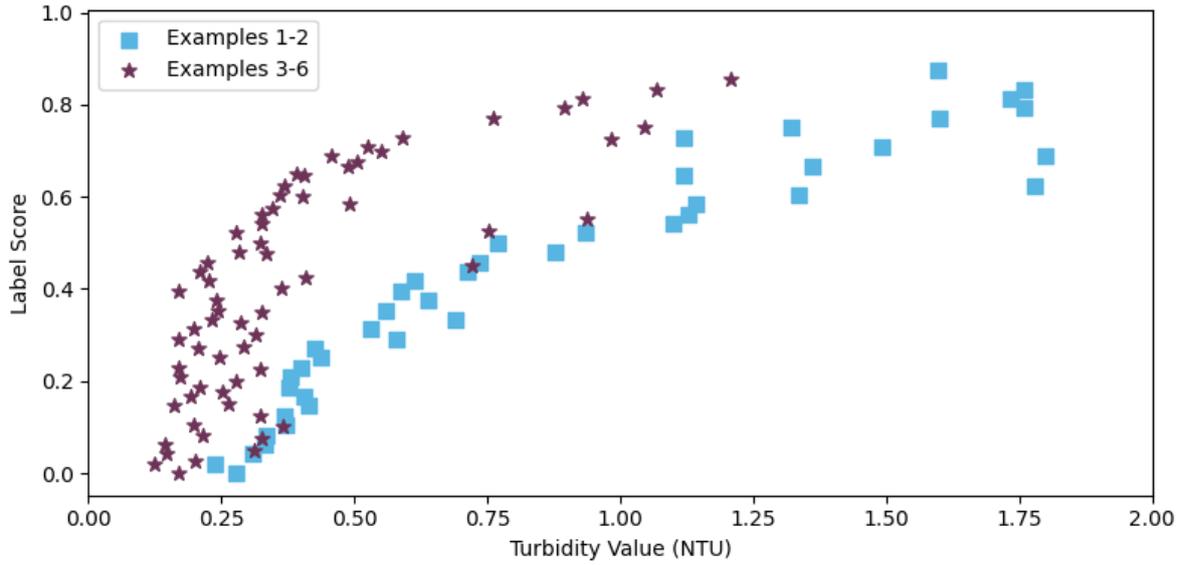


Figure 5.5. Mean absolute turbidity value for each average label value; for Examples 1 & 2 (blue squares) and Examples 3-6 (purple stars).

Table 5.3. Turbidity Event Scale.

Event Type	Turbidity Limits
Advisory	< 2 NTU
Alert	2 < NTU < 4
Alarm	>4 NTU

5.4.2 Event Analysis Results

Using the event scaling outlined in Table 5.3, this research explored whether a single algorithm could deliver all three levels, or whether combinations were required. Such as the use of simple flat-line approaches to identify and separate alert and alarm events for reactive response, and tune more sensitive algorithms with the ability to accurately identify lower-level events that could inform proactive measures. Methods that output event score time series between 0 and 1, like that of the averaged-out labels, were therefore developed and tuned for all 6 examples (all 3 event categories), and to advisory events only, examples 3-6.

5.4.2.1 Flatlines

Flat-line algorithms to detect alert and alarm events are considered separately here as they do not require the process of prediction and residual calculation. These flatline thresholds are shown applied to Example 1 in Figure 5.6 below. This simple approach was effective at identifying both alarm and alert events. Attempts to use a flat line approach for identification of advisory events was very poor. Unlike alert and alarm events, applying a flat-line at lower turbidity levels would make detection strongly dependent on background turbidity levels. For example, applying a 0.5 NTU flat-line would result in 1 advisory event in Example 3 and 39 in Example 5, due to Example 5 having higher background turbidity. Therefore, analysis of advisory events required consideration of background turbidity which the calculation of residual values achieved, prior to conversion to event scores.

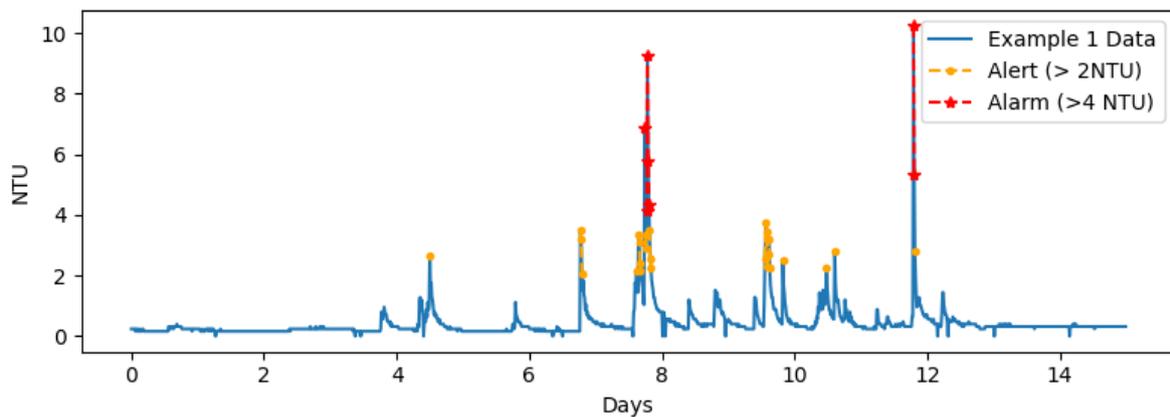


Figure 5.6. Flatline thresholds differentiating Alert (2-4 NTU) and Alarm (>4 NTU) events in Example 1.

5.4.2.2 Calculating Event Scores

Figure 5.7 illustrates the application of the ARIMA algorithm as an example of the approach adopted. The top plot shows an expanding window ARIMA forecast for Example 1, while the bottom plot shows an expanding time-based averaging forecast for Example 3. Figure 5.8 illustrates how the obtained residuals were then transformed into an event score time series. A sigmoid function was used to understand and optimise the relationship between the residual values and the labelled data. The RHS plots show the outputs of the sigmoid function compared to the labels when averaged, demonstrating how the obtained human interpretation can be mimicked using this sigmoidal function.

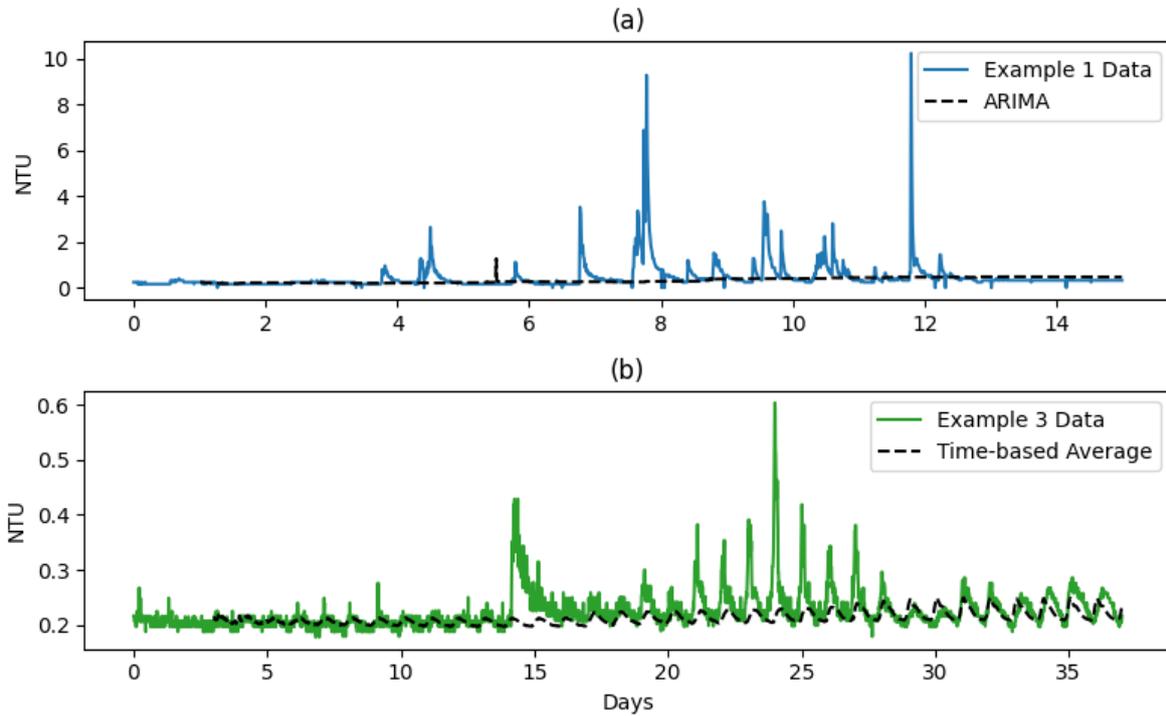


Figure 5.7. Example 1 with expanding ARIMA forecast values (a) and Example 3 with expanding time-based average forecast values (b).

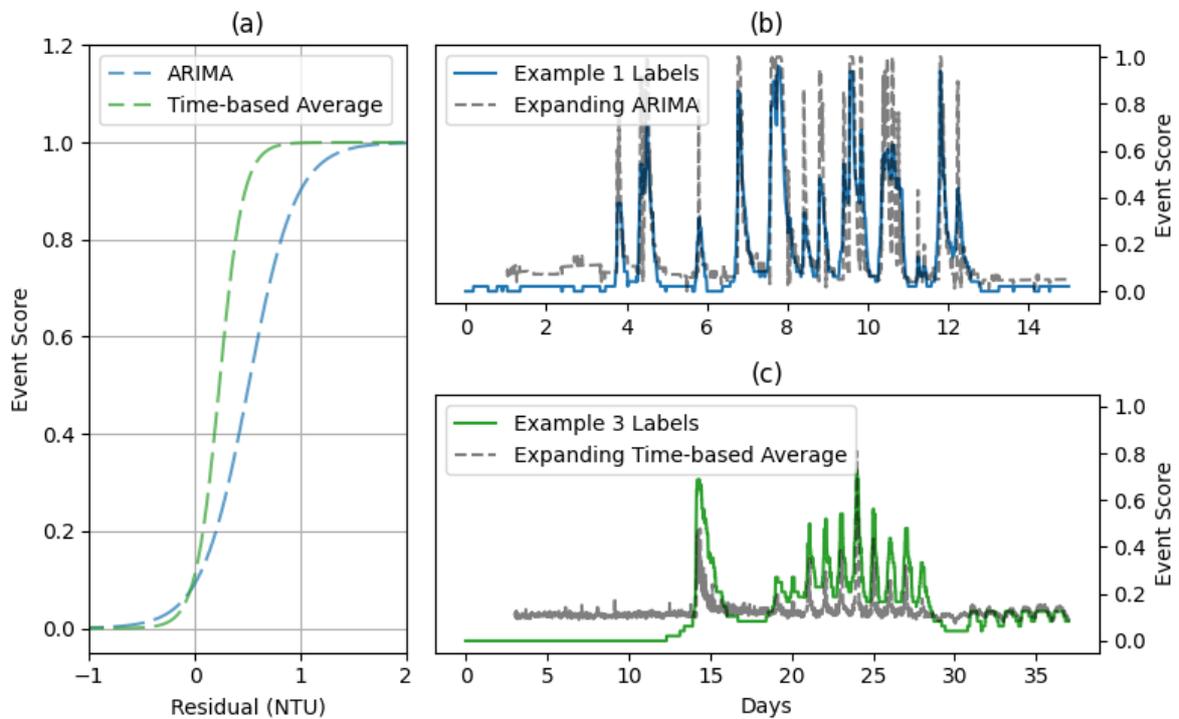


Figure 5.8. Optimised sigmoid function for ARIMA tuned on all events and time-based (tb) average method tuned on advisory events (a), with corresponding outputs compared to Example 1 and 3 labels, (b) and (c) respectively.

5.4.2.2.1 Forecasting Methods

Each approach (other than flat-lines) included in this research had its calculated residual optimised to minimise errors against the averaged-out labels, hence the forecasting methods can be compared to each other using the optimised error values. The solution with the lowest RMSE for both ‘all events’ and for ‘advisory only events’, for averaging and time-based averaging approaches, is shown in Table 5.4. An expanding window had better results than any sliding window approach, while forecast horizons of 12 and 24 hours were best for averaging, for ‘all events’ and ‘advisory only events’ respectively. For time-based averaging, forecast horizons can only be in multiples of days, with 1 and 3 days found to work well for ‘all events’ and ‘advisory only events’, respectively. Daily window sizes of 6 and 3 hours were found to work better than exact time-based values, showing that including data before and after each timestamp was useful.

The optimal parameters for ARIMA and SARIMA are displayed in Table 5.5. An expanding-type window in combination with 24-hour forecast horizon worked the best. For ARIMA the most useful order was (1,0,0) which represents a first-order autoregressive model without any differencing or moving averaging. SARIMA was found to be the most time-consuming approach, meaning not all possible combinations were completed. Of those that were, the best approach had an order of (1,0,0)x(0,0,0) meaning no seasonality terms were employed, suggesting the (1,0,0) order ARIMA was adequate. The optimal parameters for the exponential smoothing approaches are listed in Table 5.6. Using a half-life of 14-days worked well for EWM, while the optimal solutions found using ETS both involved no trend or seasonal components, meaning simple exponential smoothing was found to work best. The optimal parameters for Prophet are shown in Table 5.7. For Prophet the same residual was found to be the best solution for all events and for only advisory events. Even the sigmoid function parameters are like each other. For LPCF, parameters are shown in Table 5.8 for solutions using both the optimised sigmoid approach and CANARY’s BED function. CANARY does not allow expanding windows, nor does it include forecast horizons other than single point. The optimised sigmoid parameters have noticeable different parameters to other methods due to the larger magnitudes seen in CANARY LPCF residual time series. When using BED, the best solution for all event levels and advisory event levels were identical.

Table 5.4. Optimal parameters found for averaging and time-based averaging.

	Averaging		Time-based averaging	
	Tuned on all events	Tuned on advisory only events	Tuned on all events	Tuned on advisory only events
Window	expanding	expanding	expanding	expanding
Forecast Horizon	12-hour	24-hour	1 day	3 day
Other parameters	averaging method = mean	quantile=0.8	daily window size = 6-hour, averaging method = median	daily window size = 3-hour, averaging method=mean
Sigmoid Parameters	b = 0.51, c = 4.47	b = 0.24, c = 7.72	b = 0.59, c = 4.04	b = 0.23, c = 8.98

Table 5.5. Optimal parameters found for ARIMA and SARIMA.

	ARIMA		SARIMA	
	Tuned on all events	Tuned on advisory only events	Tuned on all events	Tuned on advisory only events
Window	expanding	expanding	expanding	expanding
Forecast Horizon	24-hour	24-hour	24-hour	24-hour
Other parameters	order = (1,0,0)	order = (1,0,0)	order = (1,0,0) x (0,0,0)	order = (1,0,0) x (0,0,0)
Sigmoid Parameters	b = 0.51, c = 4.50	b = 0.32, c = 6.55	b = 0.51, c = 4.49	b = 0.31, c = 6.57

Table 5.6. Optimal parameters found for EWM and ETS.

	EWM		ETS	
	Tuned on all events	Tuned on advisory only events	Tuned on all events	Tuned on advisory only events
Window	expanding	expanding	24-hour	24-hour
Forecast Horizon	24-hour	24-hour	12-hour	48-hour
Other parameters	half-life = 14 days	half-life = 14 days	error=additive, trend=None, damped=False, seasonal=None	error=multiplicative, trend=None, damped=False, seasonal=None
Sigmoid Parameters	b = 0.52, c = 4.42	b = 0.34, c = 5.98	b = 0.63, c = 3.89	b = 0.41, c = 5.25

Table 5.7. Optimal parameters found for Prophet.

	Prophet	
	Tuned on all events	Tuned on advisory only events
Window	expanding	expanding
Forecast Horizon	24-hour	24-hour
Other parameters	growth=logistic, growth cap=0.5, daily seasonality=False, seasonality mode=additive	growth=logistic, growth cap=0.5, daily seasonality=False, seasonality mode=additive
Sigmoid Parameters	b = 0.55, c = 4.17	b = 0.48, c = 4.42

Table 5.8. Optimal parameters for LPCF using sigmoid function and BED.

	LPCF		LPCF + BED	
	Tuned on all events	Tuned on advisory only events	Best for all events	Best for advisory only events
Window	48-hour	9-hour	72-hour	72-hour
Forecast Horizon	N/A	N/A	N/A	N/A
Other parameters	outlier threshold = 0.5	outlier threshold = 1.0	outlier threshold = 1.5	outlier threshold = 1.5
Sigmoid (or BED) Parameters	b = 6.41, c = 0.36	b = 11.30, c = 0.19	BED window = 4	BED window = 4

5.4.2.2.2 Comparison to Labels

Figure 5.9 plots the lowest RMSE found for each of the forecasting methods investigated. The methods were tuned for lowest RMSE and assessed both for all event levels (using all 6 examples) and for advisory events only (using examples 3-6). The residuals from CANARY’s LPCF algorithm were passed through the same sigmoidal function, as well as using CANARY’s in-built BED function (though this was not tuned in the same way as the sigmoid approach). An expanding ARIMA approach with a (1,0,0) order and 24-hour forecast horizon resulted in the lowest RMSE of 0.1137 across all examples. The second-best approach across all examples was the simple averaging (RMSE of 0.1140) though there were ten different ARIMA combinations that resulted in an overall RMSE of 0.1140 or less, including a (1,0,0) order at a 48-hour horizon, (0,0,0) order, which represents white noise, with shorter 6 and 12-hour horizons, and orders (0,0,q) with q=1,2,3 for forecast horizons of 6 and 12-hours. This suggests ARIMA has strong applicability to this research and that

including autoregressive or moving average terms can be useful for calculating the residual and subsequent event score time series. The time-based averaging approach worked the best for advisory events with a RMSE of 0.1095, but this approach did not work effectively at generalising across all event types and performed worse than the averaging approach for all examples. Figure 5.10 shows this approach applied to data from example 2, before and after a larger alarm type event, and compared to the averaged-out labels in example 2, with the turbidity data clipped to exclude the larger 6 NTU event. The averaged-out label, shown in the context of this NTU y-axis scale, did not have any score above 0.2 outside of this larger alarm event. By contrast the time-based average approach, tuned on advisory data, is not biased by the presence of the alarm event, and returned event scores of increasing magnitude in the days leading up to the alarm event, with a value of 0.7 seen about half a day before the alarm event occurred. This demonstrates the promise of this approach in being part of a proactive management system.

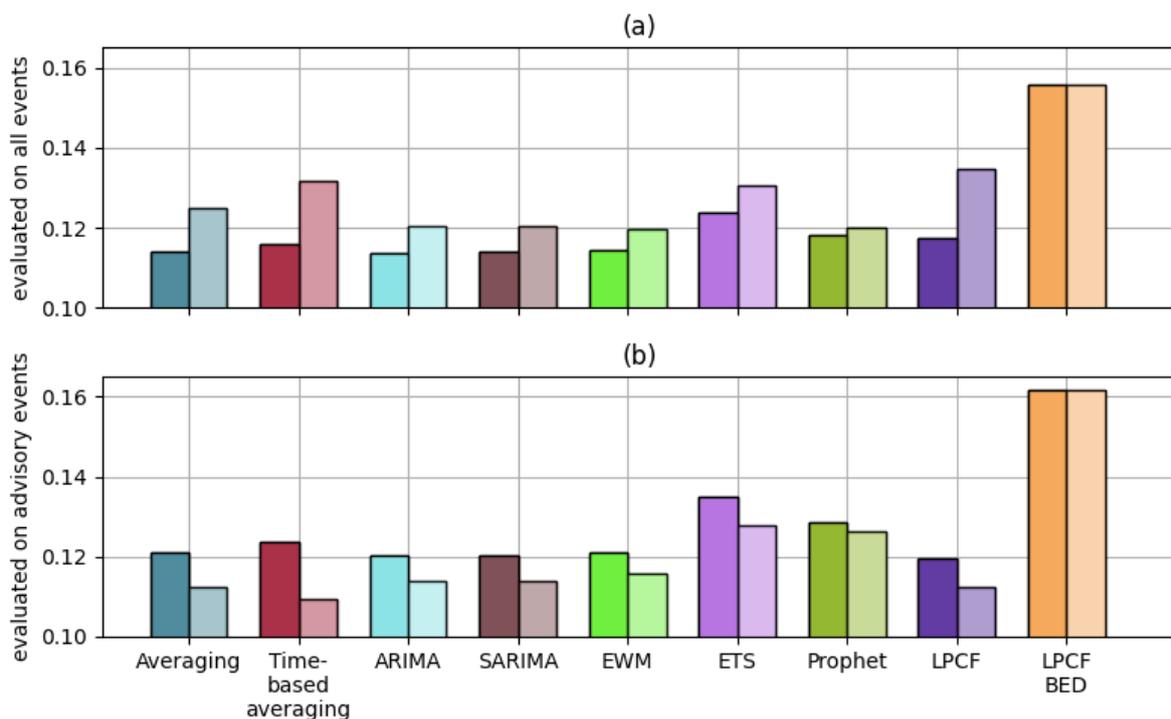


Figure 5.9. Lowest RMSE for each type of forecasting methods, divided into tuning for all event types (darker shades) and advisory only events (lighter shades) and evaluated on all event types (a) and advisory events (b). Note that the y-axis is clipped from 0.1 for visual interpretation.

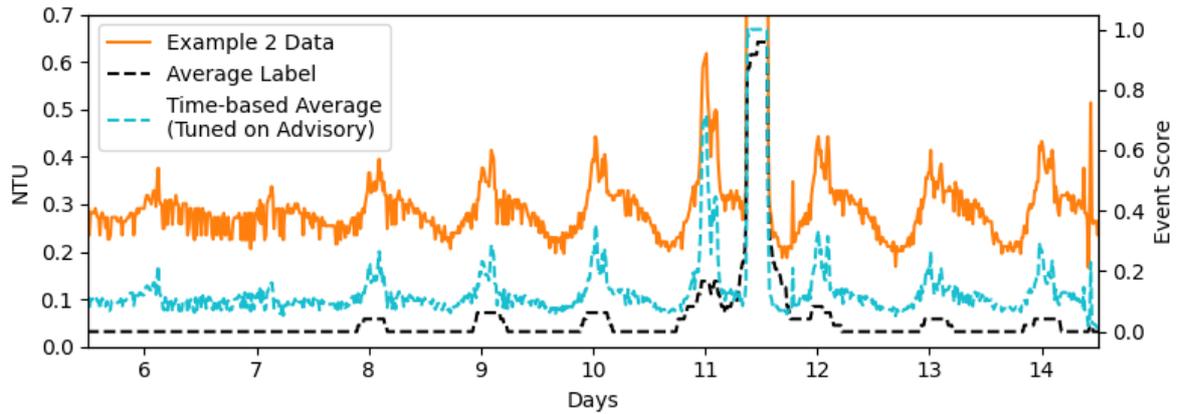


Figure 5.10. Example 2 turbidity data (clipped exclude the alarm event on day 11) and averaged-out labels versus sigmoid output using time-based averaging method, optimised for advisory events.

5.5 Discussion

This research presents an evaluation of approaches to analyse and understand events in DWDS turbidity time series data and uses a crowd-sourced labelled dataset as a benchmark. This evaluation process presents an alternative approach to overcome the difficulty of linking turbidity data with confirmed real-world events. With six turbidity examples and 48 participants in total, but covering a wide range of companies/organisations, conclusions should be considered with this relatively small sample size in mind. The number of examples included was limited by how long these sessions could be run, while obtaining more participants would be challenging without reducing the level of domain expertise. Following reflection on the results of the four labelling exercises, a three-level turbidity event scale was defined as advisory, alert and alarm (Table 5.3). The presence of alert and alarm turbidity events in examples 1 and 2 impacted labellers interpretation of advisory events. Such advisory events were easier to interpret in examples 3-6, with participants considering many noteworthy. As highlighted in the background section, even smaller turbidity events with responses < 1 NTU can suggest in-network deterioration and may provide valuable precursor information about levels of discolouration risk within DWDS. This research identified and was subsequently able to focus on proactive discolouration approaches as well as reactive measures by differentiating through consideration of alert and alarm events. Flatline alert and alarms at 2 and 4 NTU are reliant on the turbidity sensor accuracy and overall data quality. Lower flatline approaches would be too dependent on sensor calibration accuracy and

background turbidity. The residual calculations performed in this research are however agnostic to turbidity baseline values, instead looking for increases compared to recent data. Therefore, these approaches are not reliant on sensor calibration accuracy. However, a prior data quality assessment is required, particularly to remove sensor errors such as drift that may occur in the data which, if left, could interfere with the residual calculations. This was performed on the examples in this research using a set of data quality assessment rules previously developed in Chapter 4.

The labelling and evaluation process to mimic human interpretation is something that could be repeated for different parameters, or for turbidity with additional contextual information, such as flow rate data or customer contacts. Additionally, it could be run for specific teams or companies, with the aim to develop a solution that best matches their requirements and collective intuition. This highlights the need to clearly understand what information or insight is required prior to developing automation techniques, and the need to match such techniques to the data and insight sought. The results show that many residual calculation approaches, using both statistical averaging and time series forecasting, can be used in combination with a sigmoid function to produce an event score time series. Method selection may therefore come down to decisions about computational power, processing time and number of adjustable parameters. The event score time series output can form an event detection system or can be used to understand network conditions or performance. This understanding could then be applied to compare performance across networks or allow temporal analysis to detect changing performance if mobile monitors are deployed on a shorter term but repeating basis.

5.5.1 Human Interpretation of Turbidity Events

Participants were asked to highlight “noteworthy periods of data to be flagged for further consideration”. The exercise did not provide any additional contextual information, such as sensor location or other supporting information such as flow rate data or connected sensors also measuring turbidity or other water quality parameters. Such information is important when further analysing turbidity events in DWDS but were outside of the scope of this specific research for identifying data of interest for such further interpretation. The experience of visualising each example, one by one, within the trainset application told a story and may have influenced the participants interpretation and understanding. However, this is unavoidable with any visualisation of complex data, but by averaging across multiple participants it is hoped this effect was limited. The results from the labelling exercises

demonstrate a variety in how domain experts interpret turbidity events. This highlights the complexity of the question of what a turbidity event even is; something that is often incorrectly considered as a black and white problem. Additionally, the responses show that context is everything when it comes to human interpretation. Something is only noteworthy if it stands out in the context in which it is presented. Analysis of how participants interpreted examples 1-2, compared to examples 3-6, as shown in the scatter plot in Figure 5.5, highlights how the presence of larger events impacts interpretation of lower-level data. Larger events seen in examples 1 and 2 led them to ignore the lower-level events also occurring in these examples which are less visible due to the y-axis scale, yet these are similar in magnitude to those seen in examples 3-6 that most participants acknowledged as events. This demonstrates that human interpretation is inherently subjective. By contrast, a computer will follow instructions precisely and repeatedly. It also highlights that when presented with these lower-level turbidity events unbiased by larger events, participants tended to consider them noteworthy.

Even when participants provided consistent labelled responses, there is an assumed capability that cannot be proven that a participant working in this domain is sufficiently skilled. Though all labellers are actively working in positions where they deal with and understand turbidity and discolouration, high-frequency turbidity time series data like the examples presented is relatively newly available. This means even domain experts may not necessarily be very experienced in interpreting such data. Similarly, it is not possible to determine whether participants were influenced by external opinions or factors. Some difficulties were encountered during the labelling exercises, with some participants only labelling one example, or leaving just one unlabelled. Due to the anonymity of the responses (required to meet ethics standards), it was not possible to question participants giving invalid responses. Therefore, such responses were omitted from this research. In total 48 verified labelled responses were included. This included responses from session 3 that consisted of some unlabelled examples, learned in a post labelling debrief. Session 3 was run externally, and the participants were given a slightly different event definition, where an event was considered anything requiring immediate action, so over 4 NTU or at least 2 customer contacts. This explains why there were significantly lower levels of labelled data in this session and demonstrates how easily even domain experts are influenced when given specific instruction.

The examples included in the labelling exercise were selected to include different types and magnitudes of turbidity events. However, it is not possible to include all possible scenarios with limits also required to make the labelling exercise practical for participants. The examples were checked to be clear of sensor errors, though in reality differentiating sensor errors from genuine events can be difficult. Example 4 was the only example to have been artificially concatenated, to understand how different theoretical event types may be interpreted. Looking at the corresponding averaged-out labels (Figure 5.4 (d)), the first event, representing a hydraulic-induced mobilisation type event, had the highest event score of 0.68, with the other three event types, the single point event, the baseline change event, and the changing diurnal pattern event, not exceeding a 0.3 event score. The final event in example 4 is a combination of these four events and the resulting event score shows an increase in interpreted noteworthiness due to this combination, with the start of this combined event exceeding 0.4. Ultimately it was decided that focusing on events at different scales was more useful for this research, though future research could focus more on categorising different event types.

5.5.2 Event Score Calculation

The approaches used to analyse events in turbidity in this research were performed with the understanding that turbidity events are not necessarily rare, rendering many outlier or anomaly detection methods developed in other fields unsuitable. Time series forecasting methods were explored, with the aim to obtain residual values that enable noteworthy periods of interest to be sufficiently emphasised for subsequent conversion to event score time series with values between 0 and 1, comparable to the averaged-out labels. Calculating event score time series that matched the averaged-out labels was not trivial, in particular finding a solution that generalised across the different examples. The optimisation and sensitivity analysis allowed for each method, and associated input parameters, to be compared in terms of their suitability for this task. This research did not focus on the most accurate forecasting approach, but instead investigated what approach best enables periods of noteworthy data to be highlighted. For this reason, averaging approaches worked effectively at ignoring periods of increased turbidity and in doing so these periods were revealed in the residuals. Some interesting outcomes about window type and forecast horizons that are useful for analysing turbidity events were uncovered. Expanding window types worked best across multiple methods, meaning more data tended to be beneficial in the time scales examined in this

research. Short forecast horizons run into problems during an event where the forecasts start to account for the elevated turbidity. For this research, a forecast that effectively ignores increases is beneficial with a 24-hour horizon achieving this. It has the added benefit of accounting for any seasonality present, in this case seasonality referring to repeating daily trends. As turbidity can contain diurnal trends, typically associated to hydraulic demand patterns, several methods that can account for these were included. The ETS and SARIMA methods both required at least two full cycles of data to include seasonality effects. For this reason, errors were calculated excluding the first 3 days. This meant more methods could be included in the comparison and omits potentially spurious forecasts at the very start of the time series, a problem often referred to as the ‘cold start’ problem. Due to the length of the examples included, between 15 and 75 days, seasonality effects over longer periods such as seasons or annually were not considered.

Modifying expanding average approaches to consider data at the same time of day improved performance, but only when looking at advisory events with no improvement seen across all examples. Advisory events are subtler by nature and more likely to be confused with diurnal fluctuations, which can vary by network and location. The time-based approach accounts for this factor, and matches with human interpretation that repeated diurnal fluctuations, such as those present in examples 3 and 5, are not noteworthy and that it is changes in patterns that should be the focus. ARIMA approaches resulted in the best overall performance against all 6 examples. SARIMA was found to be extremely slow when provided with a seasonality period of 96, meaning not all combinations could be explored but suggesting it is unsuited for this research. The exponential smoothing approaches within the ETS framework did not perform as well as other methods, perhaps due to putting too much weight on recent data, though EWM performed better on both advisory and across all events. Future research could include additional parameters to be used as exogenous variables to improve turbidity event analysis. The CANARY LPCF and BED output was more binary than the averaged labels were, with some complexity lost due to the additional step of determining whether each datapoint is an outlier, before counting the outliers to determine event probability. Therefore, the BED output was not well suited to this research. By instead applying a sigmoidal fuzzy logic membership directly to the residuals, the complexity of the labels averaged out could be better approximated and a better solution was found. This reinforces how well-suited the sigmoidal approach is to this research due to its ability to transform residual time series into output that matches the complexity and fuzziness found in the averages labels.

5.5.3 Proactive versus Reactive Events

Supporting the approach to mimic human interpretation of turbidity events using event score calculations, this research also examined flatline detection methods at different turbidity limits. Any event exceeding 4 NTU is exceeding regulatory limits for end-users, meaning these warrant the highest level of response, regardless of human judgement. Therefore, these events fall naturally into being categorised as alarm events. Alert events are a step lower, but represent significant deterioration compared to the 1 NTU limit at treatment works exit. For the purposes of this research, and to create a convenient division between the first two examples and the final four turbidity datasets used, a limit of 2 NTU was selected, though other values such as 1 NTU could be selected. A system that is only reactive does not prevent events from occurring and as drinking water is required to be below 1 NTU when leaving a treatment works, even low-level turbidity events are evidence of deteriorating water quality during transit. Such low-level advisory events do not come close to breaching regulatory limits and, as such, are generally ignored. However, capturing these events digitally enables whole networks and multiple sensors to be analysed automatically, meaning extra information can be used to improve strategic management of these assets.

The next step for this research is therefore incorporation into an automated event analysis system consisting of reactive alert and alarm event detection as well as novel proactive advisory alarms based on calculated event score time series. One approach to converting the event score time series into proactive advisory alarms is to simply take a threshold and report any exceedances. This in effect would be similar to applying a flatline threshold on the residual time series, though instead it would be applied to the more easily understandable event score with threshold values between 0 and 1. The similarity between these two approaches is further demonstrated and applied to example 6 in Figure 5.11, which shows how a residual-flatline threshold looking for any residuals > 0.2 NTU results in a similar outcome to a threshold of 0.5 applied to the event score time series. Note the flatline is applied to the residual, not to the turbidity data. This shows that practical application of this research may not necessarily require the sigmoidal function, though its use was essential in determining the approach that best approximated the gained insight from the labelling sessions. This research provides a platform from which such a system could be built, but the specific details of how this could be used to issue advisory alerts that aide strategic management require further understanding of what is desired.

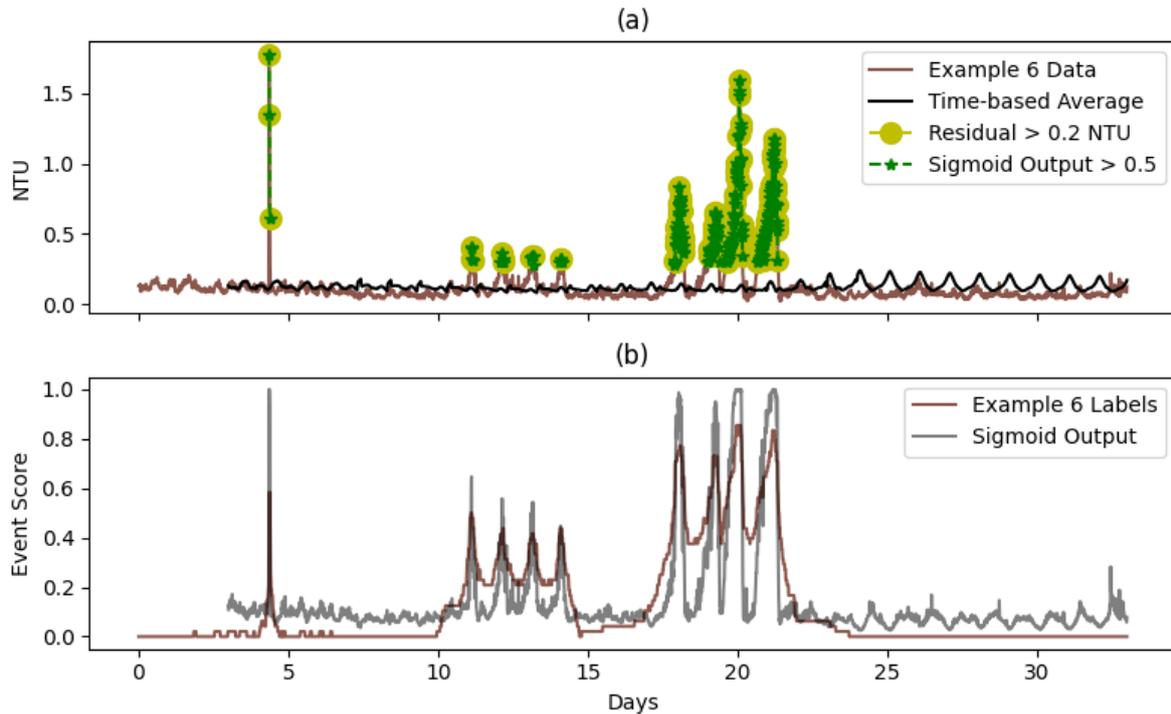


Figure 5.11. Advisory event detection on Example 6 using a 0.2 NTU limit applied to the residual, and a 0.5 threshold using the sigmoid output (a), with equivalent average labels and sigmoid output (b).

5.6 Conclusions

This research shows how complex and time-consuming human interpretation of turbidity time series data from drinking water distribution systems can be mimicked in real-time by computing algorithms. Automating such interpretation provides a rapid and more extensive capability to understand network performance, allowing for focussed strategic and operational decisions to manage in-network discolouration. The crowd-sourced labelling exercises undertaken represents a novel approach that addresses the difficulty in obtaining confirmed real-world events, while also highlighting the need to fully understand what is wanted from the data before developing analytic methods. These exercises informed a turbidity event scale that considers reactive alarm (>4 NTU) and alert (>2 NTU) events, alongside proactive advisory (<2 NTU) events. For alert and alarm events a flatline approach is considered best, assuming quality assured data is available. A time-based averaging approach was found to work best at identifying advisory events. These approaches require little computational power and could be applied in real-time.

Chapter 6: Extracting Actionable Information from Water Quality Time Series Data: From Single-Parameter Single-Sensor to Multi-Parameter Multi-Sensor

6.1 Introduction

This chapter explores how actionable information can be extracted from DWDS water quality time series datasets. While different types of actionable information may be desirable for WSP, this chapter and research focuses particularly on what can be learned from turbidity and chlorine sensor data. This is both due to these being the most commonly measured parameters in the datasets made available, and their direct importance to achieving the ultimate aim of DWDS: enabling the safe passage of drinking water to end users. It should be noted that information such as distances between sensors and pipe sizes were not generally made available but are provided in the cases where they are known. However, the purpose of this research is to explore data-driven analytics and the desire is to not rely on the availability of additional supporting network information. As identified in Section 2.2.3, the three main causes of water quality deterioration within DWDS are due to interactions with pipe surfaces, bulk water transformation, and infrastructure failures. Turbidity sensors are proxy discolouration sensors, meaning they are uniquely well-suited to help capture and understand discolouration events, the number one water quality related issue in the UK. Discolouration is often caused by hydraulic-induced pipe-wall material mobilisation, but could also result from ingress following infrastructure failure or from contamination at upstream WTW or service reservoirs. Where used as a residual disinfectant, chlorine is relied upon to provide lasting protection throughout the DWDS, meaning any changes to chlorine concentration levels indicate changes in bulk water transformation, potentially posing increased threat of contamination.

The aims of this chapter are to show how the methods developed in Chapters 4 and 5 can be applied to real-world examples, explore how they can be integrated with other analytical approaches, and examine the impact additional parameters and sensors have on the level of insight that can be derived. To achieve this, the first two examples (Sections 6.2 and 6.3) will be presented in the following sequential order, building from single parameter single sensor (SPSS) analysis, to then combining multiple locations measuring the same parameter (single

parameter multiple sensor – SPMS), combining different parameters at the same location (multiple parameter single sensor – MPSS), to eventually demonstrate the additional value from having multiple parameters measured at multiple different locations (multi parameter multi-sensor – MPMS). Table 6.1 presents analytical possibilities that will be examined at each of these stages, with the intention to highlight the presumed increase in insight when moving from SPSS to MPMS.

Table 6.1. Analysis possibilities moving from SPSS to MPMS.

SPSS	SPMS	MPSS	MPMS
Data quality assessment (framework stage 1)	Data quality assessment (entire framework is parameter dependant)	Data quality assessment (framework stage 1, other parameters may help confirm/rule out errors)	Data quality rules (entire framework)
Turbidity event analysis (proactive precursors and reactive events)	Network connectivity (parameter dependant)	Material flux (turbidity and flow rate)	Network connectivity Material flux (higher accuracy and confidence)
Drift correction	Local or global event differentiation	Event classification	Event classification (higher confidence)
Low or high chlorine event detection	Chlorine decay rates	Unsupervised anomaly detection	Event tracking

Both of the ‘SPSS to MPMS’ examples feature discolouration events detected by more than one multi-parameter sensor, but both are first reviewed with an analysis of what can be learned with only a single turbidity sensor. Example 1 (**Section 6.2**) will then move to MPSS to demonstrate the additional value from having more parameters available, before the example is fully analysed as a MPMS case. Example 2 (**Section 6.3**) will move from SPSS to SPMS to examine how additional turbidity sensors can improve understanding, before again being treated as a MPMS case. Five further case studies are then featured, this time starting with all available data, to demonstrate a variety of situations in which actionable information was derived from MPMS DWDS water quality time series. Example 3 (**Section 6.4**) again examines discolouration, this time focusing on an end of network location at which maintaining a disinfection residual was a challenge. For Example 4 (**Section 6.5**), turbidity is compared at six locations along a single 70 km long mains in order to investigate how

discolouration can change along a network section, this time without chlorine being available and only inlet flow rate. A different MPMS example is presented in Example 5 (**Section 6.6**), using dimensionality reduction and unsupervised anomaly detection to explore how information could be automatically derived from higher-dimension datasets. Example 6 (**Section 6.7**) returns to discolouration, though this time analysing a situation in which a step increase in flow rate led to increased daily turbidity, though never exceeding 2 NTU, at multiple connected sensor locations. The final Example 7 (**Section 6.8**) does not feature high-frequency DWDS data, instead focusing on weekly regulatory samples at service reservoirs, in order to highlight how improved insights could be gained through increased high-frequency online monitoring. A summary of the insights gained from each example is then provided in **Section 6.9**.

6.2 Example 1 – Building from SPSS to MPSS to MPMS

6.2.1 SPSS

The work in Chapter 4 highlighted that the first step when analysing any water quality time series is to perform a data quality assessment. Figure 6.1 is a plot of a turbidity time series over 2 years in duration, with the data quality rules developed in Chapter 4 applied. As specified in Table 6.1, only stage 1 of the developed data quality assessment framework can be applied when analysing a SPSS. The data quality rules detected a number of errors and potential errors. Four different extended periods of elevated turbidity are detected. The first of which occurs at the very start and lasts for several weeks, suggesting issues following the initial installation that are likely not representative of the DWDS water quality. The following 3, occurring in April 2020, April 2021 and May 2021 require further inspection, as these could be real network events. Other data quality issues identified include several periods of flatlining data, which was determined to have been caused by the sensor measurement resolution not being high enough. Up until May 2021, the resolution was set to 0.08 NTU, but a data quality review highlighted a number of data quality issues to the WSP using the approaches from Chapter 4. One of the outcomes of this data quality review was the resolution was changed to 0.01 NTU. Three periods of drift are also identified, and several single point outliers.

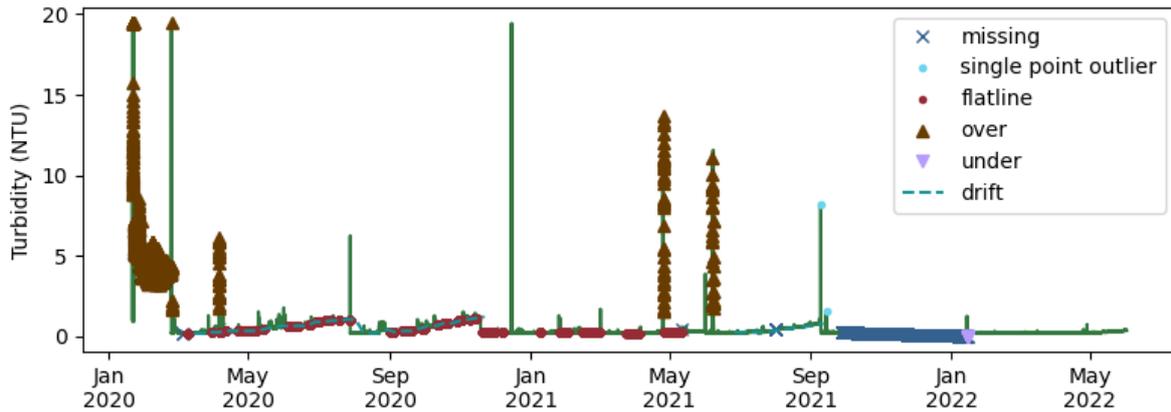


Figure 6.1. Data quality rules applied to over 2 years of turbidity time series.

As the interest here is in analysing potential discolouration incidents, the first of the three potential identified events is investigated. As this occurs within a period of detected drift, it is advisable to perform some drift correction before any analysis. This is performed using the asymmetric least squares approach previously used in Chapter 4 and the drift corrected data is plotted in Figure 6.2. The drift corrected data was then analysed using the event scale method developed in Chapter 5, with the alarm event detected on the 5th April identified and displayed in Figure 6.3(a), along with the advisory event score time series shown in Figure 6.3(b). This analysis shows that a potential precursor event was seen 10 days before the alarm event on the 27th March which resulted in an advisory score of 1.00, with another seen 3 days prior on the 2nd April with a peak advisory score of 0.77. These potentially could have been indicators of increased discolouration risk at this location, which does seem to continue into May with multiple advisory events seen again. That no alert or alarm event followed in this instance may be down to the lack of a triggering event, such as a hydraulic change. However, it is difficult to make conclusions with any degree of confidence with data from only one turbidity time series.

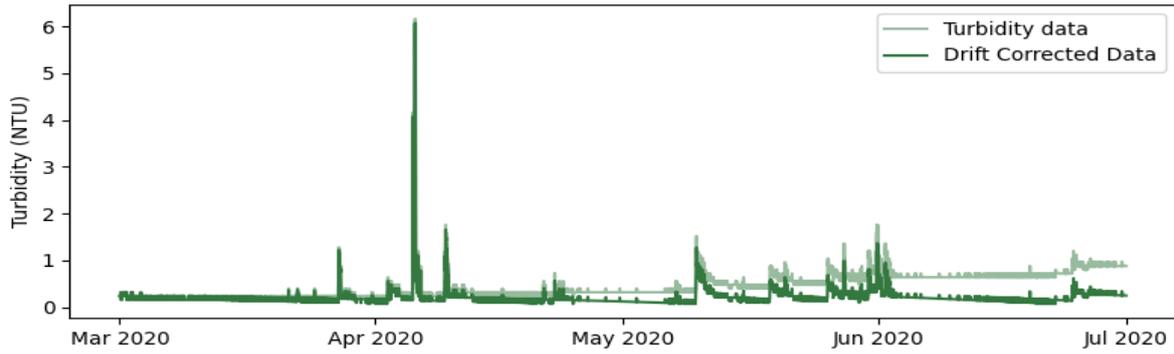


Figure 6.2. Turbidity time series containing identified event, with drift corrected version shown.

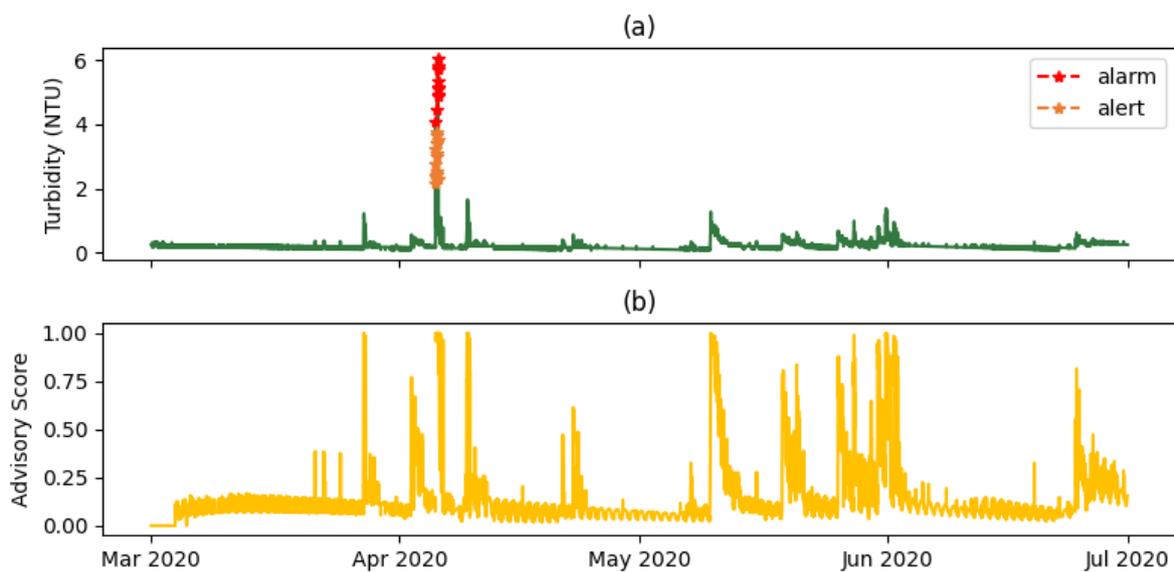


Figure 6.3. Turbidity event scale applied to example, with alert (>2 NTU) and alarm (>4 NTU) event identified in the turbidity time series (a), and an advisory time series shown in (b).

A closer inspection of the turbidity data during the alarm event is shown in Figure 6.4. There appear to be two events, the first being a short spike with the turbidity rising from 0.24 to 4.16 NTU in just 1 hour, with the peak coming at 13:15. However, it does not return to normal levels, with a larger second event with a slower rise peaking at 6.16 NTU at 19:30, before sharply declining to below 1 NTU at 22:00. With this event starting around lunchtime and is finished by the evening, this indicates it is at a time of high demand. However, with just a single turbidity sensor, it is difficult to confidently rule out a sensing error as the cause of this event or to make any assertions about the underlying causes or the extent of the network impacted by this (apparent) increased discoloration.

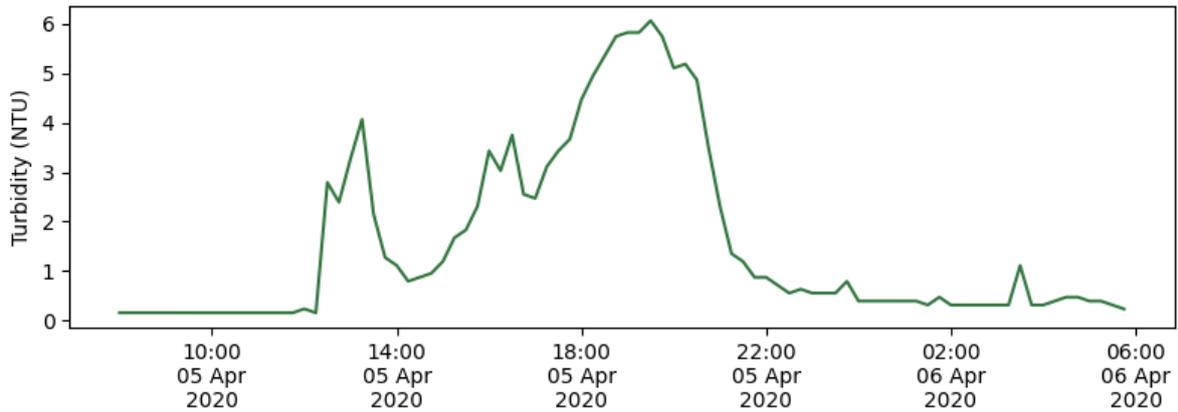


Figure 6.4. Close-up of alarm (> 4 NTU) turbidity event.

6.2.2 MPSS

In this example four additional parameters were monitored within the same sensor: free chlorine, temperature, pH, and pressure, flow rate data was also available from a meter installed within this part of the network. The data quality rules developed for turbidity and chlorine were therefore applied to these parameters, with the results shown in Figure 6.5 (a) and (b). Flatline issues are seen in both parameters. The chlorine data shows relatively low resolution of 0.02 mg/l up until May 2021, accounting for the majority of the flatline detection in this data. After May 2021 it increases to detect changes less than 0.001 mg/l. The chlorine parameter was measuring zero for the first few weeks, supporting the idea that the sensor was not properly installed, while it also shares the roughly 3 month period with missing data towards the end of 2021, meaning this was a sensor-wide issue. Reviewing the other three parameters from this sensor in Figure 6.5 (c), (d), and (e) confirms that this sensor was not functioning during these months. The pressure is also seen to be zero for the first few weeks, confirming the sensor was not connected to the pressurised supply. By viewing all 6 available parameters before, during and after the alarm event (Figure 6.6), it can be observed that a drop in chlorine from 0.74 to 0.61 ppm occurred just before the turbidity alarm event, around the same time as the 0.77 turbidity advisory event, suggesting changes in bulk water characteristics. There is also a drop in pH from 8.0 to 7.6 between the 18th March and 11th April. Reviewing the hydraulic parameters, the alarm event on the 5th April occurred alongside a pressure drop of around 4 m (from 28 m to 24 m) and a peak daily flow rate up to 68.4 m³/h, exceeding the previous peak daily flow rate of 64 m³/h on 31st March. This pressure drop and relative higher peak flow rate suggests this event was likely caused by a change in hydraulics resulting in a mobilisation of pipe-wall material.

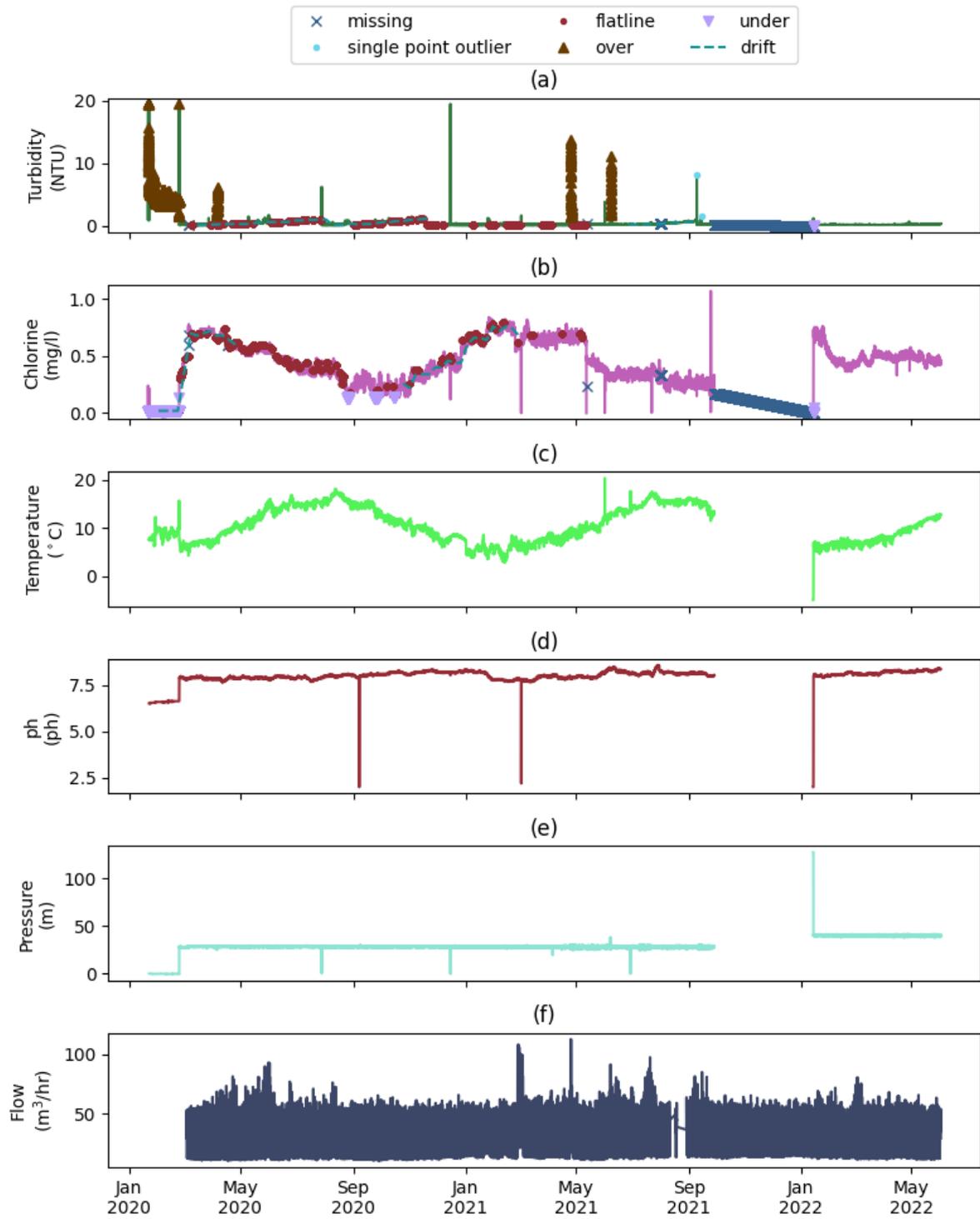


Figure 6.5. Data quality rules applied to turbidity (a) and chlorine (b) time series, with temperature (c), pH (d), pressure (e), and flow (f).

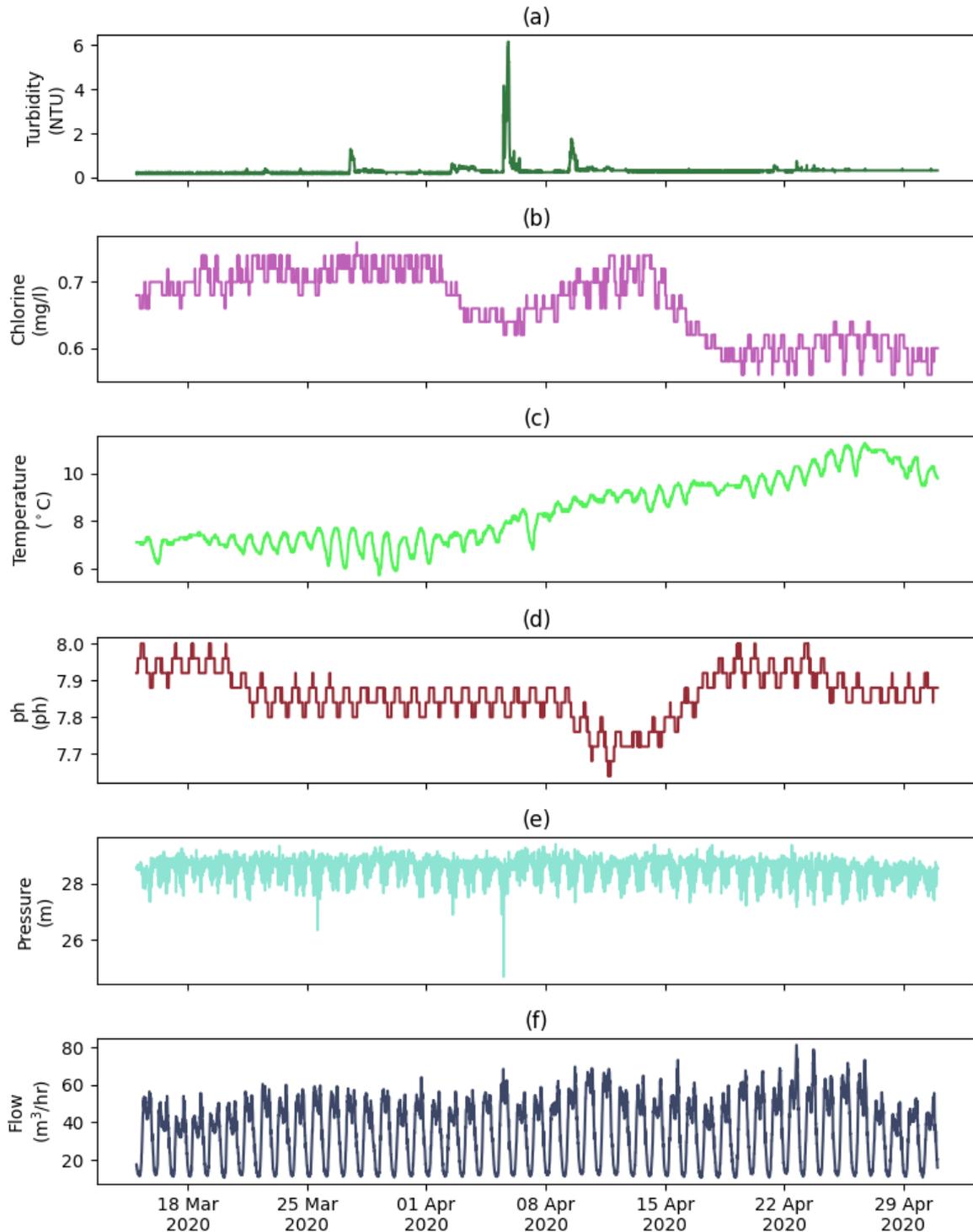


Figure 6.6. Turbidity (a), chlorine (b), temperature (c), pH (d), pressure (e), and flow (f) before and after event.

A closer inspection of the timings involved is shown in Figure 6.7, which shows an initial spike in turbidity occurs simultaneously to the flow rate increase, with the pressure drop occurring shortly after. Material flux, introduced in Background Section 2.4.3.2, was used to estimate total material mobilised during this event. Material flux was calculated by

multiplying flow rate by turbidity, as shown in Figure 6.7 where the total material was estimated to be 1380 NTU.m³, calculated as the area under the material flux curve. As indicated in this figure, the baseline turbidity was subtracted in order to only estimate the excess material movement during this event.

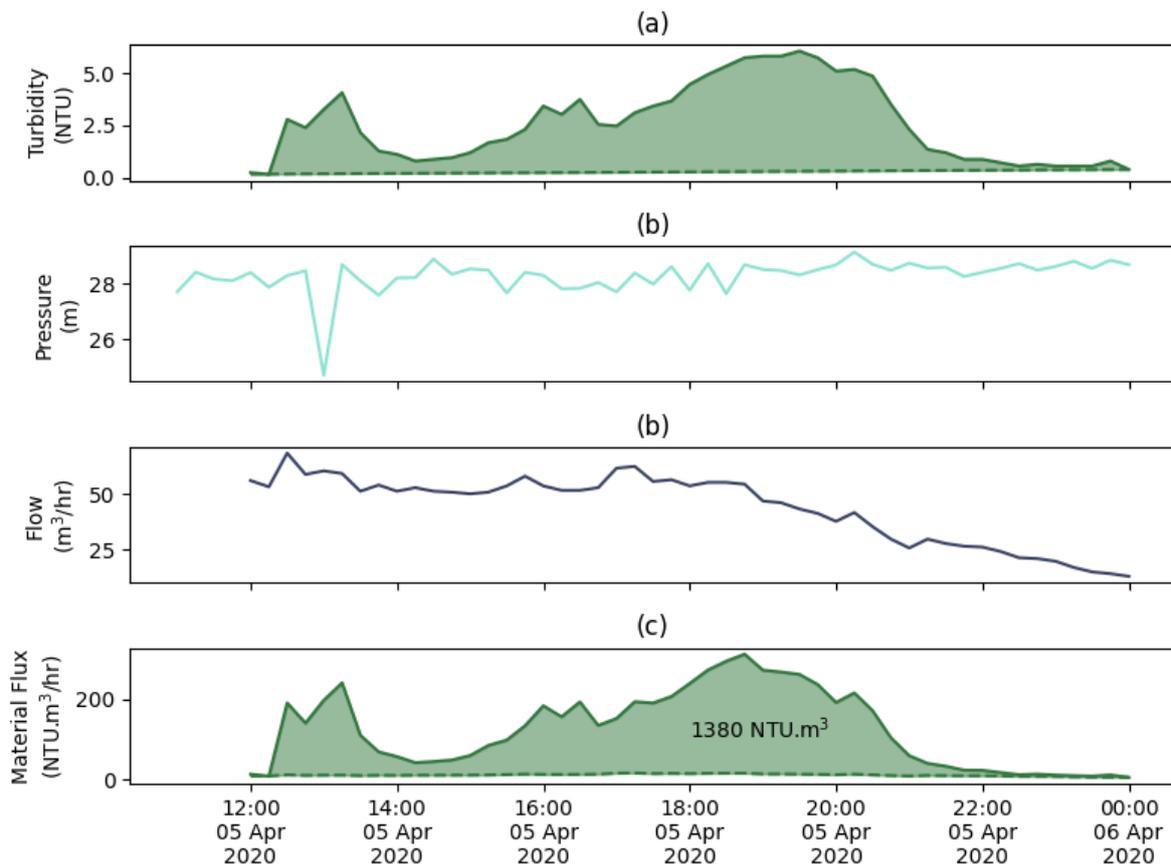


Figure 6.7. Turbidity time series during alarm event (a), with pressure (b) and flow time series (c), and calculated material flux (d), where total material is labelled within the highlighted area.

6.2.3 SPMS

Ten other multi-parameter water quality sensors were deployed across the same DWDS. Any other sensors installed at hydraulically connected locations can be used to enhance analysis of water quality at the location in question. Therefore, cross-correlation was performed to determine if any hydraulically connected locations were available. As the event in question occurred in April 2020 but very little in-network data was collected prior to this, a period of 4 months was used for the cross-correlation, running from April to the end of July 2020. Before this could be done a data quality assessment was required for the whole dataset during this period, in order to ensure no erroneous data interferes with the correlation calculations.

Figure 6.8 is a bar plot displaying the results of the data quality rules applied to each of the 11 chlorine sensors, where the ‘ref’ column on the left hand side refers to the multi-parameter *reference* sensor that has been featured thus far. Of the ten other chlorine sensors, several had performance issues during this four month window, such as missing data, periods of flatlining data, and drift. Cross-correlation was then performed between each of the eleven sensors, following the method set out in the second stage of the data quality assessment framework in Chapter 4, with the median correlation coefficients shown in Figure 6.9. This bar plot shows that the sensor at location 6 was the only one found to have a significant cross-correlation, with a median value of 0.84. The cross-correlation between these time series is shown in more detail in Figure 6.10, with the majority of the resulting sliding correlation coefficients seen to be above 0.7. The offset values were used to estimate transit time between these locations. Figure 6.10(c) shows that the offset varied between -3.5 and -4 hours, with a median of -3.5 hours, for the time period where the PCC remained above 0.7. Therefore, it can be estimated that location 6 is approximately 3.5 hours upstream of the reference sensor.

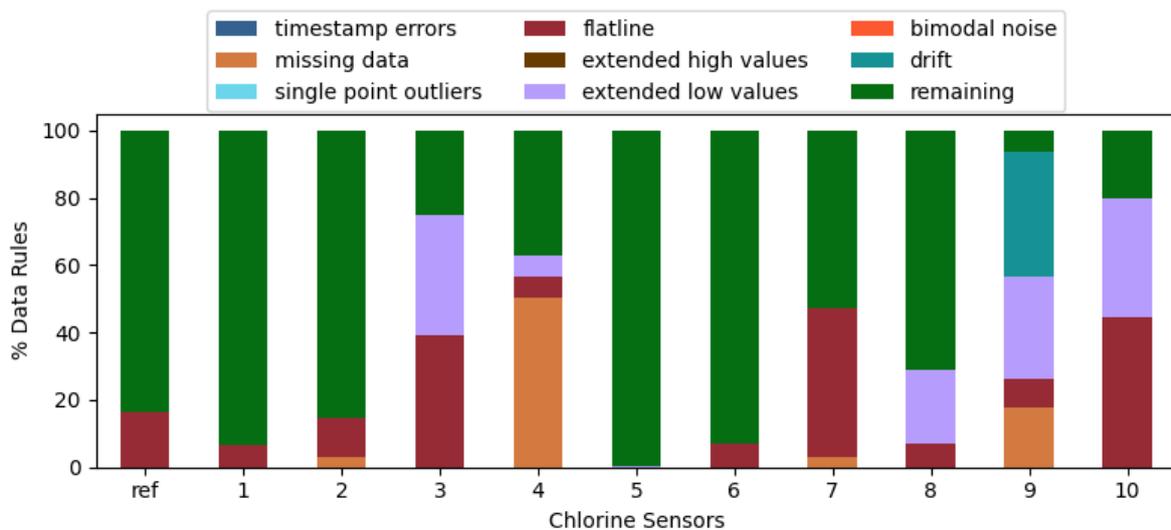


Figure 6.8. Data quality rules applied to 11 chlorine time series, with ‘ref’ referring to the chlorine sensor at the location being investigated.

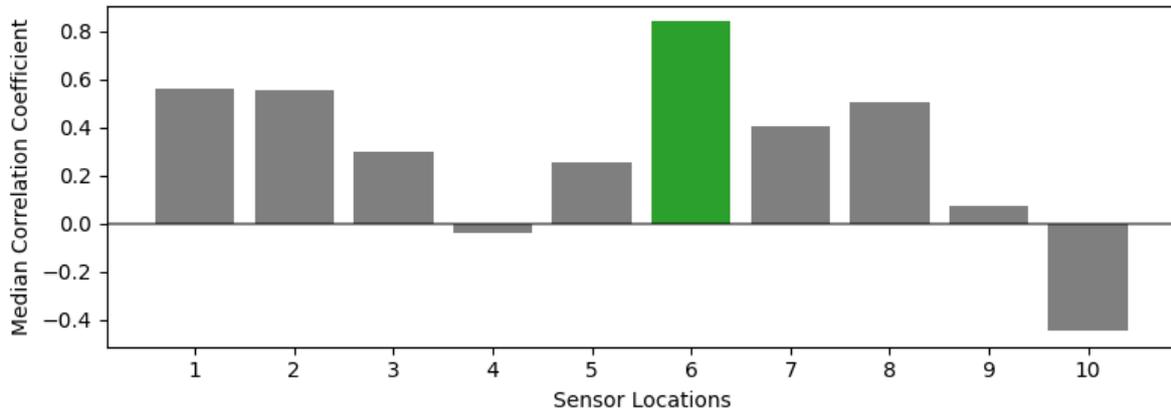


Figure 6.9. Median correlation coefficients between reference chlorine data and 10 network deployed chlorine sensors.

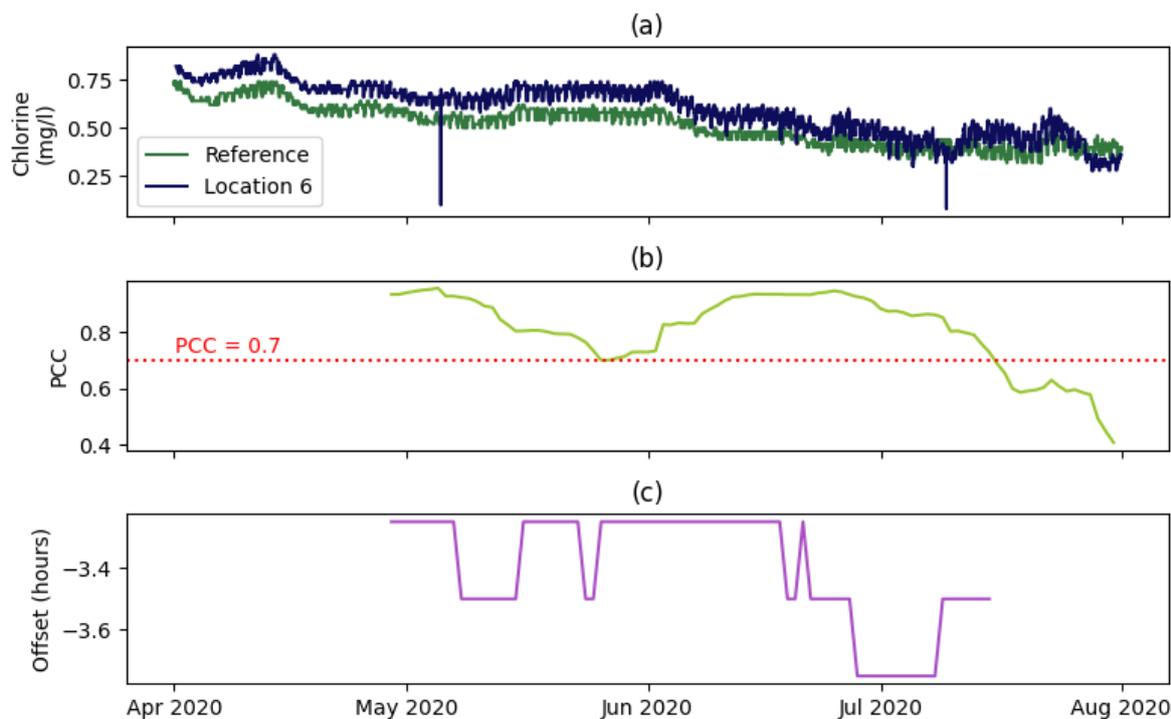


Figure 6.10. Chlorine time series at reference and location 6 (a), sliding cross-correlation PCC (b) and sliding offset (c).

6.2.4 MPMS

With a connected location identified and a transit time estimated, the other data / parameters at these locations can now be used for comparison within this context. A simplified network schematic is shown in Figure 6.11 with both locations included. Both turbidity time series are plotted around the time of the event in Figure 6.12. As an alarm event is also seen in the upstream sensor at location 6 (which was installed at the service reservoir outlet), this

supporting evidence confirms this event to be a real network alarm event and not a sensor error or an event originating solely from between the 2 sensors. The event starts with an initial spike, seen at both locations, followed by a gradual increase of turbidity spread out over 6 hours at location 6 and 12 hours downstream at the reference sensor. The simultaneous initial spike response indicates a sudden mobilising of material local to each sensor location. The turbidity peaks occur during what appears to be a second wave at 15:45 in location 6 and 19:30 in the reference, meaning an offset of 3 hours 45 mins is observed. Note that the 15 minute sampling intervals dictates that both this and the cross-correlation-derived transit times must therefore be in multiples of 15 minutes. This delay falls in line with the transit time derived by cross-correlation and indicates that this second wave relates to material upstream of sensor 6. There is also a second spike visible in the reference around 3 hours 15 minutes after the initial spike, suggesting the material initially mobilised locally to location 6 had reached the reference at this time. Interpretation of the turbidity data alone would suggest an increase in severity between location 6 and the reference when viewing the area under the curves. Comparison of the turbidity data using the event scale approach, plotted in Figure 6.13, shows that high advisory scores are seen in location 6, as well as the reference. Both locations also see multiple high advisory scores during May and an alarm event is again seen at location 6, indicating continued material movement in this network section.

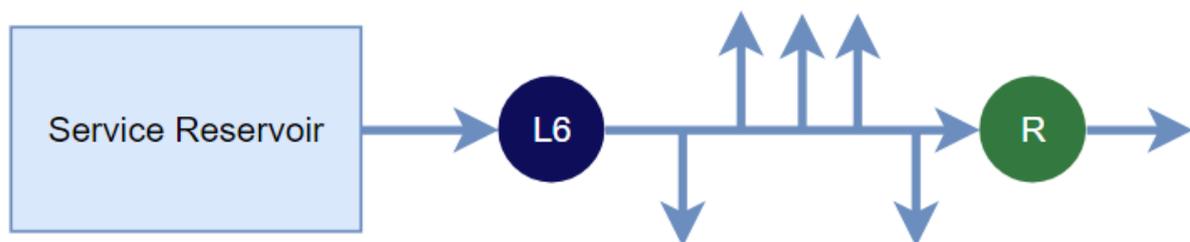


Figure 6.11. Simplified schematic showing reference location (green R) and upstream location 6 (blue L6), with both downstream of a service reservoir and separated by some offshoots.

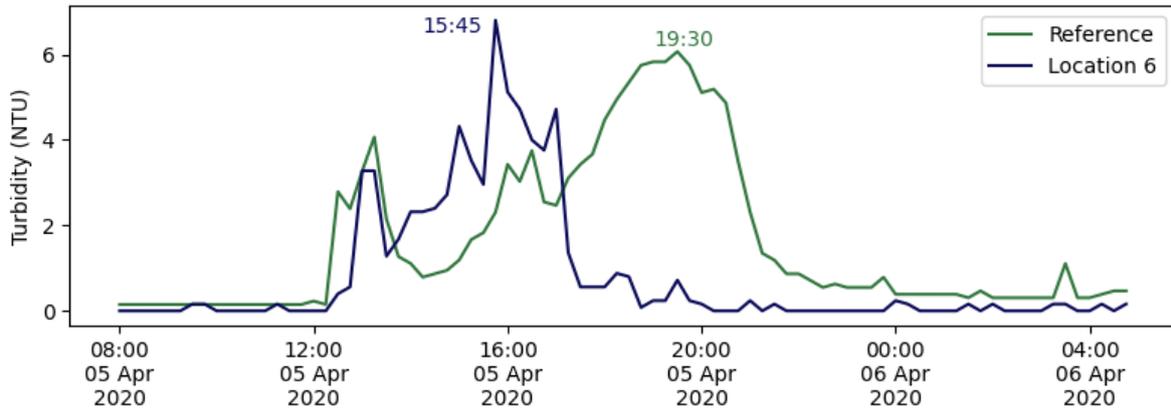


Figure 6.12. Turbidity time series at both locations during event on 5th April, with the times of each peak labelled.

Consideration of the hydraulic data at each location could provide more context regarding this event. It was previously shown in Figure 6.7 that this event occurred alongside an increased flow rate peak and was shortly followed by a pressure drop at the reference location. The flow rate and pressure time series are plotted alongside the turbidity time series from both locations in Figure 6.14. Interestingly, no pressure drop is seen at the upstream location 6 and the flow rate at this location during this event is not as high as recent daily peaks. This is not what would have been expected if this event was caused by a flow rate increase in this network section and suggests the main source of this event is upstream of both these locations.

Viewing the other available parameters, temperature, pH, and chlorine, at each location (Figure 6.15) shows that the same chlorine and pH drops observed in the reference are also seen in location 6, supporting evidence of the connectivity between these sites and the observed change in bulk water characteristics potentially linked to the increased discolouration.

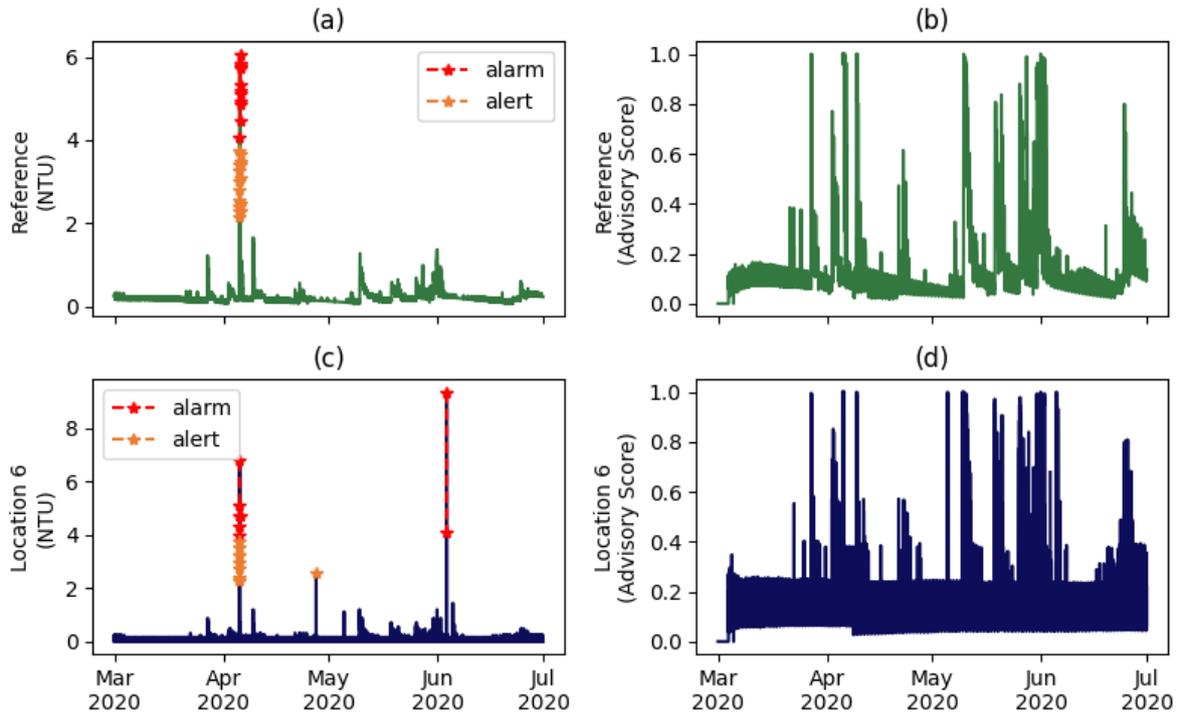


Figure 6.13. Turbidity event scale applied to the reference, with alert and alarm events shown in (a), and advisory event score in (b), and location 6, with alert and alarm events shown in (c), and advisory event score in (d).

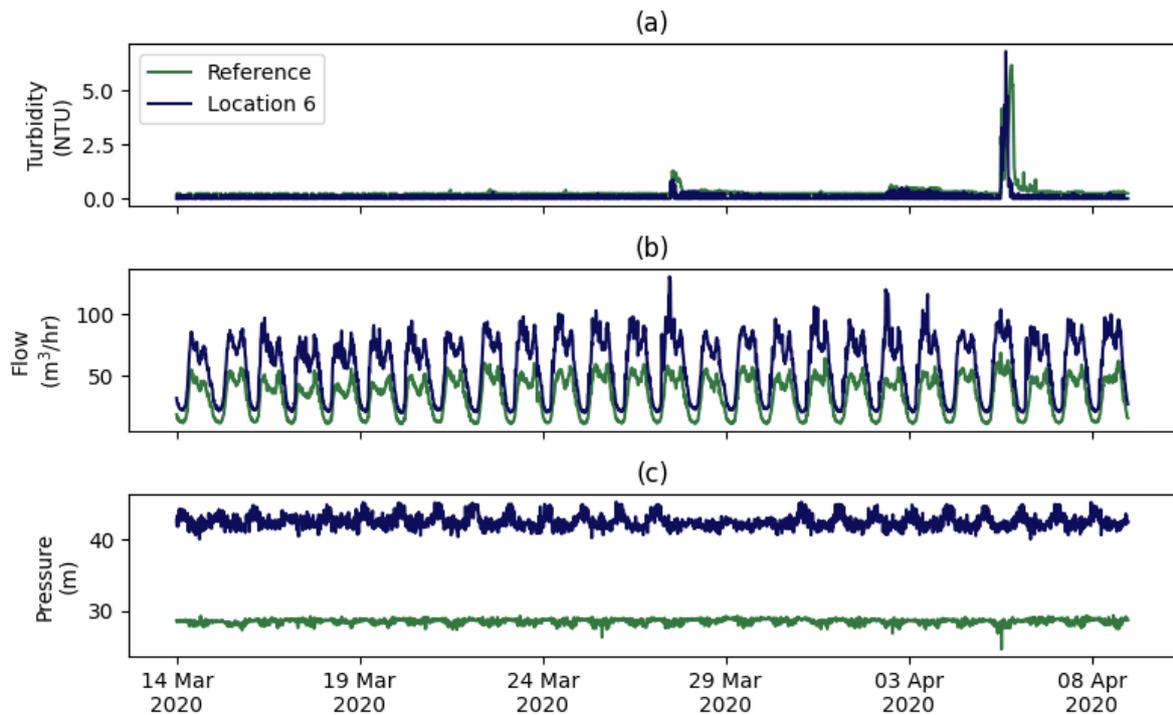


Figure 6.14. Turbidity (a), flow (b), and pressure (c) at both locations in lead up to April 5th event.

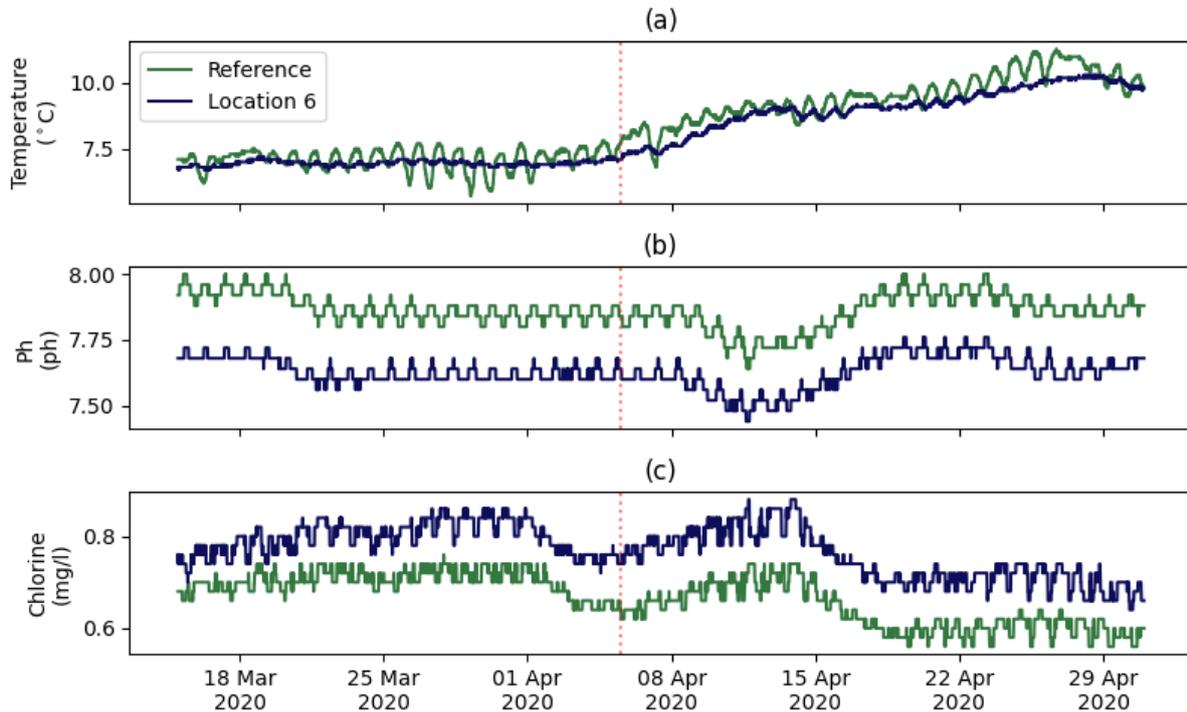


Figure 6.15. Temperature (a), pH (b), and chlorine (c) at both locations, where the alarm event is indicated by the dotted red line.

The large second turbidity wave represents a major source of concern with this event, due to its duration and appearance at two locations. Therefore, material flux was employed in order to quantify the total material passing each location during this wave and better understand this discolouration incident. The identified initial spikes and its recurrence at the reference therefore needed to be omitted in order to only focus on the material movement during this second wave. Material flux calculations were done, as illustrated in Figure 6.16, again with the baseline turbidity subtracted in order to estimate the excess material caused by this event. An overall reduction of around 10% between location 6 and the reference (from 1305 to 1163 NTU.m³) is calculated, as shown in Figure 6.16(c). This could be expected of an upstream sourced event as some material would be expected to be lost on the 3.5 hour journey between these locations. Therefore, this analysis shows that the primary material source of this event is upstream of location 6 and this can inform prioritising of maintenance. This is of value to the WSP as 4 customer contacts were made for discoloured water on the 5th April in the supplied region, between 14:49 and 20:37, indicating the significant impact of this event.

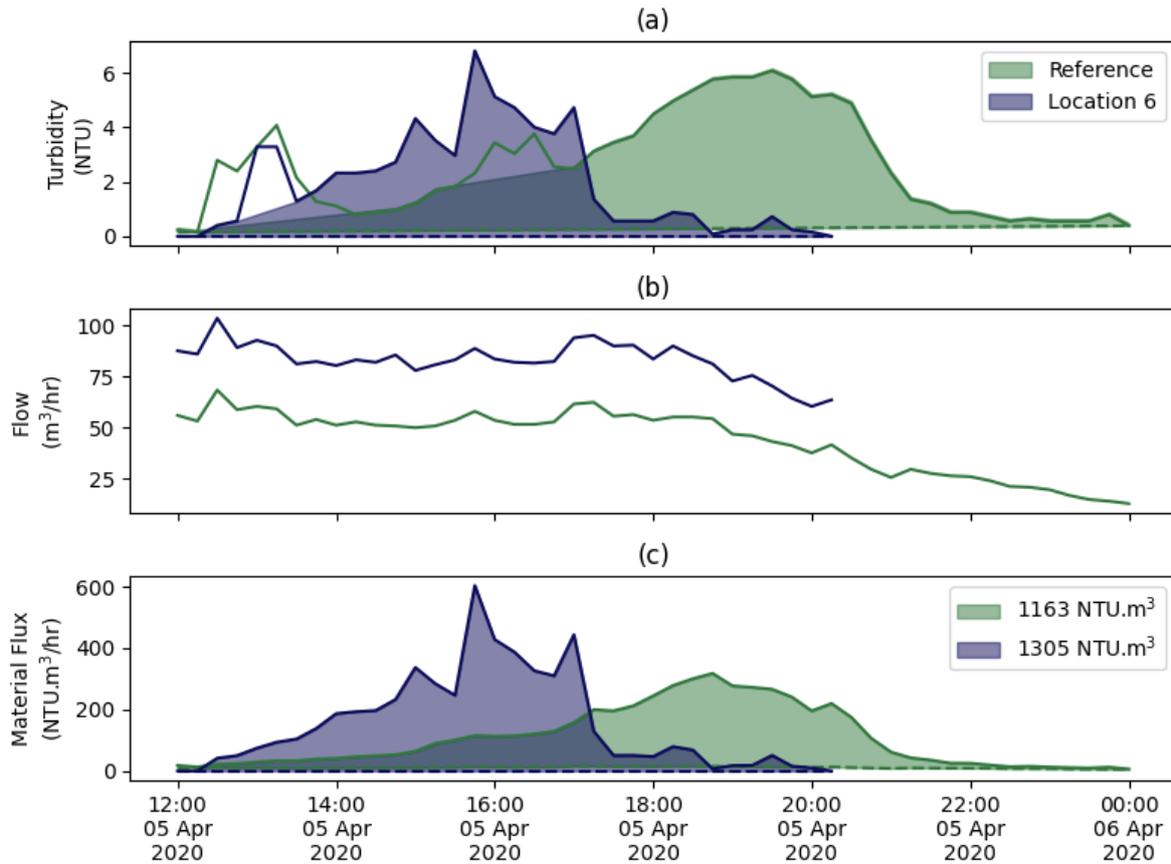


Figure 6.16. Secondary Turbidity event responses at each location with turbidity (a), flow (b), and net material flux (c).

6.3 Example 2 – Building from SPSS to SPMS to MPMS

6.3.1 SPSS

Figure 6.17 shows a turbidity time series with the rules from stage 1 of the data quality assessment framework applied. In this case, the turbidity sensor has been monitoring for about 3 years. This sensor belongs to the same overall dataset as the two featured in Example 1, and both the initial period of elevated turbidity in the first few weeks, and the flatlining issues seen in the first year, again due to the low turbidity resolution, are repeated. This sensor was also seen to have bimodal noise issues, also seen in location 6 in Example 1. Both of these issues were identified as part of this research and were largely addressed, as evidenced by the improvements seen from May 2021 onwards. Though using information from other sensors in this dataset could be said to be SPMS, this section focusses on how events can be analysed. Four different extended periods of elevated turbidity are detected: April 2020, October 2021, June 2022 and July 2022. The first identified event is investigated.

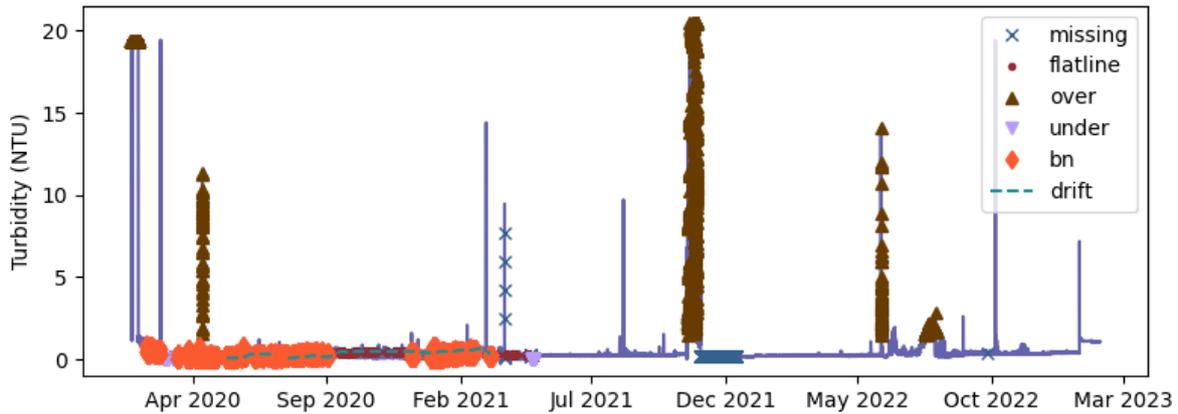


Figure 6.17. Data quality rules applied to turbidity time series.

Some drift was detected but since this started after the April event, no drift correction was applied. Figure 6.18 shows that the April 12th event was an alarm event, and Figure 6.18(b) shows that the preceding weeks had consistently high advisory scores. Though this sensor is impacted by bimodal noise issues, it is seen to reduce after the alarm event, with the median peak daily advisory score being 0.82 in the weeks leading up to the event and falling to 0.50 in the month following. This drop indicates a reduction in material movement post event. A closer inspection of the turbidity data during the alarm event is shown in Figure 6.19. The turbidity is seen to rise from 0.32 to a peak of 11.36 NTU between 22:15 on the 11th April and 00:15 on the 12th April. Turbidity levels remained above 8 NTU for another 6 hours, before they started to reduce, going below 0.5 NTU at 11:45. Without the flow rate data it is not possible to estimate the total material involved, but the sharp rise and steady decline is typical of hydraulically mobilised discolouration events (Background Section 2.2.3). That the majority of this event occurred during the night may mean lower customer demands were involved, unless in the case of a hydraulic event such as a burst pipe. Occurring during the night may also result in fewer customer observations although it did continue well into the morning when the demands typically peak. The shape and profile of this data suggests a real event, but any conclusions made without supporting information would be speculative.

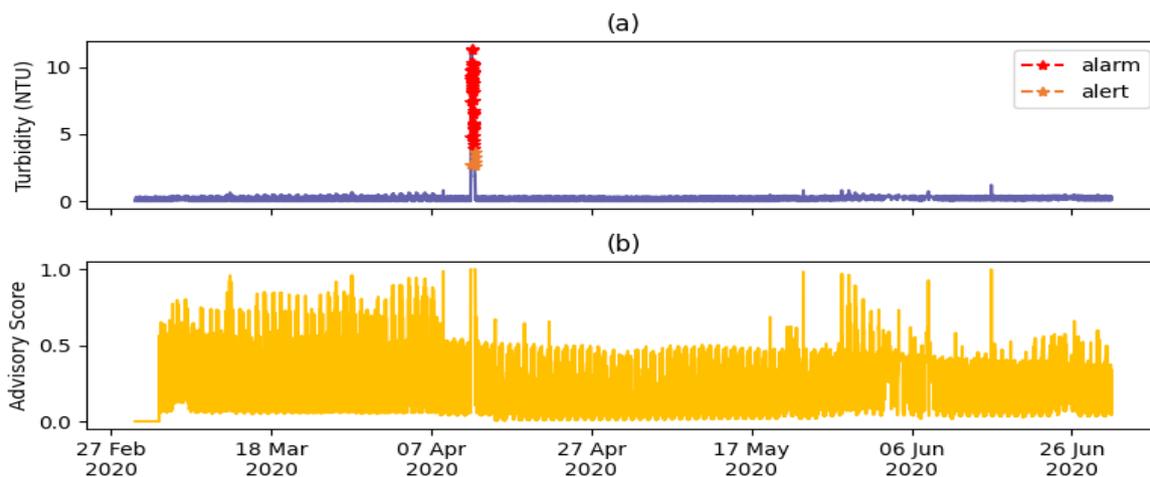


Figure 6.18. Turbidity event scale applied to example turbidity, with alert and alarm events identified (a), and the advisory score time series (b).

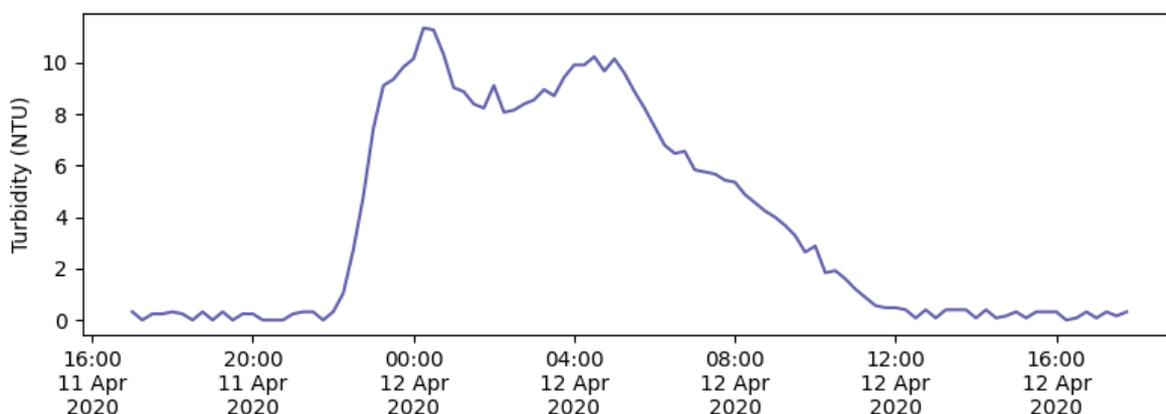


Figure 6.19. Close-up of alarm turbidity event.

6.3.2 SPMS

Moving from SPSS to SPMS explores how multiple turbidity sensors may improve understanding of network behaviour. As stated in Chapter 4, unlike chlorine, turbidity time series are generally not well suited to cross-correlation. This makes the process of determining which sensor locations are relevant less straightforward. One approach is to simply check if events are detected at other locations around the same time. The first step was to review the turbidity data quality. Figure 6.20 is a bar chart showing the data quality rules flagged for each of the 11 turbidity sensors from March to the end of June 2020, with the reference location on the LHS. Some recognised issues are shown including drift, flatlining, missing data and extended high values. To examine which of these sensors may be of use to analysing the reference alarm event, the data quality at the time of the event is reviewed in

Figure 6.21. From this figure, it appears that up to 4 locations have potentially been affected: locations 2, 5, 7 and 10, in Figure 6.21(b), (e), (g) and (j), respectively. However, location 5 is just a single data point, and has been flagged as a single point outlier, while location 10 seems to be already extremely elevated. By viewing the data quality at location 10 on a wider timeframe, shown in Figure 6.22, it appears that this turbidity sensor was measuring elevated levels up until the end of April, indicating that the sensor was malfunctioning. Therefore, this leaves just locations 2 and 5 for closer inspection, both of which contain alarm events exceeding 4 NTU.

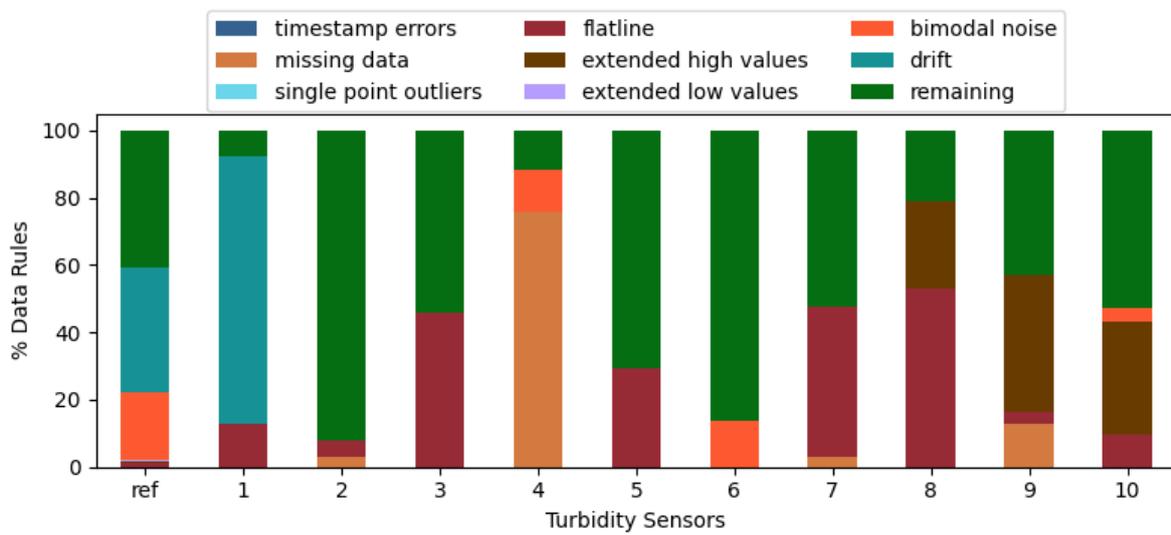


Figure 6.20. Data quality rules applied to 11 turbidity sensors.

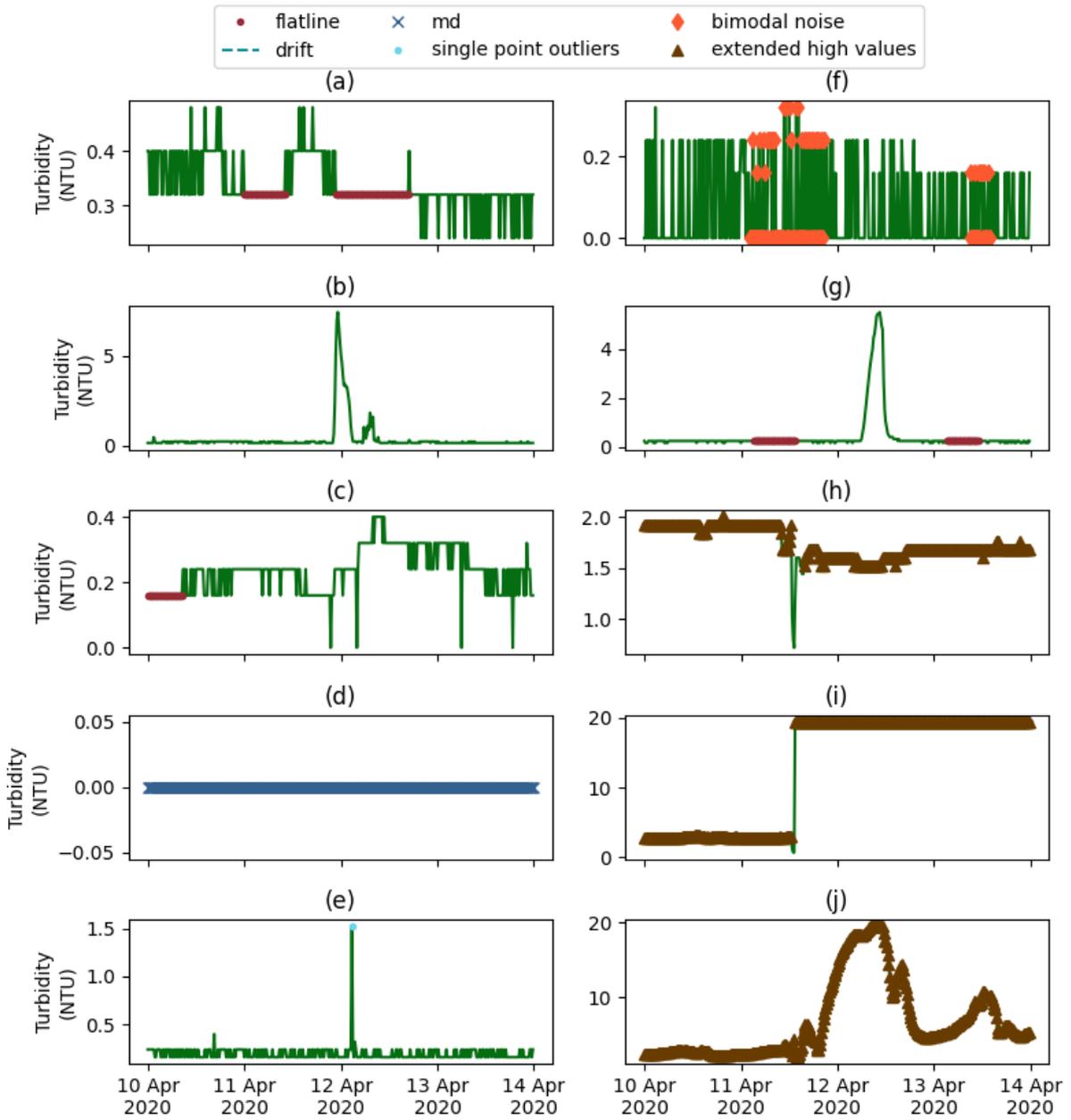


Figure 6.21. Flagged data quality rules in 10 turbidity sensors from 10th to 13th April.

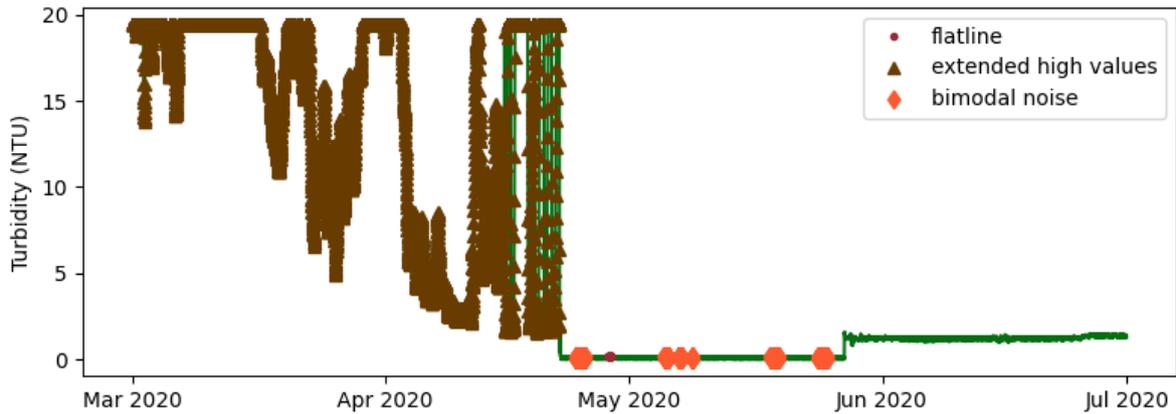


Figure 6.22. Data quality rules applied to turbidity data at location 10 from March to June.

Figure 6.23 is a plot showing the turbidity at locations 2 and 7, compared to the reference turbidity data, during the detected alarm event. The reference alarm event has an almost identical and simultaneous initial rise to that seen in location 2, indicating that this is a genuine network event and not local to one location. This suggests the cause is likely to be hydraulically induced with flow rate increases impacting both monitored sections. The event at location 2 is much shorter in duration, with turbidity seen to return to normal levels by 3am, approximately 4 hours after the peak at 23:15. The event at location 7 occurs towards the tail end of the reference event, with its peak about 10 hours later than the reference peak. More information would be needed to determine if these are linked.

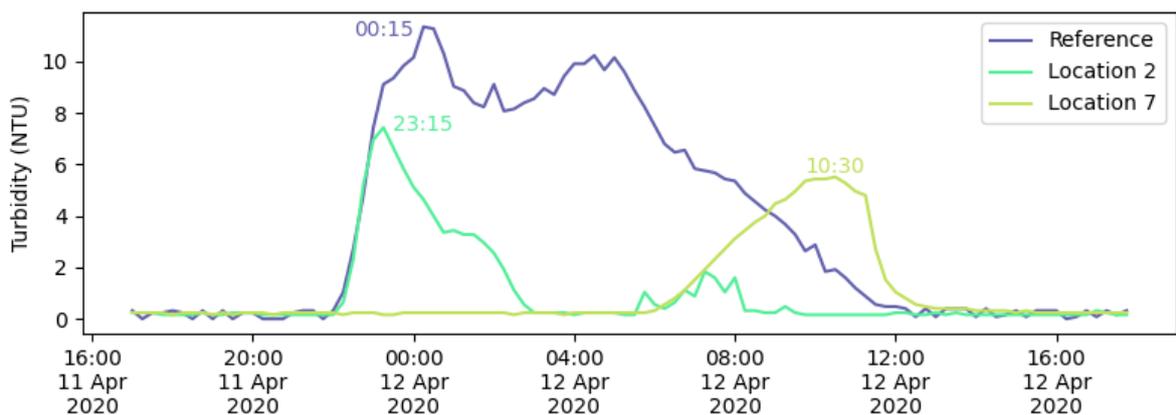


Figure 6.23. Turbidity time series at reference and locations 2 and 7 during alarm event, with the times of each peak labelled.

Assuming all three are hydraulically connected, each turbidity time series is analysed using the event scale in Figure 6.24. No other alert or alarm events are seen in this period, but viewing the peak daily advisory scores shows that location 2 was also seeing high levels of

advisory scores before this event, which reduced in the month following. This was similar, though not as extreme as was seen in the reference location, which was influenced by the bimodal noise sensor errors. This provides confidence that location 2 is linked, and indicates that this network section of network had elevated material movement prior to this alarm event. The peak daily advisory scores at location 7 are also slightly higher prior to this event than in the week following (aside from a couple of peaks). Of note is that all three locations measured some increased turbidity towards the end of May, resulting in high advisory scores of 0.97 at the reference location and 0.84 at location 2, with a 0.51 score at location 7. This alignment again suggests these locations may all be linked but further information would be need to confirm, particularly for location 7.

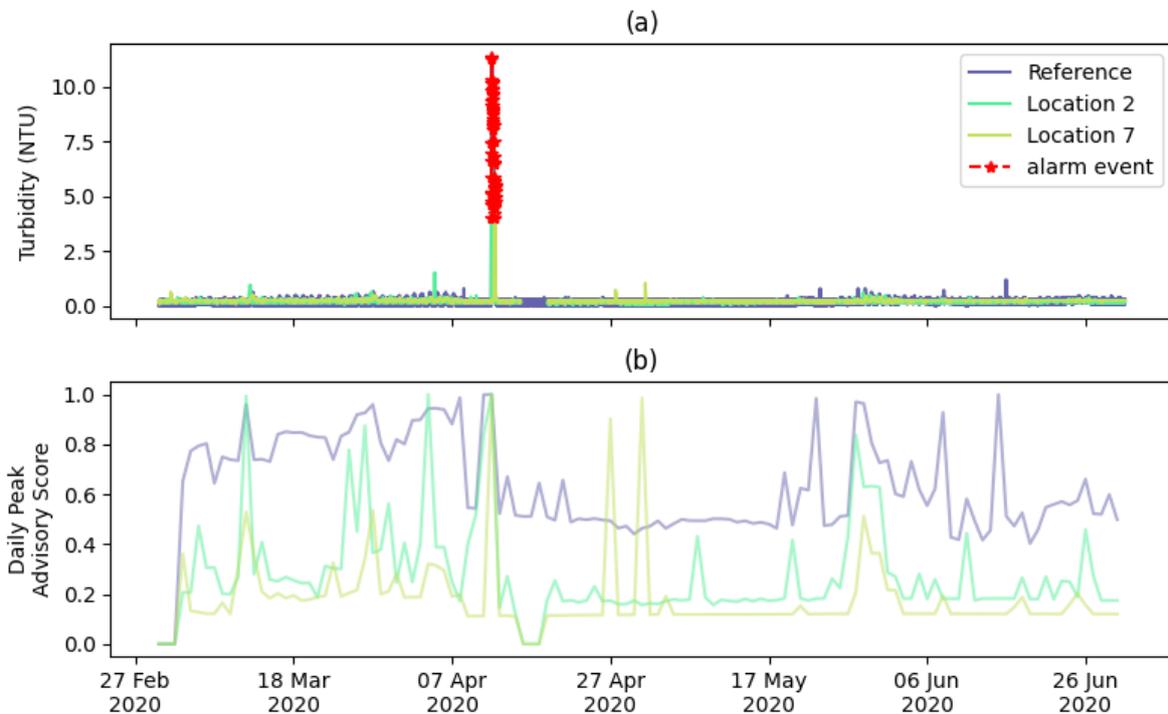


Figure 6.24. Turbidity time series at reference and locations 2 and 7, with alarm event marked (a) and daily peak advisory scores at each location (b).

After attempting to manually interpret sensor connectivity with turbidity data, the chlorine data from the same locations was used for cross-correlation, as this has been proven to be an effective method in Chapter 4 and the first example in Section 6.2. The first step involved performing a data quality assessment on the available chlorine data. As the event featured in this example occurred a week later than that featured in the first example, the same period between April and July 2020 was used to examine connectivity, with the data quality results shown in Figure 6.25 mirroring those previously shown in Figure 6.8, although with a

different reference sensor. Cross-correlation was subsequently applied to understand the relationships between the reference chlorine sensor and the remaining 10. This was done on a sliding windowed basis, using 4-week long windows and calculated every day, with the median PCC shown in a bar chart in Figure 6.26. This bar chart shows two highlighted locations, 2 and 7, as having median PCC values of 0.89 and 0.93, respectively, exceeding the significance threshold of 0.7. This supports the two locations already identified as hydraulically connected, with no other locations demonstrating having significant connectivity.

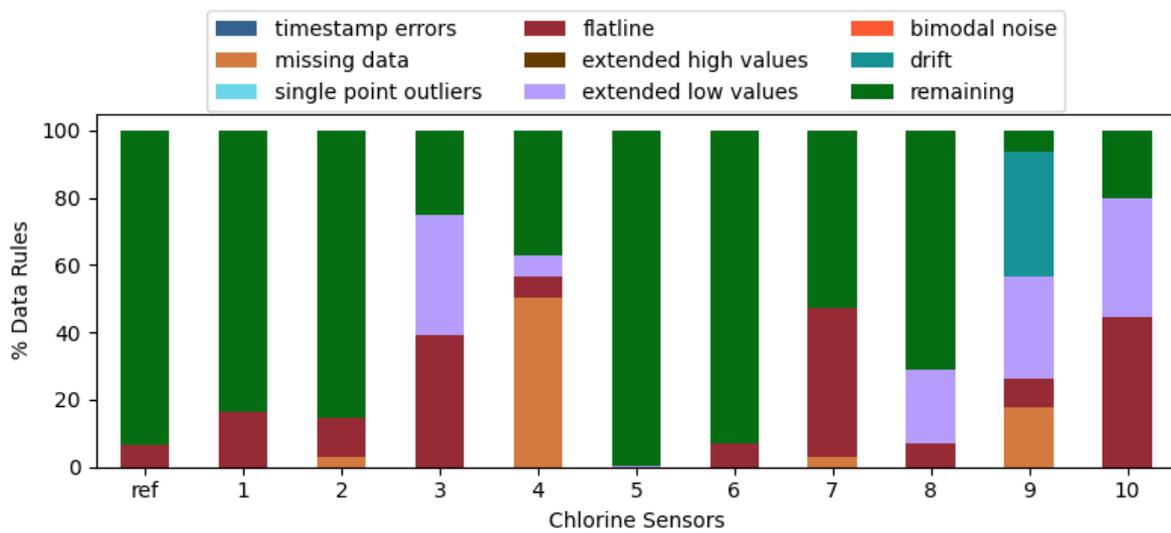


Figure 6.25. Data quality rules applied to 11 chlorine time series, with ‘ref’ referring to the chlorine at the location in question.

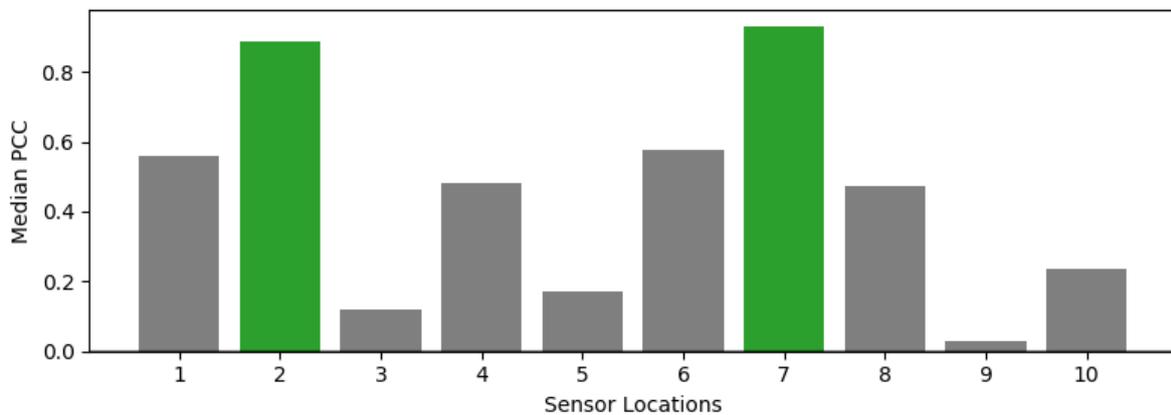


Figure 6.26. Median correlation coefficients between reference chlorine data and 10 other chlorine sensors.

The sliding cross-correlation results are shown along with the chlorine time series in Figure 6.27. The middle plot showing the sliding correlation coefficient results for both pairs, with the relationship between the reference and location 2 being strongly correlated apart from a period that was impacted by a sudden shift in the reference chlorine concentration on the 19th May. The median offset between these locations, using only results where the PCC was above 0.7, was found to be -3 hours and 45 minutes, meaning location 2 is determined to be upstream of the reference location. The cross-correlation between the reference and location 7 was found to be consistently highly correlated, with a median offset calculated as 4 hours 45 minutes, meaning location 7 was estimated to be downstream of the reference location. That this did not see the drop in correlation seen between the reference and location 2 is due to location 7 also seeing a chlorine drop on the 19th May. It is likely that both of these sensors were serviced as part of required regular maintenance that included recalibration on this day which can cause sudden changes in baseline values. This demonstrates a potential method to identify servicing without access to company operational records and also highlights a challenge for analysis. In this case using enough windows can overcome step changes like this, as well as using the median values, which is less impacted by dips in correlation like that seen between the reference and location 2.

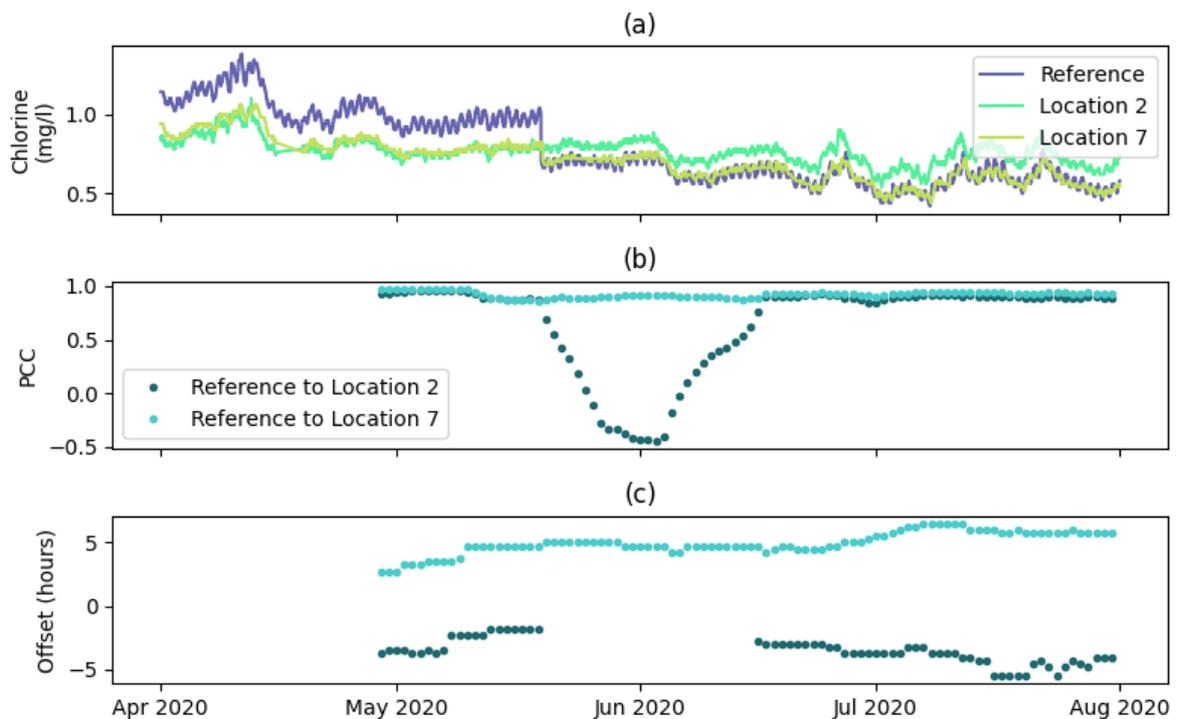


Figure 6.27. Chlorine time series at reference and locations 2 and 7 (a), sliding cross-correlation PCC's (b) and sliding offsets (c).

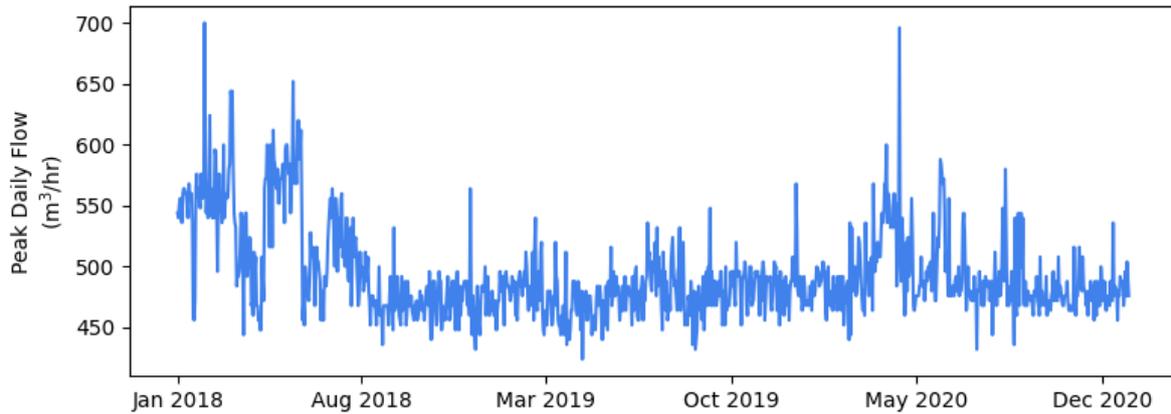


Figure 6.29. Mains peak daily flow from 2018 to 2020.

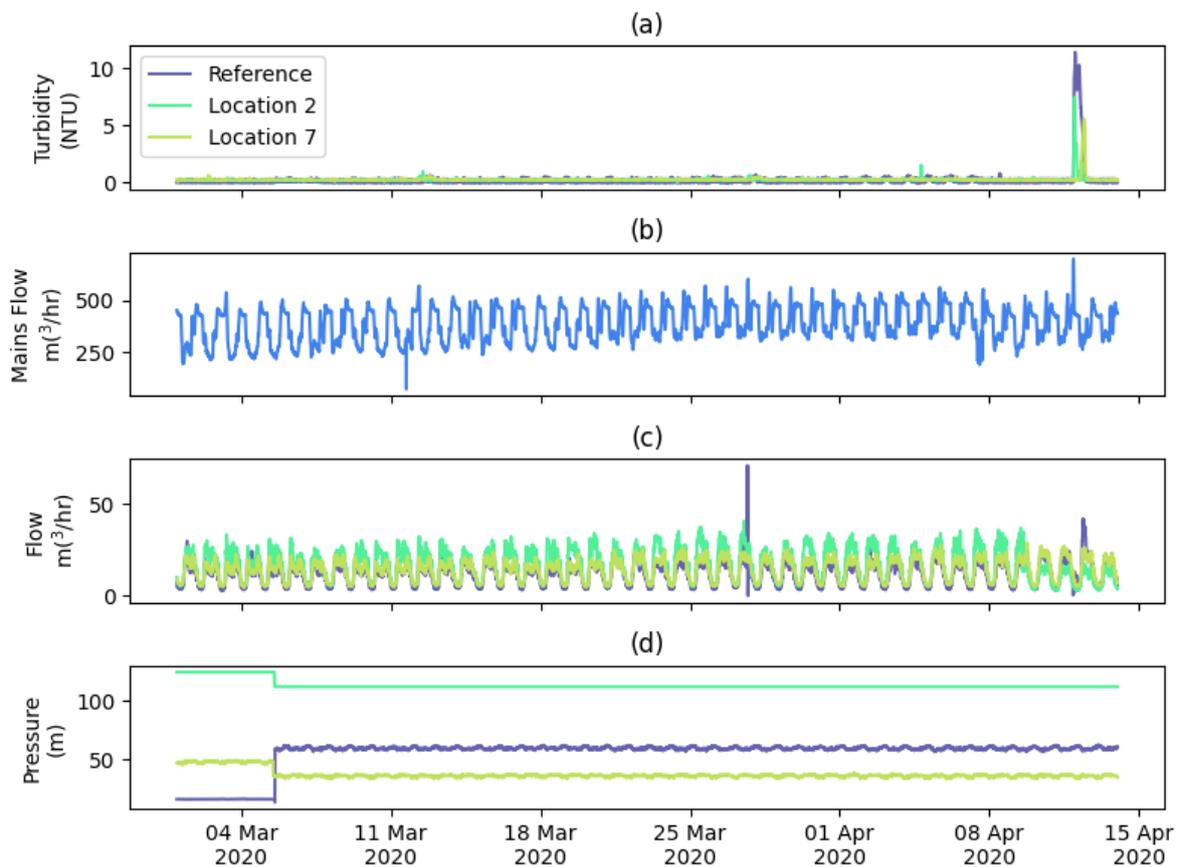


Figure 6.30. Turbidity at the three locations (a), upstream mains flow (b), flow at the three locations (c), and pressure at the three locations (d) in lead up to alarm event.

Reviewing temperature, pH and chlorine, plotted in Figure 6.31, a drop in chlorine and increase in pH is seen at all three locations in the days following this discolouration event, indicating changes in bulk water properties. Material flux calculations, illustrated in Figure

6.32, show that the reference location had more than double the material moving past location 7, and over ten times that of location 2. This suggests that the greatest risk posed by this event is related to regions downstream of the reference location. This alarm event had a notable effect on downstream customers, with over 130 discolouration contacts seen within 3 days, as shown in Figure 6.33, where (c) plots the daily discolouration contacts. Figure 6.34 separates the discolouration contacts into those related to each water quality sensor location, with 31 contacts found to be associated with customers downstream of the reference locations, compared to 2 and 8 found downstream of locations 2 and 7, respectively. This confirms the material flux analysis, that the network section where the reference sensor was installed represented the greatest risk. This event was featured in a DWI Chief Inspector’s report for drinking water, with an event risk index (ERI) score of 0.91, the 5th highest of the year. The cause of this event was concluded to be high demand from local industry, supporting the findings from this analysis.

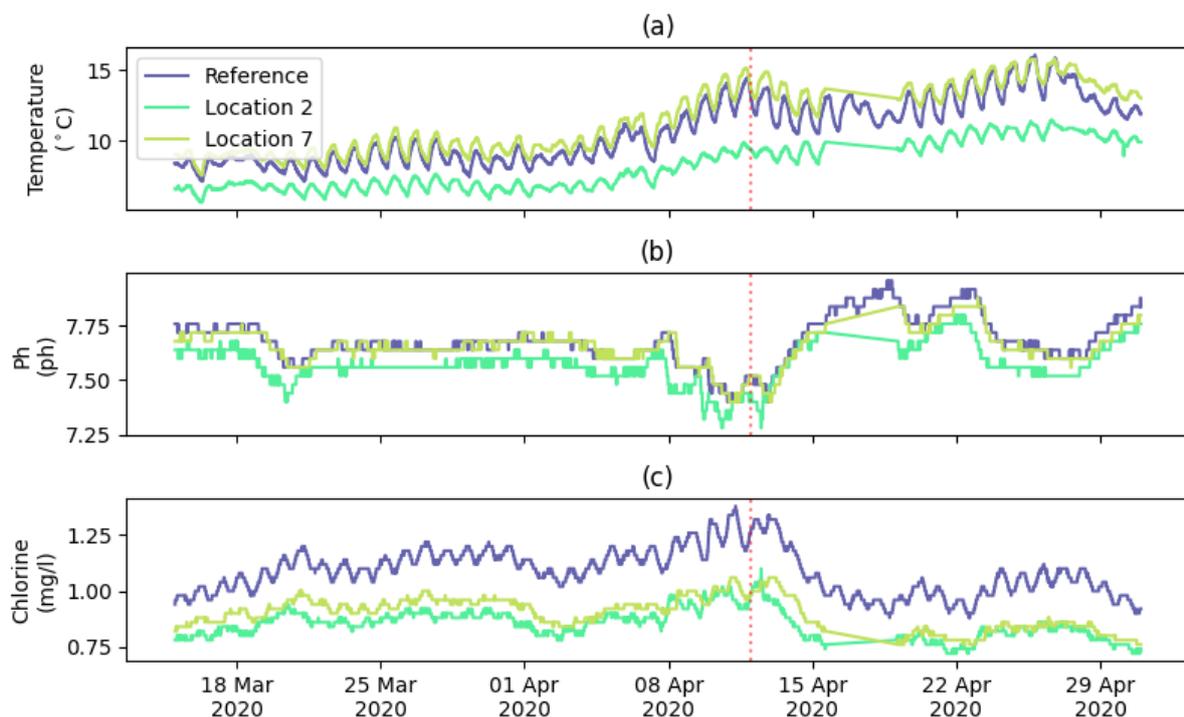


Figure 6.31. Temperature (a), pH (b), and chlorine (c) at all three locations, where the alarm event is indicated by the dotted red line.

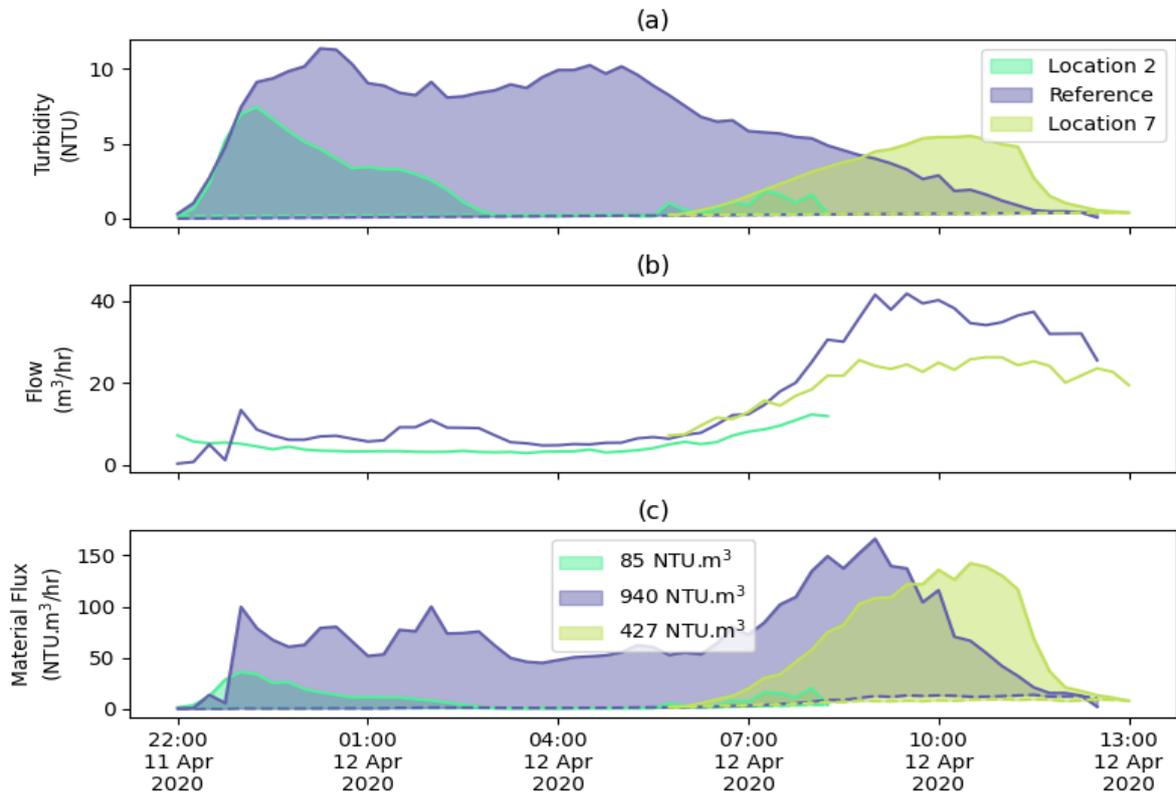


Figure 6.32. Turbidity event at all three locations with the net turbidity in (a), flow (b), and net material flux (c).

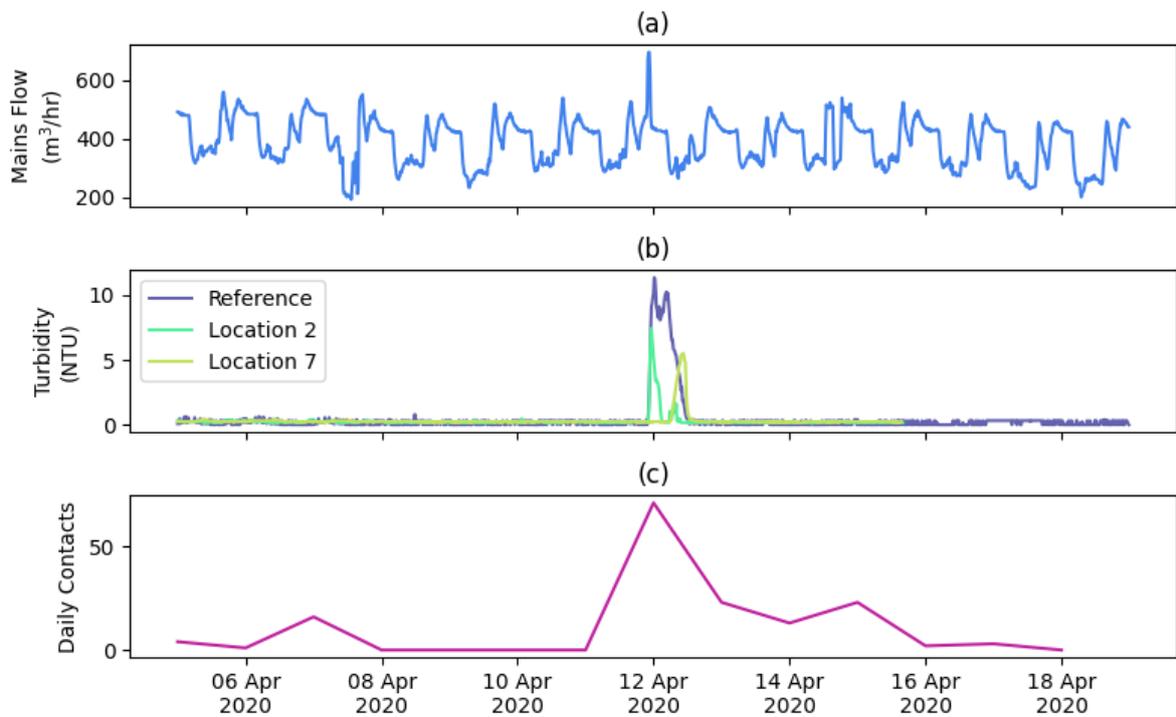


Figure 6.33. Mains flow (a), turbidity at the three locations (b), and network daily discolouration contacts (c).

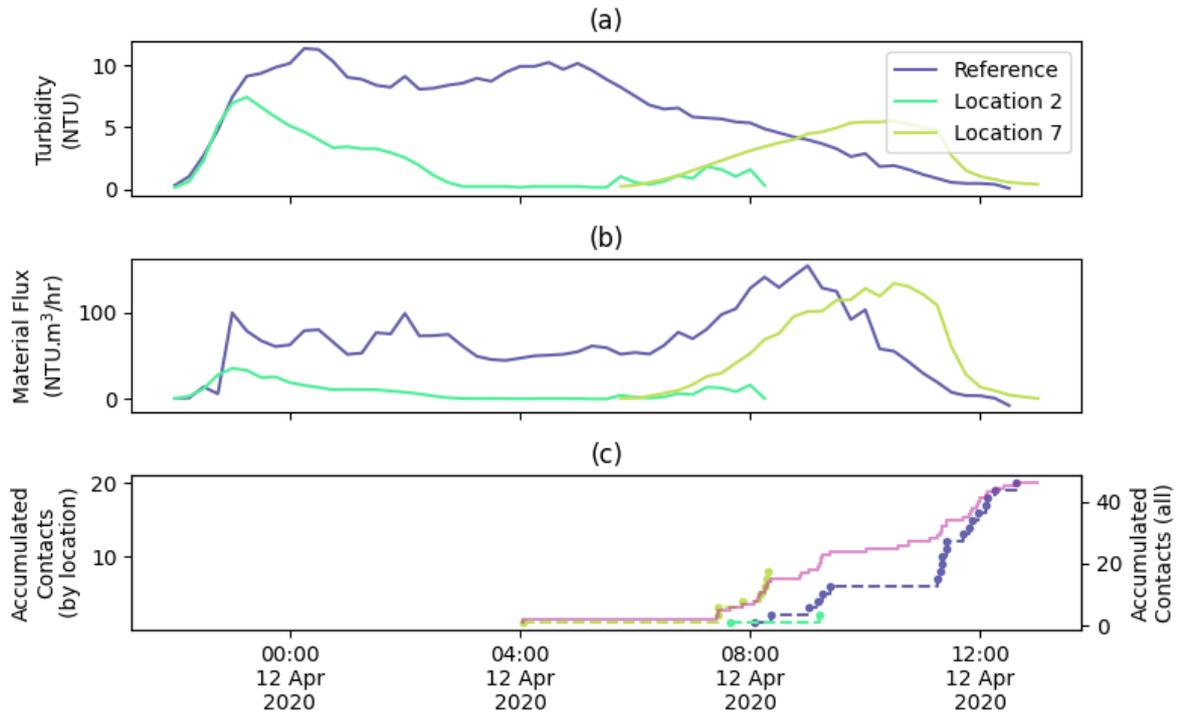


Figure 6.34. Turbidity during alarm event at three locations (a), material flux (b), and accumulated discolouration contacts per location and all (pink line and RHS y-axis) (c).

6.4 Example 3 - MPMS Discolouration and low chlorine at end of network

This example uses data from four inline multi-parameter water quality sensors, labelled A, B, C, and D connected in series with D the downstream location as shown in the simplified schematic in Figure 6.35. The location at D was towards the end of this network section, allowing for water quality changes to be examined. This section reviews the data at these locations, focusing on periods of low chlorine and increased turbidity seen in 2020 and 2021, (the latter no longer having a sensor installed at location A). The data quality rules are applied to the turbidity and chlorine data in 2020, with results shown in Figure 6.36. This shows that many of these sensors were experiencing issues during this year, with drift and flatlining very common. The sensor at location A has large amounts of missing data, due to it being removed from this location in August. The chlorine levels at C and D are both flagged as having long periods of low values (below 0.15 mg/l). A cross-correlation was done to understand the transit times between these locations, using data between May and July, with the median coefficients displayed in the heatmap in Figure 6.37 showing that only location D

did not have significant correlations with the other 3 locations. Location D was seen to have very low chlorine levels, which make it less well-suited for cross-correlation, as the signal is very flat. Transit times, using the median offset when PCC is above 0.7, was found to be 3.5 hours from A to B, 19 hours from B to C, and 23 hours from A to C.

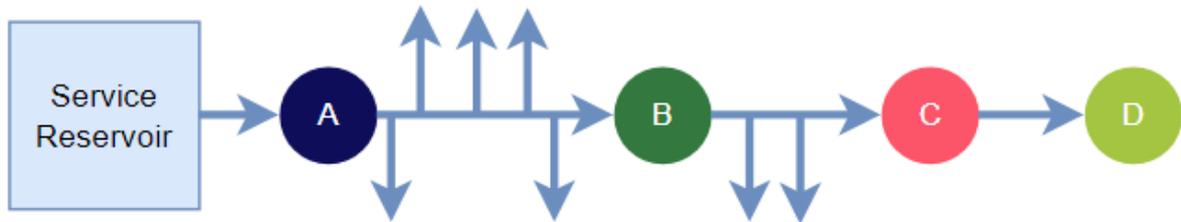


Figure 6.35. Simplified schematic showing sensor locations A, B, C and D and significant take-offs.

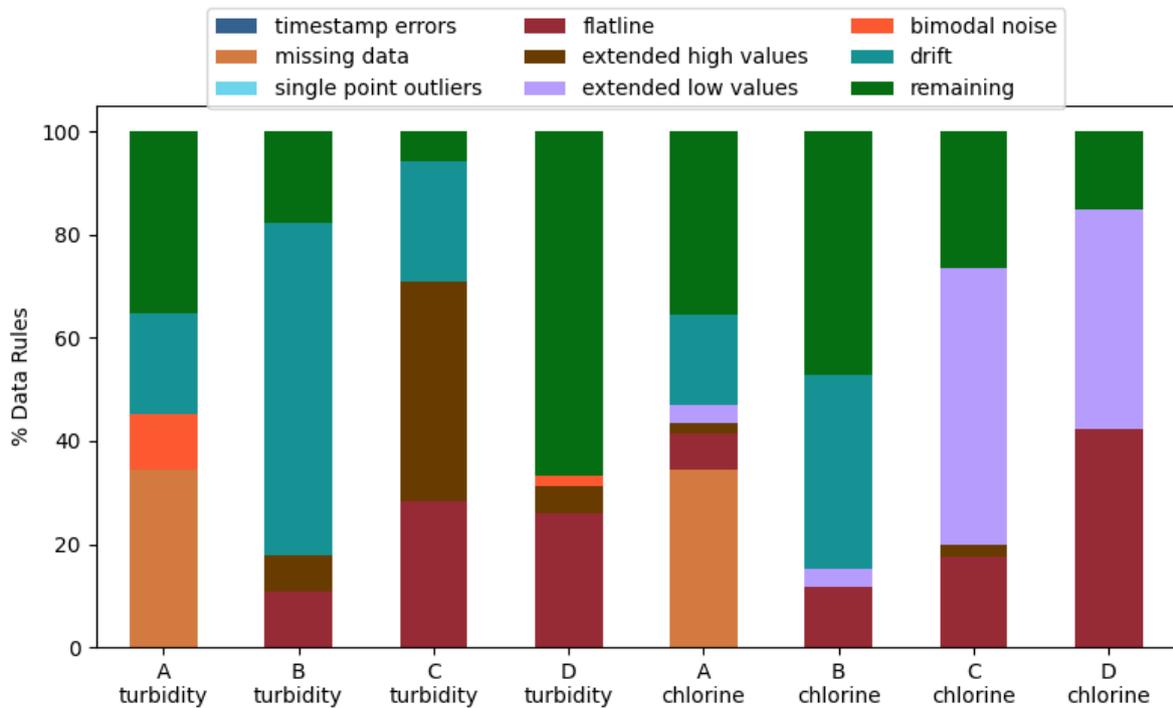


Figure 6.36. Data quality rules applied to turbidity and chlorine at locations A, B, C and D from February to December 2020.

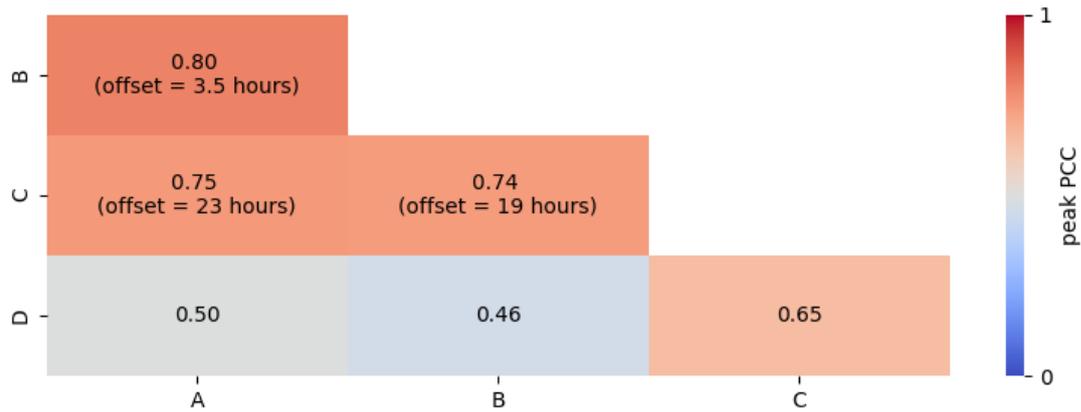


Figure 6.37. Cross correlation heatmap between locations A, B, C and D, with median offsets shown for pairs with peak PCC above 0.7.

Figure 6.38 plots the turbidity (after drift correction) and chlorine time series at each of these four locations during May and June 2020, during which multiple turbidity events are visible and the chlorine at location D is seen to drop from a peak of 0.44 to under 0.15 mg/l in about 4 days, after which a steady decline is seen until it reaches 0 mg/l. That the chlorine is seen to remain at this low level for several weeks means this network section and anything downstream may have been without protection against contamination. A site visit on the 15th June confirmed the chlorine levels to be close to zero using a handheld chlorine meter. The turbidity events seen during this period are analysed using the event scale approach, with 3 unique alarm events found at location D, compared to just one at locations A and C and none at B. Location D also had 11 unique alert events, compared to 2 at C and none at A or B. This is evidence that water quality is increasingly deteriorating as it travels downstream this network section. These events are shown in Figure 6.39, with (b) plotting the rolling daily peak advisory score at each location. This bottom plot shows that the daily peak advisory score at D was close to maximum from 24th May until the 3rd June, reflecting an increase in turbidity events of all sizes at this location. Therefore, the turbidity events precede the drop in chlorine at D, which didn't begin until the 28th May. Such an increase in turbidity, caused by elevated particles in the bulk flow, are likely to increase chlorine interactions and hence contribute to this chlorine decay. The advisory scores at the three upstream locations are also elevated during this period, in particular at location A.

Review of flow rate and pressure, plotted in Figure 6.38 (e) and (f), shows elevated flow rates during the turbidity event period at location A. That the chlorine decay observed in D did not coincide with a drop in pressure indicates the sensor remained connected to the network, i.e.

sample line not blocked. Reviewing the flow rate peaks, all locations (aside from D which did not have flow rate data) are seen to have elevated daily peaks towards the end of May, suggesting that an increased demand during this period led to mobilisation of material layers and this is most evident in the turbidity response observed at D. Analysis of temperature, pH, and chlorine show that the pH at location D was very noisy during the chlorine drop, indicating changes in bulk water characteristics. The temperature at location C is elevated during this period, reaching the maximum sensor setting of 20 °C, and a slight increase is seen at the other locations. Increased temperature can influence both chemical reactions within the bulk water, including increasing the rate of chlorine decay which may have been a factor here. Additionally water temperature increases are typically related to increased air temperature, as was highlighted in Background Section 2.2.3 (especially in surface water sourced sites such as this example) that can result in increases in drinking water demand. Peak daily temperatures are shown alongside peak daily flow rates in Figure 6.40, with peak daily air temperature also plotted, downloaded via the open-source Python weather library Meteostat (Lamprecht 2023). This figure indicates a period of high temperatures were seen in this area in the final week of May which corresponded to this period of chlorine decreasing and multiple turbidity events.

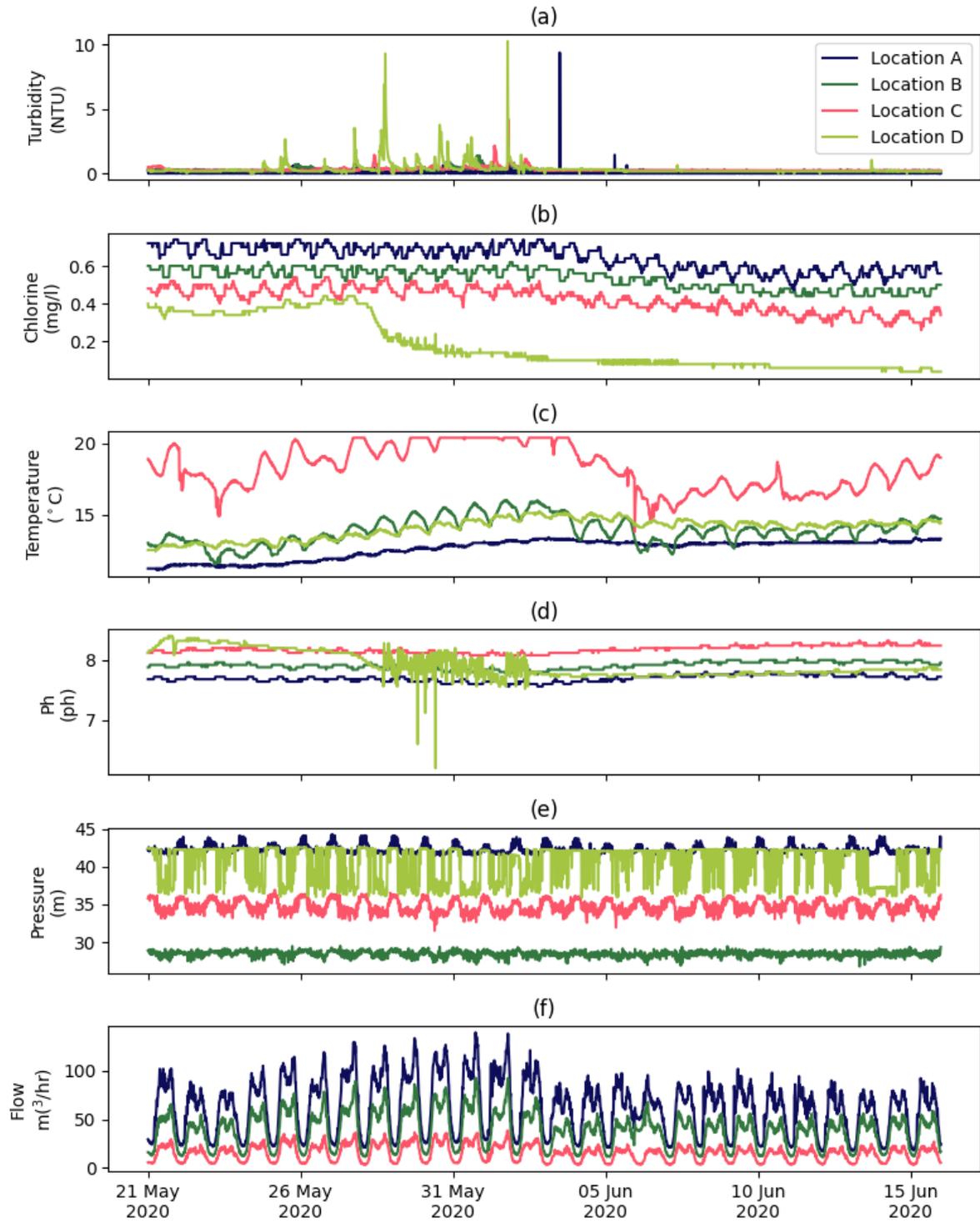


Figure 6.38. Turbidity (a), chlorine (b), temperature (c), pH (d), pressure (c)and flow (f) at locations A, B, C, and D.

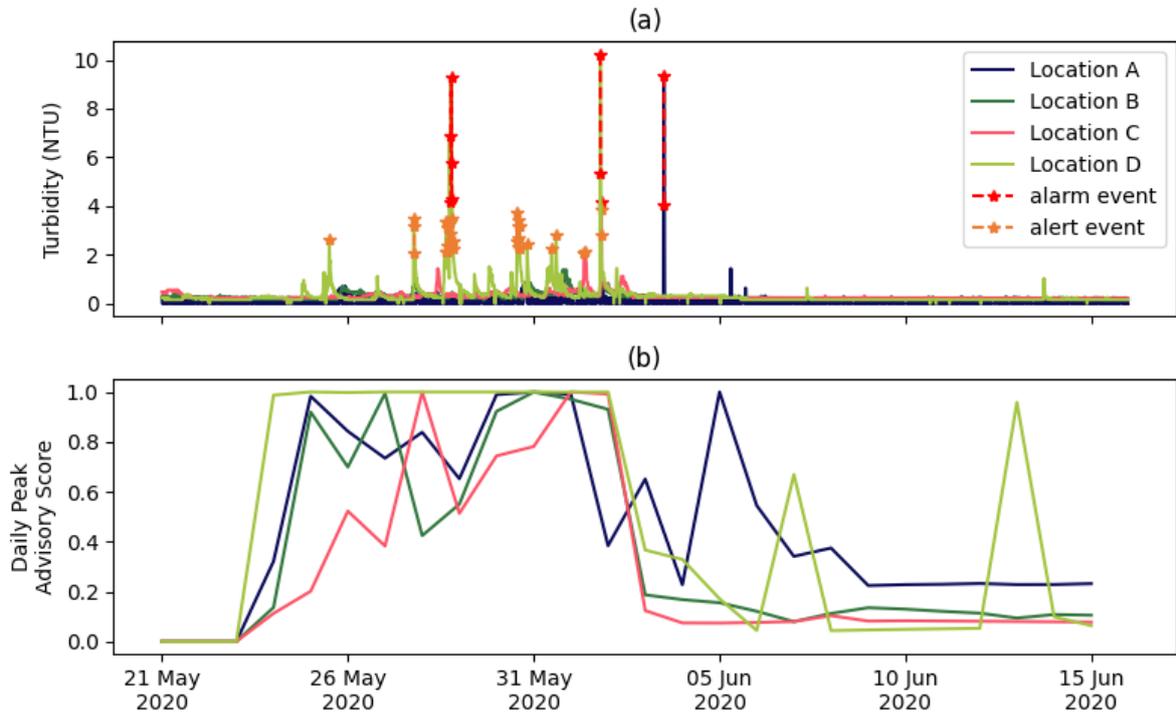


Figure 6.39. Turbidity at locations A, B, C and D, with alert and alarm events marked (a), and rolling daily peak advisory scores for the same four locations (b).

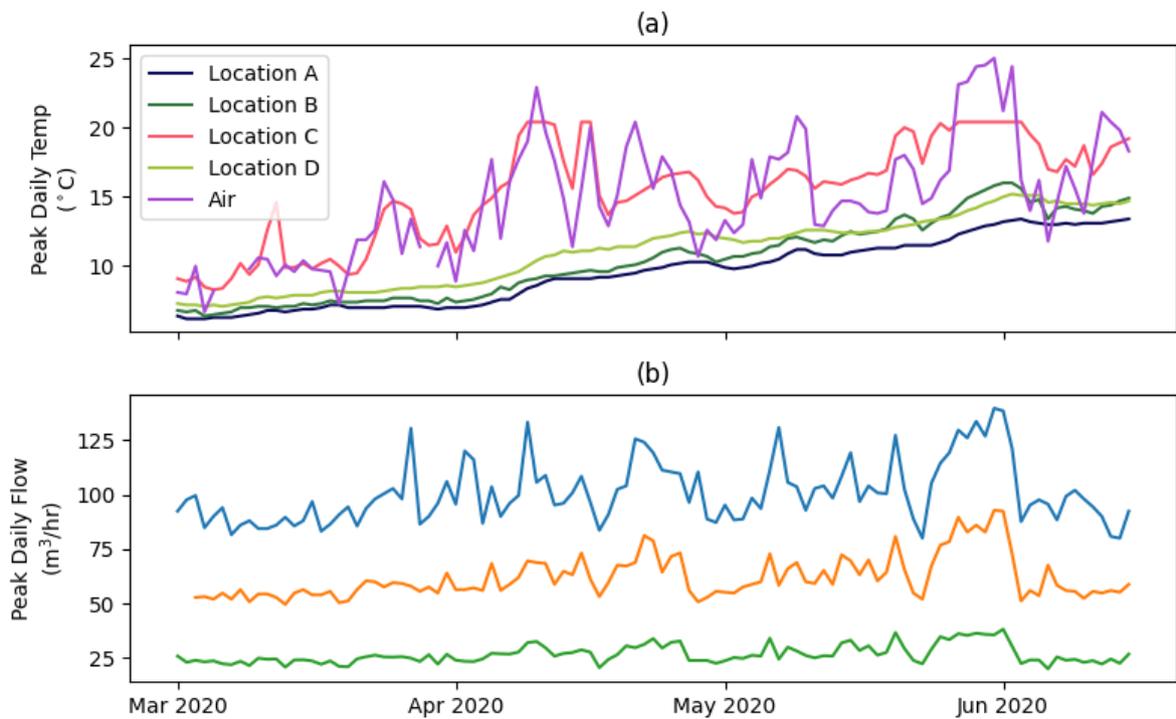


Figure 6.40. Peak daily temperature at locations A, B, C, D and air (a), and peak daily flow at A, B and C (b).

Figure 6.41 is a plot of all six parameters at locations B, C, and D from April to June 2021 (the following year, with the sensor at location A no longer available). The plot shows another reduction in chlorine residuals that appear anti-correlated to the increasing water temperatures. The pH levels at locations D are above 9 until the middle of May, before a period of noise preceded a drop below 8. This period features two alarm turbidity events seen at both B and C on April 25th and June 7th, as shown in Figure 6.43, with two additional alarm events found at location C. No alert or alarm events were detected at location D this time. The two alarm events seen at both B and C are associated with increased flow rate, indicating hydraulic causes, with the second occurring during a period of elevated water temperature mostly likely increasing consumer demand.

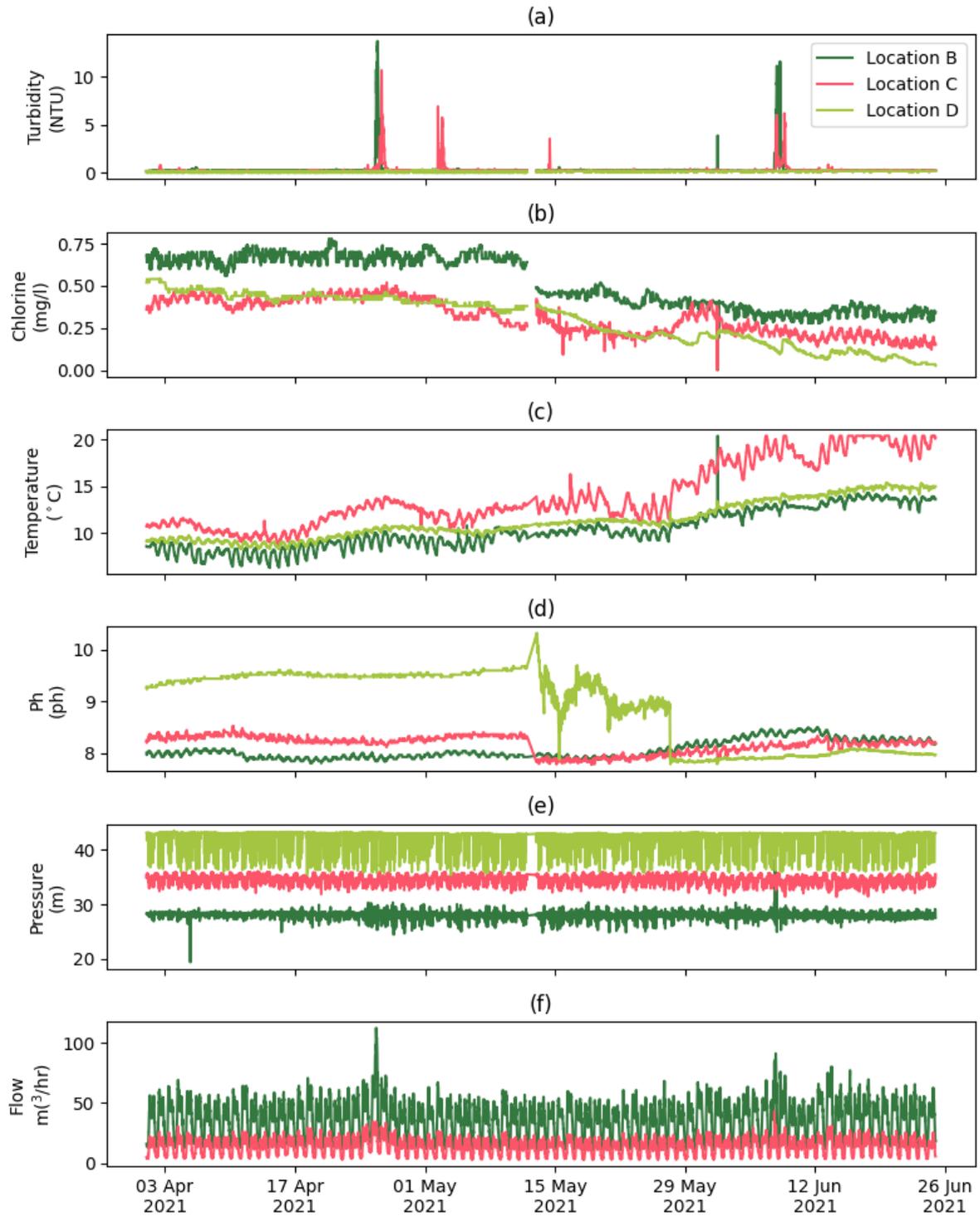


Figure 6.41. Turbidity (a) and chlorine (b) at locations B, C and D, and flow at locations A, B and C (c) for April, May and June 2021.

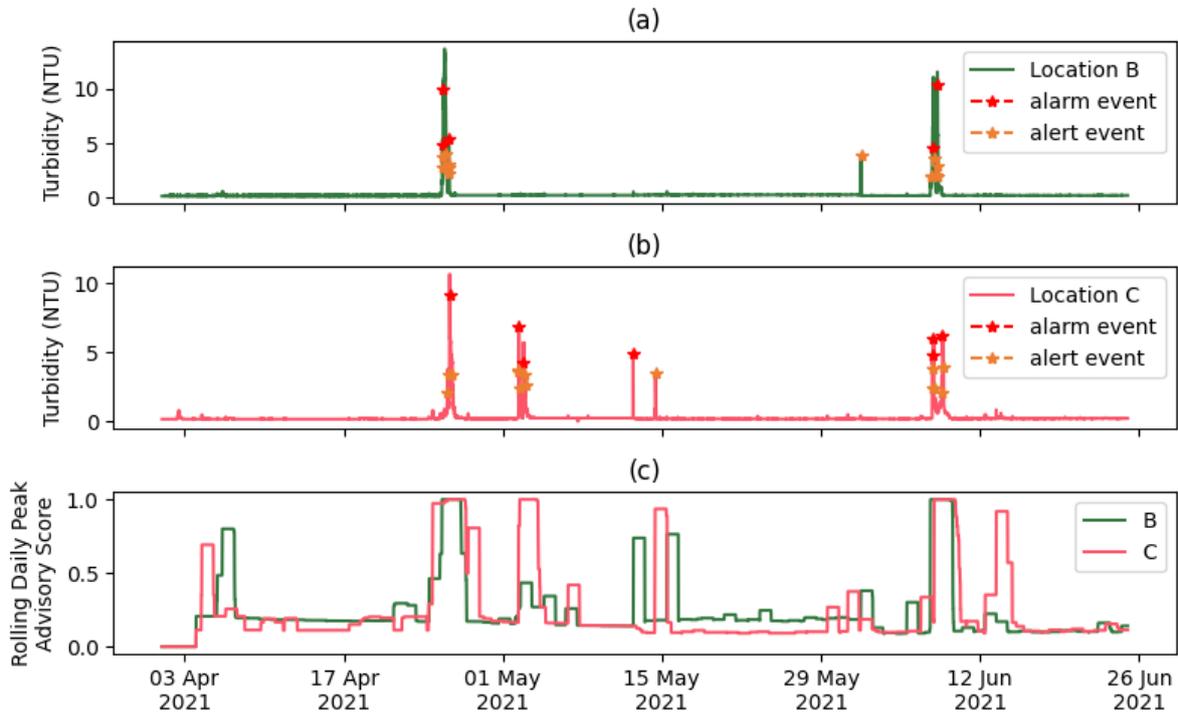


Figure 6.42. Turbidity at location B (a) and C (b) with alarm and alert events marked, and rolling daily peak advisory scores at the same locations (c).

Analysis of the two alarm events seen at both B and C (April 25th and June 7th) using material flux is shown in Figure 6.43 and Figure 6.44, respectively. Both events see significantly more material passing location B compared to C, indicating that upstream of B is the source of material mobilisation. The event on April 25th (Figure 6.43) only sees a turbidity increase at C around 11 hours after the initial increase at B at 15:00. The following day with the increased daily morning demand sees a second wave of material moving past B, and later C. That the majority of the 1127 NTU.m³ of material passing C occurred the following day indicates that this is when the material just upstream of B reached C. This occurs roughly 16 hours later, which is approximately in line with the estimated average transit time of 19 hours. Analysis of event on the 7th June using material flux (Figure 6.44) shows that significantly more material moved past location B compared to C (3254 versus 966 NTU.m³). That approximately three times the amount of discolouration material passed B compared to C is not possible to understand from just the turbidity data. This is the case for both events, with around 3-4 times as much material passing B compared to C. This informs that the take-offs between B and C are where much of the discolouration material is ending up, highlighting the value in analysing material flux where flow rate is available and the ability to determine source and destination of material. It appears both locations see the same first

wave of material on 7th June, meaning a hydraulically induced mobilisation occurred throughout this network section yet it did not impact the section towards the end of this network at D. That the increased flow rate the next morning (8th June) only seemed to initially mobilise material around location B suggests that some of the upstream material mobilised the previous day had accumulated over night upstream of location B, but not between B and C. This material is then seen at location C around 12 hours later, a much shorter transit time than the 19 hours average previously estimated but this can be explained but the significantly sustained higher flow rate than normal (Fig 6.45b). 15 customer contacts were associated with this event, indicating the impact on customers and hence the value in determining the cause and material source to facilitate preventative measures to mitigate future recurrence.

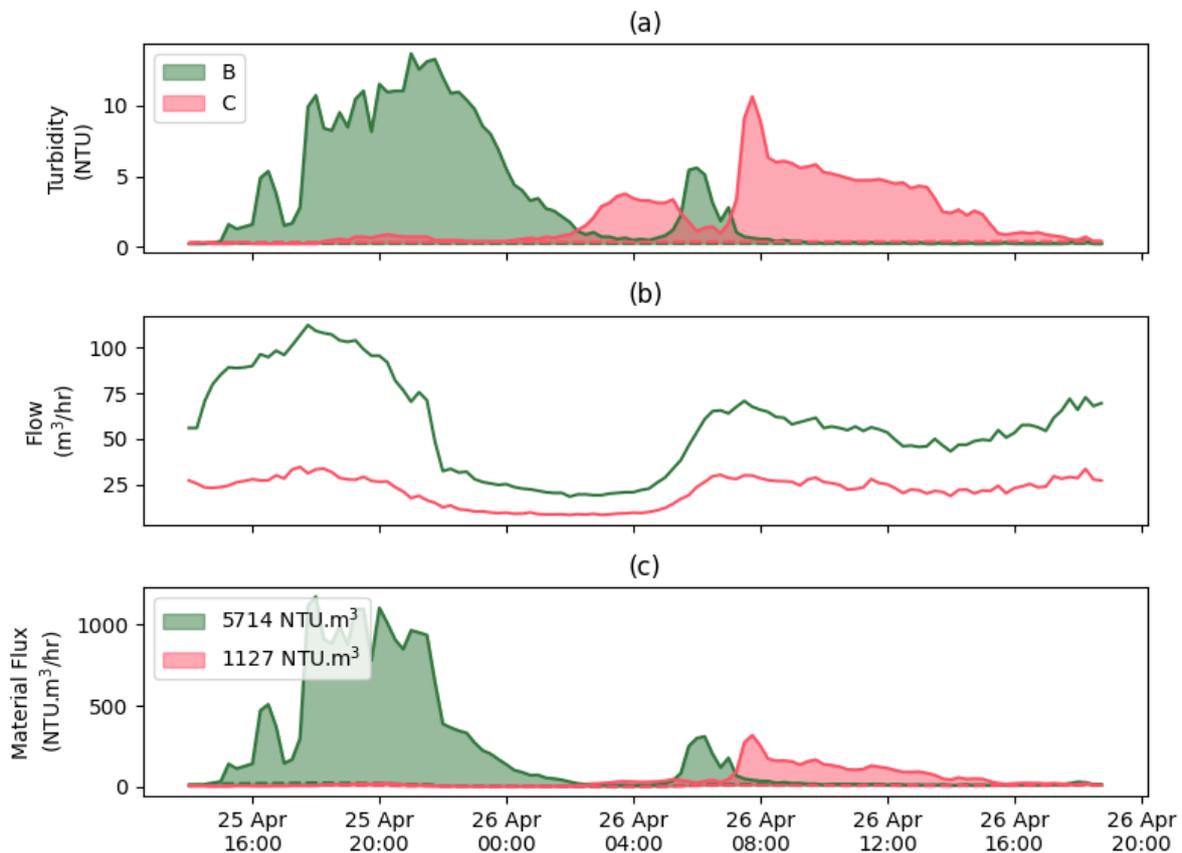


Figure 6.43. Turbidity (a), flow (b), and material flux (c) at locations B and C, with total material estimated in (c) for the alarm event on April 25th and 26th 2021.

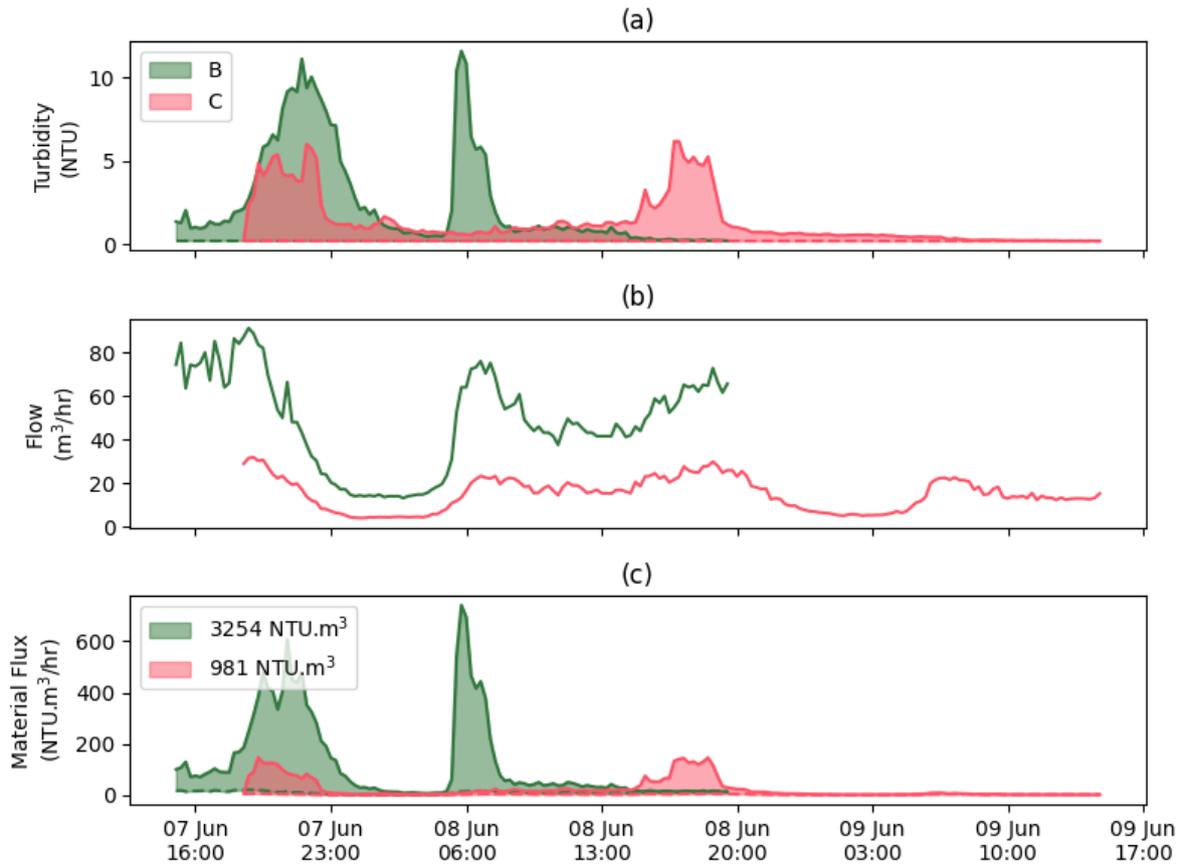


Figure 6.44. Turbidity (a), flow (b), and material flux (c) at locations B and C, with total material estimated in (c).

6.5 Example 4 - MPMS Turbidity event in a single long main

This example looks at six inline turbidity sensors on a single 70 km cast iron main with the only take-off between the 3rd and 4th monitors, illustrated as a schematic in Figure 6.45. The absence of chlorine data precludes analytical cross-correlation confirmation of connectivity or the calculation of transit times and in this case the connectivity was confirmed by the WSP. The lack of flow rate data beyond the take-off prevents precise analysis of material movement for all monitors, but flow rate data from close to the first sensor location enables some initial analysis. Figure 6.46 plots all six turbidity time series for the second half of 2021 (July to December) and multiple turbidity events are visible. As mentioned in Background Section 2.3.4, previous work has shown that the daily standard deviation is an effective metric for assessing discolouration risk in continuously deployed turbidity sensors (Cook et al. 2016) and Figure 6.47 is a bar plot comparing the daily median standard deviations at each

location. This is observed to approximately increase from locations 1-6, apart from location 3 being slightly lower than 2. Distance information provided by the WSP is displayed in the schematic Figure 6.45, with locations 2 and 3 the closest pair at just 3.8 km apart. The more disperse final three locations saw more pronounced diurnal turbidity patterns. Analysis using the event scale approach is presented in Figure 6.48, with location 6 having the most alert and alarm events, as well as the highest median peak advisory score. In general, the number of advisory, alert and alarm events are seen to increase from 1-6, indicating decreasing water quality and higher discoloration risk with distance travelled down this main.



Figure 6.45. Simplified schematic show locations 1-6 in a single straight mains.

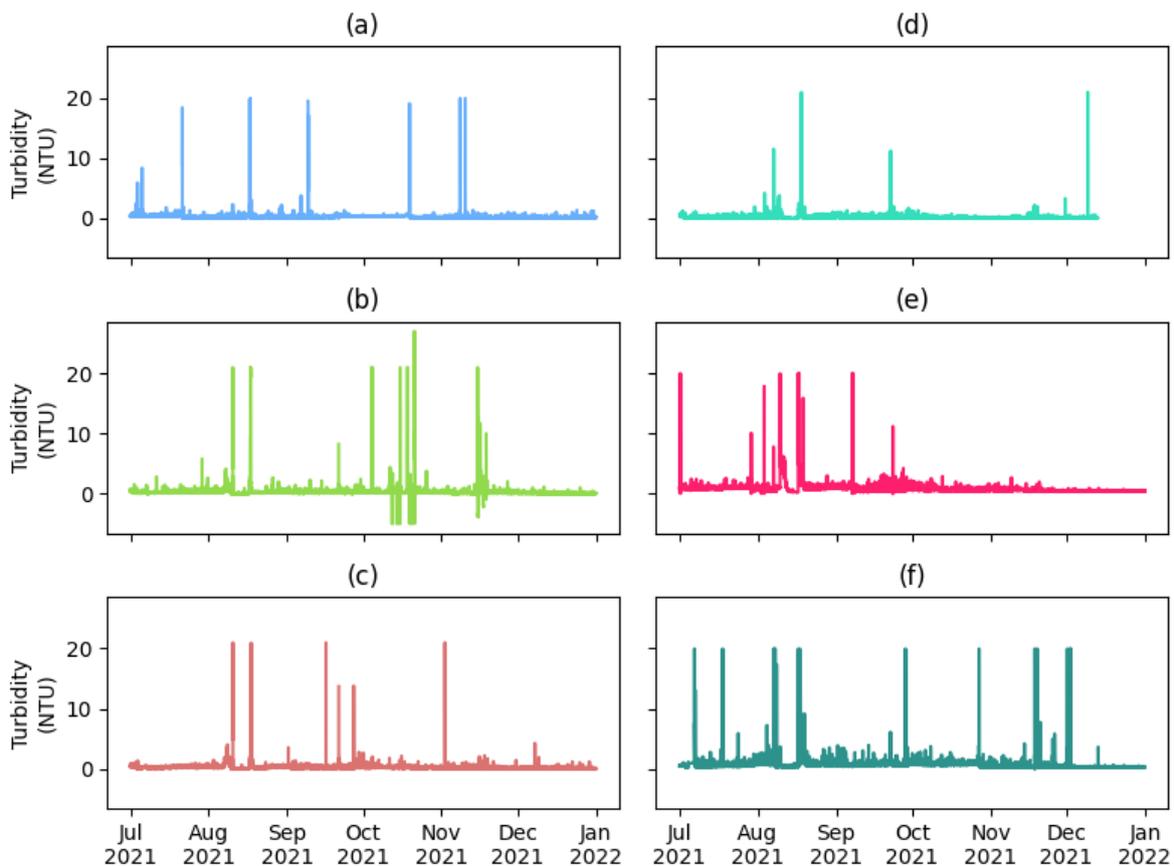


Figure 6.46. Turbidity time series from July to end of September for locations 1 (a), 2 (b), 3 (c), 4(d), 5 (e), and 6 (f).

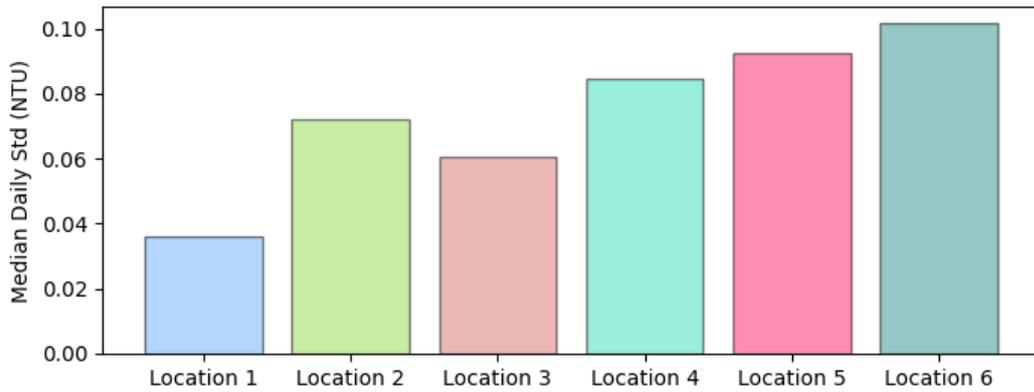


Figure 6.47. Median daily standard deviation for locations 1-6.

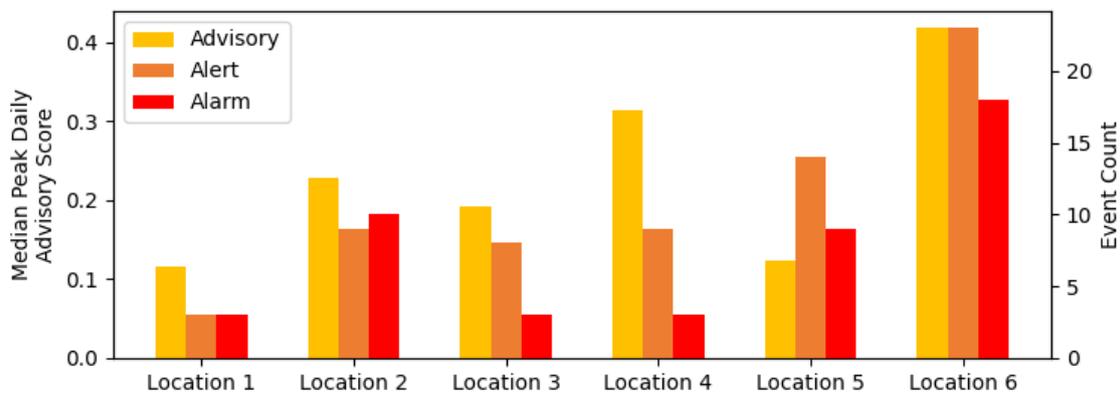


Figure 6.48. Median peak daily advisory score (LHS y-axis and yellow bars) and number of alert and alarm events (orange and red bars, respectively and RHS y-axis)

Figure 6.49 plots the mains flow rate at location 1, along with the turbidity data during a week in which flow rates were seen to increase on the 16th August from 135 to 290 m³/hr before dropping below 100 m³/hr about 2 hours later. This pattern was repeated every few hours for next few days. A clear turbidity response is seen from this first flow rate increase, and then reduces back to background levels within 2-3 days (supporting the conditioning concept covered in Background Section 2.2.4). Locations 5 and 6 appear to have two waves of turbidity events, with their second coming on the 18th and 19th of August, respectively. Visual assessment of event peaks at the locations provides approximate transit times, with the difference from location 2-3 found to be 5.5 hours, then 9 hours from 3-4, followed by two 16 hour gaps from 4-5 and finally from 5-6. That it took 5.5 hours for the event to travel from 2-3, despite these only being 3.8 km apart is explained by 3.5 hours of that time consisting of low flow rates. Using this manual tracking assessment, it appears that the first flow rate

increase has mobilised material at two different sections of this network. The first mobilisation is first seen as a single point at location 1, before moving through locations 2,3, and 4 in sequence, and reaching locations 5 and 6 in the following days, as illustrated in Figure 6.50 (a). As flow rates beyond the take-off were unavailable, the mains flow rate from near location 1 was used to estimate material fluxes. This will lead to overestimates for the downstream sensors (4,5 and 6), but is still useful to support analysis of discolouration material movement.

The second event is seen at locations 5 and 6 and shown in Figure 6.51 (a). This follows the initial flow rate increase, indicating the hydraulic impact mobilised some material located towards the end of this main. The total material mobilised during both of these events are shown in a bar plot in Figure 6.52, which suggests the second wave mobilised more material and that the section of this main close to locations 5 and 6 had a higher level of material mobilised. This analysis supports the earlier findings that more material is available for mobilisation downstream, meaning this section has a higher discolouration risk.

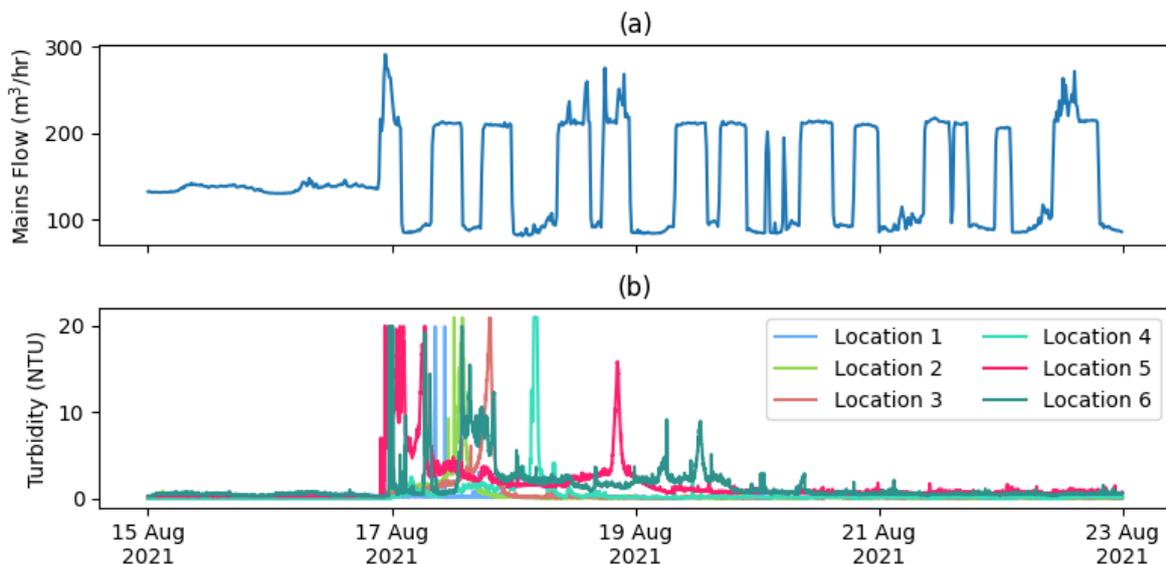


Figure 6.49. Upstream mains flow (a) and turbidity time series at each location (b) between 15th and 23rd August.

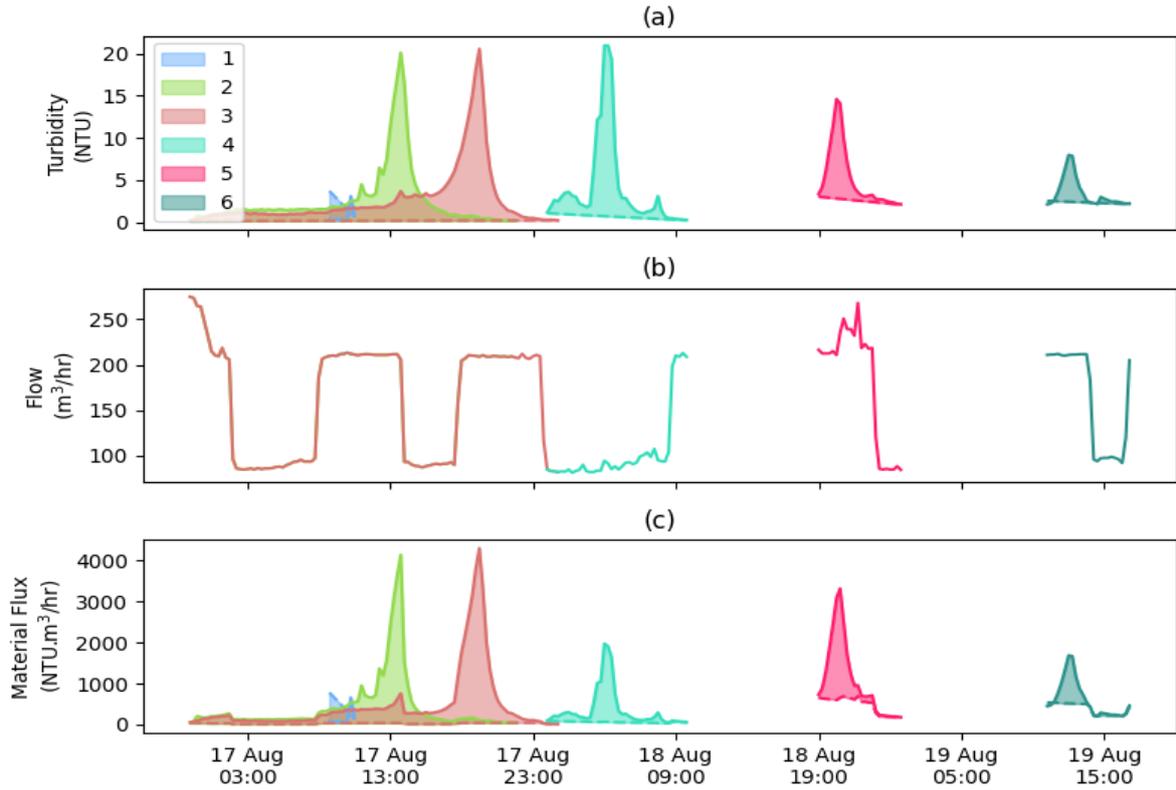


Figure 6.50. Turbidity (a), flow (b) and material flux (c) during event first wave at each location 1-6.

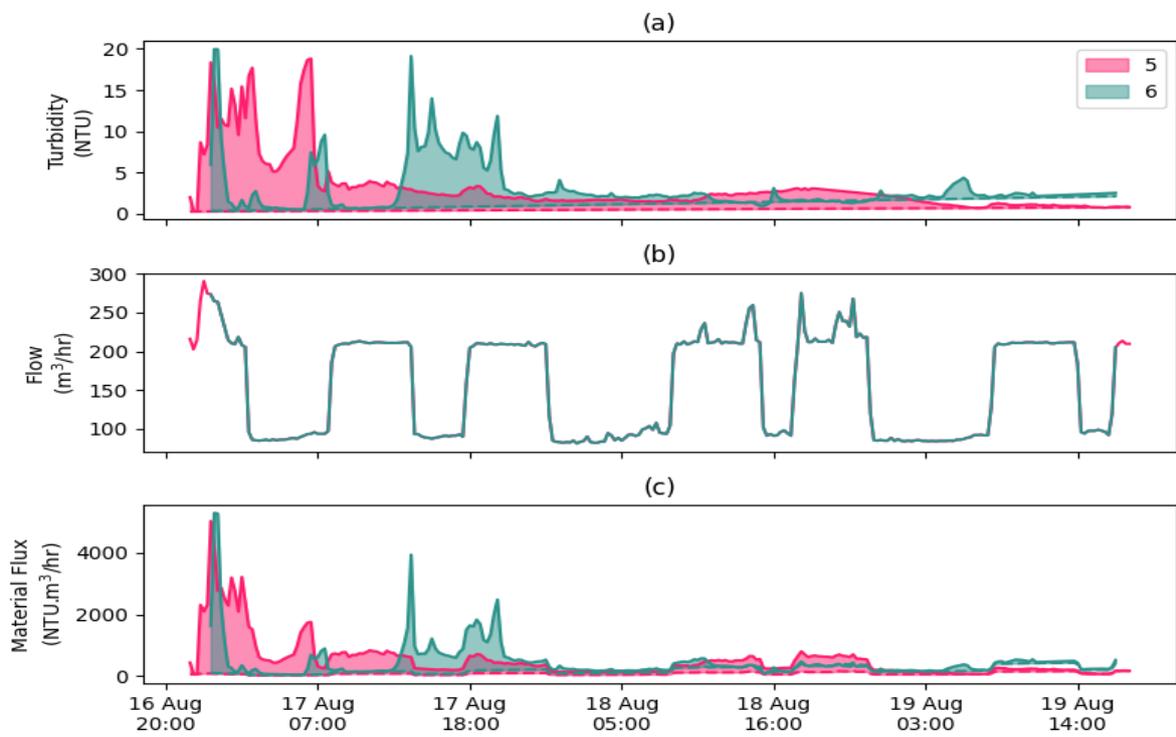


Figure 6.51. Turbidity (a), flow (b) and material flux (c) during event second wave at locations 5 and 6.

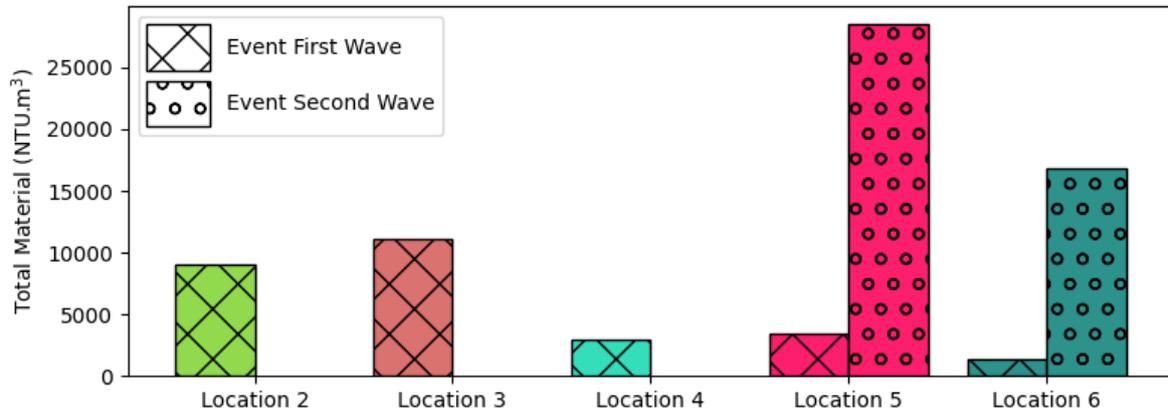


Figure 6.52. Total material passing locations 2-6 during the event first and second waves.

6.6 Example 5 – MPMS Dimensionality reduction and anomaly detection

This section explores how higher-dimensional water quality data can open up more possibilities for unsupervised anomaly detection. As highlighted in Background Section 2.4.5.1, reducing a datasets dimensions using an approach such as PCA can improve results in subsequent event detection. In this example PCA is applied to reduce the dimensions of two eight-parameter water quality sensors dimensionality and subsequently examine unsupervised anomaly detection approaches that were discussed in Background Section 2.4.5.1, such as isolation forest, elliptic envelope, local outlier factor and one-class support vector machines (OCSVM). Before these approaches, data quality assessment was performed and cross-correlation was used to find two connected locations, with location H around 2 hours 15 minutes upstream of location I. Both multi-parameter sensors were installed along a trunk mains. The two principal components are shown, with retained variance of 16% for H and 23% for I, in scatter plots in Figure 6.53, along with boundaries automatically calculated by unsupervised anomaly detection methods isolation forest, elliptic envelope, local outlier factor and one-class support vector machines (OCSVM). This process enables visual comparison of how each anomaly detection algorithm performs at splitting these datasets into normal and abnormal data. Of the four approaches reviewed, visual inspection suggests that OCSVM does a good job of capturing the majority of the data. Therefore, OCSVM was employed to detect anomalies at each of these sensor locations.

Figure 6.54 plots each of the eight water quality parameters at both locations, along with each principal component as a time series. The anomalies detected by OCSVM are included in each subplot. This approach identified two turbidity events at location H. The first coincides with a drop in ORP and, as shown in Figure 6.55, on closer inspection this event is also seen at location I, with the timings agreeing with the estimated transit time. The second turbidity event, shown in Figure 6.56, is also seen to a lesser degree at location I, though no other parameters are seen to change. As these were both below 2 NTU, these events would not have been flagged as alert or alarm events using the turbidity event scale but would have resulted in high advisory scores. However, this approach can detect events in the other parameters and a chlorine and ORP drop was detected at location I (Figure 6.57) but nothing is seen at the upstream location H. Finally, a cluster of anomalies are detected between 15th and 20th August at location I, shown in Figure 6.58. This cluster starts with a drop in pressure, ORP and pH, along with a step increase in turbidity from 0.3 to 0.47 NTU. A second larger drop in ORP is seen the next day and five days later another similar pressure drop occurs alongside a second step increase in turbidity, this time up to 0.7 NTU. That both events detected in H are also seen at downstream I and the timings are consistent confirms that the source of these events are upstream of H. The cluster of events at I that are not seen at H however indicates these are localised between H and I. These examples demonstrate the promise for PCA and OCSVM to detect different kinds of events in high-dimensional water quality time series.

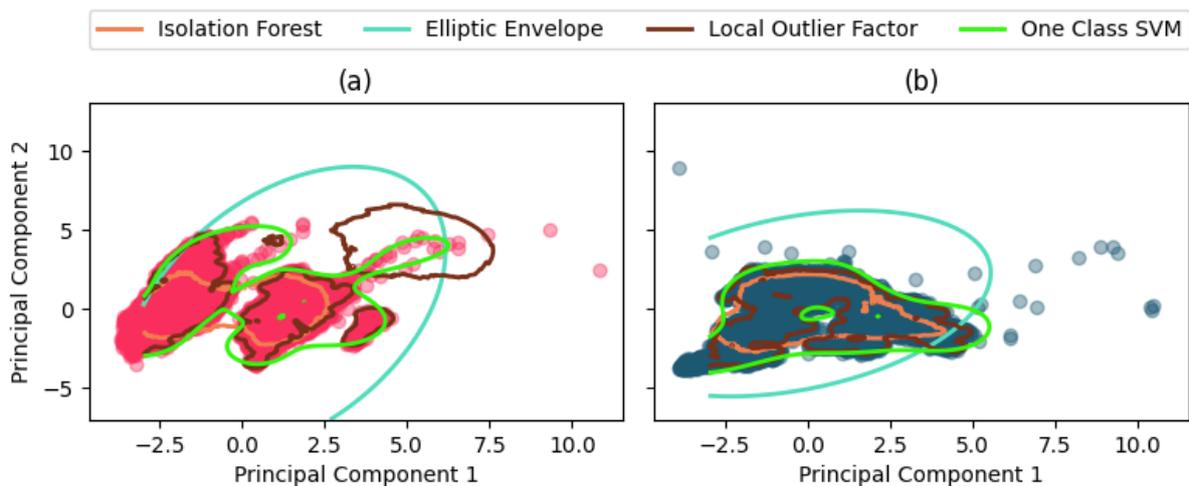


Figure 6.53. Scatter plots of two principal components for location H (a) and I (b), with boundary lines shown for unsupervised anomaly detection methods isolation forest, elliptic envelope, local outlier factor and one-class SVM.

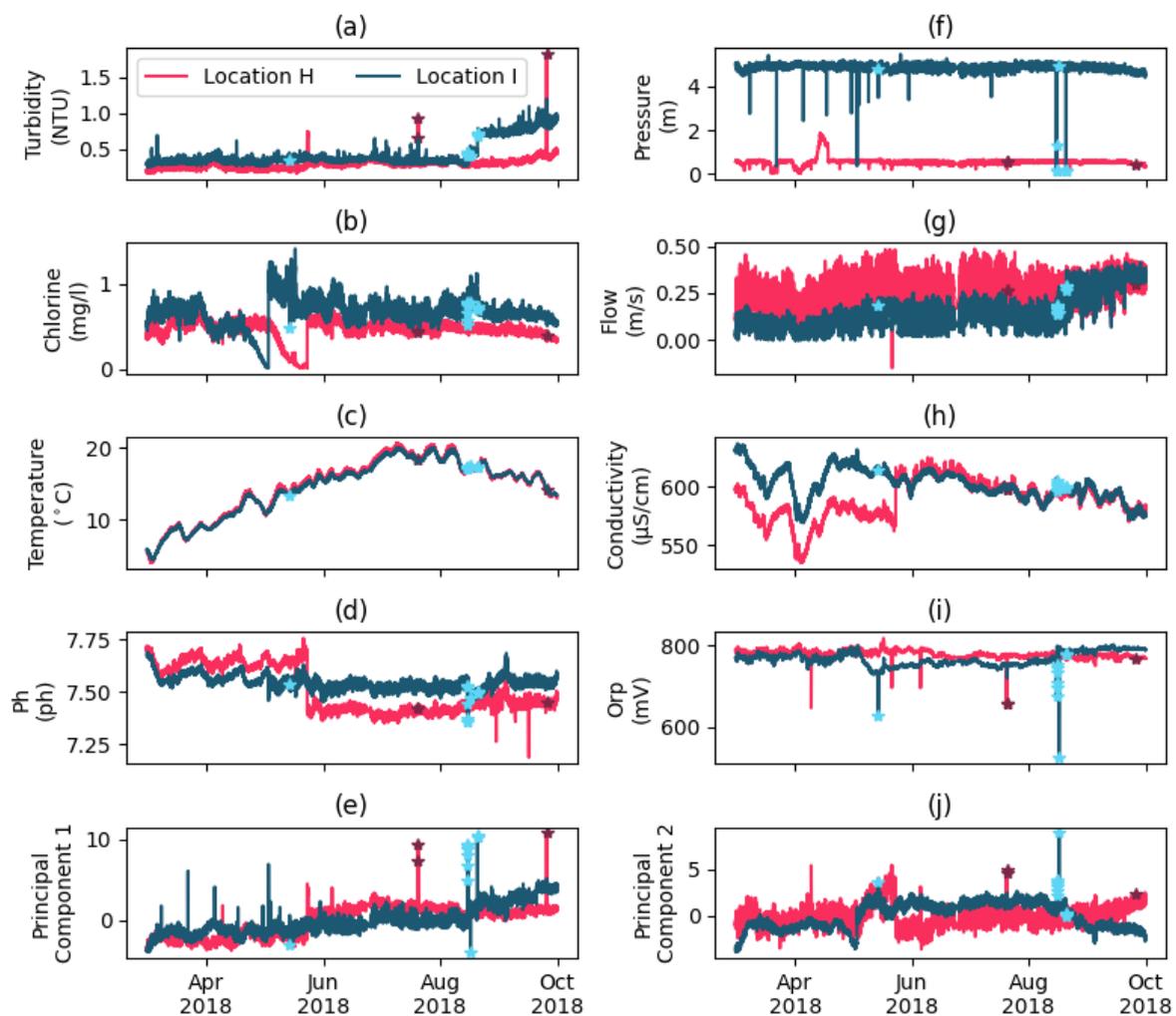


Figure 6.54. Two connected eight-parameter water quality time series from March to October s with turbidity (a), chlorine (b), temperature (c), pH (d), pressure (f), flow (g), conductivity (h), ORP (i) and two principal components in (e) and (j) respectively.

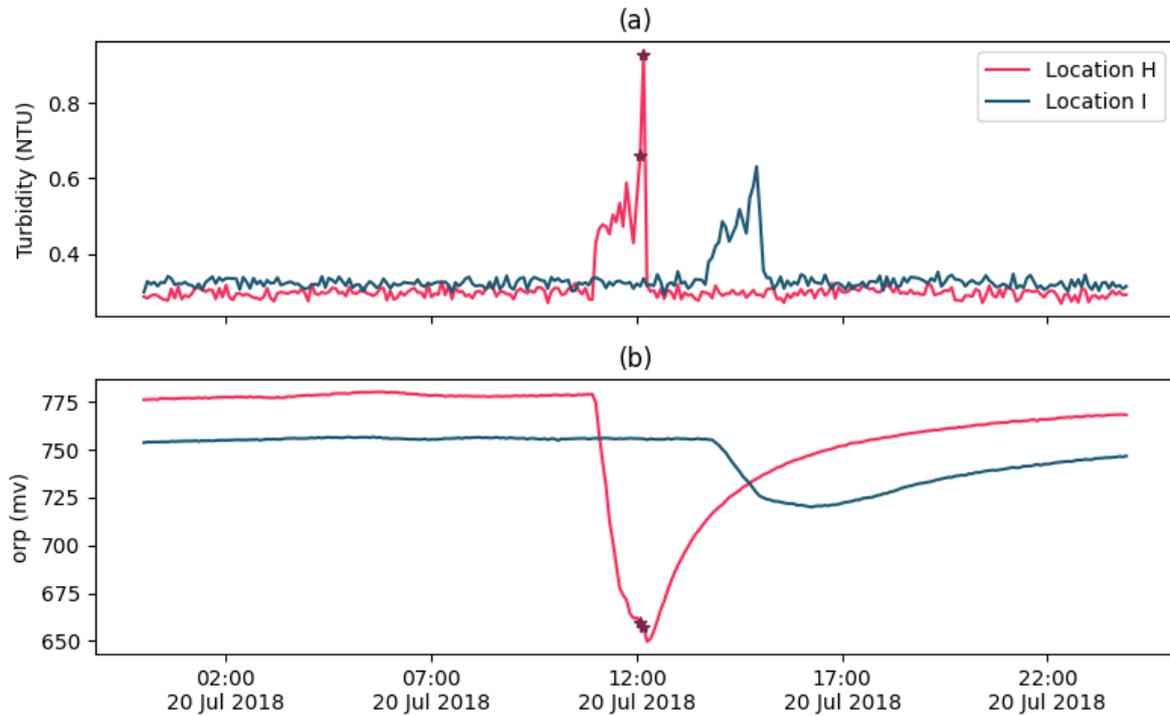


Figure 6.55. Detected anomaly at Location H shown at both locations for turbidity (a) and ORP (b).

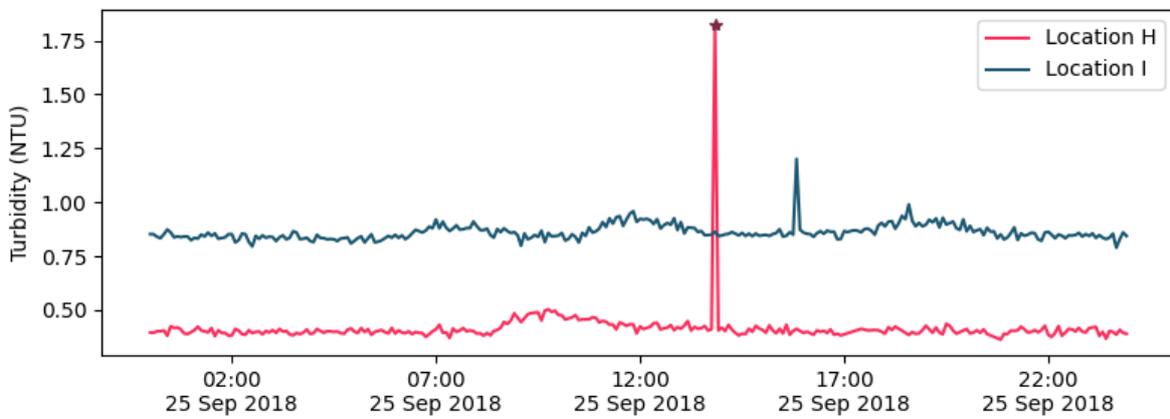


Figure 6.56. Detected anomaly at location H shown in both locations for turbidity.

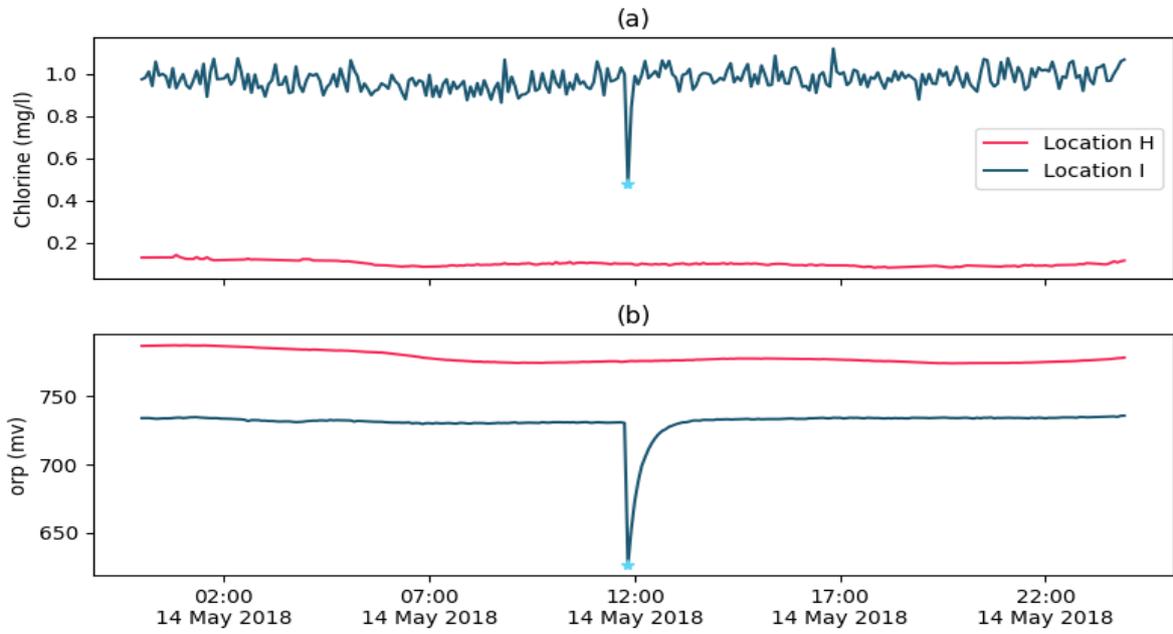


Figure 6.57. Detected anomaly at Location I shown at both locations for turbidity (a) and ORP (b).

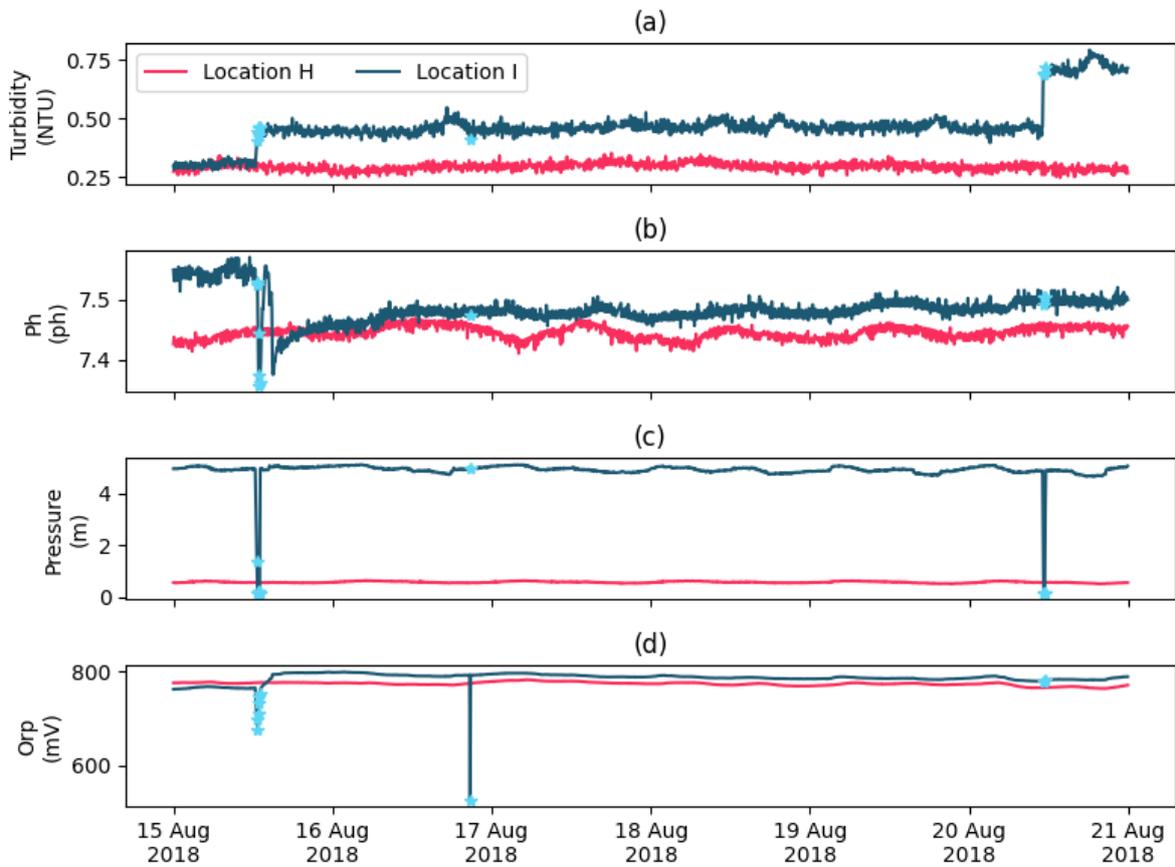


Figure 6.58. Multiple detected anomalies at location I, plotted on parameters turbidity (a), pH (b), pressure (c) and ORP (d) with location H also plotted.

6.7 Example 6 - MPMS Increase in daily turbidity profile

This example features five connected multi-parameter water quality sensors, all installed on offshoots to a main running downstream of a service reservoir, as shown in a simplified schematic in Figure 6.59. Locations X, Y and Z are the same as used in the second ‘build’ example in 6.3 (where X = Location 2, Y = Location 7, and Z = Reference), with two additional sensors later installed in this region. These sensors measured increased turbidity responses following a step increase in upstream mains flow rate in August 2021, Figure 6.60. The step increase seen towards the end of this plot attains the highest flow rates since the March 2020 spike that caused the alarm event featured in 6.3. These increased hydraulic forces cause a change in turbidity profile, with increased turbidity seen on a strong diurnal pattern at multiple locations throughout August, as seen in Figure 6.61. Unlike previous examples, this does not feature alert or alarm events, with turbidity remaining below 2 NTU throughout. The peak daily advisory score, shown in the bottom plot of Figure 6.61, was very high for all these locations at the beginning of August before reducing once the algorithm started accounting for this seemingly normalised daily fluctuations. The initial peak of this event did not occur simultaneously in X and Y, like previously for the reference and location 2 in example 6.3, instead it is only seen initially at location X, which reached just under 1 NTU at the same time as the main peak flow rate on 1st August. A larger event peaking at 1.4 NTU is then seen almost occurring identically at locations V and Y. Analysis of the transit times between these locations confirm that water would be expected to reach these locations at similar moments and previous analysis has shown the average transit time from X to Y to be 3 hours 45 mins. The delay in reaching the location Z is also expected. Location W with a very different water quality time series profile meant a transit time could not be determined and suggested only partial connectivity. The delays roughly in line with estimated transit times suggests the source of this event is upstream and propagating down the network.

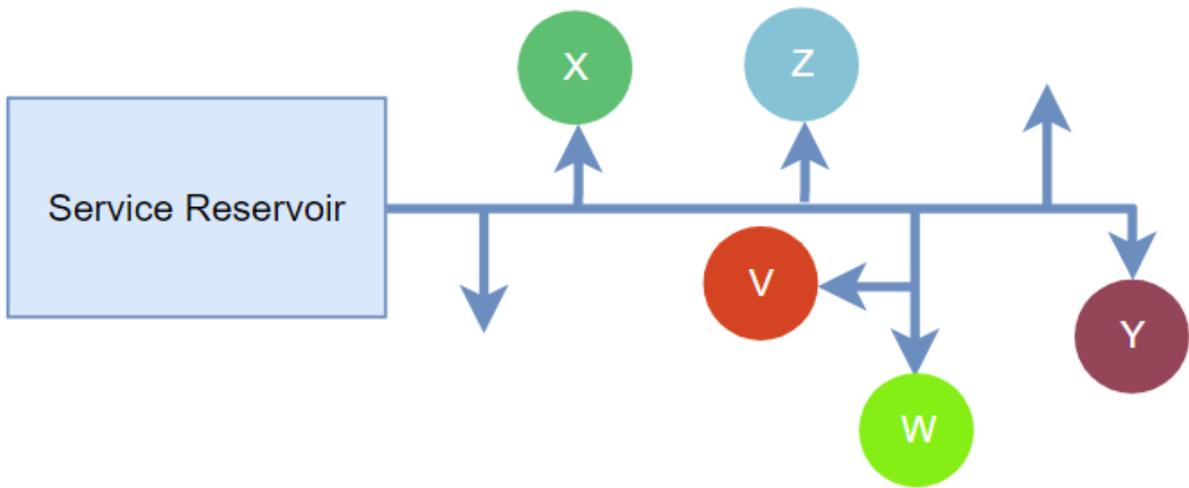


Figure 6.59. Simplified schematic showing sensors V, W, X, Y and Z (line lengths are not representative of distances).

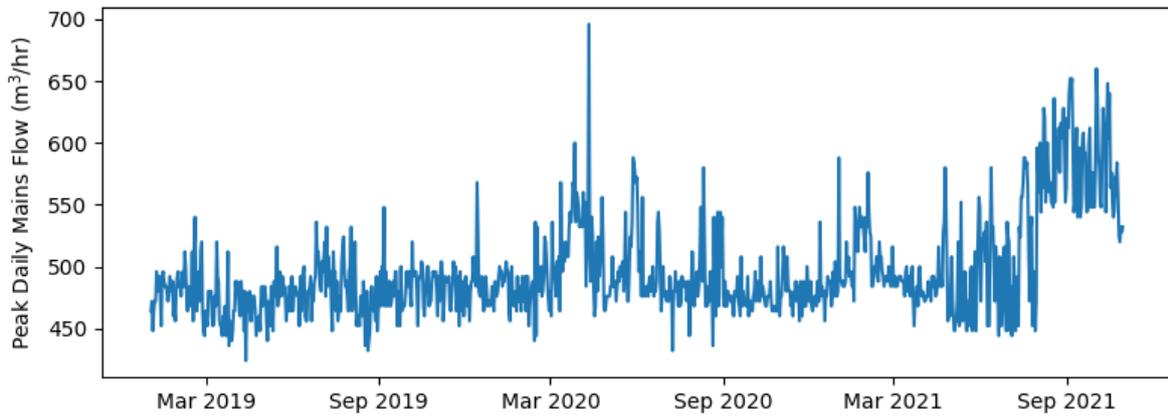


Figure 6.60. Mains peak daily flow from 2019 to 2021.

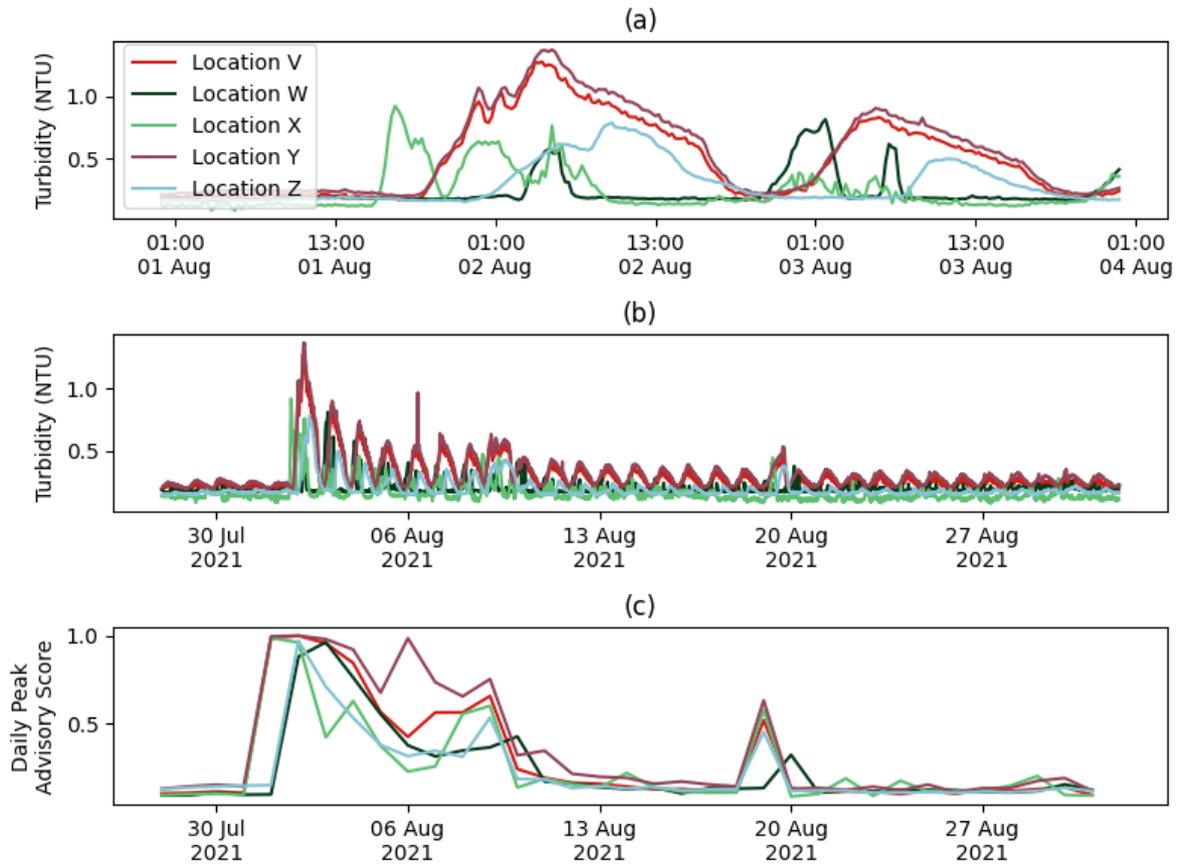


Figure 6.61. Turbidity at five locations for the first three days of August (a) and throughout August (b). The bottom plot shows the peak daily advisory scores at each location in August (c).

Material flux was calculated at each location to quantify material movement for the month of August, with net daily material throughout August plotted in Figure 6.62 along with a bar chart showing the total material moving past each location throughout this month. The flux seen in the first couple of days shows that most material is travelling down the supply offshoots related to V and Y. This suggests that the mains section downstream of X has had additional material mobilised. These locations continue to have the highest diurnal material movement, with Y having the highest total material, as shown in the bar chart in Figure 6.62(b). Just over 1000 NTU.m³ of discolouration material passes location Y during this month. This is slightly higher than the total material estimated during the alarm event in 6.3 but it is spread out over a month, with the highest in a single day just 168 NTU.m³. That location Y, previously the Reference location in Example 2 (Section 6.3), was again seen to have the largest material response indicates there is a continued discolouration risk at this location.

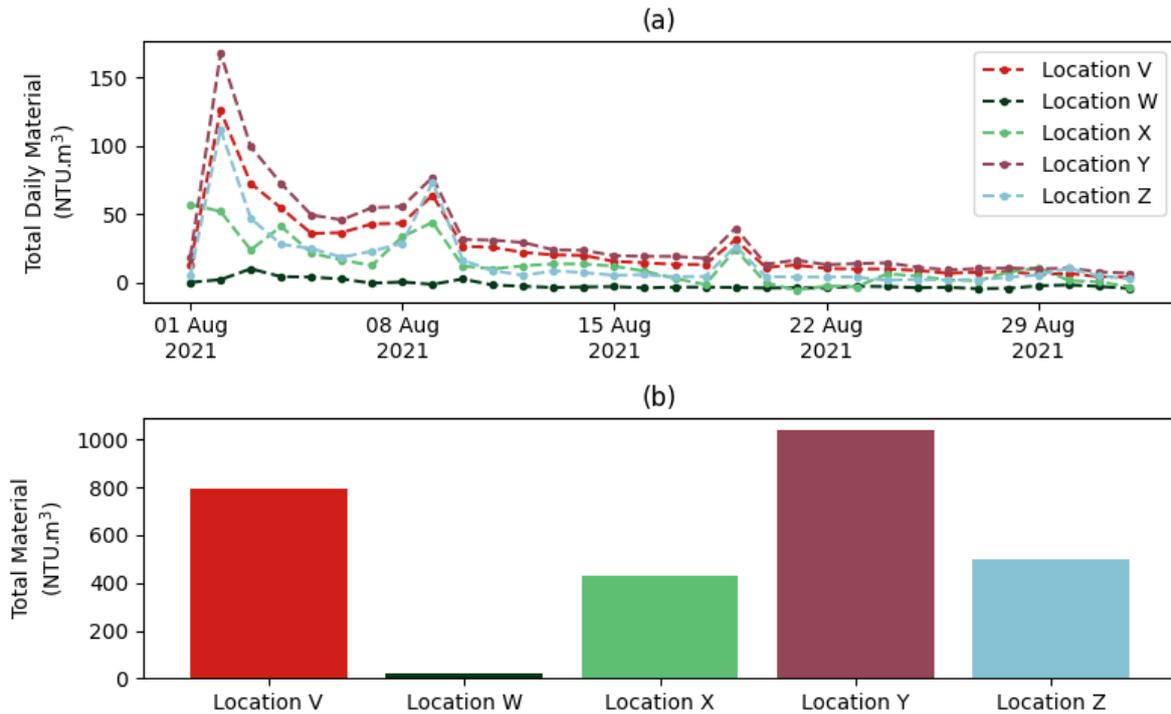


Figure 6.62. Net daily material (a) and total material passing each location (b) in August 2021.

6.8 Example 7 - MPMS Coliform failures in service reservoirs

The aim of this last example is to demonstrate again how analysis of multiparameter data from DWDS can yield deeper insight. Although this example differs from previous examples in that it uses discrete sampling data, the dataset is large, taken from 329 service reservoirs across a period of over 4 years,. The dataset analysed includes a weekly metric termed the log water quality risk score (Log WQRS), which is a proprietary metric used by the WSP and is calculated from flow cytometry data. Log WQRS is designed to determine the risk of bacteriological failure and focuses on the high nucleic acid (HNA) portion of cell populations. In general, any sample with a Log WQRS under 6 is considered clean and free of bacteriological risk, while those above 9 are more likely to be bacteriological failures. Figure 6.63 plots the average weekly Log WQRS against total chlorine, water temperature, and total coliforms across 300 service reservoirs for over 4 years. Total chlorine is used here due to some of the sites being in chloraminated systems. This plot shows a clear seasonality and inverse relationship between the Log WQRS and total chlorine, and a positive correlation

with water temperature. Additionally, it appears that most detected coliforms occur during the summer months.

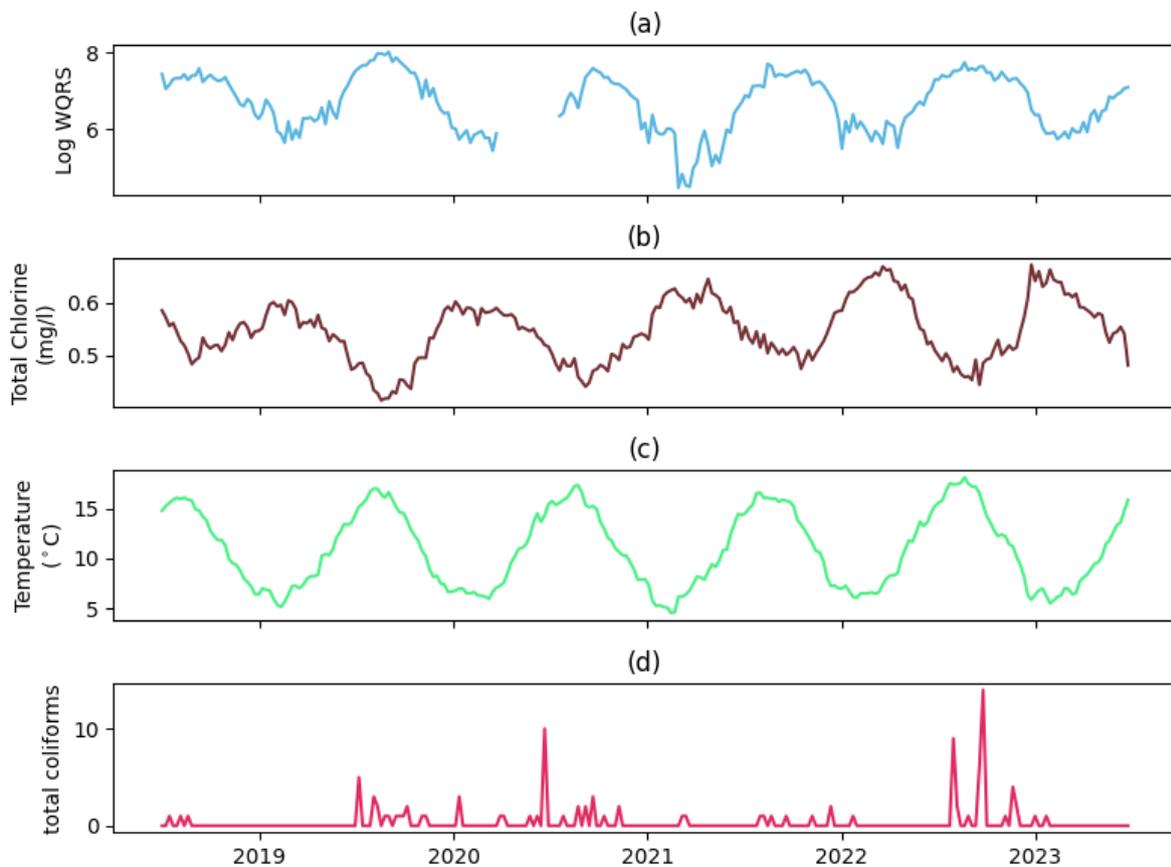


Figure 6.63. Average weekly Log WQRS (a), total chlorine (b), temperature (c) and total coliforms (d) across over 300 service reservoirs for over 4 years.

Figure 6.64 (a) is a scatter plot of the total chlorine versus temperature for all samples, with those samples returning a Log WQRS above 9 highlighted, as well as samples with at least one coliform detected. This shows a majority of the samples with high Log WQRS occurred at low chlorine levels and the same can be said for coliform failures. This again shows that low disinfection levels may be a contributing factor to increased risk of bacteriological failures. Figure 6.64 (b) is a histogram of the Log WQRS for each of the 49 samples that were coliform failures. As expected the distribution is skewed towards higher Log WQRS, though some failures also occurred with lower Log WQRS. Coliform failures not detected in bulk water by flow cytometry potentially indicate an ingress related contamination event. Two of these failures with lower Log WQRS are explored in Figure 6.65 and Figure 6.66, alongside the Log WQRS, chlorine, water and daily max air temperature, and max daily rainfall. The first example shows the weekly samples from a single service reservoir for 2.5

years, with a sudden detection of 9 coliforms seen on the 25th July 2022. This coliform failure occurred despite reasonable disinfection levels, though it was during a period of hot weather with a peak air temperature of 30.7 °C seen in the days before and after some heavy rainfall, suggesting conditions may have been well-suited for contamination via ingress. That another coliform is detected at this site at the start of November suggests the issues were ongoing. The second example shows chlorine dropping from 0.47 to below 0.1 mg/l in the weeks leading up to a coliform count of 5 on the 14th September, which also coincided with a high rainfall event, suggesting ingress and insufficient disinfection as potential contributing factors.

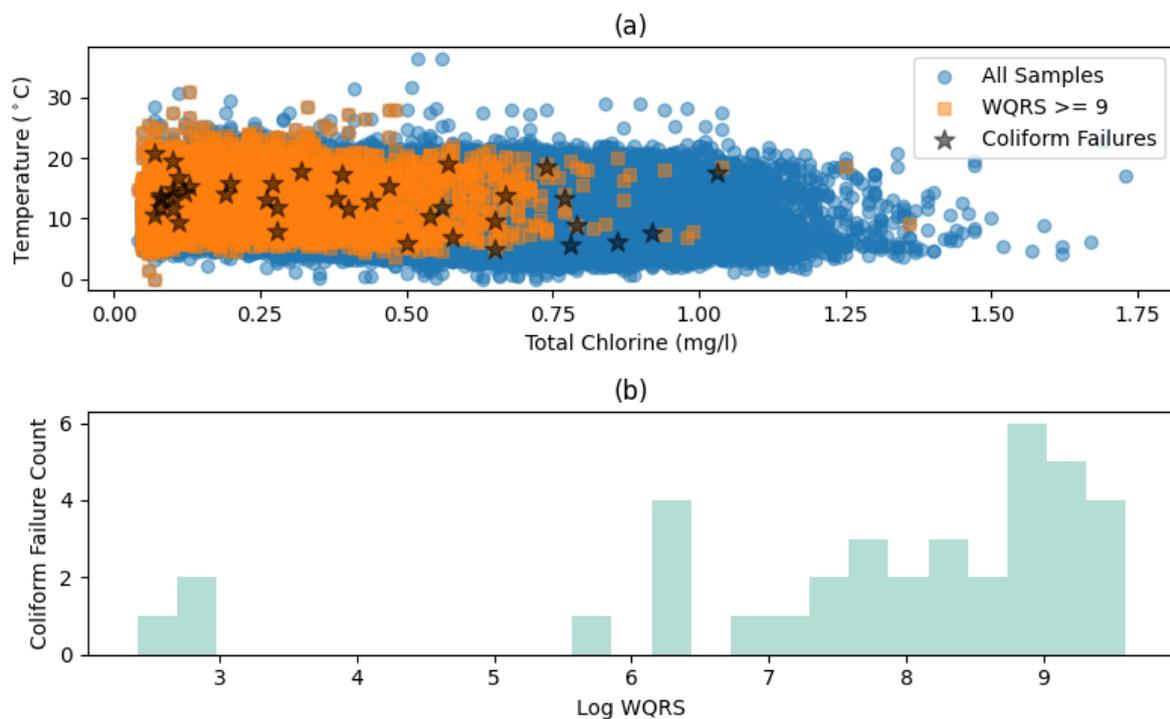


Figure 6.64. Scatter plot of total chlorine versus temperature, where blue circles represent all samples, orange squares are samples with a WQRS above 9 and black stars are samples with coliform failures (a), and a histogram showing the distribution of WQRS for samples with coliform failures (b).

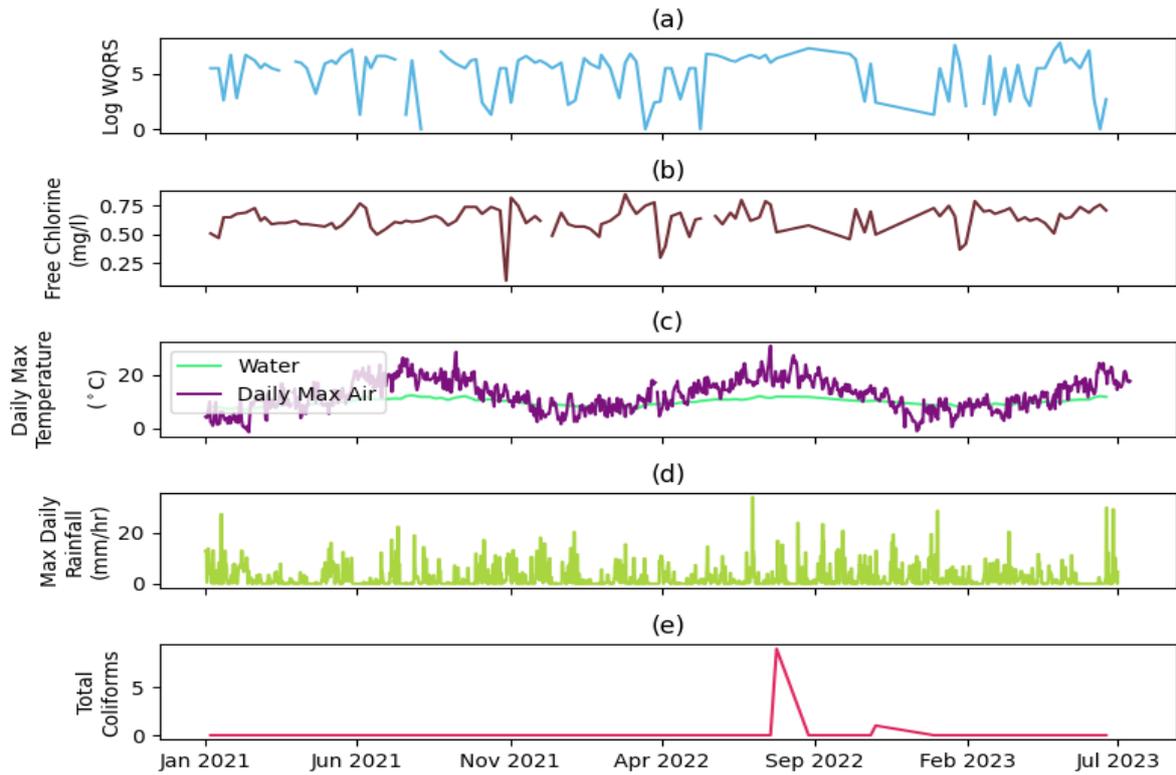


Figure 6.65. Log WQRS (a), free chlorine (b), water and air temperature (c), rainfall (d), and total coliforms (e) at a single service reservoir for over 2 years.

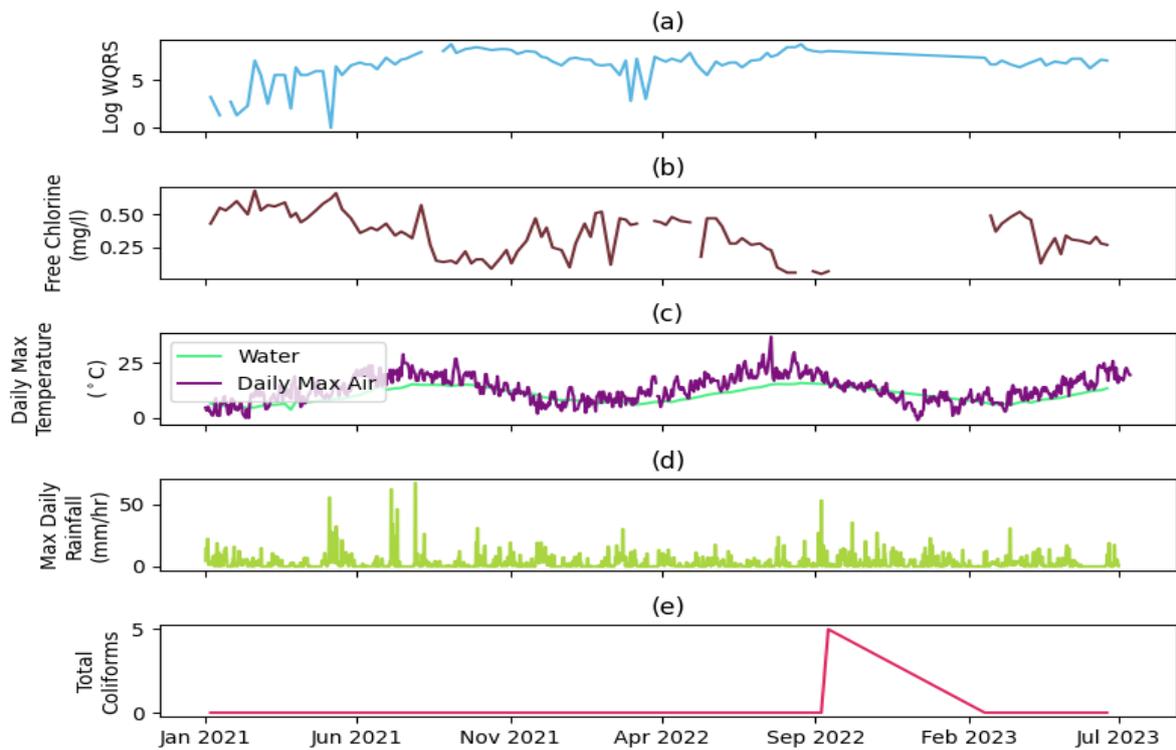


Figure 6.66. Log WQRS (a), free chlorine (b), water and air temperature (c), rainfall (d), and total coliforms (e) at a single service reservoir for over 2 years.

6.9 Summary

This chapter has demonstrated how information can be derived from MPMS water quality datasets taken from DWDS that can be used to inform operations. The methods developed in Chapters 4 and 5 have been shown to work together effectively alongside other approaches to unlock insights from the time series datasets. The first two examples moved from SPSS to MPMS, highlighting how additional sensors and parameters increase the insight gained when compared to relying on single time series data. Additionally, this increase was seen to behave in a non-linear fashion or multiplicative fashion. Table 6.2 provides a summary of the 7 examples covered in this chapter, with the analysis methods applied listed alongside the specific insights gained for each example.

Table 6.2. Summary of examples presented.

Example	Parameters	Analytic Techniques	Insight Gained
Example 1 – Building from SPSS to MPSS to MPMS (Section 6.2)	Turbidity chlorine flow rate pressure temperature pH	Data quality assessment Turbidity drift detection and correction Cross-correlation Material flux Daily peak flow rate Turbidity Event scale	The majority of material identified from upstream source highlighted by material flux assessment. This would not be possible to confirm with SPSS / MPSS alone at the reference location. Advisory events were seen at both locations in the days prior to the event suggesting this could act as a pre-cursor warning with a reduction post-event indicating likely benefits from maintenance.
Example 2 – Building from SPSS to SPMS to MPMS (Section 6.3)	Turbidity chlorine flow rate pressure temperature pH upstream flow rate discolouration contacts.	Data quality assessment Cross-correlation Daily peak flow rate Material flux Turbidity Event scale Contacts Analysis	Source of event was identified as increased flow through mains mobilising material seen at three sensors installed on offshoots. Flux estimates show most material traversing the Reference location, which also received the most discolouration contacts Event scale advisory scores also suggested the Reference location poses the greatest risk.
Example 3 - MPMS Discolouration and low chlorine at end of network (Section 6.4)	Turbidity, flow rate, chlorine	Data quality assessment Turbidity Event scale Material Flux Daily peak flow rate Contacts analysis	Shows how water quality can deteriorate during network transit and this may be linked to where chlorine residual is not maintained. Event scale analysis shows more alert and alarm events at D (end of network) in 2020, and shows advisory analysis can be used in upstream locations to infer risk further downstream.
Example 4 - MPMS Turbidity event in a single long main (Section 6.5)	Turbidity, mains flow rate	Median daily standard deviation, Turbidity Event Scale Material flux (using only available mains flow rate)	Analysis of long term turbidity shows increased diurnal fluctuations further down this network, indicating greater material movement and increased mobilisation risks. This increasing risk is supported by increased numbers of alert and alarm events, and advisory scores. The hydraulic induced turbidity event analysed showed significant material mobilisation towards the end of this network section, again highlighting the increased downstream risk.

Example	Parameters	Analytic Techniques	Insight Gained
<p>Example 5 – MPMS Dimensionality reduction and anomaly detection (Section 6.6)</p>	<p>Turbidity pH conductivity chlorine temperature ORP flow rate pressure</p>	<p>Dimensionality reduction (PCA), Unsupervised anomaly detection (IF, EE, LOF, OCSVM)</p>	<p>Demonstrates how automatic anomaly detection can work in higher dimensional water quality sensors, with OCSVM shown to provide a good fit to the 2 principal components obtained from PCA dimensionality reduction. Several anomalies are detected, including a turbidity event with an ORP drop in both locations and a step increase in turbidity along with a step drop in pH and a drop in pressure, indicating the potential value for an automated approach.</p>
<p>Example 6 - MPMS Increase in daily turbidity profile (Section 6.7)</p>	<p>Turbidity flow rate chlorine</p>	<p>Drift correction, Material flux, Longer term material quantification, Daily peak flow rate</p>	<p>This event involved a step increase in daily peak mains flow, leading to daily mobilisation of material, seen at multiple locations. The total material mobilised is similar to that previously seen in Example 2, though as spread out over a month has lower turbidity responses with reduced risk to customers. Location Y was found to have the highest daily and overall material flux, indicating this location remains at risk of discolouration.</p>
<p>Example 7 - MPMS Coliform failures in service reservoirs (Section 6.8)</p>	<p>Log WQRS (derived from flow cytometry) Chlorine temperature (air and water) rainfall total coliforms</p>	<p>Plotted as time series (averaged across all SR sites), Scatter plot and histogram to examine link between coliforms and flow cytometry</p>	<p>This example deviates from previous ones in that it does not include high-frequency water quality data, instead analyses weekly regulatory samples at service reservoirs. The seasonality of flow cytometry, chlorine and temperature are clearly seen and an inverse relationship between chlorine concentration and Log WQRS. Most coliform failures also had a high Log WQRS, with two exceptions examined and found to be potentially related to ingress with high air temperatures and rainfall preceding.</p>

Chapter 7: Discussion

This research investigates how continuously monitored water quality time series data from DWDS can be transformed into actionable information to support network operations. The focus was on developing methods for assessing sensor data quality, investigating how water quality events can be detected and understood, and exploring how multiple water quality parameters and sensor locations in combination can impact the overall level of insight derived. The work started with a review of the relevant literature to identify current techniques and knowledge gaps. To facilitate assessment, continuous time series and sampling datasets were then obtained from five supporting WSPs. A particular focus on discolouration events was driven by discolouration being the most pervasive water quality issue in DWDS, along with turbidity being the most commonly measured parameter in the datasets provided. However, the methods developed to analyse turbidity time series data also have applicability to other parameters. This research included development of novel contamination event analytics driven by domain expert interpretation of water quality time series, before applying developed algorithms to real events seen in multiple connected sensor locations. This represents a major advance from the common but unrealistic practice of artificially inserting events on top of measured data (Murray and Haxton 2010; Perelman et al. 2012; Li et al. 2019; Muharemi et al. 2019). This discussion reflects on the contributions made in respect to the identified knowledge gaps and provides recommendations regarding directions for future research to build on the contributions made in this thesis.

7.1 Multiplicative Value of Multiple Parameters and Multiple Sensors

The step increase in value with having multiple connected sensors was demonstrated first in Chapter 4, where connectivity derived from cross-correlation is used to improve the data quality assessment of each individual sensor and parameter. The increased confidence in the data quality impacts the confidence with which any subsequent analytics can be done. This is clearly seen later in Chapter 6 (best reviewed via the summary Table 6.2), where suspected water quality events in the first two examples (Sections 6.2 and 6.3) were confirmed to be real events, as opposed to local events or sensing errors, through comparison with connected sensors. This logic is applicable to parameters other than turbidity, where a sudden unexpected change or event, could indicate either a sensing error or a network event. The same point can be made for having multiple parameters measured at a single location, and

there are some particular combinations of parameters that can be combined to increase insight, such as flow rate and turbidity to determine material flux. However, it was only with the availability of both multiple parameters and multiple sensors that a multiplicative jump, not only in confidence but in obtainable insight regarding the root-cause of a contamination event, was seen as observed in the first two examples in Chapter 6 (Sections 6.2 and 6.3). The term multiplicative in this context refers to the observation that the increased value is greater than simply adding together the individual value from each single sensor and parameter. However, such a multiplicative increase is only seen when moving from single parameter single sensor to multiple parameter multiple sensor analytics and such a trend would not be expected to continue with the addition of more and more sensors and parameters. For example, a multiplicative increase in value may not necessarily be observed when moving from 10 connected sensors to 11.

Example 1 (Section 6.2) is a clear demonstration in the transformative power of having a confirmed connected location and both locations having flow rate data. The MPSS analysis strongly suggested this event was caused by increased flow rates. However, the addition of a second sensor location, which was determined to be connected through cross-correlation performed on the entire dataset, was seen to transform understanding of this event. Through analysis of the material flux at each location, the larger second wave of material was seen to decrease by around 10% to the downstream location. This precise analysis was only possible due to the calculation of an average transit time between these locations, proving the value of cross-correlation. The result suggests the material involved is travelling through the network from an upstream source, possibly from an upstream service reservoir or supplying WTWs, while the additional material local to each location is relatively minor. This is counter to the consideration of a hydraulic induced event localised to the downstream location as may have been concluded without the additional upstream location. Though it would be very challenging to automate this kind of precise analysis, it would not be difficult to automatically check known-connected sensor locations for a corresponding event, following a detection. Example 2 (Section 6.3) used material flux and analysis of peak daily flow rates to confirm that this was a network-wide event caused by mains flow rate increases, while the fact that the sensor location with the highest total discolouration material also had the most associated contacts validates this as a measure of downstream discolouration risk. Analysis of the shape of a turbidity event can also provide information regarding the dominant accumulation process. For example, a gradual increase in line with increasing flow might

suggest cohesive layers are the dominant mechanism, whereas a sharper event that does not continue to rise with increasing flow rate may indicate sedimentation-driven accumulation (Boxall et al. 2023).

Turbidity and flow at multiple connected locations was also used to confirm alarm events in Example 3 (Section 6.4) were related to upstream hydraulics, while Example 4 (Section 6.5) demonstrated how only having turbidity at multiple locations can limit accurate tracking of an event. Though in this single straight mains example connectivity could be assumed, the lack of transit time meant visual assessment was relied upon. Example 3 also demonstrated that discolouration may be more likely to occur, with or without a flow rate increase, with higher temperatures often seen in the summer time. It has been previously shown that higher air temperatures correlate with increased discolouration contacts (van Summeren et al. 2015). Increased water temperatures increase chlorine decay rates, which can be an issue in further downstream locations such as location D in this example. Therefore, in this example the root-cause is shown to be a flow rate increase, but both weather and a poorly maintained residual disinfection are likely contributing factors. The link between hot weather and increased demand is clearly shown in Figure 6.40, and suggests weather forecasts and associated parameters could play a bigger role in discolouration mitigation.

These examples demonstrate that the increase in value and insight obtainable does not increase linearly from SPSS to MPMS analysis, but instead a multiplicative increase is seen. An extreme case looking at eight-parameter sensors was included in Example 5 (Section 6.6), where dimensionality reduction and unsupervised anomaly detection was shown to be an effective way to identify unusual events in any parameter, showing the value in having high dimension water quality time series. When it comes to analysing a contamination event in a DWDS, knowing what sensors are hydraulically connected and the approximate transit times involved is transformative in its power. This does place significant importance on gaining an understanding of the hydraulic connectivity between sensor locations. The data quality assessment framework proved the suitability of chlorine time series for cross-correlation analysis, which can provide this kind of vital spatiotemporal information. Turbidity time series were not found to be well-suited, though other parameters could be investigated for their suitability. The importance of having cross-comparable sensor locations has implications for deployment strategies that will be further discussed in this section.

7.2 From Reactive to Proactive Management of Network Contamination

Events

7.2.1 Reactive Management

If it is accepted that contamination events will occur and that knowledge of these may not arise until it's too late, developing methods to effectively detect and analyse these events becomes vital. The ability to take effective action following a detected discolouration event is dependent on how quickly the event is detected and the time before the discoloured water reaches customers. An alarm event (using the event scale developed in Chapter 5), seen at two or more connected locations and therefore of high confidence, requires immediate action. However, the lead time required for utilities to actually perform mitigating actions that can halt discoloured water already on route to customers is likely greater than would be available. Therefore, in these cases perhaps the best action is simply to warn potentially impacted customers. The earlier this is done, the better the impression on customers. This is where edge computing of the turbidity event scale could function to rapidly enable such alerts and warnings to be sent out. Relying on waiting for the data to be uploaded to a central server before analysis currently means many such events will reach customers before the data is even available. As well as a rapid warning, such events should be analysed for root-cause using the techniques demonstrated in this thesis. The above approach would equally be applicable for detected alert events.

The two examples in Chapter 6 demonstrated how the data quality framework developed in Chapter 4 can function alongside the turbidity event scale developed in Chapter 5 to semi-automatically detect alert and alarm events in connected sensors, with connectivity determined using the cross-correlation method outlined in Stage of 2 of the data quality assessment framework in Chapter 4. Any period of data that is flagged for exceeding 1.5 NTU for an extended period of time (6 hours was selected, as outlined in Table 4.3), subsequently requires cross-validation against other connected turbidity sensors (as per Stage 2 and 3 of the data quality assessment framework). These examples both featured alarm events that were found to occur in connected sensor locations, therefore being confirmed as real network events. That Example 2 had over 130 reported customer contacts provides direct confirmation that this was a real discolouration event impacting customers. This research has shown the capabilities for online water quality monitoring to detect and confirm

DWDS contamination events, which represents a clear improvement on reliance on subjective customer contact information. However, estimating only total material mobilised during an event still does not provide a full picture of overall risk posed, which should also account for the number of downstream customers likely to be impacted. For example, Example 2 resulted in over 130 contacts with a total of 940 NTU.m³ discolouration material passing the reference location and a peak turbidity of 11.36 NTU, while only 15 contacts were seen as a result of the April 2021 alarm event featured in Example 3 despite a similar turbidity peak and a total of 5714 NTU.m³ discolouration material passing location B. The difference in impact on customers, despite the latter example having more discolouration material mobilised, is likely down to number of customers impacted, highlighting the importance to account for number of impacted customers when determining the appropriate response to a detected discolouration event.

7.2.2 Proactive Management

Though improved discolouration detection, root-cause determination and estimation of downstream risk represents a major improvement over the status quo, these remain reactive in nature. Though they inform about network risk regions, it is desirable to move towards more proactive discolouration management approaches that don't require a historical record of discolouration events and can be applied without context to any time series. Therefore, the potentially transformative contribution of this work to managing discolouration is related to precursor information, with the advisory score approach showing promise for flagging low level increases in turbidity that would previously have been ignored. The event detection approach investigated in Chapter 5 involved calculating a residual by subtracting forecast values from actual values, instead of previous work that sought to predict turbidity (Meyers et al. 2017; Kazemi et al. 2018). The approaches compared in this chapter focused on classical time series forecasting, which has been shown to often outperform machine learning approaches, are well-suited to univariate problems and are less computationally demanding (Makridakis et al. 2018). This makes such approaches more applicable in real-time and even well-suited for edge computing, whereby high advisory events could be detected remotely by a small computer alongside the remotely installed sensor. Such a setup could provide real-time warnings, as opposed to waiting for the data to be uploaded to a server before any analysis and/or event detection can be done. The time-based averaging advisory score approach developed was applied to the data preceding the confirmed real networks events in

Chapter 6, to investigate the potential of this method for precursor indications. As this algorithm was developed to output an event score time series, matching that of the averaged labels from the labelling exercise, some additional analysis was required to simplify the resulting time series into a useful single metric.

Example 4 showed that the number of alert and alarm events broadly was seen to increase through this single straight network section, as well as the median peak daily advisory score. Using the highest daily advisory score is an obvious way to simplify the advisory score time series, and reporting the median value of that provides information on the average level for a given day. Example 1 simply included a plot (Figure 6.13) showing the advisory score time series, which included some high advisory scores in the days leading up to the featured alarm event. Example 2 instead plotted the daily peak advisory scores for the three featured locations (Figure 6.24), which had the effect making the plot easier to visualise. This plot clearly shows increased advisory scores at all three locations before the alarm event on 12th April. The visible drop seen in the weeks following the event demonstrates that some of that risk had reduced due to the alarm event likely mobilising much of the accumulated material. A similar approach is used for the Example 3 (Figure 6.39) with the vast majority of alarm events seen at end-of-network location D, which also had the highest peak daily advisory score even a week before the first alarm event. These examples demonstrate that the event scale is an effective approach for estimating discolouration risk levels using single turbidity time series. However, as with data quality assessment and event detection, estimation of discolouration risk is also more powerful and informative when it can be compared to other connected locations.

This proactive approach assumed that sudden small increases in material mobilisation are indicative of larger mobilisation risks. This requires further work to fully validate. However, it does align well with the theory that discolouration material accumulates in cohesive layers (Husband et al. 2008), with sudden small mobilisations exposing deeper layers of material with the potential to be mobilised. Unlike previous research into detecting turbidity events that also required flow as an input (Meyers et al. 2017; Kazemi et al. 2018), the turbidity event scale was designed to be applicable to individual turbidity time series without requiring flow. It was applied in four of the examples from Chapter 6 (in all four cases at least one alarm event >4 NTU was detected). Example 6 (Section 6.7) was unique among the discolouration examples featured, in that the turbidity never exceeded 2 NTU, meaning the

five sensors all experienced only advisory events following a step increase in mains hydraulics that lasted several weeks. It is not known how this system would have reacted with an additional step increase in flow beyond that sustained for this period. Clearly there are significant levels of material present in this system and it is possible that a higher flow during this period would have mobilised further material. This would again be supported by the cohesive layers theory, whereby weaker layers of material being mobilised exposes deeper layers to be mobilised by a further increase in flow rate. This suggests that these high advisory scores are not necessarily invalidated by the lack of a subsequent alert or alarm, as the required flow increase did not occur. How to deal with advisory events is a more complicated matter than alert or alarm events. It has been shown that high advisory scores often precede alert and alarm events. However, Example 6 had high advisory scores without any alert or alarm event coming within the time frame investigated. Therefore, actions to take where high advisories are detected could range from carefully avoiding and managing any further flow rate increases, as much as is possible, to planned network interventions such as flow conditioning to safely reduce the amount of accumulated discolouration material.

7.3 Practical Considerations

Optimal use of deployed water quality sensors depends on the overall monitoring goal. This could range from investigating discolouration in a network section with perceived issues to more targeted deployment looking at a particular asset such as a service reservoir. The goal of water quality monitoring influences choice of parameter, number of sensors, installation locations, deployment time period and sampling frequency. The ultimate aim is to collect the dataset best suited to the desired insight, meaning the entire life cycle of raw data to actionable information should be considered before a deployment strategy is devised, including how the sensor will be prepared, installed, and maintained and how the data will be monitored, analysed, and what actions will be taken in response. With this in mind, there are huge benefits from accurate reporting of network events, such as maintenance and deterioration events. Collecting such data in a format easily inputted into analytics would save time and enhance analysis of deployed sensors and can add vital contextual information to supplement the time series datasets. This section briefly reviews considerations specific to taking good quality data, before discussing the merits of different sensor deployment strategies.

7.3.1 Data Quality

Data quality was defined as “how well suited data is for its intended purpose” in Section 2.4.2 and was identified as a significant challenge with water quality sensors continuously deployed in DWDS. Remotely deployed water quality sensors can foul or deteriorate, resulting in spurious measurements. Though this is inevitable, good practice both before and during deployment can minimise these occurrences and increase the value of the measured data. Pre-deployment calibration and validation enables potential issues to be learned about and remedied, while the data quality assessment framework developed in Chapter 4 provides a path to monitoring the performance of deployed sensors. This framework addresses the widely understood problem often referred to as DRIP (data rich information poor), and decreases the requirements of manual time-consuming data quality assessment that is often necessary before any analytics are possible (as outlined in Background Section 2.4.2). When a sensor issue is detected and confirmed, through its omission of data detected in other connected sensors, the sensor in question requires attention. Therefore, in order to investigate the sensor as quickly as possible, operators need to have maintenance technicians ready at relatively short notice. In many cases, the sensor may need to be either partially or completely replaced, meaning spare sensors and parts such as membrane/electrolytes for chlorine sensors should be kept at hand. In reality, the thoroughness of any maintenance plan depends on what level of accuracy is desired. If the sensors are deployed purely for large abnormal event detection, accuracy of lower-level baseline values will be of less importance. If there is interest in lower-level accuracy, then calibration and maintenance processes are more important. In this situation, the accuracy of any calibration reference sensor is essential, as well as having sensor performance checks in place while deployed. Even if ignoring sensor faults, any remotely deployed sensor relies on a battery, which will require replacement at some point. The sensor maintenance requirements are also influenced by overall deployment strategy, particularly the planned deployment duration, as longer or permanent deployments come with an inevitability about the sensor failing at some point if maintenance is not undertaken.

7.3.2 Deployment Strategies

There are a variety of different circumstances under which it may be desirable to deploy continuous water quality monitoring across DWDS. The examples included in this thesis

have entirely resulted from continuous long-term water quality monitoring projects, where each sensor largely remained at single locations for extended periods generally of several months to over a year. However, water quality monitoring may be desirable in response to a particular network change or intervention, such as during flushing activities, or at a specific location of deemed importance, such as a service reservoir. It may also be desirable to deploy water quality monitoring in a ‘lift and shift’ manner, where the same number of sensors are deployed for shorter periods of time, before being moved to another location. Such an approach would enable more network to be covered, while also ensuring the sensors themselves can be inspected regularly, such as prior to redeployment. Besides sensor deployment length, sampling frequency and spatial density and resolution are important parts of a deployment strategy.

Example 7 (Section 6.8) featuring regulatory weekly samples at service reservoirs, revealed that low chlorine levels correlate to samples with higher risk of containing potential pathogens. Figure 6.66 features a service reservoir that was found to have a coliform failure, despite a low Log WQRS suggesting nothing was apparent in flow cytometry measurements. This failure came alongside a rainfall event, following a drop in chlorine during a heatwave. It is easy to see how high-frequency monitoring of disinfection levels could have helped this issue be detected sooner, and with more clarity if a sudden drop is seen. When such targeted high-frequency monitoring of service reservoirs are desired, it is vital to monitor not only at the reservoir outlet, but also the inlet, as has been demonstrated by Dronina et al. (2020). When it comes to sampling frequency, there has traditionally been a straightforward trade-off between battery life and how frequently a sensor samples (although the frequency of data upload is considered a bigger contributing factor to battery life). The industry standard, inherited from hydraulic measurements, is to measure water quality at 15 minute intervals. However, the clear benefits of measuring at a higher frequency, such as 1 minute intervals, have been reported (Gaffney and Boulton 2012) and are particularly relevant to measuring shorter discoloration events. Such higher-frequency sampling could easily be achieved with shorter term deployment strategies, where the batteries could be replaced without any missing data.

The question of how many sensors are required and at what spatial density is more complex than the question of temporal frequency. This research has shown the transformative effect of having multiple comparable sensor locations, so focusing deployment in a specific network

section, deemed to be of risk of discolouration, is recommended. These are complex networks and a single deployed sensor only provides a small window into what is happening. Multiple sensors installed at comparable locations can result in a much more complete picture. However, there is no clear rule for how far apart sensors can be installed while remaining comparable. The low chlorine discolouration case in Example 3 showed that sensors installed 23 hours apart can still be comparable. The main criteria for this to work is that there is nothing interrupting flow between the locations, such as a service reservoir, and that the residual chlorine levels are sufficient and not flat. This will not be possible to know with any certainty before deployment. Therefore, an iterative approach is recommended, where operators test out different deployment configurations. Such an approach may be desirable anyway, in order to move these instruments to different network regions. With time, knowledge could be built up of what locations work well together. Of course there are practical considerations with deciding where to install these sensors, and operators will be limited by the difficulty of accessing much of the buried networks.

7.4 Future Research

7.4.1 From Source to Tap

The completion of this project opens up possibilities for several future research directions that could expand and build upon the findings and insights presented in this thesis. The first step would be to validate the novel methods developed in this project. To achieve this, a collaborative monitoring scheme alongside a WSP would be required as their access to details on network events would be invaluable in linking analytical outcomes to real-world incidents. Such a project would facilitate a continued examination of different sensor deployment strategies with chlorine or turbidity likely to be of key WSP interest. After such first steps are completed and the developed approaches are validated, a next step would be to develop a more complete source-to-tap understanding. This could be achieved by including data from catchments, the surrounding environment, WTW, DWDS (including service reservoirs) and customer taps. Considering turbidity, this would lead to a more comprehensive and detailed estimation of discolouration material source and destination. The event scale developed in this research could be expanded into a discolouration event categorisation and risk estimation system. Further research into proactive discolouration prevention could also explore the use of demand forecasting to supplement water quality analytics, as increased demand often leads to the type of material mobilisation that has been

heavily featured in this work. Such a source-to-tap digital monitoring scheme would also provide the kind of resilience required of drinking water systems in order to meet changing demands and aging infrastructure. The effects of climate change will be felt on both the supply and demand sides of drinking water systems, with increasing urban populations due to climate migrants and increased water shortages in certain areas. These effects are already being felt, with the practice of mixing water sources becoming more commonplace in the UK due to shortages in some areas, despite the effects on water quality not being well understood. That the DWDS assets themselves are ageing makes them more vulnerable to the increased strain they are set to come under. Enhanced system-wide monitoring offers the potential to help safeguard these vital assets against the uncertainties ahead.

7.4.2 Health Risk of Contamination Events

Integrating data all the way to customer taps would allow for water quality deterioration events and their health impacts to be better understood, as well as enable more detailed examination of the impact of different disinfection strategies. This could be achieved by supplementing the continuously monitored data with microbial grab samples looking for specific contaminants. The problem of intermittent grab samples being unlikely to take measurements during what are unpredictable water quality discolouration events could be addressed by measuring alternative parameters, such as flow cytometry alongside turbidity, flow and chlorine at multiple locations. Under such a setup, discrete flow cytometry measurements could aid detection of a discolouration event. Such a command could be issued using edge computing to process the event scale approach developed in this research. This process would enable the generation of a labelled dataset linking real discolouration events to their health risks, a dataset that could be exploited through machine learning algorithms in order to determine the key contributors to damaging discolouration events, and also could investigate what conditions lead to pathogens entering the water system. The results would allow for real-time alerts to be issued based on an estimation of health-risk associated with a given discolouration event. Additionally, the impact of different types of discolouration events could be investigated, for example if there is a difference in the health impact of an ingress-related discolouration event compared to a pipe-wall mobilisation one. Monitoring at customer properties would allow for domestic discolouration to be examined, including what kind of region and type of property is most at risk, and could tie back into research estimating

health risks of a DWDS discolouration event by accounting for distance traversed and residual chlorine.

7.4.3 Role of Artificial Intelligence

The future role of artificial intelligence, including in DWDS, is difficult to predict. Recent advancements in generative large language models (LLMs) are rapidly enhancing their ability to perform various tasks, and leading many industries to investigate increased adoption of artificial intelligence. Notably, research has shown how quickly LLMs are rapidly improving at general purpose tasks (Eloundou et al. 2023). This progress paves the way for accelerated digitalisation in many industries, including water, driven by increased LLM-powered software and automation. A future of integrated source-tap data sources promises to create a holistic digital picture capable of describing the state of a particular drinking water system. This research has shown a glimpse of how increased parameters can open up more automated analytical approaches (such as dimensionality reduction and unsupervised anomaly detection), while also demonstrating how high-frequency data can supplement and improve upon regulatory sampling. With more data integration these possibilities will increase. Different types of data could also be integrated, such as image data from satellites and site inspections, to video and audio data from customers, with multimodal artificial intelligence showing great promise for combining different data types together (Ngiam et al. 2011). A more concrete system-wide estimation of discolouration risk could be generated, building on top of the turbidity event scale developed in this research. Graph neural networks, a form of deep-learning that can capture complex dependencies between nodes and has outperformed other methods at predicting urban traffic flows (Peng et al. 2020), represents a promising research avenue that would incorporate the spatiotemporal relationships between different sensing locations and could perform both node and system wide analyses. The increasing role of artificial intelligence will be relevant for the management of these assets, and promises to deliver the digitalisation of water systems, but its success will be dependent on the availability and quality of relevant datasets.

Chapter 8: Conclusions

Through investigation of large continuous DWDS water quality datasets from multiple UK water utilities, this thesis has developed and applied innovative analytical routines to understand sensor capabilities and network performance. By demonstrating how datasets from remote sensor networks can be combined and analysed to understand and crucially track water quality changes within these complex networks, this work establishes the value of sensors in providing actionable information to support operational management and that the value increases multiplicatively as analysis moves from a single parameter single sensor to multiple parameters and multiple sensors.

The key research contributions are detailed in Chapters 4-6. The development of a data quality assessment framework, described in Chapter 4, provides a vital first step into realising the potential of continuous in-network water quality monitoring. Chapter 5 focused on arguably the most ubiquitous drinking water quality issue – discolouration, and developed algorithms to mimic a novel crowd-sourced set of labelled turbidity time series examples. Particularly unique was the focus on analysing often-ignored lower-level turbidity data to estimate risk of discolouration. Chapter 6 applied the methods developed in the previous chapters, alongside other techniques, on confirmed real-world examples. The insights gained from each example are best reviewed using Table 6.2. The use of confirmed real-world examples makes this work unique as previous research has been reliant on inserting artificial events into background water quality data. Overall, the research demonstrates how depth of insight increases multiplicatively with more parameters and sensors available.

The main contributions of this work are:

1. The data quality assessment framework developed provides novel tools necessary to assess sensor performance and flag erroneous data before further analysis.
2. Cross-correlation is demonstrated to be an effective method for determining network connectivity and estimate transit times between sensor installation locations with chlorine time series found to be well-suited.
3. Domain expert interpretation of events within turbidity time series was captured and understood through the use of an innovative crowd-sourced labelling. The results led to important insights about the importance of analysing often-ignored low-level

turbidity data, and were used to analytically compare different event detection algorithms.

4. A turbidity event scale that detects advisory (<2 NTU), alert (2-4 NTU) and alarm (>4 NTU) events enables improved proactive and reactive management of discolouration within DWDS.
5. A time-based averaging approach is shown to be effective for analysing low-level turbidity data, as demonstrated by approximating the domain expert interpretation, and was shown to outperform more complex approaches using ARIMA and exponential smoothing.
6. Turbidity, chlorine, and flow were found to be highly informative parameters for understanding water quality in DWDS, with temperature, pH, conductivity, pressure and ORP also found to be useful.
7. With turbidity, chlorine and flow available at multiple connected locations, discolouration events can be accurately tracked throughout network sections, enabling information to be derived about both the source and destination of the discolouration material.

The major overarching outcome that the level of insight obtainable increases multiplicatively with multiple parameters and sensors will help inform intelligent deployment and analysis of water quality sensor networks in order to improve understanding and management of DWDS assets. The data quality assessment framework will improve the quality of data taken, thereby increasing the value of the sensors deployed. Determining connectivity between sensor locations is significant in helping to move away from single sensor analytics to higher confidence analysis of entire network sections. Significant advances have been made with regards to utilising DWDS turbidity time series to manage discolouration. By moving away from only focusing on reacting to customer contacts or large turbidity events to a more proactive digitalised approach, discolouration risk in network sections can be estimated and larger events can be prevented. Additionally, the requirement of network analysts to manually investigate water quality datasets will be reduced, enabling them to work in other areas while further increasing the value of deployed sensors. Ultimately, the contributions made in this research facilitate a move towards an improved digitalised approach to managing these vital assets and inform the kind of monitoring that will be required to address future challenges to safeguard delivery of high quality drinking water.

References

- Aggarwal. 2016. *Outlier Analysis*. 2nd ed. New York, NY: Springer New York. doi: 10.1007/978-1-4614-6396-2.
- Ainsworth. 2013. "Safe Piped Water: Managing Microbial Water Quality in Piped Distribution Systems." *Water Intelligence Online* 12. doi: 10.2166/9781780405841.
- Allison. 2000. "Multiple Imputation for Missing Data: A Cautionary Tale." *Sociological Methods and Research* 28(3):301–9. doi: 10.1177/0049124100028003003.
- Angelakis, and Mays. 2014. *Evolution of Water Supply Through the Millennia*. IWA Publishing.
- ATi. 2022. *Metrinet Controller: O & M Manual*. Analytical Technology.
- Banna, Imran, Francisque, Najjaran, Sadiq, Rodriguez, and Hoorfar. 2014. "Online Drinking Water Quality Monitoring: Review on Available and Emerging Technologies." *Critical Reviews in Environmental Science and Technology* 44(12):1370–1421. doi: 10.1080/10643389.2013.781936.
- Bates, and Granger. 1969. "The Combination of Forecasts." *Journal of the Operational Research Society* 20(4):451–68. doi: 10.1057/jors.1969.103.
- Benesty, Chen, and Huang. 2004. "Time-Delay Estimation via Linear Interpolation and Cross Correlation." *IEEE Transactions on Speech and Audio Processing* 12(5):509–19. doi: 10.1109/TSA.2004.833008.
- Berney, Vital, Hülshoff, Weilenmann, Egli, and Hammes. 2008. "Rapid, Cultivation-Independent Assessment of Microbial Viability in Drinking Water." *Water Research* 42(14):4010–18. doi: 10.1016/j.watres.2008.07.017.
- Blázquez-García, Conde, Mori, and Lozano. 2020. "A Review on Outlier/Anomaly Detection in Time Series Data." doi: <https://doi.org/10.48550/arXiv.2002.04236>.
- Blokker, and Pieterse-Quirijns. 2013. "Modeling Temperature in the Drinking Water Distribution System." *Journal - American Water Works Association* 105(1):E19–28. doi: 10.5942/jawwa.2013.105.0011.
- Boe-Hansen, Albrechtsen, Arvin, and Jørgensen. 2002. "Bulk Water Phase and Biofilm Growth in Drinking Water at Low Nutrient Conditions." *Water Research* 36(18):4477–86. doi:

10.1016/S0043-1354(02)00191-4.

- Boker, Rotondo, Xu, and King. 2002. "Windowed Cross-Correlation and Peak Picking for the Analysis of Variability in the Association between Behavioral Time Series." *Psychological Methods* 7(3):338–55. doi: 10.1037/1082-989X.7.3.338.
- Bonnefous. 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, Bögl, Gschwandtner, and Miksch. 2017. "Visual Support for Rastering of Unequally Spaced Time Series." *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction* 53–57. doi: 10.1145/3105971.3105984.
- Bowden, Nixon, Dandy, Maier, and Holmes. 2006. "Forecasting Chlorine Residuals in a Water Distribution System Using a General Regression Neural Network." *Mathematical and Computer Modelling* 44(5–6):469–84. doi: 10.1016/j.mcm.2006.01.006.
- Box, and Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. 1st ed. San Francisco.
- Box, Jenkins, Reinsel, and Ljung. 2015. *Time Series Analysis: Forecasting and Control*. 5th ed. John Wiley & Sons, Inc.
- Boxall, Blokker, Schaap, Speight, and Husband. 2023. "Managing Discolouration in Drinking Water Distribution Systems by Integrating Understanding of Material Behaviour." *Water Research* 243:120416. doi: 10.1016/j.watres.2023.120416.
- Boxall, Machell, Dewis, Gedman, and Saul. 2011. "Operation, Maintenance and Performance." Pp. 193–226 in *Water Distribution Systems*. Thomas Telford Ltd.
- Boxall, and Saul. 2005. "Modeling Discoloration in Potable Water Distribution Systems." *Journal of Environmental Engineering* 131(5):716–25. doi: 10.1061/(ASCE)0733-9372(2005)131:5(716).
- Boxall, Saul, and Skipworth. 2004. "Modeling for Hydraulic Capacity." *Journal - American Water Works Association* 96(4):161–69. doi: 10.1002/j.1551-8833.2004.tb10607.x.
- Breunig, Kriegel, Ng, and Sander. 2000. "LOF: Identifying Density-Based Local Outliers." Pp. 93–104 in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. New York, NY, USA: ACM.
- Brown. 1959. "Statistical Forecasting for Inventory Control." *Journal of the Royal Statistical Society*.

Series A (General) 123(3):348. doi: 10.2307/2342487.

van Buuren. 2018. *Flexible Imputation of Missing Data*. 2nd ed. Utrecht: CRC Press LLC.

Camper, Burr, Ellis, Butterfield, and Abernathy. 1998. "Development and Structure of Drinking Water Biofilms and Techniques for Their Study." *Journal of Applied Microbiology* 85(S1):1S-12S. doi: 10.1111/j.1365-2672.1998.tb05277.x.

Chai, and Draxler. 2014. "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments against Avoiding RMSE in the Literature." *Geoscientific Model Development* 7(3):1247–50. doi: 10.5194/gmd-7-1247-2014.

Chandola, Banerjee, and Kumar. 2009. "Anomaly Detection." *ACM Computing Surveys* 41(3):1–58. doi: 10.1145/1541880.1541882.

Chandola, Cheboli, and Kumar. 2009. *Detecting Anomalies in a Time Series Database*.

Chatfield, Koehler, Ord, and Snyder. 2001. "A New Look at Models for Exponential Smoothing." *Journal of the Royal Statistical Society Series D: The Statistician* 50(2):147–59. doi: 10.1111/1467-9884.00267.

Clemen. 1989. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting* 5(4):559–83. doi: 10.1016/0169-2070(89)90012-5.

Colford, Roy, Beach, Hightower, Shaw, and Wade. 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.

Cook, Husband, and Boxall. 2016. "Operational Management of Trunk Main Discolouration Risk." *Urban Water Journal* 13(4):382–95. doi: 10.1080/1573062X.2014.993994.

Cooper, and Cowan. 2008. "Comparing Time Series Using Wavelet-Based Semblance Analysis." *Computers & Geosciences* 34(2):95–102. doi: 10.1016/j.cageo.2007.03.009.

Cutler, and Miller. 2004. *The Role of Public Health Improvements in Health Advances: The 20th Century United States*.

Dewan, Ay, Karim, and Beyenal. 2014. "Alternative Power Sources for Remote Sensors: A Review." *Journal of Power Sources* 245:129–43. doi: 10.1016/j.jpowsour.2013.06.081.

Doronina, Husband, Boxall, and Speight. 2020. "The Operational Value of Inlet Monitoring at

- Service Reservoirs.” *Urban Water Journal* 17(8):735–44. doi: 10.1080/1573062X.2020.1787471.
- Douterelo, Sharpe, Husband, Fish, and Boxall. 2019. “Understanding Microbial Ecology to Improve Management of Drinking Water Distribution Systems.” *Wiley Interdisciplinary Reviews: Water* 6(1):e01325. doi: 10.1002/wat2.1325.
- Du, Zhou, Guo, Guo, and Wang. 2021. “Deep Learning with Long Short-Term Memory Neural Networks Combining Wavelet Transform and Principal Component Analysis for Daily Urban Water Demand Forecasting.” *Expert Systems with Applications* 171:114571. doi: 10.1016/j.eswa.2021.114571.
- Dunia, Joe Qin, Edgar, and McAvoy. 1996. “Use of Principal Component Analysis for Sensor Fault Identification.” *Computers & Chemical Engineering* 20:S713–18. doi: 10.1016/0098-1354(96)00128-7.
- Dürrenmatt, Del Giudice, and Rieckermann. 2013. “Dynamic Time Warping Improves Sewer Flow Monitoring.” *Water Research* 47(11):3803–16. doi: 10.1016/j.watres.2013.03.051.
- DWI. 2018. *The Water Supply (Water Quality)(Amendment) Regulations 2018 SI No. 706*.
- DWI. 2022. *Drinking Water 2021: The Chief Inspector’s Report for Drinking Water in England*.
- DWI. 2023. *Drinking Water 2022: The Chief Inspector’s Report for Drinking Water in England*.
- El-Chakhtoura, Saikaly, van Loosdrecht, and Vrouwenvelder. 2018. “Impact of Distribution and Network Flushing on the Drinking Water Microbiome.” *Frontiers in Microbiology* 9. doi: 10.3389/fmicb.2018.02205.
- El-Zahab, and Zayed. 2019. “Leak Detection in Water Distribution Networks: An Introductory Overview.” *Smart Water* 4(1):5. doi: 10.1186/s40713-019-0017-x.
- Eliades, Lambrou, Panayiotou, and Polycarpou. 2014. “Contamination Event Detection in Water Distribution Systems Using a Model-Based Approach.” *Procedia Engineering* 89:1089–96. doi: 10.1016/j.proeng.2014.11.229.
- Eloundou, Manning, Mishkin, and Rock. 2023. “GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.” doi: <https://doi.org/10.48550/arXiv.2303.10130>.
- EU. 1998. *98/83/EC of Council of 3rd November 1998 on the Quality of Water Intended for Human Consumption*.

- Farman, Gardiner, and Shanklin. 1985. "Large Losses of Total Ozone in Antarctica Reveal Seasonal ClO_x/NO_x Interaction." *Nature* 315(6016):207–10. doi: 10.1038/315207a0.
- Fish, Osborn, and Boxall. 2016. "Characterising and Understanding the Impact of Microbial Biofilms and the Extracellular Polymeric Substance (EPS) Matrix in Drinking Water Distribution Systems." *Environmental Science: Water Research & Technology* 2(4):614–30. doi: 10.1039/C6EW00039H.
- Fish, Reeves-McLaren, Husband, and Boxall. 2020. "Uncharted Waters: The Unintended Impacts of Residual Chlorine on Water Quality and Biofilms." *Npj Biofilms and Microbiomes* 6(1):34. doi: 10.1038/s41522-020-00144-w.
- Fox, Shepherd, Collins, and Boxall. 2016. "Experimental Quantification of Contaminant Ingress into a Buried Leaking Pipe during Transient Events." *Journal of Hydraulic Engineering* 142(1). doi: 10.1061/(ASCE)HY.1943-7900.0001040.
- Friedman, Lakin, and Moore. 2012. "Best Practices Cleaning Mains: Clean, Pig, or Dig?" *Opflow* 38(8):10–13. doi: 10.5991/OPF.2012.38.0044.
- Furnass. 2015. "Modelling Both the Continual Accumulation and Erosion of Discolouration Material in Drinking Water Distribution Systems." The University of Sheffield.
- Furnass, Mounce, Husband, Collins, and Boxall. 2019. "Calibrating and Validating a Combined Accumulation and Mobilisation Model for Water Distribution System Discolouration Using Particle Swarm Optimisation." *Smart Water* 4(1):3. doi: 10.1186/s40713-019-0015-z.
- Gaffney, and Boulton. 2012. "Need for and Use of High-Resolution Turbidity Monitoring in Managing Discoloration in Distribution." *Journal of Environmental Engineering* 138(6):637–44. doi: 10.1061/(ASCE)EE.1943-7870.0000521.
- García-Ávila, Avilés-Añazco, Ordoñez-Jara, Guanuchi-Quezada, Flores del Pino, and Ramos-Fernández. 2021. "Modeling of Residual Chlorine in a Drinking Water Network in Times of Pandemic of the SARS-CoV-2 (COVID-19)." *Sustainable Environment Research* 31(1):12. doi: 10.1186/s42834-021-00084-w.
- García, Creus, Minoves, Pardo, Quevedo, and Puig. 2017. "Data Analytics Methodology for Monitoring Quality Sensors and Events in the Barcelona Drinking Water Network." *Journal of Hydroinformatics* 19(1):123–37. doi: 10.2166/hydro.2016.048.
- García, Puig, and Quevedo. 2020. "Prognosis of Water Quality Sensors Using Advanced Data

- Analytics: Application to the Barcelona Drinking Water Network.” *Sensors* 20(5):1342. doi: 10.3390/s20051342.
- Geocene. 2020. “Trainset.” Retrieved (<https://github.com/Geocene/trainset>).
- Geron. 2019. *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*. 2nd ed. Sebastopol, California: O’Reilly Media.
- Gibbs, Morgan, Maier, Dandy, Nixon, and Holmes. 2006. “Investigation into the Relationship between Chlorine Decay and Water Distribution Parameters Using Data Driven Methods.” *Mathematical and Computer Modelling* 44(5–6):485–98. doi: 10.1016/j.mcm.2006.01.007.
- Gleeson, Husband, Gaffney, and Boxall. 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, Vinga, and Henriques. 2021. “Spatiotemporal Correlation Feature Spaces to Support Anomaly Detection in Water Distribution Networks.” *Water* 13(18):2551. doi: 10.3390/w13182551.
- De Gooijer, and Hyndman. 2006. “25 Years of Time Series Forecasting.” *International Journal of Forecasting* 22(3):443–73. doi: 10.1016/j.ijforecast.2006.01.001.
- Gray. 2008. *Drinking Water Quality*. Cambridge University Press. doi: 10.1017/CBO9780511805387.
- Grubbs. 1969. “Procedures for Detecting Outlying Observations in Samples.” *Technometrics* 11(1):1. doi: 10.2307/1266761.
- Gschwandtner, and Erhart. 2018. “Know Your Enemy: Identifying Quality Problems of Time Series Data.” Pp. 205–14 in *2018 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE.
- Gupta, Gao, Aggarwal, and Han. 2014. “Outlier Detection for Temporal Data: A Survey.” *IEEE Transactions on Knowledge and Data Engineering* 26(9):2250–67. doi: 10.1109/TKDE.2013.184.
- Guralnik, and Srivastava. 1999. “Event Detection from Time Series Data.” Pp. 33–42 in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM.
- Halford, Baker, McCredden, and Bain. 2005. “How Many Variables Can Humans Process?”

- Psychological Science* 16(1):70–76. doi: 10.1111/j.0956-7976.2005.00782.x.
- Hart, McKenna, Klise, Cruz, and Wilson. 2007. “CANARY: A Water Quality Event Detection Algorithm Development Tool.” Pp. 1–9 in *World Environmental and Water Resources Congress 2007*. Reston, VA: American Society of Civil Engineers.
- Hillmer, and Tiao. 1982. “An ARIMA-Model-Based Approach to Seasonal Adjustment.” *Journal of the American Statistical Association* 77:63–70.
- Hochreiter, and Schmidhuber. 1997. “Long Short-Term Memory.” *Neural Computation* 9(8):1735–80. doi: 10.1162/neco.1997.9.8.1735.
- Holt. 1957. “Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages.” *Carnegie Institute of Technology, Pittsburg, USA. Reprinted in the International Journal of Forecasting* 20(1):5–10. doi: 10.1016/j.ijforecast.2003.09.015.
- Horvatic, Stanley, and Podobnik. 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, and Boxall. 2011. “Asset Deterioration and Discolouration in Water Distribution Systems.” *Water Research* 45(1):113–24. doi: 10.1016/j.watres.2010.08.021.
- Husband, and Boxall. 2016. “Understanding and Managing Discolouration Risk in Trunk Mains.” *Water Research* 107:127–40. doi: 10.1016/j.watres.2016.10.049.
- Husband, Boxall, and Saul. 2008. “Laboratory Studies Investigating the Processes Leading to Discolouration in Water Distribution Networks.” *Water Research* 42(16):4309–18. doi: 10.1016/j.watres.2008.07.026.
- Husband, Fish, Douterelo, and Boxall. 2016. “Linking Discolouration Modelling and Biofilm Behaviour within Drinking Water Distribution Systems.” *Water Supply* 16(4):942–50. doi: 10.2166/ws.2016.045.
- Husband, Whitehead, and Boxall. 2010. “The Role of Trunk Mains in Discolouration.” *Proceedings of the Institution of Civil Engineers - Water Management* 163(8):397–406. doi: 10.1680/wama.900063.
- Hyndman, and Athanasopoulos. 2021. *Forecasting: Principles and Practice*. 3rd ed. edited by OTexts. Melbourne, Australia.

- IWA. 2022. *A Strategic Digital Transformation for the Water Industry*. edited by Grievson, Holloway, and Johnson. IWA Publishing. doi: 10.2166/9781789063400.
- Johnson, Ambrose, Bassett, Bowen, Crummey, Isaacson, Johnson, Lamb, Saul, and Winter-Nelson. 1997. “Meanings of Environmental Terms.” *Journal of Environmental Quality* 26(3):581–89. doi: 10.2134/jeq1997.00472425002600030002x.
- Karamaziotis, Raptis, Nikolopoulos, Litsiou, and Assimakopoulos. 2020. “An Empirical Investigation of Water Consumption Forecasting Methods.” *International Journal of Forecasting* 36(2):588–606. doi: 10.1016/j.ijforecast.2019.07.009.
- Karkouch, Mousannif, Al Moatassime, and Noel. 2016. “Data Quality in Internet of Things: A State-of-the-Art Survey.” *Journal of Network and Computer Applications* 73:57–81. doi: 10.1016/j.jnca.2016.08.002.
- Kazemi, Mounce, Husband, and Boxall. 2018. “Predicting Turbidity in Water Distribution Trunk Mains Using Nonlinear Autoregressive Exogenous Artificial Neural Networks.” Pp. 1030–1019 in *HIC 2018. 13th International Conference on Hydroinformatics*. Vol. 3. Palermo, Italy.
- Keogh. 2012. *A Fast and Robust Method for Pattern Matching in Time Series Databases*.
- Keogh, and Pazzani. 2001. “Derivative Dynamic Time Warping.” Pp. 1–11 in *Proceedings of the 2001 SIAM International Conference on Data Mining*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Kianimajd, Ruano, Carvalho, Henriques, Rocha, Paredes, and Ruano. 2017. “Comparison of Different Methods of Measuring Similarity in Physiologic Time Series.” *IFAC-PapersOnLine* 50(1):11005–10. doi: 10.1016/j.ifacol.2017.08.2479.
- Kirchen, Schutz, Folmer, and Vogel-Heuser. 2017. “Metrics for the Evaluation of Data Quality of Signal Data in Industrial Processes.” Pp. 819–26 in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE.
- Klise, Nicholson, and Laird. 2017. *Sensor Placement Optimization Using Chama*. Albuquerque, NM, and Livermore, CA (United States). doi: 10.2172/1405271.
- Krause, Leskovec, Guestrin, VanBriesen, and Faloutsos. 2008. “Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks.” *Journal of Water Resources Planning and Management* 134(6):516–26. doi: 10.1061/(ASCE)0733-9496(2008)134:6(516).
- Krishnamachari, and Iyengar. 2004. “Distributed Bayesian Algorithms for Fault-Tolerant Event

- Region Detection in Wireless Sensor Networks.” *IEEE Transactions on Computers* 53(3):241–50. doi: 10.1109/TC.2004.1261832.
- Kyritsakas, Boxall, and Speight. 2023. “A Big Data Framework for Actionable Information to Manage Drinking Water Quality.” *AQUA — Water Infrastructure, Ecosystems and Society* 72(5):701–20. doi: 10.2166/aqua.2023.218.
- Lamprecht. 2023. “Meteostat Python.”
- LeChevallier, Gullick, Karim, Friedman, and Funk. 2003. “The Potential for Health Risks from Intrusion of Contaminants into the Distribution System from Pressure Transients.” *Journal of Water and Health* 1(1):3–14. doi: 10.2166/wh.2003.0002.
- Li, Chen, Jin, Shi, Goh, and Ng. 2019. “MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks.” Pp. 703–16 in *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series*. Springer, Cham.
- Li, and Parker. 2014. “Nearest Neighbor Imputation Using Spatial–Temporal Correlations in Wireless Sensor Networks.” *Information Fusion* 15:64–79. doi: 10.1016/j.inffus.2012.08.007.
- Lin, Keogh, Fu, and Van Herle. 2005. “Approximations to Magic: Finding Unusual Medical Time Series.” Pp. 329–34 in *18th IEEE Symposium on Computer-Based Medical Systems (CBMS’05)*. IEEE.
- Liu, Ting, and Zhou. 2008. “Isolation Forest.” Pp. 413–22 in *2008 Eighth IEEE International Conference on Data Mining*. IEEE.
- Liu, Zhang, Knibbe, Feng, Liu, Medema, and van der Meer. 2017. “Potential Impacts of Changing Supply-Water Quality on Drinking Water Distribution: A Review.” *Water Research* 116:135–48. doi: 10.1016/j.watres.2017.03.031.
- Loning, Bagnall, Ganesh, Kazakov, Lines, and Kiraly. 2019. “Sktime: A Unified Interface for Machine Learning with Time Series.” in *33rd Conference on Neural Information Processing Systems (NeurIPS)*. Vancouver, Canada.
- Machell, and Boxall. 2014. “Modeling and Field Work to Investigate the Relationship between Age and Quality of Tap Water.” *Journal of Water Resources Planning and Management* 140(9):04014020. doi: 10.1061/(ASCE)WR.1943-5452.0000383.
- Machell, Mounce, Farley, and Boxall. 2014. “Online Data Processing for Proactive UK Water Distribution Network Operation.” *Drinking Water Engineering and Science* 7(1):23–33. doi:

10.5194/dwes-7-23-2014.

- Makridakis, Spiliotis, and Assimakopoulos. 2018. “Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward” edited by Hernandez Montoya. *PLOS ONE* 13(3):e0194889. doi: 10.1371/journal.pone.0194889.
- Makridakis, Spiliotis, and Assimakopoulos. 2020. “The M4 Competition: 100,000 Time Series and 61 Forecasting Methods.” *International Journal of Forecasting* 36(1):54–74. doi: 10.1016/j.ijforecast.2019.04.014.
- Mann, Tam, Higgins, and Rodrigues. 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.
- McGuire. 2006. “Eight Revolutions in the History of US Drinking Water Disinfection.” *Journal - American Water Works Association* 98(3):123–49. doi: 10.1002/j.1551-8833.2006.tb07612.x.
- McKenna, Hart, Klise, Cruz, and Wilson. 2007. “Event Detection from Water Quality Time Series.” Pp. 1–12 in *World Environmental and Water Resources Congress 2007*. Reston, VA: American Society of Civil Engineers.
- McKinney. 2010. “Data Structures for Statistical Computing in Python.” in *Proceedings of the Python in Science Conferences*. Austin, Texas: SciPy.
- Meyers, Kapelan, and Keedwell. 2017. “Short-Term Forecasting of Turbidity in Trunk Main Networks.” *Water Research* 124:67–76. doi: 10.1016/j.watres.2017.07.035.
- Mounce. 2020. “Data Science Trends and Opportunities for Smart Water Utilities.” Pp. 1–26 in *ICT for Smart Water Systems: Measurement and Data Science*. Springer.
- Mounce, Gaffney, Boulton, and Boxall. 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.
- Mounce, Stephen, Machell, and Boxall. 2012. “Water Quality Event Detection and Customer Complaint Clustering Analysis in Distribution Systems.” *Water Science and Technology: Water Supply* 12(5):580–87. doi: 10.2166/ws.2012.030.
- Mounce, Stephen R., Mounce, and Boxall. 2012. “Identifying Sampling Interval for Event Detection in Water Distribution Networks.” *Journal of Water Resources Planning and Management*

138(2):187–91. doi: 10.1061/(ASCE)WR.1943-5452.0000170.

Mounce, Mounce, Jackson, Austin, and Boxall. 2014. “Pattern Matching and Associative Artificial Neural Networks for Water Distribution System Time Series Data Analysis.” *Journal of Hydroinformatics* 16(3):617–32. doi: 10.2166/hydro.2013.057.

Mu, Zheng, Tao, Zhang, and Kapelan. 2020. “Hourly and Daily Urban Water Demand Predictions Using a Long Short-Term Memory Based Model.” *Journal of Water Resources Planning and Management* 146(9). doi: 10.1061/(ASCE)WR.1943-5452.0001276.

Muharemi, Logofătu, and Leon. 2019. “Machine Learning Approaches for Anomaly Detection of Water Quality on a Real-World Data Set.” *Journal of Information and Telecommunication* 3(3):294–307. doi: 10.1080/24751839.2019.1565653.

Murray, and Haxton. 2010. *Water Quality Event Detection Systems for Drinking Water Contamination Warning Systems---Development, Testing, and Application of CANARY*. Cincinnati, OH.

Ngiam, Khosla, Nam, Lee, and Ng. 2011. “Multimodal Deep Learning.” in *Proceedings of the 28th International Conference on Machine Learning*. Bellevue, WA, USA.

Osman, Abu-Mahfouz, and Page. 2018. “A Survey on Data Imputation Techniques: Water Distribution System as a Use Case.” *IEEE Access* 6:63279–91. doi: 10.1109/ACCESS.2018.2877269.

P. Vatcheva, and Lee. 2016. “Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies.” *Epidemiology: Open Access* 06(02). doi: 10.4172/2161-1165.1000227.

Palma. 2016. *Time Series Analysis*. Hoboken, New Jersey: Wiley.

Pasteur. 1861. “Mémorial on the Organized Corpuscles That Exist in the Atmosphere.” *Annals of Natural Sciences* 16:5–98.

Pastorello, Agarwal, Papale, Samak, Trotta, Ribeca, Poindexter, Faybishenko, Gunter, Hollowgrass, and Canfora. 2014. “Observational Data Patterns for Time Series Data Quality Assessment.” Pp. 271–78 in *2014 IEEE 10th International Conference on e-Science*. IEEE.

Payment, Waite, and Dufour. 2003. “Introducing Parameters for the Assessment of Drinking Water Quality.” in *Assessing Microbial Safety of Drinking Water*. OECD.

Peng, Peng, Jiang, Wei, Li, and Tan. 2010. “Asymmetric Least Squares for Multiple Spectra Baseline

- Correction.” *Analytica Chimica Acta* 683(1):63–68. doi: 10.1016/j.aca.2010.08.033.
- Peng, Wang, Du, Bhuiyan, Ma, Liu, Wang, Yang, Du, Wang, and Yu. 2020. “Spatial Temporal Incidence Dynamic Graph Neural Networks for Traffic Flow Forecasting.” *Information Sciences* 521:277–90. doi: 10.1016/j.ins.2020.01.043.
- Perelman, Arad, Housh, and Ostfeld. 2012. “Event Detection in Water Distribution Systems from Multivariate Water Quality Time Series.” *Environmental Science & Technology* 46(15):8212–19. doi: 10.1021/es3014024.
- Peterson, Wanders, Horne, Collier, Alexander, Kaspi, and Maoz. 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):660–70. doi: 10.1086/316177.
- Potter, Wyble, Haggmann, and McCourt. 2014. “Detecting Meaning in RSVP at 13 Ms per Picture.” *Attention, Perception, & Psychophysics* 76(2):270–79. doi: 10.3758/s13414-013-0605-z.
- Reid Turner, Fuggetta, Lavazza, and Wolf. 1999. “A Conceptual Basis for Feature Engineering.” *Journal of Systems and Software* 49(1):3–15. doi: 10.1016/S0164-1212(99)00062-X.
- Roberts, Rossman, and Jarić. 2021. “Dating First Cases of COVID-19” edited by Lee. *PLOS Pathogens* 17(6):e1009620. doi: 10.1371/journal.ppat.1009620.
- Romano, Kapelan, and Savić. 2014. “Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems.” *Journal of Water Resources Planning and Management* 140(4):457–67. doi: 10.1061/(ASCE)WR.1943-5452.0000339.
- Rousseeuw, and Driessen. 1999. “A Fast Algorithm for the Minimum Covariance Determinant Estimator.” *Technometrics* 41(3):212–23. doi: 10.1080/00401706.1999.10485670.
- Sadiq, and Rodriguez. 2004. “Disinfection By-Products (DBPs) in Drinking Water and Predictive Models for Their Occurrence: A Review.” *Science of the Total Environment* 321(1–3):21–46. doi: 10.1016/j.scitotenv.2003.05.001.
- Sankary, and Ostfeld. 2018. “Analyzing Multi-Variate Water Quality Signals for Water Quality Monitoring Station Placement in Water Distribution Systems.” *Journal of Hydroinformatics* 20(6):1323–42. doi: 10.2166/hydro.2018.162.
- Savane, Khan, Freud, Murphy, Tesfargiogis, and Murray. 2019. “Persistent Pockets of Low Chlorine Residual in New York City’s Drinking Water Distribution System: A Case Study.” *Journal -*

- American Water Works Association* 111(10):40–50. doi: 10.1002/awwa.1379.
- Schober, Boer, and Schwarte. 2018. “Correlation Coefficients.” *Anesthesia & Analgesia* 126(5):1763–68. doi: 10.1213/ANE.0000000000002864.
- Sebri. 2016. “Forecasting Urban Water Demand: A Meta-Regression Analysis.” *Journal of Environmental Management* 183:777–85. doi: 10.1016/j.jenvman.2016.09.032.
- Shaukat, Alam, Luo, Shabbir, Hameed, Li, Abbas, and Javed. 2021. “A Review of Time-Series Anomaly Detection Techniques: A Step to Future Perspectives.” Pp. 865–77 in.
- Shortridge, and Guikema. 2014. “Public Health and Pipe Breaks in Water Distribution Systems: Analysis with Internet Search Volume as a Proxy.” *Water Research* 53:26–34. doi: 10.1016/j.watres.2014.01.013.
- Shumway, and Stoffer. 2017. *Time Series Analysis and Its Applications: With R Examples*. 4th ed. Springer.
- Speight, and Boxall. 2015. “Current Perspectives on Disinfectant Modelling.” *Procedia Engineering* 119:434–41. doi: 10.1016/j.proeng.2015.08.906.
- Speight, Mounce, and Boxall. 2019. “Identification of the Causes of Drinking Water Discolouration from Machine Learning Analysis of Historical Datasets.” *Environmental Science: Water Research & Technology* 5(4):747–55. doi: 10.1039/C8EW00733K.
- Speight, Quintiliani, Bosch, Rajnochova, Schaap, and Husband. 2020. *Guidance Manual: Water Quality in Drinking Water Distribution Systems*.
- Stańczyk, Kajewska-Szkudlarek, Lipiński, and Rychlikowski. 2022. “Improving Short-Term Water Demand Forecasting Using Evolutionary Algorithms.” *Scientific Reports* 12(1):13522. doi: 10.1038/s41598-022-17177-0.
- Stolarski, Krueger, Schoeberl, McPeters, Newman, and Alpert. 1986. “Nimbus 7 Satellite Measurements of the Springtime Antarctic Ozone Decrease.” *Nature* 322(6082):808–11. doi: 10.1038/322808a0.
- van Summeren, Raterman, Vonk, Blokker, van Erp, and Vries. 2015. “Influence of Temperature, Network Diagnostics, and Demographic Factors on Discoloration-Related Customer Reports.” *Procedia Engineering* 119:416–25. doi: 10.1016/j.proeng.2015.08.903.
- Sunny, Husband, and Boxall. 2019. “Impact of Hydraulic Interventions on Chronic and Acute

- Material Loading and Discolouration Risk in Drinking Water Distribution Systems.” *Water Research* 169:115224. doi: 10.1016/j.watres.2019.115224.
- Talagala, Hyndman, Leigh, Mengersen, and Smith-Miles. 2019. “A Feature-Based Procedure for Detecting Technical Outliers in Water-Quality Data From In Situ Sensors.” *Water Resources Research* 55(11):8547–68. doi: 10.1029/2019WR024906.
- Taylor, and Letham. 2018. “Forecasting at Scale.” *The American Statistician* 72(1):37–45. doi: 10.1080/00031305.2017.1380080.
- Teh, Kempa-Liehr, and Wang. 2020. “Sensor Data Quality: A Systematic Review.” *Journal of Big Data* 7(1):11. doi: 10.1186/s40537-020-0285-1.
- Thayanukul, Kurisu, Kasuga, and Furumai. 2013. “Evaluation of Microbial Regrowth Potential by Assimilable Organic Carbon in Various Reclaimed Water and Distribution Systems.” *Water Research* 47(1):225–32. doi: 10.1016/j.watres.2012.09.051.
- Thudumu, Branch, Jin, and Singh. 2020. “A Comprehensive Survey of Anomaly Detection Techniques for High Dimensional Big Data.” *Journal of Big Data* 7(1):42. doi: 10.1186/s40537-020-00320-x.
- Tukey. 1977. *Exploratory Data Analysis*. Addison-Wesley.
- UN-INWEH. 2020. *Water and Migration: A Global Overview*. Hamilton, Canada.
- USEPA. 2018. *2018 Edition of the Drinking Water Standards and Health Advisories Tables*.
- Ustalov, Pavlichenko, Losev, Giliazev, and Tulin. 2021. “A General-Purpose Crowdsourcing Computational Quality Control Toolkit for Python.” doi: <https://doi.org/10.48550/arXiv.2109.08584>.
- Vaidya, Ambad, and Bhosle. 2018. “Industry 4.0 – A Glimpse.” *Procedia Manufacturing* 20:233–38. doi: 10.1016/j.promfg.2018.02.034.
- Vandecar, and Crosson. 1990. “Determination of Teleseismic Relative Phase Arrival Times Using Multi-Channel Cross-Correlation and Least Squares.” *Bulletin - Seismological Society of America* 80(1):150–69.
- Virtanen, Gommers, Oliphant, Haberland, Reddy, Cournapeau, Burovski, Peterson, Weckesser, Bright, van der Walt, Brett, Wilson, Millman, Mayorov, Nelson, Jones, Kern, Larson, Carey, Polat, Feng, Moore, VanderPlas, Laxalde, Perktold, Cimrman, Henriksen, Quintero, Harris,

- Archibald, Ribeiro, Pedregosa, van Mulbregt, Vijaykumar, Bardelli, Rothberg, Hilboll, Kloeckner, Scopatz, Lee, Rokem, Woods, Fulton, Masson, Häggström, Fitzgerald, Nicholson, Hagen, Pasechnik, Olivetti, Martin, Wieser, Silva, Lenders, Wilhelm, Young, Price, Ingold, Allen, Lee, Audren, Probst, Dietrich, Silterra, Webber, Slavič, Nothman, Buchner, Kulick, Schönberger, de Miranda Cardoso, Reimer, Harrington, Rodríguez, Nunez-Iglesias, Kuczynski, Tritz, Thoma, Newville, Kümmerer, Bolingbroke, Tartre, Pak, Smith, Nowaczyk, Shebanov, Pavlyk, Brodtkorb, Lee, McGibbon, Feldbauer, Lewis, Tygier, Sievert, Vigna, Peterson, More, Pudlik, Oshima, Pingel, Robitaille, Spura, Jones, Cera, Leslie, Zito, Krauss, Upadhyay, Halchenko, and Vázquez-Baeza. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” *Nature Methods* 17(3):261–72. doi: 10.1038/s41592-019-0686-2.
- Vreeburg. 2007. “Discolouration in Drinking Water Systems: A Particular Approach.” TU Delft.
- Wang, and Strong. 1996. “Beyond Accuracy: What Data Quality Means to Data Consumers.” *Journal of Management Information Systems* 12(4):5–33. doi: 10.1080/07421222.1996.11518099.
- Weston, Scheili, Behmel, and Rodriguez. 2022. “Water Quality in Drinking Water Distribution Systems: Research Trends through the 21st Century.” *Environmental Science: Water Research & Technology*. doi: 10.1039/D2EW00491G.
- WHO. 2017. *Guidelines for Drinking Water Quality*.
- WHO. 2021. *Asbestos in Drinking Water*. Geneva.
- Yang Zhang, Meratnia, and Havinga. 2010. “Outlier Detection Techniques for Wireless Sensor Networks: A Survey.” *IEEE Communications Surveys & Tutorials* 12(2):159–70. doi: 10.1109/SURV.2010.021510.00088.
- Yu, Bambacus, Cervone, Clarke, Duffy, Huang, Li, Li, Li, Liu, Resch, Yang, and Yang. 2020. “Spatiotemporal Event Detection: A Review.” *International Journal of Digital Earth* 13(12):1339–65. doi: 10.1080/17538947.2020.1738569.
- Yu, Xu, Xie, Hou, Huang, Zhang, and Zhang. 2017. “Contamination Event Detection Method Using Multi-Stations Temporal-Spatial Information Based on Bayesian Network in Water Distribution Systems.” *Water* 9(11):894. doi: 10.3390/w9110894.
- Zhang, and Thorburn. 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.

Appendix 1 – Sensor Specifications

Table A: ATI MetriNet specifications per parameter.

<u>Parameter</u>	<u>Range</u>	<u>Resolution</u>
Free Chlorine	0-5.00 ppm	0.01 ppm
Combined Chlorine	0-5.00 ppm	0.01 ppm
Total Chlorine	0-5.00 ppm	0.01 ppm
Nitrite	0-2.000 ppm	0.001 ppm
Turbidity	0-40.00 NTU	0.01 NTU
pH	0-14.00 pH	0.01 pH
Conductivity	0-2000 μ S	1 μ S
ORP	0-1000 mv.	1 mv.
Dissolved Oxygen	0-20.00 ppm	0.01 ppm
Fluoride	0.1-10.00 ppm	0.01 ppm
Dissolved Ozone	0-5.00 ppm	0.01 ppm
Chlorine Dioxide	0-5.00 ppm	0.01 ppm
Peracetic Acid	0-200 ppm	1 ppm
Hydrogen Peroxide	0-20.00 ppm	0.01 ppm
Pressure	0-300 PSI	1 PSIG
Accuracy	1% of range	
Repeatability	0.5% of range	
Sensitivity	0.1% of range	
Non-Linearity	0.1% of range	
Warm-up Time	3 seconds to rated performance (electronics only)	
Supply Voltage Effects	+/- 0.05% of range	

Table B: Intellitect Water Intellisonde specifications per parameter.

Parameter	Performance
Flow (forward and downstream sensors, flow direction information dependent on installation)	0 to 2 m/sec \pm 10%
Temperature	-5 to +50°C \pm 0.2°C
Chlorine	0-5mg/l, \pm 5% or \pm 0.05 mg/l, whichever is the greater.
Total Chlorine	0 to 5 mg/l \pm 10% or \pm 0.5 mg/l
Dissolved Oxygen	0 to 20 mg/l \pm 10%
pH	2 to 12 pH \pm 0.2 in typical installations
ORP	-1.0 volt to +1.0 volt \pm 1mV
Conductivity	0 to 10 mS/cm \pm 5%
Colour (dependent on water quality and flow)	0 to 50 Hazen \pm 1 Hazen
Turbidity	0 to 50 NTU \pm 0.5 NTU
Pressure (4-20mA or 0-100mV input, external transducer)	0 to 10 Bar
ISE (sensor output reported directly, configurable for mV or V).	0 to 1 volt \pm 1 mV

Table C: Salamander Hydraclam specifications per parameter.

TURBIDITY SENSOR

Measurement method	Nephelometric
Range	0.1 - 10 NTU
Accuracy	± 5% of reading or ± 0.1 NTU
Resolution	0.05 NTU

CONDUCTIVITY SENSOR

Measurement method	4 pole
Range	20 – 3500 µs
Accuracy	± 2% of range
Resolution	1 µs

PRESSURE SENSOR

Measurement method	Silicon micro machined element
Range	0 – 10 bar Absolute
Accuracy	± 1.25% of full scale
Resolution	0.1 bar

CALIBRATION

Factory calibrated using standards at 1 and 10 NTU
No in-service calibration required

INTERNAL POWER

2 x LSH20 3.6V High Discharge Lithium Ion Battery

MEMORY

Up to 50,000 data points within the device
--

DATA INTERVALS

Programmable between 1 minute and 1 hour
--

ENVIRONMENTAL

Waterproofing	IP68
Operating Temperature	0 - 40 °C
Storage	-5 to +65 °C
Mains Pressure	1 - 10 bar
Sample Flow	6 l/sample

COMMUNICATIONS

Cellular data	4G network
Modem	4G Internal antenna, external option

EMC

Hydraclam	BS EN 61326-1:2006
	EN 301 489-1 v1.8.1
	EN 301 489-7 v 1.3.1

SERVICE INTERVAL

Clean sensor head	Recommended every 6 months but will be dependent on operating conditions
-------------------	--

WEIGHT

Clam unit	0.5 kg approx.
Sensor unit	0.7 kg approx.

DIMENSIONS

Clam unit	170mm height x 160mm diameter
Sensor unit	Max 220 x 130 x 160mm

DATA STORAGE

Secure web portal on AWS, data can be extracted via API
