



The
University
Of
Sheffield.

Development of a risk based approach to surface water abstraction

Vaida Suslovaite

Department of Civil and Structural Engineering

Submitted in part fulfilment of the requirements for the degree of Engineering
Doctorate in the Faculty of Engineering, University of Sheffield

31/07/2023

Intentionally Blank

Abstract

Pollution resulting from rainfall driven processes is known to adversely affect surface water quality. It arises from a variety of agricultural and urban point and non-point sources and atmospheric deposition. The issue, in addition to ecological impacts, is especially problematic in areas where surface water is used for drinking water supply. This study aims to investigate and develop tools to understand and manage water quality risks to water abstraction sites. The project focuses on risks caused by acute rainfall driven loadings and investigates short-term dynamics of water quality parameters.

First part of the thesis describes the deployment and testing of a commercially available water quality probe, intended to provide real time estimations of bacterial water quality in surface waters. The probe is evaluated based on direct comparison of *E. coli* quantified using standard techniques collected during wet weather events. It is not recommended as a current robust methodology to characterize *E. coli* loadings or provide early warning to bathing water or water abstraction sites.

Second part of the thesis proposes and tests a new modelling approach to describe the temporal dynamics of *E. coli* in the case study catchment based on Storm overflow asset and rainfall data. The developed model enables reasonable approximations of arrival times and durations of *E. coli* at the water abstraction site and is therefore judged to be fit for purpose in providing useful information to abstraction operators for decision making purposes.

The final part of the thesis presents a new methodology to reduce the impact of pesticide runoff on water abstraction sites. It is based on an inverse modelling/optimisation approach to identify priority areas for catchment mitigation. The methodology developed was found to be effective in reducing modelled pesticide levels at the water abstraction site based on the selective targeting of mitigation options in the catchment.

Keywords

Real-time sensing, *E. coli*, Bacterial pollution, FIO, Faecal Indicator Organisms, Acute diffuse pollution, SSO, Storm overflows, Surface water abstraction, Diffuse pollution modelling, Metaldehyde, Propyzamide, Pesticide, Rainfall-runoff, Hydrological forecasting, Water resources, Genetic algorithm, Land use optimisation.

Acknowledgements

First and foremost, I would like to thank my supervisors Dr. James Shucksmith and Professor Vanessa Speight for their continuous support and guidance throughout this PhD. Special thanks to my industrial supervisor Helen Pickett who has put much effort into familiarising myself with the water industry and making sure I had access to everything needed to carry out my research.

I would like to acknowledge the funding received through EPSRC Doctoral Training Partnership (EPSRC DTP) Case Award (with Severn Trent Water) Scholarship. Special thanks go to the Severn Trent staff at Church Wilne microbiology laboratory for providing me with training and assistance to carry out sampling and analysis for my research. I would like to also thank Laura Flower and Marion Perrett-Pearson at Severn Trent, who helped identify and refine the needs within the catchment I was investigating.

I would like to thank my husband Chris, for supporting me throughout the PhD, especially the last few months full of stress induced by my seemingly exponentially increasing dislike for writing. Also, a thank you to the rest of my family and friends for their encouragement and support in general.

Table of Contents

List of Abbreviations.....	X
Chapter 1 – Introduction	1
1.1 Project Aim	3
1.2 Thesis structure	3
Chapter 2 - Literature review	4
2.1 Current water management issues	4
2.2 Water quality parameters	5
2.2.1 Bacterial water quality	5
2.2.2 Pesticide pollution in river water	5
2.3 Modelling of diffuse bacterial pollution and SSOs.....	6
2.3.1 Current bacterial models	7
2.3.2 Bacterial loadings from agricultural sources	9
2.3.3 Bacterial loadings from urban sources.....	9
2.4 Modelling of pesticide impacts	10
2.5 Land use optimization.....	12
2.6 Summary of Knowledge Gaps.....	14
Chapter 3 — On site evaluation of real-time sensing technologies for characterisation of acute bacterial loads in river systems.....	16
3.1 Introduction	16
3.1.1 Study area	19
3.2 Proteus Water quality probe	20
3.3 Test Events.....	22
3.4 Sample collection	23
3.5 Sample analysis	24
3.6 Results and discussion.....	25
3.6.1 Calibration	25
3.6.1 Baseline testing	25
3.6.2 Rainfall Runoff Events.....	27

3.7 Conclusion	34
Chapter 4 - Forecasting acute rainfall driven <i>E. coli</i> impacts in inland rivers based on sewer monitoring and rainfall runoff.....	35
4.1 Introduction	35
4.2 Methodology	38
4.2.1 Study area	38
4.2.2 Development of <i>E. coli</i> modelling approach.....	39
4.2.3 Surface Runoff Modelling	40
4.2.4 Diffuse Faecal Pollution Loading and Routing	40
4.2.5 SSO Spill Volumes, Loading and Routing.....	43
4.3 Model Input Data, Water Quality Sampling and Calibration.....	45
4.3.1 Rainfall and river flow	45
4.3.2. Land use.....	45
4.3.3 Storm Sewer Overflow Data.....	46
4.3.4 Water Sampling (<i>E. coli</i> data).....	46
4.3.5 Model Calibration.....	47
4.4 Results and Discussion	47
4.4.1 Surface runoff model	47
4.4.2 Measured <i>E. coli</i> Dynamics and Model Performance	50
4.4.3 Sensitivity Analysis	55
4.4.4 Annual Simulation of <i>E. coli</i> over Jan - Dec 2022.....	56
4.5. Discussion.....	58
4.6. Conclusions	61
Chapter 5 - Genetic algorithm based land use optimisation for the mitigation of pesticide risk to potable water supply	63
5.1 Introduction	63
5.2 Development of the Propyzamide Model.....	64
5.2.1 Calibration/Validation of propyzamide model.....	66
5.3 Development of Inverse Modelling method for Designing Catchment Management Options.....	68

5.3.1 Zero-one integer programming	70
5.3.2 Rainfall event mashup.....	71
5.4 Results and discussion.....	76
5.4.1 Land-use optimisation	76
5.4.2 Validation under historical rainfall.	79
5.5 Conclusion	81
Chapter 6 – Summary and Conclusions.....	84
The overall contribution of this thesis can be summarized as follows.....	84
6.1 Real-time sensing technologies	84
6.2 Forecasting <i>E. coli</i> impacts	84
6.3 Land use optimization.....	85
6.4 Limitations and future work	85
References.....	87

List of Figures

Figure 2.1 Pathway of faecal microorganisms including surface runoff from soil to survival in surface water bodies; green lines represent the pathway of faecal microorganism release and transport and black lines point to specific fate-related process that need to be modelled from (Cho et al., 2016).....	7
Figure 2.2 Pathways of microorganism release and removal; ① and ② - above ground and belowground (soil) reservoir, respectively.; a e release in suspension to soil, b e release in suspension to runoff, c e release in suspension from soil to runoff, d e release absorbed microorganisms with soil particles to runoff, e e infiltration with runoff suspension, f e settling with sediment particles, g - overall removal (or export) from the application site from (Cho et al., 2016).	8
Figure 2.3 Calibration (B1) and validation (B2-B4) events for the Metaldehyde model (Asfaw et al., 2018)	12
Figure 3.1 Catchment average rainfall and river flow together with historical E. coli data from a nearby WTW intake provided by Severn Trent Water.	20
Figure 3.2 Top left - Installation of the sensor(April 2021); top right – installation of mesh to stop reed entanglement (June 2022); Bottom left – autosampler intake set up(July, 2021); bottom right sensors and wiper head (Jan 2022).	22
Figure 3.3 Route and duration of transporting the water samples to the lab (image from Google maps, 2023).....	24
Figure 3.4 Calibration result plot for E1 as supplied by Proteus Instruments.....	25
Figure 3.5 Low flow warm weather 24h event (E0). Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners(ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.	26
Figure 3.6 E1 rainfall event. Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners (ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.....	28
Figure 3.7 E2 rainfall event. Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners(ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.....	29
Figure 3.8 E3 rainfall event. Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners (ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.....	31

Figure 3.9 E4 rainfall event. Calibrated sensor and lab analysed <i>E. coli</i> (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners (ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.....	32
Figure 3.10 <i>Proteus</i> vs lab analysed <i>E. coli</i> log/log plot.....	33
Figure 4.1 Study catchment map showing elevation (meters above sea level) the locations of SSO's, build up areas and grassland for livestock grazing.	39
Figure 4.2 Spatial distribution of temporally averaged rainfall (mm) for the events used in hydrological model validation.....	49
Figure 4.3. Field Runoff hydrological model validation (modelled and predicted flow, spatially averaged catchment rainfall)	49
Figure 4.4 Spatial distribution of temporally averaged rainfall (mm) for the events used in <i>E. coli</i> model calibration (E1) and validation (E2-4). Note E4 is plotted using an altered scale. .	51
Figure 4.5 Results of combined model and measured (LAB) <i>E. coli</i> data showing the contribution from the field runoff and SSO models and spatially averaged rainfall ($K_1=0.4$, $K_2=0.4$).	52
Figure 4.6 Forecasted vs lab analysed <i>E. coli</i> log/log plot	53
Figure 4.7 Sensitivity of SSO model to dispersion fraction (Df) parameter for event E3.	56
Figure 5.1 Distribution of propyzamide high risk fields identified for 2021	65
Figure 5.2 Calibration results for propyzamide model against measured propyzamide data .	67
Figure 5.3 Validation results for propyzamide model against measured propyzamide data ..	67
Figure 5.4 Flow chart of the GA method. The initial solution represents the current distribution of high risk land use. Checking the objective function runs the model with the new high risk shapefile and checks the resulting forecasted total hours pesticide levels are above threshold, and redefines objective function.	70
Figure 5.5 Temporally averaged rainfall (mm) form September-December 2015-2019	73
Figure 5.6 Method of rainfall event mash-up production	74
Figure 5.7 Histograms of rainfall event spatial standard deviation for full original and sampled 'mashup' datasets.	75
Figure 5.8 Histograms of rainfall event temporal standard deviation for full original and sampled 'mashup' datasets.	76
Figure 5.9 Plot of all the objective function results for metaldehyde field distribution.....	77
Figure 5.10 Plot of all the objective function results for propyzamide field distribution.....	77
Figure 5.11 Fields removed as initial solution and resulting best genetic algorithm solution for Metaldehyde Model.....	78
Figure 5.12 Fields removed as initial solution and resulting best genetic algorithm solution for Propyzamide Model	79

Figure 5.13 Metaldehyde model time series of all 188 events run with Full HR shapefile and 5% removed best GA solution.	81
Figure 5.14 Propyzamide model time series of all 188 events run with Full HR shapefile and 5% removed best GA solution.	81

List of Tables

Table 3.1 E. coli sampling event dates, durations and sampling frequencies. Rainfall duration is based on analysis of radar rainfall data from the UK MET office (see chapter 4 for further details).....	23
Table 3.2 Stepwise regression model summary for laboratory E. coli data	33
Table 4.1 Bovine and ovine die-off constants (b) for different seasons, from Avery et al., (2004) and Oliver et al., (2009).....	41
Table 4.2 Catchment livestock densities and total grassland area derived from DEFRA County level data, utilizing deposit data from Oliver et al., 2009 and assumed grazing periods in the catchment (Oliver et al., 2018).....	42
Table 4.3 Characteristics of Leam catchment SSOs included in E. coli model, based on monitored period between April 2021 - March 2023 and EDM data is a sum of annual return data from 2021 and 2022.....	46
Table 4.4 Summary of rainfall events used to re-evaluate the surface runoff model. Quoted durations are based on presence of rainfall at any position in the catchment. Intensities are based on temporal and spatial averaged values. Initial and peak flow rates during each event based on EA gauging station data.....	48
Table 4.5 Flow simulation model error statistics.....	48
Table 4.6 E. coli sampling event dates, durations, sampling frequencies, and associated catchment rainfall statistics. Estimated 1 year R.P. rainfall depths for each event duration are also provided based on the UKCEH web service. Initial and peak flow rates during each event based on EA gauging station data.....	53
Table 4.7 Combined model error statistics	53
Table 4.8 Total calculated E.coli loads (CFU) over each event.	54
Table 5.1 Propyzamide sampling event dates, durations and sampling frequencies, and associated rainfall statistics. Quoted durations are based on presence of rainfall at any position in the catchment. Intensities are based on temporal and spatial averaged values...	66
Table 5.2 Propyzamide model error statistics	67
Table 5.3 Validation of optimization methodology under full dataset	80

List of Abbreviations

ADZ	Aggregated Dead Zone
ATP	Adenosine triphosphate
BOD	Biochemical Oxygen Demand
CEH	Centre of Ecology and Hydrology
CDOM	Coloured Dissolved Organic Matter
CFU	Colony Forming Units
CN	Curve Number
DEFRA	Department for Environment, Food & Rural Affairs
DEM	Digital Elevation Model
DO	Dissolved Oxygen
DOM	Dissolved Organic Matter
DWI	Drinking Water Inspectorate
EA	Environment Agency
FIOs	Faecal Indicator Organisms
FWR	Foundation for Water Research
GA	Genetic Algorithm
NTU	Nephelometric Turbidity Unit
ppb	parts per billion
SCS	Soil Conservation Service
SSOs	Combined/Storm Sewer Overflows; Storm water Overflows
STW	Severn Trent Water
STWs	Sewage Treatment Works
TLF	Tryptophan-Like Fluorescence
UPM	Urban Pollution Manual
VCI	volume conservation index
WHO	World Health Organization
WTW	Water Treatment Works
WWTW	Waste Water Treatment Works

Chapter 1 – Introduction

Pollution resulting from rainfall driven processes is known to adversely affect surface water quality. It arises from a variety of agricultural and urban point and non-point sources and atmospheric deposition. Nutrients, pesticides, faecal bacteria, sediment and other chemicals are transported to watercourses when runoff processes driven by rainfall mobilise pollutants accumulated in the soils (FAO and IWMI, 2018) or when rainfall causes overloading of urban drainage networks which results in spills from combined drainage systems (Giakoumis and Voulvoulis, 2023). Sewerage undertakers are required by S.82 of the Environment Act 2021 to continuously monitor the quality of the receiving water upstream and downstream of their assets. Monitoring for pH and temperature, turbidity, dissolved oxygen, ammonia, and anything else specified in regulations made by the Secretary of State must be every hour or every 15 minutes during high-risk periods (DEFRA, 2023). The issue, in addition to ecological impacts, is especially problematic in areas where surface water is used for drinking water supply. Clean up of certain pollutants, before the drinking water is ready and suitable for potable supply, is expensive and time consuming to carry out. Between 2004-05 and 2008-09, water companies in England spent £189 million on nitrate, £92 million on pesticide and additional unquantified costs relating to bacterial contamination removal from their water supplies (National Audit Office, 2010). Health risks associated with drinking water can be severe, with a large number of waterborne pathogens able to cause mild to severe illness and death (Magana-Arachchi and Wanigatunge, 2020). In England and Wales, where drinking water treatment practices are well established, overloaded water treatment works can still lead to waterborne disease cases, with cryptosporidium accounting for majority of outbreaks (Smith et al., 2006; Chalmers et al., 2019). Pesticide treatment options and their effectiveness for drinking water are limited (Tröger et al., 2021; Saleh et al., 2020). Pesticide removal techniques are divided into chemical, physical, and biological methods. Chemical wastewater treatment consists of a variety of chemical reactions that help in hydrolysing contaminants into safer chemicals. Physical treatment techniques include membrane filtration. While biological treatment methods are designed to treat pesticide polluted water using microorganisms capable of digesting pesticides. Most effective removal techniques for pesticide removal combine these methods into hybrid technologies. While health risks posed by chronic exposure to pesticide residue in drinking water after treatment are unknown (Damalas and Eleftherohorinos, 2011). The indicators used to assess the potential risk of pesticides to human health and the environment show reduced certainty, suggesting the need for development of alternative indicators to increase the accuracy and reliability of pesticide risk assessment. An improved understanding of water quality risks to surface drinking water abstraction and treatment processes could

improve the efficiency of treatment processes and reduce both asset operational costs and risks to public health.

Risks to water abstraction systems can arise from 'acute' water quality impacts. For example, impacts from Combined/Storm Sewer Overflows (SSO) have significant detrimental effects on water quality, but can be highly intermittent with durations ranging from minutes to hours. Similarly diffuse agricultural runoff loadings can vary significantly at over hourly timescales. Short term loadings of highly contaminated water into abstraction sites, can cause specific risks to drinking water treatment processes, with potential for poor quality water to enter potable water supply systems (Kyritsakas et al., 2023). Effective understanding and management of these risks to water supply systems requires a detailed understanding of catchment processes at high temporal resolution. Unfortunately, standard regulatory monitoring of water quality parameters in surface waters commonly occurs at weekly\monthly timescales, with the purpose of understanding the background or overall health of the river network.

Specific risks to water abstraction systems can come from a variety of sources, these may include pesticides (from cultivated land), bacterial loading such as *E. coli* or cryptosporidium from livestock or human waste (¹WHO, 2017). Many of these specific parameters are time consuming and technologically challenging to monitor (Boxal et al., 2021; Lundqvist et al., 2019). Datasets describing fluxes of these parameters within surface waters are scarce. Whilst water quality probes to characterize standard water quality parameters in 'real time' (e.g. pH, turbidity, DO) are widespread, the effectiveness of using such parameters for understanding the behaviour of bacterial or other emerging pollutants of concern is limited (Saalidong et al., 2022; ²WHO, 2017; Sánchez et al., 2007). Due to a lack of datasets, predictive models for forecasting risks to water abstraction points (e.g. Asfaw, 2019; Dorner et al., 2006; Ferguson et al., 2006) are also uncommon. With most widely used models calibrated and validated based on low resolution datasets (Cho et al., 2016; Quilbé et al., 2006), such models are commonly applied to understand trends over long terms timescales, with limited representation of acute impacts.

Currently, the UK water industry utilises catchment management options to reduce diffuse water pollution and reduce costs associated with drinking water treatment. Current options, such as STW Farm to tap (Cooke et al., 2020) work with farmers to help mitigate the impact from pesticides, (specifically metaldehyde) through encouragement to use alternative options and reward participants based on water quality results. The scheme is available to anyone within priority catchments who meet the criteria of growing winter wheat and/or oil seed rape. Within STW the scheme paid out £749,206 in 2016-2018, with average reduction in metaldehyde concentrations of 46% at water treatment works (Cooke et al., 2020). However,

there is no current methodology for identifying specific areas within the priority catchments for targeted intervention. This would reduce the costs associated with this type of blanket mitigation measures, and hence allow a greater range of pollutants and/or to be targeted.

1.1 Project Aim

This study aims to investigate and develop tools to understand and manage water quality risks to water abstraction sites. The project focuses on risks caused by acute rainfall driven loadings and investigates short-term dynamics of water quality parameters at a case study water abstraction site in response to rainfall. The case study site and several datasets are provided by the project industrial sponsor (Severn Trent Water) who have also assisted with the collection of further datasets described in this thesis. Within this work, two specific pesticides as well as bacterial water quality parameters are considered.

1.2 Thesis structure

Chapter 2 provides a literature review concerning water quality risks to water abstraction, water quality monitoring and modelling of acute water quality impacts, and the use of catchment management approaches.

Chapter 3 describes the case study catchment used in this work, as well as the deployment and testing of a commercially available water quality probe which is intended to provide estimations of bacterial water quality (*E. coli*) in surface waters in real time. The probe is evaluated based on direct comparison of *E. coli* quantified using standard techniques collected during wet weather events.

Chapter 4 proposes and tests a new modelling approach to describe the temporal dynamics of *E. coli* in the case study catchment based on SSO asset and rainfall data. The methodology is validated against the datasets collected in Chapter 3. This chapter is written as a stand-alone piece of work and has been submitted for publication.

Chapter 5 describes a new methodology to inform catchment management approaches to reduce the impact of pesticide runoff on water abstraction sites. This is based on an inverse modelling/optimisation approach to identify priority areas for catchment mitigation.

Chapter 6 provides an overarching summary and conclusions of the of the work in this thesis.

Chapter 2 - Literature review

The literature review provides a general context of current water management issues, a summary of water quality parameters considered in this thesis and well as modelling approaches and computational techniques for managing catchments.

2.1 Current water management issues

Water resources are under pressure from population growth, pollutant emergence and climate change. The Office for National Statistics (2019) predicts that the UK population will increase by 3 million (4.5%) by mid-2028. Global population growth rate is projected to be higher with global population increasing by 0.8 billion (10%) between 2019 and 2030 (United Nations, 2019) leading to increased water demand. General climate change trends projected over UK land for the 21st century in UK Climate projections (UKCP18) show an increased chance of warmer, wetter winters and hotter, drier summers along with an increase in the frequency and intensity of extremes (Met Office, 2019).

UK Governments 25 year environment plan shows commitment to take steps to achieve, among other environmental objectives, clean and plentiful water. The EU legislation on water quality intended for human consumption will be assimilated into UK law as retained EU law with amendments (as stated as an amendment to The Floods and Water Act 2019) and came into force on EU exit day. Associated guidance to water companies have been issued by Drinking Water Inspectorate (DWI, 2020) on the implementation of The Water Supply (Water Quality) Regulations 2016 (as amended) in England and The Water Supply (Water Quality) Regulations (Wales) 2018. Under these regulations water suppliers have a statutory duty to supply wholesome water. Water intended for human consumption should not contain any micro-organism, parasite or substance at a concentration that would constitute a potential danger to human health.

The effects of climate change with more frequent and intense extremes in rainfall and droughts will result in increased occurrence of acute pollution events (Graydon et al., 2022). This may arise due to both increased pollutant wash off into river systems, and/or increasing periods of low river flows with corresponding reduced capacity for dilution of pollutants. The likely increased pressure on water treatment facilities will lead to increased costs and higher risks of drinking water quality failures, which will in turn heighten risks to human health. Growing risks to drinking water resources in both quality and quantity will be capital intensive to counteract.

There is therefore a driver to better manage surface water resources through the deployment of increased and improved sensing and monitoring, as well as improved modelling and forecasting of surface waters, water demand and infrastructure (Yassin et al. 2021).

2.2 Water quality parameters

2.2.1 Bacterial water quality

The effects of microbial waterborne pollution on human health can be acute, with some cases resulting in death (WHO, 2017). Faecal indicator organisms (FIOs) are a group of organisms that indicate the presence of faecal contamination, such as the bacterial groups thermotolerant coliforms or *E. coli*. Hence, they only infer that pathogens may be present (WHO, 2001). In the UK, drinking water monitoring for microbial pollutants include: *Clostridium perfringens*, Coliform bacteria, Colony counts at 22 C, Enterococci and *Escherichia coli* (*E. coli*) DWI (2020). These tests are indicators of microbiological water quality. Although *Cryptosporidium* is not a parameter to be monitored under current legislation, it is still monitored due to severe health issues it can cause.

There are two major sources of faecal contamination – human origin and agricultural sources (WHO, 2012). Human origin pollution comes from point sources such as SSOs and waste water treatment works (WWTWs), that discharge directly into the river. Animal origin agricultural source pollution can originate over a widespread area of the catchment from slurry spreading on arable land or livestock grazing on grassland. The risk from agricultural sources will differ from catchment to catchment as hydrological transfer pathways and phases are catchment specific (Murphy et al., 2015). Different stocking densities, type of livestock, land use within the catchment all impact the bacterial load in the river and many studies have attempted to quantify the risk and implement the mitigation action to combat the said risk of bacterial pollution (Kay, 2007; Oliver et al., 2009, Bragina et al., 2017; Dwivedi et al, 2013). Due to the difficulty of measurement, regulatory sampling and published datasets are generally of coarse resolution Zan et al., 2023. This has implications for water quality modelling as it is difficult to directly characterise acute impacts because they are highly intermittent and difficult to capture.

2.2.2 Pesticide pollution in river water

Pesticides are used to control insects and weeds and rainfall driven runoff transports them into surface water bodies. In the United Kingdom water supply regulations set the maximum current legal UK limit for drinking water at 0.1 µg/l of one particular pesticide or not more than 0.5µg/l of all pesticides present in total. The limits are considerably lower than the concentration level at which it is considered to cause any health impacts. There is a large number of pesticides

available for use across UK and Europe. In turn a large number of pesticides is found in surface waters at a wide range of concentrations (Tröger et al., 2021; Claver, 2006; Masia, 2013). Tröger et al. 2021 found the effectiveness of pesticide removal from drinking water varied from country to country and a number of pesticides were still present in drinking water after treatment.

Metaldehyde is a soluble molluscicide used in agriculture to control slugs and snails (Li et al., 2010). Its low sorption coefficient of active ingredient to organic carbon (KOC) value (34 and 240 L/kg) (Kay and Grayson, 2014) combined with its relatively long half-life in soil (3.17 and 223 days) allows for it to be readily leached into surface runoff during rainfall events. Metaldehyde is typically applied to winter crops such as winter wheat, potatoes and oilseed rape, between September and December, when the conditions are most favorable for slugs (Asfaw, 2018).

Metaldehyde has previously been identified to be responsible for majority of all cases of pesticide exceedances in drinking water in England and Wales. In 2016 it accounted for 87% of all pesticide exceedances recorded that year (DWI, 2017). Severn Trent Water has reported exceedances at 17% of water treatment works (WTW) in 2017 and at 8% of WTWs in 2018 (Cooke, 2020). Although the sale and use of metaldehyde has been banned in the UK since 2022, the existing metaldehyde model (Asfaw et al., 2018) can be utilized to investigate the behavior of other pesticides under different application scenarios and weather conditions at catchment scale.

Propyzamide has also been identified to be of concern to the water industry (Cooke 2020) with implementation of some of the catchment management approaches for propyzamide reduction in surface waters proving challenging (Stoate et al., 2017). It is a herbicide used to control a range of weeds in oilseed rape, field beans and other crops (Corteva, 2023). The sorption coefficient of active ingredient to organic carbon is higher than that of metaldehyde at KOC of 840 with shorter half-life of active ingredient in soil at 47 days (ADAMA, 2015). However, the application rate of propyzamide is over twice the amount at 0.425 g per 5 square meter based on typical application of 1.7 litre/ha application using 500 g/litre (43.86% w/w) propyzamide suspension concentrate (Corteva, 2023) compared to 0.19 g per 5 m² of metaldehyde (Asfaw et al., 2018).

2.3 Modelling of diffuse bacterial pollution and SSOs

Bacterial surface water pollution can arise from a variety of point and nonpoint sources. It is the diffuse, nonpoint sources that are especially challenging, as they are difficult to pinpoint and hence mitigate (Hubbart et al., 2022). Examples of diffuse pollutant sources include

intensely grazed grasslands where bacteria from livestock will accumulate and be transported to watercourses through rainfall-runoff processes. The importance of characterizing water quality in rivers during baseflow conditions have been highlighted and well studied (Ferguson and Kay, 2012; Cho et al., 2016). However, acute pollution incidents – short-term ‘spikes’ in pollution directly following rainfall are of specific concern to water utilities due to the pressures it brings to water treatment facilities if this water is abstracted for drinking water supply. High resolution hourly/sub-hourly water quality datasets needed to accurately characterise acute pollution incidents are scarce.

2.3.1 Current bacterial models

Cho et al (2016) has provided a comprehensive review of watershed-scale modeling of microbial water quality of surface waters to date. Key processes are summarized below. Figure 2.1 presents a diagram from the same paper demonstrating summary of pathways involved in the modelling of bacteria transport into river systems.

Most current the models lack high resolution input data and forecasting that is necessary for successful identification of acute pollution events and therefore are of limited use to utilization in abstraction management systems. Although a model proposed by Wilkinson et al. (2011) did use short time steps (1/4 to 1 hour), it simplified catchment delivery processes, loads and the accumulation and flushing of land-based faecal reservoirs. The simplified approach is suitable for modelling microbial pollution of large catchments. However, accounting for spatial rainfall and pollution source distribution is desirable in smaller catchments where it has greater impact on resulting water pollution.

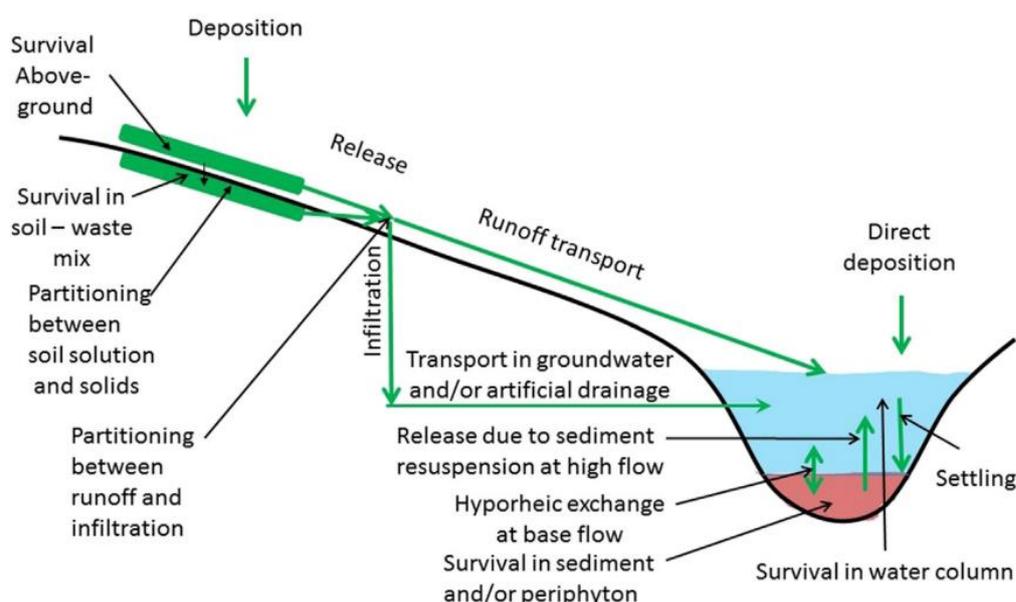


Figure 2.1 Pathway of faecal microorganisms including surface runoff from soil to survival in surface water bodies; green lines represent the pathway of faecal microorganism release and

transport and black lines point to specific fate-related process that need to be modelled from (Cho et al., 2016)

Both bacteria survival in manure and animal waste, and survival in soil are commonly modelled with the use of the first order kinetics equation known as Chick's law to find the rate of overall bacteria die-off (Sadeghi and Arnold, 2002; Collins and Rutherford, 2004; Dorner et al., 2006; Ferguson et al., 2007; Walker and Stedinger, 1999; Whelan et al., 2014; Haydon and Deletic, 2006; Wilkinson et al., 2011; Schijven et al., 2015; Sterk et al., 2016). Alternately some models (Ferguson et al., 2007; Brannan et al., 2002) use daily removal rate for modelling bacterial survival in soil. Only a third of the reviewed models in Cho et al. (2016) chose to simulate transport in the soil column. In models that did account for this process it was estimated as a loss from the soil mixing layer proportional to bacteria concentration and infiltration with exception of WHATFLOOD (Dorner et al., 2006) which included it as mass balance in unsaturated zone. Release from animal waste and manure reservoirs is summarised in figure 2.2. Release from the soil reservoir is mostly modelled as fraction of microbial reservoirs proportional to bacteria number and runoff depth in soil solution or bound to sediment.

All microbial release and delivery to streams via overland transport is mostly assumed to happen on the same time step. As for most models operating time step is daily or more, transport is assumed to be faster than this time step. Some models apply partitioning for part of pollutants to be delivered on the day of release with the rest stored for future delivery. WAM model (Collins and Rutherford, 2004) has taken the most elaborate approach where it is determined by slope, proximity to stream and flow accumulation. Fate within the in stream water column is determined by the first-order kinetic with convective in-stream microbial transport in majority of the models. Only a few models account for microbial fate in sediment, exchange between sediment and water column or transport via groundwater.

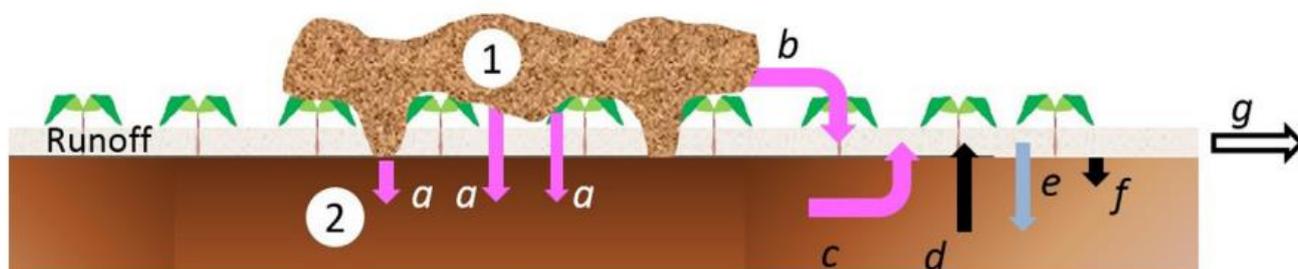


Figure 2.2 Pathways of microorganism release and removal; ① and ② - above ground and belowground (soil) reservoir, respectively.; a e release in suspension to soil, b e release in suspension to runoff, c e release in suspension from soil to runoff, d e release absorbed microorganisms with soil particles to runoff, e e infiltration with runoff suspension, f e settling with sediment particles, g - overall removal (or export) from the application site from (Cho et al., 2016).

Sylvestre et al. (2021) fitted parametric distribution to routine monitoring data to predict daily mean protozoan pathogen concentrations following hydrometeorological events. Cyterski et al. (2022) utilised least-angle regression modelling using a number of environmental covariates (such as rainfall, turbidity, pH, etc.) to forecast bacteria. The mobile-immobile model proposed by Drummond et al. (2022) incorporates transport, immobilisation, that increases during baseflow, and vastly increased remobilisation of bacteria and fine particles during stormflow.

The existing models for bacterial pollution forecasting operate on either long time steps or coarse spatial resolution on large catchments. They lack the fine temporal and spatial detail acute diffuse pollution events in smaller catchments need. For forecasting arrival times of acute impacts at abstraction points, a bacterial model working on a fine spatial and temporal (hourly/sub-hourly) scale is needed that accounts for spatial rainfall and land use variability.

2.3.2 Bacterial loadings from agricultural sources

Agricultural bacterial loadings can stem from either direct deposition by livestock during grazing period or slurry/manure spreading of the stores accumulated when livestock is removed from the fields. The bacterial loading through grazing is quantified by the number and type of livestock present, the associated livestock specific faecal excretion rates and *E. coli* shedding potential (Coffey et al., 2016).

Oliver et al. (2017) have developed and evaluated (Oliver et al., 2018) a catchment-scale model of *E. coli* burden to map spatial patterns of *E. coli* accumulation on land based on a cross-disciplinary toolkit for assessing farm scale contributions to *E. coli* risk (Olliver et al., 2009). The empirical model reported in Oliver et al. (2009), estimates *E. coli* loadings on fields based on 'a worst case scenario' which represented a realistic upper level of stocking densities in the UK.

2.3.3 Bacterial loadings from urban sources

Discharges from urban wastewater sources represent a potentially significant source of pollutants to surface waters (Shepherd et al., 2023). Much effort has been put into modelling of impacts from SSOs through major research and development programme that resulted in publication of Urban Pollution Manual (UPM) by Foundation for Water Research (FWR,2019). The manual and the associated standards have been adopted by the regulators as guidance and best practice for planning the improvements of SSOs to reduce impacts on water quality.

Water utilities can be manage compliance risks by investing in new infrastructure and/or applying new system management strategies. These decisions are based on the assessment of proposed and alternative schemes tested using hydrodynamic network models to simulate

the hydraulic performance (Delelegn et al., 2011). However, hydrodynamic network models can contain considerable uncertainty (Vezzaro et al., 2013; Thorndahl and Willems 2008). Srivastava et al. (2018) describes an objective uncertainty quantification process in simulation of sewer overflow volume to enable water utilities to evaluate and report the uncertainty in their modelling predictions.

2.4 Modelling of pesticide impacts

An array of models have been developed to forecast pesticide impacts on surface waters and inform water quality management decisions. These water quality models are capable of operating at different spatial and time scales. A model running at a fine spatial and temporal resolution is required for accurate prediction of the arrival of short-term peak pesticide concentrations at catchment outlets following rainfall events to inform water abstraction decisions. MACRO (Larsbo and Jarvis, 2003) is an example of a detailed water quality model, however its large data requirements can be computationally intensive. It is therefore typically applied at small spatial scale but run at larger time-steps. Large spatial scale models such as the FOCUS (2000) and FOCUS (2001) are not suitable to operate at the finer catchment scale as FOCUS does not mimic specific fields and should be viewed as representing major agricultural areas. Catchment scale water quality models include the Soil and Water Assessment Tool (SWAT) (Neitsch et al., 2002; Abbasi et al, 2019; Dogan and Karpuzcu, 2023; Cambienet et al., 2020; Zhang et al., 2020; Janney & Jenkins, 2022), Hydrological Simulation Program Fortran (HSPF) (Donigian et al., 1995), AnnAGNPS (Bingner & Theurer 2005), Integrated Model for Pesticide Transport (IMPT) (Pullan et al., 2016), diffusion–advection process models (Cardenas et al., 2023; Márquez-Romance et al., 2022), and accumulation–utilization model (Cárdenas-Izaguirre, 2022). However, most of these models predict long-term trends of diffuse pollution and thus forecast concentrations at large time scales. While SWAT provides an option to run with small time-steps it would require large amounts of data and a large number of parameters would need to be calibrated (Benaman et al. 2005). The Dynamic Watershed Simulation Model (DWSM) (Borah et al., 2002), the Agricultural Nonpoint-Source Pollution Model (AGNPS) (Young et al., 1989) and MIKE SHE (Refshaard et al. 1995) are capable of predicting pollutant loadings and transport from single rainfall events run at small time-steps are interest for informing short-term water quality management.

Asfaw et al. (2018) has presented a new validated, travel time based, physically distributed model used to predict metaldehyde levels after a rainfall event accounting for variations in rainfall and distribution of land use. It is comprised of surface runoff generation, surface runoff routing and pollutant build-up/wash-off components. Many of the processes could be simplified

because metaldehyde is quite stable compound and high resolution calibration data was available.

The surface runoff component calculates the cumulative excess rainfall depth I^t (mm) at each timestep t based on the differential form of the Soil Conservation Service (SCS) curve number (CN) method (Mancini and Rosso, 1989). Initial CN values are determined based on hydrologic soil group (HSG), land use and hydrologic conditions data (Mishra and Singh, 1999). The effect of soil moisture on runoff generation is incorporated by adjusting CN values based on antecedent moisture condition (AMC) categories. The surface runoff routing component uses a spatially distributed time variant direct runoff travel time technique to account for spatial and temporal variability of runoff generation and flow routing through overland flows and stream networks (Melesse and Graham, 2004; Du et al., 2009). The pollutant build-up/wash-off component estimates metaldehyde build-up through pesticide applications on identified high risk areas. Metaldehyde wash-off was based on the “simplified formula for indirect loadings caused by runoff” (SFIL) (Berenzen et al. 2005; Reus et al. 1999) used to calculate percentage loss from high risk areas at each timestep.

The model operates at 1 h time step, input spatial rainfall data at 1 km² and calculates runoff at 5m² resolution. Model validation (Figure 2.3) has returned an average coefficient of determination of 0.75 and model efficiency of 0.46. It is currently used in drinking water abstraction management at STW to suspend abstraction from surface water when a peak in metaldehyde concentrations is forecasted. It would be useful to develop an approach where

this model is used to facilitate identification of priority catchment areas to inform catchment management.

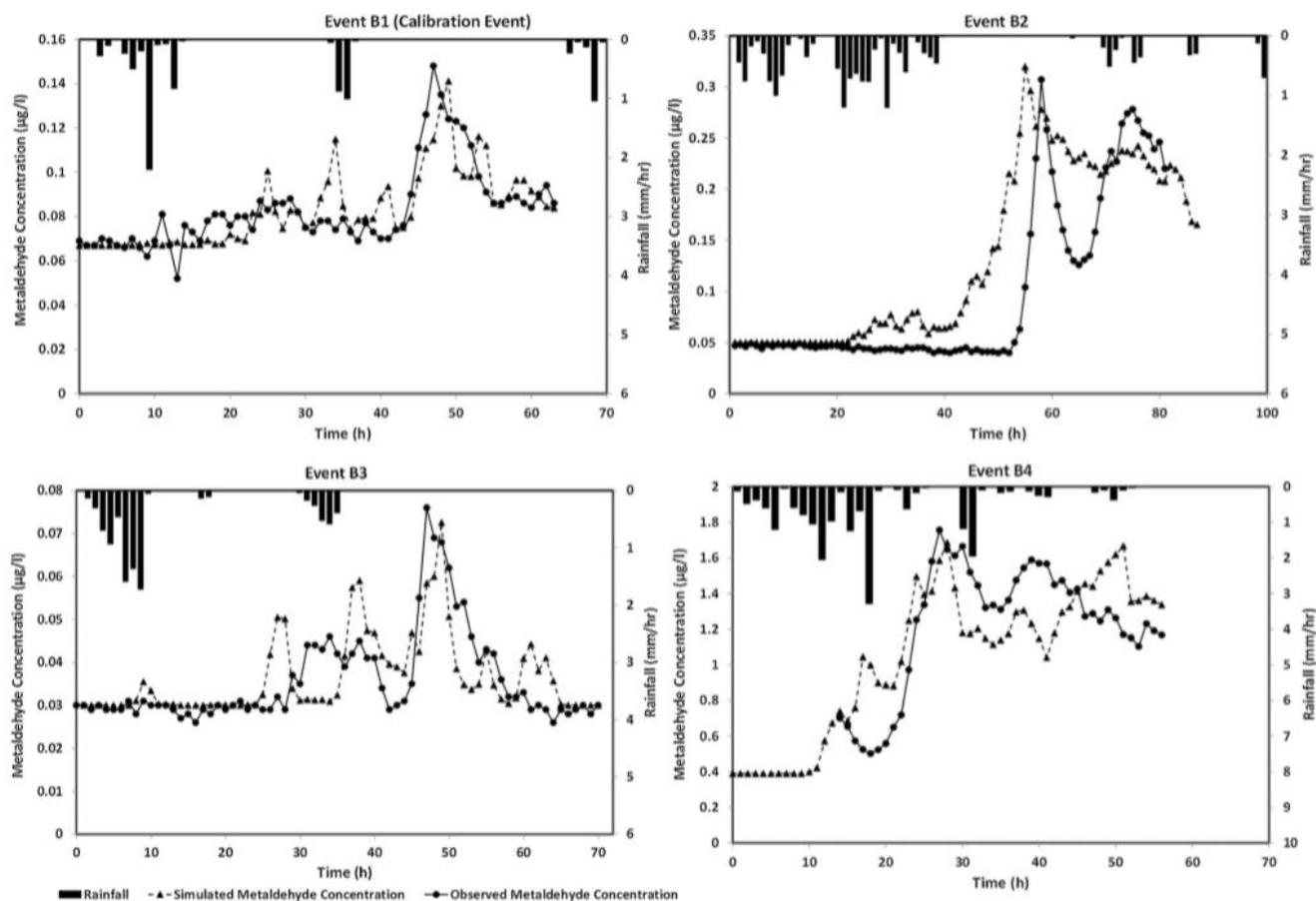


Figure 2.3 Calibration (B1) and validation (B2-B4) events for the Metaldehyde model (Asfaw et al., 2018)

2.5 Land use optimization

Land use pattern optimization is a useful method to inform catchment management for non-point pollution. It involves simulation of different land use scenarios and evaluation of the resulting pollutant loads. Such that catchment management resources can be targeted to specific areas which are most effective in reducing pollutant impacts.

A number of optimization methods have been created for agricultural best management practices (BMPs) (Liu et al., 2019; Geng and Sharpley, 2019; Aslani et al., 2023; Yousefi and Moridi, 2022). Nie et al. (2019) and Li et al. (2020) concentrated on optimizing land use in terms of tradeoffs between crop productivity and resources used (such as water and energy). Other optimization methodologies looked into addressing non-point pollution from specific contaminants such as nitrogen and phosphate (Milne et al., 2020; Ouyang et al., 2020; Zhang et al., 2021). Zhang et al. (2011) carried out land use optimization by coupling the CLUE-S (the Conversion of Land Use and its Effect at Small regional extent) and SWAT(Soil and Water

Assessment Tool) models to produce and evaluate two land use scenarios. A 'business as usual' land use case and a changed forecast based on agricultural non-point source pollution control. SWAT model was calibrated and validated at daily time step for water quality, thus forecasting long-term trends and missing out on short-term events that can peak within hours but result in high levels of pollution. While the pollution control scenario has resulted in significant decrease in pollutant loads, the method does not conduct a search for the areas that pose highest risk and should be concentrated on when designing the mitigation measures. Zhu et al. (2020) also set two scenarios, which correspond to the adjustments of transforming the farmland with the distances of 50 and 100 m from water bodies, respectively, to forest land to analyse the changes of risk levels. Aerts et al. (2002), Sadeghi et al. (2009), Li et al. (2019) and Cui et al. (2019) used linear programming for land use allocation optimization. While linear programming guarantees an optimal solution it comes with a high computational time costs and is therefore less suitable for large search spaces. Srivastava et al. (2002) used an optimization algorithm for the selection of best management practices (BMP) on a field-by-field basis for the entire catchment. The AnnAGNPS model which simulates long-term runoff, sediment, nutrient, and pesticide transport from agricultural catchments is coupled with genetic algorithm to find a number of near optimal solutions that maximize pollutant reduction while minimizing implementation costs/maximizing net return. However, the long-term simulation nature of the AnnAGNPS model does not consider the acute impacts driven by individual rainfall events. Pollutant reduction for the use in objective function was quantified as pollutant load of BMP scheme normalized with respect to the load for the baseline case where the percent changes in various pollutants were treated equally. This approach however, does not indicate to the objective function which individual pollutant is reduced or how useful the reduction of that pollutant is in relation to water quality guidelines.

Srivastava et al. (2002) used a genetic algorithm (GA) with a continuous simulation, watershed-scale, Annualized Agricultural Non-Point Source Pollution model (AnnAGNPS) to optimize the selection of best management practices (BMPs) (e.g. corn-corn-soybeans-alfalfa crop rotation, permanent alfalfa, etc.) on a field-by-field basis for an entire watershed. The goal was to minimize long-term water quality degradation and maximise net farm return on an annual basis. Arabi et al., 2006 also used genetic algorithm in combination with a watershed model (SWAT) to optimize the selection of best management practices on a field-by-field basis. The BMPs in this case are field border, parallel terrace, grassed waterway and grade stabilization structure and the optimization goal was to select which BMPs should be allocated where to minimize sediment and nutrient loads, and implementation costs.

A land use optimization approach could be adapted to integrate pollutant models and genetic algorithm to identify high risk areas within the catchment under current land use and investigate how removal of these areas (classing them as a non-contributing area) affect pollutant loads at the abstraction site. This would allow better targeting of available resources. A land use optimization approach could be adapted to integrate pollutant models and genetic algorithm to identify high risk areas within the catchment under current land use and investigate how removal of these areas (classing them as a non-contributing area) affect pollutant loads at the abstraction site. This would allow better targeting of available resources. Genetic Algorithm is a search algorithm based on natural selection. It works with parameter sets of a model while checking the outcome of the model as its objective function. The parameter values that produce the most optimal model outcome are then selected to produce next set of parameters ('offspring') through crossover and mutation. It is therefore especially suited as an inverse modelling method. Because it searches from populations rather than a single point, it can provide more than one solution. In practical applications, there is a need to have a number of near optimal solutions as alternatives because not all solutions can be implemented due to practical reasons, specific to the catchment in question. Therefore, GA is especially suited to mitigation measure allocation searches (Srivastava et al., 2003; Srivastava et al., 2002; Arabi et al., 2006; Perez-Pedini et al., 2005). A summary of Genetic Algorithms and their applications is detailed in Tang et al. (1996). All the solutions are saved for later reference to enable the analysis of near optimal solutions and thus be of more use for the catchment management as a tool.

2.6 Summary of Knowledge Gaps

Based on the literature review, key knowledge gaps and associated research objectives can be summarized as follows.

1. There is a general paucity of high-resolution water quality datasets capturing acute impacts from diffuse or intermittent rainfall driven events in surface waters (section 2.2). This is especially relevant for water quality parameters which are difficult to characterize in 'real time' such as FIOs, pesticides and emerging contaminants. This limits understanding of key transport mechanisms and water quality processes relevant to applications relevant to the short term management of surface water abstraction systems. The work in this thesis will focus on the most common FIO used in the UK water industry (*E. coli*) as well as two specific pesticides. Further, whilst sensing technologies are continuously under development, there is a lack of robust testing of new sensors in live drinking water catchments.

Objective 1: Sample river water at high-resolution to obtain 4 sampling events that encapsulate bacterial pollution following a rainfall event. Use the data obtained to calibrate and test a water quality probe.

2. Existing surface water quality models are most commonly focused on the characterization of long term, or representative conditions in surface waters (sections 2.3 and 2.4). Commonly operating at daily temporal scales or above, they frequently neglect or heavily simplify explicit representation of time varying SSO impacts of diffuse runoff. This focus is linked to the paucity of high resolution datasets for model calibration and validation. Current modelling tools to characterize impacts of SSOs at high resolution are based on detailed hydrodynamic modeling of the urban drainage network. Such models are time consuming to run, have high data requirements as well as high levels of predictive uncertainty. This limits the application of current modelling tools to the direct management of water abstraction systems, which cannot be normally closed for longer than a few hours in periods of water stress.

Objective 2: Develop a novel, practically applicable process-based forecasting approach to characterize short term bacterial dynamics in catchment scale river networks, considering both inputs from SSO discharges and diffuse agricultural runoff at an hourly resolution.

3. Land use optimization techniques have previously been developed to target catchment interventions to reduce pollutant impacts (section 2.5). However similarly to above, most approaches are based on the optimization of representative, long term water quality parameters. The effectiveness of these approaches to short term, acute impacts characterized by spatially and temporally variable sources has not yet been considered in detail.

Objective 3: Build a land-use optimisation framework based on short term acute impacts to aid catchment mitigation of rainfall driven non-point pollution impacts.

Chapter 3 — On site evaluation of real-time sensing technologies for characterisation of acute bacterial loads in river systems

Chapter 3 focuses on the testing of a commercial real-time *E. coli* sensor in a UK catchment for the purpose of informing drinking water abstraction management. Further, the chapter also describes the case study catchment, fieldwork and water quality datasets obtained for the subsequent chapters in this thesis. Methods of data collection to achieve the aims as stated in the introduction chapter are outlined. Results are presented and discussed, and the chapter ends with a conclusion section describing future recommendations.

3.1 Introduction

Bacterial water quality is of concern where surface water is abstracted for drinking water supply and/or where the water body is used as a bathing site by the general public (National Audit Office, 2010; Collier et al., 2021). When the drinking water is abstracted, it needs to undergo an extensive cleansing process before it is supplied to the public. Acute pollution events are of concern due to the possibility of treatment not being sufficient and contamination passing through to the end user. Bathing water quality is important from both the aesthetic and health points of view as swimming in heavily polluted rivers poses health risks (WHO, 2012). Therefore, it is desirable to monitor the levels of bacterial pollutants in surface waters to comply with regulations and manage risk associated with acute impacts.

E. coli is a commonly used faecal indicator organism (¹WHO, 2017) for bacterial water quality, to identify the levels of faecal contamination of surface water and thus infection risk to humans when assessing water quality. The indicator organisms are used for regulatory purposes to comply with the needs of monitoring drinking and bathing water quality. Culture based bacterial analysis of river water is labour and resources intensive, with a delay of at least 24 hours before the results can be retrieved. Total coliforms and *E. coli* Isolation and Enumeration from Water by Membrane Filtration method is industry standard and is carried out as stated in The Standing Committee of Analysts (2016) based on Sartory and Howard (1992). As such, these culture-based approaches cannot provide real-time warning of adverse water quality impacts. They are also labour and time intensive, which makes it logistically difficult to understand short-term temporal water quality dynamics using these techniques.

Sensing technology has advanced such that monitoring systems are commercially available to quantify many water quality parameters (ammonia, turbidity, optical brighteners, Biochemical Oxygen Demand (BOD), Coloured Dissolved Organic Matter (CDOM), etc). However, the monitoring of bacterial pollution in real time is not routinely undertaken as the technology is

unproven in commercial applications. Attempts to use more easily measurable surrogate parameters (such as turbidity) to inform bacterial water quality had limited success (Vincent et al., 2022; Herrig et al 2019). High frequency in-situ monitoring of bacterial water quality has the potential to facilitate the understanding of acute short-term rainfall driven bacterial peaks in the river water and enable a more dynamic control of water resource assets (Yassin et al. 2021). Deployment of a real-time water quality sensor could allow continuous observation of water quality over a full range of antecedent conditions and diurnal as well as seasonal variations. Potential real-time bacterial water quality monitoring technologies can be divided into 3 groups: Indirect (e.g. Adenosine triphosphate assays, Tryptophan-Like Fluorescence), direct (e.g. Optical Imaging, Flow Cytometry) and molecular(e.g. DNA, Immuno- Recognition). The DNA and Immuno- Recognition within the molecular monitoring methods group are not suitable for the continuous real-time monitoring use due to the technology being not yet developed for continuous high frequency monitoring (Boxall et al, 2021). While the Optical Imaging and Flow Cytometry in direct monitoring methods group are unsuitable due to the nature of raw river water receiving high particle loads as it can clog fluidics (Safford and Bischel, 2019).

Indirect monitoring technologies such as Adenosine triphosphate (ATP) assays, Enzyme based assays and Tryptophan-Like Fluorescence (TLF) measure substances associated with bacteria to quantify the relative number of bacteria in question are potentially more suitable due to lower costs and higher practicality. The basis of ATP assay technology existed since 1987 (Stanfield and Jago, 1987), an enzyme uses ATP to fuel a chemical reaction and produce light where the amount of light produced is proportional to the amount of ATP in a sample. Laboratory-based trials have been carried out for applications of ATP quantification for bacterial quality monitoring in drinking water (Standfield & Jago, 1987; Delahaye et al., 2003). A variety of commercially available kits in drinking water samples, have been used to compare total plate counts and ATP assays, reporting low correlation of the average ATP measurements to cell counts (Hammes, 2010, van der Kooij 2017; de Vera & Wert 2019). Enzyme assays detects enzymes expressed by a microbe in the environment, producing a signal that is representative of a population. The method is based on the addition of a tag to the substrate that, when cleaved of its tag will release fluorescent, luminescent, or spectrophotometric endpoint signal and be detected quantitatively. These devices however, claim a range of management needs, including maintenance and regular refills for reagents depending on frequency of sampling and disposal of waste. The detection takes from minutes to hours with sensitivity from 1 organism per sample volume (Stadler et al., 2016). Testing of the real-time monitors in surface waters (Stadler et al., 2017) has shown significant differences between

culture and enzymatic assays. It was also suggested that the source type and age of contamination affect the underlying correlation.

Tryptophan is an amino acid present in organisms that, when excited with a wavelength of ~280 nm, fluoresces at ~340nm (Reynolds, 2003; Hudson et al., 2008). The detection of tryptophan-like fluorescence has been shown to be an indication of an active community degrading organic matter, similar to BOD measurements, and so analysis of drinking water sources using fluorimeters at these wavelengths shows that a higher TLF is reported when contaminated with faecal coliforms or farm waste (Cammack et al., 2004; Hudson et al., 2008; Sorensen et al., 2018; Henderson, et al., 2009). Tryptophan-Like Fluorescence (TLF) sensors are relatively cheap, require no reagents (and hence produce no waste) and involve relatively little maintenance/labour. Previous testing of TLF sensors has been conducted in a limited range of environments. Khamis et al. (2015), Khamis et al. (2016) and Khamis et al. (2021) considered performance of a TLF sensor against lab analysed samples of Tryptophan-like fluorescence, Dissolved Organic Matter DOM and Biochemical Oxygen demand BOD respectively. Baker et al. (2015), Mendoza et al. (2020) and Fox et al. (2022) investigated other tryptophan-like fluorescence sensors to evaluate the performance on measuring bacterial pollution. Mendoza et al. (2020) looked sensor performance measuring untreated wastewater added to natural creek water in a laboratory sewage spill simulation and in a natural setting during a single storm event. The sensor was deployed in a small river catchment, representative of urban Mediterranean systems, with no WWTW inputs but 11 different sanitary sewer overflow events in the two years prior to the study period. TLF and *E. coli* levels were found to be significantly correlated for both laboratory simulated sewage spill simulations and the deployment in the river during a single storm event. Fox et al., (2022) deployed sensor and collected single samples from a number of locations along an urban, highly polluted river. Strong significant correlations were found between TLF and bacterial enumeration. Baker et al. (2015) sampled a number of locations within two poor water quality river catchments and found that a TLF sensor performed better in less polluted conditions, accuracy falling with the increasing range and variability in *E. coli* and complexity of pollutant sources. They have highlighted the possibility of optical interferences, such as pH, thermal quenching, and scattering of emitted fluorescence in turbid samples, to have an effect on the *E. coli*-fluorescence intensity relationship. However, Baker et al. (2015) data has also suggested the optical interferences to likely be catchment specific, such as microbial activity that increases fluorescence intensity without the presence of *E. coli* and pollution sources unrelated to faecal pollution, which emit fluorescence in the region of interest.

During this study, a newly commercial TLF sensor Proteus Instruments Ltd (2021), supplied by R.S. instruments was acquired by the project sponsor with a view to potential use in real time water resource applications. To date, the performance of such sensors has not been independently verified in live drinking water catchment and on-site validation of this technique has been not previously been made available. This study provides a detailed testing of the Proteus sensors ability to derive *E. coli* levels from real-time measurements of TLF. The sensor is evaluated via direct comparison of *E. coli* concentrations evaluated using traditional, culture based methods based on samples from the river system collected both during dry weather as well as following a number of rainfall events over a range of seasonal conditions.

As initially recommended by the manufacturer, one calibration event across the entire measurement range of the conditions expected is supplied to calibrate the algorithms used to derive the *E. coli* data and another three rainfall events covering a range of flow and seasonal conditions to validate sensor performance under acute impacts. Current high resolution (sub-daily) *E. coli* datasets from surface water catchments are scarce, hence the collection of such samples during wet weather events is in itself a valuable resource.

3.1.1 Study area

The River Leam is a 300 km² sub-catchment of River Severn with elevation ranging from 46m to 232 m above sea level. River flow at the abstraction site is monitored using a EA flow gauging site with data available every 15 mins. The normal flow depth of the River Leam at the gauging station ranges between 0.24 m and 1.16 m with an average flow of 1.55 m³/s and mean annual catchment rainfall of 649 mm. A surface water abstraction site is maintained by the utility operator for potable water supply (figure 4.1, chapter 4), situated at the catchment outlet. Based on long term routine monitoring at the abstraction site, the utility operator has identified large short-term increases in bacterial pollution after rainfall events as a further water quality concern.

Severn Trent water conduct routine monitoring of *E. coli* levels in the raw water supply to the WTW. Historical data of *E. coli* obtained through the Severn Trent monitoring programme (figure 3.1) shows the pattern of short term *E. coli* peaks to be coinciding with large cold autumn\winter rainfall events. Large rainfall events may wash-off *E. coli* from the agricultural fields and result in SSO spills as the sewer network struggle to cope with addition of rainwater to the system. However, some warm weather rainfall events, which do not coincide with significant increases in flow do still display peaks in *E. coli* levels. A small and/or localised rainfall event coinciding with recent slurry spreading on agricultural fields close to the water body may result in wash-off of *E. coli* into the river while not resulting in significant increase in

river flow. A localised rainfall event over SSO catchment may also result in a SSO spill without largely affecting river flow. These patterns should be interpreted with caution due to a large variation in schedule and number of samples taken, ranging from 1 to 24 samples per month. Within the catchment there is a significant proportion of grazing land and a number of SSOs which may contribute to short term increases in *E. coli* following rainfall events due to diffuse runoff as well as SSO spill events. These sources will be discussed and explored further in chapter 4. The regular occurrence of these events provides evidence that this is a feasible site for the testing of a real time sensor for the characterisation of short-term acute impacts within surface waters.

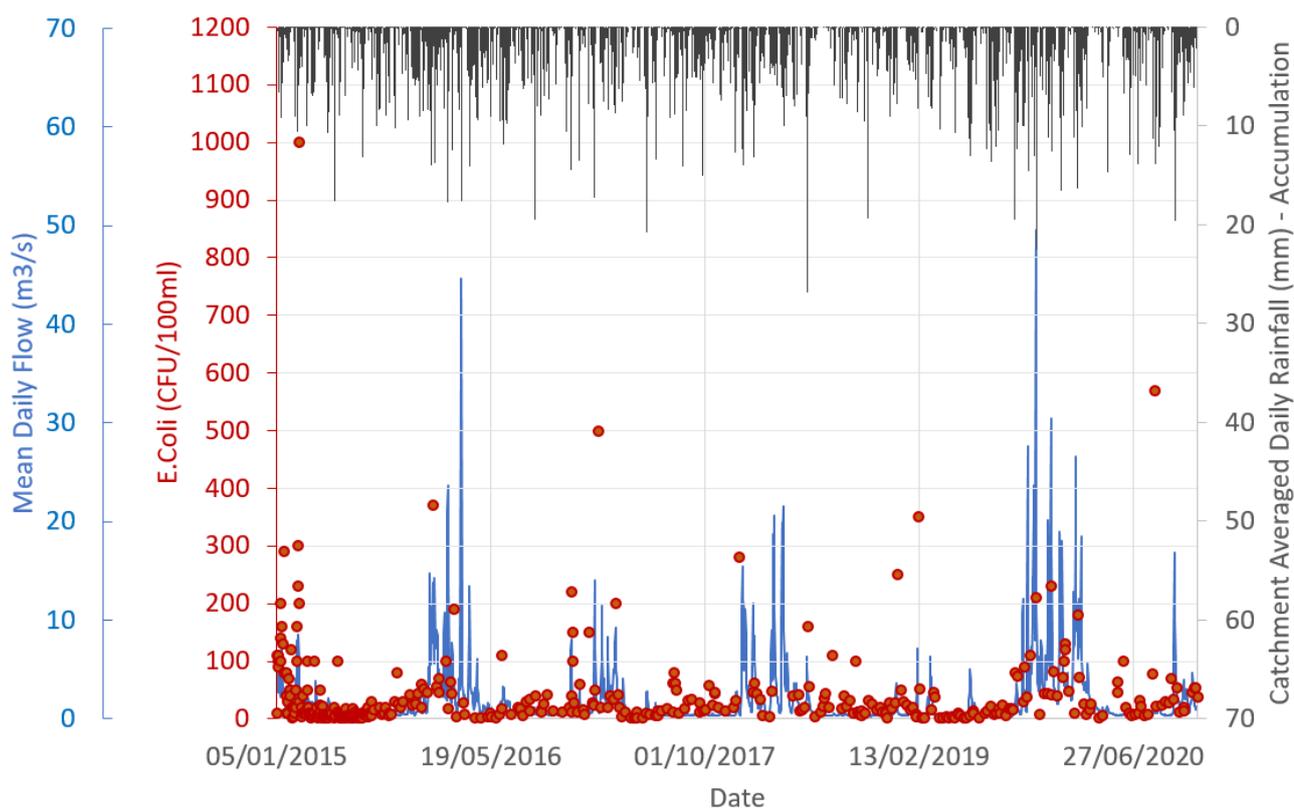


Figure 3.1 Catchment average rainfall and river flow together with historical *E. coli* data from a nearby WTW intake provided by Severn Trent Water.

3.2 Proteus Water quality probe

The Proteus is a multi-parameter water quality probe (figure 3.2) procured by the Severn Trent Water and deployed in the Leam catchment within the river system, directly next to the drinking water abstraction intake (figure 4.1, chapter 4). The probe was installed to evaluate the opportunity to provide real time warning of high bacterial loads /poor water quality to the WTW. The initial installation of the probe was carried out in April 2021.

The sensor utilises a fluorometer to measure tryptophan-like fluorescence in-situ to derive the Total Coliforms/*E. coli* levels, using a relationship/correlation derived through calibration. Other parameters include Turbidity and Temperature both of which are used for Tryptophan

compensation. Optical Brighteners are of interest for SSO spill identification as it is added to laundry detergent and can therefore be used to track wastewater spills from urbanized areas (Lockmiller et al., 2019). The manufacturers recommended calibration protocol for accurate and repeatable measurement of the derived parameters was followed to attain a site-specific calibration. An initial test was carried out in dry weather flow conditions, where the probe readings were compared to reference values obtained from repeated sampling at the abstraction site, analysed utilising standard methods with the STW laboratory (see section 4.3.4 for further details). Following this initial test, sampling during a rainfall runoff event was carried out at the same location as the Proteus sensor, to cover bacterial loading across the entire measurement range of the conditions expected. Manufacturers guidelines required a minimum of 15 paired samples to be analysed at an accredited laboratory of users choice. In this case, river water samples were analysed for presence of Total Coliforms and *E. coli* at a main Severn Trent microbiology laboratory (see section 4.3.4). An event is triggered by the presence of rainfall anywhere in the catchment, while aiming to start sampling a few hours before the onset of rainfall. This encapsulated taking samples every 2 hours from just before the start of rainfall and for the duration of the high-flow event following a rainfall event. Following calibration, 3 further events were collected for the purposes of sensor evaluation/testing. These were collected under a range of different rainfall events under different seasonal conditions. Table 3.1 provides the characteristics of each testing event.

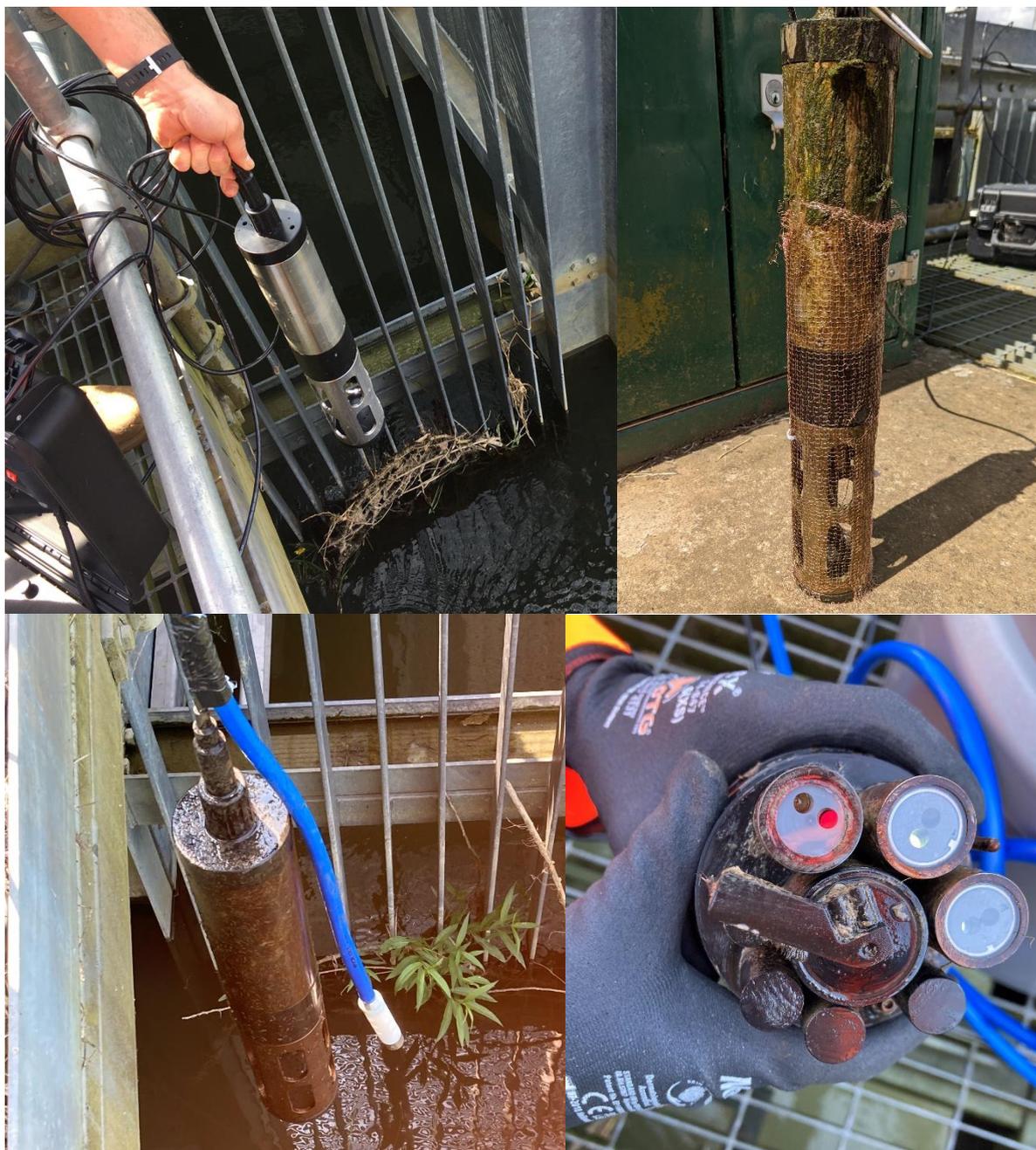


Figure 3.2 Top left - Installation of the sensor(April 2021); top right – installation of mesh to stop reed entanglement (June 2022); Bottom left – autosampler intake set up(July, 2021); bottom right sensors and wiper head (Jan 2022).

Live data from the probe is provided by telemetry and hosted online. This project had access to the post-processed data. Specific calibration relationships are currently commercially sensitive and are the property of RS instruments. Hence all calibration and quality assurance was undertaken by RS instruments staff, with limited input by the project team.

3.3 Test Events

During this project an additional monitoring campaign was conducted to characterise acute bacterial (*E. coli*) loading during rainfall runoff events, as well as one baseline testing to characterise levels during dry weather flow. These events coincided with the deployment of the

Proteus probe. The data was therefore used to both test the performance of the probe in characterising acute bacterial peaks, and for further model development work undertaken in this thesis (see chapter 4).

The first sampling event was made up of hourly samples over a 24 hour period of low flow, warm weather conditions (E0). This was used as baseline testing to identify the bacterial levels to be expected when acute impacts should not be present. It was followed by collection of one high flow/rainfall event for site-specific calibration (E1) with 3 further high flow events to validate the calibrated sensor data (table 3.1). The wet weather events were timed to coincide with incidents of high rainfall over the catchment, as anticipated using STWs in house rainfall forecasting system (WeatherQuest, 2023). Wet weather event samples were collected every 1-2 hours for 3-5 days to encapsulate the event from start of rainfall to when the flow hydrograph has approximately returned to low flow conditions.

Table 3.1 *E. coli* sampling event dates, durations and sampling frequencies. Rainfall duration is based on analysis of radar rainfall data from the UK MET office (see chapter 4 for further details)

Event No.	Start date	Sampling duration (h)	Sampling frequency (h)	Rainfall event duration (h)	River Flow (m ³ /s)	
					Max	Min
E0	19.09.2021	24h	1	0	0.240	0.226
E1	03.12.2021	82h	2	8	5.850	0.369
E2	05.02.2022	114h	2	32	2.530	0.575
E3	16.08.2022	46h	2	9	0.389	0.225
E4	12.03.2023	68h	1	41	31.800	0.272

3.4 Sample collection

An autosampler was used to collect all water samples for subsequent culture analysis and its intake was set up as close to the location of the sensors as possible without obstructing the flow of water to and from the sensors (figure 3.2). As the autosampler bottles were made of plastic and could not be autoclaved for sterilisation, it was deemed that thorough rinsing with tap water should be sufficient in the light of the initial low flow sample collected in a sterile bottle having high numbers of *E. coli*. The autosampler automatically purged the internal piping before each sample was taken. Ice was used to keep samples cool while in the autosampler and a refrigerated cool box with ice was used during the 1 hour, twice a day, transportation of samples for analysis to the STW laboratory at Church Wilne (figure 3.3).

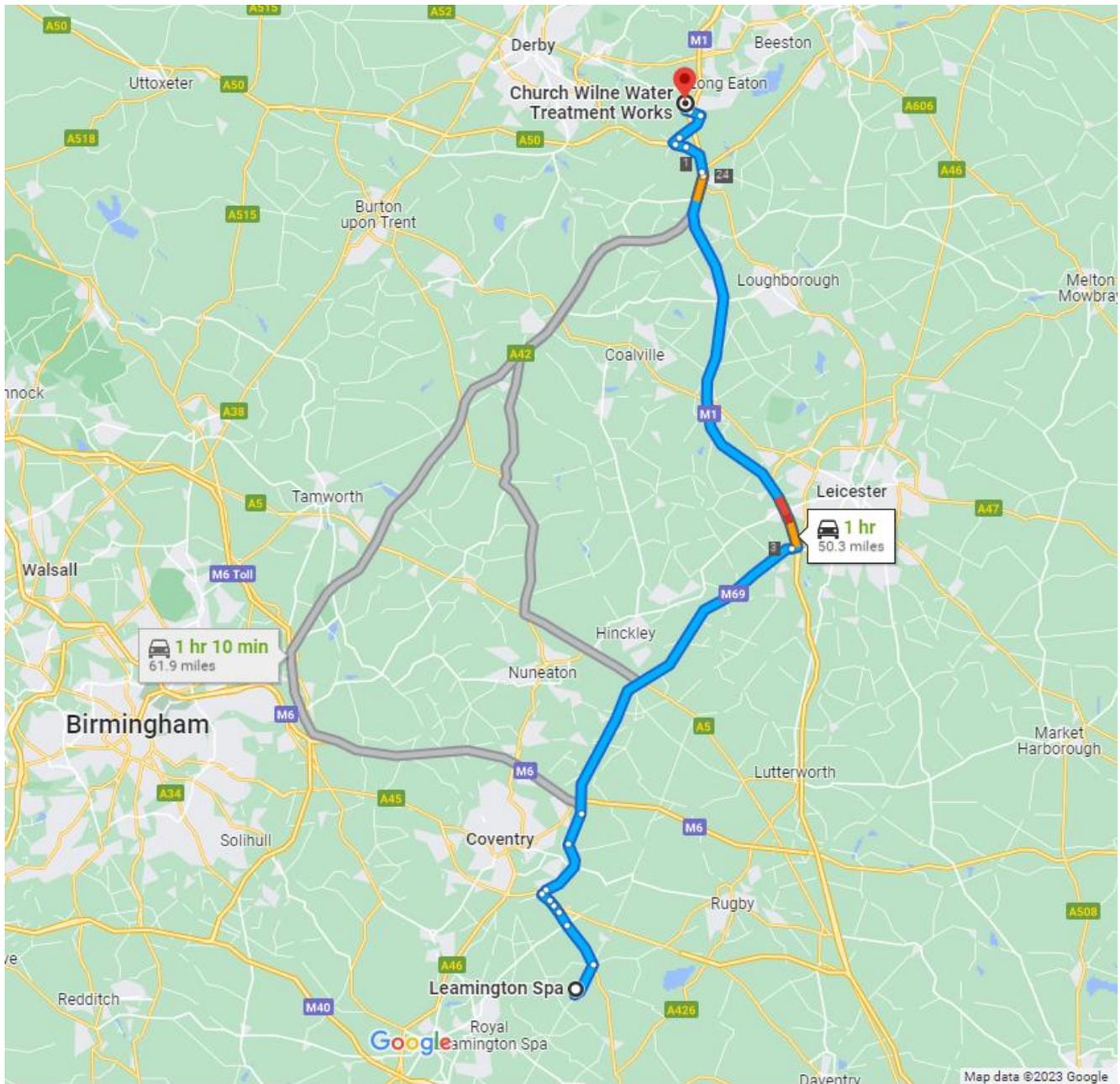


Figure 3.3 Route and duration of transporting the water samples to the lab (image from Google maps, 2023)

3.5 Sample analysis

The samples were analysed using membrane filtration method as stated in The Standing Committee of Analysts (2016) based on Sartory and Howard (1992). Raw river water at the study site is very high concentrations of bacteria and particulates, as demonstrated by a grab sample analysed to assess the range of dilutions needed, prior to the main sampling event. Therefore, it was judged necessary to analyse sample volumes ranging from 10ml down to 0.01 ml. For serial dilutions of 0.1 and 0.01 ml a portion of the sample is added to a Maximum Recovery Solution (MRD) to prepare a serial 10 fold dilution series.

The water sample is then filtered through a sterile, white gridded 47 mm diameter cellulose acetate membrane filter, 0.45 µm pore size upon which bacteria are entrapped. The filter is

then placed on a Membrane Lactose Glucuronide Agar selective growth medium and incubated at $30\text{ }^{\circ}\text{C} \pm 1.0\text{ }^{\circ}\text{C}$ for 4 ± 0.25 hours followed by $37\text{ }^{\circ}\text{C} \pm 1.0\text{ }^{\circ}\text{C}$ for 17 ± 3 hours. After incubation is complete the colonies, which are characteristic of Coliforms, and *Escherichia coli* are counted. The total yellow blue and green colony count is the presumptive total coliform count, while the green colony count alone is the presumptive *E. coli* count.

3.6 Results and discussion

3.6.1 Calibration

Calibration result, as undertaken by RS instruments staff, for high flow event E1, supplied for sensor calibration can be seen in figure 3.4. The plot displays sensor data plotted against lab data with resulting R-Squared of 0.932 showing a good fit of calibrated sensor data to the lab data. The time series plots of lab analysed *E. coli* and calibrated sensor data for the calibration event and the subsequent validation events can be seen in section 3.5.2.

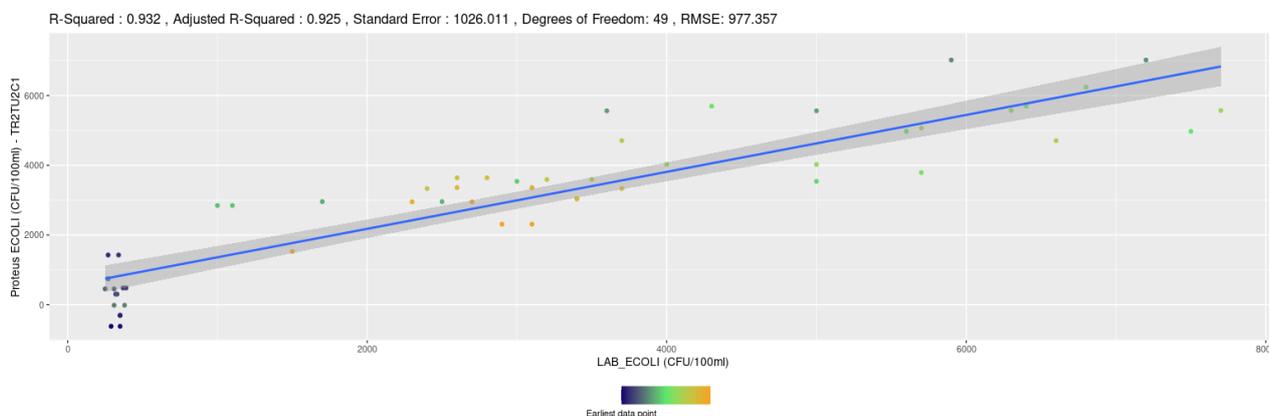


Figure 3.4 Calibration result plot for E1 as supplied by Proteus Instruments

3.6.1 Baseline testing

Event E0 (figure 3.5), was used to identify baseline bacterial levels to be expected, when acute impacts arising after a rainfall event should not be present. Hourly samples were taken mid-September 2021 for 24 hours.

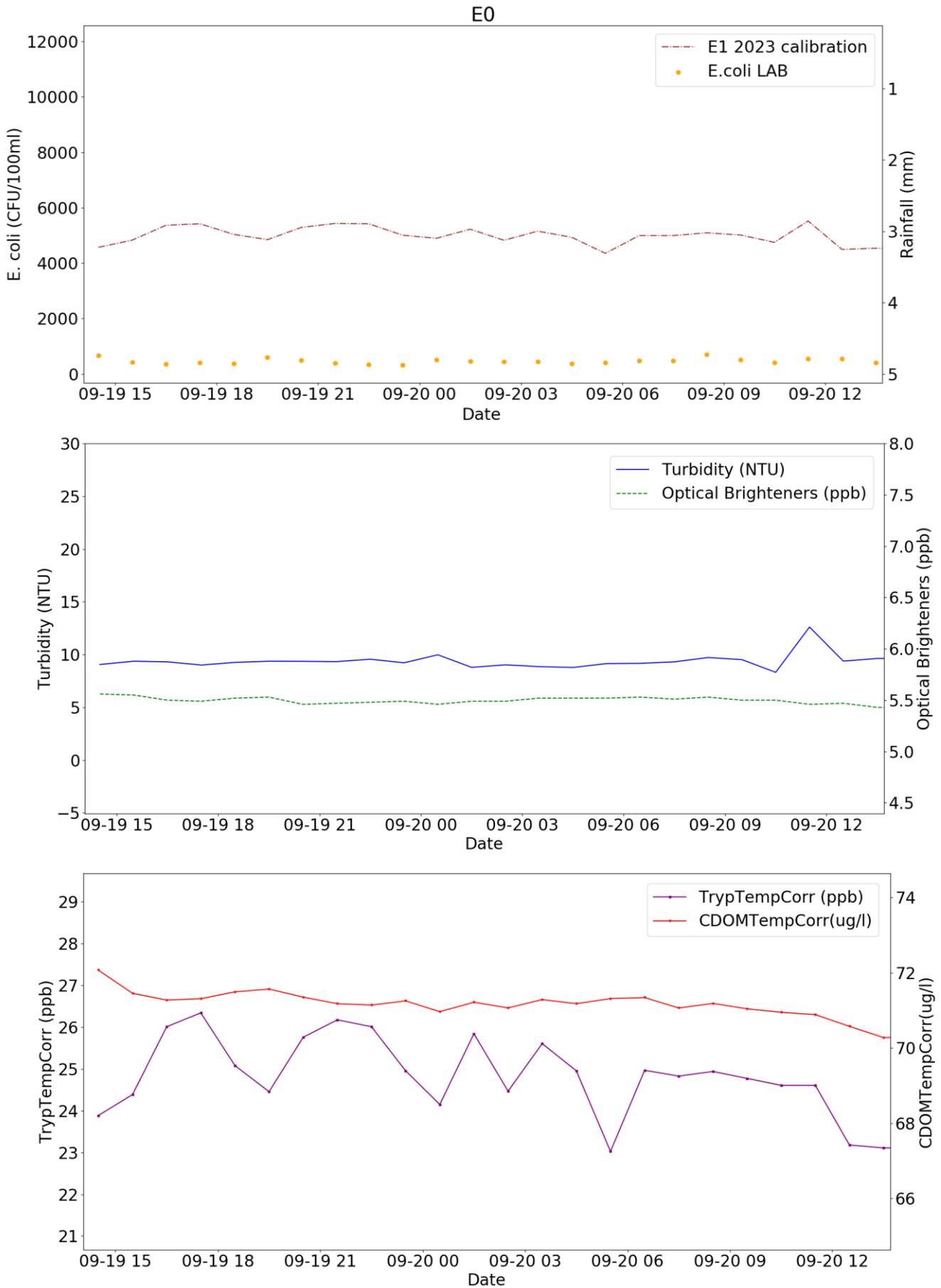


Figure 3.5 Low flow warm weather 24h event (E0). Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners(ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.

The retrospectively calibrated sensor data (figure 3.5) displays large offset during baseline conditions. This could be a result of the models used to derive the calibrated data being trained using an event where *E. coli* peak levels are acute (figure 3.6). Thus the calibration model might not be able to represent baseline conditions. As expected, there is no real trend due to lack of rainfall inputs into the river. The lab analysed *E. coli* falls between 330 and 700 CFU/100ml which places the water quality at baseline conditions between excellent and good classification for inland bathing waters.

3.6.2 Rainfall Runoff Events

Figure 3.6 shows the event (E1) provided to the supplier of the sensor to use for the calibration of the models used to derive *E. coli* data. Lab analysis data shows the event start with *E. coli* levels as low as 130 CFU/100ml peaking at 12000 CFU/100ml and 6800 CFU/100ml during the second peak. The calibrated data seem to mainly follow the trend displayed by the turbidity. As turbidity and *E. coli* peaks, the other parameters plummet, rising again as *E. coli* peaks for the second time. E1 and the subsequent high-flow/rainfall sampling events (figure 3.7 and 3.8) have displayed dual peaks suggesting possibility of more than one source of *E. coli* in this catchment.

Calibrated sensor data did not successfully identify the large magnitude of the peaks present in lab analysed results of the validation events (figures 3.7, 3.8 and 3.9). Only a slight elevation is visible in event E2 and E3 sensor *E. coli* data, thus pointing to the probe derived data to be more suitable for identifying the periods of high risk rather than being used to indicate the precise magnitude of the pollution event.

Event 2 lab analysed data record the event starting with lows of 2100 CFU/100ml peaking at 43000 CFU/100ml followed by the second peak of 17000 CFU/100ml. Turbidity for event 2 follows the trend of elevated *E. coli* levels fairly well. Tryptophan has fallen throughout the event, rising back up again as the event nears its end.

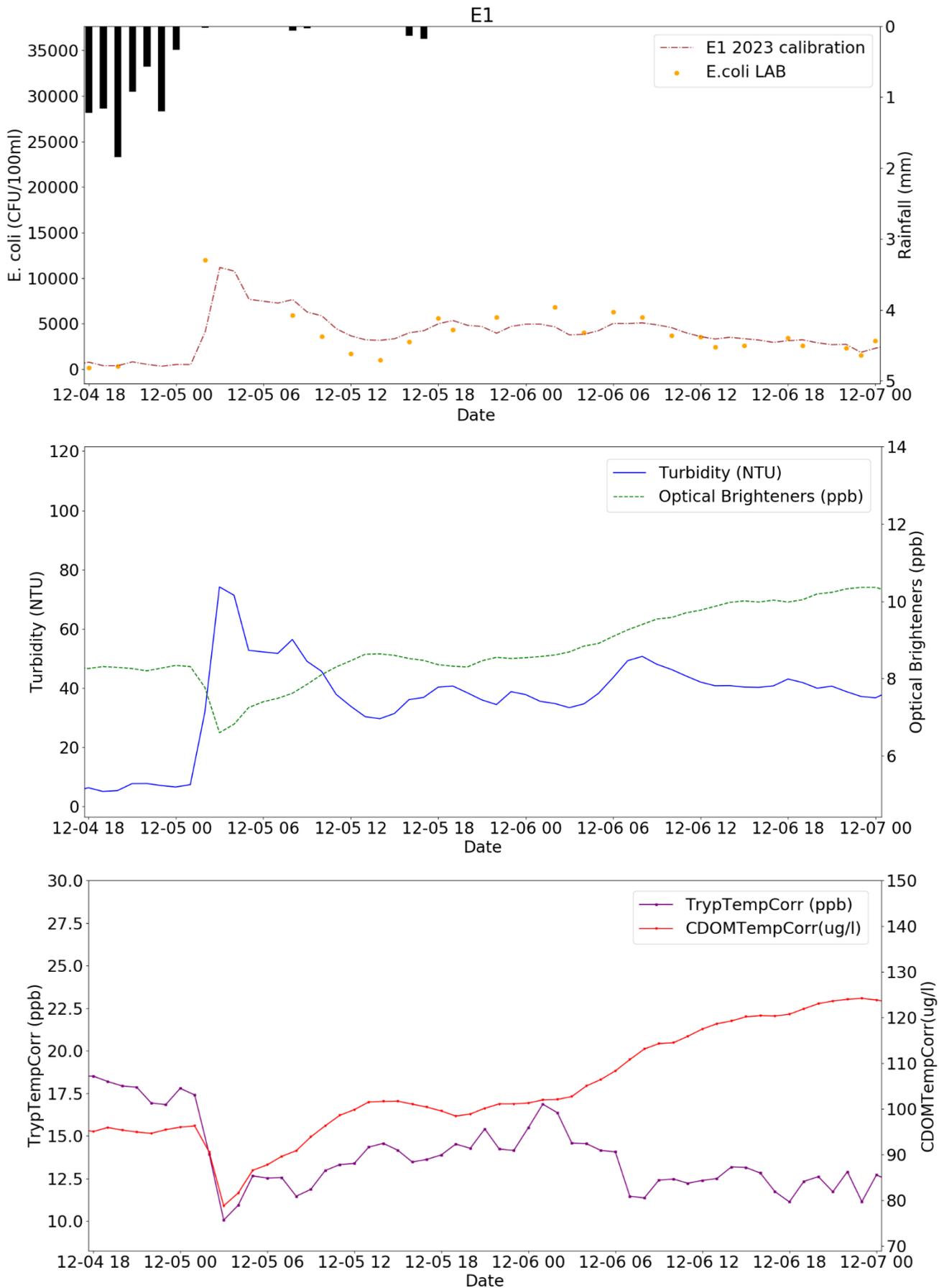


Figure 3.6 E1 rainfall event. Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners (ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom

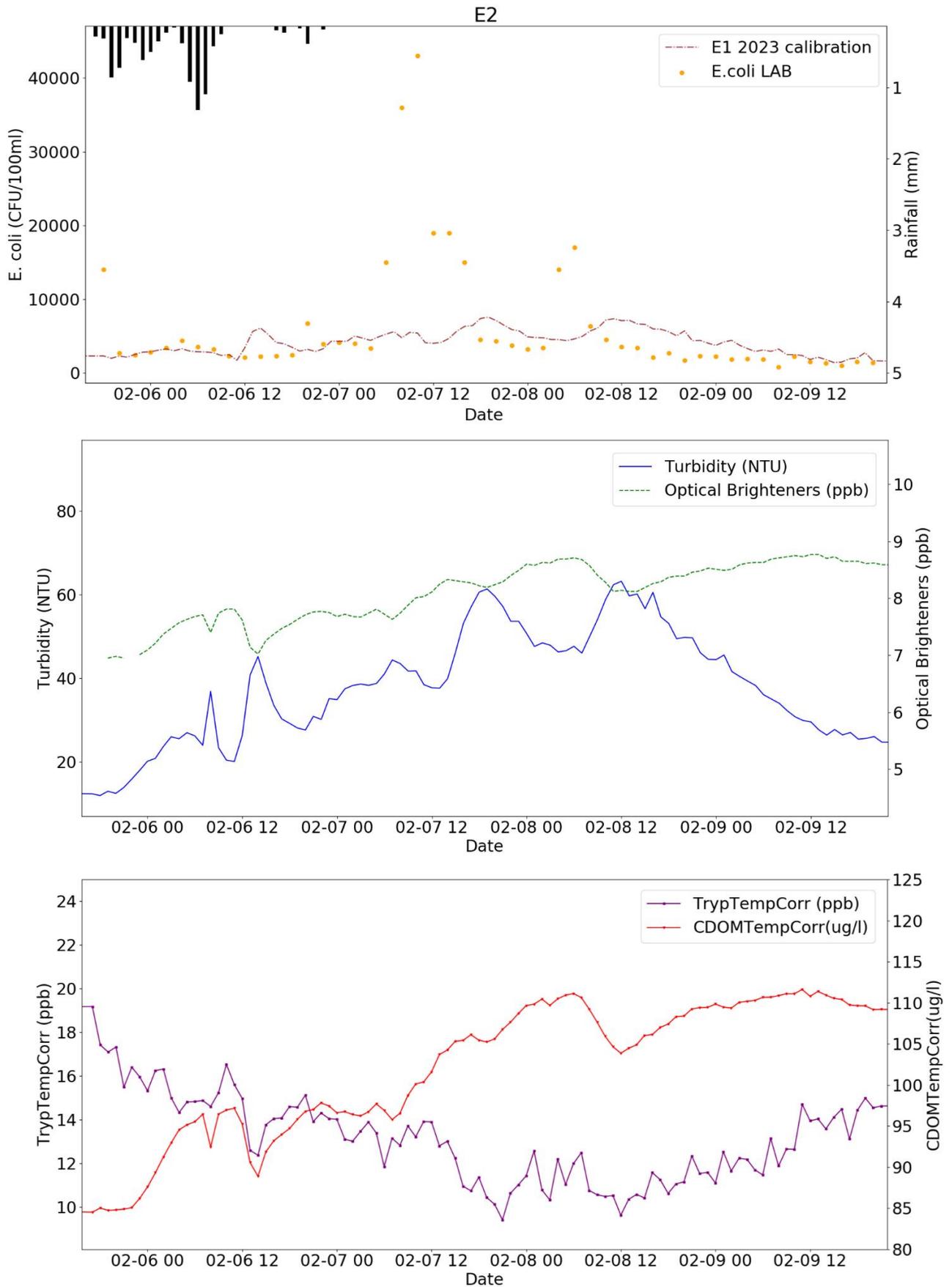


Figure 3.7 E2 rainfall event. Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners(ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.

Event 3 has again produced two *E. coli* peaks. Initial *E. coli* concentrations were low at 200 CFU/100ml but quickly escalated to 17000 CFU/100ml during the first peak and 5800 CFU/100ml during the second peak. While the turbidity has followed the trend of first *E. coli* peak it did not show the matching rise for the second peak. The optical brighteners, tryptophan and CDOM have all displayed clear increases in timing with the two peaks.

Event 4 lab analysed *E. coli* has shown the event to be complex with high levels of *E. coli* and more than the previously evidenced two peaks. The event had lows of 3800 CFU/100ml and highs of 42000 CFU/100ml. The calibrated sensor data misses the first half of the elevated *E. coli* levels as the turbidity only increases during the second half of the event. However, the optical brighteners, tryptophan and CDOM shows a large increase in first half and a decrease in the second half.

The contrast in the ability of sensor to derive *E. coli* data from surrogate parameters between events E1-E2 and E3-E4 could point to the difficulty of the sensor to perform under conditions where the supply sources are complex and highly variable. Therefore, deriving a set relationship between *E. coli* levels and the surrogate parameters from just a single calibration event may be insufficient. While the probe manufacturer recommended one calibration event only, the data analysis suggests possible need for more regular or seasonal site specific calibrations.

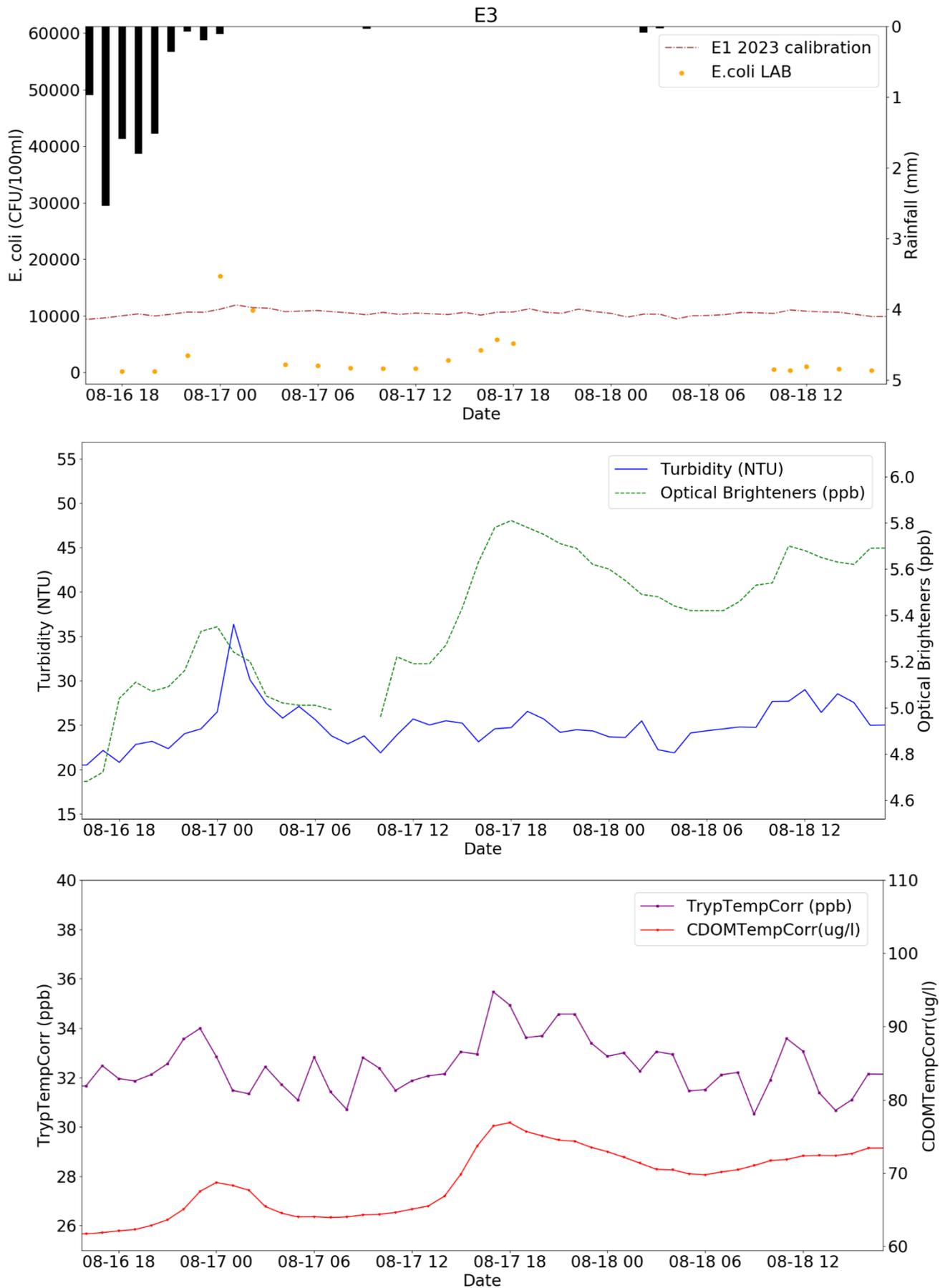


Figure 3.88 E3 rainfall event. Calibrated sensor and lab analysed *E. coli* (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners (ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.

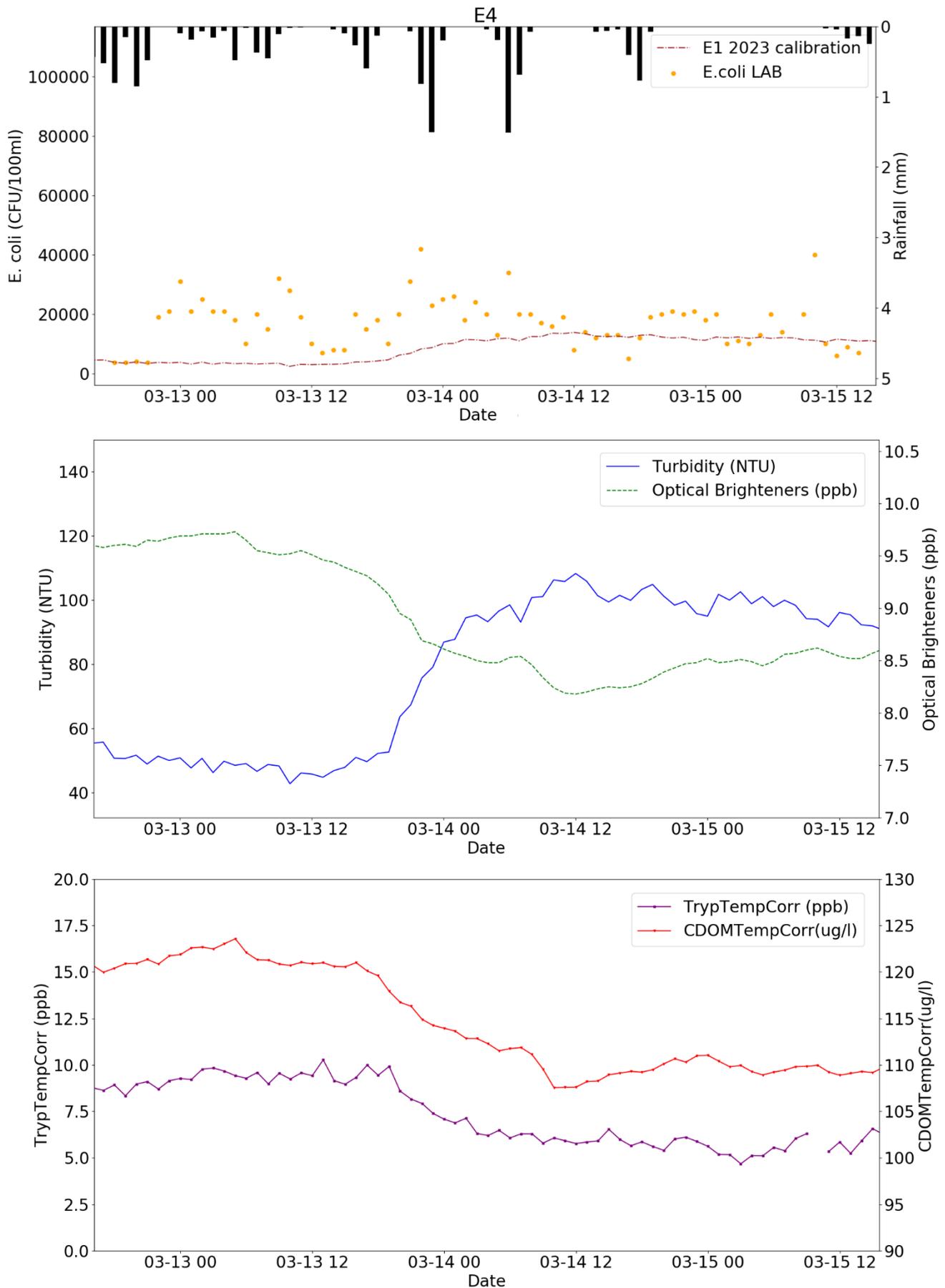


Figure 3.9 E4 rainfall event. Calibrated sensor and lab analysed E. coli (CFU/100ml) data – top graph ; turbidity (NTU) and optical brighteners (ppb) - middle; tryptophan (ppb) and coloured dissolved organic matter (CDOM) ($\mu\text{g L}^{-1}$) – bottom.

The data from all the sampled events was combined into a single database. A log\log plot (figure 3.10) was then produced of laboratory *E. coli* data versus sensor derived data. The flatness of the trend data displays and its distribution that is not uniform points towards a systematic error of *Proteus E.coli* data. Results of stepwise regression analysis used to study correlations between parameters measured by the sensor and lab analysed *E. coli* data are shown in Table 3.2. Turbidity (NTU), CDOM (ppb) and optical brighteners (ppb) were the parameters selected for the model while tryptophan (ppb) and temperature (C°) were rejected. R² of the model indicate that 39.7% of the variation in lab measured *E.coli* is explained by the model, turbidity (NTU) on its own accounted for 34.1%.

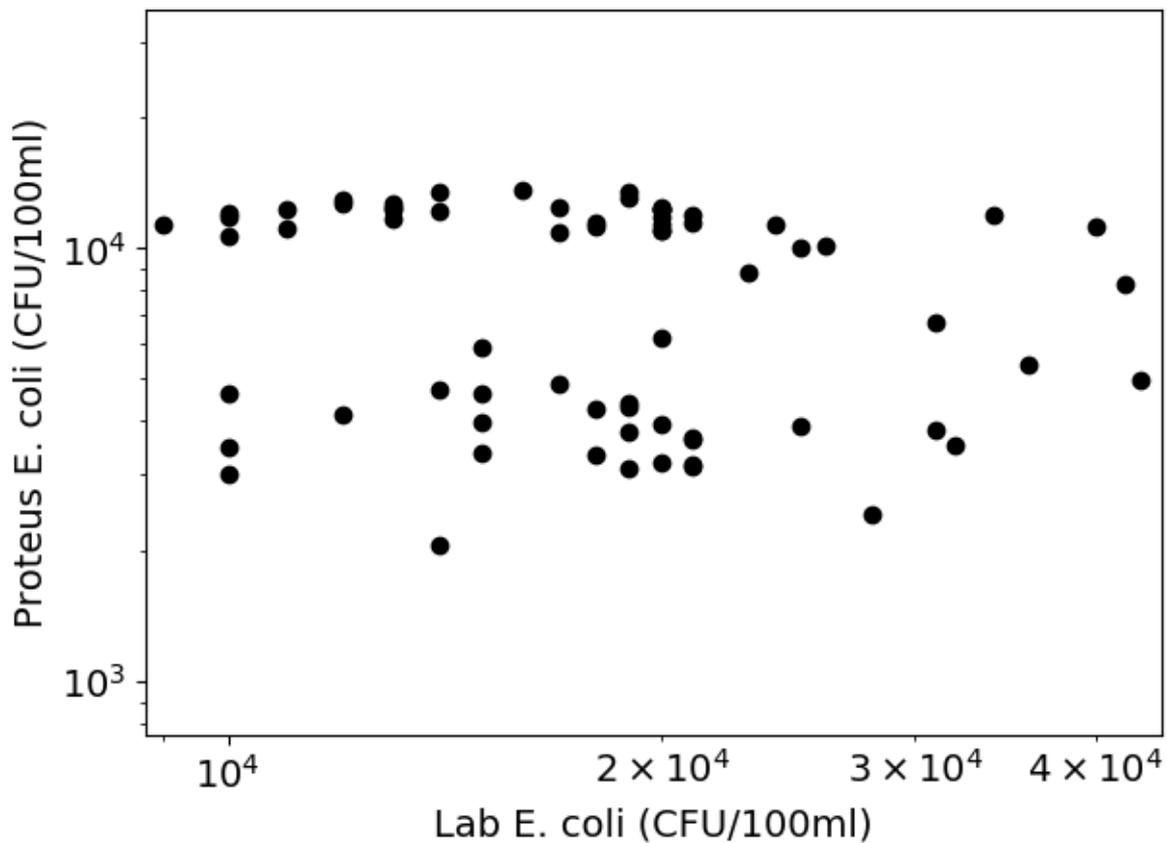


Figure 3.10 *Proteus* vs lab analysed *E. coli* log/log plot

Table 3.2 Stepwise regression model summary for laboratory *E. coli* data

Model Summary		
Model	Variables Entered	R ²
1	Turbidity (NTU)	0.341
2	CDOM (ppb)	0.370
3	Optical Brighteners (ppb)	0.397

3.7 Conclusion

High frequency sampling of *E. coli* levels in river water following a rainfall event has shown acute rises in *E. coli* levels. Surface water quality is significantly affected by intense rainfall events in the catchment. Standard water quality monitoring would not have necessarily given indication of these impacts as it would have unlikely captured the very acute but brief peaks of *E. coli*. This reinforces the need of high-resolution sampling to capture the impact of acute events.

Overall, the probe performance is varied, with some events identified better than others. It's use for real time warning would therefore be limited. The sensor calibration for acute pollution events with complex sources needs improvement. Given that probe does not perform well this gives more weight to developing the modelling approach.

The implementation of continuous water quality monitoring of sewerage undertaker assets is a legal requirement for water companies with minimum guidelines outlined in Continuous Water Quality Monitoring Programme Provisional technical guidance for sewerage undertakers on implementing s.82 of the Environment Act 2021 (DEFRA, 2023). Minimum monitoring parameters detailed in this guide are levels of dissolved oxygen, temperature, pH values, turbidity and levels of ammonia. The findings of the study in this chapter shows that at least the parameters measured by the water quality probe in this case do not accurately represent *E.coli* pollution in surface water. Bacterial pollution in surface water may therefore not be accurately represented by the minimum parameters listed in the DEFRA guide thus failing to identify the periods of high risk.

Chapter 4 - Forecasting acute rainfall driven *E. coli* impacts in inland rivers based on sewer monitoring and rainfall runoff

.4.1 Introduction

Developing understanding of the fate, transport and survival of faecally derived microorganisms in river systems is a requirement for improving the effective and safe management of water resources (DWI, 2020; Dienus et al., 2016, Graydon et al 2022), and for health risk assessments associated with recreational activities undertaken in water bodies (Bathing Water Regulations 2013; Marsalek and Rochfort, 2004, Boehm and Soller, 2020). In many countries, the quality of surface water bodies has come under increasing recent focus due to increased spill frequency monitoring of storm sewer overflows (SSOs) and public demand for designated bathing water sites (Zan et al. 2023). For example, in the UK the Environment Act (2021) has recently increased requirements for the direct monitoring of water quality impacts of SSOs. Whilst the robust direct real time measurement of microbial water quality remains unproven (Demeter et al. 2020; Burnet, J-B., 2021), modelling tools can potentially consider and provide warning of periods of elevated risk to surface water sources and public health. However, the development of widely applicable, generalized tools to understand faecal pollution and associated risks in surface waters remains challenging, especially those caused by acute impacts with high temporal variability (Taghipour et al. 2019, Kammouna et al., 2023). A number of studies have conducted detailed monitoring and/or small scale modelling to understand spatial and temporal dynamics of *E. coli* at individual river reaches, or in small agricultural sub catchments (e.g. Hellweger and Masopust, 2008; Sokolova et al. 2013; Gao et al. 2015; Neill et al. 2020). However, there is a current lack of well validated modelling methodologies for acute impacts that can be applied in mixed use (i.e. urban and rural) catchments, without extensive characterisation of sources and the use of detailed 2D/3D hydrodynamic modelling (and associated topographic surveys). Further, forecasting models for early warning applications (such as water abstraction management or bathing water alerts) require input datasets which characterise source loadings that can be collected and communicated remotely and be available in near real time (Seis et al. 2018, Yassin et al. 2021).

Many studies utilize *E. coli* counts as an indicator of faecal contamination in waterbodies (Madoux-Humery et al. 2013). However, the conventional microbial analysis of water quality samples is relatively time/resource intensive. For example, the membrane filtration method (standard used by the UK water industry), involves dilution (if needed), filtration and incubation of the sample for a minimum of 18 hours (The Standing Committee of Analysts, 2016). The characterization of the microbial quality of surface waters is therefore commonly based on sampling and analysis conducted at relatively coarse resolution in relation to the temporal

dynamics of potential rainfall driven sources, and hence can neglect the full influence of SSOs which may only discharge for a few hours (Seis et al., 2018; Madoux-Humery et al. 2016; Jalliffier-Verne et al., 2017, Shepherd et al. 2023). Further, a number of previous studies have shown that a significant source of faecal contamination in rivers within mixed catchments is diffuse, rainfall-driven runoff, with risks particularly high during elevated flow events due to contaminated runoff from pastoral farmland which also commonly exhibits high temporal variability at sub hourly timescales (Oliver et al, 2009; Ghimire and Deng, 2013; McKergow and Davies-Colley, 2010; Jovanovic et al. 2017, Buckerfield et al. 2019, Hubbart et al., 2022). The mismatch in monitoring practice and timescales of key water quality processes mean that significance of many accumulation and transport processes is currently poorly understood, particularly those which may dominate acute impacts over shorter timescales, such as mixing and dispersion processes in river and streams (Camacho Suarez et al., 2019), spatially variable rainfall runoff and associated processes (Jovanovic at al., 2017) and volumes and loadings from individual highly intermittent SSOs (Madoux-Humery et al, 2015, Owolabi et al. 2022). Many existing approaches widely applied to predict diffuse pollution exposure in surface water bodies are developed with a view to analysing long-term effects of catchment management practices and are often calibrated and validated with relatively coarse datasets (daily and above). (Sadeghi and Arnold, 2002; Collins and Rutherford, 2004; Dorner et al., 2006; Ferguson et al., 2007; Walker and Stedinger, 1999; Whelan et al., 2014; Haydon and Deletic, 2006; Schijven et al. 2015; Sterk et al., 2016; Brannan et al., 2002). As a result, many existing catchment scale water quality models lack detailed representation of spatio-temporally distributed surface runoff generation from source areas, intermittent point loadings (e.g. from SSOs) and transport processes and hydrological pathways throughout the catchment, which are likely to be required for accurate prediction of the arrival of temporally variable short-term (sub daily) peak concentrations of pollutants following rainfall events (Asfaw et al, 2018, Buckerfield et al. 2019). This limits such process based models applicability and viability for forecasting applications such as active water abstraction management (Yassin et al., 2021) or real time bathing water condition modelling/early warning (Seis et al. 2018). There is therefore a need to develop and validate new, practically applicable forecasting tools for faecal contamination that can be applied at catchment scales and consider acute inputs from both agricultural and urban sources. In addition to a lack of water quality data at appropriate resolution for calibration and validation, further challenges associated with the modelling of faecal contamination include high inherent parametric and structural uncertainties associated with modelling loadings from inputs such as SSOs and diffuse agricultural runoff (Srivastava et al., 2018, Tscheikner-Gratl et al., 2019).

The lack of integrated monitoring and modelling capabilities of acute impacts across the urban drainage and catchment domains, means that quantifying the relative scale and nature of risks to water resource systems from different potential sources (e.g. SSOs vs rural diffuse runoff) is also practically difficult (Camacho Suarez et al., 2019, Derx et al. 2023). The recent use of microbial source tracking techniques has been shown to successfully elucidate potential sources (Joseph et al. 2021, Wiesner-Fridman et al. 2022, Zan et al. 2023). However, the present cost and complexity of such techniques means that they are generally only applied to a limited number of samples, which may not provide a representative apportionment of source loadings over longer timescales.

Despite challenges, in recent years the quality and quantity of spatially distributed environmental datasets of concern to water applications has increased, including radar rainfall data, remote sensing of soil condition and land use, and in some cases such as in the U.K., datasets concerning timing of SSO discharges for urban drainage networks. Further to this, the use of automated sampling techniques has simplified the logistics concerning the collection of higher resolution water quality samples. The potential for real time sensing/monitoring of faecal pollution is also a subject of current research, for example based on fluorescence-based detection of the enzymatic activity (Demeter et al. 2020, Burnet et al. 2021), however the reported performance of such techniques for *E. coli* measurement is variable across different waterbody types. Whilst these tools provide potential to improve understanding of short-term dynamics in surface runoff-based generation and transport, many of approaches/datasets have yet to be integrated into river impact models or fully deployed to characterise and assess the significance of acute loadings of Faecal Indicator Organisms (FIOs) into receiving waters. For example, to the authors knowledge the use of directly monitored SSO water levels as an alternative to hydrodynamic and water quality modelling of an urban drainage network to estimate SSO impacts in near real time has not been previously attempted.

The aim of this work is to develop a novel, practically applicable process-based forecasting approach to characterize short term *E. coli* dynamics in catchment scale river networks, considering both inputs from SSO discharges and diffuse agricultural runoff. The model application is focused on providing advanced warning of water quality issues at a water abstraction site, although a similar model structure may be considered for forecasting the quality of recreational waters. As such, understanding the arrival time and duration of elevated loadings within the river network following commonly occurring rainfall events is the primary objective of the model. The approach is based on the temporal routing of individual source areas (based on land use) within the catchment through the river network, considering spatially variable rainfall runoff processes (for agricultural areas), and the novel use of hydraulic

monitoring data from individual SSO sites provided by the water infrastructure operator. The model is calibrated and validated against new hourly/bi-hourly datasets of *E. coli* concentrations in a UK case study catchment featuring both agricultural and SSO inputs, collected during and after rainfall events over a range of seasonal conditions. The model outputs are used to consider the relative significance of urban and rural sources in the catchment area.

4.2 Methodology

This section describes the case study catchment area, sampling and microbial water quality analysis procedure as well as the development of a modelling approach to describe short-term fluxes of *E. coli* in response to individual rainfall runoff and SSO discharge events.

4.2.1 Study area

The River Leam is a 300km² sub catchment of River Severn with elevation ranging from 46m to 232m above sea level. A surface water abstraction site is maintained by the utility operator for potable water supply (figure 4.1), situated at the catchment outlet. Agriculture is the dominant catchment land use with predominantly clayey and loamy soils. Several urban, suburban and rural developments are also present in the catchment, totalling 12.83 km² of built-up area (Ordnance Survey, 2023), with predominantly combined urban drainage systems also maintained by the utility operator alongside a number of associated SSO outfalls. A UK Environment Agency flow gauging station is situated at the outlet of the catchment to monitor abstraction license restrictions. The normal flow depth of the River Leam at the gauging station ranges between 0.24 m and 1.16 m with an average flow of 1.55 m³/s (Q70; 0.319 m³/s, Q50; 0.441 m³/s, Q10; 3.573 m³/s) and mean annual catchment rainfall of 649mm (NRFA, 2023). The catchment has previously been used to develop a rainfall runoff model to forecast the arrival of pesticides at the abstraction point caused by field runoff (Asfaw et al. 2018). Based on long term routine monitoring at the abstraction site, the utility operator has identified acute faecal pollution after rainfall events as a further water quality concern.

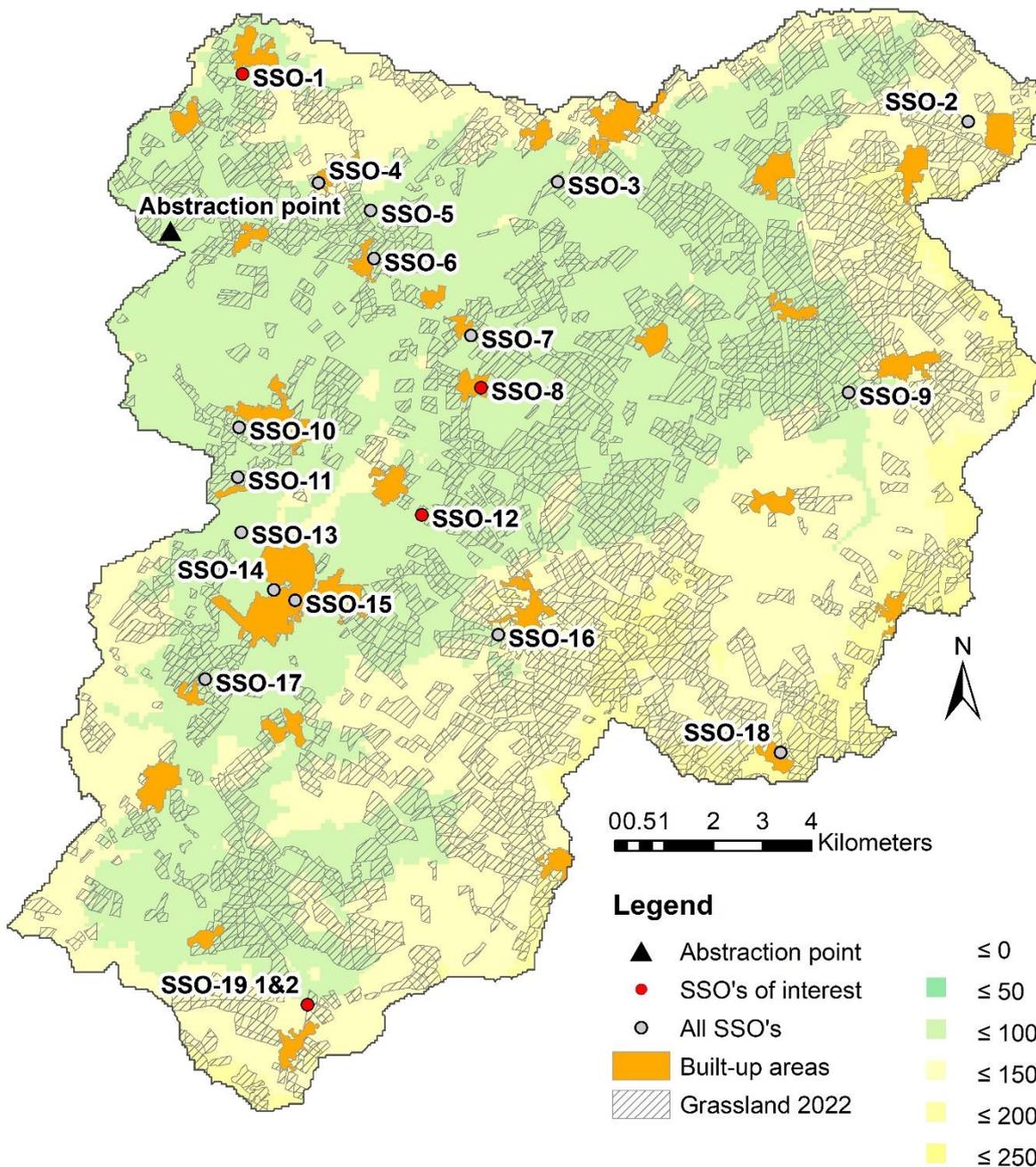


Figure 4.1 Study catchment map showing elevation (meters above sea level) the locations of SSO's, build up areas and grassland for livestock grazing.

4.2.2 Development of *E. coli* modelling approach

Based on the available catchment information, the major sources of acute rainfall driven FIOs in the catchment are assumed to be field runoff from pastoral agricultural land, and SSO spills. The proposed model therefore accounts for SSO loadings and agricultural runoff sources for given rainfall events as identified by catchment land use and asset data. During rainfall events travel times from sources to a monitoring point at the catchment outlet are based on a travel time approach utilizing an existing surface runoff model of the catchment presented in Asfaw et al (2018). Surface runoff is calculated based on overland flow generated from 5 m² grid cells in the catchment utilizing radar rainfall data. The travel time based surface runoff routing

method estimates storm runoff transport from catchment grid cells to the outlet of the catchment based on a Geographic Information System (GIS) method. The spatially distributed time variant direct runoff travel time technique employed in the model accounts for spatial and temporal variability of runoff generation and flow routing through overland flows and stream networks (Melesse and Graham, 2004, Du et al., 2009) following rainfall events at a 1 hr resolution.

Diffuse *E. coli* loadings are estimated based on build-up functions associated with grazing animals in high risk areas (grasslands) and its wash-off to water courses during surface runoff processes (Oliver et al. 2009). SSO impacts are based on level data from Storm Overflow monitors collected at 15 min resolution and used to estimate volumes and loadings entering the surface waters at each timestep from SSO sites. Loadings from significant SSOs and grassland areas are then routed to the catchment outlet. Diffuse and SSO impacts are integrated and combined model to enable rainfall event based prediction of *E. coli* concentrations at the catchment outlet after rainfall events.

The underlying surface runoff, diffuse pollution and SSO modeling approaches are described in further detail in the following sections.

4.2.3 Surface Runoff Modelling

A hydrological model based on the differential form of the Soil Conservation Service (SCS) curve number (CN) method (Mancini and Rosso (1989) has been previously developed and tested within the same catchment (Asfaw et al. 2018), and hence is not reproduced here in detail. Runoff routing is performed using a time varying travel time computation technique, based on flow pathways defined via a GIS flow direction tool based on the catchment digital elevation model (DEM). Output surface flow hydrographs at the catchment outlet are based on cumulative excess rainfall travel times from each grid cell, based on kinematic wave theory Wong (2003). Further details of the model setup and initial validation can be found in Asfaw et al. (2018). To ensure robustness of the approach for this study, the model was evaluated during three further wet weather events, during which the model was compared against monitored EA gauging station data (see section 4.4.1.). Based on this testing, model antecedent moisture conditions were modified to be evaluated based on the preceding 25 days of cumulative rainfall data.

4.2.4 Diffuse Faecal Pollution Loading and Routing

The diffuse modelling component estimates the build-up of *E. coli* loading on grazing land within the catchment, and subsequent wash-off during surface runoff events following each rainfall event rainfall. Grassland/grazing areas were derived from satellite imagery, acquired

from the Centre of Ecology and Hydrology (CEH, 2023) for the period covered in this work (2021-2023). The methodology is based on the approach of Oliver et al. (2009), who developed a method to estimate *E. coli* loadings on fields based on ‘a worst case scenario’ which represented a realistic upper level of stocking densities in the UK. The concentration of *E. coli* (CFU/m²) on grassland for a given Julian day (E_x) is calculated as the sum of the daily fresh input of *E. coli* (Ein_x) by grazing livestock and the previous *E. coli* burden, which estimated as a declining due of first-order die-off relationship (see Table 1):

$$E_x = Ein_x + E_{x-1} * e^{-b} \quad \text{(Equation 1)}$$

Where Ein_x (Colony Forming Units, CFU) are fresh *E. coli* deposits, E_{x-1} (CFU) is the *E. coli* store from the previous day, and b is the appropriate seasonal exponential die-off constant. The ovine and bovine die-off constants (Table 4.1) are higher for the summer (Avery et al., 2004) and lower for the winter months (Oliver et al., 2009).

Table 4.1 Bovine and ovine die-off constants (b) for different seasons, from Avery et al., (2004) and Oliver et al., (2009).

Season	Bovine die-off constant (day ⁻¹)	Ovine die-off constant (day ⁻¹)
Autumn/winter	0.0606	0.0640
Spring/summer	0.0909	0.0920

E. coli deposits are estimated using livestock numbers supplied by DEFRA at UK county level (DEFRA, 2022), multiplied by daily load of *E. coli* excreted by each livestock type during the assumed grazing period (based on the method of Oliver et al., 2009, see table 4.2). The number and type of animals is assumed to be equally distributed over the entire grassland area of the catchment. The daily *E. coli* burden in each 5m² cell is summed up for each livestock type present and used to calculate daily fresh deposit totals for livestock present during its grazing period (Table 4.2).

Table 4.2 Catchment livestock densities and total grassland area derived from DEFRA County level data, utilizing deposit data from Oliver et al., 2009 and assumed grazing periods in the catchment (Oliver et al., 2018)

Livestock type	Livestock Count – County level	<i>E. coli</i> (CFU) contribution per livestock	<i>Ein_x</i> (CFU) per 5 m ² of grassland	Grazing period
Dairy cow	9681.50	8.99x10 ⁸	85195.24	1 Apr–31 Oct
Beef cow	17359.50	2.54x10 ⁹	431602.31	1 Apr–31 Oct
Calves	23644.00	2.10x10 ¹⁰	4662237.30	1 Apr–31 Oct
Sheep	151061.14	7.74x10 ⁸	968451.65	1 Jan–14 Apr ^a ; 1 May–31 Dec
Lambs	127827.44	1.01x10 ¹⁰	14934373.36	1 May–1 Nov
Grassland total (ha)	51080.73			

^a removed for lambing.

In addition to direct deposits, key risk times for slurry spreading in the catchment are in the autumn and spring. To account for slurry spreading contribution to *E. coli* store in this catchment, *E. coli* store on grassland between 31 January and 1 April is assumed to be 2.1x10⁸ CFU per m², based on the findings of McGechan and Vinten, (2003).

E. coli detachment or washout rate from each cell at each timestep (t) during rainfall events is estimated based on the method of Collins and Rutherford (2004), applied here at hourly resolution.

$$Z, t = C_p \frac{O, t}{K} \quad (\text{when } O, t < K) \quad (\text{Equation 2})$$

$$Z, t = C_p, \quad (\text{when } O, t \geq K) \quad (\text{Equation 3})$$

Where Z, t is the *E. coli* detachment or washout rate (*E. coli*/hr) during the timestep, O, t is the cell surface runoff rate (mm/h) during the timestep (from the surface runoff model) and K is threshold a runoff coefficient, taken as 1.04 mm/hour (Collins and Rutherford, 2004). C_p is the available *E. coli* store (E_x), modified by a calibration constant ($K1$), discussed further in section 3.5.

$$C_p = E_x * K1 \quad (\text{Equation 4})$$

The calculated travel time from each high-risk cell is calculated based on the surface runoff model for each model time step. This is then used to route *E. coli* load at each hourly timestep from each cell (Z, t) to the outlet of the catchment. Time series of river flow based on the hydrological model (Q, t , m³/s) and total *E. coli* load in surface runoff (*E. coli*/m³) can then be

used to determine concentrations water arriving at the outlet of the catchment from field sources. Thus, the concentration of *E. coli* from diffuse runoff field sources at each model time step ($E. coli_F, t$), can be expressed as:

$$E. coli_{F, t} = \frac{\Sigma(Z,t)}{(Q,t)} \quad (\text{Equation 5})$$

4.2.5 SSO Spill Volumes, Loading and Routing

SSO monitoring equipment has recently been installed within the catchment as part of the current commitment to provide event duration monitoring data of all operational SSOs to the UK public (DEFRA, 2023). In the study catchment, spill event durations are currently estimated based on monitored level data within chambers connected to outflow pipes (discharging to surface waters), with start and stop times logged as when water level exceeds the outflow weir crest/pipe invert level. Although monitoring systems are not designed to estimate volumes or pollutant loadings to receiving waters, a simple approach is proposed here to make estimates of flow rate and loadings based on sensor information.

Raw water level data (collected via ultrasonic probes) at 15 min resolution data is provided at each of the 20 SSO sites within the catchment (figure 4.1). Based on asset data (weir/pipe dimensions) and monitored level information for the analysis period, SSO spill volumes at each site are calculated every 15 mins where the water level exceeds the outflow weir crest or pipe invert level based on standard equations for hydraulic structures and pipe flows. A similar approach has been used by Fachs et al. (2008) to estimate flow rates from urban drainage systems overflows.

At sites where the outflow is controlled by a weir, the SSO spill flowrate ($Q_{spill,x}$ m³/s), is calculated every 15 minutes as:

$$Q_{spill,x} = \frac{2}{3} C_{dw} L \sqrt{2gh}^3 \quad (\text{Equation 6})$$

Where C_{dw} is the coefficient of discharge for a weir, taken at 0.6. L is the effective length of weir (m), g acceleration due to gravity (m/s²) and h is the height of water surface above weir crest (m). For sites where the outflow is controlled by a pipe, two states are simulated to consider when the pipe is surcharged or flowing with a free surface, defined at each time step by the by the monitored water level relative to the pipe soffit level. When the pipe is in surcharged condition, the spill flowrate is calculated based on an orifice condition:

$$Q_{spill,x} = C_{do} a_o \sqrt{2gh} \quad (\text{Equation 7})$$

Where a_o (m²) is the area of the orifice and h (m) is the height of water surface above the outlet. Where C_{do} is the coefficient of discharge of the orifice, taken at 0.57. When the flow in the pipe has a free surface, the flow rate is based on Manning's equation:

$$Q_{spill,x} = \frac{1}{n} \frac{A^{\frac{5}{3}}}{P^{\frac{2}{3}}} S_o^{\frac{1}{2}} \quad (\text{Equation 8})$$

Where A is the cross-sectional area of the portion of the channel occupied by the flow (m²), n is the Gauckler–Manning coefficient (s/[m^{1/3}]), taken as 0.014 for vitrified clay, P is the wetted perimeter of the channel occupied by the flow (m), S is the stream slope (based on asset data). Equation 6 assumes that the flow in the pipe is uniform, whilst unlikely to be the case the short duration of free surface pipe flow conditions in most cases means that the uncertainty arising from this assumption is unlikely to be significant. The total spill volume per hourly timestep ($V_{spill,x} t$) at each SSO is calculated by the integration of the calculated flowrates. Currently there is no sampling of *E. coli* of storm overflow sites to estimate loadings within SSO spill volumes. Therefore for the purposes of this model *E. coli* concentrations of 40000 *E. coli*/(CFU)\100ml have been utilized based on the previous observations found in literature (Ellis and Yu, 1995, García-García at al., 2021; Hamel et al., 2016; USEPA, 2008) to calculate the *E. coli* load from each SSO ($SSO_{x,t}$). The implications of this assumption are discussed further in section 4.4.

As point source discharges, loadings from SSOs are subject to considerable dispersion effects within the receiving water (Rutherford, 1994). To account for this, SSO loadings from each site at each timestep are routed to the catchment outlet using an Aggregated Dead Zone (ADZ) transport and mixing model (Beer and Young, 1984, Wallis et al 1989). The ADZ is a simple two parameter routing approach which accounts for mixing processes within surface waters. Unlike the (more commonly used) Advection Diffusion Equation the ADZ accounts for skewed distributions commonly observed during mixing studies conducted in surface waters (Rutherford, 1994). The ADZ model provides loadings at the downstream catchment outlet from each individual SSO site ($SSO_{Dx,t}$) as:

$$SSO_{Dx,t} = -\alpha(SSO_{Dx,t-1}) + (1 + \alpha)(SSO_{x,t} - \delta) \quad (\text{Equation 9})$$

Where $\alpha = -e^{\left(\frac{-\Delta t}{\bar{t}-\tau}\right)}$ and $\delta = \frac{\tau}{\Delta t}$. The parameter \bar{t} is mean traveltime over the reach (s) and τ is an initial reach time delay (s). The two ADZ parameters (\bar{t} , τ) can be expressed as the dispersive fraction D_f , as defined by Young and Wallis (1986), and used to scale the mixing effects within a reach.

$$D_f = (\bar{t} - \tau)/\bar{t} \quad (\text{Equation 10})$$

To deploy the ADZ model to each SSO spill, the mean reach travel time (\bar{t}) is estimated based on the surface runoff model described in section 4.2.3. By applying a series of uniform rainfall events (from 0.08 to 1mm/hr) over the catchment, travel time against catchment outlet river flow relationships for each SSO were extracted from the hydrological model. In each case a uniform rainfall intensity was applied to the catchment until the modelled river flow at the outlet stabilized. This allowed representative mean travel times (\bar{t}) from each SSO to be determined over a range of measurable catchment flow conditions (from 6.21 - 79.57 m³/s), based on the coordinates of each SSO as identified based on asset records. As the time delay parameter (τ) cannot be directly established by the hydrological model, τ is calculated for each timestep and SSO based on the traveltime (from above), according to equation 10. In this case a fixed value of Df of 0.2 is taken in all cases, based on the database values of dispersive fractions from UK rivers found in Guymer (2002). Given the uncertainty induced by the use of a single representative Df value, a sensitivity analysis of this parameter on SSO predictions was also carried out (see section 4.4.3). Routed *E. coli* loadings from each SSO are summed for each model timestep and diluted by the calculated river flow volume at the catchment outlet to determine the SSO *E. coli* component ($E. coli_{S,t}$) of the model (equation 11). Similarly to the diffuse component, a calibration parameter ($K2$) is also applied, discussed further in section 4.3.5.

$$E. coli_{S,t} = K2 \frac{\sum_{x=1}^{x=20} (SSO_{Dx,t})}{(Q,t)} \quad (\text{Equation 11})$$

4.3 Model Input Data, Water Quality Sampling and Calibration

4.3.1 Rainfall and river flow

Radar rainfall at 1km² spatial resolution, 15min temporal resolution, used as field runoff model input, was acquired from the UK met-office's NIMROD system. Rainfall was aggregated into hourly intervals to be used with the runoff generation and pollutant wash-off components of the model. A set of rainfall events was selected for validation of the hydrological component of the field runoff model (table 4.4). Summary of the statistics for the four events used for the diffuse component of the *E. coli* model (calibration and validation) are provided in table 4.6.

River flow (m³/s) data from a flow gauging station situated at the outlet of the study catchment was obtained from the UK Environment Agency. It was used as initial baseflow input for field runoff model and for the validation of the hydrological model.

4.3.2. Land use

UKCEH Land Cover® plus: Crop maps were used to create an *E. coli* high-risk area map. In this case, grasslands were selected due to livestock grazing throughout most of the year,

creating *E. coli* stores that replenish and die off with time (as described in section 4.2.4). Figure 4.1 shows the distribution of grasslands in the catchment during the study period.

4.3.3 Storm Sewer Overflow Data

Monitored Storm Sewer Overflow Data (water level time series from 20 catchment SSO's) and asset location data was obtained from the local water utilities Event Duration Monitoring (EDM) analytics platform. An initial screening of the calculated volume from each SSO site (based on equations 6-8) was performed. This showed that 5 out of the 20 SSOs contributed approximately 90% of the total spill volume/load over the 24 month study period (April 2021 – March 2023). Therefore, to simplify the model and further analysis, calculated catchment SSO loadings included only these 5 SSOs. Information on each of these SSO's is included in table 4.3, alongside published EDM return data from Environment Agency (2023). Calculation of traveltimes to the abstraction site is based on the application of the hydrological model and SSO location (river distance to sampling site)

Table 4.3 Characteristics of Leam catchment SSOs included in *E. coli* model, based on monitored period between April 2021 - March 2023 and EDM data is a sum of annual return data from 2021 and 2022.

Name	River distance to sampling site (km)	Modelled traveltime under 1mm/hr uniform rainfall (h)	% of total spill volume in catchment	EDM -Total Duration (hrs)	Outflow type
SSO ₁	4.40	7	65.67	2857.66	Pipe
SSO ₈	15.45	21	5.02	37.52	Pipe
SSO ₁₂	21.24	29	6.01	782.13	Weir
SSO _{19.1}	29.14	35	9.66	3086.83	Pipe
SSO _{19.2}	29.14	35	3.51	371.60	Pipe

4.3.4 Water Sampling (*E. coli* data)

Water samples were taken from the River Leam at the water abstraction site using autosamplers during and shortly after four monitored rainfall events in the catchment. This enabled the continuous collection of hourly/bihourly water samples during storm runoff events, which successfully captured the short-term fluctuations of *E. coli* concentrations at the abstraction site. The auto-samplers were manually triggered before the arrival of forecasted rainfall events. For each event sampling was carried out for a period of 1–5 days, which enabled the acquisition of water samples during the full surface runoff period following the rainfall events. During the sampling campaign a range of seasonal conditions and rainfall events of varying intensity and duration were captured over the period September 2021-February 2023 (table 4.6).

During each event, designated compartments within the autosamplers were filled with ice to keep the adjacently stored collected sample temperature low and stable. Samples were placed in a controlled environment (3-5°C) within 12 hours and analysed within 24 hours of collection. The samples were analysed using Total coliforms and *E. coli*- Isolation and Enumeration from Water by Membrane Filtration method as stated in The Standing Committee of Analysts (2016) based on Sartory and Howard (1992). The water sample is filtered through a cellulose acetate membrane filter upon which bacteria are entrapped. The filter is then placed on a selective growth medium and incubated at 30°C ± 1.0°C for 4 ± 0.25 hours followed by 37°C ± 1.0°C for 17 ± 3 hours. After incubation is complete the colonies, which are characteristic of Coliforms, and *Escherichia coli* are counted.

4.3.5 Model Calibration

Understanding *E. coli* loadings within surface waters is subject to considerable uncertainty. Whilst information concerning the arrival and duration of microbiological loadings into river systems can be directly characterized using monitoring or input data from rainfall radar or SSO sensors, due to a lack of direct monitoring of loadings within field runoff and storm overflows, the model utilizes literature values. However, it is known that these values can be highly variable between sites and with time (Madoux-Humery et al., 2015). Further to this, to maintain a simple model structure, processes such as *E. coli* decay/dies off in the river network are neglected. To mitigate this, two calibration parameters ($K1$, $K2$) are introduced to scale loadings from field runoff and SSOs respectively. It should be noted that these parameters are used to adjust magnitude of *E. coli* loadings and do not affect the arrival times and durations of *E. coli* events (i.e. the primary model application). Calibration of the model is based on initial monitored event (E1), and then validated on the remaining 3 events (E2-4). The sampling events are distributed over the year to cover a range of seasonal and hydrological conditions (winter and summer, with initial to peak river flows ranging well over the Q70 - Q10 range for the 4 events), with rainfall durations ranging from 8 to 41hrs. This provides some indication of the scale of uncertainties to be expected if the processes approximated by the calibration parameters are assumed to be constant throughout the year.

4.4 Results and Discussion

4.4.1 Surface runoff model

A set of events chosen for the validation of the hydrological component of field runoff model are listed in table 4.4. The events were selected to cover a range of Initial and peak flow conditions. Event A3 is also used to calibrate the *E. coli* model (E1). Spatial distribution of

temporally averaged rainfall for events A1 and A2 can be seen in figure 4.2, event A3's rainfall can be seen in figure 4.4 under event E1.

Table 4.4 Summary of rainfall events used to re-evaluate the surface runoff model. Quoted durations are based on presence of rainfall at any position in the catchment. Intensities are based on temporal and spatial averaged values. Initial and peak flow rates during each event based on EA gauging station data.

Event No.	Start date	Duration	Rainfall intensity (mm/hr)		River flow data	
			Average	Peak	Initial flow (m3/s)	Peak flow (m3/s)
A1	03.12.2021	8h	1.10	2.10	5.45	19.2
A2	03.03.2021	15h	0.72	2.33	2.02	7.72
A3 (= E1)	04.12.2020	16h	0.71	1.48	0.43	5.85

The results of hydrological model calibration and validation can be seen in figure 4.3 with the performance statistics listed in table 4.5. The performance of the surface runoff model was evaluated using, R-Squared (R^2) volume conservation index (VCI), calculated using equation 12, model efficiency coefficient (E) as shown in equation 13 and prediction error of time to peak (ΔT).

$$VCI = \sum_{t=1}^T Q_m^t / \sum_{t=1}^T Q_o^t \quad \text{Equation (12)}$$

$$E = 1 - \frac{\sum_{t=1}^T (Q_m^t - Q_o^t)^2}{\sum_{t=1}^T (Q_o^t - \bar{Q}_o)^2} \quad \text{Equation (13)}$$

Table 4.5 Flow simulation model error statistics

Event	R^2	E	VCI	$\Delta T(h)$
A1	0.90	0.88	1.06	1
A2	0.80	0.79	0.94	2
A3	0.68	0.63	0.84	3

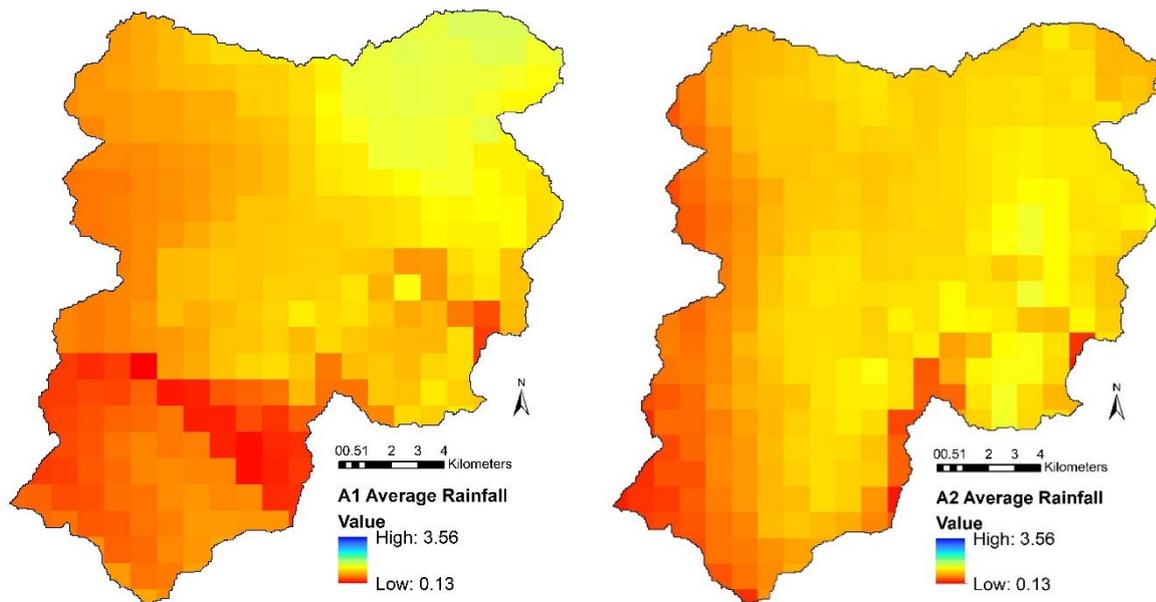


Figure 4.2 Spatial distribution of temporally averaged rainfall (mm) for the events used in hydrological model validation.

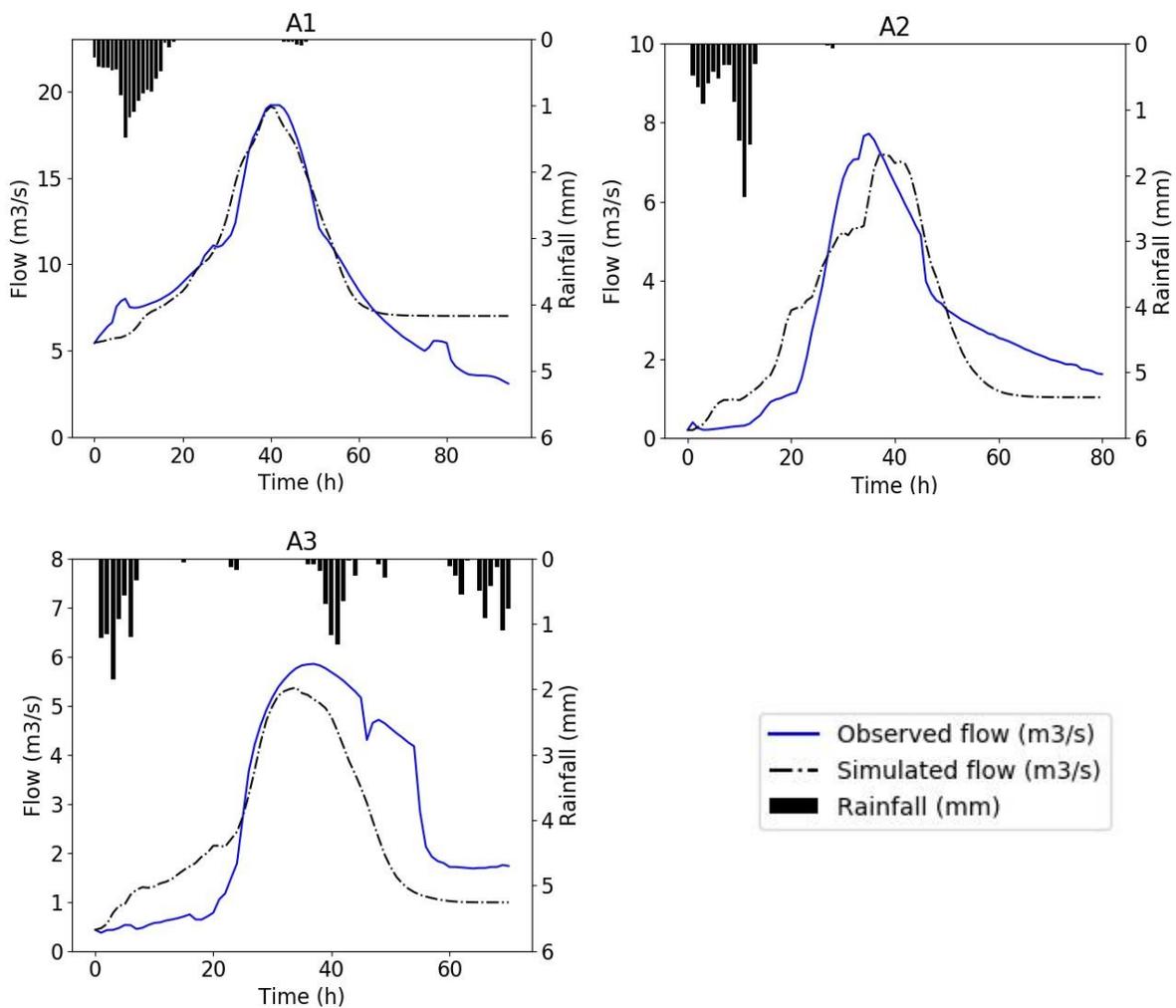


Figure 4.3. Field Runoff hydrological model validation (modelled and predicted flow, spatially averaged catchment rainfall)

In Asfaw et al. (2018), VCI ranged from 0.82 to 0.99 compared to 0.84-1.06 here and $\Delta T(h)$ range of 1-5 hours was slightly reduced to 1-3 hours during this validation. While the efficiency coefficient has seen a reduction from 0.83-0.91 to 0.63-0.88, the error statistics show the hydrological model to still be valid and suitable to use as a basis for travel time estimation of the water quality model.

4.4.2 Measured *E. coli* Dynamics and Model Performance

The integrated SSO-Field runoff model was calibrated and validated using a set of rainfall events and related statistics are listed in table 4.6. For each event, the estimated rainfall depth for a 1 year return period (R.P.) storm has been calculated using the UKCEH recent FEH 22 model (UKCEH, 2023; Vesuviano, 2022). The events used for model validation and calibration are well within the 1 year return period, and are therefore reasonably typical in terms of overall magnitude. Event E1 was used for calibration of parameters $K1$ and $K2$ with the remaining 3 events utilized for validation. Following calibration, a value of 0.4 was used for both $K1$ and $K2$. Figure 4.4 displays the spatial distribution of temporally averaged rainfall for each event. Figure 4.5 presents the measured *E. coli* for each event, spatially averaged rainfall within the catchment, and the outputs of the *E. coli* model. The SSO, field and combined (i.e., summation of SSO and Diffuse components) model outputs are presented. Initial base and peak river flow measured at the gauging station are also provided for each event.

The data from all the sampled events was combined into a single database. A log\log plot (figure 4.6) was then produced of laboratory *E. coli* data versus model forecasted data. The data distribution shows as many under predictions as over predictions. This points to error being present but not systematic. As the model aim is to be used as a tool to identify periods of high risk rather than exact magnitudes forecasted *E.coli* levels should be sufficient to fulfil this purpose.

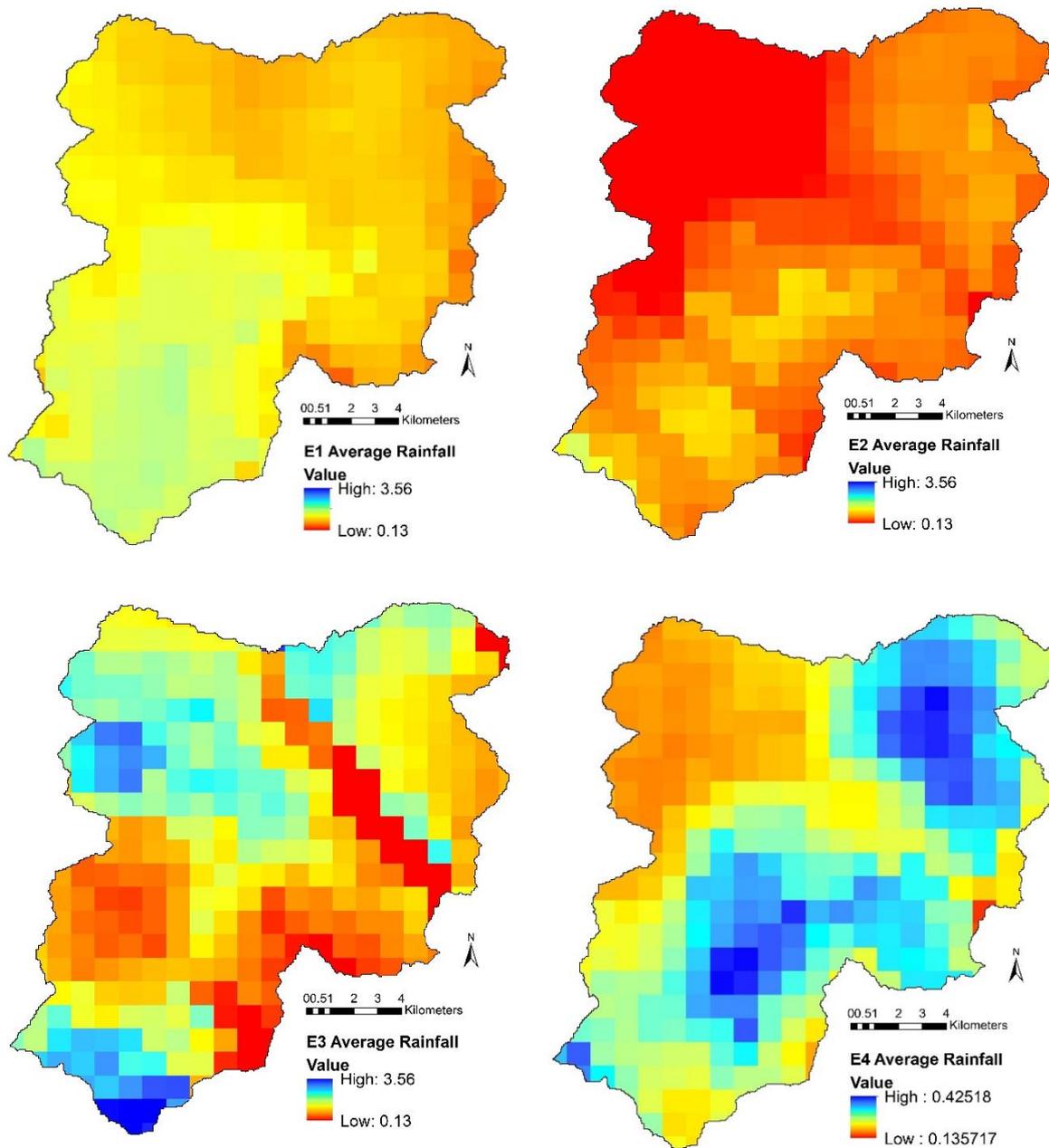


Figure 4.4 Spatial distribution of temporally averaged rainfall (mm) for the events used in *E. coli* model calibration (E1) and validation (E2-4). Note E4 is plotted using an altered scale.

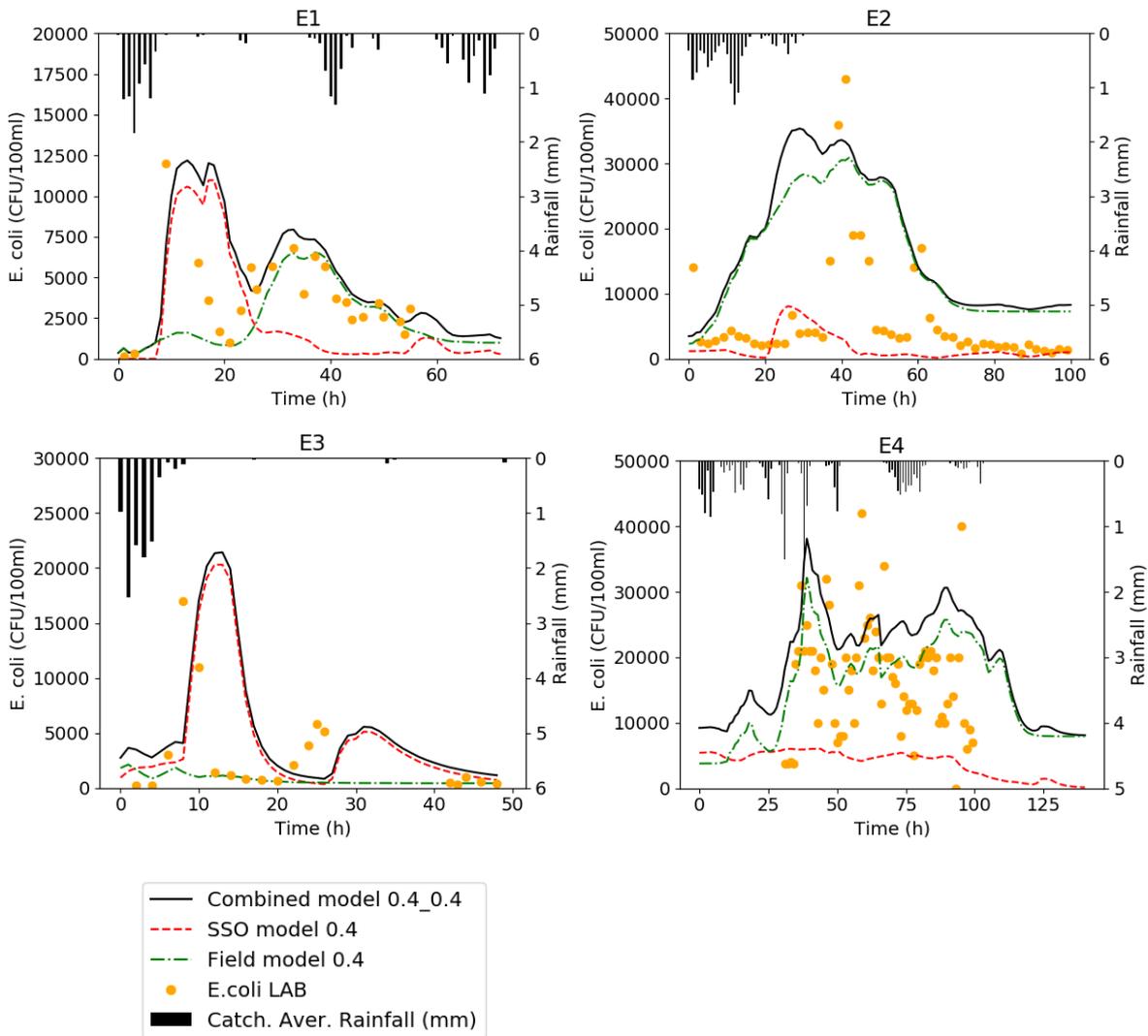


Figure 4.5 Results of combined model and measured (LAB) *E. coli* data showing the contribution from the field runoff and SSO models and spatially averaged rainfall ($K_1=0.4$, $K_2=0.4$).

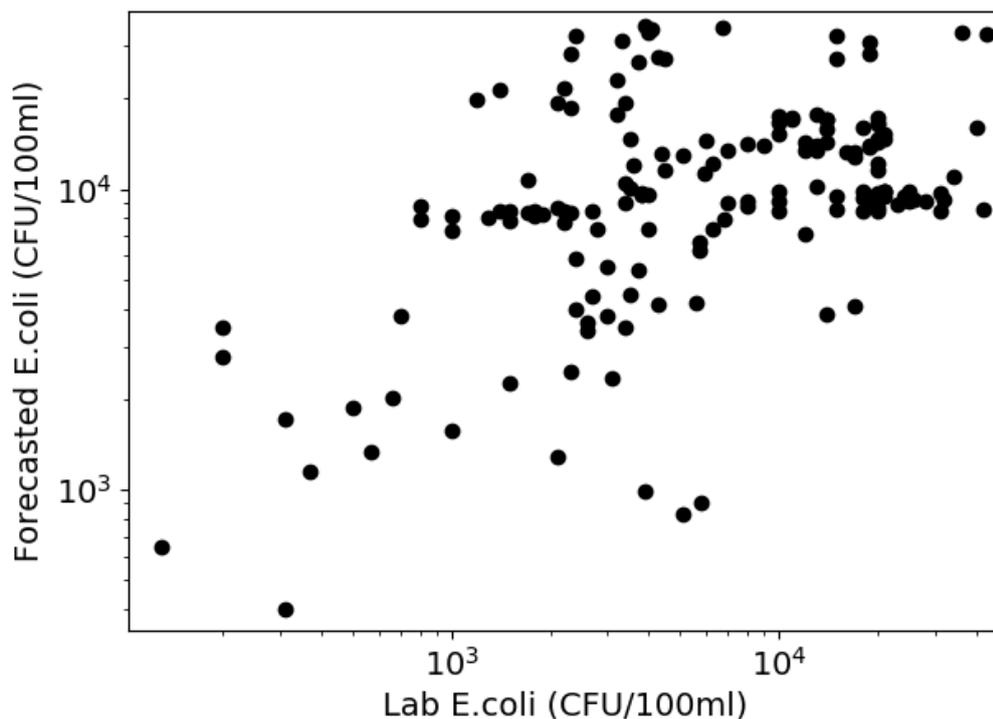


Figure 4.6 Forecasted vs lab analysed *E. coli* log/log plot

Table 4.6 *E. coli* sampling event dates, durations, sampling frequencies, and associated catchment rainfall statistics. Estimated 1 year R.P. rainfall depths for each event duration are also provided based on the UKCEH web service. Initial and peak flow rates during each event based on EA gauging station data.

Event No.	Start date	Sampling duration and frequency ^a (hrs)	Catchment Averaged Rainfall Statistics				River Flow Data	
			Duration (hrs)	Depth (mm)	Peak intensity (mm/hr)	1 year R.P. depth (mm)	Initial Flow (m ³ /s)	Peak Flow (m ³ /s)
E1(A3)	03.12.21	82(2)	8	7.26	1.85	21.82	0.43	5.85
E2	05.02.22	114(2)	32	10.68	1.32	31.55	0.72	2.53
E3	16.08.22	46(2)	9	9.15	2.53	22.63	0.25	0.39
E4	12.03.23	68(1)	41	11.51	1.51	33.62	0.29	2.30

^a in brackets

Table 4.7 presents goodness of fit statistics for the combined *E. coli* forecasting model for each event. $\Delta A(h)$ –prediction error of arrival time, $\Delta D(h)$ - prediction error of event duration, $\Delta T(h)$ - prediction error of time to peak.

Table 4.7 Combined model error statistics

Event	$\Delta A(h)$	$\Delta D(h)$	$\Delta T(h)$
E1	2	2	7
E2	11	19	9
E3	1	1	4
E4	24	45	6

Table 4.8 presents the calculated total *E.coli* load at the Leam abstraction point over each event from field, SSO and combined sources.

Table 4.8 Total calculated E.coli loads (CFU) over each event.

Event	SSO CFU	Field CFU	Total CFU
E1	1.07×10^{13}	2.75×10^{13}	3.82×10^{13}
E2	6.61×10^{12}	7.58×10^{13}	8.24×10^{13}
E3	2.51×10^{12}	4.20×10^{11}	2.93×10^{12}
E4	3.34×10^{13}	1.20×10^{14}	1.53×10^{14}

For all events, observed and modeled *E. coli* exhibit large rises in the sampling period following rainfall. Although the forecasted peak durations are over predicted at times, these over predictions are on the order of a few hours and a small fraction of storm durations. The model suggests that event E1 is characterized by significant contributions from both SSO and diffuse runoff sources. The SSO being responsible for the initial spike (due to the SSO spill from SSO1 relatively close by the abstraction site), and the tail (from 20-60 hours after sampling commenced) being due to the slower diffuse runoff. The preceding rainfall event is moderate and of 8 hrs duration, and sufficient to cause approximately equivalent loadings from both SSO and field runoff sources. Whilst this this event is used for calibration, the model structure correctly predicts the arrival time and duration of the event.

The model prediction for event E2 significantly underestimates the arrival time of the *E. coli* peak concentrations; however, the end of the event is predicted reasonably well. In this case, the event is of lower intensity, but of longer duration, with a relatively high initial river flow. Hence, in this case predicted SSO volumes and loadings are significantly lower and the predicted *E. coli* contributions are mainly from diffuse runoff. The spatial distribution of rainfall intensity suggests a lower rainfall closer to the catchment outlet, and in this case, the model may be over predicting wash off from these areas leading to higher *E. coli* loadings at the start of the event than is observed.

Event E3 is an example of a high intensity, shorter duration event typical of summer rainfall with a low initial and peak river flow. In this case the model predicts that the runoff and loadings from field areas and corresponding diffuse pollution impacts are relatively minor, with the main source of pollution from SSOs which are more likely to overflow during the sudden inundation from such rainfall events. In this case the model gives a generally good estimation of the arrival time and overall duration of the observed *E. coli* concentrations. Two peaks are observed in both measured and modeled *E. coli* values, which are a result of inputs from SSOs at different locations in the catchment. The model predicts a longer duration initial peak than is observed and the arrival times for the second peak are overestimated by approximately 5 - 6 hrs. This

discrepancy may be caused by residual errors within the hydrological model when predicting a short, flashy event, a lack of calibration of the mixing parameters within the routing methodology, or sensing errors within the SSOs themselves.

Event E4 is a prolonged and complex rainfall event resulting in multiple *E. coli* peaks. With significantly larger river flow rates than E1-3. Some SSO loading is present throughout the event but the majority of *E. coli* load supplied via diffuse sources from agricultural runoff. Despite the complexity of the event, the model gives a reasonable approximation of the arrival time of elevated *E. coli* levels commencing shortly after the initiation of sampling. Due to the length of the event, it is unlikely the sampling period covered the end of the event in this case.

The overall results suggest that the sampling campaign has captured events with a diverse range and a variation of sources (SSOs and field runoff). Despite the logistical challenges in measuring *E. coli* at high resolution, this demonstrates that value of measuring events over a range of seasonal conditions, such that relatively short summer rainfall events, as well as longer rainfall events in winter are captured. Further, sampling over winter and summer provides evidence that the model is fit for purpose over a good range of hydrological conditions, which is significant due to the influence of river travel time calculations on the model output.

4.4.3 Sensitivity Analysis

Mixing and dispersion processes in rivers can have significant effects on arrival times and duration that pollutants remain over given thresholds (Camacho Suarez et al., 2019). Given that the application of the model is to forecast arrival times and duration of *E. coli* peaks it is important to understand the uncertainty introduced into model outputs due to the lack of direct quantification of mixing processes, and associated use of standard literature values of mixing parameters. Based on the survey of UK rivers (Guymer, 2002), dispersive fraction commonly falls within the range $0.05 < D_f < 0.4$. Figure 4.7 shows results of a sensitivity analysis carried out on the SSO *E. coli* model for event E3 based on these values as upper and lower bounds. At this site, the analysis shows a relatively small change in arrival and peak timings over this range of D_f , with a more significant effect on peak concentrations (table 4.9). Given this result, it is likely that some improvements in model performance could be achieved in this case by calibration of dispersive fraction, with higher D_f values leading to earlier arrival times which may positively affect performance of events 1 and 3. However at this site the use of a representative D_f value provides an acceptable level of model performance. In this case it is noted the most significant SSO (SSO₁) is relatively close to the catchment outlet (see table 4.3), which may reduce the significance of the mixing processes. In other catchments, with more spatially disrupted SSO loadings results are likely to be more sensitive to the D_f parameter, and hence

direct calibration may be required. Further understanding of mixing processes at sensitive sites may also be improved by undertaking simple solute tracing experiments.

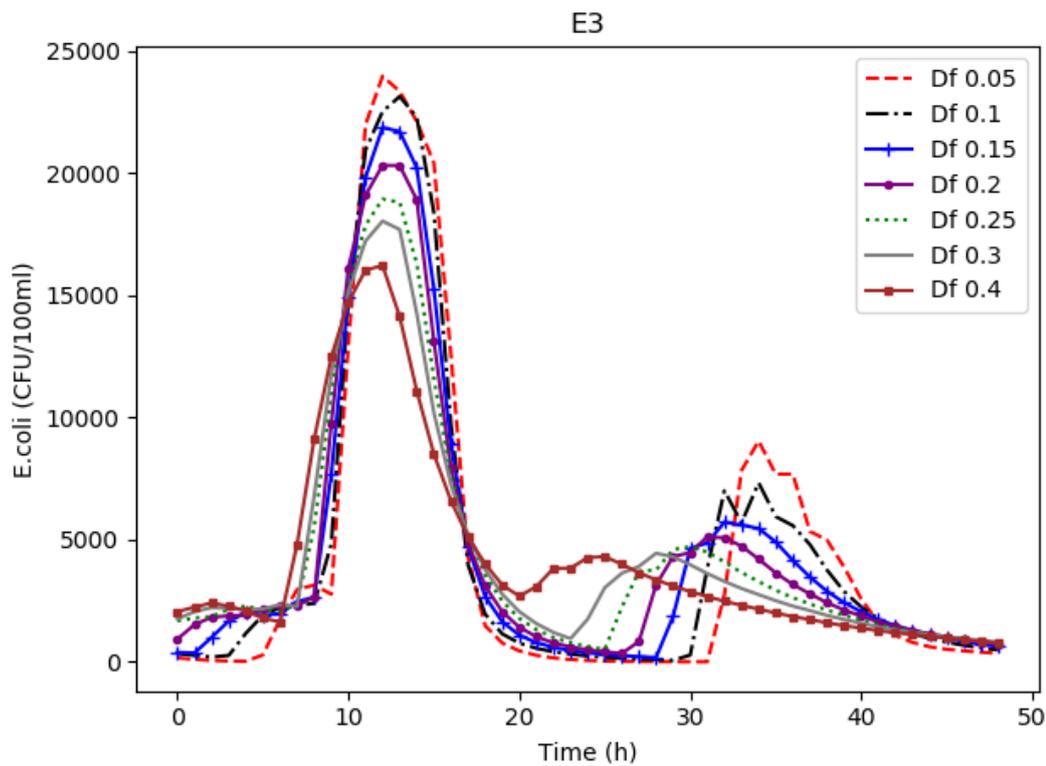


Figure 4.7 Sensitivity of SSO model to dispersion fraction (Df) parameter for event E3.

Table 4.9 Difference in arrival and peak times and concentrations in relation to dispersive fraction value of 0.2, as used in the model.

Dispersive fraction value	Difference in arrival time (h)	Difference in peaks (h)	Difference in peak <i>E. coli</i> concentrations (CFU/100ml)
0.05	1	0	3700
0.1	0	1	2800
0.15	0	0	1500
0.2	0	0	0
0.25	-1	0	-1300
0.3	-1	0	-2300
0.4	-2	0	-4100

4.4.4 Annual Simulation of *E. coli* over Jan - Dec 2022

Whilst the primary objective of the work is to validate an event-based forecasting methodology for *E. coli* peaks under common rainfall events, it is also informative to consider the results from a yearly simulation of rainfall driven acute impacts and consider the relative modelled loading from different sources in the catchment. The full yearly record of spatially distributed rainfall as well as SSO water level data over the 2022 calendar year was therefore taken as model input with resulting time series outputs of modeled *E. coli* used to derive percentile values and relative loadings. It is important to note that the proposed approach does not model *E. coli* during dry

weather/base flow conditions. Hence, for the purposes of this simulation *E.coli* concentrations in the absence of modelled loadings are taken as the mean of sampled measurements taken in dry weather flow conditions (395 CFU/100ml based on 36 measurements).

Table 10 presents 90th percentile *E.coli* values resulting from the simulation, which may be considered in the context of current UK/EU bathing water standards (90th percentile of 900 CFU/100ml for the minimum ‘sufficient’ classification. EU, 2006). However, it should be noted that such assessments are based on a low number of sampled measurements (commonly 12-16 per year) conducted within the bathing water season only. To show relative contributions, results are presented in terms of the total *E.coli* as well as results from the separate field runoff and SSO *E.coli* sub models.

Table 4.10. Forecasted 90th percentiles for a full year (2022) from the full and sub-models.

Total (CFU/100ml)	Field Only (CFU/100ml)	SSOs Only (CFU/100ml)
7725.99	3678.92	3242.36

Results from the simulation show that the modelled water quality falls short of current bathing water standard classifications. Considering the full calendar year contributions from both field and SSO sources are significant (with field runoff being marginally higher), and contributions from either source independently are sufficient to exceed the minimum bathing water threshold. It is notable that current official assessments based on infrequent measurements are unlikely to provide comparable results to a model considering short term dynamics in which runoff/SSOs causes *E.coli* to rise significantly after rainfall events.

Table 4.11 presents calculated total and apportioned *E.coli* loadings (CFU) over different seasons throughout 2022. To consider potential SSO mitigation (i.e. via the installation increased sewer storage or surface runoff mitigation), a simulation in which the contribution from SSO₁ (i.e. the most significant point source) is removed is also considered.

Table 4.11. Forecasted total loads over 2022 and catchment averaged rainfall depth.

	Rainfall depth (mm)	Total Load (CFU)	SSO Load (CFU)	Field Load (CFU)	Total without SSO₁ (CFU)
Jan-Mar 2022	130.02	2.05x10 ¹⁵	1.73x10 ¹⁴	1.87x10 ¹⁵	1.94 x10 ¹⁵
Apr-Jun 2022	115.95	5.03x10 ¹³	1.76x10 ¹³	3.27x10 ¹³	4.29 x10 ¹³
Jul-Sept 2022	82.99	2.70x10 ¹³	1.37x10 ¹³	1.32x10 ¹³	2.12 x10 ¹³
Oct-Dec 2022	219.47	2.68x10 ¹⁴	1.65x10 ¹⁴	1.04x10 ¹⁴	1.46 x10 ¹⁴
Full year 2022	548.43	2.39x10 ¹⁵	3.69x10 ¹⁴	2.02x10 ¹⁵	2.15 x10 ¹⁵

Table 4.11 shows that overall modelled field loadings are larger than SSO loadings in this catchment, although the relative significance changes over the year. Winter/spring seasons are dominated by larger and longer rainfall events causing significant field runoff volumes, increasing the relative loadings from diffuse sources above those from SSO's. This also corresponds to the period in which field loadings are assumed to be higher due to increased grazing. In Summer/Autumn, rainfall volumes are lower with reduced field runoff volumes. However, the increased occurrence of low-duration high intensity rainfall events in summer (e.g. E3) increases the relative significance of SSO loadings in the catchment, as these events are still likely to cause SSO spills (Srivastava et al. 2018).

The removal of SSO₁ contribution from the simulation has resulted in reduction of total bacterial loads throughout the year. Notably, the largest reduction was forecasted between the months of October to December. Therefore, the significance of this SSO as *E. coli* source in the catchment is further reiterated by the results of the annual simulation.

4.5. Discussion

Similar to past studies of which collected high resolution measurements of FIOs in surface waters following precipitation (e.g. Hellweger et al., 2008, Oliver et al. 2015), all four events monitored in this work exhibit significant (order of magnitude) increases in observed *E. coli* after moderate (< 1 year return period) rainfall events. This supports past work which has called for enhanced monitoring and/or modeling of microbial water quality for regulatory classification of waterbodies and/or health risk assessment (Zan et al., 2023). Current characterization of waterbodies for EU/UK bathing water assessments can be based on as few 12 samples per year (EU, 2006), such sampling is highly unlikely to effectively characterize the effects of rainfall driven impacts which can vary significantly at sub daily timescales.

Whilst increases in faecal pollution after rainfall events are expected, this study has also considered how the duration and distribution of elevated periods of *E. coli* can be better understood by the characterization of sources, hydrological pathways and travel times facilitated by the use of spatially distributed rainfall, land use and distributed monitoring at SSOs. For example, where field runoff combines with significant SSO spill contributions (as suggested during event E1) multiple distinct peaks are observed. This supports previous evidence that at this spatial scale the characterization of the spatio-temporal hydrological response of the catchment and the associated pollutant sources, pathways and dilution potential is significant when aiming to model acute impacts (Asfaw et al., 2018, Neill et al. 2020). I.e., rainfall events with similar return periods, but with varying spatial and temporal distributions may result in significant different pollutant responses due to the distribution and

characteristics of source areas across the catchment and associated travel times, hydrological pathways as well as the assimilative capacity of the receiving water (dilution). For *E. coli*, this includes consideration of both the distribution and density of livestock (Oliver et al. 2018, Neill et al. 2020), but also the variations in condition and performance of sewer networks (and associated SSO's) which may be affected by localized factors such as network blockages and sewer maintenance (Shepherd et al., 2023). In general, shorter more intense events (e.g. such as in E3) may tend to have higher contributions from SSOs as the intense localized rainfall can overwhelm the urban drainage network. Longer, less intense events (e.g. E4) see higher contributions from field runoff sources. Considering the variation in the relative contributions of different sources over the duration of a rainfall runoff event may also be significant for when designing future studies considering microbial source tracking techniques for source identification (e.g. Wiesner-Friedman et al. 2022).

The proposed model developed in this work is developed with the intention of describing acute, rainfall driven events for forecasting applications such as short-term water resource management (Yassin et al., 2021) or bathing water alerts (Seis et al. 2018). To enhance practical application, it is also desirable to minimize required data collection beyond existing datasets which are available to water infrastructure operators via remote and/or distributed sensing. As such the model neglects several processes more relevant to understanding longer term/background pollution levels such as groundwater flow, sediment/water interactions and in stream microbial processes (e.g. Afolabi et al., 2023, Jiang, et al., 2023) and utilizes literature values to characterize sources (which are effectively modified during model calibration). A key innovation of this work is the characterization of SSO impacts utilizing spatially distributed water level monitoring. Whilst traditional integrated catchment models characterize sewer impacts using complex sewer network models, these require extensive sewer asset records, detailed calibration and frequently suffer from high levels of predictive uncertainties in the prediction of spill volumes (Srivastava et al. 2018) and pollutant loads (Moreno-Rodenas et al. 2019). It is important to recognize the quantification of loadings by such means is subject to measurement errors (as well as further uncertainties associated with the calculation of flow rate, Leonhardt et al. 2014). Further work is required to better quantify such uncertainties as the direct monitoring of catchments is likely to increase in the future, with further potential to integrate modelling tools and data collection to overcome traditional challenges associated with modelling water quality in complex catchments.

Despite simplifications, results from the validation events suggest that expected peak *E. coli* magnitudes are predicted reasonably well by the proposed modelling approach. The calibrated model parameters $K1$ and $K2$ are both lower than unity, suggesting that initial source loading

values used in this work may overestimate the *E. coli* burden in this catchment from both field and SSO sources. It is noted that the use of constant calibration parameters is a relatively simplified approach to account for uncertainties associated with source loadings and the omission of a number of complex microbiological processes from the model structure (e.g. in stream *E. coli* die-off). However, in this 300km² mixed use catchment, the model accuracy is adequate to provide useful information to the utility operator regarding likely peaks and durations of acute *E. coli* impacts arriving at the water abstraction site. As expected, some residual errors are present in the predictions, and overall there is a tendency to overestimate the duration of *E. coli* peaks. As the model outputs are sensitive to travel time predictions, further refinement of the underlying hydrological model has the potential to improve performance (specifically calculated arrival time and peak durations), and further enhancement to the SSO model may be achieved by a direct calibration of the dispersive fraction parameter. However, given typical uncertainties in the measurement of *E. coli* itself (Harmel et al., 2016), as well as the limited number of measured events, it was considered preferable to avoid risks associated with over-parameterizing or over-calibrating the model (e.g. see Beven, 2006).

Analysis of model outputs over the 2022 calendar year has demonstrated the relative contribution of field and SSO sources, with both having significant contributions in this mixed-use catchment. It is notable that relative contributions change over the seasons due to the nature of the rainfall events and the changes in field source loading due to grazing. There is therefore potential for further use of the model to explore potential mitigation options (i.e. simulating the effects of reducing field runoff, or reducing SSO spill volumes). However, it is recommended that further validation of the model is undertaken over a greater range (magnitude) of storm events to provide increased confidence that the size as well as duration of peaks can be predicted during more significant events (i.e. for 1 year return period).

For transfer to larger, more complex catchments (e.g. for those with longer timescales, or with significant WWTW impacts), the model may require further development to account for these processes and additional calibration. However, in smaller catchments a relatively simple model structure appears sufficient given the model application. This reduces calibration requirements and hence costs for model setup, which can be a significant burden for water quality models (Tscheikner-Gratl et al., 2019). In more complex catchments, a potential option is to integrate travel time-based modelling approaches and high resolution measurements with microbial source tracking techniques (e.g. Zan et al. 2023), to provide enhanced identifiability and validation of travel times from the variety of source areas.

4.6. Conclusions

This paper presents a novel approach to forecasting *E. coli* dynamics in surface waters under commonly occurring, acute rainfall events. To the best of the authors knowledge, no other validated methodologies are currently available in the scientific literature for the description of short term *E. coli* dynamics in mixed catchments (featuring significant diffuse and urban point sources) at comparable scales, utilizing equivalent input datasets. The methodology is based on the determination of travel times from source areas based on hydrological routing, radar rainfall and the novel use of distributed SSO water level monitoring and as such does not require the setup and calibration of a detailed high order hydrodynamic model of the river system or sewer networks. As the primary application is the forecasting of arrival times and durations of periods of elevated *E. coli* levels, understanding travel/arrival times is of primary importance, with factors that control the overall magnitude of *E. coli* peaks of secondary importance. As such, in the absence of monitoring data characterizing catchment source loadings, the methodology is based on assumed concentrations which are calibrated based on model outputs. Despite simplifications, the model provides a reasonably good representation of *E. coli* dynamics in most cases, with calibration parameters not varying significantly over the study period. This suggests the value in accounting for the temporal and spatial variability of sources (diffuse and SSO) when accounting for *E. coli* dynamics, particularly over short time periods in the order of hours. Further, the work provides a new demonstration of how distributed sewer monitoring and rainfall data can be utilized for water resource and surface water management. As the approach is not dependent on complex integrated hydrodynamic modeling and/or direct measurement of source loadings, it has potential to be deployed to water resource management applications such as water abstraction management and bathing water quality forecasting in real time.

The results from the monitoring campaign show significant differences in *E. coli* dynamics between the four monitored events as a function of spatial and temporal rainfall variability causing mobilization of different sources. This finding demonstrates the value of source characterization using remote sensing and spatially distributed sensors and the significance of spatially distributed runoff. The proposed modelling approach can also be used as a source apportionment tool as it allows the effects of different sources to be disaggregated. Further work may consider identifying the significance of individual SSOs or field areas on high *E. coli* periods over longer timescales.

There is significant scope for development to identify and reduce modelling uncertainties, in particular, in larger more complex catchments it is likely that the model complexity will need to be increased to account for additional processes which are less significant in this case (e.g. *E.*

coli die off). However, this effort would likely increase the number of datasets required for robust model calibration to overcome parameter identifiability issues. In this initial application, a simple model structure is preferred given the purpose of the model.

Chapter 5 - Genetic algorithm based land use optimisation for the mitigation of pesticide risk to potable water supply

This chapter presents the development of land-use optimisation method to aid catchment mitigation of rainfall driven non-point pollution impacts. The optimisation is applied separately to consider two different pesticides to compare the performance of the approach under different application scenarios. A novel methodology to characterise rainfall inputs for optimisation-based approaches is developed for this purpose. The effectiveness of the methodology in finding land-use patterns which minimize the time concentration of pesticide is above a given threshold at the abstraction site is evaluated under historic rainfall events.

5.1 Introduction

Land use pattern optimization is a useful method to inform catchment management for the mitigation of non-point pollution (Srivastava et al., 2002; Arabi et al., 2006). It involves simulation of different land use scenarios and evaluation of the resulting pollutant loads as a function of the spatial characteristics of pollutant sources. This allows catchment management resources to be targeted to specific areas/sites which are expected to be most effective in reducing pollutant impacts. For example, the efficient targeting of farmer subsidies for the encouragement of alternate, less damaging pesticides or herbicides (Cooke et al., 2020).

As discussed in chapter 2, a number of land use optimization techniques have been proposed for catchment management (e.g. Zhang et al. (2011); Sadeghi et al. (2009)). However, such techniques have commonly been applied to consider long term, or averaged, or 'characteristic' water quality indicators, rather than for the specific mitigation of acute impacts caused by surface runoff, which feature significant short term dynamics at hourly timescales.

This chapter develops and proposes a land use optimization approach which can be adapted to the mitigation of acute pollutant pesticide loadings from rainfall driven field runoff. The methodology can be used to prioritize the mitigation of high-risk areas within the catchment, and further investigates how removal of these areas (classing them as a non-contributing area) affects acute pollutant loads in river systems.

Land use targeting based methodologies are normally based on inverse modelling. This involves an optimisation-based framework in which an objective function (defined based on some aspect of a water quality model output), is minimized as a function of the spatial properties of the pollutant sources in the catchment. An important aspect is the selection of model inputs for the optimization routine which enable the model to produce a sample 'output' of the system dynamics that suitably represents the application under investigation (Arabi et al., 2006; Perez-Pedini et al., 2005).

In the case of developing an optimization approach for acute rainfall driven pollution sources this presents a specific challenge. Past work has shown that temporal dynamics are sensitive to the spatial and temporal distribution of rainfall over the catchment (Asfaw et al. 2018). The rainfall input used within the optimization routine should therefore account for these variations. Whilst using a long term historical time series of measured rainfall within the catchment would successfully account for these dynamics, this approach would be infeasible for reasons of practical computational resource. This chapter therefore proposes an approach to develop a representative rainfall event to account for these processes which is feasible for use in an optimization framework.

The work in this chapter is based datasets and models collected and developed in the same Leam case study catchment as described in chapter 3. In this case the pollutants under investigation are two specific pesticides of concern to Severn Trent Water in terms of risks to raw drinking water supply (Metaldehyde and Propyzamide, described further in chapter 2). Previous work has developed a validated model to describe acute metaldehyde dynamics in the catchment, and this chapter describes how this model can be coupled to a Genetic algorithm-based optimization technique for land use planning and mitigation targeting. Here a tool is developed to prioritize which existing high risk fields in the catchment should be considered for runoff mitigation, based on an objective function defined as hours the pesticide remains above a given threshold at the water abstraction site. Following optimization, the performance of the approach is evaluated based on a time series of historical rainfall and the proposed land use patterns. In the case of propyzamide, the procedure is similar, however prior to optimization the existing metaldehyde model is adapted and tested on new high resolution propyzamide datasets collected in the catchment following rainfall events.

5.2 Development of the Propyzamide Model

In order to apply the optimisation approach to the distribution of propyzamide, a model must be developed, calibrated and validated that can represent the dynamics of acute propyzamide loadings at the case study catchment outlet. Based on the literature, no currently available models were found to describe surface runoff driven propyzamide dynamics, hence the proposed propyzamide model used here has been adapted from the existing metaldehyde model (Asfaw et al, 2018). This was achieved by amending the necessary variables in the pollutant wash-off component of the model based on the “simplified formula for indirect loadings caused by runoff” (SFIL) (Berenzen et al. 2005; Reus et al. 1999), and revising the high risk field distribution in view of the likely sources of the specific pesticide in the catchment. The specific parameters changed where the sorption coefficient of active ingredient to organic carbon (taken as 840), half-life of active ingredient in soil (47 days) (ADAMA, 2015) and build-

up on soil surface through applications (taken as 0.425 g per 5 square meter based on typical application of 1.7 litre/ha application using 500 g/litre (43.86% w/w) propyzamide suspension concentrate (²Corteva, 2023). The high-risk areas where propyzamide is applied (oilseed rape and field beans (¹Corteva, 2023), figure 5.1) were derived from satellite imagery, acquired from the Centre of Ecology and Hydrology (CEH, 2023). The herbicide can also be used on other crops, ranging from sugar beet and clover seed crops to fruit and forestry crops (²Corteva, 2023). Most of these crops would potentially be grouped under ‘other crops’ within the satellite imagery derived crop land use maps. The model was then recalibrated and validated against hourly propyzamide sampling results collected by the STW. catchment team following monitored rainfall events (table 5.1) in the Leam case study catchment. Rainfall statistics are derived based on rainfall radar data using the same methodology as presented in chapter 4.

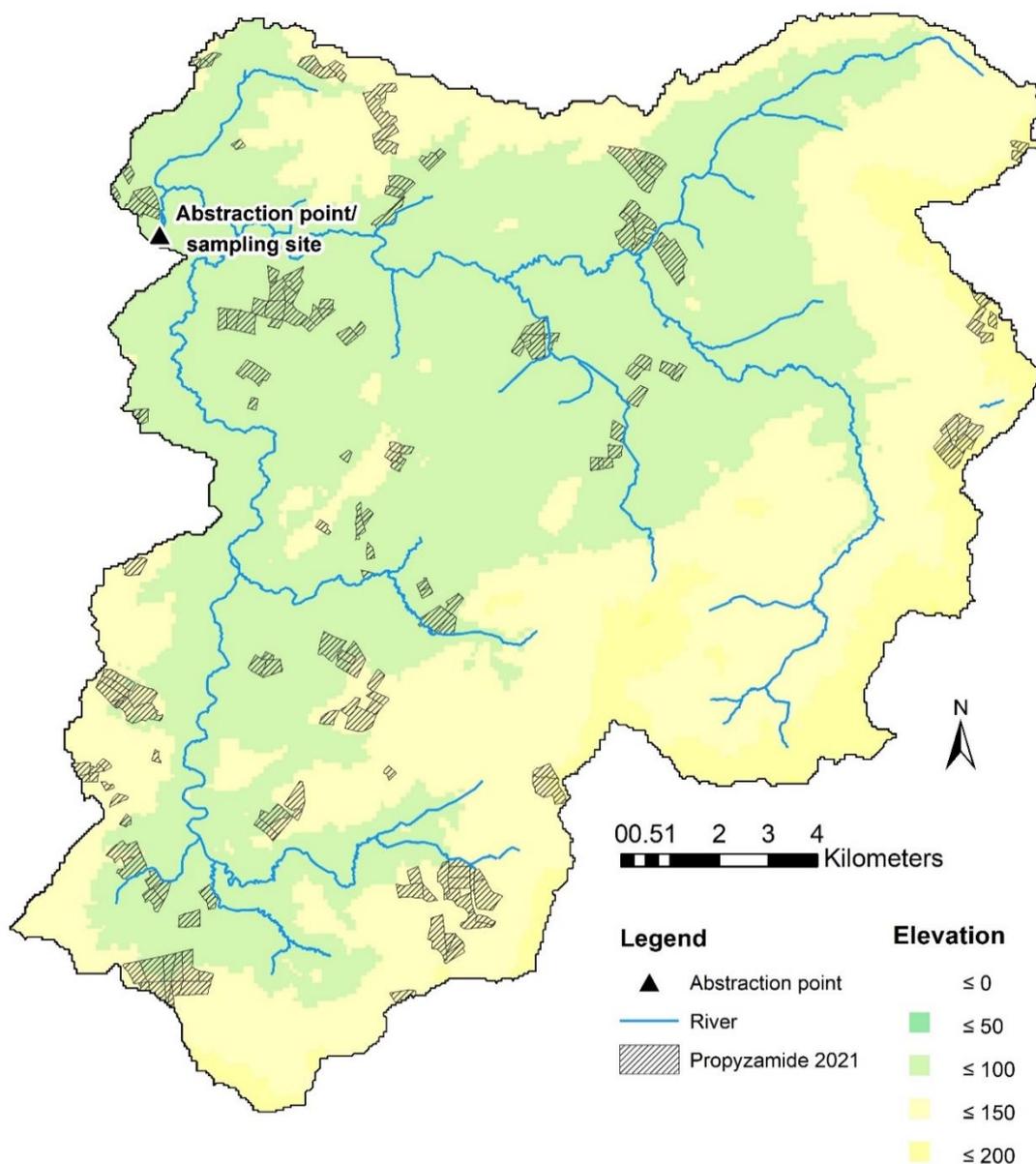


Figure 5.1 Distribution of propyzamide high risk fields identified for 2021

Table 5.1 Propyzamide sampling event dates, durations and sampling frequencies, and associated rainfall statistics. Quoted durations are based on presence of rainfall at any position in the catchment. Intensities are based on temporal and spatial averaged values.

Event No.	Start date	Sampling duration (h)	Rainfall duration (h)	Rainfall intensity (mm/hr)	
				Average	Peak
P1	04.12.2020	120h	19	0.61	4.10
P2	31.10.2021	41h	18	0.57	3.57
P3	26.11.2021	42h	8	0.35	2.33

5.2.1 Calibration/Validation of propyzamide model

Event P1 was used to calibrate the Propyzamide model, using the same procedure as the original metaldehyde model. The calibration parameter is applied to the propyzamide load calculations (equation 14)

$$P_t = KL_tB \quad \text{(Equation 14)}$$

Where P_t – propyzamide load (g) in surface runoff at timestep t , $K = C_o * K_b$, where C_o is propyzamide concentration in the river prior to each rainfall event ($\mu\text{g/l}$), K_b is a calibration parameter ($\text{l}/\mu\text{g}$), B – propyzamide build-up on soil surface through applications (taken as 0.425 g per 5 m^2). Figure 5.2 shows the results of the model using different of K_b values against measured propyzamide levels in the river for event P1. The arrival time of propyzamide was calibrated to fit reasonably well, while the end of the first peak is over predicted and second peak is missed. K_b value of 3.5 produces best fit to the magnitude of the peaks. High K_b value combined with missing second peak suggests the number and location of fields propyzamide was identified to be used on to be underestimated, thus indicating a need for additional investigation into the sources of propyzamide.

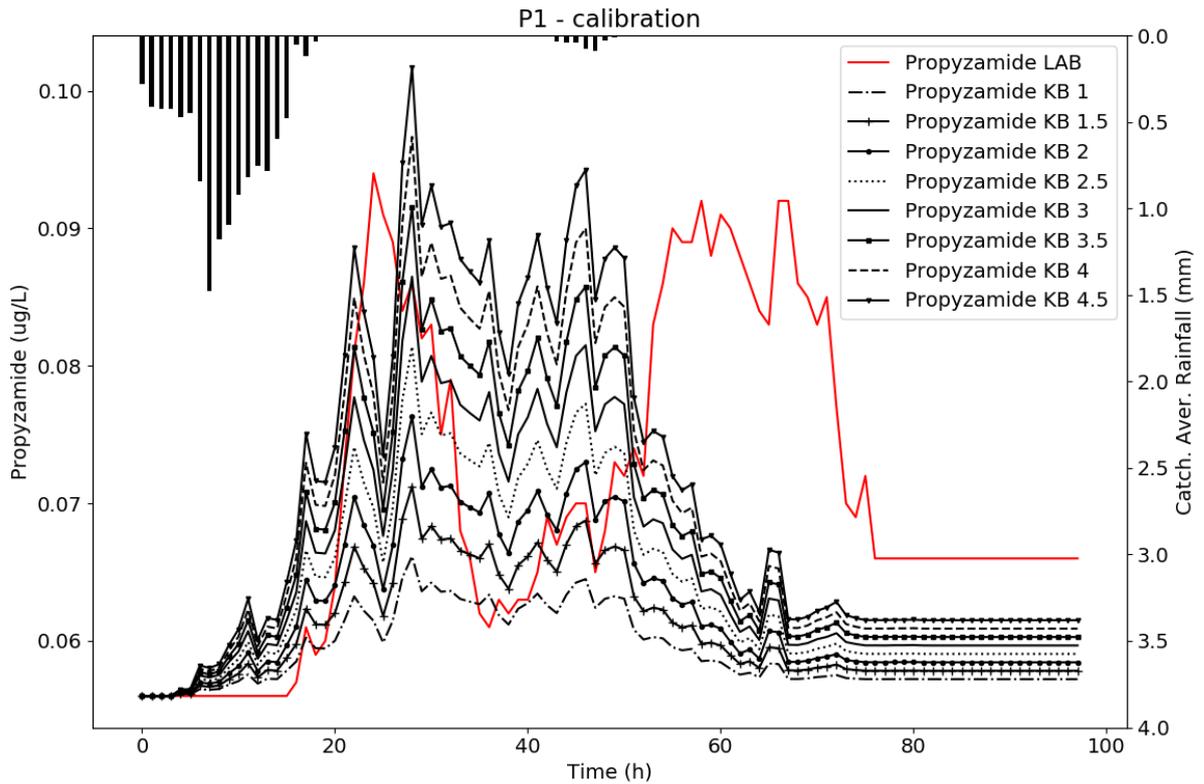


Figure 5.2 Calibration results for propyzamide model against measured propyzamide data

Following calibration a Kb value of 3.5 was applied and the model applied to the two remaining monitored events (figure 5.3). Error statistics for all events are presented in table 5.2.

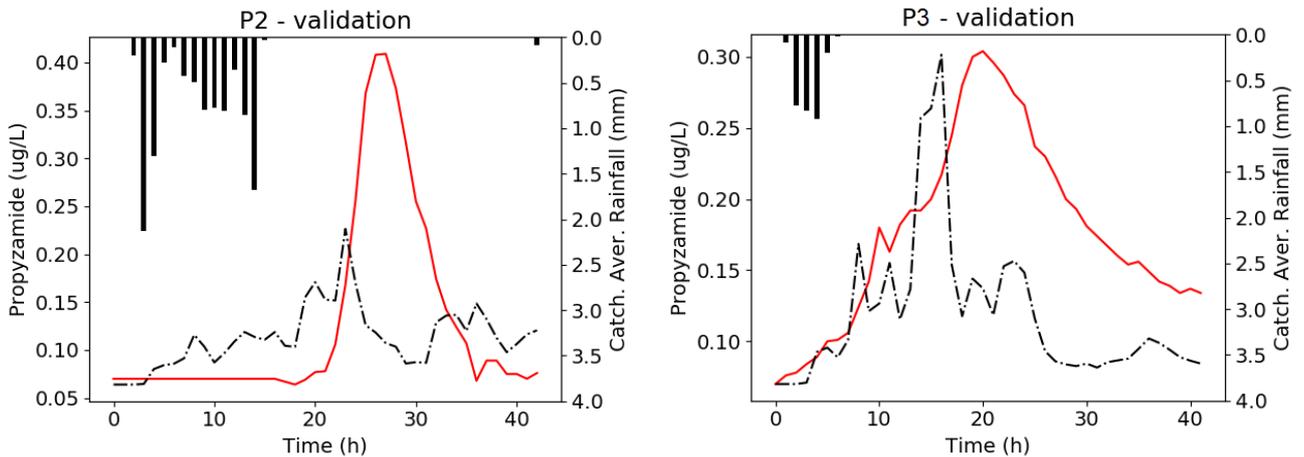


Figure 5.3 Validation results for propyzamide model against measured propyzamide data

Table 5.2 Propyzamide model error statistics

Event	R ²	E	VCI	ΔT(h)
P1	0.07	-0.32	0.95	4
P2	0.01	-0.06	0.88	-4
P3	0.17	-0.78	0.67	-4

Following calibration and validation of propyzamide model, the model shows forecasted peak magnitudes and event start time to match well, however the length of the event is under predicted. The second peak within the calibration event is also missed. Event P2 start time and peak magnitude are both under predicted. Event P3 forecasted first arrival times and peak magnitude well. However, the peak timing of P3 event was under predicted. Overall, the performance of the model is inferior to the original model for metaldehyde. Hence this indicates that some significant, time varying processes which are specific to this pesticide are being neglected by this reasonably simple modelling approach. However, it was judged that for the purposes of this application it was suitable to describe approximate arrival times and event durations and therefore identify time periods of high risk for propyzamide pollution in the river.

5.3 Development of Inverse Modelling method for Designing Catchment Management Options

An inverse modelling methodology was developed and applied to the case study catchment. It aims to determine which spatial areas have most influence on peak pesticide (metaldehyde and propyzamide) concentrations at the abstraction location.

The inverse modelling approach searches for model input, in this case a distribution of catchment land use/high-risk fields, that result in desired model output (pesticide levels in river water). The goal is to minimise the objective function, which is set to be the number of predicted hours that pesticide levels exceed the specified EU and UK threshold of $0.1 \mu\text{g L}^{-1}$ for pesticides in drinking water. It keeps all the other variables (Rainfall, AMC initial pesticide concentrations, approximate total area of pesticide application) fixed and only amends the high risk land use distribution within the catchment. By iteratively running the appropriate pesticide model the land use distribution with that minimised the objective function is sought. There are vast amounts of possible land use patterns and so a guided search algorithm is needed. In this case, Genetic algorithm (GA) was used to carry out land use optimisation (see section 2.5). GA is widely used to solve optimization problems in water resources planning and management (Nicklow et al., 2010, Eulogi et al., 2022).

Within the case study catchment, metaldehyde was applied to 1114 fields in 2018, while propyzamide to only 336 fields. The large difference in fields available for selection between the two pesticides enables evaluation of optimisation method performance when the pool of possible solutions is large vs small. While GA python code was adapted from an existing script, all of the associated processes (zero-one integer programming, checking the solutions total field area, creating high risk shapefiles and checking of the objective function) were written original scripts, see code flow diagram in figure 5.4.

A significant complicating factor is that the spatial and temporal variations in rainfall have a significant effect on the dynamics of pesticide concentrations at the abstraction site (Asfaw et al 2018). Therefore, an appropriate model rainfall input for the inverse modelling approach needs to be carefully considered. Utilising long-term rainfall datasets within the optimisation would potentially give a valid approach, as this would inherently capture the variations of temporal and spatial rainfall over the catchment area. However, this is infeasible in practice due to the time consuming nature of the optimisation routine. Hence, to account for the influence of spatial variability of rainfall patterns and intensities, a shorter 'compilation' of rainfall events representative of the historic catchment rainfall was compiled. This rainfall event 'mashup' contain historical rainfall events representative of catchment rainfall spanning a number of years (2015-2019). This is further explained in section 5.3.2.

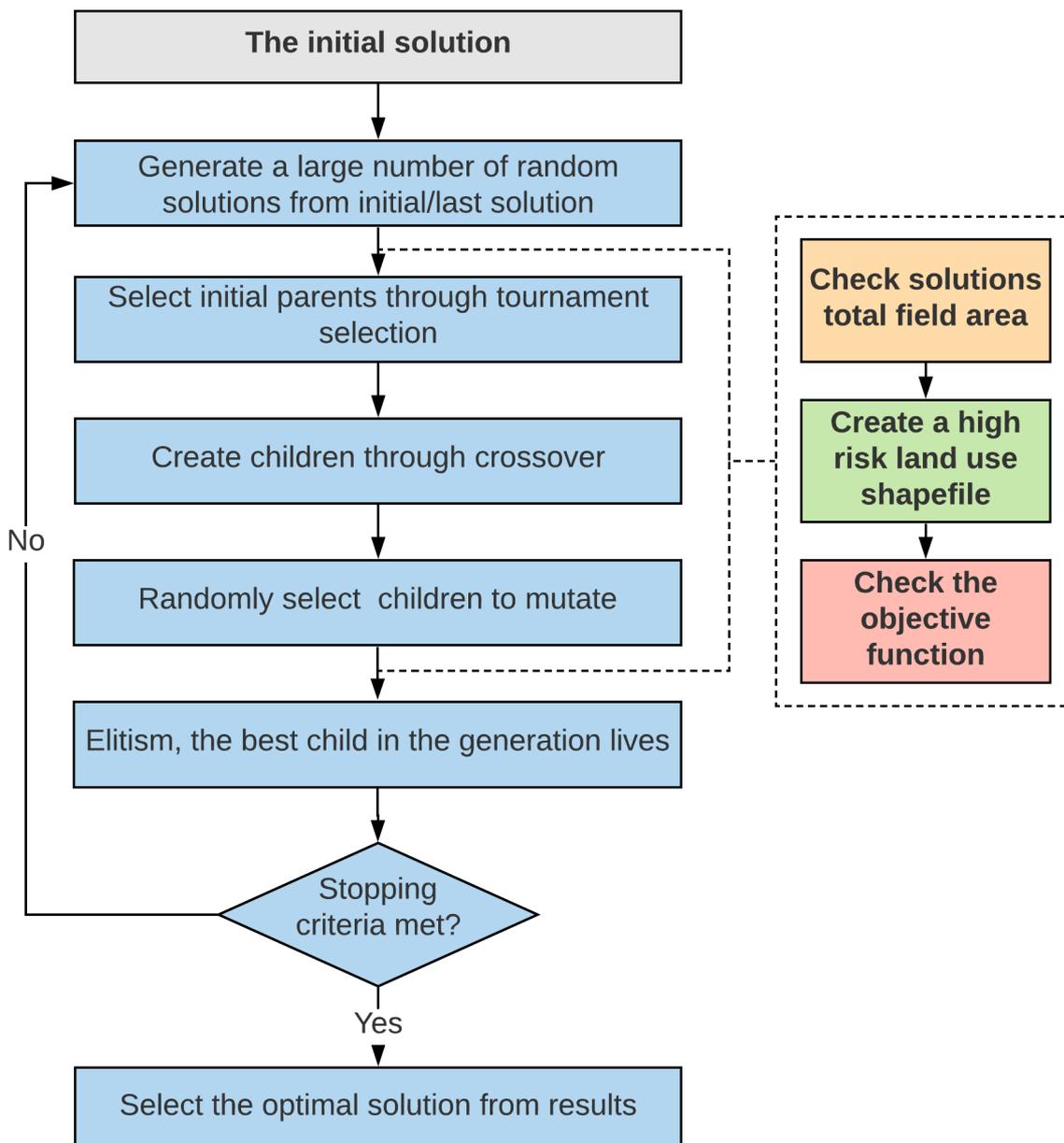


Figure 5.4 Flow chart of the GA method. The initial solution represents the current distribution of high risk land use. Checking the objective function runs the model with the new high risk shapefile and checks the resulting forecasted total hours pesticide levels are above threshold, and redefines objective function.

5.3.1 Zero-one integer programming

Combinatorial genetic algorithm problem requires an input as a list of values that can be presented in different combinations which genetic algorithm can optimise. Zero-one integer programming is used to represent fields within the catchment as present (1) or not present (0). The technique has been used in solving allocation problems where the method (or land use) is either implemented (1) or not implemented (0) (Aerts et al., 2002), (Wang et al., 2019).

To start running GA, an initial solution is created. In this case, it is a list the length of the number of fields present in the high-risk shapefile filled with 1's, with each index of the list corresponding

to a particular field in the shapefile. The model is first run with current land use shapefile (initial solution) and the outcome forms the initial objective function that GA uses to compare to its subsequent objective functions.

Since fields have a non-uniform area, maintaining the exact land use area for each iteration is not possible in this case. Therefore, GA intends to find a combination of high risk fields, which constitute a maximum of 5% removal of initial/original high risk shapefile area, that minimises the number of hours forecasted total pesticide levels exceed the threshold of $0.1 \mu\text{g L}^{-1}$. Hence, every new solution created in GA is at least 95% of total original high-risk area. In essence this simulates a 'targeted' mitigation approach in which up to 5% of the land area can be considered.

All the new solutions are a list of same length as the initial solution but where fields are selected to be removed, the 1s are replaced with 0s. When GA creates a new solution through crossover and mutation, it checks if the area selected for removal does not exceed or is below 5% of the current high risk area. A new high risk shapefile is then created. To check if the solution GA created has performed better or worse than the last one, it needs to check its objective function. First, it externally runs the model with rainfall events, associated variables and the newly created shapefiles as inputs. Once the model completes the run, it reads the total pesticide output file and counts how many hours the pesticide levels have exceeded the threshold of $0.1 \mu\text{g L}^{-1}$.

5.3.2 Rainfall event mashup

Ideally a number of years worth of spatial rainfall radar data needs to be input into the model when running GA to represent the overall rainfall patterns and intensities of the catchment, and therefore find the optimal land use distribution based on a pattern of rainfall that may be typically be expected over a pesticide application season. However, this method would be extremely time consuming due to the model operating at hourly resolution, and hence become infeasible for optimisation purposes. Therefore, a shorter dataset needs to be compiled by analysing full rainfall record, identifying and removing any gaps in rainfall and refining rainfall events to a limited number representative of the study catchment. Automation of rainfall event recognition and selection is needed. Hence a shorter, representative mashup of rainfall events must be created that will allow the identification of targeted land use interventions that are not specific to a single given rainfall event. Such an approach has been used previously for optimisation based approaches which are sensitive to temporal variation in rainfall inputs (Mounce et al. 2020). In this case, due to the nature of the rainfall runoff, it is important to retain elements that capture both the temporal and spatial distribution of rainfall within the catchment.

To achieve this, the statistical characteristics (temporal and spatial variability) of the historic catchment rainfall patterns were analysed and 'recreated' in the mashup event.

Python code was written to analyse the 1 km² spatial and 5 min temporal resolution spatial rainfall radar data from the Met Office Nimrod System (Met Office, 2003). First, every 5-minute time step data was catchment averaged to produce a single value. This produced a time series of 5-minute catchment averaged values for several years (2015-2019). As the metaldehyde and propyzamide models are mostly used for September-December months, yearly data for these months only was analysed (figure 5.5). For every 5-minute time step spatial standard deviation, defined as standard deviation of all the values within the spatial rainfall file for that 5 minute time step, was calculated to see how even/uneven spatial rainfall distribution for that time step was. It was used when considering which events to include in rainfall mashup as it provides a measure of spatial rainfall distribution which is a key process in determining pollutant concentrations in river water.

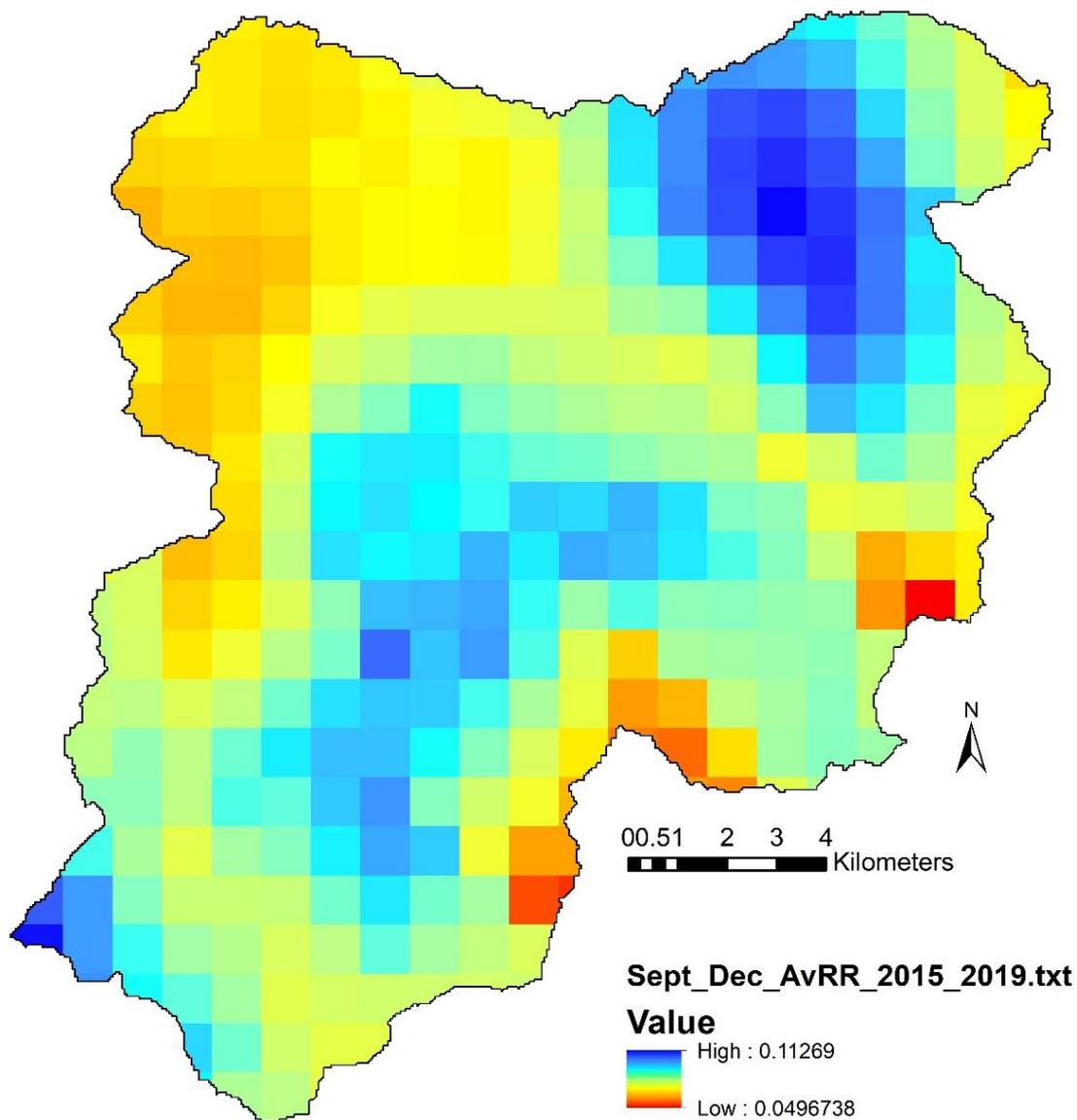


Figure 5.5 Temporally averaged rainfall (mm) from September-December 2015-2019

Code was then written to loop through the resulting time series to automate the recognition of storm events (figure 5.6). The code identifies a gap in rainfall, takes it as a start of a storm event, loops through rainfall and the next gap in rainfall is recorded as the end of a storm event. It then skips through the no rain gap to find next rainfall, when it does this is marked as the start of the next event. It assigns a storm ID to the event, records its start and end date/time, calculates its length (time in hours), total event rainfall, average spatial standard deviation, and antecedent moisture condition for 15 days prior to the start of the rainfall event (AMC15). The identified events were further refined so that each event was at least 1 hour long and have produced any increase in pesticide. Over the full time series (months September to December in the years 2015-2019) this resulted in 188 refined storm events in the catchment.

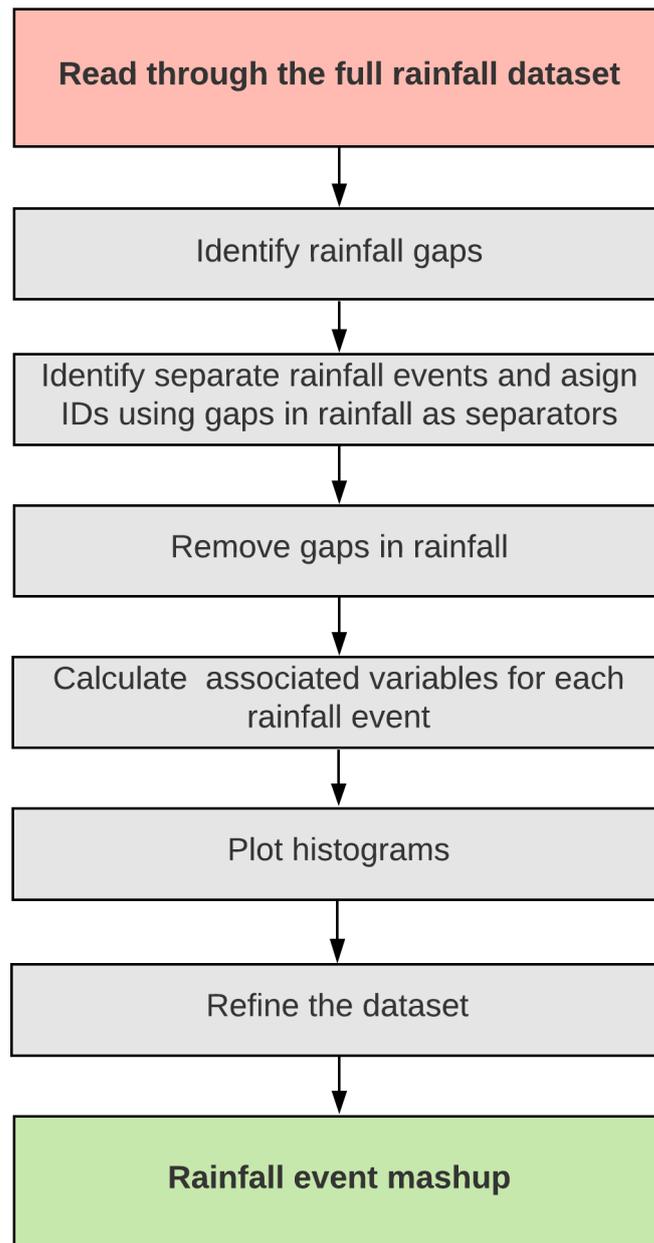


Figure 5.6 Method of rainfall event mash-up production

A multivariate stratified sampling method (Speight et al., 2004) was then used to select a subset of these events which characterised the overall temporal and spatial variability of rainfall events in the catchment. The rainfall events were assigned into strata by spatial standard deviation, each strata was then stratified by temporal standard deviation. At sub-strata level, random numbers were assigned to elements and sorted largest to smallest. Single element at the top of each sorted sub-strata was then selected. Based on this analysis sixteen events were selected as the mashup subset.

A two-sample Kolmogorov-Smirnov test was used to check if the full rainfall dataset and rainfall mashup dataset obtained through multivariate stratified sampling have the same distribution of

temporal and spatial standard deviation. If the Kolmogorov–Smirnov test statistic exceeds critical D (D_α , equation 15) the null hypothesis of both samples come from a population with the same distribution can be rejected.

$$D_\alpha = c(\alpha) \sqrt{\frac{m+n}{mn}} \tag{Equation 15}$$

Where $c(\alpha)$ is the inverse of the Kolmogorov distribution at significance level α , m is the first sample size and n is the second sample size. As test statistic was lower than critical D at $\alpha = 0.05$, the null hypothesis cannot be rejected. Therefore the two datasets can be assumed to be from the same distribution when checked both by spatial and temporal standard deviation distributions. The histograms below show the distributions for spatial standard deviation (figure 5.7) and temporal standard deviation (figure 5.8).

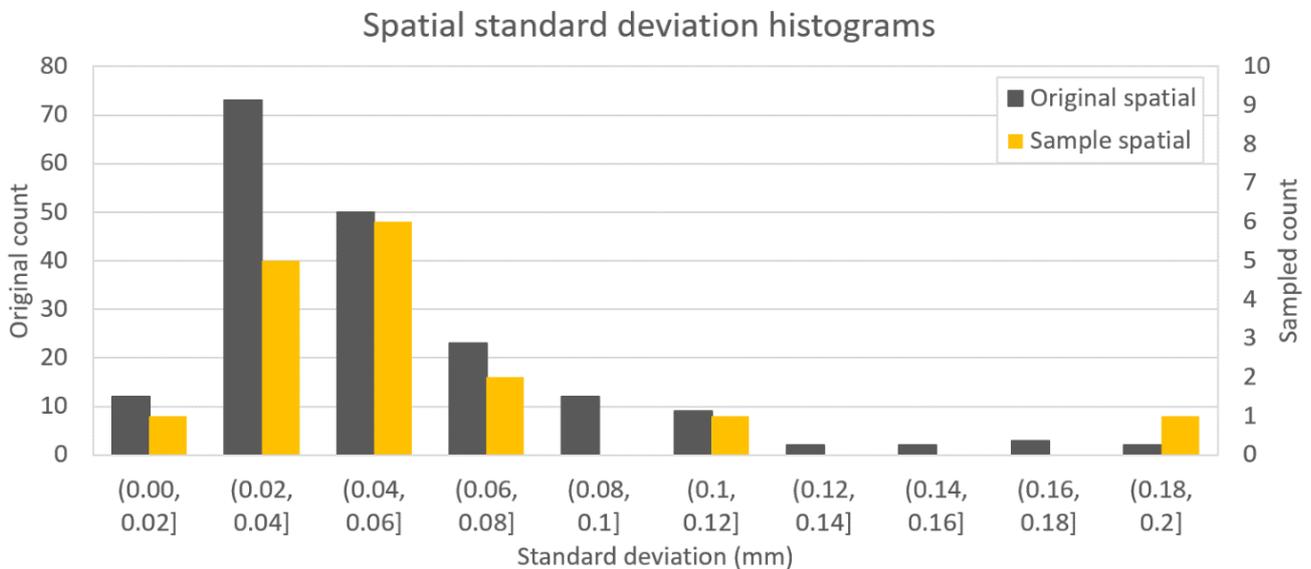


Figure 5.7 Histograms of rainfall event spatial standard deviation for full original and sampled 'mashup' datasets.

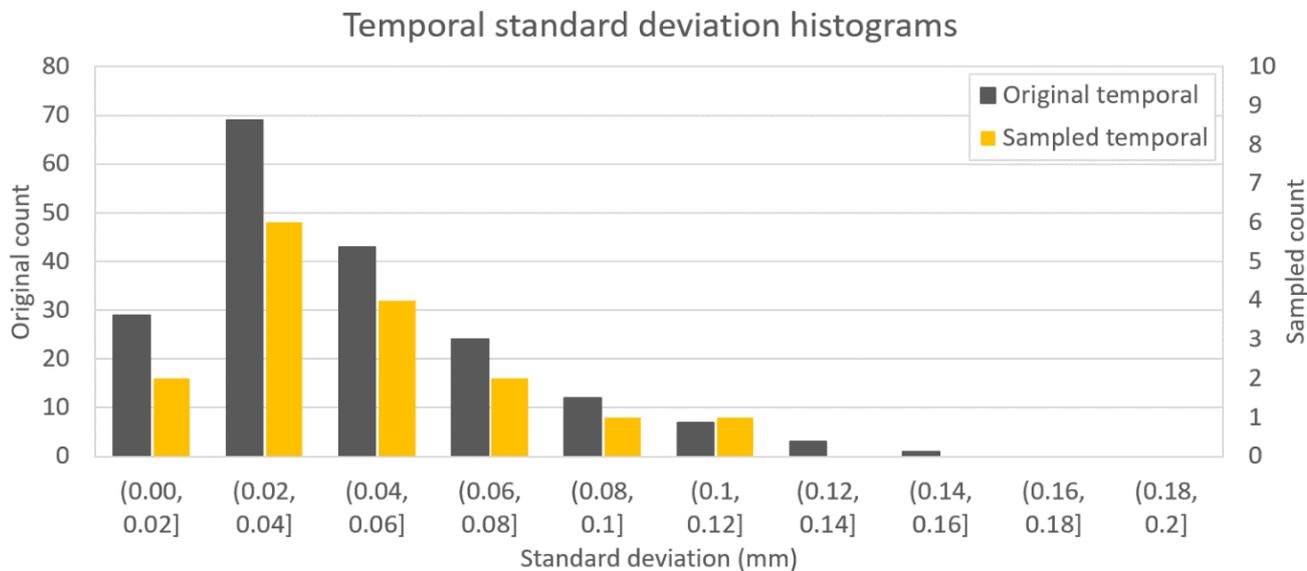


Figure 5.8 Histograms of rainfall event temporal standard deviation for full original and sampled 'mashup' datasets.

5.4 Results and discussion

5.4.1 Land-use optimisation

The optimisation algorithm for metaldehyde land use was evaluated to have reached the best solution after 3023 runs (reduction of 39 hours), propyzamide land use optimisation did so after 2887 runs (reduction of 32 hours). Each 'run' involves running the algorithm with the mashup event selected with the multivariate stratified sampling method (figure 5.9 and 5.10). Most of the objective function results for metaldehyde fall between 307 and 322 hours. For propyzamide majority of objective function results fall between 146 and 174 hours. Propyzamide exhibits much wider variation in objective function result, which may be due to it being applied to the field at much higher concentrations. Variation in the location of the high-risk fields within the catchment and temporal variation in rainfall would therefore impact the propyzamide levels in river water more than it would for metaldehyde.

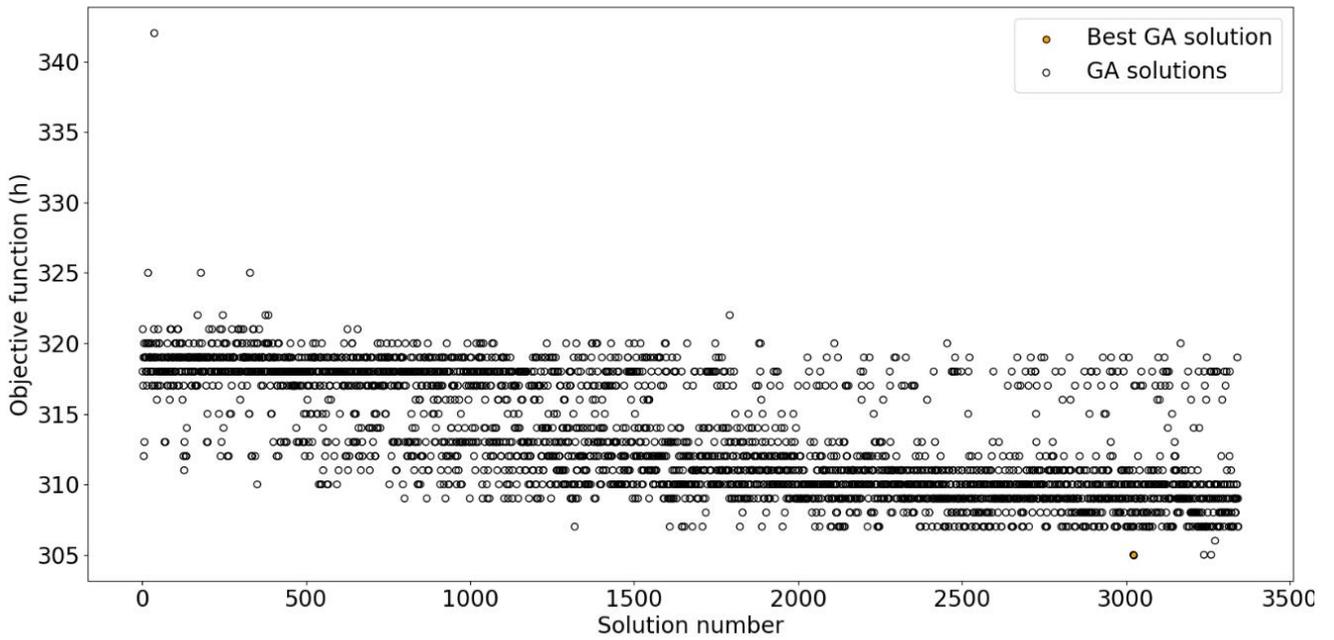


Figure 5.9 Plot of all the objective function results for metaldehyde field distribution

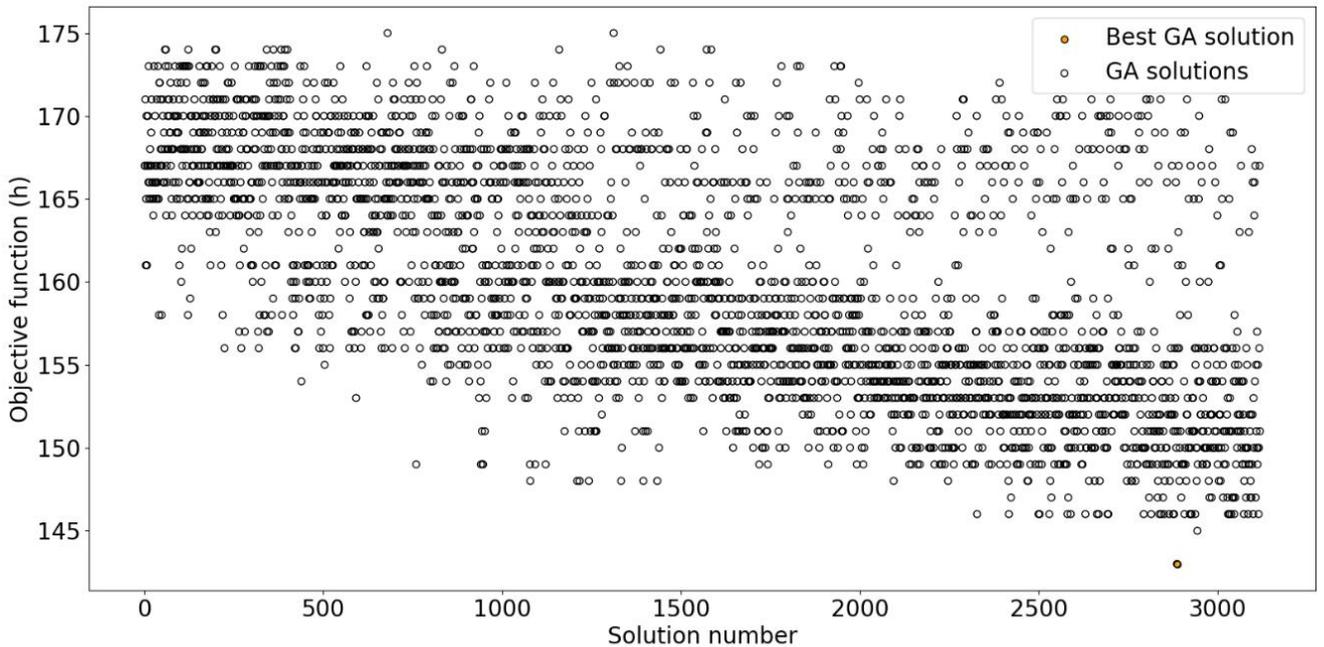


Figure 5.10 Plot of all the objective function results for propyzamide field distribution

Fields removed for initial solution as well as fields removed for the best solution for metaldehyde and propyzamide are plotted in figures 5.11 and 5.12 respectively. There is no discernible clustering or other clear pattern in the removed fields of best solutions for either metaldehyde or propyzamide. The GA may not have had sufficient time to reach a solution that would show a pattern related to variables such as rainfall patterns, distance to watercourse, etc. The lack of apparent pattern in the removed fields may prove the implementation of mitigation measures more difficult to carry out as the number of stakeholders increases. Future work should consider either extending the time GA runs by altering stopping criteria or introduce additional conditions to the algorithm such as clustering of the fields.

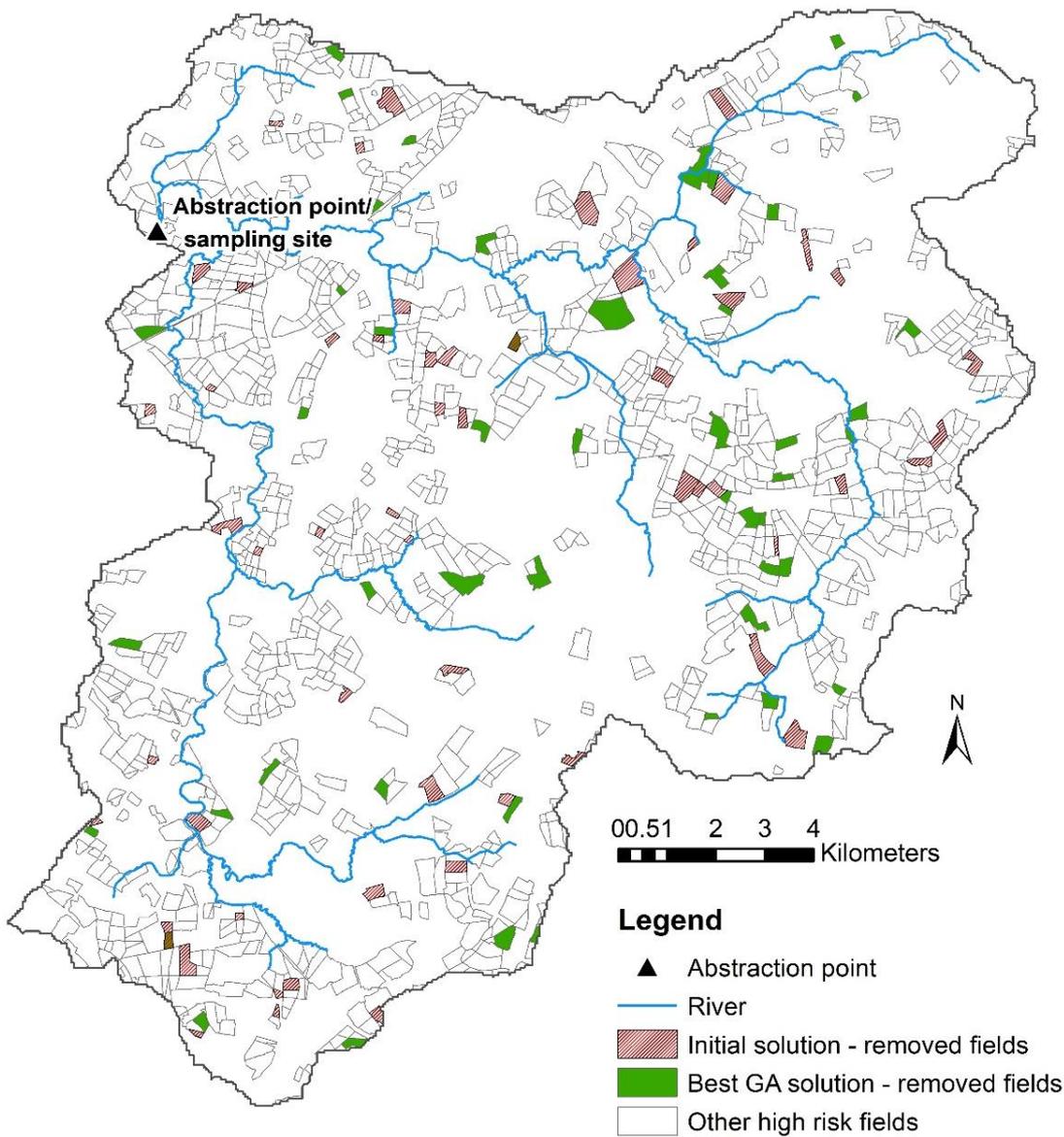


Figure 5.11 Fields removed as initial solution and resulting best genetic algorithm solution for Metaldehyde Model

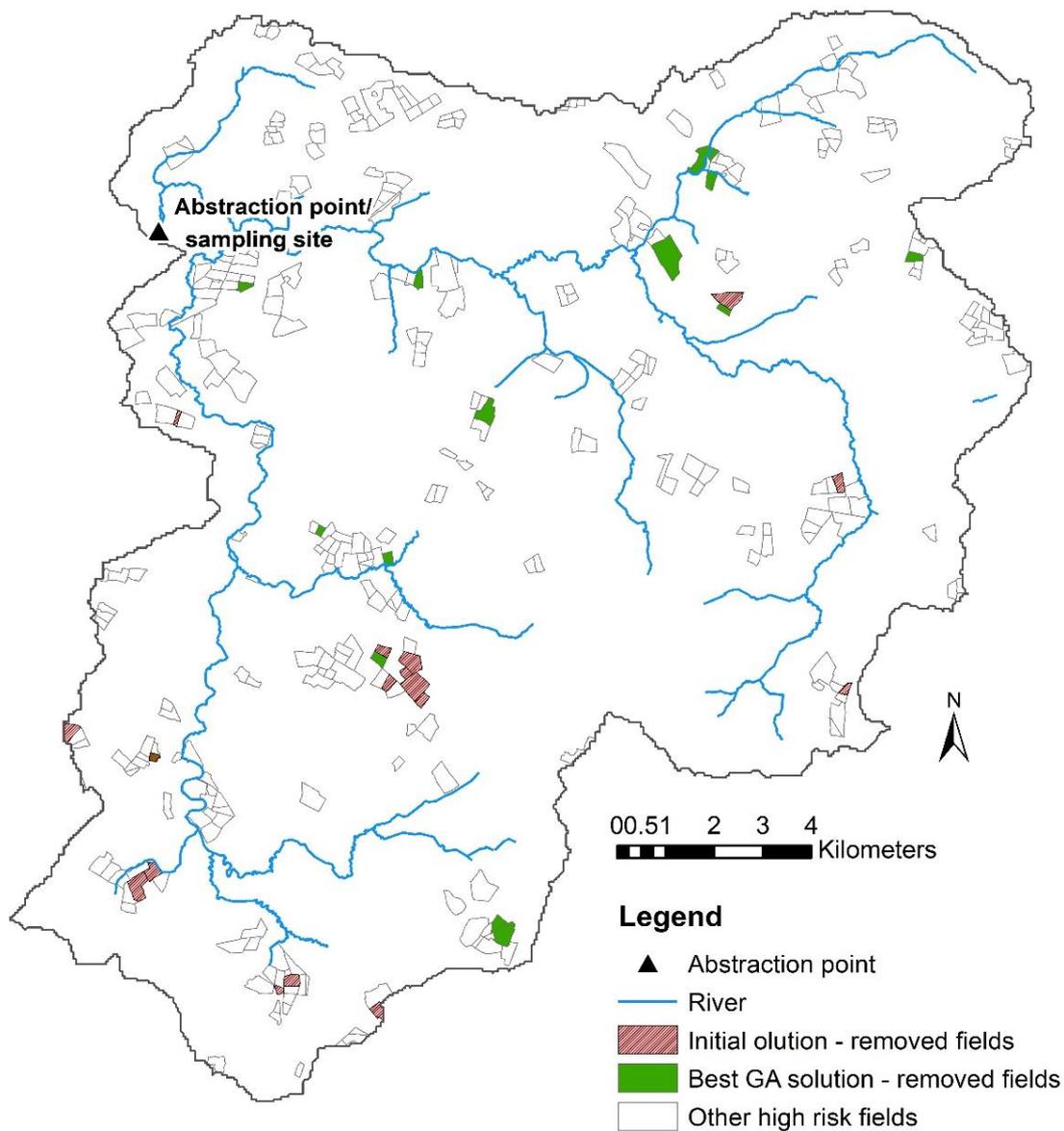


Figure 5.12 Fields removed as initial solution and resulting best genetic algorithm solution for Propyzamide Model

5.4.2 Validation under historical rainfall.

To validate the performance of the optimisation approach the best land use solutions for both pesticides were validated by running the model with all the 188 rainfall events identified for September-December 2015-2019. This was then compared to a simulation under the original land use i.e. no fields removed (Full HR) and a removal of 5% of fields by area (m^2) with based on the shortest traveltime (identified after running the hydrological model with 1mm uniform rainfall for 1 hour). The total time that pesticides remained over threshold for each of these scenarios is presented in table 5.3.

Table 5.3 Validation of optimization methodology under full dataset

Shapefile used to run all events	Metaldehyde hours-above-threshold all events	Reduction from Full HR	Propyzamide hours-above-threshold all events	Reduction from Full HR
Full HR	3248	N/A	1523	N/A
5% random HR fields removed	3157	91	1444	79
5% removed best GA solution	2997	251	1344	179
5% shortest traveltime HR fields removed	3169	79	1486	37

The original land use for metaldehyde has produced more than twice the amount of hours spend above threshold compared to propyzamide. The GA solution has reduced metaldehyde time above threshold by 251 hours or 7.7%, however propyzamide has outperformed this at 11.8%. Best solutions for both pesticides performed much better than removing an equivalent area of high-risk fields closest to the abstraction point. This indicates variables other than traveltime being of importance when trying to identify the areas for implementation of mitigation measures to reduce pesticide concentrations in river water. The GA solution has consistently reduced peak concentrations for both pesticides (figures 5.13 and 5.14). This shows the ability of the optimisation approach to reduce pesticide concentrations in river water when the impact is most severe and pesticide pollution poses highest risk to water abstracted for drinking water supply.

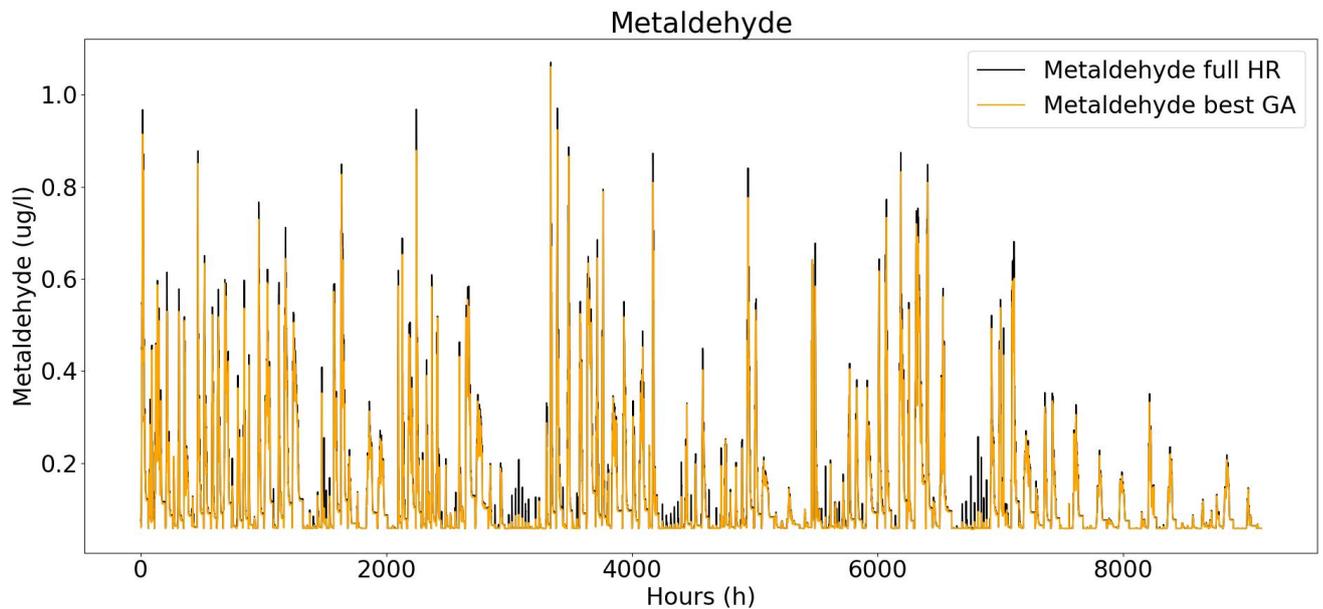


Figure 5.13 Metaldehyde model time series of all 188 events run with Full HR shapefile and 5% removed best GA solution.

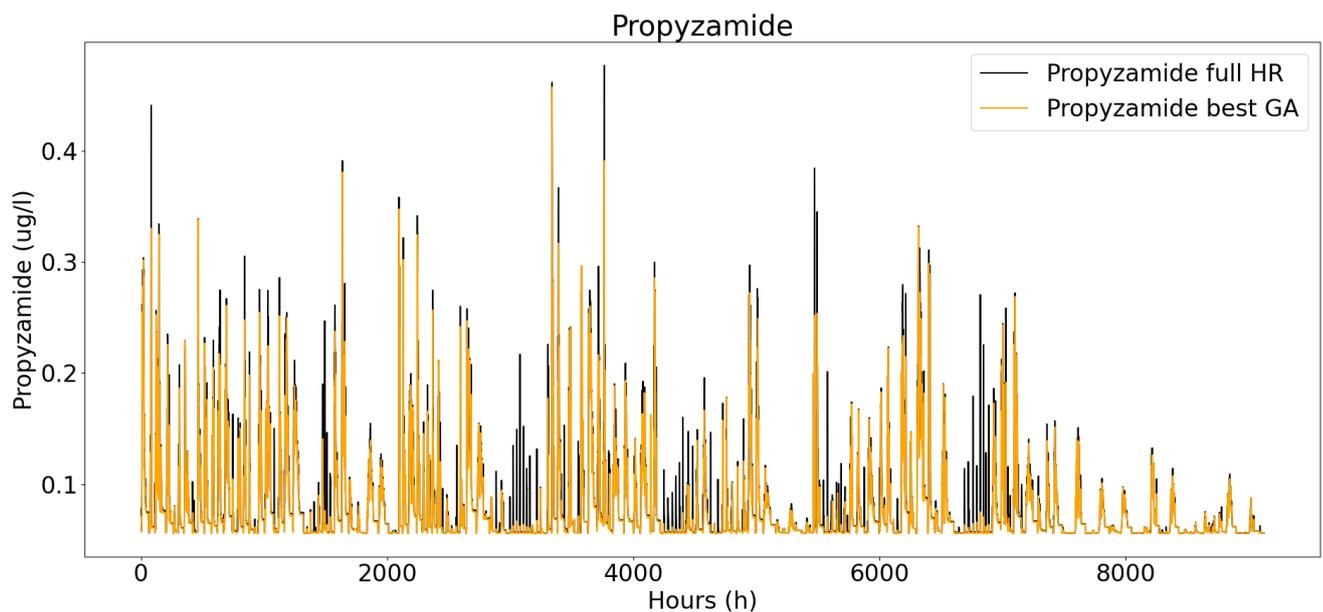


Figure 5.14 Propyzamide model time series of all 188 events run with Full HR shapefile and 5% removed best GA solution.

5.5 Conclusion

This chapter has developed a new methodology for targeting catchment mitigation options for the reduction of impacts from acute rainfall events on water abstraction systems. This is based on the simulation of two distinct pesticides of concern to STW. Whilst the metaldehyde model as already been developed and validated by previous work, there is no current equivalent model to forecast acute impacts from propyzamide. Hence the initial challenge was to develop and validate a propyzamide model. Because similar rainfall driven runoff processes were expected to be significant, the Metaldehyde existing model was therefore adapted with

adjustments to relevant application and runoff characteristics. This therefore preserved the ability of the model to account for temporal and spatial variations in rainfall over the catchment. Testing was conducted on high resolution datasets collected and analysed by the STW catchment team. In this case it is noted that the performance of the model was found to be generally inferior to the equivalent model for Metaldehyde. It is noted that this pesticide is applied over smaller areas at higher concentrations than Metaldehyde, which therefore makes the model highly sensitive to the accurate identification of application areas. Further to this there is some uncertainty as to the identification of application areas using high risk maps (as propyzamide can be applied to a wide variety of crop types). Therefore, one option to improve the model is to undertake more extensive ground or camera-based investigation of application areas. Further potential sources of error include the simplification of transport processes associated with propyzamide which are less significant to metaldehyde.

The main contribution of this chapter is the development of a land use optimisation approach for acute impacts. Whilst land use optimisation to address water quality problems is a common topic in the literature, to date such methods have focused on an evaluation of long term water quality trends, 'averaged' or representative conditions. When considering the mitigation of acute impacts a key challenge is the representation of the effects of spatial and temporal variations in rainfall within the optimisation framework. This work has proposed the selection of a subset of rainfall events based on statistical interrogation of the historic rainfall record (based on statistical properties related to the spatial and temporal variability of rainfall events). The use of this mashup within a GA optimisation algorithm was able to identify priority areas for catchment intervention which resulted in a significant reduction in pesticide risk to the water abstraction site. The approach also is feasible to be applied on a desktop computer. Further work may consider the application of different statistical rainfall characteristics for the selection of the mashup event.

The application of this approach to the full rainfall record demonstrated a significant improvement when compared to a simple selection/prioritisation of fields closed to the abstraction site. Further work may interrogate properties of the selected fields themselves to explore links such as proximity to water courses or land slope. Application of the technique to propyzamide also resulted in larger reduction in the objective function. As previously mentioned, due to the smaller number of fields under propyzamide application, the temporal dynamics within the watercourse are highly sensitive to land use distribution.

This study shows the importance of accurate identification of source areas when modelling for pesticide pollution, especially so when the importance of individual fields is exaggerated by the high application concentration. The method of optimisation can be transferred to other

catchments by using existing catchment specific models and for a range of pollutants. The water companies may use catchment interventions such as subsidies to use alternative pesticides, cultural controls, different drainage systems, etc. The use of optimization method can help obtain the most return for the money invested by replacing blanket mitigation measures with a targeted approach.

Chapter 6 – Summary and Conclusions

The overall contribution of this thesis can be summarized as follows

6.1 Real-time sensing technologies

The thesis collected and presented new high resolution datasets concerning *E. coli* dynamics in UK catchment following wet weather events. The four events were collected over a range of seasonal conditions including short intense rainfall events as well as high volume longer duration flow events. The work includes the first independent testing and validation of a commercial probe for providing warning of bacterial contamination in surface waters based on the continuous monitoring of TLF. Current results suggest that the probe has a high degree of uncertainty when compared to conventional sampling and analysis. Maintenance and calibration requirements are extensive, and results should be interpreted with care. Whilst this technology is continually under development the author would not recommend it as a current robust methodology to characterize *E. coli* loadings or provide early warning to bathing water or water abstraction sites. To improve Proteus water quality probe performance more regular calibrations may be needed as well as more validation and testing in river networks and non-potable water systems to see if different inferences can be corrected.

6.2 Forecasting *E. coli* impacts

The thesis presents a new approach to modelling acute *E. coli* impacts in drinking water catchments. The method further developed from the initial approach of Asfaw et al. (2018) for pesticide modelling, which accounts for the spatial and temporal variation in rainfall over the catchment and the distribution of pollutant sources. In this work, the model structure is significantly further developed to be applicable to *E. coli* by modelling diffuse runoff and loadings from farmland (based on livestock data) as well as including contributions from SSO. Uniquely, the approach does not utilize hydrodynamics modelling of sewer networks to understand SSO timings and loadings, instead making use of newly installed spatially disturbed SSO depth sensors. This significantly simplifies the modelling approach. Despite inherent uncertainties associated with bacterial modelling, the model provided reasonable approximations of arrival times and durations of *E. coli* at the water abstraction site following acute catchment loadings, and is therefore judged to be fit for purpose in providing useful information to abstraction operators for decision making purposes. A relatively simple two parameter calibration of the model was also found to be sufficient in this case. The model can be used by water companies to inform when to change the amount/timing of abstraction or when to change the water treatment processes. It may also be a useful tool as a warning system in bathing waters for general public.

6.3 Land use optimization

Finally, the thesis has developed an inverse modelling approach for targeting catchment intervention. In this case the methodology was applied to the targeting of two distinct pesticides. Whilst further work and development of the model for propyzamide is recommended before future use, the optimization methodology developed in this thesis was found to be effective in reducing modelled pesticide levels at the water abstraction site based on the selective targeting of mitigation options in the catchment. Unlike previous land use optimization based approaches, this methodology is specifically applicable to consider acute impacts. To enable a feasible optimization approach and novel analysis for historical rainfall was conducted to identify a shorter but representative rainfall sample. The optimization method developed here could be applied to a range of pollutant types, for example application in conjunction with the E coli model (from chapter 4) would enable targeting of both land areas for runoff mitigation, as well as problematic SSO sites.

6.4 Limitations and future work

Overall, this thesis has produced several distinct contributions concerning the development of new approaches to water abstraction management. It is anticipated that these methods will enable operators to better understand and mitigate water quality-based risk to drinking water supplies. Limitations of the work should however be acknowledged. Firstly, for reasons of time and resource only a limited number of water quality events were collected in this work. Ideally tools and methods would be developed using a greater number of events, the inherent logistical difficulty of sample collection and analysis mean that this is a common issue with bacterial water quality studies or those considering 'non-standard' water quality parameters. Further chapter 3 and 4 have (as is common) taken E coli as an indicator of bacterial water quality. It should be acknowledged that there are a number of different bacterial parameters which can cause water quality concerns (e.g. cytosporidium), however due to time and cost reasons it would be unfeasible to consider the full range. More significantly, is the limitation concerning the testing catchment. The model testing and validation would benefit from application and testing on a wider range of catchments with different land types, sizes and pollutant characteristics.

It is noted that the work in this thesis concerns a number of applications of sensor technology and novel data sources (i.e. SSO level data, satellite land use). As such it is likely that future improved sensing and monitoring technology will make possible the deployment and use of increasingly novel methodologies and technologies. There is significant future potential to link

water quality based concerns to water resource management tools and datasets such as water storage levels, predicted demand use and datasets from water distribution networks.

References

Abbasi, Y., Mannaerts, C.M., Makau, W. (2019) 'Modeling Pesticide and Sediment Transport in the Malewa River Basin (Kenya) Using SWAT', *Water*, 11(87), pp. 1-20, doi:10.3390/w11010087

ADAMA (2015) 'SAFETY DATA SHEET', https://www.agrigem.co.uk/documents/Cohort%20SDS_tcm105-43679.pdf (Accessed at 16.07.2023)

Aerts, J. C. J. H., Eisinger, E., Heuvelink, G. B. M. and Stewart T. J. (2002) 'Using Linear Integer Programming for Multi-Site Land-Use Allocation', *Geographical Analysis*, 35(2), pp. 148-169.

Arabi, M., Govindaraju, R. S. and Hantush, M. M. (2006) 'Cost-effective allocation of watershed management practices using a genetic algorithm', *Water Resources Research*, 42, pp. 1-14.

Asfaw, A. , Maher, K., Shucksmith, J.D. (2018) 'Modelling of metaldehyde concentrations in surface waters: A travel time based approach', *Journal of Hydrology*, 562, pp. 397-410.

Avery, S.M., Moore, A., Hutchison, M.L., (2004) 'Fate of Escherichia coli originating from livestock faeces deposited directly onto pasture' *Lett. Appl. Microbiol.* 3, pp.355–359.

Baker, A., Cumberland, S. A., Bradley, C., Buckley C., Bridgeman, J. (2015), 'To what extent can portable fluorescence spectroscopy be used in the real-time assessment of microbial water quality?' , *Science of the Total Environment*, 532, 14-19.

Bathing Water Regulations (2013) <https://www.legislation.gov.uk/uksi/2013/1675/contents/made> (Accessed at 08.06.2023).

Beer, T and Young, P.C. (1984) 'Longitudinal Dispersion in Natural Streams' *Journal of Environmental Engineering*, 109 (5), pp. 1049-1069.

Benaman, J., Shoemaker, C.A., Haith, D.A. (2005) 'Calibration and Validation of Soil and Water Assessment Tool on an Agricultural Watershed in Upstate New York', *Journal of Hydrologic Engineering*, 10(5), pp.363–374.

Berenzen, N., et al., (2005) 'A comparison of predicted and measured levels of runoff related pesticide concentrations in small lowland streams on a landscape level', *Chemosphere*, 58 (5), pp. 683–691.

Bingner, R.L. & Theurer, F.D. (2005) 'AnnAGNPS TECHNICAL PROCESSES', Gaithersburg, MD.

Borah, D.K., Xia, R. J.; Bera, M. (2002) 'DWSM - a Dynamic Watershed Simulation Model' In *Mathematical models of small watershed hydrology and applications*. Water Resources Publications, pp. 113–166.

Boxall, J., Court, E, Speight, V. (2021) 'Real-time monitoring of bacteria at water treatment works and in downstream networks', *UK Water Industry Research*, London, Webree.com Ltd. (<https://ukwir.org/water-research-reports-publications-viewer/b972bc73-ee22-40a0-8dfc-5ad4900d42b1>)

- Bragina, L., Sherlock, O., Rossum, A.J., Jennings, E., (2017) 'Cattle exclusion using fencing reduces *Escherichia coli* (*E. coli*) level in stream sediment reservoirs in northeast Ireland', *Agriculture, Ecosystems and Environment*, 239, pp. 349–358.
- Brannan, K.M., Mostaghimi, S., Dillaha, T.A., Heatwole, C.D., Wolfe, M.L. (2002) 'Fecal Coliform TMDL for Big Otter River, Virginia : a case study', *Total Maximum Daily Load (TMDL) Environmental Regulations: Proceedings of the March 11-13, 2002 Conference*, (Fort Worth, Texas, USA) 701P0102, pp. 367-376.
- ¹Camacho Suarez, V. V., Schellart, A. N. A., Brevis, W., & Shucksmith, J. D. (2019). Quantifying the impact of uncertainty within the longitudinal dispersion coefficient on concentration dynamics and regulatory compliance in rivers. *Water Resources Research*, 55, 4393-4409. <https://doi.org/10.1029/2018WR023417>
- ²Camacho Suarez, V.V , Brederveld, R.J., Fennema, M., Moreno-Rodenas, A., Langeveld, J., Korving, H., Schellart, A.N.A., Shucksmith, J.D. (2019) 'Evaluation of a coupled hydrodynamic-closed ecological cycle approach for modelling dissolved oxygen in surface waters', *Environmental Modelling and Software*, 119, pp. 242–257.
- Cambien, N., Gobeyn, S., Nolivos, I., Forio M.A.E., Arias-Hidalgo, M., Dominguez-Granda, L., Witing, F., Volk, M., Goethals, P.L.M. (2020); 'Using the Soil and Water Assessment Tool to Simulate the Pesticide Dynamics in the Data Scarce Guayas River Basin, Ecuador', *Water*, 12(696), doi:10.3390/w12030696
- Cammack, W.K.L., Kalff, J., Prairie, Y.T., Smith, E.M. (2004) 'Fluorescent dissolved organic matter in lakes: relationships with heterotrophic metabolism', *American Society of Limnology and Oceanography*, 49(6), pp. 2034-2045.
- Cardenas, S., Márquez, A., Guevara, E. (2023) 'Diffusion–advection process modeling of organochlorine pesticides in rivers, *Journal of Applied Water Engineering and Research*, 11(1), pp. 1-22, <https://doi.org/10.1080/23249676.2021.1982029>
- Cárdenas-Izaguirre, S. F., Márquez-Romance, A. M., Guevara-Pérez, E., & Pérez-Pacheco, S. A. (2022) 'An approach to models for transport and transformation of organochlorine pesticides in rivers', *Environmental Quality Management*, 31(3), 369–391, <https://doi.org/10.1002/tqem.21791>
- CEH (2023) 'UKCEH Land Cover® plus: Crops', <https://www.ceh.ac.uk/data/ceh-land-cover-plus-crops-2015#specs> (Accessed at 14.05.2023)
- Chalmers, R.M., Robinson, G., Elwin, K., Elson, R. (2019) 'Analysis of the *Cryptosporidium* spp. and gp60 subtypes linked to human outbreaks of cryptosporidiosis in England and Wales, 2009 to 2017', *Parasites Vectors*, 12, 95. <https://doi.org/10.1186/s13071-019-3354-6>
- Cho, K.H., Pachepsky, Y.A., Oliver, D.M., Muirhead, R.W., Park, Y., Quilliam, R.S. and Shelton, D. R. (2016) 'Modelling fate and transport of fecally-derived microorganisms at the watershed scale, *Water Research*, 100, pp. 38–56.
- Claver, A., Ormad, P., Rodriguez, L., Ovelleiro, J.L., (2006) 'Study of the presence of pesticides in surface waters in the Ebro river basin (Spain)', *Chemosphere*, 64, pp. 1437–1443.
- Coffey, R., Benham, B., Wolfe, M.L., Dorai-Raj, S., Bhreathnach, N., O'Flaherty, V., Cormican, M., Cummins, E. (2016) 'Sensitivity of streamflow and microbial water quality to

future climate and land use change in the West of Ireland', *Reg. Environ. Chang.*, 16, pp. 2111–2128.

Collier, S. A., Deng, L., Adam, E. A., Benedict, K. M., Beshearse, E. M., Blackstock, A. J., Beach, M. J. (2021). Estimate of Burden and Direct Healthcare Cost of Infectious Waterborne Disease in the United States. *Emerging Infectious Diseases*, 27(1), 140-149. <https://doi.org/10.3201/eid2701.190676>.

Collins, R., Rutherford, K. (2004) 'Modelling bacterial water quality in streams draining pastoral land', *Water Research*, 38 (3), pp. 700-712.

Cooke, A., Rettino, J., Flower, L., Filby, K. and Freer, A. (2020) 'Farming for water; catchment management initiatives for reducing pesticides', *Water and Environment Journal*, 34, pp. 679-691. <https://doi-org.sheffield.idm.oclc.org/10.1111/wej.12609>

¹Corteva, 'Kerb® Flo 500 Label', <https://www.corteva.co.uk/products-and-solutions/crop-protection/kerb-flo-500.html> (Accessed at 16.07.2023)

²Corteva, 'Kerb® Flo Label', <https://www.corteva.co.uk/products-and-solutions/crop-protection/kerb-flo.html> (Accessed at 16.07.2023)

Cumberland, S., Bridgeman, J., Baker, A., Sterling, M., Ward, D., 2012. Fluorescence spectroscopy as a tool for determining microbial quality in potable water applications. *Environ. Technol.* 33, 687–693.

Cui, H., Guo, P., Li, M., Guo, S., Zhang, F. (2019) 'A multi-risk assessment framework for agricultural land use optimization', *Stochastic Environmental Research and Risk Assessment*, 33, pp. 563–579, <https://doi.org/10.1007/s00477-018-1610-5>

Cyterski, M., Shanks, O.C., Wanjugi, P., McMinn, B., Korajkic, A., Oshima, K., Haugland, R. (2022) 'Bacterial and viral fecal indicator predictive modeling at three Great Lakes recreational beach sites', *Water Research*, 223, 118970, <https://doi.org/10.1016/j.watres.2022.118970>.

Damalas, C.A. and Eleftherohorinos, I.G. (2011) 'Pesticide Exposure, Safety Issues, and Risk Assessment Indicators', *Int. J. Environ. Res. Public Health*, 8, pp. 1402-1419. doi:10.3390/ijerph8051402

DEFRA (2022) 'Structure of the agricultural industry in England and the UK at June – County', Statistical data set, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1084972/structure-england-june21-county-23jun22.ods (Accessed at 11.04.2023).

Delahaye, E., Welte, B., Levi, Y., Leblon, G., Montiel, A. (2003) 'An ATP-based method for monitoring the microbiological drinking water quality in a distribution network', *Water Research*, 37(15), pp. 3689-3696

Deleegn, S. W., Pathirana, A., Gersonius, B., Adeogun, A. G., Vairavamoorthy, V.(2011) 'Multi-objective optimisation of cost-benefit of urban flood management using a 1D2D coupled model', *Water Sci. Technol.* 63 (5), pp. 1053–1059. <https://doi.org/10.2166/wst.2011.290>.

De Vera, G.A., Wert, E.C. (2019) 'Using discrete and online ATP measurements to evaluate regrowth potential following ozonation and (non)biological drinking water treatment', *Water Research*, 154, pp. 377-386

Dienus, O., Sokolova, E., Nystrom, F., Matussek, A., Lofgren, S., Blom, L., Pettersson, T.J.R., Lindgren, P.-E. (2016) 'Norovirus dynamics in wastewater discharges and in the recipient drinking water source: Long-term monitoring and hydrodynamic modeling', *Environ. Sci. Technol.*, 50, pp. 10851–10858.

Dogan, F.N. and Karpuzcu, M.E. (2023) 'Modeling fate and transport of pesticides from dryland agriculture using SWAT model', *Journal of Environmental Management*, 117457, <https://doi.org/10.1016/j.jenvman.2023.117457>

Donigian Jr, A. S., Bicknell, B. R., Imhoff, J. C. (1995) 'Hydrological Simulation Program-Fortran (HSPF)' *Computer models of watershed hydrology.*, pp. 395-442.

Dorner, S.M., Anderson, W.B., Slawson, R.M., Kouwen, N., Huck, P.M. (2006) 'Hydrologic modeling of pathogen fate and transport', *Environ. Sci. Technol.*, 40 (15), pp. 4746-4753.

Du, J., Xie, H., Hu, Y., Xu, Y., Xu, C-Y. (2009) 'Development and testing of a new storm runoff routing approach based on time variant spatially distributed travel time method' *J. Hydrol.*, 369 (1–2), pp. 44-54.

Drummond, J. D., Aquino, T., Davies-Colley, R. J., Stott, R., & Krause, S. (2022) 'Modeling contaminant microbes in rivers during both baseflow and stormflow', *Geophysical Research Letters*, 49, e2021GL096514, <https://doi.org/10.1029/2021GL096514>.

DWI (2020) 'Guidance on implementing the Water Supply (Water Quality) Regulations 2016 (as amended) in England and the Water Supply (Water Quality) Regulations in (Wales) 2018', Part 4 Monitoring of water supplies, <http://www.dwi.gov.uk/stakeholders/+guidanceand-codes-of-practice/wswq/index.html> (Accessed at 15.06.2020).

Dwivedi, D., Mohanty, B.P., Lesikar B.J. (2013) 'Estimating Escherichia coli loads in streams based on various physical, chemical, and biological factors', *Water Resources Research*, 49, pp. 2896–2906.

Ellis, J.B., Yu, W., (1995) 'Bacteriology of urban runoff: the combined sewer as a bacterial reactor and generator', *Wat.Sci.Tech*, 31(7), pp. 303-310.

Environment Act (2021) <https://www.legislation.gov.uk/ukpga/2021/30/contents/enacted> (Accessed at 15.06.2020).

Eulogi M, Ostojin S, Skipworth P, Kroll S, Shucksmith JD, Schellart A. (2022) 'Optimal Positioning of RTC Actuators and SuDS for Sewer Overflow Mitigation in Urban Drainage Systems', *Water*, 14(23):3839. <https://doi.org/10.3390/w14233839>

Fach, S., Sitzenfrei, R., Rauch, W., (2008) 'Assessing the relationship between water level and combined sewer overflow with computational fluid dynamics', *Water Sci Technol*, 60 (12), pp. 3035–3043.

FAO and IWMI, (2018) 'More people, more food, worse water? a global review of water pollution from agriculture', Edited by Javier Mateo-Sagasta (IWMI), Sara Marjani Zadeh (FAO) and Hugh Turrall, ISBN 978-92-5-130729-8 (FAO), Published by the Food and Agriculture Organization of the United Nations, and the International Water Management.

Ferguson, C. and Kay, D. (2012) '5. Transport of Microbial pollution in catchment systems' in *World Health Organization (WHO)*, Edited by Al Dufour, Jamie Bartram, Robert Bos and Victor Gannon, *Animal Waste, Water Quality and Human Health* ISBN: 9781780401232. Published by IWA Publishing, London, UK, pp. 157-193.

- Ferguson, C.M., Croke, B.F.W., Beatson, P.J., Ashbolt, N.J., Deere, D.A. (2007) 'Development of a process-based model to predict pathogen budgets for the Sydney drinking water catchment', *J. Water Health*, 5 (2), pp. 187-208.
- FOCUS (2000) 'FOCUS groundwater scenarios in the EU review of active substances', Report of the FOCUS Groundwater Scenarios Workgroup, EC Document Reference Sanco/321/2000 rev.2, 202pp.
- FOCUS (2001) 'FOCUS Surface Water Scenarios in the EU Evaluation Process under 91/414/EEC' Report of the FOCUS Working Group on Surface Water Scenarios EC Document Reference SANCO/4802/2001-rev2, 245 pp.
- Fox, B.G., Thorn, R.M.S., Dutta, T.K., Bowes, M.J., Read, D.S., Reynolds, D.M. (2022) 'A case study: The deployment of a novel in situ fluorimeter for monitoring biological contamination within the urban surface waters of Kolkata, India', *Science of the Total Environment*, 842, pp. 1-10.
- FWR (2019) 'Urban Pollution Manual', 3rd Edition, <http://www.fwr.org/UPM3/>.
- Gao, G.H., Falconer, R.A. Lin, B.L. (2015) 'Modelling the fate and transport of faecal bacteria in estuarine and coastal waters', *Mar. Pollut. Bull.*, 100, pp. 162-168.
- García-García, L.M., Campos, C.J.A., Kershaw, S., Younger, A., Bacon, J., (2021) 'Scenarios of intermittent *E. coli* contamination from sewer overflows to shellfish growing waters: The Dart Estuary case study', *Marine Pollution Bulletin*, 167, pp. 1-17.
- Geng R. and Sharpley, A.N. (2019) 'A novel spatial optimization model for achieve the trade-offs placement of best management practices for agricultural non-point source pollution control at multi-spatial scales', *Journal of Cleaner Production*, 234, pp. 1023-1032, <https://doi.org/10.1016/j.jclepro.2019.06.277>
- Ghimire, B. and Deng, Z. (2013) Hydrograph-based approach to modeling bacterial fate and transport in rivers', *Water Research*, 47(3), pp. 1329-1343.
- Giakoumis, T., Voulvoulis, N., (2023) 'Combined sewer overflows: relating event duration monitoring data to wastewater systems' capacity in England', *Environ. Sci.: Water Res. Technol.*, 9, pp. 707-722.
- Gunter, H., Bradley, C., Hannah, D. M., Manaseki-Holland, S., Stevens, R., Khamis, K. (2022) 'Advances in quantifying microbial contamination in potable water: Potential of fluorescence-based sensor technology', *WIREs Water*, e1622, pp. 1–19.
- Graydon, R.C., Mezzacapo, M., Boehme, J., Foldy, S., Edge, T.A., Brubacher, J., Chan, H.M., Dellinger, M., Faustman, E.M., Rose, J.B., Takaro; T.K. (2022) 'Associations between extreme precipitation, drinking water, and protozoan acute gastrointestinal illnesses in four North American Great Lakes cities (2009–2014)', *J Water Health*, 20 (5), pp. 849–862. doi: <https://doi.org/10.2166/wh.2022.018>.
- Guymer, I. (2002) 'A National Database of Travel Time, Dispersion and Methodologies for the Protection of River Abstractions', Environment Agency, Bristol, ISBN : 1 85705 821 6.
- Harmel, R.D., Hathaway, J.M., Wagner, K.L., Wolfe, J.E., Karthikeyan, R., Francesconi, W., McCarthy, D.T. (2016) 'Uncertainty in monitoring *E. coli* concentrations in streams and stormwater runoff', *Journal of Hydrology*, 534, pp. 524–533.

- Hammes, F., Goldschmidt, F., Vital, M., Wang, Y., Egli, T. (2010) 'Measurement and interpretation of microbial adenosine triphosphate (ATP) in aquatic environments', *Water Research*, 44(13), pp. 3915-3923.
- Hashemi Aslani, Z., Nasiri, V., Maftei, C., Vaseashta, A. (2023) 'Synergetic Integration of SWAT and Multi-Objective Optimization Algorithms for Evaluating Efficiencies of Agricultural Best Management Practices to Improve Water Quality', *Land*, 12, 401, <https://doi.org/10.3390/land12020401>
- Haydon, S., Deletic, A. (2006) 'Development of a coupled pathogen-hydrologic catchment model', *J. Hydrol.*, 328 (3-4), pp. 467-480.
- Hellweger, F.L. and Masopust, P. (2008) 'INVESTIGATING THE FATE AND TRANSPORT OF ESCHERICHIA COLI IN THE CHARLES RIVER, BOSTON, USING HIGH-RESOLUTION OBSERVATION AND MODELING', *Journal of the American Water Resources Association*, 44(2), pp. 509-522.
- Henderson, R.K., Baker, A., Murphy, K.R., Hambly, A., Stuetz, R.M., Khan, S.J. (2009) 'Fluorescence as a potential monitoring tool for recycled water systems: a review', *Water Research*, 43(4), pp. 863-881.
- Herrig, I., Seis, W., Fischer, H., Regnery, J., Manz, W., Reifferscheid, G., Böer, S. (2019) 'Prediction of fecal indicator organism concentrations in rivers: the shifting role of environmental factors under varying flow conditions', *Environ Sci Eur* 31, 59 <https://doi.org/10.1186/s12302-019-0250-9>
- Hubbart, J.A., Kellner, E., Petersen, F.A. (2022) '22-Site Comparison of Land-Use Practices, E-coli and Enterococci Concentrations', *International Journal of Environmental Research and Public Health*, 19(21):13907, <https://doi.org/10.3390/ijerph192113907>.
- Hudson, N., Baker, A., Ward, D., Reynolds, D.M., Brunson, C., Carliell-Marquet, C., Browning, S. (2008) 'Can fluorescence spectrometry be used as a surrogate for the biochemical oxygen demand (BOD) test in water quality assessment? An example from the south west England', *Science of the Total Environment*, 391(1), pp. 149-158.
- Jalliffier-Verne, I., Leconte, R., Huaranga-Alvarez, U., Heniche, M., Madoux-Humery, A.-S., Autixier, L., Galarneau, M., Servais, P., Prévost, M., Dorner, S. (2017) 'Modelling the impacts of global change on concentrations of Escherichia coli in an urban river', *Adv. Water Resour.*, 108, pp. 450-460.
- Janney, P., and Jenkins, J., (2022) 'Passive sampling and ecohydrologic modeling to investigate pesticide surface water loading in the Zollner Creek watershed, Oregon, USA', *Science of the Total Environment*, 819(152955), <http://dx.doi.org/10.1016/j.scitotenv.2022.152955>.
- Jovanovic, D., Hathaway, J., Coleman, R., Deletic, A., McCarthy, D. (2017) 'Conceptual modelling of *E. coli* in urban stormwater drains, creeks and rivers', *Journal of Hydrology*, 555, pp. 129-140.
- Kammouna, R., McQuaid, N., Lessard, V., Prévost, M., Bichai, F., Dorner, S., (2023) 'Comparative study of deterministic and probabilistic assessments of microbial risk associated with combined sewer overflows upstream of drinking water intakes', *Environmental Challenges* 12(100735), pp. 1-12.

- Kay, D., Aitken, M., Crowther, J., Dickson, I., Edwards, A.C., Francis, C., Hopkins, M., Jeffrey, W., Kay, C., McDonald, A.T., McDonald, D., Stapleton, C.M, Watkins, J, Wilkinson, J., Wyer, M.D, (2007) 'Reducing fluxes of faecal indicator compliance parameters to bathing waters from diffuse agricultural sources: the Brighthouse Bay study, Scotland', *Environ. Pollut.* 147, pp. 138–149.
- Kay, P. & Grayson, R. (2014) 'Using water industry data to assess the metaldehyde pollution problem', *Water and Environment Journal*, 28(3), pp.410–417.
- Khamis, K., Sorensen, J. P. R., Bradley, C., Hannah, D. M., Lapworth, D. J., Stevens, R. (2015) 'In situ tryptophan-like fluorometers: assessing turbidity and temperature effects for freshwater applications', *Environ. Sci.: Processes Impacts*, 17, pp. 740-752.
- Khamis, K., Bradley, C., Stevens, R., Hannah, D. M. (2016) 'Continuous field estimation of dissolved organic carbon concentration and biochemical oxygen demand using dual-wavelength fluorescence, turbidity and temperature', *Hydrological Processes*, 31, pp. 540–555.
- Khamis, K., Bradley, C., Gunter, H. J., Basevi, G., Stevens, R., Hannah, D. M. (2021) 'Calibration of an in-situ fluorescence-based sensor platform for reliable BOD5 measurement in wastewater'. *Water Science and Technology*, 83.12, pp. 3075-3091.
- Kyriitsakas G, Boxall J., Speight V. (2023) 'Forecasting bacteriological presence in treated drinking water using machine learning', *Front. Water*, 5:1199632. doi: 10.3389/frwa.2023.1199632
- Larsbo, M. & Jarvis, N. (2003) 'MACRO 5.0 : a model of water flow and solute transport in macroporous soil : technical description', Department of Soil Sciences, Swedish Univ. of Ag
- Li, C., Wu Y.L., Yang, T., Zhang, Y. (2010) 'Determination of Metaldehyde in Water by SPE and UPLC – MS – MS', *Chromatographia*, 72(9-10), pp.987–991.
- Li, M., Li, J., Singhd, V.P., Fua, Q., Liu, D., Yange, G. (2019) 'Efficient allocation of agricultural land and water resources for soil environment protection using a mixed optimization-simulation approach under uncertainty', *Geoderma*, 353, pp. 55-69, <https://doi.org/10.1016/j.geoderma.2019.06.023>
- Li, M., Zhou, Y., Wang, Y., Singh, V.P., Li, Z., Li, Y. (2020) 'An ecological footprint approach for cropland use sustainability based on multi-objective optimization modelling, *Journal of Environmental Management*, 273, 111147, <https://doi.org/10.1016/j.jenvman.2020.111147>
- Liu, Y., Guo, T., Wang, R., Engel, B.A., Flanagan, D.C., Li, S., Pijanowski, B.C., Collingsworth, P.D., Lee, J.G., Wallace, C.W. (2019) 'A SWAT-based optimization tool for obtaining cost-effective strategies for agricultural conservation practice implementation at watershed scales', *Science of The Total Environment*, 691, pp. 685-696, <https://doi.org/10.1016/j.scitotenv.2019.07.175>.
- Lockmiller, K. A., Wang, K., Fike, D.A., Shaughnessy, A.R., Hasenmueller, E.A. (2019) 'Using multiple tracers (F-, B, δ11B, and optical brighteners) to distinguish between municipal drinking water and wastewater inputs to urban streams', *Science of The Total Environment*, Volume 671, pp. 1245-1256, <https://doi.org/10.1016/j.scitotenv.2019.03.352>.
- Lundqvist, J., von Brömssen, C., Rosenmai, A.K., Ohlsson, A., Le Godec, T., Jonsson, O., Kreuger, J., Oskarsson, A. (2019) 'Assessment of pesticides in surface water samples from

Swedish agricultural areas by integrated bioanalysis and chemical analysis'. *Environ Sci Eur*, 31, 53. <https://doi.org/10.1186/s12302-019-0241-x>

Madoux-Humery, A., Dorner, S., Se´bastien, S., Aboulfadl, K., Galarneau, M., Servais, P., Pre´vost, M. (2013) 'Temporal variability of combined sewer overflow contaminants: Evaluation of wastewater micropollutants as tracers of fecal contamination', *Water Research*, 47, pp. 4370-4382.

Madoux-Humery, A.S., Dorner, S.M., Sauve, S., Aboulfadl, K., Galarneau, M., Servais, P., Prevost, M. (2015) 'Temporal analysis of *E. coli*, TSS and wastewater micropollutant loads from combined sewer overflows: implications for management', *Environ. Sci. Process. Impacts*, 17 (5), pp. 965-974.

Madoux-Humery, A-S., Dorner, S., Se´bastien, S., Aboulfadl, K., Galarneau, M., Servais, P., Pre´vost, M. (2016) 'The effects of combined sewer overflow events on riverine sources of drinking water', *Water Research*, 92, pp. 218-227.

Magana-Arachchi, D.N., Wanigatunge, R.P. (2020) 'Chapter 2 - Ubiquitous waterborne pathogens', Editor(s): Majeti Narasimha Vara Prasad, Anna Grobelak, *Waterborne Pathogens*, Butterworth-Heinemann, pp. 15-42, <https://doi.org/10.1016/B978-0-12-818783-8.00002-5>.

Mancini, M., Rosso, R. (1989) 'Using GIS to assess spatial variability of SCS Curve Number at the basin scale. In *New directions for surface water modelling*', *Proceedings of the Baltimore Symposium*. IAHS, 181, pp. 435–444.

Márquez-Romance, A., Cárdenas-Izaguirre, S., Guevara-Pérez, E., Pérez-Pacheco, S. (2022) 'Calibration of diffusion-advection process models for organochlorine pesticides in a tropical river', *Environ Qual Manage.*, 31(4), pp. 403-424, <https://doi.org/10.1002/tqem.21827>

Marsalek, J, and Rochfort, Q. (2004) 'Urban wet-weather flows: sources of fecal contamination impacting on recreational waters and threatening drinking-water sources.' *Journal of Toxicology and Environmental Health, Part A* 67.20-22, pp. 1765-1777.

Masia, A., Campo, J., Vázquez-Roig, P., Blasco, C., Picó, Y. (2013) 'Screening of currently used pesticides in water, sediments and biota of the Guadalquivir River Basin (Spain)', *Journal of Hazardous Materials*, 263P, pp. 95-104.

McGechan, M.B. and Vinten, A.J.A. (2003) 'Simulation of transport through soil of *E. coli* derived from livestock slurry using the MACRO model', *Soil Use and Management*, 19, pp. 321-330.

McKergow, L.A., Davies-Colley, R.J. (2010) 'Stormflow dynamics and loads of *Escherichia coli* in a large mixed land use catchment', *Hydrological Processes*, 24 (3), pp. 276-289.

Melesse, A.M., Graham, W.D. (2004) 'Storm runoff prediction based on a spatially distributed travel time method utilizing remote sensing and GIS', *J. Am. Water Resour. Assoc.* 40 (4), 863–879.

Mendoza, L.M., Mladenov N., Kinoshita, A. M., Pinongcos, F., Verbyla, M. E., Gersberg, R. (2020) 'Fluorescence-based monitoring of anthropogenic pollutant inputs to an urban stream in Southern California, USA', *Science of the Total Environment*, 817, pp. 1-8.

Met Office (2003) '1 km Resolution UK Composite Rainfall Data from the Met Office Nimrod System', NCAS British Atmospheric Data Centre. Available at: <http://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350> (Accessed 11.07.2019).

- Met Office (2019) 'Met Office UK Climate Projections: Headline Findings' Version 2, <https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp-headline-findings-v2.pdf>, (Accessed at 15.06.2020).
- Milne, A.E., Coleman, K., Todman, L.C., Whitmore, A.P., (2020) 'Model-based optimisation of agricultural profitability and nutrient management: a practical approach for dealing with issues of scale', *Environ Monit Assess*, 192: 730, <https://doi.org/10.1007/s10661-020-08699-z>
- Mishra, S.K. & Singh, V.P. (1999) 'Another Look at SCS-CN Method. *Journal of Hydrologic Engineering*', 4(3), pp.257–264.
- Mounce, S. R., Shepherd, W. , Ostojin, S. , Abdel-Aal, M. , Schellart, A. N. A. , Shucksmith, J. D. , Tait, S. J. (2020) 'Optimisation of a fuzzy logic-based local real-time control system for mitigation of sewer flooding using genetic algorithms', *Journal of Hydroinformatics*, 22 (2), pp. 281–295. doi: <https://doi.org/10.2166/hydro.2019.058>
- Murphy, P.N.C., Mellander, P.-E., Melland, A.R., Buckley, C., Shore, M., Shortle, G., Wall, D.P., Treacy, M., Shine, O., Mehan, S., Jordan, P. (2015) 'Variable response to phosphorus mitigation measures across the nutrient transfer continuum in a dairy grassland catchment', *Agriculture, Ecosystems & Environment*, 207, pp. 192-202, <https://doi.org/10.1016/j.agee.2015.04.008>.
- National Audit Office, (2010) 'Tackling diffuse water pollution in England', Report by the Comptroller and Auditor General, London, National Audit Office, HC 188 Session 2010-2011.
- Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011) 'Soil and water assessment tool theoretical documentation version 2009', Texas Water Resources Institute.
- Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A. and Zechman, E. (2010) 'State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management*', 136(4), pp.412-432.
- Nie, Y., Avraamidou, S., Xiao, X., Pistikopoulos, E.N., Li, J., Zeng, Y., Song, F., Yu, J., Zhu, M. (2019) 'A Food-Energy-Water Nexus approach for land use optimization', *Science of The Total Environment*, 659, pp. 7-19, <https://doi.org/10.1016/j.scitotenv.2018.12.242>.
- Oliver, D.M., Fish, R.D., Hodgson, C.J., Heathwaite , A.L., Chadwick, D.R., Winter, M. (2009) 'A cross-disciplinary toolkit to assess the risk of faecal indicator loss from grassland farm systems to surface waters', *Agriculture, Ecosystems and Environment* ,129, pp. 401–412.
- Oliver, D.M., Bartie, P.J., Heathwaite, A.L., Pschetz, L., Quilliam, R.S. (2017) 'Design of a decision support tool for visualising *E. coli* risks on agricultural land using a stakeholder driven approach', *Land Use Policy*, 66, pp. 227–234.
- Oliver, D.M., Bartie, P.J., Heathwaite, A.L., Reaney, S.M., Parnell, J.A.Q., Quilliam, R.S. (2018) 'A catchment-scale model to predict spatial and temporal burden of *E. coli* on pasture from grazing livestock', *Science of the Total Environment*, 616-617, pp. 678-687.
- Ouyang, W., Wei, P., Gao, X., Srinivasan, R., Yen, H., Xie, X., Liu, L., Liu, H. (2020) 'Optimization of SWAT-Paddy for modeling hydrology and diffuse pollution of large rice paddy fields', *Environmental Modelling & Software*, 130, 104736, <https://doi.org/10.1016/j.envsoft.2020.104736>.

- Perez-Pendini, C., Limbrunner, J. F. and Vogel, R.M. (2005) 'Optimal Location of Infiltration-Based Best Management Practices for Storm Water Management', *Journal of Water Resources Planning and Management*, 131(6), pp 441-448.
- Proteus Instruments Ltd (2021) 'Proteus User Manual', V2.2, pp. 1-46.
- Pullan, S.P., Whelan, M.J., Rettino, J., Filby, K., Eyre, S., Holman, I.P. (2016) 'Development and application of a catchment scale pesticide fate and transport model for use in drinking water risk assessment', *Science of the Total Environment*, 563–564, pp.434–447.
- Quilbé, R. Rousseau, A.N., Lafrance, P., Leclerc, J., Amrani, M. (2006) 'Selecting a pesticide fate model at the watershed scale using a multi-criteria analysis', *Water Quality Research Journal of Canada*, 41(3), pp.283–295.
- Refshaard, J.C., Storm, B. & Singh, V.P. (1995) 'MIKE SHE. In Computer Models of Watershed Hydrology', Water Resources Publications, Colorado, USA, pp. 809–846.
- Reus, J., Middleton, D., Bockstaller, C. (1999) 'Comparing environmental risk indicators for pesticides : results of the European CAPER project', *CLM (Netherlands)*, 426, pp.80–82.
- Reynolds, D.M. (2003) 'Rapid and direct determination of tryptophan in water using synchronous fluorescence spectroscopy', *Water Research*, 37(13), pp. 3055-3060.
- Rutherford, J. C., (1994) John Wiley & Sons, Chichester, ISBN 0-471-94282-0
- Saalidong B.M., Aram S.A., Otu S., Lartey P.O. (2022) 'Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems' *PLoS One*, 25;17(1):e0262117. doi: 10.1371/journal.pone.0262117.
- Sadeghi, A.M., Arnold, J.G. (2002) 'A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales', *Total Maximum Daily Load (TMDL) Environmental Regulations: Proceedings of the March 11-13, 2002 Conference*, (Fort Worth, Texas, USA) 701P0102, pp. 56-63.
- Sadeghi, S.H.R., Jalili, Kh., Nikkami, D. (2009) 'Land use optimization in watershed scale, *Land Use Policy*', 26(2), pp. 186-193. <https://doi.org/10.1016/j.landusepol.2008.02.007>.
- Safford, H.R., Bischel, H.N. (2019) 'Flow cytometry applications in water treatment, distribution, and reuse: A review', *Water Research*, 151, pp.110-133.
- Saleh, I.A., Zouar, N., Al-Ghouti, M.A. (2020) 'Removal of pesticides from water and wastewater: Chemical, physical and biological treatment approaches', *Environmental Technology & Innovation*, 19, 101026. <https://doi.org/10.1016/j.eti.2020.101026>
- Sánchez, E., Colmenarejo, M.F., Vicente, J., Rubio, A., García, M.J., Travieso, L., Borja, R. (2007) 'Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution', *Ecological Indicators*, 7(2), pp 315-328, <https://doi.org/10.1016/j.ecolind.2006.02.005>
- Sartory, D. P. and Howard, L. (1992). 'A medium detecting beta-glucuronidase for the simultaneous membrane filtration enumeration of *Escherichia coli* and coliforms from drinking water'. *Letters in Applied Microbiology*, 15, 273-276.
- Schijven, J., Derx, J., de Roda Husman, A.M., Blaschke, A.P., Farnleitner, A.H. (2015) 'QMRACatch: microbial quality simulation of water resources including infection risk assessment', *J. Environ. Qual.*, 44 (5), pp. 1491-1502.

- Seis, W., Zamzow, M., Caradot, N., Rouault, P. (2018) 'On the implementation of reliable early warning systems at European bathing waters using multivariate Bayesian regression modelling', *Water Research*, 143, pp. 301-312.
- Shepherd, W., Mounce, S., Sailor, G., Gaffney, J., Shah, N., Smith, N., Cartwright, A., Boxall, J. (2023) 'Cloud-Based Artificial Intelligence Analytics to Assess Combined Sewer Overflow Performance', *Journal of Water Resources Planning and Management*, 149(10), <https://doi.org/10.1061/JWRMD5.WRENG-5859>.
- Smith, A., Reacher, M., Smerdon, W., Adak, G., Nichols, G., Chalmers, R. (2006) 'Outbreaks of waterborne infectious intestinal disease in England and Wales, 1992–2003', *Epidemiology & Infection*, 134(6), pp. 1141-1149. doi:10.1017/S0950268806006406
- Sokolova, E., Pettersson, T.J.R., Bergstedt, O., Hermansson, M. (2013) 'Hydrodynamic modelling of the microbial water quality in a drinking water source as input for risk reduction management', *J. Hydrol.*, 497, pp. 15-23.
- Sorensen, J.P.R., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G., Ward, J.S.T. (2018) 'Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values', *Science of the Total Environment*, 662-3, pp. 1250-1257.
- Sorensen, J.P.R., Vivanco, A., Ascott, M.J., Goody, D.C., Lapworth, D.J., Read, D.S., Rushworth, C.M., Bucknall, J., Herbert, K., Karapanos, I., Gumm, L.P., Taylor, R.G. (2018) 'Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water', *Water Research*, 137, pp. 301-309.
- Speight, V. L., Kalsbeek, W. D., DiGiano, F.A. (2004) 'Randomized Stratified Sampling Methodology for Water Quality in Distribution Systems', *Journal of Water Resources Planning and Management*, 130(4), pp. 330-338.
- Srivastava, P., Hamlett, J. M., Robillard, P.D. and Day, R.L. (2002) 'Watershed Optimization of best management practices using AnnAGNPS and a genetic algorithm', *Water Resources Research*, 38(3), pp. 3.1-3.14.
- Srivastava, P., Hamlett, J. M., and Robillard, P.D. (2003) 'Watershed Optimization of agricultural best management practices: Continuous simulation versus design storms', *Journal of American Water Resources Association*, 39(5), pp. 1043-1054. <https://doi.org/10.1111/j.1752-1688.2003.tb03691.x>
- Srivastava A.K., Tait, S., Schellart, A., Kroll, S., Van Dorpe, M., Van Assel, J., Shucksmith, J. (2018), 'Quantifying Uncertainty in Simulation of Sewer Overflow Volume', *ASCE J. Environ. Eng.*, 144 (7), pp. 1-10.
- Stadler, P., Bloschl, G., Vogl, W., Koschel, J., Epp, M., Lackner, M., Oismuller, M., Kumpan, M., Nemeth, L., Strauss, P., Sommer, R., Ryzinska-Paier, G., Farnleitner, A.H., Zessner, M. (2016) 'Real-time monitoring of beta-D-glucuronidase activity in sediment laden streams: a comparison of prototypes', *Water Research*, 101, pp. 252-261.
- Stadler, P., Farnleitner, A., Zessner, M. (2017) 'Development and evaluation of a self-cleaning custom-built auto sampler controlled by a low-cost RaspberryPi microcomputer for online enzymatic activity measurements', *Talanta*, 162, pp. 390-397.

- Stanfield, G., Jago, P.H. (1987) The development and use of a method for measuring the concentration of assimilable organic carbon in water. Report PRU 1628-M. Medmenham: Water Research Center.
- Sterk, A., Schijven, J., de Roda Husman, A.M., de Nijs, T. (2016) 'Effect of climate change on runoff of *Campylobacter* and *Cryptosporidium* from land to surface water, *Water Res.*, 95, pp. 90-102.
- Stoate, C., Brown, C., Vilamizar Velez, Z., Jarratt, S., Morris, C., Biggs, J., Szczur, Z., Crotty, F. (2017) 'The use of a herbicide to investigate catchment management approaches to meeting Sustainable Intensification (SI) objectives', *Aspects of Applied Biology*, 136, pp. 115-120.
- Sylvestre, É., Burnet, J-B., Dorner, S., Smeets, P., Medema, G., Villion, M., Hachad, M., Prévost, M., (2021) 'Impact of Hydrometeorological Events for the Selection of Parametric Models for Protozoan Pathogens in Drinking-Water Sources', *Risk Analysis*, 41(8), DOI: 10.1111/risa.13612.
- Taghipour, M., Shakibaeinia, A., Sylvestre, E., Tolouei, S., Dorner, S. (2019) 'Microbial risk associated with CSOs upstream of drinking water sources in a transboundary river using hydrodynamic and water quality modeling' *Science of The Total Environment*, 683, pp. 547-558.
- Tang, K.S., Man, K.F., Kwong, S., He, Q. (1996) 'Genetic Algorithms and their Applications', *IEEE Signal Processing Magazine*, 13(6), pp. 22-37. DOI:10.1109/79.543973
- The Floods and Water (Amendment etc.) (EU Exit) Regulations 2019, <https://www.legislation.gov.uk/ukxi/2019/558/contents/made> , (Accessed at 31.07.2023).
- The Office for National Statistics (2019) 'National population projections: 2018-based', <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2018based>, (Accessed at 15.06.2020).
- The standing Committee of Analysts (2016) 'The Microbiology of Drinking Water (2016) - Part 4 Methods for the isolation and enumeration of coliform bacteria and *Escherichia coli* (including *E. coli* O157:H7)'.
- The Water Supply (Water Quality) Regulations 2016, <https://www.legislation.gov.uk/ukxi/2016/614/contents/made> , (Accessed at 31.07.2023).
- England and The Water Supply (Water Quality) Regulations (Wales) 2018, <https://www.legislation.gov.uk/wsi/2018/647/made> , (Accessed at 31.07.2023).
- Thorndahl, S., Willems, P. (2008) 'Probabilistic modelling of overflow, surcharge and flooding in urban drainage using the first order reliability method and parameterization of local rain series.', *Water Res.*, 42 (1–2), pp. 455–466. <https://doi.org/10.1016/j.watres.2007.07.038>.
- Tian, Y.Q., Gong, P., Radke, J.D., Scarborough, J., (2002) 'Spatial and temporal modeling of microbial contaminants on grazing farmlands', *J. Environ. Qual.*, 31, pp. 860–869.
- Tröger, R., Ren, H., Yin, D., Postigo, C., Nguyen, P.D., Baduel, C., Golovko, O., Been, F., Joerss, H, Boleda, M.R., Polesello, S., Roncoroni, M., Taniyasu, S., Menger, F., Ahrens, L., Lai, F., Wiberg, K. (2021) 'What's in the water? – Target and suspect screening of contaminants of emerging concern in raw water and drinking water from Europe and Asia', *Water Research*, 198, 117099.

Tscheikner-Gratl, F., Bellos, V., Schellart., A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Angel Rico-Ramirez, M., Fernandes de Carvalho, R., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, T. (2019), 'Recent insights on uncertainties present in integrated catchment water quality modelling', *Water Research*, 150, pp. 368-379.

United Nations, Department of Economic and Social Affairs, Population Division (2019) 'World Population Prospects 2019: Highlights', ST/ESA/SER.A/423.

USEPA (2008) 'Results of the interlaboratory testing study for the comparison of methods for detection and enumeration of enterococci and *Escherichia coli* in Combined Sewer Overflows (CSOs)', EPA-821-R-08-006, Environmental Protection Agency, Office of Water, Washington, DC.

Van der Kooij, D., Veenendaal, H.R., van der Mark, E.J., Dignum, M. (2017) 'Assessment of the microbial growth potential of slow sand filtrate with the biomass production potential test in comparison with the assimilable organic carbon method', *Water Research*, (125), pp. 270-279, <https://doi.org/10.1016/j.watres.2017.06.086>.

Vezzaro, L., Mikkelsen, P. S., Deletic, A., McCarthy D. (2013) 'Urban drainage models—simplifying uncertainty analysis for practitioners' *Water Sci. Technol.*, 68 (10), pp. 2136–2143 <https://doi.org/10.2166/wst.2013.460>.

Vincent, K., Starrs, D., Wansink, V., Waters, N., Lal, A. (2022) 'Relationships between extreme flows and microbial contamination in inland recreational swimming areas', *J Water Health*, 20 (5): pp. 781–793. doi: <https://doi.org/10.2166/wh.2022.294> .

Wallis, S.G., Young, P.C., Beven, K.J. (1989) 'Experimental investigation of the aggregated dead zone model for longitudinal solute transport in stream channels', *Proc. Instn Ciu. Engrs.*, 87, pp.1-22.

Walker, F.R., Stedinger, J.R. (1999) 'Fate and transport model of *Cryptosporidium*', *J. Environ. Eng.*, 125 (4), pp. 325-333.

Wang, M., Wang, Y., Gao, X. and Sweetapple, C. (2019) 'Combination and placement of sustainable drainage system devices based on zero-one integer programming and schemes sampling', *Journal of Environmental Management*, 238, pp. 59-63.

Whelan, G., Kim, K., Pelton, M.A., Soller, J.A., Castleton, K.J., Molina, M., Pachepsky, Y., Zepp, R. (2014) 'An integrated environmental modeling framework for performing quantitative microbial risk assessments', *Environ. Model. Softw.*, 55, pp. 77-91.

Weatherquest (2023) <https://weatherquest.co.uk/> (Accessed at 17.07.2023).

WHO (2001) 'Water Quality: Guidelines, Standards and Health', Edited by Lorna Fewtrell and Jamie Bartram. Published by IWA Publishing, London, UK. ISBN: 1 900222 28 0

WHO (2012) 'Animal Waste, Water Quality and Human Health', Edited by A Dufour, J Bartram, R Bos and V Gannon, IWA Publishing, London, UK, ISBN: 9781780401232.

¹WHO (2017) 'Guidelines for drinking-water quality: fourth edition incorporating the first addendum.' Geneva: World Health Organization, Licence: CC BY-NC-SA 3.0 IGO.

²WHO (2017) 'WATER QUALITY AND HEALTH - REVIEW OF TURBIDITY: Information for regulators and water suppliers', Technical Brief,

(Accessed at 30.07.2023)

Wilkinson, R.J., McKergow, L.A., Davies-Colley, R.J., Ballantine, D.J., Young, R.G., (2011) 'Modelling storm-event *E. coli* pulses from the Motueka and Sherry Rivers in the South Island', New Zealand. N. Z. J. Mar. Freshw. Res. 45 (3), 369-393.

Wong, T.S.W. (2003) 'Comparison of celerity-based with velocity-based time-of-concentration of overland plane and time-of-travel in channel with upstream inflow', Adv. Water Resour., 26 (11), 1171–1175

Yassin, M., Asfaw , A., Speight, V., Shucksmith, J.D., (2021) 'Evaluation of Data-Driven and Process-Based Real-Time Flow Forecasting Techniques for Informing Operation of Surface Water Abstraction', J. Water Resour. Plann. Manage., 147(7), pp.1-14.

Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P. (1989) 'AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds', Journal of Soil and Water Conservation, 44(2), pp.168–173.

Yousefi, H. and Moridi, A. (2022) 'Multiobjective Optimization of Agricultural Planning Considering Climate Change Impacts: Minab Reservoir Upstream Watershed in Iran', American Society of Civil Engineers, 148(4): 04022007, DOI: 10.1061/(ASCE)IR.1943-4774.0001675.

Zan, R., Blackburn, A., Plaimart, J., Acharya, K., Walsh, C., Stirling, R., Kilsby, C.G., Werner, D. (2023) 'Environmental DNA clarifies impacts of combined sewer overflows on the bacteriology of an urban river and resulting risks to public health', Science of The Total Environment, 889(164282), ISSN 0048-9697, <https://doi.org/10.1016/j.scitotenv.2023.164282>.

Zhang, P., Liu, Y., Pan, Y., Yu, Z. (2011) 'Land use pattern optimization based on CLUE-S and SWAT models for agricultural non-point source pollution control', Mathematical and Computer Modelling, 58, pp. 588–595.

Zhang, B., Zhang Q.-Q., Zhang, S.-X., Xing, C., Ying, G.-G., (2020) 'Emission estimation and fate modelling of three typical pesticides in Dongjiang River basin, China', Environmental Pollution, 256(113660), <https://doi.org/10.1016/j.envpol.2019.113660>.

Zhang, F., Sun, Q., Mehrabadi, M., Khoshnevisan, B., Zhang, Y., Fan, X., Zhai, L., Xia, Y., Wu, M., Liu, D., Pan, J., Rafiee, S., Liu, H. (2021) 'Joint analytical hierarchy and metaheuristic optimization as a framework to mitigate fertilizer-based pollution', Journal of Environmental Management, 278, Part 1, 111493, <https://doi.org/10.1016/j.jenvman.2020.111493>.

Zhu, K., Chen, Y., Zhang, S., Yang, Z., Huang, L., Lei, B., Li, L., Zhou, Z., Xiong, H., Li, X. (2020) 'Identification and prevention of agricultural non-point source pollution risk based on the minimum cumulative resistance model', Global Ecology and Conservation, 23, e01149, <https://doi.org/10.1016/j.gecco.2020.e01149>