Rapid detection of faecally contaminated drinking water with in-situ fluorescence spectroscopy

James P.R. Sorensen

PhD Thesis
University College London
November 2022

Declaration

I, James Peter Robert Sorensen, confirm that the work presented in this thesis entitled "Rapid detection of faecally contaminated drinking water with in-situ fluorescence spectroscopy" is my own work. Where information has been derived from other sources, I confirm that this has been indicated in the thesis with an appropriate reference.

……

James Sorensen
22 October 2021

Abstract

Two billion people consume drinking water contaminated with human and animal faeces. The resulting infections are a major source of disease globally, with an estimated 500,000 deaths per year from diarrhoea alone. This thesis explores the viability of in-situ fluorescence spectroscopy as a simpler, instantaneous, more temporally resilient alternative to faecal indicator organisms (FIOs) to indicate faecal contamination risk in drinking water sources across a range of hydrological and climatological settings

TLF was a significant predictor in logistic regression models of the presence-absence of FIOs and moderately to very strongly correlated with FIO enumeration, including in real-time data. HLF was also shown to be a similarly effective indicator to TLF of FIO presence-absence and enumeration. TLF/HLF were superior to other rapid approaches, such as turbidity, sanitary risk scores or total bacterial cell counts, to indicate FIOs. Seasonal sampling demonstrated TLF/HLF were more associated with FIOs during the wet season than dry season. The ranking of sources using TLF/HLF was more resilient with time than that using FIOs, with dry season TLF/HLF more related to wet season FIOs, when FIOs are elevated, then dry season FIOs.

In groundwater, tryptophan-like and humic-like fluorophores were shown to be predominantly extracellular and hence will have different transport properties in comparison to FIOs and larger pathogens. Fluorophores were also demonstrated to accumulate in highly contaminated intergranular aquifers where TLF/HLF intensity related to other indicators of faecal contamination, such as on-site sanitation density, but not to FIO enumeration.

In-situ fluorescence spectroscopy is a simple, instantaneous approach to screen water sources for faecal contamination risk. The technique offers global scope to reduce population exposure to enteric pathogens in drinking water: from providing real-time data across municipal supply infrastructure to assessing the relative risks between drinking water sources in low-income settings.

Impact statement

The seminal paper (Chapter 3) produced by this thesis was the first evidence of a relationship between tryptophan-like fluorescence and faecal indicator bacteria in drinking water. It has directly led to further BGS funding through NERC (£86k), an NGO (£25k), REACH, and Innovate UK (£100k). Following the initial paper, fluorescence spectroscopy as an indicator of microbial risks has subsequently formed a substantial element of several full-time PhD students, some of which have finished during this extended part-time PhD. These included a groundwater focus at Karlsruhe Institute of Technology (Simon Frank) and Surrey University (Jade Ward), and a sensor development project at the University of Colorado (Emily Bedell). The six papers produced during this thesis have, as of 18^{h} October 2022, received 256 citations on Google Scholar, including policy citations for "Innovations in WASH" for the World Bank and "Disruptive Technologies that may Transform the Water Sector in the Next 10 Years" for the Inter-American Development Bank.

The start of the thesis fortuitously coincided with a call from UNICEF and the WHO, under the Joint Monitoring Programme (JMP), for more rapid approaches to identify faecal contamination in drinking water (https://www.unicef.org/innovation/rapid-water-quality-testing). I was invited to an initial stakeholder consultation in 2016 following publication of Chapter 3, where I presented a poster, and have stayed engaged in the process, including delivering an invited talk at Tryptophan 2020, where a UNICEF representative was present. In-situ fluorescence spectroscopy is now being considered as an alternative rapid method by the JMP in upcoming independent WHO lab tests. UNICEF alone conducts at least a million tests annually, and if the technique performs well in their tests, could be used in behaviour changes programmes targeting over two billion people.

In-situ fluorescence spectroscopy has been deployed by Portsmouth Water as a real-time indicator of microbial contamination, and others within the water industry have also expressed an interest in running a trial with a microbial focus, including Affinity Water and United Utilities. I have presented

at various conferences with a strong water company presence, including the Sensors for Water Interest Group on two occasions, to engage water companies. I also attend the regular water industry led biannual Flow Cytometry conference where there appears an effort to combine flow cytometry and fluorescence spectroscopy to provide an early-warning of faecal breakthroughs, infrastructure integrity, and treatment failure across the municipal water supply network.

The work has generated commercial benefits for fluorimeter manufacturers. Chelsea Technologies, who manufacturer the fluorimeter used throughout the thesis (https://chelsea.co.uk/achieving-un-sustainable-development-goal-6-with-chelsea-technologies-tryptophan-sensor/), estimate a directly attributable increase in gross sales of around £150k. Another supplier of tryptophan-like fluorimeters, Proteus, now markets their real-time coliform detection capability (https://www.proteus-instruments.com/parameters/total-coliform-sensors/). These Proteus fluorimeters are deployed in several locations, including the Chicago River as part of the "first real-time water quality monitoring project in the U.S. to measure microbial pollutants in an urban waterway" to allow people "to make decisions about how to recreate on the river" (https://www.h2nowchicago.org/).

Acknowledgments

Of course, I am indebted to my supervisors Richard Taylor, Lena Ciric (UCL), and Alan MacDonald (BGS) whom have always found the time to listen, advice, and find resources where necessary. Most importantly, Richard has provided a lovely lunch in the Housman Room on every visit to UCL. I should point out that due to the COVID-19 pandemic, and the lack of visits to UCL since 2019, that I should have at least two years' of banked lunch credits to use at my leisure, and I'm feeling hungry!

During the PhD, I have been fortunate to have worked directly with four UCL MSc students: Andrew Carr, Robert Lyness, Laura van der Marel, and Raphaëlle Roffo. It was a pleasure travelling and working with each of these individuals on their first trips to Africa (Kenya, Senegal, Uganda), all of whom helped generate data for this PhD. There was a particularly chaotic and memorable 2018 involving back-to-back trips with Andrew (Uganda), Laura (Kenya) and, finally, Raphaëlle (Senegal), which started the day after flying in from Armenia. Slightly exhausting 8 weeks, and the memories of running around as many field sites as possible in a day, then rushing to an airport to catch a night flight, so we'd have fresh samples for analysis the following day in Wallingford, UK, still get my heart pounding. Thanks to Dan Read (UKCEH) and Peter Williams (BGS) for kindly taking those samples off my hands and analysing them on each of those bleary-eyed mornings - they could both clearly see I was incapable of anything, despite me turning up with the intention of doing the analyses.

None of the overseas work would have been possible without fantastic and good-humoured in-country support. My thanks go out to Abdoulaye Pouye, Mor Talla Diaw, Djim Diongue, Seynabou Faye, and Cheikh Gaye (Université Cheikh Anta Diop de Dakar); Jacintha Nayebare, Rita Nakazibwe, Michael Owor, and Robinah Kulabako (Makerere University); Japhet Kanoti and Daniel Olago (University of Nairobi); Meki Chirwa, Joel Kabika, Daniel Nkhuwa (University of Zambia); Mirriam Chibesa and Moses Liemisa (Lukanga Water and Sewerage Company). Unfortunately, I cannot

remember the names of the numerous university drivers, I'm ashamed to say, but Samuel, Robert *et al.*, thank you.

The PhD was only possible through finance support from my employers, the British Geological Survey, who covered my tuition fees and provided 20 days per year of training time to work on the thesis. I also acknowledge support from:

- FDCO, the Economic and Social Research Council (ESRC), and the National Environmental Research Council (NERC), Mapping groundwater quality degradation beneath growing rural towns in sub-Saharan Africa, NE/002078/1. This project was the launchpad for the thesis and I'm grateful to Dan Lapworth for involving me in this initial work.
- NERC funding, TryGGER: Application of tryptophan Florescence Sensors for improved raw water quality monitoring of Faecal Contamination in Groundwater sources, NE/M021939/1.
- Royal Society and UK UK Foreign, Commonwealth & Development Office (FCDO) Africa Capacity Building Initiative, AfriWatSan project, AQ140023.
- British Geological Survey NC-ODA grant NE/R000069/1: Geoscience for Sustainable Futures.
- Innovate UK, Development of a low-cost and portable optical sensor for the instantaneous indication of pathogens in drinking water, project 104479.

Thank you, Ria, for letting me travel around the world whilst I left you in a house full of unwanted lodgers! The good news is: I didn't die, the lodgers have gone, you now get to have Mondays off, and the PhD is finished. It was all worth it.

To the parents, money well spent!

Contents

List of Figures

List of Tables

List of acronyms

BGS (British Geological Survey)

BOD (Biological oxygen demand)

DO (Dissolved oxygen)

DOC (Dissolved organic carbon)

EEM (Excitation-emission matrix)

FIO (faecal indicator organism)

FNR (False-negative rate)

FPR (False-positive rate)

HI (High-income)

HLF (Humic-like fluorescence)

LED (Light emitting diode)

LMI (Low-middle-income)

MDL (Minimum detection limit)

MLSB (Membrane lauryl sulphate broth)

NOM (Natural organic matter)

PARAFAC (Parallel Factor Analysis)

ROC (Receiver operating characteristic)

RSD_{RC} (Relative standard deviation of reproducibility)

SEC (Specific electrical conductivity)

SRP (Soluble reactive phosphorus)

SRS (Sanitary risk score)

TDP (Total dissolve phosphorus)

TLF (Tryptophan-like fluorescence)

TTC (Thermotolerant coliform)

UKWIR (United Kingdom Water Industry Research)

UNICEF (United Nations Children's Fund)

UV (Ultraviolet)

WHO (World Health Organisation)

Thesis structure

This thesis contains a series of interrelated chapters that address the main aim of evaluating the utility of fluorescence spectroscopy for the instantaneous detection of faecally contaminated drinking water risks. Chapter 1 details the global prevalence and impacts of faecally contamination drinking water, current methodologies to assess faecal contamination risks, introduces fluorescence spectroscopy as a potential alternative, and leads to the research aims and objectives. Chapter 2 outline the reasoning for the study areas and details methods used consistently through the thesis. Chapter 3 demonstrates the first application of in-situ fluorescence spectroscopy as a real-time predictor of the presence-absence and enumeration of faecal indicator organisms (FIOs) in drinking water supplies. Chapter 4 derives fluorescence thresholds to determine faecal contamination from a multicounty country dataset and compares the technology against broad indicator assessment criteria defined by UNICEF/WHO. Chapter 5 establishes online fluorescence spectroscopy as a real-time indicator of faecal contamination risks at public water supplies that is correlated with both *E. coli* and total bacteria, and a superior bacteriological indicator than turbidity. Chapter 6 confirms tryptophan-like and humic-like fluorophores are predominantly extracellular in groundwater and discusses the implications in terms of their use as faecal indicators. Chapter 7 shows that tryptophan-like (TLF) and humic-like fluorescence (HLF) are most related to total bacterial cell counts and dissolved organic carbon, not FIOs, in a highly contaminated intergranular aquifer. Chapter 8 illustrates that TLF and HLF fluorescence are more related to FIOs than other rapid approaches and that both TLF and HLF are more temporally resilient indicators of faecal contamination risk than FIOs. Chapter 9 summarises the key results, and looking forward, articulates how in-situ fluorescence spectroscopy could reduce exposure to faecally contaminated drinking water.

Publications

All research chapters (3-8) are published as six separate papers:

Sorensen, J.P.R., Lapworth, D.J., Marchant, B.P., Nkhuwa, D.C.W., Pedley, S., Stuart, M.E., Bell, R.A., Chirwa, M., Kabika, J., Liemisa, M. and Chibesa, M., 2015. In-situ tryptophan-like fluorescence: a real-time indicator of faecal contamination in drinking water supplies. Water Research, 81, 38-46. (Chapter 3)

Sorensen, J.P., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G. and Ward, J.S., 2018. Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values. Science of the Total Environment, 622, 1250-1257. (Chapter 4)

Sorensen, J.P.R., Vivanco, A., Ascott, M.J., Gooddy, D.C., Lapworth, D.J., Read, D.S., Rushworth, C.M., Bucknall, J., Herbert, K., Karapanos, I. and Gumm, L.P., 2018. Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. Water Research, 137, 301-309. (Chapter 5)

Sorensen, J.P.R., Carr, A.F.; Nayebare, J., Diongue, D.M.L., Pouye, A., Roffo, R., Gwengweya, G., Ward, J.S.T., Kanoti, J., Okotto-Okotto, J., van der Marel, L., Ciric, L., Faye, S.C., Gaye, C.B., Goodall, T., Kulabako, R., Lapworth, D.J., MacDonald, A.M., Monjerezi, M., Olago, D., Owor, M., Read, D.S., Taylor, R.G.. 2020 Tryptophan-like and humic-like fluorophores are extracellular in groundwater: implications as real-time faecal indicators. Scientific Reports, 10, 15379, 9. (Chapter 6)

Sorensen, J.P.R., Diaw, M.T., Pouye, A., Roffo, R., Diongue, D.M.L., Faye, S.C., Gaye, C.B., Fox, B.G., Goodall, T., Lapworth, D.J., MacDonald, A.M., Read, D.S., Ciric, L., Taylor, R.G.. 2020 In-situ fluorescence spectroscopy indicates total bacterial abundance and dissolved organic carbon. Science of the Total Environment, 738, 139419, 10. (Chapter 7)

Sorensen, J.P.R., Nayebare, J., Carr, A.F., Lyness, R., Campos, L.C., Ciric, L., Goodall, T., Kulabako, R., Rushworth Curran, C.M., MacDonald, A.M., Owor, M., Read, D.S., Taylor, R.G.. 2021. In-situ fluorescence spectroscopy is a more rapid and resilient indicator of faecal contamination risk in drinking water than faecal indicator organisms. Water Research, 206, 117734. 11. (Chapter 8)

Data availability

Data generated by this PhD were produced under multiple funded projects involving a host of organisations, with varying intellectual property ownership. All data are available through the principal investigator(s) of each project:

Kenya - Prof. Daniel Olago (University of Nairobi) olagodan@gmail.com or Prof. Richard Taylor (UCL) richard.taylor@ucl.ac.uk

Malawi - Prof. Alan MacDonald (BGS) amm@bgs.ac.uk or Dr. Dan Lapworth (BGS) djla@bgs.ac.uk

Senegal - Prof. Cheikh Gaye (Université Cheikh Anta Diop de Dakar) cheikhbecayegaye@gmail.com or Prof. Richard Taylor (UCL) richard.taylor@ucl.ac.uk

Uganda, Dr. Michael Owor (Makerere University) mmowor@gmail.com or Prof. Richard Taylor (UCL), richard.taylor@ucl.ac.uk

UK sensor development, Dr. James Sorensen, jare1@bgs.ac.uk

UK water company, Dr. Dan Lapworth, djla@bgs.ac.uk

Zambia, Dr. Dan Lapworth, djla@bgs.ac.uk

Authorship statement

All work for this thesis was undertaken using resources provided by multiple projects that involved numerous individuals whom are co-authors in all publications where they contributed. These projects were conceived by a combination of Matt Ascott, Luiza Campos, Seynabou Cisse Fay, Cheikh Gaye, Daren Gooddy, Robinah Kulabako, Dan Lapworth, Alan MacDonald, Maurice Monjerezi, Daniel Nkhuwa, Dan Olago, Michael Owor, Steve Pedley, Dan Read, Cathy Rushworth Curran, Marianne Stuart, Richard Taylor, and myself. Logistical and sampling support was provided by Japhet Kanoti, Laura van der Marel, Joseph Okotto-Okotto (Kenya); Gloria Gwengweya (Malawi); Mor Talla Diaw, Djim Diongue, Abdoulaye Pouye, and Raphaëlle Roffo (Senegal); Andrew Carr, Robert Lyness, Jacintha Nayebare (Uganda); James Bucknall, Tim Goodall, Kaylee Herbert, Ilias Karapanos, Anibal Vivanco (UK); and Rachel Bell, Mirriam Chibesa, Meki Chirwa, Joel Kabika, and Moses Liemisa (Zambia). Ben Marchant and Lee Gumm performed the logistic and linear regression modelling in Chapter 3, including their description, and the PARAFAC modelling in Chapter 5, respectively. The PhD was also an opportunity to reach out to other researchers (Andy Baker, Susan Cumberland, Bethany Fox, and Jade Ward) in the fluorescence spectroscopy field to incorporate additional datasets and discuss ideas. All researchers reviewed the chapter text and offered revisions where they are listed as co-authors of the respective paper. The lower cost fluorimeter discussed in Chapter 8, was engineered and produced by a team led by Cathy Rushworth-Curran, as part of an Innovate UK project with myself as principal investigator. I input ideas into the design and led the laboratory and field testing of the sensor.

Chapter 1 - Introduction

1.1 Drinking water as a source of disease

Microbial contamination of drinking water remains the primary water quality concern in both developing and developed countries (WHO, 2017b). The greatest public health risk relates to the consumption of drinking water contaminated with human and animal faeces to which two billion people are currently exposed worldwide (WHO, 2019). Consumption of such water is linked to income and is most prominent in low-income countries, for example over half of the African population (Table 1-1). Nevertheless, there are still large populations exposed in high-income countries. This is typically a result of consuming ineffectively treated groundwater from unregulated private supplies in rural areas (Charrois, 2010; DeFelice et al., 2016). For example, recent US figures indicate 44.5 million residents (14% of the population) consume untreated groundwater (Maupin et al., 2014). In municipal supplies, there also remain sporadic incidences of contamination due to failures within the water supply chain (e.g. Adler et al., 2017; Smith et al., 2010).

Table 1-1 Estimated regional and global population exposure to faecally contaminated drinking water in 2012 (adapted from Bain et al., 2014a). Note: HI – high-income, LMI – low-middle-income.

WHO Region, by income group	Exposure (%)
Africa	52.2
Americas (HI)	1.2
Americas (LMI)	14.6
Eastern Mediterranean (HI)	12.2
Eastern Mediterranean (LMI)	28.8
Europe (HI)	0.6
Europe (LMI)	14.0
South-East Asia	35.1
Western Pacific (HI)	1.5
Western Pacific (LMI)	23.8
Global	26.0

The consumption of enteric pathogens in faecally contaminated drinking water constitutes a major burden on public health due to the elevated incidence of infectious diseases (Bain et al., 2014b). Enteric pathogens encompass bacteria, viruses, protozoa, and helminths; their consumption can result in well-known diseases such as diarrhoea, cholera, typhoid, hepatitis, dysentery, giardiasis and guinea worm. The major impact of enteric pathogens is the more than half a million deaths per year in low- and middle-income countries from diarrhoea, with children under five particularly at risk (Prüss-Ustün et al., 2014). Early childhood diarrhoea is also associated with decremented child growth, cognitive development, and school performance (Berkman et al., 2002; Dillingham and Guerrant, 2004; Esrey et al., 1988). Outside the developing world, impacts remain appreciable. For example, there are an estimated 750,000 to 5,900,000 cases of waterborne illness and 1400 to 9400 deaths per year in the USA from consuming faecally-contaminated groundwater (Macler and Merkle, 2000). To reduce the proportion of the world's population drinking faecally contaminated water, the United Nations have established Sustainable Development Goals (SDGs) for the universal access to safe drinking water for all and improvements in water quality by 2030.

1.2 Current indicators of faecally contaminated drinking water

Routine microbiological analyses of drinking water are undertaken using faecal indicator organisms (FIOs) that share similar characteristics to enteric pathogens and are used to infer their presence (Savichtcheva and Okabe, 2006). Ideal indicators should be present in faeces, not multiply outside the intestinal tract, share similar resilience to environmental conditions and disinfection to enteric pathogens, and can be quantified relatively simply in the laboratory (Hurst et al., 2007). Currently, the WHO Guidelines for Drinking Water Quality (WHO, 2017b), and regulation in many countries, include the indicator group thermotolerant (faecal) coliforms (TTC), including *Escherichia coli* that is considered the best indicator of faecal contamination by warm-blooded mammals.

Drinking water compliance monitoring using FIOs alone provides ineffective protection of public health (Stelma Jr and Wymer, 2012; WHO, 2017a) and waterborne outbreaks remain common, even in high-income countries (Collier et al., 2021). The main concerns relating to FIOs are that microbial

contamination is highly variable temporally, which is not characterised by infrequent (e.g. quarterly/annual) FIO sampling in many circumstances (Hrudey and Hrudey, 2004), and results are delivered after exposure has occurred. Furthermore, FIO analysis requires well-trained personnel, restricting the extent of nationally representative surveys. Because no result is provided in-situ at the source, communication of risks and behavioural change is also inhibited (UNICEF/WHO, 2017).

To address some of these limitations with FIO monitoring, the World Health Organisation (WHO) recommends a risk-based management approach to ensure water safety (WHO, 2017b). A risk-based approach often includes sanitary inspections of the source (Kelly et al., 2020) and operational monitoring of parameters that can be quantified rapidly to indicate changes in source water quality (WHO, 2017b), notably turbidity (WHO, 2017d), in addition to FIO culturing. However, turbidity is not a reliable indicator of faecal contamination (Jung et al., 2014; Pronk et al., 2006; UKWIR, 2012) and there is a dearth of suitable alternative online detection methods that can be used (Besmer and Hammes, 2016). There is also a current drive by UNICEF/WHO (2017) for the development of new water quality approaches enabling more rapid, in situ detection of faecal contamination in drinking water.

1.3 Fluorescence spectroscopy

1.3.1 Principles of fluorescence

Fluorescence is produced via three stages: excitation of a molecule, vibrational relaxation and internal conversion, and emission of light (Figure 1-1) (Lakowicz, 2006; Reynolds, 2014). A molecule reaches an excited state by gaining energy through the adsorption of an appropriate incoming photon and happens over femtoseconds (10^{-15}). Vibrational relaxation is a non-radiative process whereby the gained energy is lost as kinetic energy within the same molecule or to surrounding molecules allowing transition between vibration levels over picoseconds (10^{-12}). Internal conversion is a further loss of energy at similar timescales to vibrational relaxation allowing relaxation across vibration levels and electronic energy states. Fluorescence is a slower process (10^{-9} seconds) whereby gained energy is lost

through the emission of light. This light is typically of longer wavelength than the original absorbed light because of the loss of energy through vibrational relaxation and internal conversion - termed the Stokes shift (Stokes, 1852). Emitted light is also generally distributed over a range of wavelengths because of the array of vibrational levels between electronic states that are returning to ground state (S0, Figure 1-1). Other processes such molecular collisions and quenching, energy transfer between molecules due to spectral overlap in absorption and fluorescence, are competing pathways with fluorescence to return an excited an electron to the ground state.

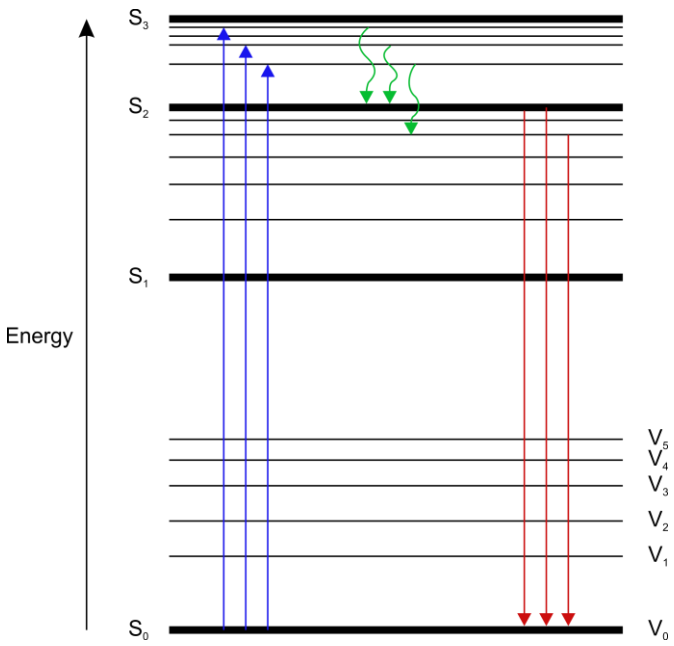

Figure 1-1. Jablonski Diagram with electronic energy states (S_0-S_3) and vibration levels (V_0-V_5) demonstrating varied excitation through absorption (blue), relaxation through internal conversion and vibrational relaxation (green) and relaxation to a ground state through fluorescence (red).

1.3.2 Fluorescence of natural organic matter

Natural organic matter (NOM) is defined as the assortment of organic, carbon-based materials that are present in all natural waters (Matilainen et al., 2010). NOM can be broadly be split into non-humic solutes (e.g. fats, carbohydrates, and proteins) and complicated heterogeneous humics, though the two groups are not completely distinct as non-humic solutes can be contained in the structure of complicated humics (Fabris et al., 2008). NOM can be either allochthonous, derived from the surrounding environment, or autochthonous, created in-situ through microbial processes (Hudson et

al., 2007). The specific characteristics of NOM are frequently related to its origin (e.g. vegetation, soil, wastewater) (Baker and Curry, 2004; Baker and Spencer, 2004; Reynolds, 2002; Spencer et al., 2008).

Fluorescence spectroscopy is a rapid, non-invasive technique commonly used to characterise fluorescent NOM over the past 50 years (Baker, 2002a; Bieroza et al., 2009; Carstea et al., 2010; Fellman et al., 2010; Hudson et al., 2007; Lapworth et al., 2008; Stedmon and Markager, 2005; Stryer, 1968). Aromatic organic compounds commonly fluoresce because of the energy sharing, unpaired electron structure of the carbon ring (Hudson et al., 2007). Fluorescence spectroscopy allows the discrimination between different types of NOM because the wavelengths of adsorption (excitation) and emission are specific to a molecule (Lakowicz, 2006). In the study of fluorescent NOM, compounds that fluoresce or only absorb light are termed fluorophores and chromophores, respectively (Mopper et al., 1996).

Fluorescence spectroscopy is commonly conducted on water samples using a range of excitation and emission wavelengths. This approach allows the creation of excitation-emission matrices (EEMs) to visualise the intensity of fluorescence across the range of wavelengths (Figure 1-2). A range of peaks have been observed in these EEMs that are often split into protein-like and humic-like fluorescence (Coble et al., 1993; Yamashita and Tanoue, 2003). Protein-like fluorescence is considered autochthonous and includes tryptophan- (Figure 1-2A) and tyrosine-like fluorescence (Coble et al., 2014). These peaks resemble the fluorescent properties of the equivalent amino acids but are not necessarily representative of the presence of the amino acid itself.

Tryptophan-like fluorescence (TLF), or peak T, is the most commonly detected protein-like fluorescence in natural waters (Yang et al., 2015a), with one peak centred around $\lambda_{ex/em}$ of 275/340 nm (Coble, 1996) (Figure 1-2A). TLF is considered an indicator of biological activity in water (Cammack et al., 2004; Elliott et al., 2006a) and has been characterised in high concentrations in both human and animal wastewater (Baker, 2001; 2002b). In wastewater and sewage impacted rivers, multiple studies have successfully correlated TLF with biological oxygen demand (BOD5) tests (Hudson et al., 2008;

Reynolds and Ahmad, 1997), which is the quantity of oxygen required by the native microbial population to degrade the labile organic matter over five days. Consequently, researchers have recommended TLF as a rapid, indicative tool of bioavailable organic matter for monitoring effluent quality, river water quality, and pollution events (Baker and Inverarity, 2004; Hudson et al., 2008).

Humic-like fluorescence (Figure 1-2B) has traditionally been considered allochthonous, outside of the marine environment, and resembles the fluorescent properties of humic acids (Coble et al., 2014). Nevertheless, HLF is also elevated in wastewater (Hur et al., 2010; Sihan et al., 2021), is an excellent proxy for DOC in rivers (r^2 0.74-0.99) (Carstea et al., 2020), and recent laboratory studies suggest it can also be microbially-derived (Fox et al., 2019; Fox et al., 2017; Fox et al., 2021). Other observed NOM fluorophores include polyaromatic hydrocarbons and chlorophyll (Moberg et al., 2001; Nahorniak and Booksh, 2006).

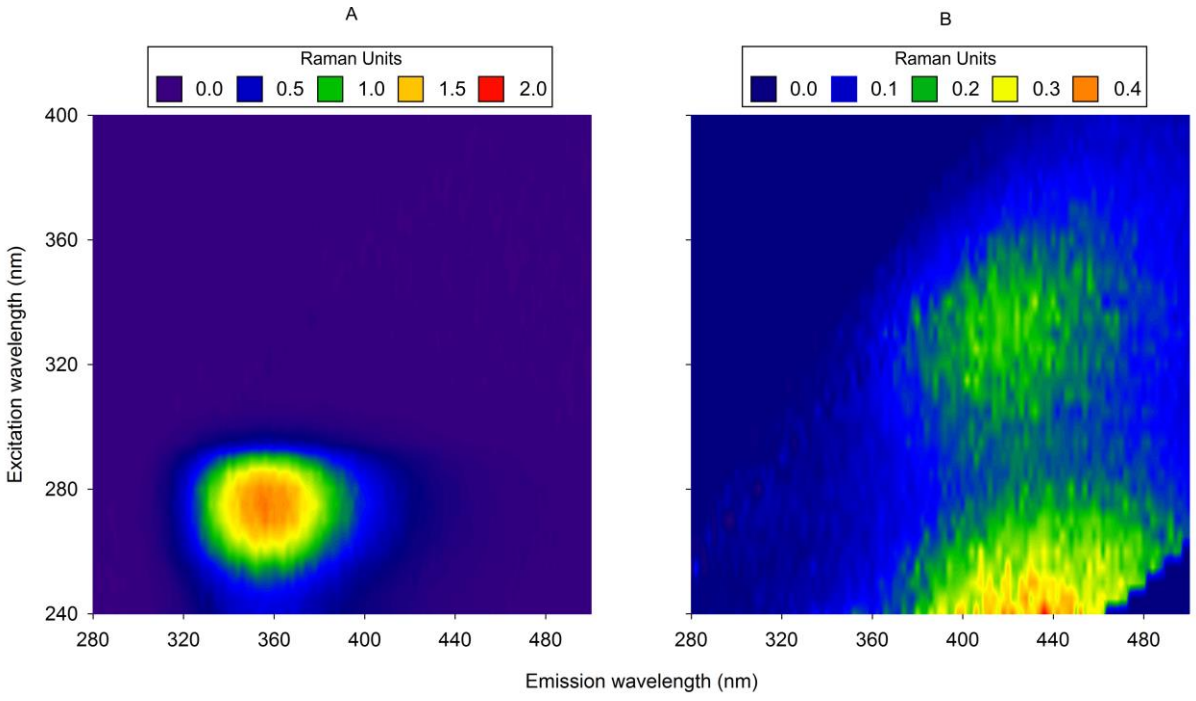

Figure 1-2. (A) Excitation-emission matrix (EEM) of a 100 ppb tryptophan standard (B) mean EEM from repeated sampling of a UK public water supply demonstrating no clear protein-like peak, but dual humic-like peaks. Both EEMs have been masked to remove Raleigh and Raman scattering.

1.3.3 Data analysis

Peak-picking and PARAllel FACtor (PARAFAC) analyses are commonly used to interpret EEMs. Peak picking involves identifying components based on their maximum intensity at specific excitation-emission wavelength pairs (Coble, 1996) (Figure 1-2). However, peak-picking is limited by overlapping and interference between peaks and peak shifts (Carstea et al., 2016; Murphy et al., 2008; Yang et al., 2015a). Consequently, authors usually additionally report intensity ratios between identified components (e.g. Baker, 2002b). Recently, it has become standard practice to additionally undertake PARAFAC analysis (Bro, 1997; Murphy et al., 2013). PARAFAC is a mathematical tri-linear model that identifies and separates the contribution of individual fluorophore components to EEMs without any assumptions regarding their excitation-emission spectra (Carstea et al., 2016; Cohen et al., 2014).

1.3.4 In-situ fluorescence spectroscopy

Multiple manufacturers currently produce field-deployable fluorimeters according to the principle of peak-picking which target commonly observed fluorophores such as TLF, HLF, and PAHs (e.g. Figure 1-3). The capability to take high temporal resolution fluorescence measurements in-situ has triggered a new wave of research. This work has included investigating event-driven nutrient export (Blaen et al., 2016; Tunaley et al., 2016; Wilson et al., 2013) and detection of wastewater surges/inputs (Mendoza et al., 2020; Thompson and Dickenson, 2021) in surface waters. In-situ fluorimeters have also been used as real-time indicators of DOC, BOD (Khamis et al., 2021; Khamis et al., 2017; Ruhala and Zarnetske, 2017; Saraceno et al., 2009) and cyanobacteria (Zamyadi et al., 2016) in freshwaters. In engineered water systems, in-situ fluorimeters have been used to assess cross-connections between recycled water and potable water (Hambly et al., 2015) and for real-time monitoring of organic matter in drinking water treatment plants (Mladenov et al., 2018; Shutova et al., 2016). However, in-situ measurements of fluorescence can be impacted by temperature and turbidity, which can require compensation (Downing et al., 2012; Khamis et al., 2015; Saraceno et al., 2017).

Figure 1-3 UviLux Fluorimeter developed by Chelsea Technologies Group Ltd., UK. Note that the sensor operates using UV light and is sensitive to sunlight; hence, the beaker would need to be covered for a reliable measurement.

1.3.5 An indicator of microbiological quality

Within the applied spectroscopy literature, it has been long documented that *E. coli* cells directly emit TLF and also excrete compounds that fluoresce in the TLF region in controlled laboratory studies (Bronk and Reinisch, 1993; Dalterio et al., 1986; Dalterio et al., 1987; Seaver et al., 1998). In terms of excreting TLF NOM, *E. coli* is the preferred organism for the industrial production of tryptophan by the fermentation of carbohydrates (Ikeda, 2006). Alternatively, if tryptophan is readily available in the environment then *E. coli* will import and hydrolyse tryptophan, almost quantitatively (Li and Young, 2013), into indole that is then excreted. This would also enhance any TLF signal because pure indole fluoresces within the TLF region at 33% greater intensity than tryptophan (Figure 1-4).

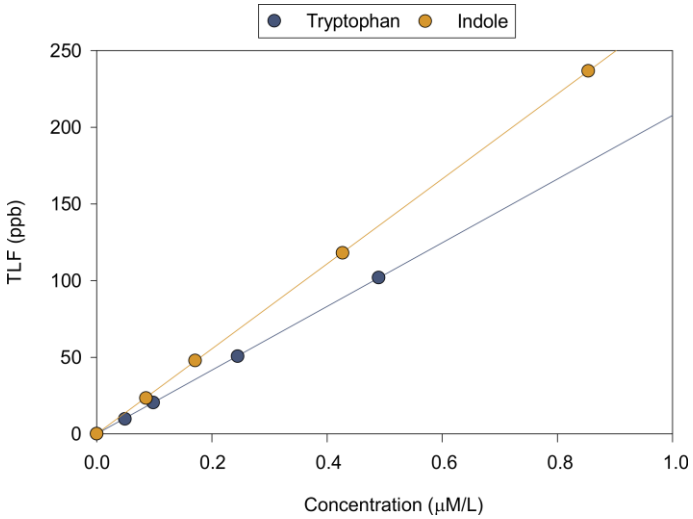

Figure 1-4. Comparison of the TLF intensity emitted by dissolved tryptophan and indole. Analyses were performed on 0, 10, 20, 50, and 100 ppb solutions of both compounds with a UviLux fluorimeter (Chelsea Technologies Group Ltd., UK). The gradients of the regression lines are 278 and 209 for indole and tryptophan, respectively.

Fox *et al*. (2017) investigated the microbial origins of both TLF and HLF. They cultured *E.coli*, amongst other species, for 6 h and collected samples every 30 min for fluorescence and optical cell density. The cultured *E. coli* cells fluoresced at TLF wavelengths and displayed a significant and very strong correlation with cell density over the bacterial growth curve (r^2 0.98). They suggested that although the correlation was strong, this was more likely to be attributed to cell activity rather than the number of cells as tryptophan will be produced as a result of cell multiplication and metabolic processing (Coble et al., 2014; Hogg, 2013). HLF was also produced by *E. coli*; the authors suggest as a metabolic by-product or a secondary meteabolite, with a correlation of r^2 0.86 across the growth curve, although HLF continued to increase during the stationary phase.

Cumberland *et al.* (2012) investigated the use of TLF as a tool for assessing the microbiological quality of river water and effluents. The authors collected a range of effluent and river water samples from the UK. These samples were used to make up a dilution series of 10:1, 100:1, 1000:1 and 10000:1, which were analysed for TLF and the bacterial indicators: total coliforms, *E. coli*, presumptive coliforms and heterotrophic bacteria and compared. Correlations were reported between TLF and heterotrophic bacteria (r^2 0.81), total coliforms (r^2 0.78) and *E. coli* (r^2 0.65). It was also noted that *E. coli* displayed the greatest TLF per unit of bacteria.

Fluorescence at TLF and HLF wavelengths is not unique to TTCs such as *E. coli*. Tryptophan residues in proteins are ubiquitous in bacteria, and many bacterial cells directly fluoresce at TLF wavelengths (Bronk and Reinisch, 1993; Dalterio et al., 1986; Dalterio et al., 1987; Dartnell et al., 2013; Fox et al., 2017; Seaver et al., 1998; Sohn et al., 2009). Furthermore, multiple species excrete TLF and HLF fluorophores including those that are omnipresent in freshwater systems such as *Pseudomonas aeruginosa* (Elliott et al., 2006a; Fox et al., 2017; Kida et al., 2019; Nakar et al., 2019). Therefore, elevated TLF/HLF may be indicative of elevated numbers of non-FIO bacteria, which are not necessarily an indicator of risk to human health. For example, Bridgeman *et al*. (2015) demonstrated a correlation between TLF and total bacterial cells (TBCs), but not *E. coli*, in predominantly treated drinking water. Additionally, various species of cyanobacteria, diatoms and green algae also fluoresce at TLF wavelengths (Determann et al., 1998; Henderson et al., 2008; Nguyen et al., 2005).

It is unclear whether the bulk of TLF or HLF fluorophores in the environment are extracellular or intracellular in groundwater. Some research indicates TLF is associated with free amino acids, i.e. extracellular, (Yamashita and Tanoue, 2003) whilst others have suggested it is partially bound in proteins or cell walls, i.e. intracellular (Determann et al., 1998). In the laboratory, Fox *et al.* (2017) reported that at least 75% of TLF was intracellular when *E. coli* were cultured under controlled conditions. This evidence was supported further by similar laboratory work using *Pseudomonas* sp. where >68% of TLF was intracellular (Fox et al., 2021). In an assortment of surface waters, Baker *et al.* (2007) showed 32-86% of TLF was lost following filtration through a 0.2 µm membrane indicating a substantial proportion was associated with particulate and cellular material. Samples from the River Leith, NW England, in an area featuring groundwater-surface interaction showed a slightly lower TLF loss following filtration of 20-40 % (Bieroza and Heathwaite, 2016a), albeit through a larger pore-size 0.45 µm membrane that would not remove all microbes. For HLF, Fox *et al.* (2021) reported >80% of fluorescence was extracellular in their laboratory model system using *Pseudomonas* sp. and Baker *et al.* (2007) reported 70-96% of the signal remained after filtration from their range of surface waters. In groundwater, we might expect an even higher proportion of TLF and HLF to be extracellular due to

natural filtration during recharge and subsurface flow, as well as, typically, a lower microbial biomass (Griebler and Lueders, 2009; Marmonier et al., 1995; Pedersen, 2000; van Driezum et al., 2018).

1.4 Aims and objectives

The principal aim of this research is to evaluate the use of in-situ fluorescence spectroscopy as an instantaneous indicator of faecal contamination risk in drinking water. This goal will be achieved by:

1. Testing TLF/HLF as indicators of faecal contamination determined by FIOs across a range of settings (Chapters 3, 4, 5, 7, 8).
2. Demonstrating online TLF/HLF as real-time indicators of faecal contamination risk at public water supplies (Chapter 5).
3. Generating TLF/HLF thresholds to infer FIO presence-absence that could be implemented widely (Chapter 4).
4. Investigating temporal dynamics in TLF/HLF-FIOs relationships (Chapters 3, 8).
5. Exploring the uniqueness of TLF/HLF to FIOs, as opposed to the wider bacterial population (Chapters 5, 7, 8).
6. Examining the intracellular/extracellular nature of TLF/HLF fluorophores (Chapter 6).
7. Evaluating TLF/HLF against UNICEF/WHO faecal indicator criteria, including limit of detection, usability, and interferents (Chapter 4).

Chapter 2 - Methodological overview

2.1 Study areas and data availability

Data were collected from multiple study areas, with a focus on Africa, the WHO region with the highest proportion of the population consuming faecally contaminated water (Table 1-1). African locations were selected opportunistically according to funding availability in Kenya (Kisumu), Senegal (Dakar), Uganda (Lukaya), and Zambia (Kabwe). A UK project solely designed around the application of fluorescence spectroscopy at public water supplies enabled a real-time technological application, which would have been challenging to undertake in Africa. Datasets were also collated for analysis from other projects operating simultaneously alongside this PhD in India (Sorensen et al., 2016) and Malawi (Ward et al., 2020), as well as a previous project in South Africa (Baker et al., 2015). These datasets from seven countries allowed in-situ fluorescence spectroscopy to be explored as an indicator of faecal contamination risk across a range of aquifers in different climates that are exposed to varying degrees of pollution pressures. Faecal contamination was primarily assessed by culturing FIOs.

2.2 Faecal indicator organisms (FIOs)

Thermotolerant (faecal) coliforms (TTCs) were selected as the FIO of contamination to assess the presence and extent of faecal contamination risks in groundwater for all research undertaken outside of the UK. TTCs include the preferred FIO *E. coli* (WHO, 2017b), which was used in the UK water industry work (Chapter 5), in addition to other genera such as *Klebsiella* spp. that are less likely to originate from a faecal source (Leclerc et al., 2001). Nevertheless, TTCs are considered acceptable FIO alternatives to *E. coli* by the WHO (2017b), as the majority of TTCs comprise *E. coli* in most circumstances. In tropical groundwater impacted by inadequate sanitation, 90-99% of all TTCs have been shown to be *E. coli* (Howard et al., 2003; Leclerc et al., 2001). The Howard *et al*. (2003) study, where 99% of TTCs were confirmed as *E. coli*, was undertaken in shallow groundwater in Kampala in

a near identical climatological and hydrogeological setting to that as the Ugandan study area (Chapter 8).

TTCs were isolated and enumerated using the membrane filtration method with Membrane Lauryl Sulphate Broth (MLSB, Oxoid Ltd, UK) as the selective medium. A volume of water (0.1-100 mL) was passed through a 0.45 µm cellulose nitrate filter (GE Whatman, UK). The filtrate volume was selected according to the corresponding TLF measurement and any previous TTC analyses at the source to ensure colonies were not too numerous to count (TNTC), whilst always attempting to maximise the volume. The filter was placed on an absorbent pad (Pall Gelman, Germany) saturated with MLSB broth in a plate and incubated at 44 °C for 18–24 h. Plates were inspected within 15 mins of removal from the incubator and all cream to yellow colonies greater than 1 mm considered TTCs. Where plates were TNTC the analysis was repeated the following day using a smaller volume of the remaining sample that had been kept refrigerated at 4°C.

2.3 In-situ fluorescence spectroscopy

There are several commercially available options available to collect in-situ fluorescence spectroscopy data (Khamis et al., 2020; Khamis et al., 2015). The UviLux fluorimeters produced by Chelsea Technologies Group Ltd were selected because they contain a photomultiplier, which improves the minimum detection limit (MDL). For example, for TLF, the MDL in the laboratory is 0.17-0.19 µg/L, which is an order of magnitude lower than competitors (Khamis et al., 2020; Khamis et al., 2015). This lower detection limit is important for groundwater fluorescence studies when fluorescence can be very low.

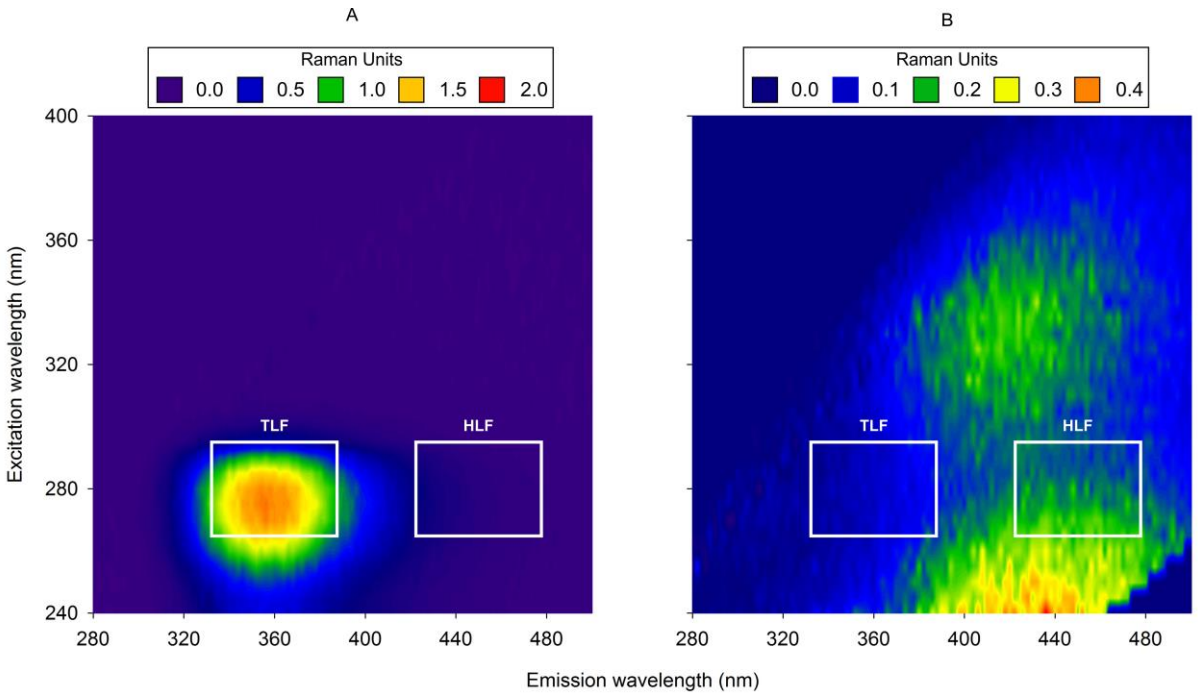

Figure 2-1 Optical regions of TLF and HLF UviLux fluorimeters shown on (A) an Excitation-emission matrix (EEM) of a 100 ppb tryptophan standard; (B) mean EEM from repeated sampling of a UK public water supply demonstrating no clear protein-like peak, but dual humic-like peaks that overlap into the TLF region.

Two UviLux fluorimeters were used targeting TLF and HLF, as indicators of faecal contamination risk. The TLF and HLF fluorimeters targeted excitation-emission values of 280/360 (λ_{ex}/ λ_{em}) and 280/450 nm (λ_{ex}/λ_{em}), respectively. Whilst the HLF λ_{em} targeted the established peak, the λ_{ex} was matched to that of TLF to also investigate evidence of optical overlap between the two regions (Figure 2-1), as opposed to centres of peaks typically observed at λ_{ex} <260 and 320-360 nm (Fellman et al., 2010). The bandpass filters for λ_{ex} and λ_{em} were ± 15 and ± 27.5 nm, respectively, for both fluorimeters (Figure 2-1).

The TLF and HLF fluorimeters are factory calibrated using dissolved tryptophan and pyrene tetrasulphonic acid in deionised water, respectively. The TLF fluorimeter units were originally equivalent to dissolved tryptophan intensity in parts per billion (ppb) but then subsequently changed to quinine sulphate units (QSU), the unit used by the HLF fluorimeter, to allow calculation of TLF:HLF ratios. QSU are a standard unit produced by standardising the factory calibrationd using the

fluorescence at λ_{ex} 347.5 nm and λ_{em} 450 nm from 1 ppb of quinine sulphate dissolved in 0.105 M perchloric acid.

Initial laboratory testing of the TLF fluorimeter using dissolved L-tryptophan (Acros Organics, USA) in ultrapure water demonstrated excellent linearity (R^2 >0.999) (Figure 2-2), repeatability of ±0.12-0.29 µg/L up to a concentration of 50 µg/L (Table 2-1), and an excellent relationship with a bench top Varian™ Cary Eclipse fluorescence spectrophotometer (R^2 >0.99).

Table 2-1 Repeatability of tryptophan-like fluorescence across 12 concentrations based on 10 measurements within each standard.

Tryptophan standard (µg/L)	3σ (µg/L)
1	0.16
2	0.16
3	0.12
4	0.18
5	0.16
10	0.29
20	0.23
50	0.20
100	0.40
200	1.23
400	2.11
1000	3.51

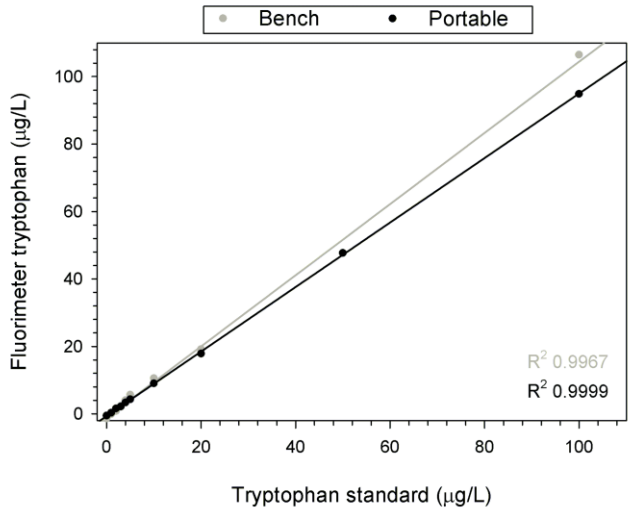

Figure 2-2 Comparison of bench and portable fluorimeters with ten synthetic tryptophan standards.

Chapter 3 - In-situ tryptophan-like fluorescence: a real-time indicator of faecal contamination in drinking water supplies

3.1 Introduction

This chapter addresses **Objective 1** and documents the first work to pilot the use of in-situ TLF fluorimeters for the rapid assessment of the microbiological quality of drinking water supplies, with a focus on groundwater in a low-income country. The aims of this work are to investigate the application of a portable fluorimeter to indicate i) whether a groundwater supply was faecally contaminated; ii) the extent of any contamination; and iii) compare TLF as an indicator of contamination against other common indicators such as NO_3, Cl, turbidity, and sanitary risk scores (SRS).

The chapter is published as:

Sorensen, J.P.R., Lapworth, D.J., Marchant, B.P., Nkhuwa, D.C.W., Pedley, S., Stuart, M.E., Bell, R.A., Chirwa, M., Kabika, J., Liemisa, M. and Chibesa, M., 2015. In-situ tryptophan-like fluorescence: a real-time indicator of faecal contamination in drinking water supplies. Water Research, 81, 38-46.

3.2 Materials and Methods

3.2.1 Study site

Kabwe is located in Zambia's Central Province approximately 150 km north of the capital Lusaka. It has a population of over 200,000, with a high proportion residing within informal settlements on the outskirts of the city such as Makululu – regarded as one of the largest slums in southern Africa with an estimated 46,000 inhabitants (LgWSC, 2014). The city is predominantly underlain by several hundred metres of either the Lower Roan Group (quartzite, schist and pelite) or Upper Roan Group (dolomite). The bedrock is concealed beneath continuous saprolite and laterite superficial deposits that are typically 5-20 m thick (Houston, 1982). Groundwater is generally encountered 5-10 m below

ground level, with superficial deposits typically in hydraulic connection with the deeper aquifer within the karstic bedrock. The local climate is sub-tropical with rainfall exhibiting strong seasonality: 95% falls between mid-November and mid-April (Nkhuwa et al., 2006). Natural surface waters are absent, as rainfall rapidly infiltrates into the subsurface.

Groundwater is the major source of drinking water supply for the city. The centralised supply system abstracts groundwater from deep boreholes within peri-urban wellfields, which is then treated and piped directly to properties, or dispensed via communal taps and water kiosks within the informal settlements. Households frequently also self-supply groundwater to some extent as the centralised supply can be unreliable and is charged on a per volume basis. Within informal settlements this is generally restricted to vulnerable shallow hand-dug wells and illegal connections to pre-treated water within the centralised supply network. In more affluent areas, self-supply includes tapping the bedrock or superficials through deeper boreholes or shallow wells, respectively, with limited use of piped supplies.

Low levels of sanitation coverage are a major cause for concern within newer parts of Kabwe. This includes the burgeoning informal settlements where coverage is estimated at less than 11% of properties and restricted to pit latrines (LgWSC, 2014). In established parts of the city, the sewerage network is more extensive, but is ageing, in need of investment, and therefore prone to leakage and overflow. Furthermore, waste collection is limited to the larger businesses in the town centre. Typically, household waste is buried within gardens, burned, or illegally dumped. It should be noted that informal settlements are beginning to encroach into the wellfield areas, with concerns over the potential threats to the city's groundwater resources in the medium to long-term.

3.2.2 Groundwater sampling and analysis

A total of 117 groundwater samples was obtained from a mixture of supplies that were distributed across the city (Figure 3-1). These comprise 55 samples in the dry season (September 2013) and 62 in the subsequent wet season (January 2014), of which 45 were obtained from the same supplies.

Dry season sampling included 25 boreholes and 30 shallow wells whereas in the wet season 26 boreholes and 36 shallow wells were investigated. These supplies included the city wellfields (K1-12), a mixture of private supplies within both higher and lower cost residential areas, as well as those in the industrial zone (K26-28).

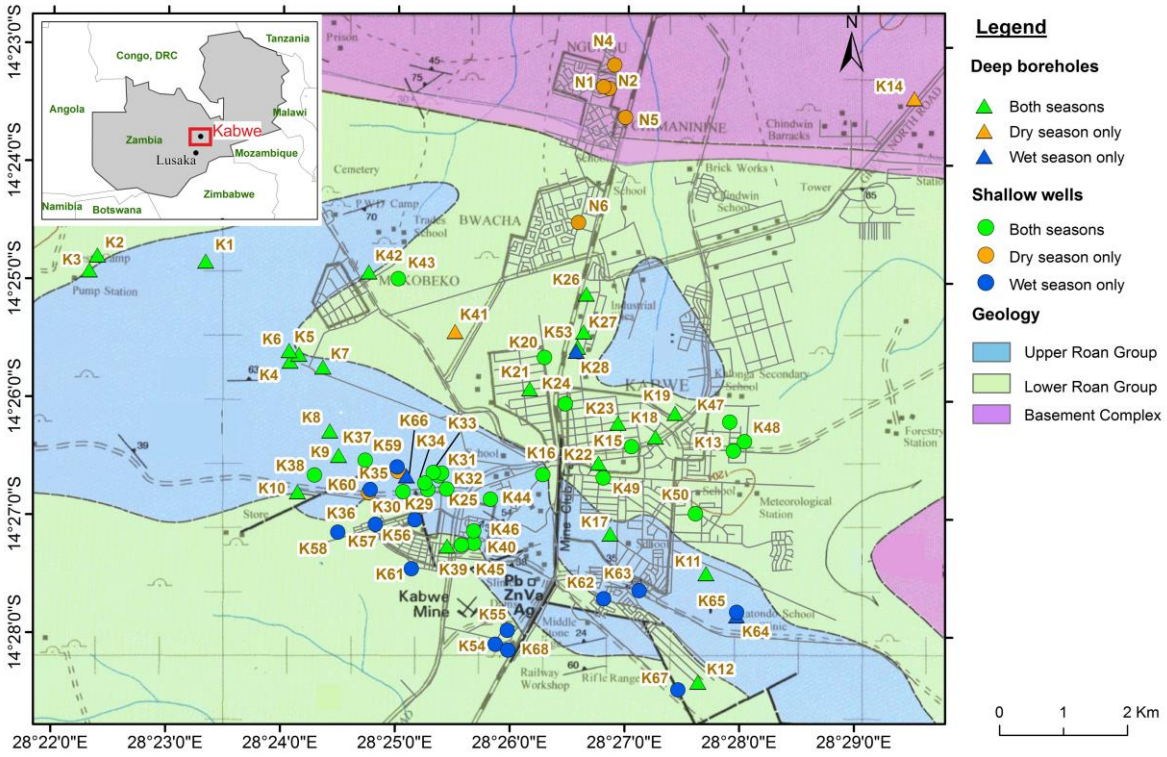

Figure 3-1 Bedrock geology and sampled groundwater supplies in Kabwe.

Groundwater samples were obtained once field measurements of pH, specific electrical conductance (SEC), Eh, dissolved oxygen (DO) and temperature had stabilised during pumping. Turbidity was also measured on an agitated pumped sample, as settling was rapid. In-situ TLF was undertaken by sampling 5 L of groundwater and immersing a portable UviLux Fluorimeter in the dark (Section 2.3).

Microbiological samples were collected in sterile 60 mL brown-glass bottles. Brief interviews with the well owners, or other informed persons, confirmed that the supplies sampled had not been recently chlorinated, or were chlorinated at a point further down the distribution system. Samples were stored in a cool box and transported back to the laboratory for analysis. Typically 50 mL (giving a limit of detection of 2 cfu/100 mL), or an appropriate dilution of the sample, were filtered and

incubated within seven hours of collection for TTCs (Section 2.2). The results from 26 duplicate measurements were used to calculate the average Relative Standard Deviation of Reproducibility (RSD_{RC}) and the Uncertainty of Measurement of the microbiological analysis.

Hydrochemical samples were collected for Cl, NO_3, NH_4, SO_4, soluble reactive P (SRP), and total dissolved P (TDP) in 60 mL HDPE bottles after passing through a 0.45 µm nitrocellulose filter, with the bottle for SRP and TDP pre-treated with 0.45 g of potassium peroxodisulfate preservative. Samples for dissolved organic carbon (DOC) were passed through 0.45 µm silver filters into sterile acid washed glass vials. All samples were stored within cool boxes before transfer to a refrigerator at the end of each day. Cl, NO_3, SO_4 were analysed by ion chromatography. NH_4 was determined by automated calorimetry using the indophenol blue method. Total dissolved phosphate and soluble reactive phosphate were determined via the methods of Eisenrech et al. (1975) after treatment with sulphuric acid. DOC analysis was conducted using a Thermalox™ C analyser after acidification and sparging.

3.2.3 Sanitary risk score

A sanitary risk assessment was undertaken at every source during the wet and dry season surveys. The assessment was carried out using the methods published by the WHO and the sanitary risk assessment forms appropriate to the supply type (WHO, 1997). The supplies and surrounding areas were assessed for the presence of defined hazards, such as latrines and waste dumps, and pathways for contaminants to reach the sample point, using pre-set questions. The sanitary risk score (SRS) was calculated from the number of positive responses.

3.2.4 Statistical analyses

Statistical models and tests were performed with MATLAB v14.1 to determine the extent to which TLF concentration and other predictors indicate the presence of TTC contamination and the accuracy with which they could be used to model TTC counts. The TTC count in water supply i was denoted y_i

and the binary variable, y_i^{I}, indicated whether or not the TTC exceeded the detection limit of 2 counts, i.e.

$$y_i^{\mathrm{I}} = \begin{cases} 1 & \text{if } y_i \geq 2 \\ 0 & \text{otherwise} \end{cases}.$$

A series of Mann-Whitney rank sum tests (Mann and Whitney, 1947) were used to assess whether or not the values of each of the predictors differed between the water supplies with TTC contamination ($y_i^{\mathrm{I}} = 1$) and those which were uncontaminated ($y_i^{\mathrm{I}} = 0$). The non-parametric Mann-Whitney rank sum test does not require the predictors to conform to a particular statistical distribution. Therefore, it was preferred to the T-test which requires the assumption of a Gaussian distribution. In each test the null hypothesis was that the expected value of the predictor was the same for contaminated and uncontaminated supplies.

The probability that a water supply was contaminated with TTCs was represented by a logistic regression model (Dobson, 2001). If p_i is the probability that there is contamination at supply i, then the logistic model is written:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^{q} \beta_j x_{ij}, \qquad (1)$$

where x_{ij} is the value of predictor j at water supply i and the β_j are coefficients which, in this study, were estimated by least squares (Draper and Smith, 1981). A stepwise regression algorithm (Draper and Smith, 1981) was used to decide which of the available predictors should be included in the model. This approach guards against too many predictors being included. In such circumstances the model is said to be overfitted, meaning that it is overly suited to the intricacies of the data set to which it has been fitted but will not achieve the same accuracy on independent validation data.

The stepwise algorithm adds one predictor to the model at a time. It considers all of the available predictors not yet included in the model in turn and selects the one which causes the largest decrease in the mean squared residuals. If this decrease is deemed significant for a p-value of 0.05, according to a F-test (Draper and Smith, 1981), this parameter is added to the model. The iterative process continues with the remaining parameters until none of them lead to a significant improvement to the model. The stepwise algorithm was implemented using the MATLAB function 'stepwiseglm'. The observed values of TLF were highly skewed (Figure A-1). Therefore these data were log transformed prior to the estimation of regression models. Note that a positive shift of 0.1 was required to ensure that the logarithm was always defined.

A classifier of TTC contamination was then formed by designating all water supplies with p_i greater than some threshold p^* as contaminated. There is uncertainty associated with such a classification. False positive errors occur when the classifier erroneously suggests that the supply is contaminated whereas false negative errors occur when the supply is erroneously classified as uncontaminated. Similarly, true positive and true negative results refer to correctly classified contaminated and uncontaminated supplies.

The effectiveness of a classifier can be assessed in terms of the receiver operating characteristic curve (Hanley and McNeil, 1983). This is a plot of the proportion of true positive results against the proportion of false positive results as p^* is varied. The area under this curve is a measure of the effectiveness of the classifier. It will be 1 for a perfect classifier and 0.5 if the classifier is performing no better than a random choice.

The TTC counts were represented by a linear regression model:

$$y_i = \beta_0 + \sum_{j=1}^{q} \beta_j x_{ij} + \varepsilon_i, \qquad (2)$$

where the ε_i were independent realizations of Gaussian random variable with zero mean and variance σ^2. The β_i and σ^2were the model parameters which in this study were estimated by least squares. Again, the optimal predictors for the model were selected by stepwise regression. When a linear regression model was estimated for TTC counts, the residuals were highly skewed, contrary to the assumptions required by the model. Therefore a model was instead estimated for the logarithm of the TTC counts.

A regression model such as Equation 2 determines the entire probability density function for the TTC count at a water supply from the values of the predictor variables from that supply. Therefore, the model can be used to determine the probability that the TTC count falls within particular intervals. In this case, the five risk classes outlined by the WHO (1997) to prioritise interventions in areas where microbiological contamination of water supplies is widespread were used.

3.3 Results

3.3.1 Characteristics of TTC contaminated water supplies

The results of the Mann-Whitney rank Sum tests suggested that the mean TLF concentration was significantly ($p<0.001$) larger in water supplies contaminated with TTCs as opposed to water supplies where they were absent (Figure 3-2). The relationship remained significant at the $p<0.001$ level when the supplies from each season were considered separately and when only boreholes were considered. No significant difference was evident when only shallow wells were considered but, since there were only five shallow wells where TTCs were absent, this finding is likely to be the result of insufficient data. Overall, the median TLF concentrations were 0.9 and 9.2 µg/L in uncontaminated and contaminated supplies, respectively.

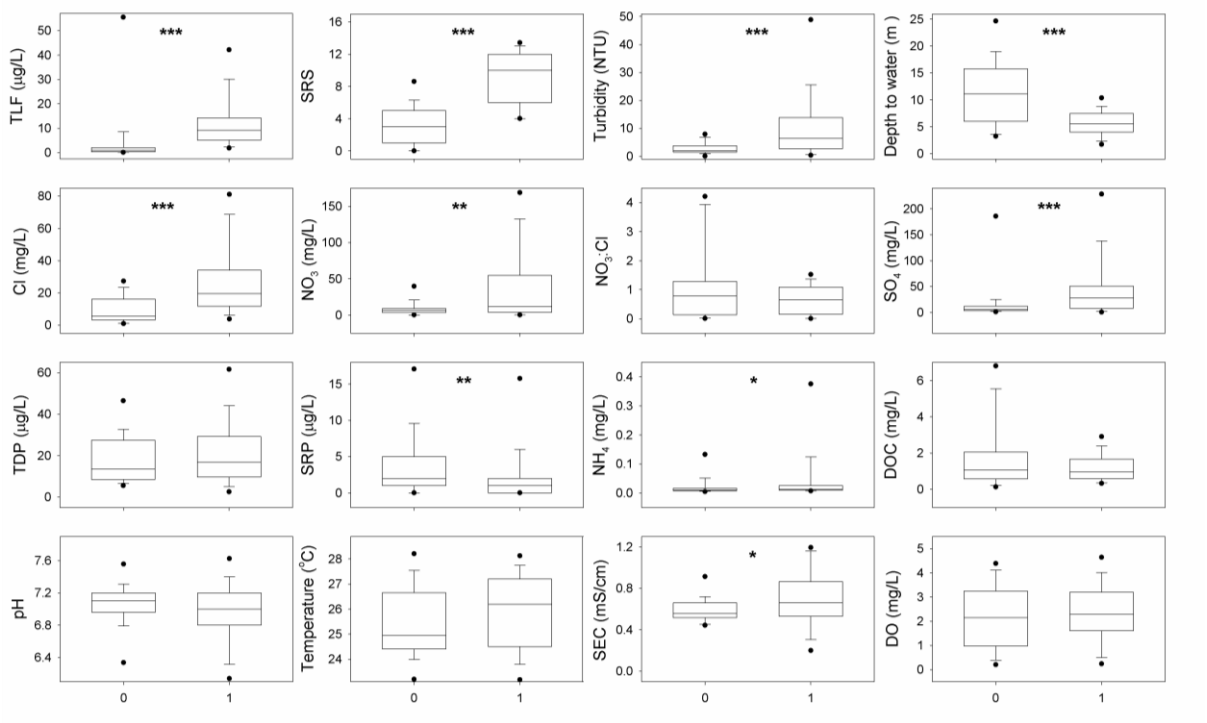

Figure 3-2 Box plots of tryptophan-like fluorescence and other variables for the presence or absence of thermotolerant coliforms. P-values are denoted as * <0.05, ** <0.01, and * <0.001. Dots indicate 5th and 95th percentiles.**

Other variables showing a highly significant difference ($p<0.001$) included SRS, turbidity, depth to water, Cl and SO_4 (Figure 3-2). Phosphate species or DOC were not appreciably elevated in contaminated water supplies, although NH_4 was significant ($p=0.019$). Temperature, pH and turbidity typically ranged between 23.8-27.7 ^oC, 6.4-7.4, and 1-20 NTU, respectively, across the dataset. All variables correlated with faecal contamination were also significantly correlated with TLF concentration (notably SRS, NO_3, and NH_4), with no relationship observed with DOC, pH, DO, or temperature (Figure A-2).

3.3.2 Predicting the presence/absence of TTCs

Transformed (and raw) TLF was a significant predictor ($p<0.001$) of the presence/absence of TTC according to the logistic regression model (Equation 1). Other indicators were significantly correlated with the probability of TTC contamination (Table 3-1), including Cl and the SRS. However, transformed TLF was the first predictor to be included in the logistic regression model and the results of the F-tests demonstrate that the addition of other predictors did not significantly improve the

model (Table 3-2). The findings were the same when data from the wet and dry seasons were considered individually, as well as when boreholes and shallow wells were considered separately. Therefore, for the predictors investigated in this study, transformed TLF appears to be the best indicator of the presence/absence of TTCs.

Table 3-1 Estimated coefficients and p-values for the hypothesis that the coefficients are zero for single-predictor models of probability of the presence and number of thermotolerant coliforms. Areas under the receiver operator characteristic curves for single predictor classifiers are also shown. Classifiers are only formed for predictors that are significantly related to contamination.

Predictor	Logistic Regression			Linear Regression	
	Area	β	p-value	β	p-value
Transformed T_1	0.92	2.12	<0.001	1.78	<0.001
SRS	0.90	0.61	<0.001	0.63	<0.001
Turbidity	0.76	0.48	0.019	1.05	<0.001
Depth to water	0.80	-2.86	<0.001	-2.42	<0.001
Cl	0.86	1.63	<0.001	1.61	<0.001
NO_3	0.73	0.35	0.019	0.86	<0.001
NO3:Cl	NA	-0.34	0.221	0.36	0.331
SO_4	0.71	0.31	0.050	0.33	0.128
TP	NA	0.25	0.382	-0.09	0.820
SRP	0.63	-0.41	0.024	-0.47	0.039
NH_4	0.64	2.88	0.174	2.04	0.016
DOC	NA	-0.74	0.057	-0.59	0.286
SEC	NA	0.00	0.502	0.00	0.101
DO	NA	-0.17	0.423	0.21	0.489

Table 3-2 P-values for the addition of predictors to the final logistic and linear regression models. NA means that the predictor is already in the model.

Predictor	Logistic regression p-values	Linear regression p-values
Transformed T_1	NA	NA
SRS	0.18	NA
Turbidity	0.47	0.44
Depth to water	0.17	0.11
Cl	0.07	0.85
NO_3	0.53	NA
NO3:Cl	0.96	0.62
SO_4	0.84	0.52
TP	0.25	0.98
SRP	0.19	0.79
NH_4	0.59	0.85
DOC	0.76	0.74
SEC	0.19	0.32
DO	0.91	0.25

When the logistic regression model with only TLF as a predictor was used to form a classifier of TTC contamination, the area under the receiver operator characteristic curve was 0.92 (Figure 3-3a). This is much closer to the perfect classifier value of 1 than the random selection value of 0.5. The corresponding areas for other single predictor models ranged from 0.63 to 0.90 (Table 3-1). There is a trade-off between false positive and false negative errors as the classifier's threshold on TLF readings is varied (Figure 3-3b). When the threshold is 3.5 µg/l both the false positive and false negative error rates are approximately 0.14. However, there are a small number of sites where the contamination status inferred from TLF and TTC differ. For example there are two water supplies where the TLF concentration is less than 1.5 µg/l but the TTC count is 2 and 10 cfu/100 mL. Note that since TLF is the only predictor and the relationship between these readings and the probability of contamination is monotonic, putting a threshold on these readings is equivalent to putting a threshold on the probability of contamination.

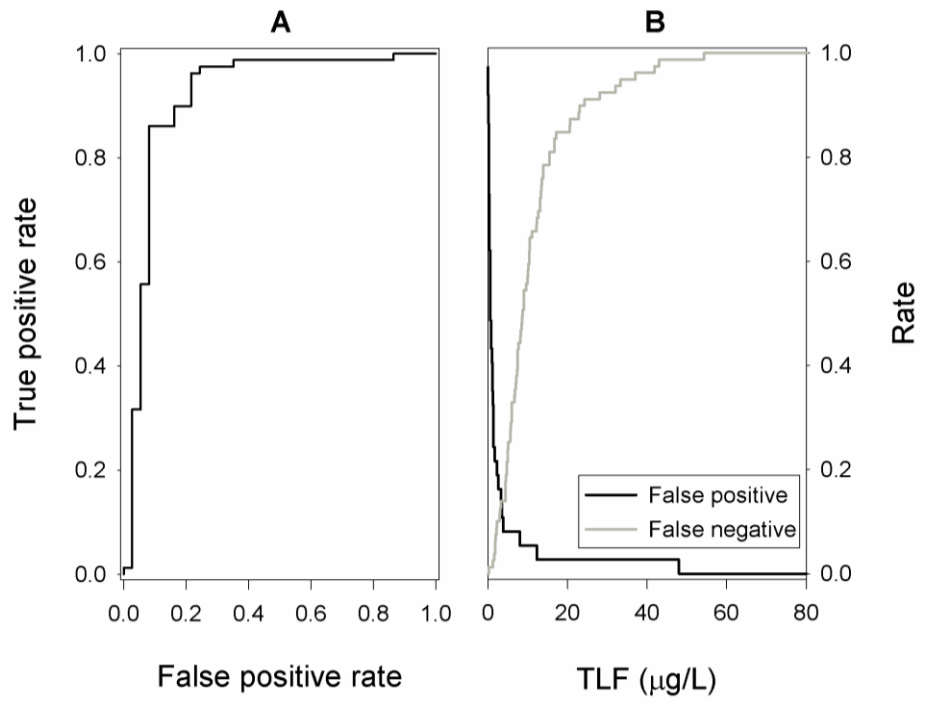

Figure 3-3 (a) Receiver operator characteristic curve for classifier of the presence or absence of TTC using transformed TLF as the only predictor; (b) error rates for this classifier.

3.3.3 Predicting the number of TTCs

Transformed TLF was a significant predictor ($p<0.001$) of the number of TTCs (Table 3-1). Other similarly significant indicators included Cl, NO_3, SRS, and turbidity. However, the stepwise regression algorithm included transformed TLF first, signifying it is the most important predictor. When transformed TLF is the only predictor the estimated model has an R^2 of 0.57 and is written:

$$\ln y_i = 1.36 + 1.78 \ \ln(\mathrm{TLF} + 0.1) \quad (3)$$

NO_3 and SRS were added at the second and third stages of the stepwise algorithm. These predictors cause the R^2 to increase to 0.67 and the final model is written:

$$\ln y_i = -0.34 + 1.11 \ \ln(\mathrm{TLF} + 0.1) + 0.35 \ \mathrm{NO_3^-} + 0.25 \ \mathrm{SRS} \quad (4)$$

The addition of other predictors did not significantly improve the model (Table 3-2). This might indicate that the predictors included in Equation 4 accounted for the other observed significant relationships.

Additionally, there is an increase in the median transformed TLF concentration with elevated WHO risk class, although the levels of scatter produce overlap between all classes (Figure 3-4). Nevertheless, the single-predictor linear regression model using TLF can be used to estimate the probability of a water supply belonging to each risk class (Figure 3-4). This suggests that as TLF concentration increases there is a greater probability of the water supply being within higher risk groups.

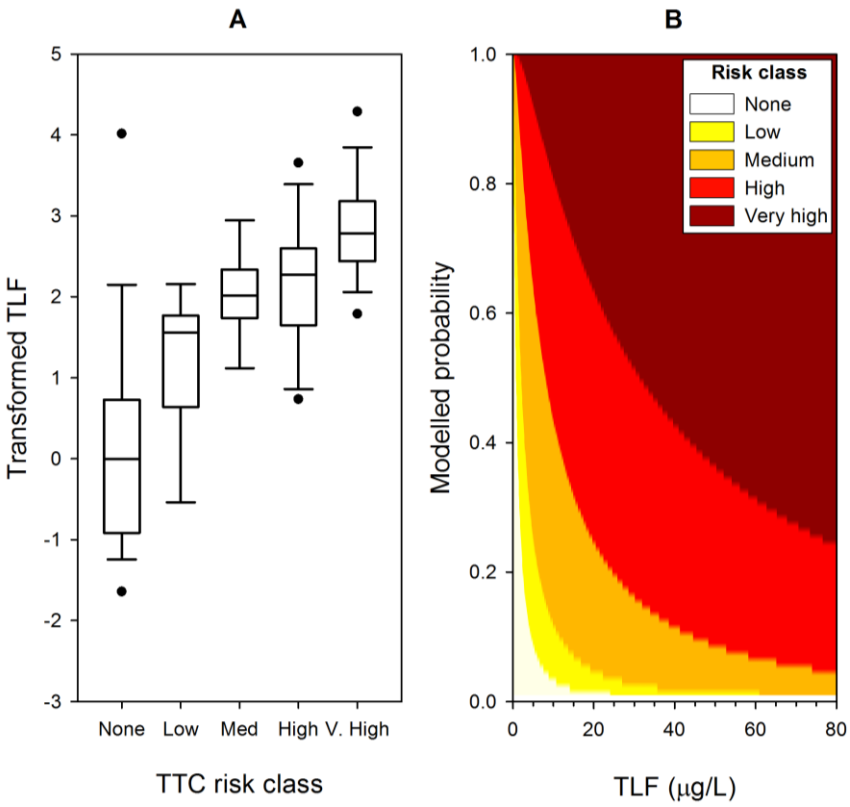

Figure 3-4 (a) Box plot of transformed TLF for each TTC risk class with dots indicating 5th and 95th percentiles where sample number exceeds nine; (b) modelled probability that a water supply belongs to each risk class for a given TLF concentration. Risk classes are defined by TTC count as <2 (none), 2 to <10 (low), 10 to <100 (medium), 100 to <1000 (high), >1000 cfu/100 mL (very high).

3.3.4 Risk classes for water supplies across Kabwe estimated by TTC count or modelled TLF

Deeper boreholes are considered less at-risk than shallow wells across the city using either modelled TLF or TTC count (Figure 3-5). Boreholes are most at-risk within the central districts, with these risks elevated in the wet season, and higher when allocated using modelled TLF compared to the TTC count. There are also notable seasonal changes in risk in the city's centralised peri-urban water supply boreholes (K1-12) using modelled TLF which are not apparent from the TTC data (Figure 3-5). In the dry season, there are no risks by TTC count or modelled TLF in any of the boreholes, with the exception of K3 and K11 being considered as low risk by modelled TLF. Following the onset of the rains, modelled TLF suggests increases in risk at the majority of these boreholes, including some located beyond the city limits (K2, K6, and K12). Furthermore, high risks are indicated at those sites closest to the urban area (K8, K9 and K11). On the other hand, TTC data only suggest elevated medium and low wet season risks at K8 and K9, respectively.

Shallow wells are typically at medium or greater risk perennially in the central districts of Kabwe by modelled TLF (Figure 3-5). However, TTC counts suggest many of these wells have no risk during the dry season, but medium or greater during the wet season. Nevertheless wells within Makululu generally remained at least at high risk during both seasons using either modelled TLF or TTC counts. Elsewhere, other wells also considered at a minimum of high risk by both TTC count and modelled TLF in the dry season, predominantly remain so in the wet season.

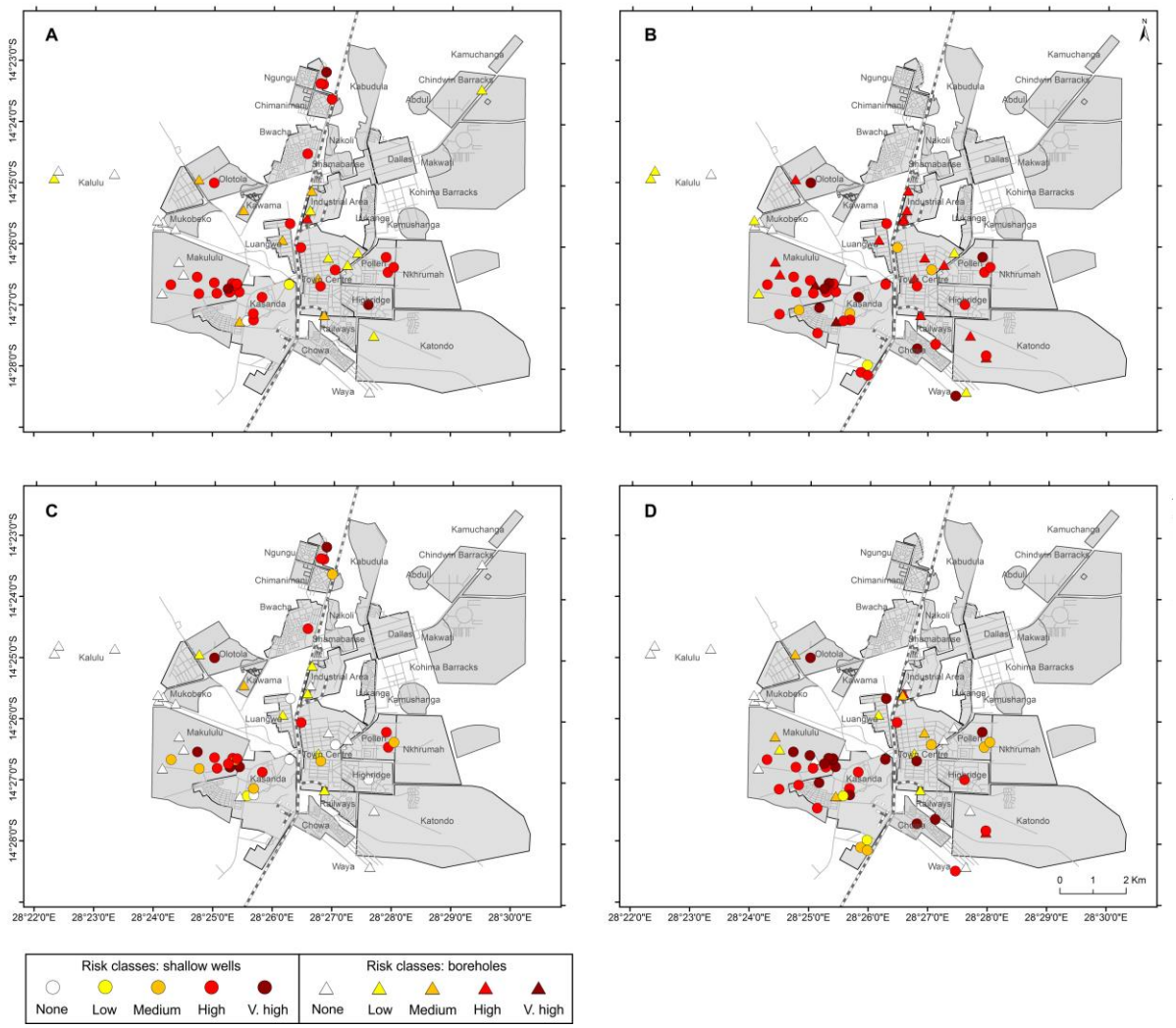

Figure 3-5 Water supply risk class across Kabwe allocated using modelled TLF (a) dry season, (b) wet season; and TTC data (c) dry season, (d) wet season. TTC risk classes are defined as <2 (none), 2 to <10 (low), 10 to <100 (medium), 100 to <1000 (high), >1000 cfu/100 mL (very high).

3.4 Discussion

3.4.1 Uncertainty in TTC counts and TLF concentrations

Doubts have been expressed about the reliability of TTCs to indicate faecal contamination in water. Although the TTC group includes the species *E.coli,* which is generally considered to be specific for faecal contamination, it also includes other genera such as *Klebsiella* and *Citrobacter* which are not necessarily of faecal origin and can emanate from alternative organic sources such as decaying plant materials and soils (WHO, 2017b). Nevertheless, studies have shown that greater than 90% of thermotolerant coliforms are *E. coli* (Dufour, 1997 cited in Leclerc *et al.* (2001)) and as high as 99% in groundwater impacted by poor environmental sanitation in Africa (Howard et al., 2003). Therefore this is not considered a significant source of uncertainty.

The uncertainty of measurement of the microbiological analysis was calculated using the RSD_{RC}, which is the average relative standard deviation of reproducibility of a series of duplicate analyses. The RSD_{RC} calculated for the duration of the sampling programme was 0.14. Using a coverage factor of 2.056, which was determined from the number of duplicate analyses used to calculate RSD_{RC}, the 95% confidence intervals for any result could be calculated. For example, a result of 100 cfu/100ml would have an upper and lower limit of 197 and 51 cfu/100ml respectively. This figure is slightly higher than might be anticipated in a laboratory accredited to International Standard ISO17025:2005, but is not unusual for microbiological analysis.

False-negatives and underestimates of TTCs may occur where environment conditions within certain groundwater supplies are unfavourable for the survival of bacteria, but TLF is less affected. There were three supplies where TTCs were absent but TLF exceeded 10 µg/L. At two of these sites the water differed from a typically neutral pH; being 5.1 and 10.5 in supplies K40 and K55, respectively. Moreover, K40 also contained high concentrations of heavy metals which have a toxic influence on bacteria (Foppen and Schijven, 2006), yet also quench TLF fluorescence to a degree (Tabak et al., 1989). These two sites actually account for two of the five shallow wells where TTCs were absent.

It was confirmed that none of the supplies had recently been chlorinated before sampling. However, a parallel study confirmed the presence of trihalomethanes (by-products of chlorination) at K16 and K30 in the dry season (Sorensen et al., 2015c), although this discrepancy could indicate the leakage of chlorinated mains water into the shallow aquifer. In addition to sterilising the supply, chlorination is generally considered to quench TLF (Henderson et al., 2009). However, the results do not appear anomalous with K16 having a TLF of 1.5 µg/L and TTC count of <2 cfu/100 mL and K30 having a TLF of 13.4 µg/L and TTC count of 700 cfu/100 mL.

Other environmental factors that may influence the intensity of TLF fluorescence in the environment did not vary appreciably. Thermal quenching may have been limited to approximately 10-20% over the typical temperature range based on previous results using bacteria cultures (Elliott et al., 2006b). This assertion is further supported by a laboratory study using dissolved tryptophan, where Khamis et al. (2015) demonstrated a 0.65 and 1.14 µg/L reduction in TLF for every 1 ^{o}C increase in temperature in 10 and 20 µg/L solutions, respectively. This factor equates to uncertainty of 22-25% across the typical temperature, and it should be re-iterated that there was no observed inverse relationship between TLF and temperature (Figure A-2). pH values are within the 5-8 bracket of minimal impact, with the exception of one supply (Patel-Sorrentino et al., 2002; Reynolds, 2003). Turbidity was less than 200 NTU, hence attenuation of TLF was considered negligible, although low-levels of signal amplification due to light scattering by particles could have been possible (Khamis et al., 2015). Finally, no correction of the data due to inner-filtering (absorption of light by the sample matrix) was necessary (Ohno, 2002), as the mean absorbance of all samples at 254 nm was only 0.04 ±0.21 (3σ).

3.4.2 Is TLF a better indicator of enteric pathogens than TTCs?

Bacteria can be effectively removed during the infiltration and percolation of water due to natural filtering and adsorption. This is also likely to remove a proportion of TLF, given its direct association with bacteria (Elliott et al., 2006a). However, a fraction of the fluorescence is likely to be in a free

dissolved form (Baker et al., 2007), which could mean it is transported more easily and rapidly through porous media. There is evidence for this in Kabwe where increases in modelled TLF risks were greater than TTC risks in the deeper boreholes following the onset of the seasonal rains. This observation suggests TLF can migrate through superficial deposits more effectively. Furthermore, once into the karstic aquifer, increases in modelled TLF risks were identified in boreholes well outside the urban limits where TTCs were absent. A more mobile indicator of enteric pathogens is favourable because viruses are transported more efficiently than bacteria through the subsurface due to their smaller sizes (Borchardt et al., 2007; Hunt et al., 2014).

The perennial persistence of high risks by modelled TLF for shallow wells in central Kabwe, as opposed to the seasonal risks by TTCs, suggests TLF is a less transient faecal indicator in groundwater. The presence of TTCs in groundwater is generally considered evidence of recent faecal contamination, with *E. coli* remaining active for 16-45 days (Taylor et al., 2004). These rates of inactivation are similar for many pathogens but enteric viruses generally inactivate at a slower rate (John and Rose, 2005). Therefore, elevated TLF could be a better indicator of pathogens which are more long-lived in the environment, thus providing a more precautionary estimate of pathogen risk.

It is unsurprising that TLF is more resilient with time than TTCs in groundwater as it is contained within organisms. Therefore, as TTCs and other organisms die-off following their release into groundwater (an unfavourable habitat) many cells will rupture and release TLF in dissolved and more complex phases into the environment. It may remain present in groundwater in these forms as groundwater ecosystems are less active in comparison to surface waters, or be taken up by other microorganisms. Either way, it will continue to contribute to a TLF signal. Furthermore, it has been demonstrated that under nutrient limited conditions *E. coli* release dissolved tryptophan, as they transit from a culturable to a dormant viable but nonculturable state (Arana et al., 2004).

3.5 Conclusions

This is the first study to investigate the use of in-situ tryptophan-like (TLF) for the rapid assessment of the biological quality of drinking water supplies. TLF was significantly elevated in supplies where thermotolerant coliforms (TTCs) were present, alongside many other traditional indicators, but was demonstrated to be the most effective indicator of TTC presence/absence. TLF was also the most significant indicator of the number of TTCs, although the linear regression model could be improved through the inclusion of NO_3 and sanitary risk scores (SRS). A single-predictor linear regression model using TLF was used to estimate the probability of a water supply belonging to each WHO risk class. This highlights that as TLF concentration increases there is a greater probability of a water supply being within the higher risk groups.

The use of in-situ TLF sensors has multiple methodology advantages over bacterial indicators for inferring the presence of enteric pathogens. TLF is likely to be more mobile and temporally resilient in groundwater and, as such a more precautionary indicator of enteric pathogens in groundwater, including where bacterial indicators are absent. Importantly, TLF has the added advantage of potentially detecting the presence of TTC in viable but non-culturable states. Furthermore, the sensors require no reagents and provide instantaneous readings. These characteristics could facilitate their inclusion in real-time pollution alert systems for drinking water supplies throughout the world, for the rapid mapping of enteric pathogen risks in developing regions, and as an initial screening tool to inform and complement further water quality investigations.

Chapter 4 - Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values

4.1 Introduction

It has been demonstrated that TLF is significantly more intense in groundwater-derived sources contaminated with TTCs than those where TTCs were absent (Chapter 3, Sorensen et al., 2016). Moreover, significant positive correlations between TLF intensity and TTC concentration have also been observed in both groundwater- and surface water-derived sources (Chapter 3, Baker et al., 2015; Sorensen et al., 2016), in addition to TLF being elevated in the presence of enteric pathogens (Sorensen et al., 2015b). This chapter collates paired TLF and TTC data from studies in four countries (India, Malawi, South Africa, Zambia) and evaluates TLF as an indicator of TTC across these different settings (**Objective 1**), including generating thresholds to infer FIO presence-absence that could be implemented widely (**Objective 3**). Finally, TLF is evaluated against defined UNICEF/WHO faecal indicator criteria covering limit of detection, usability and interferents (**Objective 7**).

The following parts of this chapter are published as:

Sorensen, J.P., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G. and Ward, J.S., 2018. Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values. Science of the Total Environment, 622, 1250-1257.

4.2 Methods

4.2.1 Available TLF-TTC data

Concurrent TLF and TTC data were collated from a mixture of unpublished and published studies of drinking water in four separate countries: India, Malawi, South Africa and Zambia (Table 4-1). The Indian study was conducted in Bihar State; 150 groundwater sources were selected to achieve spatial

coverage across four villages with an approximate split between those near (<10 m) and those away (>10 m) from recently installed on-site sanitation (Sorensen et al., 2016). The Malawian study predominantly comprised sampling a randomly selected groundwater source in 40 randomly selected villages in each of five districts of the country (Balaka, Machinga, Lilongwe, Nkhotakota, Mzimba). In total, 39 of the 200 randomly selected sources were non-functional and an additional 21 sources were opportunistically sampled near the original randomly selected source. The South African study involved repeated sampling of 28 locations selected upstream, near to, and downstream of expected contributing sources of poor water quality in two surface water catchments in KwaZulu Natal (Baker et al., 2015). The Zambian study comprised sampling 65 groundwater sources across peri-urban, industrial, and lower and higher income residential land uses in the city of Kabwe, of which 46 sources were sampled during both wet and dry seasons (Chapter 3).

Table 4-1 Details of individual TLF-TTC studies.

Study	Source	Fluorimeter	Faecal indicator	No. positive	No. negative	No. total	Published
India	Groundwater	UviLux	TTCs	45	105	150	Sorensen *et al.* 2016
Malawi	Groundwater	UviLux	TTCs	49	133	182	Unpublished
South Africa	Surface water	SMF4	*E. coli*	121	0	121	Baker *et al.* 2015
Zambia	Groundwater	UviLux	TTCs	74	37	111	Chapter 3
Total	-	-	-	289	275	564	-

All studies used a portable fluorimeter targeting TLF on an unfiltered water sample and enumerated either TTCs by membrane filtration (Section 2.2) or *E. coli* by the Colilert® (IDEXX) method (Table 4-1). Although the studies employed differing faecal indicator organisms, Hamilton *et al.* (2005) noted that *E. coli* by the Colilert method accounted for 104% of TTCs enumerated by membrane filtration in temperate surface waters. In tropical groundwaters, impacted by inadequate sanitation, 90-99% of all TTCs have been shown to be *E. coli* (Howard et al., 2003; Leclerc et al., 2001). Therefore we consider the number of faecal indicator organisms to be comparable between studies. Furthermore, relationships between TLF and the concentration of indicator organisms will be addressed using WHO risk categories (1997) that are classified based on a logarithmic scale of faecal indicator organisms that does not distinguish between TTCs and *E. coli*.

4.2.2 Assessment criteria

The ability of TLF to detect faecally contaminated drinking water was assessed using the UNICEF target product profile for a new method for the more rapid detection of *E. coli* (Table 4-2) (UNICEF/WHO, 2017). This incorporates sensor performance, design and interferences. As a minimum, any new methodology should be able to determine faecal contamination within 3 h with a 10% false-positive and false-negative error rate, have a limit of detection of 10 cfu/100 mL, and quantify faecal contamination according to the WHO risk categories (1997). Additionally, the method should be easy-to-use, suitable for fieldwork in remote locations, cost <$5 per test over its lifetime, and not be adversely affected by a range of common interferents such as turbidity and pH.

Table 4-2 Summarised UNICEF target product profile for the more rapid detection of *E.coli* drinking water (UNICEF 2017).

Attribute	Minimum performance	Ideal performance
Function	Detection of faecal contamination	Detection of faecal contamination equivalent to *E. coli*
Determining faecal contamination	False-positives <10% False-negatives <10%	False-positives <5% False-negatives < 5%
Limit of detection	Equivalent to 10 cfu/100mL	Equivalent to 1 cfu/100mL
Quantifying faecal contamination	Distinguishes WHO risk categories	Distinguishes equivalent plate count
Sensor design	Less than 3 h for result	Less than 30 mins for result
	Negative control sterile water	No negative control
	Calibration for pH	No calibration
	Easy-to-use	Easy-to-use
	Battery-powered	Battery-powered
	Lightweight and portable	Handheld
	2 year lifespan (hardware) 2 year lifespan (consumables) $1-5 per test (lifetime cost)	5 year lifespan (hardware) 5 year lifespan (consumables) $1-5 per test (lifetime cost)
Potential interferences	Turbidity (0-10 NTU) pH (5.5-8.5) salinity (drinking water range)	Turbidity (0-50 NTU) pH (4.5-8.5) salinity (drinking water range)

4.2.3 Determining faecal contamination

The extent to which TLF indicated the presence of TTCs, the indicator of faecal contamination, was investigated using a logistic regression model (Dobson, 2001) in R software version 3.2.2. The model was developed using a training dataset, then evaluated against a validation dataset. To produce these datasets, firstly, the collated dataset (n=564) was grouped into WHO risk categories (WHO, 1997). These risk categories are based on the number of faecal indicator organisms per 100 mL: Very Low

risk (0), Low risk (1-9), Medium risk (10-99), High risk (100-999), and Very High risk (1000+). Subsequently, the data were randomly sampled to split the data equally in each risk category using the R package 'caret'. This produced training and validation datasets for each risk category, which were combined to form collated training (n=286) and validation datasets (n=276).

The performance of the model was initially assessed using the receiver-operating characteristic (ROC) curve (Hanley and McNeil, 1983) implemented in the R package 'pROC' (Robin et al., 2011). The ROC curve is a plot of the false-positive rate (FPR - the proportion of uncontaminated sites falsely classified as contaminated) against the true-positive rate (TPR - the proportion of contaminated sites correctly classified) at varying TLF thresholds. The area under this curve (AUC) is a measure of the effectiveness of the classifier: it will be 1 for a perfect classifier and 0.5 if the classifier is performing no better than a random choice.

An optimal TLF threshold was selected as the value that provides the lowest sum of the FPR and the false-negative rate (FNR – the proportion of contaminated sites incorrectly classified as uncontaminated) for samples in the Very Low to Medium training risk categories. These lower risk categories were used to select the TLF threshold to ensure it was optimised around the limits of detection outlined in the assessment criteria. This threshold was then used to classify the validation dataset as a whole, then each validation risk category dataset individually. The threshold performance was assessed in terms of FPR and FNR for consistency with the assessment criteria. Finally, the threshold was used to classify data from the individual studies separately to evaluate any potential discrepancies.

4.2.4 Quantifying faecal contamination

There is a significant very strong positive correlation between TLF intensity and TTC concentration in the collated dataset (Spearman's $\rho_s = 0.80$, $p < 0.001$), but due to the amount of scatter it was considered unlikely that TLF could predict an equivalent plate count. Hence, the ability of TLF to distinguish between WHO risk categories was investigated instead. Consequently, the TLF data were

grouped into the risk categories and the non-parametric Kruskal-Wallis test was performed. A non-parametric pairwise multiple comparison was then conducted using Dunn's Method to assess differences between all individual categories. We then attempted to define TLF threshold values for each individual WHO risk category using the same training and validation datasets acquired using methods outlined in section 4.2.3. TLF thresholds were defined for each risk category by selecting the mid-point between the upper and lower quartiles of adjacent risk categories in the training dataset and then evaluated by classifying the validation datasets. These thresholds were not tested against individual studies because there are insufficient data across the risk categories to facilitate evaluation.

4.3 Results and discussion

4.3.1 Determining faecal contamination and limit of detection

Tryptophan-like fluorescence is a significant predictor of the presence of TTCs for the training dataset according to the logistic regression model (p <0.001). The area under the ROC curve was 0.90 (Figure 4-1a), which is much closer to the perfect classifier of 1 than the random selection value of 0.5. The optimal selected TLF threshold value of 1.3 ppb resulted in a FNR and FPR of 14% and 21%, respectively (Figure 4-1b). The overall FNR is slightly high, but it is biased to the Low risk category where the classification is unsuccessful (52%). Omitting this category the overall FNR reduces to within the ideal performance of 5% (5%). The Medium risk category data are classified to close to the minimum assessment criterion (12%), whilst the High and Very High risk datasets are successfully classified with FNRs of 9 and 0%, respectively (Figure 4-2).

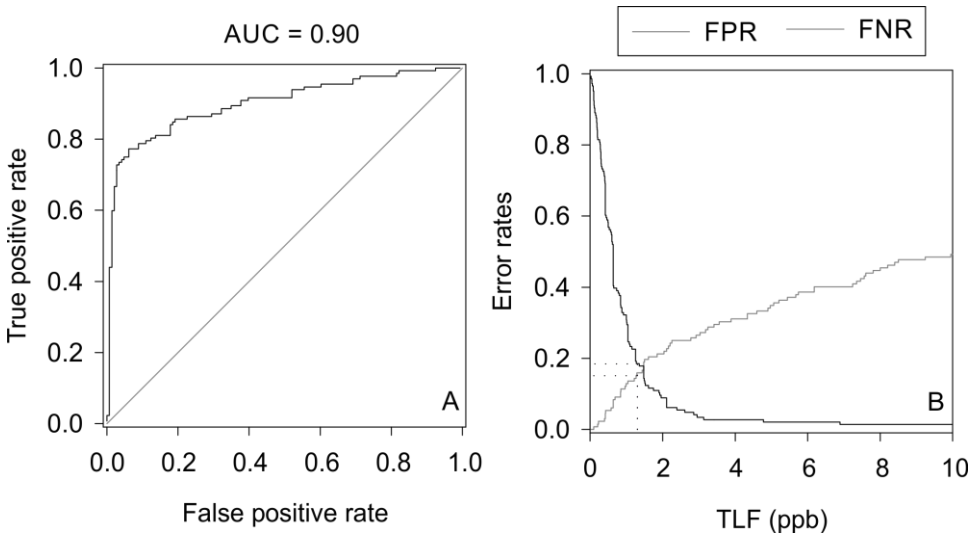

Figure 4-1 (A) Receiver operator characteristic curve for the classifier of the presence of TTCs using TLF and the performance of a random variable; (B) false-positive (FPR) and false-negative (FNR) error rates for this classifier.

The TLF threshold performs similarly, classifying the validation dataset with an overall FNR and FPR of 15% and 18%, respectively. Again, the threshold is unsuccessful at classifying sites in the Low risk category, with a FNR of 50%, which may, in part, result from limited reproducibility in the TTC counts themselves at low numbers. If we exclude these Low risk data then the overall FNR reduces to within the ideal criterion of 5% (4%). The individual Medium, High, and Very High validation datasets are classified close to or within the minimum assessment criterion with FNRs of 12%, 3% and 2%, respectively (Figure 4-2).

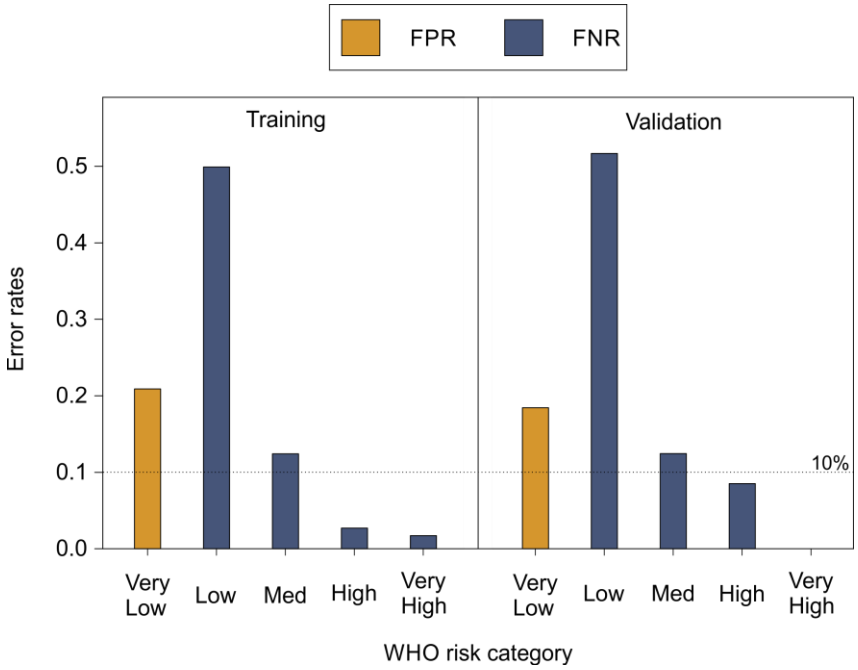

Figure 4-2 Training and validation false-positive (FPR) and false-negative (FNR) error rates for presence of TTCs using a TLF threshold of 1.3 ppb for each WHO risk category.

Given the inability of the threshold to classify Low risk sites, we deem the limit of detection is currently Medium risk, i.e. at least 10 cfu/100 mL. To refine a precise detection limit requires considerably more data around 10 cfu/100 mL. This detection limit should be considered sufficient given any relationship between *E. coli* concentration in drinking water and diarrhoea is frequently only significant above at least 10 cfu/100 mL (Brown et al., 2008; Moe et al., 1991).

The TLF threshold of 1.3 ppb is generally effective at classifying data from individual studies (Figure 4-3). TTC negative data in the Indian study are classified within the assessment criteria with a FPR of only 9%. However, FPRs for these uncontaminated sites are high in both the Malawian and Zambian datasets. The elevated FPR of 35% in the Zambian study combined with the near absence of false-negatives suggests a higher TLF threshold may be warranted for a more equal distribution of errors. However, a logistic regression model developed solely using the Very Low, Low and Medium risk Zambian datasets indicates only a marginally higher optimal threshold of 1.6 ppb. Here, the FPR still remains too high (24%) and the FNR is acceptable (7%). The majority of unsuccessfully classified Low risk data are from the Indian and Malawian datasets. Where error rates exceed the 10% minimum

criterion for Medium and High risk categories in individual studies, it is a result of only one or two incorrectly classified sources suggesting that sample sizes are too small to robustly estimate error rates for these categories. All Very High risk data are successfully classified with a FNR of only 0-1% in all studies.

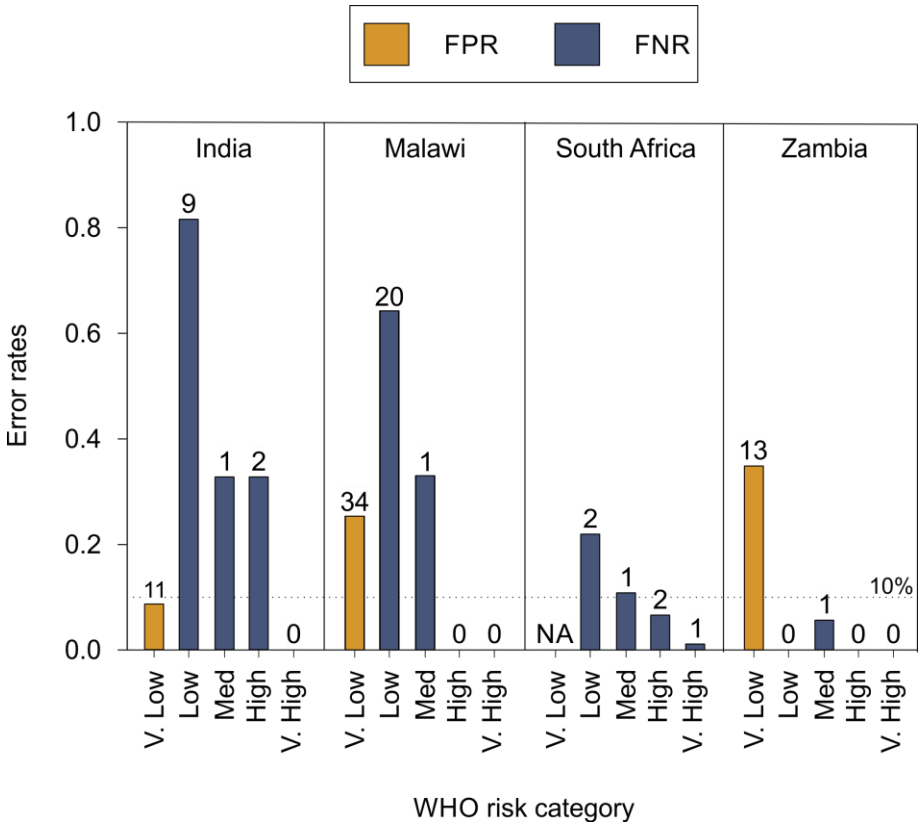

Figure 4-3 False-positive (FPR) and false-negative (FNR) error rates for presence of TTCs for each WHO risk category in each study using a TLF threshold of 1.3 ppb. Number above the bar indicates number of incorrectly classified sites.

4.3.2 Quantifying faecal contamination

There are significant differences in median TLF between risk categories indicating that multiple populations exist in the dataset (Kruskal-Wallis, $p < 0.001$). Significant differences in median TLF exist between all risk categories (Figure 4-4), with the exception of Medium and High (Dunn's Method, $p = 0.96$). The TLF thresholds, defined in Table 4-3, have error rates that are too high for practical implementation when assessed against the validation dataset (Figure 4-5). Nevertheless, incorrectly classified sites are most likely to fall in the immediately adjacent risk categories. For example, 90% of

Very Low risk sites are classified as Very Low to Low risk and 91% of Very High risk sites are classified as High to Very High risk.

Therefore, a threshold of 1.3 ppb for classifying at-risk sources (>10 cfu/100 mL) is proposed according to the logistic regression model in Section 4.3.1. This would effectively classify sites in the validation dataset according to Figure 4-4 with a FNR of only 4% for sources that are at least Medium risk. It would be unsuccessful at classifying Low risk sites (FPR = 50%) and reasonably successful at classifying Very Low risk sites (FPR = 21%).

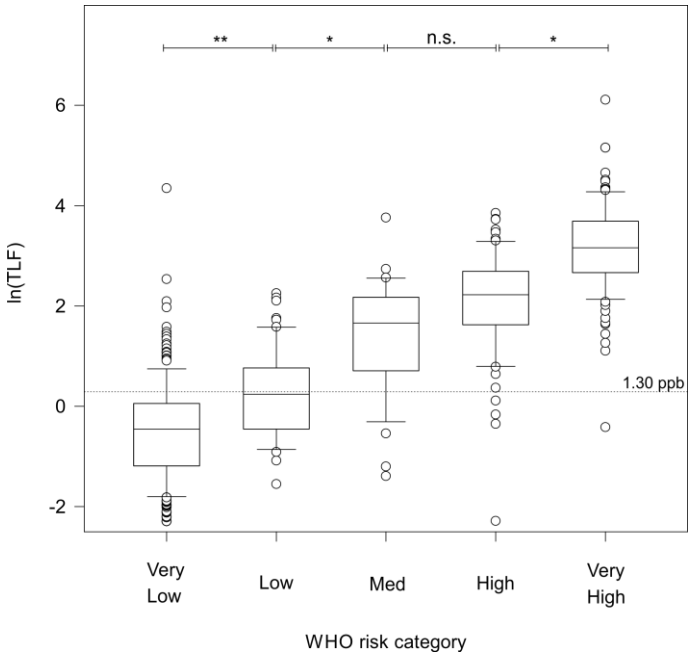

Figure 4-4 Boxplot of the natural logarithm of tryptophan-like fluorescence by WHO risk category. An addition of 0.2 ppb TLF was made to nine sites to ensure the logarithm could be defined. Significance tests were performed on untransformed TLF data using Dunn's test and are denoted by: not significant = n.s., p = 0.05 (*) p = 0.01 (**). Sample sizes: Very Low = 293, Low = 61, Medium = 32, High = 71, Very High = 107. Boxes illustrate median and interquartile range, whiskers indicate 5th and 95th percentile and all outliers are shown.

Table 4-3 Tryptophan-like thresholds for individual risk categories.

Category	TLF threshold (ppb)
Low	1.3
Medium	2.4
High	6.9
Very High	27.1

A greater threshold of 6.9 ppb is proposed for classifying higher risk sources (>100 cfu/100 mL) following the TLF thresholds defined in Table 4-3. This threshold successfully classifies sources in the validation dataset that are at least High risk with a FNR of 17% and incorrectly classifies lower risk sites with a FPR of 4%. However, due the TLF overlap between Medium and High risk categories (Figure 4-5) the threshold produces a FPR for Medium risk sites of 31% and a FNR for High risk sites of 30%. Notwithstanding, at such an elevated TLF the FNR for Very High risk sources remains low (9%) and the FPR for Very Low and Low risk sites is very low (2%).

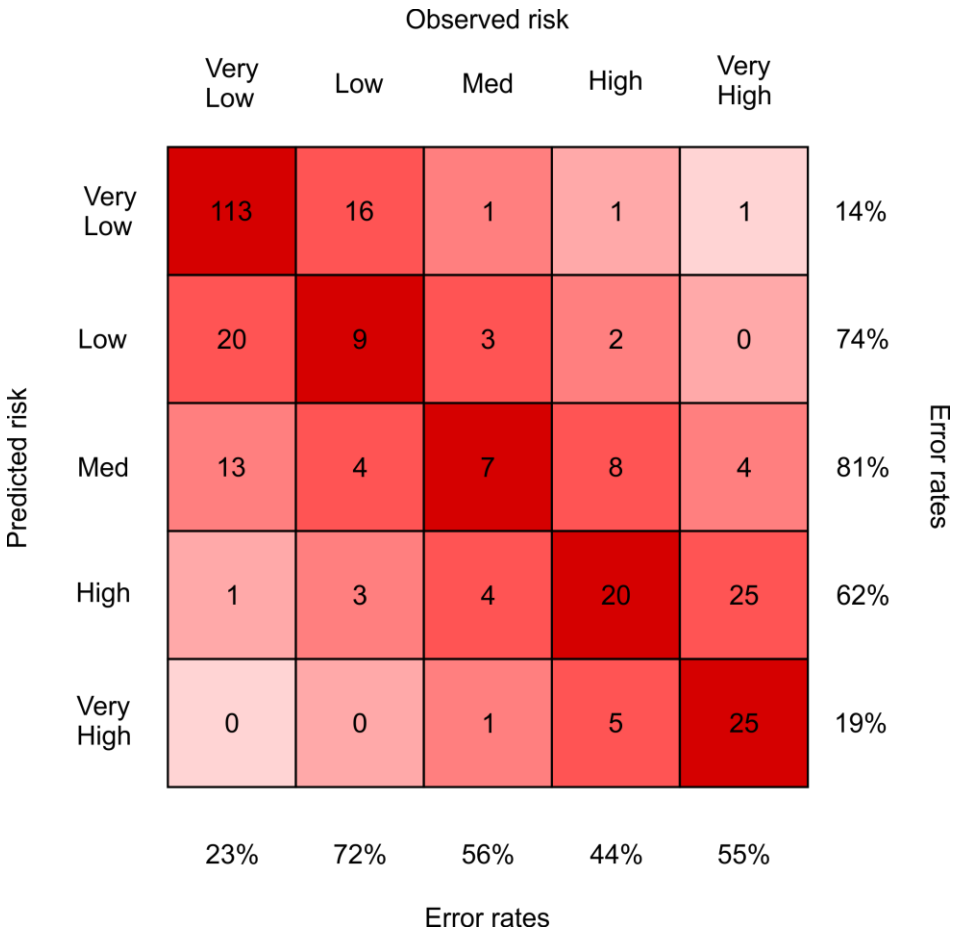

Figure 4-5 Risk matrix of observed versus predicted risk category.

4.3.3 Sensor designs

TLF can simply be measured instantaneously in a covered beaker or cuvette using a fluorimeter, with no need for reagents, consumables or incubation. There are currently several commercially available battery-powered fluorimeters on the market that are suitable for fieldwork in remote locations,

although these vary in design and method. They include submersible sensors, e.g. UviLux (Chelsea Technologies Group Ltd, UK) and Cyclops-7^{TM} (Turner Designs, USA), or cuvette-based systems, e.g. SMF4 (STS Instrument Ltd, UK). Cuvette-based systems are advantageous because they require very small volumes of water (<5 mL) and contamination of the sample is less likely. This advantage arises because the housing of submersible sensors comes into contact with the sample and the housing can be easily manually handled and can also begin to degrade allowing suspended particles to collect in surface imperfections and potentially leach residue TLF into subsequent samples. This limitation could partially be overcome by mounting submersible sensors in flow-through cells. In addition, it is notable that submersible sensors can produce an apparently very low TLF intensity when air is trapped adjacent to sensor windows. The main advantage of submersible sensors is for online deployments in a flow-cell or directly within a water body.

The most critical element to currently consider when selecting a sensor is currently the limit of detection and not whether it is submersible or cuvette-based. The proposed 1.3 ppb threshold for the detection of faecal contamination is below the specification for the majority of sensors on the market. For example, Khamis *et al.* (2015) demonstrated a minimum detection limit (MDL) of 1.99 ± 0.53 ppb for the Cyclops-7^{TM} and Khamis *et al.* (2016) reported a MDL of 1.74 ppb for a modified submersible GGUN–FL30 (Albillia Co, Switzerland). The only commercially available field fluorimeter we are aware of that is capable of sufficient sensitivity is the submersible UviLux fluorimeter, which incorporates a photomultiplier to amplify the signal and achieve a MDL of up to 0.17 ± 0.06 ppb (Khamis et al., 2015).

The present generation of field fluorimeters cost in the region of $5,000-6,000. This excludes any accessories, such as handheld readers or cables, which can inflate costs by a further $2,000-3,000. Nevertheless, once procured there are no ongoing consumable costs. The expected lifetime of these fluorimeters is estimated to be at least 10 years; hence the costs would be within the minimum criterion of $5 if used at only 100 sites a year. The sensor lifetime is determined by the LEDs and light filters. To account for a loss of output from the LED and change in the filters over time, it is

recommended to undertake a calibration check annually with dissolved tryptophan standards, which can be done by the end user. Additionally, regular negative controls should also be performed using high quality deionised water. Ideally manufacturers would supply sealed long-life containers for both calibration and negative control, but these are currently only available for cuvette-based instruments. Current cuvettes for calibration contain another fluorescent compound, quinine sulphate, which does not fluoresce directly within the TLF region. High concentration standards can be used to assess drift, but these do not reflect the range of TLF intensity encountered in the environment and these standards cannot be used for a robust calibration. In conclusion, current commercially available fluorimeters adhere to all the assessment criteria concerning sensor design.

4.3.4 Potential interferents

There are a range of matrix interferences that can impact on TLF including pH, temperature, and turbidity (Hudson et al., 2007; Khamis et al., 2015), but these have not been considered to have substantially affected previous TLF-TTC studies (Chapter 3, Baker et al., 2015; Sorensen et al., 2016). Reynolds *et al.* (2003) documents a ±3% variation in TLF from pure dissolved tryptophan when the sample matrix was modified between pH 5 and 8. Baker et al. (2007) observed an order of magnitude greater quenching in TLF for urban surface water samples for a pH range of 5 to 9.

Khamis *et al.* (2015) noted that TLF emitted from dissolved tryptophan standards was not attenuated by the addition of particles treated with hydrogen peroxide, to remove organic matter, below 50 NTU turbidity. Instead, suspended particles scattered the emitted light and increased the intensity reaching the detector. Larger silt particles were more efficient than smaller clay particles at scattering light. However, multiple high turbidity (>10 NTU) groundwater sites in India and Malawi contained low TLF (<1 ppb) suggesting that the scattering of light may be insignificant in natural settings. The issue of turbidity could be partially addressed by widening the separation between the excitation and emission wavelengths. Currently, sensors have broad filters for both excitation and emission to maximise fluorescent output and detection. For example, the Cyclops 7^{TM} has an excitation output of 285±10 nm

and emission set at 350±55 nm, i.e. there is overlap at 295 nm. Low salinity in drinking water is unlikely to have an appreciable impact on TLF. Overall, there is limited evidence, to date, that the range of interferents outlined in the assessment criteria has any substantial adverse impact on the TLF of raw drinking water.

One interferent not considered in the criteria is temperature. Baker (2005) noted quenching of TLF by 20±4% to 35±5% between 10 and 45°C in rivers and wastewaters. Such temperature quenching is generally considered to be linear (Khamis et al., 2015), hence it is possible that sensors could employ automatic corrections for temperature. Researchers have already developed such algorithms and demonstrated their value for in-situ fluorimeter deployments (Khamis et al., 2015; Shutova et al., 2016). However, such studies have explored the temperature quenching of dissolved tryptophan, whilst TLF in field studies is known to quench at different rates in different water types dependent upon its composition (Baker, 2005). The thresholds defined in this paper do not consider the influence of temperature, which was broadly similar across all studies: India (24-29 °C), Malawi (21-29 °C), South Africa (20-26 °C), Zambia (24-28 °C).

There are multiple peaks of fluorescent dissolved organic matter that overlap with the tryptophan-like region and could dominate, or give rise to an apparent, TLF signal. For example, part of the region of humic-like fluorescence can overlap with that of TLF and provide an additional baseline component or potentially mask the TLF signal completely. In addition, proteins in organic waste (Muller et al., 2011) and xenobiotic compounds, e.g. polycyclic aromatic hydrocarbons (Baker and Curry, 2004) and diesel pollution (Carstea et al., 2010), may directly fluoresce in the TLF region. Such interference that is unrelated to faecal contamination could result in false-positives.

Interference between fluorescent regions is a long-recognised issue for researchers using laboratory-based spectrofluorometers which scan over a large range of excitation-emission wavelengths, but is a real limitation of single-wavelength sensors. Studies using laboratory spectrofluorometers address the issue by investigating ratios of fluorescent peaks (Baker, 2002b; Lapworth et al., 2008) or by a

modelling approach to elucidate individual peaks known as PARAllel FACtor analysis (PARAFAC) (Baghoth et al., 2011). In the field, it is possible to use multiple single-wavelength sensors to evaluate potential interference (Sorensen et al., 2016). Additionally, there is an increasing move towards the development of dual-wavelength pair sensors that can output ratios as well as absolute TLF (Bridgeman et al., 2015; Li et al., 2016).

Other potential interferences include the absorption of light by the sample matrix (the inner filtering effect), quenching by high concentrations of metal ions, or by water treatment processes such as chlorination that is known to quench TLF. See Henderson *et al.* (2009) for a review of the impacts of water treatment processes on fluorescent dissolved organic matter.

4.3.5 Future work

Currently, false-positive error rates (18%) are too high, particularly in specific studies. Further work should better constrain the temporal relationship between TLF and TTCs at individual sources to investigate whether elevated TLF is indicative of sporadic TTC contamination at certain times of the year. The only current evidence for this is in the Zambian study where snap-shot seasonal sampling campaigns of some sources showed perennially elevated TLF whilst TTC varied substantially between dry and wet seasons (Chapter 3). If TLF is more temporally resilient than TTCs in water sources, then the use of TLF could be considered advantageous for detecting at-risk sources irrespective of the time of year selected for sampling and given that TTCs are not as long-lived as other pathogens in the environment (John and Rose, 2005).

There is uncertainty over what is actually measured when we quantify TLF in freshwater environments. Compounds emitting TLF can predominantly be: contained within bacterial cells (Determann et al., 1998; Fox et al., 2017), associated with particles (Baker et al., 2007), or entirely freely dissolved (Sorensen et al., 2016; Yamashita and Tanoue, 2003); and there remains uncertainty as to the relative contribution from each source. Further, it is questionable whether TLF is a selective indicator of TTCs. There are many other bacteria species that fluoresce in the TLF region because of

tryptophan residues in proteins (Bronk and Reinisch, 1993; Dalterio et al., 1986; Dalterio et al., 1987; Dartnell et al., 2013; Fox et al., 2017; Seaver et al., 1998; Sohn et al., 2009) and/or excrete compounds that fluoresce in the TLF region, such as *Pseudomonas aeruginosa* that is ubiquitous in freshwater systems (Elliott et al., 2006a; Fox et al., 2017). Therefore, future work should investigate whether TLF is an indicator of total bacteria cells in drinking water as opposed to specifically TTCs.

4.4 Conclusions

A tryptophan-like fluorescence (TLF) threshold equivalent to 1.3 ppb dissolved tryptophan can instantaneously predict the presence of thermotolerant coliforms (TTCs) with a false-positive error rate of 18% and a false-negative error rate of 15%. However, this TLF threshold is not effective at classifying a contaminated sample with less than 10 TTC cfu/100 mL, which we currently consider to be the limit of detection. If only contaminated sources above this limit of detection are classified then the false-negative error rate is very low at 4%. A greater threshold of 6.9 ppb is proposed for classifying higher risk sources (>100 TTC cfu/100 mL).

TLF can be quantified instantaneously using existing commercially available fluorimeters that are battery-powered and handheld for use in remote environments. Its analysis can simply be undertaken by submerging a fluorimeter in a covered container of water. The lifespan of current fluorimeters is anticipated to be in excess of ten years with annual checks and infrequent maintenance. Current procurement costs are in the order of $5000-6000 and per test costs would fall below $5 within 1000 samples, as there is an absence of ongoing consumable costs. TLF measurements are unlikely to be appreciably impaired by common interferents, such as pH, turbidity, and temperature, within typical natural ranges. Interference from other fluorescent compounds is likely to be more problematic.

The technology should be considered a viable option for the real-time detection of faecally contaminated drinking water which is still consumed by 1.8 billion people globally. It has now been successfully demonstrated in multiple different settings from surface water to groundwater and across varying climatic zones. However, all analyses herein were performed on a modest dataset

(n=564), which may contain sampling bias as the majority of data have not been drawn from an independent random sample. There is a need for a large-scale field demonstration of the technique to robustly support the results of this study.

Chapter 5 - Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water

5.1 Introduction

The aim of this chapter was to address **Objective 2** and extend the demonstration of in-situ fluorescence as an indicator of faecal contamination beyond roaming surveys at limited temporal resolution to real-time deployment online at water sources. Furthermore, HLF was investigated as an indicator of faecal contamination for the first time (**Objective 1**) and TLF/HLF are also compared with total bacterial cell counts as well as FIOs (**Objective 5**). We measured TLF and humic-like fluorescence (HLF) online at four groundwater-derived public water sources and compared these to in-situ turbidity measurements, culture-based counts of *E. coli* and total bacterial counts measured by flow cytometry. This chapter is published as:

Sorensen, J.P.R., Vivanco, A., Ascott, M.J., Gooddy, D.C., Lapworth, D.J., Read, D.S., Rushworth, C.M., Bucknall, J., Herbert, K., Karapanos, I., Gumm, L.P. and Taylor, R.G., 2018. Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. Water Research, 137, 301-309.

5.2 Methods

5.2.1 Study sites

Four groundwater-derived public water supply sites were selected in Southern England (Figure 5-1). These sites all abstract from varying types of fractured limestone, which is the largest source of groundwater in the UK. These limestone aquifers can have rapid flow paths and the selected sites are all known to experience episodic pollution events that are generally linked to heavy rainfall. Site 1 comprises Portland Stone Formation boreholes, site 2 Chalk Group springs, site 3 Chalk Group boreholes, and site 4 Jurassic Limestone springs.

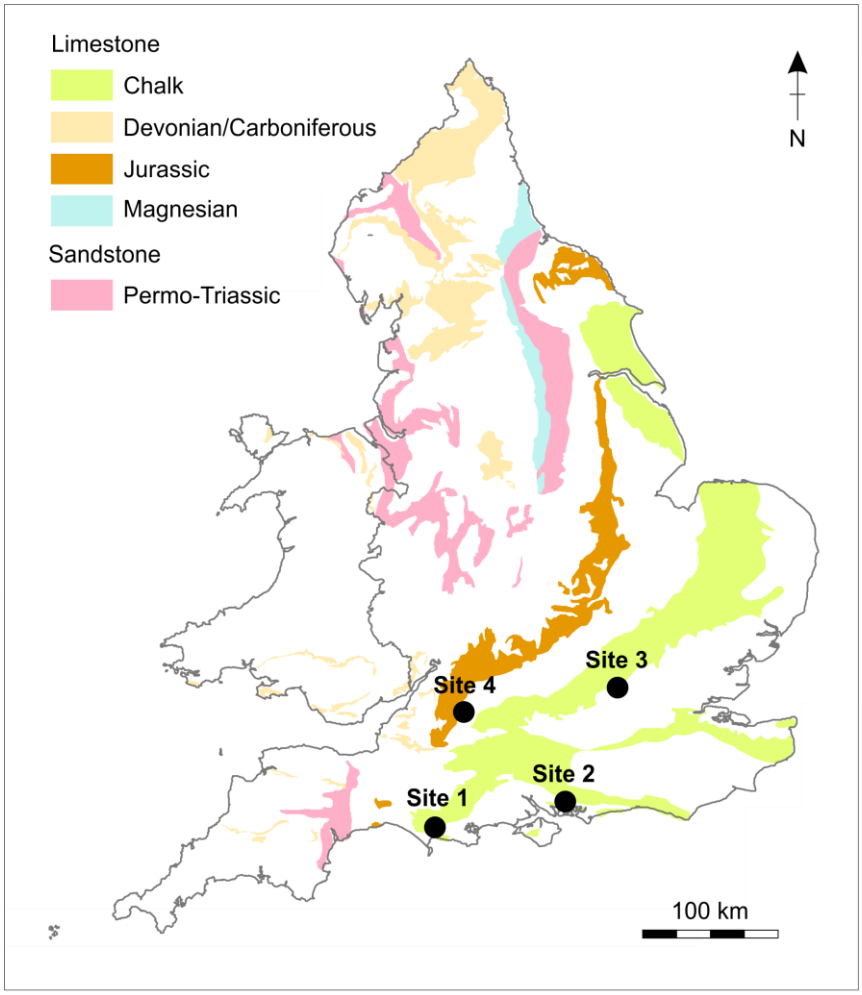

Figure 5-1 Location of public water supplies and major aquifers of the UK. Note the Portland Stone Formation is too small to be visualised at this scale.

5.2.2 Online analysis

TLF was determined at each site and HLF at sites 1, 2, and 3 every two minutes on the raw water intake for between 6 and 10 months using telemetered UviLux fluorimeters (Section 2.3). All fluorimeters were installed in flow-through cells, which excluded all natural light, connected to sample taps at each site. Factory calibrations were implemented on all UviLux sensors. For the calculation of a TLF:HLF ratio, TLF was converted to QSU by multiplication by 0.5365; this factor relates the fluorescence intensity of tryptophan to that of quinine sulphate. The temperature of water exiting each cell was monitored with a HOBO Tidbit® v2 (Onset Computer Corporation, USA).

At all sites abstracted raw groundwater is screened in real-time against defined turbidity thresholds, using nephelometric technology, before passing into a treatment plant prior to onward distribution.

If water exceeds the defined threshold, usually 1 NTU, it is considered a potential threat to the provision of clean water, as high levels of microbial contamination may be present alongside particles that could reduce treatment efficacy (DWI, 2010). In this case, abstraction may cease or raw water may be directed to waste until turbidity returns to below the threshold. In both cases, fresh sample did not pass through the flow cells containing the fluorimeters. These instances were observed regularly at sites 1, 3 and 4 and all fluorescence data were disregarded during these periods that could last from hours to weeks. Only site 2 has a complete uninterrupted record of fluorescence data across the monitoring period.

5.2.3 Sample collection and laboratory analysis

Sample collection and storage

Duplicate water samples were retrieved from each site on a regular basis for comparison with the online sensors. One was collected and analysed for *E. coli* within 8 h following standard water industry compliance procedures (EA, 2009). The second was collected in sterile 15 ml polypropylene tubes, stored in the dark at 4 ^{o}C, and analysed for fluorescence and absorbance, and total bacterial cell counts by flow cytometry. These second samples were analysed within 48–96 hours of collection. Sample stability tests were performed which confirmed the stability of samples over this timeframe, see Section 2.3.4.

Fluorescence and absorbance analysis

Fluorescence analysis was conducted on unfiltered samples at 20 ^{o}C with λ_{ex} between 200 and 400 nm (5 nm bandwidth) and λ_{em} between 280 and 500 nm (2 nm bandwidth) using a Varian™ Cary Eclipse fluorescence spectrometer (Agilent Technologies, USA). The scan rate was 9600 nm/min and the detector voltage was set to 900 V. All analysis was performed in a quartz cuvette with a path length of 1 cm. The Raman peak of ultrapure water at 348 nm was used to check for instrument stability prior

to and following analysis, for blank correction, and to standardise the fluorescence excitation-emission matrices (EEMs).

Absorbance measurements were undertaken in a 1 cm quartz cuvette at 1 nm intervals from 800 to 200 nm using a Varian Cary 50 UV–Vis spectrophotometer (Agilent Technologies, USA). All absorbance spectra were referenced to a blank of ultrapure water. Mean absorbance was 0.011 cm^{-1} (σ = 0.007 cm^{-1}) at 280 nm. This indicates that inner-filter correction was not required, since the primary inner filter effect was <3.7% across all samples (Lakowicz, 2006).

Flow cytometry

It has been proposed that the total number of bacterial cells by flow cytometry could act as an indicator for drinking water contamination, as elevated counts could indicate contamination from external sources containing higher cell densities or the growth of bacteria within the distribution network (Lautenschlager et al., 2013). Consequently, total bacterial counts were analysed for on a BD Accuri C6 flow cytometer equipped with a 488 nm solid state laser (Becton Dickinson U.K. Ltd., Oxford, U.K.). Water samples (500 µl) were stained with SYBR Green I (Sigma-Aldrich, Gillingham, UK) at a final concentration of 1:200 v/v for 20 minutes in the dark at room temperature, before running on the Accuri at a slow flow rate (14 µL/min, 10 µm core) for 5 minutes and a detection threshold of 1,500 on channel FL1. A single manually drawn gate was created to discriminate bacterial cells from particulate background, and cells per mL were calculated using the total cell count in 5 minutes divided by the reported volume run in µl. The mean absolute error (MAE) from 69 duplicate analyses across all sites was 6536 cells/mL

Sample stability

A series of samples was collected sequentially from each of sites 1, 2, and 4. These samples were destructively analysed in triplicate on a daily basis, excluding weekends, over eleven days for fluorescence and nine days for total bacterial cell counts following the procedures listed above. For

fluorescence, equivalent TLF and HLF wavelength pairs corresponding to the in-situ sensors were extracted from the excitation/emission matrix. The results are expressed in percentage change from the mean of three samples analysed after storage for one day (Figure 5-2).

There was no consistent change in either TLF or HLF over 11 days, hence variation in storage time before analysis is unlikely to have influenced the results. Nonetheless, transformations during storage have been observed within the first 24 h, which have not been considered here. For example, Bieroza and Heathwaite (2016b) demonstrated increases in TLF of 9–11% in unfiltered surface water samples stored at 10 °C over this timescale. However, potential changes in samples collected in this study are likely to be lower given the reduced storage temperature (4 °C) and because groundwater as a medium is likely to be more stable than river waters due to the generally lower concentrations of bacteria and nutrients that are present. Of greater significance is the variability between samples analysed in triplicate on the same day (Figure 5-2).

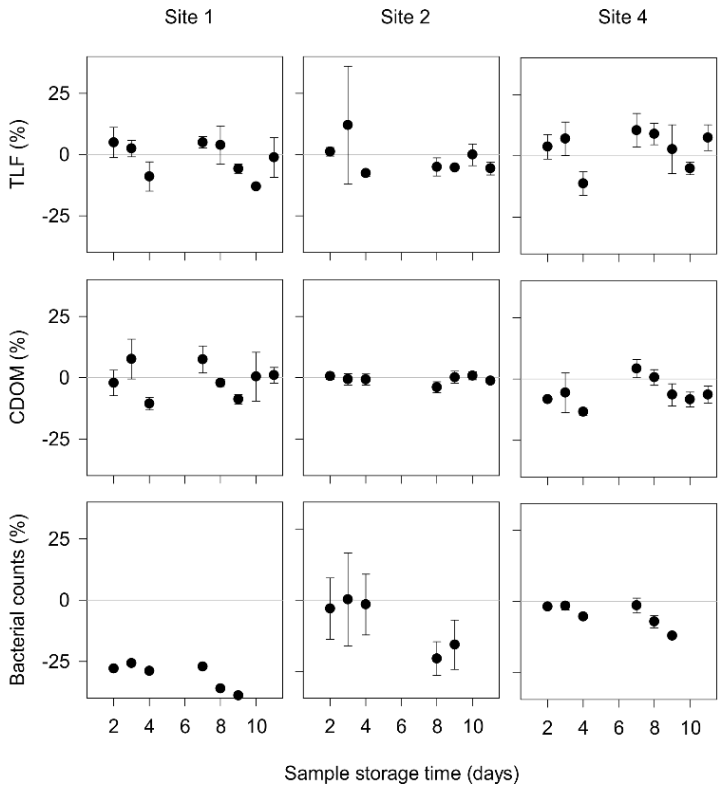

Figure 5-2 Percentage change in TLF, CDOM and bacterial counts from the initial sample analysed after one day of storage to samples stored for varying lengths of time. Error bars represent range from three separate samples analysed on the same day.

Bacterial cell counts were consistently lower after storage for one day (Figure 5-2), with an overall mean decline of 12% between storage for one and two days across all samples. Nevertheless, there was no apparent change between two and four days. Therefore, it is considered that reported bacterial cells counts are marginally lower than at the time of sampling at sites 1, 2 and 4, but variation in storage time between 2 and 4 days is unlikely to have influenced the results.

5.2.4 Statistical analysis

Correlations were assessed by Spearman's Rank (Spearman, 1904). Differences in the median between two populations were assessed by Mann-Whitney tests and between multiple populations by the Kruskal-Wallis method (Kruskal and Wallis, 1952) followed by a pairwise multiple comparison using Dunn's Method (Dunn, 1964). *E. coli* data were classified into World Health Organisation (WHO) risk categories for some analyses which defines risks based on the plate count per 100 mL (WHO, 1997) as follows: Very Low (0 cfu), Low (1-9 cfu), Medium (10-99 cfu), High (100-999 cfu) and Very High (1000+ cfu). All these statistical analyses were non-parametric, due to the non-normal distribution of all datasets, and performed in SigmaPlot version 13.

Parallel factor analysis (PARAFAC) was undertaken on blank corrected fluorescence EEMs from all sites to extract, model and quantify the spectrally overlapping components. PARAFAC uses a least-squares algorithm to decompose the data; for further details on the method and approach see Stedmon et al. (2003). The analysis was undertaken using the 'DOMFluor v1.7' toolbox (Stedmon and Bro, 2008) in Matlab version 9.1. Rapid model convergence was obtained and validated using independent split-half analysis of the dataset following removal of outlier data which had high leverage (Stedmon and Bro, 2008).

5.3 Results and Discussion

5.3.1 Online indication of E. coli

There are strong correlations between TLF and *E. coli* (ρ_s = 0.71, p <0.001, n= 134) and HLF and *E. coli* (ρ_s= 0.77, p <0.001, n= 122), but only a moderate correlation between turbidity and *E. coli* (ρ_s = 0.48,

p_s <0.001, n= 134). When grouping the *E. coli* data into WHO risk categories, TLF, HLF and turbidity all vary significantly between categories (Kruskal-Wallis, $p < 0.001$) (Figure 5-3). Significant differences exist between all categories for TLF (Dunn's Method). However, for turbidity and HLF (HLF was only monitored at sites 1–3), Low and Medium categories are not significantly different, although both are significantly different from the Very Low category. There were insufficient data to robustly include the High risk category in the analysis (n = 5), although all variables are generally most elevated in this category (Figure 5-3).

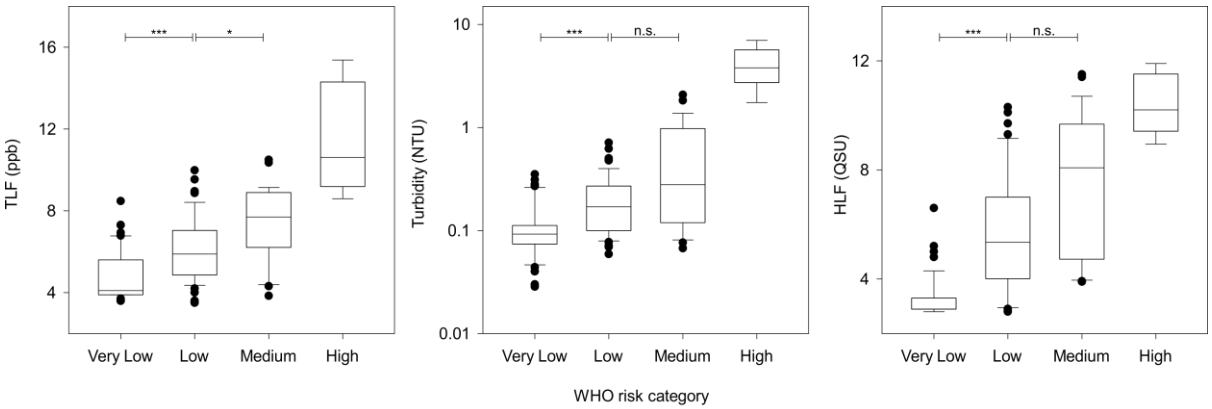

Figure 5-3 Boxplots of TLF, turbidity and HLF against WHO risk categories for E. coli. Boxes illustrate median and interquartile range, whiskers indicate 5th and 95th percentile, and all outliers are shown. Results of non-parametric Dunn's Method tests are displayed for Very Low to Medium categories: not significant = n.s =, p = 0.05 (*), p < 0.001 (***) =. Sample sizes were Very Low = 53, Low = 47, and Medium = 29 for TLF and turbidity, and Very Low = 49, Low = 44, and Medium = 24 for HLF.

Consideration of the collated dataset as a whole is important for evaluating the universal applicability of these indicators. Ideally each indicator would also be assessed at each site individually to confirm that any collated analysis is not biased by site-to-site differences. However, there is a limited spread of data across risk categories at each individual site with which to undertake such analyses (Figure 5-4). Where n > 5, there is a significant difference between Low and Medium risk categories at site 1 for TLF (Mann-Whitney, p = 0.04) and turbidity (Mann-Whitney, p = 0.01), but not HLF (Mann-Whitney, p = 0.09); there are no significant differences between Very Low and Low categories at site 2 for any variable (Mann-Whitney, p = 0.49–0.91). Similarly, correlations are only present at site 1 for all

indicators (0.49–0.54, $p < 0.001$, n=53), which has the largest number of samples that are spread across all risk categories.

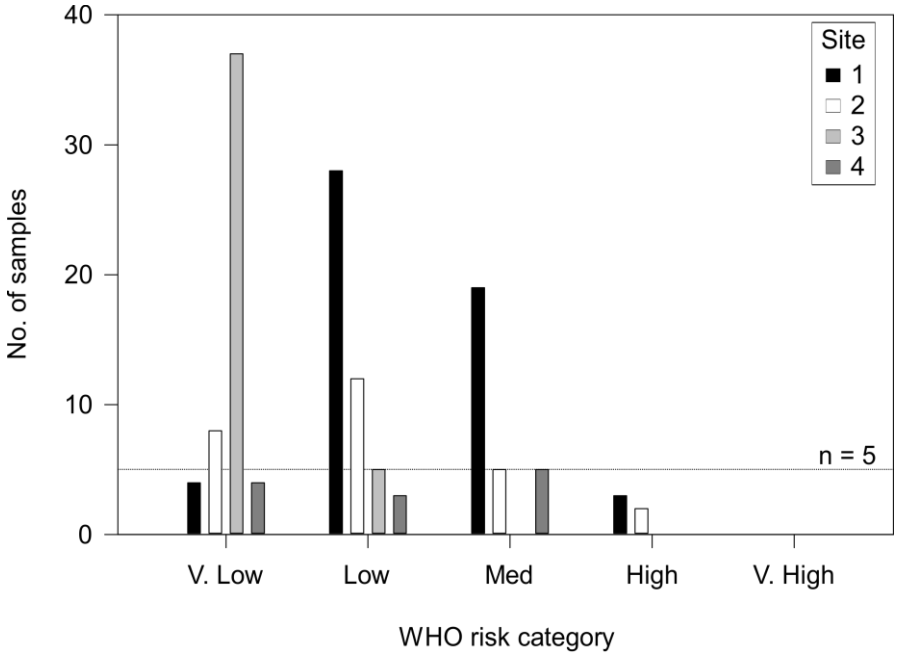

Figure 5-4 Frequency of data within each WHO risk categories for each site.

At all sites, TLF remains in excess of the proposed threshold of 1.3 ppb (Chapter 4) to indicate the presence of faecal contamination in untreated drinking water. Indeed, all sites experience *E. coli* contamination, but this ranges from commonplace occurrence at site 1 (93% of samples) to episodic occurrence at site 3 (12% of samples). TLF always remains above the threshold because the sites appear to have a reasonably consistent TLF baseline of around 4 to 8 ppb (Figure 5-5). Comparatively, previous groundwater studies linking TLF to faecal indicator organisms in boreholes have suggested a zero TLF baseline (Chapter 3; Sorensen et al., 2016).

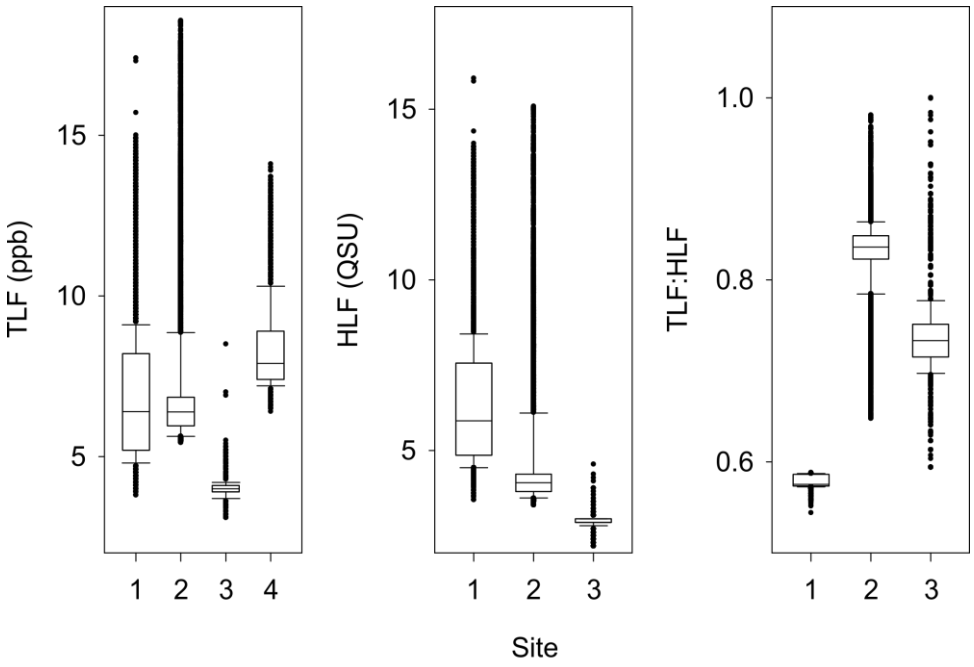

Figure 5-5 Boxplots of online TLF, HLF and TLF:HLF ratio subsampled to 15 minute resolution data at relevant sites.

5.3.2 Online indication of total bacterial cell counts

A strong positive non-linear correlation is observed between TLF and total bacterial cell counts, although there is noticeable scatter (Figure 5-6). This positive correlation holds at several individual sites: site 1 (ρ_s = 0.82, n = 17, p <0.001), site 2 (ρ_s = 0.68, n = 91, p <0.001), and site 4 (ρ_s = 0.70, n = 5, p = 0.23). No correlation is evident at site 3 (ρ_s = 0.11, n = 11, p = 0.73) because there is limited variation in TLF (σ = 0.1 ppb) and bacterial cell counts (σ = 5800 cells/mL). In fact, if site 3 is removed from the overall analysis, the overall correlation becomes even stronger (ρ_s = 0.86, n = 113, p <0.001). It is also noted that there is a discrepancy in the order of 2 ppb TLF at similar low bacterial cell counts between sites 2 and 3.

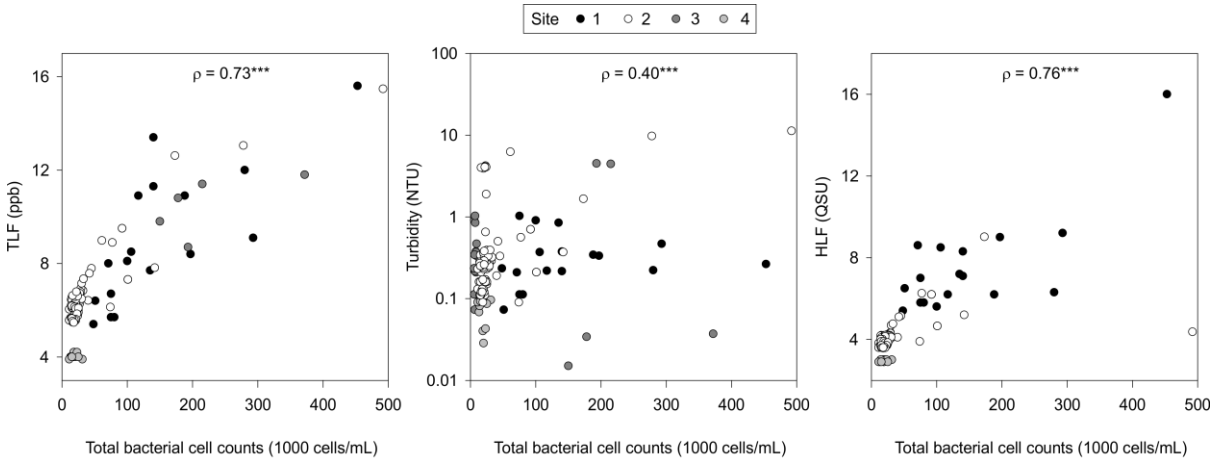

Figure 5-6 Scatterplots of TLF, turbidity and HLF against total bacterial cell counts for all sites. Spearman's rank correlation coefficients (ρ) and significance are displayed (* indicates p < 0.001), n = 124 for TLF and turbidity and n = 119 for HLF.**

HLF is strongly correlated with total bacterial cell counts in a collated dataset from sites 1-3 (Figure 5-6). Again, this positive correlation is observed at site 1 (ρ_s = 0.77, n = 17, *p* <0.001) and site 2 (ρ_s = 0.65, n = 91, *p* = <0.001), but not site 3 where there is limited variation in counts. The correlation coefficients for sites 1 and 2 are very similar to the correlation coefficients between TLF and total bacterial cell counts.

Only a weak correlation is noted between turbidity and total bacterial cell counts across all sites (Figure 5-6). The only statistically significant correlation is actually observed at site 2 (ρ_s = 0.72, n = 91, *p* <0.001), with no relationships at the other sites (ρ_s = 0.01 to -0.18).

5.3.3 Complete time series of online indicators at site 2

The complete dataset from site 2 shows the relationship between online indicators and bacteriological variables (Figure 5-7). There is a consistent TLF baseline (IQR = 0.9 ppb) where *E. coli* are generally absent, or few in number, and total bacterial cell counts are reasonably stable (IQR = 10 000 cells/mL). Following rainfall, increases in TLF, *E. coli* and total bacterial cell counts are noted. These increases are greatest during the main groundwater recharge season ending in March, and include several smaller peaks between late-May and mid-July. HLF displays a near identical relationship to TLF and bacteriological variables over the time period. Examination of the TLF:HLF ratio demonstrates it is

reasonably stable at around 0.85, except during events when there is a decrease in the ratio due to a proportionally greater increase in HLF. There is a noticeable decline in baseline TLF and HLF of around 0.8 ppb and 0.4 QSU, respectively, over the record.

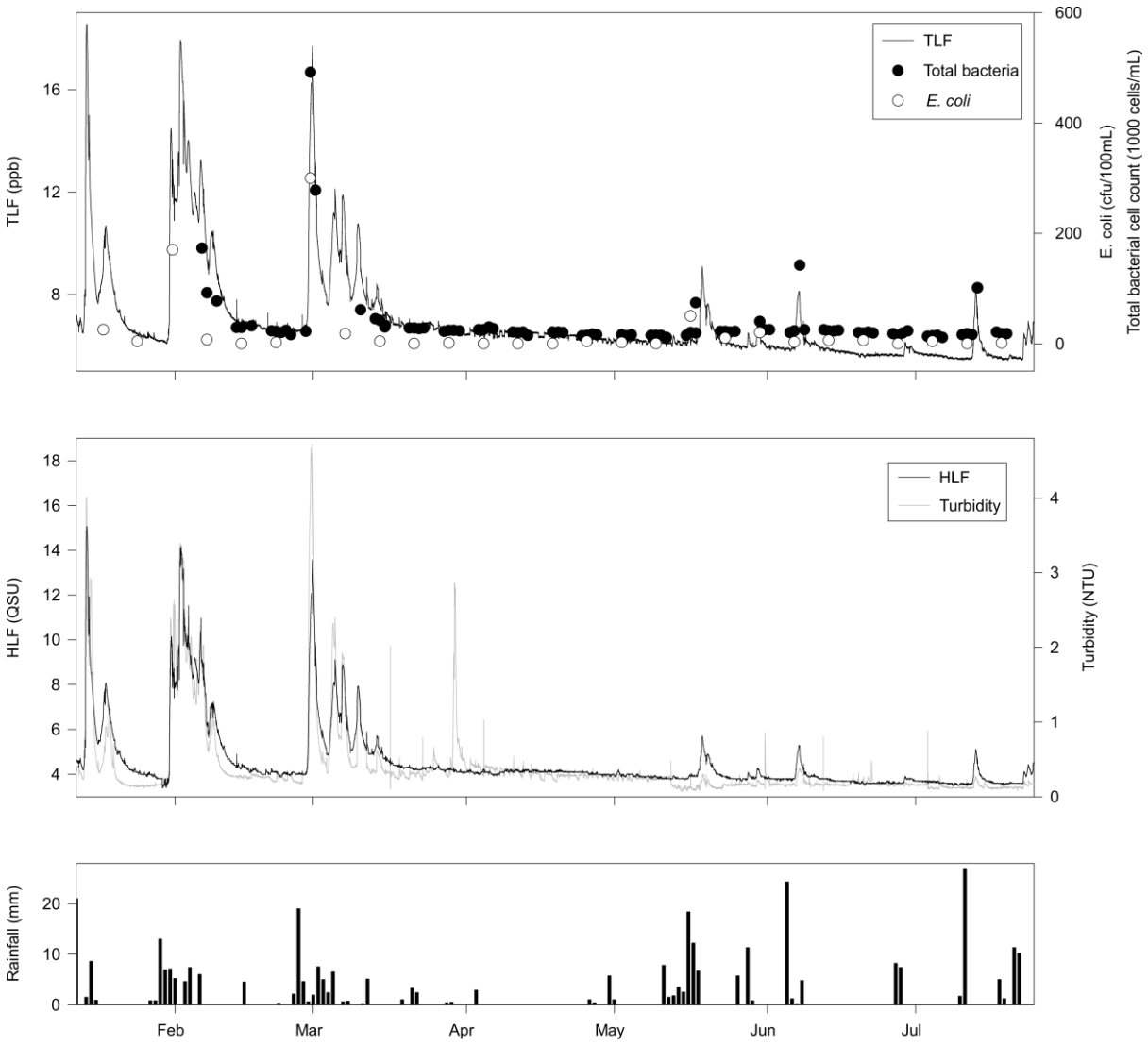

Figure 5-7 Comparison of TLF, HLF, and turbidity with bacteriological variables at site 2. All data are displayed at 15 minute resolution. No samples were taken for total bacterial cell counts prior to 6 Feb and between 1-10 Mar. Daily rainfall values are from Havant, Hampshire.

Turbidity increases appreciably during events in the main groundwater recharge season, but returns back to baseline more quickly than the bacteriological and fluorescence variables. This partially mirrors the results of Pronk *et al.* (2006) who demonstrated that following rises in turbidity, DOC and *E. coli* counts in springs after rainfall, turbidity rapidly returned to baseline values whereas both DOC and *E. coli* remained elevated for several more days. These observations may result from turbidity

predominantly being associated with larger particles, which rapidly settle out, whereas smaller particles also associated with bacteria and fluorescent DOM continue to be transported through the system. It is also noted that turbidity changes during late Spring and Summer rainfall events, are minimal despite sizeable increases in bacteriological variables. These observations indicate why turbidity is an inferior indicator of both *E. coli* and total bacterial cells compared to fluorescence at site 2.

5.3.4 HLF bleed-through into the TLF spectral region

The sensor data from sites 1 to 3 do not indicate a clear TLF peak, with relatively constant TLF:HLF ratios (IQR = 0.01–0.03) that never exceed 1.0 (Figure 5-5). Mean laboratory fluorescence data also do not indicate a visual TLF peak (Figure 5-9). The TLF region appears completely masked by the neighbouring HLF peak centred at λ_{em} 420–460 nm, which bleeds through and causes an apparent amplified TLF signal. Deconvoluting the EEMs statistically using PARAFAC analysis demonstrates that a two component model, based on HLF peaks (Figure 5-8), explains 98.3% of the variability. A validated three component model, including an additional TLF component (Figure 5-8), provides only a very marginal improvement to 98.6%. It is, hence, unsurprisingly that HLF and TLF were similarly correlated to both *E. coli* and total bacterial cells.

In other studies, HLF has proven to be similarly or better correlated than TLF to: BOD at wastewater treatment works (Cohen et al., 2014), total coliforms and *E. coli* at springs (Frank et al., 2017) and *E. coli* concentrations during controlled laboratory experiments (Fox et al., 2017). Therefore, a HLF fluorimeter alone may suffice for monitoring the microbial quality of untreated drinking water sources in many instances. Alternatively, a sensor that excites at λ_{ex} 280 nm and has a broader emission filter of λ_{em} 335-500 nm could also provide comparable performance. The advantage of this device would be a reduction in cost due to the replacement of an expensive narrow bandpass filter.

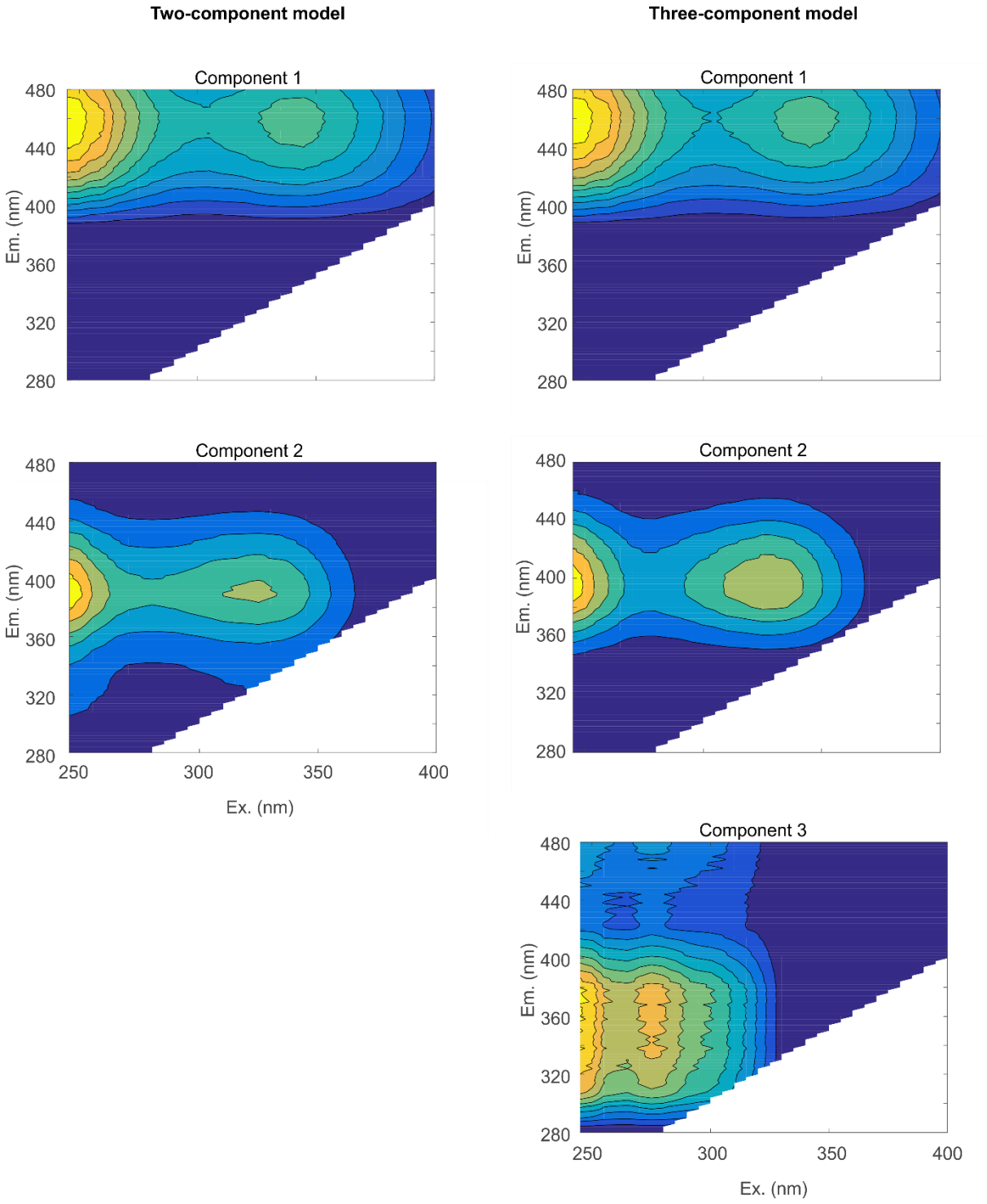

Figure 5-8 Two- and three-component models produced by PARAFAC analysis.

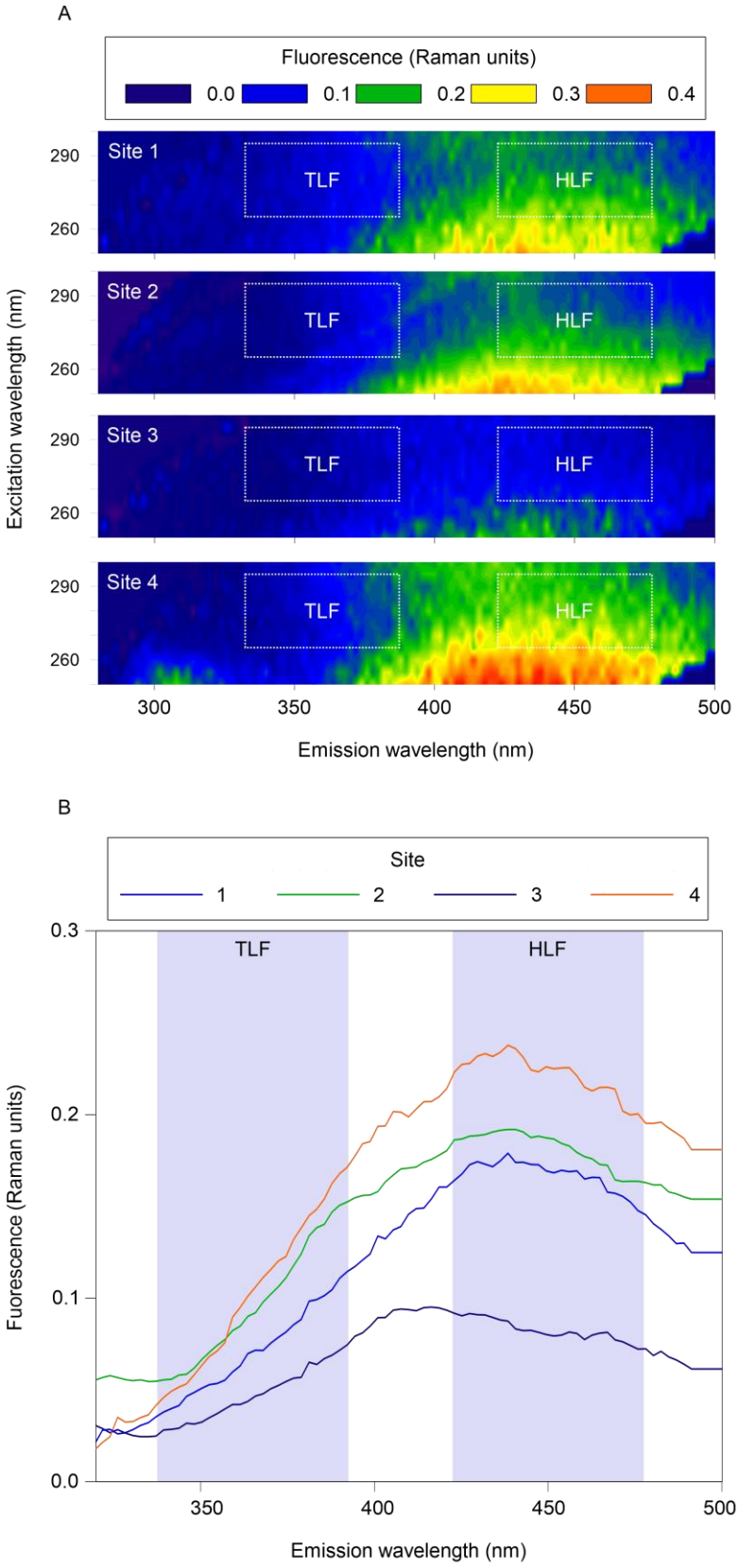

Figure 5-9 A) Mean fluorescence excitation/emission matrix between λex 250–300 nm and λem 280–500 nm for all sites with TLF and TLF sensor spectral regions indicated; B) Mean fluorescence intensity at λex 280 nm, λem 320–500 nm for all water samples collected at sites 1 to 4.

The ratios of fluorescent peaks are commonly used to address the issue of optical bleed-through (e.g. Baker, 2002b; Lapworth et al., 2008). Variability in TLF:HLF ratios between sites demonstrates that there is not a consistent ratio in the absence of a clear TLF peak at these study sites (Figure 5-5). This observation arises because fluorescence intensity is always low and relatively consistent at wavelengths slightly shorter than the TLF region (λ_{em} 320-340 nm); therefore, an intense HLF peak will typically result in a lower TLF:HLF ratio as the recession in fluorescence intensity from the peak is steeper (Figure 5-9b). Hence, site 1 with a relatively intense HLF peak has a median ratio of 0.58, as opposed to a median ratio of 0.73 at site 3 which has the lowest intensity HLF peak. Consequently, using TLF:HLF ratios does not allow comparisons between sites, although it may indicate relative changes in the fluorescent components at an individual site.

The lack of a clear TLF peak indicates the sites are not likely to have been appreciably impacted by wastewater during the monitoring period, which is the most likely attribution of TLF in groundwater. HLF is generally considered to be of allochthonous origin (Coble et al., 2014) and is likely to derive from the near-surface in groundwater. Therefore increases in fluorescence during rainfall events are indicative of water from the near-surface rapidly impacting the sites.

5.3.5 Operational constraints

At all water company sites, operating procedures impacted fluorescence monitoring. Borehole pump start-up resulted in spikes in fluorescence as well as turbidity at all borehole sites. These spikes typically lasted for minutes to hours until returning to expected levels. However, fluorescence remained elevated for longer periods (up to several days) when the primary water source was rotated to another borehole at the same site that had remained on standby for days/weeks. Boreholes contain elevated concentrations of planktonic bacteria cells and DOC compared with the surrounding aquifer when they are not pumped due to the accumulation of organic material (Sorensen et al., 2013). The longer a borehole is not pumped, the greater the accumulation of this organic material within the borehole and adjacent aquifer. Consequently, a greater load is mobilised by pump start-up and it

takes longer for fluorescence readings to stabilise. Furthermore, there is likely to be build-up of organic material within the pipe network with time when there is no active flow that would then only begin to clear upon pump start-up.

Variations in the active pumping rate within a borehole are also likely to impact fluorescence monitoring. A higher flow rate increases flow velocities and shear stress within the borehole and aquifer, increasing disturbance, entrainment, and shearing of biofilms. Previous research has demonstrated an increase in DOC mobilisation and bacterial cell counts at higher pumping rates (Graham et al., 2015; Kwon et al., 2008). A higher pumping rate through the pipe network could similarly increase biofilm mobilisation (Cloete et al., 2003), with a resultant increase in TLF and HLF.

Minor sensor drift of 0.3–0.8 ppb TLF over 6 to 10 months was evident at all sites and is the underlying cause of the change in baseline fluorescence over time at Site 2. Sensor drift was evaluated by testing the fluorimeters in ultrapure water and tryptophan standards before installation and following removal at the end of the project. There was no visible sign of material on any sensor window, but the apparent drift could be corrected for by cleaning the windows with a lens cloth. It is recommended that suppliers supplement these optical sensors with self-cleaning wipers on the windows for online installations. This should resolve the problem, but infrequent manual maintenance will still be required to validate the sensor readings. More problematic was an installation at a riverbank filtration site that was not included in this paper. Here, there was a dramatic >40% loss of signal in the first two weeks following installation (Figure 5-10). The sensor was cleaned and re-installed on multiple occasions, but similar rates in signal loss were always observed. Hydrochemical analysis of deposits removed from the windows indicated a build-up of metals, most notably iron. The sensor windows are constructed from silica, which has a surface comprising mainly silanol (Si-O-H) groups. Any loss of the proton could allow the windows to become negatively charged, attracting cations, and there could also be exchange of the proton with cations in the sample.

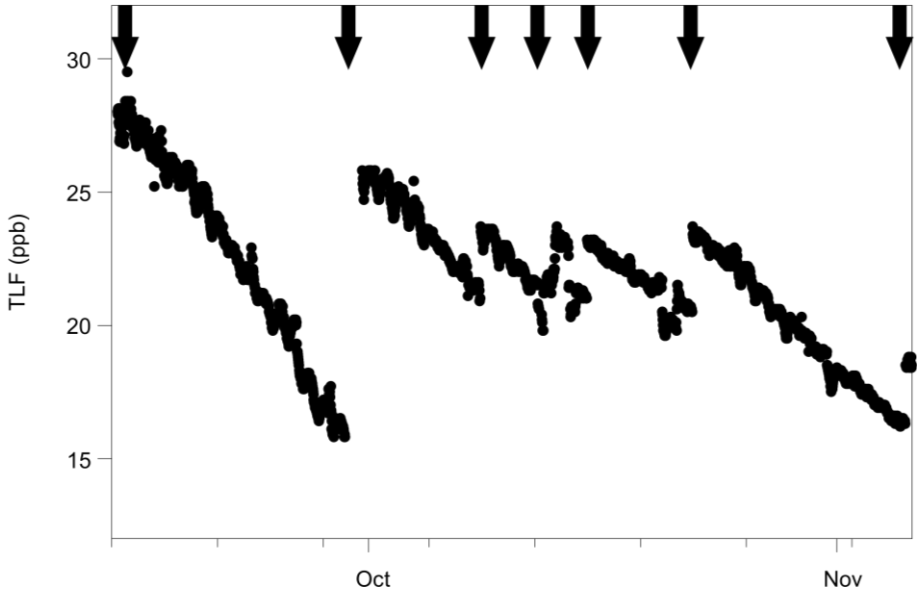

Figure 5-10 TLF variation at a river bank filtration scheme (site 5). Arrows indicate a site visit when the sensor was cleaned and re-installed.

There are a range of other variables that interfere with fluorescence spectroscopy measurements (Hudson et al., 2007), notably temperature and turbidity (Khamis et al., 2015). Turbidity was generally very low with the 95th percentile being 1.0, 0.1, 0.1 and 1.6 NTU at sites 1 to 4, respectively. Temperature was similar at all sites, with median values of between 10.6 and 12.2 ^{o}C, with variations typically limited to only 1–2 ^{o}C, and only very occasional greater variation. Therefore, neither temperature nor turbidity are likely to have had appreciable influence on the fluorescence results. It is possible to generate correction factors to account for variations in temperature and turbidity, which could be automated online (e.g. Shutova *et al*. 2016*)*, and these may be necessary for surface water sources. However, Khamis *et al*. (2015) demonstrated only marginal improvements by correcting groundwater TLF data and suggest that corrections factors are likely to be unnecessary for groundwater sources universally, as generally turbidity is very low and temperature is perennially stable.

5.3.6 Beyond an indicator of microbial quality in untreated water

Online fluorescence has viable applications beyond purely as an indicator of the microbial quality of untreated drinking water. Bridgeman *et al*. (2015) identified a correlation between TLF and total

bacterial cell counts within the treated drinking water distribution network; albeit slightly weaker (r^2 = 0.56) than in this study. The correlations between fluorescence and total bacterial counts, observed herein and in Bridgeman et al. (2015), indicate that online fluorescence spectroscopy could represent a more practical, cheaper, and robust alternative to flow cytometry for monitoring total bacteria throughout the water supply network. However, flow cytometry has the potential to record other properties of bacterial communities, including cell size, viability (the proportion of live and dead cells) and the metabolically active proportion of the community (Hammes and Egli, 2010).

Within the treated water distribution network it has been demonstrated that HLF is relatively stable (Heibati et al., 2017) so that online fluorescence would be a potentially sensitive indicator of any water quality changes. Potential changes could include cross-connection detection, which many researchers have highlighted as a future use for online fluorescence (e.g. Hambly et al., 2010), or failure in the water treatment system. Further uses could be strategic deployments at the outlets of service reservoirs and large diameter trunk mains, which can both suffer from structural integrity issues. Nevertheless, the use of online fluorescence spectroscopy for treated water is likely to be complicated by water treatment processes that remove fluorescent DOM to various degrees (Carstea et al., 2016). Further work is need to appraise the utility of the technique on treated water.

Fluorescence targeting HLF ($\lambda_{ex}/\lambda_{em}$ of around 350/450 nm) has been correlated with DOC (Shutova et al., 2016) and TOC (Bieroza et al., 2009; Bridgeman et al., 2015; Stedmon et al., 2011) in untreated and treated drinking water. If peak C is related to the total organic matter load in water, then online HLF fluorimeters could be useful to forewarn of the generation of harmful amounts of carcinogenic by-products during certain types of disinfection such as trihalomethanes (Yang et al., 2015b).

Fluorescence spectroscopy could also be a useful indicator for the water industry, or other abstractors, for rapidly characterising the vulnerability of groundwater sources to contamination. The higher the fluorescence (either TLF or HLF) then the stronger the likely link with the near-surface, which is the

source of the majority of microbial and chemical contaminants. Therefore, fluorescence could be considered an instantaneous assessor of risk for groundwater sources.

5.4 Conclusions

- Online tryptophan-like fluorescence (TLF) and humic-like fluorescence (HLF) were both strongly correlated with *E. coli* concentration and total bacterial cell counts at public water supplies.
- The current commonly employed microbial indicator in the UK, turbidity, was more weakly correlated with both bacterial variables. Nevertheless, turbidity is still an essential indicator of suspended solids in raw water that can affect treatment efficacy.
- Monitoring both TLF and HLF was unnecessary in this environment with minimal change in the ratio. A HLF sensor alone would be sufficient for evaluating when water from the near-surface was impacting a groundwater supply.
- Fluorescence data were strongly influenced by pump start-up at borehole sources and, at one site, continuous build-up of ferric deposits on the sensor reduced intensity by up to >40% within two weeks. There is a need for manufacturers to supply fluorimeters that are more resistant to fouling, for example through the provision of wiper systems.
- Online fluorescence could be a more practical approach for monitoring total bacterial cell counts than flow cytometry.
- Online fluorescence is an effective indicator of the microbial quality of untreated drinking water and could be effective throughout the water supply chain for identifying changes in water quality that could signify the presence of enteric pathogens.

Chapter 6 - Extracellular nature of common fluorophores in groundwater

6.1 Introduction

This chapter addresses **Objective 6** and hypothesises that tryptophan-like and humic-like fluorophores in groundwater are predominantly extracellular. This hypothesis is tested using data collected from four countries with contrasting hydrogeological settings and pollution pressures, and the implications for TLF and HLF as faecal indicators are discussed.

This chapter is published as:

Sorensen, J.P.R.; Carr, A.F.; Nayebare, J.; Diongue, D.M.L.; Pouye, A.; Roffo, R.; Gwengweya, G.; Ward, J.S.T.; Kanoti, J.; Okotto-Okotto, J.; van der Marel, L.; Ciric, L.; Faye, S.C.; Gaye, C.B.; Goodall, T.; Kulabako, R.; Lapworth, D.J. ; MacDonald, A.M. ; Monjerezi, M.; Olago, D.; Owor, M.; Read, D.S.; Taylor, R.G.. 2020 Tryptophan-like and humic-like fluorophores are extracellular in groundwater: implications as real-time faecal indicators. Scientific Reports, 10, 15379, 9.

6.2 Materials and Methods

6.2.1 Study areas

The four study areas together comprise varying degrees of urbanisation from a large city to rural context where pollution sources and pressures vary considerably. Dakar is the large capital city of Senegal with over three million inhabitants constrained within the Cap-Vert peninsular, Kisumu is a medium-sized city of 600,000 residents in Kenya, Lukaya is a small town of around 24,000 people in Uganda, and the rural communities were located in the Lilongwe & Balaka Districts of Malawi (Figure 6-1).

The four study areas have contrasting hydrogeological settings. The unconfined shallow Thiaroye aquifer of Dakar comprises Quaternary fine- and medium-grained sands with a shallow water table less than 2 m below ground level (bgl) (Faye et al., 2019; Faye et al., 2004). The heterogeneous

volcano-sedimentary Kisumu aquifer system is a suite of Archaean age metasediments, Tertiary volcanics, Quaternary sediments, colluvium and lateritic soils (Consultants, 1988; Okotto-Okotto et al., 2015; Olago, 2019). It is multi-layered with unconfined (5-30m) and confined (>30m) horizons. Lukaya predominantly sits on Precambrian Basement with aquifers developed within the weathered overburden and fractured bedrock, in addition to alluvial aquifers towards Lake Victoria. Rest water levels are between 0.5 and 9 m bgl. Lilongwe District is Precambrian Basement with deeper rest water levels between 15-25 m bgl (Chavula, 2012). The alluvial aquifers in Balaka District are dominated by clays with significant subordinate sand horizons and a typical rest water level of 5-10 m bgl (Chavula, 2012). Sanitation in all the study areas consists of mainly onsite sanitation (Cole et al., 2013; Diédhiou et al., 2012; Nayebare et al., 2020; Wright et al., 2013), which has the potential to faecally contaminate underlying groundwater resources that are used by communities in all settings (Sorensen et al., 2016).

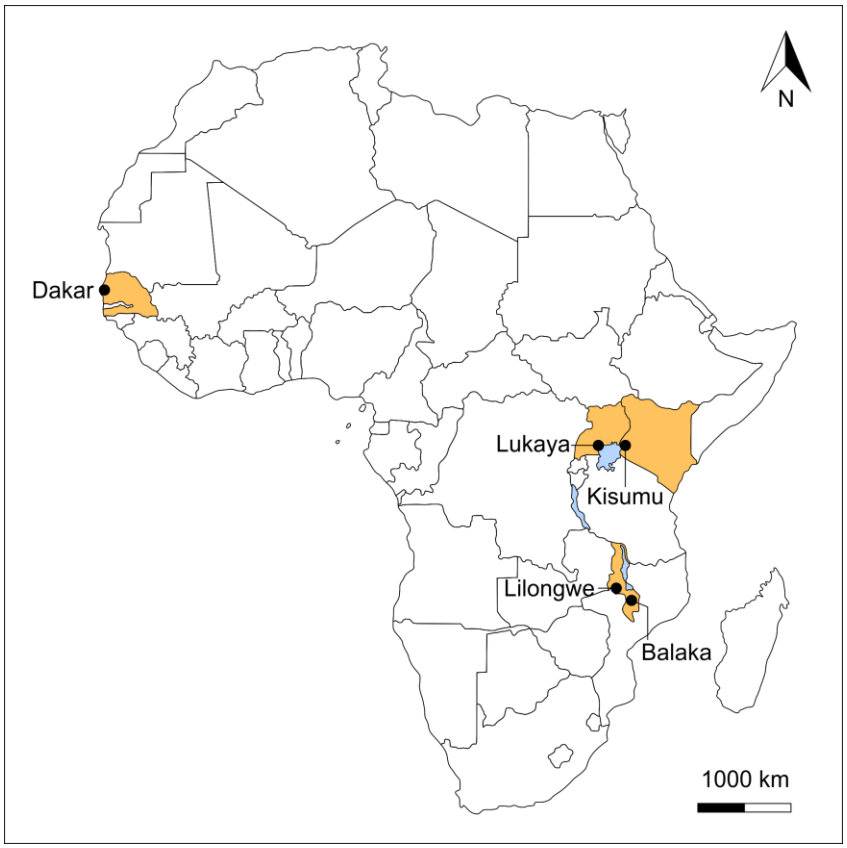

Figure 6-1 Study site locations in Africa. Continental map modified from https://online.seterra.com/pdf/africa-countries.pdf.

6.2.2 Groundwater sampling and analysis

A total of 140 groundwater sources were sampled: 29 in Dakar, 38 in Kisumu, 32 in Lukaya, and 41 in Lilongwe & Balaka Districts. Sampling was undertaken in the dry season in Dakar and Kisumu and the wet season in Lukaya and Lilongwe & Balaka Districts. The sources comprised a mixture of pumped boreholes, hand pumped boreholes, open wells, and springs. Samples were obtained from open wells using a 12V submersible WaSP-P5 pump, except for Lilongwe & Balaka Districts where a rope and bucket were used. Prior to sampling, boreholes, open wells (except Lilongwe & Balaka Districts) and hand pumps flowed for at 1-2 minutes to ensure pipework was flushed and the sample was representative of the source. Springs were sampled directly from the outlet: either a discharge pipe in a protected setting or from the surface water channel in an unprotected setting.

TLF and HLF were determined using portable UviLux fluorimeters reporting in QSU (Section 2.3). Fluorescence analysis was conducted in a HDPE beaker placed within a covered black container to prevent interference from sunlight. Analysis was conducted on unfiltered water to indicate total fluorescence, then passed through a low-protein binding 0.22 µm PVDF membrane (Sterivex, Merck KGaQ, Germany) to sterilise the water and quantify extracellular fluorescence.

Repeatability of TLF fluorimeter data was previously investigated in the laboratory using dissolved tryptophan standards (Chapter 3). This study indicated that repeatability was approximately 0.2-0.6 QSU up to 100 QSU, with evidence that absolute repeatability decreased with increasing intensity. To address repeatability in a field situation, including HLF, we calculated 3σ of 74 duplicated measurements in Lukaya. These data ranged between 0-16.5 and 0-49.2 QSU for TLF and HLF, respectively. Repeatability was 0.5 QSU for TLF and 0.3 QSU for HLF.

Specific electric conductivity (SEC), pH, temperature and turbidity were quantified using multiparameter Manta-2 sondes (Eureka Waterprobes, USA) in Dakar, Kisumu, and Lukaya. In

Lilongwe & Balaka Districts, temperature, SEC, and turbidity were measured using a HI766EIE1 thermocouple with HI935005 thermometer (Hanna Instruments, USA), S3 portable conductivity meter (METTLER TOLEDO, USA), and 2100Q turbidimeter (Hach Company, USA), respectively.

Thermotolerant coliform (TTC) samples were collected in sterile 250 mL polypropylene bottles and stored in a cool box (up to 8 h) before analysis (Section 2.2). Samples in Dakar and Lukaya for total (planktonic) bacterial cells were collected in 4.5 mL polypropylene cryovials (STARLAB, UK) that were pre-treated with the preservative glutaraldehyde and the surfactant Pluronic F68 (Marie et al., 2014) at final concentrations of 1% and 0.01%, respectively. The samples were frozen at -18^{o}C within 8 h of collection, defrosted overnight during transit to the UK, and analysed the following morning on a BD Accuri C6 flow cytometer equipped with a 488 nm solid state laser (Becton Dickinson UK Ltd, UK). Water samples (500 mL) were stained with SYBR Green I (Sigma-Aldrich, UK) at a final concentration of 0.5% for 20 min in the dark at room temperature, before running on the Accuri at a slow flow rate (14 mL/min, 10 mm core) for 5 min and a detection threshold of 1500 on channel FL1 (Chapter 5). A single manually drawn gate was created to discriminate bacterial cells from particulate background, and cells per mL were calculated using the total cell count in 5 min divided by the reported volume run in µL.

6.2.3 Statistical analyses

The non-parametric paired Wilcoxon test was used to assess the impact of filtration using the null hypothesis that the median difference following filtration is zero (Hollander and Wolfe, 1973). Relationships between loss in fluorescence and various independent variables (total bacterial cells, TTCs, SEC, and turbidity) were assessed using the non-parametric Spearman's Rank test (Spearman, 1904). Non-parametric techniques were used because of the non-Gaussian distribution of the datasets. All analyses were undertaken in R version 3.4.0 using core commands *wilcox.test* and *cor.test*(R Core Team, 2020). Boxplots display the median, the interquartile range, whiskers denote

that 10th and 90th percentiles, and dots the 5th and 95th percentiles; these were produced in SigmaPlot version 13.0.

6.3 Results

6.3.1 Variation in TLF and HLF in the study areas

The intensity of total TLF/HLF in groundwater relates to the degree of urbanisation (Figure 6-2a). Median TLF reduces from 17.4 QSU in the large city of Dakar, to 7.0 QSU in the medium-sized city of Kisumu, and 1.1-1.2 QSU in the small town of Lukaya and rural Lilongwe & Balaka Districts. Furthermore, median HLF in Dakar is almost 90-fold that of Lukaya.

Neither variations in water temperature nor turbidity is considered to appreciably impact the fluorescence results (Figure 6-2b-d). Water temperature across all data vary between 23.2 and 29.6 ^{o}C and within individual study areas by 2.5-6.0 ^{o}C. Therefore, uncertainty relating to temperature is likely to be limited to a maximum of <6 and <9% for TLF and HLF, respectively (Baker, 2005; Downing et al., 2012; Watras et al., 2011). Any optical attenuation relating to suspended solids is also likely to have limited influence on TLF/HLF with a median turbidity of 0.3-4.6 NTU and 97% all of data <46.2 NTU (Khamis et al., 2015; Saraceno et al., 2017).

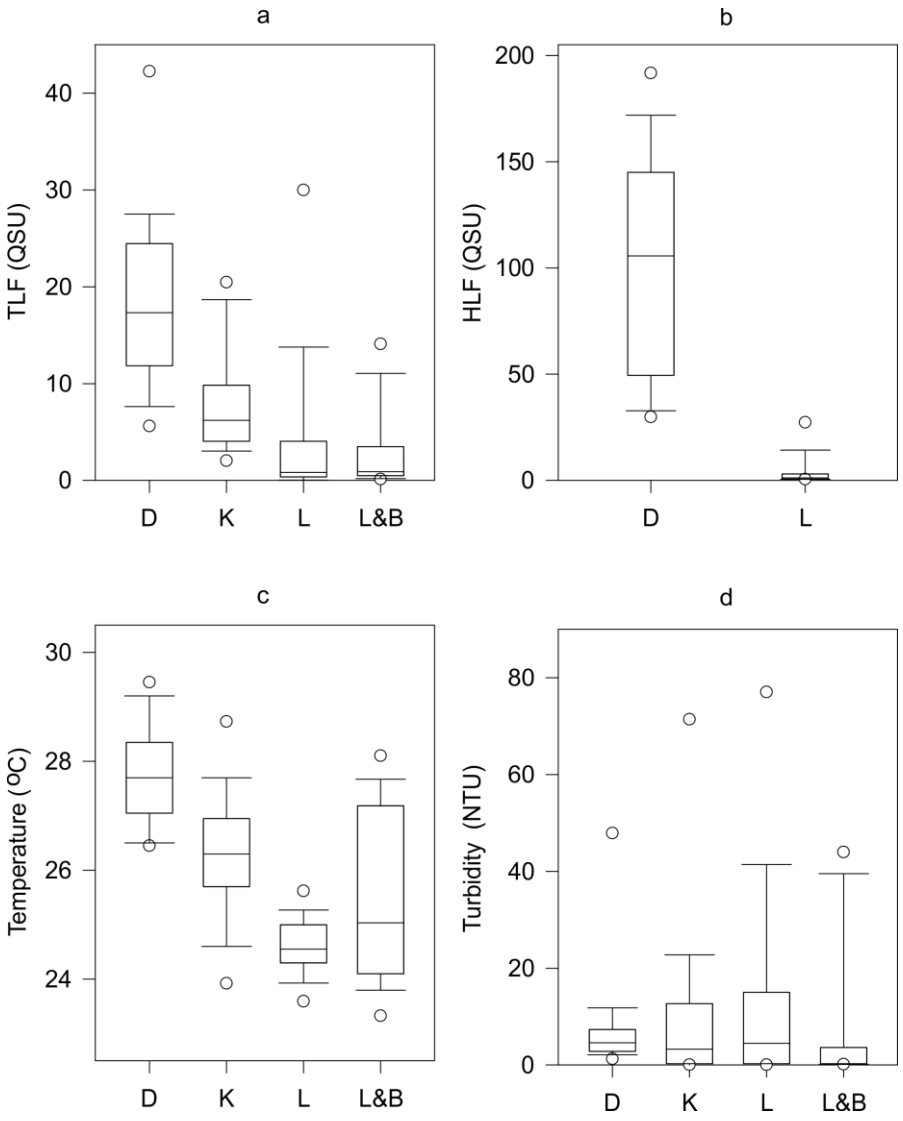

Figure 6-2 Boxplots of (a) total TLF, (b) total HLF, (c) temperature and (d) turbidity across D = large city of Dakar, K = medium-sized city of Kisumu, L = small town of Lukaya, and L&B = rural Lilongwe & Balaka Districts.

6.3.2 TLF/HLF are predominantly extracellular in groundwater

There is a significant change in median TLF following filtration (ΔTLF) across the whole dataset (paired Wilcoxon, $p < 0.001$). Median ΔTLF is a decline of 0.2 QSU (3.1%), which is within the error of repeatability (Figure 6-3). The majority (68.6%) of the supplies show no change when considering repeatability uncertainty, with 28.6% declining, and 2.9% increasing. The lower and upper quartiles highlight limited changes of only -12.2 and 2.6%, respectively.

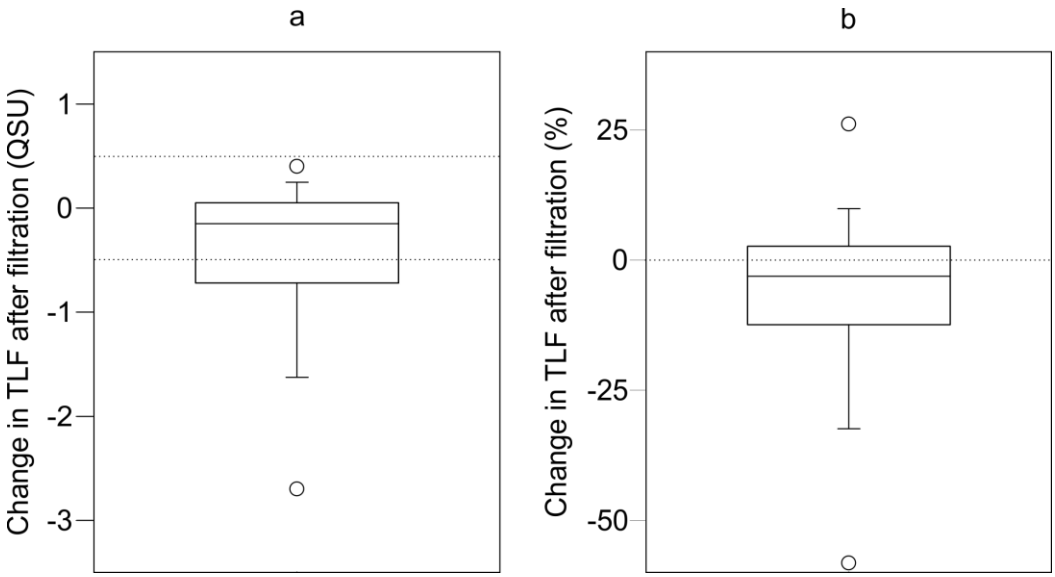

Figure 6-3 Change in TLF following filtration in (a) QSU and (b) as a percentage. Dotted lines in (a) denote error in repeatability and (b) is a reference zero line.

Within the individual country datasets, significant changes following filtration were observed in Dakar and Kisumu, but not Lukaya nor Lilongwe & Balaka Districts (Figure 6-4). The lack of a significant change in Lukaya and Lilongwe & Balaka Districts could be a result of the low unfiltered TLF intensities (median 1.1-1.2 QSU). Consequently, any changes following filtration could be harder to detect amongst repeatability uncertainty (±0.5 QSU). Overall median changes following filtration in Dakar, Kisumu, Lukaya and Malawi were -0.4 (-2.8%), -0.3 (-4.2%), -0.1 (-6.2%), 0 (0%) QSU, respectively. The greatest loss in TLF following filtration across the entire dataset was at a spring in Lukaya 28.6 QSU (78.6%), with a further spring in the town experiencing a loss of 2.7 QSU (69%). These springs were gentle seepages, and could be considered more representative of slow-moving surface waters; algae were also visible in the channel where samples were obtained. Given any changes following filtration are insignificant or minimal within country datasets collected in both the dry and the wet seasons, TLF is likely to be predominantly extracellular throughout the year.

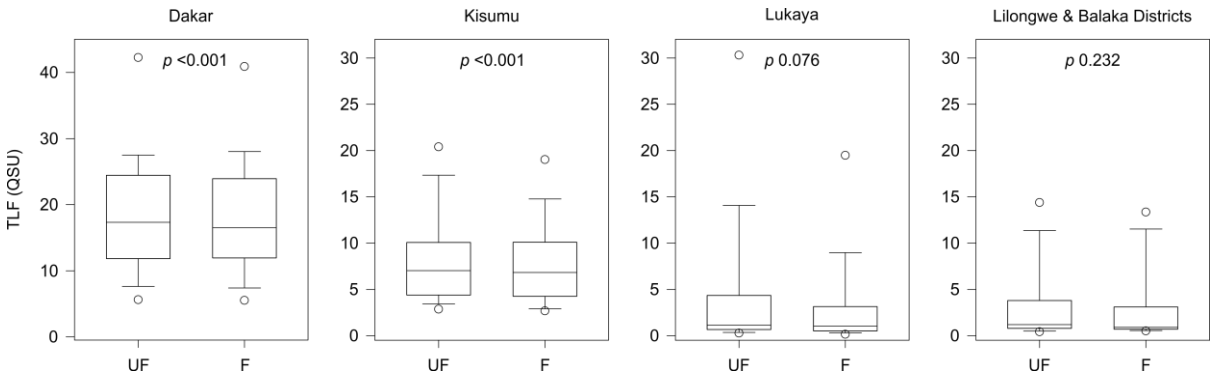

Figure 6-4 Comparative boxplots of unfiltered (UF) and filtered (F) TLF data for each study area. Displayed p-values are the results of paired Wilcoxon signed rank tests.

There was no significant change in HLF following filtration (Figure 6-5) across the entire dataset (paired Wilcoxon, p 0.861), or within both the Dakar (paired Wilcoxon, p 0.522) and Lukaya (paired Wilcoxon, p 0.704) datasets individually (Figure 6-7). Although TLF and HLF are correlated in Dakar (ρ_s 0.678, p <0.001), contrasting filtration effects suggest that fluorescence is, in part, emanating from two different sources.

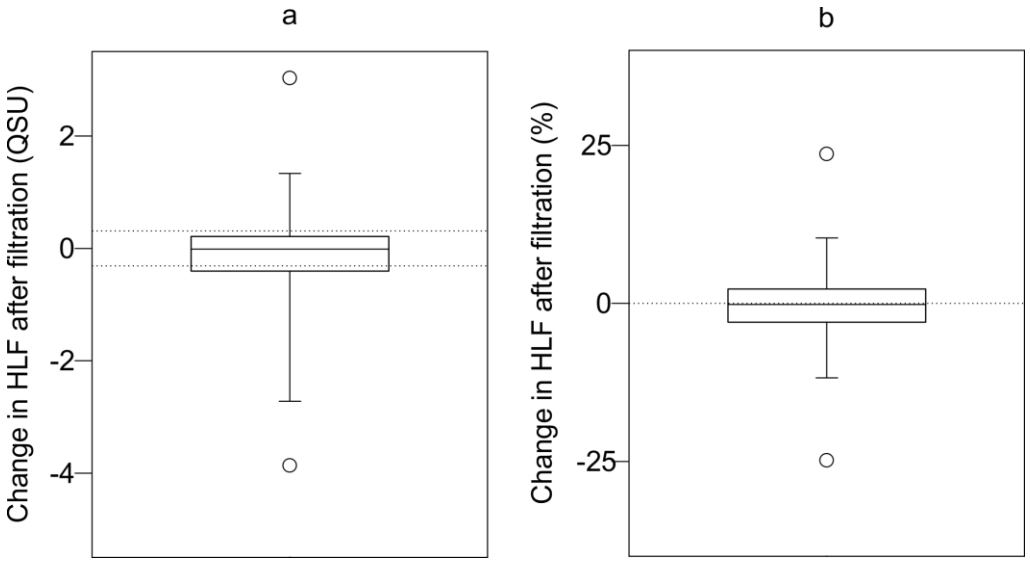

Figure 6-5 Change in HLF following filtration in (a) QSU and (b) as a percentage. Dotted lines in (a) denote error in repeatability and (b) is a reference zero line.

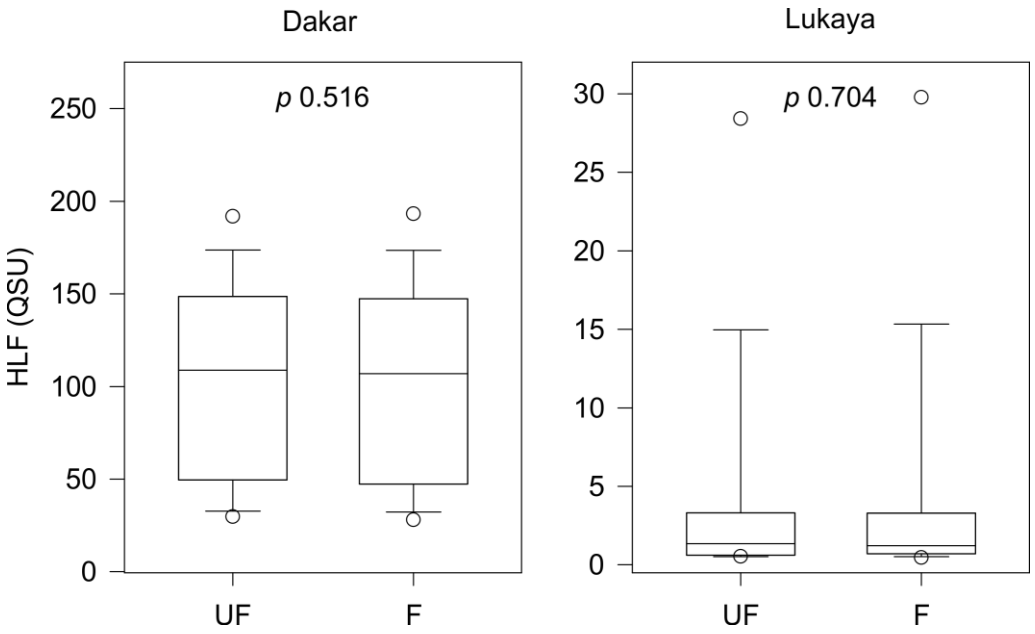

Figure 6-6 Comparative boxplots of unfiltered (UF) and filtered (F) HLF data for each study area. Displayed p-values are the results of paired Wilcoxon signed rank tests.

6.3.3 Relationships between TLF loss, total bacterial cells, TTCs, SEC, and turbidity

There is a moderate positive correlation between ΔTLF after filtering and total bacterial cells in Dakar (Figure 6). Note, though, that there is a significant tendency for ΔTLF to decrease with increasing TLF intensity (Figure 6-8). No other significant correlations between ΔTLF and other variables in any country, including turbidity (Figure 6-7). It is unsurprising there is no significant correlations in Lukaya and Lilongwe & Balaka Districts, where no significant ΔTLF was observed, but coefficients are included for completeness.

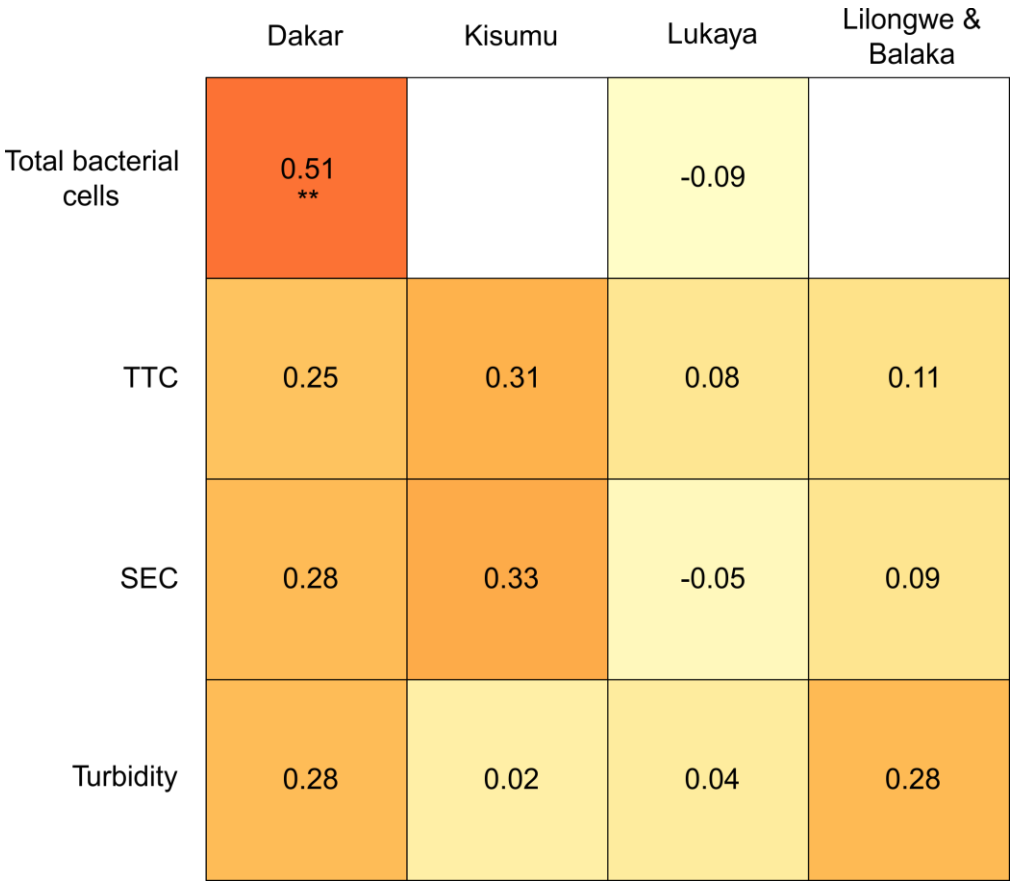

Figure 6-7 Correlation matrix between four independent variables and TLF change following filtration in each of the four countries. Displayed values are Spearman's ρ_s and ** denotes a p value of 0.01.

6.4 Discussion

6.4.1 Implications for TLF/HLF as faecal indicators

TLF and HLF are predominantly extracellular in groundwater, which supports findings from our earlier TLF pilot study in rural India (Sorensen et al., 2016). Extracellular TLF/HLF will have different transport properties to larger faecal indicator bacteria and enteric pathogens, such as bacteria, *Cryptosporidium* oocysts, and *Giardia* that are around 1, 5, and 10 µm, respectively (Hunt and Johnson, 2017). These organisms will be more readily strained between a faecal source and a groundwater source through both the unsaturated and saturated zones. This behaviour could potentially result in false-positives when organisms are removed completely, whilst an elevated TLF/HLF signal remains. Indeed, false-positives have been highlighted as an issue when defining TLF thresholds to indicate the presence of TTCs (Chapter 4). False-positives will be more likely in aquifers

exhibiting matrix flow where faecal indicator bacteria and pathogens are typically restricted to only metres or tens of metres from sources such as pit latrines (Graham and Polizzotto, 2013). Co-transport of TLF/HLF and both TTCs and pathogens is more likely in fractured aquifers where rapid transport of bacteria can occur over several kilometres within a few days, for example Heinz *et al.* (Heinz et al., 2009), or alternatively, irrespective of the aquifer, where a source's integrity is compromised at or near the surface. Co-occurrence of TLF/HLF and TTCs is also more probable where there is a very shallow water table (<1 m).

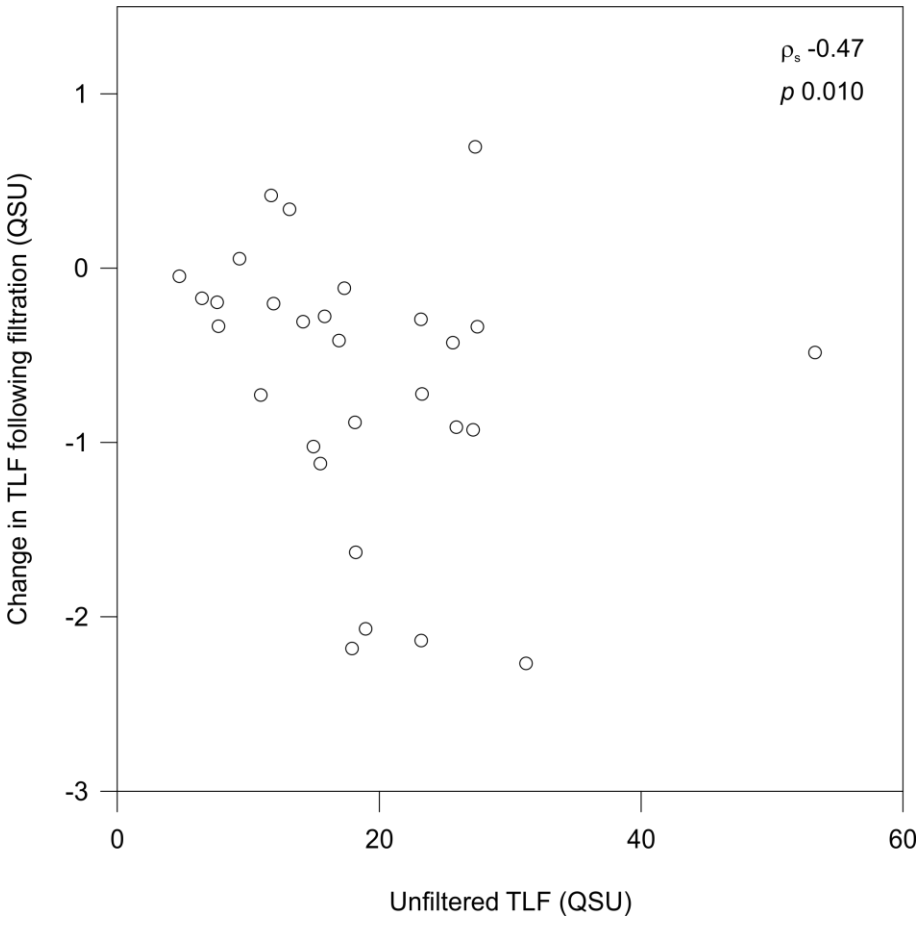

Figure 6-8 Relationship between unfiltered TLF and change in TLF following filtration beneath Dakar. Spearman's Rank and associated p-value displayed.

Elevated TLF/HLF in the absence of TTCs may still indicate a groundwater source is at-risk. Viruses are the smallest pathogens (27-75 nm) and are often also found in the absence of faecal indicator bacteria and larger pathogens because they can be more mobile in the subsurface (Borchardt et al., 2004; Wu et al., 2011). Future work could explore TLF as an indicator of virus contamination where

faecal indicator bacteria are ineffective. We also previously demonstrated through seasonal sampling that TLF tended to remain perennially elevated in some sources whilst TTCs were more transient (Chapter 3). Finally, elevated TLF/HLF is likely to mean a source is better connected to the near surface and potential sources of contamination, even if the fluorescent OM may currently relate to non-faecal sources such as plant litter and soil organic matter.

Our results demonstrate a significant correlation between ΔTLF and total bacterial cells in Dakar, and not between ΔTLF and TTCs. These observations may suggest a proportion of tryptophan-like fluorophores are bound within larger protein molecules within cells, although the relationship could be coincidental and a result of other particulate OM being filtered out. Furthermore, it is supportive of TLF being associated with total microbial biomass and activity, as opposed to purely TTCs, such as *E. coli*. Previous work has shown a vast assortment of microbes exhibit TLF (Bridgeman et al., 2015; Dartnell et al., 2013; Determann et al., 1998), including ubiquitous species in the environment (Elliott et al., 2006a; Fox et al., 2017; Nakar et al., 2019).

6.4.2 Implications for fluorescence sampling

The extracellular nature of TLF/HLF means there are minimal concerns over comparing groundwater fluorescence data collected from in-situ field sensors and samples that have been filtered prior to laboratory analysis. The filtration of laboratory samples is undertaken in many studies to remove suspended solids and sterilise the water to minimise biologically driven OM transformation prior to analysis.

There is no evidence for turbidity attenuating TLF/HLF in groundwater (Figure 6). Samples with the highest turbidity in each country (46-153 NTU) also showed no evidence of appreciable change in TLF following filtration (1, 2, -3, 5%, respectively). This result further supports Khamis *et al. (2015)* in suggesting turbidity is unlikely to have any impact on the in-situ fluorescence monitoring of groundwater. With no observed relationship between ΔTLF and turbidity, and no significant ΔHLF, it is unconfirmed why some ΔTLF/ΔHLF are positive beyond any repeatability error. Many of these

positive data are at greater fluorescent intensity, than the repeatability study conducted herein, with previous evidence that error increases with greater intensity; hence, these positives values may not indicate an appreciable change. However, it is also possible some of these data are indicative of either anomalously low total or high extracellular fluorescence. Anomalous low readings could occur due to air bubbles trapped within the sensor and high readings due to contamination resulting from sensor handling (Chapter 4).

6.5 Conclusions

Tryptophan-like fluorophores are predominantly extracellular in groundwater. Significant changes in TLF following filtration were only observed beneath the cities of Kisumu, Kenya, and Dakar, Senegal, where TLF was elevated in comparison to the small town of Lukaya, Uganda and rural Lilongwe & Balaka Districts, Malawi. Nevertheless, TLF was still 93.2-97.2% extracellular on average. In Dakar, tryptophan-like fluorophores associated with the unfiltered >0.22 µm fraction were moderately correlated with total bacterial cells. Humic-like fluorophores are extracellular in groundwater.

The extracellular nature of TLF/HLF means they will have different transport properties in comparison to faecal indicator bacteria and larger pathogens, which will be more readily strained in both the unsaturated and saturated zones of the subsurface. Co-transport is least likely in intergranular aquifers where the water table exceeds a metre depth below ground level resulting in a higher likelihood of false-positives. TLF/HLF should be considered more precautionary indicators of microbial risks than faecal indicator bacteria in groundwater-derived drinking water.

Chapter 7 – Relationship between in-situ fluorescence and total bacterial abundance

7.1 Introduction

This chapter primarily addresses **Objective 5** and explores the relationships between in-situ fluorescence spectroscopy and total bacterial abundance, as well as common parameters associated with faecal contamination such as TTCS (**Objective 1**), nitrate and dissolved organic carbon, across a wide-range of drinking water sources. The study area is a highly contaminated aquifer beneath a suburb of Dakar (Senegal), Thiaroye, where on-site sanitation (OSS) comprises the only means of sewage disposal.

This chapter is published as:

Sorensen, J.P.R.; Diaw, M.T.; Pouye, A.; Roffo, R.; Diongue, D.M.L.; Faye, S.C.; Gaye, C.B.; Fox, B.G.; Goodall, T.; Lapworth, D.J. ; MacDonald, A.M. ; Read, D.S.; Ciric, L.; Taylor, R.G.. 2020 In-situ fluorescence spectroscopy indicates total bacterial abundance and dissolved organic carbon. Science of the Total Environment, 738, 139419, 10.

7.2 Methods

7.2.1 Study area

The unconfined Thiaroye aquifer beneath a suburb of Dakar is located on the Cap-Vert Peninsula (Figure 7-1). The aquifer comprises Quaternary fine- and medium-grained aeolian sands overlying low-permeability Tertiary marl deposits (Faye et al., 2019; Faye et al., 2004), with no humiferous layers (Fall, 1986; Martin, 1970). The sands are 5 to 75 m thick depending on the morphology of the Tertiary deposits. It is bounded by the Atlantic Ocean to the North and Southwest, the Tertiary marl deposits that outcrop to the South, a piezometric ridge separating it from the infrabasaltic aquifer to the west, and the Tanma depression and its seasonal lake in the east. The water table is typically within 2 to

15 m of the ground surface and the hydraulic gradient is generally from southeast towards the Atlantic Ocean.

Monsoonal rainfall occurs between July and October and provides the only annual precipitation that is typically 450 to 500 mm (Faye et al., 2019). The semi-arid peninsula has an absence of perennial surface water features with the exception of hypersaline Retba Lake located below sea level. However, numerous seasonal lakes are expressions of a rise in the shallow water table. Tritium (^{3}H), stable isotope ratios of O ($\delta^{18}O$) and H ($\delta^{2}H$), and piezometric data suggest groundwater is predominantly modern (post-1963), and diffuse recharge occurs during the latter part of the monsoon once soil moisture deficits are overcome (Faye et al., 2019).

The Thiaroye aquifer contributed ~50% of Dakar's water supply in the 1980s (Faye et al., 2019; Faye et al., 2004). Groundwater withdrawals have now decreased to ~5% of total municipal supply due to exorbitantly high nitrate (NO_3^-) concentrations that had increased to an average of 450 mg/L by 2008 (Diédhiou et al., 2012). Stable isotopes of dissolved nitrate ($\delta^{15}N$ and $\delta^{18}O$) indicate an organic source of contamination (Diédhiou et al., 2012; Re et al., 2011) that is likely to relate to the vast network of septic tanks (Diaw et al. 2020, Diédhiou et al., 2012), as there is a lack of mains sewerage outside the historic city centre in the far west of the peninsula. Furthermore, Diaw *et* al. (2020) demonstrate a significant linear relationship between the density of OSS and NO_3^- concentrations. Notwithstanding high levels of contamination, the aquifer remains essential for self-supply due to limited access to piped water.

7.2.2 Groundwater sampling and analysis

Sampling sites

Eighty-four groundwater samples were collected from 70 water sources between 29th May and 3rd July 2018 (Figure 7-1). Fourteen of the sources were resampled after 27th June following the only precipitation event (19 mm at Yoff) in the sampling period, as part of another study investigating temporal dynamics in water quality in the aquifer. These 14 resampled data were included following

confirmation that the outcomes of the statistical analysis presented in the results remained unchanged. The 70 sources comprise: 41 drilled hand-pumped boreholes, 1 borehole equipped with a submersible pump, 22 dug-wells, which are all used for water supply, and 6 research piezometers. Samples were obtained directly from the source where there was a pump *in-situ*, using a 12V submersible WaSP-P5 pump (In-Situ Europe, Redditch, UK) at hand dug-wells, and an MP1 pump (Grundfos, Bjerringbro, Denmark) set at around 1 m^3/h in piezometers. All sources flowed for at least a minute before sampling to ensure all pipework associated with the pump was flushed, but the sources were not purged so the water was representative of what was actually being used for supply.

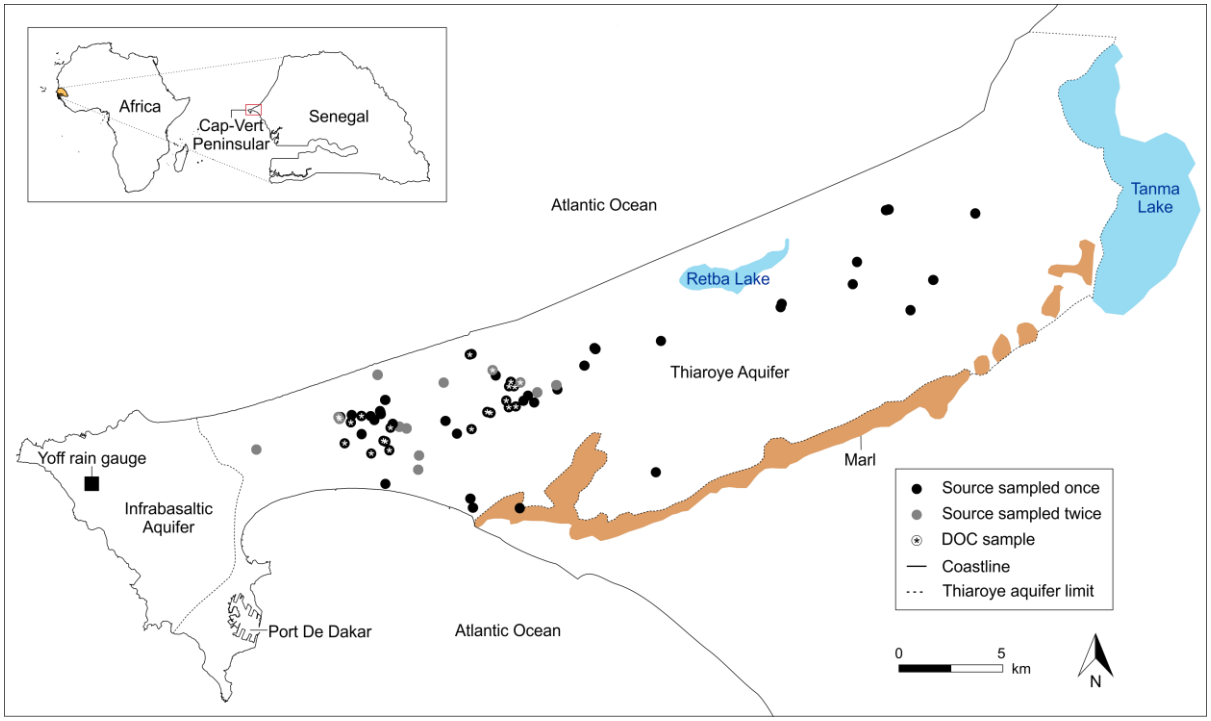

Figure 7-1 Location of Thiaroye aquifer on the Cap-Vert Peninsula and sources sampled in this study.

Sanitary inspection surveys were undertaken at all sources to assess potential risks of contamination observable at the surface (WHO, 1997). These surveys consist of a list of proscribed yes-no questions pertaining to the presence of potential sources of contamination (e.g. on-site sanitation (OSS), animal faeces, trash) and pathways for contamination to migrate rapidly into groundwater sources. The sanitary risk score (SRS) is the total number of positive responses to these ten questions. SRS is interpreted as 9-10 (very high), 6-8 (high), 3-5 (medium), and 0-3 (low).

The density of OSS in proximity to each source was retrieved from a 100x100 m grid of the entire Thiaroye aquifer previously devised by Diaw *et al.* (2020). The grid was produced using both object-oriented classification and photo-interpretation or visual interpretation of high-resolution Quickbird satellite images, which were validated by ground-truthing surveys. Population density was retrieved for the district of each source from the National Census of Senegal in 2016 (ANSD, 2016).

In-situ analysis

TLF and HLF were determined on unfiltered samples using portable UviLux fluorimeters (Section 2.3). Both TLF and HLF fluorimeters expressed intensity using in-built factory calibrations in quinine sulphate units (QSU) allowing calculation of TLF:HLF ratios. TLF data herein are reported in ppb tryptophan, for comparison to previous work, through calibration in standards of laboratory grade L-tryptophan (Acros Organics, USA) dissolved in ultrapure water (ppb TLF = 2.4461 QSU TLF -1.5086, r^2 1.00). Fluorescence analysis was conducted in a HDPE beaker placed within a covered black container to prevent interference from sunlight. Specific electric conductivity (SEC), pH, temperature and turbidity were quantified using a multi-parameter Manta-2 sonde (Eureka Water Probes, USA).

Laboratory analysis

Thermotolerant coliform (TTC) samples were collected in sterile 250 mL polypropylene bottles and stored in a cool box (up to 8 h) before analysis (Section 2.2). Samples for total (planktonic) bacterial cells (TBCs) were collected in 4.5 mL polypropylene cryovials (Starlab, UK) that were pre-treated with the preservative glutaraldehyde and the surfactant Pluronic F68 (Marie et al., 2014) at final concentrations of 1% and 0.01%, respectively. The samples were frozen at -18^{o}C within 8 h of collection, defrosted overnight during transit to the UK, and analysed the following morning on a BD Accuri C6 flow cytometer equipped with a 488 nm solid state laser (Becton Dickinson UK Ltd., UK). Water samples (500 mL) were stained with a 1:50 v/v solution of SYBR Green I (Sigma-Aldrich, UK) to a final concentration of 1:10,000 v/v for 20 min in the dark at room temperature. Samples were run with the Accuri at a slow flow rate (14 mL/min, 10 mm core) for 5 min and a detection threshold of

1500 on channel FL1. A single manually drawn gate was created to discriminate bacterial cells from particulate background, and cells per mL were calculated using the total cell count in 5 min divided by the reported volume run in µL.

Samples for major anions were filtered through 0.45 µm cellulose nitrate membranes (GE Whatman®, UK) into 30 mL Nalgene bottles (Thermo Fisher Scientific, USA) in the field. Analysis for chloride, nitrite, nitrate and sulphate was conducted by ion chromatography (Dionex AS50, Thermo Fisher Scientific, USA). Samples for dissolved organic carbon (DOC) were filtered through 0.22 µm hydrophilic polyvinylidene fluoride membranes (Sterivex, Merck KGaQ, Germany) into 15 mL polypropylene centrifuge tubes (Merck KGaA, Germany). DOC was quantified by thermal oxidation using an Elementar Vario Cube (Elementar Analysensysteme GmbH, Germany). Only a subset of 22 samples was collected and analysed for DOC due to budget constraints. All hydrochemical samples were refrigerated at 4^{o}C between collection and analysis.

7.2.3 Statistical analyses

Linear regression and correlation analyses were performed in R version 3.4.0 using base commands unless described otherwise. A forward stepwise linear regression algorithm was implemented (*train,* car package) using 10-fold cross validation. The algorithm adds one predictor to the model at a time according to whichever predictor will yield the largest decrease in the root mean square error (RMSE), until no further improvement can be achieved. Once an optimal set of predictors was retrieved, standardised beta coefficients (β) were estimated using *lm.beta* (lm.beta package) to allow quantitative comparisons between predictors. This command multiplies each unstandardized coefficient by the standard deviation of the associated predictor over the standard deviation of the dependent variable, hence a β refers to how many standard deviations a dependent variable changes, per standard deviation increase in the predictor.

The normality of model residuals were evaluated using Q-Q plots (*qqPlot*, car package), and Shapiro-Wilk's tests that employ the null hypothesis that the population is of Gaussian distribution (Royston,

1982). Multicollinearity (i.e. correlation between predictors) within multi-predictor models was investigated by calculating variance inflation factors (VIF) (*vif*, car package) (Alin, 2010). Prior to linear regression modelling, all variables with a skewness (*skewness*, moments package) greater than one were natural log transformed. In the case of TTCs and NO_2^-, additions of 1 and 0.1, respectively, were made prior to transformation to ensure the logarithm could be defined. Transforming the data was necessary as initial modelling produced skewed residuals, in contrary to the assumptions required by the model. Histograms of all variables used in the linear regression modelling are presented in Figure B-1.

Correlation matrices for variables of interest were produced by *rcorr* (Hmisc package) using Spearman's Rank incorporating mid-ranks in the case of ties (Myles and Wolfe, 1973), due to the non-Gaussian distribution of some data. These matrices were displayed using *corrplot* (corrplot package) and ordered by hierarchical clustering (Friendly, 2002; Murdoch and Chow, 1996).

7.3 Results

7.3.1 Faecal contamination in the Thiaroye aquifer

Water quality and sanitary risks

Groundwater beneath Dakar displays clear evidence of faecal contamination (Figure 7-2). TTCs are present in 80% of samples with a median count of 115 cfu/100mL. The median NO_3^- concentration (257 mg/L) is five times the WHO drinking water guideline value and exceeds this value in 90% of samples. A subset of sources also shows that the concentration of dissolved organic carbon is elevated (median 6.8 mg/L). Median TLF (70 ppb) and HLF (86.9 QSU) are similarly high. There is a large variation in TLF:HLF from 0.11 and 0.61 (5-95th percentiles); there is no clear TLF peak as the ratio is <1, with the exception of a single site. Median water temperature was 28.0^{o}C with variations limited to 2.2^{o}C (2σ) indicating minimal influence upon fluorescence data (Baker, 2005; Khamis et al., 2015). Turbidity is also likely to have limited influence on TLF/HLF with a median of 1.5 NTU and 2σ of 20.9 NTU (Khamis et al., 2015; Saraceno et al., 2017).

Sanitary risk scores show that these groundwater sources are at moderate risk of contamination from the surface (mean 4.7). In terms of potential sources of faecal contamination within 10 m, only one site had on-site sanitation (OSS), 64% had animal faeces, and 50% had trash.

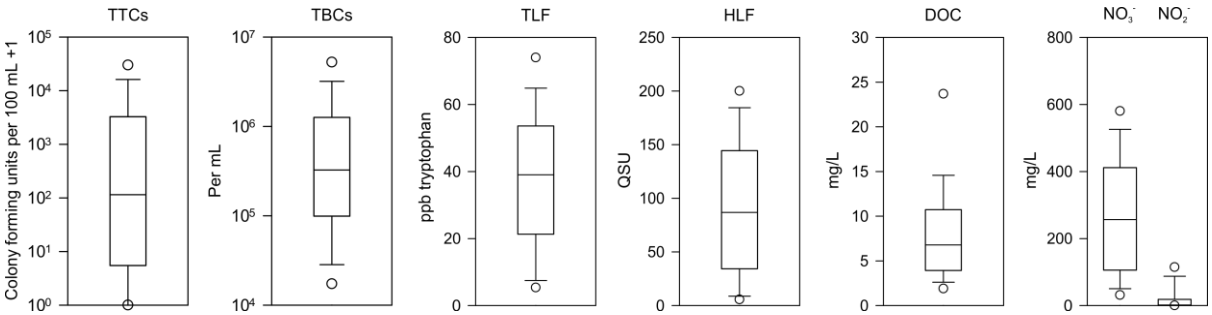

Figure 7-2 Boxplots of TTCs, TBCs, TLF, HLF, DOC, NO3- and NO2-. Box boundaries illustrate the 25th and 75th percentiles, the line within the box is the median, the whiskers are the 10th and 90th percentiles, and the circles are the 5th and 95th percentiles (n=84).

Table 7-1 Estimated coefficients and p-values for single predictor linear regression models of ln(TTCs+1). Predictors are ordered by coefficient.

Predictor	Adj. beta	p-value
ln(turbidity)	0.276	0.011
ln(NO_2^-+0.1)	0.224	0.041
ln(SEC)	0.211	0.054
OSS density	-0.182	0.097
ln(Cl^-)	0.173	0.115
Latrines	-0.153	0.165
Population density	-0.142	0.197
NO_3^-	0.102	0.356
ln(TBCs)	0.073	0.511
Animal faeces	0.070	0.526
SRS	-0.057	0.608
Trash	0.042	0.707
HLF	0.037	0.737
ln(TLF+1)	0.029	0.793
Temperature	-0.004	0.971
SO_4^{2-}	0.004	0.974

Predicting thermotolerant coliform counts

There is no relationship between fluorescent OM fractions and TTCs (Table 7-1). Significant linear regression models cannot be developed using either TLF (p-value 0.793) or HLF (p-value 0.737) as predictors of TTC counts. TLF is far in excess of the 1.3 ppb threshold proposed to indicate faecal

contamination and typically (>95%) exceeds the 6.9 ppb threshold to classify high risk sources ≥ 100 cfu/100 mL TTCs (Figure 7-2, thresholds from Chapter 4). TTC counts are, in contrast, comparably low (Figure 7-2) and absent in 20% of the samples.

The only significant predictors of TTCs are turbidity (p-value 0.011) and NO_2^- (p-value 0.041) though relationships are very weak (r^2 0.07 and 0.04, respectively). Potential explanatory variables relating to common sources of TTCs observed at the surface including OSS within 10 m, density of OSS, the presence of either animal faeces or trash are not significant (Table 7-1). The lack of relationships between TTCs and TBCs is also notable.

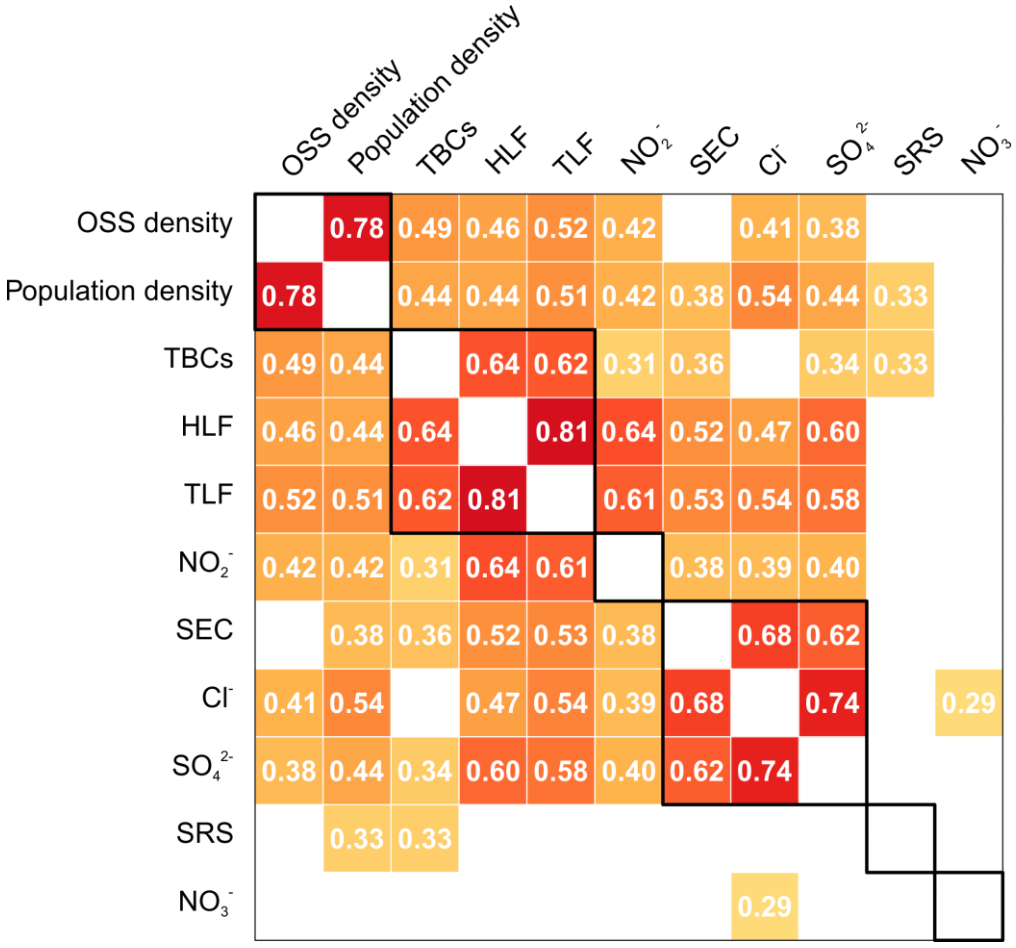

Figure 7-3 Correlation matrix of TLF and all significant single-predictors of TLF. Only significant (p-value < 0.01) Spearman's ρ_s are shown. Variables are ordered by hierarchical clustering and black squares enclose six clusters. If the number of clusters is increased to seven, SEC forms a cluster of its own.

Table 7-2 Estimated coefficients and p-values for single predictor linear regression models of ln(TLF). Predictors are ordered by coefficient.

Predictor	Adj. beta	p-value
HLF	0.767	<0.001
ln(TBCs)	0.647	<0.001
OSS density	0.553	<0.001
ln(SEC)	0.533	<0.001
$ln(NO_2^-+0.1)$	0.531	<0.001
SO_4^{2-}	0.521	<0.001
Population density	0.430	<0.001
$ln(Cl^-)$	0.423	<0.001
NO_3^-	0.249	0.022
SRS	0.241	0.027
Animal faeces	0.209	0.057
Temperature	-0.207	0.059
Trash	0.104	0.346
ln(turbidity)	-0.049	0.657
ln(TTCs+1)	0.029	0.793
pH	0.011	0.920
Latrines	0.010	0.928

Predicting tryptophan-like fluorescence

Significant single-predictor linear regression models can be estimated using either: HLF, TBCs, OSS density, SEC, SO_4^{2-}, population density, Cl^-, or NO_2^- with a p-value of <0.001, or NO_3^- or SRS with a p-value of <0.05 (Table 7-2). The superior model uses HLF as the predictor and has an r^2 of 0.58, with the next best models using TBCs and OSS density obtaining an r^2 of 0.41 and 0.30, respectively. Many of these ten TLF predictors are significantly correlated to each other with varying strengths (ρ_s 0.29-0.78, Figure 7-3), although they can be split into six hierarchical clusters: HLF and TBCs (alongside TLF); OSS and population densities; SO_4^{2-}, SEC, and Cl^-; NO_2^-; SRS; and NO_3^- (Figure 7-3).

Table 7-3 Stepwise linear regression model RMSE and r^2 following the addition of each predictor with ln(TLF) as the dependent variable

Predictor	RMSE	r2
Model 1		
HLF	0.52	0.58
NO_3^-	0.48	0.64
ln(TBCs)	0.42	0.72
OSS density	0.40	0.74
Model 2		
ln(TBCs)	0.61	0.41
NO_3^-	0.52	0.56
$ln(NO_2^-+0.1)$	0.48	0.63
OSS density	0.46	0.66
SO_4^{2-}	0.44	0.68

Implementing the stepwise linear regression algorithm, the model incorporates HLF, followed by the addition of NO_3^-, then TBCs and, finally, sanitation density (p-value <0.001) (Equation 1). HLF is the most important predictor in the model: being added first and having the greatest β. The additional predictors improve both the RMSE and r^2 to final values of 0.40 and 0.74, respectively (Table 3, Figure 7-4A).

$$\ln(TLF) = 0.480\ HLF + 0.335\ NO_3^- + \ 0.306 \ln(TBCs) + 0.196\ OSS\ density + 0.401$$

(Equation 1)

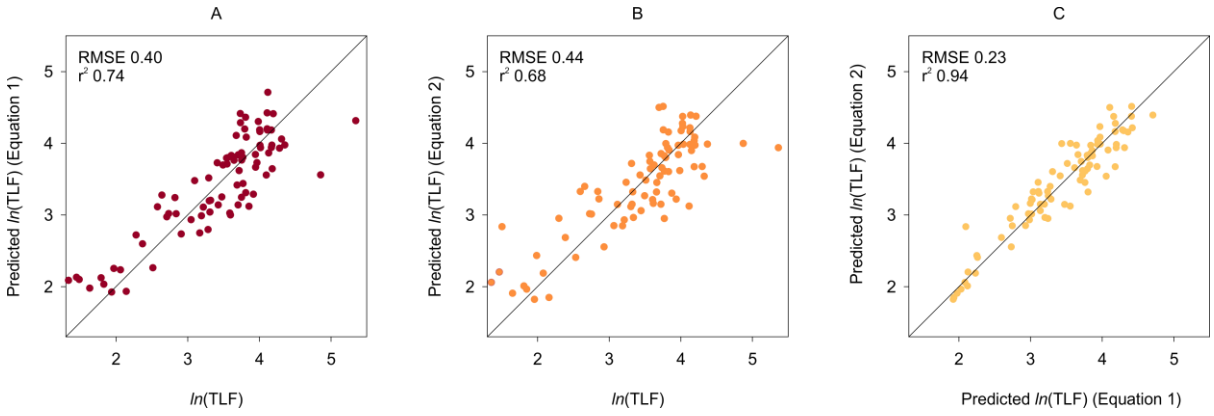

Figure 7-4 Scatterplots of observed against predicted ln(TLF) using (A) equation 1 and (B) equation 2, and (C) a scatterplot comparison of predicted ln(TLF) from both equations. A 1:1 line is shown in all plots.

The stepwise algorithm was also applied excluding HLF as a predictor. TBCs is the most important predictor and included first, followed by NO_3^-, OSS density, NO_2^- and SO_4^{2-} (p-value <0.001) (Equation 2). The final model has an r^2 of 0.68 and an RMSE of 0.44 (Table 3, Figure 7-4B). Equation 2 predictions are similar to Equation 1 at lower TLF with increasing differences between predictions as TLF increases (Figure 7-4C).

$$\ln(TLF) = 0.489 \ln(TBCs) + \ 0.304\ NO_3^- + 0.205\ \ln\ (NO_2^- + 0.1)\ + 0.187\ OSS\ density + 0.176\ SO_4^{-2} - 0.376$$

(Equation 2)

Both multiple linear regression models have Gaussian residuals (Figure B-2, Shapiro-Wilk, p-values = 0.103-0.330), no systematic spatial patterns in model residuals (Figure B-3), and no evidence of multicollinearity (VIF <2.06, Table B-1). The models are robust to the two highest TLF values; if we consider them outliers, identical predictors enter in the same order producing marginal improvements in the RMSE and r^2 (Equations B1 and B2). Furthermore, both models are also significant (p-value < 0.001) for the two main types of groundwater source: hand-pumped boreholes and dug wells, individually, where the models have an r^2 of between 0.67 and 0.87.

Relationships within DOC

Significant predictors of *ln*(TLF) are generally also predictors of *ln*(DOC) for the subset of 22 sources (Table 7-4). The optimal single-predictor models are HLF (r^2 0.84) and *ln*(TLF) (r^2 0.71). The predictors and TLF cluster identically to the complete dataset with similar correlations between them (Figure 7-5). DOC clusters with TLF, HLF and TBCs and there is a very strong or strong tendency for DOC to increase with each of these variables (Figure 7-5). TLF and HLF remain strongly associated with TBCs (ρ_s 0.67-0.68, p-value <0.001) in this subset of sources. No significant relationships exist between either DOC, NO_3^- or OSS/population densities.

Table 7-4 Estimated coefficients and p-values for single predictor linear regression models of ln(DOC). Predictors are ordered by coefficient.

Predictor	Adj. beta	p-value
HLF	0.921	<0.001
ln(TLF)	0.853	<0.001
ln(Cl^-)	0.685	<0.001
ln(TBCs)	0.650	0.001
SO_4^{2-}	0.615	0.002
ln(SEC)	0.553	0.008
ln(NO_2^-+0.1)	0.551	0.008
Trash	-0.479	0.024
NO_3^-	0.234	0.294
Population Density	0.212	0.344
Temperature	-0.199	0.375
Latrines	0.163	0.469
SRS	-0.127	0.575
Cattle	0.118	0.600
pH	0.097	0.668
ln(TTC+1)	-0.091	0.688
OSS Density	0.061	0.788
ln(turbidity)	-0.057	0.800

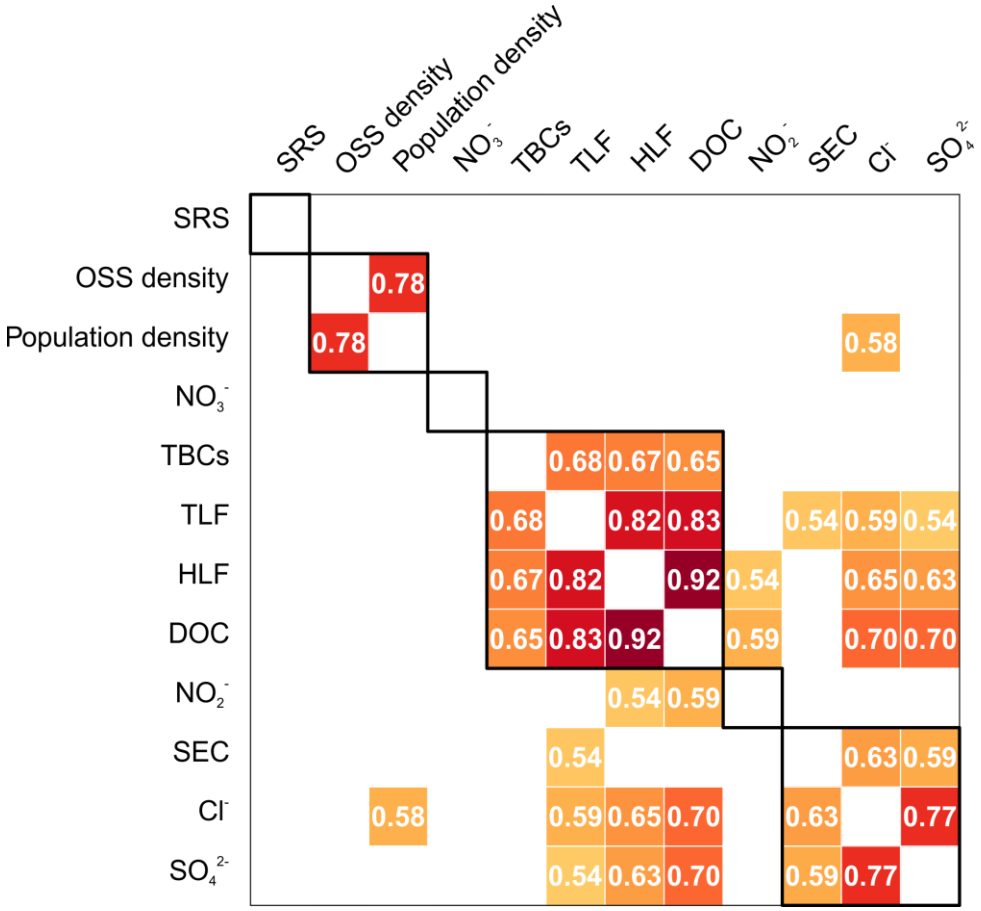

Figure 7-5 Correlation matrix of DOC, TLF, and all significant single-predictors of TLF identified in Table 7-2 for the subset of 22 groundwater sources. Only significant (p-value < 0.01) Spearman's ρ_s are shown. Variables are ordered by hierarchical clustering and black squares enclose six selected clusters.

The stepwise algorithm was employed using *ln*(DOC) as the dependent variable. HLF enters the model first, followed by TLF, which improves performance at lower concentrations (<2), and finally OSS density (Figure 7-6), with the final model (Equation 5) having an RMSE and r^2 of 0.19 and 0.90, respectively. When *ln*(TLF) was designated the dependent variable, the optimal model includes only *ln*(DOC) (Equation 4), with an RMSE and r^2 of 0.28 and 0.71, respectively. The unstandardized model coefficient is 0.712.

$$\ln(DOC) = 0.686\,\mathrm{HLF} + 0.344\ln(TLF) - 0.171\,OSS\ density - 0.202 \quad \text{(Equation 3)}$$

$$\ln(TLF) = 0.853\ln(DOC) + 2.211 \quad \text{(Equation 4)}$$

Both multiple linear regression models (Equations 3 and 4) have Gaussian residuals (Figure B-2, Shapiro-Wilk, p-values = 0.189-0.642) and Equation 3 has no evidence of multicollinearity (VIF <2.76, Table B-1).

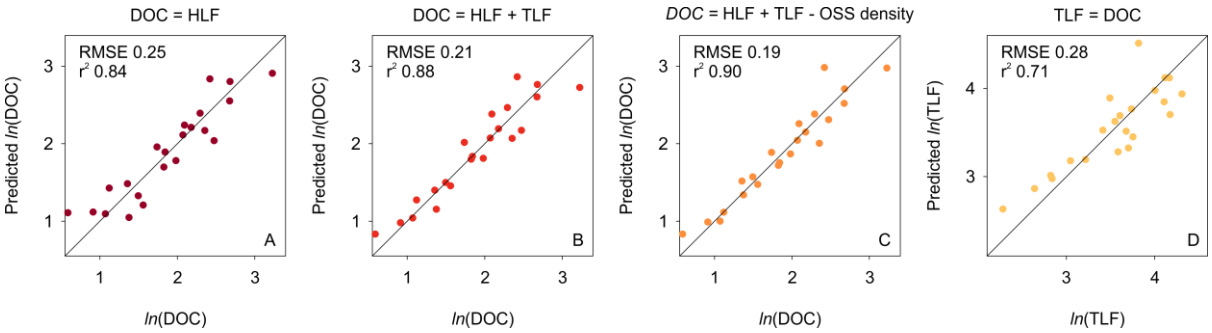

Figure 7-6 Scatterplots of observed against predicted ln(DOC) using (A) HLF; (B) HLF and TLF; (C) HLF, TLF and OSS density (Equation 3); and (D) observed against predicted ln(TLF) using ln(DOC) (Equation 4). A 1:1 line is shown in all plots.

7.4 Discussion

7.4.1 Fluorescent OM as an in-situ indicator of TTCs

This is the first groundwater study to demonstrate no relationship between TLF/HLF and TTCs. We consider several potential explanations for this that relate to the environment of the Thiaroye aquifer, timing of the study, and uniqueness of TLF to TTCs. Firstly, the Thiaroye aquifer has been subject to high loading of faecal waste from a dense network of OSS facilities since settlement began in the 1970s. In intergranular aquifers such as this, groundwater movement is relatively slow and porosity high, facilitating the accumulation of pollutants over time if they do not breakdown as rapidly as they arrive. We estimate a groundwater velocity of around 30 to 200 m/year, here, indicating a flow path from outcrop in the southeast to the Atlantic Ocean in the North of up to 50 to 120 years (assuming a hydraulic conductivity of 10-60 m/d (Henry, 1972), hydraulic gradient of 0.003 (Faye et al., 2019), and effective porosity of 0.30 (Diedhiou, 2011)). Consequently, median NO_3^- is five times the WHO drinking water quality guideline value, median DOC is almost six times the global median (McDonough et al., 2020), and median TLF/HLF are at least an order of magnitude greater than previous groundwater studies (Chapters 3 and 5; Sorensen et al., 2016). In terms of TLF, the groundwater resembles poor quality surface waters in South Africa (Baker et al., 2015). Against this elevated

historical baseline of fluorescent OM that is potentially spatially heterogeneous across the city, detecting deviations in recent faecal contamination, determined by TTCs, may not be possible. Additionally, there are also likely to be fluorophores unrelated to faecal contamination in this complex urban environment, which fluoresce at either TLF or HLF wavelengths such as proteins in solid organic waste and xenobiotic compounds (Baker and Curry, 2004; Muller et al., 2011).

The fine- and medium- grained sands of the aquifer are also likely to be effective at straining bacteria during vertical and lateral groundwater flow whilst predominantly extracellular TLF and HLF fluorophores are expected to be transported more readily. Indeed, we have demonstrated that the majority of TLF (~97%) was extracellular and there was no evidence of intracellular HLF in this aquifer (Chapter 6). There are no specific studies investigating bacterial transport in this aquifer, though Weaver *et al*. (2013) showed bacterial transport was limited to only a few metres in moderate and coarse sandy aquifers.

Our study was conducted outside of the recharge season following nine months of no rainfall, which is often a key driver of the microbiological contamination of groundwater (Hynds et al., 2012; Worthington and Smart, 2017). Furthermore, OSS, which could provide a perennial source of recharging water contaminated with TTCs, was typically absent in close proximity to water sources. Consequently, any TTCs present during the recharge season may have perished or become non-culturable while elevated fluorescent OM remained present. This seasonal pattern was previously reported in Chapter 3 where we demonstrated that at some groundwater sources, TLF remained perennially elevated whereas TTCs were only elevated during the recharge season. Future sampling during the recharge season could examine whether seasonal TLF/HLF-TTC relationships exist.

Finally, fluorescence at TLF and HLF wavelengths is not unique to TTCs such as *E. coli* (Bronk and Reinisch, 1993; Dalterio et al., 1986; Dalterio et al., 1987; Dartnell et al., 2013; Fox et al., 2017; Seaver et al., 1998; Sohn et al., 2009). Here, we demonstrate, despite the lack of TLF/HLF-TTC relationship,

that TBCs are the best predictor of TLF and are strongly correlated to HLF. Hence, non-TTC bacteria may be the source of TLF/HLF.

7.4.2 Fluorescent OM as an in-situ indicator of total bacterial cells

Relationships between TLF/HLF and TBCs may indicate that we are fluorescing autochthonous compounds produced by bacteria, given the associated fluorophores are predominantly extracellular (Chapter 6). Therefore, we are enumerating cells indirectly because of their activity, which is consistent with surface water (Cammack et al., 2004; Hudson et al., 2008; Parlanti et al., 2000) and laboratory studies (Fox et al., 2019; Fox et al., 2017) linking fluorescent OM and microbial activity. Despite this, abundance and activity are likely to be broadly interlinked in environmental systems.

Alternatively, we can consider that fluorescent OM is allochthonous and derives mainly from anthropogenic activity at the surface, in addition to a subordinate baseline relating to naturally occurring fluorescent OM. The statistical relationship between TLF/HLF and OSS density supports the theory that septic tanks are a key source of allochthonous OM here. This anthropogenic fluorescent OM is associated with nutrients (C and N) that provide resources to the bacterial community in a habitat, which typically has low-availability of organic carbon and nutrients (Griebler and Lueders, 2009). It then follows that the greater the OM supply to the system, the higher the bacterial biomass that can be supported. This hypothesis has been confirmed in multiple groundwater studies that have specifically related DOC inputs from anthropogenic sources to increases in bacterial abundance and activity (Findlay et al., 1993; Foulquier et al., 2011; Smith et al., 2015; Sobczak and Findlay, 2002). Therefore, the demonstrated relationship may not be a result of bacteria producing TLF/HLF fluorophores *in-situ*, but rather fluorophores being associated with OM inputs from the surface that provide a microbial resource. The abundance of this resource is then quantitatively related to the abundance of bacteria.

A final explanation may be the bacterial population is transported alongside fluorescent OM from the same source at the surface because many sources, particularly those that are faecal, contain elevated

fluorescent OM and bacteria. We consider this to be the least plausible explanation as bacterial transport through the fine- and medium- grained sand aquifer is likely to be limited. Hence, the bacterial population that relates to TLF/HLF is likely to be a subsurface community that develops *in-situ*. Furthermore, the study was conducted nine months after the termination of the recharge season, which is likely to provide the greatest influx of organic matter and bacteria; any foreign bacteria from that point in time are expected to have perished.

Future microbiology studies could use high throughout DNA sequencing to reveal the community structure and function of the groundwater communities to indicate the primary source of these communities. For example the Bayesian tool SourceTracker (Knights et al., 2011) has been used to identify sources of faecal microbes in a variety of freshwater habitats (Baral et al., 2018; Brown et al., 2017; Henry et al., 2016). In addition, it is possible to use flow cytometry to distinguish what proportion of the community is live versus dead (Berney et al., 2007) and the extent of microbial activity (Léonard et al., 2016). When combined, such analyses could disentangle the relationships between total bacterial cells and TLF/HLF by elucidating the origins of the bacterial community and whether the fluorescent OM is likely to be autochthonous or allochthonous.

7.4.3 Fluorescent OM as an in-situ indicator of DOC

This is the first groundwater study demonstrating a strong relationship between *in-situ* fluorescent OM and laboratory DOC data. Previous *in-situ* fluorescence research has demonstrated similarly strong relationships (r^2 0.80-1.0) in surface waters (Downing et al., 2012; Khamis et al., 2017; Lee et al., 2015; Snyder et al., 2018; Tunaley et al., 2016). In these studies, researchers were deploying fluorescent sensors targeting the excitation peak of HLF around 350 nm and, with the exception of Khamis *et al*. (2017), did not investigate DOC-TLF relationships. Our analysis mirrors the linear regression modelling of Khamis *et al*. (2017): HLF is a slightly better predictor of DOC but the model is marginally improved by the addition of TLF. The inclusion of both HLF and TLF suggests that each wavelength pair results in the fluorescence of slightly different components of the available DOC.

7.5 Conclusions

Fluorescent OM is significantly associated with parameters linked to faecal contamination that include on-site sanitation density (OSS), NO_2^-/NO_3^-, total bacterial cells (TBCs), and dissolved organic carbon (DOC); but not with thermotolerant coliforms (TTCs), which tend to be indicative of recent contamination. It is unclear whether the lack of TTC association is because of: the intergranular nature of the aquifer, sampling being conducted nine months after the cessation of the recharge season, or the non-uniqueness of tryptophan-like (TLF) and humic-like fluorescence (HLF) to TTCs. *In-situ* fluorescence spectroscopy instantly indicates a drinking water source is impacted by faecal contamination, although it is unclear how that may relate to microbial risk in this setting.

TBCs is the most related environmental parameter to both TLF and HLF across all samples. We consider this relationship is a result of either: (1) OM and its embedded nutrients providing a resource for bacteria in the subsurface; or (2) the *in-situ* production of fluorescent OM by subsurface bacteria. Irrespective, the relationships support the case of fluorescence spectroscopy as a more practical, cheaper, and robust alternative for indicating TBCs than quantification by flow cytometry. In remote settings, there is the additional advantage that samples can be measured *in-situ* rather than being preserved for subsequent laboratory analysis using hazardous chemicals (e.g. formaldehyde, glutaraldehyde, ethanol), that can modify the sample and are not always effective (Kamiya et al., 2007).

This groundwater study demonstrates strong relationships between *in-situ* fluorescent OM and DOC. HLF is the most strongly correlated peak to DOC and is likely to be sufficient in providing high-resolution data that could provide novel insights into DOC and C fluxes within aquifers – an ecosystem that is generally C limited. Furthermore, online fluorimeters could facilitate more cost-effective chlorination of potable groundwater and forewarn of the generation of harmful disinfection by-products within the water industry.

Chapter 8 - Temporal relationships between fluorescence spectroscopy and bacteria

8.1 Introduction

In this chapter, we evaluate the utility of both TLF and HLF as instantaneous, in-situ indicators of faecal contamination risk in groundwater, the world's largest store of freshwater and the primary source of drinking water for up to two billion people (Gleeson et al. 2010). We repeatedly sampled 40 groundwater sources in a community in Uganda across a period of fourteen months for TTCs and alternative rapid approaches that could be used to indicate faecal contamination. The rapid indicators included standard approaches of sanitary inspections, turbidity and electrical conductivity, alongside the more novel indicators of in-situ fluorescence spectroscopy and total (planktonic) bacterial cells (TBCs) by flow cytometry. We aim to demonstrate: (1) in-situ TLF and HLF are the superior rapid indicators of TTCs; and (2) the seasonal nature of the associations among TLF, HLF, TTCs, and TBCs.

This chapter is published as:

Sorensen, J.P.R.; Nayebare, J.; Carr, A.F.; Lyness, R.; Campos, L.C.; Ciric, L.; Goodall, T.; Kulabako, R.; Rushworth Curran, C.M.; MacDonald, A.M.; Owor, M.; Read, D.S.; Taylor, R.G.. 2021. In-situ fluorescence spectroscopy is a more rapid and resilient indicator of faecal contamination risk in drinking water than faecal indicator organisms. Water Research, 206, 117734. 11.

8.2 Methods

8.2.1 Study area

Lukaya is a town in central Uganda around 100 km southwest of the capital city Kampala and close to the shores of Lake Victoria (Figure 8-1A). The town's population was 24,000 in the last census (UBOS, 2014), with a density of *c.* 640 inhabitants per km^2 within the built-up area that is growing at 3% per year (PDP, 2017). To the east of town is the Lweera Swamp where commercial rice farming is practised (Figure 8-1B). The climate is humid with a mean annual rainfall of 890 mm (Nayebare et al., 2020) that

is bimodal and focussed within the rainy seasons of March to May and September to November (Figure 8-1C).

The town predominantly sits on Precambrian basement rocks with aquifers developed within the weathered overburden and fractured bedrock with a shallow water table between 0.5 and 9 m below ground level (bgl). Groundwater is the primary source of water for the town, with the majority obtained from hand pumped wells and springs. Piped water is used by <1% of households (Nayebare, 2021), which is obtained from a borehole operated by the National Water Sewerage Corporation (NWSC) to the south of the town (Figure 8-1B) (Nayebare et al., 2020).

The town possesses neither a sewer network nor a wastewater treatment facility. On-site sanitation facilities number around 2,100 and predominantly comprise partially lined pit latrines that are elevated because of the shallow water table. The pits are not emptied: when full, faecal matter is moved from one pit to another, or a new pit is dug. Many pits also have overflow outlets in case of inundation during the rainy seasons (Nayebare et al., 2020).

8.2.2 Hydrological monitoring

Tipping bucket rainfall gauges, Lambrecht meteo model 15189 (Lambrecht meteo GmbH, Germany), were installed in two locations (1 and 2, Figure 8-1B) and data were aggregated to daily sums. A rainfall timeseries for the study period was produced using the data from station 1, in the centre of Lukaya, unless records were absent or failed quality checks, in which case data were replaced using records from station 2. Groundwater levels in the weathered overburden were monitored in three boreholes screened at the following depth intervals: 10.2-16.1, 11.4-17.3, and 23.5-29.4 m bgl. All boreholes are located within 20 m of each other at the surface (Figure 8-1B). Levels were monitored using Rugged TROLL 100 data loggers (In-Situ, USA).

8.2.3 Water sampling and analysis

Water sources and sampling rounds

An exhaustive survey of all water sources in the town was previously undertaken by Nayebare *et al.* (2020) and identified 56 shallow hand-dug wells equipped with hand pumps (shallow), 4 boreholes (deep) and 7 unprotected springs. The shallow sources vary between 3 and 8 m depth, and boreholes are at least 30 m deep, including the NWSC water supply well drilled to 61 m bgl (Nayebare et al., 2020). All shallow and deep sources are protected and considered *improved* water sources, and the springs are *unimproved* (WHO, 2017c).

A stratified sampling approach was implemented to sample 40 of the water sources (Figure 8-1B). The selected sources included: three deep sources, with the fourth having no accessible sampling location, and five springs, with the other two springs being such gentle seepages that groundwater inputs were not visible and these springs were not heavily utilised by the community. Finally, 32 of the 56 shallow sources were selected to maximise the spatial spread of shallow sources across the town, whilst accounting for some sources which had become non-functional.

Sources were sampled in six rounds (R) across 14 months. R1 to 4 were undertaken in 2018 from late-April to late-May when monthly rainfall typically peaks (Figure 8-1C); each round was separated by six to nine days. R5 and 6 were undertaken in 2019 from mid- to late-June when monthly rainfall is close to its annual minimum and the rounds were separated by four days. In R5 and 6, four of the hand pumps on the shallow sources had become non-functional and only 36 sources were sampled. Note that rainfall had progressed through two wet and dry seasons between R4 and R5.

Water sampling and analysis

All shallow and deep water sources were unlocked and in frequent use by the community or owner throughout daylight hours. Nevertheless, all sources were allowed to flow for an additional minute before sampling to ensure all pipework was adequately flushed. All unprotected springs were sampled from the surface water channel as close to the point of groundwater discharge as possible.

Each source was sampled for a range of possible in-situ rapid indicators of faecal contamination. TLF and HLF were quantified using separate UviLux fluorimeters (Section 2.3). The TLF fluorimeter was calibrated using eight standards (0, 1, 2, 5, 10, 50, and 100 ppb) of L-tryptophan dissolved in ultrapure water. The factory calibration was implemented for the HLF sensor, which expresses intensity in quinine sulphate units (QSU). The TLF ppb dissolved tryptophan data can be converted to QSU by division by 2.5037 or 2.3696 in rounds 1-4 and 5-6, respectively, to allow calculation of TLF:HLF ratios.

Fluorescence spectroscopy measurements were taken by submerging the fluorimeter in 150 mL of groundwater contained in a polypropylene beaker. Each measurement was taken in the dark by placing the beaker and fluorimeter within a covered stainless steel container. Given the sensitivity of the fluorimeters, all measurements were taken in duplicate, or repeated further to obtain reproducible data. Field repeatability (σ) of TLF and HLF measurements were calculated as 0.4 ppb and 0.1 QSU, respectively, across all data in R5 and 6. Specific electrical conductivity (SEC), pH and temperature were monitored using a multi-parameter Manta-2 sonde (Eureka Waterprobes, USA). Turbidity was measured using a DR/890 portable colorimeter (HACH, USA), including blank correction with deionised water before each measurement, except during R4 when the Manta-2 was used. To account for absolute differences between the turbidimeters, the turbidity data were min-max normalised. Fluorescence data did not require linear correction for temperature quenching (Khamis et al., 2015), with only a range between 22.4-25.6^{o}C. The pH of the samples was 4.6-6.7, hence pH would not appreciably have impacted the fluorescence (Reynolds, 2003).

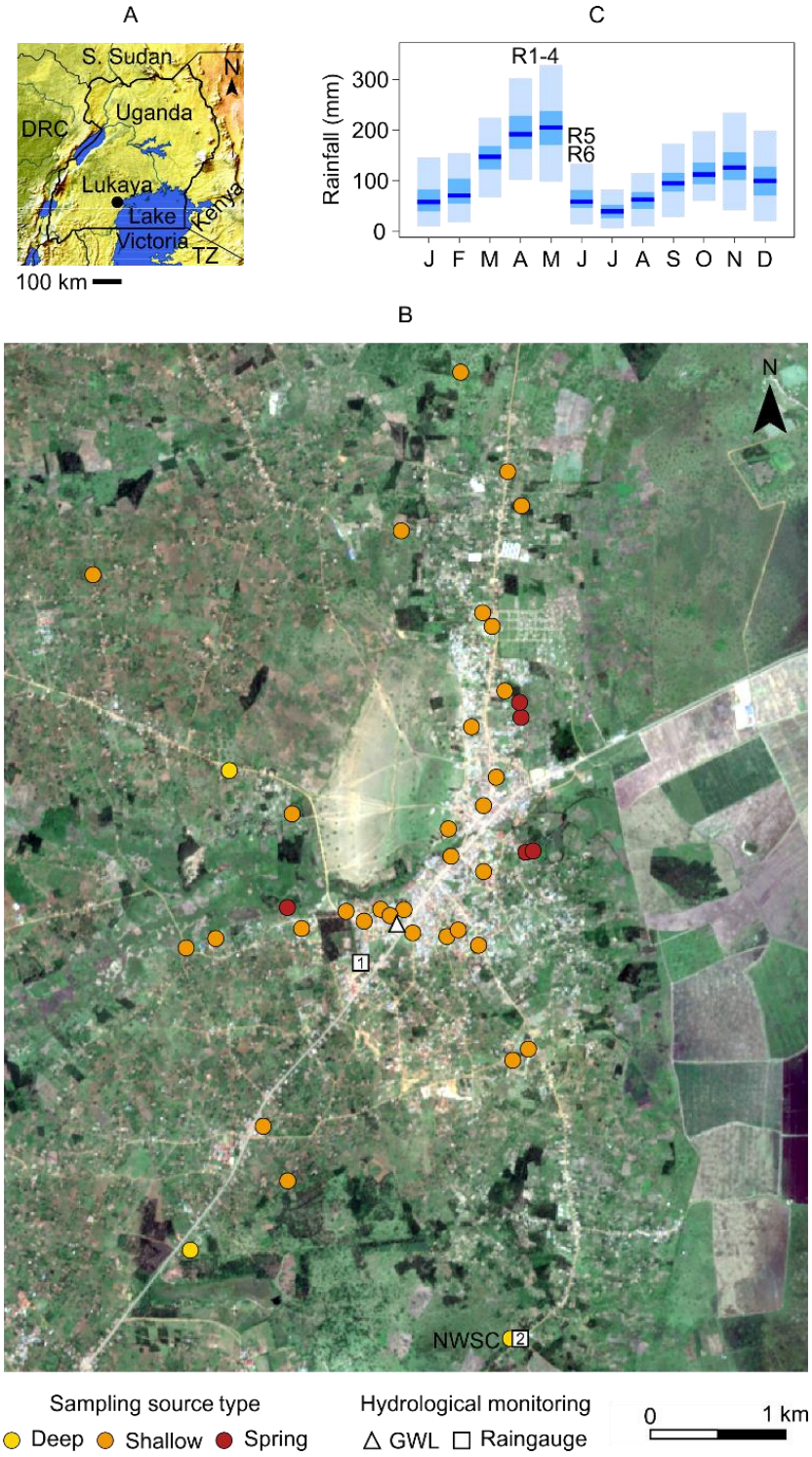

Figure 8-1 (A) Location of Lukaya within Uganda (TZ = Tanzania); (B) Sampling sources and hydrological monitoring in Lukaya mapped on Copernicus Sentinel data (2020), with the piped water source labelled NWSC; (C) Tukey boxplot, excluding outliers, of CRU monthly rainfall data (1900-2019) for grid cell 0.25 S, 31.75 E (Harris et al. 2020) indicating the timing of sampling rounds R1-6.

Sanitary risk inspections were undertaken at each source by the same assessor during sampling in R5 (WHO, 2020) (Table C-1). The surveys consisted of a list of nine yes-no questions: to identify sources of contamination observable at the surface, pathways for contaminants to the enter source, and breakdowns in barriers to contamination (Kelly et al., 2020). The questions differed for the shallow/deep sources and springs because of different potential pathways leading to contamination. The total number of positive responses to the questions equates to the sanitary risk score (SRS).

Flow cytometry analysis for total (planktonic) bacteria cells (TBCs) was conducted in the laboratory on preserved samples, but the analysis can also be undertaken rapidly and online at a water source (Safford and Bischel, 2019). Samples (2 mL) were collected in 4.5 mL polypropylene cryovials (STARLAB, UK) that were pre-loaded with the preservative glutaraldehyde (Sigma-Aldrich, UK) and the surfactant Pluronic F68 (Gibco, USA) (Marie et al., 2014) at final concentrations of 1% and 0.01%, respectively. The samples were kept in a cool box for up to 8 h, then frozen at -18^{o}C, defrosted overnight during transit to the UK in a cool box, and then analysed the following morning. Analysis was conducted using a BD Accuri C6 flow cytometer utilising a 488 nm solid state laser (Becton Dickinson UK Ltd., UK). Water samples (500 µL) were stained with a 1:50 v/v solution of SYBR Green I (Sigma-Aldrich, UK) to a final concentration of 1:10,000 v/v for 20 min in the dark at room temperature. Samples were run at a slow flow rate (14 mL/min, 10 mm core) for 5 min and a detection threshold of 1500 on channel FL1. A single manually drawn gate was created to discriminate bacterial cells from particulate background, and cells per mL were calculated using the total cell count in 5 min divided by the reported volume run in µL (Chapter 5).

Thermotolerant (faecal) coliforms (TTCs) were selected as the FIO of contamination (Section 2.2). TTC samples were collected in sterile 250 mL polypropylene bottles and stored in a cool box (up to 8 h) before analysis.

8.2.4 Statistical analysis and modelling

Rapid approaches to assess faecal contamination were tested against the benchmark FIO of TTCs using R v4.0.3 (R Core Team, 2020) and base commands unless otherwise stated. Logistic regression models were developed for each rapid approach as a predictor of ≥10 cfu/100 mL TTCs. There were insufficient data (n=1) where TTCs <1 cfu/100 mL to develop models for TTC presence-absence. Model performance was assessed using the area under the receive operating curve (AUC) (Mandrekar, 2010), which is a plot of the proportion of true positive results against the proportion of false positive results as the threshold of the predictor is varied. A perfect classifier has an AUC of 1 and a random classifier has a value of 0.5. Furthermore, we consider AUC values of 0.7 to 0.8, 0.8 to 0.9, and 0.9 and greater as acceptable, excellent, and outstanding, respectively (Hosmer Jr et al., 2013). Rank correlations between rapid approaches and TTCs were estimated using the non-parametric Spearman's rank (ρ_s) (Spearman, 1904), given the non-Gaussian distribution of many of the variables. Coefficients of 0.80-1.00, 0.60-0.79, 0.40-0.59, 0.20-0.39, 0.00-0.19 were considered very strong, strong, moderate, weak, and very weak, respectively.

Multiple linear regression was applied to investigate what combination of rapid approaches was optimal for the prediction of TTC enumeration. A forward stepwise algorithm was used using 10-fold cross validation within the R package *car* (Fox and Weisberg, 2018). One predictor is added to the model at a time to achieve the largest decrease in the root mean square error (RMSE), until no further reduction can be yielded. The normality of model residuals was evaluated using Q-Q plots. Initial models produced non-Gaussian residuals, in violation of the assumptions, so all variables with a skewness >1 were natural log transformed. An addition of 1 was made to TTCs to ensure the logarithm could be defined.

Differences in both rapid approaches and TTCs between sampling rounds were explored using the Friedman test in the R package *PMCMR* (Pohlert, 2014), with post-hoc Nemenyi tests (Demšar, 2006). The Friedman test is a non-parametric alternative to the repeated-measures ANOVA and tests the null

hypothesis that at least one group does not belong to the same population. If the Friedman test is significant ($p < 0.05$), the subsequent multiple comparison Nemenyi tests report significant ($p < 0.05$) differences between each pair of groups, if their corresponding mean ranks differ by at least the critical difference (Demšar, 2006; Pohlert, 2014).

The comparative stability of rapid approaches and TTCs were evaluated by cross-correlating each variable with itself between sampling rounds. Additionally, TLF and HLF were cross-correlated with TTCs and TBCs across the sampling rounds to explore the seasonal nature of any associations. Spearman's Rank was used because the variables were non-Gaussian and because we were most interested in the rank-order of the sources as an indicator of relative risk across the community.

Groundwater levels (GWLs) for the most complete record, BH ALP-3, were hindcasted by 24 days to contextualise groundwater conditions before and during R1 and 2 where GWL observations were not collected. Hindcasting was conducted using a forward model implementing the water table fluctuation method and assuming diffuse recharge from daily rainfall observations (Cuthbert et al., 2019). The model was parameterised using a linear rainfall-recharge relationship with a rainfall threshold of 10 mm, an exponential recession coefficient of 1.1×10^{-3} day^{-1} to a base of 1149 m asl, and a specific yield of 5%. The model effectively captures which rainfall events result in groundwater recharge, the timing of GWL responses, and the rate of recession, with an r^2 of 0.85 and RMSE of 0.23 m (Figure C-1).

8.3 Results

8.3.1 Widespread prevalence and high variability of TTCs

All sources show evidence of at least intermittent faecal contamination, inferred through the presence of TTCs (Figure 8-2A&B). Fifty percent of the sources have median TTCs of at least 88 cfu/100 mL (Figure 8-2A), with a range in median counts between <1 and 5,101 cfu/100 mL. The shallow sources cover the entire range in median counts, with median TTCs at springs and deep sources are in the upper 50% and lower 53% of all sources, respectively (Figure 8-2A).

TTCs vary widely at each source with 50% of sources having a range of at least 720 cfu/100 mL (Figure 8-2B). The range in TTCs at a source is at least 8 cfu/100 mL and up to 34,000 cfu/100 mL, with all but two sources varying between risk categories, based upon the order of magnitude of TTCs, previously defined by WHO (1997). A third of the sources transit between testing negative and positive for TTCs, including the only source with a median count of <1 cfu/100 mL (Figure 8-2D). There is a tendency for spring sources to have greater ranges in TTCs, than the other types of source (Figure 8-2B).

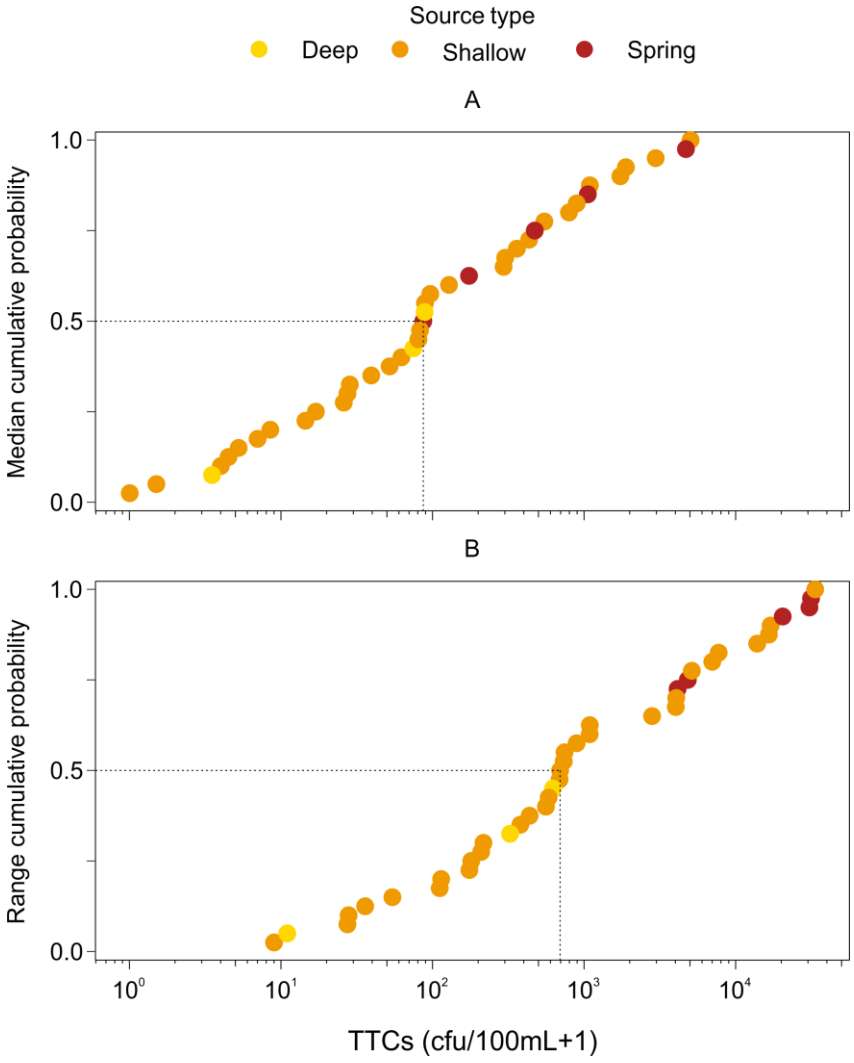

Figure 8-2 Figure 2 Empirical cumulative distribution functions of (A) median TTCs and (B) range in TTCs for each water sources (n = 40).

8.3.2 TLF and HLF are superior rapid approaches to indicate TTCs using source medians

Median TLF is the only significant predictor of median TTCs ≥10 cfu/100 mL according to logistic regression models (β = 1.09, p-value = 0.042) (Figure 8-3A; Table C-2). The AUC using TLF as a classifier

is 0.88, which is closest to the perfect classifier value of 1 than the random selector of 0.5 (Figure 8-3A) and considered “excellent”. An optimal TLF threshold of 2.2 ppb can be defined to classify TTCs ≥10 cfu/100 mL, with associated false-negative and false-positive rates of 16% and 25%, respectively (Figure 8-3C).

The AUC when classifying median TTCs ≥10 cfu/100 mL using HLF was 0.85 and considered “excellent”, with the logistic regression model being borderline significant (β = 1.90, p-value 0.059) (Figure 8-3A). A HLF threshold of 0.85 QSU can classify median TTCs of ≥10 cfu/100 mL with identical error rates to those as the proposed TLF thresholds. In fact, if the TLF ppb threshold is converted into QSU then they are almost equivalent to these HLF thresholds.

The AUC was “acceptable” for Sanitary risk scores (SRS) and median total bacterial cells (TBCs), and demonstrated that SEC and turbidity performed no better than a random classifier (Figure 8-3A). Considering only the shallow sources (n=32), SRS was a significant predictor (β = 0.90, p-value 0.038) with an “acceptable” AUC of 0.79. Only one individual sanitary inspection question, whether drainage was inadequate, was a significant predictor (p-value < 0.05) of median TTCs ≥10 cfu/100 mL for shallow and deep sources where sanitary inspection questions were identical. The AUC for inadequate drainage as a classifier was 0.75 and considered “acceptable” (Table C-3).

Median TLF is very strongly correlated with median TTCs (ρ_s 0.81, Figure 8-3D), being the most correlated rapid approach (Figure 8-3B). All types of water source follow the same rising trend (Figure 8-3D), with two notable outliers. One outlier is a shallow source, which has a median TLF of 65.9 ppb, more than three times the TLF intensity of any other source, although median TTCs are also high at 296 cfu/100 mL. The second outlier is a deep source with median TTCs of 89 cfu/100 mL, yet the lowest median TLF of 0.5 ppb, as well as being the only site with a zero SRS. Median HLF is similarly correlated with median TTCs (ρ_s 0.79) as TLF, with identical outliers. Median TBCs correlate moderately with median TTCs, but other in-situ indicators are only weakly related to median TTCs (Figure 8-3B).

No other rapid approaches provide additive performance to *ln*(TLF) for the prediction of median *ln*(TTCs) using the stepwise forward linear regression algorithm. The linear regression model has an r^2 of 0.51 and p-value <0.001 (Equation 1). Omitting *ln*(TLF), only *ln*(HLF) is included by the algorithm and the model has an r^2 of 0.48 and p-value <0.001 (Equation 2). Natural log transforms of TTCs, TLF, and HLF where required in the linear regression models to ensure the model residuals were Gaussian (see Figure C-2 for Q-Q plots).

$$\ln(TTC + 1) = 1.66 \ln TLF + 2.19 \quad \text{Equation 1}$$

$$\ln(TTC + 1) = 1.44 \ln HLF + 3.73 \quad \text{Equation 2}$$

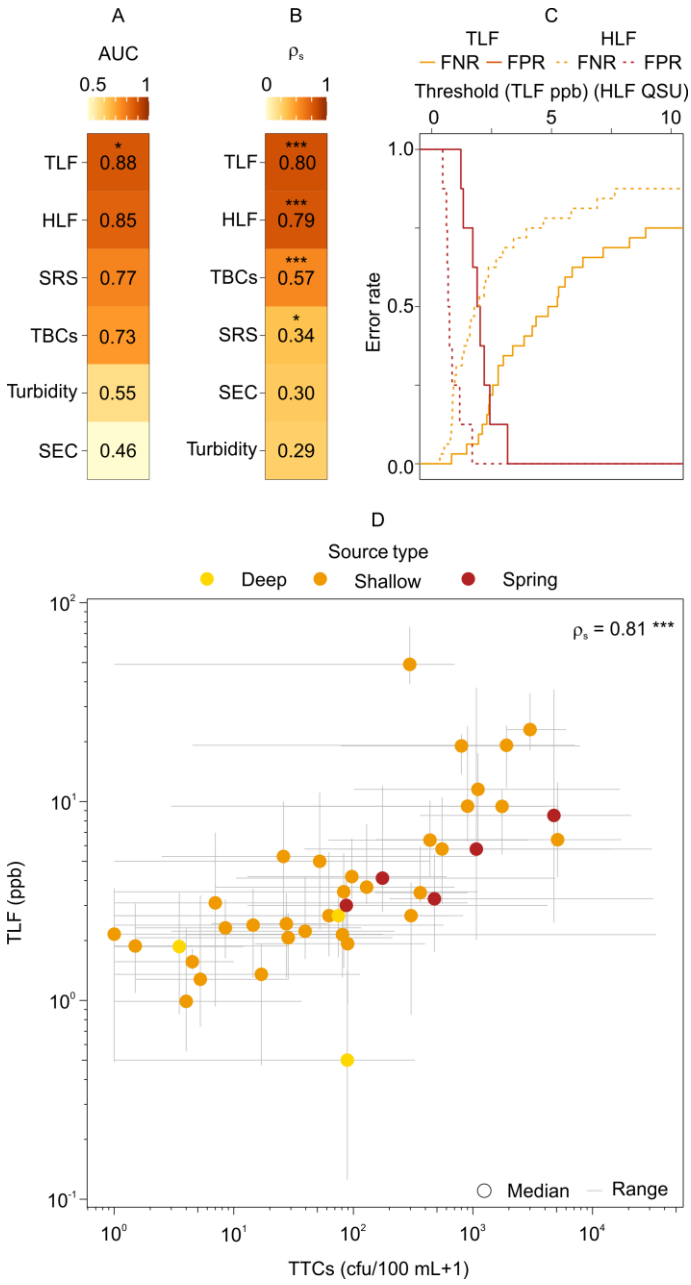

Figure 8-3 (A) Area under curve (AUC) and significance of the logistic regression models for each in-situ parameter as a classifier of TTCs ≥10 cfu/100 mL; (B) Spearman's Rank correlation coefficients and significance for TTCs and each in-situ parameter; (C) False-negative (FNR) and false-positive (FPR) rates for TLF thresholds (0-10 ppb) as classifiers of TTCs ≥10 cfu/100 mL and an optimal TLF threshold highlighted with a dotted line; (D) Scatterplot of median TTCs and TLF for each source, illustrating ranges in both variables. The median of both in-situ parameters and TTCs at each source (n = 40) are used in all statistics. p-values of <0.05, <0.01 and <0.001 are denoted by '*','', and '***', respectively.**

8.3.3 Relationships between in-situ approaches and TTCs by sampling round

TTCs were significantly different between the wet season rounds of R1, R2 and R4 and the dry season rounds of R5 and R6 (Figure 8-4A & B). Median TTCs were higher during than wet season (up to 382 cfu/100mL, R1) than the dry season (as low as 13 cfu/100mL, R5) (Figure 8-4B). TTCs rapidly reduced in the absence of large rainfall events, for example, median TTCs reduced from 382 to 55 cfu/100 mL within 17-23 days between rounds R1 and R3. The two large successive daily rainfall events of 40 mm preceding round R4 resulted in substantial groundwater recharge, an almost five-fold increase in median TTCs to 262 cfu/100 mL, and increases in TTCs at 73% sources.

TLF shows a similar trend to TTCs across sampling rounds (Figure 8-4B). Significant differences exist between wet and dry season rounds, with median TLF being highest in round R1 and lowest in rounds R5 and R6. HLF and TBCs also show significant differences only between wet and dry season rounds, with both at minima in the dry season. Turbidity and SEC show the least variability by sampling round, with fewest significant differences between rounds.

TLF is generally the most strongly correlated in-situ approach with TTCs in each sampling round (Figure 8-4C). Positive correlations are very strong or strong during the wet season rounds R1-4, but only moderate or weak in dry season rounds R5 and R6, respectively (Figure 8-4D). The strongest coefficient is during round R4, following the two large successive rainfall events. Correlation coefficients between HLF and TTCs are similar or marginally lower, notably in round R4, than between TLF and TTCs, with significant correlations in all rounds (Figure 8-4C), apart from R6 where significance is borderline ($p = 0.051$). TLF and HLF are also better predictors of TTCs ≥10 cfu/100 mL in the wet season rounds (mean AUC 0.84 and 0.73, respectively) than dry season rounds (mean AUC 0.68 for both).

TBCs are intermittently significantly correlated with TTCs, with strong (ρ_s 0.70) and moderate (ρ_s 0.56) correlations during rounds R4 and R1, respectively. Other in-situ approaches are rarely significantly correlated within TTCs and coefficients are typically weak or very weak (Figure 8-4C). Only TBCs in the

dry season and SRS in the wet season have an AUC > 0.70 for classifying TTCs ≥10 cfu/100 mL (both mean seasonal AUC 0.71).

There are also notable associations between in-situ approaches. During rounds R5 and R6 there is an almost perfect positive correlation between TLF and HLF (mean ρ_s 0.97); ρ_s remains very strong, but is lower in rounds R1-4 (mean r^2 0.88) (Figure C-3). The TLF:HLF ratio is higher in rounds R1-4 (median 0.95) than R5-6 (median 0.70), with the percentage of samples having a ratio >1 also decreasing from 45 to 7%. The lower TLF:HLF ratio in R5-6 is a result of a greater reduction in TLF relative to HLF (Figure 8-4C). In rounds R5 and R6, when the relationships between TLF/HLF and TTCs weaken, TLF and HLF are both strongly positively correlated with TBCs (mean ρ_s 0.62).

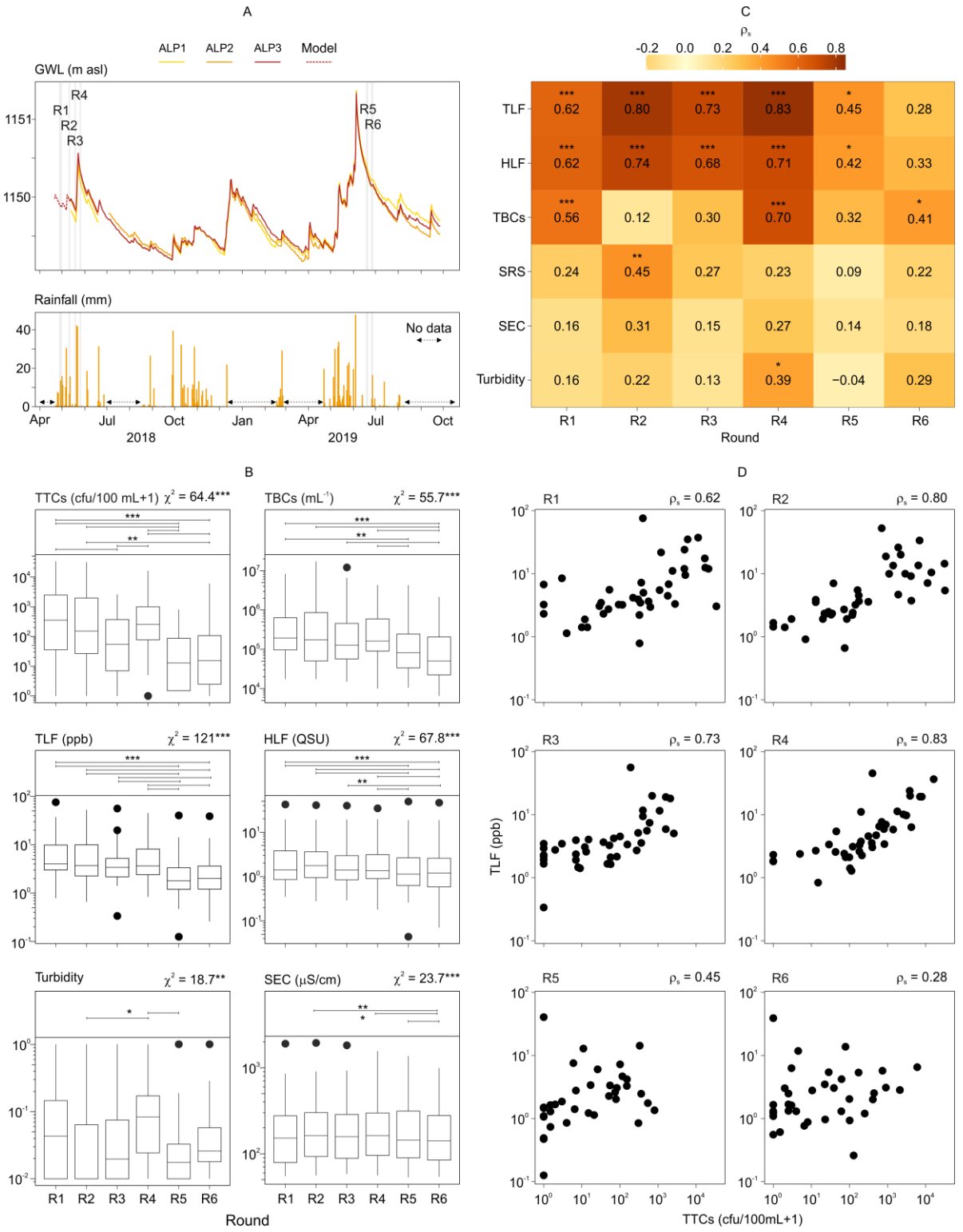

Figure 8-4 (A) Relationship between groundwater levels (GWLs) and rainfall in Lukaya illustrating timing of all sampling rounds in grey; (B) Tukey boxplots of TTCs and in-situ approaches by sampling round with χ2and significance of Friedman tests above each subplot and significant differences between rounds from post-hoc Nemenyi tests marked by ends of horizontal lines; (C) Spearman's Rank correlation coefficients and significance between TTCs and each in-situ approach for all sampling rounds; (D) Scatterplots of TTCs and TLF for each sampling round with corresponding Spearman's Rank correlation coefficients shown. p-values of <0.05, <0.01 and <0.001 are denoted by '*','', and '***', respectively.**

8.3.4 Cross-correlations between in-situ approaches and TTCs across sampling rounds

There are very strong positive rank cross-correlations for both HLF and TLF between sampling rounds at the 36 sources, but rank cross-correlations are weaker and more varied for TTCs (Figure 8-5A, B, D). The source rank-order by HLF is most consistent with a mean ρ_s of 0.91 (σ 0.04) and, remarkably, a ρ_s of 0.95 between rounds R1 and R6 (Figure 8-5B), separated by 14 months. The mean ρ_s for TLF is 0.86, with consistently very strong correlations between all rounds (σ 0.03). The rank-order of sources by TTCs is inconsistent, moderately correlated on average (ρ_s mean 0.57, σ 0.11), but with only a weak correlation between rounds R1 and R6 (Figure 8-5D). Bulk hydrochemistry rank-order of sources, as indicated by SEC, is also consistent between rounds (ρ_s mean 0.90, σ 0.07) (Figure 8-5D), but SEC is unrelated to TTCs (Figure 8-4C).

A survey of TLF or HLF across the community in either the dry or the wet season relates to TTCs during the wet season when TTCs are most elevated. Ranking the sources based on HLF intensity during any sampling round correlates well (ρ_s mean 0.68, σ 0.06, all p-values <0.001) with the rank-order of the sources by TTCs during the wet season rounds R1-4 (Figure 8-5F). TLF cross-correlates similarly to HLF with TTCs over the same time period (ρ_s mean 0.67, σ 0.09, 92% p-values <0.001), although some coefficients for round R1 with TTCs are weaker (Figure 8-5E). Note, because of the very strong rank-correlation between HLF and TLF in the dry season rounds (mean ρ_s 0.97), both rank-correlate near-identically and strongly with TTCs during the wet season rounds. Importantly, dry season TLF and HLF (both mean ρ_s 0.68, σ 0.05, all p-values <0.001) both correlate more strongly than dry season TTCs (mean ρ_s 0.50, σ 0.10, 38% p-values <0.001) with wet season TTCs.

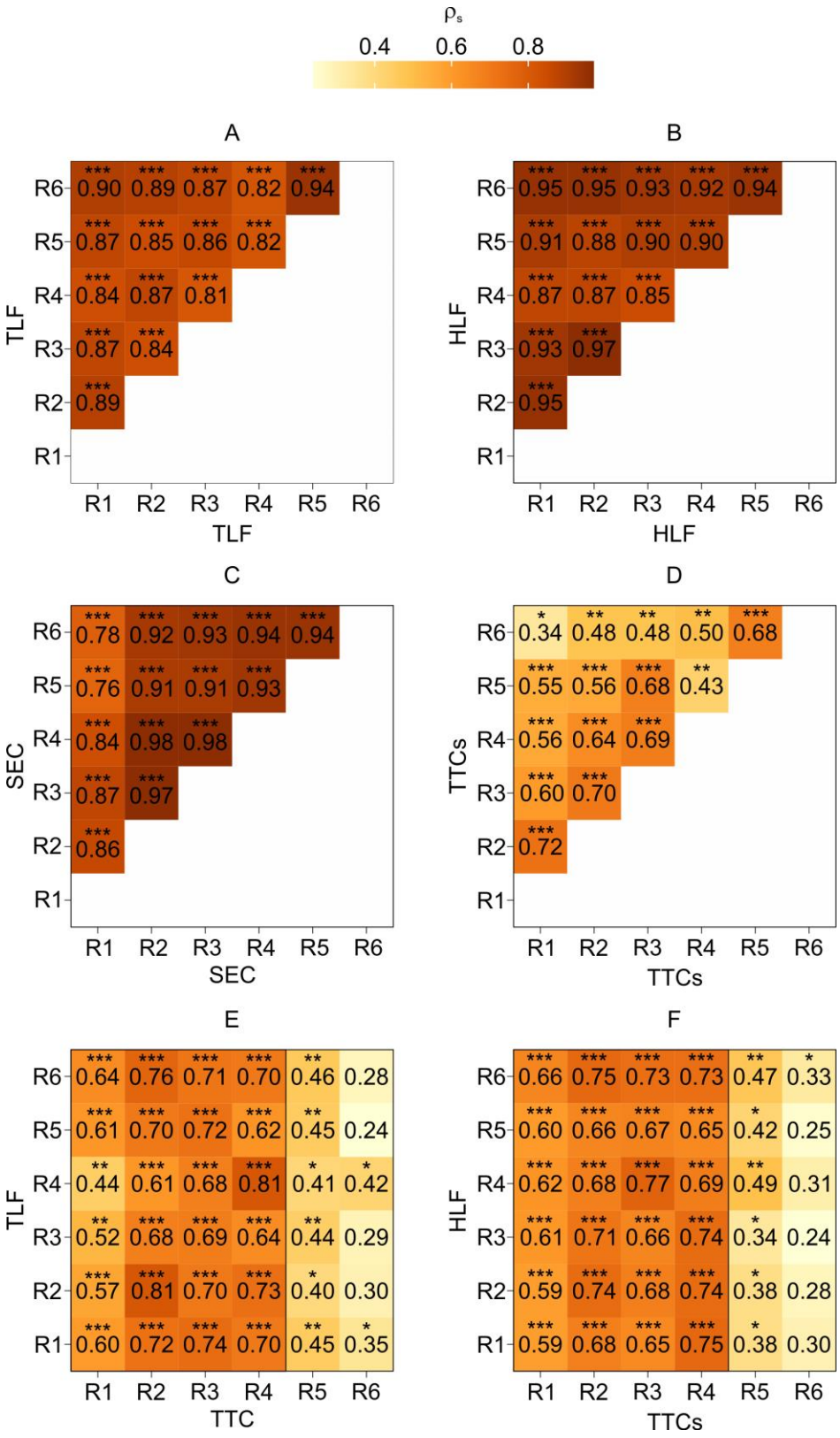

Figure 8-5 Cross-correlations between variables in each sampling round illustrated by Spearman's Rank correlation coefficients for (A) TLF; (B) HLF; (C) SEC (D) TTCs; (E) TLF and TTCs; (F) HLF and TTCs (n = 36). p-values of <0.05, <0.01 and <0.001 are denoted by '*','', and '***', respectively.**

8.4 Discussion

8.4.1 In-situ TLF/HLF as rapid approaches to indicate faecal contamination

In our study, TLF/HLF are the superior rapid approaches to indicate faecal contamination of groundwater sources, as determined by TTCs. To set these results in a wider context, we re-analysed existing published datasets from contrasting hydrogeological settings following the same statistical approach (Figure 8-6). The datasets were collated from: i) boreholes drilled to a consistent depth in an alluvial aquifer in Bihar, India (n = 145) (Sorensen et al., 2016); and ii) boreholes (n = 50) and shallow hand-dug wells (n = 61) tapping either quartzite/dolomite or the overlying weathered saprolite/laterite, respectively, in Zambia (Chapter 3).

The re-analysis of these datasets demonstrates TLF is an effective significant predictor of the presence-absence of TTCs in a 100 mL sample in these other settings (Figure 8-6A). Logistic regression models using TLF are significant ($p < 0.001$) and the AUC is 0.89-0.94. SRS and turbidity perform no better than a random classifier in India; whilst both are significant predictors ($p < 0.05$) in boreholes in Zambia, their AUCs are much lower than TLF (Figure 8-6A). The shallow wells in Zambia were typically always contaminated with TTCs present in all but 5 of the 61 samples, so AUCs were not estimated.

TLF is the most correlated in-situ indicator of the number of TTCs in our study, and the re-analysis of other published data (Figure 8-6B). In India, there is a significant relationship between TLF and TTCs, but not between either SRS or turbidity and TTCs. In Zambia, TLF is strongly correlated with TTCs in boreholes, but only weak relationships exist between SRS or turbidity and TTCs. Correlation coefficients with TTCs also remain strongest for TLF, from the rapid approaches, in the shallow wells. An online application of TLF in groundwater-derived public water supplies in the UK has also demonstrated that TLF was better correlated (ρ_s 0.71) with *E. coli* than online turbidity (ρ_s 0.48) (Chapter 5).

There are alternative groundwater studies that have presented evidence that TLF has been unrelated to FIOs in groundwater. Nevertheless, TLF has still served as an effective in-situ indicator of

contamination deriving from faecal matter in these studies. For example, Chapter 7 showed TLF was related to the density of on-site sanitation and associated nitrate but not TTCs beneath Dakar, Senegal. This study was also undertaken during the dry season and the results of our study suggests TLF/HLF relationships with FIOs are seasonal, and it is possible that during the wet season TLF/HLF could relate to FIOs in Dakar. Alternatively, the fluorophores in Dakar could relate to historic faecal contamination, as also observed at a source adjacent to an abandoned pit latrine in Malawi containing perennially high TLF, but sporadic and low TTC counts (Ward et al., 2021).

There remains inconsistent evidence regarding the use of turbidity and SEC (Buckerfield et al., 2019; Jung et al., 2014; Pronk et al., 2006; Pronk et al., 2009; Valenzuela et al., 2009), or sanitary inspections (Bain et al., 2014b; Kelly et al., 2020; Misati et al., 2017) to determine faecal contamination risk in groundwater. Turbidity and SEC can derive from a variety of common sources, including the re-mobilisation of particles within the aquifer, and the relationship with faecal indicator bacteria in the literature is consequentially inconsistent (WHO, 2017d). We consider that TLF/HLF are more appropriate indicators of variations in source water quality that relate to faecal contamination. A recent review by Kelly *et al*. (2020) suggested it was inappropriate to use sanitary inspections as indicators of microbial water quality. They argued that microbial samples from the same source are highly varied, whereas a sanitary inspection serves as a "lasting condition of the water source". Other limitations of SRS are they only represent conditions local to the source, whereas rapid subsurface transport of enteric pathogens can occur over large distances in fracture flow aquifers (Worthington and Smart, 2017), and it is not possible to assess failure of the sanitary seal subsurface. Nevertheless, sanitary inspections are undoubtedly invaluable irrespective of whether they are indicative of microbial water quality, particularly as they provide information about potential risks and causes of contamination to inform interventions.

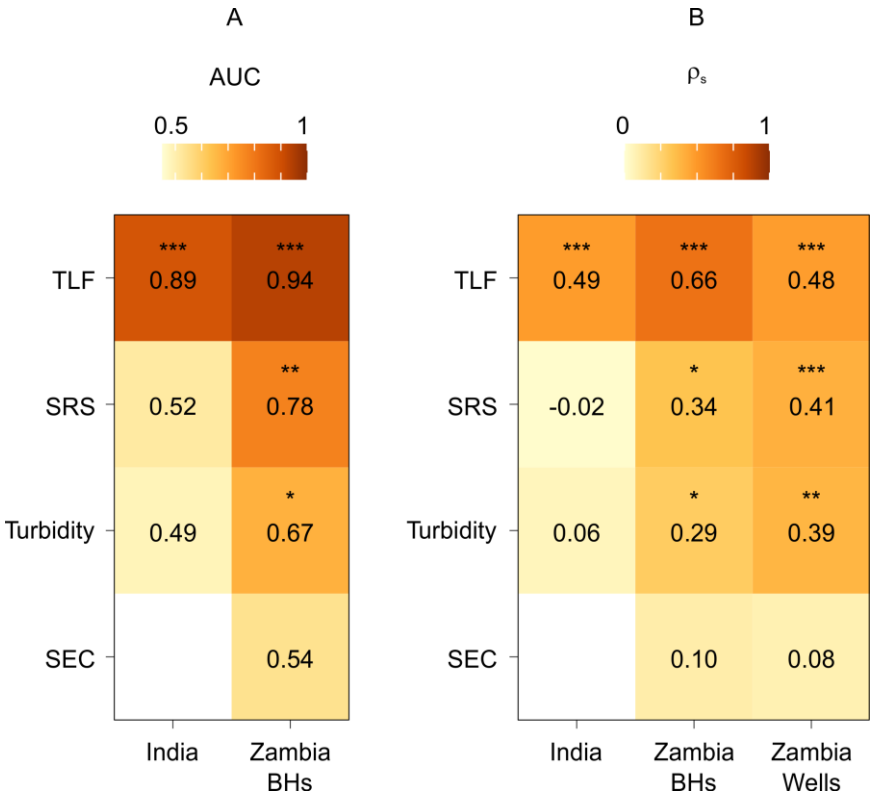

Figure 8-6 (A) Area under curve (AUC) and significance of the logistic regression models for each in-situ rapid approach as a classifier of TTCs ≥1 cfu/100 mL and (B) Spearman's Rank correlation coefficients and significance for in-situ rapid approaches against TTCs. Data are from previous studies in India (Sorensen et al. 2016), and Zambia split by source type: borehole (BH) and Wells (Shallow well) (Chapter 3). AUC is not shown for Zambia Wells because of only five from 61 samples where TTCs <1 cfu/100 mL. p-values of <0.05, <0.01 and <0.001 are denoted by '*','', and '***', respectively.**

8.4.2 TLF and HLF are more temporally resilient indicators of faecal contamination risk than TTCs (thermotolerant coliforms)

TLF and HLF are more temporally resilient faecal contamination indicators in groundwater than TTCs within our study. We highlight comparable observations in Zambia where TLF remained elevated in several shallow sources over a period of four months, whereas TTCs were only elevated in the wet season (Chapter 3); this dynamic was also recently suggested at five water sources in Malawi by Ward *et al.* (2021). Despite TLF remaining seasonally elevated in several Zambian sources, there was an overall trend towards higher median TLF and TTCs in the wet season (7.1 ppb and 48 cfu/100 mL) relative to the dry season (2.8 ppb and 2 cfu/100 mL). Re-analysis of the Zambia data demonstrates that the relationship between TLF and TTCs is stronger in the wet (ρ_s 0.82) than the dry season (ρ_s 0.67). Moreover, there is also a stronger cross-correlation between dry season TLF and elevated

wet season TTCs (ρ_s 0.80), than dry and wet season TTCs (ρ_s 0.60). In summary, both TLF and TTC vary seasonally in Uganda and Zambia, but ranking the sources within a community by faecal contamination risk using TLF is a more temporally robust approach than using TTCs.

Contrasting seasonal variations and relationships between TLF/HLF and TTCs suggest TLF/HLF differ from TTCs in one or more properties: (i) their source, (ii) their transport properties, and/or (iii) their persistence in the subsurface. The dominant source term for both types of faecal indicator is likely to be effluent from on-site sanitation in urban low-income settings, where present, except potentially where there are naturally high levels of sedimentary fluorescent NOM, or water is contaminated with fluorescent xenobiotic compounds, such as diesel (Carstea et al., 2010). Inputs from on-site sanitation are likely to be greatest during the wet season, particularly following large rainfall events, when latrines can be inundated and overflow (Nayebare et al., 2020), and accumulated faecal matter on the ground surface can be mobilised (Howard et al., 2003). There is also likely to be a continuous input function from on-site sanitation, as pit latrines and septic tanks, leak year-round.

The faecal indicators have different transport properties. Frank *et al.* (2021) demonstrated that dissolved tryptophan was comparable in transit time and recovery to the conservative dye tracer uranine over short distance, <2h tracer tests, with no evidence of retardation. Frank *et al*. also demonstrated similar recovery for a humic acid, although there was some evidence of retardation and the tracer peak was marginally delayed by five minutes, in comparison to uranine. It is unclear what proportion of TLF/HLF can be attributed to dissolved pure tryptophan or the humic acid used in any given setting, although TLF/HLF fluorophores are predominantly extracellular in groundwater (Chapter 6). Nevertheless, there will also be an element of sorption and desorption of dissolved OM between groundwater and the aquifer matrix and soils (Shen et al., 2015), particularly for more hydrophobic molecules, which may have a TLF/HLF component, as well as a minor component contained within cells. TTCs can be transported more rapidly than solutes, notably in heterogeneous media such as weather crystalline rocks, but are subject to appreciable attenuation (Taylor et al.,

2004). TTCs tend to accumulate and be transported laterally when flow velocities increase (WHO, 2017b), such as during a rainfall event generating groundwater recharge. Therefore, TLF/HLF are likely to be more readily and continuously transported than TTCs in groundwater.

The persistence of TLF/HLF fluorophores and TTCs are likely to differ in groundwater. HLF is expected to be the most persistent indicator, demonstrating the strongest rank-order cross-correlation between sampling rounds. Furthermore, although HLF decreases in the dry season, there is a proportionally greater loss in TLF indicating either preferential breakdown or more efficient lateral transport of TLF fluorophores. HLF has been demonstrated to be more recalcitrant, resistant to breakdown, than TLF in surface water and wastewater (Cory and Kaplan, 2012; Ignatev and Tuhkanen, 2019) and a greater proportion of fluorophores are like to be recalcitrant in groundwater where NOM is typically less bioavailable (Chapelle, 2021; Shen et al., 2015). The refractory nature of some HLF fluorophores led Zheng *et al*. (2020) to suggest that HLF is an effective tracer of wastewater in groundwater. There is also potential for the in-situ production of TLF or HLF from NOM entering an aquifer system (Fox et al., 2017; Yang et al., 2020). Therefore, fluorophore persistence in the subsurface could be a result of the continuous recycling and microbial transformation of NOM arriving in the system as opposed to the accumulation of recalcitrant molecules (Benk et al., 2019; Roth et al., 2019). The dry season relationships between TLF/HLF and TBCs when faecal inputs are more limited, as also observed in Senegal (Chapter 7), suggest bacteria are using the NOM as a substrate and potentially generating fluorophores in-situ. Irrespective of the relative persistence of either TLF or HLF, either wavelength pair would provide a similar indicator of faecal contamination risk given their co-correlation in our study and the optical overlap between the peaks. TTCs are generally only considered indicative of recent contamination with die-off within 16-45 days (Taylor et al., 2004), in contrast to the more persistent fluorescence indicators.

The more efficient transport of TLF/HLF fluorophores and their greater persistence in the subsurface in comparison to TTCs could explain why dry season TLF/HLF relates to wet season TTCs. Firstly, more

efficient transport could facilitate the perennial transport of TLF/HLF fluorophores from a faecal source to a water source, whereas TTCs are predominantly mobilised following rainfall in the wet season. It should also be re-iterated that TLF/HLF does also respond to rainfall with higher intensity in the wet season, indicating higher seasonal risks. Secondly, faecal contamination events at a water source would remain detectable for a longer period by fluorescing more persistent TLF/HLF fluorophores than TTCs. If these events are focussed in the wet season, as observed here, then the proportion of TLF/HLF persisting into the dry season may relate to wet season TTCs, given the two types of indicator correlate very strongly after heavy rainfall (e.g. TLF, ρ_s 0.83, R4).

8.4.3 Remaining uncertainties, instrumentation improvements, and future work

There are a range of potential interferents with in-situ fluorescence measurements that are discussed in a review by Carstea *et al.* (2020) but these have not adversely impacted previous TLF-FIO studies (Chapter 4) or this study across a range of settings. Corrections for temperature, turbidity, and absorbance of light by the sample matrix (the inner-filtering effect) are not likely to be necessary in the majority of groundwater settings (Khamis et al., 2015). Moreover, the next generation of commercially available portable fluorimeters are now capable of automatic corrections. pH does not have an appreciable impact on TLF/HLF between values of 5 and 8 (Reynolds, 2003; Spencer et al., 2007), and groundwater outside this range is unlikely to be suitable for drinking. High concentrations of metal ions could quench fluorescence (Yang et al., 2018), which is most likely where water is contaminated by mining and industry. Certain water treatments, including chlorination, also quench fluorescence (Henderson et al., 2009) so the well owner or other informed individuals should be interviewed to assess if the water has been treated prior to testing, as would be undertaken before FIO sampling. Alternatively, a chlorine residual test could be performed.

There is the potential for TLF or HLF fluorophores to originate from contamination unrelated to faecal sources such as diesel and fuel derivatives, food waste, paper mills, and pesticides (Carstea et al., 2016). In these instances, a source displaying high TLF/HLF should still be considered a higher faecal

contamination risk than one displaying low TLF/HLF, as there would be evidence that a pathway is present to a source of anthropogenic waste. Sedimentary fluorescent NOM contained within the aquifer could also potentially be problematic when comparing faecal contamination risks determined by TLF/HLF, particularly between study areas. In which case, deviation from baseline fluorescence intensity in uncontaminated sources would be more important than the absolute value for determining risk.

Relatively high upfront costs undoubtedly constrain widespread adoption of in-situ fluorescence spectroscopy. The present generation of single peak fluorimeters cost in the region of US$5000-7000, before considering accessories that can also cost a further US$2000-3000 (Chapter 4). However, there is substantial scope to reduce these costs through the development of lower-cost portable fluorimeters, engineered specifically to provide an in-situ indication of faecal contamination risk at a water source. Multiple researchers have developed prototype fluorimeters with various benefits over commercial alternatives (Bedell et al., 2020; Bridgeman et al., 2015; Simões et al., 2021), but field validation and a discussion of indicative costs are absent or limited. As part of our study, we successfully developed and demonstrated the efficacy of a lower-cost prototype portable multi-wavelength LED-based fluorimeter on duplicate samples in rounds R5 and R6 (Figure C-4). The prototype provided comparable results to the UviLux sensors in both the laboratory (Table C-5) and field. For example, the prototype derived HLF data from R5 and R6 both correlate strongly with TTCs in R4 (mean ρ_s 0.69). Therefore, a low-cost, portable fluorimeter to indicate faecal contamination risk could be produced for a total component cost of $1100. Further details are provided in the supplementary information (Section C.1). In addition to reducing costs, future development should investigate the production of low-cost sealed long-life containers of TLF/HLF standards. These containers would enable calibration checks, ideally annually, and negative controls to be performed by the end-user without return to the manufacturer or access to a well-equipped laboratory with reagents and high quality deionised water (Chapter 4).

It remains unclear how TLF/HLF relate to the presence of enteric pathogens or risks posed to human health. There is one published study showing a relationship between TLF and DNA markers of enteric pathogens, although this study is limited to 22 sources in one town (Sorensen et al., 2015b). Future work should explore the potential link between TLF/HLF and enteric pathogens using molecular approaches, as well as exploring the viability of pathogens where possible. Furthermore, studies should investigate if and how TLF/HLF could effectively be used for on-site risk communication to induce behavioural change in communities and reduce the disease burden relating to the consumption of faecally contamination drinking water.

8.5 Conclusions

In-situ fluorescence spectroscopy provides an instantaneous assessment of water source quality that relates to faecal contamination risk determined by faecal indicator organisms (FIOs). Consequently, faecal contamination risks can be assessed immediately, including in real-time, and could be communicated on-site to consumers to reduce exposure to contamination, whilst confirmative regulatory FIO analysis is undertaken. Furthermore, in-situ fluorescence can extend FIO sampling programs because data can be collected rapidly by users who require minimal training; nor are there consumable costs for additional samples.

TLF and HLF are more temporally resilient indicators of faecal contamination risk than FIOs. Both types of indicator respond to rainfall and contamination events, with the strongest relationships between the indicators observed in the wet season, notably immediately after heavy rainfall (e.g. TLF-TTC ρ_s 0.83). However, ranking the sources across a community by risk using FIOs is more variable (cross-correlations ρ_s 0.34-72) between sampling rounds, than using TLF or HLF (cross-correlations ρ_s 0.81-97). This ranking of sources using TLF/HLF at any point in time relates to TTCs during the wet season, when TTCs are significantly elevated and risks to human health would consequently also be expected to be greatest. Furthermore, the source rank-orders in the dry season using TLF/HLF cross-correlate more strongly (both mean ρ_s 0.68) with wet season TTCs than dry season TTCs (mean ρ_s 0.50).

Therefore, the comparative faecal contamination risks between sources generated by a dry season survey of TLF/HLF would be more accurate than using highly transient FIOs to indicate the comparative risks occurring in the wet season when risks are elevated. This characteristic is advantageous given water quality surveys are infrequent for private water supplies globally, as well as across low-income countries. TLF/HLF provide a more repeatable and temporally robust approach than FIOs to ranking sources by faecal contamination risk across a community to strategise prioritisation of sources for drinking or interventions.

Chapter 9 - Conclusions

9.1 A rapid indicator of faecal contamination

In-situ fluorescence spectroscopy provides an instantaneous assessment of source water quality that relates to faecal contamination risk, determined by faecal indicator organisms (FIOs), across a range of hydrogeological and climatological settings (Objectives 1&2). Tryptophan-like fluorescence (TLF) was a significant predictor in logistic regression models of the presence-absence of FIOs, with the area under the receiver-operator curve (AUC) considered “excellent” to “outstanding” (0.88-0.92) (Chapters 3, 4, 8). Furthermore, the correlation coefficients between TLF intensity and FIO enumeration were significant, with ranked correlation coefficients of between 0.49 and 0.80 (Chapters 3, 4, 5, 8). In Uganda, humic-like fluorescence (HLF) was also shown to be a similarly effective indicator to TLF of FIO presence-absence and enumeration (Chapter 8).

Fluorescence thresholds can successfully be determined to indicate the presence and extent of faecal contamination (Objective 3, Chapters 3, 4, 8). Analysis of pooled datasets from India, Malawi, South Africa, and Zambia showed that a TLF threshold of 1.3 ppb dissolved tryptophan can instantaneously predict FIOs >= 10 cfu/100 mL, with a false-positive error rate of 18% and a false-negative error rate of 4% (Objective 7, Chapter 4). The threshold is not effective at classifying contaminated sources with <10 cfu/100 mL. The repeatability of field TLF measurements is 0.4 ppb (Chapter 8), so such a low threshold value of 1.3±0.4 ppb is effectively suggesting that any source exhibiting only traces of TLF is at risk of contamination. A higher threshold of 6.9 ppb has been proposed to indicate higher faecal contamination risk sources (Chapter 4). Nevertheless, in certain hydrogeological settings, there will be naturally high levels of sedimentary fluorescent organic matter, and elevation above the natural baseline would indicate contamination, as opposed to fluorescence thresholds devised from other settings. Thresholds to indicate likely presence of FIOs are also not appropriate in highly contaminated slow-moving intergranular aquifers (Chapter 7).

TLF or HLF are superior to other rapid approaches to indicate faecal contamination as determined by FIOs (Chapters 3, 5, 8). TLF was the superior single-predictor of FIO contamination in logistic regression models, compared with other rapid approaches including turbidity and sanitary risk assessments (Chapters 3, 8). Furthermore, no other rapid approaches provided additive performance to this model in Zambia (Chapter 3). TLF and HLF were more strongly correlated with FIOs than other rapid approaches on each occasion (Chapters 3, 5, 8), and TLF was incorporated first into multiple linear regressions of FIOs (Chapters 3, 8). In Zambia, nitrate and sanitary risk scores improved the multiple linear regression model (Chapter 3), but in Uganda TLF was the only independent variable included (Chapter 8).

9.2 Advantages over faecal indicator organisms

In-situ fluorescence spectroscopy has multiple methodology advantages over FIOs for inferring faecal contamination risks and protecting public health:

1. Instantaneous results. Sampling and culturing of FIOs takes 18-48 h and consequently exposure, potentially by large numbers of people in piped supplies, can occur before a positive result is received and communicated to consumers. Fluorescence can be quantified in-situ at the source during roaming studies (Chapters 3, 4, 6, 7, 8) or in real-time during inline deployments (Objective 2, Chapter 5). These instant results allow rapid evidence-based communication of faecal contamination risks that could reduce enteric pathogen exposure.
2. Target fluorophores are more temporally resilient indicators of faecal contamination risk than FIOs (Objective 4). FIOs are episodic and, generally, do not survive for long in the subsurface and, therefore, to understand whether a source is at risk using FIOs, regular data need to be collected over a sustained period of time, which is uncommon. TLF and HLF are more persistent than FIOs in the subsurface. For example, some sources displaying elevated FIO contamination in the wet season but an absence of FIO in the dry season, maintain perennially elevated TLF, although the intensity does vary and mirror that of FIOs (Chapter 3). Ranking

sources across a community using TLF or HLF during the dry season is more strongly related to the wet season ranking by TTCs, when TTCs are most elevated, than dry season TTCs (Chapter 8). Fluorophores may more temporally resilient indicators as they are predominantly extracellular and less prone to straining than FIOs (Objective 6, Chapter 6). Equally, once fluorophores are emitted into the subsurface, predominantly during the wet season (Objective 4, Chapters 3, 8), there can be a continuous cycle of uptake, recycling and transformations by microbes, which could explain dry-season associations between fluorophores and total bacterial cells (Objective 5, Chapters 7, 8). A proportion of fluorophores, particularly some producing HLF, will be refractory in nature and could also explain how fluorophores can accumulate in some aquifers (Chapter 7). Finally, it is also known that *E. coli* excrete tryptophan as they move from a culturable to a dormant viable but nonculturable state (Arana et al., 2004).

3. Analytical simplicity that does not require an aseptic environment. During all roaming surveys (Chapters 4, 6, 7, 8), fluorescence data were collected by submerging a portable, battery-powered fluorimeter, in a beaker of water by the source, after both the fluorimeter and beaker had been rinsed in the sample water. No consumables nor reagents were used in the field, following prior factory or laboratory calibrations. In-field training was successfully provided to MSc and overseas PhD students who were also able to collect high quality, repeatable fluorescence data (Objective 7, Chapters 6, 7, 8).

9.3 Looking forward

9.3.1 An approach to improve drinking water quality and reduce exposure to waterborne pathogens

In-situ fluorescence spectroscopy offers an instantaneous approach to screen water sources for faecal contamination risk. Within municipal supply networks, real-time fluorimeter deployments offer scope to identify and provide an early-warning of faecal breakthroughs at groundwater sources, as well as across the entire treated distribution network. Sudden increases in intensity in

treated water may signify ineffective water treatment, cross-connections, or a loss of integrity within the system. Fluorimeters can also be used in a roaming capacity to check for integrity within the distribution network and allow technicians to rapidly diagnose the location of any faults. Instant results also enable rapid communication to consumers to potentially encourage behavioural change. From a treatment perspective, HLF is strongly related to DOC in groundwater (Chapter 7) and real-time fluorimeters could facilitate more cost-effective chlorination of potable groundwater and forewarn of the generation of harmful disinfection by-products

In all settings, in-situ fluorescence spectroscopy can be used to assess the relative risks of faecal contamination between groundwater-fed drinking water sources. Fluorescence spectroscopy is a more temporally resilient indicator of faecal contamination risk, which has great utility for assessing risks between sources when historic, frequent, records of FIOs are unavailable, which is common to private water supplies, and most supplies in low-income settings.

In-situ fluorescence spectroscopy remains a long way from being introduced from a regulatory perspective and is likely to be used conjunctively with FIO monitoring. In the municipal water supply network, detected changes in TLF/HLF could be used to target additional FIO compliance monitoring from both spatial and temporal perspectives. In low-income settings, in-situ TLF/HLF could prioritise sources for FIO sampling and analysis, extend drinking water microbial risk assessments given there are no consumable costs, and increase the impact of FIO surveys by allowing an immediate communicate of microbial risks to consumers.

9.3.2 Addressing key limitations

All analyses, herein, assessing in-situ fluorescence spectroscopy as an indicator of faecal contamination risk, determined using FIOs, have been performed on modest datasets up to a maximum of n = 564 (Chapter 4). Furthermore, the vast majority of datasets have not been drawn from an independent random sample. There is a need for a large-scale randomised field demonstration to robustly support the results of this thesis. Furthermore, evidence linking in-situ

fluorescence spectroscopy to waterborne pathogen presence is limited to a single case study in one town (Sorensen et al., 2015b) and there is no robust evidence linking the technique to risks posed to human health. More work is need to interrogate such relationships, rather than simply focusing on in-situ fluorescence spectroscopy as an alternative indicator to an imperfect indicator of pathogens.

Fluorimeters are optical sensors and are impacted by suspended particles in water. Consequently, in some settings real-time deployments may not provide reliable fluorescence data. For example, in a bank-filtration setting, continuous build-up of ferric deposits on the sensor reduced intensity by up to 40% within two weeks (Chapter 5). However, suspended particles have not been detrimental to roaming studies (Chapters 3, 4, 6, 7, 8), given turbidity is generally low in groundwater. Furthermore, the next generation of fluorimeters are now including corrections for turbidity and wipers to improve performance in both real-time and roaming applications. High upfront costs will continue to constrain the uptake of the technique, but we have demonstrated that lower cost fluorimeters (component cost $1100) are feasible (Chapter 8). It remains unclear how best to engineer a fluorimeter or reader to specifically communicate risks onsite, although the proposed thresholds (Chapter 4) could be utilised to inform risks from a categorical perspective, potentially using coloured LEDS.

A. Appendix

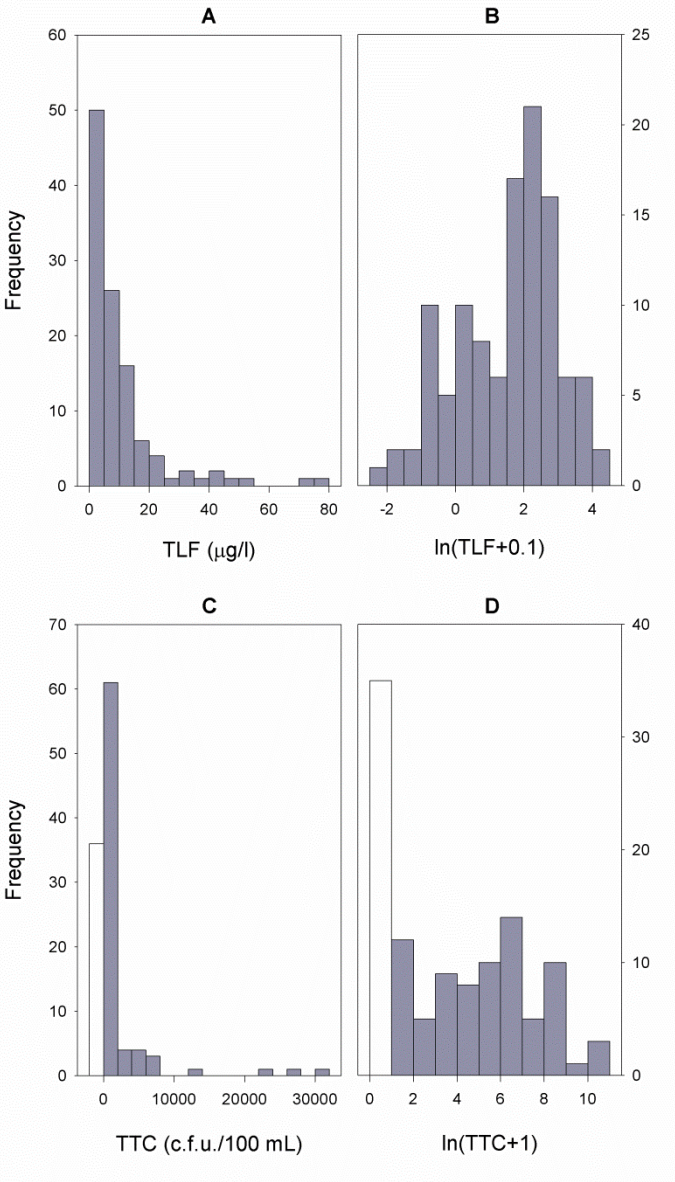

Figure A-1 Histograms of (a) raw and (b) transformed TLF data; and (c) raw and (d) transformed TTC counts. Unfilled bars refer to negative TTC counts.

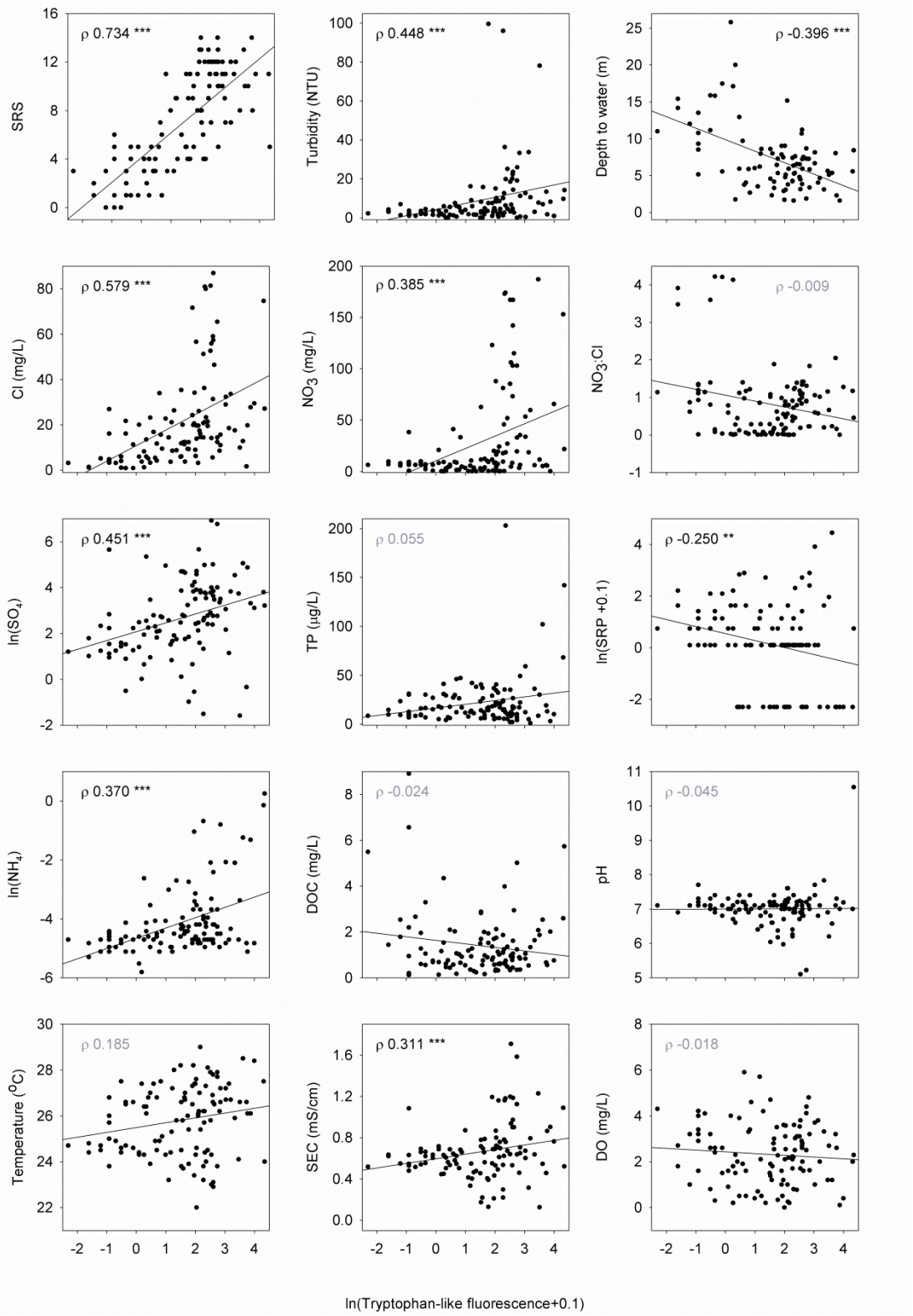

Figure A-2 Correlations between TLF and all physio-chemical parameters. Spearman's rank correlation coefficients are shown (ρ) with significance denoted as * <0.05, ** <0.01, and * <0.001. Certain parameters have been logged for illustrative purposes.**

B. Appendix

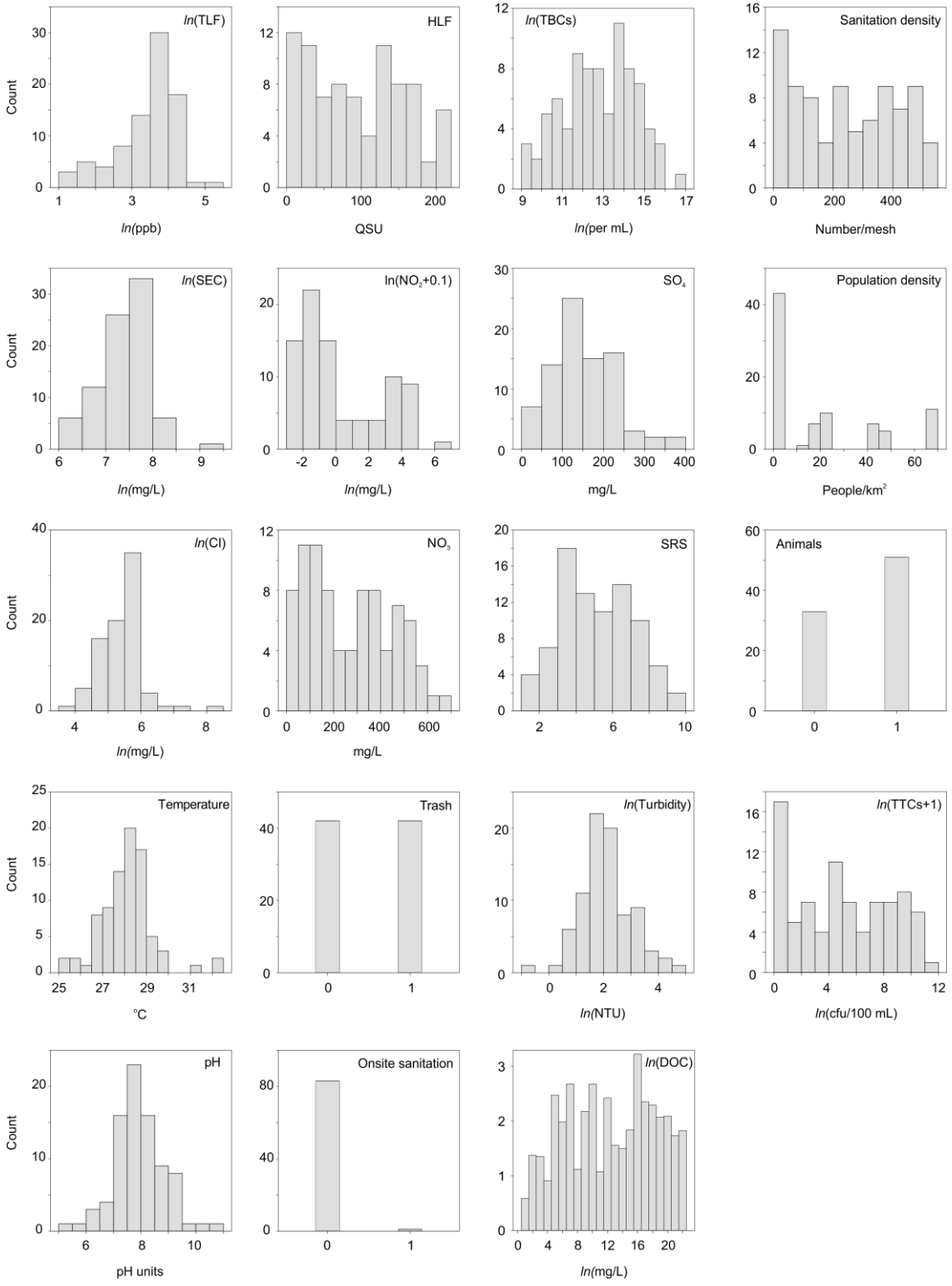

Figure B-1 Histograms of all continuous variables and bar plots of binary variables used for linear regression modelling

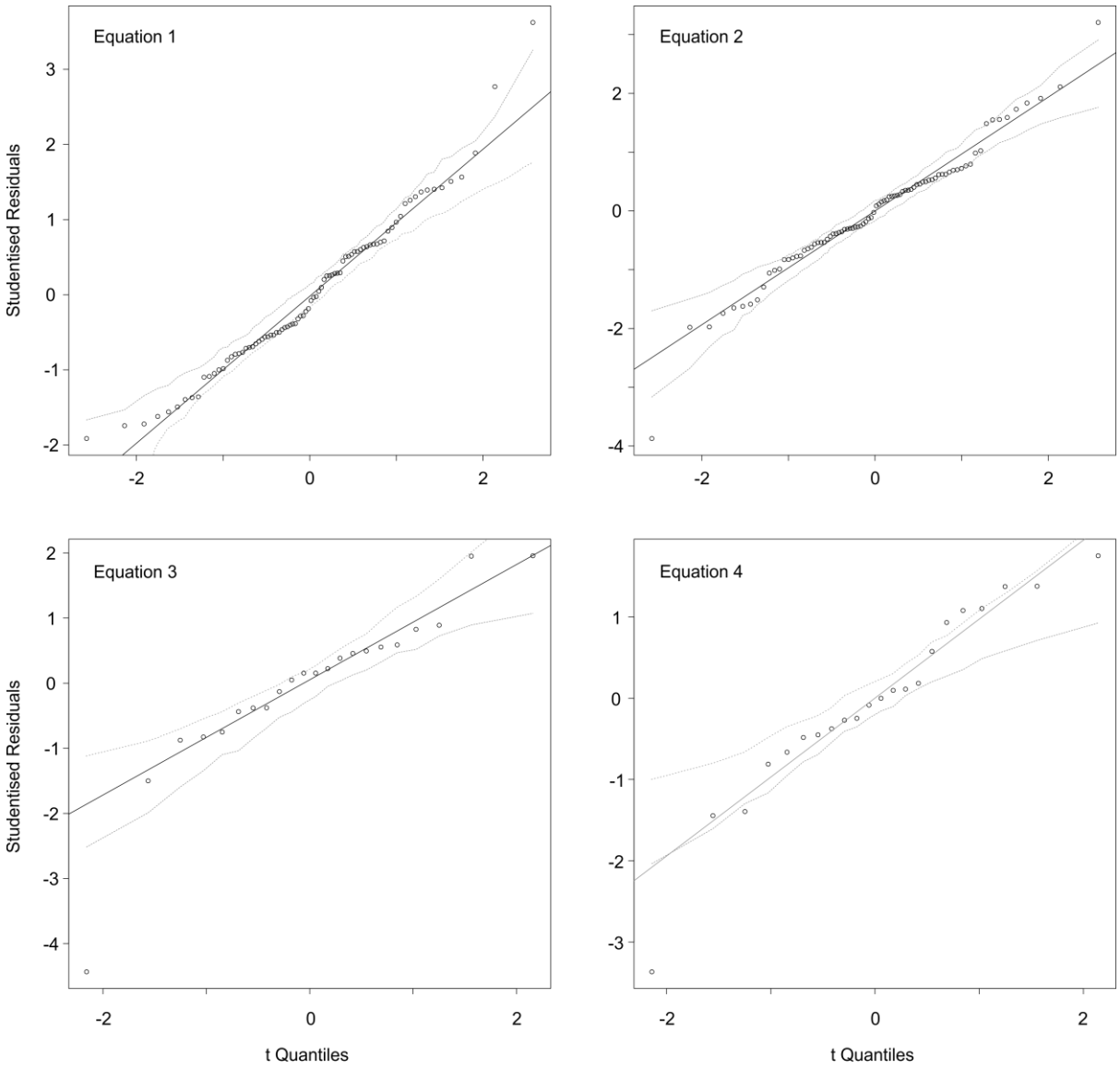

Figure B-2 Q-Q plots illustrating Gaussian residuals for Equations 1-4. A 95% confidence envelope is shown as a dotted grey line.

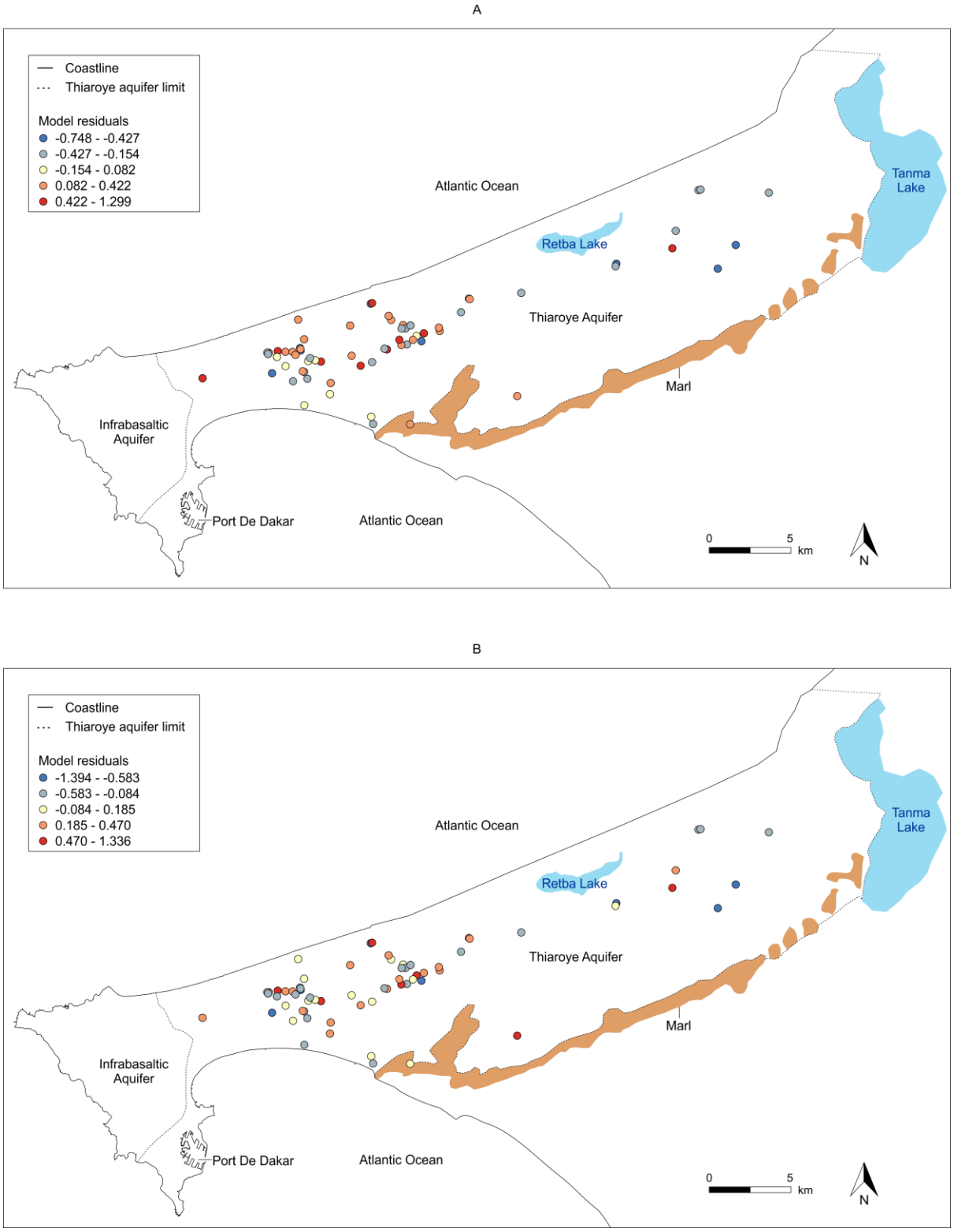

Figure B-3 Spatial plots of model residuals for (A) Equation 1 and (B) Equation 2

Table B-1 Estimates of the variance inflation factor (VIF) for all predictors in each model

Predictor	VIF
Equation 1	
HLF	1.84
NO_3^-	1.09
ln(TBCs)	2.06
OSS density	1.39
Equation 2	
ln(TBCs)	1.58
NO_3^-	1.12
ln(NO_2^-+0.1)	1.31
OSS density	1.45
SO_4^{2-}	1.29
Equation 3	
ln(TLF)	2.76
HLF	2.76
OSS density	1.06

Equations

$$\mathbf{ln}(\boldsymbol{TLF}) = \mathbf{0.559}\ \boldsymbol{HLF} + \mathbf{0.349}\ \boldsymbol{NO_3^-} + \mathbf{0.190}\ \mathbf{ln}(\boldsymbol{Total\ bacteria}) + \mathbf{0.211}\ \boldsymbol{Sanitation\ density} + \mathbf{0.974}$$

(Equation B-1)

$$\ln(TLF) = 0.387 \ln(Total\ Bacteria) + 0.331\ NO_3^- + 0.270\ NO_2^- + 0.207\ Sanitation\ density + 0.193\ SO_4^{2-} + 0.168$$

(Equation B-2)

The RMSE/r^2 are 0.35/0.78 and 0.40/0.71 for Equations S1 and S2, respectively.

C. Appendix

C.1 Development of a low-cost fluorimeter

The low-cost fluorimeter targeted $\lambda_{ex}/\lambda_{em}$ peaks at 275/320-375 (TLF) and 365/400-490 nm (HLF). Additionally, absorbance was monitored to enable correction for the inner-filtering effect (Ohno, 2002). Absorbance was low; the median at 275 nm was 0.04 cm^{-1} and corrections were not implemented. The fluorimeter utilised a sample holder to minimise contamination associated with manual handling of submersible fluorimeters (Chapter 4). In place of a standard quartz cuvette (US$85) typically used in fluorescence studies, we used transparent 3.5 mL borosilicate glass test tubes (US$0.07). Emission signals were quantified with photodiodes (US$7 for HLF and US$70 for TLF), rather than more expensive and higher sensitivity photomultipliers (US$400). To account for the lower sensitivity of the photodiodes compared to a photomultiplier, the sampling frequency was reduced from 10 to 0.0027 Hz and the signal output was averaged over six minutes to minimise noise. A touch screen was built into the device to enable control and display of results. Before field deployment, the prototype compared favourably with TLF and HLF UviLux sensors in the laboratory (Table S5). In Lukaya, both TLF and HLF datasets obtained with the lower-cost prototype cross-correlated similarly to the equivalent UviLux data with wet season TTCs; for example the prototype derived HLF data from R5 and R6 both correlate strongly with TTCs in R4 (mean ρ_s 0.69).

The total component costs of the low-cost fluorimeter to target only HLF (λ_{ex} 365 nm) is currently US$1100. This price includes all optics, electronics, screen, case, power, and assumes procurement for 50 units, with the potential to reduce costs further if scaled-up commercially. Our data suggest that such a sensor can provide similar information regarding faecal contamination risk in groundwater as a deep UV TLF sensor, given the co-correlation of HLF and TLF observed both in Lukaya, as well as in groundwater experiencing frequent faecal contamination in the UK (Chapter 5). The main advantage of the HLF sensor is cost, with the equivalent component costs of the prototype TLF sensor being 60% higher, due to required excitation within the deep-UV spectrum. Furthermore, use of deep UV light

typically requires an expensive quartz glass cuvette to minimise absorbance as opposed to borosilicate. Finally, a HLF signal is much more intensive than a TLF signal with the same amplifier circuit (Li et al., 2016), enabling greater sensitivity and precision.

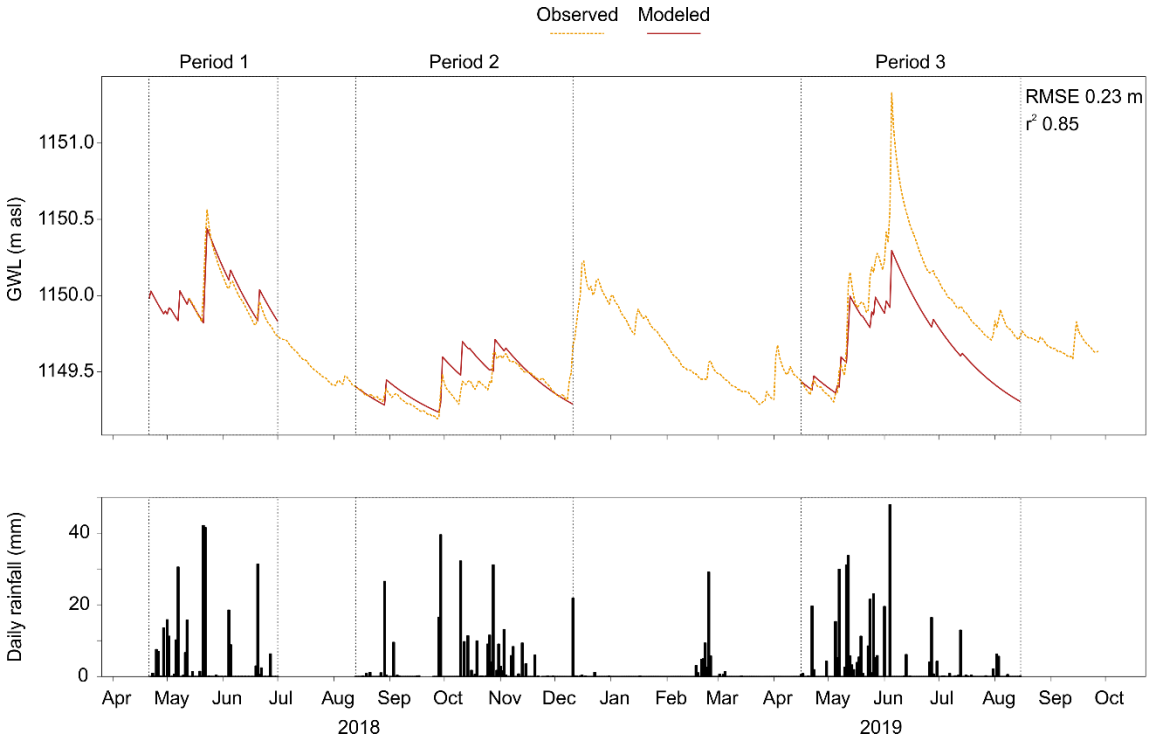

Figure C-1 Comparison of observed and modelled groundwater levels (GWL) in borehole ALP-3. Three separate periods were modelled due to gaps in the driving rainfall observations.

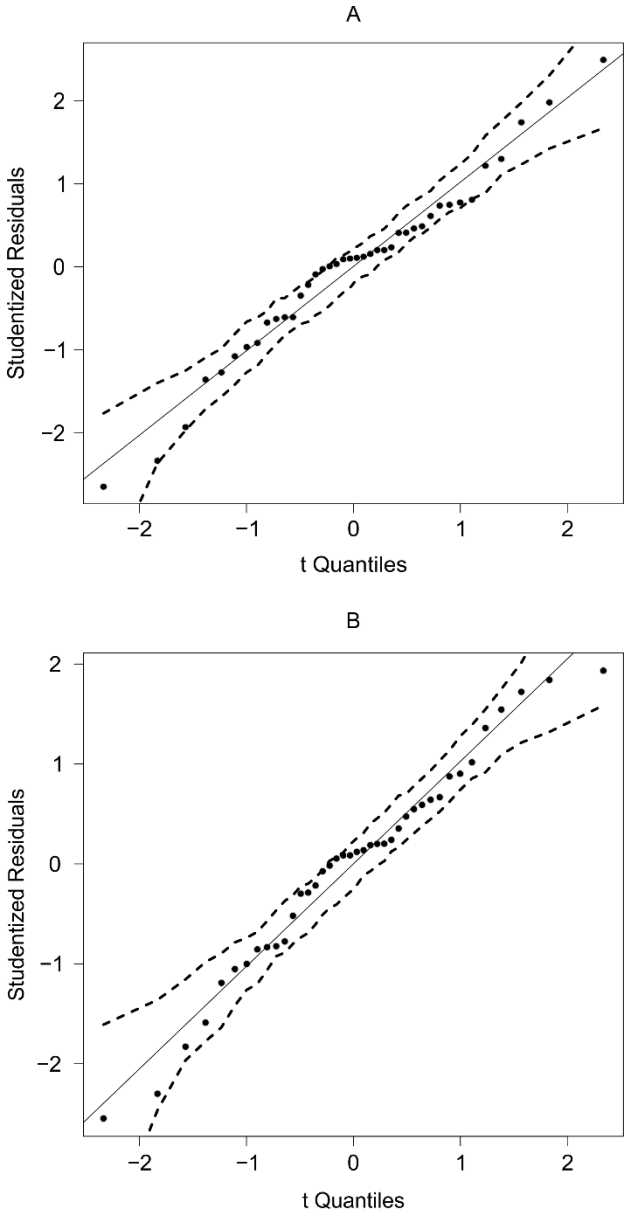

Figure C-2 Q-Q plots for (A) Equation 1; (B) Equation 2. A 95% confidence envelope is shown as dotted lines.

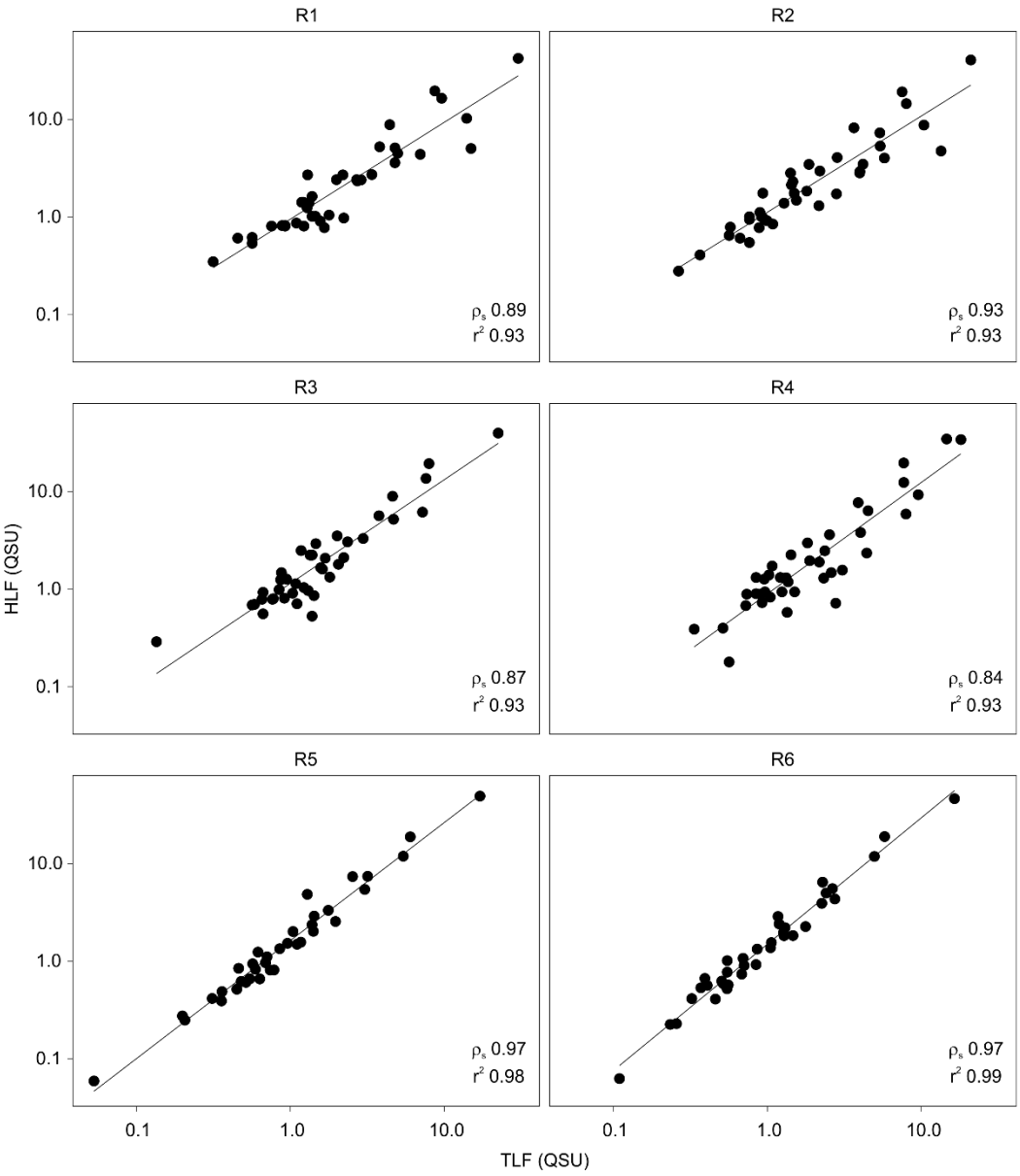

Figure C-3 Relationship between TLF and HLF in sampling rounds R1 to R6. Spearman's Rank (ps) and Pearson (r^2) correlation coefficients shown. Pearson coefficients were calculated on the natural log of the data to ensure a Gaussian distribution.

Figure C-4 Low-cost portable fluorimeter used to test the quality of water sampled from a shallow well by co-author Jacintha Nayebare in Lukaya, Uganda; the fluorimeter is contained within the open black hardcase in the foreground of the photo.

Table C-1 Combined sanitary inspection questions from hand pump and spring WHO forms. Note that questions 3 and 4 were not answered because no inspection ports were present for the hand pumps, so the respective questions on the spring form were also removed.

Question Number	Sanitary inspection question
1	Is the pump damaged or loose at the point of attachment/a protective wall or spring box structure inadequate so that contaminants could enter the source?
2	Is the cover slab or outlet pipe inadequate to prevent contaminants entering the source?
3	N/A
4	N/A
5	Is the apron or diversion ditch inadequate to prevent contaminants from entering the source?
6	Is the drainage inadequate, which may result in stagnant water around the source?
7	Is the fencing or barrier around the source inadequate to prevent animals entering the area?
8	Is there sanitation infrastructure within 15 m of the source?
9	Is there sanitation infrastructure on higher ground within 30 m of the source?
10	Can signs of other sources of pollution be seen within 15 m of the source (e.g. animals, rubbish, human settlement, open defecation, fuel storage)?
11	Is there any point of entry to the aquifer that is unprotected within 100 m of the source?

Table C-2 Beta coefficients and p-values for the hypothesis that the coefficients are zero for single-predictor logistic regression models of the probability of TTCs ≥10 cfu/100 mL. Source median values used for both independent and dependent variables.

Predictor	β	p-value
TLF	1.09	0.042
HLF	1.90	0.059
SRS	0.58	0.056
TBCS	4.48×10^{-6}	0.130
Turbidity	0.97	0.763
SEC	0.003	0.356

Table C-3 Beta coefficients, p-values, area under the receive operating curve (AUC) for the hypothesis that the coefficients are zero using sanitary inspection questions as single-predictor logistic regression models of the probability of TTCs ≥10 cfu/100 mL. Source median values used for both independent and dependent variables.

Predictor	β	p-value	AUC
Q1	16.47	0.994	0.56
Q2	0.23	0.799	0.52
Q5	0.88	0.285	0.61
Q6	0.92	0.021	0.75
Q7	-15.39	0.995	0.48
Q8	16.47	0.994	0.56
Q9	-0.53	0.513	0.56
Q10	0.53	0.513	0.56

Table C-4 Beta coefficients and p-values for the hypothesis that the coefficients are zero using sanitary inspection questions as single-predictor logistic regression models of the probability of TTCs ≥10 cfu/100 mL in each sampling round. p-values of <0.05, <0.01 and <0.001 are denoted by '*','', and '***', respectively.**

Round	1	2	3	4	5	6
Q1	0.42**	0.45**	0.13	0.15	0.27	0.45**
Q2	-0.18	-0.22	0.12	-0.03	-0.08	-0.01
Q5	0.17	0.45**	0.33*	0.29	0.15	0.26
Q6	0.11	0.16	0.02	0.09	-0.01	0.15
Q7	-0.16	-0.08	0.12	0.03	0.06	0.18
Q8	0.33*	0.53***	0.3	0.12	0.2	0.15
Q9	-0.08	-0.19	-0.2	-0.19	-0.12	-0.27
Q10	0.07	0.27	0.09	0.12	-0.13	-0.05

Table C-5 Comparison of lower-cost prototype and Uvilux fluorimeters. Limit of detection is 3σ of triplicate analysis of ultrapure water blanks. Precision is the mean of 3σ of triplicate analysis of a range of dissolved tryptophan (0, 1, 2, 5, 10, 20, 50, 100 ppb) and pyrene (0, 0.2, 0.5, 1, 2, 5, 10, 20) standards. Accuracy is the root mean square error (RMSE) for the mean of triplicate analysis of these laboratory standards.

Parameter	Prototype TLF (ppb tryptophan)	UviLux TLF (ppb tryptophan)	Prototype HLF (ppb pyrene)	Uvilux HLF (ppb pyrene)
Limit of detection	0.7	0.2	0.2	0.1
Precision	1.6	0.1	0.2	0.1
Accuracy	1.5	0.6	0.2	0.2
Calibration response (r^2)	0.999	0.999	0.999	0.999

References

Adler, S., Widerström, M., Lindh, J. and Lilja, M. 2017. Symptoms and risk factors of Cryptosporidium hominis infection in children: data from a large waterborne outbreak in Sweden. Parasitology Research, 1-6.

Alin, A. 2010. Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics 2(3), 370-374.

ANSD 2016. National census of the population of Senegal in 2016. National Agency for Statistics and Demography, Dakar, Senegal.

Arana, I., Seco, C., Epelde, K., Muela, A., Fernández-Astorga, A. and Barcina, I. 2004. Relationships between Escherichia coli cells and the surrounding medium during survival processes. Antonie van Leeuwenhoek 86(2), 189-199.

Baghoth, S., Sharma, S. and Amy, G. 2011. Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation–emission matrices and PARAFAC. Water research 45(2), 797-809.

Bain, R., Cronk, R., Hossain, R., Bonjour, S., Onda, K., Wright, J., Yang, H., Slaymaker, T., Hunter, P. and Prüss-Ustün, A. 2014a. Global assessment of exposure to faecal contamination through drinking water based on a systematic review. Tropical Medicine & International Health 19(8), 917-927.

Bain, R., Cronk, R., Wright, J., Yang, H., Slaymaker, T. and Bartram, J. 2014b. Fecal contamination of drinking-water in low-and middle-income countries: a systematic review and meta-analysis. PLoS medicine 11(5), e1001644.

Baker, A. 2001. Fluorescence excitation-emission matrix characterization of some sewage-impacted rivers. Environmental Science & Technology 35(5), 948-953.

Baker, A. 2002a. Fluorescence excitation− emission matrix characterization of river waters impacted by a tissue mill effluent. Environmental science & technology 36(7), 1377-1382.

Baker, A. 2002b. Fluorescence properties of some farm wastes: implications for water quality monitoring. Water Research 36(1), 189-195.

Baker, A. 2005. Thermal fluorescence quenching properties of dissolved organic matter. Water research 39(18), 4405-4412.

Baker, A., Cumberland, S.A., Bradley, C., Buckley, C. and 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.

Baker, A. and Curry, M. 2004. Fluorescence of leachates from three contrasting landfills. Water Research 38(10), 2605-2613.

Baker, A., Elliott, S. and Lead, J.R. 2007. Effects of filtration and pH perturbation on freshwater organic matter fluorescence. Chemosphere 67(10), 2035-2043.

Baker, A. and Inverarity, R. 2004. Protein-like fluorescence intensity as a possible tool for determining river water quality. Hydrological Processes 18(15), 2927-2945.

Baker, A. and Spencer, R.G. 2004. Characterization of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy. Science of the Total Environment 333(1-3), 217-232.

Baral, D., Speicher, A., Dvorak, B., Admiraal, D. and Li, X. 2018. Quantifying the relative contributions of environmental sources to the microbial community in an urban stream under dry and wet weather conditions. Appl. Environ. Microbiol. 84(15), e00896-00818.

Bedell, E., Sharpe, T., Purvis, T., Brown, J. and Thomas, E. 2020. Demonstration of Tryptophan-Like Fluorescence Sensor Concepts for Fecal Exposure Detection in Drinking Water in Remote and Resource Constrained Settings. Sustainability 12(9), 3768.

Benk, S.A., Yan, L., Lehmann, R., Roth, V.-N., Schwab, V.F., Totsche, K.U., Küsel, K. and Gleixner, G. 2019. Fueling diversity in the subsurface: Composition and age of dissolved organic matter in the critical zone. Frontiers in Earth Science 7, 296.

Berkman, D.S., Lescano, A.G., Gilman, R.H., Lopez, S.L. and Black, M.M. 2002. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. The Lancet 359(9306), 564-571.

Berney, M., Hammes, F., Bosshard, F., Weilenmann, H.-U. and Egli, T. 2007. Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry. Appl. Environ. Microbiol. 73(10), 3283-3290.

Besmer, M.D. and Hammes, F. 2016. Short-term microbial dynamics in a drinking water plant treating groundwater with occasional high microbial loads. Water research 107, 11-18.

Bieroza, M., Baker, A. and Bridgeman, J. 2009. Relating freshwater organic matter fluorescence to organic carbon removal efficiency in drinking water treatment. Science of the Total Environment 407(5), 1765-1774.

Bieroza, M.Z. and Heathwaite, A.L. 2016a. Unravelling organic matter and nutrient biogeochemistry in groundwater-fed rivers under baseflow conditions: uncertainty in in situ high-frequency analysis. Science of the Total Environment 572, 1520-1533.

Bieroza, M.Z. and Heathwaite, A.L. 2016b. Unravelling organic matter and nutrient biogeochemistry in groundwater-fed rivers under baseflow conditions: uncertainty in in situ high-frequency analysis. Science of the Total Environment.

Blaen, P.J., Khamis, K., Lloyd, C.E., Bradley, C., Hannah, D. and Krause, S. 2016. Real-time monitoring of nutrients and dissolved organic matter in rivers: Capturing event dynamics, technological opportunities and future directions. Science of the Total Environment 569, 647-660.

Borchardt, M.A., Bradbury, K.R., Gotkowitz, M.B., Cherry, J.A. and Parker, B.L. 2007. Human enteric viruses in groundwater from a confined bedrock aquifer. Environmental science & technology 41(18), 6606-6612.

Borchardt, M.A., Haas, N.L. and Hunt, R.J. 2004. Vulnerability of drinking-water wells in La Crosse, Wisconsin, to enteric-virus contamination from surface water contributions. Appl. Environ. Microbiol. 70(10), 5937-5946.

Bridgeman, J., Baker, A., Brown, D. and Boxall, J. 2015. Portable LED fluorescence instrumentation for the rapid assessment of potable water quality. Science of the Total Environment 524, 338-346.

Bro, R. 1997. PARAFAC. Tutorial and applications. Chemometrics and intelligent laboratory systems 38(2), 149-171.

Bronk, B.V. and Reinisch, L. 1993. Variability of steady-state bacterial fluorescence with respect to growth conditions. Applied spectroscopy 47(4), 436-440.

Brown, C.M., Staley, C., Wang, P., Dalzell, B., Chun, C.L. and Sadowsky, M.J. 2017. A high-throughput DNA-sequencing approach for determining sources of fecal bacteria in a Lake Superior estuary. Environmental science & technology 51(15), 8263-8271.

Brown, J., Proum, S. and Sobsey, M. 2008. Escherichia coli in household drinking water and diarrheal disease risk: evidence from Cambodia. Water Science and Technology 58(4), 757-763.

Buckerfield, S.J., Quilliam, R.S., Waldron, S., Naylor, L.A., Li, S. and Oliver, D.M. 2019. Rainfall-driven E. coli transfer to the stream-conduit network observed through increasing spatial scales in mixed land-use paddy farming karst terrain. Water research X 5, 100038.

Cammack, W.L., Kalff, J., Prairie, Y.T. and Smith, E.M. 2004. Fluorescent dissolved organic matter in lakes: Relationships with heterotrophic metabolism. Limnology and Oceanography 49(6), 2034-2045.

Carstea, E.M., Baker, A., Bieroza, M. and Reynolds, D. 2010. Continuous fluorescence excitation–emission matrix monitoring of river organic matter. Water research 44(18), 5356-5366.

Carstea, E.M., Bridgeman, J., Baker, A. and Reynolds, D.M. 2016. Fluorescence spectroscopy for wastewater monitoring: A review. Water research 95, 205-219.
Carstea, E.M., Popa, C.L., Baker, A. and Bridgeman, J. 2020. In situ fluorescence measurements of dissolved organic matter: a review. Science of the Total Environment 699, 134361.
Chapelle, F.H. 2021. The Bioavailability of Dissolved, Particulate, and Adsorbed Organic Carbon in Groundwater Systems. Groundwater 59(2), 226-235.
Charrois, J.W. 2010. Private drinking water supplies: challenges for public health. Canadian Medical Association Journal 182(10), 1061-1064.
Chavula, G.M.S. 2012 Groundwater availability and use in Sub-Saharan Africa: a review of 15 countries. Pavelic, P., Giordano, M., Keraita, B.N., Ramesh, V. and Rao, T. (eds), International Water Management Institute (IWMI).
Cloete, T., Westaard, D. and Van Vuuren, S. 2003. Dynamic response of biofilm to pipe surface and fluid velocity. Water science and technology 47(5), 57-59.
Coble, P.G. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Marine chemistry 51(4), 325-346.
Coble, P.G., Lead, J., Baker, A., Reynolds, D.M. and Spencer, R.G. 2014 Aquatic organic matter fluorescence, Cambridge University Press.
Coble, P.G., Schultz, C.A. and Mopper, K. 1993. Fluorescence contouring analysis of DOC intercalibration experiment samples: a comparison of techniques. Marine Chemistry 41(1-3), 173-178.
Cohen, E., Levy, G.J. and Borisover, M. 2014. Fluorescent components of organic matter in wastewater: efficacy and selectivity of the water treatment. Water research 55, 323-334.
Cole, B., Pinfold, J., Ho, G. and Anda, M. 2013. Exploring the methodology of participatory design to create appropriate sanitation technologies in rural Malawi. Journal of Water, Sanitation and Hygiene for Development 4(1), 51-61.
Collier, S.A., Deng, L., Adam, E.A., Benedict, K.M., Beshearse, E.M., Blackstock, A.J., Bruce, B.B., Derado, G., Edens, C. and Fullerton, K.E. 2021. Estimate of burden and direct healthcare cost of infectious waterborne disease in the United States. Emerging infectious diseases 27(1), 140.
Consultants, D. 1988. Rural domestic water resources assessment Kisumu District. Available at: https://www.ircwash.org/sites/default/files/824-5816.pdf.
Cory, R.M. and Kaplan, L.A. 2012. Biological lability of streamwater fluorescent dissolved organic matter. Limnology and Oceanography 57(5), 1347-1360.
Cumberland, S., Bridgeman, J., Baker, A., Sterling, M. and Ward, D. 2012. Fluorescence spectroscopy as a tool for determining microbial quality in potable water applications. Environmental technology 33(6), 687-693.
Cuthbert, M.O., Taylor, R.G., Favreau, G., Todd, M.C., Shamsudduha, M., Villholth, K.G., MacDonald, A.M., Scanlon, B.R., Kotchoni, D.V. and Vouillamoz, J.-M. 2019. Observed controls on resilience of groundwater to climate variability in sub-Saharan Africa. Nature 572(7768), 230-234.
Dalterio, R., Nelson, W., Britt, D., Sperry, J., Psaras, D., Tanguay, J. and Suib, S. 1986. Steady-state and decay characteristics of protein tryptophan fluorescence from bacteria. Applied spectroscopy 40(1), 86-90.
Dalterio, R., Nelson, W., Britt, D., Sperry, J., Tanguay, J. and Suib, S. 1987. The steady-state and decay characteristics of primary fluorescence from live bacteria. Applied spectroscopy 41(2), 234-241.
Dartnell, L.R., Roberts, T.A., Moore, G., Ward, J.M. and Muller, J.-P. 2013. Fluorescence characterization of clinically-important bacteria. PloS one 8(9), e75270.
DeFelice, N.B., Johnston, J.E. and Gibson, J.M. 2016. Reducing emergency department visits for acute gastrointestinal illnesses in North Carolina (USA) by extending community water service. Environmental health perspectives 124(10), 1583-1591.

Demšar, J. 2006. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research 7, 1-30.
Determann, S., Lobbes, J.M., Reuter, R. and Rullkötter, J. 1998. Ultraviolet fluorescence excitation and emission spectroscopy of marine algae and bacteria. Marine Chemistry 62(1), 137-156.
Diedhiou, M. 2011. Approche multitraceur géochimique et isotopique à l'identification des sources de la pollution nitratée et des processus de nitrification et dénitrification dans la nappe de Thiaroye. Univ. Cheikh Anta Diop de Dakar, 210.
Diédhiou, M., Cissé Faye, S., Diouf, O., Faye, S., Faye, A., Re, V., Wohnlich, S., Wisotzky, F., Schulte, U. and Maloszewski, P. 2012. Tracing groundwater nitrate sources in the Dakar suburban area: an isotopic multi-tracer approach. Hydrological Processes 26(5), 760-770.
Dillingham, R. and Guerrant, R.L. 2004. Childhood stunting: measuring and stemming the staggering costs of inadequate water and sanitation. The Lancet 363(9403), 94-95.
Dobson, A.J. 2001 An introduction to generalized linear models, CRC press.
Downing, B.D., Pellerin, B.A., Bergamaschi, B.A., Saraceno, J.F. and Kraus, T.E. 2012. Seeing the light: The effects of particles, dissolved materials, and temperature on in situ measurements of DOM fluorescence in rivers and streams. Limnol. Oceanogr.: Methods 10, 767-775.
Draper, N.R. and Smith, H. 1981. Applied regression analysis 2nd ed.
Dunn, O.J. 1964. Multiple comparisons using rank sums. Technometrics 6(3), 241-252.
DWI 2010. Guidance on the implementation of the water supply (water quality) regulations 2000 (as amended) in England.
EA 2009 The Microbiology of Drinking Water (2009) - Part 4 - Methods for the isolation and enumeration of coliform bacteria and Escherichia coli (including E. coli O157:H7).
Eisenreich, S., Bannerman, R. and Armstrong, D. 1975. A simplified phosphorus analysis technique. Environmental letters 9(1), 43-53.
Elliott, S., Lead, J. and Baker, A. 2006a. Characterisation of the fluorescence from freshwater, planktonic bacteria. Water Research 40(10), 2075-2083.
Elliott, S., Lead, J. and Baker, A. 2006b. Thermal quenching of fluorescence of freshwater, planktonic bacteria. Analytica chimica acta 564(2), 219-225.
Esrey, S.A., Habicht, J.-P., Latham, M.C., Sisler, D.G. and Casella, G. 1988. Drinking water source, diarrheal morbidity, and child growth in villages with both traditional and improved water supplies in rural Lesotho, southern Africa. American journal of public health 78(11), 1451-1455.
Fabris, R., Chow, C.W., Drikas, M. and Eikebrokk, B. 2008. Comparison of NOM character in selected Australian and Norwegian drinking waters. Water research 42(15), 4188-4196.
Fall, M. 1986. Environnements sédimentaires quaternaires et actuels des tourbières des Niayes de la grande côte du Sénégal. Thèse de 3ème cycle UCAD, ORSTOM, 130pp.
Faye, S.C., Diongue, M., Pouye, A., Gaye, C.B., Travi, Y., Wohnlich, S., Faye, S. and Taylor, R.G. 2019. Tracing natural groundwater recharge to the Thiaroye aquifer of Dakar, Senegal. Hydrogeology Journal 27(3), 1067-1080.
Faye, S.C., Faye, S., Wohnlich, S. and Gaye, C.B. 2004. An assessment of the risk associated with urban development in the Thiaroye area (Senegal). Environmental Geology 45(3), 312-322.
Fellman, J.B., Hood, E. and Spencer, R.G. 2010. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. Limnology and Oceanography 55(6), 2452-2462.
Findlay, S., Strayer, D., Goumbala, C. and Gould, K. 1993. Metabolism of streamwater dissolved organic carbon in the shallow hyporheic zone. Limnology and oceanography 38(7), 1493-1499.
Foppen, J. and Schijven, J. 2006. Evaluation of data from the literature on the transport and survival of< i> Escherichia coli</i> and thermotolerant coliforms in aquifers under saturated conditions. Water Research 40(3), 401-426.

Foulquier, A., Mermillod-Blondin, F., Malard, F. and Gibert, J. 2011. Response of sediment biofilm to increased dissolved organic carbon supply in groundwater artificially recharged with stormwater. Journal of Soils and Sediments 11(2), 382-393.
Fox, B., Thorn, R., Anesio, A., Cox, T., Attridge, J. and Reynolds, D. 2019. Microbial processing and production of aquatic fluorescent organic matter in a model freshwater system. Water 11(1), 10.
Fox, B., Thorn, R., Anesio, A. and Reynolds, D. 2017. The in situ bacterial production of fluorescent organic matter; an investigation at a species level. Water Research.
Fox, B.G., Thorn, R. and Reynolds, D.M. 2021. Laboratory In-Situ Production of Autochthonous and Allochthonous Fluorescent Organic Matter by Freshwater Bacteria. Microorganisms 9(8), 1623.
Fox, J. and Weisberg, S. 2018 An R companion to applied regression, Sage publications.
Frank, S., Goeppert, N. and Goldscheider, N. 2017. Fluorescence-based multi-parameter approach to characterize dynamics of organic carbon, faecal bacteria and particles at alpine karst springs. Science of The Total Environment.
Frank, S., Goeppert, N. and Goldscheider, N. 2021. Field tracer tests to evaluate transport properties of tryptophan and humic acid in karst. Groundwater 59(1), 59-70.
Friendly, M. 2002. Corrgrams: Exploratory displays for correlation matrices. The American Statistician 56(4), 316-324.
Graham, J.P. and Polizzotto, M.L. 2013. Pit latrines and their impacts on groundwater quality: a systematic review. Environmental health perspectives 121(5), 521-530.
Graham, P., Baker, A. and Andersen, M. 2015. Dissolved organic carbon mobilisation in a groundwater system stressed by pumping. Scientific reports 5, 18487.
Griebler, C. and Lueders, T. 2009. Microbial biodiversity in groundwater ecosystems. Freshwater Biology 54(4), 649-677.
Hambly, A., Henderson, R., Storey, M., Baker, A., Stuetz, R. and Khan, S. 2010. Fluorescence monitoring at a recycled water treatment plant and associated dual distribution system–Implications for cross-connection detection. water research 44(18), 5323-5333.
Hambly, A.C., Henderson, R.K., Baker, A., Stuetz, R.M. and Khan, S.J. 2015. Application of portable fluorescence spectrophotometry for integrity testing of recycled water dual distribution systems. Applied spectroscopy 69(1), 124-129.
Hamilton, W.P., Kim, M. and Thackston, E.L. 2005. Comparison of commercially available Escherichia coli enumeration tests: Implications for attaining water quality standards. Water Research 39(20), 4869-4878.
Hammes, F. and Egli, T. 2010. Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. Analytical and bioanalytical chemistry 397(3), 1083-1095.
Hanley, J.A. and McNeil, B.J. 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148(3), 839-843.
Heibati, M., Stedmon, C.A., Stenroth, K., Rauch, S., Toljander, J., Säve-Söderbergh, M. and Murphy, K.R. 2017. Assessment of drinking water quality at the tap using fluorescence spectroscopy. Water Research 125, 1-10.
Heinz, B., Birk, S., Liedl, R., Geyer, T., Straub, K., Andresen, J., Bester, K. and Kappler, A. 2009. Water quality deterioration at a karst spring (Gallusquelle, Germany) due to combined sewer overflow: evidence of bacterial and micro-pollutant contamination. Environmental Geology 57(4), 797-808.
Henderson, R., Baker, A., Murphy, K., Hambly, A., Stuetz, R. and Khan, S. 2009. Fluorescence as a potential monitoring tool for recycled water systems: A review. Water research 43(4), 863-881.
Henderson, R.K., Baker, A., Parsons, S.A. and Jefferson, B. 2008. Characterisation of algogenic organic matter extracted from cyanobacteria, green algae and diatoms. Water research 42(13), 3435-3445.

Henry, J. 1972. Etude sur modèle mathématique du système aquifère de la presqu'île du Cap-Vert. Rapport final, GEOHYDRAULIQUE/OMS.
Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., Grant, T., Cottam, D., O'Brien, P. and Deletic, A. 2016. Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters. Water research 93, 242-253.
Hogg, S. 2013 Essential microbiology, John Wiley & Sons.
Hollander, M. and Wolfe, D.A. 1973. Nonparametric statistical methods John Wiley & Sons. Inc. New York.
Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X. 2013 Applied logistic regression, John Wiley & Sons.
Houston, J. 1982. Rainfall and recharge to a dolomite aquifer in a semi-arid climate at Kabwe, Zambia. Journal of Hydrology 59(1), 173-187.
Howard, G., Pedley, S., Barrett, M., Nalubega, M. and Johal, K. 2003. Risk factors contributing to microbiological contamination of shallow groundwater in Kampala, Uganda. Water research 37(14), 3421-3429.
Hrudey, S.E. and Hrudey, E.J. 2004 Safe drinking water, IWA publishing.
Hudson, N., Baker, A. and Reynolds, D. 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review. River Research and Applications 23(6), 631-649.
Hudson, N., Baker, A., Ward, D., Reynolds, D.M., Brunsdon, C., Carliell-Marquet, C. and 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 South West England. Science of the Total Environment 391(1), 149-158.
Hunt, R.J., Borchardt, M.A. and Bradbury, K.R. 2014. Viruses as Groundwater Tracers: Using Ecohydrology to Characterize Short Travel Times in Aquifers. Groundwater 52(2), 187-193.
Hunt, R.J. and Johnson, W.P. 2017. Pathogen transport in groundwater systems: contrasts with traditional solute transport. Hydrogeology Journal 25(4), 921-930.
Hur, J., Lee, B.-M., Lee, T.-H. and Park, D.-H. 2010. Estimation of biological oxygen demand and chemical oxygen demand for combined sewer systems using synchronous fluorescence spectra. Sensors 10(4), 2460-2471.
Hurst, C.J., Crawford, R.L., Garland, J.L. and Lipson, D.A. 2007 Manual of environmental microbiology, American Society for Microbiology Press.
Hynds, P.D., Misstear, B.D. and Gill, L.W. 2012. Development of a microbial contamination susceptibility model for private domestic groundwater sources. Water Resources Research 48(12).
Ignatev, A. and Tuhkanen, T. 2019. Monitoring WWTP performance using size-exclusion chromatography with simultaneous UV and fluorescence detection to track recalcitrant wastewater fractions. Chemosphere 214, 587-597.
Ikeda, M. 2006. Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. Applied microbiology and biotechnology 69(6), 615.
John, D.E. and Rose, J.B. 2005. Review of factors affecting microbial survival in groundwater. Environmental science & technology 39(19), 7345-7356.
Jung, A.-V., Le Cann, P., Roig, B., Thomas, O., Baurès, E. and Thomas, M.-F. 2014. Microbial contamination detection in water resources: interest of current optical methods, trends and needs in the context of climate change. International Journal of Environmental Research and Public Health 11(4), 4292-4310.
Kamiya, E., Izumiyama, S., Nishimura, M., Mitchell, J.G. and Kogure, K. 2007. Effects of fixation and storage on flow cytometric analysis of marine bacteria. Journal of Oceanography 63(1), 101-112.

Kelly, E.R., Cronk, R., Kumpel, E., Howard, G. and Bartram, J. 2020. How we assess water safety: A critical review of sanitary inspection and water quality analysis. Science of The Total Environment 718, 137237.
Khamis, K., Bradley, C., Gunter, H., Basevi, G., Stevens, R. and Hannah, D. 2021. Calibration of an in-situ fluorescence based sensor platform for reliable BOD5 measurement in wastewater. Water Science and Technology.
Khamis, K., Bradley, C. and Hannah, D. 2020. High frequency fluorescence monitoring reveals new insights into organic matter dynamics of an urban river, Birmingham, UK. Science of the Total Environment 710, 135668.
Khamis, K., Bradley, C., Stevens, R. and 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.
Khamis, K., Bradley, C., Stevens, R. and Hannah, D.M. 2017. Continuous field estimation of dissolved organic carbon concentration and biochemical oxygen demand using dual-wavelength fluorescence, turbidity and temperature. Hydrological Processes 31(3), 540-555.
Khamis, K., Sorensen, J., Bradley, C., Hannah, D., Lapworth, D.J. and Stevens, R. 2015. In situ tryptophan-like fluorometers: assessing turbidity and temperature effects for freshwater applications. Environmental Science: Processes & Impacts 17(4), 740-752.
Kida, M., Kojima, T., Tanabe, Y., Hayashi, K., Kudoh, S., Maie, N. and Fujitake, N. 2019. Origin, distributions, and environmental significance of ubiquitous humic-like fluorophores in Antarctic lakes and streams. Water research 163, 114901.
Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R. and Kelley, S.T. 2011. Bayesian community-wide culture-independent microbial source tracking. Nature methods 8(9), 761.
Kruskal, W.H. and Wallis, W.A. 1952. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association 47(260), 583-621.
Kwon, M.J., Sanford, R.A., Park, J., Kirk, M.F. and Bethke, C.M. 2008. Microbiological response to well pumping. Groundwater 46(2), 286-294.
Lakowicz, J.R. 2006 Principles of fluorescence spectroscopy, Springer.
Lapworth, D.J., Gooddy, D., Butcher, A. and Morris, B. 2008. Tracing groundwater flow and sources of organic carbon in sandstone aquifers using fluorescence properties of dissolved organic matter (DOM). Applied Geochemistry 23(12), 3384-3390.
Lautenschlager, K., Hwang, C., Liu, W.-T., Boon, N., Köster, O., Vrouwenvelder, H., Egli, T. and Hammes, F. 2013. A microbiology-based multi-parametric approach towards assessing biological stability in drinking water distribution networks. water research 47(9), 3015-3025.
Leclerc, H., Mossel, D., Edberg, S. and Struijk, C. 2001. Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. Annual Reviews in Microbiology 55(1), 201-234.
Lee, E.-J., Yoo, G.-Y., Jeong, Y., Kim, K.-U., Park, J.-H. and Oh, N.-H. 2015. Comparison of UV–VIS and FDOM sensors for in situ monitoring of stream DOC concentrations. Biogeosciences 12(10), 3109-3118.
Léonard, L., Bouarab Chibane, L., Ouled Bouhedda, B., Degraeve, P. and Oulahal, N. 2016. Recent advances on multi-parameter flow cytometry to characterize antimicrobial treatments. Frontiers in microbiology 7, 1225.
LgWSC 2014. Concept paper for Kabwe sanitation interventions. 10p.
Li, G. and Young, K.D. 2013. Indole production by the tryptophanase TnaA in Escherichia coli is determined by the amount of exogenous tryptophan. Microbiology 159(2), 402-410.
Li, W.-T., Jin, J., Li, Q., Wu, C.-F., Lu, H., Zhou, Q. and Li, A.-M. 2016. Developing LED UV fluorescence sensors for online monitoring DOM and predicting DBPs formation potential during water treatment. Water research 93, 1-9.

Macler, B.A. and Merkle, J.C. 2000. Current knowledge on groundwater microbial pathogens and their control. Hydrogeology Journal 8(1), 29-40.
Mandrekar, J.N. 2010. Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology 5(9), 1315-1316.
Mann, H.B. and Whitney, D.R. 1947. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 50-60.
Marie, D., Rigaut-Jalabert, F. and Vaulot, D. 2014. An improved protocol for flow cytometry analysis of phytoplankton cultures and natural samples. Cytometry Part A 85(11), 962-968.
Marmonier, P., Fontvieille, D., Gibert, J. and Vanek, V. 1995. Distribution of dissolved organic carbon and bacteria at the interface between the Rhône River and its alluvial aquifer. Journal of the North American Benthological Society 14(3), 382-392.
Martin, A. 1970 Les nappes de la presqu'île du Cap Vert (République du Sénégal): leur utilisation pour l'alimentation en eau de Dakar, Bureau de recherches géologiques et minières.
Matilainen, A., Vepsäläinen, M. and Sillanpää, M. 2010. Natural organic matter removal by coagulation during drinking water treatment: a review. Advances in colloid and interface science 159(2), 189-197.
Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L. and Linsey, K.S. 2014 Estimated use of water in the United States in 2010, US Geological Survey.
McDonough, L.K., Santos, I.R., Andersen, M.S., O'Carroll, D.M., Rutlidge, H., Meredith, K., Oudone, P., Bridgeman, J., Gooddy, D.C. and Sorensen, J.P. 2020. Changes in global groundwater organic carbon driven by climate change and urbanization. Nature Communications 11(1), 1-10.
Mendoza, L.M., Mladenov, N., Kinoshita, A.M., Pinongcos, F., Verbyla, M.E. and Gersberg, R. 2020. Fluorescence-based monitoring of anthropogenic pollutant inputs to an urban stream in Southern California, USA. Science of The Total Environment 718, 137206.
Misati, A.G., Ogendi, G., Peletz, R., Khush, R. and Kumpel, E. 2017. Can sanitary surveys replace water quality testing? Evidence from Kisii, Kenya. International Journal of Environmental Research and Public Health 14(2), 152.
Mladenov, N., Bigelow, A., Pietruschka, B., Palomo, M. and Buckley, C. 2018. Using submersible fluorescence sensors to track the removal of organic matter in decentralized wastewater treatment systems (DEWATS) in real time. Water Science and Technology 77(3), 819-828.
Moberg, L., Robertsson, G. and Karlberg, B. 2001. Spectrofluorimetric determination of chlorophylls and pheopigments using parallel factor analysis. Talanta 54(1), 161-170.
Moe, C., Sobsey, M., Samsa, G. and Mesolo, V. 1991. Bacterial indicators of risk of diarrhoeal disease from drinking-water in the Philippines. Bulletin of the World Health Organization 69(3), 305.
Mopper, K., Feng, Z., Bentjen, S.B. and Chen, R.F. 1996. Effects of cross-flow filtration on the absorption and fluorescence properties of seawater. Marine Chemistry 55(1-2), 53-74.
Muller, M., Milori, D.M.B.P., Déléris, S., Steyer, J.-P. and Dudal, Y. 2011. Solid-phase fluorescence spectroscopy to characterize organic wastes. Waste management 31(9), 1916-1923.
Murdoch, D. and Chow, E. 1996. A graphical display of large correlation matrices. The American Statistician 50(2), 178-180.
Murphy, K.R., Stedmon, C.A., Graeber, D. and Bro, R. 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. Analytical Methods 5(23), 6557-6566.
Murphy, K.R., Stedmon, C.A., Waite, T.D. and Ruiz, G.M. 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. Marine Chemistry 108(1-2), 40-58.
Myles, H. and Wolfe, D.A. 1973. Nonparametric statistical methods. Ed. John Wiley and Sons. New York, NY. 503pp.

Nahorniak, M.L. and Booksh, K.S. 2006. Excitation-emission matrix fluorescence spectroscopy in conjunction with multiway analysis for PAH detection in complex matrices. Analyst 131(12), 1308-1315.
Nakar, A., Schmilovitch, Z.e., Vaizel-Ohayon, D., Kroupitski, Y., Borisover, M. and Sela, S. 2019. Quantification of bacteria in water using PLS analysis of emission spectra of fluorescence and excitation-emission matrices. Water Research, 115197.
Nayebare, J. 2021 Loading and mobility of faecal and chemical contaminants to shallow groundwater: evidence from Lukaya Town, Central Uganda, Makerere University.
Nayebare, J., Owor, M., Kulabako, R., Campos, L., Fottrell, E. and Taylor, R. 2020. WASH conditions in a small town in Uganda: how safe are on-site facilities? Journal of Water, Sanitation and Hygiene for Development 10(1), 96-110.
Nguyen, M.-L., Westerhoff, P., Baker, L., Hu, Q., Esparza-Soto, M. and Sommerfeld, M. 2005. Characteristics and reactivity of algae-produced dissolved organic carbon. Journal of Environmental Engineering 131(11), 1574-1582.
Nkhuwa, D., Xu, Y. and Usher, B. 2006. Groundwater quality assessments in the John Laing and Misisi areas of Lusaka. Groundwater pollution in Africa, 239-251.
Ohno, T. 2002. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. Environmental science & technology 36(4), 742-746.
Okotto-Okotto, J., Okotto, L., Price, H., Pedley, S. and Wright, J. 2015. A longitudinal study of long-term change in contamination hazards and shallow well quality in two neighbourhoods of Kisumu, Kenya. International journal of environmental research and public health 12(4), 4275-4291.
Olago, D.O. 2019. Constraints and solutions for groundwater development, supply and governance in urban areas in Kenya. Hydrogeology Journal 27(3), 1031-1050.
Parlanti, E., Wörz, K., Geoffroy, L. and Lamotte, M. 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. Organic geochemistry 31(12), 1765-1781.
Patel-Sorrentino, N., Mounier, S. and Benaim, J. 2002. Excitation–emission fluorescence matrix to study pH influence on organic matter fluorescence in the Amazon basin rivers. Water Research 36(10), 2571-2581.
PDP 2017. Lukaya Town Council Physical Development Plan of Lukaya Town, Kalungu District. Final Physical Development Plan Report, 2017–2027, 256p.
Pedersen, K. 2000. Exploration of deep intraterrestrial microbial life: current perspectives. FEMS microbiology letters 185(1), 9-16.
Pohlert, T. 2014. The pairwise multiple comparison of mean ranks package (PMCMR). R package 27(2019), 9.
Pronk, M., Goldscheider, N. and Zopfi, J. 2006. Dynamics and interaction of organic carbon, turbidity and bacteria in a karst aquifer system. Hydrogeology Journal 14(4), 473-484.
Pronk, M., Goldscheider, N., Zopfi, J. and Zwahlen, F. 2009. Percolation and particle transport in the unsaturated zone of a karst aquifer. Groundwater 47(3), 361-369.
Prüss-Ustün, A., Bartram, J., Clasen, T., Colford, J.M., Cumming, O., Curtis, V., Bonjour, S., Dangour, A.D., De France, J. and Fewtrell, L. 2014. Burden of disease from inadequate water, sanitation and hygiene in low-and middle-income settings: a retrospective analysis of data from 145 countries. Tropical Medicine & International Health 19(8), 894-905.
R Core Team 2020 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.
Re, V., Faye, S.C., Faye, A., Faye, S., Gaye, C.B., Sacchi, E. and Zuppi, G.M. 2011. Water quality decline in coastal aquifers under anthropic pressure: the case of a suburban area of Dakar (Senegal). Environmental monitoring and assessment 172(1-4), 605-622.
Reynolds, D. and Ahmad, S. 1997. Rapid and direct determination of wastewater BOD values using a fluorescence technique. Water Research 31(8), 2012-2018.

Reynolds, D.M. 2002. The differentiation of biodegradable and non-biodegradable dissolved organic matter in wastewaters using fluorescence spectroscopy. Journal of Chemical Technology and Biotechnology 77(8), 965-972.
Reynolds, D.M. 2003. Rapid and direct determination of tryptophan in water using synchronous fluorescence spectroscopy. Water Research 37(13), 3055-3060.
Reynolds, D.M. 2014 Aquatic organic matter fluorescence. Andy Baker, D.M. and Reynolds, J.L., Paula G Coble, and Robert G M Spencer (eds), pp. 3-34, Cambridge University Press.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. and Müller, M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC bioinformatics 12(1), 77.
Roth, V.-N., Lange, M., Simon, C., Hertkorn, N., Bucher, S., Goodall, T., Griffiths, R.I., Mellado-Vázquez, P.G., Mommer, L. and Oram, N.J. 2019. Persistence of dissolved organic matter explained by molecular changes during its passage through soil. Nature Geoscience 12(9), 755-761.
Royston, J.P. 1982. An extension of Shapiro and Wilk's W test for normality to large samples. Journal of the Royal Statistical Society: Series C (Applied Statistics) 31(2), 115-124.
Ruhala, S.S. and Zarnetske, J.P. 2017. Using in-situ optical sensors to study dissolved organic carbon dynamics of streams and watersheds: A review. Science of the Total Environment 575, 713-723.
Safford, H.R. and Bischel, H.N. 2019. Flow cytometry applications in water treatment, distribution, and reuse: A review. Water research 151, 110-133.
Saraceno, J.F., Pellerin, B.A., Downing, B.D., Boss, E., Bachand, P.A. and Bergamaschi, B.A. 2009. High-frequency in situ optical measurements during a storm event: Assessing relationships between dissolved organic matter, sediment concentrations, and hydrologic processes. Journal of Geophysical Research: Biogeosciences (2005–2012) 114(G4).
Saraceno, J.F., Shanley, J.B., Downing, B.D. and Pellerin, B.A. 2017. Clearing the waters: Evaluating the need for site-specific field fluorescence corrections based on turbidity measurements. Limnology and Oceanography: Methods 15(4), 408-416.
Savichtcheva, O. and Okabe, S. 2006. Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Research 40(13), 2463-2476.
Seaver, M., Roselle, D.C., Pinto, J.F. and Eversole, J.D. 1998. Absolute emission spectra from Bacillus subtilis and Escherichia coli vegetative cells in solution. Applied optics 37(22), 5344-5347.
Shen, Y., Chapelle, F.H., Strom, E.W. and Benner, R. 2015. Origins and bioavailability of dissolved organic matter in groundwater. Biogeochemistry 122(1), 61-78.
Shutova, Y., Baker, A., Bridgeman, J. and Henderson, R. 2016. On-line monitoring of organic matter concentrations and character in drinking water treatment systems using fluorescence spectroscopy. Environmental Science: Water Research & Technology 2(4), 749-760.
Sihan, P., Xiaowen, C., Chenyue, C., Zhuo, C., NGO, H.H., Qi, S., Wenshan, G. and Hong-Ying, H. 2021. Fluorescence analysis of centralized water supply systems: indications for rapid cross-connection detection and water quality safety guarantee. Chemosphere, 130290.
Simões, J., Yang, Z. and Dong, T. 2021. An ultrasensitive fluorimetric sensor for pre-screening of water microbial contamination risk. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 258, 119805.
Smith, R.J., Paterson, J.S., Sibley, C.A., Hutson, J.L. and Mitchell, J.G. 2015. Putative effect of aquifer recharge on the abundance and taxonomic composition of endemic microbial communities. PloS one 10(6).
Smith, S., Elliot, A., Mallaghan, C., Modha, D., Hippisley-Cox, J., Large, S., Regan, M. and Smith, G. 2010. Value of syndromic surveillance in monitoring a focal waterborne outbreak due to an unusual Cryptosporidium genotype in Northamptonshire, United Kingdom, June–July 2008. Eurosurveillance 15(33), 19643.

Snyder, L., Potter, J.D. and McDowell, W.H. 2018. An evaluation of nitrate, fDOM, and turbidity sensors in New Hampshire streams. Water Resources Research 54(3), 2466-2479.
Sobczak, W.V. and Findlay, S. 2002. Variation in bioavailability of dissolved organic carbon among stream hyporheic flowpaths. Ecology 83(11), 3194-3209.
Sohn, M., Himmelsbach, D.S., Barton, F.E. and Fedorka-Cray, P.J. 2009. Fluorescence spectroscopy for rapid detection and classification of bacterial pathogens. Applied spectroscopy 63(11), 1251-1255.
Sorensen, J., Lapworth, D., Marchant, B., Nkhuwa, D., Pedley, S., Stuart, M., Bell, R., Chirwa, M., Kabika, J. and Liemisa, M. 2015a. In-situ tryptophan-like fluorescence: a real-time indicator of faecal contamination in drinking water supplies. Water research 81, 38-46.
Sorensen, J., Lapworth, D., Read, D., Nkhuwa, D., Bell, R., Chibesa, M., Chirwa, M., Kabika, J., Liemisa, M. and Pedley, S. 2015b. Tracing enteric pathogen contamination in sub-Saharan African groundwater. Science of the Total Environment 538, 888-895.
Sorensen, J., Sadhu, A., Sampath, G., Sugden, S., Gupta, S.D., Lapworth, D., Marchant, B. and Pedley, S. 2016. Are sanitation interventions a threat to drinking water supplies in rural India? An application of tryptophan-like fluorescence. Water research 88, 923-932.
Sorensen, J., Vivanco, A., Ascott, M., Gooddy, D., Lapworth, D., Read, D., Rushworth, C., Bucknall, J., Herbert, K. and Karapanos, I. 2018a. Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. Water research 137, 301-309.
Sorensen, J.P., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G. and Ward, J.S. 2018b. Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values. Science of the Total Environment 622, 1250-1257.
Sorensen, J.P., Carr, A.F., Nayebare, J., Diongue, D.M., Pouye, A., Roffo, R., Gwengweya, G., Ward, J.S., Kanoti, J. and Okotto-Okotto, J. 2020a. Tryptophan-like and humic-like fluorophores are extracellular in groundwater: implications as real-time faecal indicators. Scientific reports 10(1), 1-9.
Sorensen, J.P., Diaw, M.T., Pouye, A., Roffo, R., Diongue, D.M., Faye, S.C., Gaye, C.B., Fox, B.G., Goodall, T. and Lapworth, D.J. 2020b. In-situ fluorescence spectroscopy indicates total bacterial abundance and dissolved organic carbon. Science of The Total Environment, 139419.
Sorensen, J.P., Maurice, L., Edwards, F.K., Lapworth, D.J., Read, D.S., Allen, D., Butcher, A.S., Newbold, L.K., Townsend, B.R. and Williams, P.J. 2013. Using boreholes as windows into groundwater ecosystems. PloS one 8(7), e70264.
Sorensen, J.P.R., Lapworth, D.J., Nkhuwa, D.C.W., Stuart, M.E., Gooddy, D.C., Bell, R.A., Chirwa, M., Kabika, J., Liemisa, M., Chibesa, M. and Pedley, S. 2015c. Emerging contaminants in urban groundwater sources in Africa. Water Research 72, 51-63.
Spearman, C. 1904. The proof and measurement of association between two things. The American journal of psychology 15(1), 72-101.
Spencer, R.G., Aiken, G.R., Wickland, K.P., Striegl, R.G. and Hernes, P.J. 2008. Seasonal and spatial variability in dissolved organic matter quantity and composition from the Yukon River basin, Alaska. Global Biogeochemical Cycles 22(4).
Spencer, R.G., Bolton, L. and Baker, A. 2007. Freeze/thaw and pH effects on freshwater dissolved organic matter fluorescence and absorbance properties from a number of UK locations. Water research 41(13), 2941-2950.
Stedmon, C.A. and Bro, R. 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. Limnology and Oceanography: Methods 6(572 - 579).
Stedmon, C.A. and Markager, S. 2005. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. Limnology and Oceanography 50(2), 686-697.

Stedmon, C.A., Markager, S. and Bro, R. 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Mar. Chem. 82, 239 - 254.
Stedmon, C.A., Seredyńska-Sobecka, B., Boe-Hansen, R., Le Tallec, N., Waul, C.K. and Arvin, E. 2011. A potential approach for monitoring drinking water quality from groundwater systems using organic matter fluorescence as an early warning for contamination events. Water research 45(18), 6030-6038.
Stelma Jr, G.N. and Wymer, L.J. 2012. Research considerations for more effective groundwater monitoring. Journal of water and health 10(4), 511-521.
Stokes, G.G. 1852. On the change of refrangibility of light. Philosophical transactions of the Royal Society of London (142), 463-562.
Stryer, L. 1968. Fluorescence spectroscopy of proteins. Science 162(3853), 526-533.
Tabak, M., Sartor, G. and Cavatorta, P. 1989. On the interactions of metal ions with tryptophan and glyciltryptophan: a fluorescence study. Journal of Luminescence 43(6), 355-361.
Taylor, R., Cronin, A., Pedley, S., Barker, J. and Atkinson, T. 2004. The implications of groundwater velocity variations on microbial transport and wellhead protection–review of field evidence. FEMS Microbiology Ecology 49(1), 17-26.
Thompson, K.A. and Dickenson, E.R. 2021. Using machine learning classification to detect simulated increases of de facto reuse and urban stormwater surges in surface water. Water Research 204, 117556.
Tunaley, C., Tetzlaff, D., Lessels, J. and Soulsby, C. 2016. Linking high-frequency DOC dynamics to the age of connected water sources. Water Resources Research 52(7), 5232-5247.
UBOS 2014. Central Region - Parish Level Profiles (Census 2014). Available from: https://www.ubos.org/explore-statistics/20/.
UKWIR 2012. Turbidity in groundwater: understanding cause, effect and mitigation measures. Report Ref. No. 12/DW/14/5.
UNICEF/WHO 2017 Target Product Profile: Rapid E. coli Detection. Version 2.0.
Valenzuela, M., Lagos, B., Claret, M., Mondaca, M.A., Pérez, C. and Parra, O. 2009. Fecal contamination of groundwater in a small rural dryland watershed in central Chile. Chilean Journal of Agricultural Research 69(2), 235-243.
van Driezum, I.H., Chik, A.H., Jakwerth, S., Lindner, G., Farnleitner, A.H., Sommer, R., Blaschke, A.P. and Kirschner, A.K. 2018. Spatiotemporal analysis of bacterial biomass and activity to understand surface and groundwater interactions in a highly dynamic riverbank filtration system. Science of the Total Environment 627, 450-461.
Ward, J.S., Lapworth, D.J., Read, D.S., Pedley, S., Banda, S.T., Monjerezi, M., Gwengweya, G. and MacDonald, A.M. 2020. Large-scale survey of seasonal drinking water quality in Malawi using in situ tryptophan-like fluorescence and conventional water quality indicators. Science of The Total Environment 744, 140674.
Ward, J.S., Lapworth, D.J., Read, D.S., Pedley, S., Banda, S.T., Monjerezi, M., Gwengweya, G. and MacDonald, A.M. 2021. Tryptophan-like fluorescence as a high-level screening tool for detecting microbial contamination in drinking water. Science of The Total Environment 750, 141284.
Watras, C., Hanson, P., Stacy, T., Morrison, K., Mather, J., Hu, Y.H. and Milewski, P. 2011. A temperature compensation method for CDOM fluorescence sensors in freshwater. Limnology and Oceanography: Methods 9(7), 296-301.
Weaver, L., Sinton, L.W., Pang, L., Dann, R. and Close, M. 2013. Transport of microbial tracers in clean and organically contaminated silica sand in laboratory columns compared with their transport in the field. Science of the total environment 443, 55-64.
WHO 1997. Guidelines for Drinking-Water Quality (2nd Edition)

WHO 2017a. Drinking Water Parameter Cooperation Project: Support to the revision of Annex I Council Directive 98/83/EC on the Quality of Water Intended for Human Consumption (Drinking Water Directive). WHO: Geneva, Switzerland.
WHO 2017b. Guidelines for drinking-water quality (4th edition).
WHO 2017c. Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.
WHO 2017d. Water quality and health-review of turbidity: information for regulators and water suppliers.
WHO 2019 Drinking-water. https://www.who.int/news-room/fact-sheets/detail/drinking-water.
WHO 2020. Sanitary Inspection Form. Available from https://www.who.int/water_sanitation_health/water-quality/safety-planning/sanitary-inspection-packages-for-drinking-water/en/.
Wilson, H.F., Saiers, J.E., Raymond, P.A. and Sobczak, W.V. 2013. Hydrologic drivers and seasonality of dissolved organic carbon concentration, nitrogen content, bioavailability, and export in a forested New England stream. Ecosystems 16(4), 604-616.
Worthington, S.R. and Smart, C.C. 2017. Transient bacterial contamination of the dual-porosity aquifer at Walkerton, Ontario, Canada. Hydrogeology Journal 25(4), 1003-1016.
Wright, J.A., Cronin, A., Okotto-Okotto, J., Yang, H., Pedley, S. and Gundry, S.W. 2013. A spatial analysis of pit latrine density and groundwater source contamination. Environmental monitoring and assessment 185(5), 4261-4272.
Wu, J., Long, S., Das, D. and Dorner, S. 2011. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. Journal of water and health 9(2), 265-278.
Yamashita, Y. and Tanoue, E. 2003. Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids. Marine Chemistry 82(3), 255-271.
Yang, L., Hur, J. and Zhuang, W. 2015a. Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: a review. Environmental Science and Pollution Research 22(9), 6500-6510.
Yang, L., Kim, D., Uzun, H., Karanfil, T. and Hur, J. 2015b. Assessing trihalomethanes (THMs) and N-nitrosodimethylamine (NDMA) formation potentials in drinking water treatment plants using fluorescence spectroscopy and parallel factor analysis. Chemosphere 121, 84-91.
Yang, Y., Ma, X., Yang, X. and Xu, H. 2018. Influence of heavy metal ions on the spectra and charge characteristics of DOM of municipal sewage secondary effluent. Water Science and Technology 77(4), 1098-1106.
Yang, Y., Yuan, X., Deng, Y., Xie, X., Gan, Y. and Wang, Y. 2020. Seasonal dynamics of dissolved organic matter in high arsenic shallow groundwater systems. Journal of Hydrology 589, 125120.
Zamyadi, A., Choo, F., Newcombe, G., Stuetz, R. and Henderson, R.K. 2016. A review of monitoring technologies for real-time management of cyanobacteria: Recent advances and future direction. TrAC Trends in Analytical Chemistry 85, 83-96.
Zheng, Y., He, W., Li, B., Hur, J., Guo, H. and Li, X. 2020. Refractory Humic-like Substances: Tracking Environmental Impacts of Anthropogenic Groundwater Recharge. Environmental Science & Technology 54(24), 15778-15788.