

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

In-situ fluorescence spectroscopy indicates total bacterial abundance and dissolved organic carbon

James P.R. Sorensen^{a,b,*}, Mor Talla Diaw^c, Abdoulaye Pouye^c, Raphaëlle Roffo^b, Djim M.L. Diongue^c, Seynabou C. Faye^c, Cheikh B. Gaye^c, Bethany G. Fox^d, Timothy Goodall^e, Daniel J. Lapworth^a, Alan M. MacDonald^f, Daniel S. Read^e, Lena Ciric^g, Richard G. Taylor^b

^a British Geological Survey, Maclean Building, Wallingford OX10 8BB, UK

^b Department of Geography, University College London, London WC1E 6BT, UK

^c Department of Geology, Université Cheikh Anta Diop, Senegal

^d Centre for Research in Biosciences, University of the West of England, Bristol BS16 1QY, UK

^e UK Centre for Ecology & Hydrology (UKCEH), Maclean Building, Wallingford OX10 8BB, UK

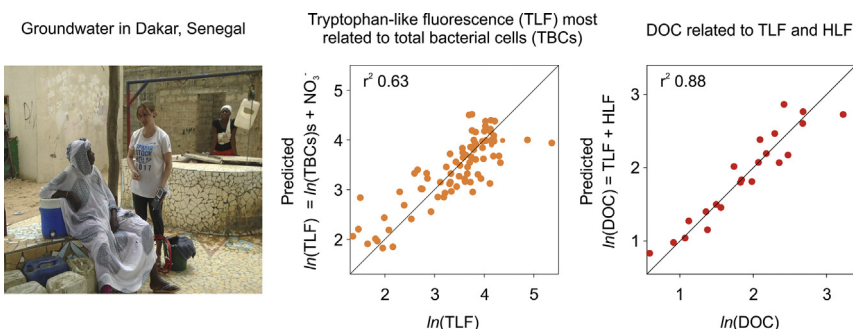
^f British Geological Survey, Lyell Centre, Research Avenue South, Edinburgh EH14 4AP, UK

^g Department of Civil, Environmental & Geomatic Engineering, University College London, London WC1E 6BT, UK

HIGHLIGHTS

- Total bacterial cells most related variable to tryptophan-like fluorescence (TLF)
- TLF and humic-like fluorescence strongly correlate with dissolved organic carbon.
- Thermotolerant coliforms are not strongly related to other variables.
- TLF and HLF relate to faecal contamination.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 18 March 2020

Received in revised form 11 May 2020

Accepted 11 May 2020

Available online 19 May 2020

Editor: Damia Barcelo

Keywords:

Fluorescence spectroscopy

Faecal contamination

Drinking water

Flow cytometry

Groundwater

Total bacterial cells

ABSTRACT

We explore *in-situ* fluorescence spectroscopy as an instantaneous indicator of total bacterial abundance and faecal contamination in drinking water. Eighty-four samples were collected outside of the recharge season from groundwater-derived water sources in Dakar, Senegal. Samples were analysed for tryptophan-like (TLF) and humic-like (HLF) fluorescence *in-situ*, total bacterial cells by flow cytometry, and potential indicators of faecal contamination such as thermotolerant coliforms (TTCs), nitrate, and in a subset of 22 samples, dissolved organic carbon (DOC). Significant single-predictor linear regression models demonstrated that total bacterial cells were the most effective predictor of TLF, followed by on-site sanitation density; TTCs were not a significant predictor. An optimum multiple-predictor model of TLF incorporated total bacterial cells, nitrate, nitrite, on-site sanitation density, and sulphate (r^2 0.68). HLF was similarly related to the same parameters as TLF, with total bacterial cells being the best correlated (ρ_s 0.64). In the subset of 22 sources, DOC clustered with TLF, HLF, and total bacterial cells, and a linear regression model demonstrated HLF was the best predictor of DOC (r^2 0.84). The intergranular nature of the aquifer, timing of the study, and/or non-uniqueness of the signal to TTCs can explain the significant associations between TLF/HLF and indicators of faecal contamination such as on-site sanitation density and nutrients but not TTCs. The bacterial population that relates to TLF/HLF is likely to be a subsurface community that develops *in-situ* based on the availability of organic matter originating from faecal sources. *In-situ*

* Corresponding author at: British Geological Survey, Maclean Building, Wallingford OX10 8BB, UK.
E-mail address: jare1@bgs.ac.uk (J.P.R. Sorensen).

fluorescence spectroscopy instantly indicates a drinking water source is impacted by faecal contamination but it remains unclear how that relates specifically to microbial risk in this setting.

© 2020 United Kingdom Research and Innovation, as represented by its component body, the British Geological Survey. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Natural organic matter (OM) is the assortment of organic, carbonaceous materials that are ubiquitous in hydrological systems (Matilainen et al., 2010). Aquatic OM is traditionally separated into that transported into a system (allochthonous) and that produced *in-situ* (autochthonous) (Coble et al., 2014; Stedmon et al., 2007). The amount and characteristics of OM in water are related to soil, geology, hydrology, land use, climate, and biogeochemical processes (Chapelle et al., 2009; Fellman et al., 2011; Larsen et al., 2010; Thottathil et al., 2013; Williams et al., 2010) and can be characterised using fluorescence spectroscopy (Carstea et al., 2016; Fellman et al., 2010; Hudson et al., 2007).

Fluorescence spectroscopy is a rapid, reagentless technique that can now be undertaken *in-situ* using portable fluorimeters (Carstea et al., 2019; Sorensen et al., 2018b). These fluorimeters are typically configured to target specific excitation-emission wavelength ($\lambda_{ex}/\lambda_{em}$) pairs associated with either tryptophan-like (TLF) or humic-like fluorescence (HLF) (e.g. Li et al., 2016). TLF, at $\lambda_{ex}/\lambda_{em}$ of 280/350 nm, is the most commonly detected protein-like fluorescence in natural waters (Yang et al., 2015). TLF is often autochthonous, elevated in human and animal waste (Baker, 2001, 2002), and is considered an indicator of biological activity (Cammack et al., 2004; Elliott et al., 2006; Reynolds and Ahmad, 1997). HLF is typically associated with more complex allochthonous material that derives from the breakdown of vegetation, and is the dominant fluorophore in freshwater (Coble et al., 2014; Fellman et al., 2010). Studies in marine environments (Romera-Castillo et al., 2011; Shimotori et al., 2012), soils (Kallenbach et al., 2016), and the laboratory (Fox et al., 2017) also show HLF can be microbially derived.

Fluorescence spectroscopy has been proposed as a surrogate indicator of many environment variables in freshwater, including biological oxygen demand, dissolved organic carbon (DOC), and other nutrients (Baker and Inverarity, 2004; Henderson et al., 2009; Hudson et al., 2008). More recently, it has been suggested the technique could be a rapid indicator of faecal contamination (Stedmon et al., 2011), the most widespread health risk associated with drinking water (WHO, 2011). TLF has been associated with the presence/absence and enumeration of thermotolerant (faecal) coliforms (TTCs), including *Escherichia coli*, in groundwater and surface water (Baker et al., 2015; Frank et al., 2017; Mendoza et al., 2020; Sorensen et al., 2015a,b, 2016, 2018a,b). HLF has also been similarly or better correlated than TLF to *E. coli* enumeration in UK public water supply boreholes (Sorensen et al., 2018a), Austrian springs (Frank et al., 2017), and controlled culture experiments (Fox et al., 2017).

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, 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., 2006; Fox et al., 2017; Nakar et al., 2020). Therefore, elevated TLF/HLF may be indicative of elevated total bacterial cells (TBCs), which are not necessarily an indicator of risk to human health, as opposed to purely TTCs. Previously, Bridgeman et al. (2015) demonstrated a correlation between TLF and TBCs, but not *E. coli*, in predominantly treated drinking water. Further, Sorensen et al. (2018a) correlated TLF/HLF with TBCs, as well as *E. coli*, in untreated groundwater from four UK public water supply boreholes.

Here, we investigate both TLF and HLF as indicators of total bacterial cells (TBCs) as well as common parameters associated with faecal contamination such as TTCs, nitrate and dissolved organic carbon, across a wide-range 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.

2. Methods

2.1. Study area

The unconfined Thiaroye aquifer beneath a suburb of Dakar is located on the Cap-Vert Peninsula (Fig. 1). The aquifer comprises Quaternary fine- and medium-grained aeolian sands overlying low-permeability Tertiary marl deposits (Faye et al., 2004, 2019), 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 (^3H), stable isotope ratios of O ($\delta^{18}\text{O}$) and H ($\delta^2\text{H}$), 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., 2004, 2019). 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 ($\delta^{15}\text{N}$ and $\delta^{18}\text{O}$) 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 (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. (submitted to journal) 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.

2.2. Groundwater sampling and analysis

2.2.1. Sampling sites

Eighty-four groundwater samples were collected from 70 water sources between 29th May and 3rd July 2018 (Fig. 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

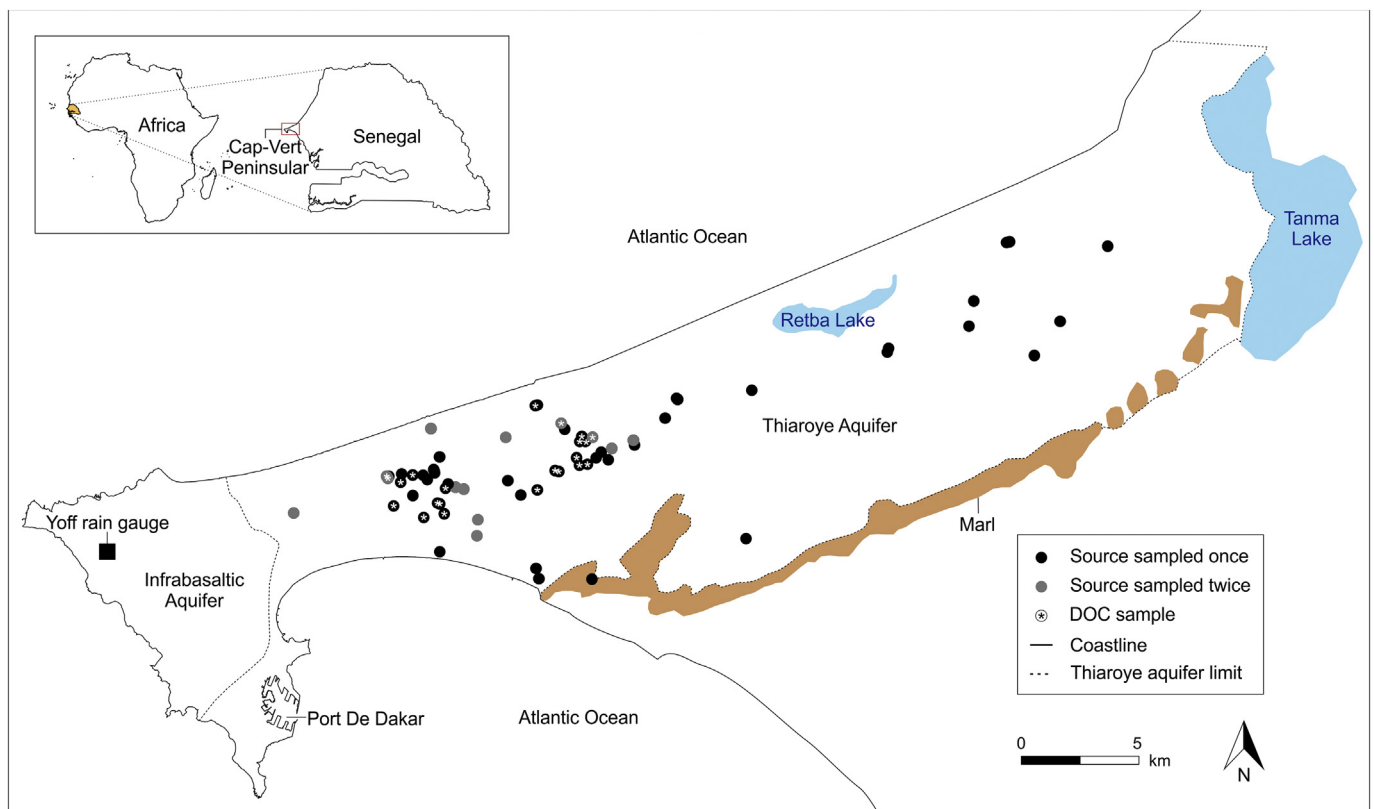


Fig. 1. Location of Thiaroye aquifer on the Cap-Vert Peninsula and sources sampled in this study.

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 12 V submersible WaSP-P5 pump (*In-Situ* Europe, Redditch, UK) at hand dug-wells, and an MP1 pump (Grundfos, Bjerringbro, Denmark) set at around 1 m³/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.

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 100 × 100 m grid of the entire Thiaroye aquifer previously devised by Diaw et al. (submitted to journal). 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).

2.2.2. In-situ analysis

TLF was determined on unfiltered samples using a portable UviLux fluorimeter targeting the excitation-emission peak at λ_{ex} 280 ± 15 nm and λ_{em} 360 ± 27.5 nm (Chelsea Technologies Group Ltd., UK).

The minimum detection limit for the fluorimeter is 0.17 ± 0.18 mg/L (3σ) dissolved tryptophan (Khamis et al., 2015). HLF was measured using a UviLux fluorimeter configured at an excitation-emission peak at λ_{ex} 280 ± 15 nm and λ_{em} 450 ± 27.5 nm (Chelsea Technologies Group Ltd., UK). This fluorimeter did not target the centres of the HLF peaks but was aligned at the same excitation as TLF because of the optical overlap between TLF and HLF regions. 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).

2.2.3. 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. TTCs were isolated and enumerated using the membrane filtration method with Membrane Lauryl Sulphate Broth (MLSB, Oxoid Ltd., UK) as the selective medium. Between 1 and 100 mL of sample was passed through a 0.45 µm cellulose nitrate membrane (GE Whatman®, UK) to ensure colonies were not too numerous to count, whilst maintaining a limit of detection of 1 cfu/100 mL. The membrane was placed on an absorbent pad (Pall Gelman, Germany) saturated with MLSB broth in an aluminium petri dish and incubated at 44 °C for 18–23 h. Subsequently, all cream to yellow colonies >1 mm counted as TTCs.

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 $-18\text{ }^{\circ}\text{C}$ 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 $0.45\text{ }\mu\text{m}$ 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 $0.22\text{ }\mu\text{m}$ 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 $4\text{ }^{\circ}\text{C}$ between collection and analysis.

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 Fig. S1.

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).

3. Results

3.1. Faecal contamination in the Thiaroye aquifer

3.1.1. Water quality and sanitary risks

Groundwater beneath Dakar displays clear evidence of faecal contamination (Fig. 2). TTCs are present in 80% of samples with a median count of 115 cfu/100 mL. 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 $28.0\text{ }^{\circ}\text{C}$ with variations limited to $2.2\text{ }^{\circ}\text{C}$ (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.

3.1.2. Predicting thermotolerant coliform counts

There is no relationship between fluorescent OM fractions and TTCs (Table 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

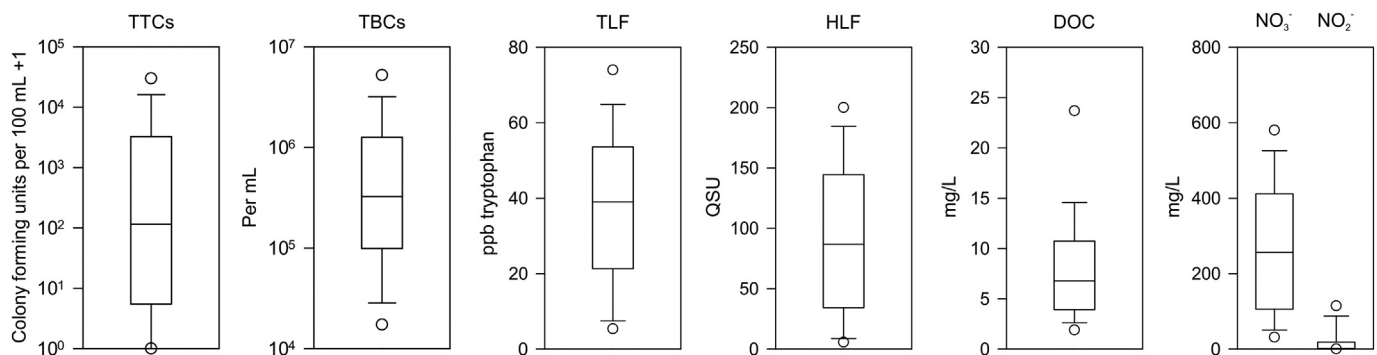


Fig. 2. Boxplots of TTCs, TBCs, TLF, HLF, DOC, NO_3^- and NO_2^- . Box boundaries illustrate the 25th and 75th percentiles, the line within the box is the median, the whiskers are the 10th and 90th percentiles, and the circles are the 5th and 95th percentiles ($n = 84$).

Table 1

Estimated coefficients and p-values for single predictor linear regression models of $\ln(\text{TTCs} + 1)$. Predictors are ordered by coefficient.

Predictor	Adj. beta	p-Value
$\ln(\text{turbidity})$	0.276	0.011
$\ln(\text{NO}_2^- + 0.1)$	0.224	0.041
$\ln(\text{SEC})$	0.211	0.054
OSS density	-0.182	0.097
$\ln(\text{Cl}^-)$	0.173	0.115
Latrines	-0.153	0.165
Population density	-0.142	0.197
NO_3^-	0.102	0.356
$\ln(\text{TBCs})$	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
$\ln(\text{TLF} + 1)$	0.029	0.793
Temperature	-0.004	0.971
SO_4^{2-}	0.004	0.974

(Fig. 2, thresholds from (Sorensen et al., 2018b)). TTC counts are, in contrast, comparably low (Fig. 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 1). The lack of relationships between TTCs and TBCs is also notable.

3.1.3. 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 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, Fig. 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^- (Fig. 3).

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) (Eq. (1)). HLF is the most important predictor in the model: being added first and having the

Table 2

Estimated coefficients and p-values for single predictor linear regression models of $\ln(\text{TLF})$. Predictors are ordered by coefficient.

Predictor	Adj. beta	p-Value
HLF	0.767	<0.001
$\ln(\text{TBCs})$	0.647	<0.001
OSS density	0.553	<0.001
$\ln(\text{SEC})$	0.533	<0.001
$\ln(\text{NO}_2^- + 0.1)$	0.531	<0.001
SO_4^{2-}	0.521	<0.001
Population density	0.430	<0.001
$\ln(\text{Cl}^-)$	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
$\ln(\text{turbidity})$	-0.049	0.657
$\ln(\text{TTCs} + 1)$	0.029	0.793
pH	0.011	0.920
Latrines	0.010	0.928

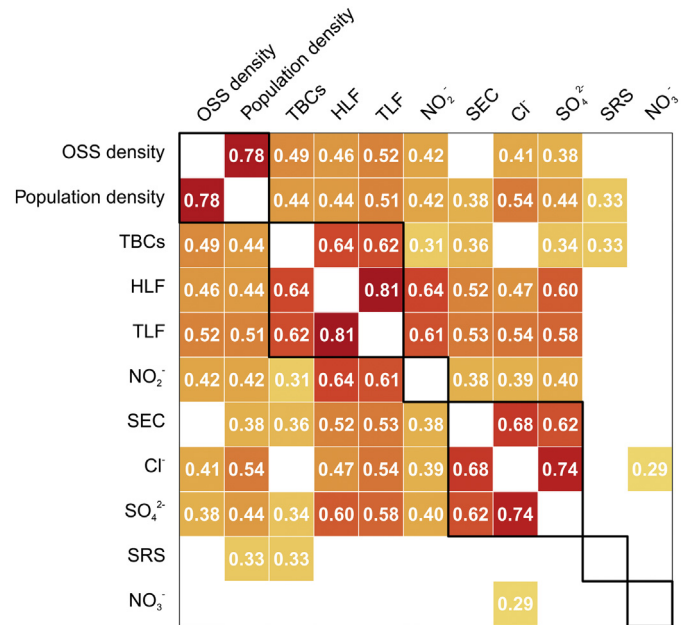


Fig. 3. Correlation matrix of TLF and all significant single-predictors of TLF. Only significant (p-value < 0.01) Spearman's ρ are shown. Variables are ordered by hierarchical clustering and black squares enclose six clusters. If the number of clusters is increased to seven, SEC forms a cluster of its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

greatest β . The additional predictors improve both the RMSE and r^2 to final values of 0.40 and 0.74, respectively (Table 3, Fig. 4A).

$$\ln(\text{TLF}) = 0.480 \text{ HLF} + 0.335 \text{ NO}_3^- + 0.306 \ln(\text{TBCs}) + 0.196 \text{ OSS density} + 0.401 \tag{1}$$

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) (Eq. (2)). The final model has an r^2 of 0.68 and an RMSE of 0.44 (Table 3, Fig. 4B). Eq. (2) predictions are similar to Eq. (1) at lower TLF with increasing differences between predictions as TLF increases (Fig. 4C).

$$\ln(\text{TLF}) = 0.489 \ln(\text{TBCs}) + 0.304 \text{ NO}_3^- + 0.205 \ln(\text{NO}_2^- + 0.1) + 0.187 \text{ OSS density} + 0.176 \text{ SO}_4^{2-} - 0.376 \tag{2}$$

Both multiple linear regression models have Gaussian residuals (Fig. S2, Shapiro-Wilk, p-values = 0.103–0.330), no systematic spatial patterns in model residuals (Fig. S3), and no evidence of multicollinearity (VIF < 2.06, Table S1). The models are robust to the two highest TLF values; if we consider them outliers, identical

Table 3

Stepwise linear regression model RMSE and r^2 following the addition of each predictor with $\ln(\text{TLF})$ as the dependent variable.

Predictor	RMSE	r^2
Model 1		
HLF	0.52	0.58
NO_3^-	0.48	0.64
$\ln(\text{TBCs})$	0.42	0.72
OSS density	0.40	0.74
Model 2		
$\ln(\text{TBCs})$	0.61	0.41
NO_3^-	0.52	0.56
$\ln(\text{NO}_2^- + 0.1)$	0.48	0.63
OSS density	0.46	0.66
SO_4^{2-}	0.44	0.68

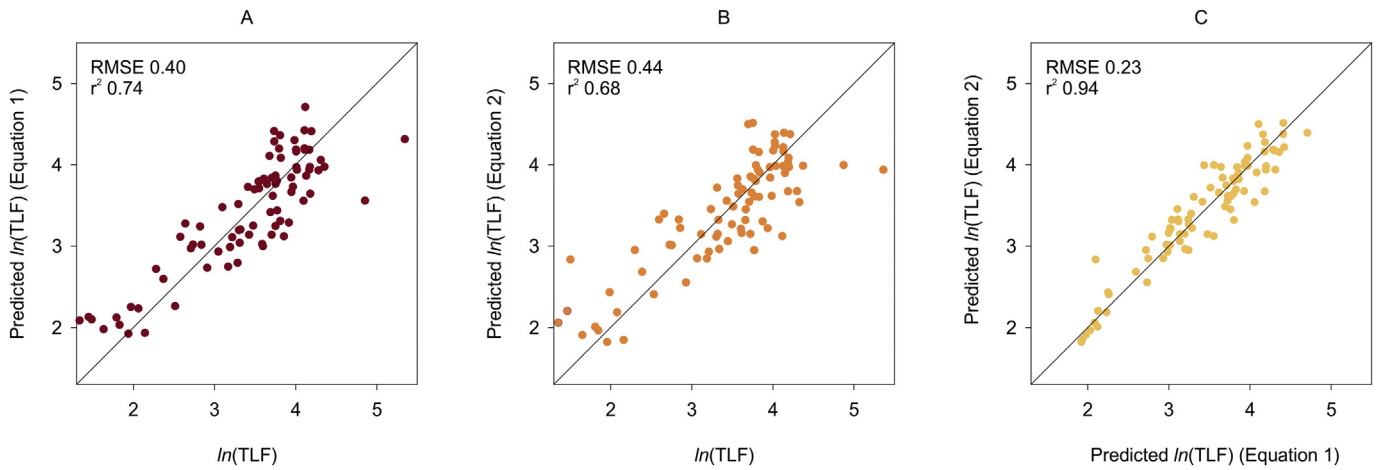


Fig. 4. Scatterplots of observed against predicted ln(TLF) using (A) Eq. (1) and (B) Eq. (2), and (C) a scatterplot comparison of predicted ln(TLF) from both equations. A 1:1 line is shown in all plots.

predictors enter in the same order producing marginal improvements in the RMSE and r^2 (Eqs. (S1) and (S2)). 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.

3.1.4. Relationships within DOC

Significant predictors of ln(TLF) are generally also predictors of ln(DOC) for the subset of 22 sources (Table S2). 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 (Fig. 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 (Fig. 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.

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 (Fig. 6), with the final model (Eq. (3)) 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) (Eq. (4)), with an RMSE and r^2 of 0.28 and 0.71, respectively. The unstandardized model coefficient is 0.712.

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

$$\ln(TLF) = 0.853 \ln(DOC) + 2.211 \quad (4)$$

Both multiple linear regression models (Eqs. (3) and (4)) have Gaussian residuals (Fig. S2, Shapiro-Wilk, p -values = 0.189–0.642) and Eq. (3) has no evidence of multicollinearity (VIF < 2.76, Table S1).

4. Discussion

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 (Sorensen et al., 2015a, 2016, 2018a). 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

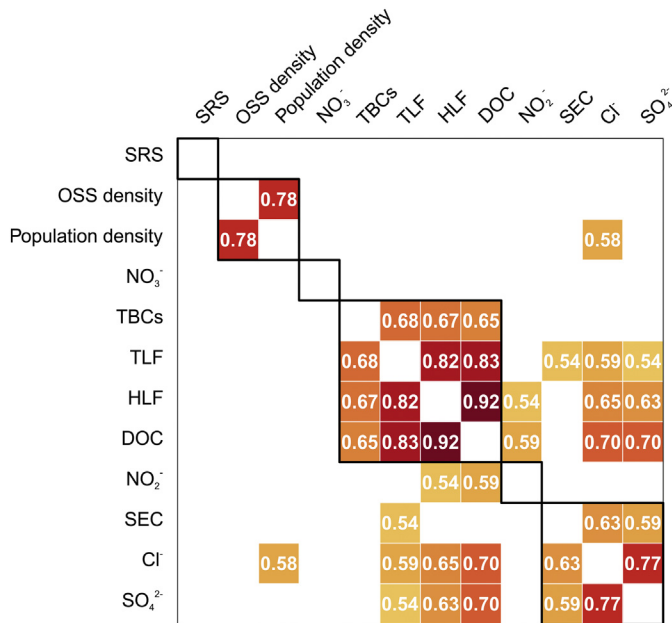


Fig. 5. Correlation matrix of DOC, TLF, and all significant single-predictors of TLF identified in Table 2 for the subset of 22 groundwater sources. Only significant (p -value < 0.01) Spearman's ρ are shown. Variables are ordered by hierarchical clustering and black squares enclose six selected clusters.

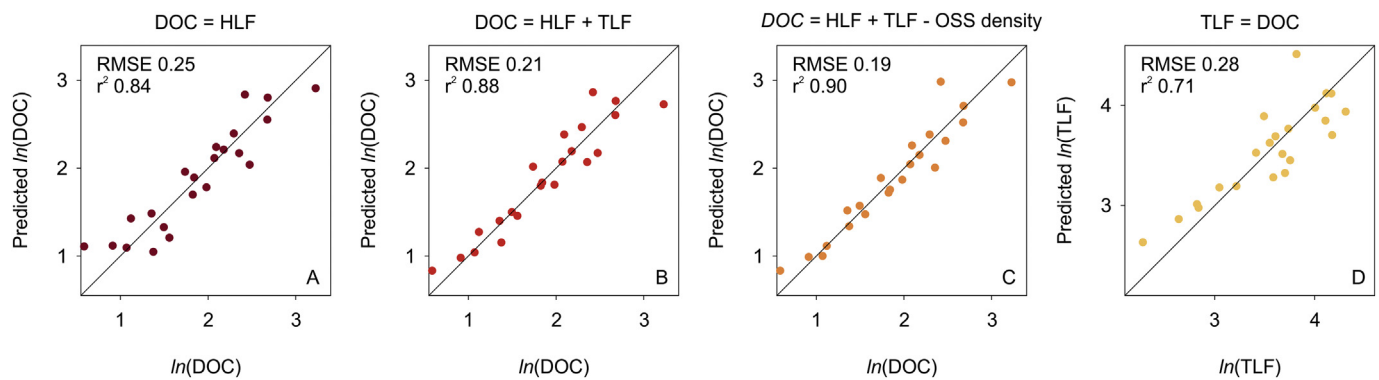


Fig. 6. Scatterplots of observed against predicted $\ln(\text{DOC})$ using (A) HLF; (B) HLF and TLF; (C) HLF, TLF and OSS density (Eq. (3)); and (D) observed against predicted $\ln(\text{TLF})$ using $\ln(\text{DOC})$ (Eq. (4)). A 1:1 line is shown in all plots.

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 (Sorensen et al., submitted to journal). 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 the water sources. Consequently, any TTCs present during the recharge season may have perished or become non-culturable whilst elevated fluorescent OM remained present. This seasonal pattern was previously reported by Sorensen et al. (2015a) who 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, 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.

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 (Sorensen et al., submitted to journal). 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 throughput 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.

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.

5. Conclusions

Fluorescent OM is significantly associated with parameters linked to faecal contamination that include on-site sanitation density (OSS), $\text{NO}_2^-/\text{NO}_3^-$, 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.

CRediT authorship contribution statement

James P.R. Sorensen: Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft. **Mor Talla Diaw:** Data curation, Methodology, Writing - review & editing. **Abdoulaye Pouye:** Data curation, Methodology, Writing - review & editing. **Raphaëlle Roffo:** Data curation, Methodology, Writing - review & editing. **Djim M.L. Diongue:** Data curation, Writing - review & editing. **Seynabou C. Faye:** Methodology, Funding acquisition, Project administration, Writing - review & editing. **Cheikh B. Gaye:** Methodology, Funding acquisition, Project administration. **Bethany G. Fox:** Writing - review & editing. **Timothy Goodall:** Data curation, Resources. **Daniel J. Lapworth:** Funding acquisition, Writing - review & editing. **Alan M. MacDonald:** Funding acquisition, Writing -

review & editing. **Daniel S. Read:** Data curation, Resources, Writing - review & editing. **Lena Ciric:** Supervision, Writing - review & editing. **Richard G. Taylor:** Supervision, Funding acquisition, Project administration, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

BGS authors publish with the permission of the Director, British Geological Survey (UKRI). Any identification of equipment does not imply recommendation or endorsement by the authors and their respective employers.

JPRS, AP, MTD, MLD, RR, SCF, CBG, and RGT acknowledge support from the Royal Society/UK Department for International Development (DFID) Africa Capacity Building Initiative AfriWatSan project [AQ140023] for fieldwork in Kenya, Senegal and Uganda, and co-author contributions.

JPRS and DJL acknowledge support from the DFID/Natural Environment Research Council (NERC) Future Climate for Africa (FCFA) HyCRISTAL project [NE/M020452/1] for laboratory support.

JPRS was supported by the British Geological Survey NC-ODA grant NE/R000069/1: Geoscience for Sustainable Futures for data analysis and writing of the paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2020.139419>.

References

- Alin, A., 2010. Multicollinearity. *Wiley Interdiscip. Rev. Comput. Stat.* 2 (3), 370–374.
- ANSD, 2016. National Census of the Population of Senegal in 2016. National Agency for Statistics and Demography, Dakar, Senegal.
- Baker, A., 2001. Fluorescence excitation-emission matrix characterization of some sewage-impacted rivers. *Environ. Sci. Technol.* 35 (5), 948–953.
- Baker, A., 2002. Fluorescence properties of some farm wastes: implications for water quality monitoring. *Water Res.* 36 (1), 189–195.
- Baker, A., 2005. Thermal fluorescence quenching properties of dissolved organic matter. *Water Res.* 39 (18), 4405–4412.
- Baker, A., Curry, M., 2004. Fluorescence of leachates from three contrasting landfills. *Water Res.* 38 (10), 2605–2613.
- Baker, A., Inverarity, R., 2004. Protein-like fluorescence intensity as a possible tool for determining river water quality. *Hydrol. Process.* 18 (15), 2927–2945.
- Baker, A., Cumberland, S.A., Bradley, C., Buckley, C., Bridgeman, J., 2015. To what extent can portable fluorescence spectroscopy be used in the real-time assessment of microbial water quality? *Sci. Total Environ.* 532, 14–19.
- Baral, D., Speicher, A., Dvorak, B., Admiraal, D., 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.
- Berney, M., Hammes, F., Bosshard, F., Weilenmann, H.-U., 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.
- Bridgeman, J., Baker, A., Brown, D., Boxall, J., 2015. Portable LED fluorescence instrumentation for the rapid assessment of potable water quality. *Sci. Total Environ.* 524, 338–346.
- Bronk, B.V., Reinisch, L., 1993. Variability of steady-state bacterial fluorescence with respect to growth conditions. *Appl. Spectrosc.* 47 (4), 436–440.
- Brown, C.M., Staley, C., Wang, P., Dalzell, B., Chun, C.L., Sadowsky, M.J., 2017. A high-throughput DNA-sequencing approach for determining sources of fecal bacteria in a Lake Superior estuary. *Environ. Sci. Technol.* 51 (15), 8263–8271.
- Cammack, W.L., Kalf, J., Prairie, Y.T., Smith, E.M., 2004. Fluorescent dissolved organic matter in lakes: relationships with heterotrophic metabolism. *Limnol. Oceanogr.* 49 (6), 2034–2045.
- Carstea, E.M., Bridgeman, J., Baker, A., Reynolds, D.M., 2016. Fluorescence spectroscopy for wastewater monitoring: a review. *Water Res.* 95, 205–219.
- Carstea, E.M., Popa, C.L., Baker, A., Bridgeman, J., 2019. *In situ* fluorescence measurements of dissolved organic matter: a review. *Sci. Total Environ.* 134361.

- Chapelle, F.H., Bradley, P.M., Goode, D.J., Tiedeman, C., Lacombe, P.J., Kaiser, K., Benner, R., 2009. Biochemical indicators for the bioavailability of organic carbon in ground water. *Groundwater* 47 (1), 108–121.
- Coble, P.G., Lead, J., Baker, A., Reynolds, D.M., Spencer, R.G., 2014. *Aquatic Organic Matter Fluorescence*. Cambridge University Press.
- Dalterio, R., Nelson, W., Britt, D., Sperry, J., Psaras, D., Tanguay, J., Suib, S., 1986. Steady-state and decay characteristics of protein tryptophan fluorescence from bacteria. *Appl. Spectrosc.* 40 (1), 86–90.
- Dalterio, R., Nelson, W., Britt, D., Sperry, J., Tanguay, J., Suib, S., 1987. The steady-state and decay characteristics of primary fluorescence from live bacteria. *Appl. Spectrosc.* 41 (2), 234–241.
- Dartnell, L.R., Roberts, T.A., Moore, G., Ward, J.M., Muller, J.-P., 2013. Fluorescence characterization of clinically-important bacteria. *PLoS One* 8 (9), e75270.
- Diaw, M.T., Faye, S.C., Gaye, C.B., Niang, S., Pouye, A., Campos, L.C., Taylor, R.C. (Submitted to *Journal of Water, Sanitation and Hygiene for Development*). On-site sanitation density and groundwater quality: evidence from remote sensing and in situ observations in the Thiaroye aquifer, Senegal.
- Diédhiou, M., 2011. Approche multitraceur géochimique et isotopique à l'identification des sources de la pollution nitrée et des processus de nitrification et dénitrification dans la nappe de Thiaroye. Univ. Cheikh Anta Diop de Dakar, p. 210.
- Diédhiou, M., Cissé Faye, S., Diouf, O., Faye, S., Faye, A., Re, V., Wöhllich, S., Wisotzky, F., Schulte, U., Maloszewski, P., 2012. Tracing groundwater nitrate sources in the Dakar suburban area: an isotopic multi-tracer approach. *Hydrol. Process.* 26 (5), 760–770.
- Downing, B.D., Pellerin, B.A., Bergamaschi, B.A., Saraceno, J.F., 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.
- Elliott, S., Lead, J., Baker, A., 2006. Characterisation of the fluorescence from freshwater, planktonic bacteria. *Water Res.* 40 (10), 2075–2083.
- 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., Faye, S., Wöhllich, S., Gaye, C.B., 2004. An assessment of the risk associated with urban development in the Thiaroye area (Senegal). *Environ. Geol.* 45 (3), 312–322.
- Faye, S.C., Diongue, M., Pouye, A., Gaye, C.B., Travi, Y., Wöhllich, S., Faye, S., Taylor, R.G., 2019. Tracing natural groundwater recharge to the Thiaroye aquifer of Dakar, Senegal. *Hydrogeol. J.* 27 (3), 1067–1080.
- Fellman, J.B., Hood, E., Spencer, R.G., 2010. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: a review. *Limnol. Oceanogr.* 55 (6), 2452–2462.
- Fellman, J.B., Dogramaci, S., Skrzypek, G., Dodson, W., Grierson, P.F., 2011. Hydrologic control of dissolved organic matter biogeochemistry in pools of a subtropical dryland river. *Water Resour. Res.* 47 (6).
- Findlay, S., Strayer, D., Goumbala, C., Gould, K., 1993. Metabolism of streamwater dissolved organic carbon in the shallow hyporheic zone. *Limnol. Oceanogr.* 38 (7), 1493–1499.
- Foulquier, A., Mermillod-Blondin, F., Malard, F., Gibert, J., 2011. Response of sediment biofilm to increased dissolved organic carbon supply in groundwater artificially recharged with stormwater. *J. Soils Sediments* 11 (2), 382–393.
- Fox, B., Thorn, R., Anesio, A., Reynolds, D., 2017. The in situ bacterial production of fluorescent organic matter; an investigation at a species level. *Water Res.* 125, 350–359.
- Fox, B., Thorn, R., Anesio, A., Cox, T., Attridge, J., Reynolds, D., 2019. Microbial processing and production of aquatic fluorescent organic matter in a model freshwater system. *Water* 11 (1), 10.
- Frank, S., Goepfert, N., Goldscheider, N., 2017. Fluorescence-based multi-parameter approach to characterize dynamics of organic carbon, faecal bacteria and particles at alpine karst springs. *Sci. Total Environ.* 615, 1446–1459.
- Friendly, M., 2002. Corgrams: exploratory displays for correlation matrices. *Am. Stat.* 56 (4), 316–324.
- Griehler, C., Lueders, T., 2009. Microbial biodiversity in groundwater ecosystems. *Freshw. Biol.* 54 (4), 649–677.
- Henderson, R., Baker, A., Murphy, K., Hambly, A., Stuetz, R., Khan, S., 2009. Fluorescence as a potential monitoring tool for recycled water systems: a review. *Water Res.* 43 (4), 863–881.
- 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., Deletic, A., 2016. Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water Res.* 93, 242–253.
- Hudson, N., Baker, A., Reynolds, D., 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review. *River Res. Appl.* 23 (6), 631–649.
- Hudson, N., Baker, A., Ward, D., Reynolds, D.M., Brunsdon, C., Carliell-Marquet, C., Browning, S., 2008. Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England. *Sci. Total Environ.* 391 (1), 149–158.
- Hynds, P.D., Misstear, B.D., Gill, L.W., 2012. Development of a microbial contamination susceptibility model for private domestic groundwater sources. *Water Resour. Res.* 48 (12).
- Kallenbach, C.M., Frey, S.D., Grandy, A.S., 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nat. Commun.* 7, 13630.
- Kamiya, E., Izumiya, S., Nishimura, M., Mitchell, J.G., Kogure, K., 2007. Effects of fixation and storage on flow cytometric analysis of marine bacteria. *J. Oceanogr.* 63 (1), 101–112.
- Khamis, K., Sorensen, J., Bradley, C., Hannah, D., Lapworth, D.J., Stevens, R., 2015. In situ tryptophan-like fluorometers: assessing turbidity and temperature effects for freshwater applications. *Environ Sci Process Impacts* 17 (4), 740–752.
- Khamis, K., Bradley, C., Stevens, R., Hannah, D.M., 2017. Continuous field estimation of dissolved organic carbon concentration and biochemical oxygen demand using dual-wavelength fluorescence, turbidity and temperature. *Hydrol. Process.* 31 (3), 540–555.
- Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., Kelley, S.T., 2011. Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8 (9), 761.
- Larsen, L.G., Aiken, G.R., Harvey, J.W., Noe, G.B., Crimaldi, J.P., 2010. Using fluorescence spectroscopy to trace seasonal DOM dynamics, disturbance effects, and hydrologic transport in the Florida Everglades. *J. Geophys. Res. Biogeosci.* 115 (G3).
- Lee, E.-J., Yoo, G.-Y., Jeong, Y., Kim, K.-U., Park, J.-H., 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., Oulhal, N., 2016. Recent advances on multi-parameter flow cytometry to characterize antimicrobial treatments. *Front. Microbiol.* 7, 1225.
- Li, W.-T., Jin, J., Li, Q., Wu, C.-F., Lu, H., Zhou, Q., Li, A.-M., 2016. Developing LED UV fluorescence sensors for online monitoring DOM and predicting DBPs formation potential during water treatment. *Water Res.* 93, 1–9.
- Marie, D., Rigaut-Jalabert, F., Vault, D., 2014. An improved protocol for flow cytometry analysis of phytoplankton cultures and natural samples. *Cytom. Part A* 85 (11), 962–968.
- 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., Sillanpää, M., 2010. Natural organic matter removal by coagulation during drinking water treatment: a review. *Adv. Colloid Interf. Sci.* 159 (2), 189–197.
- McDonough, L.K., Santos, I.R., Andersen, M.S., O'Carroll, D.M., Rutledge, H., Meredith, K., Oudone, P., Bridgeman, J., Gooddy, D.C., Sorensen, J.P., 2020. Changes in global groundwater organic carbon driven by climate change and urbanization. *Nat. Commun.* 11 (1), 1–10.
- Mendoza, L.M., Mladenov, N., Kinoshita, A.M., Pinongcos, F., Verbyla, M.E., Gersberg, R., 2020. Fluorescence-based monitoring of anthropogenic pollutant inputs to an urban stream in Southern California, USA. *Sci. Total Environ.* 718, 137206.
- Muller, M., Milori, D.M.B.P., Déléris, S., Steyer, J.-P., Dudal, Y., 2011. Solid-phase fluorescence spectroscopy to characterize organic wastes. *Waste Manag.* 31 (9), 1916–1923.
- Murdoch, D., Chow, E., 1996. A graphical display of large correlation matrices. *Am. Stat.* 50 (2), 178–180.
- Myles, H., Wolfe, D.A., 1973. *Nonparametric Statistical Methods*. Ed. John Wiley and Sons, New York, NY (503 pp.).
- Nakar, A., Schmilovitch, Z., Vaizel-Ohayon, D., Kroupitski, Y., Borisover, M., Sela, S., 2020. Quantification of bacteria in water using PLS analysis of emission spectra of fluorescence and excitation-emission matrices. *Water Res.* 169, 115197.
- Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M., 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Org. Geochem.* 31 (12), 1765–1781.
- Re, V., Faye, S.C., Faye, A., Faye, S., Gaye, C.B., Sacchi, E., Zuppi, G.M., 2011. Water quality decline in coastal aquifers under anthropic pressure: the case of a suburban area of Dakar (Senegal). *Environ. Monit. Assess.* 172 (1–4), 605–622.
- Reynolds, D., Ahmad, S., 1997. Rapid and direct determination of wastewater BOD values using a fluorescence technique. *Water Res.* 31 (8), 2012–2018.
- Romera-Castillo, C., Sarmento, H., Alvarez-Salgado, X.A., Gasol, J.M., Marrasé, C., 2011. Net production and consumption of fluorescent colored dissolved organic matter by natural bacterial assemblages growing on marine phytoplankton exudates. *Appl. Environ. Microbiol.* 77 (21), 7490–7498.
- Royston, J.P., 1982. An extension of Shapiro and Wilk's W test for normality to large samples. *J. R. Stat. Soc. Ser. C: Appl. Stat.* 31 (2), 115–124.
- Saraceno, J.F., Shanley, J.B., Downing, B.D., Pellerin, B.A., 2017. Clearing the waters: evaluating the need for site-specific field fluorescence corrections based on turbidity measurements. *Limnol. Oceanogr. Methods* 15 (4), 408–416.
- Seaver, M., Roselle, D.C., Pinto, J.F., Eversole, J.D., 1998. Absolute emission spectra from *Bacillus subtilis* and *Escherichia coli* vegetative cells in solution. *Appl. Opt.* 37 (22), 5344–5347.
- Shimotori, K., Watanabe, K., Hama, T., 2012. Fluorescence characteristics of humic-like fluorescent dissolved organic matter produced by various taxa of marine bacteria. *Aquat. Microb. Ecol.* 65 (3), 249–260.
- Smith, R.J., Paterson, J.S., Sibley, C.A., Hutson, J.L., Mitchell, J.G., 2015. Putative effect of aquifer recharge on the abundance and taxonomic composition of endemic microbial communities. *PLoS One* 10 (6).
- Snyder, L., Potter, J.D., McDowell, W.H., 2018. An evaluation of nitrate, fDOM, and turbidity sensors in New Hampshire streams. *Water Resour. Res.* 54 (3), 2466–2479.
- Sobczak, W.V., 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., Fedorka-Cray, P.J., 2009. Fluorescence spectroscopy for rapid detection and classification of bacterial pathogens. *Appl. Spectrosc.* 63 (11), 1251–1255.
- Sorensen, J.P.R., Carr, A.F., Nayebar, J., Diongue, D.M.L., Pouye, A., Roffo, R., Gwengweya, G., Ward, J.S.T., Kanoti, J., Okotto-Okotto, J., van der Marel, L., Cirić, 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. (Submitted to *Scientific Reports*). Tryptophan-like and humic-like fluorophores are extracellular in groundwater: implications as real-time faecal indicators.
- Sorensen, J., Lapworth, D., Marchant, B., Nkhuwa, D., Pedley, S., Stuart, M., Bell, R., Chirwa, M., Kabika, J., Liemisa, M., 2015a. In-situ tryptophan-like fluorescence: a real-time indicator of faecal contamination in drinking water supplies. *Water Res.* 81, 38–46.

- Sorensen, J., Lapworth, D., Read, D., Nkhuwa, D., Bell, R., Chibesa, M., Chirwa, M., Kabika, J., Liemisa, M., Pedley, S., 2015b. Tracing enteric pathogen contamination in sub-Saharan African groundwater. *Sci. Total Environ.* 538, 888–895.
- Sorensen, J., Sadhu, A., Sampath, G., Sugden, S., Gupta, S.D., Lapworth, D., Marchant, B., Pedley, S., 2016. Are sanitation interventions a threat to drinking water supplies in rural India? An application of tryptophan-like fluorescence. *Water Res.* 88, 923–932.
- Sorensen, J., Vivanco, A., Ascott, M., Goody, D., Lapworth, D., Read, D., Rushworth, C., Bucknall, J., Herbert, K., Karapanos, I., 2018a. Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. *Water Res.* 137, 301–309.
- Sorensen, J.P., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G., Ward, J.S., 2018b. Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values. *Sci. Total Environ.* 622, 1250–1257.
- Stedmon, C.A., Thomas, D.N., Granskog, M., Kaartokallio, H., Papadimitriou, S., Kuosa, H., 2007. Characteristics of dissolved organic matter in Baltic coastal sea ice: allochthonous or autochthonous origins? *Environ. Sci. Technol.* 41 (21), 7273–7279.
- Stedmon, C.A., Sereďyńska-Sobecka, B., Boe-Hansen, R., Le Tallec, N., Waul, C.K., 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 Res.* 45 (18), 6030–6038.
- Thottathil, S.D., Hayakawa, K., Hodoki, Y., Yoshimizu, C., Kobayashi, Y., Nakano, S.-i., 2013. Biogeochemical control on fluorescent dissolved organic matter dynamics in a large freshwater lake (Lake Biwa, Japan). *Limnol. Oceanogr.* 58 (6), 2262–2278.
- Tunaley, C., Tetzlaff, D., Lessels, J., Soulsby, C., 2016. Linking high-frequency DOC dynamics to the age of connected water sources. *Water Resour. Res.* 52 (7), 5232–5247.
- Weaver, L., Sinton, L.W., Pang, L., Dann, R., Close, M., 2013. Transport of microbial tracers in clean and organically contaminated silica sand in laboratory columns compared with their transport in the field. *Sci. Total Environ.* 443, 55–64.
- WHO, 1997. Guidelines for Drinking-water Quality. 2nd edition. WHO Press, Geneva, Switzerland.
- WHO, 2011. Guidelines for Drinking-water Quality. 4th edition. WHO Press, Geneva, Switzerland.
- Williams, C.J., Yamashita, Y., Wilson, H.F., Jaffé, R., Xenopoulos, M.A., 2010. Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. *Limnol. Oceanogr.* 55 (3), 1159–1171.
- Worthington, S.R., Smart, C.C., 2017. Transient bacterial contamination of the dual-porosity aquifer at Walkerton, Ontario, Canada. *Hydrogeol. J.* 25 (4), 1003–1016.
- Yang, L., Hur, J., Zhuang, W., 2015. Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: a review. *Environ. Sci. Pollut. Res.* 22 (9), 6500–6510.