Science of the Total Environment 928 (2024) 172285

Contents lists available at ScienceDirect

Science of the Total Environment

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

Temporal drivers of tryptophan-like fluorescent dissolved organic matter along a river continuum

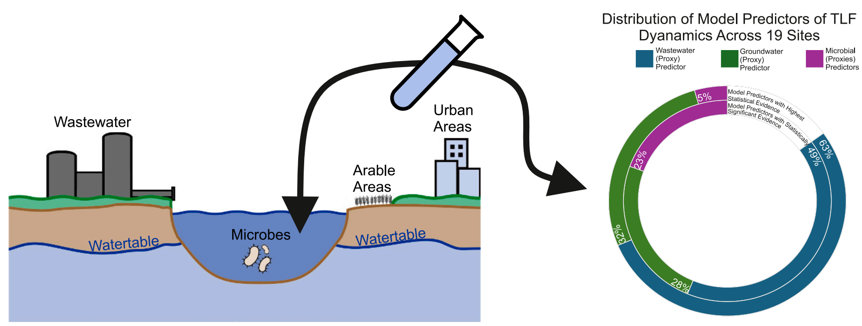
N.A. Harris [a,*], J.P.R. Sorensen [a], B. Marchant [a], G.H. Old [b], P.S. Naden [b], M.J. Bowes [b], P.M. Scarlett [b], D.J.E. Nicholls [b], L.K. Armstrong [b], H.D. Wickham [b], D.S. Read [b], D. Lapworth [a], T. Bond [c], K. Pond [c]

[a] British Geological Survey, Maclean Building, Wallingford OX10 8BB, UK
[b] Centre for Ecology & Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK
[c] Centre for Environmental Health and Engineering, Department of Civil and Environmental Engineering, University of Surrey, Guildford GU2 5XH, UK

HIGHLIGHTS

- Wastewater omnipresent significant predictor of TLF dynamics, but importance varies
- Groundwater is a dilutionary control on TLF in groundwater-dominated sub-catchments.
- Microbial sources significant in 52 % of sub-catchments
- Complex interplay of wastewater, baseflow and microbial sources drive TLF dynamics.
- Importance of different sources depends on sub-catchment characteristics.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Fernando Pacheco

Keywords:
Fluorescent organic matter
Tryptophan-like
Microbial processing
In situ microbial production
Catchment characteristics

ABSTRACT

Tryptophan-like fluorescence (TLF) is used to indicate anthropogenic inputs of dissolved organic matter (DOM), typically from wastewater, in rivers. We hypothesised that other sources of DOM, such as groundwater and planktonic microbial biomass can also be important drivers of riverine TLF dynamics. We sampled 19 contrasting sites of the River Thames, UK, and its tributaries. Multivariate mixed linear models were developed for each site using 15 months of weekly water quality observations and with predictor variables selected according to the statistical significance of their linear relationship with TLF following a stepwise procedure. The variables considered for inclusion in the models were potassium (wastewater indicator), nitrate (groundwater indicator), chlorophyll-a (phytoplankton biomass), and Total bacterial Cells Counts (TCC) by flow cytometry. The wastewater indicator was included in the model of TLF at 89 % of sites. Groundwater was included in 53 % of models, particularly those with higher baseflow indices (0.50–0.86). At these sites, groundwater acted as a negative control on TLF, diluting other potential sources. Additionally, TCC was included positively in the models of six

Abbreviations: WFD, Water Framework Directive; DOC, Dissolved Organic Matter; OM, Organic Matter; BOD, Biological Oxygen Demand; TCC, Total Cell Count; EA, Environment Agency; TLF, Tryptophan Like Fluorescence; VE, Variance Explained; PCA, Principle Component Analysis; SPE, Sewage Population Estimate; BFI, Baseflow Index; WwTW, Wastewater Treatment Works; CSO, Combined Sewerage Overflow; NRFA, National River Flow Archive; (NH_4), Ammonia; (PARAFAC), Parallel factor; (RU), Raman Units.

* Corresponding author.
E-mail address: nhar1@bgs.ac.uk (N.A. Harris).

https://doi.org/10.1016/j.scitotenv.2024.172285
Received 15 December 2023; Received in revised form 4 April 2024; Accepted 5 April 2024
Available online 8 April 2024

(32 %) sites. The models on the Thames itself using TCC were more rural sites with lower sewage inputs. Phytoplankton biomass (Chlorophyll-a) was only used in two (11 %) site models, despite the seasonal phytoplankton blooms. It is also notable that, the wastewater indicator did not always have the strongest evidence for inclusion in the models. For example, there was stronger evidence for the inclusion of groundwater and TCC than wastewater in 32 % and 5 % of catchments, respectively. Our study underscores the complex interplay of wastewater, groundwater, and planktonic microbes, driving riverine TLF dynamics, with their influence determined by site characteristics.

1. Introduction

Anthropogenic effects are having increasingly serious impacts on surface water quality globally, threatening aquatic ecosystems and public health (Akhtar et al., 2021; Dodds et al., 2013; Huang et al., 2021; du Plessis, 2022). These detrimental impacts are predominantly attributed to wastewater discharges and agriculture (Van Drecht et al., 2009; Li et al., 2022; du Plessis, 2022; UN DESA, 2022). Whilst increased and enhanced legislation, monitoring and management has improved river water quality and ecological metrics, with respect to some parameters, water quality remains unacceptably poor in many locations (du Plessis, 2022; UN DESA, 2022; Whelan et al., 2022). For example, only 38 % of European surface waters are in "good" chemical status, under the Water Framework Directive (WFD) standards. In addition, 100 % of rivers in Belgium, England, Germany, and Sweden currently fail WFD standards. (Environment Agency, 2023; Hannah et al., 2022; Kristensen et al., 2018).

Anthropogenic activity delivers a vast quantity of Organic Matter (OM) into many rivers (Stanley et al., 2012; Wagner et al., 2015). For example, Dissolved Organic Carbon (DOC) concentrations have doubled (1884–2014) in the River Thames, and 90 % of this trend is attributed to increased urban land cover and the resultant increased input of sewage effluent (Noacco et al., 2017). This organic pollution can be detrimental to ecosystems and increases risk to human health due to the introduction of pathogens (Sirota et al., 2013; Strokal et al., 2019; Wen et al., 2017) Such organic matter tends to be labile (Stanley et al., 2012) and can drive increases in microbial biomass and activity (Lambert et al., 2017; Williams et al., 2010). It can also lead to phytoplankton blooms (Bowes et al., 2014), alter microbial community composition (Zhang et al., 2020), and contribute to reduced biological diversity and ecological integrity (Arthington et al., 2010).

Fluorescence spectroscopy can characterise and quantify fluorescent fractions of organic matter and help understand their origin, riverine processing, and fate (Fellman et al., 2010; Hudson et al., 2007). The technique is sensitive, reagentless, non-destructive, and provides rapid results (Bridgeman et al., 2011), enabling *in situ* deployments (Carstea et al., 2020). Fluorescent peaks can be broadly split into those associated with humic-like and protein-like OM (Hudson et al., 2007). It is the protein-like peaks that are more commonly associated with anthropogenic waste, notably those observed at excitation and emission wavelengths of 275 nm and 340 nm, respectively (Coble, 1996). This peak is termed peak T or Tryptophan-Like Fluorescence (TLF). It resembles the fluorescent properties of the amino acid tryptophan, although it is commonly associated with protein residues or high-molecular weight DOM, as well as free amino acids (Fellman et al., 2010).

The dominant source of TLF in rivers is typically considered to be wastewater (Ahmad and Reynolds, 1999; Baker, 2001; Khamis et al., 2017). Consequently, strong positive relationships have been observed between TLF and both Biological Oxygen Demand (BOD) (Baker and Inverarity, 2004; Hudson et al., 2008; Hur et al., 2008; Hur and Cho, 2012; Khamis et al., 2017) and faecal indicator organisms (Baker et al., 2015; Sorensen et al., 2018a). Laboratory studies have also shown that tryptophan-like fluorophores can be produced and consumed by bacteria (Bridgeman et al., 2015; Cammack et al., 2004; Elliott et al., 2006; Fox et al., 2017, 2018) and phytoplankton (Henderson et al., 2008; I. Khan et al., 2019; Ly et al., 2019; Nguyen et al., 2005; Villacorte et al., 2015). Positive relationships between TLF and Total bacterial Cell Counts (TCC) have also been observed in groundwater (Sorensen et al., 2020, 2021) and drinking water supply networks (Bridgeman et al., 2015). Nevertheless, we are not aware of previous studies directly linking TLF dynamics to microbial biomass within riverine environments. Riverine TLF dynamics can also be influenced by dilution (Baker, 2002; Pellerin et al., 2011; Saraceno et al., 2009), by sources of water containing less labile OM, such as groundwater (Chen et al., 2010).

Our study investigates controls of key potential environmental drivers; wastewater, groundwater, and for the first-time, TCC and chlorophyll-a, on riverine TLF dynamics in a range of sites of the River Thames, UK. We hypothesise that wastewater is not the only important driver of TLF dynamics in all anthropogenically impacted sites and seek to explore other drivers such as land use, groundwater and aquatic microbes.

2. Methods

2.1. Study area

The River Thames is the longest river wholly in England at 354 km to its tidal limit, and has a catchment area of 9948km^2.The catchment is home to 15 million people (Fig. 1) (Marsh et al., 2008) and has a temperate climate, with mean average (1985–2014) total annual rainfall and average air temperature of 730 mm (maximum 950 mm in 2000) and 11 °C (maximum 12 °C in 2014), respectively (Bussi et al., 2016a). Our study area is defined as the Thames upstream of London, incorporating its tributaries of the Coln, Leach, Windrush, Evenlode, Cherwell, Ray, Thame, Ock, Pang, Enborne, Loddon and The Cut.

The River Thames flows across limestone in the upper reaches, mudstones in the Oxford area, chalk over the middle reaches south of Oxford and north of Reading, and finally, clays covering the Reading and London area (Bloomfield et al., 2009; M.G Sumbler and British Geological Survey, 1996). This change in geology dictates the change in the groundwater regime and hence baseflow index across the catchment. The porous chalk and limestone are highly productive aquifers and the Base Flow Index (BFI) of the Thames and tributaries on these carbonate rocks are generally high (>0.8) (Bloomfield et al., 2009; M.G Sumbler and British Geological Survey, 1996). However, the bedrocks that underlie Swindon to Oxford and Reading to London are more impermeable. Here, the Thames and its tributaries have lower BFI values (<0.65) ((Bloomfield et al., 2009; M.G Sumbler and British Geological Survey, 1996).

There are a wide range of land uses within the catchment, including agriculture in the upper reaches, more forested land between Oxford and Reading, and increasing urban coverage from Reading to London (Marston et al., 2022). Due to the wide range of land uses, geology, and hydrogeology, the River Thames also has a rich diversity of wildlife (Environment Agency, 2021). For example, the network of chalk streams across the catchment creates unique and rare ecosystems which are globally and nationally important (Environment Agency, 2021).

Nineteen monitoring sites were selected within the catchment (Fig. 1), including six sites along the Thames and 13 tributaries, primarily to encompass a range of upstream catchment characteristics (Bowes et al., 2018). In addition, the locations were selected to be easily/safely accessible by road, at bridges where possible to enable easier access to the centre of the river, and to be close or near to Environment Agency (EA) flow gauging stations (Bowes et al., 2018). Characteristics upstream of each site are shown in Table S1.1. Catchment area was determined (Bowes et al., 2018) using the flood estimation handbook. The land use percentage cover was calculated using the Centre for Ecology and Hydrology (CEH) intelligent river network (Dawson et al., 2002) and the UK Land and Cover Map 2000 (Fuller et al., 2002) using River and Catchment Query and Extraction Layer (RAC-QUEL) (Bowes et al., 2018). Sewerage Population Estimate (SPE) is an estimated load given to the sewerage treatment works (STW) calculated from a typical *per capita* biological oxygen demand (BOD) and the population served by the STW (Keller et al., 2006). This is then standardised to the catchment area. BaseFlow Index (BFI) for each site was taken from the UK Hydrometric Register (Marsh et al., 2008).

2.2. Fieldwork and laboratory methods

For 62 weeks from June 2012 to August 2013, the 19 sites were sampled weekly from the centre of the river using a bucket between 9 am–6 pm (Bowes et al., 2018a; Old et al., 2019b). The bulk sample was subdivided onsite into samples for chemistry, microbiology, and fluorescence analysis. All samples were immediately stored in cool boxes and subsequently refrigerated within 8 h of collection on return to the laboratory.

2.2.1. Fluorescent dissolved organic matter: TLF

Samples for fluorescence and absorbance analysis were filtered through 0.45 μm cellulose nitrate filters into 15 mL polypropylene centrifuge tubes (Old et al., 2019). Old et al., 2019 validated the performance of these filters and tubes, demonstrating no contamination in field blanks, and negligible retention of dissolved tryptophan by the filters (<5 % at 0.3 Raman Units (RU) to <1.5 % at 1RU). Analysis was carried out within 24- 48 h of sampling using a Varian Cary Eclipse spectrophotometer with slit width of 5 nm, path length of 10 mm, integration time of 12.5 ms, excitation wavelength of 200-500 nm (5 nm steps) and emission wavelength of 280-500 nm (2 nm steps)(Old et al., 2019). Instrument corrections, following manufacturer instructions, were conducted to account for lamp output and instrument sensitivity (Holbrook et al., 2006). Absorbance was measured in a 10 mm cuvette on a Varian Cary 50 UV–vis spectrophotometer at 1 nm intervals from 800 to 200 nm. These data were then corrected to account for long-wavelength scatter using the Blough et al., 1993 methodology. The absorbance data was then used to correct for inner filtering effects using the Lakowicz, 2006 methodology. Finally, the fluorescence data was converted to Raman units using Lawaetz and Stedmon, 2009 methods.

Parallel factor (PARAFAC) analysis had previously been undertaken on the fluorescence of 1505 Excitation Emission matrices (EEMs) by Old et al. (2019) (See SI1.11 for further details). PARAFAC was completed using DOMFluor toolbox in MATLAB following Stedmon and Bro, 2008. The validated PARAFAC model contained 4-components, including a peak centred at 285 nm/325-355 nm (Ex/Em), which is consistent with a TLF peak from Coble, 1996, Hudson et al., 2008 and Parlanti et al., 2000. As this analysis had previously been completed by Old et al. (2019) and due to the advantages of PARAFAC, namely disentangling

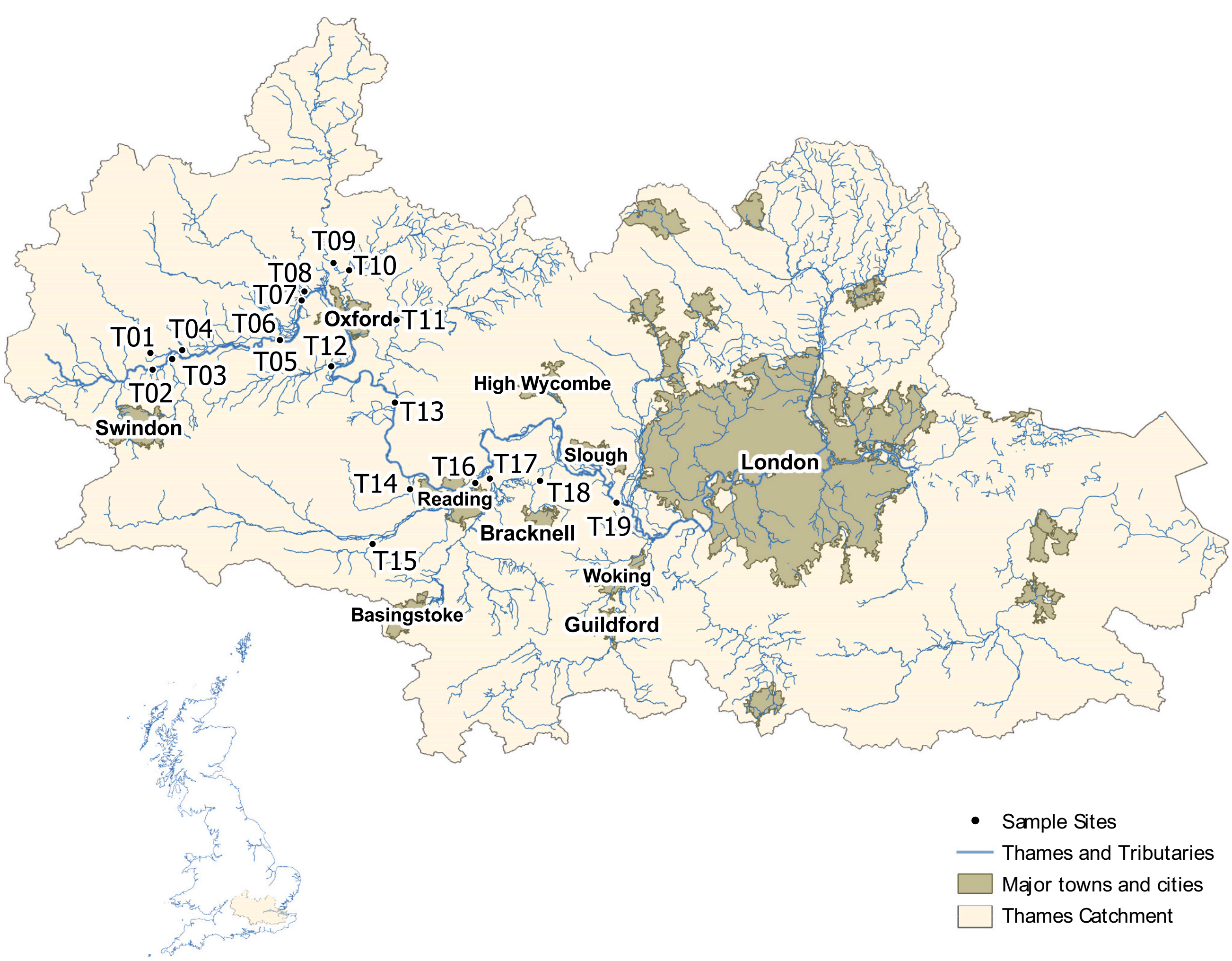


Fig. 1. Site locations across the River Thames catchment.

potential overlapping fluorophores, this TLF contribution data is used in subsequent analysis rather than running an addition peak picking algorithm. Further details of PARAFAC modelling can be found in Old et al. (2019), and the excitation and emission loading plots from Old et al. (2019) can be found in Fig. S1.11.

2.2.2. Total bacterial cell counts (TCC)

The samples for TCC analysis were collected in autoclaved polypropylene bottles (Read et al., 2015). Flow cytometry (FCM) was then used to count total bacterial cells. SYBR Green I stain was used, which reacts with bacterial nucleic acids to give off a green fluorescence which is detected by FCM. Each sample was then analysed for 1 min at a flow rate of approximately 5ul per minute using a Gallios flow cytometer (Beckman Coulter, High Wycombe, UK). A 488 nm laser was used, and TCC was determined using manually drawn gates in Kaluza 1.2 software (Beckman Coulter, High Wycombe, UK) on a cytogram of side scatter *vs* FL1. Further detail can be found in (Read et al., 2015).

2.2.3. Water chemistry

The sub-sample was collected in a 125 mL amber glass bottle, which was also used for pH and alkalinity analysis. Similar sub-samples for calcium, nitrate, total dissolved nitrogen (TDN), ammonia (NH_4), sodium, potassium, boron, and total dissolved phosphorus (TDP) analysis were collected and filtered through 0.45 μm cellulose nitrate membrane filter into three 60 mL polypropylene bottles. Unfiltered bulk water samples were filtered and soaked overnight in 90 % acetone while stored in a refrigerator to extract chlorophyll-a, which was then quantified by spectrophotometry. Details of the analytical methods used can be found in Bowes et al. (2018).

2.2.4. Selecting wastewater, groundwater, and in situ microbial indicators

Single water quality parameters were selected to indicate wastewater, groundwater (which we assume to be the dominant source of baseflow at these sites), and *in situ* microbial processes that could potentially drive TLF dynamics. The selected parameters were informed by previous studies (Fox et al., 2017, 2018; Henderson et al., 2008; Khamis et al., 2020) and by investigating parameter interrelationships. For each site, parameter interrelationships were analysed using Spearman's Rank and hierarchical clustering (SI 1.2). Hierarchical clustering was performed on the median of the Spearman's rank coefficients across all sites, which form four clusters that we consider representative of: wastewater (potassium, TDP, boron, sodium), groundwater (nitrate, TDN, calcium, pH), microbial (TCC, Chlorophyll-a) and NH_4 (shown in Fig. 2). Four clusters were chosen for the hierarchical clustering method, as additional clusters created single variable clusters. For example, a 5-cluster hierarchical clustering algorithm had a fifth cluster consisting solely of Sodium.

Potassium was selected to indicate potential wastewater contributions of TLF. Potassium consistently clustered alongside other wastewater variables, such as boron, TDP, and sodium, (Fig. 2; Fig. S1.2). The median Spearman's Rank correlation coefficient between potassium and TDP was 0.79, additionally the median potassium has a Spearman's Rank correlation with SPE of $r = 0.760$ ($p < 0.005$). More common wastewater indicators like TDN, TDP, and NH_4 were not suitable due to the variation of tertiary treatment of wastewater across the catchment (Bowes et al., 2018). For example, in the River Leach (site T04), TDP has no significant correlation with either potassium or ammonium (see Fig. S1.2). TDN does not relate to other typical wastewater variables in the median correlation matrix in Fig. 2, and rarely does at any individual site (Fig. S1.2). The majority of TDN consists of nitrate, with the mean proportion between 0.84 and 0.95 across the sites.

Nitrate was chosen to indicate potential groundwater contributions of TLF *via* baseflow. Nitrate clusters near to and positively correlates with other variables associated with groundwater emanating from calcareous bedrock such as calcium, and pH (Fig. 2). Furthermore, previous research has attributed the vast majority of nitrate inputs to the Thames upstream of London to groundwater (Bowes et al., 2018; Stuart et al., 2016). Nitrate was chosen over pH and calcium because it had the

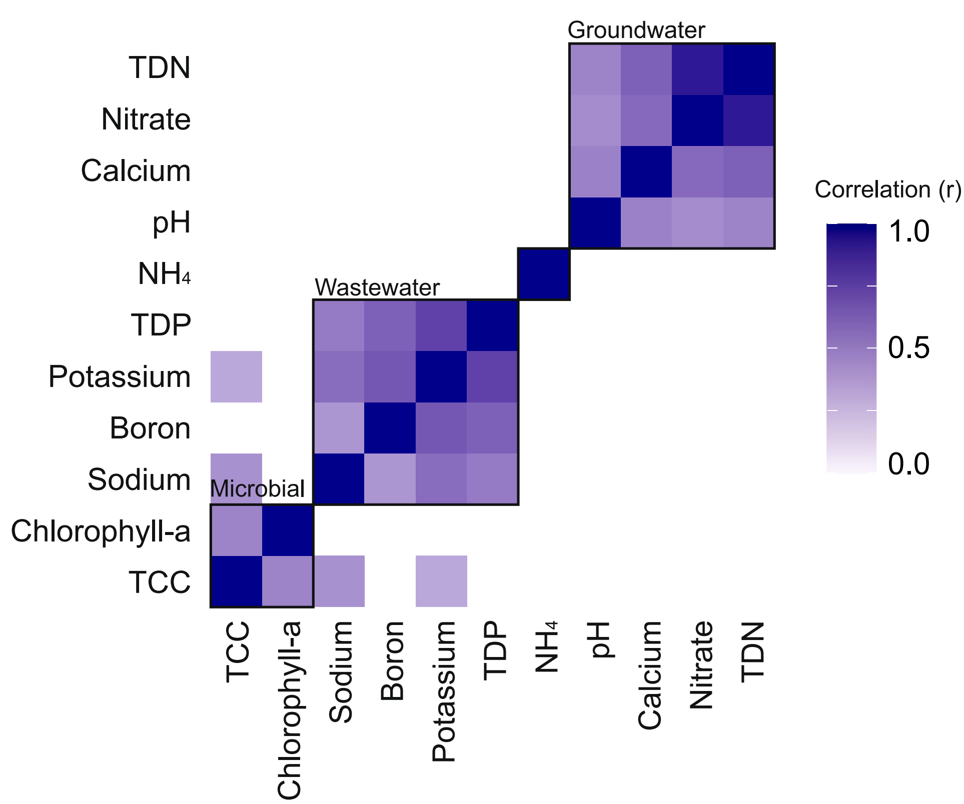


Fig. 2. Correlation Matrix with *p* values <0.05 of the median Spearman's Rank correlation of each variable across each site and arranged into four clusters using hierarchical clustering.

weakest correlation with potassium, the chosen wastewater variable. The median nitrate, pH and calcium Spearman-rank coefficient with potassium were 0.219, 0.242, and 0.317, respectively.

Finally, TCC and chlorophyll-a measurements were used as indicators of *in situ* microbial processes potentially affecting TLF. Chlorophyll-a measurements were used as an indication of phytoplankton biomass (Bowes et al., 2012, 2018) and TCC represented the total planktonic bacterial cell counts.

2.2.5. Disentangling relative contributions of wastewater, groundwater, and in situ microbial indicators on TLF dynamics at each site

The four indicator variables were used as possible predictors in a series of linear mixed models (Marchant, 2018) of the temporal variation of TLF at each site. This model type was chosen as multiple predictor models will allow for a more detailed understanding and analysis of how the different potential sources of TLF vary between sites.

This model can be written in Eq. (1):

$$y = M\beta + \varepsilon, \tag{1}$$

where y is a vector of n measured and transformed (see below) TLF values, M is an $n \times q$ matrix containing the values of q predictor variables corresponding to the n TLF measurements, β is a vector of q regression coefficients and ε is a vector of n residuals.

In standard linear regression, the linear model residuals are assumed to be independent which can lead to the significance of predictors being over-stated if the residuals are, in fact, temporally correlated. (*i.e.* if measurements made a week apart, for example, are more likely to be similar than those made several months apart). The residuals of a linear mixed model are permitted to be temporally correlated and this correlation is accounted for when quantifying the uncertainty associated with estimated regression coefficients. The degree of temporal correlation is estimated as part of the model fitting procedure.

In this case we assume that the temporal correlation, $C(\tau)$, between a pair of measurements made τ time units apart can be represented by a nested nugget and exponential function in Eq. (2) (Webster and Oliver, 2007):

$$C(\tau) = c_0 + c_1\left(1 - exp\left(\frac{\tau}{a}\right)\right). \tag{2}$$

We estimated the three parameters of this function (the nugget variance c_0, the partial sill variance c_1 and the temporal parameter a) by maximum likelihood (Marchant, 2018). A Box Cox transformation is applied to the TLF data to ensure that the residuals are consistent with a Gaussian distribution (Marchant, 2018). This procedure also leads to estimates of the linear regression coefficient of each potential predictor β_i and the corresponding standard error (*i.e.* the uncertainty) of each of these estimates σ_i (Marchant, 2018).

The *Z*-score for each predictor, defined as β_i/σ_i, is a measure of our confidence that the true value of β_i is not zero. If the true value were zero, then the Z-score would be expected to be drawn from a Gaussian distribution with zero mean and variance one. Hence, a Z-score value with magnitude >1.96 would occur with probability <0.05. When the calibrated model leads to a Z-score with magnitude >1.96, the corresponding predictor can be considered to be significant at the $p = 0.05$ level. In this paper, we consider the significance of multiple predictor variables. In this multiple hypothesis testing situation and the case where all the true β_i were zero, the probability that at least one *Z*-score had magnitude >1.96 would be >0.05. Bonferroni (1936) showed that if q hypotheses were being tested at level p, a threshold on the test statistic (in our case the *Z*-score) based on the p/q level would lead to a conservative adjustment for this problem.

We used an iterative or stepwise modelling procedure to decide which predictors should be included in the model for a particular site. The initial model included all four potential predictors. This model then underwent leave-one-out cross validation and any outliers (defined as the measured value being more than four standard deviations from the predicted value) were identified and removed. The model was then refitted and any predictors with a significant *Z*-score were included in the model. We selected a p level of 0.05. Since the four predictors used in modelling were selected from an initial list of 11, we conservatively adjusted for multiple hypothesis tests by basing the *Z*-score threshold on a p value of 0.05/11. The resultant threshold was 2.84.

In addition to the magnitude of the Z-score, the proportion of Variance Explained (VE) by a predictor can also be seen as a measure of the strength of the relationship between that predictor and TLF. We approximate the VE of the model by Eq. (3):

$$1 - \frac{\text{variance}(\varepsilon)}{\text{variance}(y)}. \tag{3}$$

Furthermore, we approximate the VE by a single predictor variable, as the variance explained by the full model minus the variance explained by the full model without the predictor of interest. We note that the VE by a model can be negative. This should not occur when all the predictor variables are significant, but could occur following the removal of a predictor, if that leads to one of the remaining predictors no longer being significant. In that case it is possible that the approximate VE is >1.

2.2.6. Understanding how catchment characteristics influence the drivers of TLF dynamics

Principle Component Analysis (PCA) was used to investigate how catchment characteristics, shown in Table S1.1, impacted the relative contributions of wastewater, groundwater, and *in situ* microbial processes on TLF dynamics. PCA was conducted using the prcomp function in the stats package within R (R Core Team, 2023).

3. Results

3.1. Seasonal TLF trends

Across the catchment, most sites show lower TLF values during the winter and higher values during the summer (See Fig. S1.10). Sites T14, T04 and T01 are exceptions where there is little variation seasonally, evidenced by lower than average variance (T14 = 0.005, T04 = 0.006, T01 = 0.003, Median Variance = 0.009). These sites also have a higher BFI than average (T14 = 0.723, T04 = 0.865, T01 = 0.842, Median BFI = 0.642). Overall, there is also a strong negative correlation between TLF variance and BFI ($r = -0.781$).

3.2. Wastewater drivers of TLF dynamics

There was a statistically significant relationship between the wastewater predictor, potassium, and TLF at 17 out of 19 site (Fig. 3). Potassium had the highest mean *Z*-score (9.074) and was the predictor with the highest Z-score in 12 models (Fig. 3). There was a trend for increasing potassium Z-scores at sites further downstream within the catchment (Fig. 3). Potassium also, had the highest median VE explained (0.614) and had the highest VE in 13 models (Fig. 4). Models with a higher Z-score and VE for the potassium predictor tended to be located in sites with higher SPE and more urban land. This is shown on the PCA plot (Fig. 5A1–2).On average, sites on the Thames tended to have a higher Z-score and VE than on tributaries (Thames median Z-score = 11.3, tributary median Z-score = 5.6, Thames median VE = 0.0588, Tributary median VE = 0.0395). An exception to this is at site T18, which is a tributary with the highest *Z*-score and highest SPE (lower left of Fig. 5A1 and top right on Fig. 5B2).

3.3. Groundwater drivers of TLF dynamics

The groundwater predictor, nitrate, was the second most common predictor and was used negatively on all occasions (Fig. 3). Nitrate had the second-highest median Z-score magnitude (−5.32). For six out of 19 models, the evidence for including groundwater was stronger than all

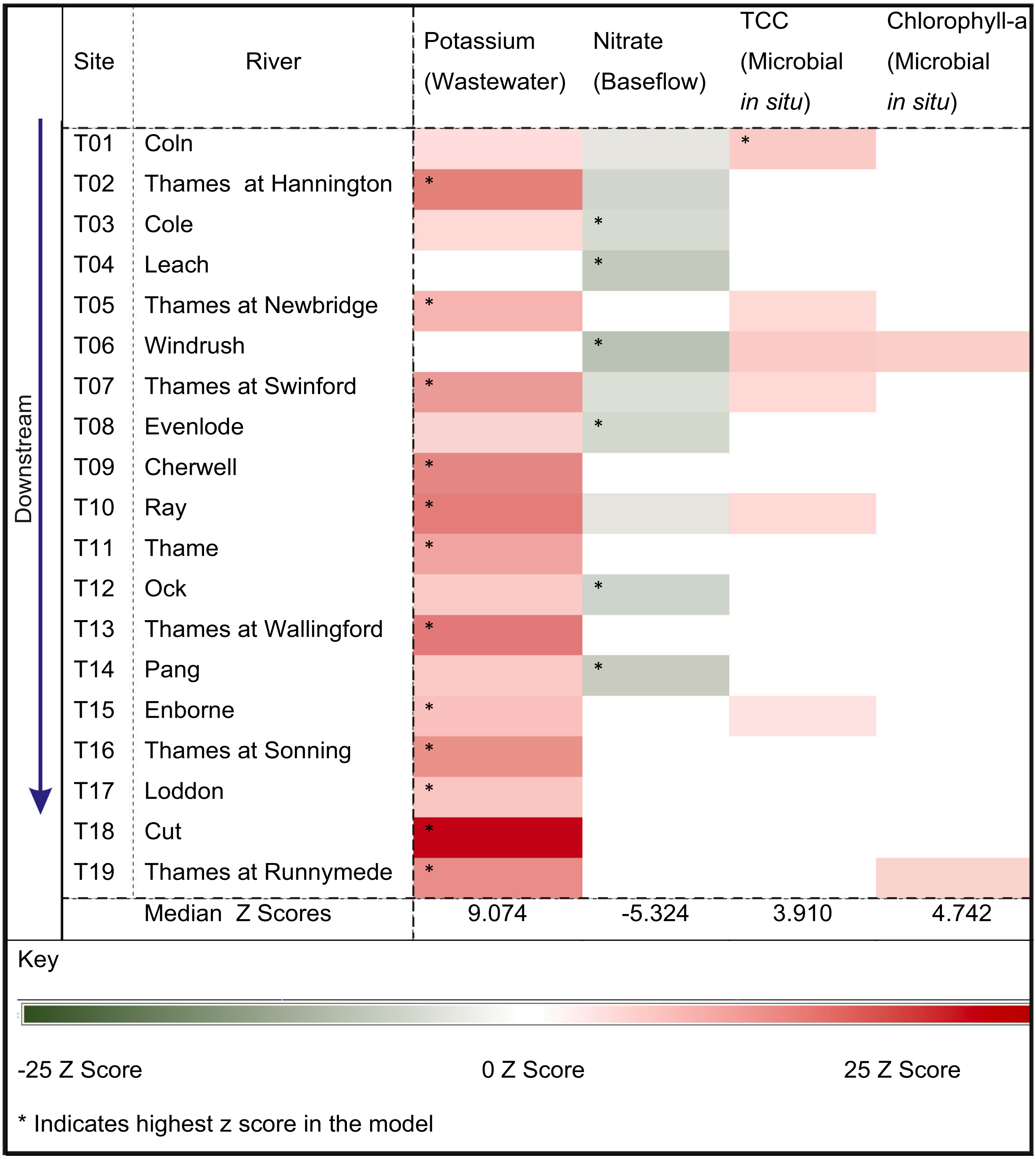


Fig. 3. Results of linear mixed models for each site, where TLF is the objective variable and Potassium, Nitrate, TCC and Chlorophyll-a possible predictors. Starting with the site furthest upstream and ending with the site further downstream in the catchment. The colour scale denotes the *Z*-score of the predictor (Z Score < −2.84, Z Score > 2.84), with dark red detonating a positive Z-score near 25, dark blue denoting a negative Z-score near −25, and white indicating no presence in the model.

other predictors (Fig. 3). Models which used nitrate had the second highest median variability explained (0.123) and tended to be located further upstream in the Thames catchment (Fig. 4). Nitrate had the highest VE in five out of 19 models (Fig. 4).

Models with nitrate as a predictor and the most negative Z-scores but the higher VE tended to be located in areas with higher BFI and more arable land (Fig. 6A1–2). Finally, more negative Z-scores were found on tributaries rather than the Thames sites (Thames median z score = −2.72, Tributary median z score = −4.57, Thames median VE = 0.054, Tributary median VE = 0.123). These Thames sites had a higher median SPE than the tributary sites at 244 and 71, respectively.

3.4. *In situ microbial drivers of TLF dynamics*

The *in situ* microbial predictors TCC and chlorophyll-a were included in seven out of 19 models with median average Z-scores of 3.91 and 4.74, respectively (Fig. 3). TCC was the predictor with the strongest evidence for inclusion in one model (Fig. 3 and Fig. 7A). More models

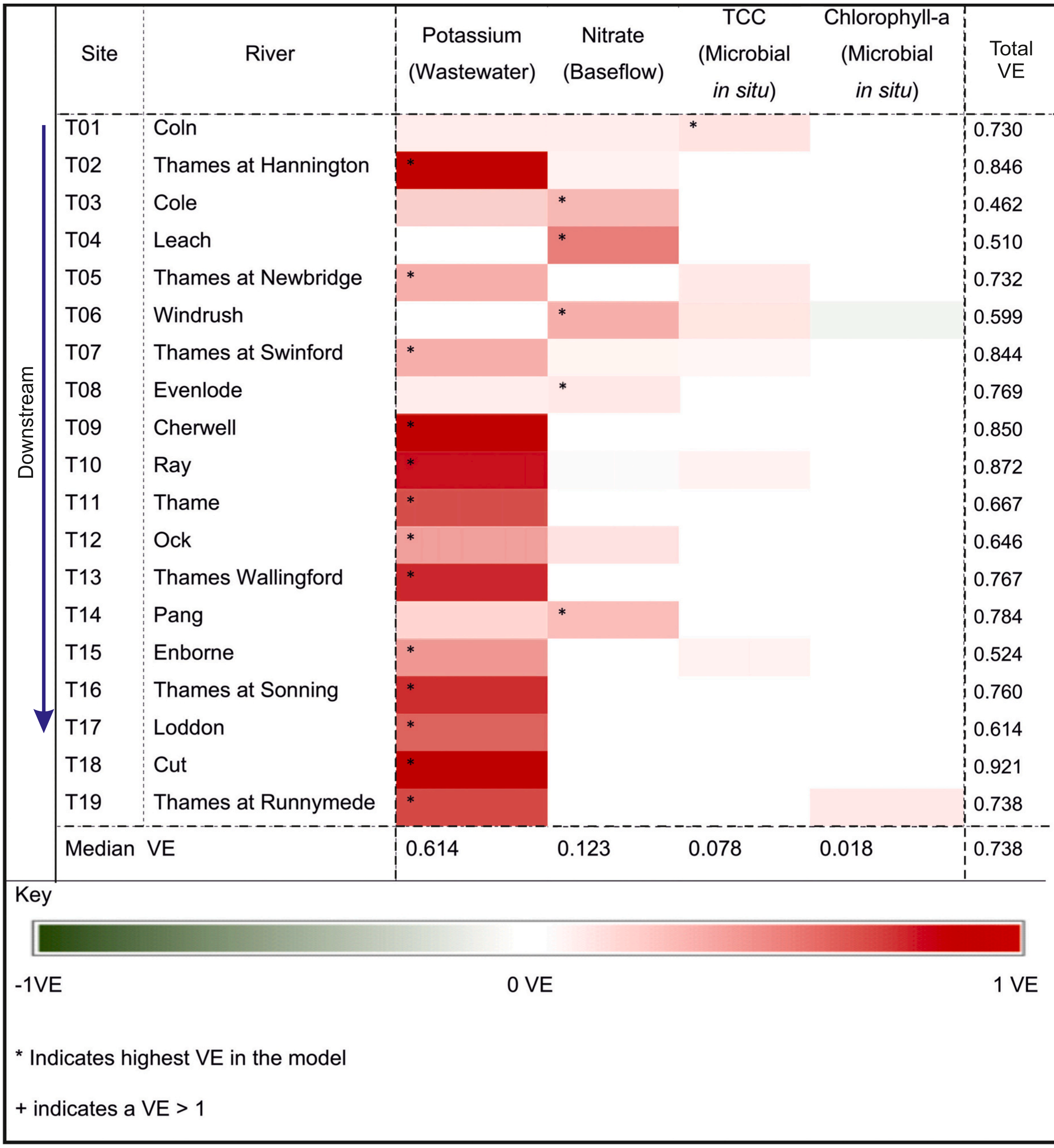


Fig. 4. Results of linear mixed models for each site, where TLF is the objective variable and Potassium, Nitrate, TCC and Chlorophyll-a possible predictors. Starting with the site furthest upstream and ending with the site further downstream in the catchment. The colour scale denotes the variance of TLF explained(VE) by each predictor in the model with dark orange denoting near 1 VE, yellow indicating a VE less than or equal to 0, and white denoting no occurrence in the model The final column contains total model VE.

with statistically significant evidence for inclusion in the model for TCC were tributary sites (four *versus* two). On top of this there was an appreciable difference between the median Z-scores of Thames and tributary models that use TCC (median Z-Score = 3.74, 4.59). Additionally, two sites that use chlorophyll-a as a predictor, one is situated on the Thames and one is situated on a tributary.

For VE, TCC has the third highest median VE (Fig. 4) and had the highest VE in one out of 19 models again site T01 (Fig. 4). Again, there was an appreciable difference between the VE of Thames and tributary models (median VE = 0.059, 0.088). Chlorophyll-a VE had the lowest median VE across all models (0.018) and didn't have the highest VE for any model (Fig. 4).

The models of Thames sites using TCC were the Thames sites that tended to have larger catchment areas and moderate BFI and arable land cover, and lower urban land cover and SPE (Fig. 7A). There is also a cluster of three of the four models of tributary sites that had lower flows, smaller catchments, and more woodland/grassland land cover (Fig. 7A). There was no consistent trend for models using chlorophyll-a as a predictor and catchment characteristics (Fig. 7B).

4. Discussion

4.1. Drivers of TLF dynamics and links to catchment characteristics

Wastewater (estimated using potassium concentrations as a proxy) had the strongest evidence for inclusion in models for surface water TLF in 63 % of sites and explained the most variance of TLF in 68 %. These were generally models for sites with the highest sewerage loading and

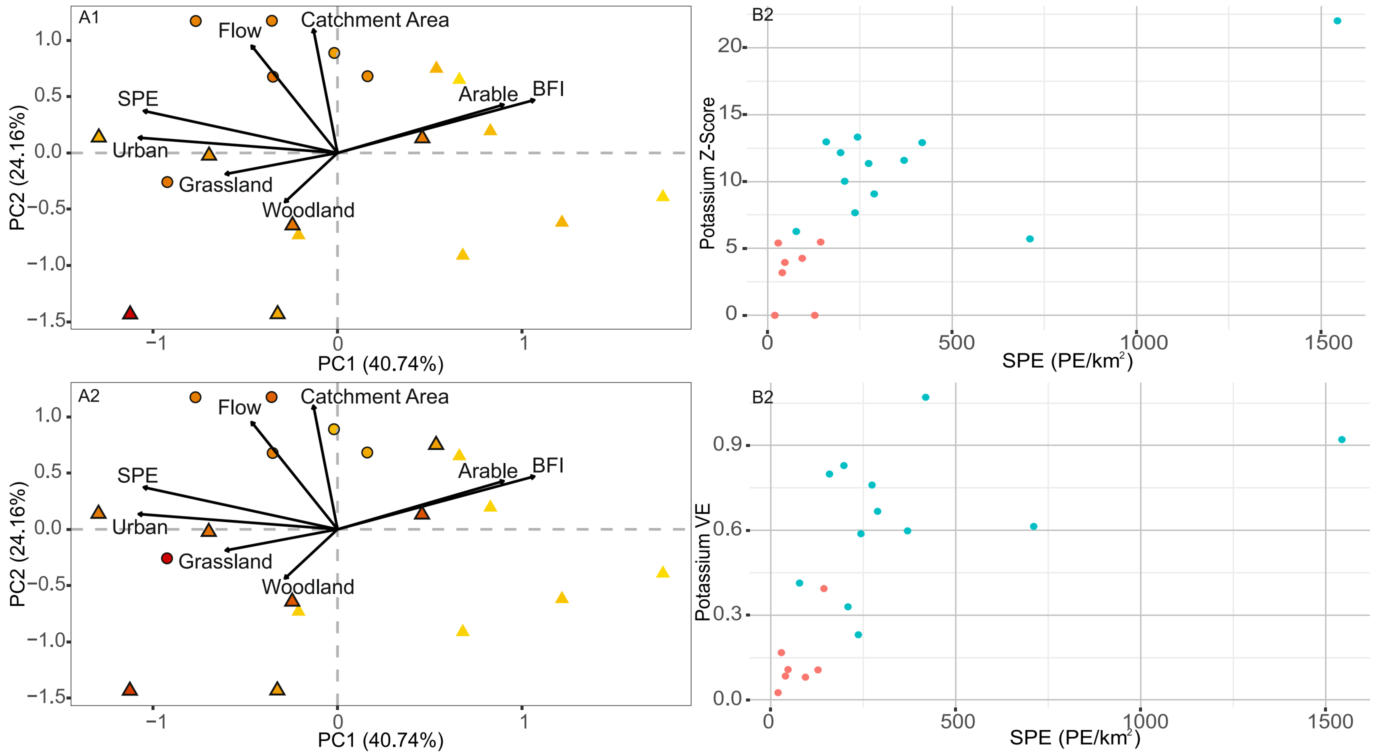


Fig. 5. A: Principal component analysis of site characteristics where a colour scale from yellow(0) to red (high) indicates (A1) potassium Z-score and (A2) potassium VE. Circular points denote a Thames site and triangular a tributary. Points with an outline indicate potassium has the highest Z-score or VE at that site's model. B: Scatterplot of Potassium Z-score (B1) VE (B2) verses SPE. Blue denotes Potassium has the highest Z-score or VE at that site's model.

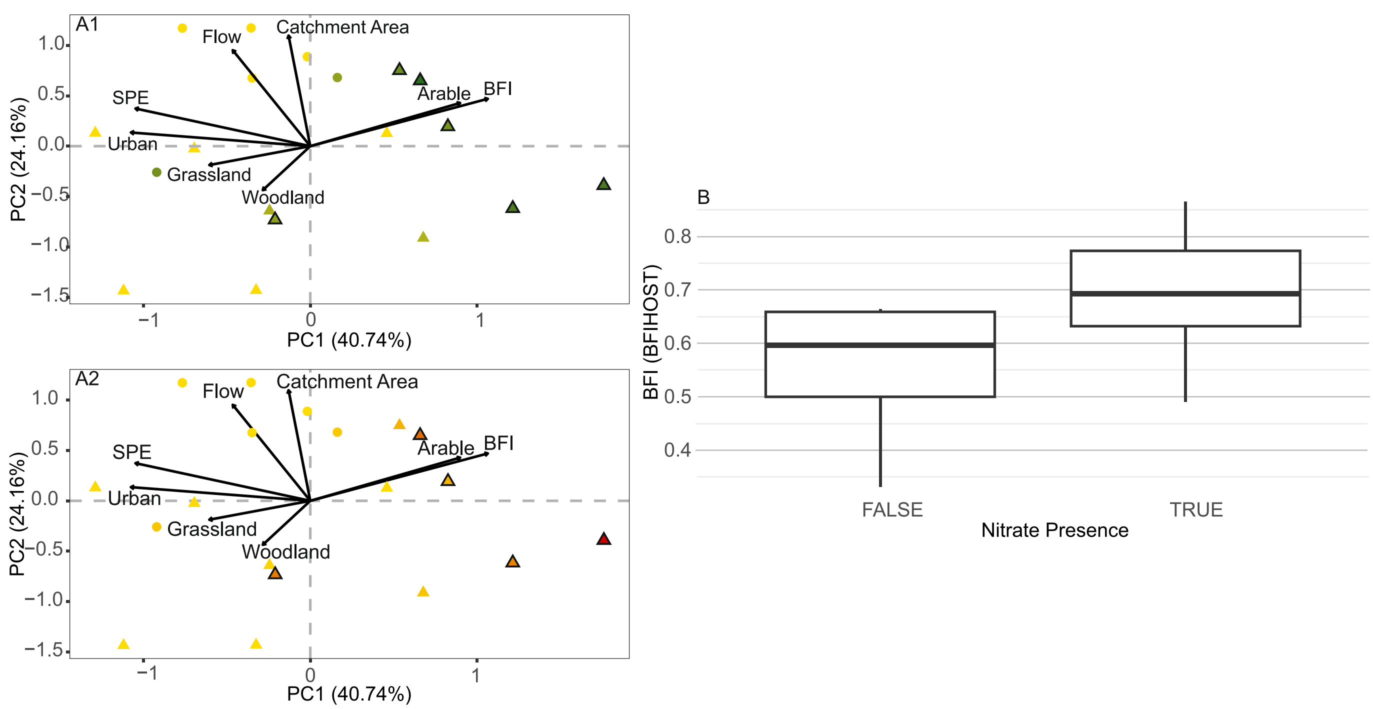


Fig. 6. A: Principal component analysis of site characteristics with site points coloured using colour scales according to nitrate mixed linear model results. For A1 a colour scale from yellow(near 0) to dark green(−10) indicates a nitrate Z-score and for A2 a colour scale from red(1) to yellow(0) VE. Circular points denote a Thames site and triangular denotes a tributary site. Points with an outline indicate potassium has the highest Z-score or VE at that site's model. B: Boxplots of BFI are split into sites that use Nitrate as a predictor and those that don't.

proportion of urban area (Fig. 5). Indeed, there was a strong positive (r > 0.7) Spearman's Rank correlations between SPE and potassium Z-score and VE. Whilst wastewater was a significant predictor of TLF dynamics in all sites, it was in 89 % which supports the traditional view in literature (Baker et al., 2003; Carstea et al., 2010; Hudson et al., 2007; Old et al., 2019).

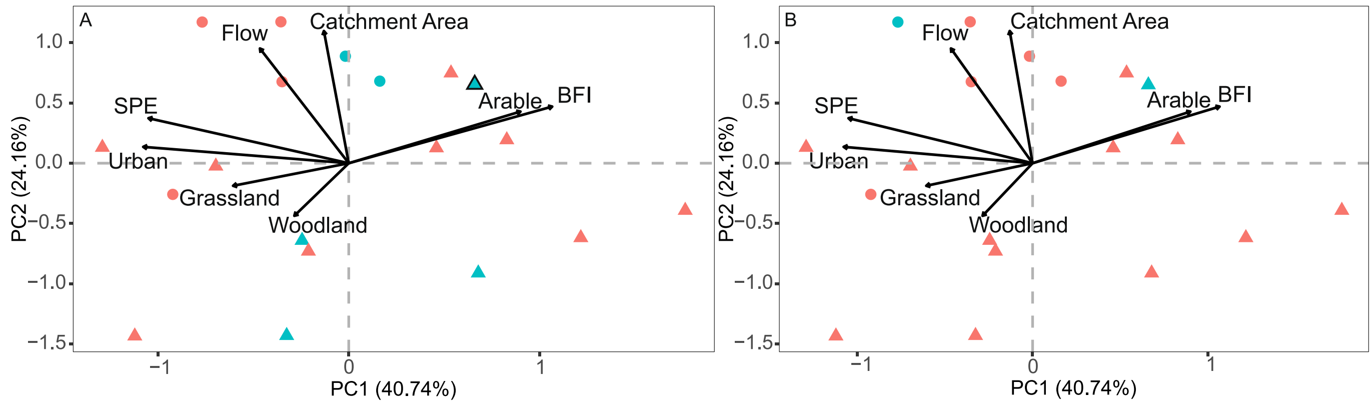


Fig. 7. Principal component analysis of site characteristics. Blue denotes the site uses A: TCC or B: Chlorophyll-a within the model. Red indicates no occurrence of TCC or Chlorophyll-a within that site's model. Circular points denote a Thames site and triangular denotes a tributary site. Points with an outline denote that TCC has the highest Z-score at that site's model.

Nitrate from groundwater sources had statistically significant evidence for inclusion in the TLF models for 53 % of sites and had the strongest evidence for inclusion in models for TLF dynamics in 32 % of sites and explained the most variance in 26 %. Indeed, at T04 it was the sole variable included in the model for TLF. These sites tend to be further upstream where groundwater indices are higher and land use is dominated by arable agriculture (Fig. 6A1–2). These sites are not pristine and do all receive some wastewater inputs from Wastewater Treatment Works (WwTW) but to a lesser extent than those further downstream. Groundwater serves as a negative control on TLF in surface water catchments, where higher contributions of groundwater dilute the wastewater source. Groundwater typically contains lower amounts of dissolved organic matter than surface water (Harjung et al., 2023). Furthermore, groundwater DOM is considered to be more recalcitrant than surface waters, as more labile DOM is preferentially broken down by biotic and abiotic processes as water passes through soils (Roth et al., 2019). For example, even vulnerable groundwater-derived public water sources experiencing frequent faecal breakthroughs in the UK, including in the Thames Catchment, had limited evidence of a TLF peak following PARAFAC analysis (Sorensen et al., 2018b). This is similar to a previous study on this data, which looked at the median data across each site (Old et al., 2019).

Microbial sources emerged as a significant predictor of TLF dynamics in 37 % of sites. Phytoplankton biomass was only a significant predictor at two sites, and never had the strongest evidence for inclusion in the models, despite the seasonal occurrence and subsequent breakdown of phytoplankton blooms across the catchment (See Fig. S1.3). Both models use the phytoplankton biomass predictor with positive Z-Scores, suggesting a positive relationship between phytoplankton abundance and riverine TLF, as seen in marine studies (Chari et al., 2019). However, there was no consistent trend between sites that had statistically significant evidence for the inclusion of chlorophyll-a in the models and catchment characteristics (Fig. 7B, Fig. S1.7). Bacterioplankton abundance (TCC) was used in 32 % of models and had the strongest evidence of inclusion in the models in 5 % of models and explained the most variance in 5 %. Similarly, for phytoplankton, all had positive Z-Scores, again suggesting greater bacterial abundance is related to TLF in the river. Importantly, there is a tendency for sites where TCC is a significant predictor to have a lower sewage input (Median SPE for sites without inclusion of TCC = 244PE/km^2, Median SPE for sites with inclusion of TCC = 143.2 PE/km^2). Also, there was not any trends with other site characteristics (Fig. 7a, Fig. S1.6). Indeed, at these sites TCC correlates weakly with potassium, thus suggesting that bacterial cells did not directly originate from WwTWs.

Our study demonstrates a balance of processes is at play across most sites, as 63 % of site models use more than one predictor despite our adjustment to account for multiple hypothesis test. Indeed, the consistency or dynamism of the source will influence its dominance in the model. For instance, the 37 % of sites (seven models) that only use one predictor, six use the wastewater predictor potassium. Two of these sites have the largest SPE (T17 = 711PE/km^2, T18 = 1522 PE/km^2) and all surpass the median SPE for all sites (198 PE/km^2). Therefore, the wastewater source of TLF is likely to be predominant and overshadow other sources, processing, or dilution of TLF. This impacts how well we can separate out or even find evidence for other less dominant sources. For instance, the *in situ* microbial drivers, are not likely to be constant and are highly seasonal. However, when we separate these from wastewater, as done in this study, we can see that *in situ* microbial drivers are observed and even can have stronger evidence of inclusion in the models when wastewater inputs are lower. Overall, this means the relative balance of the sources of TLF are highly dependent on individual site characteristics which affect potential for the autochthonous and allochthonous dominance of the TLF source (Wilson and Xenopoulos, 2008).

Under future environmental and population changes, the drivers of TLF may alter in the Thames Basin. The predicted combination of lower flows and increased urbanisation may mean a larger input of sewage into The Thames, unless alternative disposal pathways are implemented (Bussi et al., 2016b; Hutchins et al., 2018; Johnson et al., 2009). This could mean an increase in wastewater as a dominant driver of TLF and reduced dilution through groundwater, as groundwater contribution to river flows is also predicted to decline (Hutchins et al., 2018). Both of these processes are likely to increase TLF intensity. Lastly, some of our rivers are warming in response to environmental changes (Johnson et al., 2009). Further increases in river temperatures combined with lower flow could lead to greater riverine microbial activity, resulting in microbial processes becoming more important as a driver of TLF dynamics (Bussi et al., 2016b; Johnson et al., 2009). In the laboratory Elliott et al. (2006) and Fox et al. (2017) both found that increased incubation temperature increased TLF production. Indeed, for algae, Bussi et al. (2016b) and Hutchins et al. (2018) concluded environmental, population and land use changes will lead to an extended phytoplankton growing season on The Thames. Taken together, it is likely that we will see an increase in the sources and intensity of TLF in the Thames at its tributaries.

4.2. Limitations and future work

There are two limitations to using bacterioplankton cell counts as a predictor for TLF dynamics. Firstly, laboratory studies have found stronger relationships between TLF and certain microbial taxa (Bridgeman et al., 2015; Fox et al., 2017; Villacorte et al., 2015). Secondly, Fox et al. (2017) found the *in-situ* production of organic matter displaying TLF could be better quantified by bacterial activity rather than bacterial enumeration in laboratory experiments. Therefore, it is possible that *in situ* microbial processes may have a more influential impact on riverine TLF dynamics than described herein and future work should investigate the influence of microbial activity and microbial community composition on TLF dynamics.

Furthermore, there are limitations of a weekly data resolution. Short-term pulses of TLF fluorophores could be missed, *e.g.*, from precipitation, Combined Sewerage Overflow (CSO) pollution or phytoplankton activity. *In situ* fluorescence sensors have revealed that pulses of fluorescent OM can be driven by precipitation events with less than a week's duration (Carstea et al., 2009; Croghan et al., 2021; Khamis et al., 2020), notably in response to more short-term precipitation events (Carstea et al., 2009). CSO pollution events can last a few hours, frequently in response to a WwTWs being overloaded by rainfall or blockages in the distribution network (Giakoumis and Voulvoulis, 2023). This means short-term pollution events could be missed at weekly sampling resolution, which could have a critical impact on the TLF measurements, as research has shown an event like this would have a high TLF signal (Baker et al., 2003).

Many factors play into how quickly a phytoplankton bloom forms and how quickly it ends; for example, hydrology (Bowes et al., 2012; Hardenbicker et al., 2014; Reynolds and Descy, 1996), temperature (Desortová and Punčochář, 2011) and nutrient loading (Bowes et al., 2012; Tavernini et al., 2011; Wu et al., 2011). Phytoplankton blooms can be short-lived and weekly sampling may not adequately capture the dynamics of the bloom, missing peaks, and subsequent breakdown (Dubelaar et al., 2004; Thyssen et al., 2008). An illustrative case is site T01 (see Fig. S1.3), where there seems to be a bloom at the beginning of May. However, this is confirmed by only two high data points and the data resolution does not capture the subsequent breakdown, during which large quantities of organic matter can be mobilised (Stedmon and Markager, 2005; Villacorte et al., 2015). Therefore, future work should improve the temporal resolution of TLF dynamics during particular events such as phytoplankton blooms and/or CSO pollution events.

When examining the sampling period of the study, the conditions were fairly typical for the majority of sites. Across the sites, the median percentage difference between mean flow across the sampling period and mean flow from National River Flow Archive (NRFA) period of record was −5.3 %. However, there were higher flows than the maximum previously recorded by the NFRA across the summer and winter of 2012 (See Fig. S1.8). There was some flooding that occurred in the winter of 2012 at sites T01 – T12, and spring of 2013 at sites T13-T14, T16-T19. Higher seasonal flows across the catchment could have potentially altered the relationship found between groundwater and TLF at some of the sites.

5. Conclusions

Wastewater proxy (potassium) had statistically significant evidence of inclusion into the models of riverine dissolved TLF dynamics in 17 out of 19 sites of the anthropogenically impacted River Thames, UK. However, wastewater only had the strongest evidence for inclusion in the models of only 63 % sites. Groundwater proxy (nitrate) emerged as having the strongest evidence of inclusion in 32 % of site models and was included negatively in 53 % of models. Microbial sources (TCC and chlorophyll-a) were included positively in the models of seven sites, with bacterial cells surpassing wastewater or baseflow as having the strongest evidence for inclusion in one site model. Bacterial cell counts were utilised in more models than phytoplankton biomass, which was only used in two sites, and never had the strongest evidence for inclusion in the models, despite the seasonal phytoplankton blooms across the Thames site.

The relative importance of these predictors for the model of each site's TLF dynamics was determined by the characteristics of the sites. For example, the higher the sewage loading, the stronger the evidence of inclusion in the model the wastewater predictor had, with the four site models only using wastewater as a predictor having above the median SPE in the data set. There was no evidence to suggest that the bacterial cells systematically emanate from wastewater treatment works. Therefore, the bacterial contribution of dissolved TLF may relate to *in situ* processing of organic matter. Our study underscores the complex interplay of wastewater, baseflow, and microbial sources, driving TLF dynamics in riverine environments, with their influence determined by site characteristics.

CRediT authorship contribution statement

N.A. Harris: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **J.P.R. Sorensen:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **B. Marchant:** Writing – review & editing, Methodology, Investigation, Formal analysis. **G.H. Old:** Writing – review & editing, Methodology, Data curation. **P.S. Naden:** Writing – review & editing, Methodology, Data curation. **M.J. Bowes:** Writing – review & editing, Methodology, Data curation. **P.M. Scarlett:** Writing – review & editing, Methodology, Data curation. **D.J.E. Nicholls:** Data curation. **L. K. Armstrong:** Data curation. **H.D. Wickham:** Data curation. **D.S. Read:** Writing – review & editing, Supervision, Methodology, Data curation. **D. Lapworth:** Writing – review & editing, Supervision. **T. Bond:** Writing – review & editing, Supervision. **K. Pond:** Writing – review & editing, Supervision.

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.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by Scenario Doctoral Training Partnership and University of Surrey Doctoral Programme. The work was completed as part of a PhD thesis funded and supported by NERC Scenario DTP. JPRS and DJL publish with the permission of the Executive Director of the British Geological Survey (UKRI). JPRS acknowledges NERC large grant SMARTWATER: Diagnosing controls of pollution hot spots and hot moments and their impact on catchment water quality (NE/X018857/1). The Centre for Ecology and Hydrology (CEH) supported *via* two Natural Environment Research Council (NERC) funded activities: (1) the Thames Initiative research platform, and (2) the CEH River Lambourn Observatory enabled the production of the water quality data reported in this paper. G. Old also acknowledges support from the research programme NE/N018125/1 LTS-M ASSIST - Achieving Sustainable Agricultural Systems (www.assist.ceh.ac.uk), funded by NERC and the Biotechnology and Biological Sciences Research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.172285.

References

Environment Agency, 2021. Thames River Basin District flood risk management plan 2021 to 2027. https://www.gov.uk/government/publications/thames-river-basin-district-flood-risk-management-plan.

Ahmad, S.R., Reynolds, D.M., 1999. Monitoring of water quality using fluorescence technique: Prospect of on-line process control. In: Water Research, Vol. 33 No. 9.

Elsevier Science Ltd, pp. 2069–2074. https://doi.org/10.1016/S0043-1354(98)00435-7.

Akhtar, N., Syakir Ishak, M.I., Bhawani, S.A., Umar, K., 2021. Various Natural and Anthropogenic Factors Responsible for Water Quality Degradation: A Review. In: Water 2021, Vol. 13 No. 19. Multidisciplinary Digital Publishing Institute, p. 2660. https://doi.org/10.3390/W13192660.

Arthington, A.H., Naiman, R.J., McClain, M.E., Nilsson, C., 2010. Preserving the biodiversity and ecological services of rivers: new challenges and research opportunities. In: Freshwater Biology, Vol. 55 No. 1. John Wiley & Sons, Ltd, pp. 1–16. https://doi.org/10.1111/J.1365-2427.2009.02340.X.

Baker, A., 2001. Fluorescence excitation−emission matrix characterization of some sewage-impacted rivers. Environ. Sci. Technol. 35 (5), 948–953. https://doi.org/10.1021/es000177t.

Baker, A., 2002. Spectrophotometric discrimination of river dissolved organic matter. Process 16, 3203–3213. https://doi.org/10.1002/hyp.1097.

Baker, A., Inverarity, R., 2004. Protein-like fluorescence intensity as a possible tool for determining river water quality. Hydrol. Process. 18 (15), 2927–2945. https://doi.org/10.1002/hyp.5597.

Baker, A., Inverarity, R., Charlton, M., Richmond, S., 2003. Detecting river pollution using fluorescence spectrophotometry: case studies from the Ouseburn, NE England. In: Environmental Pollution, 124 No. 1. Elsevier, pp. 57–70. https://doi.org/10.1016/S0269-7491(02)00408-6.

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?. In: Science of The Total Environment, 532. Elsevier, pp. 14–19. https://doi.org/10.1016/j.scitotenv.2015.05.114.

Bloomfield, J.P., Allen, D.J., Griffiths, K.J., 2009. Examining geological controls on baseflow index (BFI) using regression analysis: an illustration from the Thames Basin, UK. J. Hydrol. 373 (1–2), 164–176. https://doi.org/10.1016/j.jhydrol.2009.04.025.

Blough, N.V., Zafiriou, O.C., Bonilla, J., 1993. Optical absorption spectra of waters from the Orinoco River outflow: terrestrial input of colored organic matter to the Caribbean. J. Geophys. Res. 98 (C2), 2271–2278. https://doi.org/10.1029/92JC02763.

Bonferroni, C.E., 1936. Teoria statistica delle classi e calcolo delle probabilità. In: Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di. Seeber, Firenze Italy.

Bowes, M.J., Gozzard, E., Johnson, A.C., Scarlett, P.M., Roberts, C., Read, D.S., Armstrong, L.K., et al., 2012. Spatial and temporal changes in chlorophyll-a concentrations in the river Thames basin, UK: are phosphorus concentrations beginning to limit phytoplankton biomass? Science of The Total Environment, Elsevier 426, 45–55. https://doi.org/10.1016/j.scitotenv.2012.02.056.

Bowes, M.J., Jarvie, H.P., Naden, P.S., Old, G.H., Scarlett, P.M., Roberts, C., Armstrong, L.K., et al., 2014. Identifying priorities for nutrient mitigation using river concentration–flow relationships: the Thames basin, UK. Journal of Hydrology, Elsevier 517, 1–12. https://doi.org/10.1016/j.jhydrol.2014.03.063.

Bowes, M.J., Armstrong, L.K., Harman, S.A., Wickham, H.D., Nicholls, D.J.E., Scarlett, P.M., Roberts, C., et al., 2018. Weekly water quality monitoring data for the river Thames (UK) and its major tributaries (2009–2013): the Thames initiative research platform. Earth System Science Data, Copernicus GmbH 10 (3), 1637–1653. https://doi.org/10.5194/essd-10-1637-2018.

Bridgeman, J., Bieroza, M., Baker, A., 2011. The application of fluorescence spectroscopy to organic matter characterisation in drinking water treatment. Rev. Environ. Sci. Biotechnol. 10 (3), 277–290. https://doi.org/10.1007/s11157-011-9243-x.

Bridgeman, J., Baker, A., Brown, D., Boxall, J.B., 2015. Portable LED fluorescence instrumentation for the rapid assessment of potable water quality. Science of The Total Environment, Elsevier 524–525, 338–346. https://doi.org/10.1016/j.scitotenv.2015.04.050.

Bussi, G., Dadson, S.J., Prudhomme, C., Whitehead, P.G., 2016a. Modelling the Future Impacts of Climate and Land-Use Change on Suspended Sediment Transport in the River Thames (UK). https://doi.org/10.1016/j.jhydrol.2016.09.010.

Bussi, G., Whitehead, P.G., Bowes, M.J., Read, D.S., Prudhomme, C., Dadson, S.J., 2016b. Impacts of climate change, land-use change and phosphorus reduction on phytoplankton in the river Thames (UK). Science of The Total Environment, Elsevier 572, 1507–1519. https://doi.org/10.1016/j.scitotenv.2016.02.109.

Cammack, W.K.L., Kalff, J., Prairie, Y.T., Smith, E.M., 2004. Fluorescent dissolved organic matter in lakes: relationships with heterotrophic metabolism. Limnol. Oceanogr. 49 (6), 2034–2045. https://doi.org/10.4319/lo.2004.49.6.2034.

Carstea, E.M., Baker, A., Pavelescu, G., Boomer, I., 2009. Continuous fluorescence assessment of organic matter variability on the Bournbrook River, Birmingham, UK. Hydrological Processes 23 (13), 1937–1946. https://doi.org/10.1002/hyp.7335. John Wiley & Sons, Ltd.

Carstea, E.M., Baker, A., Bieroza, M., Reynolds, D., 2010. Continuous fluorescence excitation–emission matrix monitoring of river organic matter. Water Research, Pergamon 44 (18), 5356–5366. https://doi.org/10.1016/j.watres.2010.06.036.

Carstea, E.M., Popa, C.L., Baker, A., Bridgeman, J., 2020. In situ fluorescence measurements of dissolved organic matter: A review. In: Science of The Total Environment, vol. 699. Elsevier B.V., p. 134361. https://doi.org/10.1016/j.scitotenv.2019.134361

Chari, N.V.H.K., Pandi, S.R., Kanuri, V.V., Basuri, C.K., 2019. Structural variation of coloured dissolved organic matter during summer and winter seasons in a tropical estuary: A case study. Marine Pollution Bulletin, Pergamon 149, 110563. https://doi.org/10.1016/j.marpolbul.2019.110563.

Chen, M., Price, R.M., Yamashita, Y., Jaffé, R., 2010. Comparative study of dissolved organic matter from groundwater and surface water in the Florida coastal Everglades using multi-dimensional spectrofluorometry combined with multivariate statistics. Appl. Geochem. 25 (6), 872–880. https://doi.org/10.1016/j.apgeochem.2010.03.005.

Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Marine Chemistry, Elsevier 51 (4), 325–346. https://doi.org/10.1016/0304-4203(95)00062-3.

Croghan, D., Khamis, K., Bradley, C., Van Loon, A.F., Sadler, J., Hannah, D.M., 2021. Combining in-situ fluorometry and distributed rainfall data provides new insights into natural organic matter transport dynamics in an urban river. Science of The Total Environment, Elsevier 755, 142731. https://doi.org/10.1016/j.scitotenv.2020.142731.

Dawson, F.H., Hornby, D.D., Hilton, J., 2002. A method for the automated extraction of environmental variables to help the classification of rivers in Britain. Aquat. Conserv. Mar. Freshwat. Ecosyst. 12 (4), 391–403. https://doi.org/10.1002/aqc.534.

Desortová, B., Punčochář, P., 2011. Variability of phytoplankton biomass in a lowland river: response to climate conditions. Limnologica 41 (3), 160–166. https://doi.org/10.1016/j.limno.2010.08.002.

Dodds, W.K., Perkin, J.S., Gerken, J.E., 2013. Human impact on freshwater ecosystem services: A global perspective. Environmental Science and Technology, American Chemical Society 47 (16), 9061–9068. https://doi.org/10.1021/es4021052.

Dubelaar, G.B.J., Geerders, P.J.F., Jonker, R.R., 2004. High frequency monitoring reveals phytoplankton dynamics. Journal of Environmental Monitoring, Royal Society of Chemistry 6 (12), 946–952. https://doi.org/10.1039/B409350J.

Elliott, S., Lead, J.R., Baker, A., 2006. Characterisation of the fluorescence from freshwater, planktonic bacteria. Water Research, Pergamon 40 (10), 2075–2083. https://doi.org/10.1016/j.watres.2006.03.017.

Environment Agency, 2023. Water Framework Directive 2000/60/EC (WFD) Classification Status Cycle 2 Dataset (25 September).

Fellman, J.B., Hood, E., Spencer, R.G.M., 2010. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. Limnology and Oceanography 55 (6), 2452–2462. https://doi.org/10.4319/lo.2010.55.6.2452. John Wiley & Sons, Ltd.

Fox, B.G., Thorn, R.M.S., Anesio, A.M., Reynolds, D.M., 2017. The in situ bacterial production of fluorescent organic matter; an investigation at a species level. Water Research, Elsevier Ltd 125, 350–359. https://doi.org/10.1016/j.watres.2017.08.040.

Fox, B.G., Thorn, R.M.S., Anesio, A.M., Cox, T., Attridge, J.W., Reynolds, D.M., 2018. Microbial processing and production of aquatic fluorescent organic matter in a model freshwater system. Water (Switzerland), MDPI AG 11 No. 1. https://doi.org/10.3390/w11010010.

Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., 2002. The UK land cover map 2000: construction of a parcel-based vector map from satellite images. The Cartographic Journal, University of Aberdeen 39 (1), 15–25. https://doi.org/10.1179/caj.2002.39.1.15.

Giakoumis, T., Voulvoulis, N., 2023. Combined sewer overflows: relating event duration monitoring data to wastewater systems' capacity in England. Environmental Science: Water Research & Technology, Royal Society of Chemistry 9 (3), 707–722. https://doi.org/10.1039/d2ew00637e.

Hannah, D.M., Abbott, B.W., Khamis, K., Kelleher, C., Lynch, I., Krause, S., Ward, A.S., 2022. Illuminating the 'invisible water crisis' to address global water pollution challenges. Hydrological Processes 36 (3), e14525. https://doi.org/10.1002/hyp.14525. John Wiley & Sons, Ltd.

Hardenbicker, P., Rolinski, S., Weitere, M., Fischer, H., 2014. Contrasting long-term trends and shifts in phytoplankton dynamics in two large rivers. Int. Rev. Hydrobiol. 99 (4), 287–299. https://doi.org/10.1002/iroh.201301680.

Harjung, A., Schweichhart, J., Rasch, G., Griebler, C., 2023. Large-scale study on groundwater dissolved organic matter reveals a strong heterogeneity and a complex microbial footprint. Science of The Total Environment, Elsevier 854, 158542. https://doi.org/10.1016/j.scitotenv.2022.158542.

Henderson, R.K., Baker, A., Parsons, S.A., Jefferson, B., 2008. Characterisation of algogenic organic matter extracted from cyanobacteria, green algae and diatoms. Water Research, Pergamon 42 (13), 3435–3445. https://doi.org/10.1016/j.watres.2007.10.032.

Holbrook, R.D., DeRose, P.C., Leigh, S.D., Heckert, N.A., Rukhin, A.L., 2006. Excitation–Emission Matrix Fluorescence Spectroscopy for Natural Organic Matter Characterization: A Quantitative Evaluation of Calibration and Spectral Correction Procedures. Applied Spectroscopy 60 (7), 791–799. https://doi.org/10.1366/000370206777886973. Society for Applied Spectroscopy.

Huang, J., Zhang, Y., Bing, H., Peng, J., Dong, F., Gao, J., Arhonditsis, G.B., 2021. Characterizing the river water quality in China: recent progress and on-going challenges. Water Research, Pergamon 201, 117309. https://doi.org/10.1016/j.watres.2021.117309.

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. https://doi.org/10.1002/RRA.1005.

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. Science of The Total Environment, Elsevier 391 (1), 149–158. https://doi.org/10.1016/j.scitotenv.2007.10.054.

Hur, J., Cho, J., 2012. "Prediction of BOD, COD, and Total nitrogen concentrations in a typical Urban River using a fluorescence excitation-emission matrix with PARAFAC and UV absorption indices", sensors 2012, Vol. 12, pages 972-986. Molecular Diversity Preservation International 12 (1), 972–986. https://doi.org/10.3390/S120100972.

Hur, J., Hwang, S.-J., Shin, J.-K., 2008. Using synchronous fluorescence technique as a water quality monitoring tool for an Urban River. Water, Air, and Soil Pollution 191 (1), 231–243. https://doi.org/10.1007/S11270-008-9620-4, 2007 191:1, Springer.

Hutchins, M.G., Abesser, C., Prudhomme, C., Elliott, J.A., Bloomfield, J.P., Mansour, M. M., Hitt, O.E., 2018. Combined impacts of future land-use and climate stressors on water resources and quality in groundwater and surface waterbodies of the upper Thames river basin, UK. Science of The Total Environment, Elsevier 631–632, 962–986. https://doi.org/10.1016/j.scitotenv.2018.03.052.

I. Khan, S., Zamyadi, A., Rao, N.R.H., Li, X., Stuetz, R.M., Henderson, R.K., 2019. Fluorescence spectroscopic characterisation of algal organic matter: towards improved in situ fluorometer development. Environ. Sci.: Water Res. Technol. 5 (2), 417–432. https://doi.org/10.1039/c8ew00731d.

Johnson, A.C., Acreman, M.C., Dunbar, M.J., Feist, S.W., Giacomello, A.M., Gozlan, R.E., Hinsley, S.A., et al., 2009. The British river of the future: how climate change and human activity might affect two contrasting river ecosystems in England. Science of The Total Environment, Elsevier 407 (17), 4787–4798. https://doi.org/10.1016/j.scitotenv.2009.05.018.

Keller, V., Fox, K., Rees, H.G., Young, A.R., 2006. Estimating population served by sewage treatment works from readily available GIS data. Science of The Total Environment, Elsevier 360 (1–3), 319–327. https://doi.org/10.1016/j.scitotenv.2005.08.043.

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. Hydrological Processes, John Wiley and Sons Ltd 31 (3), 540–555. https://doi.org/10.1002/hyp.11040.

Khamis, K., Bradley, C., Hannah, D.M., 2020. High frequency fluorescence monitoring reveals new insights into organic matter dynamics of an urban river, Birmingham, UK. Science of The Total Environment, Elsevier 710, 135668. https://doi.org/10.1016/j.scitotenv.2019.135668.

Kristensen, P., Whalley, C., Zal, Ner Nihat, F. and Christiansen, T., 2018. European Waters Assessment of Status and Pressure 2018, Copenhagen.

Lakowicz, J.R., 2006. "Principles of Fluorescence Spectroscopy", Principles of Fluorescence Spectroscopy. Springer, pp. 1–954. https://doi.org/10.1007/978-0-387-46312-4.

Lambert, T., Bouillon, S., Darchambeau, F., Morana, C., Roland, F.A.E., Descy, J.P., Borges, A.V., 2017. Effects of human land use on the terrestrial and aquatic sources of fluvial organic matter in a temperate river basin (the Meuse River, Belgium). Biogeochemistry, Springer International Publishing 136 (2), 191–211. https://doi.org/10.1007/s10533-017-0387-9.

Lawaetz, A.J., Stedmon, C.A., 2009. Fluorescence intensity calibration using the Raman scatter peak of water. Appl. Spectrosc. 63 (8), 936–940. https://doi.org/10.1366/000370209788964548.

Li, Y., Wang, M., Chen, X., Cui, S., Hofstra, N., Kroeze, C., Ma, L., et al., 2022. Multi-pollutant assessment of river pollution from livestock production worldwide. Water Res. 209, 117906 https://doi.org/10.1016/j.watres.2021.117906.

Ly, Q.V., Lee, M.-H., Hur, J., 2019. Using fluorescence surrogates to track algogenic dissolved organic matter (AOM) during growth and coagulation/flocculation processes of green algae. J. Environ. Sci. 79, 311–320. https://doi.org/10.1016/j.jes.2018.12.006.

Marchant, B.P., 2018. Model-Based Soil Geostatistics. Springer, Cham, pp. 341–371. https://doi.org/10.1007/978-3-319-63439-5_11.

Marsh, T.J., Hannaford, J. (Eds.), 2008. UK Hydrometric Register. Hydrological data UK series. Centre for Ecology & Hydrology, 210 pp.

Marston, C., Rowland, C., O'Neil, A., Morton, R., 2022. Land Cover Map 2021 (Land Parcels, GB). https://doi.org/10.5285/398dd41e-3c08-47f5-811f-da990007643f.

Nguyen, M.-L., Westerhoff, P., Baker, L., Hu, Q., Esparza-Soto, M., Sommerfeld, M., 2005. Characteristics and reactivity of algae-produced dissolved organic carbon. Journal of Environmental Engineering, American Society of Civil Engineers 131 (11), 1574–1582. https://doi.org/10.1061/(ASCE)0733-9372(2005)131:11(1574).

Noacco, V., Wagener, T., Worrall, F., Burt, T.P., Howden, N.J.K., 2017. Human impact on long-term organic carbon export to rivers. Journal of Geophysical Research: Biogeosciences, Blackwell Publishing Ltd 122 (4), 947–965. https://doi.org/10.1002/2016JG003614.

Old, G.H., Naden, P.S., Harman, M., Bowes, M.J., Roberts, C., Scarlett, P.M., Nicholls, D. J.E., et al., 2019. Using dissolved organic matter fluorescence to identify the provenance of nutrients in a lowland catchment; the River Thames, England. Science of The Total Environment, Elsevier 653, 1240–1252. https://doi.org/10.1016/j.scitotenv.2018.10.421.

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. Organic Geochemistry, Pergamon 31 (12), 1765–1781. https://doi.org/10.1016/S0146-6380(00)00124-8.

Pellerin, B.A., Saraceno, J.F., Shanley, J.B., Sebestyen, S.D., Aiken, G.R., Wollheim, W. M., Bergamaschi, B.A., 2011. Taking the pulse of snowmelt: in situ sensors reveal seasonal, event and diurnal patterns of nitrate and dissolved organic matter variability in an upland forest stream. Biogeochemistry 108 (1), 183–198. https://doi.org/10.1007/S10533-011-9589-8, 2011 108:1, Springer.

du Plessis, A., 2022. Persistent degradation: global water quality challenges and required actions. One Earth, Cell Press 5 (2), 129–131. https://doi.org/10.1016/j.oneear.2022.01.005.

R Core Team, 2023. "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing. Vienna, Austria.

Read, D.S., Gweon, H.S., Bowes, M.J., Newbold, L.K., Field, D., Bailey, M.J., Griffiths, R. I., 2015. Catchment-scale biogeography of riverine bacterioplankton. ISME Journal, Nature Publishing Group 9 (2), 516–526. https://doi.org/10.1038/ismej.2014.166.

Reynolds, C.S., Descy, J.-P., 1996. The production, biomass and structure of phytoplankton in large rivers. Large Rivers, Schweizerbart'sche Verlagsbuchhandlung 10 (1–4), 161–187. https://doi.org/10.1127/LR/10/1996/161.

Roth, V.N., Lange, M., Simon, C., Hertkorn, N., Bucher, S., Goodall, T., Griffiths, R.I., et al., 2019. "Persistence of dissolved organic matter explained by molecular changes during its passage through soil", nature geoscience 2019 12:9. Nat. Publ. Group 12 (9), 755–761. https://doi.org/10.1038/s41561-019-0417-4.

Saraceno, J.F., Pellerin, B.A., Downing, B.D., Boss, E., Bachand, P.A.M., 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 114 (G4), 0–09. https://doi.org/10.1029/2009JG000989. John Wiley & Sons, Ltd.

Sirota, J., Baiser, B., Gotelli, N.J., Ellison, A.M., 2013. Organic-matter loading determines regime shifts and alternative states in an aquatic ecosystem. Proc. Natl. Acad. Sci. 110 (19), 7742–7747. https://doi.org/10.1073/pnas.1221037110.

Sorensen, J.P.R., Baker, A., Cumberland, S.A., Lapworth, D.J., MacDonald, A.M., Pedley, S., Taylor, R.G., et al., 2018a. Real-time detection of faecally contaminated drinking water with tryptophan-like fluorescence: defining threshold values. Science of The Total Environment, Elsevier 622–623, 1250–1257. https://doi.org/10.1016/j.scitotenv.2017.11.162.

Sorensen, J.P.R., Vivanco, A., Ascott, M.J., Gooddy, D.C., Lapworth, D.J., Read, D.S., Rushworth, C.M., et al., 2018b. Online fluorescence spectroscopy for the real-time evaluation of the microbial quality of drinking water. Water Research, Pergamon 137, 301–309. https://doi.org/10.1016/j.watres.2018.03.001.

Sorensen, J.P.R., Diaw, M.T., Pouye, A., Roffo, R., Diongue, D.M.L., Faye, S.C., Gaye, C. B., et al., 2020. In-situ fluorescence spectroscopy indicates total bacterial abundance and dissolved organic carbon. Science of The Total Environment, Elsevier 738, 139419. https://doi.org/10.1016/j.scitotenv.2020.139419.

Sorensen, J.P.R., Nayebare, J., Carr, A.F., Lyness, R., Campos, L.C., Ciric, L., Goodall, T., et al., 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, Pergamon 206, 117734. https://doi.org/10.1016/j.watres.2021.117734.

Stanley, E.H., Powers, S.M., Lottig, N.R., Buffam, I., Crawford, J.T., 2012. Contemporary changes in dissolved organic carbon (DOC) in human-dominated rivers: is there a role for DOC management? Freshwater Biology 57 (Suppl. 1), 26–42. https://doi.org/10.1111/J.1365-2427.2011.02613.X. John Wiley & Sons, Ltd.

Stedmon, C.A., Bro, R., 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. Limnology and Oceanography: Methods 6 (11), 572–579. https://doi.org/10.4319/LOM.2008.6.572. John Wiley & Sons, Ltd.

Stedmon, C.A., Markager, S., 2005. Tracing the production and degradation of autochthonous fractions of dissolved organic matter by fluorescence analysis. Limnology and Oceanography 50 (5), 1415–1426. https://doi.org/10.4319/LO.2005.50.5.1415. John Wiley & Sons, Ltd.

Strokal, M., Spanier, J.E., Kroeze, C., Koelmans, A.A., Flörke, M., Franssen, W., Hofstra, N., Langan, S., Tang, T., van Vliet, M.T., Wada, Y., Wang, M., van Wijnen, J., Williams, R., 2019. Global multi-pollutant modelling of water quality: scientific challenges and future directions. Curr. Opin. Environ. Sustain. 36, 116–125. https://doi.org/10.1016/j.cosust.2018.11.004.

Stuart, M.E., Wang, L., Ascott, M., Ward, R.S., Lewis, M.A., Hart, A.J., 2016. Modelling the Groundwater Nitrate Legacy.

Sumbler, M.G., British Geological Survey, 1996. London and the Thames Valley. Fourth Edition 1996. Date Code 5/96, 4th ed.

Tavernini, S., Pierobon, E., Viaroli, P., 2011. Physical factors and dissolved reactive silica affect phytoplankton community structure and dynamics in a lowland eutrophic river (Po river, Italy). Hydrobiologia 669 (1), 213–225. https://doi.org/10.1007/s10750-011-0688-2.

Thyssen, M., Tarran, G.A., Zubkov, M.V., Holland, R.J., Grégori, G., Burkill, P.H., Denis, M., 2008. The emergence of automated high-frequency flow cytometry: revealing temporal and spatial phytoplankton variability. Journal of Plankton Research, Oxford Academic 30 (3), 333–343. https://doi.org/10.1093/PLANKT/FBN005.

UN DESA, 2022. The Sustainable Development Goals Report, New York.

Van Drecht, G., Bouwman, A.F., Harrison, J., Knoop, J.M., 2009. Global nitrogen and phosphate in urban wastewater for the period 1970 to 2050. Global Biogeochem. Cycles 23 (4). https://doi.org/10.1029/2009GB003458 p. n/a-n/a.

Villacorte, L.O., Ekowati, Y., Neu, T.R., Kleijn, J.M., Winters, H., Amy, G., Schippers, J. C., et al., 2015. Characterisation of algal organic matter produced by bloom-forming marine and freshwater algae. Water Research, Pergamon 73, 216–230. https://doi.org/10.1016/j.watres.2015.01.028.

Wagner, S., Riedel, T., Niggemann, J., Vähätalo, A.V., Dittmar, T., Jaffé, R., 2015. Linking the molecular signature of heteroatomic dissolved organic matter to watershed characteristics in world rivers. Environmental Science and Technology, American Chemical Society 49 (23), 13798–13806. https://doi.org/10.1021/acs.est.5b00525.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 2nd edn. Wiley, Chichester, UK, Print. https://doi.org/10.1002/9780470517277. ISBN: 9780470028582, Online ISBN:9780470517277.

Wen, Y., Schoups, G., van de Giesen, N., 2017. Organic pollution of rivers: combined threats of urbanization, livestock farming and global climate change. Sci. Rep. 7 (1), 43289. https://doi.org/10.1038/srep43289.

Whelan, M.J., Linstead, C., Worrall, F., Ormerod, S.J., Durance, I., Johnson, A.C., Johnson, D., et al., 2022. Is water quality in British rivers 'better than at any time since the end of the industrial revolution'? Science of The Total Environment, Elsevier 843, 157014. https://doi.org/10.1016/j.scitotenv.2022.157014.

Williams, C.J., Yamashita, Y., Wilson, H.F., Jaffe, R., Xenopoulos, M.A., 2010. Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. Limnology and Oceanography 55 (3), 1159–1171. https://doi.org/10.4319/LO.2010.55.3.1159. John Wiley & Sons, Ltd.

Wilson, H.F., Xenopoulos, M.A., 2008. Effects of agricultural land use on the composition of fluvial dissolved organic matter. Nature Geoscience 2 (1), 37–41. https://doi.org/10.1038/ngeo391, 2009 2:1, Nature Publishing Group.

Wu, N., Schmalz, B., Fohrer, N., 2011. Distribution of phytoplankton in a German lowland river in relation to environmental factors. J. Plankton Res. 33 (5), 807–820. https://doi.org/10.1093/plankt/fbq139.

Zhang, L., Fang, W., Li, X., Lu, W., Li, J., 2020. Strong linkages between dissolved organic matter and the aquatic bacterial community in an urban river. Water Research, Pergamon 184, 116089. https://doi.org/10.1016/j.watres.2020.116089.