



Spatial and seasonal controls on dissolved organic matter composition in shallow aquifers under the rapidly developing city of Patna, India

George J.L. Wilson^a, Chuanhe Lu^{a,1}, Dan J. Lapworth^b, Arun Kumar^c, Ashok Ghosh^c,
Vahid J. Niasar^d, Stefan Krause^{e,f}, David A. Polya^a, Daren C. Gooddy^b, Laura A. Richards^{a,*}

^a Department of Earth and Environmental Sciences and Williamson Research Centre for Molecular Environmental Science, The University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL, United Kingdom

^b British Geological Survey, Maclean Building, Wallingford, Oxfordshire OX10 8BB, United Kingdom

^c Mahavir Cancer Sansthan and Research Center, Phulwarisharif, Patna 801505, Bihar, India

^d Department of Chemical Engineering, The University of Manchester, M13 9PL Manchester, United Kingdom

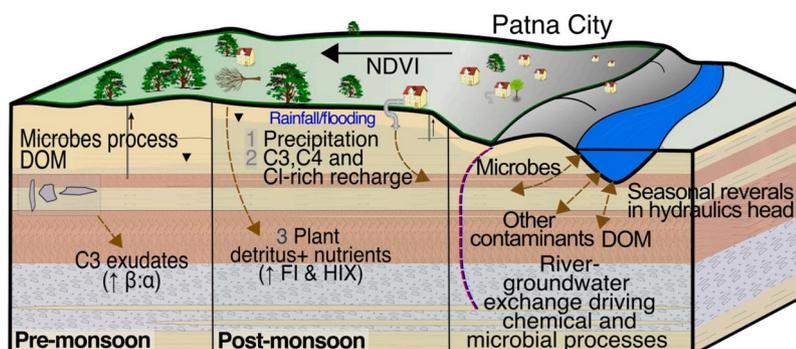
^e School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, B15 2TT Birmingham, United Kingdom

^f LEHNA- Laboratoire d'écologie des hydrosystèmes naturels et anthropiques, University of Lyon, France

HIGHLIGHTS

- Characterisation of fluorescent dissolved organic matter (fDOM) around Patna, Bihar
- Seasonal differences observed in fDOM, particularly tryptophan-like fluorescence.
- Novel technique indicates relationship between vegetation index and fDOM parameters.
- Evidence suggests ingress of surface-derived contamination under the city.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: José Virgílio Cruz

Keywords:

Groundwater contamination
Dissolved organic matter characterisation
Fluorescence spectroscopy
Parallel factor analysis (PARAFAC)
Surface-groundwater interactions

ABSTRACT

The distribution and composition of dissolved organic matter (DOM) affects numerous (bio)geochemical processes in environmental matrices including groundwater. This study reports the spatial and seasonal controls on the distribution of groundwater DOM under the rapidly developing city of Patna, Bihar (India). Major DOM constituents were determined from river and groundwater samples taken in both pre- and post-monsoon seasons in 2019, using excitation-emission matrix (EEM) fluorescence spectroscopy. We compared aqueous fluorescent DOM (fDOM) composition to satellite-derived land use data across the field area, testing the hypothesis that the composition of groundwater DOM, and particularly the components associated with surface-derived ingress, may be controlled, in part, by land use. In the pre-monsoon season, the prominence of tryptophan-like components likely generated from recent biological activity overwhelmed the humic-like and tyrosine-like fluorescence

* Corresponding author.

E-mail address: laura.richards@manchester.ac.uk (L.A. Richards).

¹ (current) Cooling Circuit Department, Safety Research Division, Gesellschaft für Anlagen, und Reaktorsicherheit (GRS) gGmbH, Boltzmannstr. 14, 85748 Garching, Germany.

<https://doi.org/10.1016/j.scitotenv.2023.166208>

Received 21 March 2023; Received in revised form 10 July 2023; Accepted 8 August 2023

Available online 10 August 2023

0048-9697/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

signals. Evidence from fluorescence data suggest groundwater in the post-monsoon season is composed of predominantly i) plant-derived matter and ii) anthropogenically influenced DOM (e.g. tryptophan-like components). Organic tracers, as well as *Eh* and Cl^- , suggest monsoonal events mobilise surface-derived material from the unsaturated zone, causing dissolved organic carbon (DOC) of more microbial nature to infiltrate to >100 m depth. A correlation between higher protein:humic-like fluorescence and lower vegetation index (NDVI), determined from satellite-based land use data, in the post-monsoon season, indicates the ingress of wastewater-derived OM in groundwater under the urban area. Attenuated protein:humic-like fluorescence in groundwater close to the river points towards the mixing of groundwater and river water. This ingress of surface-derived OM is plausibly exacerbated by intensive groundwater pumping under these areas. Our approach to link the composition of aqueous organics with land use could easily be adapted for similar hydrogeochemical settings to determine the factors controlling groundwater DOM composition in various contexts.

1. Introduction

The distribution and chemical composition of dissolved organic matter (DOM) affects many reaction pathways which are significant to human health (Coble et al., 2014; Guo et al., 2019; Postma et al., 2007). The composition of groundwater DOM has implications for other water quality parameters influenced by the microbial utilisation of DOM as a substrate, including dissolved oxygen (Borsuk et al., 2001; Song et al., 2000), nutrients (Burford and Bremner, 1975; Hofmann and Griebler, 2018; Siemens et al., 2003) and trace metal(loid)s such as ferrous iron (Abdelrady et al., 2020; Mladenov et al., 2010) and arsenic (Aftabtalab et al., 2022; Erban et al., 2013; Guo et al., 2019; Kulkarni et al., 2017, 2018; Neumann et al., 2014; Postma et al., 2007; Richards et al., 2019). In addition, groundwater organics reflecting sewage-derived ingress have potential implications on microbial aspects of water quality, particularly in drinking water supplies (Bain et al., 2014; Bivins et al., 2020). The processes affecting the distribution of such constituents in groundwater is particularly relevant in the context of rapidly developing cities in South and Southeast Asia (Lawrence et al., 2000; Richards et al., 2021), who increasingly rely on groundwater potentially influenced by both surface contamination and monsoon weather patterns (Farooq et al., 2010; Milledge et al., 2018; Yang et al., 2020). Quantifying the temporal and spatial heterogeneity in DOM composition therefore provides insight into the vulnerability of groundwater resources and the potential implications to human and ecosystem health. The characterisation of DOM composition can furthermore provide crucial information about its source and subsequent reaction history.

DOM may be characterised using mass spectrometry and chromatographic techniques (Matilainen et al., 2011), though resolving the detailed components can be difficult and expensive (Leenheer and Croué, 2003). Alternatively, a proportion of organic matter is known to fluoresce upon irradiation with UV light, and fluorescent dissolved organic matter (fDOM) measurements provide a powerful, fast and relatively cheap method of characterising DOM in waterbodies (Baker, 2002; Coble et al., 2014; Kowalczyk et al., 2010). Excitation-emission matrices (EEMs) consist of a three-dimensional matrices of fluorescence intensities at defined ranges of excitation/emission wavelengths. Information from EEMs can be used to identify characteristic groupings of organics, and to indicate the source, relative “freshness”, lability and relative processing extent of DOM (McKnight et al., 2001; Parlanti et al., 2000). Due to the relative inexpensiveness of fluorescence measurements, this technique can enable higher resolution (spatially and/or temporally) sampling and interpretation (Stedmon and Markager, 2005) compared to some other more resource-intensive methods of organics characterisation (e.g. nuclear magnetic resonance or liquid chromatography-mass spectrometry; Dittmar and Paeng, 2009; Hertkorn et al., 2013; Matilainen et al., 2011).

Fluorescence can be utilised to distinguish between different fractions of DOM through various EEM post-processing methods. Specific excitation/emission wavelength pairs and ratios between reference regions of the EEM are used to make a number of inferences on the nature of the DOM (Guo et al., 2019; Kulkarni et al., 2017; Murphy et al., 2013; Richards et al., 2019). “Peak-picking” of pre-defined wavelengths

provides a useful proxy for the quantification of humic and fulvic acids, as well as fluorescent amino acids, particularly tryptophan and tyrosine (Coble et al., 2014). The selection of reference wavelength values to represent particular components of DOM presents the problem of superimposed signals from multiple fluorophores of similar fluorescence signatures (Murphy et al., 2013); trilinear methods of signal deconvolution, such as parallel factor analysis (PARAFAC), help to address this potential source of error (Coble et al., 2014; Murphy et al., 2013).

Fluorescence techniques are able to distinguish between DOM of terrestrial and microbial nature – terrestrially derived DOM originates from dispersed sedimentary organic carbon (SOC), subsurface peat deposits or other recent lignaceous plant detritus (Bianchi, 2011; Hedges, 1992), whilst microbially-derived DOM refers to leachate and products of extracellular release from microbes (McKnight et al., 2001). Previous research of DOM fluorescence have inferred hydrological connections between groundwater DOM nature and recharge and inferred connectivity into groundwater and river system (Lapworth et al., 2009; Shen et al., 2014). This linkage may be particularly evident in monsoon (or seasonal rain) affected regions of the world; Sorensen et al. (2021) showed strong precipitation controls on tryptophan-like fluorescence and microbial activity in an aquifer in East Africa. Similarly, Yang et al. (2020) observed an abundance of humic-like, terrestrially derived DOM in groundwaters after periods of heavy rainfall in the Jiangnan Plain, whilst more microbially derived DOM was found during the dry season. Since groundwater may be replenished by dilute rainwater or surface water (which may be comparatively rich in organics, nutrients and/or other solutes), DOM of humic-like, terrestrial nature in groundwater in some areas is thought to be increased through precipitation-driven recharge, including in association with paddy cultivation and irrigation of crops (Chen et al., 2018; Yang et al., 2020). Changes in temperature likely influence microbial activity and net primary productivity in aquifers and subsequently may affect the temporal variation in DOM composition (McDonough et al., 2020). Hofmann et al. (2020) demonstrated that the historical flux of bioavailable DOM influences the available stock of microbial biomass in an aquifer.

In some geochemical settings, a large proportion of the bioavailable DOM flux may originate from surface-derived sources (Guo et al., 2019). Surface-derived DOM under urban areas often originates from industrial effluent (Rodríguez-Vidal et al., 2022), sewage leakage (Held et al., 2006) and bodies of surface water (Wallis et al., 2020). Understanding the factors underpinning the composition of surface-derived organics under urban areas, including the infiltration of effluent and river-groundwater interactions, may be achieved through fDOM characterisation (Rodríguez-Vidal et al., 2022). Amino-like fluorescence can serve as a tracer and relic of anthropogenic waste and has previously been used as a proxy of DOM reactivity in groundwater (Baker et al., 2003; Lapworth et al., 2008; Mladenov et al., 2010; Reynolds, 2003). The ratio of amino-like to recalcitrant DOM components can be quantified to further discriminate uncontaminated from slurry/wastewater-contaminated water. Indeed, Baker (2002) shows that farm waste increases the tryptophan-like to fulvic-like fluorescence ratio in streams. This ratio is also used as a proxy of DOM bioavailability in other groundwater quality studies (Kulkarni et al., 2017; Richards et al.,

2019). The relationship between amino-like:recalcitrant fluorescent DOM and urban land use is something yet to be investigated, particularly in the context of a rapidly developing city, such as Patna, the capital of the State of Bihar, India, where ingress of surface-derived organics may be impacted, in part, by increasing populations and incomplete wastewater infrastructure (Ahmad et al., 2019; Lu et al., 2022; Richards et al., 2021).

Land use is of interest in this context because it may indicate the source of the organic loading to groundwaters (Harjung et al., 2023; Salman et al., 2018; Singh et al., 2020; Williams et al., 2010). The use of satellite remote sensing has been established as a robust method of identifying land cover where land use information is not readily available (Roberts et al., 2003). Indeed, a multitude of operational sensors of varying characteristics provide effective tools for characterising land use and land cover at regional scales (Boyle et al., 2014). The Landsat-8 Operational Land Imager (L8/OLI) sensor captures accurate and timely information; it is considered a flagship method of medium-resolution (10–30 m) remote sensing and has been used in >3000 peer-reviewed pieces of literature (Chaves et al., 2020; USGS, 2019). Conventional vegetation indices are widely applied to the multispectral L8/OLI bands as a proxy for land use (Chaves et al., 2020; Zaitunah et al., 2018). Though, to the authors' knowledge, relationships between groundwater fDOM composition and surface land cover, in terms of satellite-derived vegetation indices, have not yet been established or investigated.

This study aims to decipher the spatial and seasonal controls on DOM distribution in groundwater under Patna, India, a rapidly developing city on the banks of one of the world's most iconic rivers, through bulk DOM quantification and excitation-emission matrix spectroscopy techniques. Previous studies on groundwater bulk DOM concentration in the Indo-Gangetic Plain have reported values of <2 mg/L (Kulkarni et al., 2019; Mohanta and Goel, 2014; Richards et al., 2022b), though few studies have characterised DOM composition (Kulkarni et al., 2019). We quantify the relative amounts of protein-like and humic-like DOM in local surface and groundwaters, using fluorescence indices and modelled PARAFAC components. In addition, we compare the seasonal composition of fDOM in groundwater in a 3D spatial sampling frame designed to investigate the potential influence of groundwater-surface water interaction near the Ganges (Ganga) River as well as urban influences on groundwater fDOM composition. By investigating the relationships between satellite image-derived vegetation indices to the composition of fDOM in groundwater, this study illustrates a unique approach which could be similarly applied in other settings to better understand the spatial and seasonal dynamics of groundwater processes particularly in the context of rapidly developing cities.

2. Materials and methods

2.1. Site description

The study site covers a 900 km² area around the rapidly developing city of Patna (Bihar, India), which has one of the fastest growing populations of any city in the world — increasing 76 % between 1991 and 2011 (Government of India, 2011). Growth of the Patna metropolis over recent decades has lacked commensurate expansions to infrastructure (Alakshendra, 2019) and access to sanitation facilities in north-eastern India is far below the national average (IIPS, 2022). In addition to this, Bihar is heavily dependent on the abstraction of groundwater for drinking supplies, making such resources inherently important to characterise and understand (Ramanathan et al., 2009; Singh et al., 2014).

The aquifer system beneath Patna consists of a low hydraulic conductivity, discontinuous clay unit to 220 m depth, which is interbedded from 30 to 220 m depth with medium-coarse, unconsolidated sand layers of transmissivity between 5500 and 9200 m² day⁻¹ (Saha et al., 2014). The estimated groundwater abstraction rate from the sand-

dominated aquifer is 212 Mm³ year⁻¹ (Saha et al., 2014). Sample sites in this study are orientated along the inferred regional net groundwater flowpath direction and of increasing distance from the Ganges River (i.e. SW to NE; Fig. 1; Lu et al. (2022)). More context on the field area and sampling design can be found in Richards et al. (2022b). Based on Resourcesat-1 LISS III satellite data, the predominant land uses of the Patna district (3202 km²) consists of agricultural land (76 %), built up area (9 %), fallow land (7 %), waterbodies (3–7 %), forest (<1 %) and wasteland (<1 %) (NWIC, 2017).

2.2. Groundwater and surface water sampling

Groundwater samples (from 5 to 120 m depth) were collected in Patna and the surrounding area from private and government hand-pumps (Richards et al., 2022b). All wells used for groundwater samples were in frequent use and were pumped for ~90 s before taking samples, to purge standing water from the standpipes. Patna pre- and post-monsoon groundwater sampling took place during 2019 (June–July and December, respectively). Where possible, identical wells (~70 % of sites) were sampled between pre- and post-monsoon season. Due to changes in well functionality and/or accessibility in a limited number of cases (~30 % of sites), the sampling team were required to find a suitable alternative (i.e. a handpump of similar depth in as close of proximity as possible) in the post-monsoon season re-sampling (Fig. 1). The groundwater sampling methods utilised have previously been described by Richards et al. (2022b).

Surface water samples were collected in November–December 2019 from the Ganges adjacent to Patna, from the middle of the channel and south and north riverbanks ($n = 17$). A managed aquifer recharge reservoir ($n = 1$) was also sampled. Surface water sampling details have previously been published by Richards et al. (2022a).

Samples for fDOM analysis were filtered in the field (0.45 µm cellulose and polypropylene syringe filters) into 20 mL amber glass bottles, filled to the top and sealed using Parafilm. Samples for DOC and ion chromatography (IC, for analysis of chloride) analysis were filtered (0.45 µm) and stored separately in 60 mL glass bottles. All bottles used for subsamples for EEM, DOC and IC analysis had been pre-cleaned prior to use by acid washing (10 % nitric acid for a minimum of ~8 h), thorough rinsing with MilliQ®-grade deionized water, followed by furnacing at 400 °C. All samples were stored at 4 °C until analysis, except where it was not possible, including during transit from Patna to Manchester, to prevent potential changes in organic matter composition as much as feasible (Richards et al., 2019).

2.3. Excitation-emission matrix (EEM), dissolved organic carbon (DOC) and ion chromatography (IC) analytical details

Fluorescence measurements were taken using a Varian Cary Eclipse fluorescence spectrophotometer at the British Geological Survey (Wallingford, UK). The excitation (Ex.) and emission (Em.) scanning ranges were between 250 and 400 nm (5-nm bandwidth) and 250 to 500 nm (2-nm bandwidth), respectively, using methods previously described in Richards et al. (2019). The voltage of the photomultiplier detector was set at 725 V and analysis was undertaken in quartz cuvettes with a 1-cm path length. Fluorescence is reported following blank subtraction and reported in Raman Units (RU), after normalisation to the Raman peak value of ultrapure water blanks at Ex. 350 nm, Em. 397 nm (Murphy, 2011). Ultraviolet absorbance at 254 nm (Abs₂₅₄) was measured using a Varian UV-vis spectrophotometer with a 1-cm cell path length. Specific UV absorbance at 254 nm (SUVA₂₅₄) was calculated by dividing Abs₂₅₄ by [DOC] and is reported in units of m⁻¹ mg⁻¹ L (Lakowicz, 2006). Reagent-grade Ultrapure water (ASTM Type 1) was used for blanks and to clean the quartz cuvettes between samples. A number of samples (~5 %) were run as repeat measurements.

Quantification of reference wavelength areas of EEMs provides useful indices for OM source and nature, without the need for complex

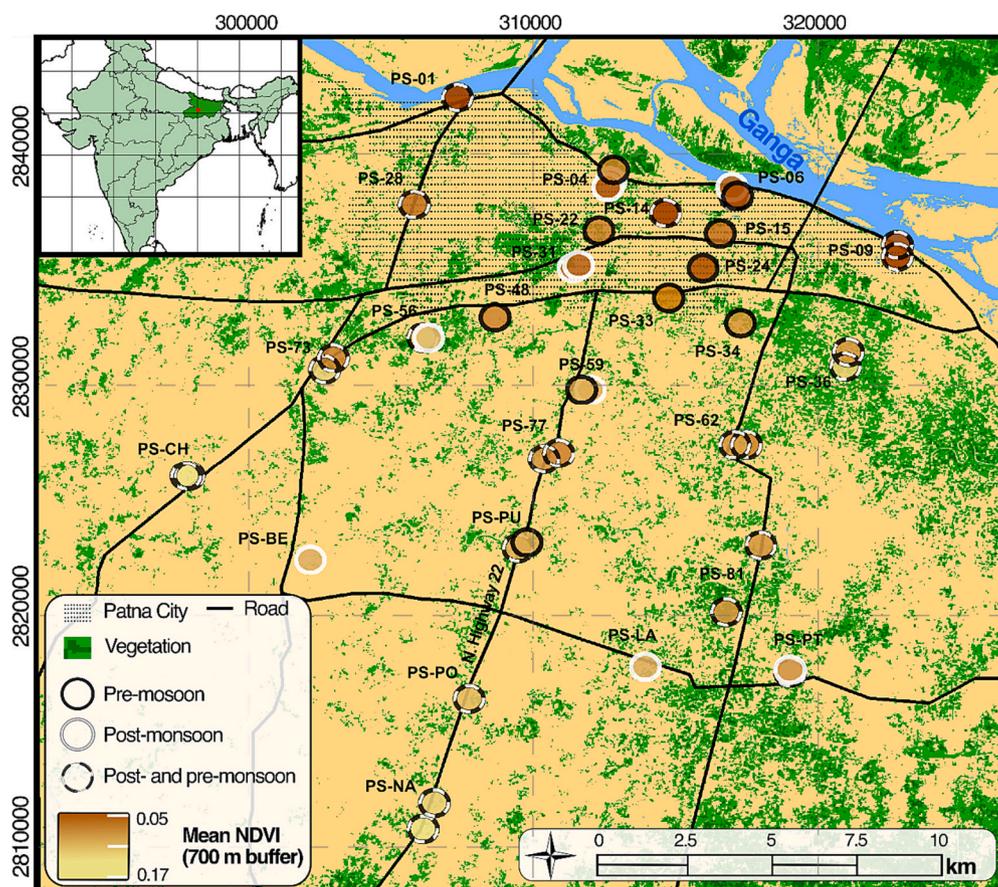


Fig. 1. Location of pre- and post-monsoon groundwater sampling sites in Patna sampled for organic characterisation. Colour within circles indicates the mean NDVI of surrounding land use (700 m buffer) from Landsat-8 satellite data (USGS, 2019). A higher NDVI is suggestive of more vegetation cover within the buffer zone, compared to low NDVI, which suggests the buffer area contains more barren and/or urbanised land (Akbar et al., 2019).

modelling procedures. After absorbance correction of the data, using a method from Lakowicz (1994), the following regions of interest were quantified by peak picking: i) humic-like fluorescence (HA-like), ii) fulvic-like fluorescence (FA-like), iii) tyrosine-like (Tyr-like), iv) tryptophan-like (Tryp-like), as well as fluorescent indices i) the $\beta:\alpha$ ratio, or 'freshness index', which has been used in previous studies to indicate the ratio between compounds of recent biological origin (β components) and more recalcitrant organic compounds (α components) (Kulkarni et al., 2017; Parlanti et al., 2000) and ii) the fluorescence index (FI), which indicates the microbial to terrestrial constituents of DOM, operationally defined as a positive ratio between an emission wavelength of 450 and that at 500 nm, obtained with an excitation of 370 nm (McKnight et al., 2001).

Bulk DOC, as non-purgeable organic carbon (NPOC), was measured at the Manchester Analytical Geochemical Unit (MAGU) laboratories at The University of Manchester using a Shimadzu® Total Organic Carbon Analyser (TOC-VCPN) coupled with an ASI-V autosampler, fitted with 24-mL glass vials. The detection of organic compounds was achieved through the 680 °C combustion catalytic oxidation method. Several samples (~5 % of the dataset) were run as repeat measurements. Chloride was measured using IC in the MAGU laboratories using methods previously described in Richards et al. (2022b).

2.4. Parallel factor analysis (PARAFAC) analysis

A PARAFAC model of the EEM data was used to deconvolute the EEM data array into several underlying units – the overall aim was to explain every sample EEM within the dataset in terms of a set of modelled

components. Using a PARAFAC model utilises a greater proportion of data available within the EEMs, compared to picking wavelength regions (Lakowicz, 2006). The PARAFAC model was built in the *staRdom* R package (Pucher et al., 2019) and served to characterise pre- and post-monsoon groundwater and river water composition of the Patna district. A 'full' dataset of 139 groundwater and 56 surface water samples from Bihar and Uttar Pradesh was used to overcome convergence issues (Murphy et al., 2013). Additional surface sites included in the modelling were from the Ganges (between Varanasi and Begusarai; $n = 25$), the Gandak ($n = 3$) and Ghaghara ($n = 9$) tributaries.

The model was developed using blank subtracted, absorbance and spectral corrected data, which was fitted to a non-negatively constrained (i.e. negative wavelengths set to zero) model, as specified by Stedmon and Bro (2008). An emission wavelength >460 nm was influenced by the first-order Rayleigh line and thus this spectral region was removed from the processing. Outliers were examined and removed in an iterative manner, such that individual samples had minimal leverage on the PARAFAC model – in total, 5 outliers were removed from the extended dataset.

After the creation of the PARAFAC model, samples were characterised in terms of i) PARAFAC components C1–C4; ii) total PARAFAC fluorescence (TPF; the sum of C1–C4), a parameter used to quantify total fluorescence of the samples in terms of the PARAFAC components and iii) the ratio of protein to humic- and fulvic-like PARAFAC components ($C4 + C3$)/($C2 + C1$) to represent microbial to recalcitrant DOM (Prasad et al., 2023; Wells et al., 2022) – a similar index has previously been used as an indicator of DOM bioavailability (Baker, 2002; Kulkarni et al., 2017).

2.5. Land use characterisation based on normalised difference vegetation index (NDVI)

To quantify contributing factors to groundwater fDOM signature, land use was characterised around each site, using satellite data. Classifying land use using normalised difference vegetation index (NDVI) is a widely recognised technique for identifying land usage to a high resolution (Akbar et al., 2019; Tittebrand et al., 2009; Vermote et al., 2016; Zaitunah et al., 2018). NDVI consists of a continuous parameter, which means comparison to other environmental variables can easily be assessed. Surface reflectance L8/OLI data of 30 m × 30 m resolution was obtained from the United States Geological Society (USGS, 2019). NDVI is calculated as the difference between near infrared (NIR) and red (RED) reflectance, divided by their sum (Eq. 1, Anabitarte et al., 2020):

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

Temporal variations in NDVI due to climatic factors are well documented (Nischitha et al., 2014; Revadekar et al., 2012). To select a Landsat scan representative of recent land use around sampling sites, the variation in NDVI, from the start of the data record (2013) until the last groundwater sampling date, was quantified. QGIS 3.16 (QGIS Development Team, 2020) was used to calculate mean NDVI in 36 cloud-free Landsat-8 scans for the 900-km² study area; the scan of highest mean NDVI score (Oct 2018) was used to reflect land use in this study, on the premise that more distinct vegetation can determine a more accurate representation of land use from satellite data (also noting that current land use activities may not affect groundwater composition until many years after (Lawson et al., 2016)). To accurately capture the surrounding terrestrial environment around each site, surface water (NDVI < 0.02) was excluded from the final NDVI parameter.

For the selection of an appropriate buffer size, from which to calculate a NDVI score to compare to groundwater DOM bioavailability, mean NDVI scores for a range of buffer radii (50–800 m) were calculated. It was important that the buffer accounted for the grid size of the land use data, as well as the groundwater zone of influence around the sites. The buffer size (700 m) was therefore selected based on the highest correlation coefficient between (C4 + C3)/(C2 + C1) and mean NDVI. As sites < 1 km from the river Ganges may be comprised of > 10 % river water in this setting (Lu et al., 2022), these were excluded from the correlation analysis.

2.6. Statistical tests

All statistics used in this study were conducted using SPSS Statistics 26 and Microsoft Excel 2020. Regression statistics were reported using a Pearson correlation in the format “ r (degrees of freedom) = r value; $p = p$ value”. Mann-Whitney-Wilcoxon rank-sum tests, for comparing the mean of non-parametric populations, were reported as “ $U =$ Mann-Whitney U , $p = p$ value”, where U is the difference between the two rank totals. All statistics tests are reported at 95 % confidence and Bonferroni correction, $\frac{\alpha}{n}$, was applied as a more conservative test in the instances of several statistical tests being performed simultaneously, where $\alpha = 0.05$ and $n =$ the number of tests performed.

3. Results

3.1. Excitation-emission matrix-parallel factor analysis (EEM-PARAFAC)

3.1.1. PARAFAC components

The sources of organic carbon in river and groundwater around Patna were characterised using excitation-emission matrix (EEM) fluorescence techniques. Statistical deconvolution of the extended fluorescence dataset from Bihar and Uttar Pradesh samples, using PARAFAC modelling, revealed that four components adequately represented the

fDOM composition of surface and groundwater samples. Validation of the four-component model was achieved through split-half analysis; data was split into four quarters and recombined in six different ways to verify that these returned similar results (Murphy et al., 2013). The model was considered stable since all Tucker’s congruency coefficients (TCCs) were close to one (Wünsch et al., 2019). When more than four components were used, the results of the sub-samples varied significantly between themselves, revealing that the model was not robust beyond four components. Split-half analysis could only validate a PARAFAC model using the extended dataset; it was not possible to validate a model using sub-datasets (i.e. groundwater/surface water; Harjung et al., 2023; Murphy et al., 2013; Wünsch et al., 2019). Analysis of residual samples revealed that the model was adequate in representing the DOM fractions; the fluorescence in these samples was non-systematic and likely a product of spectrophotometer noise.

The four components were i) Component 1 (C1) – terrestrial humic-like fluorescence; ii) Component 2 (C2) – microbial humic-like fluorescence; iii) Component 3 (C3) – tryptophan-like fluorescence and iv) Component 4 (C4) – tyrosine-like fluorescence (Fig. 2). Detailed peak excitation/emission characterisation is described in Table S1. Classification rationale of components has been based on comparisons to the OpenFluor database and previous fDOM studies (Table S1). Components in this study are broadly similar to those identified previously in West Bengal (Kulkarni et al., 2017). The broad occurrence of the PARAFAC components, by relative contribution to PARAFAC fluorescence, is summarised in Table 1.

3.1.2. PARAFAC/peak-picking comparison and DOM measurement data quality

To compare two different methods of fluorescence data post-processing, bivariate correlations and data quality of PARAFAC and peak-picked methods were tested. Humic-like PARAFAC components C1 and C2 showed a strong correlation with peak-picked HA-like fluorescence ($r(98) = 0.99$, $p < 0.05$, both) (Fig. S1a & b). Amino-like components C3 and C4 also showed a strong correlation with peak-picked tryptophan- and tyrosine-like fluorescence ($r(98) = 0.98$ & 0.95 , $p < 0.05$, respectively) (Fig. S1c & d). The strong association between the two methods of fDOM characterisation is also documented in other studies (Richards et al., 2019).

Repeat fluorescence measurements were taken for surface and groundwater samples used in this study. The mean peak-picked fluorescence error/95 % C.I. for groundwater sample repeats ($t = 5$) was 17.5 % (± 22.4) and 2.9 % (± 2.3) for surface samples ($t = 3$) (for HA_{mean} , FA_{mean} , $Tryp_{mean}$, Tyr_{mean} , HIX , FI and $\beta:\alpha$; full dataset). The mean repeat error/95 % C.I. for the full PARAFAC dataset was 5.1 % (± 6.7) for groundwater samples ($t = 8$) and 1.1 % (± 3.7) for surface water samples ($t = 2$), for components C1–C4. Though both methods yielded similar results, the comprehensiveness associated with PARAFAC modelling, compared to peak-picking of fluorescence, mean PARAFAC data will hereafter be used for fDOM characterisation in this study. For completeness, the reported DOC (NPOC) measurement error/95 % C.I. for the full dataset ($t = 5$) was 4.4 % (± 3.6).

3.2. Geospatial and seasonal variability in dissolved organic carbon (DOC) and fluorescent dissolved organic matter (fDOM)

3.2.1. Fluorescence characterisation: seasonal differences in groundwater depth profiles

DOM was characterised in (inferred) vertical and lateral flowpath directions to understand the geospatial and seasonal variation in organic composition under Patna. Bulk DOC in groundwater (5–120 m; typically < 2 mg/L) was consistent to groundwater in West Bengal (Kulkarni et al., 2019; Mohanta and Goel, 2014). A spike in groundwater DOC, as well as fluorescent DOM at ~15 m depth (Fig. 3a–c) in both seasons is consistent with previous studies on groundwater fluorescence characterisation (Mladenov et al., 2010; Richards et al., 2019). All PARAFAC components

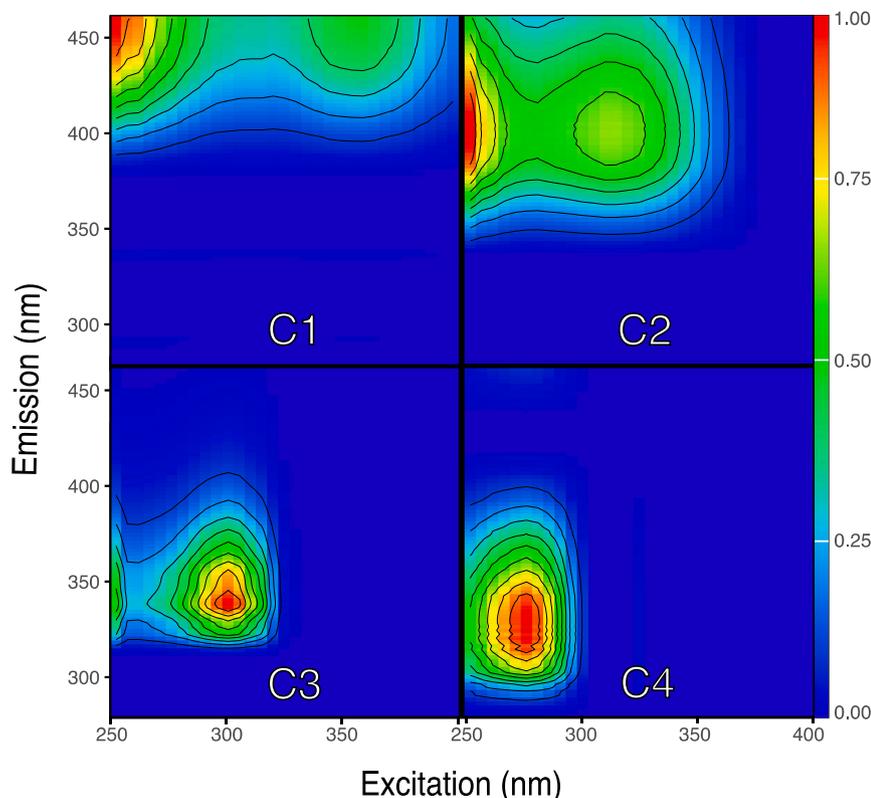


Fig. 2. Modelled excitation-emission matrix-PARAFAC (EEM-PARAFAC) components C1–C4, derived from pre- and post-monsoonal surface and groundwater samples in the Bihar and Uttar Pradesh region. C1 = terrestrial humic-like; C2 = microbial humic-like; C3 = tryptophan-like and C4 = tyrosine-like fluorescence (justification based on comparisons to similar components found in other fluorescence studies). Component fluorescence maxima have been scaled to 1 and the colour denotes the relative fluorescence. Fluorescence components were used to quantify the source and nature of DOM in the study area.

Table 1

Relative component contributions to total PARAFAC fluorescence (TPF) in groundwater and surface water around Patna. The 95% confidence intervals are reported in brackets.

PARAFAC component	Pre-monsoon groundwater (%)	Post-monsoon	
		Groundwater (%)	Surface water (%)
C1	10.6 (±3.8)	30.1 (±3.4)	37.1 (±2.3)
C2	10.9 (±3.5)	28.0 (±3.0)	33.4 (±1.9)
C3	69.8 (±8.2)	9.2 (±2.1)	9.7 (±2.0)
C4	8.7 (±1.7)	32.7 (±5.5)	19.8 (±2.2)

displayed modest negative (to $p < 0.05$ significance, though $p > 0.05$ after Bonferroni correction) with increasing depth in both measured seasons, except for C3 in the pre-monsoon period, which did not show a relationship with depth to the 0.05 level.

Fluorescence indices were utilised to further comprehend the source and nature of OM. As a proxy for DOM bioavailability, the $(C4 + C3)/(C2 + C1)$ ratio displayed spikes in the pre-monsoon period but was not significantly correlated with depth ($r(35) = 0.26$, $p < 0.05$) (Fig. 3e), though in post-monsoon the bioavailability ratio showed a weak negative correlation with increasing depth ($r(40) = 0.42$, $p < 0.05$, though $p > 0.05$ after Bonferroni correction). The $\beta:\alpha$ ratio displayed a weak positive relationship with increasing depth in the pre-monsoon period ($r(37) = 0.37$, $p < 0.05$, though $p > 0.05$ after Bonferroni correction) (Fig. 3f), which suggests there was more DOM of recent biological origin at depth than for shallow soil horizons, whilst there was no statistical relationship with depth in the post-monsoon season ($r(42) = -0.12$, $p > 0.05$). The $\beta:\alpha$ ratio was seen to closely track C3 fluorescence and to a lesser extent C4 (Fig. S2a & b).

3.2.2. Fluorescence characterisation in the lateral flowpath direction

Other than from vertical exchange, the predominant groundwater net flow direction is thought to be from SW to NE (Lu et al., 2022;

Mukherjee et al., 2018). Groundwater DOM composition as a function of increasing SW-NE distance from the Ganges was therefore explored. None of the DOM parameters (DOC, C1–C4, $[C4 + C3]/[C2 + C1]$, FI, $\beta:\alpha$, HIX and $SUVA_{254}$) were found to correlate with increasing distance from the Ganges (to $p < 0.05$; Fig. 4a–f). It was notable that C1 (of shallow-to-mid depth) and DOC peaked close to the Ganges, particularly after the onset of the monsoon (Fig. 4b). In both seasons, there were also elevated concentrations of DOC at 10 km and at >20 km (after scatterplot smoothing; Fig. 4a), though the latter observation is largely influenced by the PS-NAD-160 sample site, which was potentially hydrologically connected to nearby organic-rich surface bodies of water. The FI was elevated in samples 4–16 km from the Ganges, particularly in the post-monsoon season (Fig. 4e).

3.3. Comparison of surface water and groundwater

Organic properties of groundwater and surface water from the post-monsoon season were compared by a Mann Whitney U test to distinguish differences in DOC and fDOM between the two groups (Table S2). All identified PARAFAC components (C1–C4) were higher in the surface samples, compared to groundwater samples ($U = 72, 74, 63$, and 55 , respectively, $p < 0.05$, all). These differences in fDOM composition are consistent with Holocene sediments and overlying surface water in West Bengal (Kulkarni et al., 2017). DOC was not significantly different ($U = 98$, $p > 0.05$) between post-monsoon groundwater (0.2–4.6 mg/L) and surface water (1.5–2.2 mg/L) in this study, though DOC sample size was small in the surface sites ($n = 5$). The $(C4 + C3)/(C2 + C1)$ ratio and McKnight was lower in the surface water, compared to groundwater ($U = 221, 155$, $p < 0.05$, respectively), whilst the $SUVA_{254}$ ratio was higher in the surface water ($U = 29$, $p < 0.05$). The $\beta:\alpha$ ratio was not significantly different between the two populations ($U = 343$, $p > 0.05$), whilst HIX was significantly lower in surface water than in groundwater ($U = 233$, $p < 0.05$). A comparison between groundwater and surface water characteristics is given in Table S2.

These results show that the dissolved organic composition of post-

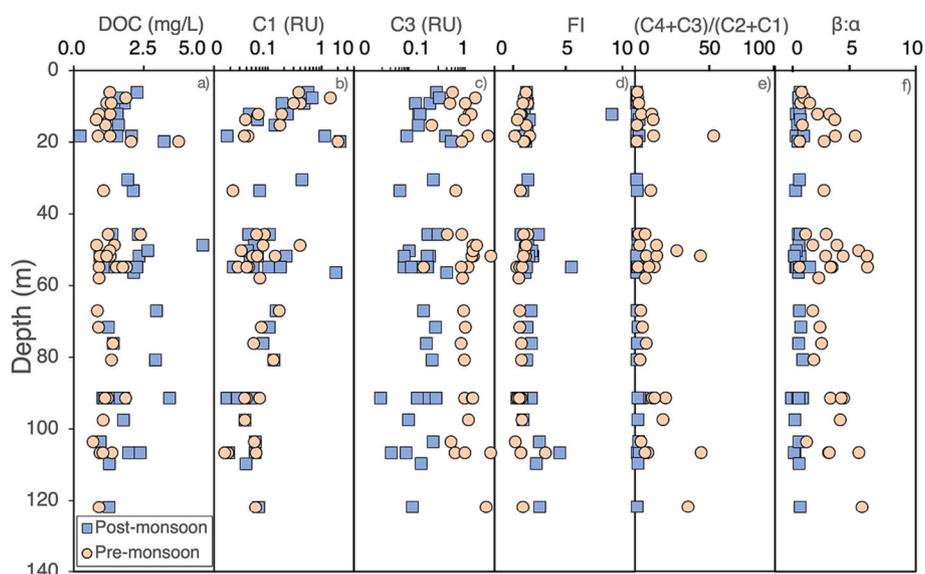


Fig. 3. Depth profiles of DOM composition for pre- and post-monsoon groundwater under Patna, Bihar for a) DOC, modelled PARAFAC components; b) C1; c) C3 and fluorescence ratios; d) fluorescence index (FI); e) $(C4 + C3)/(C2 + C1)$; and f) $\beta:\alpha$ ratio. Note the use of a log scale in b and c.

monsoon surface water is, to a great degree, distinct from post-monsoon groundwater under Patna. Although the characteristics of measured surface water was found to be fairly consistent from sample to sample, compared to groundwater, there can be substantial changes to the composition of surface waters on various (e.g. daily or seasonal) time-scales (Richards et al., 2022a; Stedmon and Markager, 2005), and it is important to note that the number of surface water samples was relatively limited in this investigation.

3.4. Overall seasonal and spatial distribution of dissolved organic matter (DOM) and parallel factor analysis (PARAFAC) components' prominence

Systematic seasonal changes in DOC and fDOM were investigated. C3 was the only PARAFAC component in groundwater that was significantly different between seasons, decreasing substantially from pre- to post-monsoon ($U = 33, p < 0.05$). In the pre-monsoon season, C3 typically contributed $>60\%$ to TPF (mean = $70 \pm 8\%$, $n = 36$) and was particularly prevalent at >30 m in depth. This seasonal variation in protein-like fluorescence is consistent with Stedmon and Markager (2005).

The large increase in C3 fluorescence from post- to pre-monsoon also resembles other fDOM indices. The $(C4 + C3)/(C2 + C1)$ ratio and HIX display positive and negative relationships to the C3 component in the pre-monsoon season ($r(37) = 0.88$ & $-0.26, p < 0.05$), respectively, because the tryptophan-like component contributed to a large proportion of TPF before the onset of monsoon. The $(C4 + C3)/(C2 + C1)$ and $\beta:\alpha$ were also higher, whilst HIX was lower, before the onset of the monsoon ($U = 159, 43$ and 315 , respectively, $p < 0.05$, all; Table S2), than under post-monsoon conditions.

Though the relative abundances of PARAFAC components generally varied a lot between seasons, for sites <30 m in depth in the pre-monsoon season, the predominant components were C1 and C2, which resembles an fDOM ratio similar to most post-monsoon season samples. High TPF, C1 and C2 hotspots were apparent adjacent to the river at <30 m depth in the post-monsoon (Fig. 5). DOC was higher in the post-monsoon than the pre-monsoon ($U = 465, p < 0.05$), in particular at ≥ 50 m depth (Fig. 3a). An increase in FI from pre-monsoon to post-monsoon ($U = 444, p < 0.05$), particularly at >15 m depth (Fig 3d), indicates a shift from allochthonous (has been transported away from its source) to autochthonous (produced *in-situ* within the body of water/sediment) fDOM between seasons, respectively – an observation which

is inconsistent with that documented in the monsoon influenced Jian-gan Plain (Yang et al., 2020). In comparison to other PARAFAC components, the fluorescence signal of C4 was low throughout both seasons, though C4 comprised the major fluorescence component in several sites in the post-monsoon season, as a proportion of TPF (Fig. 5; Table 1).

It was difficult to disentangle the effects of potential surface-groundwater mixing and proximity to the urban center given their co-variability in the context of Patna located very close to a major river. In sites that were identical between pre- and post-monsoon season ($n = 31$), PARAFAC C4 increased in sites closest to the river, whilst it decreased in sites further away from the river (Fig. 6c). Decreases in $(C4 + C3)/(C2 + C1)$, $\beta:\alpha$ and TPF between pre- and post-monsoon season were more apparent further from the river, than for proximal sites (Fig. 6a, b & d). A comparison in fDOM characteristics between identical boreholes (Fig. 6), as opposed to comparing the two seasonal data subsets (Section 2.2), also confirms that seasonal controls were more significant to DOM composition than, for example, systematic changes with distance from the River Ganges and the urban center, noting that sub-surface heterogeneity also may impact spatial observations. (CGWB, 2015; Harjung et al., 2023). In order to further untangle the concurrence of the river and urban area within the sampling domain, the protein: humic-like fDOM of groundwater and surface water was compared to land use patterns and the DOM composition of the river.

3.5. Groundwater and surface water DOM comparison and relationship to normalised difference vegetation index (NDVI)

A mean normalised difference vegetation index (NDVI) was calculated around each sample site (700 m buffer) to assess the impact of land use on groundwater DOM composition. The land use characterisation was compared to PARAFAC components C1, C3 and $(C4 + C3)/(C2 + C1)$ of groundwater under Patna (Fig. 7b, d & f). The occurrence of C1 and C3 PARAFAC components did not relate to the land use proxy. For all sample sites, NDVI was not significantly correlated to the $(C4 + C3)/(C2 + C1)$ ratio ($p > 0.05$), though modelling by Lu et al. (2022) suggested that sample sites <1 km from the river Ganges could potentially be comprised of $>10\%$ river water as dependent on pumping rates and well location. For sample sites >1 km away from the river, NDVI expressed a negative correlation with $(C4 + C3)/(C2 + C1)$ ($p < 0.05$) (Fig. 7f), which suggests that urban land use may be associated with a higher protein signature in groundwater, compared to background

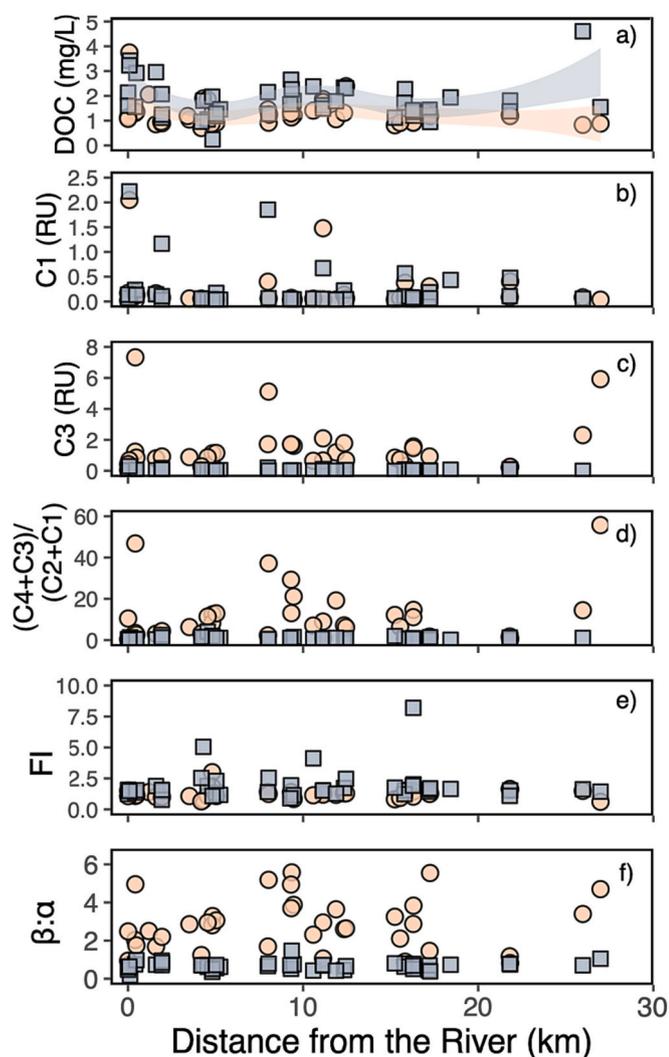


Fig. 4. Spatial trends in DOM composition for pre- and post-monsoon groundwater under Patna, Bihar for a) DOC; modelled PARAFAC components b) C1; c) C3; and fluorescence ratios d) $(C4 + C3)/(C2 + C1)$; e) fluorescence index (FI); and f) $\beta:\alpha$ ratio. Envelope trend lines show confidence limits (95 %) for the locally estimated scatterplot smoothing (LOESS) trendline in the ggplot2 package for R.

levels of DOM.

A spike in fDOM in some samples may be associated with the predominantly urban land use around these sites, or the proximity of the river to these sites; there is a similarity between the surface water and near-bank fDOM composition (Fig. 7a–d). There was also similarity in $(C4 + C3)/(C2 + C1)$ ratio between river samples and sites in the most urbanised part of Patna (Fig. 7e & f).

4. Discussion

4.1. Dynamics of dissolved organic matter (DOM) composition under Patna

Large variations in DOM fluorescence properties are observed both spatially and temporally across the Patna study area. A small change in DOC between seasons (relative to changes in fDOM signature), accompanied by large changes in fDOM signature, suggest that the DOM composition is more seasonally influenced than DOM abundance. Higher McKnight/FI in the post-monsoon (~ 1.9) than pre-monsoon (~ 1.2) may suggest a shift from more terrestrial (i.e. lignaceous, plant-derived DOM) to more microbially sourced DOM (i.e. extracellular release

from microorganisms) after the monsoon rain (McKnight et al., 2001). The addition of monsoonal influxes of water likely contains a mixture of precipitation, contaminated surface water and plant matter (Chen et al., 2018; Ghosh et al., 2022; Yang et al., 2020), which provides precursor material for microbial activity. Fluorescence in the pre-monsoon season, dominated by the C3 component of recent biological origin (Fig. 5a), is likely the product of the bacterial biodegradation, which succeeds inputs of organic material into the aquifer. Microbial activity may be promoted most significantly where surface sources of DOM are added to low fertility soil horizons (i.e. at depth; Chen et al., 2018), noting that here, a recurring seasonal supply of labile DOM (Fig. 3a) may contribute to a large initial stock of microbial biomass (Hofmann et al., 2020). The observed C3 fluorescence is plausibly quenched during high precipitation events. In the monsoon period, there is also evidence for both lateral and vertical inflows of organics into the aquifer under Patna. A contradiction between FI and PARAFAC C3 provokes further investigation on the interpretation of fluorescence parameters as indicators of microbial activity.

4.2. Comparison of river and groundwater and evidence for river ingestion of organics

Comparison of river and groundwater fluorescence characteristics was suggestive of surface-groundwater mixing. Post-monsoon river water was characterised by a similar concentration of bulk DOC, yet much higher fDOM, than groundwater. An attenuated protein:humic ratio in post-monsoonal near-bank sites that is distinctive of river water suggests that this is a result of lateral inflows (i.e. from the river) of DOM. This theory is supported by spikes in humic-dominated TPF next to the Ganges in the post-monsoon season (Fig. 5b). Further work to sample at a higher temporal and spatial resolution would be beneficial to further untangle this matter.

Elevated chloride is often associated with leachates of anthropogenic source (Abd El-Salam and Abu-Zuid, 2015; Lyngkilde and Christensen, 1992; McArthur et al., 2012). Given the much lower Cl^- in river water compared to groundwater (post-monsoon; Table S2), it is likely that industrial or domestic effluent from the city was responsible for high Cl^- in some of the near-bank samples, whilst dilution from river water was responsible for low Cl^- in others (Fig. 8). In fact, modelling by Lu et al. (2022) and Wallis et al. (2020) suggest monsoonal reversals in hydraulic head in this setting could result in the ingress of river water in the order of hundreds of meters. Heterogeneous permeability and pumping of the underlying aquifer could plausibly create groundwater zones with contrasting vulnerability to river water ingress. The mixing of polluted groundwater at the river-groundwater interface has potential to impact harmful chemical and microbial processes, such as the mobilisation of metals (Wallis et al., 2020).

4.3. Vertical infiltration of dissolved organic matter

Evidence for vertical infiltration of surface-derived organics under the city was more pronounced than evidence of lateral inflow. Spatially ubiquitous increase in microbially derived DOC from pre- to post-monsoon points to the flushing of surface-derived organics to >100 m depth (Lapworth et al., 2018). High protein-like fluorescence in both seasons, relative to matured terrestrial components, suggests that much of the infiltration is impacted by slurry or wastewater (Baker, 2002; Rodríguez-Vidal et al., 2022). This conclusion is consistent with higher SUVA_{254} at shallower than deeper depths, thus lower molecular weight, less aromatic DOM is likely more prevalent at deeper depth, relative to recalcitrant DOC (Castan et al., 2020).

An increase in E_h from pre-monsoon to post-monsoon samples points to a general increase in more oxygenated groundwater over this time period (data published in Richards et al., 2022b), likely confirming that monsoonal precipitation influences the composition of local groundwater and subsequent leachate throughout the vertical profile.

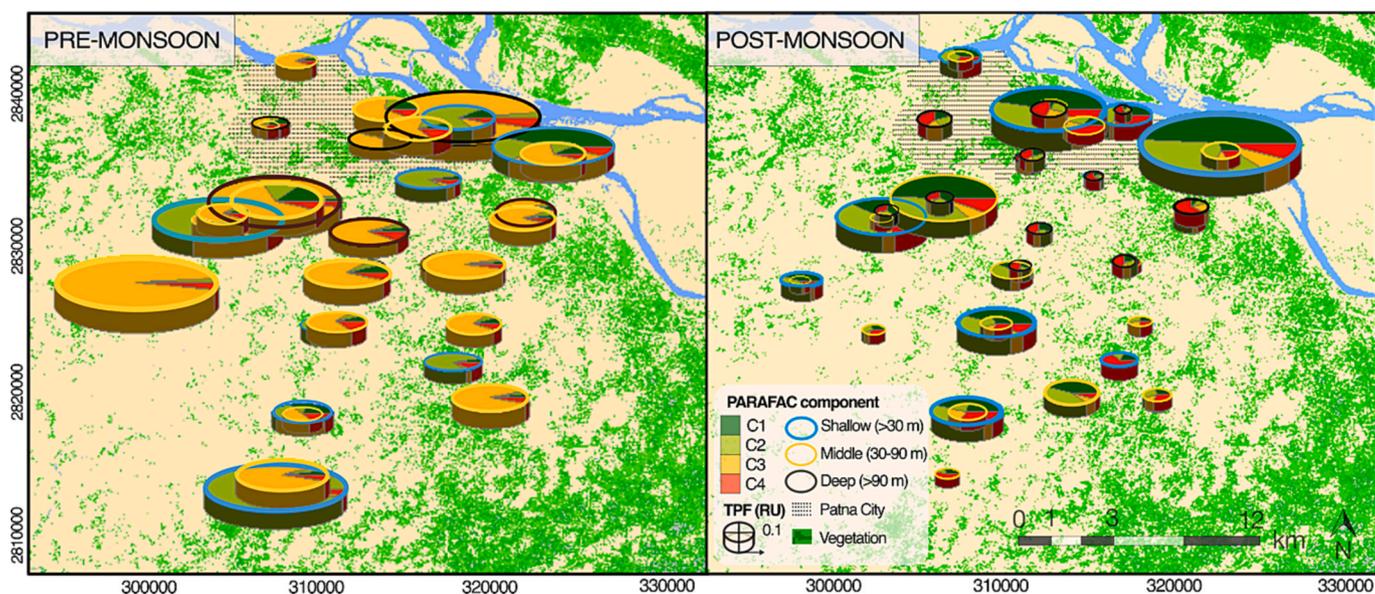


Fig. 5. Relative importance of PARAFAC C1–C4 fluorescence of groundwater DOM under Patna in the pre-monsoon (left) and post-monsoon (right) season, based on excitation-emission matrix (EEM) techniques. Fluorescence was measured in groundwaters of 6–120 m depth; C1–C4 are the modelled parallel factor analysis (PARAFAC) components (C1 = terrestrial humic-like; C2 = microbial humic-like; C3 = tryptophan-like and C4 = tyrosine-like fluorescence). The size of the pie chart indicates the total PARAFAC fluorescence (TPF; sum of PARAFAC components C1–C4). The colour of the boundary of each pie chart is an indicator of the depth of the associated sample.

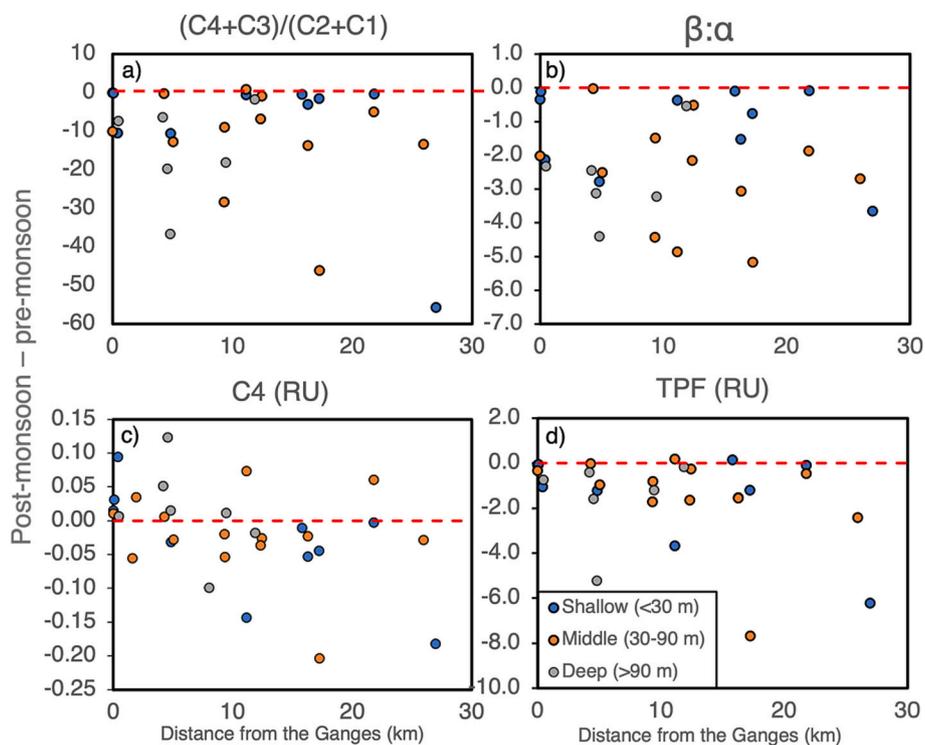


Fig. 6. Seasonal differences in selected fDOM parameters from groundwater samples in Patna, collected from identical wells ($n = 31$) calculated as post-monsoon – pre-monsoon (absolute values) for a) $(C4 + C3)/(C2 + C1)$ PARAFAC components; b) $\beta:\alpha$; c) PARAFAC C4 (RU; Raman Units); and d) TPF (Total PARAFAC fluorescence; $C1 + C2 + C3 + C4$; RU). The red dashed line represents '0' (i.e. no change between pre- and post-monsoon season).

Consistent with the mean residence time of groundwater below Patna reported to be generally ~ 40 – 70 years (likely reflecting a mixture of waters of different age) (Richards et al., 2022b), rapid ingress of oxic recharge to depth, as suggested by Richards et al. (2022b), during the monsoon season is plausible and supported by distinct seasonal differences in C3, $(C4 + C3)/(C2 + C1)$, FI, $\beta:\alpha$ and chloride (Table S2).

Similar increase in Cl^- over this time period may suggest precipitation events during the monsoon season facilitate the leaching of agrochemical-derived solutes to depth, particularly in agricultural settings (McArthur et al., 2012; Richards et al., 2022b). Likewise, Ghosh et al. (2022) noted that greater runoff contributed to higher sediment yield in runoff in the Ganges River catchment.

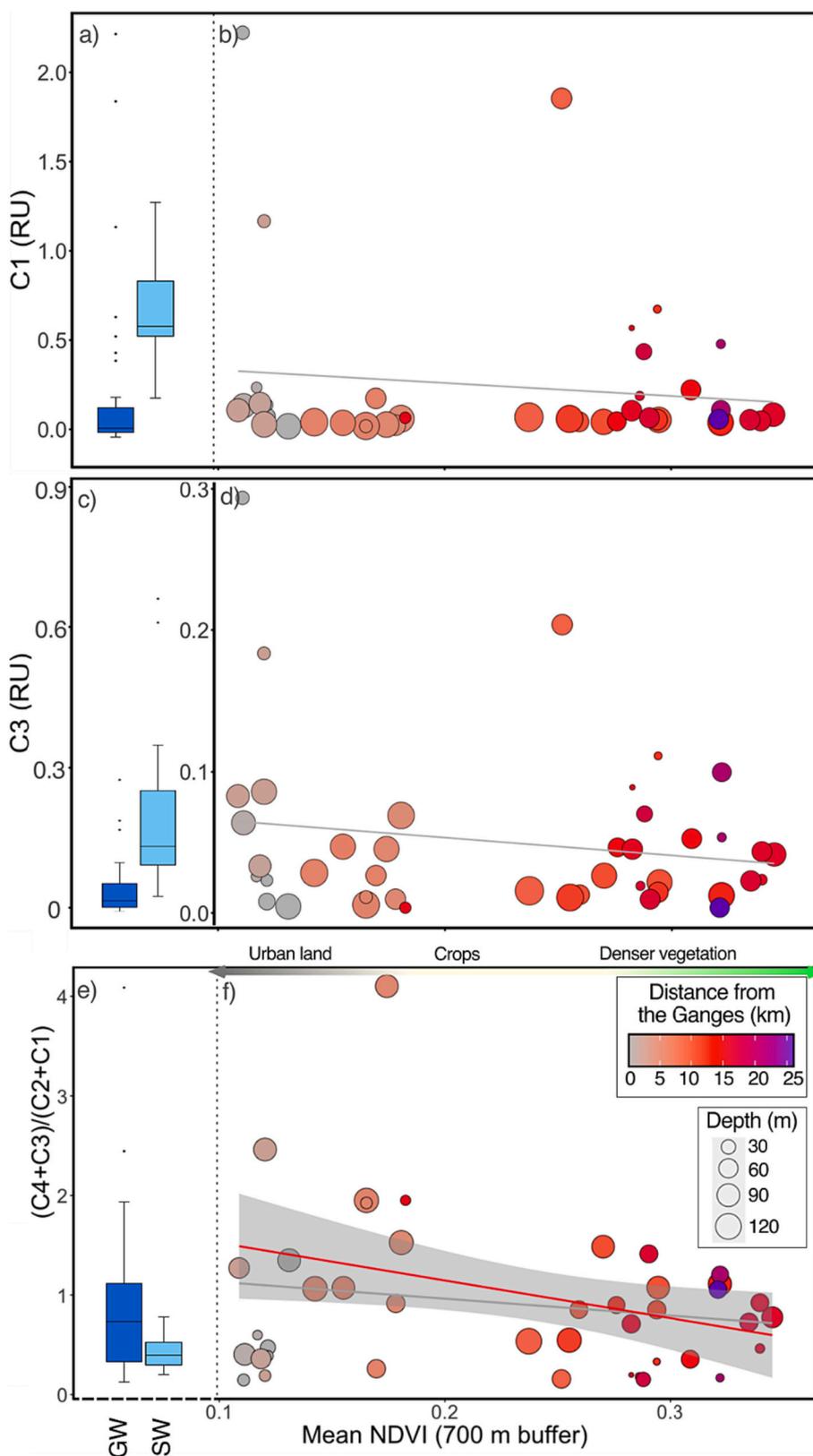


Fig. 7. Correlation between post-monsoon PARAFAC components and land use buffer zones and river water fluorescence signal indicate hydrological linkage between land use and groundwater composition, as well as river-groundwater interactions. Plots show river and groundwater distributions and a comparison between mean normalised difference vegetation index (NDVI; 700 m buffer) for a) & b) C1 (humic-like fluorescence); c) & d) C3 (tryptophan-like fluorescence); and e) & f) $(C4 + C3)/(C2 + C1)$ (protein to humic-like fluorescence ratio). Solid lines relate to trendlines fitted to all points (grey line) and excluding sample points < 1 km from the river (significant correlation; red line), as the organic composition of the proximal water may be influenced by river water ingression (Lu et al., 2022); a 95 % C.I. for the latter is indicated by the grey shaded area.

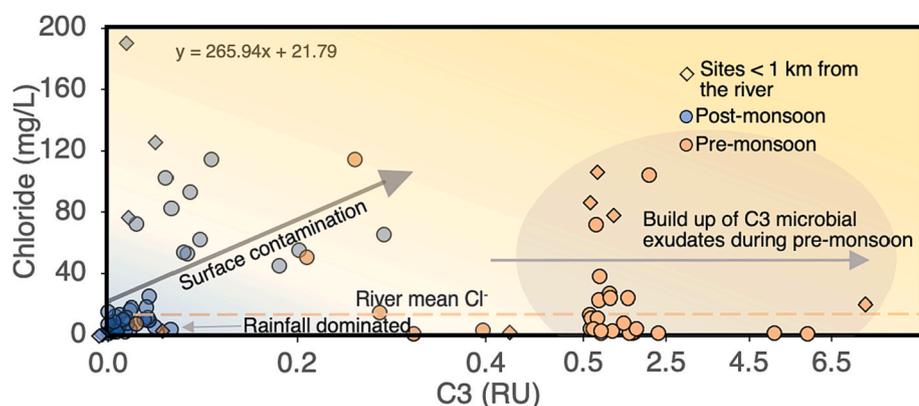


Fig. 8. Chloride concentration compared to modelled PARAFAC component C3 (tryptophan-like fluorescence) indicates the degree of groundwater contamination from components recently derived from the surface. Fluorescent dissolved organic matter and chloride measurements were measured in pre- and post-monsoon groundwaters from Bihar from 6 to 120 m depth. Trendline plots post-monsoon samples only – *in-situ* production of tryptophan is thought to govern fDOM composition in the pre-monsoon season, rather than surface contamination, which would also exhibit high Cl^- . Sites < 1 km from the river are shown as diamonds. The shaded zone indicates sites which display a large spike in C3 fluorescence in the pre-monsoon season.

Samples with high tryptophan and high chloride plausibly reflect areas that are most contaminated with surface components (McArthur et al., 2012). Though, monsoonal infiltration of plant residues and nutrients from rural areas likely provides precursor material for extensive *in-situ* processing and subsequent tryptophan production in the pre-monsoon season (Fig. 8; Chen et al., 2018). In the post-monsoon season, the distribution of protein-rich groundwater was concentrated in urban areas; further spatial analysis indicated that groundwater organics composition was directly related to land use.

4.4. Dissolved organic matter (DOM) relationship to land use

There was no direct relationship between the land use characterization and PARAFAC C1 nor C3 (Fig. 7b & d, respectively). Though, the relationship between NDVI and the protein:humic ratio (Fig. 7f), for post-monsoon samples at >1 km distance from the Ganges may help to untangle two confounding variables, i) the infiltration of proteinaceous, labile DOM in groundwater under urban areas (Rodríguez-Vidal et al., 2022) and ii) the attenuation of protein-like fluorescence in the hyporheic zone, plausibly where labile DOM is adsorbed at the surface-groundwater interface (Lapworth et al., 2009), though low protein:humic in near-bank sites is not found in the pre-monsoon season. Ingression of humic-rich river water is therefore more likely to impose the relationship between land use and protein:humic fDOM. Crucially, the spike in protein:humic DOM below Patna City (Fig. 7f) is consistent with groundwater EOC concentrations, of which maximum values were observed within 5–10 km of the river (Richards et al., 2021).

Whilst a predominant control on the distribution of protein:humic-like fluorescence in Patna's post-monsoon groundwater is considered to be the infiltration of sewage derived organic components derived from surface sources, the control of tryptophan-like fluorescence changes in the pre-monsoon season. Here, the *in-situ* processing of monsoon-recharged plant residue and nutrients likely leads to ubiquitous production of protein-like exudates in the groundwater (Stedmon and Markager, 2005), as evident by spikes in tryptophan-like fluorescence regardless of depth (Fig. 3c) and crucially, not accompanied by increases in Cl^- concentration (Fig. 8), which would otherwise point towards a dilution/concentration effect. These results are consistent with other studies which utilise tryptophan-like fluorescence as an indicator of microbial contamination (Baker, 2002; Ward et al., 2021), demonstrating that fDOM can be an informative predictor variable in considering groundwater vulnerability more widely.

Intensive groundwater pumping likely exacerbates the relationship between urban land use and elevated C3 and C4 fluorescence; the extraction of groundwater through private and municipal pumping under highly populated areas could exacerbate drawdown of immature protein and surface-derived organics to depth (e.g. Lapworth et al., 2018). Amongst other factors, high pumping rates of groundwater in Patna may help to explain why the seasonal variation in DOM was so

pronounced, compared to other studies of low DOM aquifers (Hofmann et al., 2020; Lu et al., 2022). Further characterisation of groundwater DOM subjected to intensive pumping regimes would help to understand the processes that underpin groundwater quality under rapidly developing and groundwater-reliant areas.

5. Conclusion

Groundwater and surface waters were characterised around the rapidly developing city of Patna, Bihar to understand the spatial and temporal dynamics of DOM. A novel technique of characterising land use was used to explain, in part, the observed groundwater composition.

Seasonal changes were observed in the abundance of fDOM, though the compositional changes were more distinct, particularly tryptophan-like fluorescence. In the dry season the high tryptophan-like signature is likely the product of some degree of sewage-derived ingress, though predominantly *in-situ* processing of monsoonal material; the pre-monsoon fluorescence signature is then quenched by compositionally different material after heavy rainfall and onset of monsoon. Combining evidence from organic tracers, as well as redox potential (*Eh*) and chloride, suggests that monsoonal likely recharges groundwater with relatively high i) plant residues and ii) anthropogenic surface contamination, including agrochemicals and wastewater and iii) river-derived contamination, in near-bank sites. Elevated protein:humic ratios of DOM in the post-monsoon season, particularly underneath urban areas, indicate ingression of proteinaceous surface organics into groundwater. The authors note that this relationship is likely related to the composition of surface-derived DOM in urban areas and extensive groundwater pumping. High DOC of humic-like nature and attenuated protein:humic fDOM in near-bank samples could also suggest mixing of river water and groundwater close to the river.

Our approach to analyse the relationship of the composition of aqueous organics with land use in the vicinity of sampling sites demonstrate promising potential for future adaptation to other locations with similar hydrogeochemical settings to determine the factors controlling groundwater DOM composition in various contexts. The results in this study improve our mechanistic understanding of the potential processes controlling the distribution and seasonal dynamics of DOM around Patna, as an example of a rapidly developing city.

CRediT authorship contribution statement

George J.L. Wilson: Conceptualization, Formal analysis, Investigation, Data curation, Methodology, Software, Validation, Visualization, Writing – original draft. **Chuanhe Lu:** Methodology, Software, Writing – review & editing. **Dan J. Lapworth:** Conceptualization, Methodology, Software, Writing – review & editing. **Arun Kumar:** Conceptualization, Resources, Investigation, Project administration, Writing – review & editing. **Ashok Ghosh:** Conceptualization, Funding acquisition,

Resources, Project administration, Writing – review & editing. **Vahid J. Niasar**: Conceptualization, Methodology, Writing – review & editing. **Stefan Krause**: Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **David A. Polya**: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Daren C. Gooddy**: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Laura A. Richards**: Conceptualization, Funding acquisition, Investigation, Data curation, Methodology, Project administration, Resources, Supervision, 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.

Data availability

Data will be made available on request.

Acknowledgements

This research reported here was part of the FAR-GANGA Indo-UK Water Quality project supported by Newton Fund NERC and DST (NE/R003386/1 and DST/TM/INDO-UK/2K17/55(C) & 55(G); 2018–2022) (www.farganga.org). A Dame Kathleen Ollerenshaw Fellowship is acknowledged for LAR and for GJLW's PhD studentship. DJL was supported by the BGS International NC programme 'Geoscience to tackle Global Environmental Challenges'. We thank individuals who contributed to relevant field campaigns including Neha Parashar, Rupa Kumari, Prabhat Shankar, Aman Gaurav and Siddhu Kumar (all Mahavir Cancer Sansthan) and Sam Addison (University of Manchester). The following individuals are also acknowledged for analytical support: Peter Williams (British Geological Survey Wallingford) and Rosie Byrne (University of Manchester). BGS authors publish with the permission of the Executive Director of the British Geological Survey. The views expressed here do not necessarily represent those of the institutions, funders or individuals whose support is acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2023.166208>.

References

- Abd El-Salam, M.M., Abu-Zuid, G.I., 2015. Impact of landfill leachate on the groundwater quality: a case study in Egypt. *J. Adv. Res.* 6 (4), 579–586. <https://doi.org/10.1016/j.jare.2014.02.003>.
- Abdelrady, A., Sharma, S., Sefelnasr, A., Kennedy, M., 2020. Characterisation of the impact of dissolved organic matter on iron, manganese, and arsenic mobilisation during bank filtration. *J. Environ. Manag.* 258, 110003.
- Aftabtalab, A., Rinklebe, J., Shaheen, S.M., Niazi, N.K., Moreno-Jiménez, E., Schaller, J., Knorr, K.-H.H., 2022. Review on the interactions of arsenic, iron (oxy)(hydr) oxides, and dissolved organic matter in soils, sediments, and groundwater in a ternary system. *Chemosphere* 286 (December 2020), 131790. <https://doi.org/10.1016/j.chemosphere.2021.131790>.
- Ahmad, R., Kaushik, H., Ranjan, R.K., 2019. Assessment of microbial communities and heavy metals in urban soils of Patna, Bihar (India). *Arab. J. Geosci.* 12 (2) <https://doi.org/10.1007/s12517-018-4188-9>.
- Akbar, T.A., Hassan, Q.K., Ishaq, S., Batool, M., Butt, H.J., Jabbar, H., 2019. Investigative spatial distribution and modelling of existing and future urban land changes and its impact on urbanization and economy. *Remote Sens.* 11 (2) <https://doi.org/10.3390/rs11020105>.
- Alakshendra, A., 2019. City profile: Patna, India. *Environ. Urban. ASIA* 10 (2), 374–392. <https://doi.org/10.1177/0975425319859132>.
- Anabitarte, A., Subiza-Pérez, M., Ibarluzea, J., Azkona, K., García-Baquero, G., Miralles-Guasch, C., Irazusta, J., Whitworth, K.W., Vich, G., Lertxundi, A., 2020. Testing the multiple pathways of residential greenness to pregnancy outcomes model in a sample of pregnant women in the metropolitan area of Donostia-San Sebastián. *Int. J. Environ. Res. Public Health* 17 (12), 1–23. <https://doi.org/10.3390/ijerph17124520>.
- Bain, R., Cronk, R., Hossain, R., Bonjour, S., Onda, K., Wright, J., Yang, H., Slaymaker, T., Hunter, P., Prüss-Ustün, A., Bartram, J., 2014. Global assessment of exposure to faecal contamination through drinking water based on a systematic review. *Tropical Med. Int. Health* 19 (8), 917–927. <https://doi.org/10.1111/tmi.12334>.
- Baker, A., 2002. Fluorescence properties of some farm wastes: implications for water quality monitoring. *Water Res.* 36 (1), 189–195. [https://doi.org/10.1016/S0043-1354\(01\)00210-X](https://doi.org/10.1016/S0043-1354(01)00210-X).
- Baker, A., Inverarity, R., Charlton, M., Richmond, S., 2003. Detecting river pollution using fluorescence spectrophotometry: case studies from the Ouseburn, NE England. *Environ. Pollut.* 124 (1), 57–70. [https://doi.org/10.1016/S0269-7491\(02\)00408-6](https://doi.org/10.1016/S0269-7491(02)00408-6).
- Bianchi, T.S., 2011. The role of terrestrially derived organic carbon in the coastal ocean: a changing paradigm and the priming effect. In: *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108. National Academy of Sciences, pp. 19473–19481. <https://doi.org/10.1073/pnas.1017982108>. Issue 49.
- Bivins, A., Lowry, S., Murphy, H.M., Borchardt, M., Coyte, R., Labhasetwar, P., Brown, J., 2020. Waterborne pathogen monitoring in Jaipur, India reveals potential microbial risks of urban groundwater supply. *npj Clean Water* 3 (1). <https://doi.org/10.1038/s41545-020-00081-3>.
- Borsuk, M.E., Higdon, D., Stow, C.A., Reckhow, K.H., 2001. A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. *Ecol. Model.* 143 (3), 165–181.
- Boyle, S.A., Kennedy, C.M., Torres, J., Colman, K., Pérez-Estigarribia, P.E., De La Sancha, N.U., 2014. High-resolution satellite imagery is an important yet underutilized resource in conservation biology. *PLoS ONE* 9 (1), 86908. <https://doi.org/10.1371/JOURNAL.PONE.0086908>.
- Burford, J.R., Bremner, J.M., 1975. Relationships between the denitrification capacities of soils and total, water-soluble and readily decomposable soil organic matter. *Soil Biol. Biochem.* 7 (6), 389–394.
- Castan, S., Sigmund, G., Hüffer, T., Tepe, N., Von Der Kammer, F., Chefetz, B., Hofmann, T., 2020. The importance of aromaticity to describe the interactions of organic matter with carbonaceous materials depends on molecular weight and sorbent geometry. *Environ. Sci. Process. Impacts* 22 (9), 1888–1897. <https://doi.org/10.1039/DOEM00267D>.
- CGWB, 2015. Central Ground Water Board Department of Water Resources, River Development and Ganga Rejuvenation Ministry of Jal Shakti Pilot Studies on Aquifer Mapping Bhu-Jal News Editorial Board.
- Chaves, M.E.D., Picoli, M.C.A., Sanches, I.D., 2020. Recent applications of Landsat 8/OLI and Sentinel-2/MSI for land use and land cover mapping: a systematic review. *Remote Sens.* 12 (18) <https://doi.org/10.3390/rs12183062>.
- Chen, X., Liu, M., Kuzyakov, Y., Li, W., Liu, J., Jiang, C., Wu, M., Li, Z., 2018. Incorporation of rice straw carbon into dissolved organic matter and microbial biomass along a 100-year paddy soil chronosequence. *Appl. Soil Ecol.* 130, 84–90. <https://doi.org/10.1016/j.apsoil.2018.06.004>.
- Coble, P.G., Lead, J., Baker, A., Reynolds, D.M., Spencer, R.G.M., 2014. Aquatic organic matter fluorescence. In: *Aquatic Organic Matter Fluorescence*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139045452>.
- Dittmar, T., Paeng, J., 2009. A heat-induced molecular signature in marine dissolved organic matter. *Nat. Geosci.* 2 (3), 175–179. <https://doi.org/10.1038/ngeo440>.
- Erbán, L.E., Gorelick, S.M., Zebker, H.A., Fendorf, S., 2013. Release of arsenic to deep groundwater in the Mekong Delta, Vietnam, linked to pumping-induced land subsidence. *Proc. Natl. Acad. Sci. U. S. A.* 110 (34), 13751–13756. <https://doi.org/10.1073/pnas.1300503110>.
- Farooq, S.H., Chandrasekharam, D., Berner, Z., Norra, S., Stüben, D., 2010. Influence of traditional agricultural practices on mobilization of arsenic from sediments to groundwater in Bengal delta. *Water Res.* 44 (19), 5575–5588. <https://doi.org/10.1016/j.watres.2010.05.057>.
- Ghosh, A., Roy, M.B., Roy, P.K., 2022. Analysing LULC Change on Runoff and Sediment Yield in Urbanizing Agricultural Watershed of Monsoonal Climate River Basin in West Bengal, India. Springer, Singapore. https://doi.org/10.1007/978-981-16-6966-8_2.
- Government of India, 2011. *Census of India 2011. Provisional Population Totals*. Government of India, New Delhi, pp. 409–413.
- Guo, H., Li, X., Xiu, W., He, W., Cao, Y., Zhang, D., Wang, A., 2019. Controls of organic matter bioreactivity on arsenic mobility in shallow aquifers of the Hetao Basin, P.R. China. *J. Hydrol.* 571, 448–459. <https://doi.org/10.1016/j.jhydrol.2019.01.076>.
- 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. *Sci. Total Environ.* 854 (July 2022), 158542 <https://doi.org/10.1016/j.scitotenv.2022.158542>.
- Hedges, J.I., 1992. Global biogeochemical cycles: progress and problems. *Mar. Chem.* 39 (1–3), 67–93. [https://doi.org/10.1016/0304-4203\(92\)90096-S](https://doi.org/10.1016/0304-4203(92)90096-S).
- Held, I., Wolf, L., Eiswirth, M., Hötzel, H., 2006. Impacts of sewer leakage on urban groundwater. In: *Urban Groundwater Management and Sustainability*. Springer, pp. 189–204.
- Hertkorn, N., Harir, M., Koch, B.P., Michalke, B., Schmitt-Kopplin, P., 2013. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* 10 (3), 1583–1624. <https://doi.org/10.5194/bg-10-1583-2013>.
- Hofmann, R., Griebler, C., 2018. DOM and bacterial growth efficiency in oligotrophic groundwater: absence of priming and co-limitation by organic carbon and phosphorus. *Mol. Plant-Microbe Interact.* 31 (3), 311–322. <https://doi.org/10.3354/ame01862>.

- Hofmann, R., Uhl, J., Hertkorn, N., Griebler, C., 2020. Linkage between dissolved organic matter transformation, bacterial carbon production, and diversity in a shallow oligotrophic aquifer: results from flow-through sediment microcosm experiments. *Front. Microbiol.* 11 (November) <https://doi.org/10.3389/fmicb.2020.543567>.
- IIPS, 2022. National Family Health Survey (NFHS-5), 2019–21. International Institute for Population Sciences.
- Kowalczyk, P., Cooper, W.J., Durako, M.J., Kahn, A.E., Gonsior, M., Young, H., 2010. Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: relationships between fluorescence and its components, absorption coefficients and organic carbon concentrations. *Mar. Chem.* 118 (1–2), 22–36. <https://doi.org/10.1016/j.marchem.2009.10.002>.
- Kulkarni, H., Mladenov, N., Datta, S., 2019. Effects of acidification on the optical properties of dissolved organic matter from high and low arsenic groundwater and surface water. *Sci. Total Environ.* 653 (November 2018), 1326–1332. <https://doi.org/10.1016/j.scitotenv.2018.11.040>.
- Kulkarni, H.V., Mladenov, N., Johannesson, K.H., Datta, S., 2017. Contrasting dissolved organic matter quality in groundwater in Holocene and Pleistocene aquifers and implications for influencing arsenic mobility. *Appl. Geochem.* 77, 194–205. <https://doi.org/10.1016/j.apgeochem.2016.06.002>.
- Kulkarni, H.V., Mladenov, N., Datta, S., Chatterjee, D., 2018. Influence of monsoonal recharge on arsenic and dissolved organic matter in the Holocene and Pleistocene aquifers of the Bengal Basin. *Sci. Total Environ.* 637–638, 588–599. <https://doi.org/10.1016/j.scitotenv.2018.05.009>.
- Lakowicz, J.R., 1994. *Topics in Fluorescence Spectroscopy: Volume 4: Probe Design and Chemical Sensing*, Vol. 4. Springer Science & Business Media.
- Lakowicz, J.R., 2006. *Principles of Fluorescence Spectroscopy*. <https://doi.org/10.1007/978-0-387-46312-4>.
- Lapworth, D.J., Goody, D.C., Butcher, A.S., Morris, B.L., 2008. Tracing groundwater flow and sources of organic carbon in sandstone aquifers using fluorescence properties of dissolved organic matter (DOM). *Appl. Geochem.* 23 (12), 3384–3390. <https://doi.org/10.1016/j.apgeochem.2008.07.011>.
- Lapworth, D.J., Goody, D.C., Allen, D., Old, G.H., 2009. Understanding groundwater, surface water, and hyporheic zone biogeochemical processes in a chalk catchment using fluorescence properties of dissolved and colloidal organic matter. *J. Geophys. Res. Biogeosci.* 114 (G3).
- Lapworth, D.J., Das, P., Shaw, A., Mukherjee, A., Civil, W., Petersen, J.O., Goody, D.C., Wakefield, O., Finlayson, A., Krishan, G., Sengupta, P., MacDonald, A.M., 2018. Deep urban groundwater vulnerability in India revealed through the use of emerging organic contaminants and residence time tracers. *Environ. Pollut.* 240, 938–949. <https://doi.org/10.1016/j.envpol.2018.04.053>.
- Lawrence, A.R., Goody, D.C., Kanatharana, P., Meesilp, W., Ramnarong, V., 2000. Groundwater evolution beneath Hat Yai, a rapidly developing city in Thailand. *Hydrogeol. J.* 8 (5), 564–575. <https://doi.org/10.1007/s100400000098>.
- Lawson, M., Polya, D.A., Boyce, A.J., Bryant, C., Ballentine, C.J., 2016. Tracing organic matter composition and distribution and its role on arsenic release in shallow Cambodian groundwaters. *Geochim. Cosmochim. Acta* 178, 160–177. <https://doi.org/10.1016/j.gca.2016.01.010>.
- Leenheer, J.A., Croué, J.-P., 2003. *Characterizing aquatic dissolved organic matter*. In: *Environmental Science and Technology*, Vol. 37. UTC (Issue 6). <https://pubs.acs.org/sharingguidelines>.
- Lu, C., Richards, L.A., Wilson, G.J.L., Krause, S., Lapworth, D.J., Goody, D.C., Chakravorty, B., Polya, D.A., Niasar, V.J., 2022. Quantifying the impacts of groundwater abstraction on Ganges river water infiltration into shallow aquifers under the rapidly developing city of Patna, India. *J. Hydrol. Reg. Stud.* 42, 101133. <https://doi.org/10.1016/J.EJRH.2022.101133>.
- Lyngkilde, J., Christensen, T.H., 1992. Redox zones of a landfill leachate pollution plume (Vejen, Denmark). *J. Contam. Hydrol.* 10 (4), 273–289. [https://doi.org/10.1016/0169-7722\(92\)90011-3](https://doi.org/10.1016/0169-7722(92)90011-3).
- Matilainen, A., Gjessing, E.T., Lahtinen, T., Hed, L., Bhatnagar, A., Sillanpää, M., 2011. An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. *Chemosphere* 83 (11), 1431–1442. <https://doi.org/10.1016/J.CHEMOSPHERE.2011.01.018>.
- McArthur, J.M., Sikdar, P.K., Hoque, M.A., Ghosal, U., 2012. Waste-water impacts on groundwater: Cl/Br ratios and implications for arsenic pollution of groundwater in the Bengal Basin and Red River Basin, Vietnam. *Sci. Total Environ.* 437, 390–402. <https://doi.org/10.1016/j.scitotenv.2012.07.068>.
- McDonough, L.K., Santos, I.R., Andersen, M.S., O'Carroll, D.M., Rutledge, H., Meredith, K., Oudone, P., Bridgeman, J., Goody, D.C., Sorensen, J.P.R., Lapworth, D.J., MacDonald, A.M., Ward, J., Baker, A., 2020. Changes in global groundwater organic carbon driven by climate change and urbanization. *Nat. Commun.* 11 (1), 1–10. <https://doi.org/10.1038/s41467-020-14946-1>.
- McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* 46 (1), 38–48. <https://doi.org/10.4319/lo.2001.46.1.0038>.
- Milledge, D.G., Gurjar, S.K., Bunce, J.T., Tare, V., Sinha, R., Carbonneau, P.E., 2018. Population density controls on microbial pollution across the Ganga catchment. *Water Res.* 128, 82–91. <https://doi.org/10.1016/j.watres.2017.10.033>.
- Mladenov, N., Zheng, Y., Miller, M.P., Nemergut, D.R., Legg, T., Simone, B., Hageman, C., Rahman, M.M., Ahmed, K.M., McKnight, D.M., 2010. Dissolved organic matter sources and consequences for iron and arsenic mobilization in Bangladesh aquifers. *Environ. Sci. Technol.* 44 (1), 123–128. <https://doi.org/10.1021/es901472g>.
- Mohanta, T., Goel, S., 2014. Assessment of water quality of three different aquatic environments over three seasons. In: *1st International Congress on Environmental, Biotechnology, and Chemistry Engineering*. <https://doi.org/10.7763/IPCBBE>.
- Mukherjee, A., Bhanja, S.N., Wada, Y., 2018. Groundwater depletion causing reduction of baseflow triggering Ganges river summer drying. *Sci. Rep.* 8 (1), 1–9. <https://doi.org/10.1038/s41598-018-30246-7>.
- Murphy, K.R., 2011. A note on determining the extent of the water Raman peak in fluorescence spectroscopy. *Appl. Spectrosc.* 65 (2), 233–236.
- Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Anal. Methods* 5 (23), 6557–6566. <https://doi.org/10.1039/c3ay41160e>.
- Neumann, D., Heuer, A., Hemkemeyer, M., Martens, R., Tebbe, C.C., 2014. Importance of soil organic matter for the diversity of microorganisms involved in the degradation of organic pollutants. *ISME J.* 8 (6), 1289–1300. <https://doi.org/10.1038/ismej.2013.233>.
- Nischitha, V., Ahmed, S.A., Varikoden, H., Revadekar, J.V., 2014. The impact of seasonal rainfall variability on NDVI in the Tunga and Bhadra river basins, Karnataka, India, 35(23), pp. 8025–8043. <https://doi.org/10.1080/01431161.2014.979301>.
- NWIC, N. W. I. C., 2017. No Title. Land Use and Land Cover Distribution of Bihar of Patna of Year 2017.
- 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. [https://doi.org/10.1016/S0146-6380\(00\)00124-8](https://doi.org/10.1016/S0146-6380(00)00124-8).
- Postma, D., Larsen, F., Minh Hue, N.T., Duc, M.T., Viet, P.H., Nhan, P.Q., Jessen, S., 2007. Arsenic in groundwater of the Red River floodplain, Vietnam: controlling geochemical processes and reactive transport modeling. *Geochim. Cosmochim. Acta* 71 (21), 5054–5071. <https://doi.org/10.1016/j.gca.2007.08.020>.
- Prasad, G., Ramesh, M.V., Ramesh, T., Thomas, G.M., 2023. Changing profile of natural organic matter in groundwater of a Ramsar site in Kerala implications for sustainability. *Case Stud. Chem. Environ.* 8, 100390. <https://doi.org/10.1016/j.csee.2023.100390>.
- Pucher, M., Wünsch, U., Weigelhofer, G., Murphy, K., Hein, T., Graeber, D., 2019. staRDom: versatile software for analyzing spectroscopic data of dissolved organic matter in R. *Water* 11 (2366). <https://doi.org/10.3390/w11123666>.
- QGIS Development Team, 2020. QGIS Geographic Information System. Open Source Geospatial Foundation Project.
- Ramanathan, A.L., Bhattacharya, P., Keshari, A.K., Chandrasekharan, D., 2009. *Assessment of Groundwater Resources and Management*. IK International Pvt Ltd.
- Revadekar, J.V., Tiwari, Y.K., Kumar, K.R., 2012. Impact of climate variability on NDVI over the Indian region during 1981–2010, 33(22), 7132–7150. <https://doi.org/10.1080/01431161.2012.697642>.
- Reynolds, D.M., 2003. Rapid and direct determination of tryptophan in water using synchronous fluorescence spectroscopy. *Water Res.* 37 (13), 3055–3060. [https://doi.org/10.1016/S0043-1354\(03\)00153-2](https://doi.org/10.1016/S0043-1354(03)00153-2).
- Richards, L.A., Lapworth, D.J., Magnone, D., Goody, D.C., Chambers, L., Williams, P.J., van Dongen, B.E., Polya, D.A., 2019. Dissolved organic matter tracers reveal contrasting characteristics across high arsenic aquifers in Cambodia: a fluorescence spectroscopy study. *Geosci. Front.* 10 (5), 1653–1667. <https://doi.org/10.1016/j.gsf.2019.04.009>.
- Richards, L.A., Kumari, R., White, D., Parashar, N., Kumar, A., Ghosh, A., Kumar, S., Chakravorty, B., Lu, C., Civil, W., Lapworth, D.J., Krause, S., Polya, D.A., Goody, D.C., 2021. Emerging organic contaminants in groundwater under a rapidly developing city (Patna) in northern India dominated by high concentrations of lifestyle chemicals. *Environ. Pollut.* 268, 115765. <https://doi.org/10.1016/J.ENVPOL.2020.115765>.
- Richards, L.A., Fox, B.G., Bowes, M.J., Khamis, K., Kumar, A., Kumari, R., Kumar, S., Hazra, M., Howard, B., Thorn, R.M.S., Read, D.S., Nel, H.A., Schneidewind, U., Armstrong, L.K., Nicholls, D.J.E., Magnone, D., Ghosh, A., Chakravorty, B., Joshi, H., Dutta, T.K., Hannah, D.M., Reynolds, D.M., Krause, S., Goody, D.C., Polya, D.A., 2022a. A systematic approach to understand hydrogeochemical dynamics in large river systems: development and application to the River Ganges (Ganga) in India. *Water Res.* 211, 118054. <https://doi.org/10.1016/J.WATRES.2022.118054>.
- Richards, L.A., Kumari, R., Parashar, N., Kumar, A., Lu, C., Wilson, G., Lapworth, D., Niasar, V.J., Ghosh, A., Chakravorty, B., Krause, S., Polya, D.A., Goody, D.C., 2022b. Environmental tracers and groundwater residence time indicators reveal controls of arsenic accumulation rates beneath a rapidly developing urban area in Patna, India. *J. Contam. Hydrol.* 104043. <https://doi.org/10.1016/J.JCONHYD.2022.104043>.
- Roberts, D.A., Keller, M., Soares, J.V., 2003. Studies of land-cover, land-use, and biophysical properties of vegetation in the Large Scale Biosphere Atmosphere experiment in Amazonia. *Remote Sens. Environ.* 87 (4), 377–388. <https://doi.org/10.1016/J.RSE.2003.08.012>.
- Rodríguez-Vidal, F.J., Ortega-Azabache, B., González-Martínez, Á., Bellido-Fernández, A., 2022. Comprehensive characterization of industrial wastewaters using EEM fluorescence, FT-IR and ¹H NMR techniques. *Sci. Total Environ.* 805, 150417.
- Saha, D., Dwivedi, S.N., Singh, R.K., 2014. Aquifer system response to intensive pumping in urban areas of the Gangetic plains, India: the case study of Patna. *Environ. Earth Sci.* 71 (4), 1721–1735. <https://doi.org/10.1007/s12665-013-2577-7>.
- Salman, S.A., Shahid, S., Mohsenipour, M., Asgari, H., 2018. Impact of landuse on groundwater quality of Bangladesh. *Sustain. Water Resour. Manag.* 4 (4), 1031–1036. <https://doi.org/10.1007/S40899-018-0230-Z>.
- Shen, Y., Chapelle, F.H., Strom, E.W., Benner, R., 2014. Origins and bioavailability of dissolved organic matter in groundwater. *Biogeochemistry* 122 (1), 61–78. <https://doi.org/10.1007/S10533-014-0029-4>.
- Siemens, J., Haas, M., Kaupenjohann, M., 2003. Dissolved organic matter induced denitrification in subsoils and aquifers? *Geoderma* 113 (3–4), 253–271.

- Singh, A., Patel, A.K., Deka, J.P., Das, A., Kumar, A., Kumar, M., 2020. Prediction of arsenic vulnerable zones in the groundwater environment of a rapidly urbanizing setup, Guwahati, India. *Geochemistry* 80 (4), 125590. <https://doi.org/10.1016/j.chemer.2019.125590>.
- Singh, S.K., Ghosh, A.K., Kumar, A., Kislay, K., Kumar, C., Tiwari, R.R., Parwez, R., Kumar, N., Imam, M.D., 2014. Groundwater arsenic contamination and associated health risks in Bihar, India. *Int. J. Environ. Res.* 8 (1), 49–60.
- Song, Z., Williams, C.J., Edyvean, R.G.J., 2000. Sedimentation of tannery wastewater. *Water Res.* 34 (7), 2171–2176.
- Sorensen, J.P., Nayebare, J., Carr, A.F., Lyness, R., Campos, L.C., Ciric, L., Goodall, T., Kulabako, R., Curran, C.M.R., Macdonald, A.M., Owor, M., 2021. In-situ fluorescence spectroscopy is a more rapid and resilient indicator of faecal contamination risk in drinking water than faecal indicator organisms. *Water Res.* 206, 117734 <https://doi.org/10.1016/j.watres.2021.117734>.
- Stedmon, C.A., Bro, R., 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnol. Oceanogr. Methods* 6 (11), 572–579. <https://doi.org/10.4319/lom.2008.6.572>.
- Stedmon, C.A., Markager, S., 2005. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol. Oceanogr.* 50 (2), 686–697. <https://doi.org/10.4319/LO.2005.50.2.0686>.
- Tittebrand, A., Spank, U., Bernhofer, C., 2009. Comparison of satellite- and ground-based NDVI above different land-use types. *Theor. Appl. Climatol.* 98 (1), 171–186. <https://doi.org/10.1007/S00704-009-0103-3>.
- USGS, 2019. EarthExplorer. <https://earthexplorer.usgs.gov/>.
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* 185, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>.
- Wallis, I., Prommer, H., Berg, M., Siade, A.J., Sun, J., Kipfer, R., 2020. The river–groundwater interface as a hotspot for arsenic release. *Nat. Geosci.* 13 (4), 288–295. <https://doi.org/10.1038/s41561-020-0557-6>.
- Ward, J.S.T., Lapworth, D.J., Read, D.S., Pedley, S., Banda, S.T., Monjerezi, M., Gwengweya, G., Macdonald, A.M., 2021. Tryptophan-like fluorescence as a high-level screening tool for detecting microbial contamination in drinking water. *Sci. Total Environ.* 750, 141284 <https://doi.org/10.1016/j.scitotenv.2020.141284>.
- Wells, M.J.M., Hooper, J., Mullins, G.A., Bell, K.Y., 2022. Development of a fluorescence EEM-PARAFAC model for potable water reuse monitoring: implications for inter-component protein–fulvic–humic interactions. *Sci. Total Environ.* 820, 153070.
- 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. *Limnol. Oceanogr.* 55 (3), 1159–1171. <https://doi.org/10.4319/lo.2010.55.3.1159>.
- Wünsch, U., Bro, R., Stedmon, C., Wenig, P., Murphy, K., 2019. Emerging patterns in the global distribution of dissolved organic matter fluorescence. *Anal. Methods* 11 (7), 888–893. <https://doi.org/10.1039/C8AY02422G>.
- Yang, Y., Yuan, X., Deng, Y., Xie, X., Gan, Y., Wang, Y., 2020. Seasonal dynamics of dissolved organic matter in high arsenic shallow groundwater systems. *J. Hydrol.* 589, 125120 <https://doi.org/10.1016/j.jhydrol.2020.125120>.
- Zaitunah, A., Samsuri, S., Ahmad, A.G., Safitri, R.A., 2018. Normalized difference vegetation index (NDVI) analysis for land cover types using Landsat 8 OLI in Besitang watershed, Indonesia. *IOP Conf. Ser.: Earth Environ. Sci.* 126 (1) <https://doi.org/10.1088/1755-1315/126/1/012112>.