Optimising water quality monitoring to understand biogeochemical processes in UK rivers and reservoirs

A thesis submitted for the degree of *Doctor of Philosophy*

by

Inge Elfferich

School of Earth and Environmental Sciences

Cardiff University

October 2024

Summary

This thesis explores how high-frequency sensor data can improve understanding of environmental and biogeochemical conditions in inland water bodies, particularly for predicting cyanobacteria-related taste and odour (T&O) compounds geosmin and 2-MIB in drinking water reservoirs. Since these compounds cannot be directly measured *in situ*, alternative predictors are required.

High-frequency hydrochemical sensor data from six UK rivers showed that a minimum measurement frequency of four hours was required to capture the necessary variation in the parameters. For T&O prediction, key drivers need to be identified that high-frequency sensors can measure. These drivers were studied in laboratory microcosms with reservoir water, which found that nutrient ratios (N:P and NH_4^+:NO_3^-) and concentrations impacted phytoplankton productivity, community composition and the production of 2-MIB. 2-MIB production was linked to benthic cyanobacteria (*Leptolyngbya* spp. and *Pseudanabaena* spp.) and community productivity, with total nitrogen being a stronger predictor than nutrient ratios, but P-limitation also played a role. NH_4^+ caused rapid 'boom-bust' growth, while NO_3^- led to more balanced growth four days later. Long-term manual sampling and sensor data from 2022-2023 in the same reservoir showed that optimal weather conditions led to increased productivity and a delayed T&O response. Geosmin concentrations (likely from planktic cyanobacteria) increased 1-2 weeks after high solar energy and low rainfall, while 2-MIB (likely from benthic cyanobacteria) was linked to high temperatures and low wind gusts 2-3 weeks earlier.

While N:P and NH_4^+:NO_3^- ratios were direct predictors of geosmin and 2-MIB, these are not useful for an early-warning system, because elevated NH_4^+ and PO_4^{3-} concentrations were likely related to internal loading and T&O release after cell death. The most useful predictors were environmental parameters that support phytoplankton growth (seasonality, solar energy, air temperature, precipitation and wind speed), indirect growth indicators (dissolved oxygen, vertical mixing), and measurements of inflows to the reservoir via a tributary (NO_3^--N, PO_4^{3-}-P, NH_4^+-N or EC and turbidity, alongside discharge). Appropriate *in situ* measurement of these factors, especially when combined with weather forecasts, can provide early warnings of T&O events that can guide water treatment management.

Acknowledgements

Last but not least, I thank all my friends and fellow PhD students, especially Dami, Jalila, Ece, Roberto, Jaz, Emanuela, Max, James, Lisa, Manuel, Elle, Sara, Ellie, Ana, Erin, Emma, Abi, Fiona, Laura, Gwynfor, Nico and Sophie. Thank you, Emma, Rachel, Sarah and Gemma for your friendship and support, and sharing my love for arts and craft. I thank my parents, sister and grandparents for always believing in me and encouraging me when I needed it. I am incredibly grateful for the support I had from my partner Simon. You kept me sane in tough times and cooked me dinner if I was home late, I wouldn't have been able to do this without you!

Author note and status of publications

First author thesis contributions

The following paper has contributed to this thesis (correct as of thesis submission):

- **<u>Chapter 2:</u>** Elfferich, I., Bagshaw E.A., Perkins, R.G., Johnes, P.J., Yates, C.A., Lloyd, C.E.M., Bowes, M.J., Halliday, S.J. (2024). Interpretation of river water quality data is strongly controlled by measurement time and frequency. *Science of The Total Environment, 954,* 176626.

<u>Author contributions and declarations:</u> IE: Writing – original draft, writing – review & editing, data curation, investigation, methodology, formal analysis, conceptualization, visualization. EB: Writing – review & editing, supervision, funding acquisition, methodology, conceptualization. RP: Writing – review & editing, supervision, methodology, conceptualization. PJ: Writing – review & editing, methodology, resources (access to unpublished data), supervision. CY: Writing – review & editing, supervision, resources (access to unpublished data). CL: Writing – review & editing. MB: Writing – review & editing, resources (access to unpublished data). SH: Writing – review & editing.

Co-author indirect thesis contributions

Personal contributions as co-author to the following papers, understanding of these papers has been used in this thesis but no text or data was directly reported:

- Reinl, K.L., Harris, T.D., Elfferich, I., Coker, A., Zhan, Q., Domis, L.N.D.S., Morales-Williams, A.M., Bhattacharya, R., Grossart, H.-P., North, R.L., Sweetman, J.N. (2022). The role of organic nutrients in structuring freshwater phytoplankton communities in a rapidly changing world. *Water Research*, *219,* 118573.

<u>Author contributions and declarations:</u> All authors conceptualized the paper objectives, reviewed literature and contributed to writing. All authors read and approved the final manuscript. KLR and TDH lead the development, writing, and editing of this review. IE led development of the concept figures and assisted in editing. AC and QZ assisted in table development and editing. LNSD and AMM assisted with significant editing of

manuscript drafts. RB, HPG, RN, and JS provided input and contributed material to the manuscript.

- Harris, T.D., Reinl, K.L., Azarderakhsh, M., Berger, S.A., Berman, M.C., Bizic, M., Bhattacharya, R., Burnet, S.H., Cianci-Gaskill, J.A., Domis, L.N.d.S., Elfferich, I., Ger, K.A., Grossart, H.-P.F., Ibelings, B.W., Ionescu, D., Kouhanestani, Z.M., Mauch, J., McElarney, Y.R., Nava, V., North, R.L., Ogashawara, I., Paule-Mercado, M.C.A., Soria-Píriz, S., Sun, X., Trout-Haney, J.V., Weyhenmeyer, G.A., Yokota, K., Zhan, Q. (2024). What makes a cyanobacterial bloom disappear? A review of the abiotic and biotic cyanobacterial bloom loss factors. *Harmful Algae, 133*, 102599.

Personally selected co-author contribution (see paper for full authorship contribution statement): IE: Conceptualization, Visualization, Writing – original draft, Writing – review & editing, Project administration.

Table of contents

List of Figures

List of Tables

Chapter 1: General Introduction

1.1 Cyanobacterial blooms

Anthropogenic eutrophication is a global freshwater challenge which is likely to increase due to an increasing world population (Trommer et al., 2020). Eutrophication is used to describe the symptoms that waterbodies exhibit in response to nutrient fertilisation (Schindler et al., 2008). These symptoms may include increased primary productivity, phytoplankton growth, cyanobacterial blooms, a reduction in macrophytes and anoxic deeper layers during stratification due to decaying organic matter (Moss et al., 2011). Eutrophication together with high light availability and warm temperatures are some of the environmental factors that enhance cyanobacterial growth and persistent blooms (Paerl & Otten, 2013; Richardson et al., 2019). Climate change will likely result in more optimal growth conditions for cyanobacteria because of the higher water temperature (Cottingham et al., 2015; De Senerpont Domis et al., 2007) and increased nutrient availability due to runoff from agricultural or urban areas, internal loading from sediments and longer residence times (Paerl & Otten, 2013; Paerl & Paul, 2012). Cyanobacterial blooms are a major water quality issue because they can affect animal and human health, partly due to mass cell death which results in oxygen depletion at the sediment (Alghanmi et al., 2018) but also due to cyanotoxins that some cyanobacteria can produce (Lürling & Faassen, 2013). Moreover, these blooms disrupt the use of water resources for example for fisheries, agriculture, recreation or drinking water and this results in economic loss.

1.2 Nutrient sources and dynamics

The most important macronutrients that drive eutrophication and support phytoplankton growth are nitrogen, phosphorus and carbon. The sources and dynamics of these nutrients in lakes or reservoirs can be separated into external processes (catchment inflows) and internal processes (cycling within the lake or reservoir).

1.2.1 External nutrient processes

Nitrogen

In freshwaters, the two main inorganic fractions of nitrogen are the oxidised form as nitrate (NO_3^--N) and the reduced form as ammonium (NH_4^+-N) but dissolved organic nitrogen (DON) is also important for phytoplankton (Mackay et al., 2020).

NO_3^--N enters freshwaters from agricultural land and urban areas, and due to its highly mobile nature it infiltrates quickly in groundwater where it accumulates (Elser et al., 2007; Heathwaite et al., 1996; Schmutz & Sendzimir, 2018). This frequently results in a strong dependence of NO_3^--N on hydrological connectivity in the catchment (Chen et al., 2018). Typical nitrate concentrations in pristine streams are 0.1 mg/L of NO_3^--N (Allan & Castillo, 2007; Meybeck et al., 1990) but the average NO_3^--N concentration in European rivers was 1.5 mg/L in 2016, whilst average groundwater concentrations were 18 mg/L in 2012 (Kristensen et al., 2018). Slurry runoff can result in spikes >60 mg/L of NO_3^--N (Vidal et al., 2000) and sewage inputs can cause spikes of total oxidised nitrogen (TOxN = NO_3^--N + NO_2^--N) of ±30 mg/L of N (Halliday et al., 2014; Halliday et al., 2015).

NH_4^+-N is a water-soluble ion that can turn into ammonia gas (NH_3) which is toxic to aquatic life (Dodds & Whiles, 2020). Farming practices and urbanisation can cause elevated concentrations of NH_4^+-N in freshwaters, through direct inputs of sewage or slurry, and soil runoff with NH_4^+-N adsorbed to charged soil particles (Donald et al., 2011; Schmutz & Sendzimir, 2018) as well as atmospheric deposition (Neal et al., 2012). In pristine streams the ammonium concentration is usually around 0.015 mg/L of NH_4^+-N (Allan & Castillo, 2007). Urbanised rivers affected by sewage runoff and septic leachate were found to exhibit spikes of 0.97 mg/L NH_4^+-N (Mallin & McIver, 2012), 3.49 mg/L NH_4^+-N (Halliday et al., 2014) and 5.4 mg/L NH_4^+-N (Halliday et al., 2015). Rivers affected by slurry showed spikes exceeding 2 mg/L of NH_4^+-N (Quilbé et al., 2005; Vidal et al., 2000).

Phosphorus

In freshwaters, phosphorus exists in the dissolved phase either as inorganic orthophosphate (PO_4^{3-}-P) and dissolved organic phosphorus (DOP), but is largely in the particulate phase adsorbed onto particles (Schmutz & Sendzimir, 2018) or inside

organisms, both of which are largely unavailable to phytoplankton on short-medium timescales.

The most bioavailable phosphorus fraction PO_4^{3-}-P is often approximated by measuring soluble reactive phosphate (SRP). PO_4^{3-}-P has a strong binding capacity to compounds in soils, like clay particles and metals; Aluminium (Al), Magnesium (Mg), Iron (Fe), Copper (Cu) and Calcium (Ca). Therefore, PO_4^{3-}-P enters freshwaters during runoff periods, attached to soil particles from agricultural land where fertilizer or slurry has been applied (Schmutz & Sendzimir, 2018). In pristine streams the PO_4^{3-}-P concentration is usually < 0.01 mg/L of P (Allan & Castillo, 2007), but the average PO_4^{3-}-P concentration in European rivers was 0.045 mg/L of P in 2016 (Kristensen et al., 2018). Sewage runoff was found to cause spikes of up to 5.4 and 9.3 mg/L of PO_4^{3-}-P (Halliday et al., 2014; Halliday et al., 2015) while slurry runoff resulted in spikes of 0.6 mg/L of PO_4^{3-}-P (Vidal et al., 2000).

Carbon

In freshwaters, carbon exists in the dissolved phase as inorganic carbon and organic carbon. Inorganic carbon fractions are photosynthesis-mediated equilibrium products of carbonic acid (H_2CO_3). Besides dissolved organic carbon (DOC), which are the carbon molecules in dissolved organic matter (DOM), there is also particulate organic carbon, which are a part of live or decaying organic matter (Wetzel, 2001). Total organic carbon (TOC) includes dissolved and particulate fractions. There is a range of different organic carbon-containing substances, humic and non-humic, with different molecular weights and availability for microbial decomposition (Wetzel, 2001). Phytoplankton, including cyanobacteria, obtain inorganic carbon as CO_2 through photosynthesis and convert it into organic compounds. Additionally, they can acquire organic carbon by directly absorbing certain fractions of DOC or by consuming other organisms (mixotrophy) (Reinl et al., 2022).

DOC can enter freshwaters via groundwater and runoff as part of dissolved organic matter (DOM), in seminatural catchments this is natural organic material from soils and plants etc. whereas DOM has an anthropogenic origin in catchments impacted by agricultural intensification or urbanisation (e.g. sewage treatment works) (Yates et al., 2019a). DOM inputs and its nutrient composition depend largely on the catchment; land use, soil type (Worrall et al., 2008), vegetation, hydrology and temperature

(control on microbial decomposition), nutrient cycling, soil uptake and mobility (Wetzel, 2001; Yates et al., 2019a). In pristine streams, average DOC concentrations were 2 mg/L (Allan & Castillo, 2007), whilst DOC concentrations in rivers affected by urban runoff (sewage) and septic leachate, showed spikes up to 21.2 mg C/L (Halliday et al., 2015).

Organic N and P

Despite being less bioavailable than inorganic phases, dissolved organic nitrogen (DON) and dissolved organic phosphorus (DOP), which are part of dissolved organic matter (DOM), have been suggested as an important fraction for phytoplankton growth (Mackay et al., 2020; Reinl et al., 2022). Concentrations of ammonium, urea (CH_4N_2O), dissolved organic nitrogen (DON) and dissolved organic phosphorus (DOP) are increasing in freshwaters due to sewage discharges, aquaculture and the worldwide rise of urea-based fertilizers (Glibert et al., 2016). Especially in areas of livestock farming, the inputs from manures and slurries are associated with high concentrations of DON and DOP such as urea (CH_4N_2O), amino acids and phytic acids (Turner et al., 2002).

1.2.2 Internal nutrient processes

Waterbodies with retention times of less than one month (0.07-0.08 years) are expected to limit phytoplankton biomass development, while resupplying nutrients from the catchment (Maberly et al., 2020). Lakes have longer retention times than rivers, leading to nutrient accumulation and enabling internal nutrient recycling. Bioavailable fractions of nutrients are taken up and stored as organic forms in live biomass (organic matter), whilst dead organic matter undergoes microbial decomposition at the sediment and this process breaks down organic N, P and C into bioavailable fractions again (Wetzel, 2001). Most nutrients fractions accumulate near the sediment in the stable period and get dispersed into the water column during periods of physical (wind) mixing (Jones & Welch, 1990), but some fractions become gaseous and volatilize, like CO_2, CH_4, N_2, NH_3 etc. (Santschi et al., 1990).

Phytoplankton play a major role in nutrient cycling through; I) photosynthesis-respiration cycles (uptake and release of CO_2 and dissolved oxygen), II) nitrogen fixing capacities of some cyanobacteria, and III) producing substantial biomass (organic matter) followed by decaying organic matter (Cottingham et al., 2015; Parmar et al.,

2011). Decomposition of phytoplankton debris released bioavailable P rapidly to the aquatic environment, resulting in positive feedback loops (Feng et al., 2018).

Internal loading

A stable water column (vertical stability) with warm temperatures, which often happens in summer, will increase microbial decomposition, lower the redox potential and result in anaerobic conditions at the sediment-water interface (SWI), which increases the internal loading of NH_4^+-N and PO_4^{3-}-P (Jensen & Andersen, 1992; Jeppesen et al., 2009; Jones & Welch, 1990; Yao et al., 2023). Anaerobic conditions initiate the accumulation of NH_4^+-N through ammonification and suppressed nitrification (McCarthy et al., 2016; Perkins et al., 2019). Moreover, it transforms metals such as manganese (Mn), aluminium (Al) and iron (Fe) in the sediment to the dissolved fraction (Santschi et al., 1990), and releases PO_4^{3-}-P from metal complexes, such as $Al(OH)_3$ and Fe(III) (Hupfer & Lewandowski, 2008; Søndergaard et al., 2003). Kang et al. (2018) found that anaerobic sediment conditions doubled concentrations of P (TP, TDP, SRP) and N (TN) at the sediment-water interface compared to the sediments in aerobic environments, whilst NH_4^+-N increased 12 times.

Denitrification

Anaerobic conditions at the sediment will also initiate microbial denitrification, the conversion from NO_3^--N to N_2 (g) (Seitzinger, 1988). The loss of NO_3^--N via denitrification is affected by flow speed, residence time, water temperature, oxygen availability and organic matter (Wang et al., 2019a). Seitzinger (1988) estimated that denitrification removed between one and 36% of the nitrate inputs to six studied lakes.

1.3 Taste and odour (T&O)

Some species of cyanobacteria can produce secondary metabolites that give the water an unfavourable smell and taste. So-called taste and odour (T&O) compounds have not been related to human health impacts, but they are problematic for drinking water sources as it impacts customer confidence. T&O problems in drinking water are globally increasing in frequency and magnitude (Winter et al., 2011). This causes increased customer complaints (Bai et al., 2017) and high costs of additional powdered or granular activated carbon treatment required due to T&O compounds stable nature and resistance to oxidation (Dunlap et al., 2015). Certain treatment processes can be used to remove T&O compounds as a single method or combined

approach, including granulated activated carbon (GAC) as an in-line final treatment step, manual additions of powdered activated carbon (PAC), ozonation or chlorination. T&O treatment accounted for 8% of annual plant operating costs in Saskatchewan (Kehoe et al., 2015). T&O problems occur in rivers (Chung et al., 2016; Espinosa et al., 2020; Park et al., 2018), lakes (Brownlee et al., 2004; Durrer et al., 1999; Jiang et al., 2016; Kutovaya & Watson, 2014; Li et al., 2007; Ma et al., 2013; Peter et al., 2009; Qi et al., 2020a), drinking water reservoirs (Bai et al., 2017; Billica et al., 2010; Dzialowski et al., 2009; Jähnichen et al., 2011; Jørgensen et al., 2016; Kehoe et al., 2015; Olsen et al., 2016a; Otten et al., 2016; Perkins et al., 2019; Su et al., 2013; Tsao et al., 2014; Westerhoff et al., 2005) and aquaculture (Auffret et al., 2011). T&O compounds were found to accumulate in animal tissue which can impact the economic value of fish from fisheries (Alghanmi et al., 2018).

1.4 Geosmin and 2-MIB

Among the many T&O compounds, geosmin (1,10-dimethyl-trans-9-decalol) and 2-MIB (2-Methylisoborneol) are the most common and well-studied (Figure 1.1). Geosmin and 2-MIB are secondary metabolites which have an unfavourable earthy or muddy smell and taste, and they can be detected by humans at low thresholds; 6.3 ng/L for 2-MIB and 1.3 ng/L for geosmin (Wert et al., 2014).

(-)-Geosmin   (-)-2-Methylisoborneol
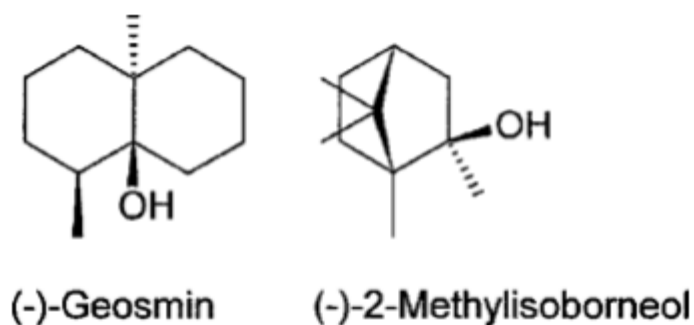
Figure 1.1 The molecular structure of geosmin and 2-MIB (Watson et al., 2016).

Geosmin and 2-MIB can be produced by a variety of microorganisms including actinomycetes, cyanobacteria, proteobacteria, myxobacteria and some fungi (Watson et al., 2016). Planktic and benthic cyanobacteria, mostly filamentous species, are the main source of geosmin and 2-MIB production in aquatic environments (Watson et al., 2016) but actinomycetes can also be present in the sediment (Asquith et al., 2018; Clercin & Druschel, 2019) or they play a role via soil runoff (Kehoe et al., 2015).

Several genera of cyanobacteria have been identified as geosmin producers, including *Dolichospermum* (formerly *Anabaena*), *Aphanizomenon*, *Nostoc*, *Cylindrospermum*, *Nodularia*, *Calothrix*, *Scytonema*, *Tychonema*, *Lyngbya*, *Leptolyngbya*, *Oscillatoria*, *Phormidium* and *Planktothrix.* Approximately 75 cyanobacterial species, growing in benthic, planktic or epiphytic habitats, are known to produce geosmin (Wang et al., 2019c; Watson et al., 2016). Cyanobacteria with the ability to produce 2-MIB are less well understood and can sometimes produce geosmin as well. So far, the genera *Oscillatoria, Lyngbya, Leptolyngbya, Phormidium, Microleus, Planktothrix, Pseudanabaena, Tychonema* and *Synechococcus* are identified as 2-MIB producers, and the 35 cyanobacterial species are mostly benthic or epiphytic (Watson et al., 2016). Certain cyanobacteria genera can be benthic or planktic, for example filamentous *Pseudanabaena* sp. has been recorded as a planktic (Yue et al., 2024) but also as a benthic (Gaget et al., 2017; Perri et al., 2024) 2-MIB producer. It is possible these genera occupy different habitats throughout their lifecycle (Cottingham et al., 2021), but there could also be a large variation in species traits between species within a genus or differences in habitat due to certain environmental conditions. Planktic species of cyanobacteria have been found to grow on the sediment for extended periods of time, if there is sufficient light availability (Boström et al., 1989).

In a Welsh drinking water reservoir, geosmin was most likely produced by planktic cyanobacteria *Dolichospermum* sp. while *Oscillatoria* sp. occurred as a benthic biofilm in shallower regions and was a likely source of both 2-MIB and geosmin (Perkins et al., 2019). A large body of research (Espinosa et al., 2020; Gaget et al., 2020; Jähnichen et al., 2011; Otten et al., 2016; Watson & Jüttner, 2019) indicates that benthic mats of cyanobacteria should not be overlooked as a source of geosmin and 2-MIB, especially in rivers, shallow lakes and shores of reservoirs (Watson & Jüttner, 2019). Some studies attributed geosmin and 2-MIB concentrations in deeper water layers to actinomycetes (Chislock et al., 2021) or diatoms (Olsen et al., 2016b).

1.5 Geosmin and 2-MIB drivers

It is not certain why cyanobacteria produce geosmin and 2-MIB, but a few theories are: 1) to dump excess energy when there are high light conditions (Behr et al., 2014) or when cells switch from oxidized NO_3^- to reduced NH_4^+ (Perkins et al., 2019), 2) to act as a carbon source for other bacteria (Guttman & van Rijn, 2012), 3) to stimulate

algae to produce alkaline phosphatases (APs) which will free up organic phosphorus (Raven, 2010), or 4) to act as allelopathic compounds that inhibit the growth of other phytoplankton (Achyuthan et al., 2017; Lee et al., 2017; Ozaki et al., 2008). It is generally accepted that T&O production happens intracellularly during the exponential growth phase, with peak production in the late exponential stage, followed by extracellular release of T&O compounds when cells enter the stationary and death phase (Alghanmi et al., 2018; Naes et al., 1989; Saadoun et al., 2001; Zhang et al., 2009). However, it is still uncertain whether cyanobacteria can *actively* release T&O compounds (Ma et al., 2013); possibly to stimulate enzymatic P release (Bar-Yosef et al., 2010; Raven, 2010), or if these compounds are *passively* released when cells die and release their contents (Zhang et al., 2016), or if both processes are involved (Wang & Li, 2015; Watson, 2003). Another discussion in the literature is whether chlorophyll *a* production affects T&O production negatively due to competing synthesis pathways (Kutovaya & Watson, 2014; Liu et al., 2009; Pattanaik & Lindberg, 2015; Saadoun et al., 2001; Shen et al., 2020; Wang & Li, 2015), or positively due to T&O accumulation as a result of increased cell metabolism (Alghanmi et al., 2018; Giglio et al., 2011; Zimba et al., 1999). Additionally, the amount of T&O compound that is produced per cyanobacterial cell may vary depending on the species or strain, as well as environmental factors (Chiu et al., 2016; Saadoun et al., 2001; Wang et al., 2016; Watson & Ridal, 2004).

Geosmin and 2-MIB can be produced by benthic and planktic cyanobacterial species, which will have different species traits that determine the environmental conditions that support their growth. Planktic cyanobacteria are often the source of geosmin, and they rely on light availability, water temperature and nutrients. Some planktic species can control buoyancy which gives them a competitive advantage during vertical stability and some species have N_2-fixation capabilities, which allows them thrive during N-limitation and at low N:P ratios (Miller et al., 2013). Benthic cyanobacteria are often the source of 2-MIB, and they rely on light availability, water temperature, a stable water column with low sediment resuspension and nutrients. However, these species are not so reliable on nutrients in the water column because they can access nutrients in the sediment (Jähnichen et al., 2011) and are low-light acclimated so they can survive periods of disturbance (Abeynayaka et al., 2018; Gao et al., 2018).

The difference in growth habitat makes it harder to determine the source of T&O production, as extracellular T&O compounds can remain present in the water column whilst they get transported away from their source (Watson & Jüttner, 2019). The underlying complexities in the environmental drivers, time-lags between intracellular T&O production and extracellular release, and challenges in representatively measuring T&O within the water column, explain why cyanobacterial biomass is not always clearly correlated with T&O concentration in field studies (Billica et al., 2010; Harris et al., 2016; Jähnichen et al., 2011; Tsao et al., 2014; Wu et al., 1991). However, there are also examples were cyanobacterial biomass did correlate with T&O concentrations (Peter et al., 2009). The location, severity and timing of T&O events is unpredictable because it depends on a multidimensional scale of interacting processes which causes spatial and temporal complexity (Watson et al., 2016). Intracellular molecular processes that involve coding, signalling and T&O production play a role as well as temperature, light, nutrients, food web interactions combined with large- and small-scale hydrodynamics. T&O drivers could be site-specific and a combination of environmental factors, which was the case for cyanotoxins in three urban lakes in Western Australia (Sinang et al., 2015).

Environmental and catchment processes determine the physical and chemical conditions of a water body, and this influences phytoplankton community structure, which determines the potential for 2-MIB and/or geosmin production (Figure 1.2). Phytoplankton growth and T&O production (C) is indirectly or directly related to the following processes: A) Nutrient dynamics: external inputs (related to catchment characteristics) and internal nutrient processes; as well as B) Hydrodynamics, temperature and light availability. These processes also impact one another, which causes complexity and frequent interdependence (Figure 1.2).
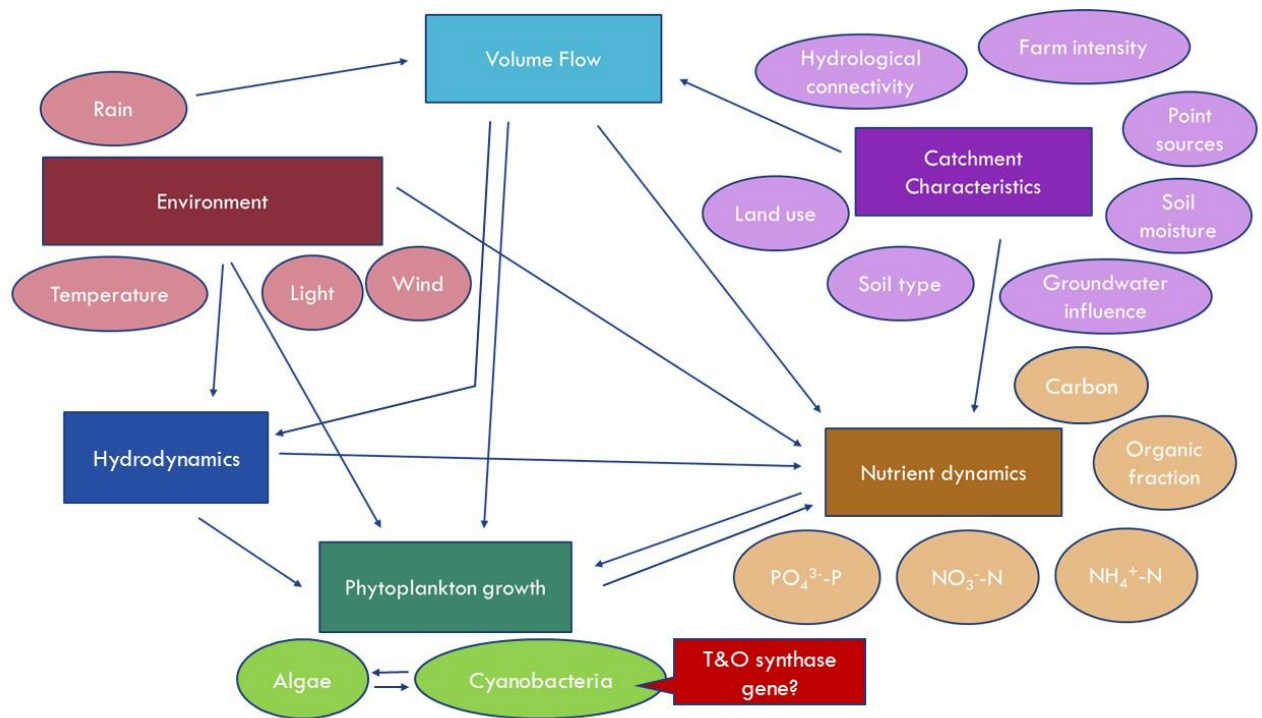
Figure 1.2 Environmental and catchment processes that influence phytoplankton growth and phytoplankton community response which can result in taste and odour production in a lake or reservoir.

1.6 Nutrient dynamics (A)

Nitrogen (N) and phosphorus (P) concentrations and their ratios play an important role in shaping phytoplankton community structure (Harris et al., 2016). Water column ratios DIN:SRP (dissolved inorganic nitrogen : soluble reactive phosphorus) and TN:TP (total nitrogen : total phosphorus) are used to approximate potential N- or P-limitation for phytoplankton growth. The Redfield ratio of 16:1 for TN:TP is considered the threshold between N-limitation (<16) and P-limitation (>16) (Redfield et al., 1963), although this can vary depending on the phytoplankton species and growth stage (Klausmeier et al., 2004). N-limitation has also been defined as TN:TP <20 and DIN:SRP <13, and P-limitation as TN:TP >38 and DIN:SRP >50 (Kosten et al., 2009). The TN:TP ratio is heavily influenced by both external inputs and internal cycling of N and P, which can lead to highly variable ratios over a short space of time (Glibert et al., 2008). PO_4^{3-}-P uptake by bloom-forming cyanobacteria happens within 15-25 minutes and increases productivity rates instantly (Aubriot & Bonilla, 2012). The form of N and P and subsequent bioavailability depends on the dominant nutrient source, which can vary depending on the seasons (Andersen et al., 2020).

10

Cyanobacterial dominance in phytoplankton communities is often associated with a low TN:TP ratio (Smith, 1983; Vrede et al., 2009), with Smith (1983) identifying a threshold of less than 64:1 (29:1 as mass ratio). P-limited conditions tend to favour chlorophytes (green algae) and diatoms, whilst cyanobacteria are more likely to dominate under N-limiting conditions (Andersen et al., 2020). Harris et al. (2014) identified a TN:TP ratio >75:1 for chlorophytes dominance and low cyanobacteria biomass. Several cyanobacterial taxa have specific species traits, such as N_2-fixation and the capability to produce enzymes that enable the uptake of otherwise unavailable forms of nitrogen and phosphorus (Chaffin & Bridgeman, 2014; Chen et al., 2020). Species capable of N_2-fixation, like *Aphanizomenon* sp., often dominate during N-limitation and at low TN:TP ratios (Miller et al., 2013).

The chemical form of inorganic nitrogen, oxidized as NO_3^--N or reduced as NH_4^+-N, affects the uptake rate among different phytoplankton taxa due to a difference in cellular energetic cost. NH_4^+-N is more readily available, whilst NO_3^--N uptake requires activation of a different transport mechanism and/or the reduction of NO_3^--N to NH_4^+-N through enzymatic processes (Erratt et al., 2020; Glibert et al., 2016; Trommer et al., 2020). Uptake of NO_3^--N can be suppressed when NH_4^+-N is available (Glibert et al., 2016). Optimal NH_4^+-N concentrations can promote faster growth compared to NO_3^--N and urea, but excessively high concentrations can inhibit growth or even become toxic, with optimal concentrations depending on the class, genus or species of phytoplankton (Collos & Harrison, 2014; Glibert et al., 2016). In lakes and reservoirs, (harmful) algal blooms have been linked to elevated concentrations of NH_4^+-N and PO_4^{3-}-P from agricultural inputs (Chen et al., 2018; Glibert, 2020; Glibert et al., 2016; Perkins et al., 2019) and internal loading from the sediment during periods of vertical stability (Hoffman et al., 2022; Yao et al., 2023). The NH_4^+:NO_3^- ratio can influence phytoplankton community structure (Glibert et al., 2016; McCarthy et al., 2009) and the abundance of T&O producing cyanobacteria. NO_3^- enrichment tends to promote diatom growth, while NH_4^+ enrichment favours dinoflagellates, chlorophytes (green algae), and (toxic) cyanobacteria (Andersen et al., 2020; Domingues et al., 2011; Donald et al., 2013; Donald et al., 2011; Erratt et al., 2020; Glibert & Berg, 2009; McCarthy et al., 2009). However, research by Kim et al. (2017) suggested that the NH_4^+:NO_3^- ratio and the N:P ratio alone did not explain the maximum growth of the

cyanobacterium *Microcystis aeruginosa*, but this was determined by the TN concentration instead when a minimum concentration of PO_4^{3-} was present.

1.7 Hydrodynamics, temperature and light availability (B)

Temperature and light availability are important factors that influence phytoplankton growth, but the optimal temperature and light availability is specific for each species or even strain. Each species has species traits and optimal growth habitats, which determine how the natural environment shapes the phytoplankton community. After a period of phytoplankton growth, a sudden change in temperature or light can cause mass cell death, which can release intracellular compounds like T&O from the cells and have large implications for a drinking water supply (Ma et al., 2013).

Hydrodynamics in a lake or reservoir are strongly associated with weather patterns; stable dry weather with a low wind speed will result in vertical stability, whilst unstable rainy weather with higher windspeeds will cause vertical mixing, catchment runoff and hydraulic flushing. Vertical stability of the water column often happens in summer during periods of warm stable weather and this process influences nutrient dynamics by increasing internal loading (PO_4^{3-}-P, NH_4^+-N, etc.) and light availability by reducing sediment resuspension in shallow lakes. When periods of vertical stability are disrupted by a weather event (rainfall, wind) this causes vertical mixing and compounds that were accumulated at the sediment will be dispersed into the water column (Qi et al., 2020b). A weather events can also result in catchment runoff, which can impact nutrient loading, as well as hydraulic flushing which can result in phytoplankton loss (Reynolds, 2006).

Short-term meteorological variability was found more important for planktic cyanobacterial bloom response than average temperature for two years with summer heatwaves (Huber et al., 2012). In one year long-lasting (>3 weeks) thermal stratification promoted blooms, whilst the other year had no blooms due to a period of lower temperatures and stronger winds mid-summer that weakened stratification (Huber et al., 2012). During a period of stratification, little chemical transport happens between the surface and bottom of the lake or reservoir, which means that nutrients can get depleted in the surface layer, whilst they accumulate at the bottom. Planktic phytoplankton that are capable of buoyancy control will have a competitive advantage as they can move to nutrient-rich lower water (Wagner & Adrian, 2009; Walsby, 1994),

while also able to access enough light near the surface (Ibelings et al., 1991). Warm, dry conditions with a stable water column were found to cause the most severe cyanobacteria blooms due to increased internal phosphorus loading, whilst summer runoff and precipitation caused bloom die-off due to reduced water temperature, light availability and water column mixing (Hecht et al., 2022). Climate change and lake warming is expected to result in stronger seasonal fluctuations as well as higher internal loading caused by longer periods of vertical stability, which will result in optimal conditions for long-lasting cyanobacterial blooms (Yindong et al., 2021).

Benthic cyanobacteria have access to nutrients in the sediment, so their growth mainly depends on light availability and temperature, which are both impacted by hydrodynamics. Hydrodynamics also impact the stability of the sediment where they grow, vertical mixing causes disturbance and potentially burial of cells. Benthic cyanobacteria need approximately 1% of surface irradiance to grow, but this is species dependent (Jähnichen et al., 2011). Increasing water levels as well as suspended particles, a high organic matter content or a planktic algal bloom in the water can reduce the light availability at the sediment. Benthic cyanobacteria can proliferate if light availability in the littoral zone is sufficiently high, which is more likely in low-nutrient conditions due to low turbidity (Jähnichen et al., 2011).

1.8 Phytoplankton growth and T&O production (C)

Multiple mesocosm and laboratory culture studies have been conducted to investigate the drivers of T&O production by cyanobacteria. An overview with details of these studies can be found in Appendix A Table S.1. The following factors were investigated; nutrient concentrations and ratios TN:TP as well as $NH_4^+:NO_3^-$, vertical mixing, light availability and temperature.

Oh et al. (2017) concluded that NO_3^--N increased geosmin and 2-MIB production, whereas most other research states that NH_4^+-N stimulated production and NO_3^--N inhibited it (Clercin & Druschel, 2019; Harris et al., 2016; Perkins et al., 2019; Saadoun et al., 2001; Shen et al., 2020). The inhibitory effect was demonstrated by Saadoun et al. (2001) and Shen et al. (2020), as elevated NO_3^--N concentrations increased cell growth and chlorophyll *a* production by *Dolichospermum smithii* and *Anabaena* sp., while geosmin synthesis decreased. The production of 2-MIB by *Nostoc* sp. and *Anabaena* sp. seemed to be suppressed by the presence of added nitrogen, but

geosmin production was not affected (Milovanović et al., 2015). Geosmin production by benthic *Oscillatoria* sp. was related to high concentrations of NH_4^+-N, NO_3^--N and PO_4^{3-}-P and a low DIN:SRP ratio of 4:1 (Espinosa et al., 2021) and geosmin production by planktic *Anabaena* sp. was also stimulated by a low N:P ratio and NH_4^+-N additions (Saadoun et al., 2001). Water temperature and NH_3-N were positively correlated with 2-MIB production by planktic *Pseudanabaena* sp. (Gao et al., 2018; Lee et al., 2023). Several field studies identified that geosmin and/or 2-MIB concentrations related to low TN:TP ratios (Winston et al., 2014; Yue et al., 2024) as well as high NH_4^+:NO_3^- ratios (Harris et al., 2016; Perkins et al., 2019). Harris et al. (2016) suggested a T&O risk threshold of TN:TP <66:1 (30:1 as mass ratio).

Ma et al. (2018) showed that high concentrations of NH_4^+-N in aquatic ecosystems can cause a stoichiometric N:P imbalance which results in a sudden P limitation. Dzialowski et al. (2009) hypothesized that limitation of inorganic phosphorus was an important determinant of geosmin production by cyanobacteria in five Kansas (USA) reservoirs. During P-limitation it is possible that some T&O producing cyanobacteria excrete alkaline phosphatases; enzymes to free up P from dissolved organic forms (Chen et al., 2020), or release T&O metabolites to stimulate other phytoplankton to produce these enzymes, which is shown for cyanotoxins (Bar-Yosef et al., 2010; Raven, 2010). Phosphorus inhibited T&O production only when it was limited but no changes in T&O production were reported due to increasing levels of phosphorus (Oh et al., 2017). Dzialowski et al. (2009) found a negative relationship between PO_4^{3-}-P concentrations and geosmin levels in Lake Huron (USA), which contained the cyanobacterial taxa *Aphanizomenon, Anabaena, Microcystis* and *Oscillatoria*. This finding emphasizes that PO_4^{3-}-P concentrations may relate to T&O concentrations differently than TP in field studies, as algal blooms convert inorganic P into organic P within its biomass, which raises TP levels (Dodds & Whiles, 2020). Moreover, algal blooms can cause stronger internal nutrient cycling, because decaying blooms release high concentrations of nitrogen and phosphorus (Wang et al., 2019b), as well as taste and odour compounds (Ma et al., 2013). Adam et al. (2016) found that colony-forming *Aphanizomenon* sp. released about half of its recently fixed N_2 (g) as NH_4^+-N, which provided an N-source to the phytoplankton community.

Vertical stratification enhanced 2-MIB concentrations produced by *Leptolyngbya* sp. in a Chinese reservoir (Yue et al., 2024). Paerl et al. (2022) found a negative correlation

between discharge and 2-MIB concentrations in a drinking water reservoir in North Carolina (USA), which suggested stratification favoured the growth of 2-MIB producers. Compounds like geosmin and 2-MIB can be released from the bottom water through the whole reservoir when mixing disrupts the vertical stratification, which was discovered in Lake Taihu, China (Qi et al., 2020b). Olsen et al. (2016b) found that nutrient enrichment (phosphorus and nitrogen) quickly boosted geosmin regardless of mixing, which was linked to cyanobacterial biovolume, whilst 2-MIB was influenced by nutrient enrichment and mixing, and was also associated with cyanobacteria. Benthic cyanobacteria (*Oscillatoria* sp.) in rivers were found to produce the highest geosmin concentrations when there were low flow conditions (0.09 L/s) and low light conditions (20% of natural light conditions; 107 µmol photons m^{-2} s^{-1}) (Espinosa et al., 2020). The expression levels of 2-MIB-related genes in *Pseudanabaena foetida* var. *intermedia* were light-dependent; unfavourable light conditions (dark period or green light) significantly reduced 2-MIB production, possibly due to demand for photopigment synthesis (Dayarathne et al., 2024). This indicates that a certain threshold of available light is required for T&O production by benthic cyanobacteria. Sudden changes in water level have been shown to reduce growth of benthic cyanobacteria (Jähnichen et al., 2011), which has been used as a management strategy for 2-MIB (Su et al., 2017). Su et al. (2017) used the reservoir shape, bathymetry and light coefficient of the water to calculate areas of potential benthic T&O producing cyanobacterial growth to inform their water level management. However, a sudden increase in water level can also release high levels of 2-MIB when conditions become unfavourable, and cells lyse (Rong et al. 2018).

In laboratory experiments with samples from Lake Taihu (China) by Huang et al., (2018a), high cyanobacteria biomass led to increased T&O concentrations. Additionally, higher temperatures were found to increase rates of cell decomposition, which triggered T&O release (Huang et al., 2018a). Yen et al. (2007) found a positive correlation between water temperature and 2-MIB production in two drinking water reservoirs in Taiwan. However, several other studies suggest that growth-inhibiting temperatures and light intensities might inhibit cyanobacteria chlorophyll *a* production whilst stimulating or forcing the production of geosmin and/or 2-MIB (Wang & Li, 2015; Zhang et al., 2009). Saadoun et al. (2001) found in laboratory cultures of *Anabaena* sp. that increasing light intensity at 20 °C boosted geosmin production and chlorophyll

a, while higher temperatures favoured chlorophyll *a* production and higher light intensity favoured geosmin production. Tsao et al. (2014) showed that geosmin production was higher in deeper water layers with low light conditions, compared to surface water with a high light intensity. Wang and Li (2015) found that growth inhibiting conditions and subsequent T&O production was species-specific; *Anabaena ucrainica* produced highest geosmin concentrations at 10 °C and 60 µmol photons m^{-2} s^{-1} and *Pseudanabaena* sp. Produced highest 2-MIB concentrations at 35 °C and 60 µmol photons m^{-2} s^{-1}. Jeong et al. (2021) found that excretion of 2-MIB by benthic *Pseudanabaena yagii* happened slowly in the growth phase but rapidly increased when the temperature dropped.

1.9 Modelling, prediction and early warning

1.9.1 Planktic cyanobacteria predictions

Predicting cyanobacterial blooms in freshwater has been attempted by many studies, but certain elements are site-specific, and it is often hard to extrapolate the findings to other lakes and reservoirs. Water column stability, hypolimnetic TP, % cyanobacteria biovolume two weeks prior and 7-day mean wind speed was found to predict % cyanobacteria biovolume in two Canadian temperate lakes (Persaud et al., 2015). Descy et al. (2016) found that phosphorus (total and soluble reactive phosphorus), dissolved inorganic nitrogen, epilimnion temperature, photoperiod and euphotic depth, were good predictive variables depending on the ecological traits of the dominant taxa (*Aphanizomenon*, *Microcystis*, *Planktothrix* and *Anabaena*), as well as meteorological factors wind, rainfall and surface irradiance. Recknagel et al. (2016) developed forecasting models for *Anabaena*, *Aphanizomenon* and *Microcystis* for Lake Müggelsee and found that water temperature and transparency, concentrations of NO$_3^-$-N and PO$_4^{3-}$-P were key predictors, but the same species had different thresholds during a hypertrophic and a eutrophic phase. Zhang et al. (2012) determined that sunshine hours and wind speed were the primary contributors to onset of the cyanobacterial blooms in Lake Taihu (China) and were also good predictors of variability in duration of annual blooms. They suggest that when nutrients are sufficiently high, climatic variables are critical in predicting cyanobacterial blooms. Meteorological factors as temperature, wind speed and wind direction were combined in an index, to determine cyanobacteria occurrence at a drinking water intake in Canada, which were found to explain 68% of a bloom that occurred (Ndong et al.,

2014). A modelling study by Hecht et al. (2022) found that cyanobacterial blooms in Lake Champlain (USA) were driven by warm, dry conditions with a stable water column (internal loading) whilst summer runoff and precipitation caused water column mixing as well as lower temperature and light availability, which caused bloom die-off.

1.9.2 Benthic cyanobacteria predictions

Fewer studies have been undertaken to predict benthic cyanobacterial growth. Sediment stability and light availability are important factors that determine benthic cyanobacterial growth, which is why cyanobacterial growth in rivers has been related to low water flow (Espinosa et al., 2020). Su et al. (2017) used bathymetry to classify areas of the littoral zone with sufficient light availability as high-risk zones for deep-living *Planktothrix* sp. in Miyun Reservoir (China) and determined that water level management could significantly reduce taste and odour problems related to this species. Benthic cyanobacteria *Lyngbya wollei* was found downstream of tributaries draining farmlands in the St. Lawrence River (Canada), and models identified that DOC (positive), TP (negative) and DIN:TDP (negative) were important variables explaining its spatial distribution (Lévesque et al., 2012). They hypothesised the growth of benthic *Lyngbya wollei* was related to high DOC, low phosphorus and nitrogen under relatively low DIN:TDP ratios, because this resulted in little competition with macrophytes. However, it is likely that nutrient concentrations were much higher within the benthic mat than water column samples revealed. Observations of *Phormidium* sp. mats in New Zealand showed that this taxa was tolerant to velocity, depth and substrate type, and that its abundance seemed to increase with increasing nitrogen concentrations (Heath et al., 2015). Perri et al. (2024) found that toxic benthic cyanobacterial mats, including *Pseudanabaena* sp. and *Leptolyngbya* sp. were associated with a low daily river inflow, high TP and NO_3^- concentrations as well as the day of the year.

1.10 T&O prediction

There are several examples of studies in which a model has been created to predict T&O events (Chong et al., 2018; Dzialowski et al., 2009; Kehoe et al., 2015; Parinet et al., 2013). Most studies to date have used simple linear regression models with a range of different water quality parameters related to phytoplankton productivity (turbidity or water transparency, chlorophyll *a* and total phosphorus) (Dzialowski et al., 2009). Some non-linear models exist that have extended model predictions based on

microbial abundance data, phytoplankton data and hydrodynamics, or they have incorporated phytoplankton data in hydrodynamic models (Chong et al., 2018). More recently artificial intelligence and machine learning have been used for T&O prediction (Kehoe et al., 2015; Parinet et al., 2013). Kehoe et al. (2015) created a random forest model to forecast T&O events in a reservoir in Saskatchewan, Canada, using algal data (T&O producing algae taxa), temperature, total phosphorus and chlorophyll *a*. The model showed highest T&O predictive capacity with a fortnight time-lag and was able to predict taste and odour threshold number levels up to 12 weeks in advance (Kehoe et al., 2015). Parinet et al. (2013) tested various models to predict geosmin concentrations in three raw water sources in Canada. The best multi-linear regression model used four variables (phaeophytin, sum of green algae, chlorophyll *a* and potential Redox), whilst the best model with artificial neural networks included ten variables (potassium, heterotrophic bacteria, organic nitrogen, total nitrogen, phaeophytin, total organic carbon, sum of green algae, potential Redox, UV absorbance at 254 nm and atypical bacteria) (Parinet et al., 2013). A non-linear Cubist model was able to predict geosmin concentrations in Cheney Reservoir (Kansas, USA), with total Kjeldahl nitrogen, day of the year and silica as most important predictors (Harris & Graham, 2017). In this model, chlorophyll *a* could predict maximum geosmin concentrations when turbidity was >22.2 FNU and silica was <10.4 mg/L. Regression analysis and self-organising maps revealed that cyanobacterial blooms of certain species, water temperature, reservoir volume and oxidised nitrogen availability were key factors to predict geosmin events in an Australian reservoir (Bertone & O'Halloran, 2016).

1.10.1 Existing sensors for water quality monitoring

Most drinking water treatment plants don't sample for T&O compounds frequently during the whole year, as the laboratory method is time consuming and expensive. There is often a decision framework in place that uses alert levels from manual sampling to identify periods of T&O risk in which they increase manual sampling and management through additional treatment (Kibuye et al., 2021). Delays between the detection of T&O events and potential management decisions are caused by the waiting time for manual laboratory results or the time passed until customers start complaining (Kibuye et al., 2021). These delays result in a reactive response at the treatment, whilst a pro-active response would be better, which highlights the

importance of developing T&O risk predictions and forecasts. Most currently available T&O models (Section 1.10) also use laboratory results, which would still not allow a pro-active response (forecasting) but can only be used for hindcasting. Real-time prediction of T&O events using high-frequency sensors is an area of research that still needs to be further explored, but it could provide direct evidence for pro-active management at the water treatment plant. For example, additional powdered activated carbon (PAC) can be added in preparation for elevated T&O concentrations.

Commercially available water quality sensors that might be useful for T&O prediction are: water temperature, water level (pressure), pH, oxidation-reduction potential (ORP), electrical conductivity (EC), turbidity, dissolved oxygen (DO), photosynthetically active radiation (PAR), dissolved organic matter (DOM) as CDOM (chromomorphic or coloured DOM) or fDOM (fluorescent DOM), photosynthetic pigments (chlorophyll *a*, phycocyanin, phycoerythrin) and nutrients (NO_3^--N, NH_4^+-N, $PO_4^{3-}-P$, TP) (Table 1.1). Standard sensors on a multiparameter probe often include EC, pH, ORP, DO, water temperature, water level and turbidity, which use well established sensing technologies (Table 1.1). Chlorophyll *a* and phycocyanin sensors measure fluorescence of the pigment *in situ*, which estimates total phytoplankton biomass and cyanobacteria-specific biomass, respectively (Painter et al., 2023) (Table 1.1). CDOM sensors measure UV-Visible absorbance and fDOM sensors measure fluorescence (Table 1.1), but they both provide estimates of DOM and can be used to estimate DOC in aquatic environments (Danhiez et al., 2017).

Table 1.1 Water quality parameters and their commercially available sensor technique and measurement method.

Sensor technique	Measurement method	Water quality parameters
Optical	Absorbance (UV-Vis)	Nitrate (NO_3^--N)/Nitrite (NO_2^--N)
		CDOM
		Turbidity
	Fluorescence	fDOM Protein-like (Tryptophan) and Humic-like
		Photosynthetic pigments: chlorophyll *a*, phycocyanin, phycoerythrin
		Dissolved oxygen (DO)
	Irradiance	Photosynthetically Active Radiation (PAR)
Electrochemical	Ion Selective Electrode	pH
		Oxidation-reduction potential (ORP)
		Nitrate (NO_3^--N)

		Ammonium (NH_4^+-N)
	Polarographic/Galvanic	Dissolved oxygen (DO)
	Inductive/Conductive	Electrical conductivity (EC)
Other	Thermistor	Temperature
	Transducer	Pressure
Wet chemistry analyser (reagent-based)	Colorimetric	Nitrate (NO_3^--N)/Nitrite (NO_2^--N)
		Ammonium (NH_4^+-N)
		Soluble reactive phosphorus (SRP)
		Total phosphorus (TP)

Recent advancements of *in situ* monitoring technology have resulted in the availability of nutrient sensors and analysers that can monitor NH_4^+-N, NO_3^--N, SRP and TP at a high frequency (Beaton et al., 2012; Beaton et al., 2017; Blaen et al., 2016; Clinton-Bailey et al., 2017; Daniel et al., 2020; Grand et al., 2017; Mowlem et al., 2021). Nitrate (NO_3^--N) can be measured with three different technologies: ion-selective electrode (electrochemical), UV-Visible absorbance (optical) and a wet chemistry analyser (reagent-based and colorimetric) (Table 1.1). Wet chemistry analysers use the same assay-based methods that are generally applied in laboratory analysis with reagents and a colour reaction, but on a microfluidic scale, which is why these analysers are also called "lab-on-chip". Ammonium (NH_4^+-N) can be measured with ion-selective electrode or a wet chemistry analyser (reagent-based and colorimetric). Soluble reactive phosphorus (SRP; indicative of PO_4^{3-}-P) and total phosphorus (TP) can currently only be measured with a wet chemistry analyser (reagent-based and colorimetric).

Different sensor technologies come with differences in cost, sensor performance (sensitivity, limit of detection or LoD, precision and accuracy; see Table 1.2), maintenance requirements, response time etc. which will determine the best choice of technology for a parameter and a certain monitoring purpose (Rozemeijer et al., 2025). Ion-selective electrodes are cheapest (few hundred pounds), but also least sensitive and are therefore mainly used in wastewater treatments or other applications where concentrations of analytes tend to be high (Bende-Michl & Hairsine, 2010). UV-Visible absorbance optical sensors are expensive (a few thousand pounds), they provide a better sensitivity but can also suffer from interferences within the water matrix and biofouling (Blaen et al., 2016; Daniel et al., 2020). Wet-chemical analysers are the most expensive (tens of thousands of pounds), they are most accurate and can measure at low LoD due to internal calibration solutions, but they also require more

maintenance (reagent replacements), produce waste and have a slower response time (Blaen et al., 2017; Daniel et al., 2020).

Table 1.2 Terms commonly used by sensor manufactures to describe sensor performance

Term used	Description of term
Limit of Detection (LoD)	The lowest amount of a substance present that can be reliably detected by a sensor
Range	The difference between the highest and lowest amount of a substance that can be detected by a sensor
Sensitivity	The smallest degree of change in the environment required to trigger a change in the sensor reading
Accuracy	The degree to which a given result given by a sensor is likely to vary from the actual amount of a substance present
Precision	The likelihood of obtaining the same result if exactly the same test is repeated by a sensor

1.10.2 Potential sensors for T&O prediction

Johnston et al. (2024) evaluated water quality and meteorological parameters that could be measured with *in situ* sensors and deployed as continuous monitoring systems for harmful algal blooms in lakes. If cyanobacterial blooms can be monitored continuously, this will provide further information on T&O risk. The overview in Figure 1.3 highlights which parameters could be measured with *in situ* sensors to capture the three key processes in lakes or reservoirs related to algal blooms: water column conditions (hydrodynamics, light availability and temperature), nutrient dynamics (external and internal) as well as phytoplankton productivity and biomass.
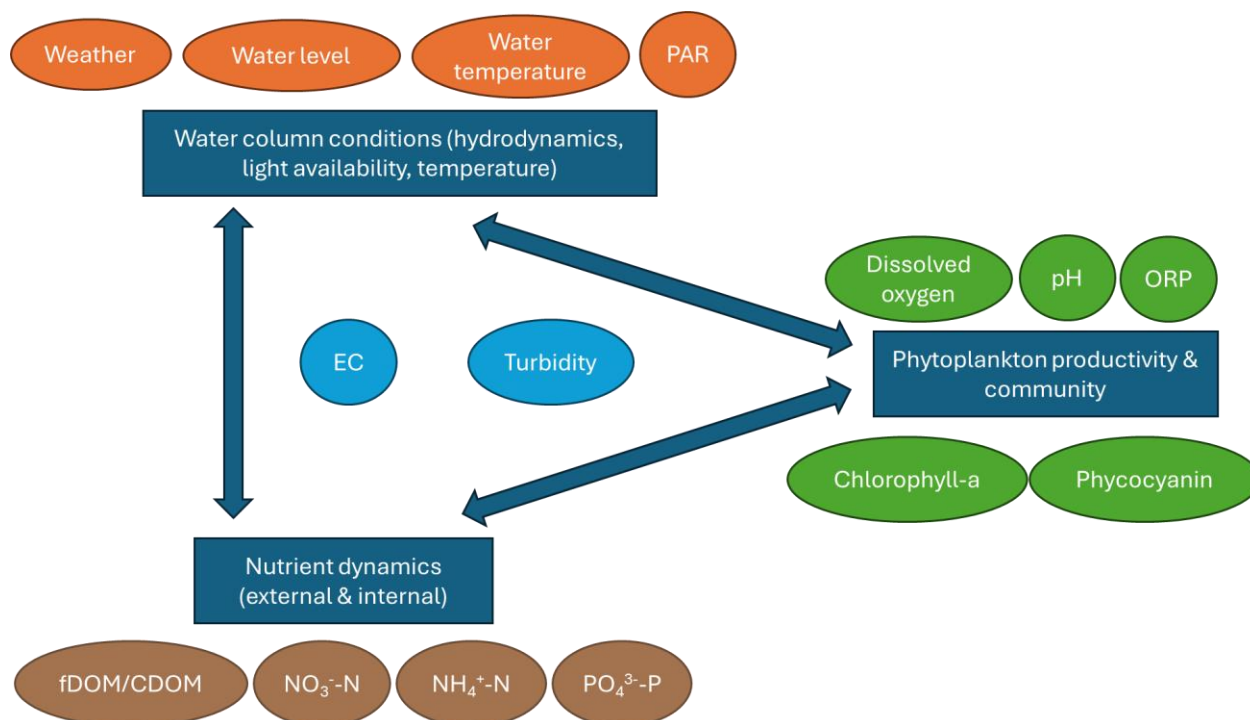
Figure 1.3 Key processes in lakes or reservoirs (dark blue squares) that are interconnected, and which can be identified in situ by measuring the parameters in the coloured circles adjacent to each blue square. The parameters in circles in the middle can be indicative of any of the three key processes.

Meteorological data from weather stations and potentially forecasts can provide insight in water column conditions, as well as measuring water level, water temperature and PAR with sensors, possibly at different depths in the water column (Figure 1.3). Nutrient dynamics are important in driving cyanobacterial blooms, and fDOM/CDOM, NO_3^--N, NH_4^+-N and PO_4^{3-}-P can be measured with sensors or analysers. Generally, a higher nutrient concentration would indicate increased risk of blooms, but DOM (fDOM/CDOM) can also affect light attenuation and limit phytoplankton growth, and has been found to interfere with phytoplankton fluorescence measurements (chlorophyll *a* and phycocyanin) (Johnston et al., 2024). Dissolved oxygen dynamics in the water indicate photosynthesis and respiration, and it can be used to estimate phytoplankton productivity as well as periods of cell death and decomposition related to oxygen consumption (Johnston et al., 2024). Measurements of pH and ORP also indicate processes related to photosynthesis processes. Fluorescence sensors can measure chlorophyll *a* to estimate phytoplankton biomass and phycocyanin to estimate cyanobacteria biomass. If they are both measured, the relative abundance of cyanobacteria in the total phytoplankton community can be estimated. Dedicated phytoplankton classification sensors exist that detect multiple fluorescence signals

and estimate the percentage of each group to the total phytoplankton community, which can provide useful insights into T&O risk, as long as manual samples are used alongside sensor data (Johnston et al., 2024).

The parameters turbidity and EC can indicate several physical and biogeochemical processes and need to be interpreted with catchment knowledge in mind. Changes in baseline measurements can indicate important events in the ecosystem, which can be further investigated using other sensors or manual sampling. Turbidity can provide an indication of suspended particles, light availability and mixing, but also interferes with fluorescence measurements (chlorophyll *a* and phycocyanin) (Johnston et al., 2024). EC can be related to the salinity of the lakes, which relates to water column stability and risk of algal cell lysis (Johnston et al., 2024), but also indicates significant pollution events in rivers (Halliday et al., 2014) or changes in chemistry of river inflows (Painter et al., 2023). Continuous real-time monitoring of the current state of the lake or reservoir can be combined with information from short-term local weather forecasts to potentially provide an early warning of cyanobacterial bloom risk and subsequent risk of T&O events.

1.10.3 Using sensors for water quality management

There are several examples of real-time sensor measurements that are used to inform water treatment management. Water temperature, pH, dissolved oxygen (DO), nitrogen, phosphate, organic matter and algal count was measured with real-time sensors at the raw water intake of a water treatment facility in Beijing, and DO could be used as an indicator for T&O events (Chen et al., 2019). A decrease in DO at the source water in late August was a result of decaying algae consuming oxygen, and the peak of T&O would happen a few weeks later (Chen et al., 2019). Chlorophyll *a*, water temperature and dissolved oxygen were dominant abiotic factors influencing T&O concentrations in Lake Taihu (China) (Xuwei et al., 2019), which could be measured with sensors. Clercin and Druschel (2019) studied Eagle Creek Reservoir (USA) and found a time-lag of 37 days between peak inflows from a river and 2-MIB and Geosmin. Investigating the site-specific lag-time between tributary inflow and T&O event is a useful tool that can be measured with sensors to provide an early warning.

Chlorophyll *a* and phycocyanin sensors are useful indicators of cyanobacteria risk (Almuhtaram et al., 2021b; Carey et al., 2021; Cotterill et al., 2019; Yang et al., 2019),

and potential for T&O compounds (Bertone et al., 2018; Painter et al., 2023). Zamyadi et al. (2016) highlighted the potential of *in situ* phycocyanin fluorescence probes in a drinking water treatment plant to optimise the dosing required to treat 2-MIB. However, without species information this does not mean T&O are produced by these species (Kibuye et al., 2021) and it won't detect benthic T&O producers (Almuhtaram et al., 2021b). Jalili et al. (2022) suggested a decision framework to manage cyanobacteria and cyanotoxins at a drinking water treatment plant combining phycocyanin sensor measurements at the raw water intake and laboratory results of toxins and T&O measurements. Dissolved oxygen and chlorophyll *a* sensors were used in a German drinking water reservoir to investigate vertical patterns in oxygen and chlorophyll *a* during stratification, which highlighted the effect of *Planktothrix* sp. on anoxia (Wentzky et al., 2019). A reservoir in Virginia (USA) was equipped with a meteorological station, DO and water temperature sensors at different depths and a UV-Visible spectrophotometer, to model hydrodynamics and predict iron and manganese concentrations (Hammond et al., 2023).

The most common algal bloom monitoring in reservoir and lakes uses real-time sensors deployed on buoys, measuring meteorological parameters, standard water quality parameters with multiparameter sondes as well as chlorophyll *a* and phycocyanin fluorescence (Painter et al., 2023). Real-time nutrient analysers have currently been used in oceans (Mowlem et al., 2021), glacial runoff (Beaton et al., 2017), rivers (Beaton et al., 2012; Bowes et al., 2015b; Halliday et al., 2014; Lloyd et al., 2015; Outram et al., 2014; Rode et al., 2016; Wade et al., 2012) and (drinking) water treatment plants (Rode et al., 2016). They are often deployed in rivers to measure nutrient patterns at a high frequency to investigate diurnal biogeochemical cycling, the response to storms and to detect sources of pollution (Bowes et al., 2015b; Chappell et al., 2017; Halliday et al., 2014; Scholefield et al., 2005). Real-time nutrient analysers have not yet been successfully utilized for monitoring algal blooms in lakes and reservoirs, but a trial in several lakes in the USA highlights the potential opportunities and challenges (Johnston et al., 2024). Given that nutrient dynamics are a likely driver of geosmin and 2-MIB production, measuring nutrients in the reservoir in real-time could help predict geosmin and 2-MIB events.

It is important to design a sensor monitoring network by considering the most efficient resource application. Depending on the sensors or analyser, their smallest

measurement frequency is between 1 minute and 30 minutes. Sensors and analysers have running costs and maintenance costs, particularly for lab-on-chip nutrient analysers which use reagents and produce waste, but even optical and electrochemical systems have duty cycles that when exceeded result in decreased sensitivity or general performance. It needs to be carefully considered what temporal frequency of measurement is necessary to capture data variability without creating excess information. Several studies have shown that nutrient loading calculations are impacted by a limited sampling frequency (Cassidy & Jordan, 2011; Johnes, 2007; Lloyd et al., 2014) and it was also found to impact Water Framework Directive (WFD) classifications (Skeffington et al., 2015). A process called 'aliasing' can happen where low sampling frequencies impact the data variability so much that the data can get wrongly interpreted (Chappell et al., 2017). Moreover, typical manual sampling strategies are also expected to cause bias due to fair weather bias (Rand et al., 2022) and sampling during working hours and typical working days (Halliday et al., 2015; Skeffington et al., 2015).

1.11 Research gaps

The environmental drivers of geosmin and 2-MIB production remain uncertain, partly due to the limited frequency of manual sampling, which fails to capture the full complexity of the physical, chemical and biological factors influencing water conditions. Additionally, these drivers are likely influenced by the species traits of the T&O producing cyanobacteria species. Furthermore, geosmin and 2-MIB may be produced by benthic or planktic cyanobacteria species, which will result in different environmental drivers. Nutrients have been suggested as a key factor in various laboratory culturing and field studies, especially considering $NH_4^+:NO_3^-$ and DIN:SRP ratios. The response to NO_3^--N, NH_4^+-N and PO_4^{3-}-P by geosmin and 2-MIB producing cyanobacteria has been tested in culture studies. However, these studies are generally unrealistic because there is no competition or interaction with other phytoplankton, bacteria or zooplankton. Field studies, which include these factors, often have limited temporal frequency and hence the lag-time between an environmental driver, the growth of cyanobacteria and geosmin and 2-MIB concentrations, cannot be determined.

To predict T&O in real-time, several commercially available sensors could be used to measure parameters that influence T&O, for example water column conditions,

nutrient dynamics, phytoplankton productivity and biomass of certain phytoplankton groups within the community. Currently high-frequency, *in situ* nutrient analysers have mainly been applied to understand biogeochemical cycles in rivers and are rarely used in lakes or reservoirs. Sensors are being used to predict cyanobacterial blooms in drinking water treatment plants and subsequent T&O events (for example, dissolved oxygen in Chen et al. (2019), phycocyanin fluorescence in Zamyadi et al. (2016) and ambient environmental conditions combined with chlorophyll *a* fluorescence in Painter et al. (2023)), but these factors seem highly site-specific, and knowledge is still limited. Questions remain how to design a monitoring network effectively which uses resources efficiently. Sampling frequency of sensor measurements needs to accurately represent biogeochemical processes, without using additional resources.

It was therefore hypothesized that:

1) Differing sampling frequencies of water quality sensors can impact interpretations of biogeochemical processes.

To explore this hypothesis, Chapter 2 investigates high-frequency hydrochemical data, including nutrients (NO_3^--N, SRP and TP), from six UK river catchments to determine the impact of measurement time and frequency on the interpretation of biogeochemical processes.

2) Nutrient concentrations and ratios are important factors in predicting geosmin and 2-MIB events.

In Chapter 3, a laboratory microcosm study was conducted with a natural phytoplankton community from a Welsh drinking water reservoir to investigate the effect of nutrient ratios (NH_4^+:NO_3^- and DIN:SRP) and concentrations on the phytoplankton productivity, community structure and subsequent T&O compound production.

3) Sensor data can provide an early warning for geosmin and 2-MIB events.

Chapter 4 investigates how 2-MIB and geosmin events in a Welsh drinking water reservoir could be predicted. A comprehensive data analysis was performed using long-term data from manual sampling and laboratory analysis, as well as high-frequency *in situ* sensor data from a deployment in 2022-2023 that measured NO_3^--N, NH_4^+-N and several standard water quality parameters.

Chapter 2: Interpretation of river water quality data is strongly controlled by measurement time and frequency

2.1 Introduction

Water quality monitoring programmes must strike a balance between resource efficiency (cost) and representation of changes in water conditions required to fulfil the monitoring purpose. Traditional water quality sampling relies on periodic sample collection and subsequent laboratory analysis. However, such manual sampling regimes cannot capture all events, and indeed biases in data can be caused by changing day and time of the week (Johnes, 2007; Skeffington et al., 2015), weather conditions (Rand et al., 2022) and extreme high or low flow conditions (Lloyd et al., 2015). Rand et al. (2022) compared manual and automated sensor data from the Belgrade Lakes, USA, where they found that manual lake sampling showed a significant likelihood to take place during "fair weather", with lower windspeeds and rainfall intensity and higher air temperature than the mean. Infrequent manual sampling of water chemistry, which most likely occurs during standard working hours at regular intervals (weekly, monthly etc.), can bias the calculation of annual average concentration, annual nutrient load and environmental quality standards (Cassidy & Jordan, 2011; Halliday et al., 2015; Johnes, 2007; Jordan et al., 2007; Skeffington et al., 2015). Extreme high or low flow conditions are important for nutrient transport; they can contribute to most of the total nutrient load in rivers with a flashy hydrology (Cassidy & Jordan, 2011). These conditions are often short-lived, only occur infrequently (Johnes, 2007; Lloyd et al., 2014) and will not be captured fully by infrequent manual sampling. High flows can promote transport of sediment-bound nutrient fractions from land to water or via in-channel remobilisation, whilst low flow conditions can be dominated by nutrient inputs from sewage effluent in urbanised rivers, due to low dilution capacity (Halliday et al., 2015), as well as nutrient delivery along throughflow pathways including from waterlogged soils when there is drizzle (Collins et al., 2010; Durand et al., 2011; Evans & Johnes, 2004; Lloyd et al., 2014; Yates & Johnes, 2013). Thus, sampling regimes that capture such conditions are critical to reflect nutrient transport processes and estimate nutrient loads accurately.

Advances in *in situ* sensing technologies have the potential to reduce bias associated with sampling periodicity. Continuous or high temporal resolution hydrochemical sampling therefore can enable an enhanced understanding of catchment processes (Bieroza et al., 2023; Blaen et al., 2017; Bowes et al., 2015b; Kirchner et al., 2004; Lloyd et al., 2015; Rode et al., 2016). This is especially relevant for transient events and short-term biogeochemical dynamics, including diurnal or other cyclic patterns that are closely linked to hydrological and biological processes (Khalil & Ouarda, 2009) such as pollutant load estimates (Johnes, 2007) and response to storm events (Chappell et al., 2017; Jordan et al., 2007), as they are based on representative measured concentrations and the discharge rate. In the UK, increased interest in high-resolution water quality monitoring is partly driven by the recent implementation (April 2023) of Section 82 of the Environment Act 2021, which requires water companies to deploy continuous water quality monitoring up and down stream of all sewage effluent discharges to a water course (DEFRA, 2023; Hanson, 2023). Simultaneously, drinking water production is moving towards smart catchment monitoring and management with high-resolution sensor technologies in source waters; for example for anoxia (Wentzky et al., 2019), iron and manganese concentrations (Hammond et al., 2023) and algal bloom related issues (Carey et al., 2021; Painter et al., 2023; Zamyadi et al., 2016).

An important consideration in monitoring, however, is that more data are not always better (Coraggio et al., 2022). The optimal sampling regime must balance the minimum frequency needed to capture fluctuations, particularly in flashy streams, and the maximum frequency that can be collected sustainably (considering power demands and data costs) without returning redundant information and increasing potential noise in the data that masks the information required (Coraggio et al., 2022; Khalil & Ouarda, 2009). The objectives of the monitoring network, for example meeting certain environmental quality standards, detecting sources of pollution or measuring a change before or after a mitigation, will determine the required data analysis, which in turn sets requirements for the temporal resolution of the data. Determining the temporal frequency of measurement is not a static process. Measurement intervals can be optimised over time or in response to external stressors (Coraggio et al., 2022), for example adaptive monitoring (Blaen et al., 2016) aims to optimise the intervals in real-time when a threshold is met, like an extreme event.

This chapter will investigate the first overarching hypothesis: Differing sampling frequencies of sensors can impact interpretations of biogeochemical processes. A systematic assessment was conducted using high resolution hydrochemical sensor data from six different UK catchments. The following research questions were addressed:

1) What is the lowest measurement frequency that can fully capture data variation in different water quality parameters?
2) How does sampling at specific times of the day impact the interpretation of biogeochemical cycles?
3) What is the effect of measurement frequency on the interpretation of biogeochemical cycles?

2.2 Materials and methods

2.2.1 Catchment characteristics

High-frequency water quality data were collected at least every hour using *in situ* sensors, in six different UK rivers (Figure 2.1): the Wylye (Hampshire Avon catchment), Enborne (Kennet catchment), Blackwater drain (Wensum catchment), Thames (Thames catchment), Hiraethlyn (Conwy catchment) and Newby Beck (Eden catchment).
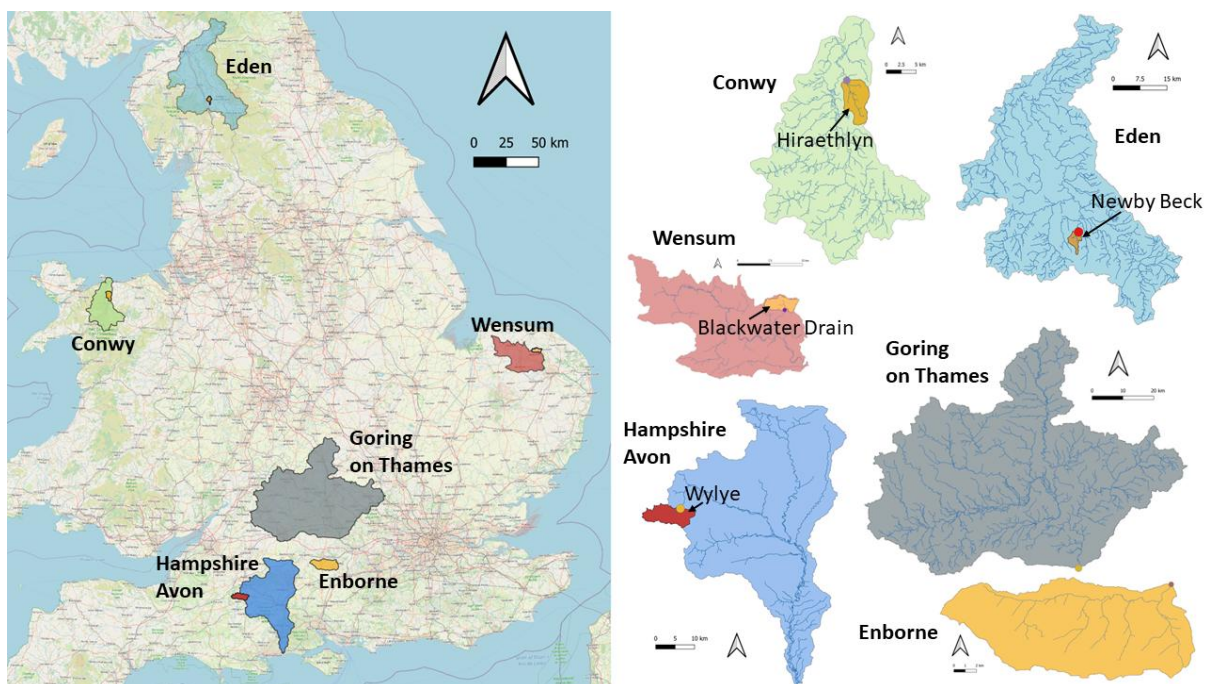
Figure 2.1. Catchments in the UK that were used for this study.

The monitoring stations in the Hampshire Avon (Lloyd et al., 2015; Lloyd et al., 2019; Outram et al., 2014), Wensum (Cooper et al., 2018; Outram et al., 2014) and Eden (Outram et al., 2014; Owen et al., 2012; Perks et al., 2015) catchments were part of the DEFRA funded Demonstration Test Catchments (DTC). The Enborne monitoring station was part of the LIMPIDS programme and UKCEH Thames Initiative (Bowes et al., 2018; Bowes et al., 2015a; Halliday et al., 2014; Wade et al., 2012) and the Conwy catchment was monitored as part of the DOMAINE programme (supplied by Chris Yates and Penny Johnes, University of Bristol, Bristol, UK; underpinning data set as referenced by Mackay et al. (2020) and Yates et al. (2023)). The Thames monitoring station at Goring-on-Thames was part of UKCEH Thames Initiative monitoring (unpublished data, supplied by Mike Bowes, UK Centre for Ecology & Hydrology, Wallingford, UK, and the UK Environment Agency; referenced in Rode et al. (2016) and Moorhouse et al. (2018)). The studied catchments cover a wide range of catchment characteristics related to geology and climate, like the base flow index (BFI) and mean flow (Table 2.1). Moreover, they vary significantly from 13 to 4634 km^2 in area, and there is a marked difference in land use (Table 2.1). Further details about the catchments can be found in the papers referenced in Table 2.1.

Table 2.1. Catchment characteristics of the six UK catchments studied and the exact period of high-frequency monitoring.

River	Hiraethlyn	Enborne	Wylye	Thames	Blackwater Drain	Newby Beck
Catchment	Conwy	Kennet	Hampshire Avon	Thames	Wensum	Eden
Monitored location	Bodnant	Brimpton	Brixton Deverill	Goring-on-Thames	Kiosk F - Park Farm	Newby
Latitude	53.2260	51.3803	51.1600	51.5235	52.7771	54.5853
Longitude	-3.7990	-1.1838	-2.1901	-1.1435	1.1491	-2.6202
Catchment size (km^2)	20.5 (g)	148.0	50.2	4633.7 (h)	19.7	12.5
Elevation of sampling point (m a.s.l.)	11 (f)	62	189 (a)	30	43 (a)	233 (i)
Annual average rainfall (mm)	1200 (f)	810 (d)	967 (d)	680 (d)	655 (a)	1167 (a)
Baseflow Index (BFI)	0.46 (f)	0.54 (d)	0.93 (a)	0.64 (d)	0.66 (c)	0.39 (a)
Mean flow (m^3/s)	0.54 (l)	1.06 (l)	0.47 (l)	23.0 (l)	0.094 (c)	0.33 (l)
Dominant land use	Improved grassland (g)	Arable and grassland	Livestock and cereals	Arable/horticulture, improved grassland (h)	Intensive arable cultivation	Livestock (dairy and meat)
% Urban	0.3 (g)	6.5	7.0	7.3	1.0	2.0 (d)
Land use distribution						
Monitoring start date	19/06/2015	01/11/2009	13/03/2012	29/12/2013	08/03/2011	14/09/2011
Monitoring end date	30/09/2017	29/02/2012	05/03/2014	13/10/2015	31/12/2014	01/01/2016

a: Robson and Reed (1999); b: https://www.landis.org.uk/soilscapes/ (accessed: 24/06/2021); c: Cooper et al. (2018); d: Marsh and Hannaford (2008); e: https://en-gb.topographic-map.com/maps/iu/United-Kingdom/ (accessed: 24/06/2021); f: Estimate based on Yates et al. (2023) and Marsh and Hannaford (2008); g: Yates et al. (2019a); h: Gauging station Thames at Reading https://nrfa.ceh.ac.uk/ (accessed: 24/06/2021); i: Outram et al. (2014); j : Lloyd et al. (2019); k : Bowes et al. (2015b); l: Calculated from dataset. Legend for land use distribution pie-charts:

■ Arable (%)   ■ Improved pasture (%)   ■ Rough grazing (%)   ■ Woodland (%)   ■ Urban (%)

The list of monitored parameters varied slightly per site, but all included temperature, water level or discharge, pH, electrical conductivity (EC), dissolved oxygen (DO), turbidity, chlorophyll *a* (Chl-a), nitrate (as N) and total reactive phosphorus (TRP). At some sites, total phosphorus (TP) and ammonium (as N) were also measured. Full details of all equipment and sampling regimes are described below for each sampling campaign.

2.2.2 Site descriptions and equipment
DOMAINE: Hiraethlyn at Bodnant
The Hiraethlyn is a tributary of the River Conwy in North-Wales, in the relatively more urbanised part of the catchment. The DOMAINE project established this monitoring station at Bodnant in June 2015 and collected data until September 2017. More information on this site can be found in Table 2.1 and Yates et al. (2019a). Water level (pressure transducer in a stilling well), velocity (acoustic Doppler flow meter) and the stream profile was used to estimate average discharge (m^3/s). Velocity measurements were not reliable enough to create discharge data for the whole dataset. Other sensors were; an OTT HydroMet Hydrolab MS5/DS5 (OTT HydroMet, Kempten, Germany) for standard water quality parameters, a Sea-Bird SUNA v2 for nitrate (NO_3^--N) and a Sea-Bird ECO-FL for CDOM & fDOM and chlorophyll *a* (Sea-Bird Scientific, Bellevue, Washington, USA) (Table 2.2). To validate the sensor measurements, spot samples were taken daily or weekly depending on the time period, which were analysed in the Bristol University laboratory, following standard procedures (Yates et al., 2019a). DOMAINE data are available on the Environmental Information Data Centre (UK Centre for Ecology and Hydrology): https://eidc.ac.uk/.

LIMPIDS: Enborne at Brimpton
The River Enborne is a tributary of the River Kennet. There are a few sewage treatment works (STW's) in the catchment serving some small towns and villages within the catchment. The LIMPIDS program installed the monitoring equipment at Brimpton in November 2009, and they measured until February 2012. Detailed information on the catchment and monitoring set-up can be found in Table 2.1 and Table 2.2 and Wade et al. (2012). Discharge was monitored by the EA flow monitoring station at Brimpton. Sensors installed were a YSI 6600 v2 multi-parameter probe (YSI Inc., Yellow Springs, Ohio, USA) for standard parameters, a Hach Nitratax Plus for nitrate (as N) (Hach, Loveland, Colorado, USA) and a Systea Micromac C (Systea

S.p.A, Anagni, Italy) to measure total reactive phosphorus (TRP) (Table 2.2). The monitoring station was housed within an insulated shed on the riverbank, to protect the equipment from extreme weather and potential theft. The Nitratax was installed in the river, but the YSI 6600 and Systea Micromac C were located in flow cells in the shed that were intermittently filled every hour with river water via a peristaltic pump. There was no filtration step in the analyses of TRP. In addition, manual samples were taken each week and analysed in the UKCEH laboratories using standard analytical methods described in Bowes et al. (2018); Wade et al. (2012). LIMPIDS data are available on the Environmental Information Data Centre (UK Centre for Ecology and Hydrology): https://eidc.ac.uk/.

Demonstration test catchments (DTC): Wylye at Brixton Deverill, Blackwater Drain at Park Farm and Newby Beck at Newby

Each DTC site had a monitoring network, with an overarching aim but run by local scientists. The Wylye at Brixton Deverill is a tributary of the Hampshire Avon (Lloyd et al., 2015; Outram et al., 2014). This monitoring station was maintained by the University of Bristol. The Blackwater Drain at Park Farm (Kiosk F) is a small agricultural stream in the Wensum catchment (Cooper et al., 2018; Outram et al., 2014). The monitoring station there was maintained by the University of East Anglia and the Wensum Alliance. Newby Beck at Newby (close to Morland) is a tributary of the River Eden (Outram et al., 2014; Owen et al., 2012). This monitoring station was maintained by Durham University. At the Wylye, discharge was monitored by EA flow monitoring stations locally. At the Blackwater Drain and Newby Beck, discharge was estimated based on water level (pressure transducer in stilling well), velocity (acoustic Doppler flow meter) and the stream profile. In the case of Blackwater Drain, the discharge data were not available for analysis. Rain gauges were installed at Blackwater Drain and Newby Beck. At all locations, sensors installed were a YSI 6600 v2 multi-parameter probe (YSI Inc., Yellow Springs, Ohio, USA) for standard parameters, a Hach Nitratax Plus for nitrate (NO_3^--N) and a Hach Sigmatax SC followed by a Hach Phosphax Sigma to measure Total Reactive Phosphorus (TRP) and Total Phosphorus (TP) after a digestion step (Hach, Loveland, Colorado, USA) (Table 2.2). The sensor data has been validated with routine spot sampling, this was done daily in the Wylye, weekly in the Blackwater Drain and monthly in Newby Beck. Samples were analysed in three different local laboratories at the related universities,

using standard methods. An inter-laboratory comparison in 2011-2012 showed that the analytical procedures were consistent in the three catchments (Outram et al., 2014). DTC data and contact details for data requests are visible on the Agricultural and Environmental Data Archive: http://www.environmentdata.org/.

UKCEH: Thames at Goring

The River Thames flows through an urbanised part of the UK. The sensors were deployed between December 2013 and October 2015. The instrumentation was housed in a mains-powered, insulated building by the riverbank upstream of Cleve Lock, near the town of Goring-on-Thames. At hourly intervals, water was pumped from the river into two flow-through tanks within the building, to reduce the impact of biofouling in the river. An EXO2 sonde (YSI Inc., Yellow Springs, Ohio, USA) within one of the flow-through tanks had probes to measure water temperature, dissolved oxygen, conductivity, turbidity, pH, chlorophyll *a* and ammonium. The other larger (10 litre) tank contained a Hach Nitratax probe for measuring nitrate concentrations. Water was sub-sampled from the tank for total phosphorus and total reactive phosphorus using a Hach Phosphax / Sigmatax spectrophotometric auto-analyser (Hach, Loveland, Colorado, USA) (Table 2.2). The YSI EXO2 was replaced with a new, fully calibrated EXO2 each month, and the Phosphax self-calibrated every 2 days. In addition, manual samples were taken each week and analysed in the UKCEH laboratories using standard analytical methods described in Bowes et al. (2018), alongside Aquacheck quality control standards (LGC Standards, Teddington, UK), to groundtruth the automatic water quality data. Discharge was recorded by the EA gauging station at Caversham, Reading, and downloaded via the National River Flow Archive (https://nrfa.ceh.ac.uk/).

Table 2.2. Detailed information on sensor equipment used at each monitoring station and three shades of green indicate the frequency (freq. in table) of each measurement, with dark green = every 1 hour, medium green = every 30 minutes and light green = every 15 minutes.

River	Hiraethlyn		Enborne		Wylye		Thames		Blackwater Drain		Newby Beck	
Detailed information	Personal communication Penny Johnes and Chris Yates		Bowes et al. (2015b); Wade et al. (2012)		Lloyd et al. (2016); Outram et al. (2014)		Personal communication Mike Bowes and Dan Read		Cooper et al. (2018); Outram et al. (2014)		Cooper et al. (2018); Owen et al. (2012)	
Parameters measured with sensors	freq.	Sensor information	freq.	Sensor information	freq.	Sensor information	freq.	Sensor information	freq.	Sensor information	freq.	Sensor information
Discharge (m³/s)		x	[dark green]	EA flow monitoring station River Enborne at Brimpton	[light green]	EA flow monitoring station River Wylye at Brixton Deverill	[light green]	EA flow monitoring River Thames at Caversham, Reading		x	[light green]	Acoustic Doppler flow meters (Argonaut-SW, Sontek) to estimate discharge using velocity, stage and stream profile
Level/stage (m)	[light green]	Pressure transducer in stilling well (FL9000)		x	[light green]	Thistle 24R Incremental Shaft Encoder with a float and counterweight in stilling well		x	[dark green]	Pressure transducers in stilling well		x
Rainfall (mm)		x		x		x		x	[dark green]	Rain gauge installed at site	[light green]	Rain gauge installed at site

Parameter						
Dissolved oxygen (mg/L)	Hydrolab MS5	YSI 6600 v2	YSI 6600 v2	YSI EXO2	YSI 6600 v2	YSI 6600 v2
Turbidity (NTU)	Hydrolab DS5	YSI 6600 v2	YSI 6600 v2	YSI EXO2	YSI 6600 v2	YSI 6600 v2
Chlorophyll *a* (µg/L)	Sea-Bird ECO-FL	YSI 6600 v2	YSI 6600 v2	YSI EXO2	YSI 6600 v2	YSI 6600 v2
pH	Hydrolab MS5	YSI 6600 v2	YSI 6600 v2	YSI EXO2	YSI 6600 v2	YSI 6600 v2
Electrical conductivity (µS/cm)	Hydrolab MS5	YSI 6600 v2	YSI 6600 v2	YSI EXO2	YSI 6600 v2	YSI 6600 v2
Water temperature (°C)	Hydrolab MS5	YSI 6600 v2	YSI 6600 v2	YSI EXO2	YSI 6600 v2	YSI 6600 v2
Ammonium as N (mg/L)	x	x	YSI 6600 v2	YSI EXO2	x	x
Nitrate as N (mg/L)	Sea-Bird SUNA v2	Hach Nitratax Plus	Hach Nitratax Plus SC	Hach Nitratax Plus SC	Hach Nitratax Plus SC	Hach Nitratax 1000 SC
Total phosphorus (mg/L)	x	x	Hach Sigmatax SC followed by Phosphax Sigma	Hach Sigmatax SC followed by Phosphax Sigma	Hach Sigmatax SC followed by Phosphax Sigma	Hach Sigmatax SC followed by Phosphax Sigma
Total reactive phosphorus (mg/L)	x	Systea Micromac C	Hach Sigmatax SC followed by Phosphax Sigma	Hach Sigmatax SC followed by Phosphax Sigma	Hach Sigmatax SC followed by Phosphax Sigma	Hach Sigmatax SC followed by Phosphax Sigma
CDOM	Sea-Bird ECO-CDOM	x	x	x	x	x

2.2.3 Data conversions

For mean flow (m^3/s), average values were calculated from datasets which included flow measurements. For the Hiraethlyn, mean flow (m^3/s) was calculated using data on level (m) and velocity (m/s), which was combined with an estimated river width of 2 m and assumption of an approximately rectangular riverbed. The Blackwater Drain had annual discharge data (Cooper et al., 2018) for several years. To calculate DO (mg/L) from % saturation for the River Enborne, daily average barometric air pressure data was used from Thatcham, a town close to the monitoring station, as well as water temperature from the Enborne dataset. The Rstudio package 'rMR' (Moulton, 2018) was used to convert the DO units.

2.2.4 Analysis

Data manipulation – artificial decimation

At each site, the sensors logged data at time intervals ranging from 15 minutes to one hour (Table 2.2). The high-resolution datasets were sub-sampled at predefined intervals to create a subset of smaller datasets. This artificial decimation (Johnes, 2007) process was executed in two different ways, to test a) the influence of reduced sampling frequency on median and range, and b) the influence of intra-daily variation. Methods are described below:

Temporal frequency effects (a)

Some data in this study were collected every 15 minutes, but for consistency the most frequent data available in all catchments was used for this comparison, which was *hourly* data. Artificial decimation was used to create one version of an *hourly* (every day at every whole hour), *four-hourly* (every day at 00:00, 04:00, 08:00, 12:00, 16:00 and 20:00), *twelve-hourly* (every day at 00:00 and 12:00), *daily* (every day at 12:00), *weekly* (every Wednesday at 12:00), and *monthly* dataset (every second week of the month, on Wednesday at 12:00). Artificially created datasets with *four-hourly, twelve-hourly, daily, weekly,* and *monthly* data were compared to the *hourly* data, to assess the influence of a reduced frequency on the percentage of the total *hourly* range captured in the data set and the percentage change in the median.

Percentage of the total range captured was calculated for each parameter accordingly (Equation 2.1):

Equation 2.1: $\dfrac{MAX(x) - MIN(x)}{MAX(hourly) - MIN(hourly)} * 100$

where x is the artificially created datasets e.g. four-hourly, twelve-hourly, daily, weekly and monthly data. Parameter behaviour is determined by the median, 25% and 75% interval and data distribution, which can be visualised by the width of a violin boxplot (the width of the boxplot depends on the number of datapoints at each value).

Percentage change in the median was calculated for each parameter accordingly (Equation 2.2):

Equation 2.2: $\dfrac{Median\ (x) - Median\ (hourly)}{Median\ (hourly)} * 100$

where x is the artificially created datasets e.g. four-hourly, twelve-hourly, daily, weekly and monthly data.

Intra-daily variation (b)

Artificial decimation was repeated for multiple initial conditions to create different versions of a daily dataset (Halliday et al., 2015; Johnes, 2007); Daily with different times of the day: *every day at 00:00, 04:00, 08:00, 12:00, 16:00, 20:00*, resulting in six different *daily* datasets.

To determine intra-daily variation, for each of these timeframes a new dataset was created which included the median for each day. The difference between the median and the corresponding datapoints in the six artificially decimated daily datasets was calculated and compiled in one dataset (Figure 2.2). For example, the intra-daily variation data consisted of a calculated difference for each of the six times of day (00:00, 04:00, 08:00, 12:00, 16:00, 20:00) for every day in the multi-year dataset. The outcome was tested for significant differences using Kruskall-Wallis analysis of variance and Dunn's post-hoc test (Rstudio version 2023.06.2+561, R version 4.2.1 (2022-06-23 ucrt)). Each dataset was then banded by significance, with data that showed no significant differences grouped together (denoted by the same colour). Variation for the multi-year datasets was plotted as boxplots (with significant outliers removed to enable better visualisation on the y-axis).
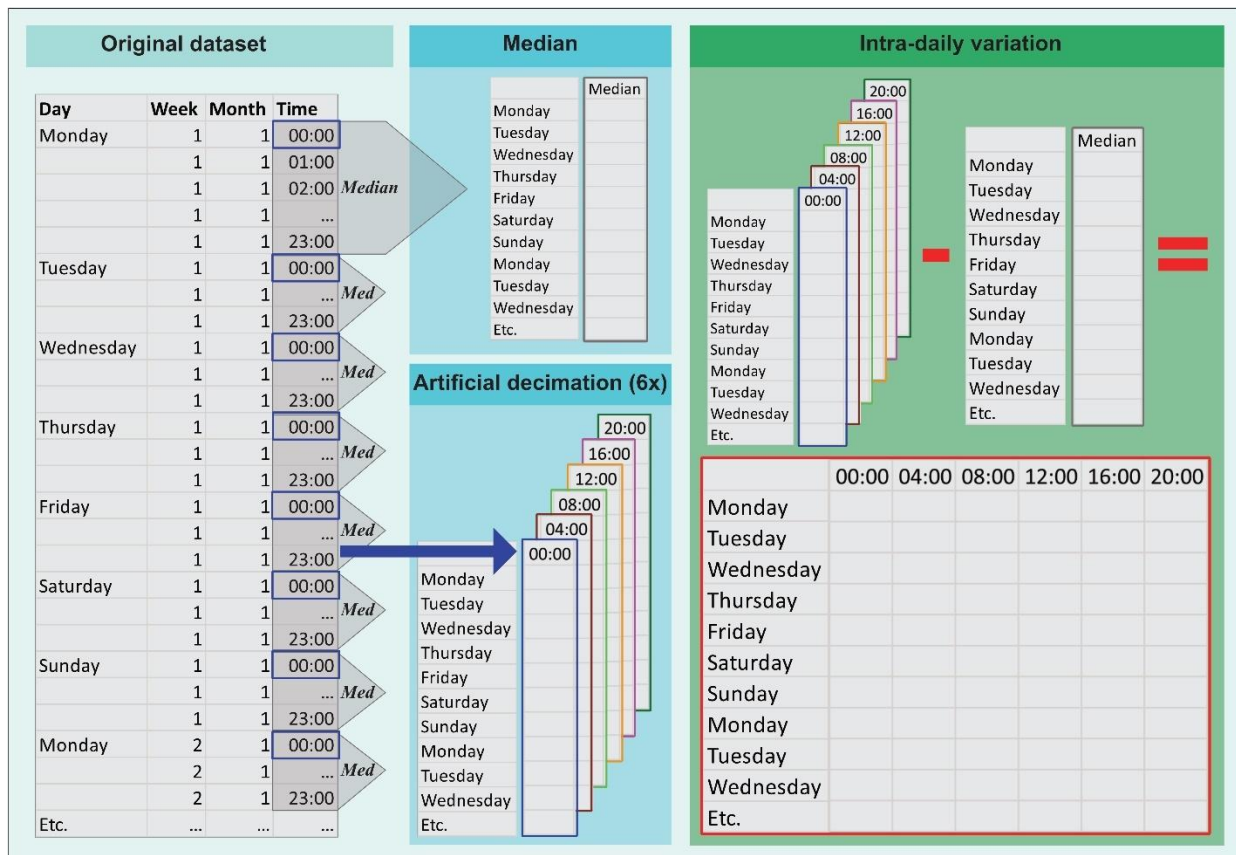
Figure 2.2. Artificial decimation process to calculate intra-daily variation. Daily median and six new daily datasets from six selected times of day were created to calculate intra-daily variation for the whole dataset.

2.3 Results

2.3.1 Seasonality

Variation in all parameters recorded in the full datasets from each site prior to artificial decimation (Figure 2.3) indicated a considerable temporal and spatial difference in range, median as well as 25% and 75% interval. The seasonal effect depended on the catchment and varied by parameter (Figure 2.3). Median nitrate, total phosphorus and total reactive phosphorus concentrations calculated per month (in multi-year datasets) highlight important biogeochemical processes and dominant transport mechanisms that occur throughout the year, which are catchment dependent (Figure 2.4).
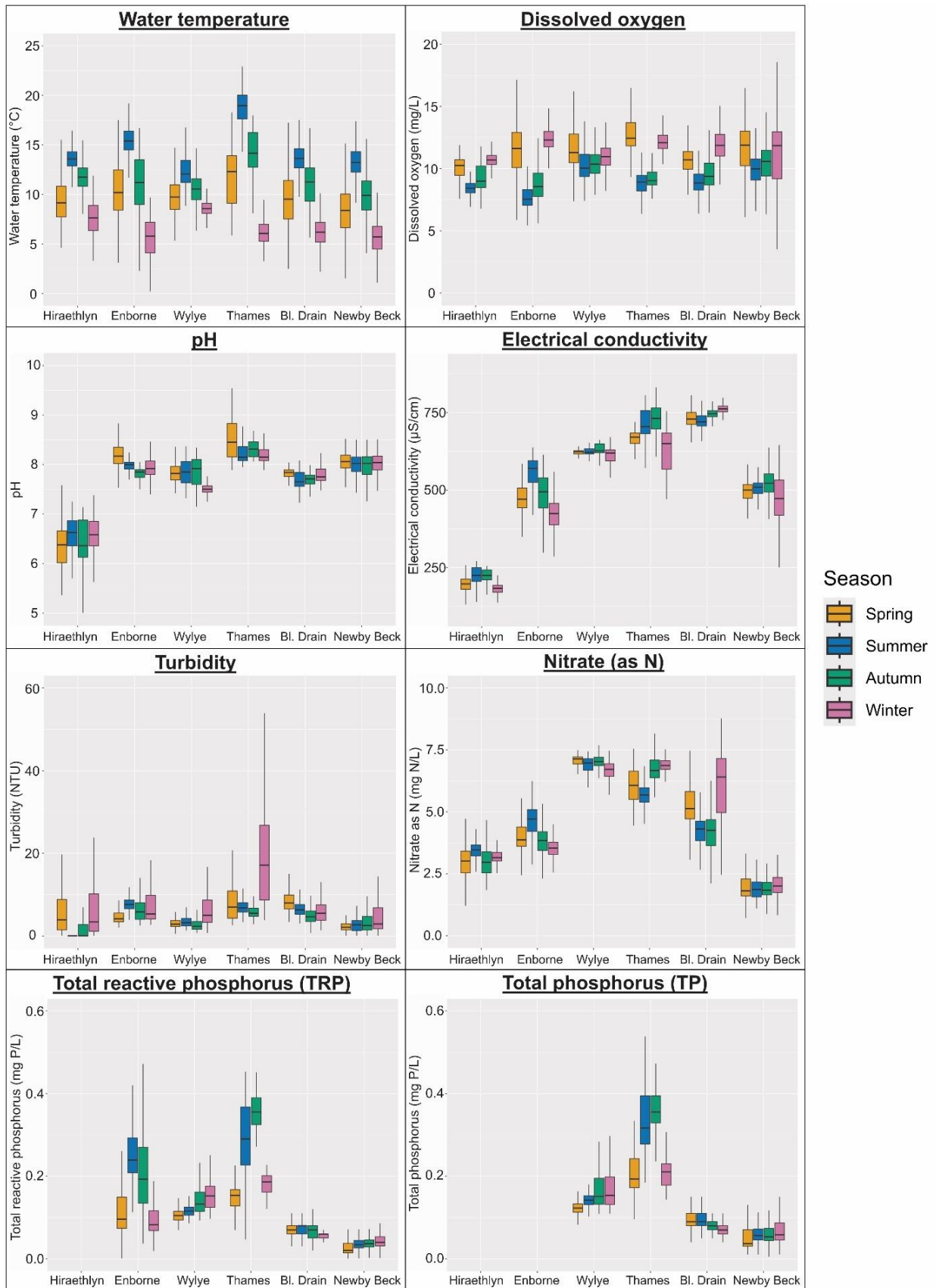
Figure 2.3. Boxplots (without outliers) for water quality data from the six study catchments. Bl. Drain = Blackwater Drain. Seasons are defined as follows; spring: March, April, May; summer: June, July, Augustus; autumn: September, October, November; winter: December, January, February.
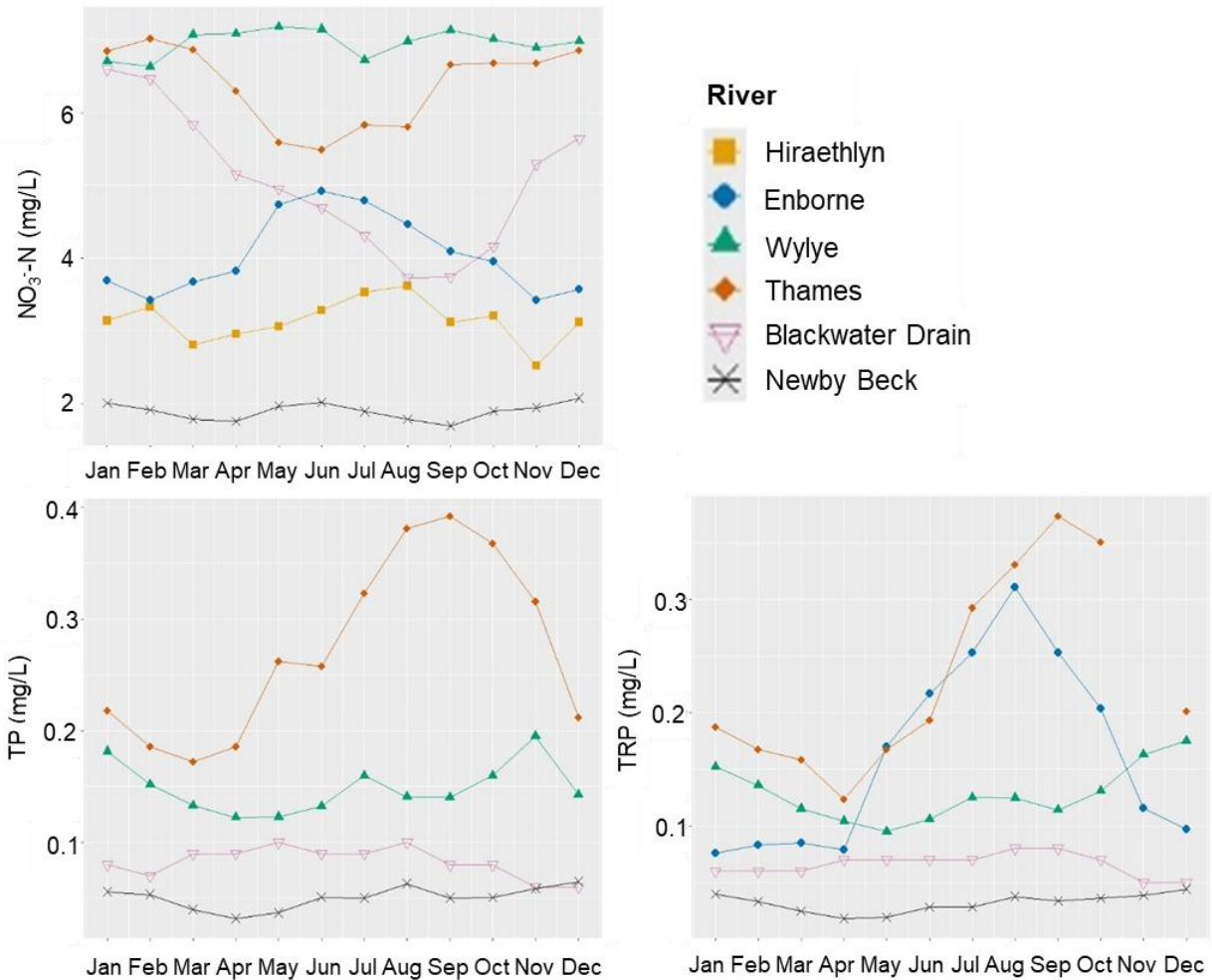
Figure 2.4. Median monthly concentrations of nitrate (as N), total phosphorus (TP) and total reactive phosphorus (TRP) in the six study catchments.

2.3.2 Temporal frequency effects

Reducing the temporal frequency had a different impact on the captured range (Table 2.3), median (Table 2.3) and data distribution (histogram; the width of the violin boxplot visualises the number of datapoints at that value, Figure 2.5), depending on the parameter and catchment. Reduced frequency showed the largest percentage change in median for turbidity, dissolved oxygen, temperature, TP and TRP, and had the largest overall impact on total range of turbidity captured (Table 2.3), but there are many nuances dependent on the catchment. Monthly frequency impacted dissolved oxygen concentrations in the Wylye and turbidity in Blackwater Drain, changing the median by >13% whilst capturing 53% and 8% of total range, respectively (Table 2.3). In general, reducing frequency had the least impact on median and range for nitrate and electrical conductivity, followed by dissolved oxygen, temperature, and pH, although this was largely catchment dependent. Reducing to monthly frequency had

a relatively small impact on nitrate concentration observations in Newby Beck and electrical conductivity in the Hiraethlyn, where the median changed by <2% whilst 88% and 90% of the total range was recorded, respectively (Table 2.3).

The percentage of the total range captured and percentage change in the median were not always similarly affected by a reduction in frequency. For example, daily data for dissolved oxygen in the Wylye almost captured the total range of the hourly data variation (99%) but had a large impact (>10% change) on the calculated median (Table 2.3). The opposite pattern, with a large impact on total range captured and relatively small impact on median, was also present in some catchments (Table 2.3), for example in weekly observations of EC in Blackwater Drain (28% of total range captured, 0% change in median).

In the six catchments, the change in median for turbidity was most consistent, decreasing (negative) at monthly compared to hourly data, but this was not the case for every temporal frequency studied (Table 2.3). There was no consistent direction (increase or decrease) of change in the median with reduced temporal frequency for any of the studied catchments (Table 2.3).

Table 2.3: Percentage of total range captured and percent median change, comparing reduced frequencies to hourly data. Reduced temporal frequency datasets were artificially created at: four-hourly, twelve-hourly, daily, weekly and monthly frequency. Figure 2.5 violin boxplots visualise this data and the data distribution. Colours in the percentage of total range table are added to clarify the trend, with a continuous green-yellow-red scale to indicate 100-50-0 percent of total range captured by the reduced frequency datasets.

River	Parameter	% Total range captured					Median % change				
		4 Hourly (%)	12 Hourly (%)	Daily (%)	Weekly (%)	Monthly (%)	4 Hourly (%)	12 Hourly (%)	Daily (%)	Weekly (%)	Monthly (%)
Hiraethlyn	Temperature	100	88	82	74	62	0.00	-0.18	-0.18	-0.26	-0.62
	DO (mg/L)	80	73	64	44	38	0.11	0.43	2.46	1.98	4.98
	pH	94	94	94	92	54	-0.15	0.00	0.31	0.62	-0.77
	EC	98	98	96	95	90	0.00	0.00	0.00	-0.73	-0.97
	Turbidity	100	19	19	3	1	33.33	100.00	200.00	0.00	-33.33
	NO$_3^-$-N	82	80	79	59	36	0.00	0.31	-0.62	-2.15	0.31
	TRP	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	TP	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Enborne	Temperature	97	93	93	87	74	0.00	0.97	0.97	-4.85	-7.28
	DO (mg/L)	99	81	80	74	65	0.06	2.02	5.03	9.74	10.60
	pH	100	88	74	65	47	0.00	0.13	0.25	0.13	0.13
	EC	83	83	75	52	48	0.00	0.00	0.63	0.95	0.95
	Turbidity	83	48	48	8	6	0.85	1.69	-1.69	-6.78	-0.85
	NO$_3^-$-N	93	90	79	67	54	0.00	0.18	0.18	0.55	-0.36
	TRP	97	85	80	80	53	0.61	0.00	-5.32	1.23	-9.20
	TP	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Wylye	Temperature	100	81	80	69	64	0.00	1.28	7.95	7.70	5.99
	DO (mg/L)	99	99	99	88	53	0.00	1.49	10.55	10.18	14.38
	pH	99	96	95	83	73	0.00	0.13	1.29	1.23	1.94
	EC	98	85	85	58	11	0.00	0.00	-0.32	-0.16	0.32
	Turbidity	100	52	51	33	3	0.00	-3.13	-9.38	-6.25	-12.50
	NO$_3^-$-N	95	95	95	55	17	-0.06	-0.05	0.12	-0.10	1.43
	TRP	93	84	52	48	11	0.00	0.00	0.00	0.00	-3.92
	TP	98	60	42	38	9	0.21	0.00	0.00	-0.22	-7.29
Thames	Temperature	99	97	97	95	78	0.07	0.07	0.59	0.63	3.97
	DO (mg/L)	97	96	75	74	68	0.00	0.27	0.27	2.71	-2.58

	pH	100	67	67	65	59	0.00	0.00	0.12	-0.12	-0.49
	EC	100	96	94	94	84	0.00	-0.04	-0.03	-0.34	-0.94
	Turbidity	68	68	68	12	12	-0.57	0.85	1.70	6.75	-12.00
	NO$_3^-$-N	99	99	99	39	31	0.00	-0.11	-0.11	-0.79	-0.07
	TRP	87	85	82	82	71	0.00	1.46	1.46	1.46	-6.58
	TP	71	69	66	59	48	0.00	0.00	1.88	1.61	-0.27
Blackwater Drain	Temperature	98	89	88	80	57	-0.19	0.84	6.00	2.44	3.61
	DO (mg/L)	98	95	91	80	73	0.00	0.60	2.05	3.50	6.70
	pH	98	77	77	55	41	0.00	0.00	0.26	0.26	0.26
	EC	100	67	66	28	19	0.00	0.00	0.00	0.00	-0.13
	Turbidity	100	89	89	15	8	0.00	-1.64	-9.84	-11.48	-13.11
	NO$_3^-$-N	99	98	98	90	39	0.00	0.21	0.64	0.64	4.06
	TRP	92	89	87	29	18	0.00	0.00	0.00	0.00	0.00
	TP	96	57	57	17	11	0.00	0.00	0.00	0.00	-12.50
Newby Beck	Temperature	99	89	87	85	71	-0.11	-0.21	0.96	0.53	0.85
	DO (mg/L)	59	59	59	43	23	0.09	0.83	4.86	4.95	6.42
	pH	97	92	92	87	73	0.00	0.12	0.87	1.00	1.06
	EC	98	96	73	72	51	0.00	0.20	0.39	0.39	-0.99
	Turbidity	100	100	100	36	11	0.00	0.00	-4.17	-6.25	-12.50
	NO$_3^-$-N	98	97	97	91	88	-0.11	0.11	0.11	0.44	-1.31
	TRP	100	82	56	44	43	0.00	0.90	1.74	-0.41	1.74
	TP	100	100	99	67	42	0.00	2.86	4.57	0.56	4.01
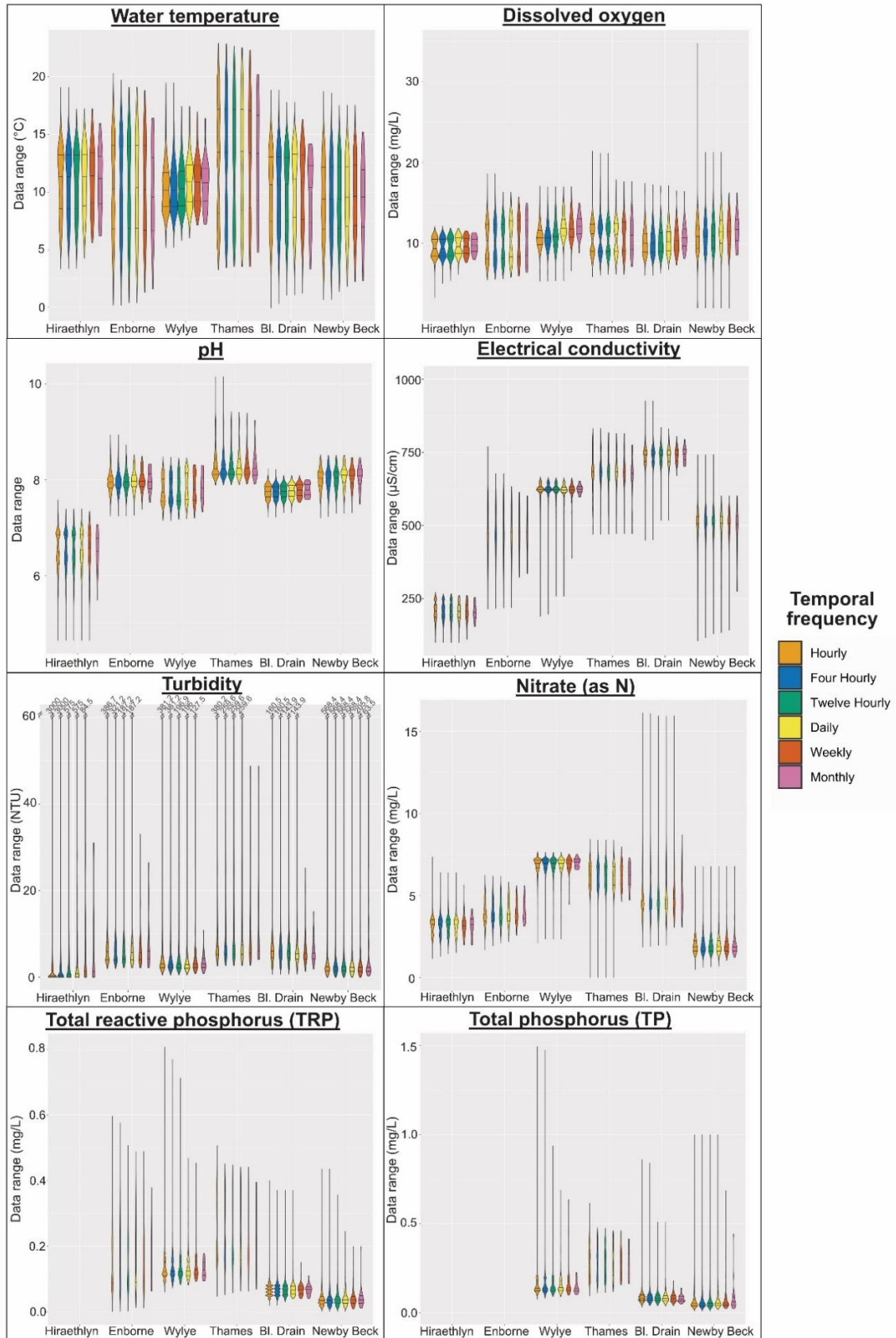
Figure 2.5. Violin boxplots showing the median, 25% and 75% interval as well as the data structure, for the six UK catchments water quality parameters, with the different artificially created datasets with lower temporal data frequency: four-hourly, twelve-hourly, daily, weekly and monthly. Bl. Drain is an abbreviation of Blackwater Drain.

Four-hourly data captured most of the parameter behaviour (Figure 2.5), as well as the percentage of total range captured and percentage change in median compared to hourly data (Table 2.3). From all 45 individual parameter-catchment combinations (six to eight parameters in six catchments), four-hourly data captured most of the hourly range (>90%) for 37 combinations, and 41 had limited impact on the median (<0.5% change). The Wylye and Blackwater Drain four-hourly datasets captured 92%-100% of the total range for all parameters in the hourly data, which is a higher overall range captured than the other four catchments at four-hourly frequency (Table 2.3). The Newby Beck four-hourly dataset captured 95-100% of the total range apart from for DO, where only 59% was recorded (Table 2.3). The other catchments captured >90% for most parameters at four-hourly measurement frequency, except for DO and nitrate at the Hiraethlyn; turbidity and EC at the Enborne; and turbidity, TP and TRP at the Thames (Table 2.3). Twelve-hourly and daily data represented >90% of the range with limited impact on the median (<0.5% change) in approximately half of the combinations. Daily measurements captured >90% of total range for certain parameters; nitrate (4 of 6 catchments), pH (3 of 6 catchments), EC (2 of 6 catchments) and DO (2 of 6 catchments). Most parameters at weekly frequency did not cover >90% of total range, except for the pH and EC at the Hiraethlyn; EC and temperature at the Thames; nitrate at the Blackwater Drain and nitrate at Newby Beck (Table 2.3), which all had <1% change in median. Monthly data frequency resulted in generally low percentages of range captured for all catchments, with some exceptions (Table 2.3). Monthly data from the Hiraethlyn revealed the lowest percentage of range captured; 1% of the hourly range in turbidity, but also the highest percentage of range captured; 90% of the hourly range in EC (Table 2.3).

2.4 Intra-daily variation

Most parameters and catchments displayed significant differences in variation between the six different times of day (denoted by differing colour bands in Figure 2.6).
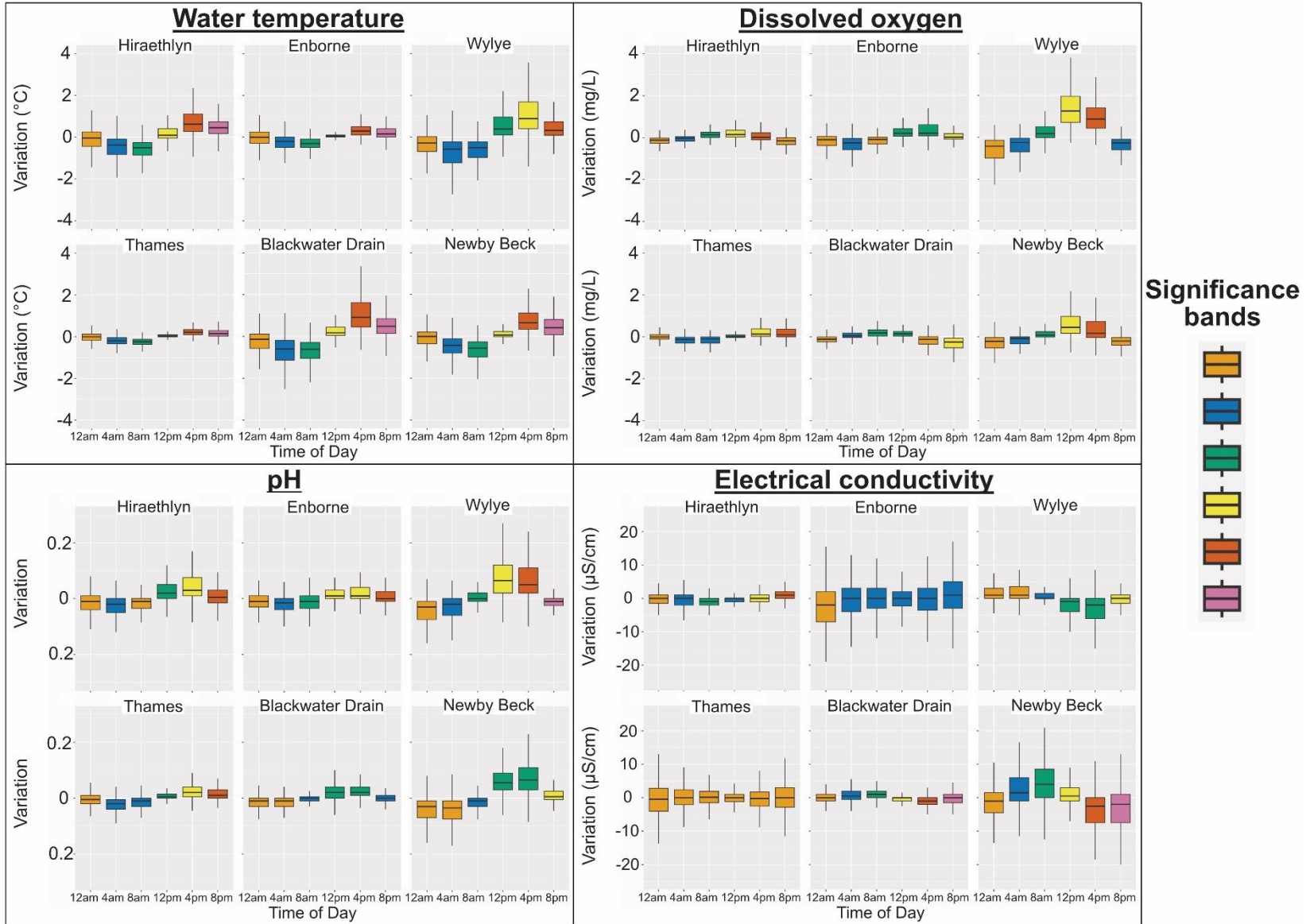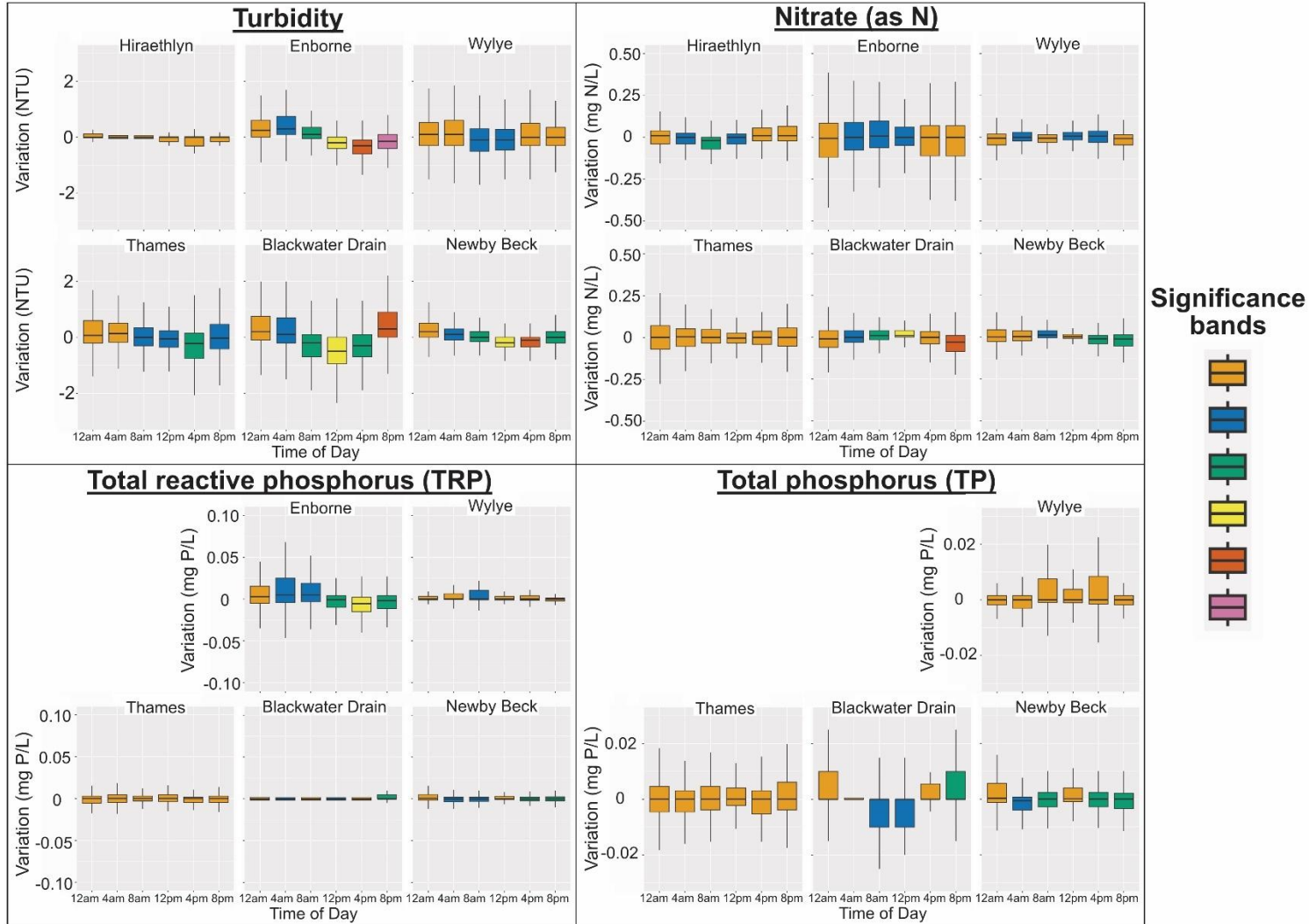
Figure 2.6. This and previous page: Intra-daily variation for all study catchments based on six versions of a daily dataset. Datapoints were selected from different times of day; 00:00, 04:00, 08:00, 12:00, 16:00, 20:00. Significance bands bar colours indicate for each individual plot (each catchment within each parameter) the significance between the six different times of day from the Kruskall-Wallis analysis of variance and Dunn's post-hoc test; bars with the same colour are not significantly different from each other, whilst different colours denote significant difference.

2.4.1 Physico-chemical parameters

The intra-daily variation in water temperature can be used to interrogate the patterns of significance shown, as this parameter has a predictable cyclic pattern throughout the day, with cooler temperatures at night and warming throughout daylight hours. This physical process persists throughout different seasons and is expected to reveal a strongly significant intra-daily variation pattern for this multi-year analysis. The variation is calculated as the parameter value at one of the six selected times of day minus the parameter median of the whole day, collated for each day in the dataset. The outcome plotted for the six selected times of day allows a comparison of variation within a day (intra-daily). Throughout the dataset there are cooler temperatures at night-time, which result in a more negative variation value (for all days in the dataset, the value at that time is lower than the daily median), and warmer temperatures at daytime which cause a more positive variation (higher values than the daily median) (Figure 2.6). The variation for the water temperature was significantly different for every time of day in almost all catchments, which means the described pattern was consistent throughout the whole dataset and all seasons (denoted by differing colours in Figure 2.6). Relative to the median temperature each day, 04:00 or 08:00 was the coldest and 16:00 was the warmest in every catchment. The Thames had the smallest range in variation, followed by the Enborne.

A cyclical day-night pattern for DO and pH was also visible in all catchments, albeit more pronounced in some, such as the River Wylye and Newby Beck (Figure 2.6). In most of the catchments there was a strong connection between DO and pH, where they both followed the same day-night trend. However, in the Blackwater Drain and Hiraethlyn, DO had maximum positive variation four hours earlier than pH. Electrical conductivity (EC) revealed a significant diurnal trend in most catchments, apart from the Enborne and Thames. The variation in EC followed the opposite trend of pH and DO in the Wylye and Newby Beck, DO in the Hiraethlyn and pH in the Blackwater Drain. Intra-daily variation for turbidity in the Enborne, Thames, Blackwater Drain and Newby Beck showed significantly more positive variation at night and significantly more negative variation during the day (Figure 2.6), whilst the Hiraethlyn and Wylye did not show any trends.

2.4.2 Total reactive phosphorus, total phosphorus and nitrate

Intra-daily variation in nutrients revealed less clear significant patterns than the physico-chemical parameters, and these patterns were catchment dependent (Figure 2.6). Nitrate had significant intra-daily variation in most catchments, apart from the Thames, with the clearest diurnal cycle (most significant differences between the timesteps) in the Blackwater Drain. There was a general trend towards more positive variation (higher values compared to the median for each day) from early morning until mid-day and more negative variation (lower values compared to the daily median) from late afternoon until midnight (Figure 2.6), except for the Hiraethlyn in which this pattern seemed to be reversed. Total reactive phosphorus (TRP) and total phosphorus (TP) showed significant intra-daily variation in some catchments, but there was often no clear diurnal trend. The Enborne showed the clearest diurnal cycle in TRP with most positive variation in early morning and most negative variation in the afternoon. Newby Beck and the Blackwater Drain had similar patterns for TRP and TP and revealed a general tendency for more negative variation in the morning. TRP in the Enborne and TP in the Blackwater Drain followed similar intra-daily variation patterns to turbidity (Figure 2.6).

Differences in intra-daily variation depending on the season will not be visible in Figure 2.6, as the datasets consisted of multiple whole years which would even out any intra-daily variation pattern that only existed seasonally. Examples for the Enborne and Newby Beck are presented here to show intra-daily variations by season (Figure 2.7), while all other results are visualised in Figure 2.8. Nitrate and TRP concentrations for the Enborne and Newby Beck, with intra-daily variation separated by season (Figure 2.7), illustrate the influence of season on intra-daily variation patterns in nutrients. Nitrate concentrations in the Enborne didn't show an impact of season on intra-daily variation, but Newby Beck had a much clearer diurnal cycle in spring, summer and autumn compared to winter (Figure 2.7). TRP concentrations in the Enborne had a clear diurnal cycle in spring, summer, and autumn but not in winter, whereas Newby Beck had only minor diurnal fluctuations in summer (Figure 2.7). TP in summer in the Wylye had a strong pattern but no other season and TRP in the Thames only had a strong pattern in winter (Figure 2.8).
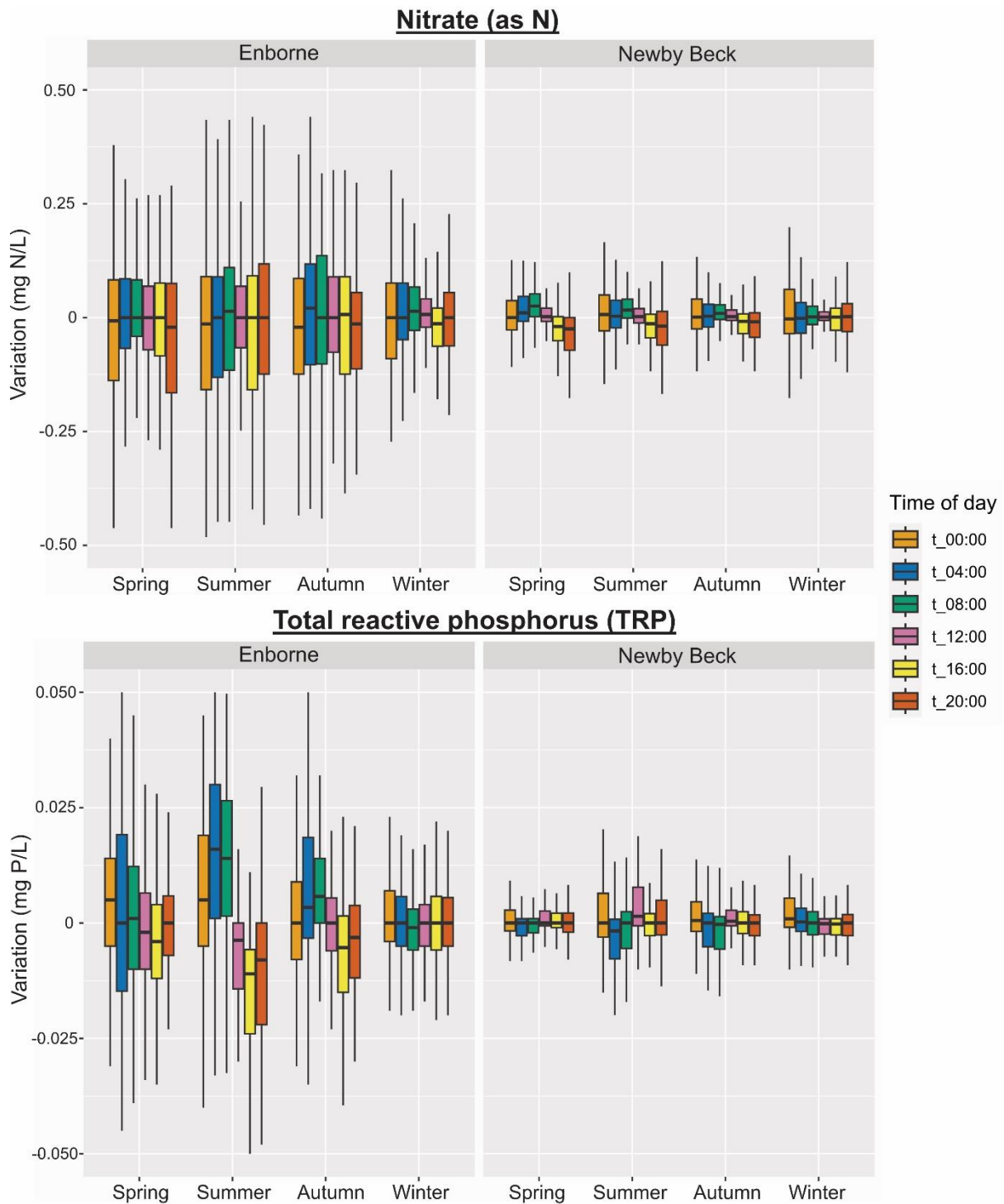
Figure 2.7. Nitrate (as N) and total reactive phosphorus (TRP) intra-daily variation separated by season for the Enborne and Newby Beck. Datapoints were selected from different times of the day; 00:00, 04:00, 08:00, 12:00, 16:00, 20:00, which are indicated by different colours.
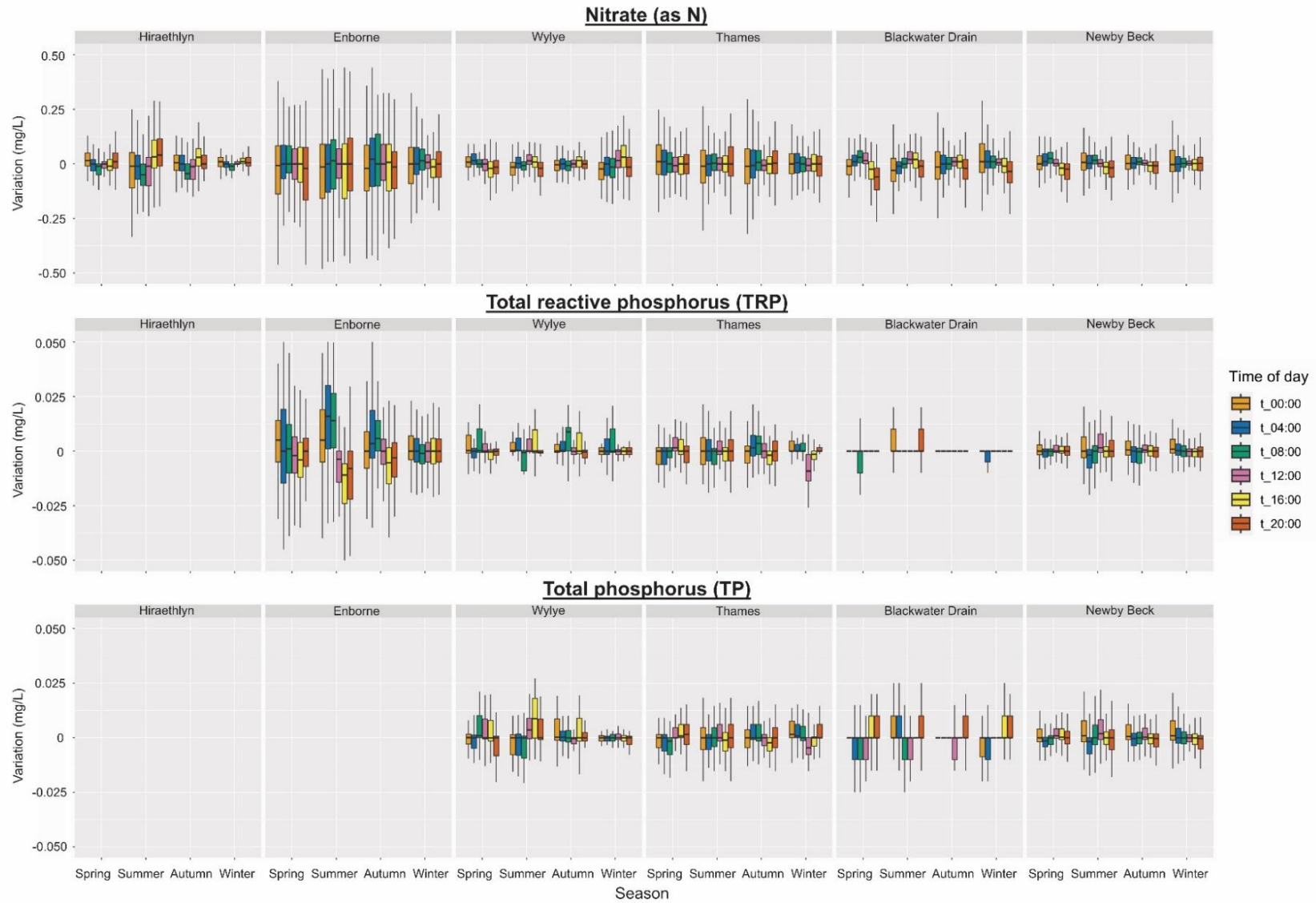
Figure 2.8. Nitrate (as N), total reactive phosphorus (TRP) and total phosphorus (TP) intra-daily variation at six selected times of the day, separated by season. Note that TP was not measured at Hiraethlyn and the Enborne, and neither TRP at Hiraethlyn.

2.5 Discussion

Catchment characteristics such as size, land use (urban and agriculture) and dominant flow paths (groundwater, throughflow or overland flow) which are primarily controlled by catchment geology, are a first order control of variation in these datasets. Previous assessments have demonstrated that monthly sampling cannot capture the full variation of physical and biogeochemical parameters, and even that monitoring at less than daily frequency can alter nutrient load assessments (Wade et al., 2012). Infrequent sampling and random sampling effects may result in the same water body being misclassified under legislation such as the Water Framework Directive (Halliday et al., 2015; Skeffington et al., 2015), with multiple classes possible depending on sampling frequency for the determinand of interest. However, each hydrochemical parameter displayed variable patterns in different catchments, seasons and times of the day, so exploring high-resolution data and signposting when, where and what frequency observation is necessary is critical for optimising sampling regimes.

2.5.1 Reduced temporal frequency effects

Reducing temporal frequency creates the risk that the data will not capture the "real" median and range, a phenomenon termed 'aliasing' (Chappell et al., 2017). Reducing measurement frequency from hourly to four-hourly, twelve-hourly, daily, weekly, and monthly in this study increasingly changed the interpretation of the data by altering data distribution, median and range, with catchment- and parameter-specific effects. In general, turbidity, dissolved oxygen, temperature, TP and TRP showed the largest percentage change in median with reduced frequency of observation, whilst the greatest overall impact on total range was for turbidity, although this effect was catchment dependent. These parameters, where reduced frequency has the largest impact, are expected to have a large data variability due to rapid rainfall response (turbidity which is controlled by sediment mobilisation and transport, and overland flow-generated phosphorus transfers such as for TP) or strong diurnal cycles (temperature and dissolved oxygen).

Reduced temporal frequency did not always affect the captured range and the median simultaneously, since the range could be impacted without any changes in the median and vice versa. Data variability for each parameter in every catchment can be influenced by some or all of; time of day (diurnal cycle), season (seasonal cycle) and extreme weather (rainfall-response and flow pathway activation and separation)

(Figure 2.9). Parameters which are less strongly controlled by the latter, and in particular with overland flow or near-surface throughflow pathways, such as nitrate and electrical conductivity in some study catchments, can potentially be measured at lower temporal frequencies without compromising the median and range, but this depends on the monitoring purpose and the catchment flow activation regime.
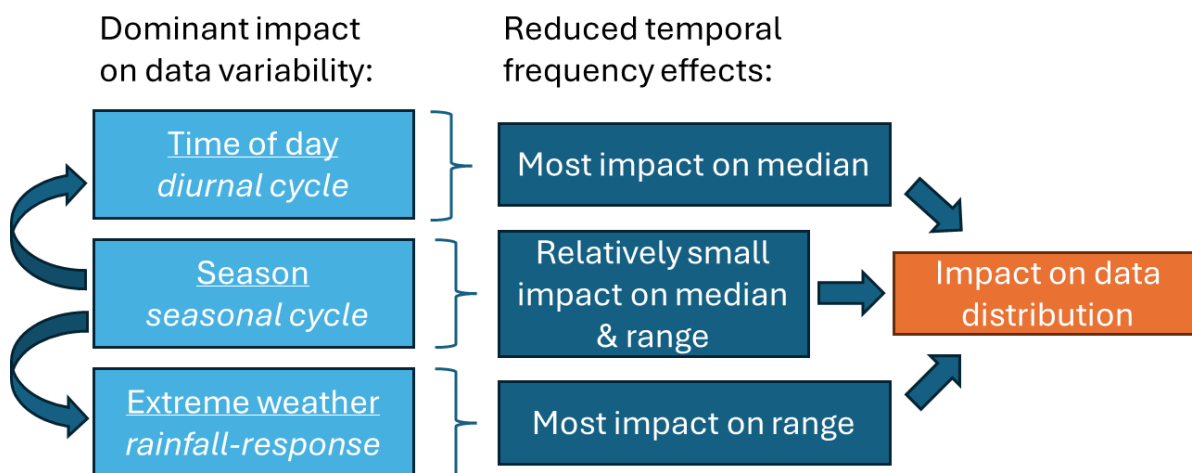
Figure 2.9. Processes that can impact data variability and their effect on median, range and data distribution.

Variability caused by time of day has the largest impact on the median, as diurnal cycles cause intra-daily variation in some parameters, which won't be fully represented in a data set with reduced measurement frequency (Figure 2.9). Variability caused by seasonality alone will have relatively little effect on median and range at reduced frequency. However, in nutrient load calculations by Williams et al. (2015), the summer season was more biased and less precise for nitrate (as N) and dissolved reactive phosphorus (DRP). Moreover, seasonality can influence the diurnal cycle, illustrated by the nitrate and total reactive phosphorus concentration presented in this study (Figure 2.7 and Figure 2.8). Variability caused by extreme weather responses will have the largest impact on range, because reduced frequency will not fully capture high concentration flux responses to short-term extreme events (Figure 2.9), unless the sample happens to accidentally capture the peak of such an event, which can then positively bias annual load estimates (Johnes, 2007; Jordan et al., 2007). Variability caused by all three factors will have an impact on data distribution (histogram), by not capturing the full width of the data variation.

Reducing the measurement frequency not only impacts the range and variability of the data, but also the distribution (Figure 2.5), as demonstrated by Cassidy and Jordan (2011) for TP, Johnes (2007) for total dissolved phosphorus (TDP) and TP and Lloyd et al. (2014) for these same fractions plus for nitrate. This result indicates that monthly or weekly sampling fails to capture important extreme events and potentially underestimates (or overestimates) the median and subsequent annual load calculations (Johnes, 2007). While the median did change in this study, the data showed no consistent under- or over-estimation. This can be partially attributed to the nature of the analysis, as sub-sampling was done at only one selected time of day (daily), day of the week (weekly) and week of the month (monthly), based on common manual sampling regimes. These conditions, however, would have had an impact on the direction of percentage change in the median, as the time of day would have skewed the results for daily, weekly and monthly frequencies, especially for parameters with strong diurnal cycles like dissolved oxygen (Rand et al., 2022).

2.5.2 Optimal frequency

From all 45 analysed parameter-catchment combinations (six to eight parameters in six catchments), four-hourly data captured most of the hourly range (>90%) for 37 combinations, and 41 out of 45 had limited impact on the median (<0.5% change). Twelve-hourly and daily data captured >90% of the range in 17 and 15 combinations respectively, with limited impact on the median in 30 and 19 combinations, respectively. Weekly data captured >90% of the hourly range in 6 combinations and 16 had limited impact on the median. Monthly data didn't capture >90% of the hourly range in any combination, whilst 10 had limited impact on the median. The individual parameters that were most affected by reducing frequency depended on the catchment.

Mathematical methods can define an optimum sampling frequency for any water quality parameter by calculating the point at which an increase in frequency does not provide an increase in information. Coraggio et al. (2022), for example, used high-frequency monitoring data from Bristol Harbour and mathematically determined the optimum sampling frequency for water temperature, electrical conductivity, dissolved oxygen, and turbidity as 9 hours, 6 hours, 5 hours, and 3 hours, respectively. Parameters with a rapid response to extreme events, such as turbidity and total or particulate phosphorus fractions, need to be monitored at a higher frequency to

capture full data variability. Parameters with a diurnal cycle, like pH, dissolved oxygen and electrical conductivity need to be monitored frequently enough to capture these cycles or could be monitored at an appropriate, but standardised time on each day to calculate an average, depending on the monitoring purpose. This study found that four-hourly frequency captured most of hourly data variability, depending on parameter-catchment responses, but the optimal frequency needs to take into account the monitoring purpose. Changing the time of day at which observations are captured, within any monitoring programme could bias the resulting data sets.

To determine the optimal monitoring frequency for a parameter, which captures sufficient data without using excess resources, the following factors need to be considered; (I) Parameter & catchment and (II) Monitoring purpose.

Parameter & catchment (I)

Parameter and catchment interaction determined the effect of reduced temporal frequency on the range, median and data distribution. No parameter in this study was found to behave consistently for the six different catchments, hence parameter behaviour was largely dependent on catchment specific characteristics that define its response to biogeochemical cycling processes and hydrological regime (Figure 2.9).

As observed in previous work on P fractions alone (Johnes, 2007; Jordan et al., 2007) catchment characteristics such as the contribution of groundwater to river flow (base flow index), land use (urban and agriculture) and size have a strong impact on water quality data variability (Table 2.1). Catchment size can strongly influence the data distribution, with biogeochemical changes damped or subject to lag times (Creed et al., 2015). Year-round high flows in the Thames (Table 2.1) were found to mask local biogeochemical effects, which is possibly a result of the large catchment size and subsequently large river flow volume (Williams et al., 2000). Diurnal biogeochemical patterns in rivers are often stronger during stable, non-turbid, low flow conditions as riverine biological processes are more prominent (Bowes et al., 2016; Scholefield et al., 2005).

Catchments with a high base flow index (BFI) have notable groundwater contributions which influence temperature and nutrient concentrations. This is illustrated in the Wylye, where groundwater nitrate inputs vary inversely with overland flow inputs (Outram et al., 2014; Yates & Johnes, 2013). Nutrient concentrations are also strongly

influenced by agriculture and urban land use (Salvia-Castellví et al., 2005). Intensive livestock farming and urban wastewater discharges cause a similar biogeochemical reaction as their effluents are both rich in ammonium (Donald et al., 2011). Rivers with a more urbanised catchment will receive a larger proportion of wastewater discharges, from sewage treatment works (STW) or septic tanks, especially during low flow conditions (Macintosh et al., 2011; Yates et al., 2019b). STW discharges are often related to increased turbidity, EC, temperature, ammonium and phosphorus, whilst stimulating microbial activity. Microbial processes include nitrification (nitrate production) and the decomposition of organic material, which can subsequently reduce dissolved oxygen (Halliday et al., 2015) and alter the composition of the nutrient pool instream (Yates et al., 2019b).

This study shows that all catchments had a clear intra-daily water temperature pattern, coldest in the early morning and warmest late afternoon. Dissolved oxygen and pH also showed intra-daily variation in every catchment, positive in the afternoon and negative in the early morning as a result of photosynthesis-respiration cycles. Driven by diurnal water temperature and solar energy cycles, daytime photosynthesis removes (acidic) carbon fractions and produces oxygen, whilst night-time respiration does the opposite (House, 2003; Scholefield et al., 2005). The amplitude of this biological diurnal cycling depends on the temperature, light availability, and the relative contribution of autotrophic and heterotrophic organisms (Nimick et al., 2011). More abundant submergent plant communities in certain catchments, particularly chalk streams like the Wylye (Evans & Johnes, 2004; Lloyd et al., 2019; Yates & Johnes, 2013), would explain its more prominent diurnal cycle for DO and pH. Electrical conductivity had intra-daily variation, negative in the afternoon and positive in the early morning, in most catchments apart from urbanised rivers Enborne and Thames, which is most likely due to uptake and release (or lack of uptake) of free ions with diurnal biological activity. Intra-daily variation for turbidity, negative (lower values than the daily median) in the afternoon and positive (higher values than the daily median) in the early morning, occurred in most catchments apart from the Wylye and Hiraethlyn, which might be a result of night-time bioturbation: sediment resuspension caused by the feeding and movement of fish and invertebrates like crayfish (Cooper et al., 2020; Cooper et al., 2016; Halliday et al., 2015). These natural biogeochemical patterns can be masked by, for example, the volume of flow, shading from bankside growth, a large

groundwater influx with lower temperatures or a large influx of non-natural water such as sewage outflows.

Nitrate (as N), total reactive phosphorus and total phosphorus can also follow diurnal cycles as a response to nutrient uptake by biological activity in the river, which results in a typical diurnal cycle of lowest concentrations in the late afternoon and highest in the early morning (Cooper et al., 2020; Nimick et al., 2011; Palmer-Felgate et al., 2008; Scholefield et al., 2005). However, in most rivers, this is not the dominant process all year round, because of minimal biological activity in the winter months and the alteration of natural cycles by anthropogenic influences (agriculture or wastewater discharges) (Jordan et al., 2007; Nimick et al., 2011; Pellerin et al., 2009). In urbanised catchments, electrical conductivity, turbidity, nitrate (as N) and phosphorus fractions (TRP, TP) can also exhibit diurnal cycles because of consistent daily patterns in wastewater effluent discharges to these rivers (Halliday et al., 2014; Palmer-Felgate et al., 2008; Withers & Jarvie, 2008). High-frequency data from the River Cut, of which 36%-90% of flow consists of STW effluent, revealed a double-peak in the daily EC signal, during midday and late evening, a delayed response to peak domestic water usage in the morning and evening (Palmer-Felgate et al., 2008; Withers & Jarvie, 2008). Such effect will become less evident in larger rivers with greater dilution capacity and when there is increased flow caused by rainfall in autumn/winter months. The same parameters can exhibit diurnal signals in agricultural catchments because of consistent daily discharges from dairy farm operations (milking) (Foy & Kirk, 1995), which might also have a delayed response.

Diurnal cycles can also be influenced by seasons, so although seasonal cycles themselves will most likely be captured with a reduced temporal monitoring frequency (monthly), it is critical to understand the influence of seasonal signals on daily, and sub-daily (for example, extreme weather) events. Moreover, certain agricultural processes only happen during spring and summer months, which could lead to a different data variation pattern throughout the year. In certain catchments, episodic short-lived extreme events can play a major role in biogeochemical processes, and it is important to fully capture their data variability.

Monitoring purpose (II)

Optimal temporal frequency depends on the purpose of monitoring; long-term trend analysis, load calculations and storm-induced solute transport modelling require different inputs and therefore have unique data frequency demands (Coraggio et al., 2022). Sub-sampling high-frequency data to pre-determined lower frequencies (artificial decimation process) can be done iteratively to contain multiple initial conditions, and determine the optimal monitoring frequency for specific purposes (Chappell et al., 2017; Coraggio et al., 2022; Crockford et al., 2017; Johnes, 2007; Reynolds et al., 2016; Skeffington et al., 2015; Williams et al., 2015). Previous analyses have suggested that seasonal variation or long-term trends can be captured with monthly or up to half-yearly frequency (Coraggio et al., 2022). For basic statistical calculations, for example to assign Water Framework Directive classifications, for phosphorus fractions, dissolved oxygen, pH and temperature (Skeffington et al., 2015), or to detect trends in nitrate data such as mean concentration, peak concentration, drinking water standard exceedance and flux (Reynolds et al., 2016), weekly or daily sampling is recommended. In annual load estimates (Bowes et al., 2009; Crockford et al., 2017; Johnes, 2007; Williams et al., 2015) daily sampling gives the more robust and reliable results but weekly is also acceptable provided the uncertainties associated with load estimates are also reported (Lloyd et al., 2014). This largely depends on the nutrient fraction, the season and catchment characteristics as those influence reaction time and variability. Williams et al. (2015) found optimal frequency for dissolved reactive phosphorus (DRP) was every 13-26 hours and nitrate (as N) every 2.7-17.5 days. When modelling biogeochemical response during storm events (Chappell et al., 2017; Lloyd et al., 2015; Outram et al., 2014), a higher measurement frequency is required to capture this accurately, with Chappell et al. (2017) arguing for sampling rates of less than 120 minutes to greater than 600 minutes. However, these studies, and this data demonstrate that minimum temporal frequency can change over time, and between catchments and parameters, with a higher frequency needed when there is more variation and depending on the variable of interest and its environmental behaviour in each catchment.

2.5.3 Sensor uncertainty implications for monitoring design

The data variability captured by any monitoring campaign is subject to the limitations of the equipment used for measurement. Where data fluctuations are within the

uncertainty bounds of a technique, or when measurements are subject to bias, limiting the availability of data points by reducing measurement frequency can be problematic. It is therefore critical that uncertainty bounds are known to ensure relevant fluctuations can be captured. The uncertainty of the sensor measurements used in this study is well-quantified, by comparison between laboratory samples for TP and nitrate (as N) and the sensor data at Brixton Deverill on the Wylye (Lloyd et al., 2015). The headline uncertainty bounds of ±0.15 mg/L for TP and ±0.75 mg/L for nitrate (as N) calculated by Lloyd et al. (2015) suggest that the intra-daily variation patterns that we have identified could fall within the range of uncertainty. The maximum daily variation without outliers that we identified in the six studied catchments is ±0.05 mg/L for TP and ±0.5 mg/L for nitrate (as N). However, where the sensor data display daily variation, the uncertainty bounds of the data also vary according to the antecedent conditions, so the signal is unlikely to fluctuate between the highest and lowest bounds at adjacent time points. This temporal autocorrelation effect means that the variations revealed in our data are likely to be a real signal, even if they fall within the overall sensor uncertainty. It is therefore imperative that data users have a strong understanding of the measurement capabilities of the chosen device.

2.5.4 Recommendations

Reliance on weekly or monthly data means the likelihood of capturing total data variability (range and median) is small for most catchments. A balance is therefore required to determine the most cost-effective yet representative sampling regimes for different catchments. High-frequency sensor data cannot be captured everywhere, so instrumentation should be selected and deployed for the target chemistries of interest. It is also important to note that sensors cannot currently measure all parameters of interest, so optimal sampling programmes are likely to combine both high resolution sensor networks with manual or automated sample collection paired with laboratory analyses where tighter quality assurance and quality control can reduce uncertainties, albeit at a lower temporal sampling resolution. If manual sample collection cannot physically be done more than weekly or monthly it is critical to calculate, report, and minimise sampling bias, while reporting data with resultant uncertainty bands is essential. Jordan and Cassidy (2022) created an overview with important considerations to select a fit-for-purpose monitoring strategy, for example stakeholder engagement and evidence for policy or land-use management changes. Moreover,

Rozemeijer et al. (2025) proposes a decision workflow for high-frequency water quality monitoring that considers monitoring data requirements, sensor selection, maintenance and calibration, and systematic data processing. It is important to evaluate the cost of every element and what the added value is of monitoring at a high frequency instead of a low frequency (Rozemeijer et al., 2025).

Our analysis of sensor datasets here shows that the size of the catchment, land use, baseflow index and the degree of urbanisation with associated sewage discharges to rivers will determine the most important biogeochemical cycles for each parameter in each season, and hence the required sampling frequency when relying on sensor-derived observations. An additional challenge is that the minimum temporal frequency is not static but can vary per season and per year. This variability might also increase in the future, with warmer, wetter years and a greater frequency of sudden, intense rainfall are predicted (Ockenden et al., 2016). Optimising the measurement frequency over time or in real-time as a response to external stressors (extreme events) with adaptive monitoring strategies (Blaen et al., 2016; Coraggio et al., 2022), can improve data collection for extreme weather driven parameters. For parameters affected by diurnal cycles, possible methods to prevent bias are to sample at standardised times of the day or taking 24 samples every seven hours (the 24/7 sampling approach), which samples every hour of the day over the course of a week (Halliday et al., 2012). The 24/7 approach was designed for the use of auto-samplers which require samples to be returned to the laboratory for analysis, but has the potential to be a cost-effective measurement frequency regime for sensor optimisation to capture dynamic river conditions (Halliday et al., 2012; Jordan & Cassidy, 2022).

In general, when deciding a minimum measurement frequency for a sensor suite, the median, 25% and 75% intervals and the data distribution as well as the range should be investigated relative to hourly data or the highest available frequency. The minimum required sampling frequency can only be determined with high-frequency observations at that location, which are often unavailable when a monitoring programme is designed. As a result, sampling frequency recommendations are typically done retrospectively, as with our analyses that suggested a minimum of four-hourly frequency. We therefore recommend flexible high-frequency monitoring installations, including sensors or autosamplers, that can be deployed for trial periods to understand the behaviour of the catchment before the long-term sampling regime commences, so

this can be optimised to reduce resource expenditure which captures representative environmental behaviours for the determinands of interest. We also caution that the data should be captured with a clear focus on understanding what questions will be asked, whether the sensors selected have uncertainty bounds beyond the expected variability, and whether capturing the full range of behaviour of all parameters is indeed necessary.

2.6 Conclusions

High resolution hydrochemical data from six different catchments in the UK was systematically analysed using data resampling techniques, which provided answers to the formulated research questions.

1) What is the lowest measurement frequency that can fully capture data variation in different water quality parameters?

All catchments in this study showed that for almost every parameter, a four-hourly data frequency was required to capture most of the hourly variation across all determinands monitored, although for some parameters most variation could be captured with twelve-hourly or daily frequency.

2) How does sampling at specific times of the day impact the interpretation of biogeochemical cycles?

Different catchments have different responses to biogeochemical and hydrological events, thus the measurement regime required to capture the true range of variation will itself be variable. Nutrient concentrations, flow regimes and temperature drive much of the in-stream biological activity and their temporal variations can in turn affect variability in other water quality parameters, such as DO and pH. Most catchments included in this study showed significant intra-daily trends in physico-chemical parameters, often clearly defined diurnal cycles, but the importance of these trends in some parameters varied depending on the season. For example, photosynthesis-related diurnal oxygen patterns are generally less prevalent in winter. If a parameter has significant intra-daily variation, the sampling time of day will determine the data distribution and alter the range and median.

3) What is the effect of measurement frequency on the interpretation of biogeochemical cycles?

The measurement frequency influences the interpretation of biogeochemical cycles, as commonly measured parameters exhibit diurnal patterns and short-lived responses to rainfall. Reducing the measurement frequency impacted the calculated range and median, but it varied by catchment and parameter. Generally, parameters that are controlled by time of day (diurnal patterns) will impact the median, whilst parameters that are controlled by extreme weather will impact the range. Parameters with a more predictable pattern can be measured at a lower frequency without impacting the interpretation of biogeochemical cycles.

High-frequency sensor data can be used to provide more insight into biogeochemical processes in freshwaters, for example to enable real-time prediction of taste and odour (T&O) events in drinking water reservoirs. This study emphasises the importance of carefully considering the monitoring frequency and the measurement time when designing programs using high-frequency sensors, with assessment needed for each catchment and parameter. The purpose of the monitoring should be clearly defined to allocate resources effectively, and a pre-monitoring sensor optimisation period is advised to understand catchment-specific responses. If data variation is small, sensors with a large uncertainty won't accurately capture the fluctuations, so sensor performance must be evaluated beforehand. Regular re-assessment of the monitoring network is essential to account for environmental changes, sensor performance, and evolving management priorities.

Chapter 3: The effect of nutrient ratios and concentrations on 2-MIB production by benthic cyanobacteria in a UK drinking water reservoir

3.1 Introduction

Water reservoirs can suffer from episodes of unfavourable taste and odour (T&O) that limits the use of drinking water assets, a problem that is globally increasing in frequency and magnitude (Winter et al., 2011). The most common T&O compounds are geosmin (trans-1,10-dimethyl-trans-9-decalol) and 2-methylisoborneol (1,2,7,7-tetramethyl-exo-bicyclo-[2,2,1]-heptan-2-ol, referred to as 2-MIB). These are secondary metabolites widely produced by soil bacteria (actinomycetes) that may be active in reservoirs (Asquith et al., 2018; Clercin et al., 2022; Kutovaya & Watson, 2014), but some strains of cyanobacteria have also acquired the genes needed to produce these metabolites. Many genera of cyanobacteria have been found to produce geosmin and/or 2-MIB in culture or molecular-based studies, and they are mostly filamentous (Watson et al., 2016). Geosmin is generally associated with planktic cyanobacterial genera like *Dolichospermum*, *Aphanizomenon* and *Planktothrix* but can also be produced by certain benthic genera. 2-MIB is almost exclusively produced by benthic cyanobacterial genera such as *Oscillatoria*, *Phormidium*, *Pseudanabaena* and *Leptolyngbya* (Watson et al., 2016) and some of these genera can produce both geosmin and 2-MIB. Recent research indicates that benthic mats of cyanobacteria should not be overlooked as a source of geosmin and 2-MIB, especially in rivers, shallow lakes and shores of reservoirs (Espinosa et al., 2020; Gaget et al., 2020; Jähnichen et al., 2011; Otten et al., 2016; Watson & Jüttner, 2019). In drinking water reservoirs, the source of T&O is difficult to pinpoint as the T&O compounds are persistent and spread through the water column easily once they are released from the cell. Moreover, standard methods of water column sampling won't detect benthic species that can sometimes grow over a large part of the shallow zones of the reservoir and near or on the drinking water inlet structure (Otten et al., 2016).

Environmental conditions (temperature and light availability, plus nutrient availability) can enhance cyanobacterial growth and persistent blooms (Paerl & Otten, 2013;

Richardson et al., 2019), and it is hypothesised that climate change will result in more optimal growth conditions for cyanobacteria (Cottingham et al., 2015; Paerl & Otten, 2013; Paerl & Paul, 2012). Nitrogen (N) and phosphorus (P) concentrations and ratios determine phytoplankton community structure (Harris et al., 2016). DIN:SRP (dissolved inorganic nitrogen : soluble reactive phosphorus) and TN:TP (total nitrogen : total phosphorus) ratios in the water can be used to determine approximations of potential N- or P limitation of phytoplankton growth. The Redfield ratio for N:P at 16:1 (molar based) is thought to be the transition between N-limitation (<16) and P-limitation (>16) (Redfield et al., 1963), but it depends on the phytoplankton taxa and growth stage (Klausmeier et al., 2004). Kosten et al. (2009) specified N-limitation as TN:TP ratios below 20 and DIN:SRP below 13, whilst TN:TP ratios above 38 and DIN:SRP above 50 indicate P limitation (Kosten et al., 2009). Cyanobacteria dominance in the phytoplankton community has often been linked to a low TN:TP ratio (Smith, 1983; Vrede et al., 2009), <64:1 (29:1 by mass) according to Smith (1983). P-limiting conditions generally favour chlorophytes and diatoms, whilst cyanobacteria can dominate under N-limiting conditions (Andersen et al., 2020).

These ratios largely depend on inputs (internal or external) and uptake of inorganic N and P concentrations which can result in highly variable nutrient ratios over a short space of time (Glibert et al., 2008). The form of N and P also depends on the dominant input, which can cause variability depending on the season (Andersen et al., 2020). External inputs like groundwater are often the dominant source of NO_3^- whilst internal recycling of organic matter under anaerobic conditions increases NH_4^+ (Andersen et al., 2020). The chemical form of inorganic nitrogen; oxidized (nitrate; NO_3^-) or reduced (ammonium; NH_4^+), influences the uptake rate by different phytoplankton taxa because they have different cellular energetic costs, ammonium being more bioavailable (Erratt et al., 2020; Glibert et al., 2016; Trommer et al., 2020). Phytoplankton taxa have specific traits (N-fixation etc.) and some produce enzymes that give them a different affinity for certain forms of N and P (Chaffin & Bridgeman, 2014). Optimal ammonium concentrations can result in higher growth rates than nitrate and urea, but concentrations that are too high can repress growth or even be toxic (Glibert et al., 2016), which depends on phytoplankton class and genus/species (Collos & Harrison, 2014). Chemically reduced forms of N, like ammonium, but also urea and dissolved organic nitrogen (DON) are increasing in freshwaters due to

sewage discharges, aquaculture and the worldwide rise of urea-based fertilizers (Glibert et al., 2016). Therefore, this is an important area of research as we might create the optimal scenario for cyanobacterial blooms and subsequent problems.

The NH_4^+:NO_3^- ratio may influence the phytoplankton community structure (Glibert et al., 2016; McCarthy et al., 2009) and the abundance of T&O producing cyanobacteria. Several studies showed that NO_3^- enrichment favoured diatom growth whilst NH_4^+ enrichment favoured dinoflagellates, chlorophytes and cyanobacteria (Andersen et al., 2020; Donald et al., 2013; Donald et al., 2011; Erratt et al., 2020; Glibert & Berg, 2009), including toxin producing species of cyanobacteria (McCarthy et al., 2009). Espinosa et al. (2021) also found that high nutrient concentrations (12 µg Ammonium-N, 110 µg Nitrate-N, 4 µg Phosphate-P) and a low DIN:SRP (dissolved inorganic nitrogen and soluble reactive phosphorus) ratio of 4:1 stimulated geosmin production by benthic *Oscillatoria* sp.. Saadoun et al. (2001) found that geosmin production by *Anabaena* sp. was directly correlated with a high Ammonium-N and a low Nitrate-N, with low N:P ratios resulting in increased geosmin production. A field study in Korea (Lee et al., 2023) found that 2-MIB correlated with water temperature and NH_3-N. Water temperature and NH_3-N also revealed a positive correlation with *Pseudanabaena* sp. growth in cascading reservoirs Xili-Tiegang-Shiyan, China (Gao et al., 2018). Harris et al. (2016) found that a low TN:TP <66:1 (30:1 as mass ratio) together with a low NO_3^-:NH_4^+ ratio resulted in the ideal conditions for cyanobacteria to produce T&O compounds. Similarly, Perkins et al. (2019) identified that ammonium (NH_4^+) can predict geosmin and 2-MIB production as well as low TN:TP ratios. Perkins et al. (2019) hypothesised that when cells switch from oxidized (NO_3^-) to reduced (NH_4^+) as nitrogen source the excess energy might be reallocated by producing geosmin and 2-MIB as an "overflow" product.

Biomass of cyanobacteria (expected T&O producers) often does not clearly relate to total T&O production in field studies (Harris et al., 2016), thus there are complexities to be revealed between environmental drivers, time-lags between intracellular T&O production and extracellular release and ability to measure within the water column. It is generally agreed that T&O producing taxa produce T&O compounds intracellularly during the exponential growth phase, with maximum production in late exponential phase and the T&O compounds are released extracellularly during the stationary and death phase (Alghanmi et al., 2018; Naes et al., 1989; Saadoun et al., 2001; Zhang et

al., 2009). It is still uncertain whether cyanobacteria release T&O compounds actively, for example to stimulate other phytoplankton to release organic phosphorus (Bar-Yosef et al., 2010; Raven, 2010), or passively due to cell death and lysis (Zhang et al., 2016), or if it is a combination of both processes (Wang & Li, 2015; Watson, 2003). Productivity and photosynthetic activity have been hypothesized to negatively (I) or positively (II) influence T&O production. I) Chlorophyll *a* and T&O metabolite synthesis compete for the same isoprenoid pathway and therefore, T&O metabolites are only produced under decreased photosynthetic activity as an "overflow" product to release excess energy (Kutovaya & Watson, 2014; Liu et al., 2009; Pattanaik & Lindberg, 2015; Saadoun et al., 2001; Shen et al., 2020; Wang & Li, 2015). II) Alternatively, increased T&O accumulation can simply be the result of increased carbon flow through the isoprenoid pathway when the cells experience increased cell metabolism (Alghanmi et al., 2018; Giglio et al., 2011; Zimba et al., 1999). What remains to be understood is how physical and chemical conditions interact to stimulate T&O production in representative drinking water reservoir phytoplankton and microbial communities. This study therefore aims to investigate the effects of different nutrient ratios ($NH_4^+:NO_3^-$ and DIN:SRP) and concentrations on the phytoplankton community structure and resulting 2-MIB production under controlled physical conditions.

This chapter aims to answer the second overarching hypothesis: Nutrient concentrations and ratios are an important factor in predicting geosmin and 2-MIB events. Laboratory microcosms with a natural phytoplankton community were used to identify the effect of different nutrient concentrations and ratios, $NH_4^+:NO_3^-$ and DIN:SRP, on 2-MIB production. Several research questions were addressed in this study:

1) How do different nutrient ratios ($NH_4^+:NO_3^-$ and DIN:SRP) and concentrations impact the community structure and productivity of phytoplankton?
2) How do phytoplankton community structure and productivity affect the production of 2-MIB?
3) Can specific nutrient ratios and concentrations be used as an early warning for the onset of 2-MIB production?

3.2 Materials and methods

3.2.1 Experimental set-up

Surface water from a drinking water reservoir in North Wales, Latitude 53.222067° N, Longitude -3.540305° W (further referred to in the text as Reservoir 1; in European Nucleotide Archive (ENA) as WR1), was collected on the 28th of June 2022 in two acid-washed (10% HCl) and sterilized (autoclaved) 20 L Nalgene bottles, after they were rinsed three times with the same surface water. Samples were transported to the laboratory on the same day and stored at room temperature. The next day, 29th of June, water from one of the 20 L Nalgene bottles was filtered through a 0.7 um GF/F filter with a vacuum pump in the laminar flow hood and collected in a third prepared 20 L Nalgene bottle, using bleached (1% sodium hypochlorite) siphoning tubes. Samples were taken from the filtered and unfiltered bottle, after which the bottles were kept in a cold room at ±4 ºC until the start of the incubation experiment, to hinder further nutrient changes. Samples were filtered through a pre-leached 0.45 μm cellulose nitrate filter with a vacuum pump and analysed for concentrations of dissolved inorganic nitrogen (DIN; NO_3^-, NH_4^+, NO_2^-) and dissolved inorganic phosphorus (DIP; PO_4^{3-}) on a Skalar San++ multichannel continuous flow autoanalyzer (Skalar Analytical B.V., The Netherlands), at the University of Bristol (detailed methods described in Section 3.3.1 and 3.3.2 and analytical performance of each run specified in Section 3.3.3). The measured nutrient concentrations in filtered and unfiltered water were used to determine mixing ratios of the two water types and calculate treatment-specific nutrient additions to ensure the required nutrient ratios and concentrations (Table 3.1 and Table 3.2). Consequently, the filtered water was 50% diluted with ultrapure Milli-Q water and this "medium" was mixed with the unfiltered water as "inoculum" in the ratio 50% medium with 50% inoculum. This new mixture was sampled in triplicate in amber glass bottles (1000 ml) for geosmin and 2-MIB analysis at Glaslyn laboratory (Dŵr Cymru Welsh Water, referred to as DCWW) (Section 3.3.7) and additional triplicate samples were collected in acid-washed (5-10% HCl) and sterilized (oven/furnace at 200 ºC for 2 hours or autoclaved) Duran bottles (250 ml) to enable Sterivex filtering for DNA analysis (Section 3.3.8).

For the incubation experiment (Figure 3.1), which started on 4th of July, 850 ml of this mixture was filled in 21 acid-washed (5-10% HCl) and sterilized (autoclaved) 1000 ml Erlenmeyer flasks with foam stoppers. Nutrient spikes were added from highly

concentrated NO_3^-, NH_4^+ and PO_4^{3-} stock solutions (1000 or 100 mg/L, shown in Table 3.2) to create seven treatments; six treatments with specific nutrient ratios and a control in which nothing was added (Table 3.1). Each treatment was done in triplicate.
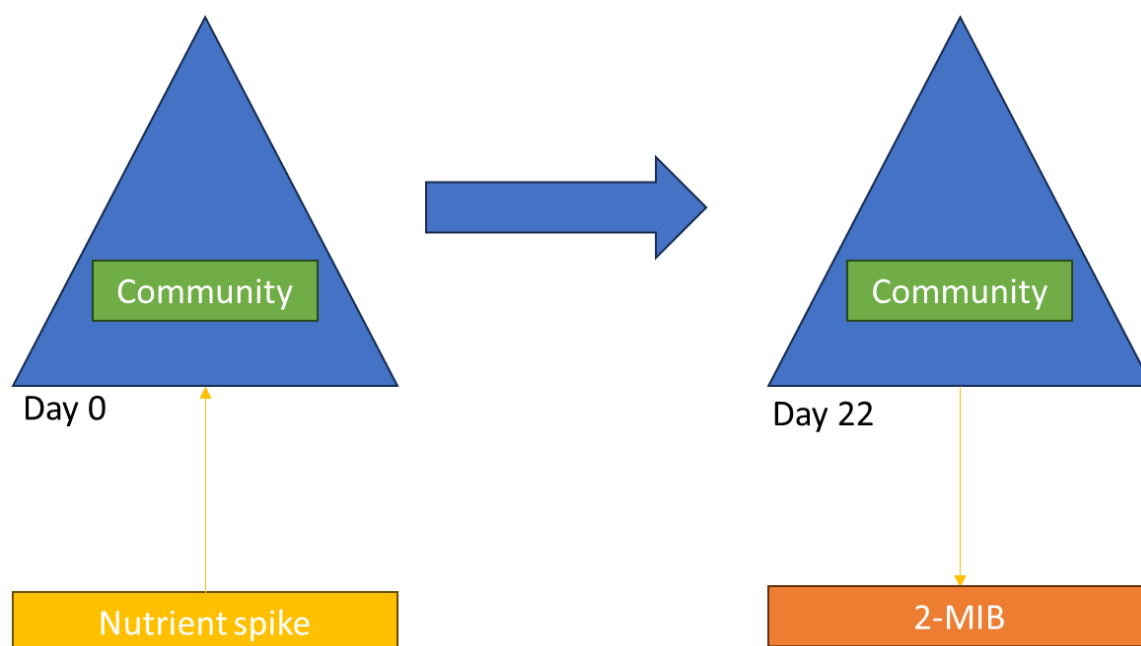
Figure 3.1. Experiment overview which highlights the different elements of the research questions that were addressed by the study.

Table 3.1. Final molar nutrient ratios NH_4^+:NO_3^{3-} (ratio in brackets if NH_4^+ was compared to 1 mol of NO_3^-) and DIN:SRP of each treatment, done in triplicate (n=3). Treatment ID was used for reference in this study.

Treatment	Treatment ID	Treatment ID meaning	NH_4^+:NO_3^-	DIN:DIP
Control	No Addition	No Addition	x	x
A	HNO3_LNH4_HPO4	High NO_3^-, Low NH_4^+, High PO_4^{3-}	1:100 (0.01)	9:1
B	MNO3_MNH4_HPO4	Medium NO_3^-, Medium NH_4^+, High PO_4^{3-}	1:10 (0.1)	9:1
C	LNO3_HNH4_HPO4	Low NO_3^-, High NH_4^+, High PO_4^{3-}	1:1 (1)	9:1
D	HNO3_LNH4_LPO4	High NO_3^-, Low NH_4^+, Low PO_4^{3-}	1:100 (0.01)	128:1
E	MNO3_MNH4_LPO4	Medium NO_3^-, Medium NH_4^+, Low PO_4^{3-}	1:10 (0.1)	128:1
F	LNO3_HNH4_LPO4	Low NO_3^-, High NH_4^+, Low PO_4^{3-}	1:1 (1)	128:1

Table 3.2. Volume (mL) of nutrients from stock solutions (concentration of 1000 mg/L or 100 mg/L) that was added to each treatment, apart from the control, as well as target concentrations (conc.) of NO_3^--N, NH_4^+-N and PO_4^{3-}-P in each treatment.

Treatment	mL of NO_3^--N stock (1000 mg/L)	Target conc. NO_3^--N (mg/L)	mL of NH_4^+-N stock (100 mg/L)	Target conc. NH_4^+-N (mg/L)	mL of PO_4^{3-}-P stock (100 or 1000 mg/L)	Target conc. PO_4^{3-}-P (mg/L)
A	7.492	7.714	0.000	0.077	1.915 (1000 mg/L)	1.915
B	1.179	1.400	0.629	0.140	0.379 (1000 mg/L)	0.379
C	0.479	0.700	6.229	0.700	0.344 (1000 mg/L)	0.344
D	7.492	7.714	0.000	0.077	1.347 (100 mg/L)	0.135
E	1.179	1.400	0.629	0.140	0.266 (100 mg/L)	0.027
F	0.479	0.700	6.229	0.700	0.242 (100 mg/L)	0.024

The dissolved oxygen and pH were measured with benchtop probes (Sections 3.3.5 and 3.3.6) in the Erlenmeyers, sterilizing them in between with 70% Ethanol. 100 ml was taken from each Erlenmeyer straight away for nutrient analysis on a Skalar San++ (Skalar Analytical B.V., The Netherlands) and an additional 5 ml was taken for measurements on a PHYTO-PAM-II multiple excitation wavelength phytoplankton & photosynthesis analyzer (Walz Photosynthesis Instruments, Germany). The 21 Erlenmeyer flasks were placed in an LMS 280 NP+ incubator (LMS, United Kingdom) set at 20 °C, with average light intensity of 33 (± 9.6 standard deviation; SD) μmol photons m^{-2} s^{-1} and a 12h:12h light:dark cycle. Every day throughout the 23 day long experiment they were swirled and randomly relocated in the incubator to ensure a random block design. Three flasks were chosen from treatments A, C and Control and these flasks (Ctrl3, A2 and C1) had continuous dissolved oxygen sensors (PreSens oxygen needle-type sensors, more information below) installed. Sub-samples were collected on days 4, 8, 12, 16 and 19 for analysis of community productivity (PHYTO-PAM-II), pH (Orion) and dissolved oxygen (PreSens Fibox 3). Sub-samples were collected on day 0 and day 22 for nutrient concentrations (Skalar San++, Skalar Analytical B.V., The Netherlands; Gallery™ Discrete Analyzer, ThermoFisher Scientific™, USA; at the University of Bristol), T&O compounds (Dŵr Cymru Welsh Water laboratory), environmental DNA (Cardiff University) and qPCR (Cardiff University) (Figure 3.2, Table 3.3). PHYTO-PAM-II measurements were deemed

unreliable due to technical issues caused by non-homogeneous phytoplankton growth in the experiment and were subsequently excluded from further analysis.
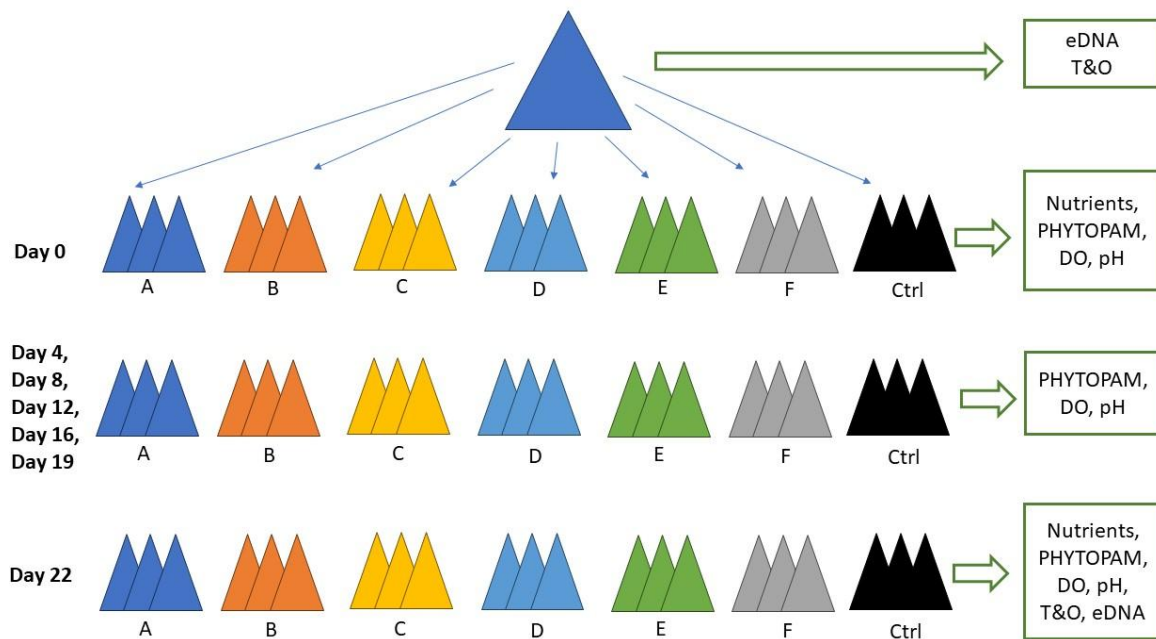
Figure 3.2. Experimental design which demonstrates the sampling days and analyses that were done at each sampling day. More details on measurements can be found in Table 3.2.

Table 3.3. Measurements and their temporal frequency during the laboratory incubation experiment

Frequency	Measurement
Continuously	Dissolved oxygen in flasks Ctrl3, A2 and C1 (primary productivity proxy)
Every 3 or 4 days	pH, dissolved oxygen, PHYTO-PAM-II (biomass proxy and yield/health)
Start and End	Nutrient composition, DOC, T&O concentration (geosmin and 2-MIB), DNA next generation sequencing (16S rRNA and *rbc*L), qPCR (16S rRNA, *geo*A, *mic*)

3.3 Laboratory methods

3.3.1 N, P, C sample preparation

100 ml of sample was collected in acid-washed (5-10% HCl) bottles and 60 ml was filtered through a pre-leached 0.45 µm cellulose nitrate filter with a vacuum pump. Half of the filtered sample was collected in a 50 ml Falcon tube and frozen at -20 °C for later DOC analysis (4). The other half of the filtered sample was put in a second 50 ml

Falcon tube and split in two, 10 ml was used as Filtered Fresh sample (1) and another 10 ml was digested, to become the Filtered Digested (2) sample. Digestion was conducted using the persulphate ($K_2S_2O_8$) oxidation method, described by Johnes and Heathwaite (1992) and modified for the CEM Mars Xpress microwave digestion unit by Yates et al. (2019b). Moreover, 10 ml of Unfiltered sample from the collection bottle was digested, creating the Unfiltered Digested (3) sample.

1)   Filtered Fresh = PO_4^{3-}-P, TOxN (nitrate and nitrite), NH_4^+-N

2)   Filtered Digested = Total dissolved phosphorus (TDP), Total dissolved nitrogen (TDN)

3)   Unfiltered Digested = Total phosphorus (TP), Total nitrogen (TN)

4)   DOC analysis

3.3.2 N and P fractions

Sample types 1, 2 and 3 (explained in Section 3.3.1) were analysed within 24 hours of sample collection using a Skalar San++ multichannel continuous flow autoanalyzer (Skalar Analytical B.V., The Netherlands), following the methods described in (Yates et al., 2019a; Yates et al., 2019b). The Skalar San++ (Skalar Analytical B.V., The Netherlands) is a continuous flow analyser with colorimetric detection which can analyse digested samples (sample types 2 and 3). The Skalar San++ (Skalar Analytical B.V., The Netherlands) was set up to simultaneously measure total oxidised nitrogen (TOxN, includes nitrate as NO_3^--N and nitrite as NO_2^--N), as well as total ammonium (NH_3-N + NH_4^+-N) and soluble reactive phosphorus (SRP, indicative of PO_4^{3-}-P). An independent stock solution was used to make calibration standards (range 0.01 – 1 mg/L for PO_4^{3-}-P , NH_4^+-N and NO_2^--N and 0.1 – 10 mg/L for TOxN) and the stock solution 6 (TOxN = 6 mg/L, NH_4^+-N = 0.6 mg/L, NO_2^--N = 0.6 mg/L, PO_4^{3-}-P = 0.6 mg/L) was run frequently throughout the analysis to enable quality control. Additionally, analytical and digest blanks were run to check the performance of the protocol and the instrument. From these results, additional nitrogen and phosphorus fractions can be determined, using calculations specified in Lloyd et al. (2019); Yates et al. (2019a):

Dissolved organic nitrogen; DON = TDN – TON – NH_4^+-N
Dissolved organic phosphorus; DOP = TDP – PO_4^{3-}-P

Particulate organic nitrogen; PON = TN – TDN

Particulate phosphorus; PP = TP - TDP

The Skalar San++ (Skalar Analytical B.V., The Netherlands) did not perform well for NH_4^+-N concentrations, possibly due to background ammonium present in the laboratory which caused sub-optimal calibration curves. The Skalar San++ (Skalar Analytical B.V., The Netherlands) NH_4^+-N measurements for the three sample runs 'preparation' (LoD 0.07 mg/L, ±3.1% precision, ±5.3% accuracy), 'day 0' (LoD 0.11 mg/L, ±5.7% precision, ±5.4% accuracy) and 'day 22' (LoD 0.11 mg/L, ±5.8% precision, ±8.5% accuracy) were not precise or accurate (Table 3.4). Therefore, 'day 22' (final day) samples were also analysed on a Gallery™ Discrete Analyzer (ThermoFisher Scientific™, USA) for NH_4^+-N with the low detection setting (LoD 0.003 mg/L, ±0.9% precision, ±1.7% accuracy) (Table 3.4), and a comparison of both sample measurements can be found in Figure S.1 (Appendix B). The Gallery™ Discrete Analyzer (ThermoFisher Scientific™, USA) is a continuous flow analyser with colorimetric detection but cannot analyse digested samples (sample types 2 and 3). Moreover, filtered water samples are frozen before they are analysed on the Gallery™ Discrete Analyzer, hence a test was done to compare between filtered samples stored in the fridge and in the freezer. Two water samples were either frozen or put in the fridge and analysed within 24 hours on the Skalar San++, whilst the frozen sample was analysed on the Gallery™ Discrete Analyzer. No notable differences that would raise concern were identified in the results, caused by freezing (Appendix B, Table S.2)

3.3.3 LoD, precision and accuracy of Skalar San++ and Gallery™

LoD, precision and accuracy were calculated for each individual nutrient fraction for every run on the Skalar San++ (Skalar Analytical B.V., The Netherlands) and Gallery™ Discrete Analyzer (ThermoFisher Scientific™, USA). LoD was determined based on the linear calibration curves, implementing a linear regression to x (actual concentration) and y (measured concentration) to calculate Sy (the standard deviation of the predicted y-value for each x in a regression) and S (slope of the regression). Following guidelines from ICH Harmonised Tripartite Guideline (2005), the limit of detection (LoD) was calculated using Equation 3.1.

Equation 3.1:

$$LoD = 3.3 * \left(\frac{Sy}{S}\right)$$

The precision was calculated based on repeated measurements of the same concentration. For the Skalar San++ (Skalar Analytical B.V., The Netherlands) this was a drift measurement (n=9) of a calibration solution that contained 0.6 mg/L of nitrite, phosphate and ammonia and 6 mg/L of TOxN. The Gallery™ Discrete Analzyer (ThermoFisher Scientific™, USA) did not execute a drift measurement but measured all calibration solutions in triplicate (n=3), so the calculations were performed on each calibration concentration separately and then averaged. For each repeated measurement, the absolute deviation from the average of all measurements was calculated, which was averaged (average deviation), then divided by the average of all measurements and multiplied by 100 (Equation 3.2). An example of hypothetical triplicate measurements y1, y2 and y3 (average value = $\bar{y}$) of actual concentration x was used to clarify Equation 3.2 and Equation 3.3.

Equation 3.2:

$$Precision = \frac{\left(\frac{|y1 - \bar{y}| + |y2 - \bar{y}| + |y3 - \bar{y}|}{3}\right)}{\bar{y}} * 100$$

The accuracy was calculated as the absolute percentage difference for repeated measurements; "average concentration ($\bar{y}$)" – "actual concentration (x)" divided by the "actual concentration (x)" and multiplied by 100 (Equation 3.3):

Equation 3.3:

$$Accuracy = \left| \frac{\bar{y} - x}{x} * 100 \right|$$

Moreover, the coefficient of variation (CV %) was calculated as the standard deviation of the measurements divided by the average of the measurements multiplied by 100 to get a percentage.

Table 3.4 demonstrates the different analytical performance of every nutrient fraction in each sample run for the Skalar San++ (Skalar Analytical B.V., The Netherlands) and the Gallery™ (ThermoFisher Scientific™, USA), to compare between instruments and runs (Table 3.4).

Table 3.4. Limit of detection (LoD, mg/L), precision (%), accuracy (%) and coefficient of variation (CV, %) for each run on the Skalar San++ (Skalar) and Gallery^{TM} (Gallery) for nitrite (NO$_2^-$-N), phosphate (PO$_4^{3-}$-P), total oxidisable nitrogen (TOxN) and ammonium (NH$_4^+$-N).

Nitrite (NO$_2^-$-N)						
Sample	**Date**	**Instrument**	**LoD (mg/L)**	**Precision (%)**	**Accuracy (%)**	**CV (%)**
Preparation	30/06/2022	Skalar	0.076	14.647	18.185	19.194
Day 0	05/07/2022	Skalar	0.015	1.120	3.574	1.419
Day 22	27/07/2022	Skalar	0.060	2.505	10.389	2.989
Phosphate (PO$_4^{3-}$-P)						
Sample	**Date**	**Instrument**	**LoD (mg/L)**	**Precision (%)**	**Accuracy (%)**	**CV (%)**
Preparation	30/06/2022	Skalar	0.030	0.584	2.130	0.733
Day 0	05/07/2022	Skalar	0.022	0.747	1.852	0.938
Day 22	27/07/2022	Skalar	0.070	0.672	5.296	0.936
Total Oxidisable Nitrogen (TOxN)						
Sample	**Date**	**Instrument**	**LoD (mg/L)**	**Precision (%)**	**Accuracy (%)**	**CV (%)**
Preparation	30/06/2022	Skalar	0.260	3.097	8.256	3.420
Day 0	05/07/2022	Skalar	0.128	1.137	18.793	1.840
Day 22	27/07/2022	Skalar	0.499	1.077	10.919	1.360
Ammonium (NH$_4^+$-N)						
Sample	**Date**	**Instrument**	**LoD (mg/L)**	**Precision (%)**	**Accuracy (%)**	**CV (%)**
Preparation	30/06/2022	Skalar	0.068	3.056	5.296	3.829
Day 0	05/07/2022	Skalar	0.108	5.680	5.352	6.465
Day 22	27/07/2022	Skalar	0.113	5.829	8.500	7.323
Day 22	29/07/2022	Gallery	0.003	0.896	1.703	1.198

3.3.4 DOC

Following methods from Yates et al. (2019a) and Yates et al. (2019b), dissolved organic carbon (DOC) concentrations were measured as non-purgeable organic carbon, on a 0.45 µm filtered sample which was acidified with HCl. A Shimadzu TOC-L series analyser (Shimadzu Corp., Kyoto, Japan) was used for coupled high temperature catalytic oxidation. The mean of three to five 100 µL injections was calculated, where the coefficient of variance between replicate injections was <2%.

3.3.5 Oxygen probes

Continuous dissolved oxygen measurements were executed with needle-type fiber-optic oxygen microsensors, Profiling Oxygen Microsensor PM-PSt1 (PreSens Precision sensing GmbH, Germany) that were connected to an autonomous data logger OXY-10 micro, a 10-channel microsensor oxygen meter (PreSens Precision sensing GmbH, Germany) using PreSens OXY10 v3 software on a desktop.

Each measurement day, an Oxygen Dipping Probe DP-PSt3 (PreSens Precision sensing GmbH, Germany) and a temperature sensor were connected to a Fibox 3 fiber optic oxygen meter and a laptop using PreSens PSt3 v6 software (PreSens Precision sensing GmbH, Germany). The software recorded dissolved oxygen as air saturation (%) and temperature (°C) every 10 seconds for the whole day. After the experiment, all the data was processed to retrieve the correct oxygen measurements in each flask by matching the measurement time in the dataset when the oxygen sensor had stabilized. This way a range of datapoints was available for each flask at every measurement day, enabling more data quality certainty and statistical calculations (mean, standard deviation etc.).

Both oxygen sensors were calibrated before the experiment and checked for drift after the experiment. The following method was used; 100% air saturation in the air above damp cotton wool and 0% in a solution with sodium sulphite dissolved in Milli-Q water (±0.1 gram in 10 ml Milli-Q). Continuous dissolved oxygen needle-type sensors (Pst1) performed well throughout the experiment with <2% drift. Final day measurements with calibration standards were; channel 1 = ctrl 3 (0% = 0.1%; 100% = 100%), channel 5 = a2 (0% = 0.1 %; 100% = 98%), channel 10 = c1 (0% = 0.1%; 100% = 98%). The intermittent oxygen dipping sensor (Pst3) also performed well, with minimal drift of < 0.2% (0% = 0.0%; 100% = 99.8%).

3.3.6 pH

pH was measured with a ThermoFisher Scientific^{TM} Orion Star series benchtop pH meter (ThermoFisher Scientific^{TM}, USA), equipped with a Sentek pH probe Type P11/ ROD/ BNC (Sentek Sensor Technologies, Australia). The probe was calibrated at the start of the experiment with a 4-point calibration (pH 2, 4, 7 and 10). After the experiment, a drift check was done and in the pH range of the experiment (7-10), the drift was approximately 0.18 pH value (7 = 6.84; 10 = 9.82).

3.3.7 Geosmin and 2-MIB

250 ml of sample was collected in Dŵr Cymru Welsh Water (DCWW) provided amber glass vials (1000 ml) and stored in the dark at 4 °C until samples were transported to Glaslyn, the DCWW accredited laboratory (ISO/IEC 17,025:2017), for analysis. Their accredited laboratory method ensures that geosmin and 2-MIB first get extracted via a solid-phase extraction (SPE) and they are subsequently eluted using 1.0 ± 0.01 mL

dichloromethane into a labelled 2 mL auto sampler vial. A deuterated internal standard was added to the sample to assess efficiency of the run. Samples were analysed by GC-MS triple quadrupole MS in MRM mode (Perkins et al., 2019). Briefly explained: Samples were injected into a gas chromatograph using a multimode inlet in solvent vent mode, where helium transported the analytes on to a 30 M Ultra Inert HP-5MS capillary column (Agilent, USA) and they were separated by boiling point. After the gas chromatograph (GC), analyte detection was done with a triple quadrupole Mass Selective detector (MS/MS) in multiple reaction monitoring (MRM) mode (Hooper, 2023a).

3.3.8 DNA sample preparation and extraction

250 ml of sample was collected in acid-washed (5-10% HCl) and sterilized (oven/furnace at 200 °C for two hours or autoclaved) Duran bottles (250 ml). A filter set-up with a sterile 50 ml syringe and a vacuum pump on the laboratory bench was used to filter 250 ml of sample through a 0.2 µm Sterivex filter unit. One ml of ATL buffer (Qiagen, Germany) was added to the filter and afterwards the filters were sealed with Parafilm and stored at -20 °C until analysis. Autoclaved Milli-Q water had also been filtered in triplicate on different days, to get an average of background DNA in the filtration environment, which was used for correcting DNA results.

The liquid from defrosted Sterivex filters was distributed equally in two sterile 15 ml Falcon tubes (one for DNA extraction, one back-up), both containing approximately 0.5 ml ATL. In a laminar flow hood and following sterile laboratory practices, the Sterivex filters were cut open with a PVC pipe cutter and the filter paper was detached from the cylindrical interior with a surgical blade, using advice from Cruaud et al. (2017). The filter paper was cut into smaller pieces and equally distributed to both Falcon tubes and back-up tubes were stored at -20 ºC. To the active Falcon tubes, 43.6 µL of proteinase-K was added, they were vortexed and spun down, sealed with parafilm and incubated in a water bath at 56 °C for 30 minutes. Tubes were placed in -20 °C freezer overnight and defrosted the next day in a water bath at 56 °C for 1 minute. Samples were dipped into liquid nitrogen for 10 seconds, then placed in a water bath at 65 °C, followed by another 2 minutes at -20 °C to ensure vigorous freeze-thaw cycles. Subsequent DNA extraction protocol was adapted from Fawley and Fawley (2004) and was described by Watson et al. (2024) for flat DNA filters, which includes both chemical and physical (bead beating) treatments and DNA purification

using the DNAeasy® Blood & Tissue kit (QIAGEN, Germany). Purified DNA was eluted in 100 µL of nuclease-free water after a 10-minute incubation to ensure maximum DNA recovery.

DNA quantity was measured with the Qubit High Sensitivity Assay Kit (Invitrogen, ThermoFisher Scientific™, USA) and a 96-well plate was created with manual dilutions for each sample to normalise to 0.5 ng/µL. Manual dilution was required, instead of a standard 1:10 dilution, due to large differences in DNA yields between the samples. Extraction controls were run for each time the extraction protocol was executed.

3.3.9 Next generation sequencing
16S rRNA

Following methods from Watson et al. (2024), bacteria-specific primer set 515F (forward) 5′-GTGCCAGCMGCCGCGGTAA-3′ and 806R (reverse) 5′-GGACTACHVGGGTWTCTAAT-3′ (Caporaso et al., 2011) was used to amplify a fragment of approximately 350 base pair (bp) of the 16S rRNA gene. For the first polymerase chain reaction (PCR), 1 µL of template DNA (manually normalised to 0.5 ng/µL) was added to 19 µL of amplification mixture, which consisted of 5 µL AllTaq Master Mix (4x concentrated) (QIAGEN, Germany), 0.525 µL of forward primer (10 pmoles/µL), 0.525 µL of reverse primer (10 pmoles/µL), 0.25 µL of BSA (20 mg/mL) and 12.70 µL of nuclease-free water. Plates with master mix and template DNA in 20 µL reaction-volumes were set up on an epMotion® robot (Eppendorf, Germany), after which a PCR was run on an Applied Biosystems SimpliAmp thermocycler (ThermoFisher Scientific™, USA) with the following conditions: 95 ∘C for 2 min (one cycle), 95 ∘C for 5 s, 55 ∘C for 15 s, 72 ∘C for 10 s (30 cycles) and 72 ∘C for 5 s (one cycle). Samples were run in triplicate and included a negative PCR control, to reduce bias and enable quality assurance. Triplicates were pooled and visualised on a QIAxcel® Advanced System to check the quality of the PCR products. Subsequent sample purification was done with a Zymo Research DNA Clean-up Kit™, following manufacturers protocol (Zymo Research, USA). A secondary PCR was prepared in plates on the epMotion® robot (Eppendorf, Germany), with a total reaction volume of 25 µL. First, 17.5 µL of amplification mixture was prepared; 12.5 µL of KAPA HiFi HotStart ReadyMix (2x concentrated) (Roche Diagnostics, Switzerland) and 5 µL nuclease-free water. This was followed by adding 2.5 µL of unique Illumina® Nextera

XT index adapter 1 (N7XX) to identify columns and 2.5 µL of unique Illumina® Nextera XT index adapter 2 (S5XX) to identify rows (Illumina®, Cambridge, UK), after which 2.5 µL of amplified DNA was added. The PCR was executed on a MultiGene™ OptiMax Thermal Cycler (Labnet International, USA) to incorporate sample-specific Illumina® Nextera XT indices (Illumina®, Cambridge, UK), following these steps: 95 ∘C for 3 min (one cycle), 95 ∘C for 30 s, 55 ∘C for 30 s, 72 ∘C for 30 s (8 cycles) and 72 ∘C for 5 s (one cycle). A SequalPrep™ Normalization Plate Kit (Invitrogen, Carlsbad, CA, USA) was used to normalise samples to equimolar concentrations, and a pooled sample was sequenced on an Illumina® MiSeq (nano cartridge: 2 × 250 bp paired-end reads) at Cardiff School of Bioscience Genome Research Hub. Next generation sequencing data used for the molecular analysis can be found on the European Nucleotide Archive (ENA), project code: PRJEB75951.

rbcL

Next-generation sequencing preparation for rbcL generally followed the same protocol as 16S rRNA, with some primer and sample-specific alterations. The primer set contained two forward reading primers; rbcL_192_Diatoms (diatom-specific) 5′-GCTACWTGGACWGTTGTWTGGAC-3′ and rbcL_192_Green (green algae-specific) 5′-GGTACTTGGACWACWGTWTGGAC-3′ and one reverse primer; rbcL_657 (non-specific) 5′-GAAACG[I]TCWCKCCAACGCAT-3', which were designed by Maya Lhoste's Final Year Project 2023. The primer set was used to amplify a fragment of approximately 550 base pair (bp) of the RuBisCO large (rbcL) gene, which is used to identify diatoms and green algae.

For the first PCR, 1 µL of template DNA (manually normalised to 0.5 ng/µL) was added to 19 µL of amplification mixture, which consisted of 5µL AllTaq Master Mix (4x concentrated) (QIAGEN, Germany), 0.2625 µL of diatom-specific forward primer (10 pmoles/µL), 0.2625 µL of green algae-specific forward primer (10 pmoles/µL), 0.525 µL of reverse primer (10 pmoles/µL), 0.25 µL of BSA (20 mg/mL) and 12.70 µL of nuclease-free water. Plates with master mix and template DNA in 20-µL reaction volumes were set up on an epMotion® robot (Eppendorf, Germany), after which a PCR was run on an Applied Biosystems SimpliAmp thermocycler (ThermoFisher Scientific™, USA) with the following conditions: 95 ∘C for 2 min (one cycle), 95 ∘C for 5 s, 55 ∘C for 15 s, 72 ∘C for 10 s (32 cycles) and 72 ∘C for 5 s (one cycle). Samples were run in triplicate and included a negative PCR control, to reduce bias and enable

quality assurance. Triplicates were pooled and visualised on a QIAxcel® Advanced System to check the quality of the PCR products. Subsequent sample purification was done with a Zymo Research DNA Clean-up Kit™, following manufacturers protocol (Zymo Research, USA). A secondary PCR was set up on an epMotion® robot (Eppendorf, Germany), with KAPA HiFi HotStart ReadyMix (2x concentrated) (Roche Diagnostics, Switzerland) and Illumina® Nextera XT indices (Illumina®, Cambridge, UK), following the same concentrations as for 16S rRNA. Altered to enhance amplification with additional cycles, the PCR was executed on a MultiGene™ OptiMax Thermal Cycler (Labnet International, USA) to incorporate sample-specific Illumina® Nextera XT indices (Illumina®, Cambridge, UK), following these steps: 95 ∘C for 3 min (one cycle), 95 ∘C for 30 s, 55 ∘C for 30 s, 72 ∘C for 30 s (10 cycles) and 72 ∘C for 5 s (one cycle).

Normalisation of samples to equimolar concentrations was done manually, as PCR product yields were too low for SequalPrep™ Normalization Plate Kit. Amplicon concentration (nM) for all peaks in the correct region of the PCR product (596 to 671 bp) was measured with the peak calling function on the QIAxcel® Advanced System (1:3 diluted samples). For each sample, all amplicon concentrations for correct peaks were summed and converted to the original concentration (nM) before 1:3 dilution. Sample concentrations (nM) were put into Excel and the volume of sample that was required to reach 1 nM was calculated, the total pool balancing out variable sample DNA concentrations. The samples were pooled according to the calculations into a single 1.5 ml Eppendorf tube. The pooled sample was cleaned with a SPRIselect™ bead-based reagent (Beckman Coulter, Inc., USA), following manufacturer's protocol. A ratio of 0.75% SPRI beads to sample was used to remove a 250 bp contamination that was visible on the QIAxcel® Advanced System. The purified pooled sample was eluted in 100 µL of nuclease-free water and this was sequenced on an Illumina® MiSeq (nano cartridge: 2 × 250 bp paired-end reads) at Cardiff School of Bioscience Genome Research Hub. Next generation sequencing data used for the molecular analysis can be found on the European Nucleotide Archive (ENA), project code: PRJEB75951.

Bioinformatics

The bioinformatics programme QIIME2 (version 2021.8) (Bolyen et al., 2019) was used to analyse the sequence data, following methods from Hooper (2023a). Fastq files for paired-end sequences were demultiplexed and quality controlled before they

were imported into QIIME2 for further processing with the in-built DADA2 pipeline (Callahan et al., 2016), which merges paired-end reads, denoises the data, removes artifacts (chimeric sequences) and obtains representative sequences of original samples. Table 3.5 contains parameters used in the DADA2 pipeline. Truncation length was selected based on quality scores for forward and reverse reads in both 16S rRNA and *rbc*L sequencing runs individually. Amplicon sequence variants (ASVs) were taxonomically assigned using compiled databases with reference sequences for 16S rRNA and *rbc*L. The silva database v138 (Callahan, 2018) was used to extract the 16S rRNA variable regions V4 & V5 (515 to 806 bp) and associated taxonomic classification. The database for *rbc*L subunit sequences and associated taxonomy was collated for diatoms (*Bacellariophyceae*) and green algae (*Chlorophyta*), explained in more detail in Hooper (2023a). In brief, diatom data was sourced from an open-access curated barcode library Diat.barcode, and this was merged with GenBank data for all *rbc*L subunit sequences available, without duplicate or primer-containing data entries and cross-referenced against NCBI taxonomic identifiers to provide complete taxonomic classification. The 16S rRNA silva database v138 and the compiled *rbc*L database were adapted to QIIME2 version 2021.2, so they were regenerated for use in QIIME2 version 2021.8. ASV assignment was done in QIIME2 (version 2021.8 - https://view.qiime2.org/) using the fit-classifier naive-bayes function to generate a Bayesian classifier. After ASV assignment, certain filters were applied to clean the data. For 16 rRNA and *rbc*L sequences, chloroplasts were removed from the dataset, because they don't represent real DNA sequences. Subsequently, samples that had reads below 2000 were filtered out for 16S rRNA, to remove the blanks with low and non-specific bacterial species reads. For *rbc*L the threshold was set to 1000 because the level of contamination with *rbc*L DNA is smaller than with 16S rRNA. These values have been determined based on DADA2 total number of reads (non-chimeric) for the highest scoring blank (Table 3.5). Moreover, proportional filtering was done per ASV (taxa) using the qiime feature-table filter-features-conditionally function with a -p-abundance of 0.001 for 16S rRNA and 0.0006 for *rbc*L. This means that ASV's with a minimum relative abundance under 0.1% for 16S rRNA (minimum number of reads for species identification between 10 and 29) and 0.06% for *rbc*L (minimum number of reads between 16 and 34) respectively, were removed (see Table 3.5) to ensure accurate species identification (ideally 10-20 reads). P prevalence was set to 0, as

sampling locations and conditions were different within the sequencing run and this prevented unique ASV's only present in one sample from being removed.

Table 3.5. Parameter settings that were used for bioinformatics and summarised information from the DADA2 table and bioinformatics filtering steps.

Parameter	16S rRNA	*rbc*L
--p-trim-left-f	19	23
--p-trim-left-r	20	21
--p-trunc-len-f	248	248
--p-trunc-len-r	248	248
--p-abundance	0.001	0.0006
Average reads in samples (from DADA2 table)	28808	34290
Average no of reads with p abundance	29	21
Minimum no of reads with p abundance	10	16
Maximum no of reads with p abundance	63	34
--p-min-frequency	2000	1000
Blanks maximum number of reads	4740	898

3.3.10 qPCR

The diluted DNA plate (normalised to 0.5 ng/L) was used directly for *geo*A/*mic* and 16S rRNA qPCRs. The qPCR reactions were done in triplicate in 384-well plates including an internal standard curve for *geo*A and *mic* as well as qPCR blanks. Exact qPCR conditions are described below, following protocol developed by Professor Peter Kille's group, adapted from Hooper et al. (2023b).

Standard curves for geoA, mic and 16S rRNA

DNA extracted from cultures of *Oscillatoria Nigro-Viridis*, a geosmin and 2-MIB producing cyanobacteria, *Streptomyces Coelicolor*, a geosmin and 2-MIB producing streptomyces, and *Escherichia coli* was used to produce standard curves (Hooper et al., 2023b). *geo*A standard curves were generated from *Oscillatoria Nigro-Viridis* DNA using the primer set GeoA F (5'-ATGCAACCCTTTAAACTGCCAG–3') and GeoA R (5'–TTAGGGATTGGTAACTGGTGACTG–3'). Standard curves for *mic* were also generated with DNA extracted from *Oscillatoria Nigro-Viridis* using primer set MIB F (5'–ATGAAAGATACCAACCTGGATGATAC-3') and MIB R (5'-TTAGGTTGATGATTGTGAATCCATCTG–3'). To generate standard curves for 16S

rRNA, equimolar concentrations of DNA extracted from *Escherichia coli*, *Oscillatoria Nigro-Viridis* and *Streptomycytes Coelicolor* were used and amplified with primer set 27F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′), as described by DeLong (1992). All amplifications were performed on a SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific, USA) using the following conditions: initial denaturation at 95 ∘C for 2 min followed by 40 cycles of 95 ∘C for 30 s (denaturation), 60 ∘C for 30 s (annealing), 72 ∘C for 2 min (extension) and a final extension step at 72 ∘C for 5 min. The amplicons were cleaned and purified using a QIAquick PCR purification kit (Qiagen Ltd, Venlo, Netherlands) and the amplicon size was checked with a 4200 TapeStation (Agilent Technologies, USA), which should be 2291 bp, 1141 bp and 1490 bp for *geo*A, *mic* and 16S rRNA, respectively.

geoA and mic qPCR

The *geo*A and *mic* qPCR was executed in the same reaction well, using probes to identify them. The *geo*A qPCR was executed with primer Geo SGF1 (5'-CATCGAATACATCGAGATGCG-3') and Geo JDR1 (5'-TCGCCTTCATCTTCCACTTC-3') using SGP5 probe FAM (5'-AGGTTGGTGGCGCACCCTGGTCA-3'), as described in (Tsao et al., 2014). qPCR for the *mic* gene was performed using primers MIBS02F (5'-ACCTGTTACGCCACCTTCT-3') and MIBS02R (5'- CCGCAATCTGTAGCACCATG-3') with a MIB probe HEX (5'- ACGACAGCTTCTACACCTCC-3'), adapted from (Chiu et al., 2016). A hydrolysis probe qPCR was done to combine *geo*A and *mic* qPCR reactions in a 10 µL reactions mixture containing 5 µL of QuantiNova probe qPCR Master Mix (2x) (Qiagen Ltd, Venlo, Netherlands), 0.4 µM of each primer, 0.2 µM of each probe, 0.05 µL of QN ROX reference dye and 2 µL of template DNA. The qPCR reactions were run on a QuantStudio™ 7 Flex Real-Time PCR System, with 384-well plates (Applied Biosystems, ThermoFisher Scientific™, USA) and the following conditions were used: initial denaturation at 95 ∘C for 2 min followed by 40 cycles of 95 ∘C for 5 s (denaturation) and 60 ∘C for 30 s (annealing and extension) which allowed fluorescence capture. The linear dynamic range of copy number detection in this qPCR was 10 and 10 x 10^6 copies µL^{-1}.

16S rRNA qPCR

qPCR for 16S rRNA was done with primers 534F (5′- GCCAGCAGCCGCGGTAAT-3′) and 907R (5′-CCGTCAATTCCTTTGAGTTT-3′), as described in Muyzer et al. (1993) and Muyzer (1998). A SYBR Green qPCR was performed in a 10 µL reactions mixture containing 5 µL of QuantiNova probe qPCR Master Mix (2x) (Qiagen Ltd, Venlo, Netherlands), 0.7 µM of each primer, 0.05 µL of QN ROX reference dye and 2 µL of template DNA. The qPCR reactions were executed on a QuantStudio™ 7 Flex Real-Time PCR System, in 384-well plates (Applied Biosystems, ThermoFisher Scientific™, USA) using the following conditions: initial denaturation at 95 ∘C for 2 min followed by 40 cycles of 95 ∘C for 5 s (denaturation) and 60 ∘C for 20 s (annealing and extension) which allowed fluorescence capture, followed by a melting curve analysis of the amplified products: 95 ∘C for 15 s (denaturation), 60 ∘C for 1 min (annealing) and 95 ∘C for 15 s (dissociation, temperature change of 0.05 ∘C/s). The linear dynamic range of copy number detection in this qPCR was 10 and 10×10^6 copies µL^{-1}.

3.3.10.1 Blanks

The Milli-Q filter blank that was executed during the Sterivex filtering was checked for 16S rRNA gene copies. 84746.4 gene copies of 16S rRNA per ml of filtered water were present on average for all the samples, whereas the Milli-Q filter blank contained 1.5 gene copies per ml of filtered water. As the 16S rRNA gene copies in the blank only make up on average 0.0018 % of the total 16S rRNA gene copies, it was chosen to not take the blank into further consideration. Additional blanks were executed on the qPCR plates for 16S rRNA and *geo*A and *mic* to check for cross-contamination.

3.4 Statistical analyses

3.4.1 Estimation of carbon produced per hour

Manual oxygen monitoring

Mean air saturation % and mean temperature was calculated for each flask. For each measurement, the hours since the start of the experiment (04/07/2022 09:58:00) was calculated. Rounded mean temperature was used to convert air saturation % into mg/L, with Equation 3.4 and the correct oxygen calculation ratio (at 0 salinity) that matched the temperature: 18 °C = 9.467, 19 °C = 9.276, 20 °C = 9.092, 21 °C = 8.914, 22 °C = 8.743, 23 °C = 8.578. Air pressure was not considered in calculations as

experiments took place in a laboratory facility and was assumed to have a negligible effect.

Equation 3.4:

$$Oxygen\left(\frac{mg}{L}\right) = \frac{Air\ saturation\ \%}{100} * oxygen\ ratio\ at\ given\ temperature$$

Dissolved oxygen (mg/L) was converted to productivity in mmoles of carbon in hours and minutes, assuming a 1:1 molar ratio in photosynthesis (1 C + 1 O_2), and using exact flask volumes throughout the experiment. The area under the curve calculation was used assuming that productivity occurred uniformly between each time interval (in hours or minutes) and results were summed to calculate the total carbon (mmoles), which resulted in mmol carbon in hours and mmol carbon in minutes.

Continuous oxygen monitoring probes

Continuous oxygen measurements every 15 minutes were first processed with R packages "zoo" (Zeileis & Grothendieck, 2005) and "xts" (Ryan & Ulrich, 2011), followed by calculating mmol carbon produced at each timestep with a constant temperature of 20 degrees in the incubator (9.092 conversion factor) and flask volume decrease, similar to the section "Manual oxygen monitoring" above. This data was interpolated using function f in R studio (approxfun) to create constant 15-minute time steps. The interpolated dataset was used to calculate the area under the curve, using R package "pracma" (Borchers, 2023), which resulted in mmoles of carbon in minutes. Calculated total carbon (mmoles in minutes) from continuous and manual oxygen sensors was comparable (Table S.3 in Appendix B).

3.4.2 Molecular data

Raw read counts at level 6 (genus) were downloaded from Qiime for 16S rRNA and *rbc*L separately. A normalisation was performed, dividing raw read counts of each ASV by the total number of reads for that sample, to determine the proportion of that ASV. This resulted in values for each ASV between 0 and 1 (0.1 indicates 10% of total reads). This normalised proportional abundance is also referred to as the relative abundance and was calculated for 16S rRNA and *rbc*L separately, as they are completely different datasets.

Raw ASV read counts were used to calculate normalised counts for 16S rRNA and *rbc*L datasets. Raw ASV read counts for each sample were multiplied by the unique

dilution factor, which was different for each sample. Additionally, normalised counts were calculated for cyanobacteria, green algae and diatoms as individual groups. The raw ASV read counts of all species within each group were summed and this was also multiplied by the unique dilution factor. This resulted in the parameters normalised_green_counts, normalised_diatom_counts and normalised_cyano_counts, which is an estimate of abundance of these phytoplankton groups in the sample. These abundance estimates should be used with caution as they are not a biomass measurement and the raw ASV read counts could be influenced by the effectiveness of the DNA extraction technique.

A list was created with all cyanobacteria genera found in the 16S rRNA sequencing indicating for each genus whether they were 'Yes, likely producer (Y)' or 'Maybe producer (M)' for 2-MIB and geosmin separately. This list was based on Watson et al. (2016) and an overview made by professor Peter Kille by searching for geosmin and 2-MIB synthase genes in cyanobacterial genomes using 'Basic Local Alignment Search Tool' (BLAST, NCBI). The list provides an overview of cyanobacteria genera that can potentially produce 2-MIB and/or geosmin.

The results from the qPCR were used to calculate gene copy numbers for *geo*A, *mic* and 16S rRNA for each sample, by normalising the mean of triplicates on the plate by individual dilution applied on that sample. Then the copy number per ml of water filtered was calculated by dividing the copy number by the filtered volume on the Sterivex filter units, which was 250 ml in this experiment.

qPCR copy numbers of *mic* were divided by the copy numbers of 16S rRNA to normalise the gene copies of *mic* over the total copies of bacterial DNA (16S rRNA). This result was transformed for further analysis using Equation 3.5:

Equation 3.5:

$$Arcsine\left(\sqrt{\frac{mic\ copy\ nr.}{16S\ rRNA\ copy\ nr.}}\right)$$

The amount of 2-MIB produced per unit of biomass was calculated for each flask using relative abundance of 2-MIB producers (Y = Yes, likely producers; from the predefined list) and the following Equation 3.6:

Equation 3.6:

$$\frac{2 - MIB\ concentration\ (\frac{ng}{L})}{Rel.abundance\ 2 - MIB\ producers\ (Y) * 16S\ rRNA\ copy\ nr.per\ ml}$$

In both 16S rRNA and *rbc*L sequencing datasets, the genus name for many ASV's was unknown or uninformative. When this was the case, the next best taxonomy level was used with an indication of which level this was; (g): genus; (f): family; (o): order; (c): class; (p): phylum and (k): kingdom, kingdom being the least detailed level. Stacked bar charts were created for 16S rRNA and *rbc*L separated by Phylum and Genus. To enable better legend visualisation, the relative abundance for all samples (start, ctrl, a, b, c, d, e, f, all in triplicate) was averaged. All genus (level 6) with average relative abundance <0.01% for 16S rRNA and <1% for *rbc*L, were combined in a category "Other", which were 39 taxa for 16S rRNA and 76 taxa for *rbc*L.

Rstudio was used (R version 4.2.1: 2022-06-23 ucrt) to perform Non-Metric Multi-Dimensional Scaling (NMDS) on the 16S rRNA and *rbc*L sequencing datasets, normalised to relative abundances. The metaMDS function from R package "vegan" (vegan 2.6-4) (Oksanen et al., 2024) with a Bray-Curtis distance matrix was applied to the data to determine the ordination. The result was plotted in the R package "ggplot2" (Wickham, 2016) and each datapoint was shaped by treatment and coloured by low, medium or high 2-MIB concentration. 2-MIB categories (referred to as 2-MIB group) were defined on final day concentrations (not the difference between start and end) as follows: Low = < 50 ng/L; Medium = 50 – 150 ng/L and High = > 150 ng/L. The category low was indicative of no 2-MIB production over the course of the experiment, because the start concentration in all flasks was ±50 ng/L. To determine whether the three 2-MIB categories were different from each other, ellipses were drawn around the 95% interval of the sample groups and statistical differences were investigated using PERMANOVA and ANOSIM tests.

The envfit function from the "vegan" package was used on the NMDS results, to visualise other parameters that could be responsible for underlying patterns in the data. Only statistically significant parameters, based on the envfit output (pvalue ≤ 0.05), were included in the final plot. Parameters that were included were: mic/16S (*mic* copy number / 16S rRNA copy number), nutrient concentrations in mg/L at the start of the experiment (NO_3^--N, TDN, TDP, NH_4^+-N), total carbon (mmoles in hours),

dissolved oxygen (DO) at peak (maximum % air saturation) and normalised counts of cyanobacteria, diatoms and greens. Another envfit function was run with all cyanobacteria species, of which statistically significant species (pvalue ≤ 0.05) were selected and plotted on the NMDS plot.

Indicator Species Analysis was performed on the 16S rRNA and *rbc*L data, using R package "indicspecies" (De Cáceres & Legendre, 2009) using the multipatt command. For 16S rRNA data, relative abundance of statistically significant cyanobacteria (pvalue ≤ 0.05) was plotted as boxplots, grouped by 2-MIB categories.

3.5 Results

3.5.1 Impact of nutrient ratios on 2-MIB production

Target and actual nutrient ratios as well as concentrations of nutrients in the flasks at the start of the experiment are shown in Table 3.2, Table 3.6, Figure 3.3, Figure 3.4 and Appendix B Figure S.2. The actual nutrient ratios were close to the target ratios for NH_4^+:NO_3^- and the high SRP treatments (A, B, C) for DIN:SRP, but the low SRP treatments (D, E, F) seemed less consistent, which was most likely due to pipetting error or measurement error at such low concentrations. The focus of the experiment was on nutrient ratios with environmentally relevant concentrations. NO_3^--N is generally an order 10 higher than NH_4^+-N in freshwater systems (see Chapter 1, Section 1.2.1), which was also confirmed with water quality data from Reservoir 1. Implementing this environmentally relevant concentration difference in the laboratory microcosm experiments, meant that the concentration of total dissolved nitrogen (TDN) for high NH_4^+-N treatments was much lower than for high NO_3^--N treatments (Table 3.6).

Nutrient uptake was calculated as the % of change between start and end of the experiment (Table 3.7, Figure 3.3, Figure 3.4 and Appendix B Figure S.2 and Table S.4). Both treatment A (HNO3_LNH4_HPO4) and D (HNO3_LNH4_LPO4) did not deplete the nitrogen pool (Table 3.7 and Figure 3.3). The % change of NH_4^+ was largest in treatments C (LNO3_HNH4_HPO4) and F (LNO3_HNH4_LPO4), which started with ±0.8 mg/L NH_4^+-N (Table 3.6) and therefore had more available to consume. Treatment A (HNO3_LNH4_HPO4) was the only treatment at the end of the experiment that did not deplete the phosphorus pool (Table 3.7 and Figure 3.4). DOP was taken up by every treatment, but D (HNO3_LNH4_LPO4) had the largest % change (Table 3.7).

Geosmin was detected in the start community at low concentrations (<5 ng/L) and higher concentrations (>10 ng/L) were found in some flasks, but it was not produced consistently throughout the experiment (Appendix B Table S.4). The *geo*A gene was not accurately detected in any of the samples (data not shown). Hence, geosmin has not been considered in further analysis. 2-MIB was produced in all treatments, except for the control where it was even degraded (Table 3.7). Treatment D (HNO3_LNH4_LPO4) had the highest 2-MIB production of 521.4 ng/L but the large standard deviation of 551.8 ng/L indicates that this was variable between the replicates (Table 3.7). Treatment A (HNO3_LNH4_HPO4) and B (MNO3_MNH4_HPO4) also revealed high 2-MIB production of 164.7 (±95.4) ng/L and 251.4 (±213.9) ng/L.

N:P and NH_4^+:NO_3^- ratios both had a significant impact on 2-MIB concentration (nested ANOVA $\log_{(10)}$(MIB+1), N:P with p = 0.00706, NH_4^+:NO_3^- with p= 0.01026). There was a significant difference between treatments and the control, and differences between ratios (Figure 3.5, Kruskall-Wallis test of variance). Treatments with a low NH_4^+:NO_3^- ratio were significantly different from the control and a high NH_4^+:NO_3^- ratio, but not from a medium NH_4^+:NO_3^- ratio. Treatments with a low N:P ratio were significantly different from the control but not from a high N:P ratio (Figure 3.5, Kruskall-Wallis test of variance).

Table 3.6. Target NH$_4^+$:NO$_3^-$ and DIN:SRP ratios for each treatment, average (mean) nutrient concentrations and standard deviation (SD) from laboratory measurements for treatment triplicates at the start of the experiment (day 0), with actual NH$_4^+$:NO$_3^-$ and DIN:SRP ratios.

Treatment	Treatment ID	Target NH$_4^+$:NO$_3^-$	NH$_4^+$:NO$_3^-$ Mean	SD	TDN (mg/L) Mean	SD	NH$_4^+$-N (mg/L) Mean	SD	DON (mg/L) Mean	SD	NO$_3^-$-N (mg/L) Mean	SD
Ctrl	No Addition	x	0.323	0.021	0.733	0.021	0.090	0.007	0.352	0.016	0.279	0.002
A	HNO3_LNH4_HPO4	0.01	0.016	0.003	9.440	0.084	0.133	0.026	0.813	0.143	8.482	0.047
B	MNO3_MNH4_HPO4	0.1	0.100	0.011	2.242	0.015	0.161	0.018	0.460	0.018	1.608	0.029
C	LNO3_HNH4_HPO4	1	0.989	0.049	2.060	0.071	0.808	0.013	0.419	0.126	0.820	0.057
D	HNO3_LNH4_LPO4	0.01	0.011	0.001	9.997	0.144	0.097	0.011	1.006	0.250	8.879	0.221
E	MNO3_MNH4_LPO4	0.1	0.093	0.005	2.252	0.090	0.149	0.005	0.490	0.048	1.599	0.047
F	LNO3_HNH4_LPO4	1	0.988	0.023	2.044	0.017	0.793	0.031	0.437	0.066	0.802	0.023

Treatment	Treatment ID	Target DIN:SRP	DIN:SRP Mean	SD	TDP (mg/L) Mean	SD	SRP (mg/L) Mean	SD	DOP (mg/L) Mean	SD
Ctrl	No Addition	x	NA	NA	0.012	0.002	0.000	0.000	0.013	0.001
A	HNO3_LNH4_HPO4	9	8.972	0.044	2.247	0.010	2.127	0.002	0.120	0.011
B	MNO3_MNH4_HPO4	9	9.153	0.083	0.466	0.001	0.431	0.002	0.035	0.001
C	LNO3_HNH4_HPO4	9	9.448	0.335	0.420	0.002	0.384	0.004	0.035	0.002
D	HNO3_LNH4_LPO4	128	143.428	1.638	0.161	0.002	0.139	0.002	0.022	0.001
E	MNO3_MNH4_LPO4	128	188.997	9.671	0.036	0.004	0.021	0.001	0.015	0.003
F	LNO3_HNH4_LPO4	128	191.029	26.515	0.033	0.003	0.019	0.003	0.014	0.000

Table 3.7. Average nutrient % change for each treatment between start and end of the experiment, as well as average concentration of 2-MIB (ng/L) produced (positive) or degraded (negative) between start and end of the experiment. Colour scale in mean % change columns indicate (negative) change 0-25% = white, 25%-50% = light green, 50%-75%= mid green, 75%-100% = dark green.

Treatment	Treatment ID	TDN % change		NH$_4^+$-N % change		DON % change		NO$_3^-$-N % change	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ctrl	No Addition	-39.04	12.73	-3.80	35.26	-20.35	31.21	-78.92	11.72
A	HNO3_LNH4_HPO4	-47.44	16.21	-70.25	8.72	-27.56	19.73	-50.87	18.67
B	MNO3_MNH4_HPO4	-82.38	0.61	-75.67	7.61	-39.66	2.11	-96.75	1.15
C	LNO3_HNH4_HPO4	-80.74	0.51	-95.59	0.21	-34.55	23.29	-91.46	5.76
D	HNO3_LNH4_LPO4	-48.96	5.81	-71.98	1.92	-40.99	10.36	-50.22	6.64
E	MNO3_MNH4_LPO4	-65.90	5.57	-71.25	3.26	-47.71	8.92	-71.93	5.97
F	LNO3_HNH4_LPO4	-71.82	9.39	-94.82	0.75	-41.88	11.11	-67.22	23.42

Treatment	Treatment ID	TDP % change		SRP % change		DOP % change		2-MIB production (ng/L)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ctrl	No Addition	-52.63	4.56	NA	NA	-39.05	3.30	-20.0	17.0
A	HNO3_LNH4_HPO4	-63.59	23.17	-64.35	23.60	-50.37	16.39	164.7	95.4
B	MNO3_MNH4_HPO4	-90.35	8.32	-93.76	8.44	-48.91	8.46	251.4	213.9
C	LNO3_HNH4_HPO4	-91.53	9.37	-94.51	8.85	-57.03	12.94	64.7	99.0
D	HNO3_LNH4_LPO4	-95.65	0.04	-100.00	0.00	-61.03	4.50	521.4	551.8
E	MNO3_MNH4_LPO4	-88.79	1.26	-100.00	0.00	-55.62	7.51	17.4	41.0
F	LNO3_HNH4_LPO4	-87.60	1.24	-100.00	0.00	-48.15	0.00	6.7	41.7
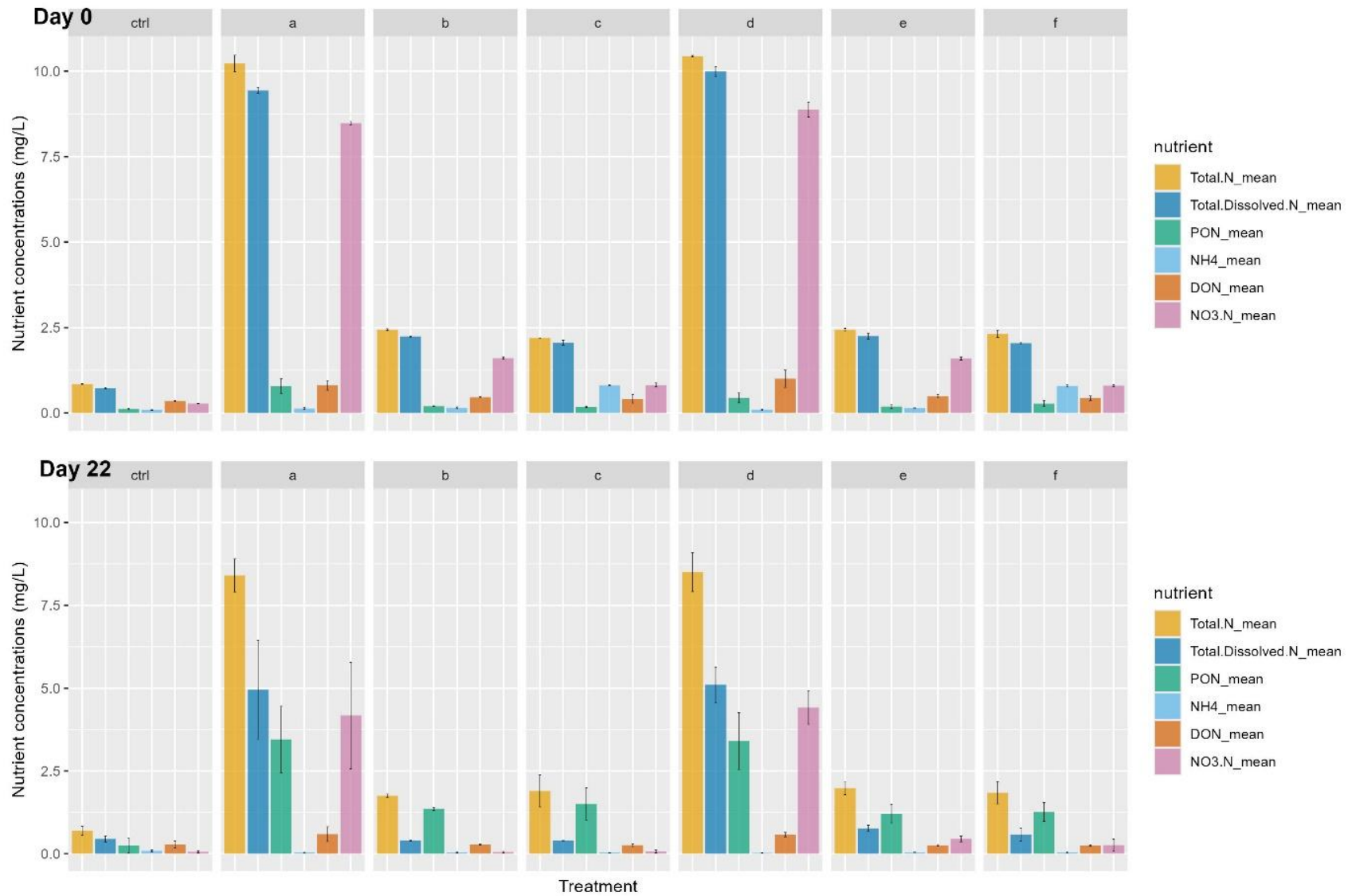
Figure 3.3. Concentrations of nitrogen fractions (mg/L) at the start (day 0) and the end of the experiment (day 22) averaged by treatment, with standard deviation indicated by error bars.
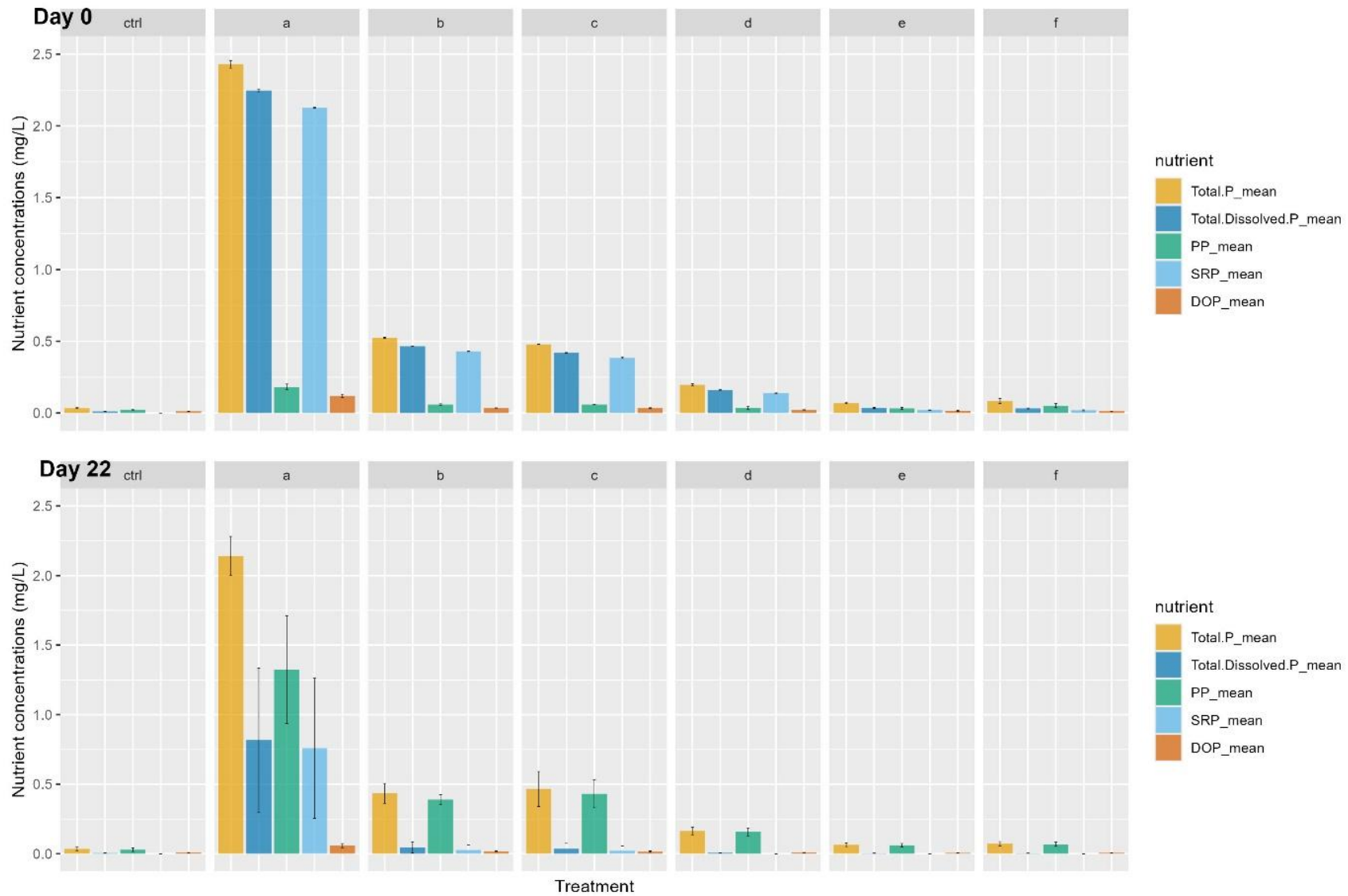
Figure 3.4. Concentrations of phosphorus fractions (mg/L) at the start (day 0) and the end of the experiment (day 22) averaged by treatment, with standard deviation indicated by error bars.
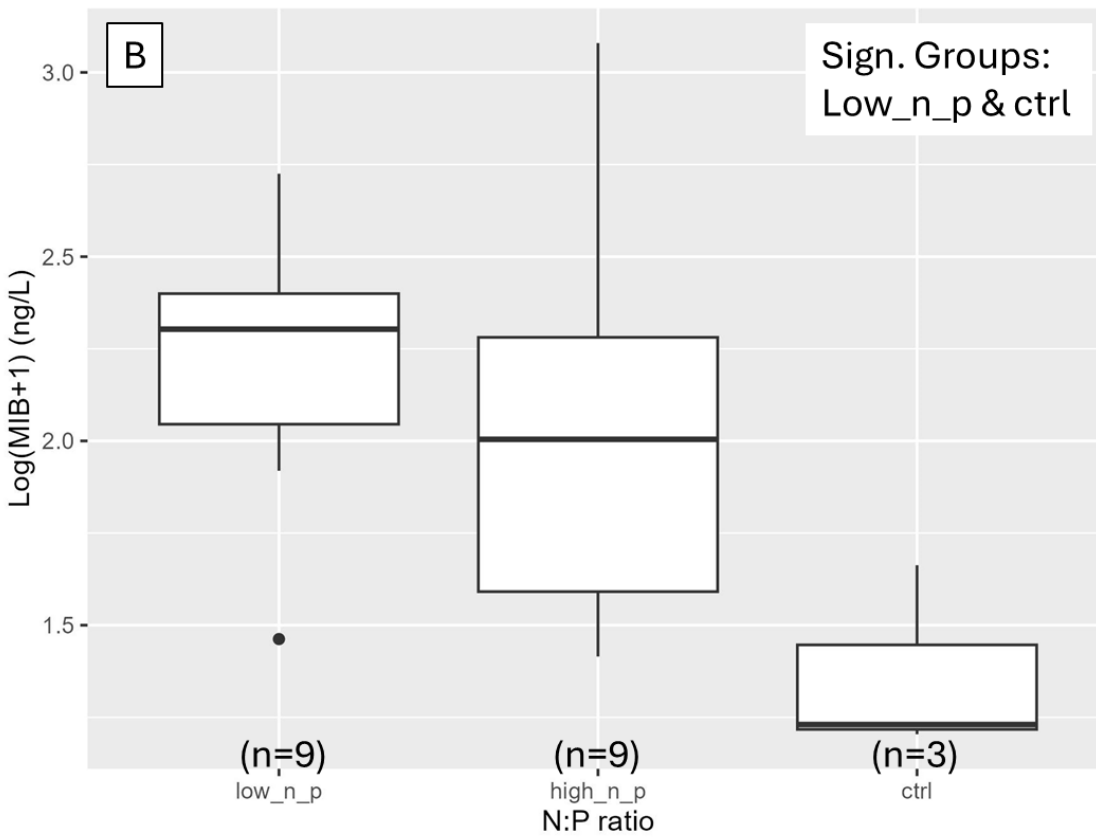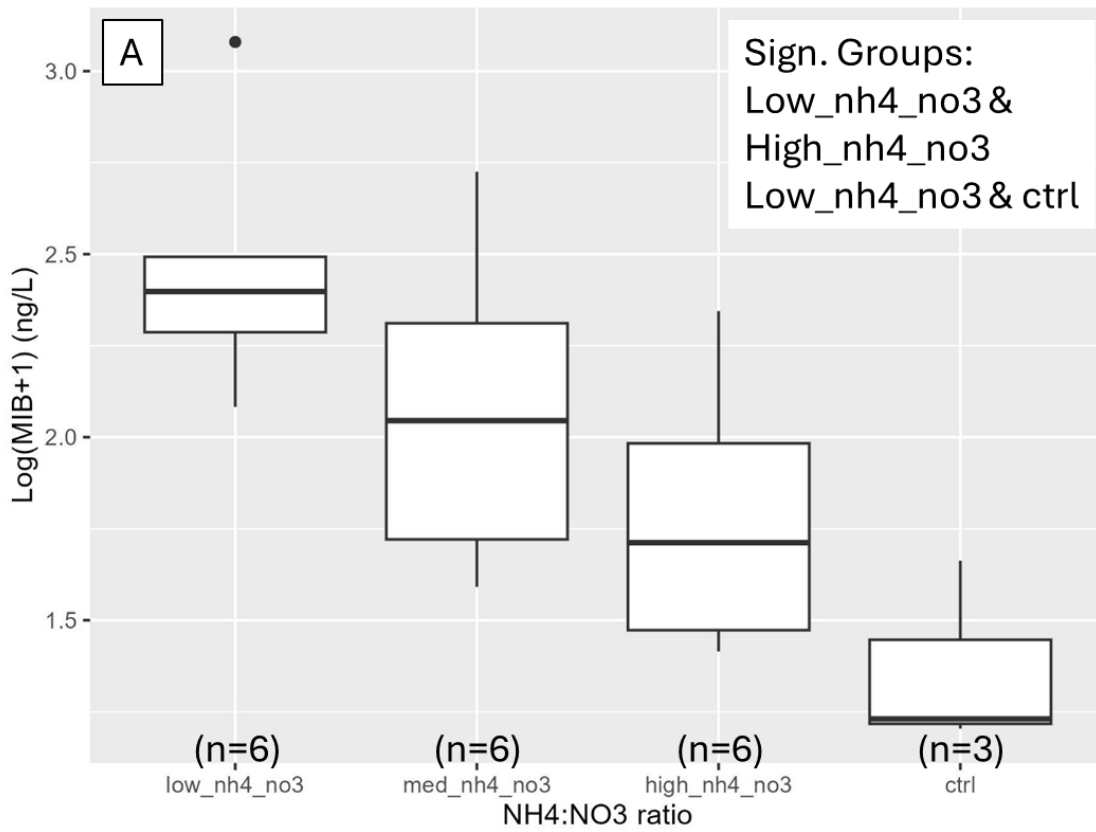
Figure 3.5. Boxplots showing the effect of $NH_4^+:NO_3^-$ (a) and N:P ratio (b) on $log_{(10)}(MIB+1)$, number of replicates in each group indicated by (n=…), significant groups from Kruskall-Wallis test of variance specified in top right.

Several treatments had significantly different concentrations of $\log_{(10)}(\text{MIB}+1)$; A and control, B and F, B and control, D and F, D and control (Figure 3.6). The low $NH_4^+:NO_3^-$ treatments had a large (8 mg/L) difference in TDN compared to medium and high $NH_4^+:NO_3^-$ treatments (Figure 3.6). This was a result of using the chosen ratios, whilst supplying environmentally relevant concentrations of NH_4^+-N and NO_3^--N; a NO_3^--N concentration of 10 mg/L is not uncommon in surface water, whilst a 10 mg/L concentration of NH_4^+-N would be toxic (Table 3.6). Treatment A, B and C for example all had a low N:P ratio, but A had a much higher P concentration than B and C, because it had to balance the amount of N that was added (Figure 3.6 and Table 3.6).
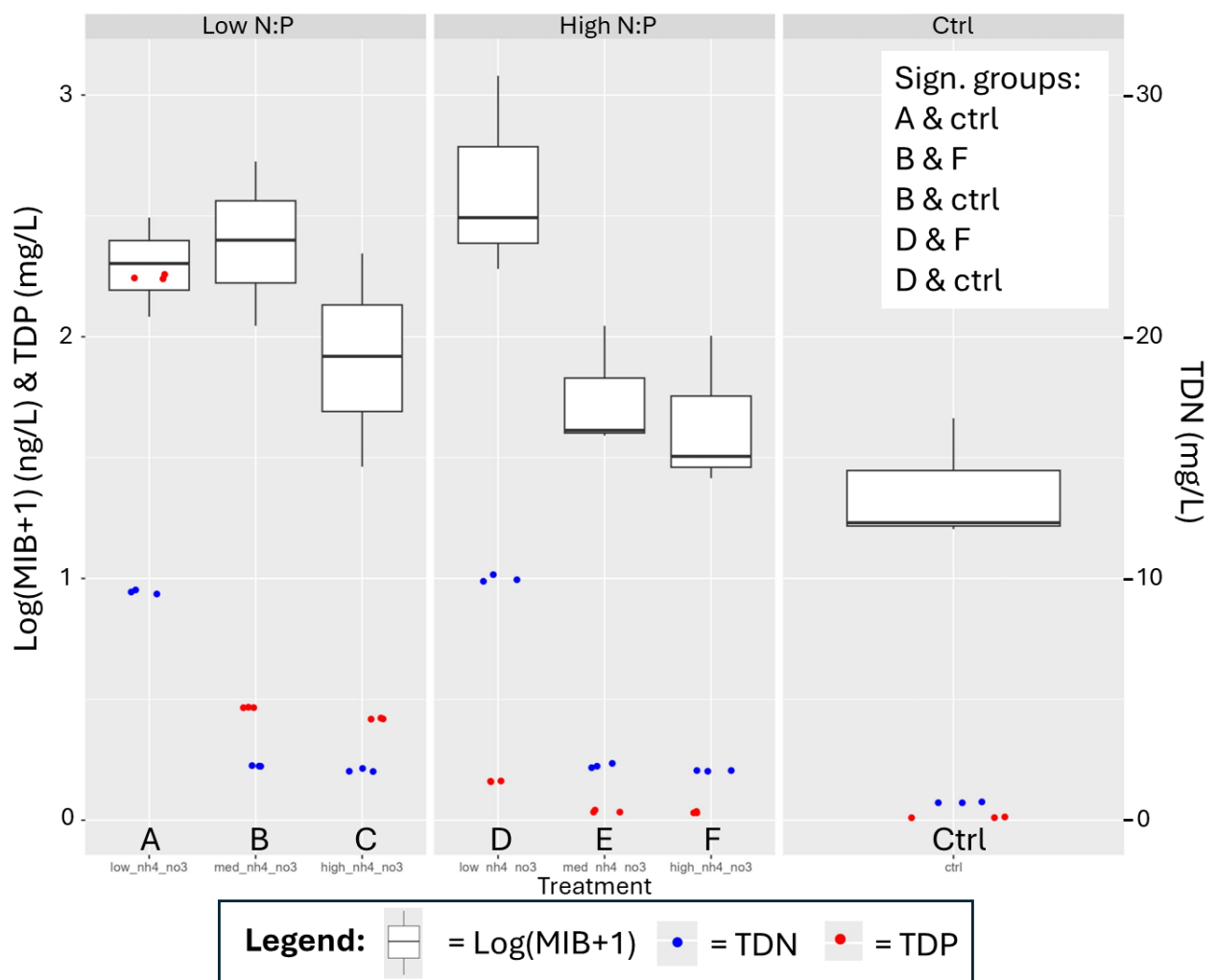
Figure 3.6. Boxplot shows the effect of each treatment on $\log_{(10)}(\text{MIB}+1)$, with triplicates for each treatment (n=3), blue scatter is total dissolved nitrogen (TDN) concentrations (mg/L) and red scatter is total dissolved phosphorus (TDP) concentrations (mg/L). Significant groups from Kruskall-Wallis test of variance specified in top right.

Treatments B and E have the same TDN, but B has 0.5 mg/L of TDP whilst E has 0.04 mg/L of TDP, with treatment B resulting in higher 2-MIB concentrations than E (Figure

3.6 and Table 3.7). The TDP concentration that was required for this community to produce elevated levels of 2-MIB was probably between 0.5 mg/L (treatment B) and 0.2 mg/L (treatment D). The concentration of TDN that was required for development of the 2-MIB producers in this experiment was probably around 2.5 mg N/L (treatment B), as long as sufficient TDP was available. This is supported by lower 2-MIB concentrations in treatment E, which had the same TDN concentration but low TDP (0.04 mg P/L).

3.5.2 Impact of nutrient ratios on community productivity

Dissolved oxygen concentrations, indicative of the productivity of the photosynthesizing phytoplankton community, revealed differences between treatments (Figure 3.7). Continuous oxygen measurements in one of the flasks of treatments A (HNO3_LNH4_HPO4), C (LNO3_HNH4_HPO4) and control, demonstrated differences in oxygen response depending on the nutrient ratios, and revealed a clear diurnal cycling of day and night. Treatment C, with high NH_4^+ and low NO_3^-, had increasing oxygen concentrations after four days, but the peak of 16 mg/L after 16 days was followed by a decrease in oxygen until the end of the experiment (Figure 3.7). At the same time, treatment A (high NO_3^- and low NH_4^+), had a slower response time of approximately eight days, but then steadily increased to 19 mg/L until the end of the experiment (Figure 3.7).

Manual oxygen measurements showed a similar pattern to the continuous sensors (Figure 3.7) although they were influenced by the time of day each flask was measured, according to diurnal oxygen dynamics. They did demonstrate that oxygen concentrations in the controls did not have a notable increase, whereas treated flasks responded according to treatment (Figure 3.7, panel B). Treatment E (MNO3_MNH4_LPO4) and F (LNO3_HNH4_LPO4) had a small increase in oxygen production compared to the start of the experiment (±9 mg/L), but on average the replicates did not exceed 12 mg/L of oxygen (Figure 3.7, panel B). Treatment B (MNO3_MNH4_HPO4) and C (LNO3_HNH4_HPO4) had a similar pattern, although B reacted a few days later to the nutrient spike it had a steeper oxygen increase than C, and they both showed an oxygen peak at day 16 which was 17 and 16 mg/L, for B and C respectively, followed by a steady decrease (Figure 3.7, panel B). Treatment A (HNO3_LNH4_HPO4) and D (HNO3_LNH4_LPO4) also had a similar pattern, although D had a steeper increase and seemed to plateau from day 19 onwards

whereas A did not reach a plateau (Figure 3.7, panel B). The general pattern for treatment A and D was a delayed oxygen increase but a steady increase in oxygen until the end of the experiment, reaching 19 and 18 mg/L, respectively (Figure 3.7, panel B).
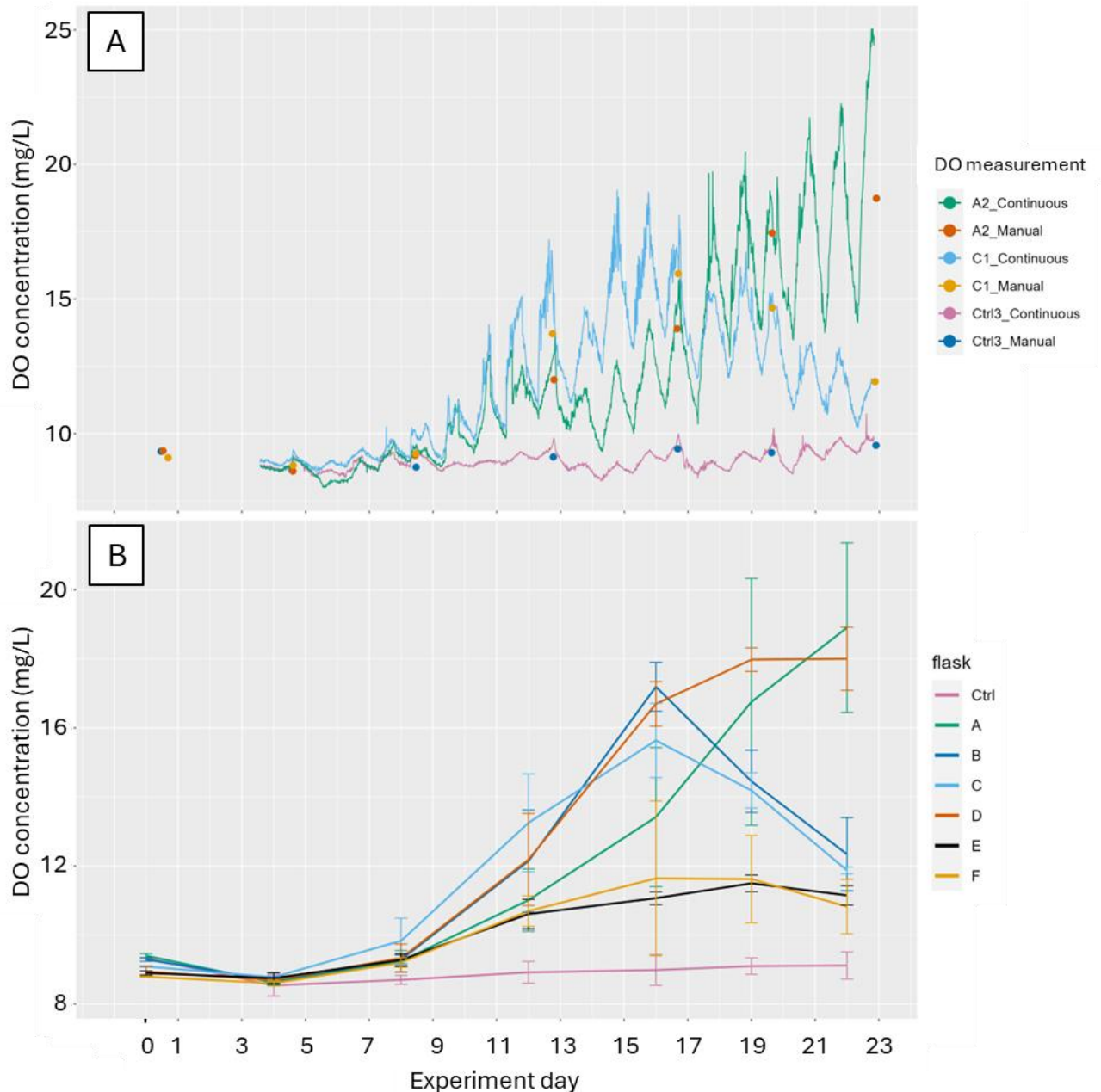
Figure 3.7. Dissolved oxygen concentration (mg/L) from continuous microsensor measurements in three selected flasks and the corresponding manual measurements with the Fibox sensor (a). Periodic dissolved oxygen measurements averaged for each treatment, error bars showing standard deviation (b).

Visual inspection of the flasks during the experiment and at the final day also gave indications of biomass and health of the community (Figure 3.8). Treatment C

(LNO3_HNH4_HPO4) was the first treatment to develop a strong green colour, indicating benthic community development at day 13 of the experiment (Figure 3.8). However, at the final day of the experiment, day 22, the biomass in treatment C had turned yellow whilst treatment A (HNO3_LNH4_HPO4) and D (HNO3_LNH4_LPO4) turned green (Figure 3.8). A green biomass colour does not necessarily indicate community health, but daily visual inspections alongside oxygen measurements did result in similar findings. The oxygen measurements and visual indicators suggested that treatment C (LNO3_HNH4_HPO4) exhibited 'boom-bust' dynamics, whereas treatments A (HNO3_LNH4_HPO4) and D (HNO3_LNH4_LPO4) had more balanced growth over the whole experiment.
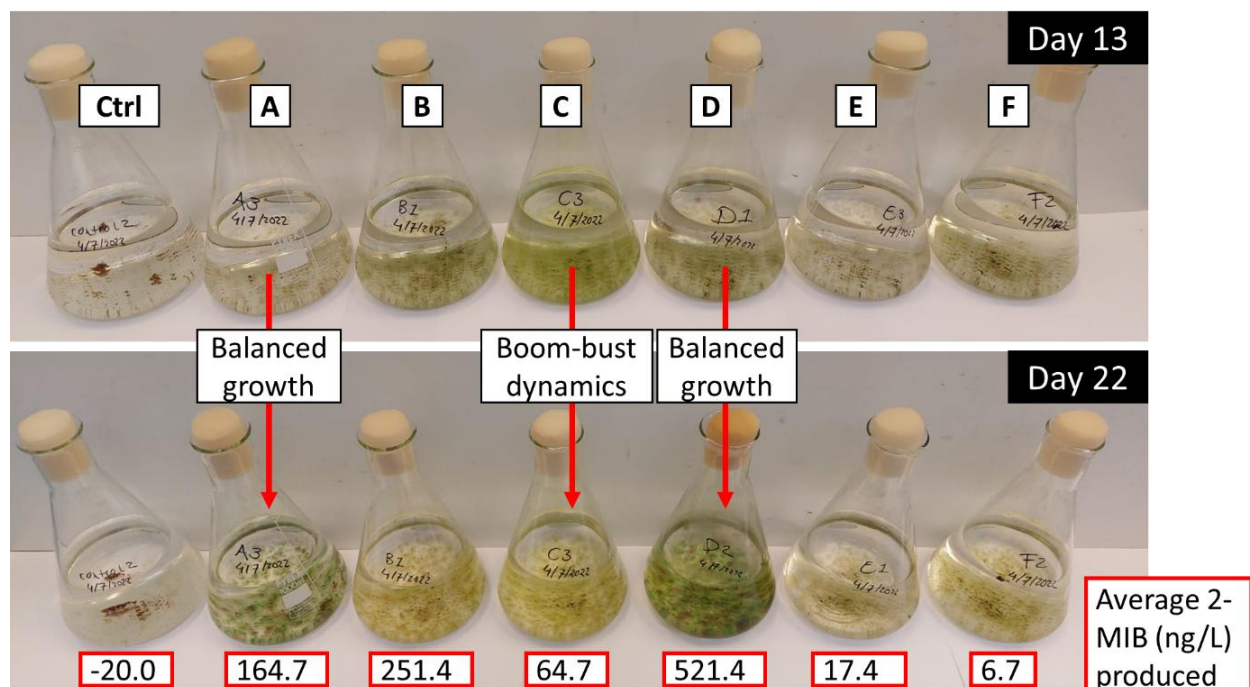
Figure 3.8. Images of all treatment flasks on day 13 and day 22, to enable visual inspection of biomass and health. Numbers at the bottom in boxes with red outline indicate average 2-MIB concentrations (ng/L) that were produced (final – start concentration) during the experiment.

Dissolved oxygen data was used to calculate the amount of carbon produced throughout the experiment (mmoles in minutes). The amount of carbon produced correlated well with $\log_{(10)}(MIB+1)$, yielding an R^2 of 0.608 in a linear model (Figure 3.9, panel A). Treatments A, B, C and D were all significantly different from the control treatment and treatment D was also significantly different from treatment E and F (Figure 3.9, panel B).
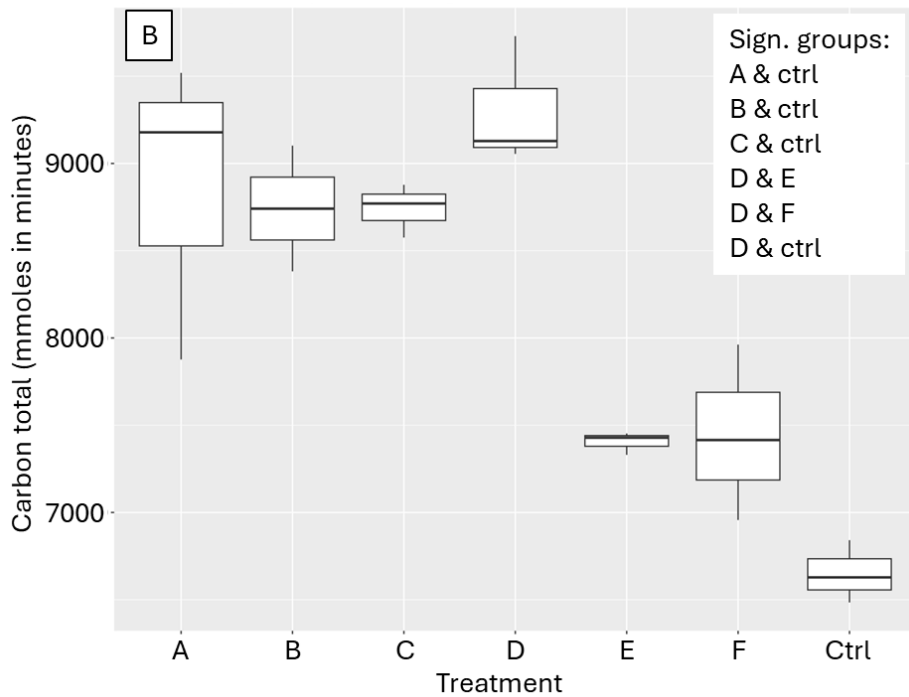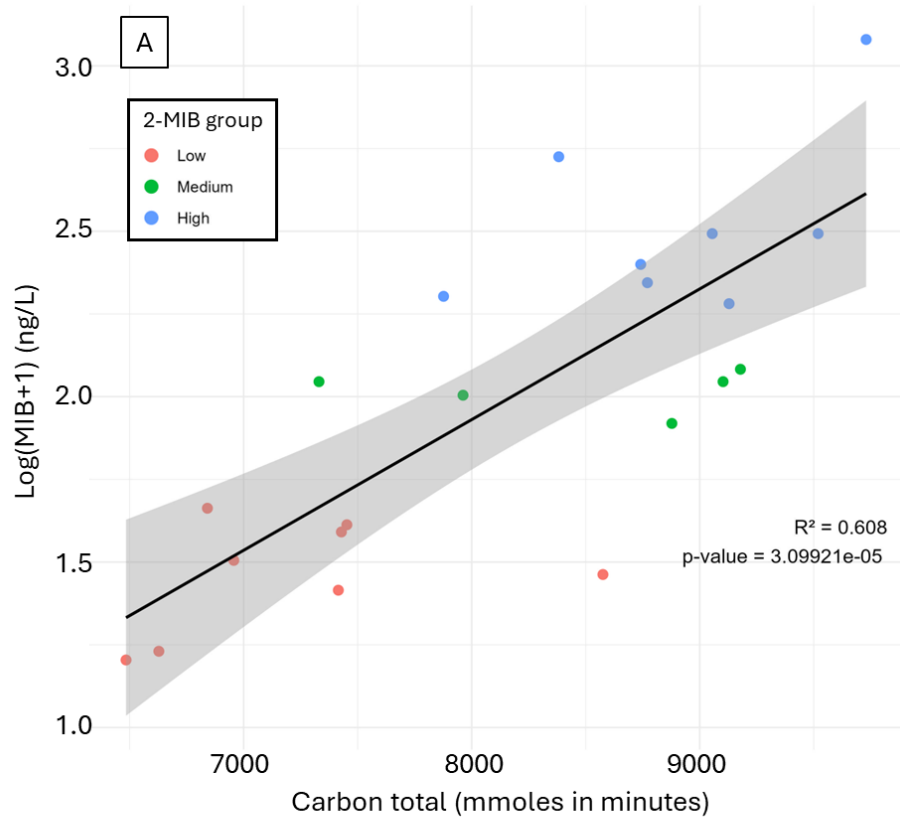
Figure 3.9. Linear model of carbon total (mmoles in minutes) and $\log_{(10)}$(MIB+1), with model performance in the plot and colour of datapoints indicate 2-MIB group (high, medium or low concentration) (a), boxplot of carbon total (mmoles in minutes) per treatment, with significant groups from Kruskall-Wallis test of variance specified in top right.

3.5.3 Impact of nutrient ratios on species composition

Nutrient ratios also influenced the relative abundance and biomass of species in the community. Normalised counts (ASV counts normalised for dilution factor) revealed that treatment A (HNO3_LNH4_HPO4) had by far the highest cyanobacteria biomass of all treatments, followed by treatment D (HNO3_LNH4_LPO4) (Figure 3.10). The biomass of diatoms was also highest in treatment A, although the differences between treatments was not as notable. Biomass of green algae was highest in treatment A, although there was high variability between the replicates of each treatment (Figure 3.11). Relative abundance for the different treatments showed the same pattern as biomass for cyanobacteria, but there was a less pronounced difference between treatment A and the rest (Figure 3.10). Relative abundance for diatoms and green algae showed the opposite pattern to one another as these groups are the only ones represented in *rbc*L sequencing data. Diatoms seemed to dominate treatments with lower nutrients, whilst green algae dominated in higher nutrient treatments and the controls (Figure 3.10 and Figure 3.11).
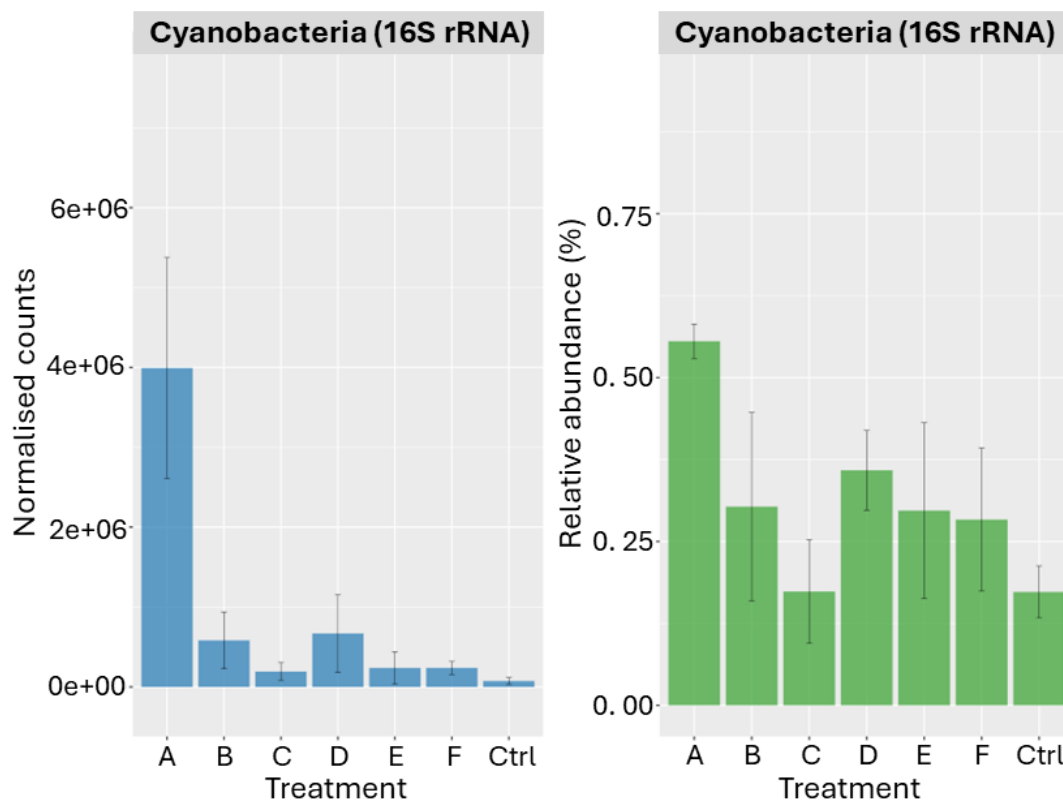
Figure 3.10. Cyanobacteria normalised counts and relative abundance (%; 1 = 100%), based on the 16S rRNA dataset.
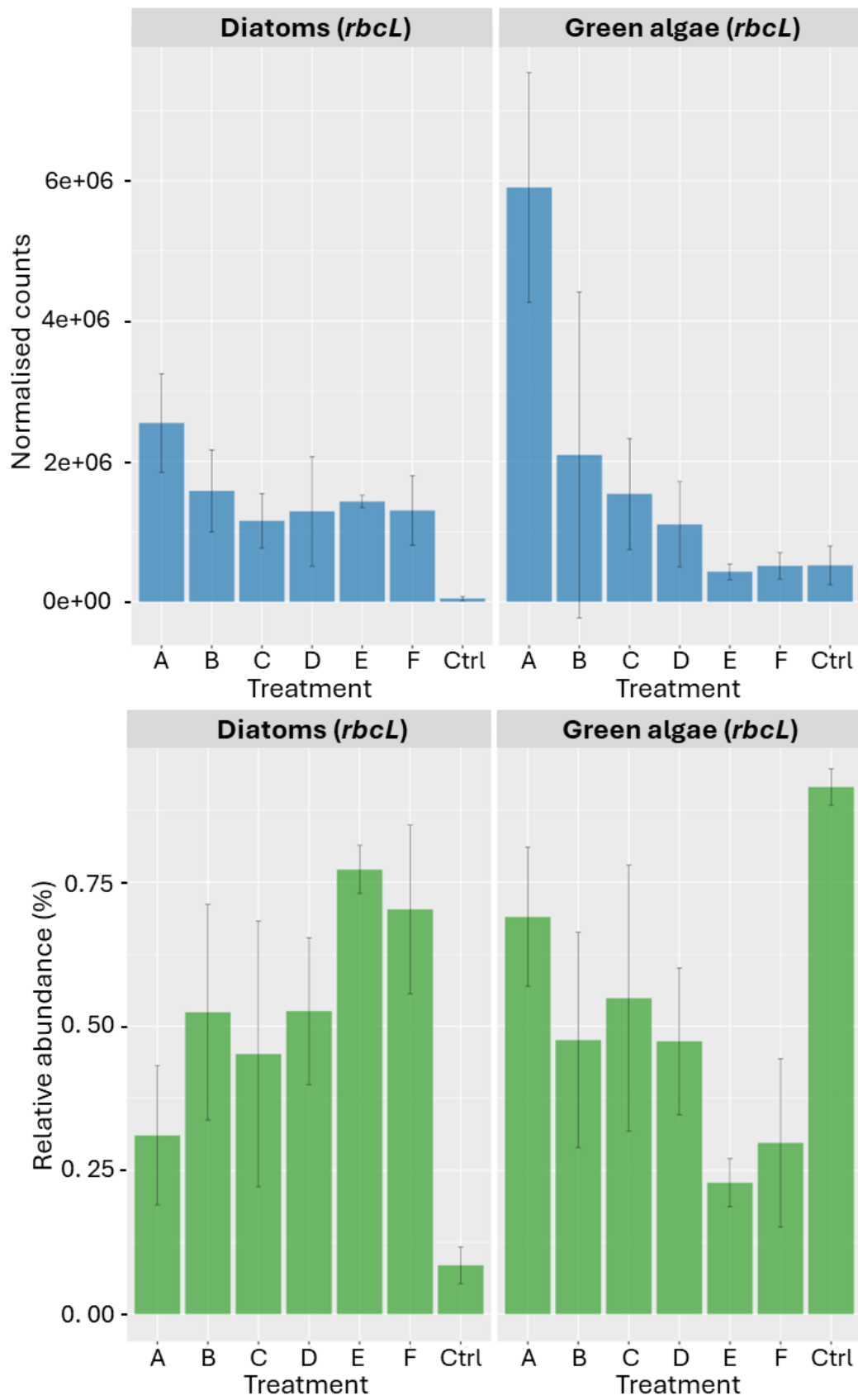
Figure 3.11. Diatoms and green algae normalised counts and relative abundance (%; 1 = 100%), based on the rbcL dataset.
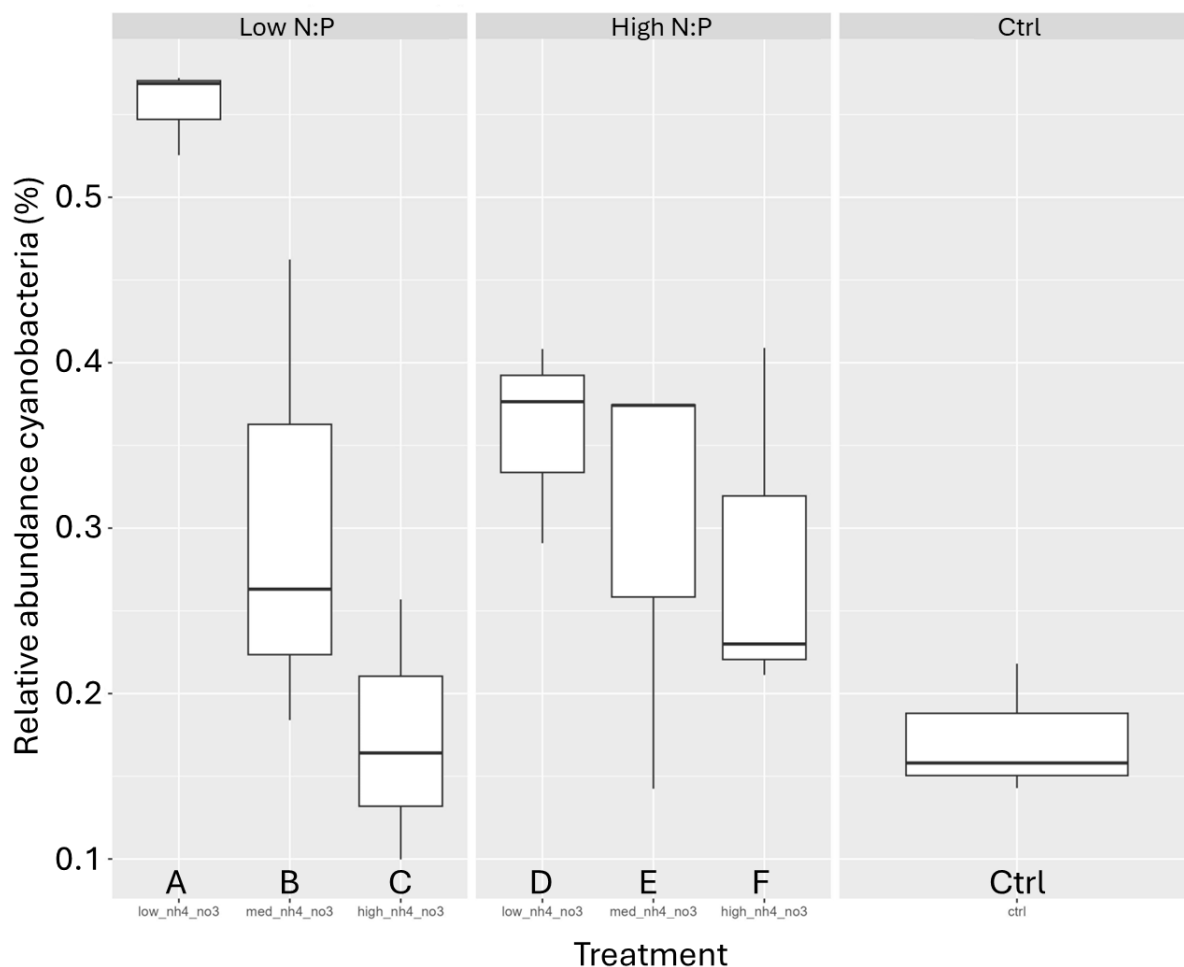
Figure 3.12. Boxplot showing effect of treatment on cyanobacteria relative abundance (%; 1 = 100%), based on the 16S rRNA dataset.

Relative abundance for phylum and normalised counts (ASV counts normalised for dilution factor) revealed that the dominant phyla were Proteobacteria, Cyanobacteria, Bacteroidota, Bacteria (k) undetermined and Actinobacteria (Figure 3.13). Treatment A had the highest relative abundance of cyanobacteria, approximately 55% of all 16S rRNA (Figure 3.13 and Figure 3.14), and also by far the highest normalised counts of 16S rRNA of all treatments (Figure 3.13). 2-MIB concentrations (ng/L) in individual samples was added to Figure 3.13 and Figure 3.14 (box with red outline) to visualise the variation in the replicates for each treatment, which could be related to biological differences.
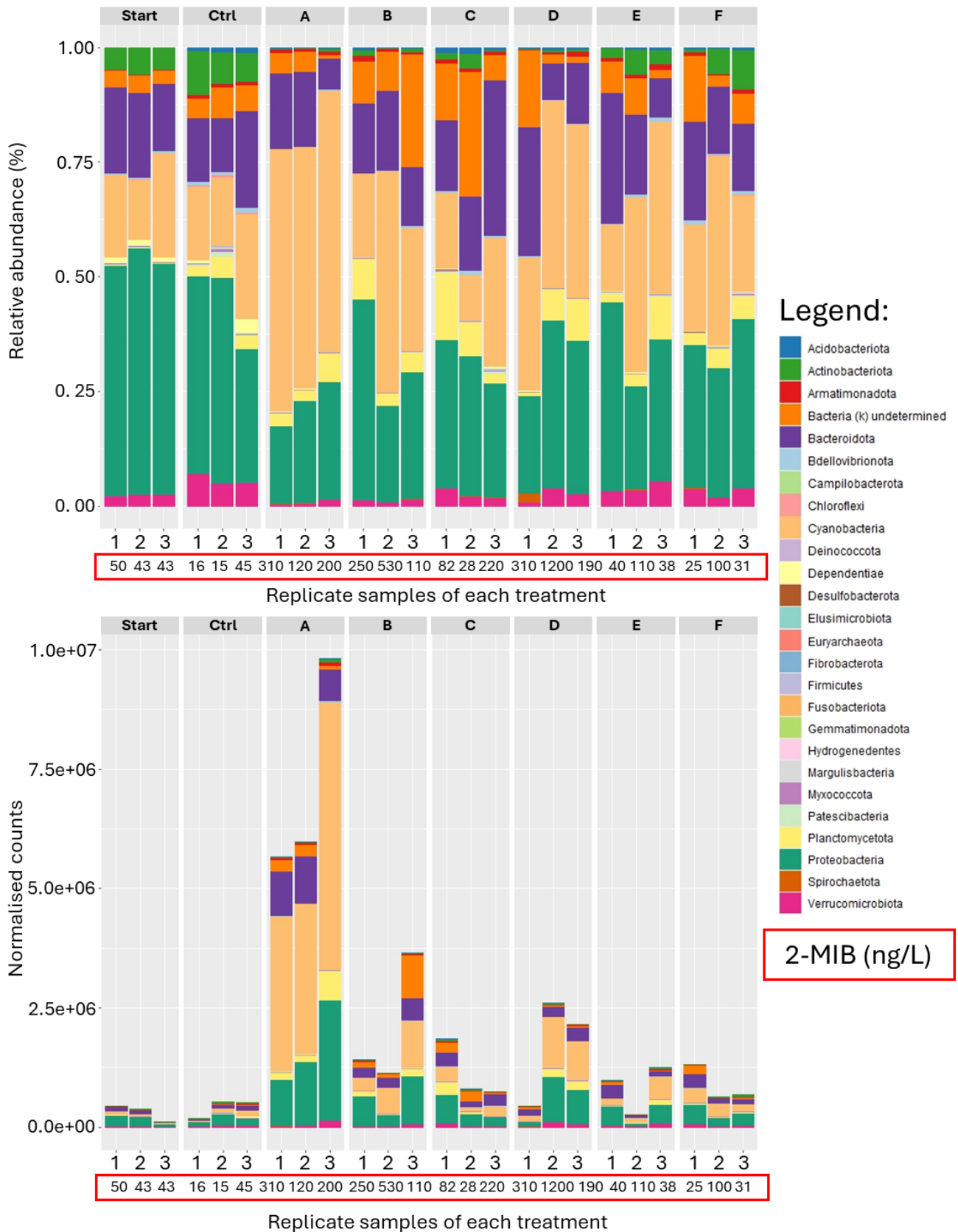
Figure 3.13. Bacterial 16S rRNA data at phylum level for each replicate (1, 2 and 3) of the start community ("Start") and final communities for every treatment, showing relative abundance (%; 1 = 100%) and normalised counts. 2-MIB concentrations in ng/L for start and flasks on the final day are visualised below x-axis in box with red outline.

The relative abundance of cyanobacterial genera in the 16S rRNA sequencing data revealed that the initial start community ("Start") had a completely different community composition than the flasks at the end of the experiment (Figure 3.14). The initial start community ("Start") was dominated by *Aphanizomenon* spp. that were most similar to strains *Aphanizomenon* sp. NIES81 and *Aphanizomenon* sp. MDT14a. The community at the end of the experiment in every flask contained mainly *Pseudanabaena* spp. (resembling strain PCC-7429), *Leptolyngbya* spp. (resembling strain SAG 2411) and *Cyanobium* spp. (resembling strain PCC 6307) (Figure 3.14). ASV counts were combined for the cyanobacterial genus and strain classification, because often species could not be determined, apart from *Pseudanabaena galeata* PCC-7429 and *Pseudanabaena foetida* PCC-7429, which were determined as closest DNA match alongside undetermined species of *Pseudanabaena* (PCC-7429). *Leptolyngbya* spp. (SAG 2411) dominated in almost every final flask and consisted of 5 - 40% of total 16S rRNA; the highest percentage was treatment A (Figure 3.14). *Pseudanabaena* spp. (PCC-7429) were present in every final flask and consisted of 5 – 20 % of total 16S rRNA, but mainly treatments A, B and D had elevated normalised counts for this genus (Figure 3.14). Normalised counts (ASV counts normalised for dilution factor) for cyanobacteria genera, showed that treatment A was also much higher than other treatments (Figure 3.14). *Aphanizomenon* sp. (NIES81) from the start community ("Start") only developed a significant relative abundance in flask F2 (Figure 3.14), which is the only community that produced a geosmin concentration >100 ng/L (Appendix B Table S.4).
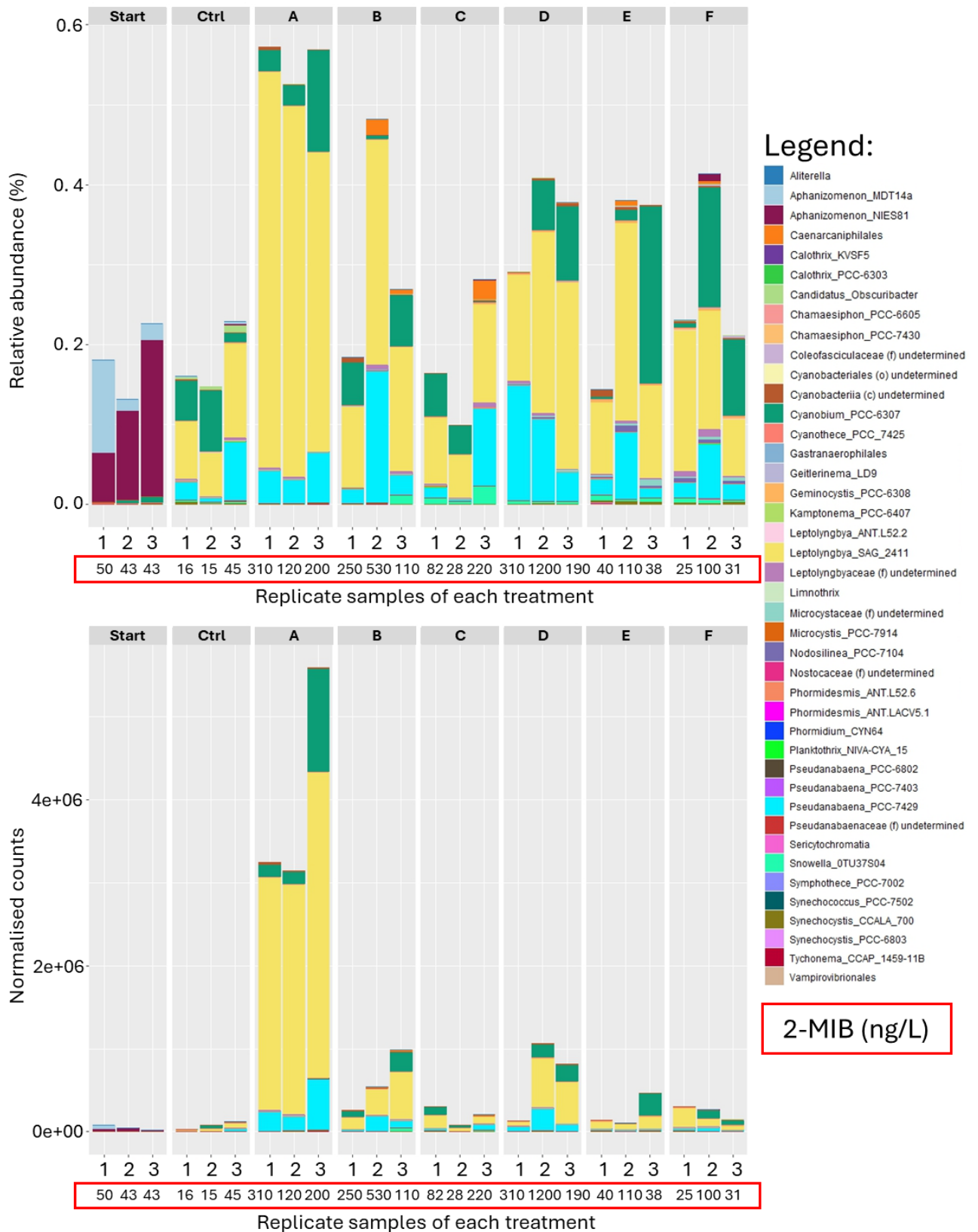
Figure 3.14. Bacterial 16S rRNA data at genus level (or next best level, reported as taxonomic rank: c = class, o = order, f = family, "undetermined") for each replicate (1, 2 and 3) of the start community ("Start") and final communities for every treatment, showing relative abundance (%; 1 = 100%) and normalised counts. For clearer visualisation, genera with <0.01% relative abundance were removed from the plot. 2-MIB concentrations in ng/L for start and flasks on the final day are visualised below x-axis in box with red outline.

The most abundant species of green algae in the samples were *Tetradesmus obliquus*, *Scenedesmacaea* (f) undetermined, *Mychonastes homosphaera*, *Desmodesmus serratus*, *Desmodesmus* sp. (resembling TAU-MAC 0810), *Choricystis* sp., *Chlorophyceae* (c) undetermined, *Botrycoccus braunii*, *Pseudopediastrum* sp. and *Ankistrodesmus fusiformis* (Figure 3.15). *Tetradesmus obliquus* was more abundant (relative abundance) in the higher nutrient treatments, with high biomass in treatment A. Control treatments had particularly high relative abundance of *Choricystis* sp., *Chlorophycaea* (c) undetermined and *Botryococcus braunii*. The initial start community ("Start") consisted of almost exclusively *Botryococcus braunii* with approximately 75% of all *rbc*L DNA consisting of that species (Figure 3.15). The most abundant species of diatom was *Melosira varians*, with a general trend of higher relative abundance at lower nutrient treatments, for example, approximately 70% of all *rbc*L DNA was from *Melosira varians* in treatments E (Figure 3.16). This trend is the opposite of green algae because they collectively make up all *rbc*L DNA. In some samples, *Stephanocyclus meneghinianus*, *Fragilaria crotonensis*, *Encyonema* sp. and *Navicula* sp. were present in small numbers (Figure 3.16).
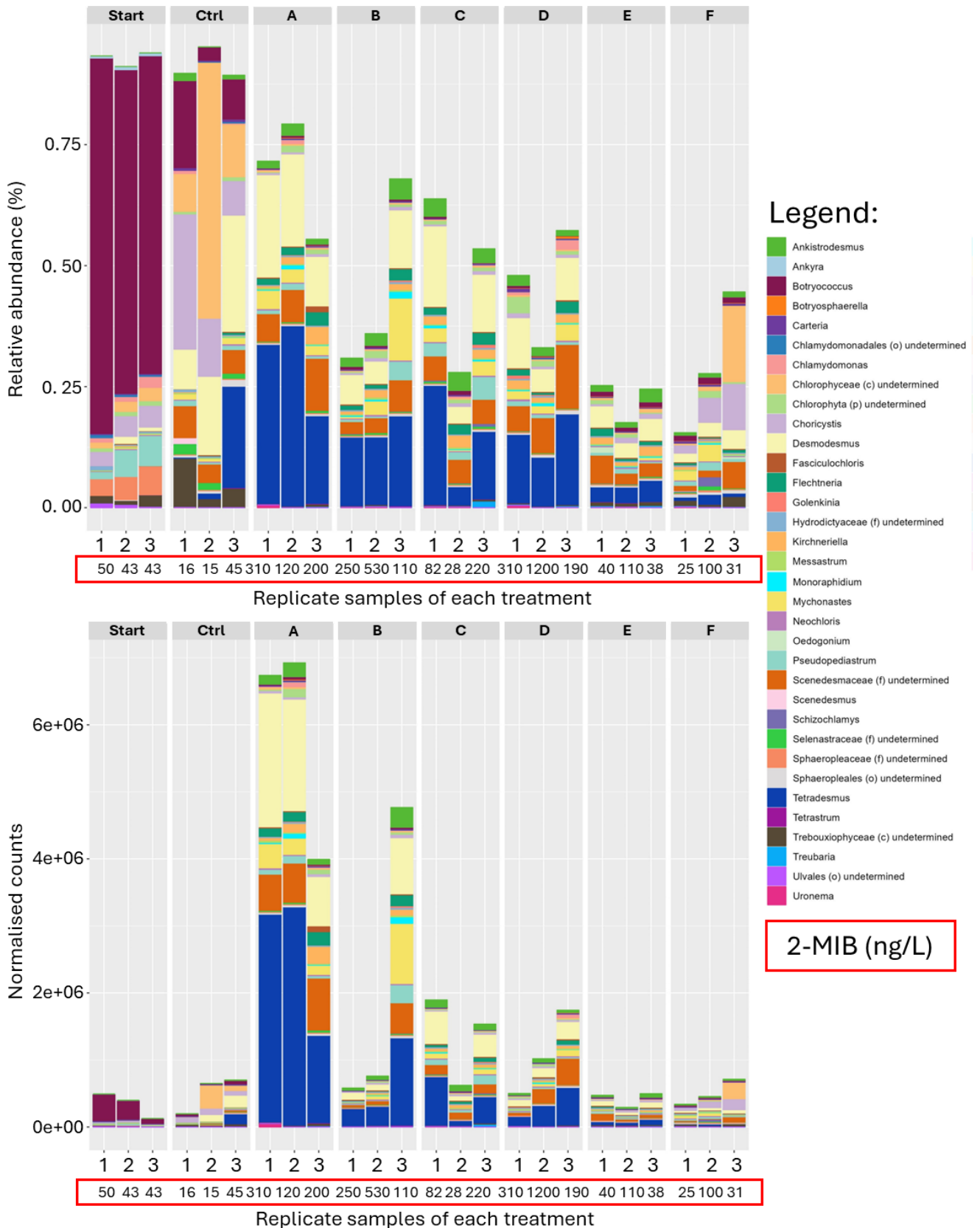
Figure 3.15. Green algae rbcL data at genus level (or next best level, reported as taxonomic rank: p = phylum, c = class, o = order, f = family, "undetermined") for each replicate (1, 2 and 3) of the start community ("Start") and final communities for every treatment, showing relative abundance (%; 1 = 100%) and normalised counts. For clearer visualisation, genera with <1% relative abundance were removed from the plot.
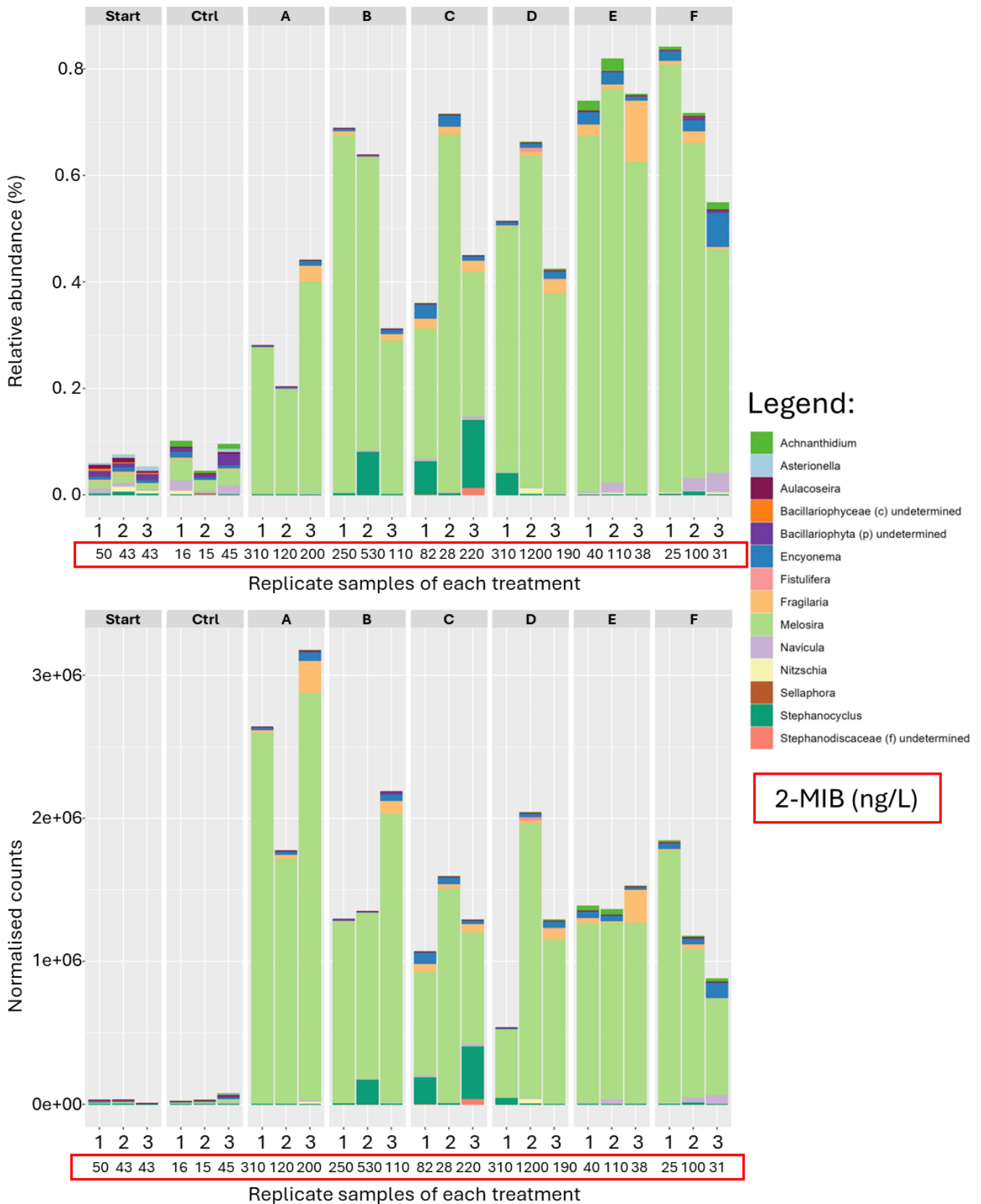
Figure 3.16. Diatom rbcL data at genus level (or next best level, reported as taxonomic rank: p = phylum, c = class, o = order, f = family, "undetermined") for each replicate (1, 2 and 3) of the start community ("Start") and final communities for every treatment, showing relative abundance (%; 1 = 100%) and normalised counts. For clearer visualisation, genera with <1% relative abundance were removed from the plot.

3.5.4 Relationships between 2-MIB production and community structure

The qPCR results of 2-MIB synthase gene (mic) prevalence (copy number) were normalised with the 16S rRNA prevalence (copy number) and arcsine square root transformed, and this factor revealed a positive linear correlation with $\log_{(10)}(\text{MIB}+1)$, with an R^2 of 0.635 (Figure 3.17).
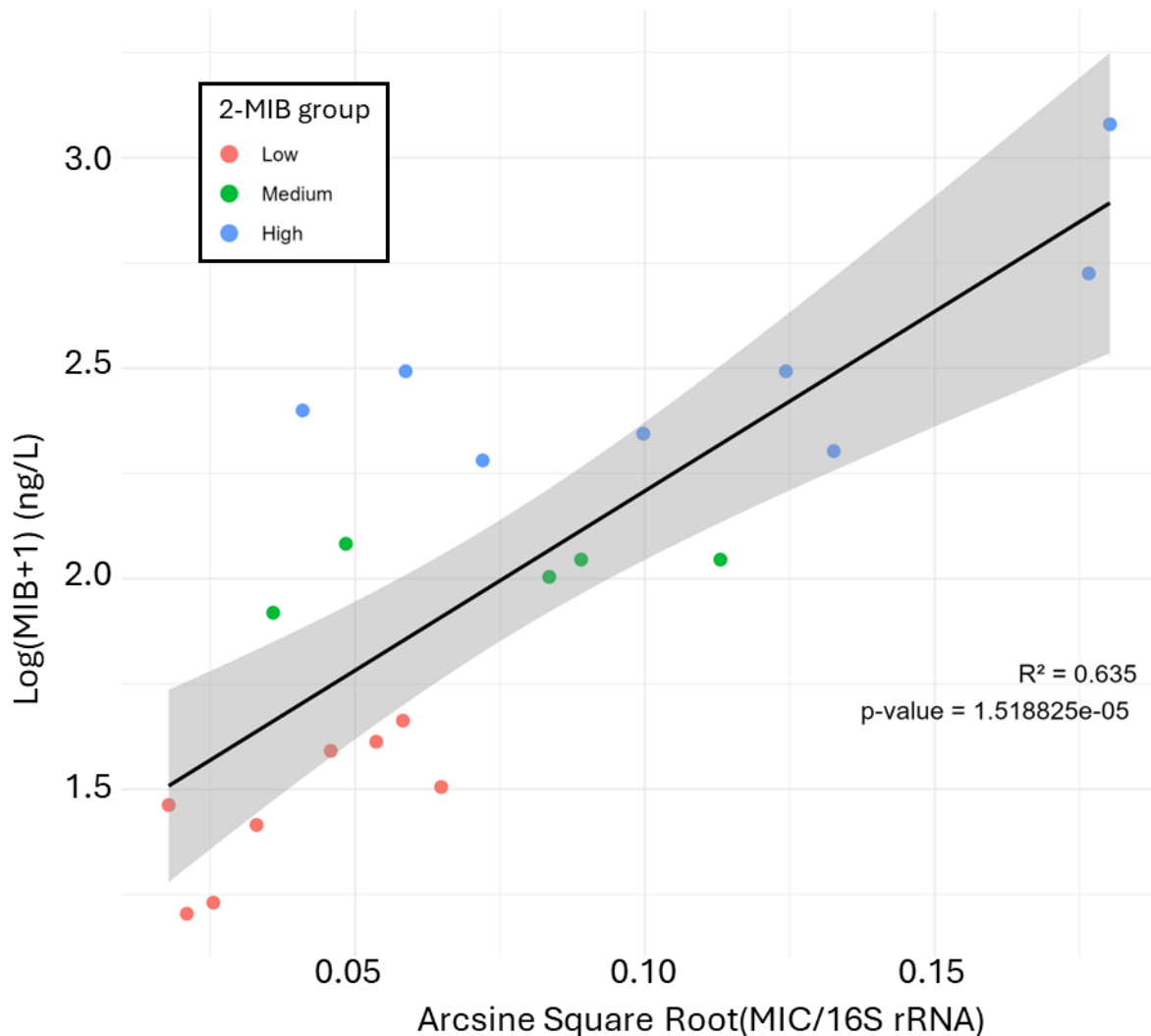
Figure 3.17. Linear model of 2-MIB synthase gene (mic) prevalence normalised with the 16S rRNA prevalence (arcsine square root transformed) and $\log_{(10)}(\text{MIB}+1)$, with model performance in the plot and colour of datapoints indicate 2-MIB group (high, medium or low concentration).

Treatment A had the highest relative abundance and normalised counts of cyanobacteria (including potential 2-MIB producers) in the 16S rRNA data (Figure 3.14 and Figure 3.10) and also had one of the highest 2-MIB concentrations of all treatments, together with treatment B and D. 2-MIB concentration normalised by approximate biomass of potential 2-MIB producers (Figure 3.18, panel A) and

Pseudanabaena spp. (PCC-7429) (Figure 3.18, panel B) showed that treatments B and D had higher 2-MIB production per unit of biomass than treatment A. This was also visible in treatment C, but there was more variability between replicates (Figure 3.18).
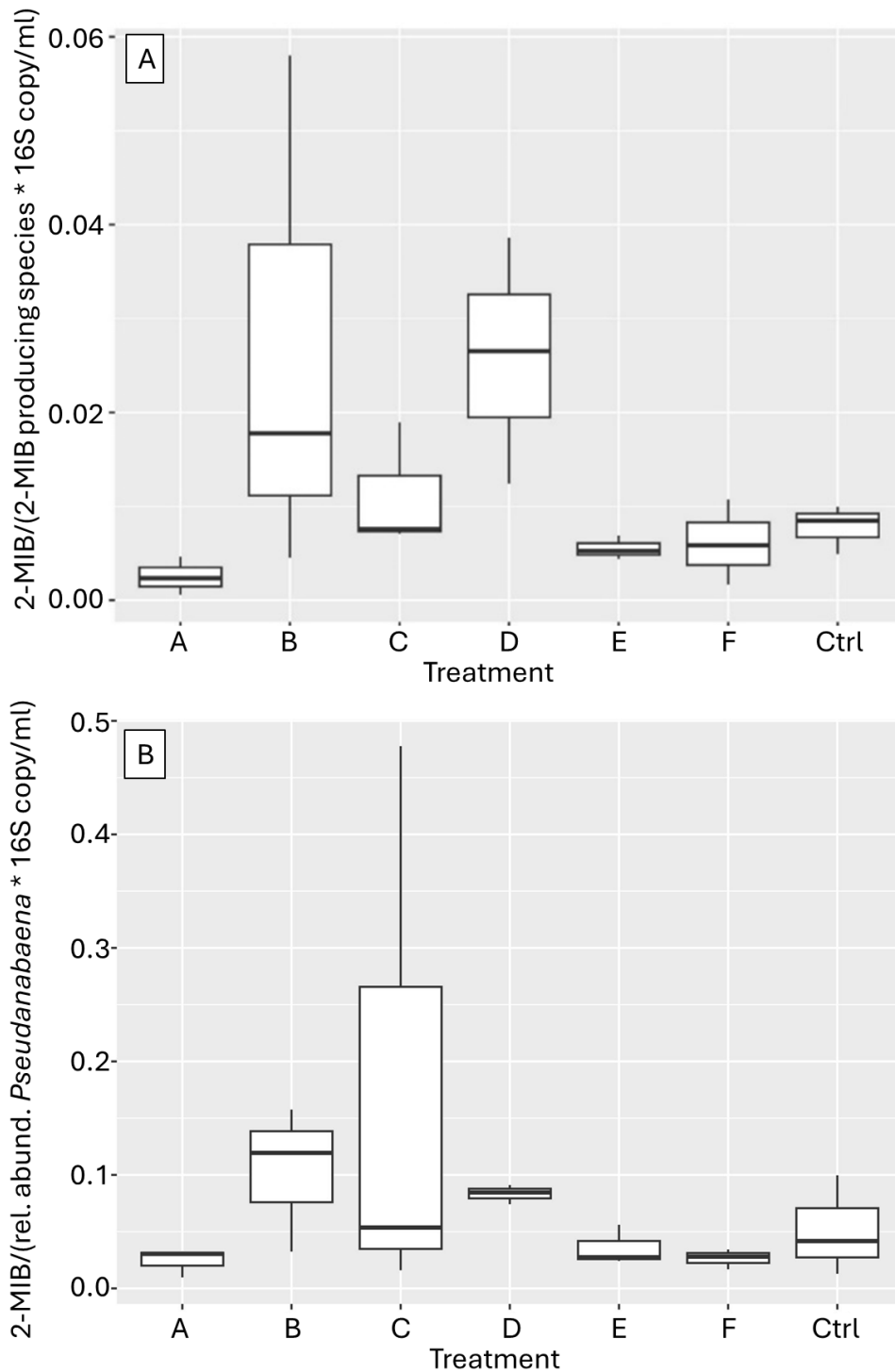
Figure 3.18. Both plots show 2-MIB concentration divided by relative abundance of (a) expected 2-MIB producing cyanobacteria * 16S copy number per ml or (b) relative abundance of Pseudanabaena spp. (PCC-7429) * 16S copy number per ml.

There was a relationship between 16S rRNA and 2-MIB concentration (divided into low (<50 ng/L), medium (50 – 150 ng/L) and high (> 150 ng/L) (Figure 3.19)), although there were sometimes differences between replicates within treatments. A statistically significant difference between the three 2-MIB groups was evidenced with 95% confidence intervals (Figure 3.19), a PERMANOVA test with 999 permutations (Pr (>F) = 0.019; R^2 = 0.916) and an ANOSIM test with 9999 permutations (ANOSIM statistic R = 0.286; significance = 0.0022). Additional variables from the experiment were plotted on the NMDS (Figure 3.20, panel A) and apart from NH_4^+, they were all significant (pvalue < 0.05) based on the RStudio envfit function. Unexpectedly, mic/16S (gene copy numbers) revealed the clearest agreement with the direction of the high 2-MIB group. Higher NH_4^+ seemed to be indicative of the low 2-MIB group (Figure 3.20, panel A). Cyanobacteria genera that were significant in the envfit analysis (pvalue < 0.05) revealed that *Pseudanabaena* spp. (PCC-7429) had a tendency towards the high 2-MIB group, *Leptolyngbya* spp. (SAG 2441) towards the area between the medium and high 2-MIB group, whilst the other nine genera pointed in the general direction of the low 2-MIB group (Figure 3.20, panel B).
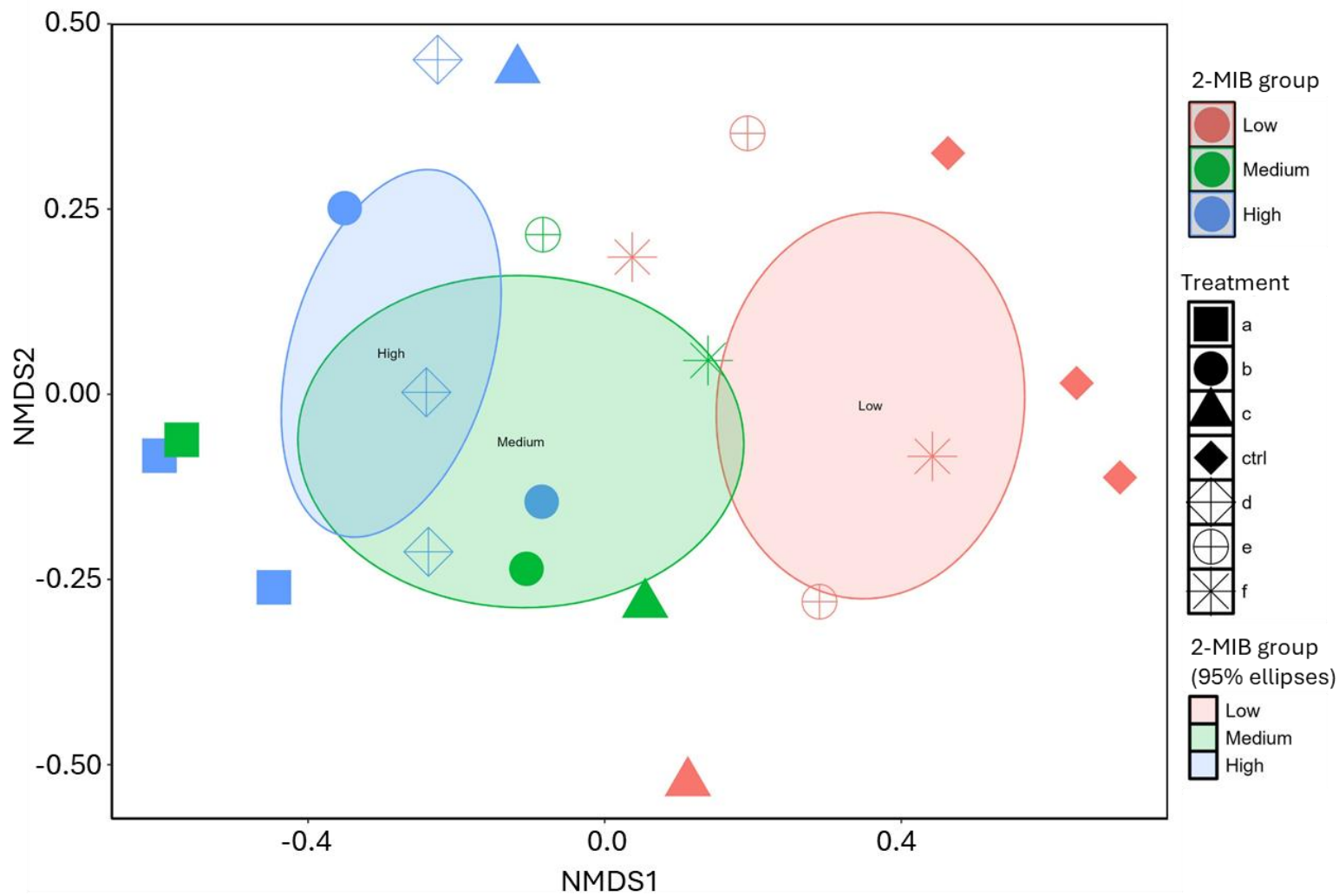
Figure 3.19. 16S rRNA sequencing data used for NMDS with three 2-MIB groups; low (<50 ng/L), medium (50 – 150 ng/L) and high (> 150 ng/L), indicated with a coloured ellipse on the NMDS as the 95% confidence interval.
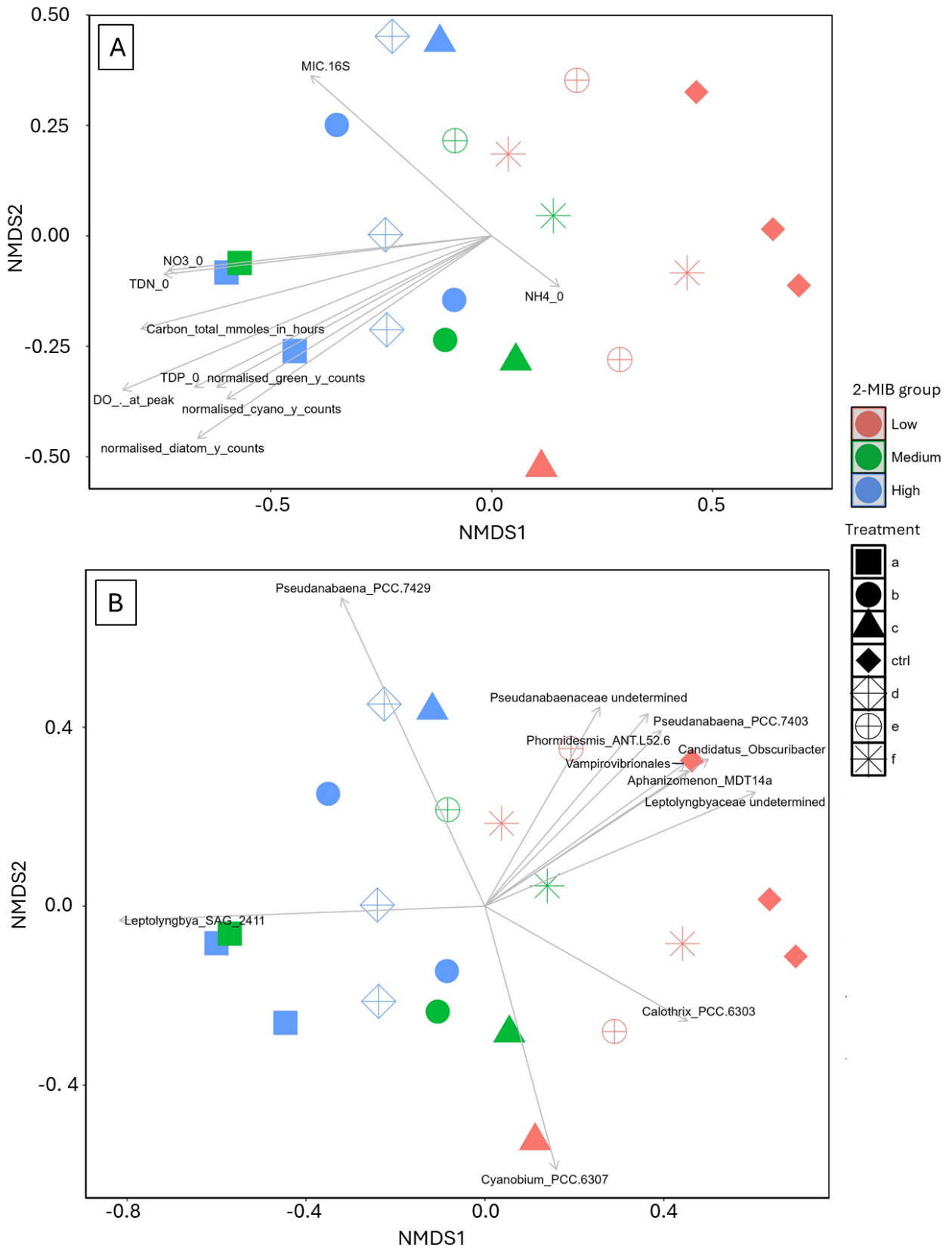
Figure 3.20. 16S rRNA sequencing data used for NMDS with three 2-MIB groups; low (<50 ng/L), medium (50 – 150 ng/L) and high (> 150 ng/L). Envfit was used to overlay the NMDS plot in Figure 3.19 with additional variables of interest (a) and relative abundance of significant (pvalue <0.05) cyanobacteria genera (b).

Indicator species analysis for 16S rRNA revealed that two genera of cyanobacteria were associated with high and medium 2-MIB groups; *Leptolyngbya* spp. (stat = 0.512, pvalue = 0.0482, Table S.5 in Appendix B) and *Pseudanabaena* spp. (stat = 0.556, pvalue = 0.0253, Table S.5 in Appendix B). Relative abundance of all significant cyanobacteria (pvalue < 0.05) from the indicator species analysis showed that these two genera indeed have higher abundance for the high 2-MIB group (Figure 3.21). The other three significant genera of cyanobacteria from the indicator species analysis (pvalue <0.05) had a significantly higher abundance at the low 2-MIB group (Figure 3.21 and Table S.5 in Appendix B). Besides cyanobacteria, several genera of other bacteria were significantly associated with certain 2-MIB groups (Table S.5 in Appendix B), but these were not investigated in this study.
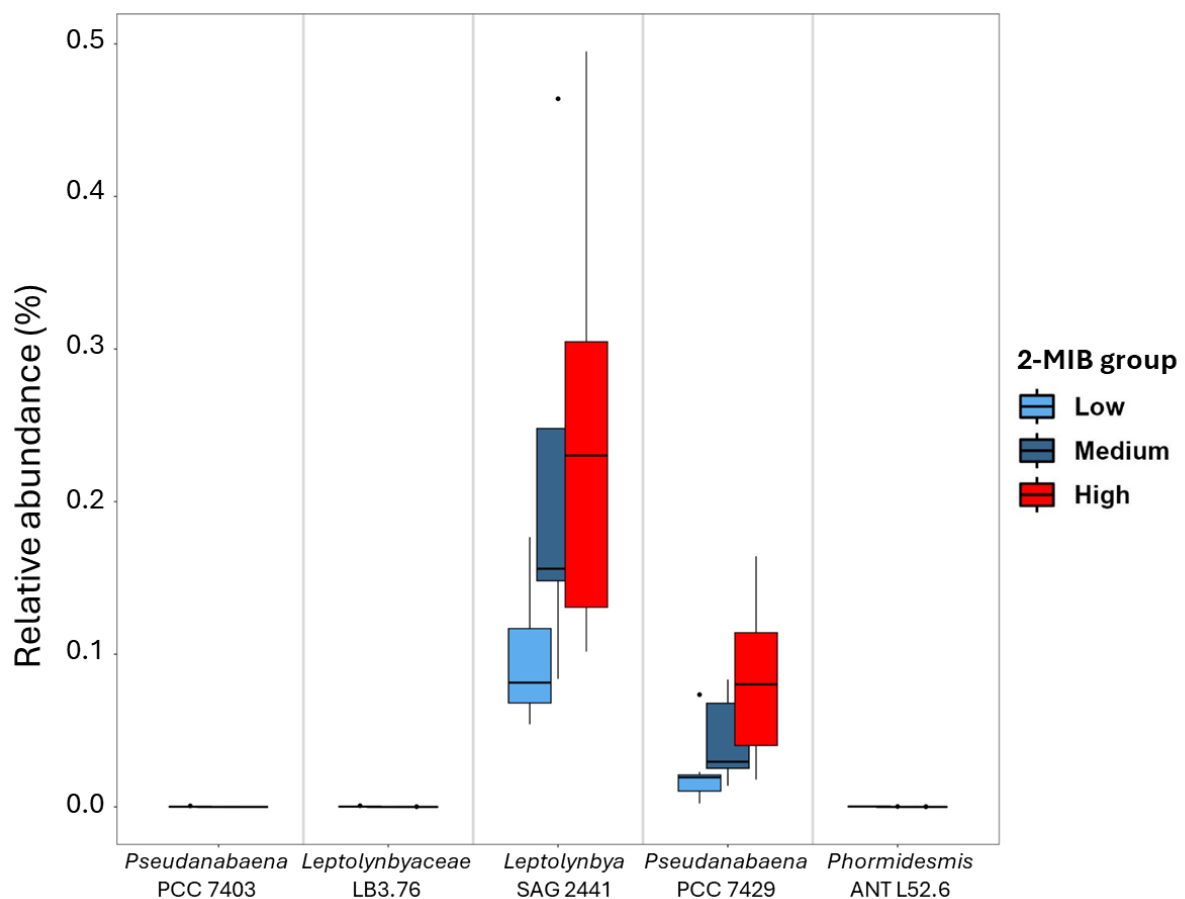
Figure 3.21. Relative abundance (%; 1 = 100%) of significant cyanobacteria from indicator species analysis (pvalue <0.05) on 16S rRNA sequencing data, separated by 2-MIB groups; low (<50 ng/L), medium (50 – 150 ng/L) and high (> 150 ng/L).

Relative abundance of *Leptolyngbya* spp. (SAG 2441) and *Pseudanabaena* spp. (PCC-7429) correlated with $\log_{(10)}$(MIB+1) yielding an R^2 of 0.286 and 0.478,
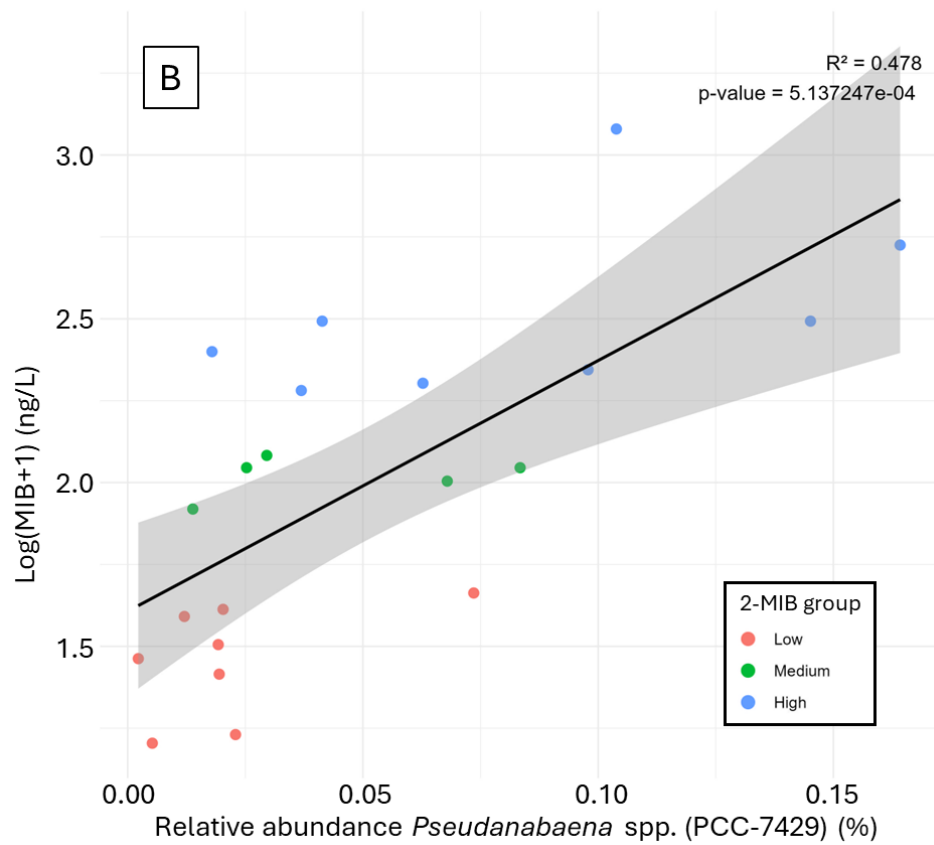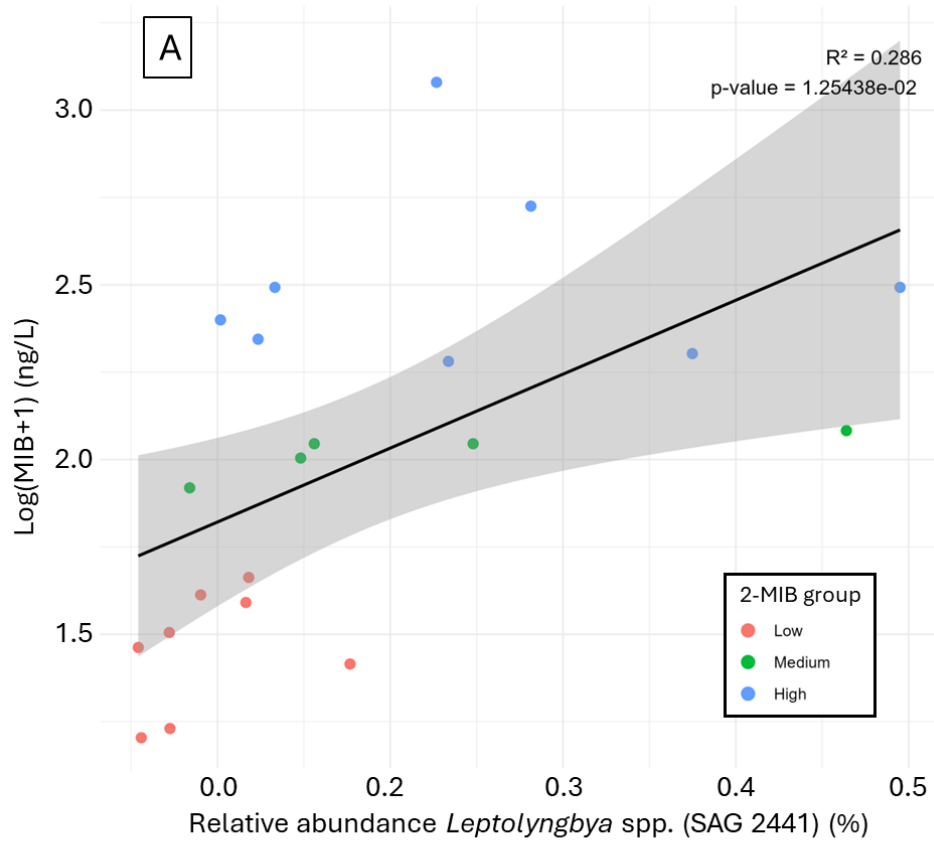
respectively (

Figure 3.22.
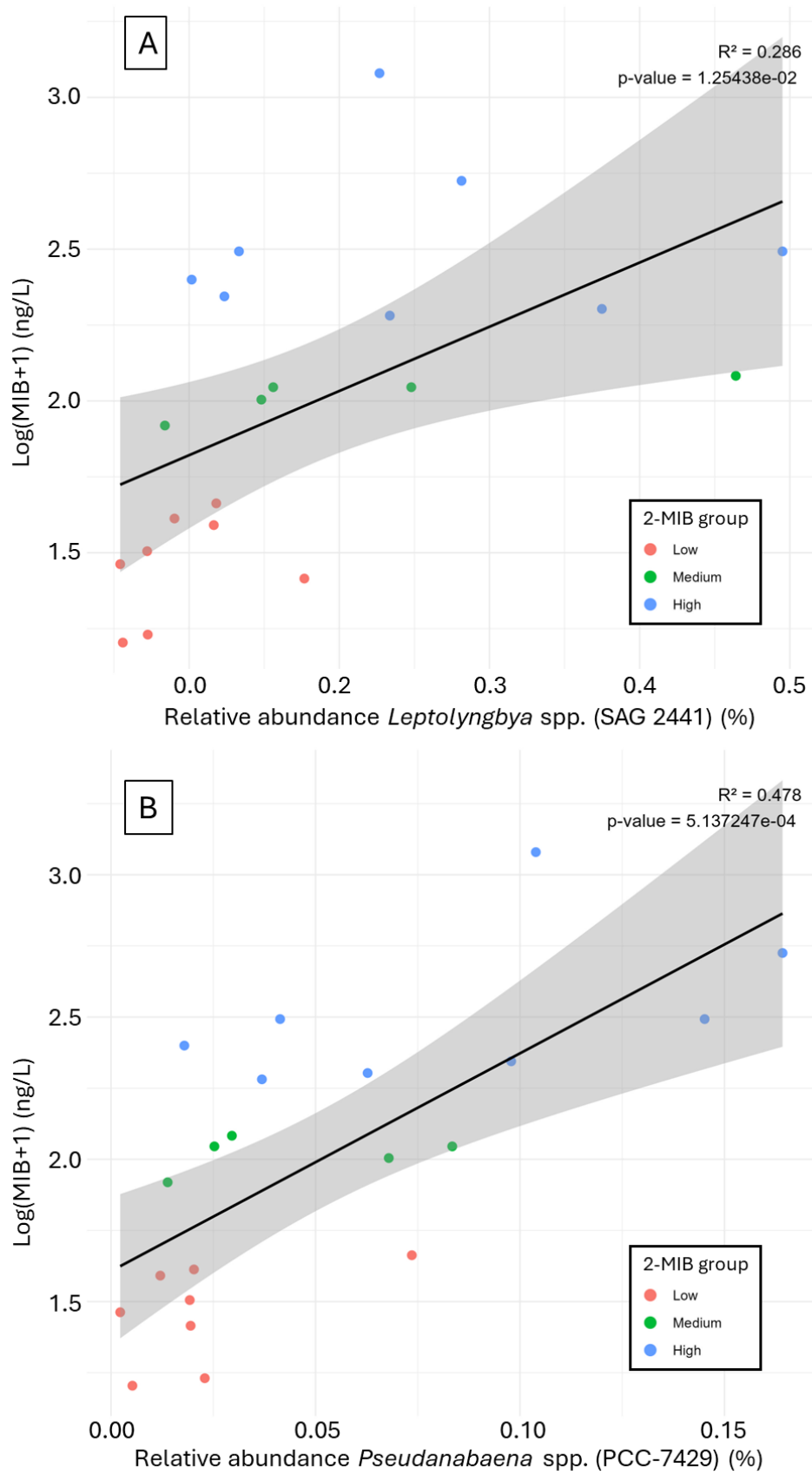
Figure 3.22. Linear models of log$_{(10)}$(MIB+1) with relative abundance (%; 1 = 100%) of Leptolyngbya spp. (SAG 2441) (a) and Pseudanabaena spp. (PCC-7429) (b) from 16S rRNA sequencing data.

The most interesting finding from indicator species analysis for *rbc*L was that green algae *Tetradesmus* sp. was significantly associated with high and medium 2-MIB groups (stat = 0.541, pvalue = 0.0368) (Table S.6 in Appendix B). Relative abundance of *Tetradesmus* sp. correlated with log$_{(10)}$(MIB+1) yielding an R^2 of 0.247 (
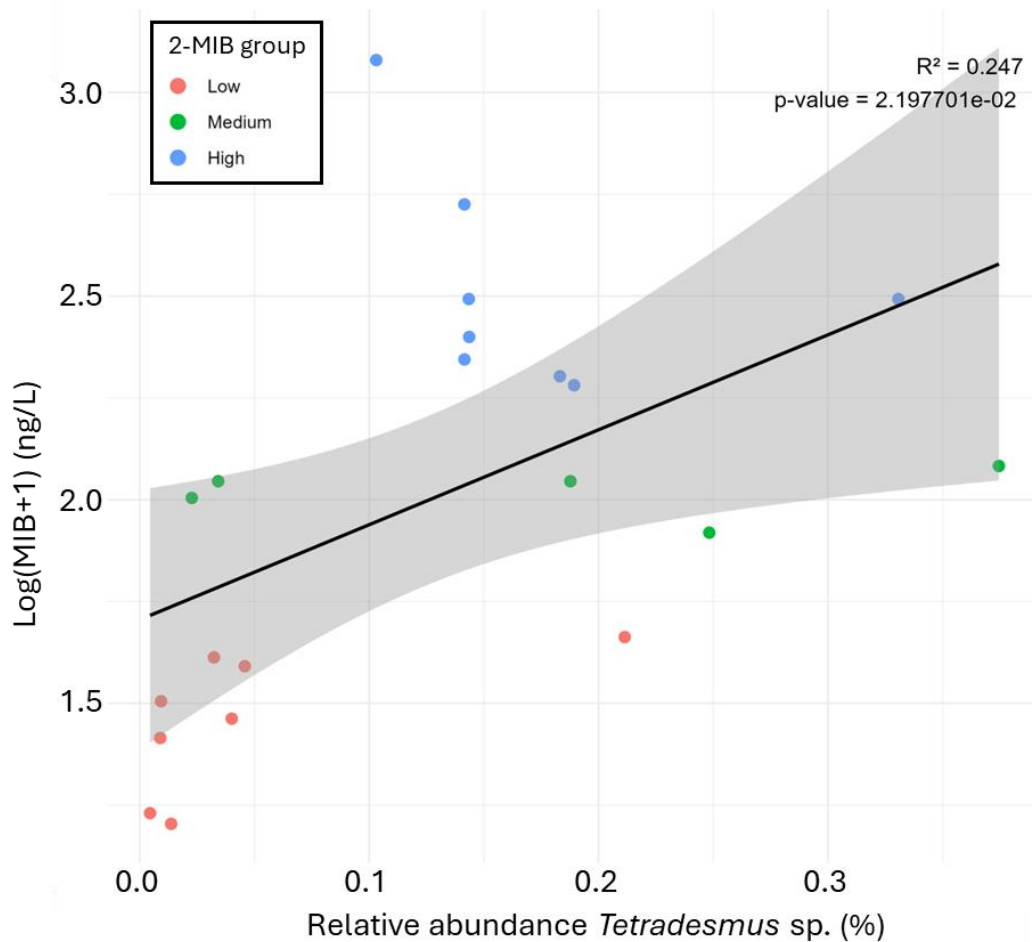
Figure 3.23).
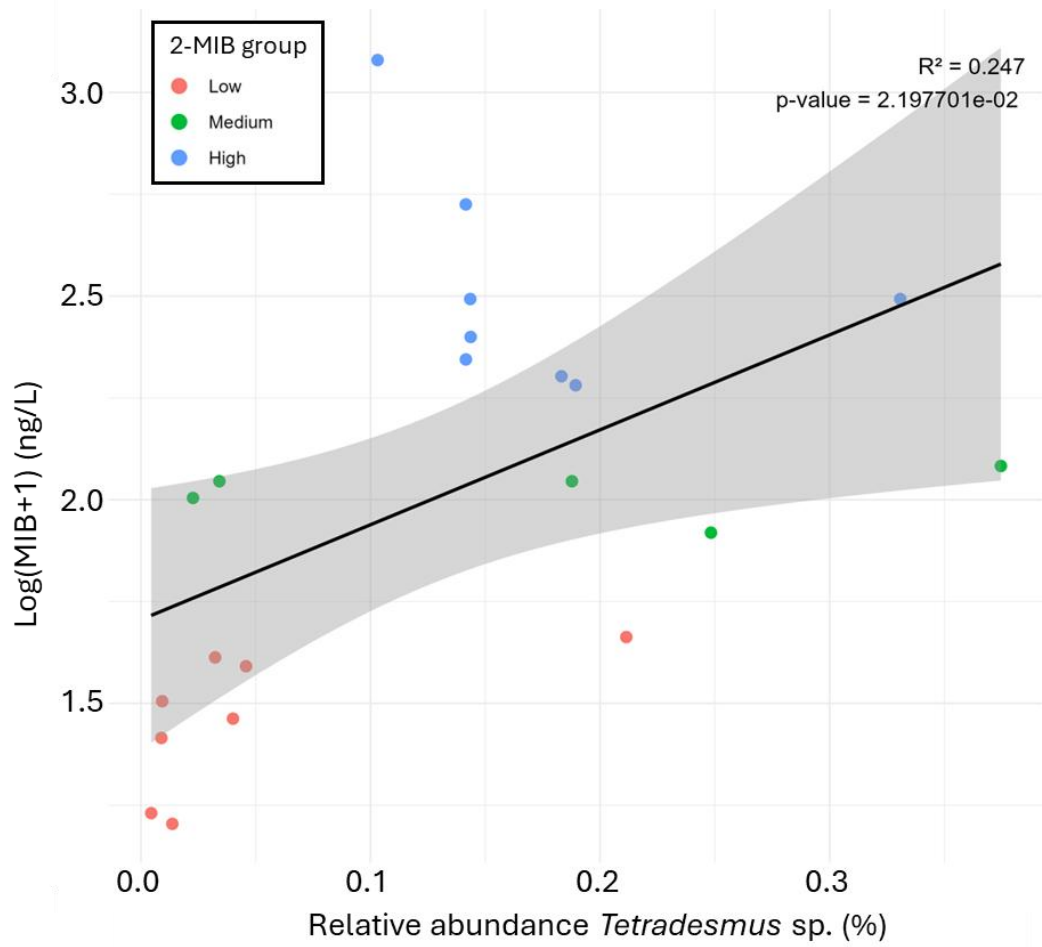
Figure 3.23. Linear model of log$_{(10)}$(MIB+1) with relative abundance (%; 1 = 100%) of green algae Tetradesmus sp. from rbcL sequencing data.

3.6 Discussion

3.6.1 Impact of nutrient ratios on community productivity and 2-MIB

Throughout the experiment, 2-MIB was produced in almost every flask, apart from the control treatments, and the different nutrient ratios had different effects. Low NH_4^+:NO_3^- ratios had significantly higher 2-MIB concentrations compared to control and high NH_4^+:NO_3^- treatments, whilst low N:P concentrations had significantly higher 2-MIB concentrations than the control. Low N:P and high NH_4^+:NO_3^- ratios were expected to result in enhanced cyanobacteria growth and 2-MIB production (Harris et al., 2016; Perkins et al., 2019; Winston et al., 2014; Yue et al., 2024), but these ratios did not seem to be the most important factor in this study.

An important finding in these experiments, is that carbon produced throughout the experiment by the whole community (not just 2-MIB producers) correlated well with 2-MIB production (R^2 = 0.608). This seems to suggest that 2-MIB production is related to increased productivity and metabolic activity. The highest 2-MIB concentrations occurred in treatments with high total nitrogen (as NO_3^-) and sufficient phosphorus that had sustained balanced growth. However, this was a total community response, and it is important to recognise that nutrient conditions driving the growth of T&O producing cyanobacteria and subsequent T&O production are likely species-specific (Watson et al., 2016).

3.6.2 Nitrogen effects on 2-MIB

The results from this experiment highlight that the concentration of total dissolved nitrogen (TDN, as NH_4^+ or NO_3^-) is more important for biomass development of 2-MIB producing cyanobacteria than the original ratio NH_4^+:NO_3^-. However, NH_4^+ and NO_3^- did have a different growth response which seemed to have impacted biomass development of 2-MIB producers. Treatments B and C had similar TDP and TDN concentrations, but treatment C had NH_4^+ whilst B had NO_3^- as N source, which resulted in rapid growth ('boom-bust' dynamics) for treatment C (top left in Figure 3.24) and balanced growth in B (top right in Figure 3.24) which started four days later but lasted longer. Phytoplankton responded rapidly with the preferred bioavailable NH_4^+-N as N source (Dortch, 1990), whilst NO_3^--N uptake was delayed because it requires specific enzymes that are more easily available to diatoms and green algae than cyanobacteria (Erratt et al., 2020; Glibert et al., 2016; Hyenstrand et al., 2000;

Trommer et al., 2020). However, the results from this experiment showed that slower balanced growth with NO_3^- as N source gave benthic 2-MIB producing cyanobacteria more time to develop than the 'boom-bust' growth in NH_4^+ treatments. It is possible that faster growing green algae, which can use a variety of N sources effectively (Donald et al., 2011), outcompeted benthic cyanobacteria in treatments with NH_4^+ that had a limited concentration of TDN (Figure 3.25). The development of 2-MIB producers in this experiment likely required a TDN concentration of around 2.5 mg N/L, provided that sufficient TDP was available.
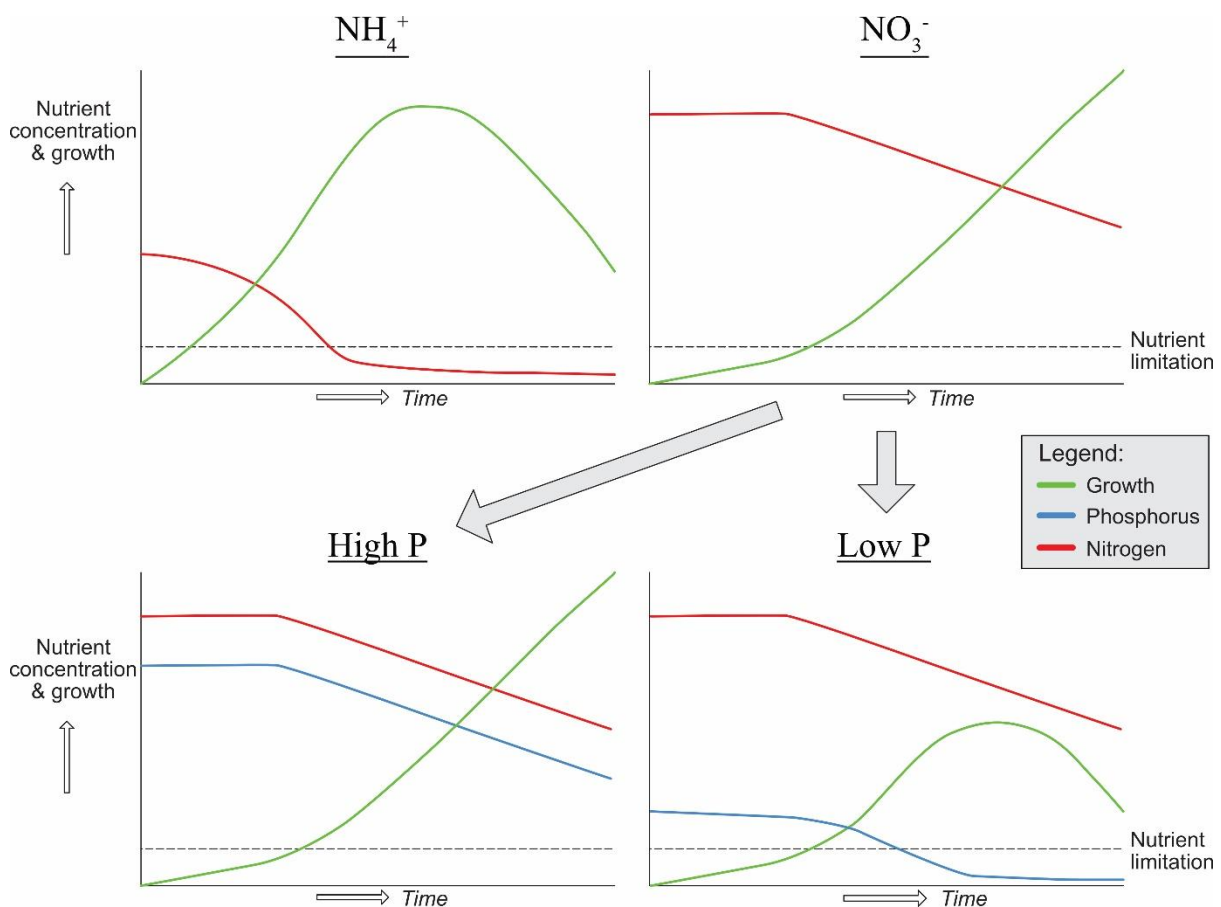
Figure 3.24. Hypothetical plots showing the different treatment types. Top left: NH_4^+-N as N source showing 'boom-bust' growth, partially due to lower TDN concentrations (treatment C). Top right: NO_3^--N as N source showing a delayed growth response but overall balanced growth, partially due to higher TDN concentrations (treatment A and D). Bottom left: High TDN and high TDP have balanced growth until the end of the experiment (treatment A). Bottom right: High TDN but low TDP have balanced growth until TDP runs out, then possible growth decline or other reasons cause enhanced 2-MIB release (treatment D).

$$\underline{NH_4^+}$$
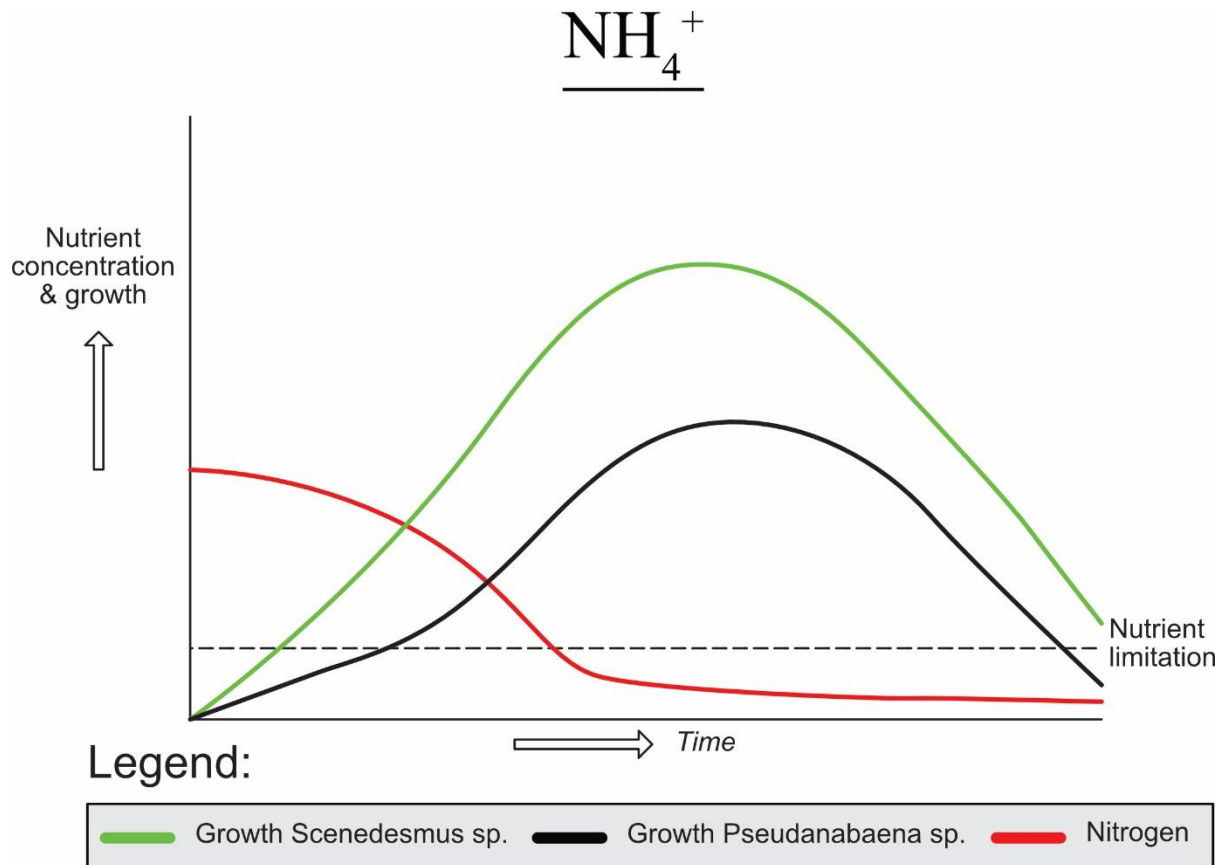
Figure 3.25. Hypothetical plot indicating how in a 'boom-bust' growth scenario with NH_4^+ but limited TDN, a potentially slower growing Pseudanabaena sp. (cyanobacteria) would produce less biomass than a potentially faster growing green algae species like Scenedesmus.

Another reason why high NH_4^+ treatments could not sustain growth was because the TDN concentrations were 10x lower than the high NO_3^- treatments and so were subsequent TDP concentrations (to balance the ratios). This was a result of basing initial NH_4^+ concentrations on realistic reservoir concentrations, which are about 10x lower than high NO_3^- concentrations (Figure 3.24). To see the effect of NH_4^+ and NO_3^- as an N source and NH_4^+:NO_3^- ratios on 2-MIB production, it would have been better to restock nutrient concentrations during the experiment, so that NH_4^+:NO_3^- ratios could be maintained and nutrient limitation would not play a role. However, it is important to consider how these biogeochemical dynamics work in an open system, such as the reservoir from which the samples were collected. In such a system, NH_4^+ may be continuously released (for example, from internal loading of the sediment layer), or could be periodically introduced (via tributary input or surface runoff), which would have very different effects on the subsequent NH_4^+:NO_3^- ratios. The experiment showed that NH_4^+ was taken up preferentially and more rapidly than NO_3^-, so short influxes of NH_4^+ and resulting impacts on the community would not be accurately

captured by periodic (twice monthly) chemical water sampling. This also emphasizes that field studies which are based on low sampling resolutions cannot capture the full biogeochemical dynamics and might find different relationships than high-frequency studies (field or laboratory) would. The suggested high $NH_4^+:NO_3^-$ ratio as a driver of T&O events (Harris et al., 2016; Perkins et al., 2019) might be a delayed effect of N and P uptake processes with differing uptake speeds, depending on the bioavailability of the nutrient fraction, and subsequent biological response. The expected lag-time between environmental drivers and T&O release highlights that initial $NH_4^+:NO_3^-$ ratio (and generally nutrient concentrations) which drive growth of T&O producing cyanobacteria can be altered by biogeochemical cycles in the meantime and result in different final $NH_4^+:NO_3^-$ ratios (and nutrient concentrations) when T&O is released (Figure 3.26). In field studies, the final $NH_4^+:NO_3^-$ ratios in a water sample would be compared to T&O concentrations, but due to the lag-time in T&O production and release, these ratios are not necessarily the same as the initial driver. The lag-time is incredibly important and requires further research, especially for the development of early warning predictors of T&O events. Knowledge from high-frequency nutrient monitoring could provide useful insights, which will be further discussed in Chapter 4.
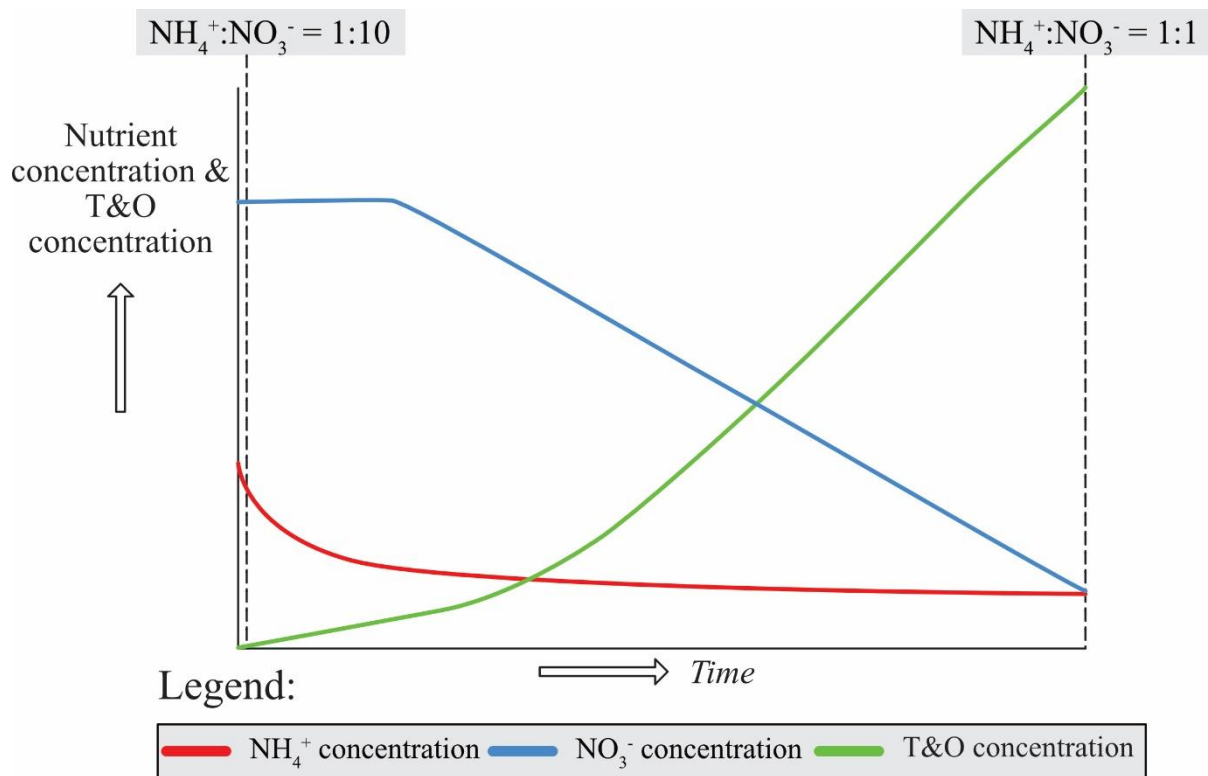
Figure 3.26. Hypothetical plot indicating how initial $NH_4^+:NO_3^-$ ratios stimulating T&O production can be different to final $NH_4^+:NO_3^-$ ratios at the time when highest T&O concentrations are detected, after a T&O production and release lag-time.

3.6.3 Phosphorus effects on 2-MIB

The evidence from this experiment suggests that total dissolved nitrogen concentrations were more important than the $NH_4^+:NO_3^-$ and N:P ratios for the establishment of 2-MIB producing benthic cyanobacteria *Pseudanabaena* spp. (resembling strain PCC-7429). Similarly, Gao et al. (2018) and Liu & Vyverman (2015) identified that relatively high concentrations of nitrogen compared to phosphorus concentrations were optimal growth conditions for planktic species of *Pseudanabaena*. This experiment showed that treatment D produced much higher concentrations of 2-MIB than E and F, which hardly produced 2-MIB. These treatments all had low TDP concentrations, but treatment D had much more TDN and slightly higher TDP than E and F. This finding suggests that TDN was more important originally as long as there was a certain threshold concentration of total dissolved phosphorus (TDP), which was sufficient to allow the development of 2-MIB producing cyanobacteria. The experiment indicates a potential threshold of TDP concentrations between 0.5 mg/L (treatment B) and 0.2 mg/L (treatment D). Similarly, a study by Kim et al. (2017) suggested that maximum growth of the cyanobacterium *Microcystis*

aeruginosa was primarily driven by total nitrogen (TN) concentrations when sufficient PO_4^{3-} was available, rather than by the NH_4^+:NO_3^- ratio and the N:P ratio.

However, phosphorus concentrations also revealed an interesting effect on 2-MIB concentrations, which seemed to be the related to the biomass of benthic 2-MIB producers that was able to develop under initial phosphorus concentrations. Treatments B and C produced on average more 2-MIB than E and F, when B, C, E and F all had very similar TDN concentrations, but E and F had much lower TDP concentrations. After the establishment of the 2-MIB producers, the experiments suggest that the depletion of the phosphorus source might have enhanced 2-MIB release. The highest 2-MIB concentrations directly and per 2-MIB producing cell (unit of biomass, estimated from DNA methods) were found in treatment D (HNO3_LNH4_LPO4), which almost completely used up the available phosphorus but still had plenty of N (bottom conceptual plots in Figure 3.24). This can be compared to treatment A (HNO3_LNH4_HPO4), which did not deplete the nitrogen or phosphorus pool and also had significant 2-MIB production directly but not such a high production per cell, as there was a much larger biomass of potential 2-MIB producers (estimated from DNA methods). 2-MIB concentrations normalised per biomass of all potential 2-MIB producers and *Pseudanabaena* spp. (PCC-7429) (likely main 2-MIB producer) revealed that treatments B, C and D had much higher 2-MIB production per unit of biomass than for example treatment A which had a high biomass of 2-MIB producers. When 2-MIB producers are prevalent in the community, biomass is not necessarily a good indicator of 2-MIB concentrations because the amount of 2-MIB produced per cell or *mic* gene copy can vary largely depending on the conditions (Chiu et al., 2016; Saadoun et al., 2001; Wang et al., 2016).

It seems that both N and P are important for the development of substantial biomass of 2-MIB producing cyanobacteria, but once there is an established community of 2-MIB producing *Pseudanabaena* spp. (PCC-7429), a decrease in P availability (such as in treatment D) might result in either passive (by cell lysis (Alghanmi et al., 2018; Naes et al., 1989; Saadoun et al., 2001; Zhang et al., 2009)) or active (enzymatic P release (Bar-Yosef et al., 2010; Raven, 2010)) extracellular release. This process is uncertain, and would require more research, but there are some indications for it, because treatment D had the highest dissolved organic phosphorus (DOP) uptake of all treatments. Organic nutrient fractions could play a role in these systems, as recent

studies suggested that organic nutrients can sometimes also support phytoplankton growth (Reinl et al., 2022). It is possible that the effect of phosphorus depletion on T&O release is species-specific. A study by Oh et al. (2017) found that geosmin production by cultures of *Anabaena* sp. NIER decreased when they were exposed to phosphorus limitation, but 2-MIB production by *Planktothrix* sp. FACHB-1374 was not affected.

3.6.4 Nutrient ratio effects on cyanobacteria, green algae and diatoms

Nutrient ratios N:P did not reveal the expected effect on the relative and normalised counts of cyanobacteria, green algae and diatoms. Green algae were expected to dominate under P-limited conditions in high N:P ratios, with a suggested TN:TP ratio > 75:1 favouring green algae dominance over cyanobacteria (Andersen et al., 2020; Harris et al., 2014). However, the largest influence factor on green algae dominance was probably TDP, as green algae showed a clear pattern of highest read counts at the highest TDP concentration and a decline with less TDP. Diatoms seemed to grow a bit better with the highest TDP concentration but also had relatively similar read counts in all the other treatments, apart from the control. Relative abundance of diatoms, however, increased towards treatments with the lowest TDP, whilst relative abundance of green algae decreased. This indicates that diatoms can cope better with low phosphorus treatments. Relative abundance of *rbc*L in the lower nutrient treatments (E and F) was dominated by diatoms, mainly *Melosira varians*, which is a species that can survive with low nutrient concentrations (Calderini et al., 2023).

Cyanobacteria had a slightly unexpected response to TDP, in which treatment A had a much higher value for normalised counts than all other flasks, but treatment D was second highest, with much lower TDP concentrations. Cyanobacteria normalised counts and relative abundance seemed to coincide with the highest TDN concentrations. This indicates that TDN is also an important factor for cyanobacteria growth, although preferential NO_3^- uptake by diatoms and NH_4^+ by cyanobacteria and green algae, was not visible.

3.6.5 Do species in the community control 2-MIB production?

The final bacterial community (16S rRNA) was significantly different between 2-MIB groups low (<50 ng/L), medium (50 – 150 ng/L) and high (> 150 ng/L), which indicates there is a certain community responsible for 2-MIB production. The cyanobacterial

genera *Leptolyngbya* spp. (resembling strain SAG-2411) and *Pseudanabaena* spp. (resembling strain PCC-7429) were associated with the medium and high 2-MIB group, and so were mic/16S qPCR results, TDN concentration, NO_3^--N concentration and total carbon (mmoles in hours). *Pseudanabaena* sp. was reported as 2-MIB producer in lakes many times (Chiu et al., 2016; Gao et al., 2018; Izaguirre & Taylor, 1998; Wang & Li, 2015; Zhang et al., 2016), but these species are described as planktic whereas the species in this study grew on the bottom of the flask as a biofilm. *Leptolyngbya* sp. has also been described to produce 2-MIB (Chiu et al., 2016; Wang et al., 2015; Watson et al., 2016; Yue et al., 2024). Moreover, *Pseudanabaena* sp. and *Leptolyngbya* sp. are capable of producing cyanotoxins and were found in toxin-producing benthic biofilms (Bauer et al., 2023; Borges et al., 2015; Cantoral Uriza et al., 2017; Catherine et al., 2013; Gaget et al., 2017; Gao et al., 2018; Perri et al., 2024; Rangel et al., 2014), which highlights the wider importance of studying these benthic cyanobacterial species.

Biomass has not always been found to indicate high 2-MIB production (Harris et al., 2016), which seemed to be the case in this study. Linear models of *Leptolyngbya* spp. (SAG 2441) ($R^2 = 0.286$) and *Pseudanabaena* spp. (PCC-7429) ($R^2 = 0.478$) relative abundance compared to 2-MIB concentrations showed that *Pseudanabaena* spp. (PCC-7429) was a better indicator of 2-MIB concentrations, which suggests it was the dominant producer. Even though *Leptolyngbya* spp. (SAG 2441) was the most dominant cyanobacteria in all flasks, relative abundance and normalised counts of *Pseudanabaena* spp. (PCC-7429) were specifically high in flasks with high 2-MIB concentrations, compared to other flasks. Relative abundance and normalised counts of *Leptolyngbya* spp. (SAG 2441) were especially high in treatment A, but this did not have the highest 2-MIB concentrations per biomass.

As this was a natural phytoplankton assemblage, cyanobacteria were in competition for nutrients with each other, diatoms, green algae and bacteria. Normalised count data revealed that cyanobacteria counts in treatment A were much higher than the other treatments, but also that green algae counts and diatom counts were highest in treatment A. Unfortunately, the DNA data doesn't allow cross comparison between results from cyanobacteria (16S rRNA) and green algae and diatoms (*rbc*L), so the phytoplankton community (cyanobacteria, diatoms and green algae) could not be analysed as a complete system. Recent research is suggesting the involvement of the

whole community in driving growth of certain phytoplankton species; a mutualistic/symbiotic relationship has been suggested with heterotrophic bacteria forming a microbiome surrounding phytoplankton cells (Cook et al., 2020; Louati et al., 2015), and interactions between heterotrophic bacteria and phytoplankton could help facilitate nutrient uptake (Reinl et al., 2022). Laboratory studies like Saadoun et al. (2001), often expose cultures of one cyanobacteria species to nutrient spikes to identify the effect on T&O production. This might be relevant to a natural situation if a species is already established but is not so relevant if there is a whole community of phytoplankton and bacteria competing for these nutrients or influencing their availability.

Species traits of cyanobacteria genera determine their optimal growth conditions and their ability to dominate the phytoplankton community (Reynolds, 2006). The species of *Leptolyngbya* and *Pseudanabaena* that originated from Reservoir 1, will most likely have different abilities to compete for nutrients and it seemed that *Leptolyngbya* spp. (SAG 2441) were able to develop more biomass than *Pseudanabaena* spp. (PCC-7429) in all treatments. However, the fact that relatively higher abundance of *Pseudanabaena* spp. (PCC-7429) was present in treatments with low phosphorus indicates that this species could potentially deal better with limited phosphorus supplies than *Leptolyngbya* spp. (SAG 2441). For example, *Pseudanabaena* sp. in Gao et al. (2018) had superior tolerance for disturbance, low light and phosphorus deficiency.

3.6.6 Species composition to detect 2-MIB producer

The dominant cyanobacteria in the initial start community were *Aphanizomenon* sp. (resembling strain NIES81) and *Aphanizomenon* sp. (resembling strain MDT14a), but those did not develop in this experiment, apart from flask F2. Geosmin concentrations in flask F2 (110 ng/L) were possibly related to the abundance of *Aphanizomenon* spp., which indicates that geosmin production in Reservoir 1 might be related to *Aphanizomenon* spp. Moreover, it is important to note that the initial start community sample had a high concentration of 2-MIB (±50 ng/L), but DNA methods only found *Aphanizomenon* spp. (not a 2-MIB producer according to Watson et al. (2016)), which indicates that common manual water sampling cannot always detect the source of 2-MIB production. Another sample in this dataset from Reservoir 1 (WR1 DNA data) in October 2022 had a high 2-MIB concentration of 160 ng/L but the most likely

responsible benthic cyanobacteria *Leptolyngbya* spp. or *Pseudanabaena* spp. (which have been studied in this experiment) were not found in the DNA of the water sample. This highlights again that benthic cyanobacteria are underestimated and often totally missed when trying to detect the source of 2-MIB. High geosmin concentrations occurred in the absence of geosmin-producing genes in water samples from three rivers in Kansas (USA) (Rider et al., 2024). Moreover, Huang et al. (2018b) found that sediment samples had a larger amount of *mic* gene copy numbers (2-MIB synthase gene) than water samples. Gaget et al. (2022) highlighted that benthic cyanobacteria are major sources of geosmin and 2-MIB in source waters, and specifically designed monitoring techniques are required to monitor them, because standard water sampling methods are inadequate (Gaget et al., 2020). The origin of 2-MIB synthase genes in Cheney reservoir in Kansas (USA) was linked to a periphytic cyanobacterial mat that was attached to the concrete drinking water inlet structure (Otten et al., 2016), which would be missed with standard water sampling techniques.

In this study, the copy number of the 2-MIB synthase gene (mic) in the whole community normalised for the total copy numbers of 16S rRNA, revealed a good correlation ($R^2 = 0.635$) and indicates that this qPCR method can be used to detect risk of 2-MIB concentration. Similarly, several field and culture studies used qPCR to detect the 2-MIB gene and found correlations with 2-MIB concentrations (Chiu et al. 2016; Wang et al. 2016; Huang et al. 2018; Rong et al. 2018; Kim et al. 2020; Devi et al. 2021; Lee et al. 2023). The fact that *mic* gene copy numbers were a good indicator of 2-MIB concentrations indicate that this is a good detection method but works best when cells of the 2-MIB producers are present in the sample, which highlights the importance of using the correct sampling protocol.

3.6.7 Other factors that influence community and 2-MIB

Examples from "single system experiments" replicated in time could, for example, show strong "year effects" and "season effects" in response to the same manipulation (Beisner & Peres-Neto, 2009; Werner et al., 2020). What is the replicability of the experiment, especially at other times of the year? Is a large part of the community that starts developing related to the history of the community and the size of the inoculum? Differences between flasks of the same treatment in this experiment, can be a result of non-homogeneous distribution of species in the source water used to initiate the experiment. There were notable floating clumps and organic material that could have

been home to many benthic cells (cyanobacteria, green algae and diatoms). If a species was prevalent in autumn and is able to survive over winter (possibly in the sediment layer), it will most likely bloom next year when temperatures, light and nutrients are sufficient (Yao et al., 2022). Species with the largest initial inoculum can grow quickly when the water conditions become optimal, and this competitive advantage over other species enables them to dominate the community. The importance on community composition of the so-called priority effect, the time of arrival of a species in the community, was identified with laboratory and field studies for many ecosystems including freshwater phytoplankton (Fukami, 2015; Rummens et al., 2018; Sefbom et al., 2015; Van Gremberghe et al., 2009). However, analysis of the initial start community revealed that later dominant *Pseudanabaena* spp. (PCC-7429) and *Leptolyngbya* spp. (SAG 2441) had low abundance at the start of the experiment. It is possible that the light and temperature conditions in the incubator and the limited mixing could have impacted the phytoplankton community that established in the treatments. *Aphanizomenon* spp., which are likely geosmin producers in this reservoir, were dominant in the initial start community but did not develop in the experiment. These planktic species might have required higher temperatures or light conditions, as benthic cyanobacteria are particularly well adapted to grow under low-light conditions (Abeynayaka et al., 2018; Gao et al., 2018).

Moreover, the experiment did not control for grazing, and it turned out that zooplankton (*Daphnia* sp.) were present, but were not recorded. Grazing experiments with geosmin producing cyanobacteria and zooplankton (*Daphnia* sp. and *Simocephalus* sp.) revealed that the grazing was a major mechanism that released geosmin from cells, probably by causing cell lysis (Durrer et al., 1999), which could influence both laboratory and field results.

3.7 Conclusions

Microcosms with a natural phytoplankton community from a Welsh drinking water reservoir were used to identify the effect of different nutrient concentrations and ratios, $NH_4^+:NO_3^-$ and DIN:SRP, on 2-MIB production. The outcomes from the experiments provided answers to the formulated research questions.

1) How do different nutrient ratios ($NH_4^+:NO_3^-$ and DIN:SRP) and concentrations impact the community structure and productivity of phytoplankton?

Elevated NH_4^+-N concentrations with low NO_3^--N lead to 'boom-bust' growth dynamics whilst elevated NO_3^--N concentrations with low NH_4^+-N resulted in a delayed response (four days later) but overall balanced growth. Treatments with high NO_3^--N and high PO_4^{3-}-P were productive until the end of the experiment, whilst treatments with high NO_3^--N and low PO_4^{3-}-P plateaued. Diatoms dominated treatment with low nutrients whilst green algae dominated in high nutrient treatments. Highest cyanobacteria biomass was in high NO_3^--N, low NH_4^+-N and high PO_4^{3-}-P treatment, followed by high NO_3^--N, low NH_4^+-N and low PO_4^{3-}-P.

2) How do phytoplankton community structure and productivity affect the production of 2-MIB?

The overall productivity of the community, the total carbon produced, correlated well with 2-MIB concentrations. Gene copies of the 2-MIB synthase gene and relative abundance of cyanobacteria *Leptolyngbya* spp. (resembling strain SAG 2441) and *Pseudanabaena* spp. (resembling strain PCC-7429) correlated well with 2-MIB concentrations, demonstrating they are the most likely producers of 2-MIB in this community.

3) Can specific nutrient ratios and concentrations be used as an early warning for the onset of 2-MIB production?

Ratios of NH_4^+:NO_3^- and DIN:SRP were less important for the development of benthic 2-MIB producing cyanobacteria in these experiments, than total concentration of nitrogen (N) and phosphorus (P). 2-MIB production by benthic cyanobacteria was related to the abundance in the community and productivity. To enable development of these species, the total concentration of N was most important. They grow better on NO_3^--N than NH_4^+-N because green algae established more quickly than the cyanobacteria in treatments with NH_4^+-N. Treatments with the same total dissolved N, but different levels of total dissolved P revealed that there was a potential threshold between 0.5 and 0.2 mg P/L for benthic 2-MIB producers to develop. Treatments with 0.2 mg P/L and high TDN concentrations developed a significant relative abundance of 2-MIB producers, and subsequent depletion of the P source caused either active (to create a P source) or passive (cell death) 2-MIB release. The concentration of TDN required for the development of the benthic cyanobacteria was probably around 2.5 mg N/L, if sufficient TDP was available.

The findings of this chapter highlight that nitrogen and phosphorus concentrations that stimulate growth of T&O producing cyanobacteria should be specified for individual species, to provide the best early warning of optimal growth. It is essential to realise that laboratory studies which identified individual species effects and field studies with limited temporal frequency of nutrient monitoring have both not yet fully grasped the nutrient drivers of T&O production. It is therefore recommended to perform site-specific field studies with the likely T&O producers, but study them whilst keeping in mind the community effect (phytoplankton and bacteria). High-frequency nutrient monitoring is possible *in situ*, which could provide a better understanding of time-lags and responses to varying nutrient concentrations and ratios, as well as cycling of specific nutrient fractions. This could be linked to community productivity, which can be measured with in-situ methods at a high frequency. Targeted monitoring should be developed to better detect benthic cyanobacteria, for example with visual inspection, fluoroprobes or ex-situ molecular techniques (Gaget et al., 2022; Gaget et al., 2020). Moreover, it is recommended to research the complete community structure with molecular methods (Almuhtaram et al., 2021a; Devi et al., 2021), as this can highlight potential indicator species or co-occurring species for T&O events.

Chapter 4: High resolution monitoring of water quality in drinking water reservoirs for early warning of taste and odour events

4.1 Introduction

Most drinking water treatment plants don't measure T&O frequently year-round, because it is a time consuming and expensive analytical method, but they often do have a decision framework in place with alert levels which set in motion additional treatment and increased monitoring (Kibuye et al., 2021). The delay in laboratory analyses combined with the time it takes for treated water to reach customers and subsequent complaints, does not allow for an early warning of T&O (Kibuye et al., 2021). This has sparked interest to create early-warning systems by measuring other parameters that could indicate a T&O event is going to happen.

4.1.1 Geosmin and 2-MIB producers

Geosmin and 2-MIB can be produced by both benthic and planktic species of cyanobacteria (Chapter 1). Chapter 3 demonstrated that geosmin in Reservoir 1 is most likely produced by planktic cyanobacteria of the genus *Aphanizomenon* whilst 2-MIB is produced by benthic cyanobacteria of the genus *Leptolyngbya* and *Pseudanabaena*. The different growth habitat and species traits influence what environmental conditions drive their growth, which provides important insights into the differences between geosmin and 2-MIB patterns in Reservoir 1.

Planktic *Aphanizomenon* spp. have been identified as geosmin-producers (Watson et al., 2016) and will rely on light availability, a certain water temperature, a stable water column, day-length and nutrients. *Aphanizomenon* spp. are tolerant to low nitrogen (due to N_2-fixation capabilities) and low carbon, but sensitive to mixing, poor light conditions and low phosphorus (Reynolds, 2006). Buoyancy regulation gives them a competitive advantage in periods of vertical stability, and N_2-fixation enables them to thrive under N-limitation and at low N:P ratios (Chen et al., 2020; Miller et al., 2013). The optimal growth temperature for several *Aphanizomenon* spp. was 14 - 32 °C (Miller et al., 2013; Pham et al., 2023; Üveges et al., 2012; Yamamoto & Nakahara, 2005), which is lower than that of toxin producing cyanobacteria *Microcystis* spp. (27.5

- 32 °C), enabling *Aphanizomenon* spp. to dominate earlier in the growing season (Miller et al., 2013). Longer light hours or continuous light (day-length) enhanced growth of *Aphanizomenon* spp., and growth limitation happened at 10:14 light:dark photoperiod (Pham et al., 2023; Yamamoto & Nakahara, 2005). Optimal light availability (irradiance) for *Aphanizomenon* spp. was found to be between 55.3 and 199 µmol photons m^{-2} s^{-1} by Yamamoto & Nakahara (2005), and >150 µmol m^{-2} s^{-1} by Üveges et al. (2012). Planktic species often grow rapidly when the conditions are optimal, forming blooms with high biomass, but they can also rapidly disappear when conditions change (Harris et al., 2024; Yamamoto & Nakahara, 2009).

Benthic *Leptolyngbya* spp. and *Pseudanabaena* spp. have been identified as 2-MIB producers (Watson et al., 2016) and they require a certain water temperature, sufficient light availability but they are also low-light acclimated, a stable water column with low sediment resuspension and nutrients (Vadeboncoeur et al., 2021). Benthic cyanobacteria can access nutrients in the pore-water from sediment (Jähnichen et al., 2011) and they are low-light acclimated so they can survive periods of environmental disturbance like resuspension and sediment burial (Abeynayaka et al., 2018; Gao et al., 2018). *Pseudanabaena* spp. and *Leptolyngbya* spp. in Reservoir 1 grew as a benthic mat (Chapter 3), but literature suggests that *Pseudanabaena* spp. can be planktic (Yue et al., 2024) or benthic (Gaget et al., 2017; Perri et al., 2024), whilst *Leptolyngbya* spp. were described as benthic (van der Grinten et al., 2005). Planktic *Pseudanabaena* spp. thrive in turbid mixed layers, capable of withstanding low light but sensitive to flushing (Reynolds, 2006). Planktic *Pseudanabaena* spp. required relatively high nitrogen compared to phosphorus and tolerated phosphorus deficiency (Gao et al., 2018; Xiao et al., 2024). Optimal growth of planktic *Pseudanabaena* spp. occurred at 20 – 30 °C (Gao et al., 2018) and 14 - 23 °C (Xiao et al., 2024), with optimal irradiance of 27 µmol photons m^{-2} s^{-1} (Gao et al., 2018). Optimal growth of benthic *Leptolyngbya* spp. occurred at 25 °C and at irradiances 5 – 200 µmol photons m^{-2} s^{-1}, with maximum growth rate at 40 µmol photons m^{-2} s^{-1} (van der Grinten et al., 2005). *Leptolyngbya* spp. and *Pseudanabaena* spp. were related to stratification and shallow waters (Xiao et al., 2024; Yue et al., 2024). The growth speed of these benthic cyanobacteria is expected to be slower than planktic species (Vadeboncoeur et al., 2021), but it is complicated to generate growth curves and estimate biomass, as they grow heterogeneously (Catherine et al., 2013).

Periods with high risk of geosmin and 2-MIB concentrations in a lake or reservoir can be predicted based on the hydrodynamics, temperature, light availability and nutrient dynamics; either related to external inputs (tributary runoff) or internal processes (Chapter 1, Section 1.5, 1.6, 1.7 and 1.8). These environmental factors will influence the phytoplankton community structure and determine if the geosmin and 2-MIB producing cyanobacteria species are present, but this does not necessarily cause a geosmin and 2-MIB event. Species traits of a cyanobacterial species determine its optimal environmental conditions for growth, but it also depends on the food web (like zooplankton grazing) and the phytoplankton and bacterial community, as cyanobacteria may rely on co-occurrence with other organisms (Louati et al., 2015; Reinl et al., 2022). Moreover, there is a lot of uncertainty with regards to geosmin and 2-MIB production and release, and how it varies between species (Chiu et al., 2016; Saadoun et al., 2001; Wang et al., 2016; Watson & Ridal, 2004). T&O production is either positively (Alghanmi et al., 2018; Giglio et al., 2011; Zimba et al., 1999) or negatively (Kutovaya & Watson, 2014; Liu et al., 2009; Pattanaik & Lindberg, 2015; Saadoun et al., 2001; Shen et al., 2020; Wang & Li, 2015) related to growth and productivity, and release might happen passively (Zhang et al., 2016), actively (Ma et al., 2013) or as a mixture of both (Wang & Li, 2015; Watson, 2003). Even though certain bacteria can degrade these compounds (Clercin et al., 2022; Westerhoff et al., 2005), the persistence of geosmin and 2-MIB in the water column causes dispersion away from the source, which complicates source-detection. Geosmin and/or 2-MIB production by benthic cyanobacteria will most likely be detected in the water column as extracellular compounds, whilst planktic cyanobacteria are present in the water column and can be detected both intracellularly and extracellularly. Chiu et al. (2016) proposed that 2-MIB compounds were produced by benthic cyanobacteria in Taiwanese reservoirs, which diffused to the water column. They did not detect the 2-MIB synthase gene (mic) as evidence of the producers, despite 2-MIB concentrations being present in water samples.

4.1.2 Potential drivers and predictors of geosmin and 2-MIB

NH_4^+-N, the reduced form of nitrogen is directly bioavailable to phytoplankton, whilst the oxidised form of nitrogen NO_3^--N requires the nitrate reductase enzyme for uptake (Glibert et al., 2016). It has been suggested that elevated concentrations of NO_3^- favour diatom growth whilst elevated NH_4^+ concentrations favour dinoflagellates,

chlorophytes and cyanobacteria (Andersen et al., 2020; Donald et al., 2013; Donald et al., 2011; Erratt et al., 2020; Glibert & Berg, 2009). PO_4^{3-}-P is highly bioavailable to cyanobacteria. Aubriot and Bonilla (2012) found that PO_4^{3-}-P is taken up by bloom-forming cyanobacteria within 15-25 minutes and increases productivity rates instantly. Low TN:TP concentrations are often used to indicate high risk of cyanobacteria, and a high NH_4^+:NO_3^- ratio is a potential driver of T&O production (Harris et al., 2016) (See Chapter 3).

Reservoir processes

In summer, periods of water column stability will increase organic matter decomposition, which lowers the redox potential. Subsequent anaerobic conditions cause enhanced denitrification (converts NO_3^--N into N_2 gas) and internal loading of PO_4^{3-}-P, NH_4^+-N and dissolved metals like manganese and iron (Jensen & Andersen, 1992; Jeppesen et al., 2009; Jones & Welch, 1990; Yao et al., 2023) (see Chapter 1, Section 1.2.2 for more details). Physical (wind) mixing and sediment resuspension weakens the physical boundary conditions at the sediment-water interface and allows dispersion of the accumulated compounds (DOC/TOC, PO_4^{3-}-P/TP, NH_4^+-N, manganese, etc.) into the surface water layer (Jones & Welch, 1990; Santschi et al., 1990).

Internal loading of nitrogen (NH_4^+-N) and phosphorus (PO_4^{3-}-P) has been linked to enhanced risk of algal blooms (Hoffman et al., 2022; Yao et al., 2023) and in turn, when these blooms die off and sink to the bottom of the lake, the process starts again as these nutrients are recycled (Zhu et al., 2013). It has been suggested that the composition of cyanobacterial organic matter, which settled on the sediment, enhanced phosphorus cycling (Wang et al., 2019b). Ma et al. (2013) found that decaying cyanobacterial blooms released high levels of nitrogen and phosphorus and simultaneously released high levels of several taste and odour compounds, but geosmin and 2-MIB were more related to the live cyanobacteria growth phase. A negative correlation between discharge and 2-MIB concentrations in a drinking water reservoir in North Carolina (USA) led Paerl et al. (2022) to suggest that 2-MIB producers are favoured under more stratified conditions.

Tributary processes

NO_3^--N in tributaries comes from agricultural land and urban areas and depends on hydrological connectivity due to its tendency to accumulate in groundwater, which

causes a delayed response to storm events (Chen et al., 2018; Heathwaite et al., 1996). High concentrations of NH_4^+-N and PO_4^{3-}-P in streams are related to a quick response to storm events through overland runoff (Chen et al., 2018), as these compounds adsorb to soil particles. Sources of NH_4^+-N and PO_4^{3-}-P are anthropogenic pollution from urban areas or agricultural catchments dominated by livestock or arable crops in which N-rich fertilisers are used (Chen et al., 2018; Dodds & Whiles, 2020; Heathwaite & Johnes, 1996; Jeppesen et al., 2011). In areas of livestock farming, the inputs from manures and slurries are associated with DON and DOP such as urea, amino acids and phytic acids (Turner et al., 2002), which are often at least partly available to phytoplankton (Chen et al., 2020; Mackay et al., 2020; Reinl et al., 2022).

Agricultural runoff strongly influences harmful algal bloom risk, and the growing use of urea-based nitrogen fertilizer in the US further heightens this risk, as urea is more bioavailable than nitrate and can cause changes in phytoplankton communities (Glibert, 2020). Chen et al. (2018) showed that 10 days after a peak runoff event, a reservoir in southeast China experienced an algal bloom with increased chlorophyll *a* concentrations, and cyanobacterial growth was related to NH_4^+-N and PO_4^{3-}-P concentrations rather than NO_3^--N. NH_4^+-N inputs in tributary inflows were a likely cause of 2-MIB and geosmin production in two UK reservoirs (Perkins et al., 2019). Painter et al. (2023) showed that flow management had a large influence on cyanobacterial blooms in a Canadian reservoir, as natural landscape runoff changed to managed flows from an upstream reservoir, which caused shifts in chemistry and increased cyanobacterial blooms. Clercin and Druschel (2019) studied Eagle Creek Reservoir (USA) and found a time-lag of 37 days between peak inflows from a river and 2-MIB and geosmin. Investigating the site-specific lag-time between tributary inflow and T&O event is a useful tool that can provide an early warning.

4.1.3 Cyanobacteria models

Models to predict planktic cyanobacterial blooms have been attempted by many studies, but predictors are often site-specific and vary between species. Commonly used predictors were; water column stability (Hecht et al., 2022; Persaud et al., 2015), water temperature and transparency (Descy et al., 2016; Hecht et al., 2022; Recknagel et al., 2016), inorganic nitrogen and phosphorus concentrations (Descy et al., 2016; Hecht et al., 2022; Persaud et al., 2015; Recknagel et al., 2016), meteorological factors like temperature, wind speed, wind direction, rainfall and surface irradiance or

sunshine hours (Descy et al., 2016; Hecht et al., 2022; Ndong et al., 2014; Persaud et al., 2015; Zhang et al., 2012) as well as cyanobacteria biovolume two weeks earlier (Persaud et al., 2015) (Chapter 1). Not many studies have attempted to predict benthic cyanobacterial growth, but they found useful predictors in; low water flow (Espinosa et al., 2020), water level related to bathymetry and light availability (Su et al., 2017), high DOC, low TP and low DIN:TP (Lévesque et al., 2012), high N (Heath et al., 2015), low river inflow, high TP, high NO_3^- and day of the year (Perri et al., 2024) (Chapter 1).

4.1.4 T&O models

Several studies attempted to use linear or non-linear models to predict T&O events (Bertone & O'Halloran, 2016; Chong et al., 2018; Dzialowski et al., 2009; Harris & Graham, 2017; Kehoe et al., 2015; Parinet et al., 2013). Key parameters in those T&O models were summarised here, but full details can be found in Chapter 1. Models often used water quality parameters like turbidity, inorganic nitrogen (NO_3^--N and NH_4^+-N), inorganic phosphorus, total phosphorus, total nitrogen, water transparency, water temperature and chlorophyll *a*, but in some cases, silica, potassium, TOC, UV absorbance and potential redox are mentioned (Harris & Graham, 2017; Parinet et al., 2013) or reservoir volume (Bertone & O'Halloran, 2016). Non-linear models can incorporate microbial abundance data, phytoplankton data and hydrodynamics, which could be modelled separately (Chong et al., 2018). Recent advancements in artificial intelligence have enabled T&O predictions based on larger datasets (Kehoe et al., 2015; Parinet et al., 2013), but this can make the result hard to interpret. Another important predictor for geosmin and 2-MIB concentrations could be the day of the year, due to strong seasonal effects of cyanobacterial growth (Harris & Graham, 2017). This strong seasonality could mean that other water quality parameters with similar seasonality are good predictors but not necessarily related to geosmin and 2-MIB production.

4.1.5 Sensors

Almost all T&O models use parameters that can currently only be measured in a laboratory. For early-warning systems, it is better to use real-time sensors that are connected to the internet as they can provide direct evidence for management (Painter et al., 2023). Commercially available water quality sensors that might be useful for T&O prediction are: water temperature, water level (pressure), pH, oxidation-reduction potential (ORP), electrical conductivity (EC), turbidity, dissolved oxygen (DO),

photosynthetically active radiation (PAR), dissolved organic matter (DOM) as CDOM (chromomorphic or coloured DOM) or fDOM (fluorescent DOM), photosynthetic pigments (chlorophyll a, phycocyanin, phycoerythrin) and nutrients (NO_3^--N, NH_4^+-N, $PO_4^{3-}-P$, TP). Standard sensors on a multiparameter probe often include EC, pH, ORP, DO, water temperature, water depth and turbidity. Chlorophyll a and phycocyanin sensors measure fluorescence of the pigment *in situ*, which estimates total phytoplankton biomass and cyanobacteria-specific biomass, respectively (Painter et al., 2023). CDOM sensors measure absorbance and fDOM sensors measure fluorescence, but they both provide estimates of DOM and can be used to estimate DOC in aquatic environments (Danhiez et al., 2017).

Nutrients NO_3^--N, NH_4^+-N, $PO_4^{3-}-P$ and TP can also be measured in real-time due to recent advancements in sensing technology (Beaton et al., 2012; Beaton et al., 2017; Blaen et al., 2016; Clinton-Bailey et al., 2017; Daniel et al., 2020; Grand et al., 2017; Mowlem et al., 2021). Nitrate (NO_3^--N) can be measured with three different technologies: ion-selective electrode (electrochemical), UV-Visible absorbance (optical) and a wet chemistry analyser (reagent-based and colorimetric). Wet chemistry analysers are called "lab-on-chip" as they do the colorimetric laboratory analysis on a microfluidic scale (Beaton et al., 2012; Beaton et al., 2017). Ammonium (NH_4^+-N) can be measured with an ion-selective electrode and a wet chemistry analyser (reagent-based and colorimetric). Phosphate ($PO_4^{3-}-P$) and total phosphorus (TP) can currently only be measured with a wet chemistry analyser (reagent-based and colorimetric). The different technologies have a different cost; ion-selective electrodes are the cheapest (few hundred pounds), UV-Visible absorbance sensors are a lot more expensive (a few thousand pounds) and wet chemistry analysers are the most expensive (tens of thousands of pounds). The different technologies also come with differences in sensor performance (sensitivity, limit of detection, precision and accuracy), maintenance requirements, response time etc. which will determine the best choice of technology for a specific parameter and a certain monitoring purpose (Blaen et al., 2017; Daniel et al., 2020).

4.1.6 Sensors for cyanobacteria and T&O prediction

The risk of T&O events depends on three key processes in lakes or reservoirs which influence each other; water column conditions (hydrodynamics, light availability and temperature), nutrient dynamics (external and internal) and phytoplankton productivity

and biomass. Johnston et al. (2024) tested several *in situ* sensors for water quality and meteorological parameters to monitor harmful algal blooms in lakes and provided useful insight in opportunities and challenges. Water column conditions can be investigated with information from a weather station (potentially forecasts) or measured in the water column with sensors for water level as well as water temperature and PAR (potentially at different depths). Nutrient dynamics can be measured with sensors or analysers for fDOM/CDOM, NO_3^--N, NH_4^+-N and PO_4^{3-}-P. Phytoplankton productivity can be measured with sensors for DO, pH and ORP, whilst chlorophyll *a* combined with phycocyanin sensors can provide an estimation of phytoplankton and cyanobacterial biomass as well as the cyanobacteria dominance within the total phytoplankton community. Sensors that can separate certain phytoplankton groups based on fluorescence also exist, which can provide insight into the phytoplankton community composition (Johnston et al., 2024). Turbidity and EC can indicate many different processes, but they can provide a baseline water quality, and periods of change combined with catchment knowledge can indicate important events in the ecosystem (see Chapter 1).

Several studies used real-time water quality sensor data for water treatment management. Dissolved oxygen (DO) sensors were used in a drinking water treatment plant to estimate phytoplankton productivity and die-off, which could predict T&O concentrations a few weeks later (Chen et al., 2019). Fluorescence probes (chlorophyll *a* and/or phycocyanin) in drinking water treatment plants can help optimise T&O treatment (Zamyadi et al., 2016). Chlorophyll *a* and phycocyanin sensors were also deployed in source waters (reservoirs) to predict cyanobacteria risk (Almuhtaram et al., 2021b; Carey et al., 2021; Cotterill et al., 2019; Painter et al., 2023), but without species information this does not directly reflect T&O risk (Kibuye et al., 2021) and it won't detect benthic T&O producers (Almuhtaram et al., 2021b). Additional DO sensors at different depths can provide insight in hydrodynamics, which can help predict anoxia caused by cyanobacteria bloom decay (Wentzky et al., 2019) and predict iron and manganese with additional water temperature sensors, weather station data and a UV-Visible spectrophotometer (Hammond et al., 2023). Discharge from major tributaries could be measured with sensors in real-time, as this was found to predict geosmin and 2-MIB in two different US reservoirs (Clercin & Druschel, 2019; Paerl et al., 2022). Combining for example phycocyanin sensor measurements at the

raw water intake with laboratory results of toxins and T&O measurements and weather forecasts could potentially provide the best evidence for cyanobacteria management at drinking water treatment plants (Jalili et al., 2022).

Algal bloom monitoring is often done with real-time sensors deployed on buoys, which measure meteorological parameters, standard water quality parameters and chlorophyll *a* and phycocyanin (Painter et al., 2023). To the best of my knowledge, real-time nutrient analysers have not yet been deployed for algal bloom monitoring and T&O prediction in lakes or reservoir, even though the literature suggests that nutrient concentrations and ratios (NH_4^+:NO_3^- and DIN:SRP) could be valuable predictors for T&O events.

This chapter aims to provide further knowledge regarding the second overarching hypothesis: Nutrient concentrations and ratios are an important factor in predicting geosmin and 2-MIB events. However, the main aim for this chapter is to answer the third overarching hypothesis: Sensor data can provide an early warning for geosmin and 2-MIB events.

Nine years of laboratory data from a Welsh drinking water reservoir was used to detect persistent trends in geosmin and 2-MIB events and identify potential processes that could be measured with real-time sensors. Additionally, several water quality sensors were deployed in the reservoir in 2022-2023, including a NH_4^+-N ISE, a NO_3^--N ISE and a NO_3^--N wet chemical analyser. This data was used to identify short-term indicators of geosmin and 2-MIB events that could potentially be used as an early warning. The following research questions were addressed:

1) What are drivers of geosmin and 2-MIB production in a drinking water reservoir?
2) What is the lag-time between environmental drivers and geosmin and 2-MIB detection?
3) Can sensor data in the reservoir be used to predict geosmin and 2-MIB events?

4.2 Methods

4.2.1 Study site

This is a case study for a Dŵr Cymru Welsh Water (DCWW) drinking water reservoir in North Wales (Figure 4.1, Figure 4.2 and Figure 4.3). Reservoir 1 is the smaller of two artificial, shallow, alkaline reservoirs that have a catchment area of 226 hectares. Reservoir 2 is the primary source for drinking water production and Reservoir 1, which is located upstream, acts as a storage reservoir in periods of drought (Figure 4.2). As Reservoir 1 is a back-up reservoir, it is not often used for abstraction, especially in recent years due to the poor water quality. Reservoir 2 is partially fed with pumped inflows from a nearby river, whilst Reservoir 1 is purely fed by surface water drainage from two small tributaries (Figure 4.3). Management decisions about water use are based on monthly manual sampling, resulting in time-lags between events in the reservoir and detection in the monitoring program. Because of this, water from Reservoir 1 gets used infrequently and inefficiently; when water quality is deemed poor, it can be several months before the monthly sampling regime determines that quality has improved sufficiently. Natural Resources Wales (NRW) also sampled both reservoirs from 2010 to 2014 (and continue to do so) to review lake eutrophication and concluded that these reservoirs have a poor water quality status and high impact of eutrophication (Hatton-Ellis, 2016). 87% of the land use around Reservoir 1 is intensive agriculture, as improved grassland and arable fields, and hence the reservoir receives frequent nutrient inputs (Figure 4.2). The source apportionment estimated that 97.2% of all catchment N contribution was from agriculture (Hatton-Ellis, 2016).
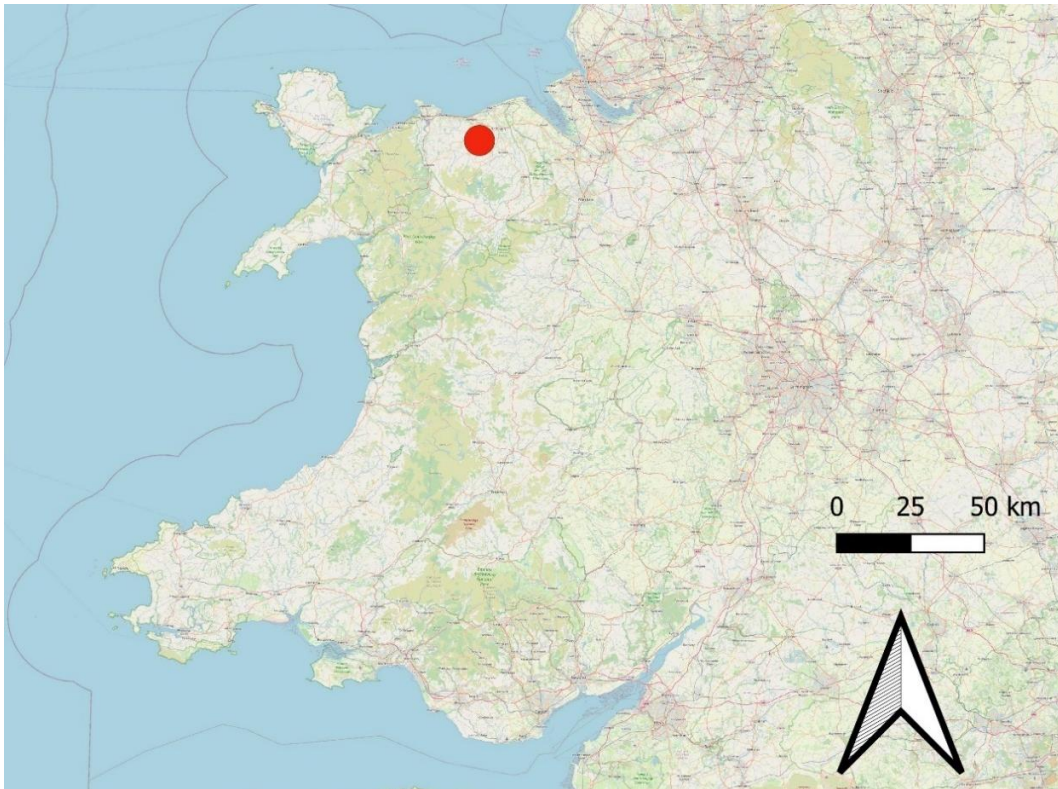
Figure 4.1. Map of Wales and reservoir location.

Land cover:
- Broadleaved woodland
- Coniferous woodland
- Arable
- Improved grassland
- Neutral grassland
- Calcareous grassland
- Acid grassland
- Fen, Marsh and Swamp
- Heather and shrub
- Heather grassland
- Bog
- Inland rock
- Saltwater
- Freshwater
- Supralittoral rock
- Supralittoral sediment
- Littoral rock
- Littoral sediment
- Saltmarsh
- Urban
- Suburban

Reservoir 2

Reservoir 1

Tributary

Farm

Figure 4.2. Land cover map 2022 which visualises the location of Reservoir 1, Reservoir 2, the major tributary and the farm upstream along the tributary. Based upon Land Cover Map 2022 © UKCEH 2022. Contains Ordnance Survey data © Crown Copyright 2007, Licence number 100017572.

142
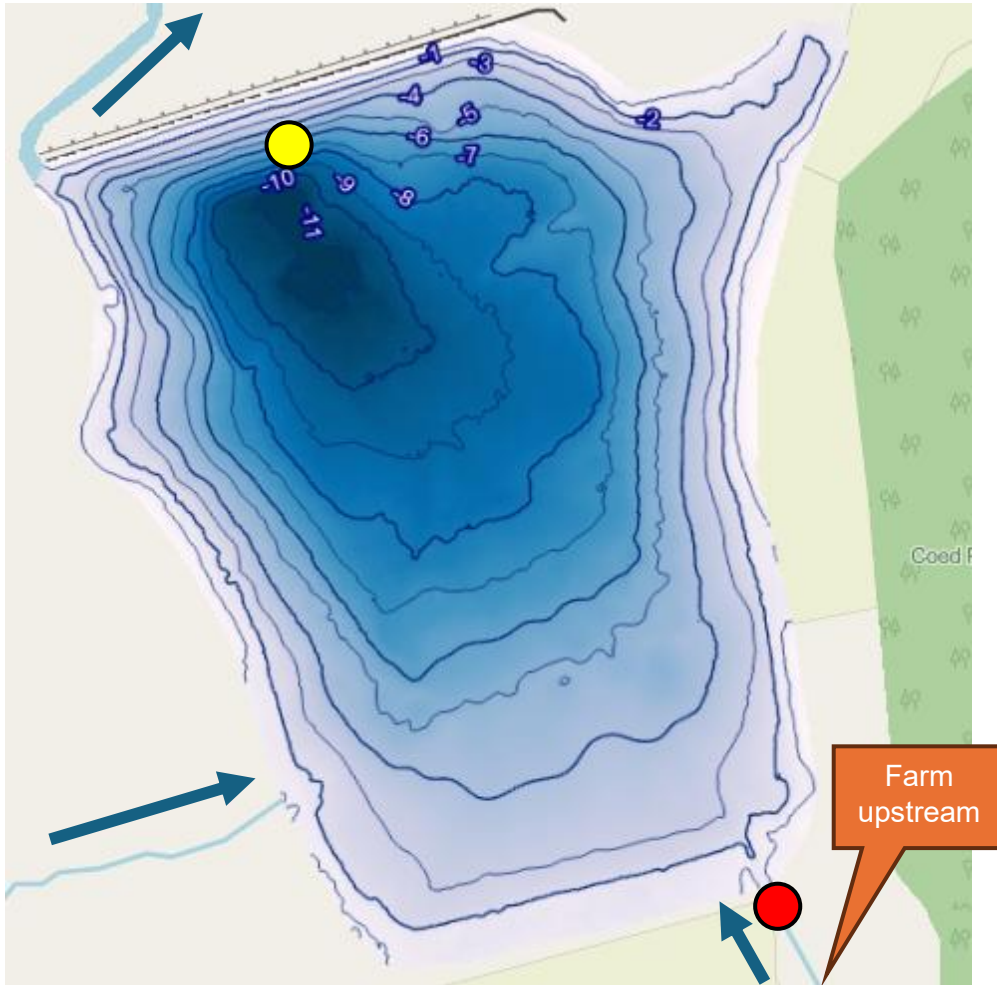
Figure 4.3. Bathymetry map of the reservoir where dark blue is deep and light blue is shallow, depth contour values are accurate when the reservoir is 100% filled. DCWW sampling location of the tributary (red dot) and abstraction point (yellow dot) are visualised. Blue arrows indicate inflow from tributaries and outflow through the spillway, farm upstream of the main tributary is shown.

4.2.2 Data sources

Several sources of data have been used in this study (Table 4.1); they are described in further detail below.

Table 4.1. Sources of data. Location of measurements in reservoir and tributary shown in Figure 4.3.

Dataset	Timespan	Frequency	Location
DCWW manual sampling abstraction	2015 – 2023	Two-weekly or Monthly	Abstraction
DCWW manual sampling tributary	2017 - 2023	Two-weekly or Monthly	Tributary
NRW manual sampling	2010 - 2023	Monthly or few samples a year	Near abstraction

Weather data	2015 - 2023	Daily	Water treatment plant (3km from Reservoir 1)
ClearWater Sensors data	June 2022 – January 2023	2 or 3 hourly	Close to abstraction, hanging from bridge
Hydrolab sensor data	June 2022 – September 2022	2 hourly	Close to abstraction, hanging from bridge
In-Situ sensor data	June 2022 – July 2023	1 hourly	Close to abstraction, hanging from bridge

4.2.3 DCWW data

DCWW manual sampling at the reservoir abstraction started in 2015, and tributary sampling started in 2017. The water samples were analysed at DCWW laboratory Glaslyn in Newport, South Wales, UK, with the following accreditation: ISO/IEC 17025:2019. Several parameters were measured, see Hooper (2023a) for detailed methods, including concentrations of the following nutrients: NH_4^+ as N (0.003 mg/L LoD; ±10%), Total Oxidisable Nitrogen (TOxN) as N (0.48 mg/L LoD; ±10%), NO_3^- as N (calculated as TOxN - NO_2^--N), NO_2^- as N (0.0045 mg/L LoD; ±10%), chloride (3.1 mg/L LoD; ±10%), PO_4^{3-} as P (0.03 mg/L LoD; ±10%), silicate dissolved (0.19 mg/L LoD; ±10%), sulphate dissolved (2.7 mg/L LoD; ±10%), TP (0.055 mg/L LoD; ±10%), DOC (0.12 mg/L; ±10%) and TOC (0.12 mg/L; ±10%). TOxN is NO_3^--N plus NO_2^--N, and subtracting the concentration of NO_2^--N provides a measure of NO_3^--N. Samples were also analysed for concentrations of the following metals: dissolved and total aluminium (0.013 mg/L LoD; ±10%), dissolved and total iron (0.011 mg/L LoD; ±10%), dissolved and total manganese (0.0022 mg/L; ±10%). Additional parameters that are also analysed in the laboratory are: turbidity (NTU), electrical conductivity at 20 °C (μs/cm), pH, algal counts and species identification as well as chlorophyll *a* concentrations (2.00 μg/L LoD; ±25%). The analysis also included concentrations of 2-MIB and geosmin (ng/L) extracted with SPME measured with a GC-MS (quadrupole MS operating in SIM mode) enabling a limit of detection (LoD) of 0.3 and 0.57 ng/L (±25%) for geosmin and 2-MIB respectively (see detailed methods in Chapter 3, Section 3.3.7).

4.2.4 NRW data

The reservoir was sampled as part of the NRW Water Framework Directive Lake Waterbodies Cycle 1 and 2 (Hatton-Ellis, 2016) and sampling has continued until

present. The data are visualised on Data Map Wales (https://datamap.gov.wales/) and the full dataset was requested from NRW. The data contains public sector information licensed under the Open Government Licence v3.0. Data was available for TOxN, TN, TP, PO_4^{3-}-P and chlorophyll *a*, which was analysed in Natural Resources Wales Analytical Services (NRWAS) in Swansea, South Wales, UK, accredited ISO/IEC 17025:2017 by UKAS (national accreditation body for the United Kingdom).

4.2.5 Meteorological data

Meteorological data are from a local weather station Rhyl Number 2, ID 03313099999, Latitude 53.25° N, Longitude -3.5° W, which is ±3 km away from the reservoir. Weather data was downloaded from www.visualcrossing.com. Daily data was used for the following parameters: maximum temperature (tempmax, °C), minimum temperature (tempmin, °C), temperature (temp, °C), precipitation (precip, mm), windgust (mph), windspeed (mph), winddirection (winddir, 0-360 °), cloudcover (0-100%), solar radiation (solarradiation, W/m^2) , solar energy (solarenergy, MJ/m^2), difference in maximum and minimum temperature (tempdiff). Daily solar radiation is calculated as the mean value of the solar radiation for the day whilst daily solar energy is calculated as the sum of the hourly values.

4.2.6 Sensor data

Three water quality sensors/analysers were deployed in the reservoir (Figure 4.4), hanging from the bridge near the abstraction point (yellow star in Figure 4.4 and photos from field deployment in Figure 4.5). The following equipment was used; an OTT HydroMet Hydrolab DS5 multiparameter sonde (OTT HydroMet, Kempten, Germany), a ClearWater Sensors Ltd. nitrate analyser (ClearWater Sensors Ltd., Southampton, UK), and an In-Situ Aqua TROLL 600 multiparameter sonde (In-Situ, Colorado, USA). Regular trips to the field site (every ±2 months in summer and ±4-6 months in autumn, winter and spring) were made to clean the equipment, calibrate the pH, nitrate and ammonium sensors and to replace the cartridge on the ClearWater Sensors nitrate analyser (Figure 4.5).
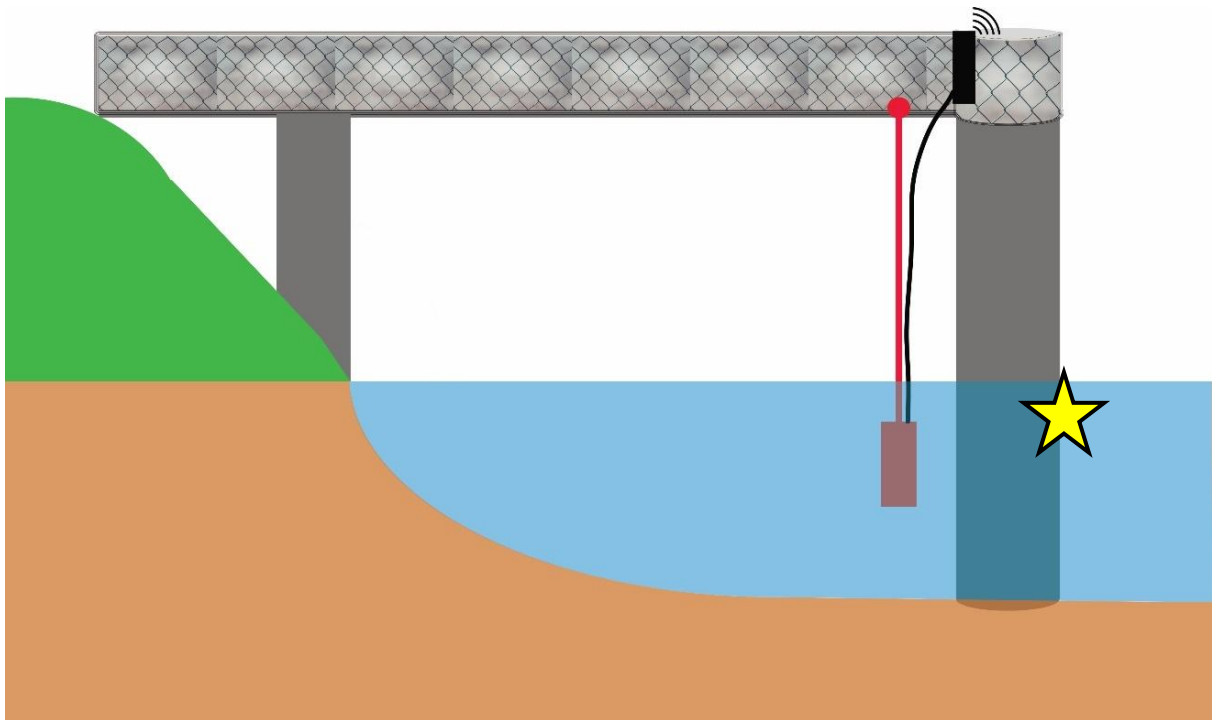
Figure 4.4. Sensor field deployment overview. Yellow star indicates the abstraction location.
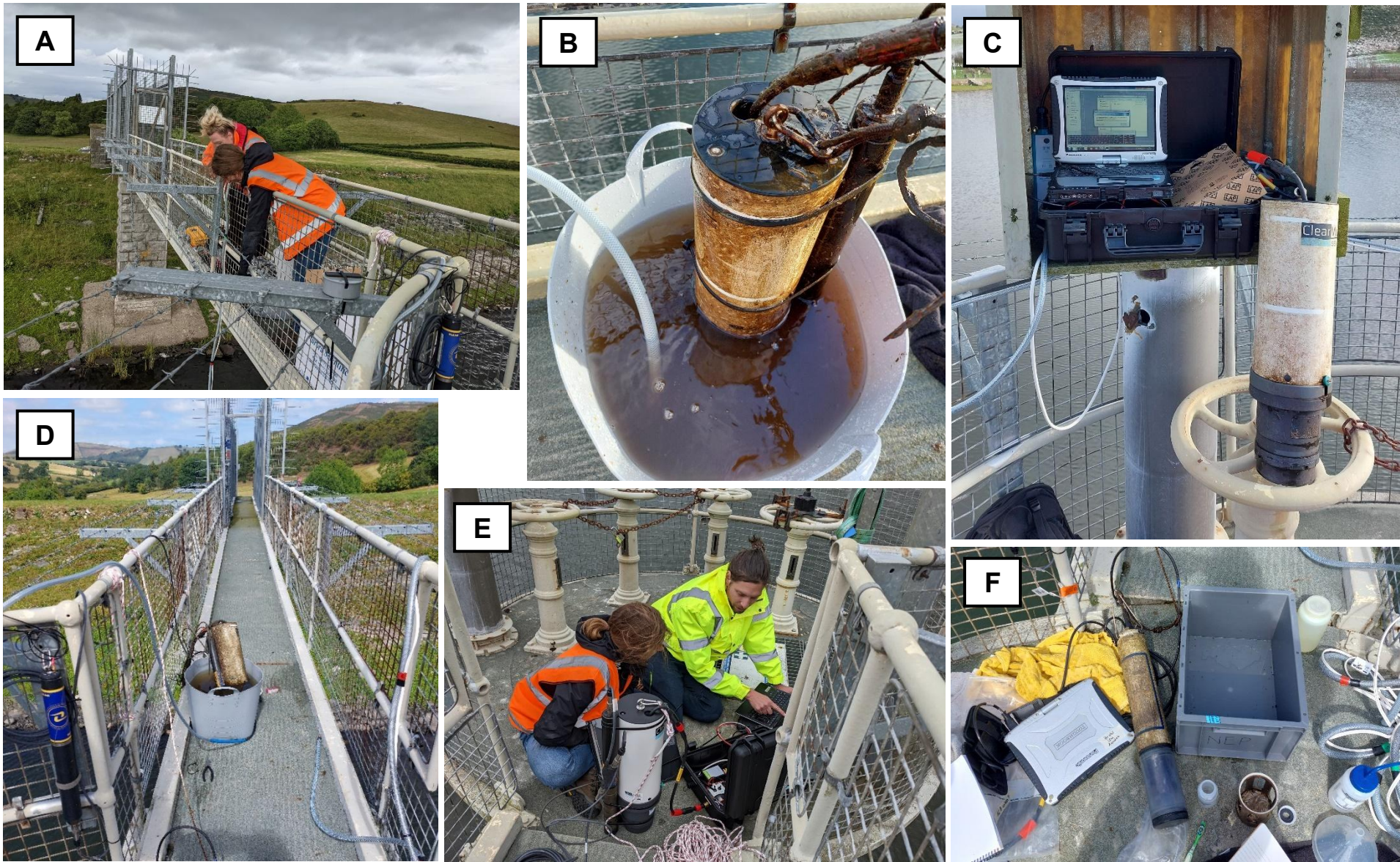
Figure 4.5. Sensor deployment from the bridge (A), Necessary cleaning (B), Communicating with Clearwater Ltd. NO$_3^-$-N analyser (C), In-Situ telemetry and view of bridge (D), Instructions how to set up the Clearwater Ltd. NO$_3^-$-N analyser (E), Hydrolab calibration (F).

147

Hydrolab multiparameter probe

The OTT HydroMet Hydrolab DS5 (OTT HydroMet, Kempten, Germany) had an internal battery pack and internal storage, with no telemetry options (Figure 4.5). This sensor measured water temperature (°C), electrical conductivity (EC, µS/cm), pH, ORP, turbidity (NTU, with a wiper brush), chlorophyll *a* (µg/L) and dissolved oxygen (DO, mg/L and % saturation). Unfortunately, the data was lost from September 2022 onwards and the dissolved oxygen sensor stopped working early on. The data was quality checked and it was decided only to further analyse the chlorophyll *a* data due to data quality concerns with the other data and In-Situ Aqua TROLL data provided a longer dataset for these parameters.

ClearWater Sensors nitrate analyser

The ClearWater Sensors Ltd. nitrate analyser (ClearWater Sensors Ltd., Southampton, UK) was powered by a solar panel and uses lab-on-chip technology and onboard calibration standards to measure nitrate as N. The NO_3^--N analyser is developed for oceanographic work, and has a limit of detection (LoD) of 0.0004 mg N/L (25 nM) with a precision of 0.0003 mg/L (20 nM), details of the technology can be found in Beaton et al. (2012) and Mowlem et al. (2021). In brief, the analyser draws in a sample through a 0.45-µm filter, then it mixes with Griess reagent and undergoes a colour reaction, which is measured by photodiodes and calculated to nitrate concentrations using an onboard nitrate calibration standard. The sensor has been tested in a variety of aquatic environments, including rivers, lakes and seawater. The novelty of this deployment was to connect the analyser via telemetry for real-time data access, which was deployed on the bridge (Figure 4.4, Figure 4.5). In the first few months of the deployment, July and August 2022, there was a lot of biofouling on the filter inlet in between maintenance days, which blocked the inlet and stopped the analyser from taking samples properly. This data has been removed from the dataset. The ClearWater Sensors nitrate analyser stopped working in January 2023 and in their laboratory, it was discovered that the chip needed replacing, which meant the analyser could not be re-deployed for the remainder of the deployment period.

In-Situ multiparameter probe

The In-Situ Aqua TROLL 600 multiparameter sonde (In-Situ, Colorado, USA) had internal batteries, a wiper brush and was connected to an In-Situ VuLink® cellular telemetry device (In-Situ, Colorado, USA) on the bridge (Figure 4.4, Figure 4.5) which

provided real-time telemetry and an extra power supply to the sonde. Sensor parameters were: Water temperature (°C), air temperature (°C), depth (m), electrical conductivity (EC, µS/cm), dissolved oxygen (DO, mg/L and % saturation), pH, ORP, turbidity (NTU) and ion-selective electrodes (ISE) for NO_3^- (mg N/L) and NH_4^+ (mg N/L). The electrical conductivity data was provided as actual conductivity (raw EC values) and specific conductivity (temperature corrected EC values to 25 °C). Specific conductivity data was used for further processing, and the specific conductivity is referred to as electrical conductivity (EC) in this thesis. Turbidity was only measured from June 2022 to mid-August 2022 because the sensor was replaced with the ammonium ion-selective sensor for the remainder of the deployment. There is a period of missing data for the whole month of October 2022, due to a battery failure in the VuLink® device. pH data was used in further analysis, but it was corrected for drift based on the calibrations of pH 7 during maintenance days.

In-Situ drift correction pH

A 3-point calibration was done each field visit with pH 4, 7 and 10. Information about pH 7 pre-measurement (In-Situ app) was used to calculate the drift correction between every calibration period (1 to 4). Values for REAL pH in this case come from the sensor calibration days, indicating the pH value the In-Situ sensor auto-detected for pH 7 buffer, so what it should be after calibration. (Table 4.2).

Table 4.2. pH drift for four periods in the sensor deployment between maintenance (CAL 1- 4) calculated from calibrations on maintenance days.

	Dates	Days	pH 7	pH7 mV	REAL pH	pH drift	pH drift per day
CAL 1	27/06/2022	0	7	-1.7	7		
CAL 1	17/08/2022	51	7.24	-15.2	7.02	0.22	0.004314
CAL 2	17/08/2022	0	7.02	-14.9	7.02		
CAL 2	29/09/2022	43	7.09	-19.1	7.04	0.05	0.001163
CAL 3	29/09/2022	0	7.04	-18.2	7.04		
CAL 3	20/01/2023	113	7.27	-30.6	7.06	0.21	0.001858
CAL 4	20/01/2023	0	7.06	-29	7.06		
CAL 4	01/08/2023	193	7.23	-42.1	7.02	0.21	0.001088

pH drift is the difference between the measurement of pH 7 buffer before and after calibration. For 17/08 that is 7.24 – 7.02 = 0.22 pH units of drift. Subsequently, the pH

drift per day is calculated over the duration of the deployment period. For calibration period 1 (CAL 1) that is 51 days and the pH 7 has drifted by 0.22 pH units, which calculates as a pH drift of 0.0043 per day (Table 4.2). Over that calibration period, each sample point is a certain number of days away from the 27/06/2022, when the calibration was done. Depending on how many days since the initial calibration, the drift in pH units is calculated for each sampling point (Equation 4.1). The first days of the CAL1 period won't have much drift but it incrementally increased until it reaches a drift of 0.22 pH units at the end of the CAL1 period. This calculation assumes that drift happened linearly over the deployment period.

Equation 4.1:

$$pH\ drift\ value = \frac{pH\ drift\ CALx}{number\ of\ days\ CALx} * number\ of\ days\ since\ start\ CALx$$

The x in CALx indicates this calculation can be done for all Calibration periods CAL 1, 2, 3 and 4. In the calculation, "pH drift CALx/number of days CALx" is the same as pH drift per day mentioned in Table 4.2.

To calculate the corrected pH value for each data point, Equation 4.2 was used because pH was drifting upwards.

Equation 4.2:

$$pH - pH\ drift\ value$$

The difference between original pH values and the corrected pH values is visualised in Figure 4.6.
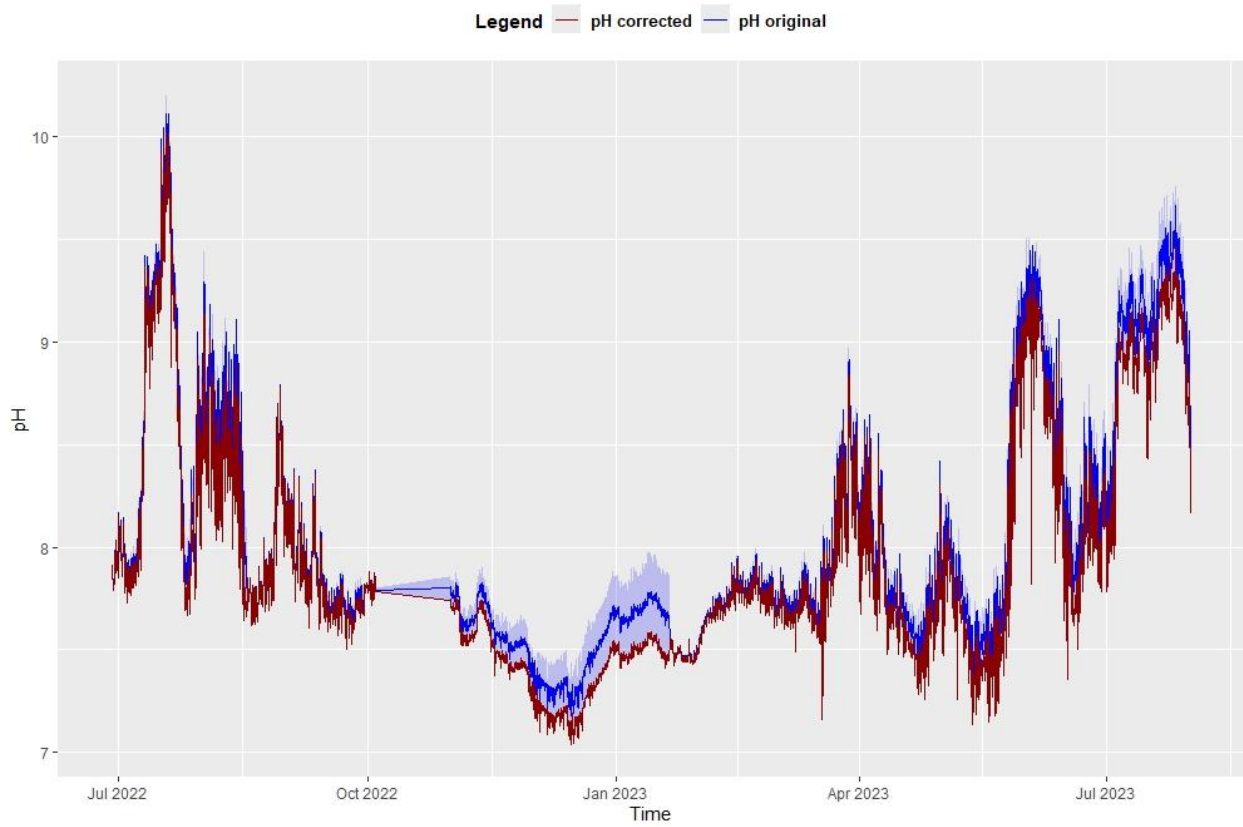
Figure 4.6. pH with uncertainty and corrected pH value

In-Situ NO_3^--N and NH_4^+-N ISE data checks

According to manufacturer's instructions, the NO_3^--N ISE should be able to measure 0.01 to 40,000 mg N/L (±10% or ±2 mg N/L, whichever is greater) and the NH_4^+-N ISE should be able to measure 0.01 to 10,000 mg N/L (±10% or ±2 mg N/L, whichever is greater). The accuracy of ±10% or ±2 mg N/L indicates the relatively low sensitivity of these sensors, which are predominantly used for wastewater monitoring. During field deployment, the NO_3^--N ISE was calibrated with a 5.6 NO_3^--N mg/L prepared calibration solution and the NH_4^+-N ISE with a 1.08 NH_4^+-N mg/L calibration solution, but the sensors did not stabilise in the field and struggled to detect the mV values. Instead of using the field calibrations, raw mV values were converted into NH_4^+-N and NO_3^--N concentrations based on laboratory-determined calibration curves after the deployment. The NO_3^--N sensor worked well in the laboratory tests at environmentally relevant concentrations and a calibration curve could be determined (R^2 = 0.9919), with a range of 0.15 mg N/L at 200 mV (log_{10} = 0.06) to 20.31 mg N/L at 67 mV (log_{10} = 1.33) (Figure 4.7). The NH_4^+-N sensor was tested at environmentally relevant concentrations from 0.021 mg N/L (-52 mV) to 1.777 mg N/L (73 mV) and resulted in an exponential calibration curve (R^2 = 0.9397), which went below 0 for mV around 0.2 mg N/L (Figure 4.7). The calibration curve would have just had positive mV values and a linear trendline if only >0.5 mg N/L concentrations were selected, as NH_4^+-N concentrations <0.5 mg N/L may not be detected by the sensor due to the low sensitivity (Figure 4.7).
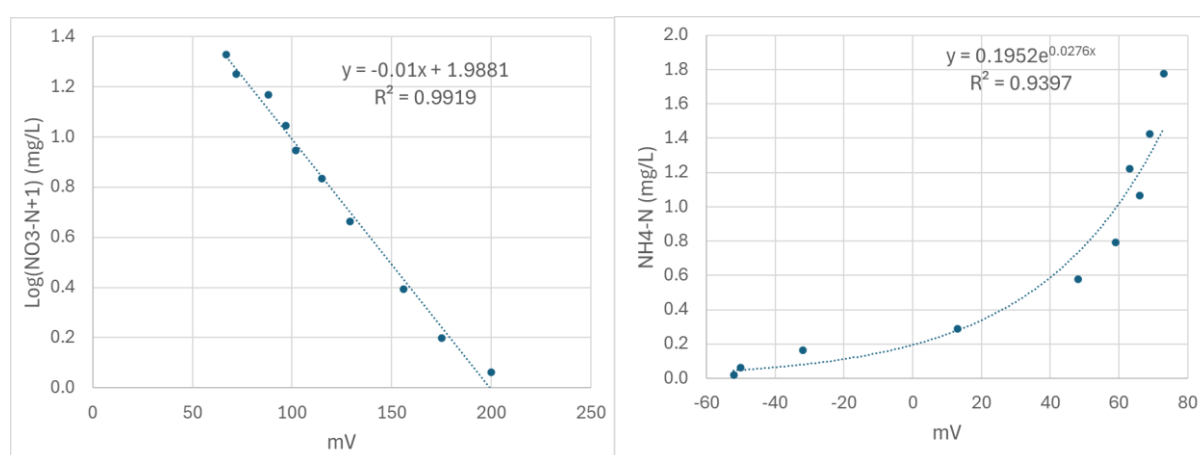
Figure 4.7. Laboratory-determined linear calibration curve for log_{10}-transformed NO_3^--N (left) and exponential calibration curve for NH_4^+-N (right).

Both calibration curves in Figure 4.7 were applied to the laboratory NO_3^--N and NH_4^+-N concentrations that were measured by DCWW throughout the deployment period,

to convert the actual reservoir concentrations into hypothetical mV values (Figure 4.8 and Figure 4.10). The NO_3^--N ISE sensor did not have signs of drift throughout the dataset, but it did not measure the raw mV that it should have been (DCWW laboratory data) consistently; the average mV difference varied between datapoints (Figure 4.8). There seemed to be a mismatch between how the sensor performed in the field compared to in the laboratory. A correction was attempted by offsetting all raw mV values by 51.12 mV (the average difference between the 'lab NO_3^--N converted to mV' and 'mV sensor' values for data that was >1 mg/L of NO_3^--N; the data is marked as a grey shaded box in Figure 4.8), prior to NO_3^--N concentration calculation with the laboratory-determined calibration curve (Figure 4.7). Even after mV correction, the NO_3^--N ISE data did not capture the low or the high NO_3^--N concentrations in the dataset accurately (Figure 4.9), so these concentrations were probably not trustworthy.
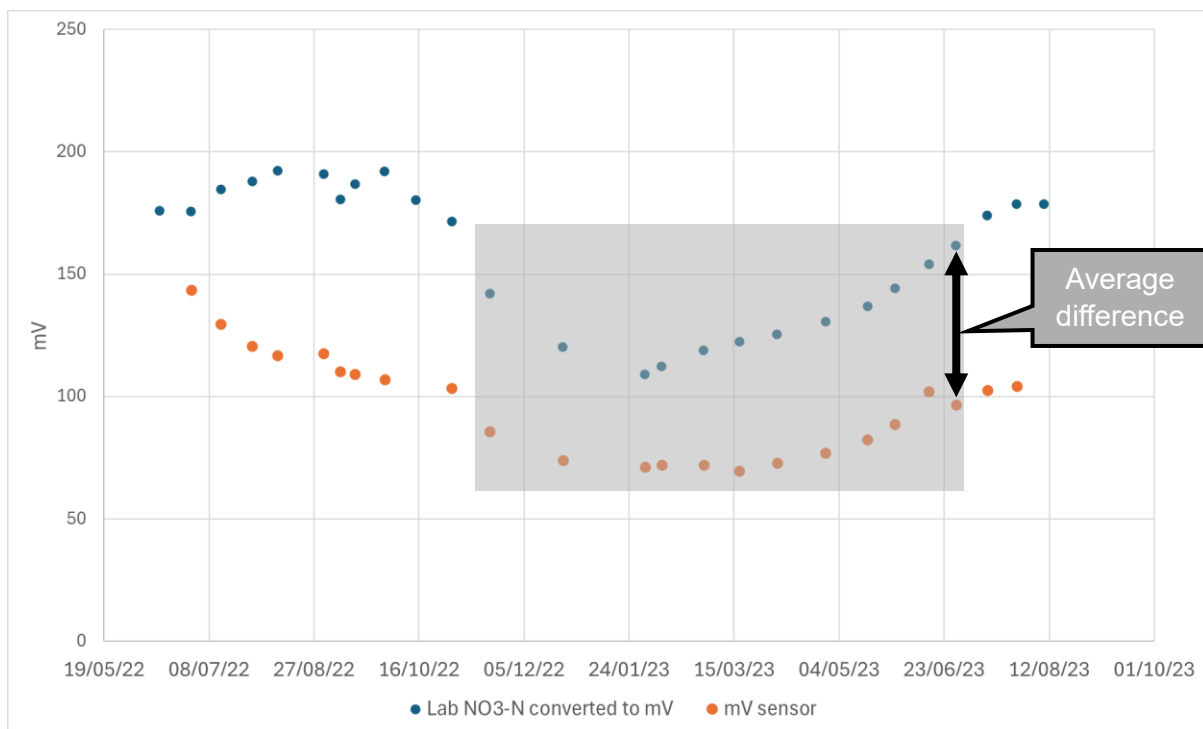
Figure 4.8. NO_3^--N laboratory converted to mV compared to actual mV the ISE sensor measured in the field.
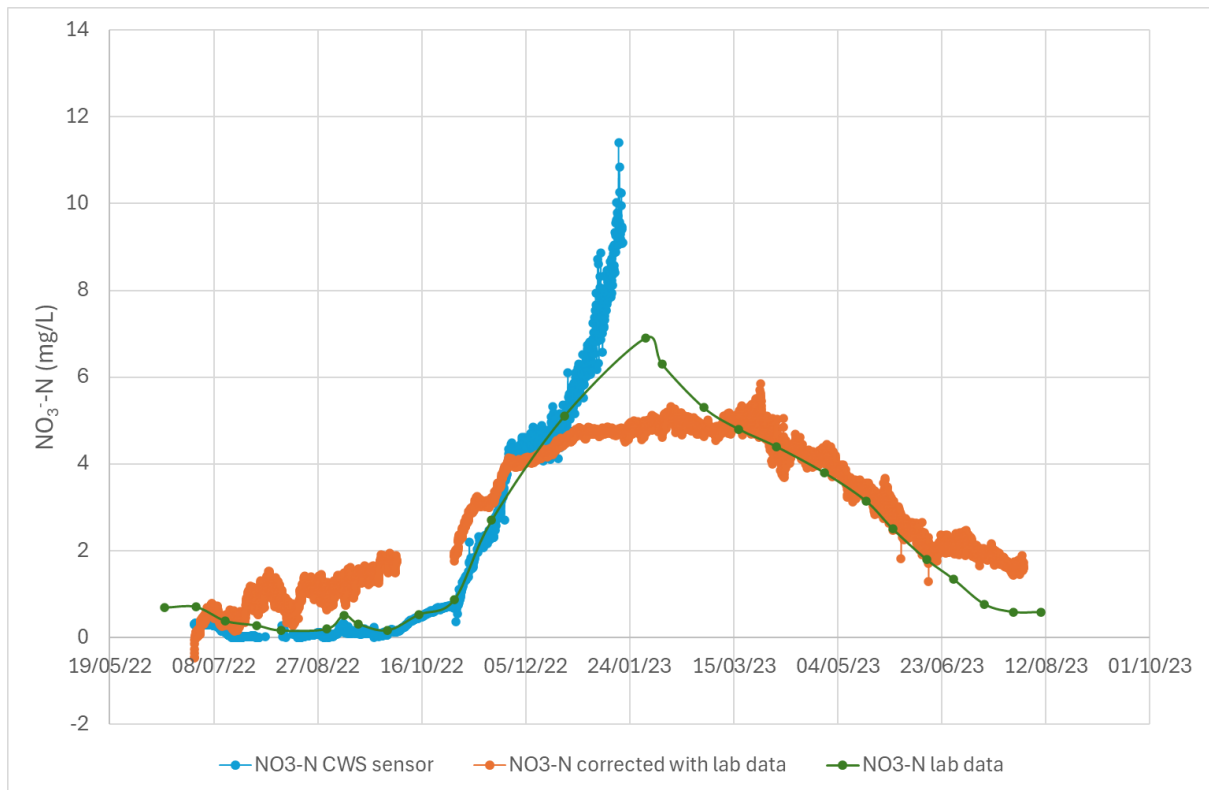
Figure 4.9. NO$_3^-$-N concentration from DCWW laboratory samples (manual), the calculated NO$_3^-$-N concentration from the ISE sensor and the NO$_3^-$-N concentration determined by the ClearWater Sensors nitrate analyser.

The NH$_4^+$-N ISE sensor was only able to measure mV up to -40 (Figure 4.10), and results from the calibration curve (Figure 4.7) showed that the concentrations during the deployment period were probably too low for the sensitivity of this sensor. When the NH$_4^+$ exponential calibration curve was applied to the raw mV data, this also revealed that the sensor was not able to detect differences at these low concentrations (Figure 4.11). This is the reason that data from ion-selective nitrate and ammonium sensors were not used in further data analysis.
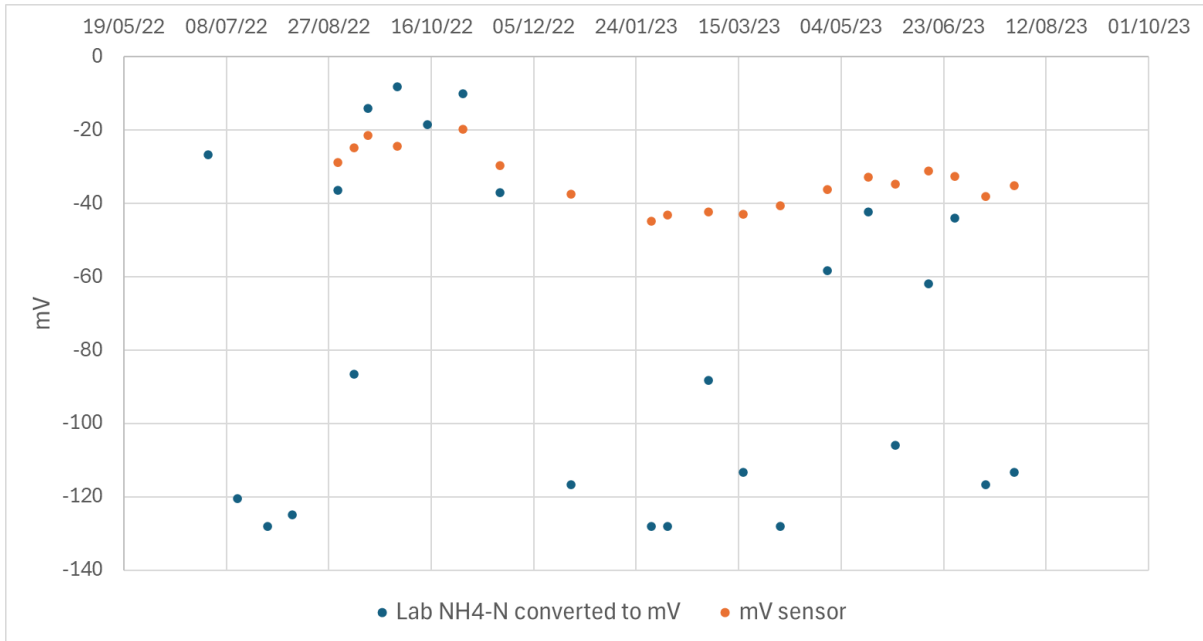
Figure 4.10. NH$_4^+$-N laboratory converted to mV compared to actual mV the ISE sensor measured in the field.
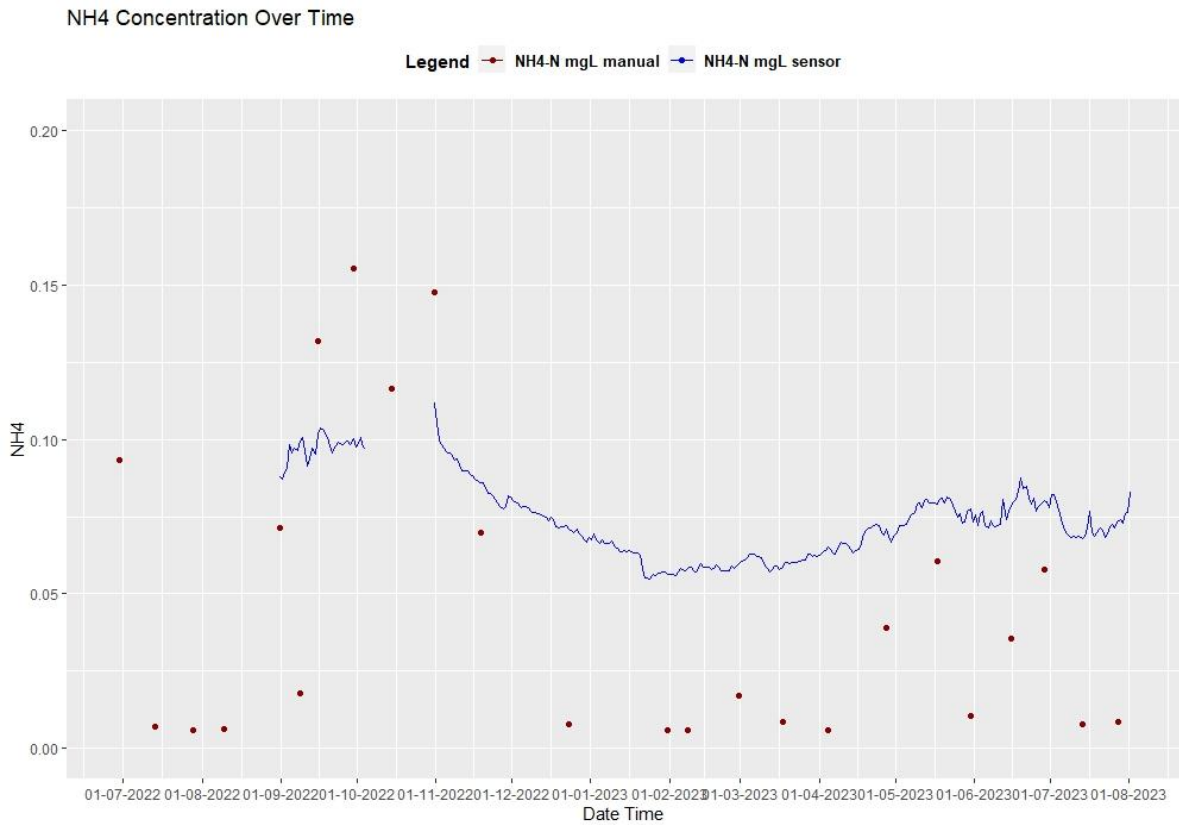
Figure 4.11. NH$_4^+$-N concentration from DCWW laboratory samples (manual) and the calculated NH$_4^+$-N concentration from the ISE sensor.

4.2.7 Pre-processing of DCWW data

All required data pre-processing was completed in R v.4.3.1 (RStudio v.2023.06.2) and associated packages as described.

Abstraction data

DCWW sampled from the abstraction tower (Figure 4.3, yellow dot) and from the tributary (Figure 4.3, red dot) during 2019 and 2019 onwards. The campaigns were not all consistent (different parameters, different sampling intervals etc.) but there was some continuity. In 2022, another sampling campaign started at Reservoir 1 which duplicated the results at certain dates. The duplicated results were compared on a case-by-case basis and merged, parameters with two values were averaged. There were five cases of duplicated results that had the following parameters averaged, with average standard deviation in brackets: NH_4^+-N (0.025 mgN/L), TOxN (0.352 mgN/L), NO_3^--N (0.362 mgN/L), NO_2^--N (0.004 mgN/L), TP (0.008 mgP/L), PO_4^{3-}-P (0.003 mgP/L), chloride (0.424 mg/L), silicate (0.554 mg/L), sulphate (1.598 mg/L), chlorophyll *a* (1.237 µg/L), algae total cell count (3341), blue-green algae cell count (1909), green algae total cell count (2120), 2-MIB (3.574 ng/L) and geosmin (2.234 ng/L).

The abstraction and tributary datasets from DCWW contained many values that were below the limit of detection (LoD) for several parameters (Table 4.3). TP and PO_4^{3-}-P datapoints that were below the LoD, were filled in with NRW data from the closest matching date using the 'zoo' package in R (Zeileis & Grothendieck, 2005). The NRW dates were maximum 15 days away from the DCWW dataset, and an average was taken if there were two close matches, one before and one after the date from the DCWW dataset. In one datapoint, data from NRW and DCWW was available for the same day for TP but not PO_4^{3-}, in which case TP for both was averaged and PO_4^{3-} was filled in from NRW data.

Table 4.3. The % of datapoints below the limit of detection (LoD) for parameters that had <LoD values in DCWW abstraction data (TOxN, TP, chlorophyll a and PO_4^{3-}-P also after gaps filled with NRW data) and DCWW tributary data.

Parameter	% data below LoD in abstraction data DCWW	% data below LoD in tributary data DCWW
NH_4^+-N	6.29 %	3.41 %
TOxN	17.48 %	1.14 %
TOxN with NRW data	12.5 %	-

NO$_3^-$-N	20.28 %	20.45 %
NO$_2^-$-N	18.18 %	19.32 %
TP	**62.68 %**	23.26 %
TP with NRW data	38.46 %	-
Aluminium	5.04 %	-
Chlorophyll *a*	5.04 %	29.41 %
Chlorophyll *a* with NRW data	2.84 %	-
2-MIB	9.22 %	48.75 %
Geosmin	0.71 %	-
PO$_4^{3-}$-P	**82.72 %**	20.29 %
PO$_4^{3-}$-P with NRW data	29.47 %	-
Silicate reactive	16.46 %	-
Sulphate	6.67 %	1.52 %
Aluminium dissolved	**66.23 %**	25.76 %
Manganese dissolved	25.64 %	13.64 %

The 'NADA' package in R (Lee, 2020) was used with to fill in <LoD values (censored data). A Regression on Order Statistics (ROS) model was used to simulate a large amount of data using the normal distribution of the data >LoD (uncensored data), of which a constrained sample, between 0 and LoD for that sample, was taken to replace the <LoD value. This method deals well with differences in the LoD for a parameter throughout the dataset, which can come from changing laboratory methods. However, when the % data below the LoD is more than 60%, the results from the NADA resampling need to be critically assessed. The DCWW abstraction point PO$_4^{3-}$-P, TP and dissolved aluminium data had more than 60% data below the LoD (red numbers in Table 4.3). Fortunately, NRW data could be used to fill values below the LoD for TP and PO$_4^{3-}$-P, which meant that TP had only 38.5% of data below the LoD instead of 62.7%, whilst PO$_4^{3-}$-P went from 82.7% to 29.5% data below the LoD (Table 4.3).

Following this resampling with the 'NADA' package, data was quality checked by ensuring NO$_2^-$-N was not higher than TOxN (no instances) and PO$_4^{3-}$-P was not higher than TP (3 instances, then TP was replaced with the value for PO$_4^{3-}$-P). Moreover, NH$_4^+$-N was calculated from NH$_4^+$ and NO$_3^-$-N was recalculated with the following formula: NO$_3^-$-N = TOxN – NO$_2^-$-N. Additionally, a correlation was established between TOxN and TN in the NRW dataset from Reservoir 1, which created the following formula: TN = 0.9636* TOxN + 0.8282 (Figure 4.12). The formula was used to calculate values for TN in the most complete version of the dataset, as TN is not measured by the DCWW laboratory, only TOxN. Nutrient concentrations were used to calculate

several nutrient ratios, including TN:TP, DIN:SRP and NH_4^+:NO_3^-, which were used in further analysis.
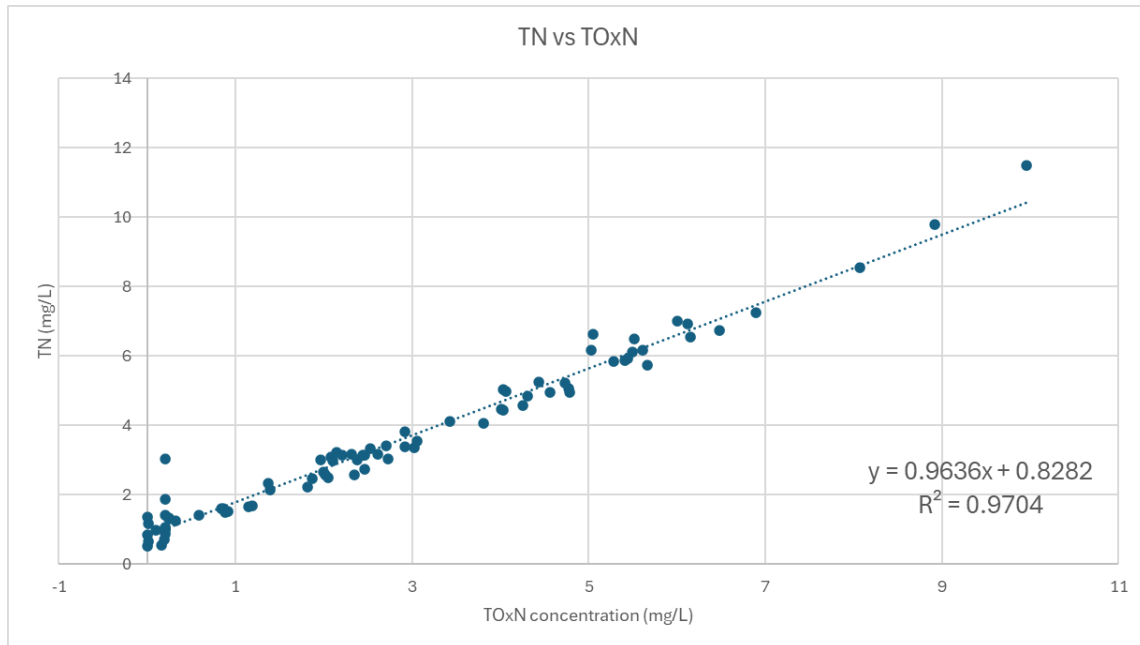
Figure 4.12. TN vs. TOxN at the abstraction point in NRW data.

Finally, to create one overall dataset, NRW and DCWW data were merged for the abstraction point. If data from DCWW and NRW was available for the same day (6 instances), the mean value was calculated. The NRW data only contained data for TN, TOxN, TP, PO_4^{3-}-P, chlorophyll *a* and nutrient ratios, hence these parameters caused duplication.

Tributary data

Tributary data has undergone a similar data pre-processing step with the Rpackage NADA (Lee, 2020) and the percentage of censored over uncensored values (% of data points below LoD) is visible in Table 4.3. None of these percentages were above 60%, so the new values were accepted. PO_4^{3-}-P exceedance of TP and NO_2^--N exceedance of TOxN was corrected similarly to abstraction data. NH_4^+-N and NO_3^--N were also calculated in the same way. TN was not calculated because NRW data was not collected at the tributary location and nutrient ratios were also not calculated due to limited available data.

Sensor data

Sensor data was cleaned by unifying time zones to UTC, removing calibration periods, and filtering out erratic values. This was done in Rstudio by computing for each row

the difference between its own value and the previous row as well as its value compared to the next row of data. For a row in which this difference for both previous and next data point exceeded the threshold (Table 4.4), it was replaced by a rolling median over a window of 5 datapoints centred around each data point, using R package 'zoo' (Zeileis & Grothendieck, 2005). Missing values were not filled in.

Table 4.4. Threshold for filtering out erratic values in the sensor data.

Parameter	Threshold for erratic value difference between two consecutive data points
pH	0.5
ORP	50
Water depth (m)	0.02
DO (mg/L)	1.5
EC (μS/cm)	7.5
NO_3^--N (mg/L)	0.8

Water depth (m) was calculated from a pressure reading measured above the In-Situ sensor in the water column, which was hanging from the bridge (Figure 4.4). For plotting, this measurement was converted to reservoir water level (m), which was measured by DCWW using a pressure sensor port hole in the abstraction tower. The DCWW water level measurements go deeper than the reservoir depth, so it cannot be directly compared to the bathymetry (Figure 4.3) but it can be used as an indication of water level changes over the years. The difference between water depth (In-Situ sensor) and DCWW water level was 11.5 meters on the 01/01/2023, when the reservoir was at its maximum volume, therefore 11.5 m was added to the In-Situ water depth measurements (Figure 4.4).

4.2.8 Analyses and plotting

All required data analysis was completed in R v.4.3.1 (RStudio v.2023.06.2) and associated packages as described. Plotting was done with R package 'ggplot2' (Wickham, 2016) and plots were saved as vector images with the R package 'Cairo' (Urbanek & Horner, 2023).

Nutrient rate of change

NRW and DCWW abstraction data were combined for the years 2015 to 2023, in order to calculate the rate of change in nutrient concentrations. The described method below was applied to the merged abstraction data ("toc", "doc", "tn", "nh4n", "toxn", "no3n",

"no2n", "tp", po4p", "chla", "algaetot", "greenalgae", "bluegreenalgae", "otheralgaetot")
and tributary data ("toc", "doc", "nh4n", "toxn", "no3n", "no2n", "tp", "po4p").

R packages 'dplyr' (Wickham et al., 2023), 'data.table' (Barrett et al., 2024) and 'zoo'
(Zeileis & Grothendieck, 2005) were used to create a function that finds for selected
columns (nutrients and algal parameters) the closest non-NA value in a previous row
(back in time) and calculates the difference between the values. It also calculated the
number of days between the rows and a threshold was applied to enable the rest of
the calculation only when there was less than 90 days between the rows. This
prevented calculation of differences from unreasonably infrequent data, with lots of NA
values. The threshold of 90 days was chosen to provide the most complete dataset,
especially for TP and PO_4^{3-}-P, because they were measured infrequently in the first
few years and gaps between samples were up to 90 days. Lapsed sampling during
the COVID-19 pandemic (that prevented staff reaching the field site) resulted in much
larger data gaps, and this was filtered out. Care needs to be taken with differences
calculated between rows that are 90 days, but fortunately there are very few.

Lastly, the calculated difference in values was divided by the number of days between
both rows, to calculate a rate of change, to compensate for differing number of days
between rows throughout the dataset using R packages 'gridExtra' (Auguie, 2017).
This was required because the number of days between observations was highly
variable, especially with the inclusion of NRW data, which only contained a selection
of the parameters. This method was chosen over taking monthly averages, because
that would have reduced the temporal frequency of the data and smoothed out
patterns in the data.

Meteorological data calculations

Time lags were investigated by calculating mean, maximum, minimum and sum of
weather parameters on the weeks prior to the sampling date. The time lags explored
were: 1 week prior to date (0-7 days from date), 2 weeks prior to date (8-14 days from
date), 3 weeks prior to date (15-21 days from date), 4 weeks prior to date (22-28 days
from date), and also 1, 2, 3, 4, 5, 6 and 7 days prior to date.

Sensor data calculations

Daily maximum, minimum, median, mean and range (difference between maximum
and minimum values) were calculated for eight sensor parameters (EC, pH, DO, ORP,

NO₃⁻-N, Chl-a, Depth, Water Temp, Air Temp). To further explore the influence of lag time, these were also calculated for the following intervals: 1 week prior to date (0-7 days from date), 2 weeks prior to date (8-14 days from date), 3 weeks prior to date (15-21 days from date), 4 weeks prior to date (22-28 days from date). This was used to calculate maximum of daily maxima, mean of daily means, minimum of daily minima, maximum of daily range, mean of daily range and minimum of daily range, median of daily medians and weekly range (difference between maximum of daily maxima and minimum of daily minima).

Event thresholds

Geosmin and 2-MIB concentrations were classified based on the level of severity: low, medium and high. The full details of the level classification were for geosmin: low = 0 – 5 ng/L, medium = 5 – 20 ng/L, high > 20 ng/L, and for 2-MIB: low = 0 - 2.5 ng/L, medium = 2.5 – 10 ng/L, high > 10 ng/L, which is the same as Hooper (2023a). The DCWW event threshold for geosmin is 10 ng/L and 5 ng/L for 2-MIB, which were halved to derive the border between the low and medium class: 5 ng/L and 2.5 ng/L for geosmin and 2-MIB, respectively. A geosmin and 2-MIB event was defined as a concentration ≥10 ng/L for geosmin and ≥5 ng/L for 2-MIB, following DCWW thresholds and Hooper (2023a). The event threshold for geosmin of 10 ng/L corresponds to the log transformed version $\log_{10}(10+1) = 1.04$ and 2-MIB value of 5 ng/L corresponds to $\log_{10}(5+1) = 0.78$, which is used for plotting.

PCA

A selection of manual sampling data and weather parameters was used for the principal component analysis (PCA): pH, turbidity, TOC (toc), TN (tn), NH₄⁺-N (nh4n), NO₃⁻-N (no3n), TP (tp), PO₄³⁻-P (po4p), manganese, chlorophyll *a* (chla), algae total cell count (algaetot), green algae total cell count (greenalgae), blue-green algae cell count (bluegreenalgae), TN:TP ratio (tntp_ratio), NH₄⁺:NO₃⁻ ratio (nh4no3_ratio), geosmin, 2-MIB (mib), precipitation (precip), maximum air temperature (tempmax), air temperature (temp), wind speed (windspeed), solar radiation (solarradiation), solar energy (solarenergy), average air temperature 2 weeks lagged (avg_temp_2weeks_prior), maximum wind gust 2 weeks lagged (max_windgust_2weeks_prior), sum of solar energy 2 weeks lagged (sum_solarenergy_2weeks_prior). These parameters were first $\log_{10}(x+1)$ transformed and scaled (zero-centred). Subsequently, a principal component analysis

(PCA) was performed using the Nipals (Nonlinear Estimation by Iterative Partial Least Squares) algorithm to deal with missing values, from the R package 'pcaMethods' (Stacklies et al., 2007). The Nipals algorithm is at the foundation of the Partial Least Squares (PLS) regression and executes PCA with missing values by excluding them from the appropriate inner products. This method uses interpolation with a least-squares-fit to find the missing values but gives these missing data no influence on the model. Each subsequent iteration improves the missing value by multiplying the score and the loading at that specific point. The Nipals algorithm can be used on datasets with approximately 5% missing values and this dataset had 5.3% missing values, which was deemed acceptable.

Datapoints in the PCA were coloured based on whether there was a geosmin event (Yes, No) and this was repeated for a 2-MIB event (Yes, No). 95% ordiellipses were plotted to indicate the difference between event and no-event on the PCA plot. To determine whether an event was statistically different from no event, permutational multivariate analysis of variance (PERMANOVA) was done with Euclidian distance and 999 permutations, using R package 'labdsv' (Roberts, 2023).

Simple GAMs

Generalised additive models (GAM) were constructed for multiple parameters in the dataset to understand the influence of each individual parameter on geosmin and 2-MIB concentrations, using R package 'mgcv' (Wood, 2017; Wood et al., 2016). For GAM analysis, geosmin and 2-MIB were log-transformed as follows: $\log_{10}(x+1)$, which are referred to as log_geosmin and log_mib. The following basic structure was used for the GAM: gam(log_mib ~ s(x), data = new_data, method = "ML", family = tw()), which uses a smoothing term s(x) in which x is the predictor parameter in dataset "new_data", a standard k of 10 (smoothing parameter), method ML (maximum likelihood) for the linearisation and a Tweedie data distribution (nonexponential) for the distribution family (Wood et al., 2016). The Tweedie distribution applies the logarithm of the expected value (log_mib) as a linear function of the predictor s(x). The same GAM was applied with log_geosmin as the modelled variable. Some parameters were flagged because they didn't have sufficient different values to use k of 10, so separate GAMs were performed with the maximum number k. In some cases, the method "ML" did not perform well, so they were executed with "REML" (restricted maximum likelihood) instead. Individual GAMs were performed on the full dataset (2015 to 2023)

with manual samples and weather parameters, as well as a selection of the dataset that included sensor parameters (2022 to 2023).

Full GAM for 2-MIB and geosmin

GAMs with multiple parameters from the whole dataset (manual and weather, 2015 to 2023) were explored to understand their ability to predict concentrations of log_geosmin and log_mib. Parameters were entered as a smoothing term based on most promising individual GAM results (highest deviance explained). The GAMs were validated by comparing several slightly different models with the Akaike Information Criterion (AIC), where a lower value indicates a better model. For example, a model with all parameters compared with subsequent models that excluded one of the parameters one-by-one, then AIC could be compared to see if excluding certain parameters improved the model. Residual plots (better if normally distributed residuals without bias) and k.check values (if smoothing parameter k wasn't over or underfitting, edf not too close to k') were checked for potential problems in the model, as well as autocorrelation and partial autocorrelation. A small amount of autocorrelation was tolerated, but if there were obvious autocorrelations in the data, then additional steps would be required. Parameters with as little missing data as possible were selected for full GAMs, as this kept more data in the model.

4.3 Results

4.3.1 Water quality & meteorological data 2015-2023

Geosmin and 2-MIB

The water quality data from 2015 to 2023 revealed that Reservoir 1 experienced yearly geosmin and/or 2-MIB events above the DCWW threshold; 10 and 5 ng/L for geosmin and 2-MIB, respectively (although there were limited 2020 data due to COVID-19 lockdown) (Figure 4.13). Geosmin levels above the threshold were consistently detected earlier in the year than 2-MIB levels above the threshold, and there were notable differences in between years for the compound behaviours (Figure 4.13). In 2019 there was no 2-MIB event whilst geosmin did have an event. In 2023, the 2-MIB event lasted the whole year because concentrations in the 2022 to 2023 winter period did not go below the threshold, whilst geosmin only had a minor event (Figure 4.13).
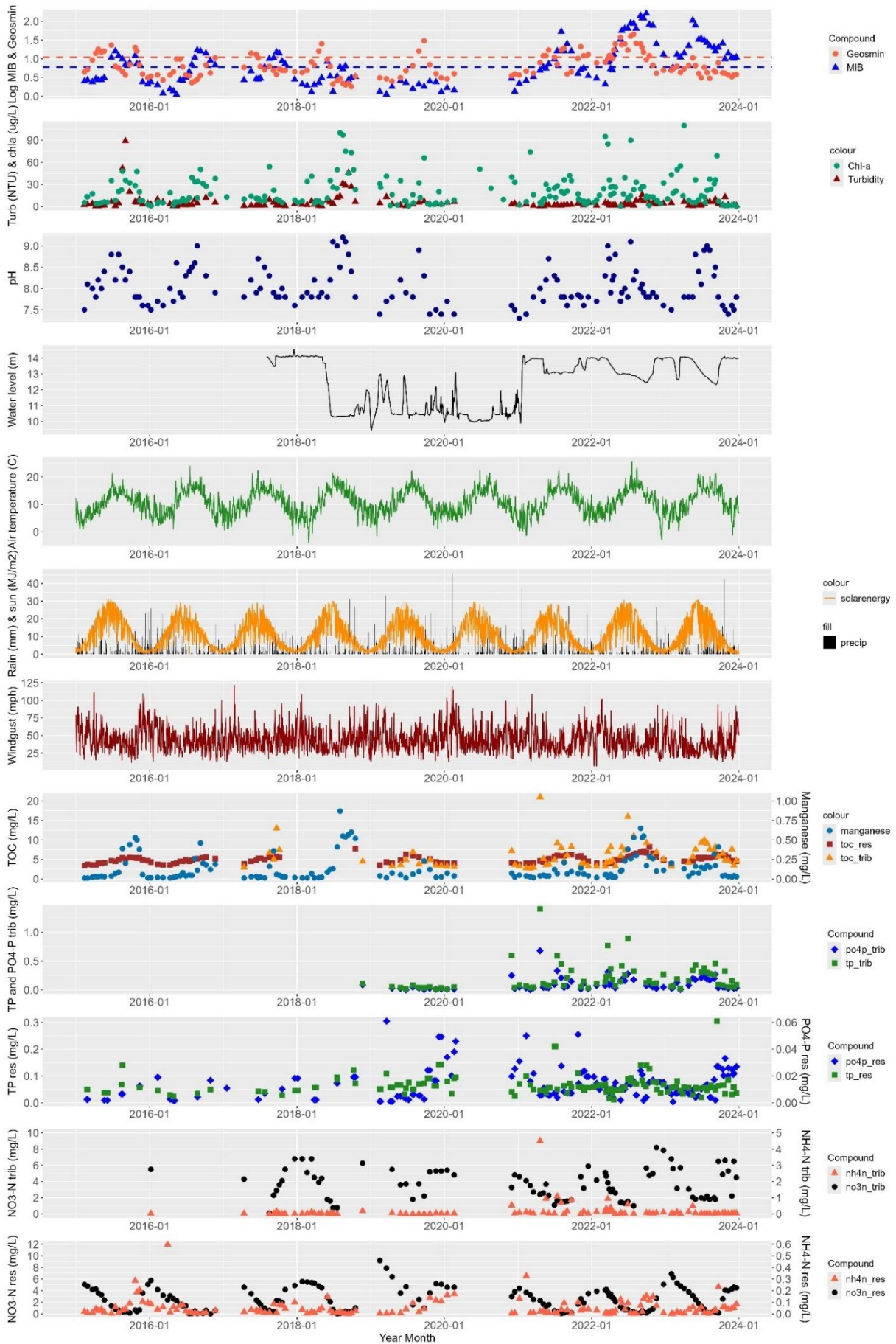
Figure 4.13. Reservoir 1 manual samples, depth sensor and meteorological data 2015-2023.

Chlorophyll a and pH

Chlorophyll *a* (µg/L) measurements generally showed a pattern of elevated concentrations in the second half of the year, approximately from June to October (Figure 4.13). However, in more recent years, chlorophyll *a* concentrations increase earlier in the year, and they can persist for long periods of time (Figure 4.13). Chlorophyll *a* did not reveal a similar pattern to geosmin or 2-MIB.

pH and chlorophyll *a* generally followed the same trend, with elevated values from late-spring to early-autumn (Figure 4.13). Due to its buffer capacity, pH showed a slower response to productivity and a more predictable yearly pattern than chlorophyll *a*. The pH pattern often overlapped with periods of geosmin events (Figure 4.13).

Water level changes

Water level in Reservoir 1 reached maximum capacity from mid-2017 until approximately May 2018, when there was a steep drop of ±4 meters. A large volume of water was abstracted to allow maintenance work on the reservoir dam, as DCWW was building a spillway. Water levels were maintained at this low level until early 2021, with sharp increases throughout this period related to precipitation which are followed by a quick drawdown to maintain low water levels (Figure 4.13). In early 2021, water levels increased ±4 meters which refilled the reservoir to its maximum level. Later in springtime 2021, the water level dropped again by 1 meter, but it was refilled to the maximum water level at the end of the year (Figure 4.13). Since 2022, Reservoir 1 has not been used for drinking water abstraction, except for a short period in February-March 2023 when water levels dropped 1.5 m over 17 days and quickly increased to maximum levels again over 6 days (Figure 4.13).

Weather

Average daily air temperature had a consistent cyclic pattern over the years with temperatures reaching <0 ºC in winter and >20 ºC in summer. 2022 is the only year in the dataset for which average daily air temperature exceeded 25 ºC (Figure 4.13). Solar energy (MJ/m^2) revealed a similar cyclic pattern every year, but maximum values happened slightly earlier in the year than for daily air temperature. Periods of precipitation and higher wind gusts indicate unstable weather and cloudy conditions, also reflected in a drop in potential solar energy (MJ/m^2). Most years had periods of stable and warm weather without precipitation in spring or summer, but prolonged dry spells occurred in 2015, 2018, 2020, 2022 and 2023. In these periods, there was

maximum potential solar energy for that time of year, little precipitation, low wind gusts (mph) and an increase in daily air temperature (ºC) (Figure 4.13).

Annual statistics for weather parameters in Table 4.5 showed that 2018 and 2022 had exceptionally low total precipitation of 626 and 660 mm, respectively, compared to other years that ranged from 760 to 942 mm. The years 2018-2019 and 2022 were officially defined as drought years in the UK by the National Centre for Atmospheric Science (National Centre for Atmospheric Science, 2023). 2015 and 2022 had exceptionally high total solar energy of 4360 and 4123 MJ/m^2, respectively, compared to a range of 3618 to 3982 MJ/m^2 (Table 4.5). This pattern was also visible for average solar radiation (W/m^2). Maximum wind gusts were highest in 2015, 2017 and 2020 with wind gusts of 111, 122 and 120 mph, respectively, whilst 2023 had the lowest maximum wind gusts of 94 mph (Table 4.5). The average windspeed was highest for the years 2015, 2017 and 2020 with windspeeds of 23, 22 and 22 mph, respectively (Table 4.5), and lowest is 2021 with 20.4 mph. Maximum temperature was highest in 2022, with 34 ºC and the lowest maximum temperature was in 2015 at 26 ºC (Table 4.5). The lowest minimum temperature also happened in 2022, at -6.3 ºC and the highest minimum temperature was -0.3 ºC in 2020 (Table 4.5). The average temperature was around 10.5 ºC from 2015 to 2021 but showed an increase to 11.2 ºC in 2022 and 2023 (Table 4.5).

Table 4.5. Annual averages of selected meteorological parameters from 2015 to 2023.

	2015	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Sum precipitation (mm)	779.7	778.0	759.4	625.6	898.9	941.5	802.9	660.3	929.1
Sum solar energy (MJ/m^2)	4359.7	3787.5	3617.6	3766.8	3738.3	3778.5	3815.1	4123.1	3981.5
Average solar radiation (W/m^2)	138.4	119.9	114.9	119.9	118.8	119.7	121.1	130.9	126.4
Maximum wind gust (mph)	111.2	100.1	121.7	108.4	96.1	119.9	109.1	101.9	94.0
Average windspeed (mph)	23.0	20.7	22.4	20.9	21.4	22.4	20.4	21.3	21.5

Maximum temperature (ºC)	26.4	30.1	27.6	28.2	30.6	29.7	28.6	34.2	27.2
Minimum temperature (ºC)	-2.8	-2.6	-5.8	-5.6	-4.4	-0.3	-3.9	-6.3	-3.8
Average temperature (ºC)	10.2	10.5	10.8	10.5	10.6	10.8	10.5	11.2	11.2

Reservoir patterns

NO_3^--N in Reservoir 1 had a strong seasonal pattern in the years 2015-2023, with highest values between 4 and 9 mg/L in the late-autumn and winter, and lowest in summer and early-autumn, when NO_3^--N concentrations were <0.5 mg/L for long periods of time (Figure 4.13).

Turbidity (NTU) measurements in this dataset were done in the laboratory rather than *in situ*. In general, turbidity peaks happened later in the year, from June to October, but they were not as pronounced every year (Figure 4.13). Two larger prolonged turbidity peaks were detected in 2015 and 2018, but the infrequency of the data means that other peaks might have been missed. The timing of the yearly turbidity peak coincided with elevated manganese (as well as aluminium and iron, Figure 4.14), NH_4^+-N, TOC and TP in the reservoir (mg/L) as well as the timing of 2-MIB events (Figure 4.13).
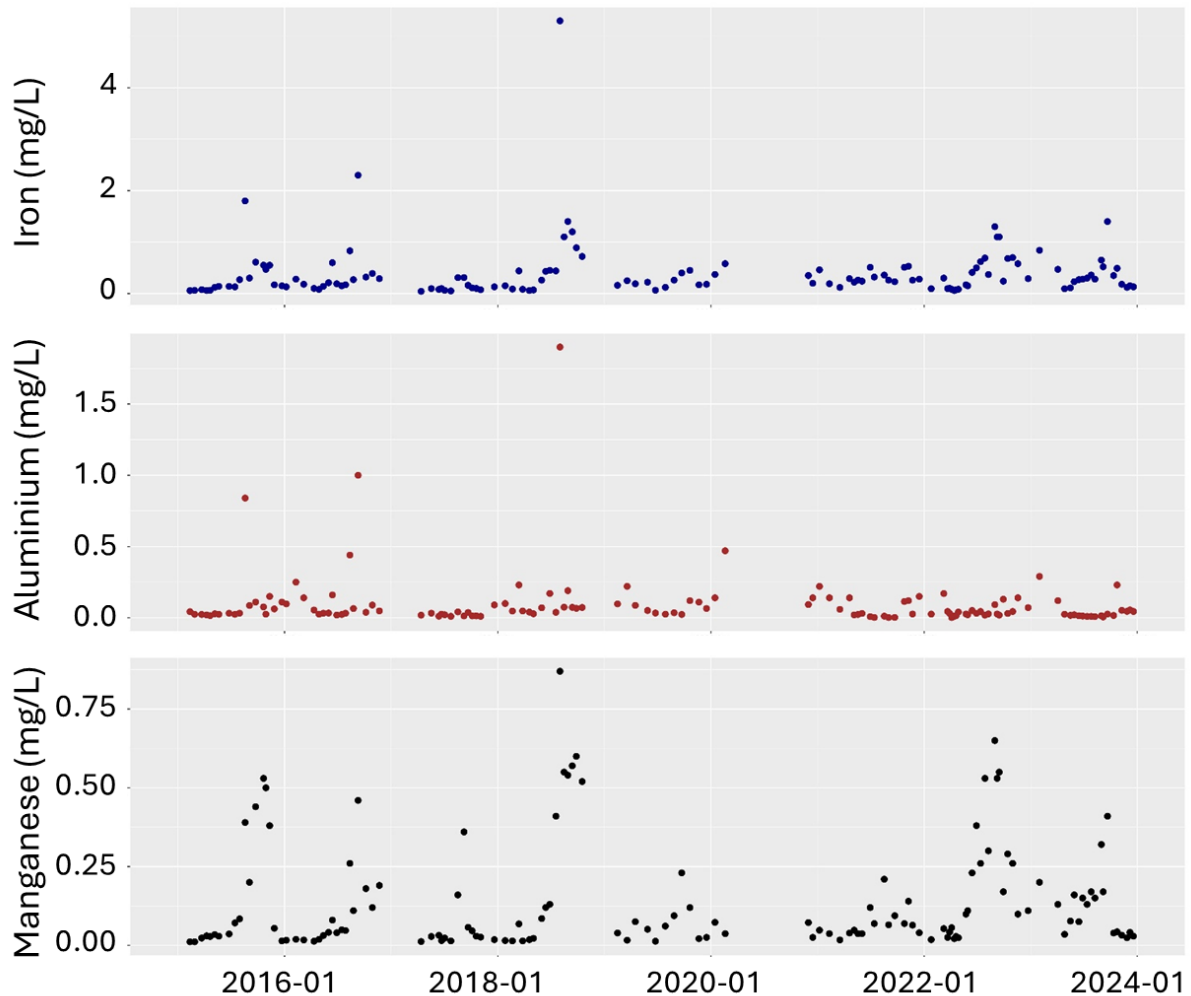
Figure 4.14. Total metal concentrations (mg/L) at abstraction in Reservoir 1 from manual DCWW sampling.

TOC and DOC in the reservoir and the tributary had very similar concentrations, which suggests that most TOC was present in dissolved form as DOC (Figure 4.15). TOC (mg/L) in the reservoir had a relatively predictable pattern of lower concentrations, ±4 mg/L, in the winter and spring, and higher concentrations in summer and autumn >5 mg/L (Figure 4.13, Figure 4.15). In October 2022, the tributary had especially high concentrations of TOC and DOC of ±15 mg/L, and the reservoir had a maximum TOC concentration of ±8 mg/L, which happened at the same time as the peak in 2-MIB. However, elevated TOC and DOC concentrations in the reservoir are not necessarily related to a 2-MIB event, which is the case in autumn 2018 (Figure 4.13, Figure 4.15).
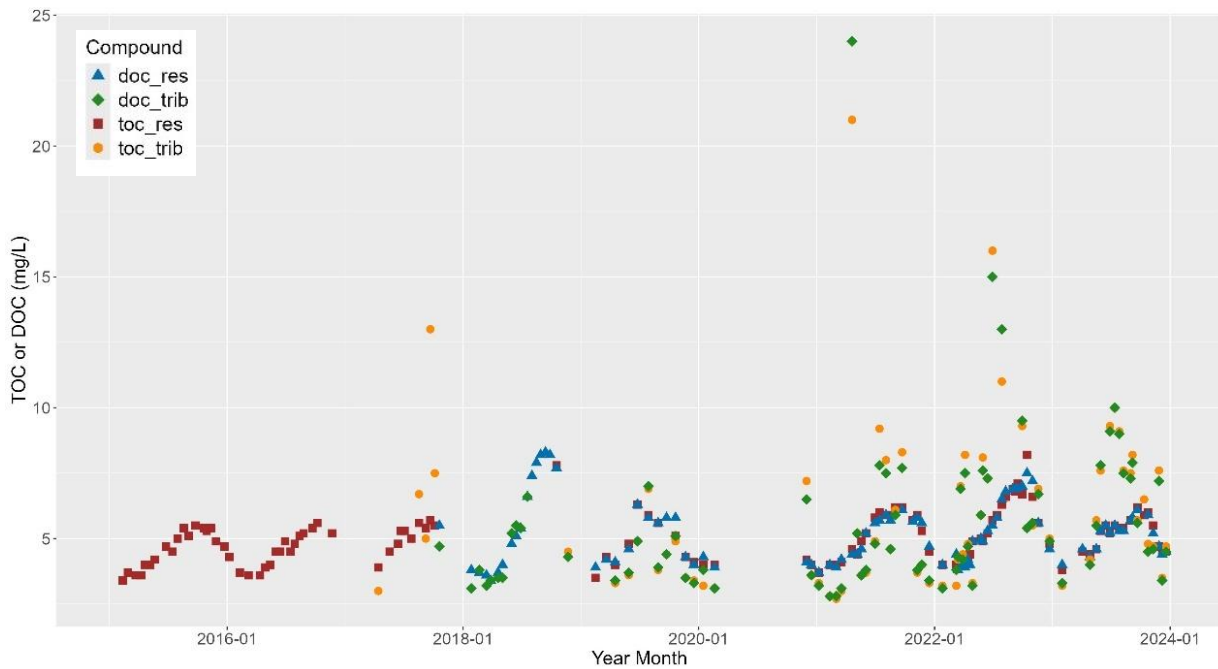
Figure 4.15. DOC and TOC comparison.

TP and PO_4^{3-}-P concentrations in the reservoir were often below the limit of detection for the DCWW laboratory (PO_4^{3-}-P: 0.03 mg/L; TP: 0.055 mg/L) and so were estimated with NRW data and the R package 'NADA' (see methods in Section 4.2.7). The estimated data should be interpreted with caution. TP seemed to follow a pattern of internal loading, visualised by elevated manganese concentrations, with peaks in late-summer and autumn (Figure 4.13). Concentrations of PO_4^{3-}-P also increased during periods of internal loading, due to the strong pattern in TP, but the data revealed that PO_4^{3-}-P concentrations were influenced by multiple processes. When water levels were low due to maintenance works (2018 to 2021), PO_4^{3-}-P concentrations were high (±0.05 mg P/L) for long periods of time, and also increased to ±0.05 mg P/L directly after the reservoir was refilled in early 2021 (Figure 4.13). PO_4^{3-}-P concentrations were also influenced by biological uptake, as this is the most bioavailable form of P. This process is visualised by late-autumn and winter maxima and summer minima, following the NO_3^--N pattern (Figure 4.13).

NH_4^+-N in the reservoir increased during periods of internal loading and had particularly high concentrations during the 2018 to 2021 period of low water levels. NH_4^+-N concentrations peaked at 0.3 mg/L in early 2021 when the reservoir was refilled, which coincided with the pattern in PO_4^{3-}-P (Figure 4.13). There were also peaks of NH_4^+-N in the reservoir at different times of the year, possibly related to short-lived influxes of high NH_4^+-N concentrations in tributary inflows (Figure 4.13).

Tributary inflows

The NO_3^--N pattern in the tributary was similar to the reservoir, with high concentrations of NO_3^--N of approximately 8 mg/L in winter, whilst concentrations drop down to around 1 mg/L in summer (Figure 4.13). 2019 was the only year in which tributary concentrations did not fall below 2 mg/L in summer (Figure 4.13) and subsequently NO_3^--N concentrations in the reservoir stayed higher than other years at >1 mg/L.

TOC concentrations in the tributary were more unpredictable and high concentrations happened any time of the year, but it generally increased in spring and decreased again in autumn (Figure 4.13). Elevated TOC concentrations in the tributary seemed to happen at the same time as 2-MIB events, particularly in recent years when sampling happened more frequently. The same pattern was visible for TP and PO_4^{3-}-P in the tributary and in some cases elevated NH_4^+-N concentrations (mg/L) were also detected (Figure 4.13). Nutrient concentrations measured (TP, PO_4^{3-}-P, TOC, NH_4^+-N) in the tributary inflow increased from 2021 onwards (Figure 4.13), with concentrations in spring-summer of around 0.5 mg/L TP, 0.25 mg/L PO_4^{3-}-P, 10 mg/L TOC and 1 mg/L NH_4^+-N. One particularly high input of organic nutrients was detected on 20th of April 2021 with >20 mg/L of TOC, 0.7 mg/L PO_4^{3-}-P, >1.3 mg/L TP and 4.5 mg/L of NH_4^+-N (whilst NO_3^--N in that same sample did not increase compared to previous measurements) (Figure 4.13).

Tributary geosmin and 2-MIB revealed that concentrations of geosmin and 2-MIB in the tributary were small concentrations that hardly exceeded the DCWW thresholds (Figure 4.16).

Figure 4.16. Abstraction and tributary geosmin and 2-MIB concentrations. Dotted lines in the same colour as the compound indicate DCWW event threshold.

4.3.2 Nutrient rate of change

Nutrient rate of change plots were created to visualise the effect of increases or decreases in nutrients at a datapoint compared to its previous datapoint. This value was divided by the number of days in between the datapoints, to get the rate of change (see methods, Section 4.2.8). Tributary and reservoir data was used, which meant that internal nutrient dynamics (internal loading) and external inputs (tributary runoff) could be compared in the same plot. Unfortunately, the first 6 years in the dataset had limited data available for nutrients. Therefore, the most interesting nutrient data from 2022 was highlighted here, and the plots for all years can be found in Appendix C (Figure S.3, Figure S.4, Figure S.5, Figure S.6, Figure S.7 and Figure S.8).

TOC and DOC had a very similar pattern in 2022, the tributary TOC and DOC increased in March and April but decreased again during April, followed by further TOC

172

and DOC inputs in June and July (Figure 4.17). Rates of change for TOC and DOC in the reservoir were much smaller than the tributary inputs, but they were elevated during the 2-MIB event and are indicative of internal loading (Figure 4.17). NH_4^+-N, TP and PO_4^{3-}-P had the same pattern in the tributary rate of change related to March and April tributary inflow (Figure 4.17). During the 2-MIB event, the levels of nutrients in the reservoir fluctuated (positive and negative rates of change).

NO_3^--N revealed an altogether different process of gradual decrease throughout the year in nutrients in both tributary and reservoir, followed by an increase in the tributary which started in October and a subsequent increase in NO_3^--N in the reservoir at the next datapoint (Figure 4.17).
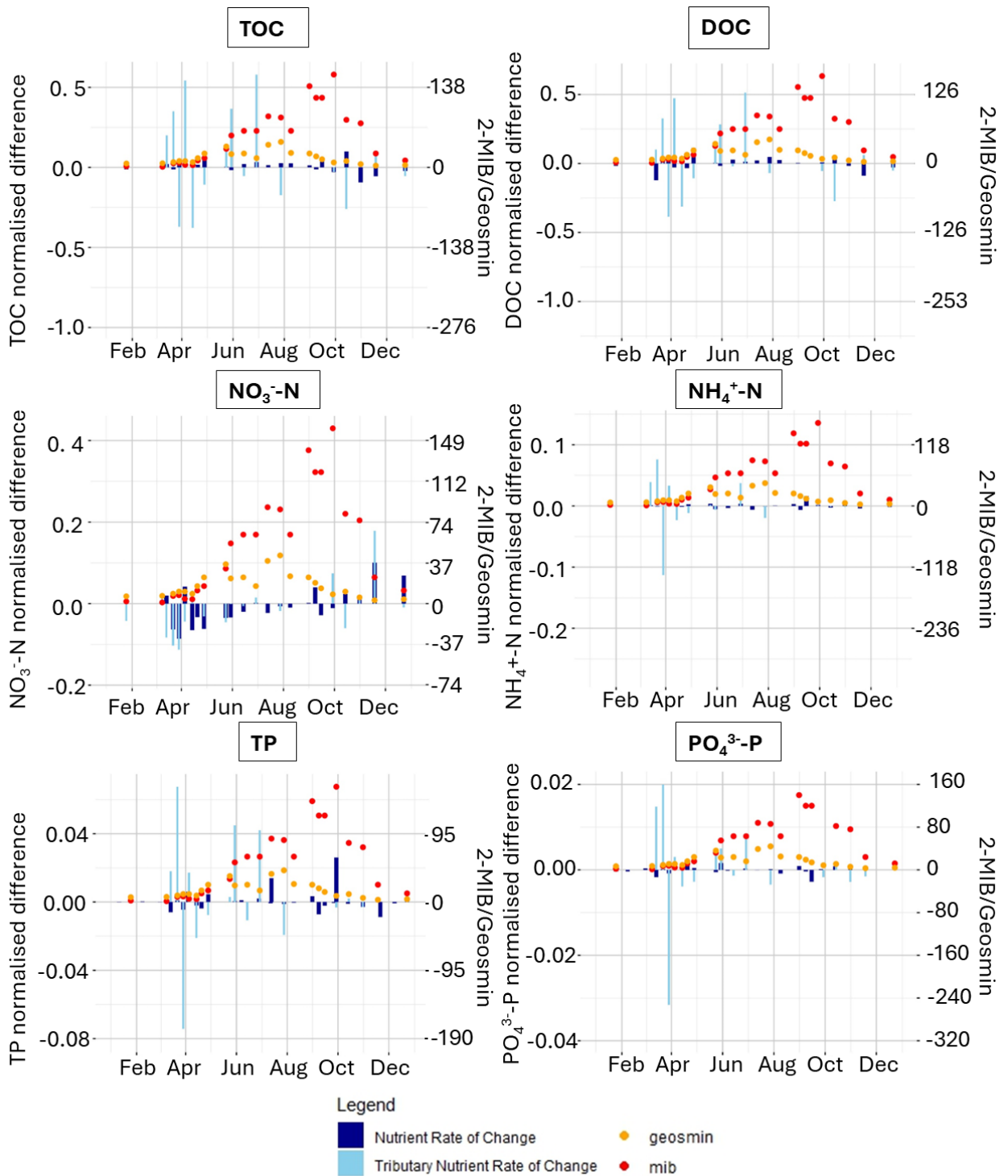
Figure 4.17. 2022 nutrient rate of change in reservoir (dark blue) and tributary (light blue), compared with 2-MIB and geosmin concentrations (ng/L).

4.3.3 PCA

To predict geosmin and 2-MIB events, it is important to investigate the underlying physical and chemical processes that drive these events. A principal component analysis (PCA) was constructed for the nine years of manual sampling data and weather parameters (some lagged), using the Nipals algorithm to deal with missing values (see methods, Section 4.2.8). PC1 explained 32.3% of the variation and PC2 explained 14.8% of the variation (Figure 4.18). The $\cos^2$ values in the PCA are all below 0.08, which is relatively low on the scale of 0 (not well representing PC axis) to 1 (perfectly representing PC axis). This indicates that none of the parameters had a strong influence on either PC axis, and that patterns are probably not consistent or easily predictable throughout the dataset. NO_3^--N and TN are both negative influences on PC1, and their direction is the exact opposite to 2-MIB in this PCA, which indicates a negative relationship between NO_3^--N/TN and 2-MIB. The TN:TP ratio is closely related to the TN concentrations, which is logical, but it indicates that TN is the driving force behind TN:TP ratio calculations, and TP has less influence. 2-MIB is closely related to air temperature (2 weeks prior), TOC, manganese, turbidity, TP and NH_4^+:NO_3^- (Figure 4.18). Generally, geosmin seems to be on the same side of the PCA as solar energy/radiation, air temperature, pH, chlorophyll *a* and various algae counts. Geosmin itself does not have a very strong influence on the PCA and high geosmin values probably do not separate well into one side of the PCA.
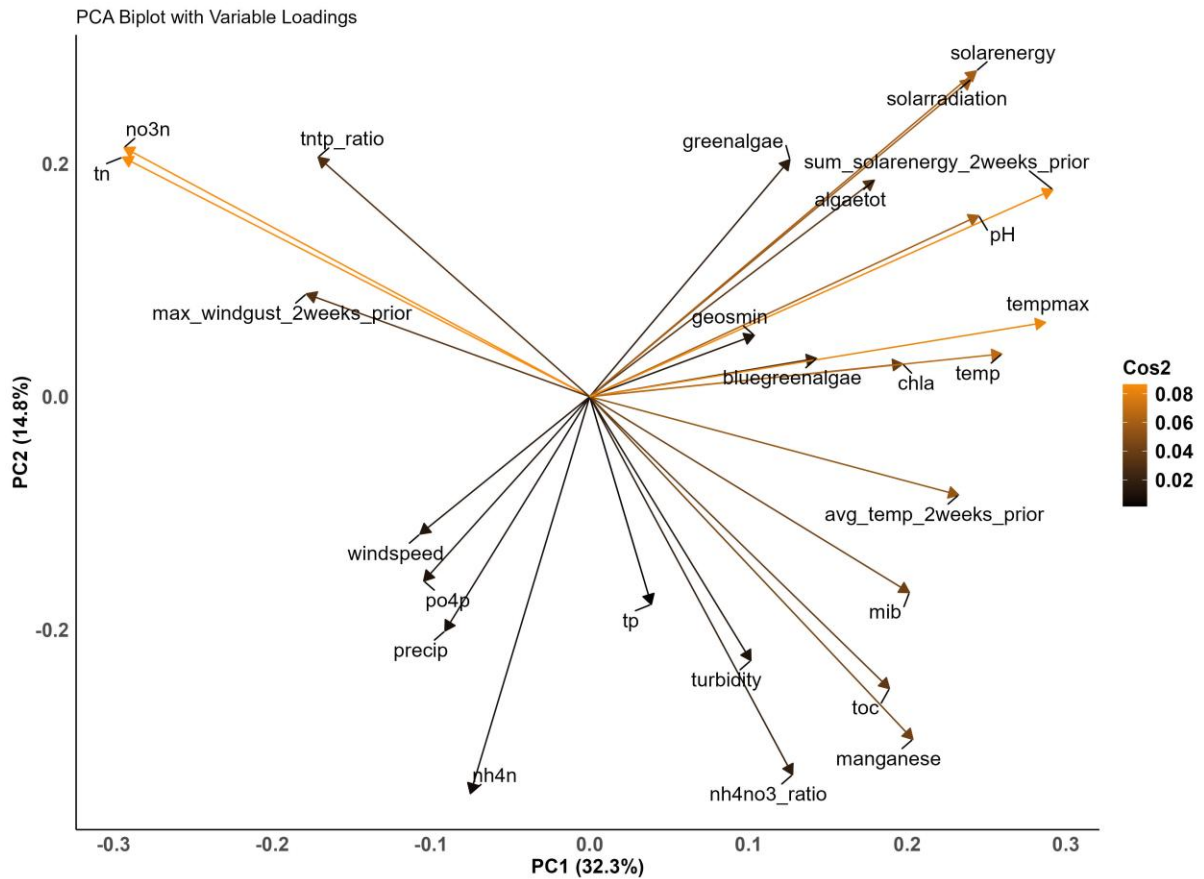
Figure 4.18. Principal component analysis (PCA) with the Nipals algorithm to deal with missing values. PC1 explains 32.3% of variation and PC2 explains 14.8% of variation. Colours of the parameter arrows represent cos^2 values; the strength of a parameters position in the PCA.

The PCA clustered datapoints with a 2-MIB event on the right side of the plot, whilst datapoints without a 2-MIB event were generally located on the left side of the plot (Figure 4.19). Results from a permutational multivariate analysis of variance (PERMANOVA) revealed that this difference explained 9.8% of the variation (R^2) with a significant difference of 0.001. Datapoints with a geosmin event were clustered on the right side of the PCA, whilst datapoints without a geosmin event tended to be in the middle and the left side (Figure 4.19). A PERMANOVA showed that 5.3% of the variation was explained (R^2) by these two groups with a significant difference of 0.001.
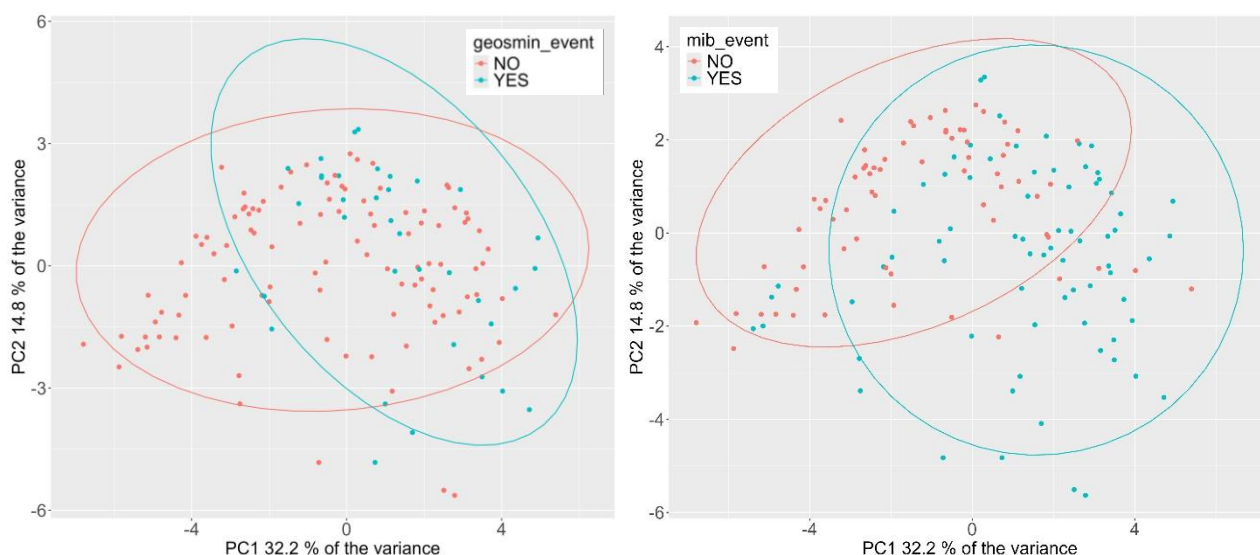
Figure 4.19. PCA datapoints coloured by 2-MIB event (left panel) and geosmin event (right panel) and ellipse drawn with the 95% confidence interval.

4.3.4 Correlations & simple GAMs full data

Correlations with a $\log_{10}$-transformed dataset were performed for geosmin as well as 2-MIB, using all parameters in the 9-year manual sampling dataset and all calculated weather parameters, including 1 to 7 days lagged and 1, 2, 3 or 4 weeks prior. A GAM with a smoothing function was executed on a selection of parameters to fit the predictor to the $\log_{10}$ transformed geosmin and 2-MIB concentrations. The correlations and GAM outcomes of the most interesting parameters for 2-MIB and geosmin were displayed in Table 4.6 and Table 4.7, respectively. A smaller subsection of parameters with significant GAM p-values was plotted in Figure 4.20 and Figure 4.21, for 2-MIB and geosmin, respectively. The full set of results can be found in Appendix C Table S.7, Table S.8 and Table S.9.

2-MIB

GAM results showed that year and day of year (DOY) explained 54% (n=141) and 25% (n=141) of the deviance in the 2-MIB pattern (Table 4.6). The GAM plot for year showed the yearly 2-MIB event was most prominent from 2015 to 2017 and then a drop in severity from 2018 to 2020, followed by a steep increase in severity from 2021 onwards with the highest 2-MIB year in 2022 (Figure 4.20). Day of year (seasonality of the event) revealed a GAM with a smooth increasing curve and a peak of 2-MIB around day 250, which is in early September (Figure 4.20).

Table 4.6. Selection of individual GAM results to predict 2-MIB. Parameters on the top rows have GAM visualised in Figure 4.20, whilst rows below the double line in italic are not visualised. Column 'short name' refers to the names used in Figure 4.20.

Predictor for 2-MIB	Short name	Log-Log correlation	GAM P_Value	GAM R2_Adj	GAM Deviance expl.	GAM obs.
DOY	Day of year	0.35	0.000	0.226	0.247	141
year	Year	0.51	0.000	0.521	0.535	141
Sum solarenergy 2weeks prior	Sum solar energy 2w	0.35	0.000	0.117	0.109	141
Min temp 2weeks prior	Min air temp 2w	0.48	0.000	0.221	0.216	141
Max windgust 2weeks prior	Max wind gust 2w	-0.44	0.000	0.188	0.188	141
Avg windspeed 1week prior	Mean wind speed 1w	-0.23	0.000	0.089	0.117	141
toc	TOC	0.67	0.000	0.458	0.467	117
no3n	NO_3^--N	-0.53	0.000	0.279	0.277	140
po4p	PO_4^{3-}—P	-0.18	0.001	0.125	0.178	93
manganese	Total manganese	0.42	0.000	0.297	0.294	137
Tntp ratio	TN:TP	-0.51	0.000	0.286	0.252	92
nh4no3 ratio	NH_4^+:NO_3^-	0.34	0.002	0.118	0.102	139
Toc trib	Tributary TOC	0.59	0.000	0.454	0.478	68
no3n trib	Tributary NO_3^--N	-0.33	0.007	0.136	0.138	85
Tp trib	Tributary TP	0.39	0.000	0.395	0.427	66
temp		*0.35*	*0.000*	*0.126*	*0.149*	*141*
tn		*-0.54*	*0.000*	*0.285*	*0.286*	*140*
nh4n		*0.08*	*0.384*	*-0.002*	*0.005*	*139*
tp		*0.13*	*0.254*	*0.005*	*0.013*	*92*
chla		*0.16*	*0.227*	*0.001*	*0.009*	*139*
nh4n trib		*0.09*	*0.805*	*-0.002*	*0.009*	*85*
po4p trib		*0.39*	*0.000*	*0.376*	*0.372*	*66*
nh4no3 ratio trib		*0.13*	*0.352*	*0.033*	*0.036*	*85*
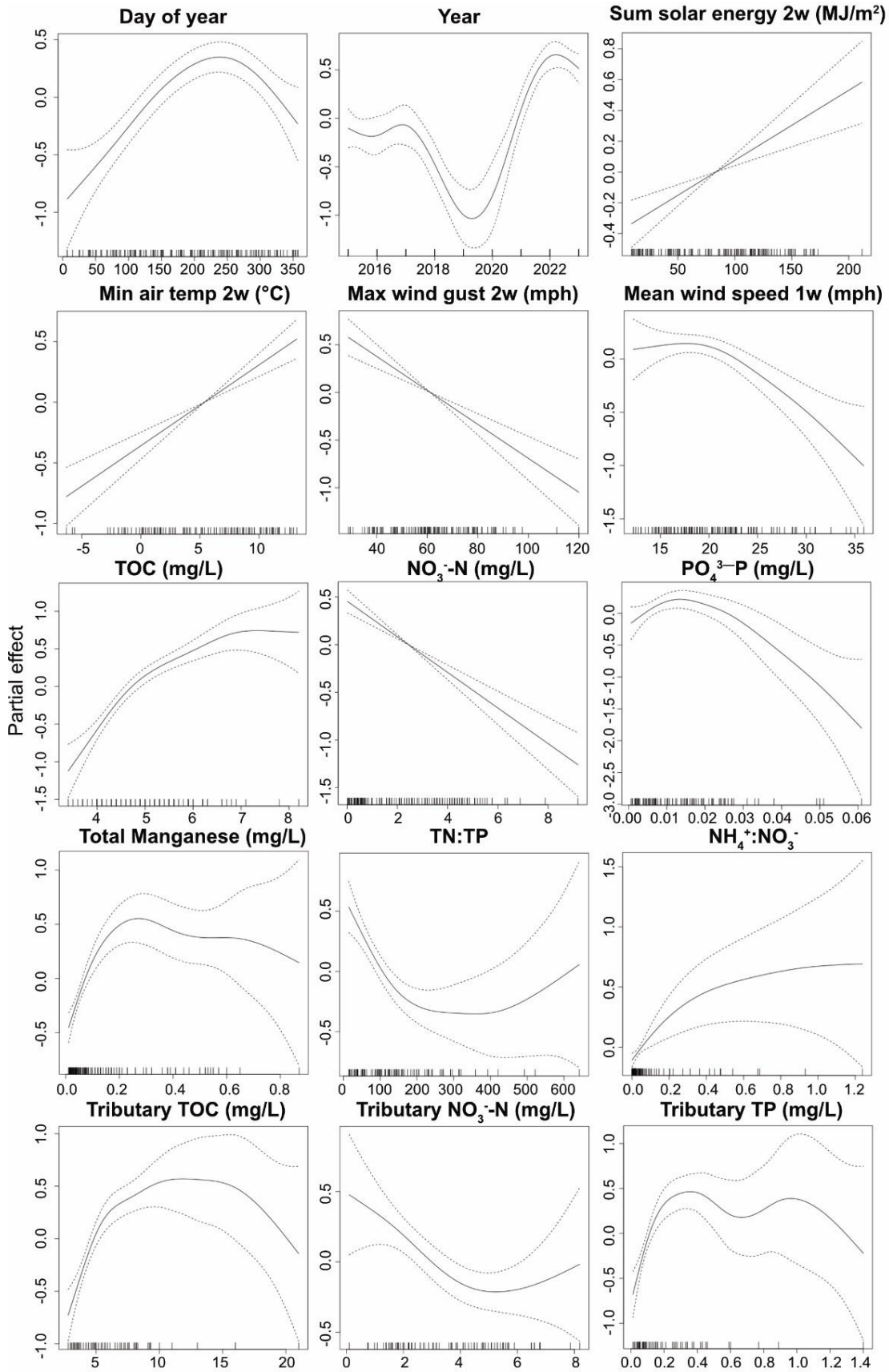
Figure 4.20. GAM selection for 2-MIB. Dotted line is 95% confidence interval. X-axis indicates all the plotted values for that parameter and vertical lines inside the plot along the x-axis indicate the distribution of those datapoints, units for the x-axis are in each plot title.

179

TOC and TP in the tributary explained 48% (n = 68) and 47% (n=66) of the deviance respectively, but the smaller dataset could have influenced these results. TOC and TP in the tributary had a steep increasing effect on 2-MIB from 0 to ±8 mg/L TOC and 0 to ±0.3 mg/L TP, which then plateaued and got much more uncertain due to limited data at these concentrations (Figure 4.20). TOC in the reservoir explained 47% (n=117) of the deviance, and also showed an increasing effect on 2-MIB with increasing values, which plateaued > 7 mg/L due to low data availability (Figure 4.20). Manganese in the reservoir explained 30% of the deviance and showed a similar pattern to TOC in the tributary of a sharp increase from 0 to ±0.25 mg/L and then a plateau with larger uncertainty (Figure 4.20).

NO_3^--N and PO_4^{3-}-P in the reservoir explained 28% and 18% of the deviance (Table 4.6), respectively, which had a negative effect on 2-MIB with increasing concentrations. The negative relationship was approximately linear for NO_3^--N, but PO_4^{3-}-P first showed a small increase between 0 and 0.015 mg/L, which was then followed by a decrease (Figure 4.20). TN:TP ratio and NH_4^+:NO_3^- ratio in the reservoir explained 25% (n=92) and 10% (n=139) of the deviance, respectively (Table 4.6). TN:TP ratio revealed a sharp decreasing effect on 2-MIB from 0 to 200, which stabilised >200 and got more uncertain (Figure 4.20). Increasing NH_4^+:NO_3^- ratios had a positive effect on 2-MIB, but the uncertainty was rather large (Figure 4.20).

Minimum air temperature 2 weeks prior was the best weather parameter (22% deviance explained), followed by maximum wind gusts 2 weeks prior (19% explained). A higher minimum air temperature 2 weeks prior, indicating warmer weather, had a positive effect on 2-MIB which was approximately linear (Figure 4.20). At the same time, an increasing maximum wind gust 2 weeks prior, indicating stormy weather, had a negative effect on 2-MIB which was also approximately linear (Figure 4.20).

Geosmin

The output from the individual GAMs showed that the tested parameters had lower overall predictive capacity (smaller deviance explained) for geosmin than they had for 2-MIB (Table 4.6 and Table 4.7). The best predictor was year, with 34% of deviance explained (n=141), which showed 2015 had a relatively high geosmin event, followed by several years (2016 to 2020) of slightly lower severity of geosmin events, and a strong increase in 2021 with the peak event in 2022 which then went back down to a

geosmin event below average (negative partial effect) in 2023 (Figure 4.21). Day of year (DOY) explained 14% of the deviance (n=141) and the seasonal effect was different from 2-MIB which peaked at day 250 (Figure 4.20) because geosmin generally peaked around 150 days (late May) and revealed a slight tendency for a second peak around day 300 (late October) (Figure 4.21).

Table 4.7. Selection of individual GAM results to predict geosmin. Parameters on the top rows have GAM visualised in Figure 4.21, whilst rows below the double line in italic are not visualised. Column 'short name' refers to the names used in Figure 4.21.

Predictor for geosmin	Short name	Log-Log correlation	GAM P_Value	GAM R2_Adj	GAM Deviance expl.	GAM obs.
DOY	Day of year	0.06	0.002	0.119	0.139	141
year	Year	0.16	0.000	0.329	0.341	141
tempmax	Max air temp	0.14	0.040	0.047	0.064	141
windspeed	Wind speed	-0.20	0.019	0.031	0.037	141
Sum solarenergy 2weeks prior	Sum solar energy 2w	0.30	0.000	0.093	0.102	141
Sum precip 1week prior	Sum precip 1w	-0.12	0.000	0.097	0.105	141
Avg windspeed 1week prior	Mean wind speed 1w	-0.24	0.000	0.057	0.078	141
pH	pH	-0.19	0.001	0.122	0.153	136
turbidity	Turbidity	-0.27	0.002	0.063	0.093	136
toc	TOC	0.26	0.005	0.054	0.064	117
no3n	NO_3^--N	-0.14	0.005	0.071	0.093	140
po4p	PO_4^{3-}-P	-0.19	0.034	0.022	0.042	93
Tntp ratio	TN:TP	-0.32	0.002	0.130	0.168	92
no3n trib	Tributary NO_3^--N	-0.27	0.004	0.073	0.092	85
Min temp 2weeks prior		*0.02*	*0.249*	*0.022*	*0.034*	*141*
Max windgust 2weeks prior		*-0.07*	*0.212*	*0.002*	*0.010*	*141*
tn		*-0.32*	*0.005*	*0.071*	*0.094*	*140*
nh4n		*0.10*	*0.257*	*0.002*	*0.009*	*139*
tp		*0.11*	*0.272*	*0.002*	*0.013*	*92*
manganese		*0.05*	*0.589*	*-0.006*	*0.002*	*137*
chla		*0.03*	*0.644*	*-0.006*	*0.001*	*139*
nh4no3 ratio		*0.14*	*0.121*	*0.009*	*0.017*	*139*
Toc rib		*0.11*	*0.377*	*-0.004*	*0.011*	*68*
nh4n trib		*0.10*	*0.488*	*-0.007*	*0.005*	*85*
Tp trib		*0.15*	*0.264*	*0.001*	*0.018*	*66*
po4p trib		*0.15*	*0.223*	*0.001*	*0.020*	*66*
nh4no3 ratio trib		*0.10*	*0.420*	*-0.006*	*0.007*	*85*
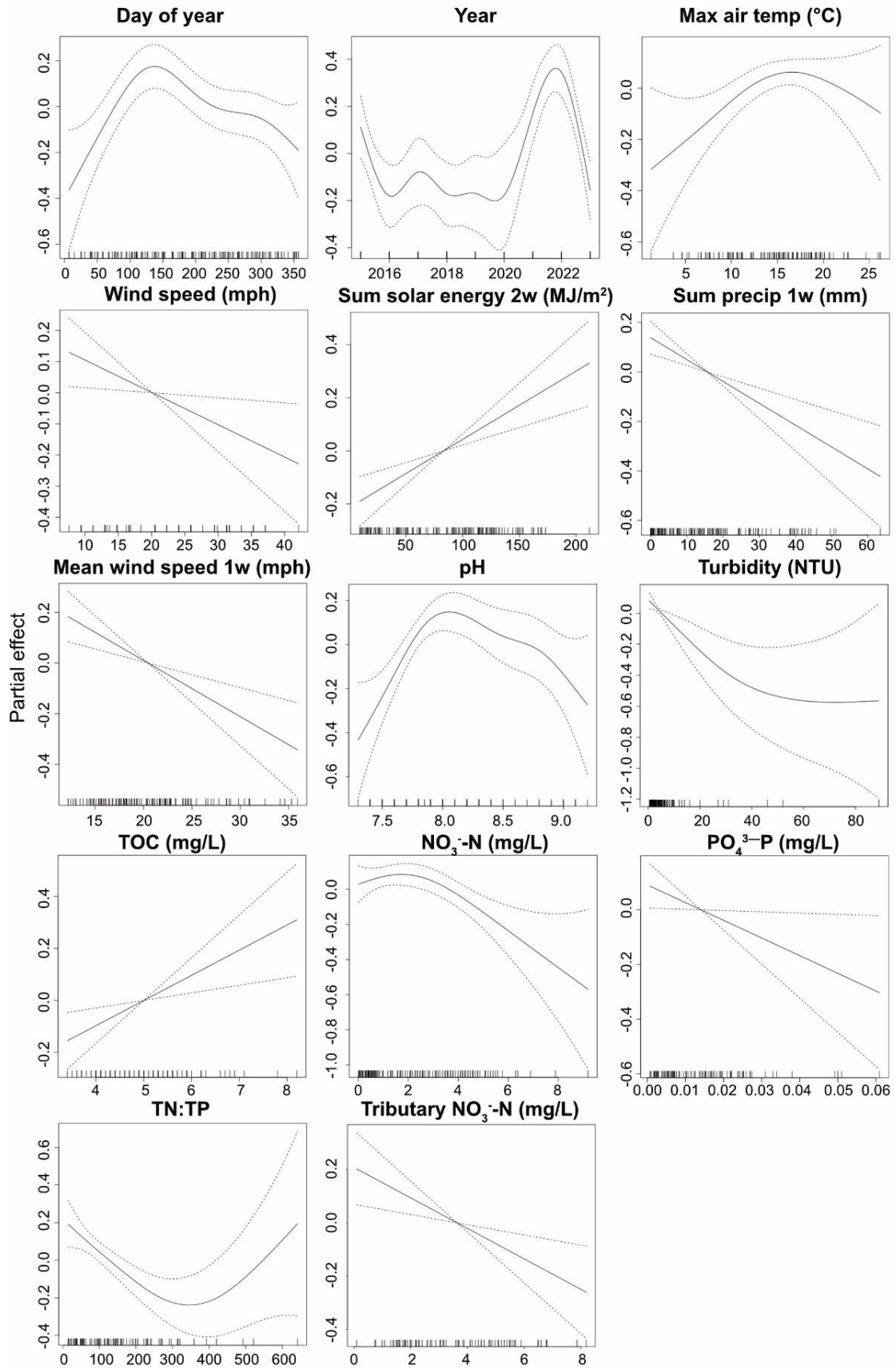
Figure 4.21. GAM selection for geosmin. Dotted line is 95% confidence interval. X-axis indicates all the plotted values for that parameter and vertical lines inside the plot along the x-axis indicate the distribution of those datapoints, units for the x-axis are in each plot title.

NO$_3^-$-N in the reservoir (9% deviance explained, n=140) and the tributary (9% deviance explained, n=85) as well as PO$_4^{3-}$-P in the reservoir (4% deviance explained, n=93) had a negative effect on geosmin (Table 4.7). Reservoir PO$_4^{3-}$-P and tributary NO$_3^-$-N had approximately linear relationships, whilst NO$_3^-$-N in the reservoir showed a slight increase up to around 2 mg N/L and then it followed a negative linear trend for increasing NO$_3^-$-N concentrations (Figure 4.21). TN:TP ratio explained 17% of the deviance (n=92) and showed a strong decreasing effect on geosmin from 0 to 350 but then increased again for values >400, but the uncertainty is quite large (Figure 4.21). NH$_4^+$:NO$_3^-$ ratio did not have a statistically significant effect on geosmin (Table 4.7).

Interestingly, for geosmin it seemed that laboratory measurements of turbidity and pH explained 9% of the deviance (n=136) and 15% of the deviance (n=136), respectively (Table 4.7). Turbidity revealed a negative trend with a decreasing effect on geosmin from 0 to 40 NTU and then a plateau with large uncertainty bounds (Figure 4.21). pH revealed an interesting trend, with a pattern that showed resemblance to the GAM with day of year (Figure 4.21). The pH pattern had a sharp increase from pH 7.3 to pH 8 where it peaked and then slowly decreased from pH 8 to pH 9.2 with a slight tendency for an increasing effect around pH 8.8 (Figure 4.21). TOC in the reservoir also explained some of the deviance (6%, n=117) with an approximately positively linear effect on geosmin, but it was nowhere near as good a predictor of geosmin as it was for 2-MIB (Table 4.7, Figure 4.21).

The sum of solar energy 2 weeks prior (10% deviance explained, n=141) and the sum of precipitation 1 week prior (11% deviance explained, n=141) were found to be the best weather parameters to explain the geosmin patterns (Table 4.7). The sum of solar energy showed a positive, approximately linear effect on geosmin, whilst the sum of the precipitation had an approximately linear negative effect on geosmin (Figure 4.21). This suggests that geosmin production is driven by calm, stable weather 1-2 weeks prior, with low rainfall and high solar energy.

Full GAM to predict 2-MIB

The best GAM to predict log_mib included smoothing terms for NO$_3^-$-N, maximum wind gust 2 weeks prior and minimum temperature 2 weeks prior as well as a tensor product which combines year and day of year (DOY); year as a cubic regression spline (cr) and DOY as a cubic cyclic spine (cc). The full model is described below:

GAM 2-MIB = gam(log_mib ~ s(no3n) + s(max_windgust_2weeks_prior) + s(min_temp_2weeks_prior) + te(year, DOY, bs = c("cr", "cc"), k=c(8,12)), data = complete_df, method='ML', family = tw(link="log")).

GAM residuals (Appendix C, Figure S.11), model parameters (Appendix C, Table S.12) and k.check results (Appendix C, Table S.13) were found to be acceptable. The impact from autocorrelation was limited (Appendix C, Figure S.12). The model explained 92.7% of the deviance, with an adjusted R^2 of 0.93 (n=136). Smoothing terms s(no3n) (p-value = 0.0422) and s(max_windgust_2weeks_prior) (p-value = 0.0492) were statistically significant, whilst smoothing term s(min_temp_2weeks_prior) was not statistically significant (p-value = 0.4134). The tensor product (year, DOY) was very statistically significant (p-value < 2E-16) (Appendix C, Table S.12). The partial effect of each smoothing term and the tensor product on log_mib, with grey 95% confidence intervals, is visualised in Figure 4.22. Increases in nitrate had a negative effect on log_mib, and this effect was stronger > 2.5 mg/L of NO_3^--N (Figure 4.22). Higher maximum wind gusts 2 weeks prior had a negative effect on log_mib, which was approximately linear (Figure 4.22). An increase in the minimum temperature 2 weeks prior had a positive approximately linear effect on log_mib, but there was a lot of uncertainty around the model (Figure 4.22). The heatmap of the tensor product with year and day of year revealed more predictable yearly periods of positive (red) and negative (blue) effect on log_mib, compared to log_geosmin (Figure 4.22). This is also visualised in the individual GAM results for year and DOY (Figure 4.20).
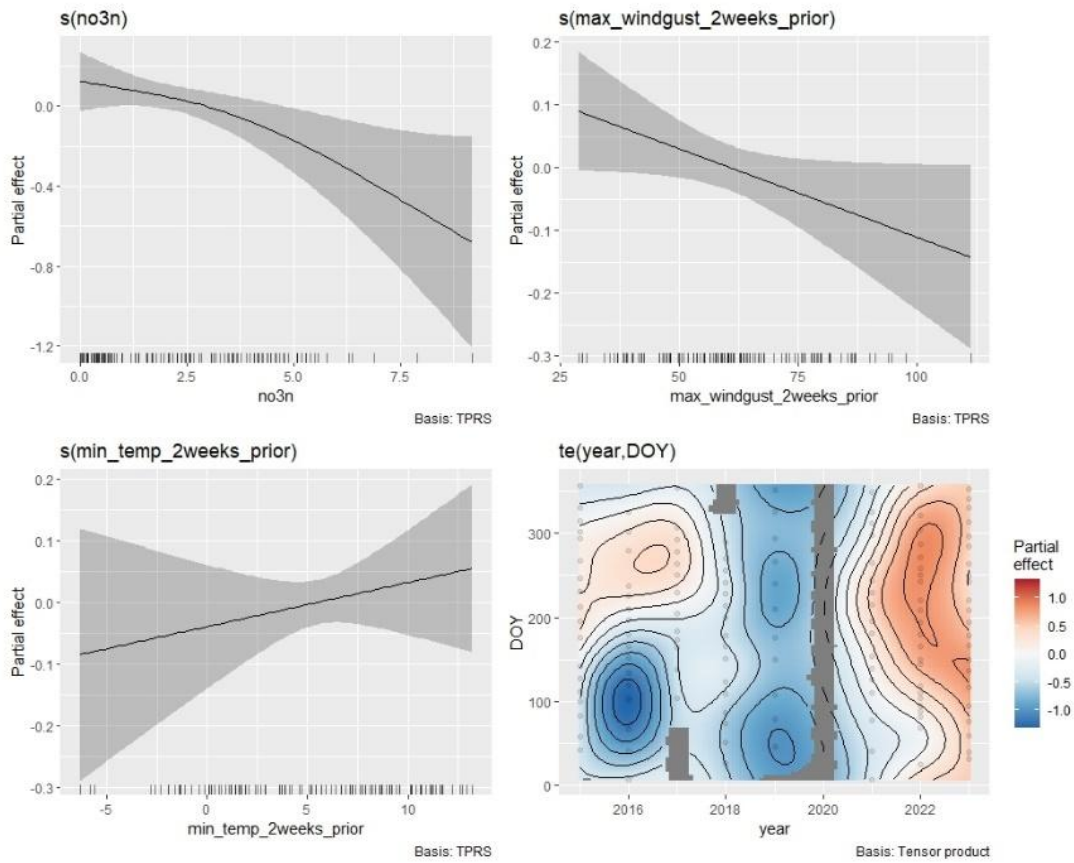
Figure 4.22. The partial effect of each smoothing term and the tensor product on log₁₀(MIB+1) in the full GAM. Grey band 95% confidence interval, lines at x-axis indicate where datapoints are. X-axis units are the same as in the GAM plots in Figure 4.20.*

The fitted values from the GAM model seemed to generally follow the patterns in 2-MIB well, except for 2018 (Figure 4.23).

GAM fit: log_mib ~ no3n + max_windgust_2weeks_prior + min_temp_2weeks_prior + (year, DOY)
R squared = 0.93, Deviance explained = 93%
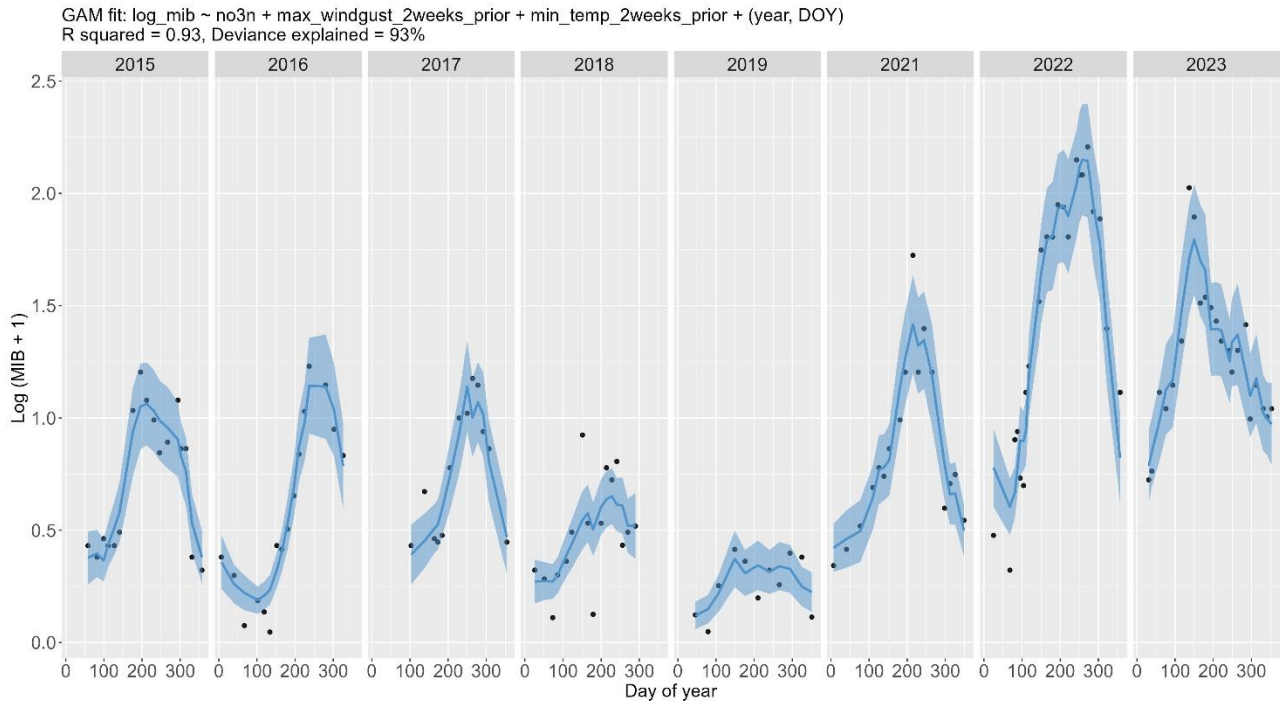
Figure 4.23. Fitted values over time from the GAM model for $\log_{10}(MIB+1)$ (blue line with shaded blue 95% confidence interval) compared to observations (dots).

Full GAM to predict geosmin

The best GAM to predict log_geosmin included smoothing terms for NO_3^--N and sum of solar energy 2 weeks prior as well as a tensor product which combines year and day of year (DOY); year as a cubic regression spline (cr) and DOY as a cubic cyclic spine (cc). The full model is shown below.

GAM geosmin = gam(log_geosmin ~ s(no3n) + s(sum_solarenergy_2weeks_prior) + te(year, DOY, bs = c("cr", "cc"), k=c(8,12)), data = complete_df, method='ML', family = tw(link="log")).

The GAM residuals (Appendix C, Figure S.9), model parameters (Appendix C, Table S.10) and the output from k.check (Appendix C, Table S.11) were acceptable. The impact from autocorrelation was limited (Appendix C, Figure S.10). The model explained 90.5% of the deviance with an R^2 adjusted of 0.85 (n=136). The smoothing term s(no3n) was not significant (p-value of 0.2443), the smoothing term s(sum_solarenergy_2weeks_prior) was close to being statistically significant (p-value = 0.0819) and the tensor product te(year, DOY) was very statistically significant (p-value < 2E-16) (Appendix C, Table S.10). The partial effect of the smoothing terms on log_geosmin, with 95% confidence intervals in grey, revealed that nitrate had a negative relationship with log_geosmin which was approximately linear, and the sum

186

of solar energy 2 weeks prior had a positive relationship with log_geosmin which was also approximately linear (Figure 4.24). The tensor product with year and day of year was plotted in Figure 4.24 as a heatmap which indicates periods of positive (red) and negative (blue) effect on log_geosmin. Figure 4.24 can be compared to the individual GAM plots for year and DOY in Figure 4.21, which showed that the timing of the yearly geosmin event and the strength of the event is not the same every year. However, the tensor product combines the information of year and DOY and therefore it was still the best predictor for log_geosmin.
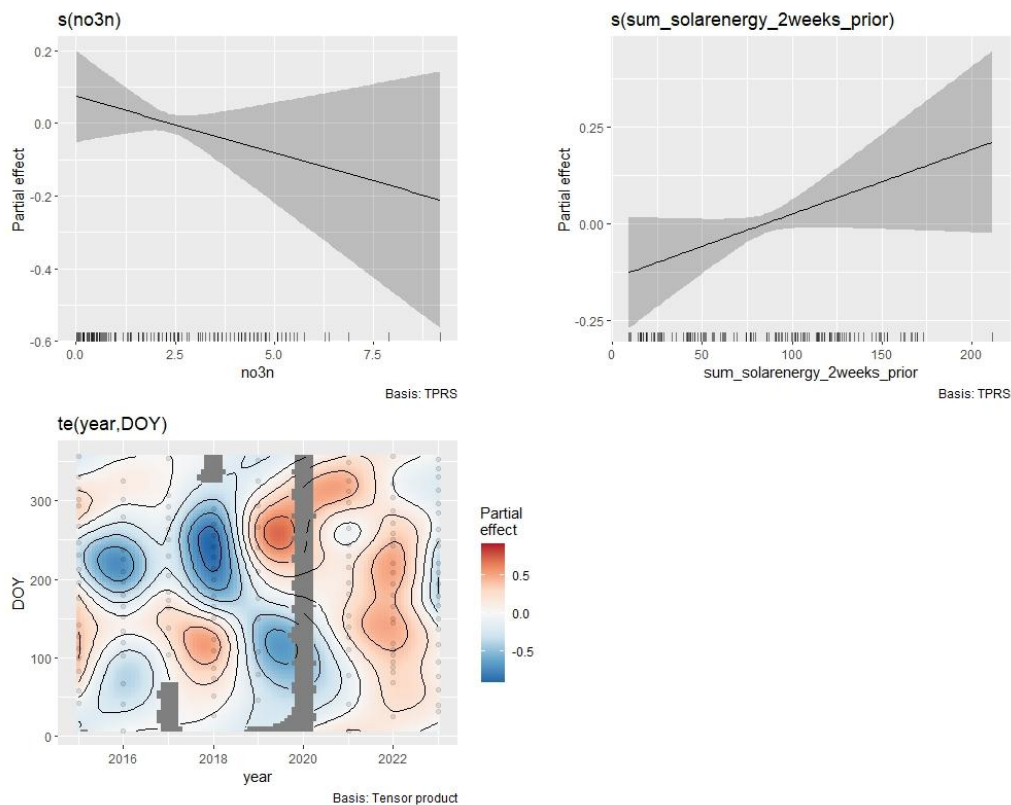
Figure 4.24. The partial effect of each smoothing term and the tensor product on log$_{10}$(geosmin+1) in the full GAM. Grey band 95% confidence interval, lines at x-axis indicate where datapoints are. X-axis units are the same as in the GAM plots in Figure 4.21.

The fitted values from the GAM model estimated the periods of increased geosmin relatively well (Figure 4.25).

GAM fit: log_geosmin ~ no3n + sum_solarenergy_2weeks_prior + (year, DOY)
R squared = 0.85, Deviance explained = 90%
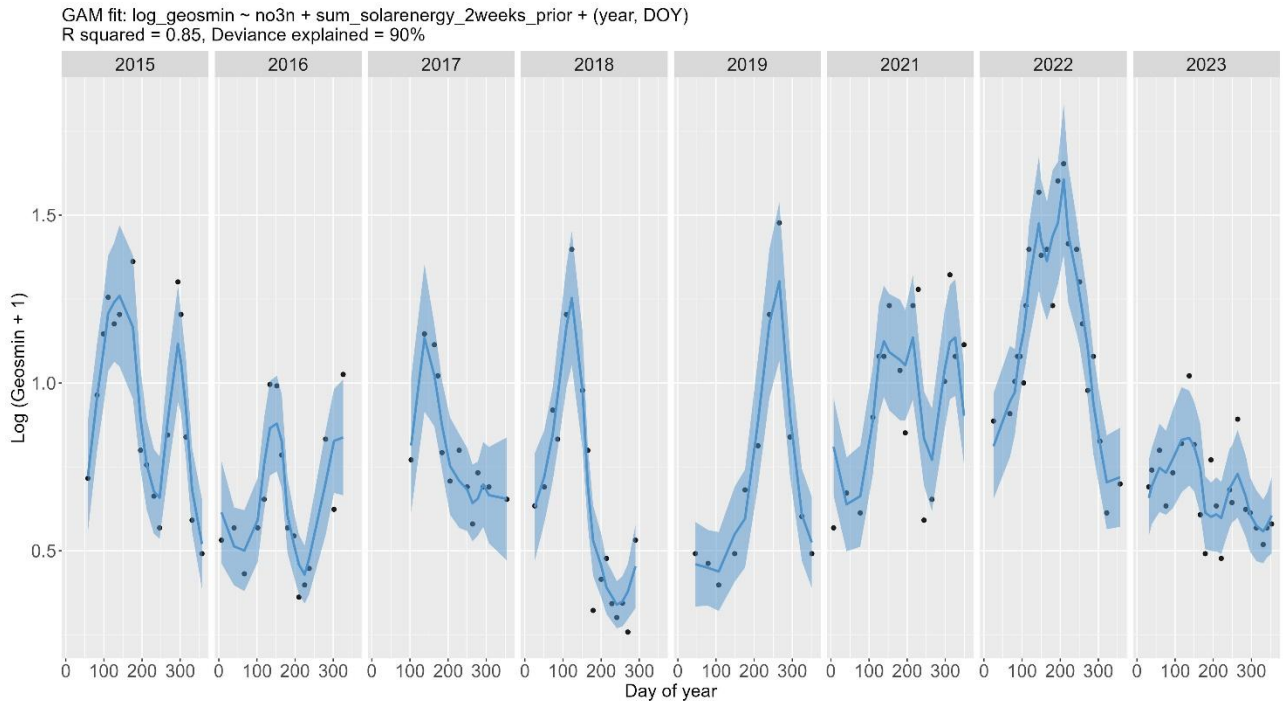
Figure 4.25. Fitted values over time from the GAM model for log_{10}(geosmin+1) (blue line with shaded blue 95% confidence interval) compared to observations (dots).

4.3.5 Case study: using sensor data to predict T&O events

Sensor data was plotted with relevant DCWW and weather parameters for the years 2022 and 2023 to better see the yearly pattern (Figure 4.26), and nutrient fractions in the tributary and reservoir were plotted for the same period (Figure 4.27).
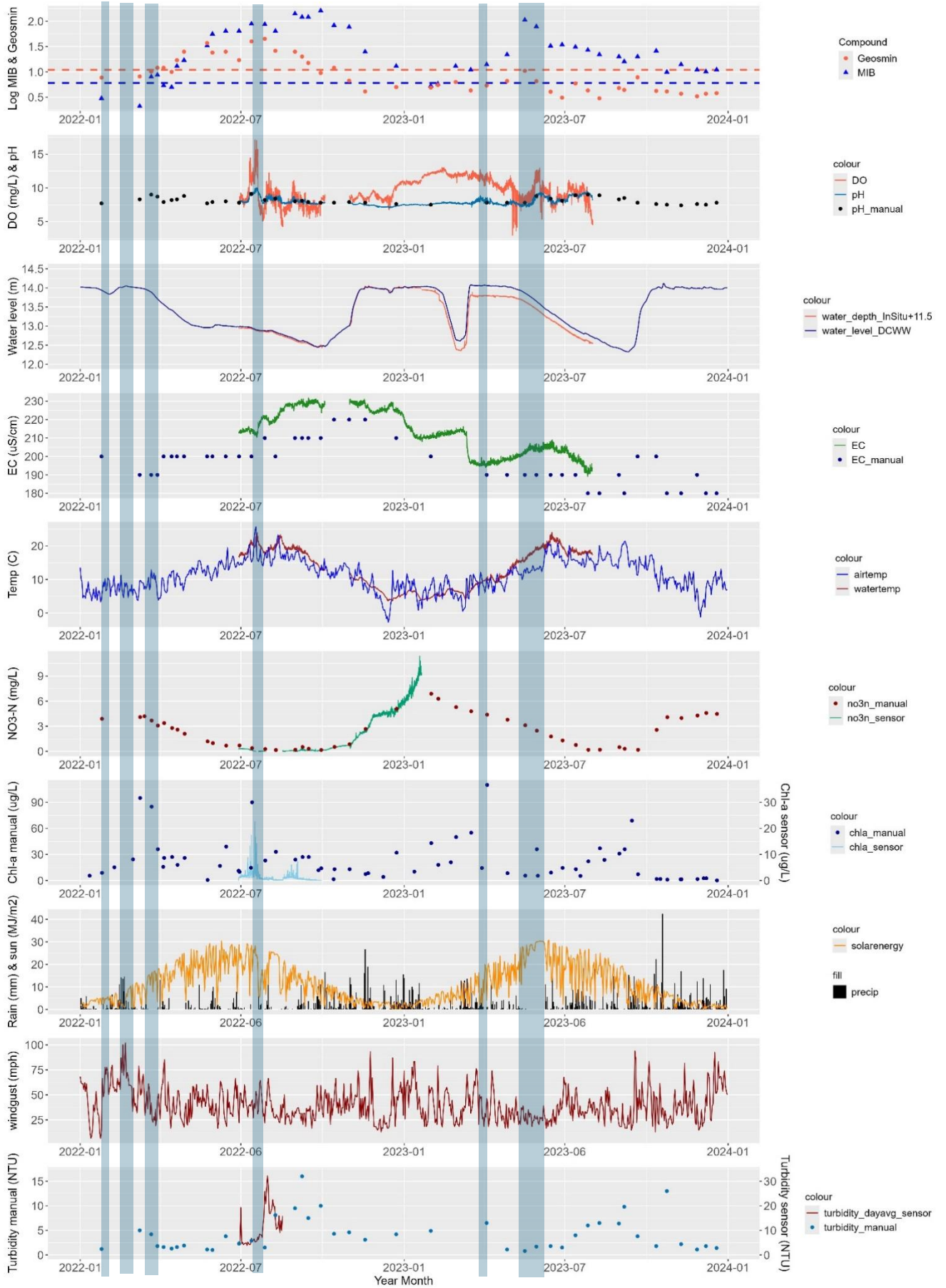
Figure 4.26. Reservoir 1 sensor, weather and lab data 2022-2023. Shaded = stable weather.
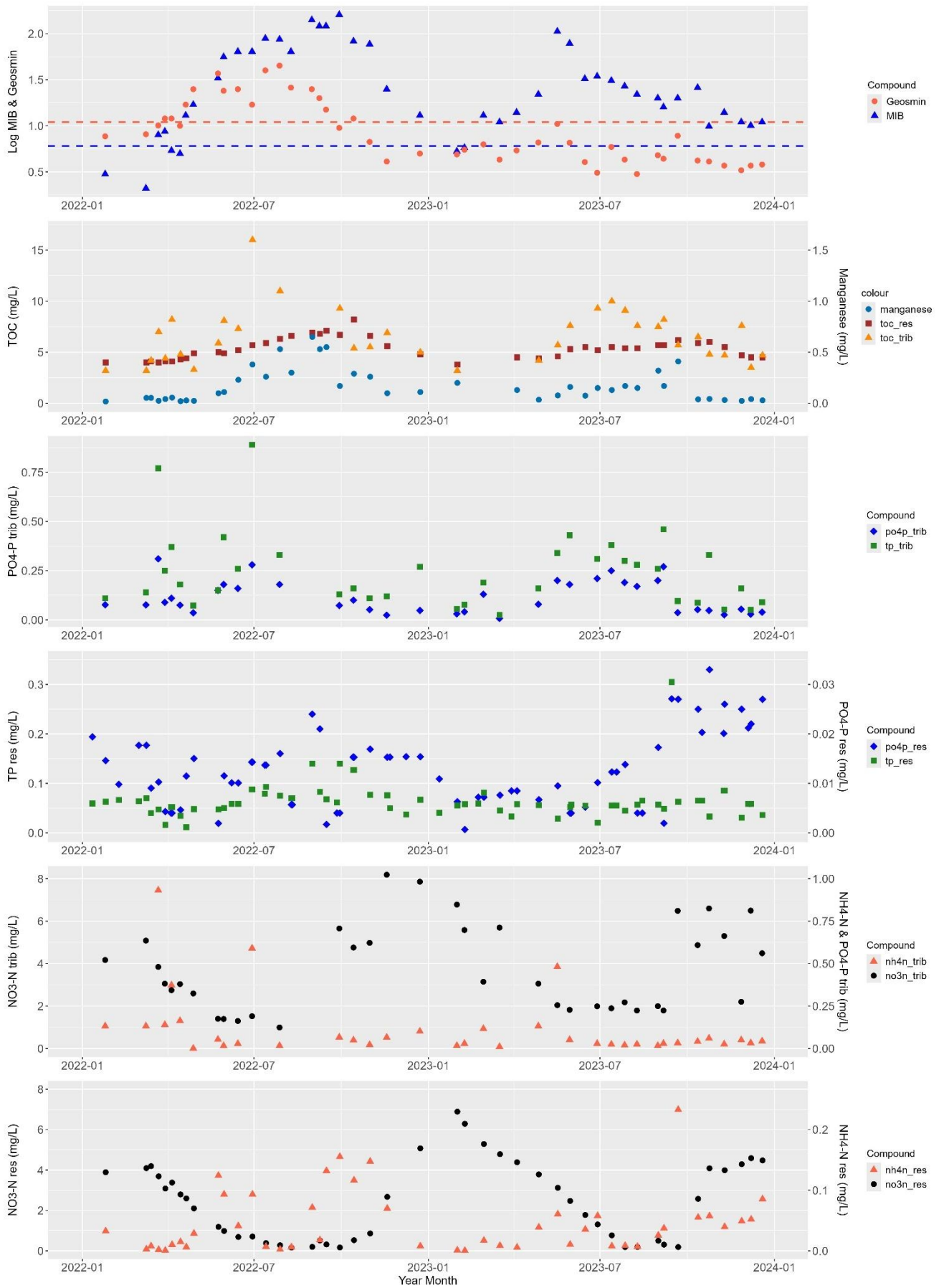
Figure 4.27. Samples from tributary & abstraction Reservoir 1 2022-2023.

NO$_3^-$ and NH$_4^+$ sensor technologies

The Clearwater NO$_3^-$-N analyser had good agreement with laboratory measured samples (Figure 4.28). The NO$_3^-$-N concentration in the reservoir showed a steep increase in autumn-winter from October/November until February and a gentle decline throughout spring and summer where it stabilized at <0.5 mg/L from mid-July until October/November (Figure 4.27 and Figure 4.28). The significant peak captured by the sensor data fell between laboratory samples, and the sensor stopped working just after this peak, so there is uncertainty on the event duration when NO$_3^-$-N peaked in January 2023 at >10 mg N/L.
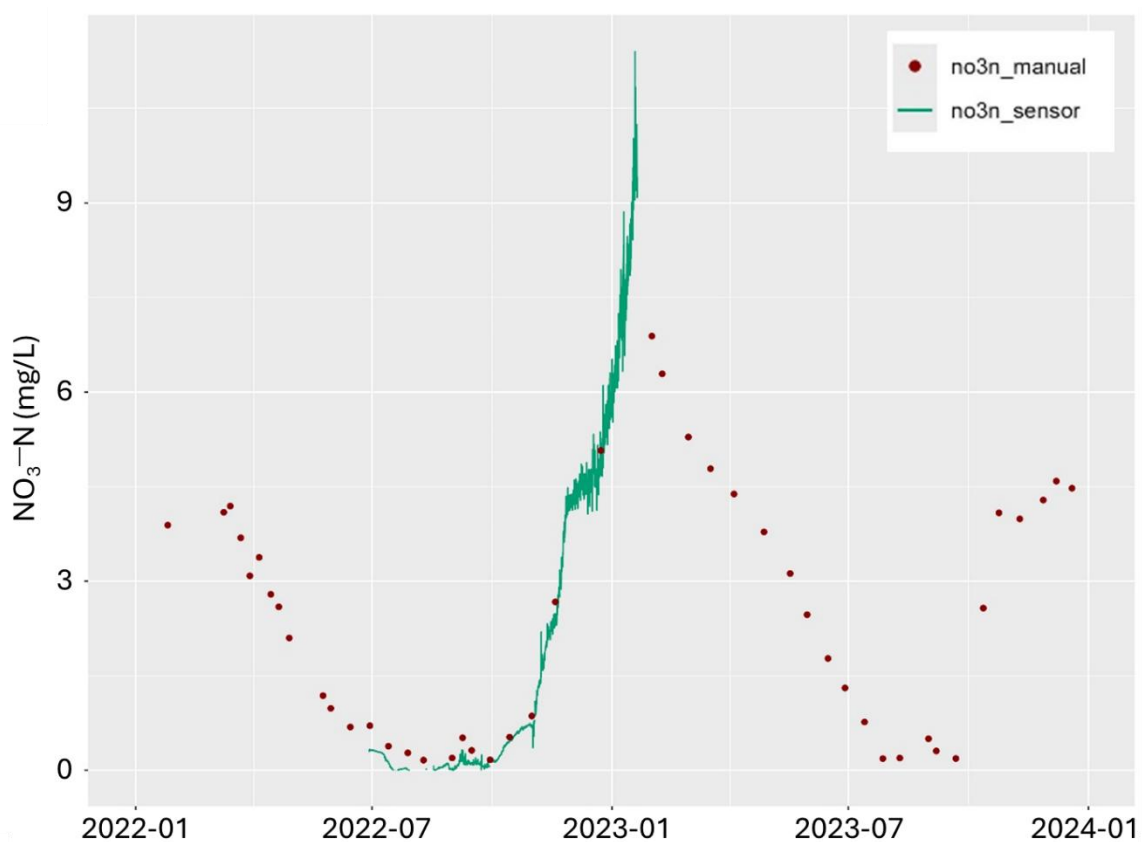
Figure 4.28. NO$_3^-$-N (mg/L) measured frequently by Clearwater analyser compared to manual laboratory samples.

The In-Situ Aqua TROLL NO$_3^-$-N sensor had poor agreement with laboratory samples, even though a similar trend in data was observed in November 2022 when NO$_3^-$-N started increasing (Figure 4.9 in methods). The sensor could not be properly calibrated in the field, as the sensor would not stabilise and identify the mV difference between concentrations of 5.6 mg/L and 1 mg/L NO$_3^-$-N. In laboratory conditions, the sensor seemed relatively stable, and a calibration curve was determined (R^2 = 0.9919), with

a range of NO_3^--N concentrations from 0.15 mg N/L to 20.31 mg N/L (Figure 4.7 in methods, Section 4.2.6). When the raw mV data was calculated to NO_3^--N concentrations with the calibration curve from the lab, it resulted in unrealistically high values of NO_3^--N, and the attempted data correction did not result in better alignment with the laboratory data (see methods, Section 4.2.6, Figure 4.9).

The In-Situ Aqua TROLL NH_4^+-N sensor had similar issues to the NO_3^--N sensor and field calibration was not possible. An exponential calibration curve was established in the laboratory for this sensor, which performed well under laboratory conditions (See methods, Figure 4.7) with a range of 0.021 mg N/L to 1.777 mg N/L (R^2 = 0.9397). However, when the calibration curve was applied to field data (raw mV) to calculate NH_4^+-N concentrations, it did not perform well (Figure 4.10 and Figure 4.11 in methods). It seemed that the NH_4^+-N sensor could only detect concentrations above 0.08 - 0.10 mg/L of NH_4^+-N, as mV values did not respond to lower concentrations (See Figure 4.11 in methods). The concentration of NH_4^+-N in the reservoir only went up to 0.15 mg/L in the measurement period, which happened largely in the period when the sensor was not working, so I don't think this dataset can justify whether it was working or not.

Chlorophyll a

The chlorophyll *a* sensor did not work for a long period of time, and it did not have a wiper brush, which meant there was a build-up of organic material on the optical sensor that reduced its sensitivity. The sensor data captured a chlorophyll *a* peak in July 2022 which was measured by the laboratory data but then underestimated the trend in chlorophyll *a* (assuming that laboratory data is correct) for the following 3 months. Chlorophyll *a* values measured *in situ* with the sensor were different from laboratory extracted chlorophyll *a*. There was a large range in the sensor data that was not fully captured by the manual samples, which indicates that manual sampling can easily miss a spike.

Turbidity

There was poor agreement between turbidity measurements in the laboratory and those captured by the *in situ* sensor (Figure 4.26). During the short period that turbidity was measured in the reservoir with a sensor, daily averages of turbidity showed a trend that was not visible from the manual samples (Figure 4.29).
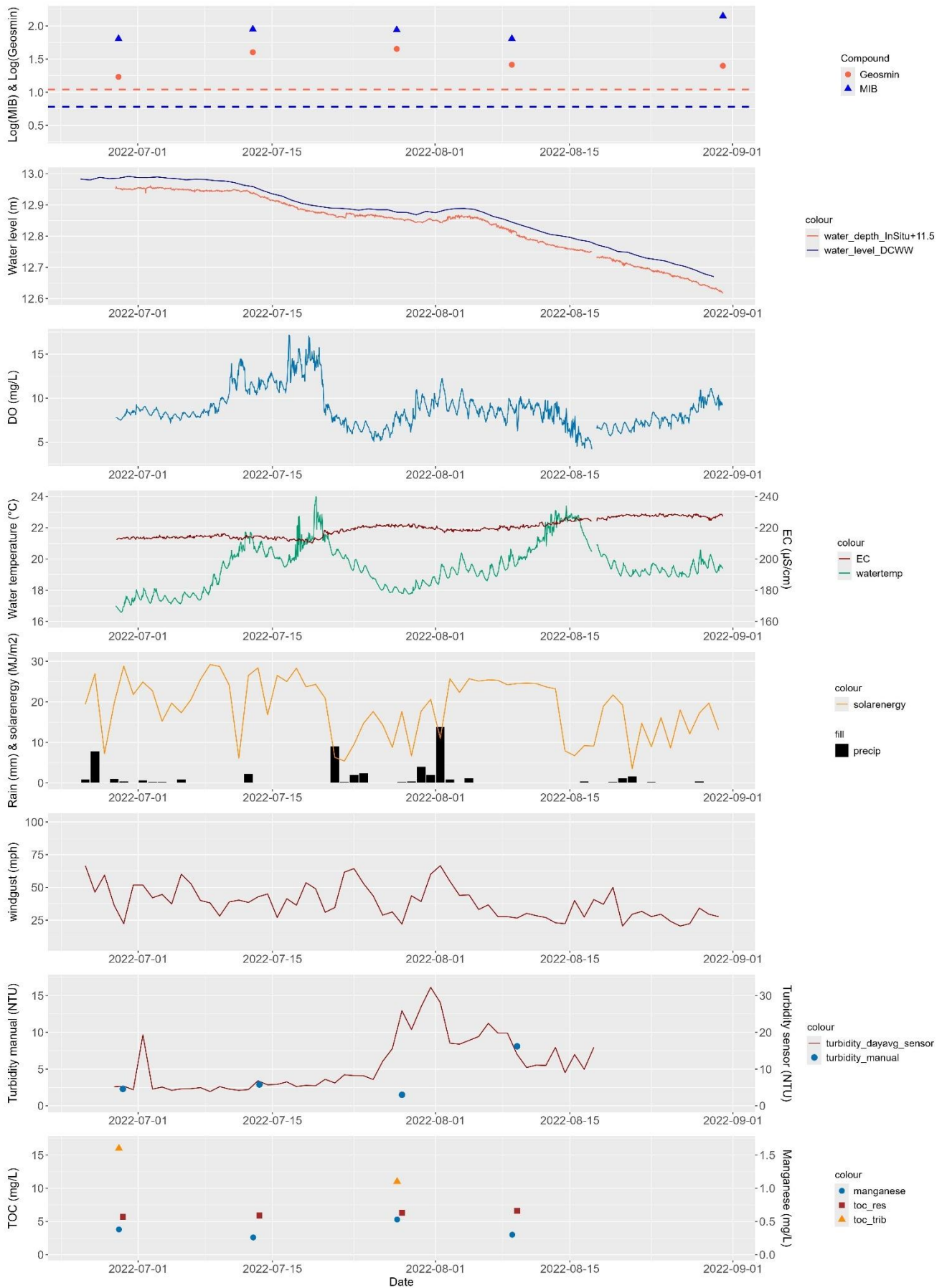
Figure 4.29. Sensor, weather and lab data in Reservoir 1 for July – September 2022.

Large fluctuations in hourly turbidity data meant that a daily averaged turbidity value was better able to capture the general trend. A week before the turbidity maximum in the second half of July, a precipitation event occurred with higher wind gusts and a dip in solar energy, which caused a rapid decrease in water temperature and DO (Figure 4.29). Around the time of the turbidity maximum, TOC in the tributary was >10 mg/L, which was the second highest concentration in 2022-2023 (Figure 4.29).

Electrical conductivity

Specific conductivity data was used in this analysis, which is temperature corrected to 25 °C, further referred to as electrical conductivity (EC). EC seemed to be dependent on water level, with decreasing volume of water causing an increase in EC. Since EC measures the concentration of dissolved ions in water, a reduced water volume would concentrate the ions and increase the EC (Figure 4.29 and Figure 4.30). EC generally increased in summer and decreased in winter (Figure 4.26), likely due to water level changes caused by summer evaporation and winter refilling. Unfortunately, the sensor data did not capture one whole year, so the seasonal pattern cannot be identified accurately. EC data did not seem to capture important events in the reservoir, for example precipitation (Figure 4.29). Laboratory EC measurements of manual DCWW samples (corrected to 20 °C) were similar to In-Situ Aqua TROLL 600 EC (25 °C) but could not capture the data variability (Figure 4.26), which indicates the importance of measuring EC with *in situ* sensors.

Water level

Water depth (m) from the sensor measurements was converted by adding 11.5 meters to compare with DCWW water level (m) that was measured in the abstraction tower (see Section 4.2.7 in methods). The data followed the same pattern, but the sensor data after January 2023 was a bit lower than DCWW water level (m), which was probably because the sensor was not deployed at exactly the same depth after maintenance (Figure 4.30). The changes in water level showed a gradual decrease of approximately 1.5 meters from springtime to autumn due to evaporation, followed by a relatively fast refilling period in autumn (Figure 4.30). There was one non-natural drawdown event in this dataset that lasted from mid-February to early March 2023, where the water levels went down 1.5 m over 17 days, which caused an elevated water temperature and EC level. This was followed by a rapid increase in water level again of 1.5 m over 6 days, which caused a relatively large drop in EC of 20 µS/cm

with a simultaneous drop in water temperature (Figure 4.30). This could be evidence of the nutrient and organic material flushing process that happened due to abstraction and refilling with cleaner rainwater.
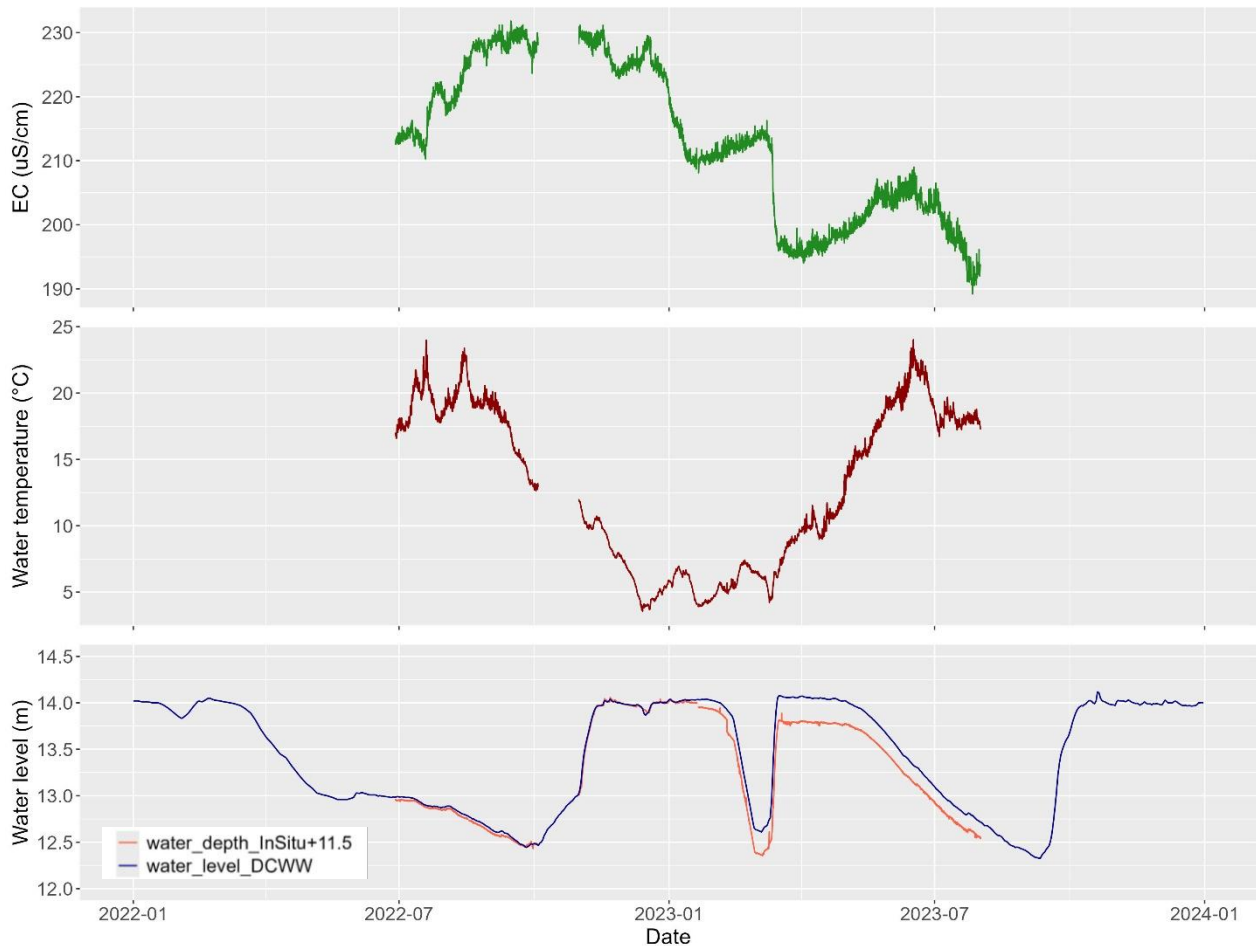
Figure 4.30. Water level (m), water temperature and EC sensor comparison.

Dissolved oxygen, pH and ORP

Dissolved oxygen (DO) data showed a general pattern of strongly fluctuating concentrations in summer with peak oxygen concentrations >15 mg/L but also the lowest concentrations <5 mg/L, whilst in winter the average dissolved oxygen concentrations are higher, and the fluctuations are much smaller (Figure 4.31). DO concentrations in summer showed a diurnal pattern of high DO during the day and low DO during the night, caused by photosynthesis-respiration cycles (Figure 4.29).

pH and oxidation-reduction potential (ORP) are related but distinct measurements. pH indicates the system's tendency to accept or donate hydrogen ions, influenced by the presence of acids and bases in the solution. ORP, on the other hand, reflects the system's tendency to gain or lose electrons, which is determined by the concentration

195

of all oxidizing and reducing agents in the solution. In oxygen rich surface water layers, the ORP will be fully determined by the behaviour of oxygen and hence the measurement of ORP does not add any value over dissolved oxygen (DO). ORP measurements can be interesting to reveal oxidation-reduction gradients when measured at different depths in the water column. Figure 4.31 showed that the ORP pattern was the same as pH and dissolved oxygen but then mirrored. pH manual samples that were done in the DCWW laboratory showed a good comparison with the sensor measurements (Figure 4.31), which is surprising considering the sample was often measured days later at a different temperature. Similar to dissolved oxygen, pH and ORP have a larger diurnal range in summer as phytoplankton productivity causes increases and decreases in dissolved ions due to photosynthesis-respiration cycles.
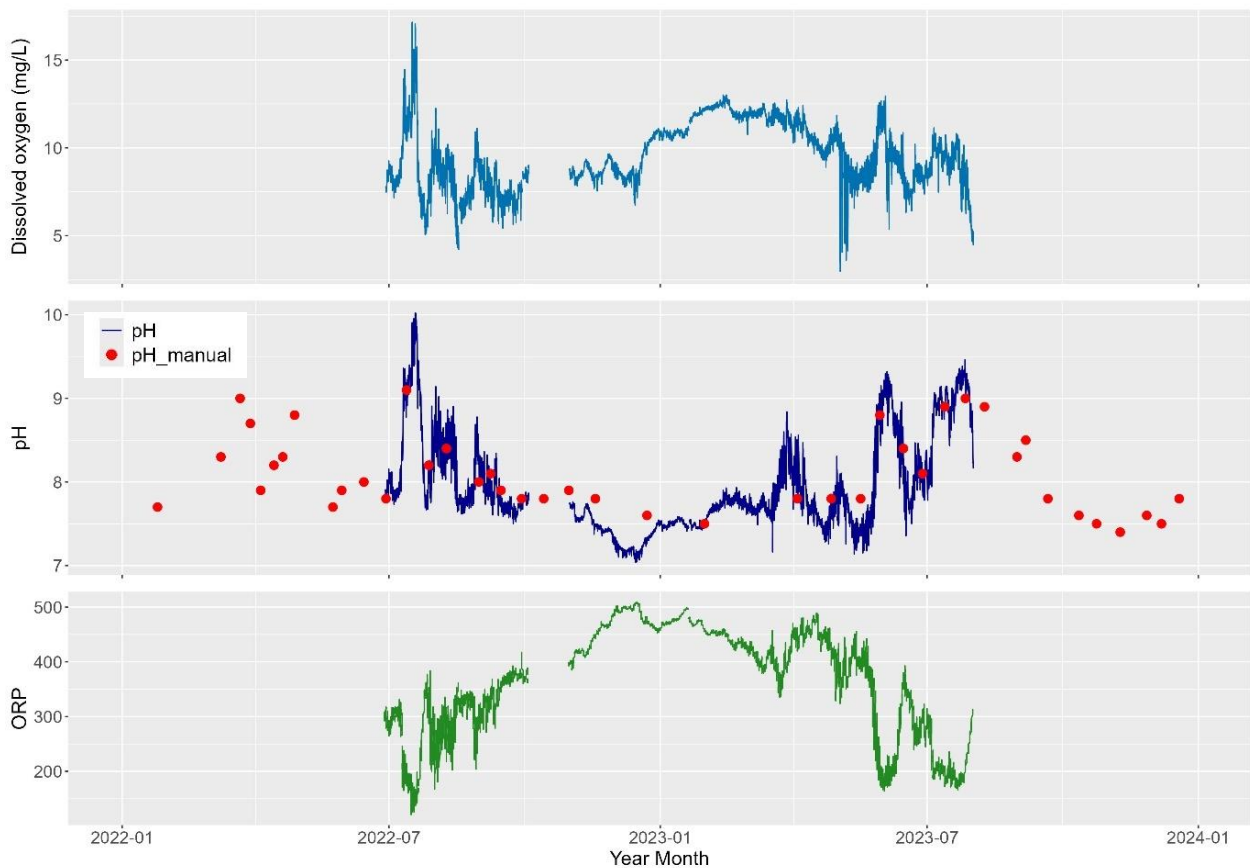
Figure 4.31. Oxidation-reduction potential (ORP), pH and dissolved oxygen (DO) sensor measurements and pH manual samples.

4.3.6 Correlations & simple GAMs sensor data

Correlations with a log_{10}-transformed dataset were performed for geosmin as well as 2-MIB, using the DCWW manual sample dataset from 2022 and 2023 merged with the calculated sensor and weather parameters (including lags of 1, 2, 3 or 4 weeks prior).

A selection of parameters was made and a GAM with a smoothing function was executed to fit the predictor to the $\log_{10}$ transformed geosmin and 2-MIB concentrations. The correlations and GAM outcomes of the most interesting parameters for 2-MIB and geosmin were displayed in Table 4.8 and Table 4.9, respectively. Chlorophyll *a* and turbidity data was too sparse to be used for this analysis. The GAMs for these parameters were plotted in Figure 4.32 and Figure 4.33, for 2-MIB and geosmin, respectively. The full set of GAM results can be found in Appendix C Table S.14, Table S.15 and Table S.16.

2-MIB

The best calculated sensor parameter to predict $\log_{10}$(MIB+1) (further referred to as 2-MIB) was average nitrate 2 weeks prior (94% deviance explained, n=12) (Table 4.8). Similarly to manual NO_3^--N measurements in the full dataset (Section 4.3.4), the sensor data showed that the strong seasonal pattern in NO_3^--N negatively correlated with 2-MIB concentrations (Table 4.8).

Maximum air temperature 3 weeks prior (80% deviance explained, n=21) and average water temperature 2 weeks prior (80% deviance explained, n=22) both showed a clear increasing effect on 2-MIB with increasing temperature, but it plateaued for maximum air temperature at 25 °C and for water temperature at 15 °C (Figure 4.32).

Average range 3 weeks prior of dissolved oxygen (77% deviance explained, n=21), pH (39%, n=21) and ORP (61% deviance explained, n=21) all have a positive effect on 2-MIB but a plateau for the highest levels >2 mg/L DO, >0.6 pH units and >60 ORP units (Figure 4.32). An increased range for these parameters indicates increased phytoplankton productivity, which seems to be related to 2-MIB concentrations. However, this is only data from the last two years of the dataset in which 2-MIB and geosmin were exceptionally high.

Maximum depth 1 week prior explained 47% of the deviance (n=22) and showed an approximately linear negative relationship with 2-MIB (Table 4.8, Figure 4.32). Maximum electrical conductivity on the day (19% deviance explained, n=23) showed a positive effect on 2-MIB (Table 4.8, Figure 4.32).

Table 4.8. Simple GAM results for 2-MIB, GAMs are visualised in Figure 4.32. Column 'short name' refers to the names used in Figure 4.32.

| Predictor for 2-MIB	Short name	Log-Log correlation	GAM P_Value	GAM R2_Adj	GAM Deviance expl.	GAM obs.
atemp_max_max_ 3weeks_prior	Max air temp 3w	0.78	0.000	0.699	0.800	21
depth_max_max_1 week_prior	Max depth 1w	-0.70	0.000	0.458	0.467	22
do_range_avg_3w eeks_prior	Range DO 3w	0.82	0.000	0.731	0.766	21
ec_max	Max EC	0.45	0.024	0.170	0.190	23
no3n_mean_avg_2 weeks_prior	Mean NO_3^--N 2w	-0.96	0.000	0.905	0.941	12
orp_range_avg_3w eeks_prior	Range ORP 3w	0.76	0.001	0.560	0.608	21
ph_range_avg_3w eeks_prior	Range pH 3w	0.55	0.018	0.322	0.387	21
wtemp_mean_avg _2weeks_prior	Mean water temp 2w	0.83	0.000	0.738	0.803	22
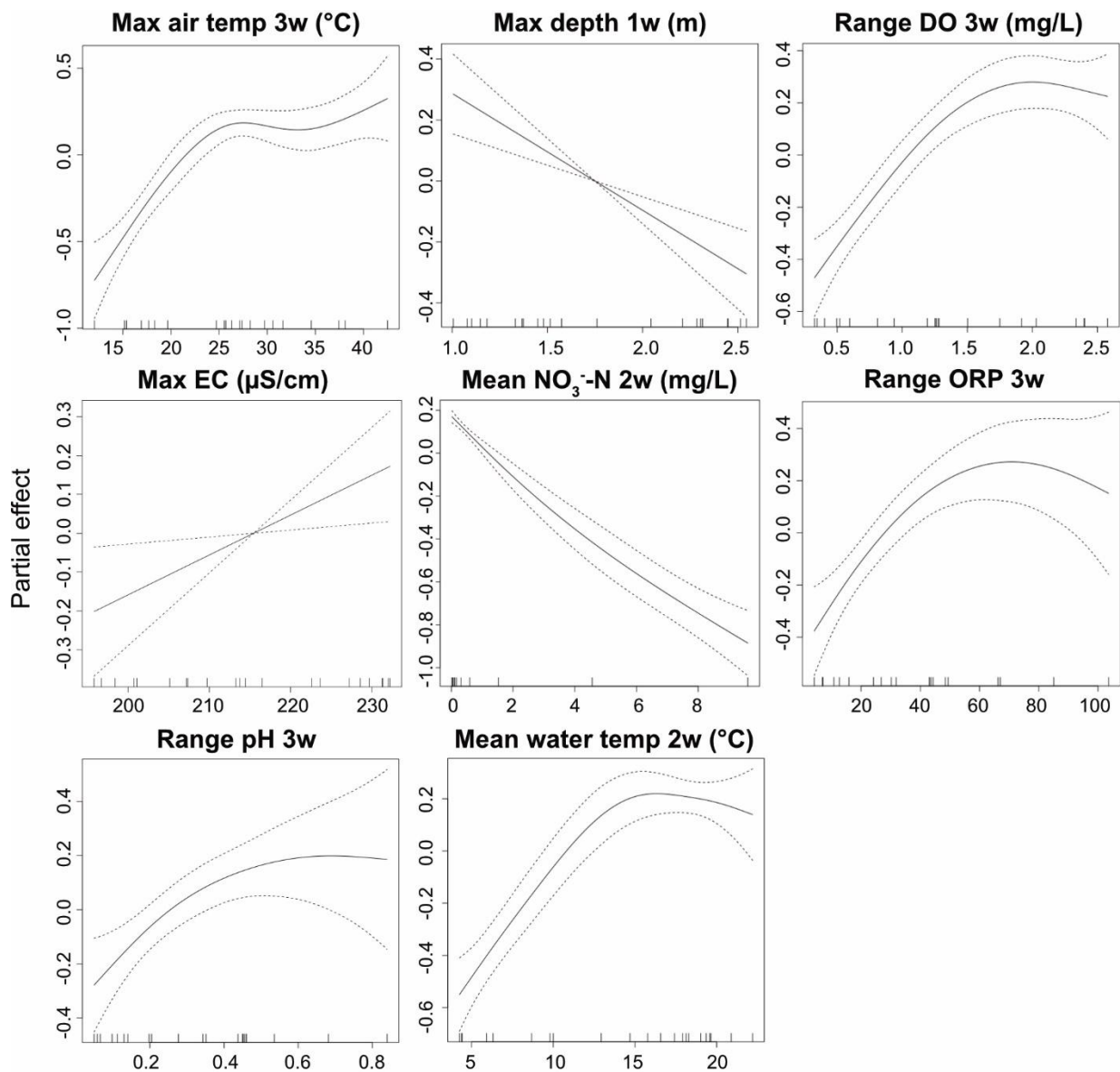
Figure 4.32. Simple GAM results for 2-MIB. Dotted lines indicate the 95% confidence interval. X-axis indicates all the plotted values for that parameter and vertical lines inside the plot along the x-axis indicate the distribution of those datapoints, units for the x-axis are in each plot title.

Geosmin

The best calculated sensor parameter to predict $\log_{10}$(geosmin+1) (further referred to as geosmin) was maximum nitrate 1 week prior (80% deviance explained, n=11) (Table 4.9). A negative effect on geosmin was visible with increasing maximum NO_3^--N concentrations up to 3 mg N/L, above which it had a large uncertainty (Figure 4.33).

Maximum air temperature 3 weeks prior (37% deviance explained, n=21) and the weekly range of water temperature 3 weeks prior (38% deviance explained, n=21) both showed a positive, approximately linear effect on geosmin (Figure 4.33).

Weekly range of dissolved oxygen 3 weeks prior (47% deviance explained, n=21), weekly range of pH 1 week prior (29% deviance explained, n=22) and maximum daily range in ORP 1 week prior (43% deviance explained, n=22) all have an approximately linear positive effect on geosmin (Table 4.9, Figure 4.33). An increased range for these parameters or a higher maximum indicates increased phytoplankton productivity, which seems to be related to geosmin concentrations.

Maximum depth 1 week prior explained 31% of the deviance (n=22) and showed an approximately linear negative relationship with geosmin (Table 4.9, Figure 4.33). Minimum electrical conductivity on the day (27% deviance explained, n=23) (Table 4.9) showed a linear positive effect on geosmin (Table 4.9, Figure 4.33).

Table 4.9. Simple GAM results for geosmin. GAMs are visualised in Figure 4.33. Column 'short name' refers to the names used in Figure 4.33.

Predictor for geosmin	Short name	Log-Log correlation	GAM P_Value	GAM R2_Adj	GAM Deviance expl.	GAM obs.
atemp_max_max_3 weeks_prior	Max air temp 3w	0.56	0.003	0.346	0.369	21
depth_max_max_1w eek_prior	Max depth 1w	-0.52	0.004	0.234	0.312	22
do_weekly_range_3 weeks_prior	Range DO 3w	0.68	0.000	0.441	0.472	21
ec_min	Min EC	0.49	0.006	0.184	0.271	23
no3n_max_max_1w eek_prior	Max NO_3^--N 1w	-0.78	0.001	0.676	0.804	11
orp_range_max_1w eek_prior	Range ORP 1w	0.58	0.001	0.380	0.431	22
ph_weekly_range_1 week_prior	Range pH 1w	0.52	0.008	0.219	0.285	22
wtemp_weekly_rang e_3weeks_prior	Range water temp 3w	0.60	0.003	0.399	0.377	21
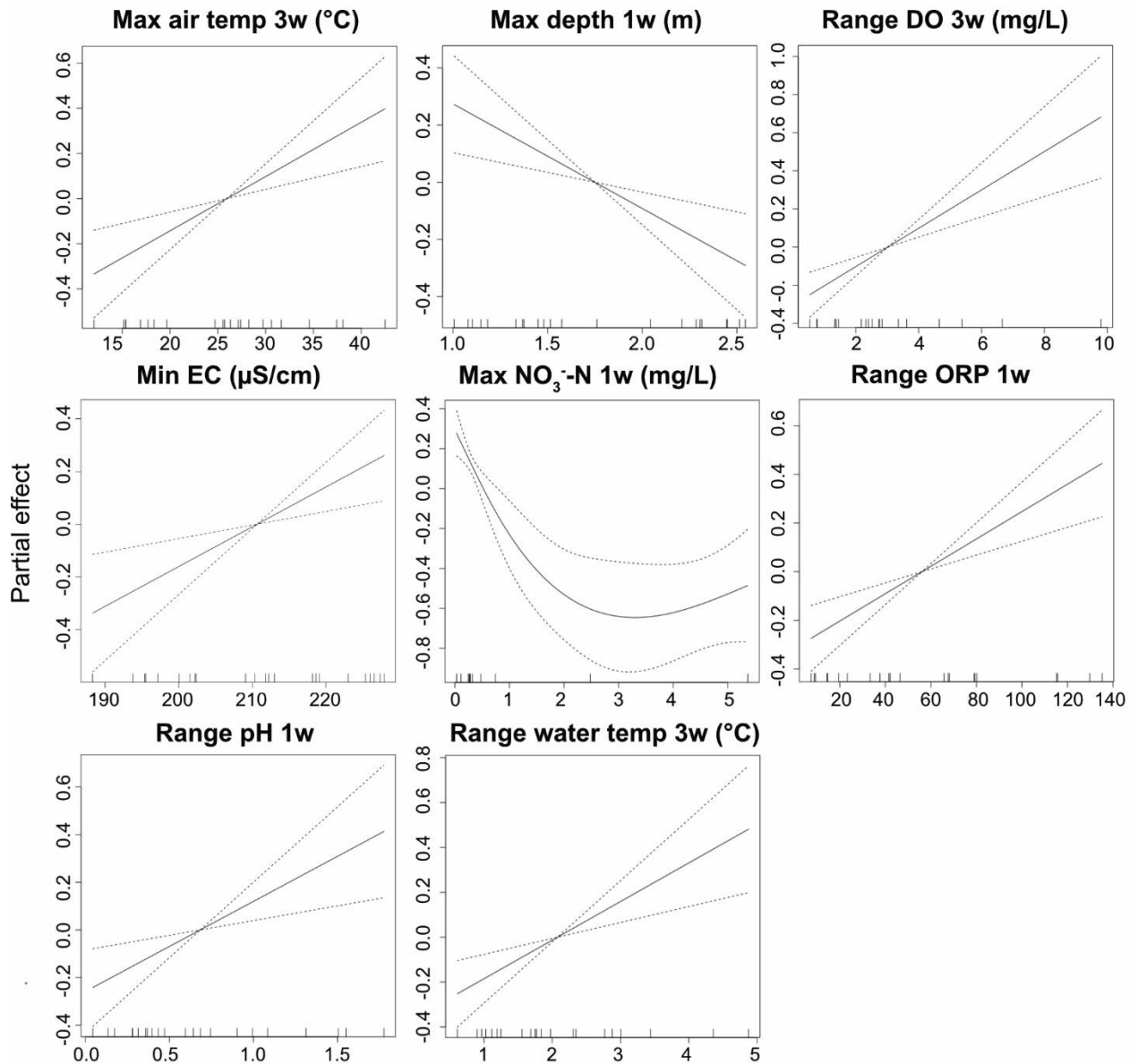
Figure 4.33. Simple GAM results for geosmin. Dotted lines indicate the 95% confidence interval. X-axis indicates all the plotted values for that parameter and vertical lines inside the plot along the x-axis indicate the distribution of those datapoints, units for the x-axis are in each plot title.

4.4 Discussion

4.4.1 Seasonal trends in environment, geosmin and 2-MIB

Manual sampling data in Reservoir 1 from 2015 to 2023 and daily weather parameters were plotted and analysed with PCA, individual GAMs and full GAMs for geosmin and 2-MIB separately, to detect patterns and potential predictors for geosmin and 2-MIB. Based on environmental conditions in the PCA, a geosmin and 2-MIB event could be effectively separated from a non-event. Geosmin had the strongest relationship with solar energy, air temperature, pH, chlorophyll *a* and various algae counts in the PCA, whereas 2-MIB was closely associated with air temperature (2 weeks prior), TOC, manganese, turbidity, TP and $NH_4^+:NO_3^-$ and directly opposite NO_3^--N, TN:TP and maximum wind gust.

The day of year was a good predictor for geosmin (14% deviance explained in GAM) and 2-MIB (25% deviance explained in GAM) events. It is not surprising there is seasonality in the growth of T&O producing cyanobacteria, but it would be more useful for T&O prediction to study the underlying environmental factors. Several environmental and weather parameters had strong seasonal relationships that could have influenced the geosmin and 2-MIB pattern. An overview of average annual reservoir patterns (2015-2023) was created to better visualise trends and explain potential relationships with biological parameters and the risk of T&O events (Figure 4.34). Figure 4.34 illustrates the strong seasonality and interdependence of these parameters. The highest risk of geosmin events was from March to August with maximum concentrations in May and June, whilst a high risk of 2-MIB events occurred from May to November with a peak in August and September (Figure 4.34). This tallies with data from a drinking water reservoir in North Carolina in the USA that had geosmin peaks (likely *Dolichospermum* sp.) in late spring to early summer and peaked earlier than 2-MIB (Paerl et al., 2022).
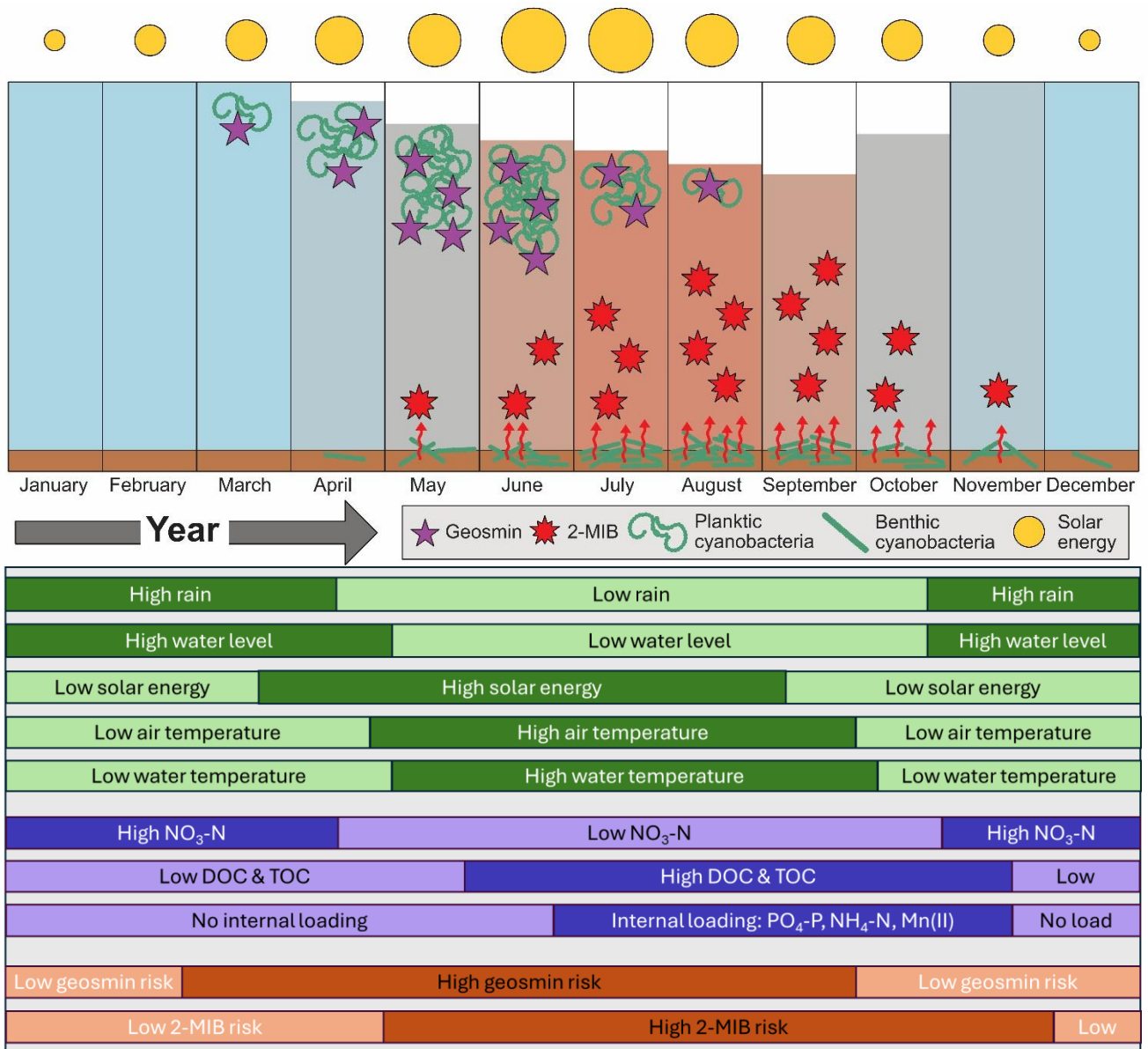
Figure 4.34. Top half of figure; Seasonal patterns in solar energy (size of yellow circles at the top or each month), water level (fill volume of monthly rectangle), water temperature (strength of red colour overlay of monthly rectangle) as well as general seasonal patterns of geosmin (purple star) produced by planktic cyanobacteria and 2-MIB (red double star) produced by benthic cyanobacteria, observed in Reservoir 1 from 2015 – 2023. Bottom half of figure; General seasonal trends in weather (green), nutrients (purple) and geosmin and 2-MIB risk (orange) in Reservoir 1 from 2015 – 2023.

Chapter 3 demonstrated that planktic *Aphanizomenon* spp. were likely geosmin producers in Reservoir 1, whilst benthic *Leptolyngbya* spp. and *Pseudanabaena* spp. were likely responsible for 2-MIB production. Planktic *Aphanizomenon* spp. generally thrive in high light conditions (>150 µmol photons m^{-2} s^{-1}), long day-length (>12 hours of daylight), medium temperature (14-32 °C), stratified conditions (buoyancy control) and low N:P ratios as they rely on phosphorus instead of nitrogen (N_2-fixation) (see

introduction, Section 4.1.1). Benthic *Leptolyngbya* spp. and *Pseudanabaena* spp. generally thrive in low-light conditions (27-40 μmol photons m^{-2} s^{-1}), medium temperatures (14-30 °C), stratified conditions (low resuspension), high nitrogen over phosphorus, they can access nutrients in the sediment and are tolerant to disturbances (see introduction, Section 4.1.1). Chapter 3 highlighted that laboratory conditions (33 μmol photons m^{-2} s^{-1}, 20 °C and 12h:12h light:dark cycle, with daily manual agitation) did not support the growth of *Aphanizomenon* spp., as it was dominant in the initial community, but mostly benthic *Leptolyngbya* spp. and *Pseudanabaena* spp. developed in the experiments. The seasonal pattern of earlier geosmin production could be related to increasing solar energy and day-length, which tallies with the environmental conditions required for planktic *Aphanizomenon* spp. growth. The growth of benthic 2-MIB producers *Leptolyngbya* spp. and *Pseudanabaena* spp. might be stimulated by increasing water temperatures and more stable weather (internal loading) later in the year. Moreover, it is likely that a planktic *Aphanizomenon* spp. bloom will increase turbidity in the reservoir which causes a shading effect that prevents benthic cyanobacteria from developing, as has been demonstrated in some eutrophic lakes (Poulíčková et al., 2008).

Solar energy, air temperature and water temperature follow a predictable pattern of lowest values in winter and highest in summer. Solar energy is directly related to the shortest and longest day, with the highest levels in June and July, whilst maximum air temperature and maximum water temperature (a shallow reservoir warms relatively quickly) are about a month later, in July and August. Reservoir 1 does not have any pumped inflows, so rainfall directly influences the water level and through runoff from the catchment. In recent years (since 2022) infrequent abstractions have caused a natural seasonal pattern of refilling from precipitation in autumn and evapotranspiration in summer.

NO_3^--N and carbon exhibited seasonal trends, associated with higher biological activity in the summer months. NO_3^--N decreased due to phytoplankton uptake, while DOC increased as carbon was cycled from atmospheric CO_2 to cellular carbon, which was subsequently decomposed (Parmar et al., 2011). NO_3^--N is restocked in autumn and winter when high rainfall results in runoff and groundwater recharge, which contains high levels of NO_3^--N (Kristensen et al., 2018).

The data implied that internal loading of PO_4^{3-}-P, NH_4^+-N and manganese (as well as other metal fractions) happened in summer and autumn in Reservoir 1. For internal loading to happen, a stable water column is required and higher water temperatures to cause elevated microbial decomposition which lead to anoxic conditions (Jeppesen et al., 2009). Under anoxic conditions, these compounds will be released and dispersed into the water column, which is enhanced by vertical mixing (Jones & Welch, 1990) (see introduction, Section 4.1.2 and Chapter 1, Section 1.2.2).

Tributary

The tributary introduces additional nutrients to the reservoir, since it passes through intensively farmed land. There are dairy farms upstream of that tributary and the slurry pit is visible from the reservoir site (personal observation). Elevated TOC, PO_4^{3-}-P, TP and occasionally NH_4^+-N concentrations in the tributary in spring to autumn are most likely related to rainfall-driven runoff from manure or slurry that has been spread on the land. One particularly high input of potentially slurry-related nutrients with >20 mg/L of TOC, 0.7 mg/L PO_4^{3-}-P, >1.3 mg/L TP and 4.5 mg/L of NH_4^+-N in April 2021 was not related to precipitation, so it might have been an accident. High TOC concentrations in the tributary coincided with high 2-MIB concentrations in the reservoir, which is potentially because TOC is the least available nutrient in slurry-related runoff and therefore more likely to be measured with a limited sampling frequency. Moreover, tributary inflows could potentially be a source of geosmin and 2-MIB in the reservoir, if soil runoff contained species of soil bacteria *Actinomycetes*, which can remain active in the water according to several studies (Franklin et al., 2023). However, the data showed that tributary geosmin and 2-MIB concentrations hardly ever exceeded the DCWW thresholds.

NO_3^--N in the tributary is high in winter and lower in summer, which follows the same pattern as the reservoir, with inputs of NO_3^--N from groundwater and runoff in winter, whilst uptake and denitrification happens in summer. The summer concentrations of NO_3^--N in the tributary were always around 1 mg N/L, when reservoir concentrations were <0.5 mg N/L. The tributary discharge is unknown so loading could not be estimated, but the data suggests that the tributary is probably a continuous NO_3^--N source throughout the summer. The lower concentrations measured at the reservoir abstraction point indicate there is active N cycling and show that nitrate is taken up prior to sampling, implying phytoplankton's high affinity for NO_3^--N.

4.4.2 Inter-annual trends in environment, geosmin and 2-MIB

GAM results showed that the year explained 54% of deviance in the 2-MIB pattern and 34% for geosmin. The following important factors seem to control inter-annual fluctuations in geosmin and 2-MIB: I) Water level changes, II) Changes in nutrient loading in tributary, III) Extreme weather events, and IV) Species behaviour. Figure 4.35 highlights some key events that happened between 2015 and 2023 and icons underneath each year indicate significant weather events that happened that year.
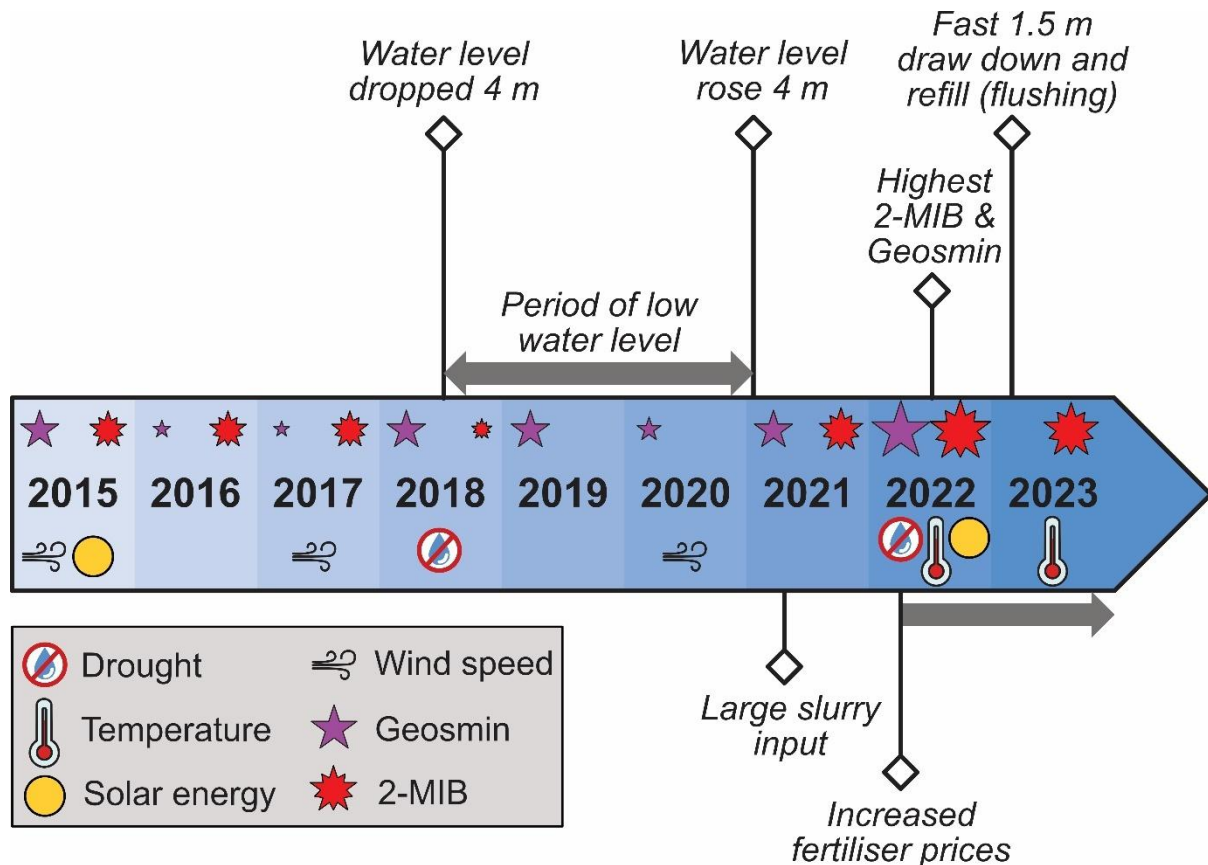
Figure 4.35. Inter-annual differences of geosmin and 2-MIB severity, the size of the geosmin and 2-MIB icon indicates the severity of the event that year. Important events and annual weather extremes are indicated with icons. Solar energy = annual sum of solar energy, temperature = annual average temperature.

Water level changes (I)

Water level patterns exhibited by natural inputs (rainfall, groundwater, tributaries) and outputs (evapotranspiration) can be overlain by human controls, for example abstraction. In Reservoir 1, abstraction for drinking water ceased in 2022, but did influence the system prior to this date (Figure 4.13). The data revealed there have been some major changes in water level management over the years which possibly influenced geosmin and 2-MIB. Water level data from 2017 to 2023 showed that in

spring 2018 the water level was drawn down 4 meters, to allow maintenance work on the spillway. The water level was low from 2018 until early 2021, but there were sharp increases and decreases in water level due to precipitation followed by draw-off. The years 2018 to 2020 had very minor or no 2-MIB events (Figure 4.35).

The habitat of benthic cyanobacteria dried up and the continuous abstraction to maintain a low water level caused sharp fluctuations in water level, and planktic phytoplankton shading, which could have prevented re-establishment of benthic cyanobacteria and subsequent 2-MIB events. Low reservoir water levels lead to large parts of the reservoir sediment drying out and occasionally rewetting, which is known to increase internal loading (PO_4^{3-}-P and NH_4^+-N), turbidity and sediment resuspension (Cantwell, 2021). Subsequent increases in eutrophication promote planktic phytoplankton blooms, which reduce light at the sediment and hinder benthic growth (Jähnichen et al., 2011). Reservoir 1 data showed evidence of the described processes as PO_4^{3-}-P, NO_3^--N and NH_4^+-N in the reservoir were often higher during these years and there were high levels of chlorophyll *a*, turbidity and geosmin, especially after the initial drop in water level in 2018. A reduced water level could have impacted biogeochemical processes like denitrification and microbial decomposition, by drastically altering the sediment and water conditions.

Ever since the low water level period from 2018 to 2021, geosmin and 2-MIB events have been higher than they were before. One contributing factor could be that, since 2021, Reservoir 1 has rarely been used for abstractions and without flushing, nutrients entering from the tributary accumulate, increasing eutrophication or becoming stored as legacy nutrients in the sediment. The effect of abstraction on nutrient flushing was visible in March 2023, when the drop in electrical conductivity and water temperature suggested that the abstraction flushed out some of these dissolved nutrients. This could be the reason there was no geosmin event that year and why 2-MIB was lower than 2022.

Changes in nutrient loading in tributary (II)

Increased severity of geosmin and 2-MIB events (higher concentrations, Figure 4.13) since 2021 could also partially be caused by increases in tributary nutrient loading since 2021. Tributary samples revealed that nutrient concentrations (NH_4^+-N, TOC, DOC, PO_4^{3-}-P, TP) have increased in recent years, starting in winter 2020 and

springtime 2021 when a substantial input of organic nutrients was measured, which might have been manure or slurry. Changes in nutrient loading in the tributary could have been caused by land-use changes or increased use of organic and urea-based nitrogen fertilizers (Glibert, 2020). Another possible reason is that fertilizers have increased in price in the last few years, in response to the Russian invasion in Ukraine which influenced gas prices and fertiliser production (Alexander et al., 2023). This might have forced farmers to use manure and slurry to fertilise the fields, which would have caused more organic nutrients to run off into the water courses.

Extreme weather events (III)

Extreme weather events will also play a role in the risk of geosmin and 2-MIB events in the reservoir. 2015 and 2022 had a high sum of solar energy, and both years had significant geosmin and 2-MIB events. 2018 was a drought year with a relatively large geosmin event but only a minor 2-MIB event because water levels dropped by four meters this year. By far the highest geosmin and 2-MIB concentrations happened in 2022; a drought year with high solar energy and a high average temperature, which also had increasing eutrophication from tributary inputs and reduced abstraction. The years 2022 and 2023 had higher average temperatures than the other years and these years had high concentrations of 2-MIB, which shows the effect that climate change might have on increasing geosmin and 2-MIB events.

Over the past 23 years cyanobacterial blooms have increased in severity in Lake Taihu (China), blooms start earlier and last longer, which seems to coincide with increasing temperature, sunshine hours, global radiation and decreasing wind speed (Zhang et al., 2012). 2-MIB production was high during a drought year, during which a low TN:TP was related to 2-MIB concentrations, and 2-MIB production was low in the following extremely wet year (Winston et al., 2014). Varying precipitation and temperature in lake Rotorua (New Zealand) were found to have interactive effects on physical, chemical and biological parameters, which led to contrasting cyanobacterial communities and microcystin concentrations (Wood et al., 2017). Higher NO_3^--N inputs occurred during a wet summer, whilst prolonged stratification and increased DRP and low DIN concentrations happened in a drier warmer summer (Wood et al., 2017).

Species behaviour (IV)

The 2-MIB event in winter 2022 did not go below the threshold and 2-MIB was found earlier in the year in 2023, which could be due to benthic overwintering of species

responsible for production and a quick re-establishment in the next year (Yao et al., 2022). A priority effect could take place if a large biomass of a species survives over winter, as the species with the largest initial inoculum has a competitive advantage over other species when the growing season starts (Van Gremberghe et al., 2009). In general, differences between the years could be caused by different species producing geosmin and 2-MIB, as different taxa have a range of production capacity per cell (Watson, 2003).

4.4.3 T&O prediction and lag-times (sensor data)

Full GAMs to predict 2-MIB and geosmin pattern had most of the data explained by the combined factor day of year and specific year. The specific year is not useful for future predictions, since it applies 'hindsight predicting'. Day of year can be used as a seasonality factor, to highlight that certain times of year have higher risk of geosmin and/or 2-MIB events, which can guide preventative water treatment or increased monitoring strategies. The day of the year was an important predictor in models for cyanobacterial growth (Perri et al., 2024) as well as geosmin production (Harris & Graham, 2017). The 2-MIB full model also included the significant terms NO_3^--N and maximum wind gust 2 weeks prior. Therefore, to develop a predictive model, the most significant parameters that can be regularly monitored should be explored. Chapter 2 demonstrated the optimum frequency of monitoring for some of the parameters used here, and this was further investigated by interrogating the effect of lag-times of easily measured parameters.

Lag-times between environmental parameters and geosmin and 2-MIB were first investigated using manual samples and weather data for the nine-year dataset, and secondly with two years of *in situ* sensor data. Nutrient lag-times in the manual data were investigated by calculating the rate of change between two manual samples, but the data was not frequent enough to show any good results.

2-MIB predictors

The most interesting significant GAM predictors for 2-MIB were summarized in Table 4.10 and organised by how many weeks prior to the manual 2-MIB result it was calculated.

Table 4.10. <u>Manual</u> and <u>weather</u> in "parameter to measure" column indicate GAMs that were performed on nine-year dataset, whilst <u>sensor</u> indicates GAMs that were performed on the one-year dataset. Strength of effect for manual and weather data (0-10% deviance explained = +, 10-20% deviance explained = ++, >20% deviance explained = +++), Strength of effect for sensor data (deviance explained; 0-30% = +, 30-40% = ++, >40% = +++). "Curve" in column effect on 2-MIB indicates that the pattern is not 100% positive or negative, but a curve.

Type of predictor	Process	Parameter to measure	Effect on 2-MIB	Strength of effect
Lagged predictors 2-3 weeks before	Stable weather	<u>Weather:</u> Maximum wind gust	Negative	++
		<u>Weather:</u> Minimum air temperature	Positive	+++
		<u>Sensor:</u> Maximum air temperature, Average water temperature	Positive	+++, +++
	Phytoplankton productivity	<u>Sensor:</u> Dissolved oxygen (range), pH (range), ORP (range)	Positive	+++, ++, +++
	Nitrate in reservoir	<u>Sensor:</u> Average NO_3^--N	Negative	+++
Lagged predictors 1 week before	Water level	<u>Sensor:</u> Maximum water depth	Negative	+++
Direct predictors	Mixing and internal loading in reservoir	<u>Manual:</u> TOC, Manganese	Positive	+++, +++
	Uncertain	<u>Sensor:</u> Maximum electrical conductivity	Positive	+
	Nitrate in reservoir	<u>Manual:</u> NO_3^--N, <u>Sensor:</u> NO_3^--N	Negative	+++, +++
	Phosphate in reservoir	<u>Manual:</u> PO_4^{3-}-P	*"Curve"*	++
	Tributary inflows	<u>Manual:</u> TOC, TP	Positive	+++, +++
		<u>Manual:</u> NO_3^--N	Negative	++
	Nutrient ratio in reservoir	<u>Manual:</u> TN:TP	*"Curve"*	+++
		<u>Manual:</u> NH_4^+:NO_3^-	*"Curve"*	++

A conceptual diagram was designed based on the findings in Table 4.10, with the hypothesised drivers of 2-MIB production and release, to highlight which parameters could act as predictors using *in situ* high-frequency sensor measurements (Figure 4.36).
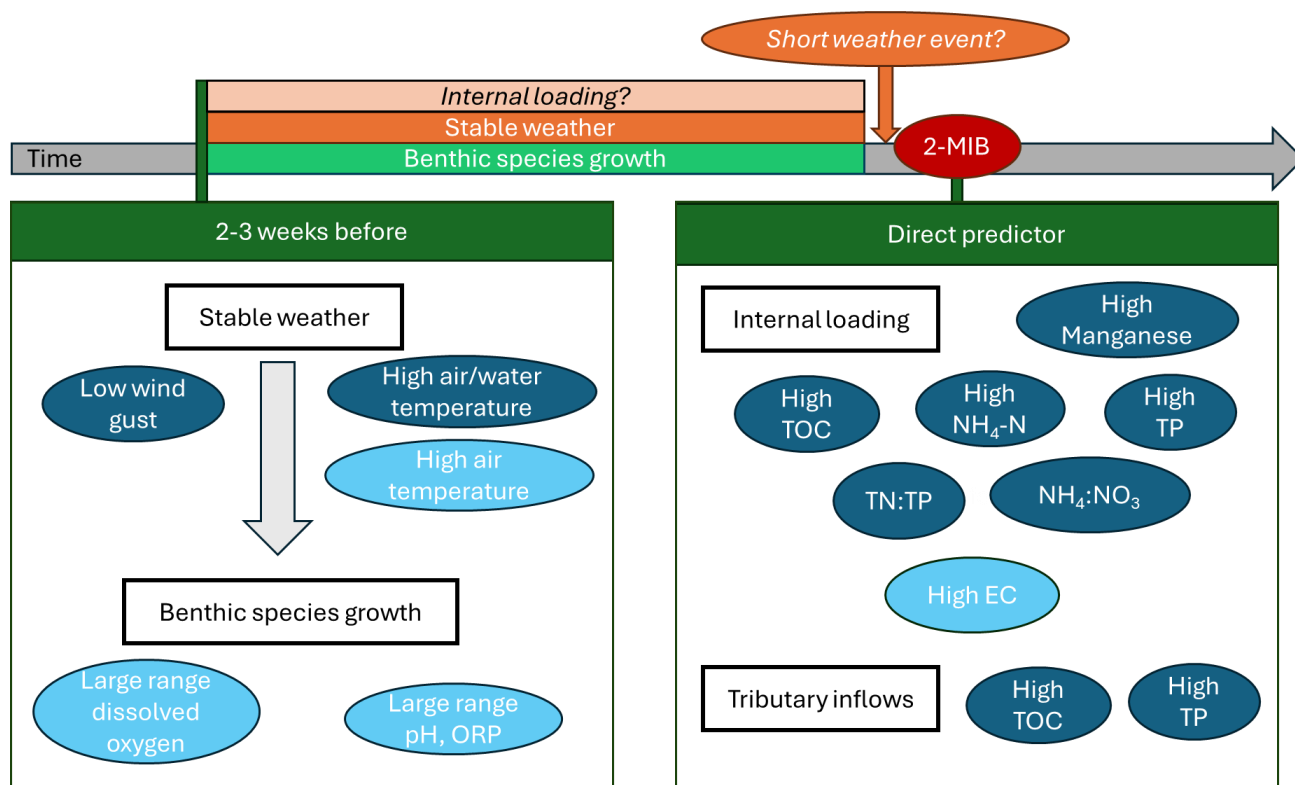
Figure 4.36. Parameters that can be measured as short-term predictors (expected seasonal variables have not been included). Light blue circles indicate sensor data (one-year dataset), and dark blue circles indicate manual or weather data (nine-year dataset). Italic text indicates hypothesized or uncertain processes.

Lagged predictors 2-3 weeks before

Periods of stable weather with low maximum wind gusts and high minimum (air/water) temperature were related to elevated 2-MIB concentrations 2-3 weeks later. Stable weather conditions (warm and dry) in the warmer months often leads to cyanobacterial growth in lakes, which can be linked to enhanced stratification (internal loading of PO_4^{3-} and NH_4^+, little resuspension) (Hecht et al., 2022; Huber et al., 2012). The overall increased productivity of the phytoplankton community in Reservoir 1 was detected 2-3 weeks before 2-MIB concentrations, with a large range in dissolved oxygen, pH and ORP (Table 4.10 and Figure 4.36). This could be related to benthic species growth as well, as the oxygen they produce rises and gets detected higher in the water column. The range of DO, pH and ORP was the best predictor, as the dataset included data from the whole year. Dissolved oxygen average is high in winter with small fluctuations (mixing and less decaying organic matter), whilst in summer, large oxygen fluctuations occur due to periods of high phytoplankton productivity (oxygen production) followed by decay (oxygen consumption), and day-night cycles (photosynthesis-respiration), but not necessarily a higher average than wintertime. Dissolved oxygen sensors were

employed to detect periods of growth and decay of cyanobacteria in a reservoir, to inform about T&O risk (Chen et al., 2019) and Williams et al. (2000) used high-frequency oxygen sensors to determine diurnal oxygen dynamics and productivity of three UK rivers.

Periods of stable weather could imply vertical stratification and internal loading, which seemed to stimulate growth of 2-MIB producing benthic cyanobacteria in Reservoir 1 and increased 2-MIB concentrations 2-3 weeks later (Figure 4.36). 2-MIB production by planktic filamentous *Pseudanabaena* sp. increased with a lower water level, moderate stratification, temperatures between 14 and 23 °C and higher total nitrogen concentrations, with a sudden disappearance after a hydrodynamic disturbance (Xiao et al., 2024). 2-MIB production in a Chinese reservoir was related to short-term hydrometeorological processes, especially in the shallow zones; concentrations increased when surface water temperature exceeded 12 °C and decreased after rainfall (Wu et al., 2022). Paerl et al. (2022) found a negative correlation between discharge and 2-MIB concentrations in a drinking water reservoir in North Carolina (USA), which suggests the 2-MIB producers preferred stratified, stable conditions, and were inhibited by increased discharge. A south-central USA reservoir had high 2-MIB concentrations during a drought year (with low TN:TP ratio influencing concentrations), whilst an extremely wet year had low 2-MIB production (Winston et al., 2014). Temperature was correlated with low TN:TP ratios and stratified conditions in a Chinese reservoir, which resulted in high 2-MIB concentrations most likely produced by *Leptolyngbya* sp. (Yue et al., 2024).

Direct predictors - Internal loading

Once benthic cyanobacteria have established, they could produce 2-MIB that accumulates together with compounds released from the sediment (internal loading: PO_4^{3-}, NH_4^+, manganese etc.) and compounds from decomposing organic matter (DOC and TOC). Decomposition of phytoplankton and cyanobacterial cells releases DOC (Peterson et al., 1995) and bioavailable phosphorus (Feng et al., 2018), which highlights that phytoplankton biomass is important in nutrient recycling. When there is a short period of unstable weather with rain, wind and lower temperatures, it will likely cause cyanobacteria die-off and additional release of 2-MIB, which will get mixed into the water column together with TOC, manganese, TP and NH_4^+-N. Hecht et al. (2022) highlighted the importance of large summer runoff events which reduced water

temperature, sunlight penetration and stratification and caused cyanobacterial bloom die-off. High 2-MIB concentrations in sediments of a Chinese reservoir suggests the likely source is the 2-MIB release from settled planktic and benthic cyanobacteria on the sediment, potentially through decomposition (Huang et al., 2018b). Similarly, Ma et al. (2013) found that decomposition of cyanobacterial blooms caused anoxic conditions and increased T&O concentrations in the water. In laboratory experiments 2-MIB release by benthic *Pseudanabaena yagii* happened slowly during growth but rapidly increased when the temperature dropped (Jeong et al., 2021).

Due to low temporal resolution of the dataset, the exact weather events responsible for 2-MIB release could not be detected, and it might be more complicated than hypothesised here. Nevertheless, manual sampling revealed that compounds related to internal loading were detected in the water at the same time as 2-MIB concentrations and a good direct predictor of 2-MIB, which indicated they could have been mixed from the sediment together. Internal loading is a process that will happen regardless of 2-MIB production, so it won't always result in increases of 2-MIB, but it is a good indicator of compounds from the sediment reaching the water column. It is also possible that TOC adsorbs 2-MIB at the sediment and acts as a vector that transports 2-MIB into the water column (Figure 4.37), which Rider et al. (2024) hypothesised for geosmin.
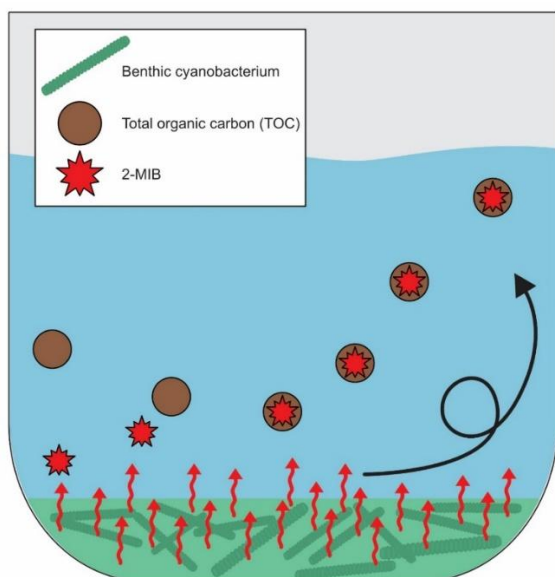
Figure 4.37. TOC might adsorb 2-MIB at the sediment and act as a vector to transport it into the water column.

Direct predictors - Tributary inflows

The hypothesised short period of unstable weather could also cause runoff from agricultural land and increase the nutrients in tributary inflows. TOC and TP in the tributary explained some of the 2-MIB patterns (Table 4.10) and could indicate moments of high organic (slurry) inflows into the reservoir, which could contain more bioavailable PO_4^{3-}-P and NH_4^+-N as well as organic P and N fractions. Tributary flow rate is unknown, so actual impacts are uncertain. Moreover, the manual data is not measured frequently enough to see the extend of the tributary inputs, because most of these compounds will flush through the system and are taken up so quickly (Aubriot & Bonilla, 2012), that a measurement every two weeks won't accurately capture these processes. It is also possible that tributary inflows happen at the same time as internal release of 2-MIB but are not a causation, as nutrient in tributary inflows are related to rainfall and unstable weather, which could have been the real reason for 2-MIB release in the reservoir.

Sensors

Sensor measurements of electrical conductivity (EC) could potentially detect the moment when 2-MIB gets released at a higher frequency, as it had a positive effect on 2-MIB in the GAM results (Table 4.10). High-frequency EC sensors can be used to detect pollution in rivers (Halliday et al., 2014), but in reservoirs it may be harder to pinpoint agricultural pollution. The EC data indicated a relationship with reservoir volume and suggested a potential dilution effect from reservoir flushing. Data plotting (Figure 4.29) revealed that high-frequency turbidity data had the potential to give insight in reservoir processes that could lead to 2-MIB release, whether that is internal loading or tributary inputs. Turbidity can indeed capture reservoir mixing (Johnston et al., 2024). Both electrical conductivity and turbidity high-frequency measurements can be difficult to interpret, as they measure the overall effect of processes that happen in the reservoir but nothing specifically, so they require a good understanding of the specific reservoir.

Seasonality - NO_3^--N, PO_4^{3-}-P, TN:TP and NH_4^+:NO_3^-

NO_3^--N had a strong seasonal pattern in the tributary and the reservoir (see Section 4.4.1), which happens regardless of 2-MIB production, but can indicate the reservoir goes from P-limited state to N-limited state (Howard, 2020). Low NO_3^--N in summer coincides with high 2-MIB, but both factors have a strong seasonal pattern and it's

possible that there is no direct causation. Howard (2020) found that low nitrate was the most important predictor of peak 2-MIB concentrations in Eagle Creek Reservoir (USA) and mentioned that close monitoring of nitrate levels could serve as an early warning for 2-MIB events. Several studies found that growth of *Pseudanabaena* spp. and *Leptolyngbya* spp. related to high concentrations of nitrogen, relative to phosphorus (Gao et al., 2018; Liu & Vyverman, 2015; Xiao et al., 2024). Based on the data from Reservoir 1, it is hypothesised that during the summertime when the NO_3^--N is low, periods of high NH_4^+-N and PO_4^{3-}-P release indicate internal loading events or tributary inputs, which might be related to high levels of 2-MIB. PO_4^{3-}-P had a complicated pattern throughout the year, because it was generally higher in winter possibly due to increased mixing, runoff and less biological uptake, whilst lower concentrations in summer might have been caused by phytoplankton uptake. However, internal loading of PO_4^{3-}-P can cause a spike in summer and autumn (Jeppesen et al., 2009), which won't be accurately defined with manual sampling as PO_4^{3-}-P is taken up by bloom-forming cyanobacteria within 15-25 minutes and increases productivity rates instantly (Aubriot & Bonilla, 2012).

The result from GAMs and the PCA for nutrient ratios TN:TP and NH_4^+:NO_3^- in the reservoir as well as the nutrient concentrations separately, indicate the importance of low NO_3^--N which means that any increases in PO_4^{3-}-P and NH_4^+-N have an impact. A lower TN:TP relates to higher 2-MIB concentrations, which means that more PO_4^{3-}-P is present during a period of low TN (NO_3^--N mainly). The GAM revealed a slight increase of the effect on 2-MIB again at very high TN:TP ratios, which was also found by Hooper (2023a), who attributed this pattern to a transition from nitrogen fixing cyanobacteria at low TN:TP to non-nitrogen fixing cyanobacteria at high TN:TP. The TN:TP ratio has often been related to increased growth of cyanobacterial dominance but does not necessarily have to be related to T&O production (Howard, 2020). The TN:TP ratio was found to be less important in determining cyanobacterial growth in eutrophic lakes, where light limitation is a more important driver (Huisman et al., 2004; Paerl & Fulton, 2006). Similarly, a higher NH_4^+:NO_3^- ratio related to a higher 2-MIB concentration, which shows that increases in NH_4^+-N during a period of low NO_3^--N were indicative of high 2-MIB. High NH_4^+:NO_3^- ratios have often been related to T&O events (Harris et al., 2016; Perkins et al., 2019), potentially from agricultural inputs or internal loading, but nutrient spiking experiments in Chapter 3 revealed that the

$NH_4^+:NO_3^-$ ratio was not a key driver of 2-MIB production by benthic *Pseudanabaena* spp. and *Leptolyngbya* spp. Findings from Chapter 3 suggest that total concentration of nitrogen (as NO_3^--N, because green algae dominated in $NH4^+$-N treatments) was most important for these species to develop relative abundance, and they required some phosphorus for growth, but a P-limitation resulted in higher 2-MIB release through cell death or active release.

Manual sampling of nutrient processes is limited by the low temporal frequency of measurement, which makes it difficult to pinpoint the effect of NH_4^+-N and PO_4^{3-}-P coming from tributary inputs versus internal loading, and the subsequent lag-time between nutrient input and 2-MIB concentration. NH_4^+-N and PO_4^{3-}-P tributary inputs during periods of high 2-MIB levels in the reservoir are most likely unrelated, because experiments in Chapter 3 indicated a potential lag-time of 4 to 8 days between nutrient spike and cyanobacterial growth, with subsequent 2-MIB production.

Geosmin predictors

The most interesting significant predictors of geosmin from GAM results were summarised in Table 4.11.

Table 4.11. <u>Manual</u> and <u>weather</u> in "parameter to measure" column indicate GAMs that were performed on nine-year dataset, whilst <u>sensor</u> indicates GAMs that were performed on the one-year dataset. Strength of effect for manual data (0-10% deviance explained = +, 10-20% deviance explained = ++, >20% deviance explained = +++), Strength of effect for sensor data (deviance explained; 0-30% = +, 30-40% = ++, >40% = +++). "Curve" in column effect on geosmin indicates that the pattern is not 100% positive or negative, but a curve.

Type of predictor	Process	Parameter to measure	Effect on geosmin	Strength of effect
Lagged predictors 3 weeks before	Stable weather	<u>Sensor:</u> Maximum air temperature, weekly range of water temperature	Positive	++, ++
	Phytoplankton productivity	<u>Sensor:</u> Dissolved oxygen (weekly range)	Positive	+++
Lagged predictors 1-2 weeks before	Stable weather	<u>Weather:</u> Sum of precipitation	Negative	++
		<u>Weather:</u> Sum of solar energy	Positive	++
	Phytoplankton productivity	<u>Sensor:</u> pH (weekly range), ORP (maximum range)	Positive	+, +++
	Nitrate in reservoir	<u>Sensor:</u> Maximum NO_3^--N	Negative	+++
	Water level	<u>Sensor:</u> Maximum water depth	Negative	++

Direct predictors	*Uncertain*	<u>Sensor:</u> Minimum electrical conductivity	Positive	+
	Uncertain	<u>Manual:</u> pH, turbidity	"Curve"	++, +
	Nitrate in reservoir	<u>Manual:</u> NO_3^--N, <u>Sensor:</u> NO_3^--N	"Curve"	+, +++
	Phosphate in reservoir	<u>Manual:</u> PO_4^{3-}-P	Negative	+
	Tributary inflows	<u>Manual:</u> NO_3^--N	Negative	+
	Nutrient ratio in reservoir	<u>Manual:</u> TN:TP	"Curve"	++

A conceptual diagram was designed based on the findings from Table 4.11, with hypothesized processes related to geosmin production and release, highlighting parameters that could predict geosmin with *in situ* high-frequency sensors (Figure 4.38).
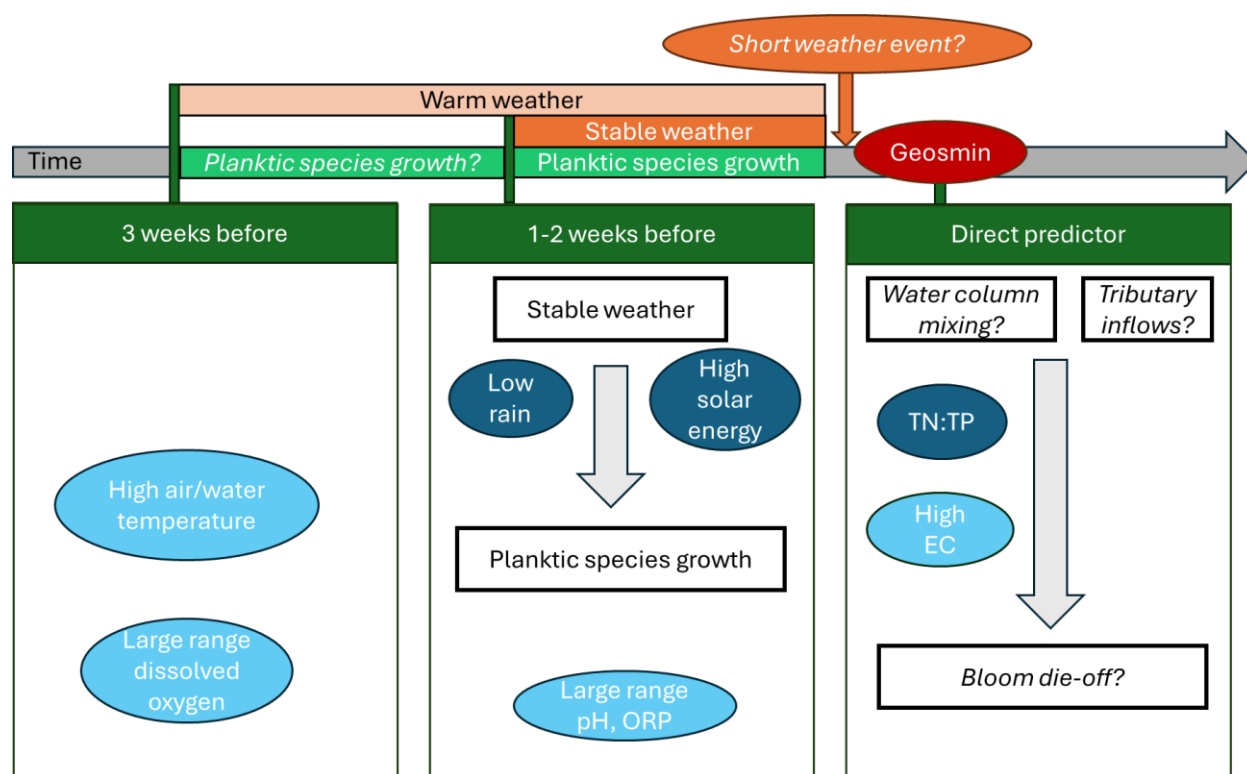
Figure 4.38. Parameters that can be measured as short-term predictors (complex or seasonal variables have not been included). Light blue circles indicate sensor data (one-year dataset), and dark blue circles indicate manual or weather data (nine-year dataset). Italic text indicates hypothesized or uncertain processes.

Lagged predictors 3 weeks before

GAM results from the sensor data (2022 to 2023 data) indicated that geosmin was related to high air and water temperature 3 weeks before, as well as a large range in

dissolved oxygen. A high overall phytoplankton productivity during warmer weather can be related to increased growth of planktic cyanobacteria that are capable of geosmin production (see Section 4.4.3).

Lagged predictors 1-2 weeks before

The manual data from 2015 to 2023 revealed a strong effect 1-2 weeks before of various weather parameters related to stable weather, like a low amount of rainfall and high solar energy for a period of time (Table 4.11). Planktic geosmin producing *Aphanizomenon* spp. require high light availability (>150 µmol photons m^{-2} s^{-1}) for optimal growth (see introduction, Section 4.1.1 and Üveges et al. (2012)), which explains why solar energy is a driver of geosmin concentrations. The range in dissolved oxygen, pH and ORP 1-2 weeks before geosmin concentrations in the water, is indicative of planktic cyanobacteria growth, similar to 2-MIB. In eutrophic lakes, a stable water column and warm weather has been related to buoyant species of cyanobacteria, like *Aphanizomenon* spp. and *Microcystis* spp., whilst a turbulent water column was related to sinking species of cyanobacteria (Huisman et al., 2004; Reynolds, 2006). Descy et al. (2016) also found that low average wind velocity, high epilimnion temperature and solar energy increased *Aphanizomenon* spp. dominance.

Direct predictors

In general, geosmin was harder to predict than 2-MIB. Elevated concentrations of geosmin could be related to a change in weather causing cyanobacterial bloom collapse, as was found by Ma et al. (2013), but data analysis and GAM results did not provide any clear evidence for geosmin release. Only a high electrical conductivity from sensor data was found to relate to geosmin, which was probably related to higher water temperatures. Planktic cyanobacteria blooms have more explosive growth than benthic species (Vadeboncoeur et al., 2021; Yamamoto & Nakahara, 2009) and there are many reasons why they might collapse (Harris et al., 2024). This could explain why there is not a straightforward predictor for the geosmin release related to mass cell death. Moreover, bloom collapse and geosmin release through cell lysis was probably not necessary for geosmin to be measured in the manual water samples. As the planktic cyanobacteria cells float in the water column, they are physically collected in the samples (Chapter 3, Section 3.3), which means that intracellular and extracellular geosmin are measured collectively in the laboratory.

Additionally, different species could have been responsible for geosmin production over the years, with potentially different environmental drivers and variable lag-times. Another possible reason is that geosmin could be more quickly biologically degraded by bacteria than 2-MIB, as Li et al. (2012) found that 2-MIB had a larger turnover time than geosmin.

Complex patterns - pH, turbidity

Manual pH data showed that, over the years, pH was a good indicator of geosmin (Table 4.11), but a little complex. Increasing levels of pH are a result of photosynthetic activity by the community, which generally results in the highest pH in summer when most activity happens (Wetzel, 2001). There is a seasonal succession of species, and geosmin production generally happens relatively early in the year when pH is still increasing (DOY plot showed geosmin peak around day 150, in May). Therefore, the effect of pH could just be a seasonal effect, or it is an indication of the environmental conditions that favours planktic producers.

Manual turbidity data had a negative relationship with geosmin concentrations (Table 4.11). It is possible that turbidity in the reservoir indicated periods of unstable weather, runoff and water column mixing (internal loading), which would have resulted in collapse of the planktic cyanobacterial bloom by shading from particles and diluting extracellular geosmin compounds. Periods of increased turbidity might mark the lack of geosmin events and will probably not be useful as a direct predictor, unless high-frequency sensor data can provide direct insight when certain mixing processes occur.

Seasonal patterns - NO_3^--N, PO_4^{3-}-P and TN:TP

The strong seasonal NO_3^--N pattern in the reservoir and tributary seems to be related to geosmin concentrations but is less indicative than for 2-MIB (Table 4.11). There might be a threshold of NO_3^--N, below which *Aphanizomenon* spp. can dominate. Findings by Descy et al. (2016) suggest that *Aphanizomenon* spp. development was driven by dissolved inorganic nitrogen (DIN) levels going below 1.5 mg/L in summer and TP exceeding 0.030 mg/L. Miller et al. (2013) discovered that blooms of *Aphanizomenon* spp. were primarily related to higher NO_3^--N concentrations. In Reservoir 1, PO_4^{3-}-P was negatively correlated with geosmin, which is most likely related to the seasonal effect of elevated PO_4^{3-}-P concentrations in winter and lower concentrations over summer (see seasonality section for 2-MIB). The PO_4^{3-}-P patterns

are not well captured due to limited sampling, so the direct effect on planktic cyanobacteria and geosmin production is hard to predict.

Interestingly, GAM results for TN:TP revealed that geosmin production was related to higher levels of TP compared to TN, even though higher PO_4^{3-}-P had a negative effect on geosmin. Low TN:TP was related to high geosmin concentrations and increasing ratios had a negative effect on geosmin, whilst the effect from a ratio of > 400 was also related to higher geosmin concentration. The pattern is the same as for 2-MIB, which could also indicate the shift from non-nitrogen fixing cyanobacteria (high TN:TP) to nitrogen-fixing cyanobacteria (low TN:TP). The expected source of geosmin in this reservoir, the cyanobacterial genus *Aphanizomenon*, has nitrogen-fixing capabilities and therefore would be most influenced by increases in phosphorus, hence the relationship with low TN:TP. However, other processes in this reservoir might be related to increases in PO_4^{3-}-P, like internal loading or runoff caused by a weather event, which could impact planktic cyanobacteria blooms and subsequent geosmin concentrations. Moreover, *Aphanizomenon* might be able to take up organic phosphorus fractions, as Raven (2010) found that *Aphanizomenon ovalisporum* released the cyanotoxin cylindrospermopsin to induce the production of alkaline phosphatases in other phytoplankton, to free up inorganic phosphate.

4.4.4 NH_4^+-N and NO_3^--N sensor technology

The sensitivity of the NO_3^--N and NH_4^+-N ISE sensors was not sufficient to capture the changes in biogeochemistry required to predict T&O in this reservoir. Blaen et al. (2016) mentioned that ISE are often subject to significant drift, ionic interferences and can have issues with systematic errors when the ionic activity in the calibration solution is notably different to the ionic activity of the field sample. This might explain why there was a mismatch between how the sensor performed in the field compared to the laboratory. The precision of ISEs was specified by the manufacturer as ±10% or ±2 mg N/L (but they suggest an LoD of 0.01 mgN/L). However, Bende-Michl and Hairsine (2010) concluded that in practise, no field-deployable ISE exists that can detect nutrient concentrations lower than ±0.1 mg/L, and that they required frequent re-calibration. These results highlight potential drawbacks of using ISE sensors for nutrients at environmentally relevant concentrations.

The lab-on-chip wet chemical NO_3^--N analyser from Clearwater sensors had a precision of 0.0003 mg N/L (with a 0.0004 mg N/L LoD). The Clearwater sensors NO_3^--N analyser had good agreement with the laboratory data but suffered from rapid blocking of the filter due to biofouling when the water system was very productive in July-August 2022. Appropriate maintenance schedules need to be designed depending on the site and time of year, which could prevent this from happening.

4.4.5 Sensor recommendations

Based on the findings of this study, it is advisable to focus monitoring efforts on the identified processes that might drive cyanobacterial growth in the reservoir and cause risk of geosmin and 2-MIB events. Several parameters can be measured at a high frequency with *in situ* sensors or analysers to better quantify these processes, the most important parameters being:

- Tributary nutrient inflows: NO_3^--N, NH_4^+-N (ISE or lab-on-chip) or PO_4^{3-}-P (lab-on-chip) or indicators of pollution like turbidity & EC.
- Tributary discharge to calculate loading.
- Reservoir conditions related to phytoplankton growth: water temperature & dissolved oxygen
- Reservoir vertical stability: water temperature at different depths
- Meteorological factors: Solar energy, air temperature, precipitation, wind speed
- Forecasts of meteorological factors: Solar energy, air temperature, precipitation, wind speed

Tributary inflows are an important source of nutrients (nitrogen, phosphorus and carbon) to the reservoir, as potentially slurry-related runoff occurred in this catchment. Current manual sampling does not capture the impact on the reservoir and the exact timing when they occur, which is crucial to understanding the biogeochemical processes and subsequently predicting T&O events. Nutrients can be measured with ISE sensors or lab-on-chip analysers (see Section 4.1.5 and Section 4.4.4) and simultaneous tributary discharge measurements can provide insight on nutrient loading. Tributary inflows of NO_3^--N could be captured with lab-on-chip analysers or less sensitive ISEs, as concentrations are higher (>1 mg/L all summer) than the reservoir (<0.5 mg/L). High-frequency measurements of NH_4^+-N and PO_4^{3-}-P in the tributary could provide valuable insights into the response of cyanobacterial growth

and subsequent T&O production to slurry-related runoff events, including any lag-time. Alternatively, indicators of agricultural runoff can be measured in the tributary using turbidity or EC sensors. The turbidity sensor in this reservoir deployment detected periods of high turbidity when there was precipitation and higher wind speeds, which indicate it could capture reservoir mixing or runoff from the tributary, similar to Johnston et al. (2024). The comparison with laboratory turbidity data showed that only *in situ* sensor measurements at a high frequency could capture these events. High-frequency EC data from this reservoir visualised a potential dilution effect from reservoir flushing, but it did not have a clear pattern related to unstable weather, or T&O events. EC has the potential to detect agricultural pollution, as it can successfully detect urban pollution in rivers (Halliday et al., 2014), but it might be better to target the tributary inflows. Additionally, this runoff could potentially be detected with other commercially available sensors like tryptophan-like fluorescence, fDOM, CDOM (Chapter 1) etc. but these won't provide the same insight in nutrient loading.

Measurements of dissolved oxygen, pH and ORP were found to serve a similar purpose in this deployment. Monitoring ORP is valuable primarily for vertical depth profiles, as dissolved oxygen sensors provide equivalent information in oxygen-rich surface waters. Additionally, dissolved oxygen sensor technology requires the least maintenance, making it the preferred choice over pH and ORP sensors. Dissolved oxygen data in the reservoir captured the response of the phytoplankton community to a weather event with precipitation, higher wind gusts and a drop in solar energy. The rapid decline in DO concentration related to phytoplankton cell death could highlight an important period with risk of geosmin and 2-MIB release into the reservoir. High-frequency measurements of DO in the source water of a water treatment plant in Beijing is used as early warning for T&O risk (Chen et al., 2019).

Stable weather conditions, water column stratification and mixing were identified as key processes for geosmin and 2-MIB production, release and/or distribution in Reservoir 1. Vertical profiles of water temperature can indicate periods of stratification and mixing (Yang et al., 2019). Meteorological data was arguably the most important driver of geosmin and 2-MIB events in Reservoir 1, particularly solar energy, air temperature, precipitation and wind speed. Similarly, high-frequency vertical water quality profiles and meteorological data identified that cyanobacterial blooms were related to calm and sunny weather which followed periods of strong wind which had

released nutrients (Yang et al., 2019). Air temperature, precipitation and wind speed were meteorological factors determined by Johnston et al. (2024) to help monitor harmful algal blooms. Forecasts of these meteorological factors can serve as an early warning for periods of optimal conditions for phytoplankton growth, followed by an unstable weather event which causes bloom die-off and release of geosmin and 2-MIB.

Other processes that are interesting, but less important for predicting geosmin and 2-MIB events are:

- Reservoir chlorophyll *a* and phycocyanin
- Reservoir water level
- Reservoir light availability
- Reservoir NO_3^--N
- Reservoir internal loading: NH_4^+-N, PO_4^{3-}-P (lab-on-chip, requires high sensitivity)

Chlorophyll *a* laboratory data in Reservoir 1 sometimes coincided with geosmin and 2-MIB events, but chlorophyll *a* peaked more regularly throughout the year which was independent of geosmin and 2-MIB. This makes sense, as all *planktic* species of phytoplankton (diatoms, chlorophytes and cyanobacteria, which were present in Reservoir 1 water samples, as identified in Chapter 3) will contribute to elevated chlorophyll *a* concentrations, regardless of their T&O producing capabilities, and this won't detect *benthic* producers. Moreover, when geosmin and 2-MIB compounds are released extracellularly, they get dispersed away from the producers. Nevertheless, chlorophyll *a* data has been used extensively (alongside phycocyanin) in lakes and reservoirs to estimate algal blooms or cyanobacterial blooms (with phycocyanin) (Carey et al., 2021; Painter et al., 2023; Yang et al., 2019), so it could be used as an indicator of potential T&O risk. The chlorophyll *a* sensor that was deployed in the reservoir highlighted the difference between laboratory and *in situ* chlorophyll *a* measurement methods, as the sensor detects the pigment without extraction, which must be considered when interpreting data. Moreover, studies suggest that chlorophyll *a* sensor measurements at night-time or in deployments where they are sheltered from light, are the best indicator of actual phytoplankton biomass, because there is no light effect on the pigment (Lucius et al., 2020). A high-frequency chlorophyll *a* sensor and

a more targeted sensor for cyanobacteria that measures phycocyanin, can be a useful indicator of high phytoplankton and cyanobacteria productivity. This can be used to prompt an investigation into the species and subsequently provide a possible warning for T&O events (Painter et al., 2023).

Water level could be a useful indicator, but mainly for benthic 2-MIB producers and during periods when the reservoir was used for abstraction and had larger fluctuations in water level. Inter-annual trends in Reservoir 1 highlighted the effect of changes in abstraction management on benthic 2-MIB producers, with larger fluctuations due to abstraction limiting their growth. Water level can be used to manage benthic cyanobacteria (Su et al., 2017), and in certain reservoirs water level was a good indicator of T&O compounds (Xiao et al., 2024).

Light availability in the water column could be measured with sensors for photosynthetically active radiation (PAR) at different depths, to calculate the extinction coefficient. This can be a useful parameter for cyanobacteria prediction (Descy et al., 2016) but would especially be useful if the optimal light conditions were known for the geosmin and 2-MIB producing cyanobacteria.

NO_3^--N seasonal pattern needs to be measured but doesn't have to be done with sensors as the pattern is predictable. High-frequency sensor measurements of NO_3^--N in Reservoir 1 are not required, because manual data seemed to predict the general seasonal trend. Internal loading should ideally be investigated, because it directly related with 2-MIB concentrations in Reservoir 1, but more standard parameters can be measured to indicate this (water temperature at different depths). Lab-on-chip nutrient analysers (NH_4^+-N and PO_4^{3-}-P) should be able to detect internal loading, but this won't provide an early warning of 2-MIB as it is 'hindsight' predicting. Moreover, high-frequency measurements of NH_4^+-N in the reservoir could indicate internal sediment release and, consequently, the release of 2-MIB. However, the deployed ISE was not sensitive enough to detect this. If NH_4^+-N and/or PO_4^{3-}-P internal loading is identified as a key driver of T&O, lab-on-chip analysers for NH_4^+-N and/or PO_4^{3-}-P could be deployed to capture this process, potentially closer to the sediment.

4.4.6 Parameter frequency and data elements for T&O prediction

The optimal measurement frequency of each parameter largely depends on the catchment, the parameter and what element of the sensor data you require to identify

the process of interest (Chapter 2). Many water quality parameters have seasonal and diurnal cycles, and some parameters have a quick response to extreme weather events which influence the data variability (Chapter 2). For example, a turbidity sensor in the tributary to pinpoint the exact timing of runoff events might require a high frequency of hourly to four-hourly (Chapter 2). On the other hand, the seasonal nitrate cycle in the reservoir is predictable and daily or weekly measurements, with a sensor or manually, can be enough to capture this pattern and calculate the load (Lloyd et al., 2014).

It is an important consideration what element of sensor data is most useful as a predictor of geosmin and 2-MIB events and related processes. For parameter with seasonality and diurnal cycles, a daily average, maximum, minimum or range can be more informative. For example, the daily range in dissolved oxygen was a better predictor of phytoplankton productivity than for example the daily average. When using a dataset that includes all seasons, it is important to be aware of seasonality and diurnal cycles in the parameters, to pick the best element of the sensor data to identify the process of interest.

It is essential that geosmin and 2-MIB in the reservoir are measured with manual sampling as frequently as possible during the high-risk period, to best use the high temporal resolution of sensor measurements and detailed weather data for geosmin and 2-MIB prediction. This will result in a better understanding of the processes that causes the initial production and release of geosmin and 2-MIB early in the season.

4.5 Conclusions

Nine years of Reservoir 1 data from manual sampling was merged with calculated weather data (lagged 1,2,3 or 4 weeks) to investigate the general environmental patterns in the reservoir and how they related to geosmin and 2-MIB events. Sensors were deployed from June 2022 until July 2023 and data was merged with the manual sampling data for this period to identify short-term indicators of geosmin and 2-MIB events that could be used as an early warning. The following research questions were addressed in this study:

1) What are drivers of geosmin and 2-MIB production in a drinking water reservoir?

Day of year and year were the best predictors for geosmin and 2-MIB, which emphasizes that certain environmental parameters might only be related because they also have a strong seasonal pattern. Geosmin and 2-MIB have a different seasonal pattern; geosmin was detected earlier in the year than 2-MIB. Geosmin in the water column was likely produced by planktic cyanobacteria and the data suggested that geosmin events were driven by a period of stable weather with high solar energy and low rainfall, which could have caused internal NH_4^+-N and PO_4^{3-}-P loading. Geosmin was directly related to low TN:TP ratios which could indicate elevated TP concentrations from cyanobacterial biomass or internal PO_4^{3-}-P loading in summer when there is low TN. Cyanobacteria bloom collapse is probably not necessary to detect geosmin with manual sampling, as geosmin inside planktic cyanobacteria cells will be physically collected. Planktic cyanobacteria can result in shading of the benthic environment and delay growth of benthic producers. 2-MIB was likely produced by benthic cyanobacteria and the production was driven by a period of stable weather with low wind gusts and high air/water temperatures, causing internal NH_4^+-N and PO_4^{3-}-P loading. Most likely, a short weather event can cause runoff from the catchment and vertical mixing which might lead to cyanobacteria cell death. Simultaneously, mixing will bring all the released 2-MIB and other compounds from the sediment to the surface, where it can be detected with manual sampling. There was also a yearly difference in geosmin and 2-MIB severity, most likely impacted by a combination of: I) water level changes, II) changes in tributary nutrient concentrations, III) extreme weather events and IV) species behaviour.

2) What is the lag-time between environmental drivers and geosmin and 2-MIB detection?

Geosmin had a 1-2 week lag-time between environmental drivers and geosmin peak, whilst 2-MIB had a 2-3 week lag-time. The lag-time could be different for different species and different events. The lag-time for 2-MIB could be longer due to the added time of extracellular release and mixing to the water surface, or differences in growth speed of benthic cyanobacteria compared to planktic species.

3) Can sensor data in the reservoir be used to predict geosmin and 2-MIB events?

Geosmin was harder to predict than 2-MIB, possibly due to more explosive growth and quicker biological degradation. Sensors can identify certain processes that relate to

high geosmin and 2-MIB risk in Reservoir 1. Both geosmin and 2-MIB were related to periods of stable conditions in the reservoir. 2-MIB concentrations in the water column seemed to relate to vertical mixing and release of benthic compounds, whilst geosmin is already present in the water column and does not depend on vertical mixing. The following processes were recommended as most important to measure with sensors to provide early warning for geosmin and 2-MIB events: A) Tributary nutrient runoff: nutrient sensors (NH_4^+-N, PO_4^{3-}-P, NO_3^--N), turbidity, EC and discharge; B) Reservoir conditions related to phytoplankton growth: water temperature and dissolved oxygen; C) Reservoir vertical stability: water temperature at different depths; D) Meteorological parameters: precipitation, wind speed, solar energy, air temperature; E) Forecasts of meteorological parameters: precipitation, wind speed, solar energy, air temperature.

Knowledge from Chapter 2 on optimal design of a high-frequency sensor network should be combined with knowledge of species-specific growth factors (Chapter 3) and catchment-specific processes that could identify T&O events (this chapter), to create the best early-warning predictors for T&O for Reservoir 1. High-frequency nutrient sensors have still not been effectively used to predict phytoplankton blooms in lakes, but this study highlights the potential of using them in T&O prediction, as long as the appropriate sensor technology is selected, and maintenance schedules are designed properly.

Chapter 5: General discussion

This is the first study that has used a comprehensive assessment of laboratory and field studies to identify biogeochemical processes that are important for freshwater monitoring and highlight considerations regarding high-frequency sensor networks. High-frequency sensors and molecular measurement techniques were utilised to identify key biogeochemical processes that can be measured with sensors, which drive taste and odour (T&O) events in a drinking water reservoir. Findings from this thesis inform stakeholders on how to establish and utilise an effective sensor monitoring network. It also recommends how high-frequency sensors in the catchment can be used to predict and manage the presence of operationally significant concentrations of T&O in the drinking water treatment. Hence, this study is of importance to river and reservoir management.

In the UK and worldwide, drinking water reservoirs are under threat of increasing eutrophication due to agricultural activity and effects of climate change, which will likely increase the risk of cyanobacterial blooms (Cottingham et al., 2015; Paerl & Paul, 2012). T&O compounds are just an example of problematic compounds produced by cyanobacteria in drinking water reservoirs, but an increasing prevalence of cyanobacteria has much wider implications for water quality and can have substantial health impacts. Cyanobacteria species that can produce geosmin and 2-MIB can often also produce cyanotoxins, so it possibly has wider health impacts if these species start to dominate more in drinking water reservoirs. It is therefore critical that these reservoirs and the surrounding catchments are managed appropriately to prevent pollutants in the reservoir that can cause issues at the drinking water treatment and affect customers. Additionally, it is more sustainable if catchments and reservoirs are managed appropriately, because less chemical dosing and additional carbon is required to treat the water for drinking water production.

To guide management decisions, it is important to understand site-specific meteorological and biogeochemical processes that drive T&O events. Biological mechanisms behind cyanobacterial T&O production remain unclear and are likely to be species-specific, which emphasizes that drivers of T&O events should be investigated on a case-by-case basis. Indicators of meteorological and

biogeochemical processes need to be monitored at an environmentally relevant temporal frequency to provide sufficient evidence to predict T&O events. Manual sampling in the reservoir or treatment plant inlet is most often at a weekly to monthly cadence with additional days before laboratory analyses report results, which is insufficient to provide insight in the important drivers and cannot easily be used as an early warning. Deploying high-frequency sensors or analysers in a drinking water reservoir can provide a real-time indication of meteorological and biogeochemical processes, which can potentially enable T&O early warning and prediction. Careful consideration is required to identify the optimal sensor suite and temporal frequency of measurement, which in turn requires a good understanding of catchment processes and the site-specific drivers of T&O events.

The summarised findings from each chapter will be used to answer the overarching hypotheses, and the interaction between the chapters will be discussed linking into the wider of the area of research. Lastly, the implications of the findings will be explained with a focus on the drinking water industry, and future directions of research will be recommended.

5.1 Hypothesis 1

Hypothesis 1: Differing sampling frequencies of water quality sensors can impact interpretations of biogeochemical processes: Accepted.

The systematic assessment of high resolution hydrochemical data from six UK catchments in Chapter 2 revealed that a minimum four-hourly measurement frequency was required to capture the necessary variation. The monitoring frequency influenced the calculated median and range of a parameter, but this was parameter and catchment specific. The most important factors that influence data variability determined the impact of reduced measurement frequencies on the calculated median and range. A reduced measurement frequency impacted the median for parameters with a strong intra-daily variation like temperature and dissolved oxygen, whilst it impacted the range for parameters with a strong rainfall-response such as turbidity. Manual sampling at specific times of the day would not accurately capture the full data variability of parameters with significant intra-daily variation. Parameters with a predictable pattern or seasonal cycle, like NO_3^--N, could probably be measured at lower frequencies without impacting the data interpretation. A low measurement

frequency can potentially still capture the values around the median for more variable parameters, but it will most likely underestimate extreme values and similarly any extreme values it does measure will affect the interpretation of the data. An optimal monitoring design should consider the monitoring purpose and select the required sampling frequency based on the required accuracy of the result, for example a median or model calculation, which should be assessed for each catchment and parameter separately. A period of pre-monitoring sensor optimisation where data is captured at high resolution is recommended to understand catchment-specific responses. Moreover, the sensor uncertainty needs to be evaluated beforehand to determine if the sensors can accurately measure data fluctuations of interests beyond the uncertainty bounds of the measurement technique.

High-frequency hydrochemical data from sensors and analysers can be used to understand water quality and biogeochemical processes in freshwaters. This chapter highlighted the effect of monitoring frequency and the measurement time on interpretation of these processes, which are catchment and parameter specific. The meteorological and biogeochemical processes that drive T&O need to be explored first, to determine the correct sensor suite that will enable T&O prediction.

5.2 Hypothesis 2

Hypothesis 2: Nutrient concentrations and ratios are important factors in predicting geosmin and 2-MIB events: Accepted.

Chapter 3 explored the results of a laboratory microcosm study which used a natural phytoplankton community from a Welsh drinking water reservoir (Reservoir 1) to investigate the effect of varying nutrient concentrations and ratios, $NH_4^+:NO_3^-$ and DIN:SRP, on 2-MIB production. The data revealed that a higher overall community productivity (the total carbon produced) related to higher 2-MIB concentrations. Moreover, gene copies of the 2-MIB synthase gene and relative abundance of cyanobacteria *Leptolyngbya* spp. and *Pseudanabaena* spp. correlated well with 2-MIB concentrations, demonstrating they are the most likely producers of 2-MIB in this community.

These findings highlighted the need to identify factors that promote optimal growth of these species. Optimal growth parameters for geosmin and 2-MIB producers are often studied on laboratory cultures of the same species, but this prevents the competition

for resources within the phytoplankton community from having an impact. The microcosm experiment in this study was more realistic because it tested the response of a natural phytoplankton community, but it was more difficult to interpret the specific effect that nutrients and ratios had on the 2-MIB producers individually.

Despite variability in the replicates which made some interpretation challenging, the experiments suggested that total available nitrogen concentration was more important for 2-MIB production than the ratio of $NH_4^+:NO_3^-$, total phosphorus and DIN:SRP ratio. Elevated concentrations of NH_4^+-N with high PO_4^{3-}-P and low NO_3^--N caused rapid growth ('boom-bust' dynamics) of mostly green algae. Elevated NO_3^--N concentrations with low NH_4^+-N, in both PO_4^{3-}-P treatments, resulted in a delayed response in productivity (4 days later than high NH_4^+-N treatments) but overall balanced growth with more time for benthic cyanobacteria to develop. Total dissolved nitrogen (TDN) concentrations of around 2.5 mg N/L seemed to be required for the development of benthic 2-MIB producers, if sufficient total dissolved phosphorus (TDP) was available. Phosphorus did not have a clear effect on 2-MIB production, because also low PO_4^{3-}-P treatments had high 2-MIB production if they had enough total nitrogen. In this experiment, the DIN:SRP ratio was less important for 2-MIB concentrations than total phosphorus concentrations. Productivity in treatments with high NO_3^--N and high PO_4^{3-}-P was still increasing at the end of the experiment, whilst productivity in treatments with high NO_3^--N and low PO_4^{3-}-P plateaued, but both produced high 2-MIB concentrations. This finding suggested that benthic cyanobacteria required a critical PO_4^{3-}-P concentration (TDP between 0.2 and 0.5 mg P/L) to establish dominance in the community, before P depletion might have triggered 2-MIB release through active or passive mechanisms. Diatoms dominated in the treatment with low nutrients whilst green algae dominated in the high nutrient treatments. Highest cyanobacteria biomass was present in the treatment with high NO_3^--N, low NH_4^+-N and high PO_4^{3-}-P, followed by the treatment with high NO_3^--N, low NH_4^+-N and low PO_4^{3-}-P.

There was a difference in lag-times between initial nutrient concentrations and benthic cyanobacterial growth and detection of productivity. The lag-time between the nutrient spike and community productivity was four days with NH_4^+ as N source due to rapid 'boom-bust' growth, while NO_3^--N as N source resulted in slower balanced growth, with increasing productivity after eight days but lasting longer. The different response times for NH_4^+-N and NO_3^--N suggest that their relative availability to phytoplankton will

influence how NH_4^+:NO_3^- fluctuate over time. Therefore, the complexities in community response to different N sources needs to be considered when determining the drivers of T&O production in a natural aquatic environment.

The findings from the microcosm experiment emphasize that nutrient concentrations (particularly NO_3^--N and PO_4^{3-}-P) as well as ratios are important factors that result in high abundance and productivity of benthic cyanobacteria *Leptolyngbya* spp. and *Pseudanabaena* spp., which was related to 2-MIB concentrations. The findings of this experiment are specific for these benthic 2-MIB producing cyanobacteria species at that time of the year (the phytoplankton community in the inoculum). There are many different species of cyanobacteria that can produce geosmin and 2-MIB, planktic and benthic, which have different species traits and, therefore, require different growth stimuli. In a natural phytoplankton community with constant competition for resources, species traits determine which species can co-occur and have a higher relative abundance. There is a general lack of understanding related to benthic species (Catherine et al., 2013; Gaget et al., 2020), but this study is the first (to the best of my knowledge) that demonstrated drivers of 2-MIB production in a natural phytoplankton community.

In Chapter 4, nine years of manual nutrient data from Reservoir 1 and a major tributary were analysed to identify drivers of geosmin and 2-MIB events. Unfortunately, concentrations of nutrients, geosmin and 2-MIB were measured at insufficient temporal frequency to fully understand how nutrients influenced geosmin and 2-MIB events, which is often the case in field studies (Dzialowski et al., 2009). Infrequent water column sampling will probably not capture cycling processes (release-uptake) or external inputs (in this case, the tributary runoff) of highly bioavailable nutrient fractions like PO_4^{3-}-P and NH_4^+-N, because they have fast biological uptake rates. In the natural environment the nutrient ratios will be variable due to continuous uptake and release (nutrient cycling) with the possibility of external inputs from tributary flows. The concentrations and ratios in reservoirs have seasonal patterns and can have diurnal patterns at certain times of the year.

Interestingly, the statistical analysis revealed that 2-MIB was directly related to high levels of TOC and TP in the tributary, as well as high levels of TOC, TP, NH_4^+-N, manganese and low ratios of NH_4^+:NO_3^- and TN:TP in the reservoir. These findings

support the conclusions from other studies (Harris et al., 2016; Perkins et al., 2019). However, a detailed investigation of the data led to the conclusion that seasonality (day of the year), year, weather and water column stability could be more important drivers of geosmin and 2-MIB production and release. NO_3^--N, for example, had a strong seasonal pattern (high in winter, low in summer) regardless of geosmin and 2-MIB events, but because geosmin and 2-MIB had a strong seasonal pattern themselves (low in winter, high in summer), they correlated with each other. In summer, calm weather can result in a stable water column and increased internal loading of PO_4^{3-}-P, NH_4^+-N, manganese and other metals near the sediment, which will mix throughout the water column when a short weather event disrupts water column stability. In the data analysis, it was clear these periods of internal loading overlapped with high 2-MIB concentrations, which in this reservoir most likely originates at the sediment and might also be mixed to the water surface during mixing with PO_4^{3-}-P, NH_4^+-N, etc. from internal loading. However, this process also happened independently in years without 2-MIB events. Geosmin was generally harder to correlate with nutrient concentrations and ratios and didn't relate to internal loading as a direct predictor, but low TN:TP concentrations had the same relationship as with 2-MIB, which could indicate the same seasonal effect.

The findings from Chapter 3 and 4 posed questions with regards to nutrient ratios and concentrations as T&O drivers. It is possible that the nutrient ratios TN:TP, DIN:SRP and NH_4^+:NO_3^- are driven by seasonality and water column stability, and just happen to coincide with 2-MIB and geosmin events. During the summer, nitrogen concentrations are low and when there is a stable water column it causes internal loading of PO_4^{3-}-P and NH_4^+-N from sediments, which might play a role in 2-MIB and geosmin production by cyanobacteria. When a short weather event (such as a storm) disrupts the water column, it may cause mixing of sediment compounds (including 2-MIB) and trigger cyanobacterial cell lysis, which further increases geosmin and 2-MIB concentrations in the water column and also results in a low TN:TP, low DIN:SRP and high NH_4^+:NO_3^- ratio. The TN:TP ratio is possibly not the best indicator of T&O events, because the TP measurements capture intracellular particulate organic phosphorus within the phytoplankton biomass, which might cause TP to correlate with planktic phytoplankton blooms.

5.3 Hypothesis 3

Hypothesis 3: Sensor data can provide an early warning for geosmin and 2-MIB events: Accepted.

High-frequency water quality sensors were deployed in 2022-2023 in Reservoir 1 to identify short-term physical and biogeochemical indicators of geosmin and 2-MIB events that could be used as for early warning in the future. Moreover, insights from the nine years of laboratory data for this reservoir was used to pinpoint meteorological and biogeochemical processes that could potentially be measured as T&O predictors with *in situ* sensors.

Nine years of manual data revealed that geosmin and 2-MIB have a different seasonal pattern; geosmin probably originates from planktic cyanobacteria and was detected earlier in the year than 2-MIB, which was likely produced by benthic cyanobacteria. 2-MIB was related to low wind gusts and high water temperature, whilst geosmin patterns were better explained by high solar energy and low rainfall. The different species will require different growth stimuli, but it is also important to note that geosmin is more readily detected in the water column than 2-MIB, as geosmin inside planktic cells will be sampled, whilst 2-MIB first needs to be dispersed from the sediment before it can be measured. Yearly differences highlighted that fluctuating water levels can prevent benthic 2-MIB production, that tributary nutrient concentrations increased over the years (potentially causing increased eutrophication and legacy P storage) and that years with heatwaves have a high geosmin and 2-MIB risk (potentially more internal loading).

Geosmin was harder to predict than 2-MIB within the dataset, which might be due to more explosive growth of planktic cyanobacteria and quicker compound degradation in the water by specialist bacteria. The data analysis in Chapter 4 suggested that high-frequency sensors in this reservoir can be used to identify certain processes that relate to high geosmin and 2-MIB risk, but there is still insufficient knowledge about T&O specific drivers. To predict geosmin and 2-MIB events, it is important to investigate the underlying physical and chemical processes that could drive these events. However, biological processes that will ultimately determine growth of T&O producing cyanobacteria and T&O production, are much harder to predict. Nevertheless, using *in situ* sensors to predict T&O risk periods (not actual concentrations) is already

valuable for drinking water management, as preventative measures can be put in place.

It is important to consider the lag-times between environmental drivers and geosmin and 2-MIB production, release and detection. An added complication is that the exact process of production and release of 2-MIB and geosmin is still uncertain; it might occur actively or passively, or both, and is potentially species-specific. Chapter 3 microcosm experiments demonstrated that NH_4^+-N as a source of N resulted in a quicker response in productivity, whilst NO_3^--N as a N source had a four-day delayed growth response. In the reservoir, stable weather conditions and vertical stability had a 1-2 week lag-time with geosmin concentrations, whilst the same conditions had a 2-3 week lag-time for 2-MIB concentrations. This can be related to slower growth of benthic species compared to planktic species or the added time required for benthic compounds to be dispersed into the water column through vertical mixing. The manual reservoir data couldn't shed light on the influence of short-term changes in nutrients and the response in 2-MIB and geosmin, as laboratory measurements had limited data frequency. Unfortunately, the high-frequency sensor data did not provide this insight either, due to insensitivity of NH_4^+-N and NO_3^--N ion-selective electrodes and the NO_3^--N pattern being uninformative due to its predictable seasonal pattern. Lag-time variability based on cyanobacterial species or nutrient fraction highlights the need for field studies that measure the nutrient fractions and geosmin and 2-MIB concentrations at a high frequency, at least daily.

5.4 Recommendations for *in situ* sensor deployment strategies to provide early warning of T&O events

In Reservoir 1, a number of factors could be measured *in situ* to help predict geosmin and 2-MIB events. The most important processes that can be measured with the mentioned high-frequency sensors are:

A) Tributary nutrient inflows

These could be measured at high precision with lab-on-chip analysers for NO_3^--N and PO_4^{3-}-P, but the higher concentrations present in the tributary may mean that less sensitive ISEs (available for NH_4^+-N and NO_3^--N) could be used. Proxy measurements such as EC or turbidity might also be useful since they would indicate significant transport of pollutants into the reservoir.

B) Tributary discharge to calculate loading

As well as sensors measuring when the tributary flows are active, the volume of ingress measured by a simple discharge monitor would also be valuable.

C) Reservoir optimal growth conditions for phytoplankton

This could be measured with water temperature & dissolved oxygen sensors.

D) Reservoir vertical stability

A simple temperature sensor string, deployed at different depths, could be a useful indicator of stability and note when vertical mixing (and potential sediment remobilisation) is occurring.

E) Meteorological factors

These factors are already measured by local weather stations; solar energy, air temperature, precipitation and wind speed are the most critical. Moreover, forecasts of meteorological factors with solar energy, air temperature, precipitation and wind speed, can help predict T&O events further in the future and provide an early-warning system.

Some other interesting factors can be measured with sensors, but they are less critical: 1) Reservoir chlorophyll *a* and phycocyanin, fluorescence sensors can only indicate planktic cyanobacteria; 2) Reservoir water level with sensors; 3) Reservoir light availability, with PAR sensor; 4) Reservoir NO_3^--N (lab-on-chip), but manual samples can capture predictable seasonal pattern; 5) Reservoir internal loading, exact concentrations can be measured with NH_4^+-N and/or PO_4^{3-}-P lab-on-chip analysers but can be estimated from reservoir vertical stability.

When designing a high-frequency monitoring network to predict T&O events, the findings from Chapter 2 need to be considered to ensure the most effective design for the monitoring purpose. A thorough data analysis like Chapter 4, or a high-frequency test period can be done to help determine the optimal sensor suite and measurement time and frequency. It is therefore recommended that adaptive drinking water treatment management is considered in the context of the data available. Meteorological and physical parameters are most frequently and easily measured, thus better use of these data should be made, in conjunction with specific biological

and biogeochemical parameters in at-risk locations. Field observations and laboratory studies like the microcosm experiments in Chapter 3 could provide critical insight in the cyanobacterial species responsible for T&O production. This knowledge can add to the design of the optimal sensor suite, as certain parameters might be more appropriate and specific locations in the reservoir or tributaries should be targeted for monitoring.

It needs to be carefully considered whether high-frequency sensors are indeed necessary. Sensors and analysers are expensive, they have limitations and require good maintenance practices to provide useful data. A cost-benefit analysis needs to be done which can help determine the priorities for a certain monitoring purpose. For a drinking water company, it could be best to outsource the sensor monitoring to a third-party, as this assures good maintenance practices and network design by sensor engineering experts. In that case, it is essential to ensure the monitoring network is designed for the right purpose and the data can feed directly into existing workflows.

5.5 Reflections & conclusions

The problem for most of the current T&O predicting models, is that many parameters are laboratory-based and are not measured frequently enough to accurately capture lag-times. It is also important to realise that the water quality parameters that can be used to predict geosmin and 2-MIB are site-specific and depend on the species of cyanobacteria which produces the compounds. Ideally, laboratory studies with isolated cultures from a reservoir should be conducted to provide insights in the exact growth stimuli (temperature, light availability, nutrients) and their potential geosmin and 2-MIB production. However, laboratory simulations cannot capture competition or co-occurrence with other bacteria or phytoplankton, which might be more important for geosmin and 2-MIB production than is presently acknowledged (Cook et al., 2020; Hooper, 2023a; Louati et al., 2015; Reinl et al., 2022).

Seasonal occurrence and yearly severity of geosmin and 2-MIB has been investigated in this specific Welsh drinking water reservoir (Reservoir 1) and differences were explained based on species-traits and important changes in the catchment. These novel insights are critical to improve reservoir management to prevent T&O events and can highlight potential environmental drivers in other reservoirs. Moreover, meteorological and biogeochemical processes were critically assessed to determine

which elements were most important for geosmin and 2-MIB prediction, and which high-frequency sensors would be recommended. A high-frequency sensor network could effectively be combined with information from weather forecasts, as weather factors played an important role in driving these events, to potentially predict geosmin and 2-MIB in the near-future (a few weeks before) as an early warning. To aid management decisions at the water treatment plant, it is essential to create a decision diagram with critical threshold values based on high-frequency sensor data from the reservoir, weather data and forecasts (using catchment knowledge) and potentially include measured (laboratory) parameters at WTW (Kibuye et al., 2021).

Moreover, investments should be made to better survey and identify benthic cyanobacteria species. *In situ* methods are available, such as fluoroprobes that use *in situ* fluorescence or ex situ molecular techniques (Gaget et al., 2022; Gaget et al., 2020). Manual water sampling in this reservoir did not detect benthic species or even their DNA, even during periods of high 2-MIB production. Investing in molecular techniques will provide more insight in the biological processes (Almuhtaram et al., 2021a; Devi et al., 2021) and can potentially highlight important species or co-occurring groups within the phytoplankton or bacterial community that indicate geosmin and 2-MIB events.

It is important to consider biogeochemical processes that occur at the sediment in catchment management decisions. Benthic cyanobacterial species can access sediment nutrients and legacy nutrients stored in sediment layers, especially in highly eutrophic lakes, can cause high productivity of benthic and planktic cyanobacteria even when nutrients in the water column are reduced (Paerl et al., 2020). Internal loading of PO_4^{3-}-P and NH_4^+-N during periods of vertical stability can stimulate cyanobacterial growth, which emphasizes that catchment nutrient reductions alone may not be sufficient to tackle cyanobacterial blooms since the ecological legacy of past nutrient inputs still plays a role in present day biogeochemical cycling (Barcala et al., 2021). Climate change will most likely increase cyanobacterial growth in freshwaters (Paerl & Huisman, 2009), and it is important to appropriately monitor and manage drinking water sources to protect them for the future.

References

Abeynayaka, H. D. L., Asaeda, T., & Rashid, M. H. (2018). Effects of elevated pressure on Pseudanabaena galeata Böcher in varying light and dark environments. *Environmental Science and Pollution Research, 25*(21), 21224-21232. https://doi.org/10.1007/s11356-018-2218-5

Achyuthan, K. E., Harper, J. C., Manginell, R. P., & Moorman, M. W. (2017). Volatile metabolites emission by in vivo microalgae—an overlooked opportunity? *Metabolites, 7*(3). https://doi.org/10.3390/metabo7030039

Adam, B., Klawonn, I., Svedén, J. B., Bergkvist, J., Nahar, N., Walve, J., Littmann, S., Whitehouse, M. J., Lavik, G., Kuypers, M. M. M., & Ploug, H. (2016). N2 -fixation, ammonium release and N-transfer to the microbial and classical food web within a plankton community. *ISME Journal, 10*(2), 450-459. https://doi.org/10.1038/ismej.2015.126

Alexander, P., Arneth, A., Henry, R., Maire, J., Rabin, S., & Rounsevell, M. D. A. (2023). High energy and fertilizer prices are more damaging than food export curtailment from Ukraine and Russia for food prices, health and the environment. *Nature Food, 4*(1), 84-95. https://doi.org/10.1038/s43016-022-00659-9

Alghanmi, H. A., Alkam, F. M., & Al-Taee, M. M. (2018). Effect of light and temperature on new cyanobacteria producers for geosmin and 2-methylisoborneol. *Journal of Applied Phycology, 30*(1), 319-328. https://doi.org/10.1007/s10811-017-1233-0

Allan, J. D., & Castillo, M. M. (2007). *Stream Ecology: Structure and function of running waters*. Springer Netherlands.

Almuhtaram, H., Kibuye, F. A., Ajjampur, S., Glover, C. M., Hofmann, R., Gaget, V., Owen, C., Wert, E. C., & Zamyadi, A. (2021a). State of knowledge on early warning tools for cyanobacteria detection. *Ecological Indicators, 133*, 108442. https://doi.org/10.1016/j.ecolind.2021.108442

Almuhtaram, H., Zamyadi, A., & Hofmann, R. (2021b). Machine learning for anomaly detection in cyanobacterial fluorescence signals. *Water Research, 197*, 117073. https://doi.org/10.1016/j.watres.2021.117073

Andersen, I. M., Williamson, T. J., González, M. J., & Vanni, M. J. (2020). Nitrate, ammonium, and phosphorus drive seasonal nutrient limitation of chlorophytes, cyanobacteria, and diatoms in a hyper-eutrophic reservoir. *Limnology and Oceanography, 65*(5), 962-978. https://doi.org/10.1002/lno.11363

Asquith, E., Evans, C., Dunstan, R. H., Geary, P., & Cole, B. (2018). Distribution, abundance and activity of geosmin and 2-methylisoborneol-producing Streptomyces in drinking water reservoirs. *Water Research, 145*, 30-38. https://doi.org/10.1016/j.watres.2018.08.014

Aubriot, L., & Bonilla, S. (2012). Rapid regulation of phosphate uptake in freshwater cyanobacterial blooms. *Aquatic Microbial Ecology, 67*(3), 251-263. https://doi.org/10.3354/ame01596

Auffret, M., Pilote, A., Proulx, É., Proulx, D., Vandenberg, G., & Villemur, R. (2011). Establishment of a real-time PCR method for quantification of geosmin-producing Streptomyces spp. in recirculating aquaculture systems. *Water Research, 45*(20), 6753-6762. https://doi.org/10.1016/j.watres.2011.10.020

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics.* In (Version R package version 2.3) https://CRAN.R-project.org/package=gridExtra

Bai, X., Zhang, T., Wang, C., Zong, D., Li, H., & Yang, Z. (2017). Occurrence and distribution of taste and odor compounds in subtropical water supply reservoirs and their fates in water treatment plants. *Environmental Science and Pollution Research, 24*(3), 2904-2913. https://doi.org/10.1007/s11356-016-7966-5

Bar-Yosef, Y., Sukenik, A., Hadas, O., Viner-Mozzini, Y., & Kaplan, A. (2010). Enslavement in the water body by toxic aphanizomenon ovalisporum, inducing alkaline phosphatase in

phytoplanktons. *Current Biology, 20*(17), 1557-1561.
https://doi.org/10.1016/j.cub.2010.07.032

Barcala, V., Rozemeijer, J., Osté, L., Van der Grift, B., Gerner, L., & Behrends, T. (2021). Processes controlling the flux of legacy phosphorus to surface waters at the farm scale. *Environmental Research Letters, 16*(1), 015003. https://doi.org/10.1088/1748-9326/abcdd4

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., & Hocking, T. (2024). *data.table: Extension of `data.frame`*. In (Version R package version 1.15.4) CRAN. https://CRAN.R-project.org/package=data.table

Bauer, F., Wolfschlaeger, I., Geist, J., Fastner, J., Schmalz, C. W., & Raeder, U. (2023). Occurrence, Distribution and Toxins of Benthic Cyanobacteria in German Lakes. *Toxics, 11*(8). https://doi.org/10.3390/toxics11080643

Beaton, A. D., Cardwell, C. L., Thomas, R. S., Sieben, V. J., Legiret, F.-E., Waugh, E. M., Statham, P. J., Mowlem, M. C., & Morgan, H. (2012). Lab-on-Chip Measurement of Nitrate and Nitrite for In Situ Analysis of Natural Waters. *Environmental science & technology, 46*(17), 9548-9556. https://doi.org/10.1021/es300419u

Beaton, A. D., Wadham, J. L., Hawkings, J., Bagshaw, E. A., Lamarche-Gagnon, G., Mowlem, M. C., & Tranter, M. (2017). High-Resolution in Situ Measurement of Nitrate in Runoff from the Greenland Ice Sheet. *Environmental Science and Technology, 51*(21), 12518-12527. https://doi.org/10.1021/acs.est.7b03121

Behr, M., Serchi, T., Cocco, E., Guignard, C., Sergeant, K., Renaut, J., & Evers, D. (2014). Description of the mechanisms underlying geosmin production in Penicillium expansum using proteomics. *Journal of Proteomics, 96*, 13-28. https://doi.org/10.1016/j.jprot.2013.10.034

Beisner, B. E., & Peres-Neto, P. R. (2009). Seasonal trophic dynamics affect zooplankton community variability. *Freshwater Biology, 54*(11), 2351-2363. https://doi.org/10.1111/j.1365-2427.2009.02266.x

Bende-Michl, U., & Hairsine, P. B. (2010). A systematic approach to choosing an automated nutrient analyser for river monitoring. *Journal of Environmental Monitoring, 12*(1), 127-134. https://doi.org/10.1039/b910156j

Bertone, E., Burford, M. A., & Hamilton, D. P. (2018). Fluorescence probes for real-time remote cyanobacteria monitoring: A review of challenges and opportunities. *Water Research, 141*, 152-162. https://doi.org/10.1016/j.watres.2018.05.001

Bertone, E., & O'Halloran, K. (2016). Analysis and Modelling of Taste and Odour Events in a Shallow Subtropical Reservoir. *Environments, 3*(4). https://doi.org/10.3390/environments3030022

Bieroza, M., Acharya, S., Benisch, J., ter Borg, R. N., Hallberg, L., Negri, C., Pruitt, A., Pucher, M., Saavedra, F., Staniszewska, K., van't Veen, S. G. M., Vincent, A., Winter, C., Basu, N. B., Jarvie, H. P., & Kirchner, J. W. (2023). Advances in Catchment Science, Hydrochemistry, and Aquatic Ecology Enabled by High-Frequency Water Quality Measurements. *Environmental science & technology, 57*(12), 4701-4719. https://doi.org/10.1021/acs.est.2c07798

Billica, J. A., Oropeza, J., & Elmund, G. K. (2010). Monitoring to determine geosmin sources and concentrations in a Northern Colorado reservoir. Proceedings of the National Water Quality Monitoring Council 7th National Monitoring Conference, City of Fort Collins Utilities, Fort Collins, Colorado,

Blaen, P. J., Khamis, K., Lloyd, C., Comer-Warner, S., Ciocca, F., Thomas, R. M., MacKenzie, A. R., & Krause, S. (2017). High-frequency monitoring of catchment nutrient exports reveals highly variable storm event responses and dynamic source zone activation. *Journal of Geophysical Research: Biogeosciences, 122*(9), 2265-2281. https://doi.org/10.1002/2017JG003904

Blaen, P. J., Khamis, K., Lloyd, C. E. M., Bradley, C., Hannah, D., & Krause, S. (2016). Real-time monitoring of nutrients and dissolved organic matter in rivers: Capturing event dynamics, technological opportunities and future directions. *Science of the Total Environment, 569-570*, 647-660. https://doi.org/10.1016/j.scitotenv.2016.06.116

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., Cope, E. K., Da Silva, R., Diener, C., Dorrestein, P. C., Douglas, G. M., Durall, D. M., Duvallet, C., Edwardson, C. F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J. M., Gibbons, S. M., Gibson, D. L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G. A., Janssen, S., Jarmusch, A. K., Jiang, L., Kaehler, B. D., Kang, K. B., Keefe, C. R., Keim, P., Kelley, S. T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M. G. I., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B. D., McDonald, D., McIver, L. J., Melnik, A. V., Metcalf, J. L., Morgan, S. C., Morton, J. T., Naimey, A. T., Navas-Molina, J. A., Nothias, L. F., Orchanian, S. B., Pearson, T., Peoples, S. L., Petras, D., Preuss, M. L., Pruesse, E., Rasmussen, L. B., Rivers, A., Robeson, M. S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S. J., Spear, J. R., Swafford, A. D., Thompson, L. R., Torres, P. J., Trinh, P., Tripathi, A., Turnbaugh, P. J., Ul-Hasan, S., van der Hooft, J. J. J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K. C., Williamson, C. H. D., Willis, A. D., Xu, Z. Z., Zaneveld, J. R., Zhang, Y., Zhu, Q., Knight, R., & Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology, 37*(8), 852-857. https://doi.org/10.1038/s41587-019-0209-9

Borchers, H. W. (2023). *pracma: Practical Numerical Math Functions*. In *Practical numerical math functions, version* (Version R package version 2.4.4) CRAN. https://CRAN.R-project.org/package=pracma

Borges, H. L. F., Branco, L. H. Z., Martins, M. D., Lima, C. S., Barbosa, P. T., Lira, G. A. S. T., Bittencourt-Oliveira, M. C., & Molica, R. J. R. (2015). Cyanotoxin production and phylogeny of benthic cyanobacterial strains isolated from the northeast of Brazil. *Harmful Algae, 43*, 46-57. https://doi.org/10.1016/j.hal.2015.01.003

Boström, B., Pettersson, A.-K., & Ahlgren, I. (1989). Seasonal dynamics of a cyanobacteria-dominated microbial community in surface sediments of a shallow, eutrophic lake. *Aquatic Sciences, 51*(2), 153-178. https://doi.org/10.1007/BF00879300

Bowes, M. J., Armstrong, L. K., Harman, S. A., Wickham, H. D., Nicholls, D. J. E., Scarlett, P. M., Roberts, C., Jarvie, H. P., Old, G. H., Gozzard, E., Bachiller-Jareno, N., & Read, D. S. (2018). Weekly water quality monitoring data for the River Thames (UK) and its major tributaries (2009–2013): the Thames Initiative research platform. *Earth Syst. Sci. Data, 10*(3), 1637-1653. https://doi.org/10.5194/essd-10-1637-2018

Bowes, M. J., Gozzard, E., Newman, J., Loewenthal, M., Halliday, S., Skeffington, R. A., Jarvie, H. P., Wade, A., & Palmer-Felgate, E. (2015a). *Hourly physical and nutrient monitoring data for the River Enborne, Berkshire (2009-2012)* NERC Environmental Information Data Centre.

Bowes, M. J., Jarvie, H. P., Halliday, S. J., Skeffington, R. A., Wade, A. J., Loewenthal, M., Gozzard, E., Newman, J. R., & Palmer-Felgate, E. J. (2015b). Characterising phosphorus and nitrate inputs to a rural river using high-frequency concentration-flow relationships. *Science of the Total Environment, 511*, 608-620. https://doi.org/10.1016/j.scitotenv.2014.12.086

Bowes, M. J., Loewenthal, M., Read, D. S., Hutchins, M. G., Prudhomme, C., Armstrong, L. K., Harman, S. A., Wickham, H. D., Gozzard, E., & Carvalho, L. (2016). Identifying multiple stressor controls on phytoplankton dynamics in the River Thames (UK) using high-

frequency water quality data. *Science of the Total Environment, 569-570*, 1489-1499. https://doi.org/10.1016/j.scitotenv.2016.06.239

Bowes, M. J., Smith, J. T., & Neal, C. (2009). The value of high-resolution nutrient monitoring: A case study of the River Frome, Dorset, UK. *Journal of Hydrology, 378*(1), 82-96. https://doi.org/10.1016/j.jhydrol.2009.09.015

Brownlee, B., MacInnis, G., Charlton, M., Watson, S., Hamilton-Browne, S., & Milne, J. (2004). An analytical method for shipboard extraction of the odour compounds, 2-methylisoborneol and geosmin. *Water Science and Technology, 49*(9), 121-127. https://doi.org/10.2166/wst.2004.0550

Calderini, M. L., Pääkkönen, S., Salmi, P., Peltomaa, E., & Taipale, S. J. (2023). Temperature, phosphorus and species composition will all influence phytoplankton production and content of polyunsaturated fatty acids. *J Plankton Res, 45*(4), 625-635. https://doi.org/10.1093/plankt/fbad026

Callahan, B. J. (2018). *Silva taxonomic training data formatted for DADA2 (Silva version 132)*. https://doi.org/10.5281/zenodo.1172783

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods, 13*(7), 581-583. https://doi.org/10.1038/nmeth.3869

Cantoral Uriza, E. A., Asencio, A. D., & Aboal, M. (2017). Are We Underestimating Benthic Cyanotoxins? Extensive Sampling Results from Spain. *Toxins, 9*(12). https://doi.org/10.3390/toxins9120385

Cantwell, H. (2021). *Determining the drivers of harmful algal blooms and their impact on public water supply resilience during droughts* Cardiff University].

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, 108*(supplement_1), 4516-4522. https://doi.org/10.1073/pnas.1000080107

Carey, C. C., Woelmer, W. M., Lofton, M. E., Figueiredo, R. J., Bookout, B. J., Corrigan, R. S., Daneshmand, V., Hounshell, A. G., Howard, D. W., Lewis, A. S. L., McClure, R. P., Wander, H. L., Ward, N. K., & Thomas, R. Q. (2021). Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*, 1-14. https://doi.org/10.1080/20442041.2020.1816421

Cassidy, R., & Jordan, P. (2011). Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. *Journal of Hydrology, 405*(1), 182-193. https://doi.org/10.1016/j.jhydrol.2011.05.020

Catherine, Q., Susanna, W., Isidora, E.-S., Mark, H., Aurélie, V., & Jean-François, H. (2013). A review of current knowledge on toxic benthic freshwater cyanobacteria – Ecology, toxin production and risk management. *Water Research, 47*(15), 5464-5479. https://doi.org/10.1016/j.watres.2013.06.042

Chaffin, J. D., & Bridgeman, T. B. (2014). Organic and inorganic nitrogen utilization by nitrogen-stressed cyanobacteria during bloom conditions. *Journal of Applied Phycology, 26*(1), 299-309. https://doi.org/10.1007/s10811-013-0118-0

Chappell, N. A., Jones, T. D., & Tych, W. (2017). Sampling frequency for water quality variables in streams: Systems analysis to quantify minimum monitoring rates. *Water Research, 123*, 49-57. https://doi.org/10.1016/j.watres.2017.06.047

Chen, C., Zamyadi, A., Lin, T.-F., & Gallagher, D. (2019). Removal of odourants from drinking water. In *Taste and Odour in Source and Drinking Water: Causes, Controls, and Consequences* (pp. 167-209). IWA Publishing London, UK.

Chen, N., Mo, Q., Kuo, Y.-M., Su, Y., & Zhong, Y. (2018). Hydrochemical controls on reservoir nutrient and phytoplankton dynamics under storms. *Science of the Total Environment, 619-620*, 301-310. https://doi.org/10.1016/j.scitotenv.2017.09.216

Chen, X., Dolinova, I., Sevcu, A., Jurczak, T., Frankiewicz, P., Wojtal-Frankiewicz, A., Wan, L., Deng, Q., Song, C., Zhou, Y., & Cao, X. (2020). Strategies adopted by Aphanizomenon flos-aquae in response to phosphorus deficiency and their role on growth. *Environmental Sciences Europe, 32*(1). https://doi.org/10.1186/s12302-020-00328-3

Chislock, M. F., Olsen, B. K., Choi, J., Abebe, A., Bleier, T. L., & Wilson, A. E. (2021). Contrasting patterns of 2-methylisoborneol (MIB) vs. geosmin across depth in a drinking water reservoir are mediated by cyanobacteria and actinobacteria. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-021-12973-z

Chiu, Y. T., Yen, H. K., & Lin, T. F. (2016). An alternative method to quantify 2-MIB producing cyanobacteria in drinking water reservoirs: Method development and field applications. *Environmental Research, 151*, 618-627. https://doi.org/10.1016/j.envres.2016.08.034

Chong, S., Lee, H., & An, K.-G. (2018). Predicting Taste and Odor Compounds in a Shallow Reservoir Using a Three–Dimensional Hydrodynamic Ecological Model. *Water, 10*(10), 1396. https://doi.org/10.3390/w10101396

Chung, S. W., Chong, S. A., & Park, H. S. (2016). Development and Applications of a Predictive Model for Geosmin in North Han River, Korea. *Procedia Engineering, 154*, 521-528. https://doi.org/10.1016/j.proeng.2016.07.547

Clercin, N. A., & Druschel, G. K. (2019). Influence of Environmental Factors on the Production of MIB and Geosmin Metabolites by Bacteria in a Eutrophic Reservoir. *Water Resources Research, 55*(7), 5413-5430. https://doi.org/10.1029/2018WR023651

Clercin, N. A., Koltsidou, I., Picard, C. J., & Druschel, G. K. (2022). Prevalence of Actinobacteria in the production of 2-methylisoborneol and geosmin, over Cyanobacteria in a temperate eutrophic reservoir. *Chemical Engineering Journal Advances, 9*, 100226. https://doi.org/10.1016/j.ceja.2021.100226

Clinton-Bailey, G. S., Grand, M. M., Beaton, A. D., Nightingale, A. M., Owsianka, D. R., Slavik, G. J., Connelly, D. P., Cardwell, C. L., & Mowlem, M. C. (2017). A Lab-on-Chip Analyzer for in Situ Measurement of Soluble Reactive Phosphate: Improved Phosphate Blue Assay and Application to Fluvial Monitoring. *Environmental science & technology, 51*(17), 9989-9995. https://doi.org/10.1021/acs.est.7b01581

Collins, A. L., Walling, D. E., Stroud, R. W., Robson, M., & Peet, L. M. (2010). Assessing damaged road verges as a suspended sediment source in the Hampshire Avon catchment, southern United Kingdom. *Hydrological Processes, 24*(9), 1106-1122. https://doi.org/10.1002/hyp.7573

Collos, Y., & Harrison, P. J. (2014). Acclimation and toxicity of high ammonium concentrations to unicellular algae. *Marine Pollution Bulletin, 80*(1), 8-23. https://doi.org/10.1016/j.marpolbul.2014.01.006

Cook, K. V., Li, C., Cai, H., Krumholz, L. R., Hambright, K. D., Paerl, H. W., Steffen, M. M., Wilson, A. E., Burford, M. A., Grossart, H.-P., Hamilton, D. P., Jiang, H., Sukenik, A., Latour, D., Meyer, E. I., Padisák, J., Qin, B., Zamor, R. M., & Zhu, G. (2020). The global Microcystis interactome. *Limnology and Oceanography, 65*(Suppl 1), S194-S207. https://doi.org/10.1002/lno.11361

Cooper, R. J., Hiscock, K. M., Lovett, A. A., Dugdale, S. J., Sünnenberg, G., Garrard, N. L., Outram, F. N., Hama-Aziz, Z. Q., Noble, L., & Lewis, M. A. (2018). Application of high-resolution telemetered sensor technology to develop conceptual models of catchment hydrogeological processes. *Journal of Hydrology X, 1*. https://doi.org/10.1016/j.hydroa.2018.100007

Cooper, R. J., Hiscock, K. M., Lovett, A. A., Dugdale, S. J., Sunnenberg, G., & Vrain, E. (2020). Temporal hydrochemical dynamics of the River Wensum, UK: Observations from long-term high-resolution monitoring (2011-2018). *Science of the Total Environment, 724*, 138253. https://doi.org/10.1016/j.scitotenv.2020.138253

Cooper, R. J., Outram, F. N., & Hiscock, K. M. (2016). Diel turbidity cycles in a headwater stream: evidence of nocturnal bioturbation? *Journal of Soils and Sediments, 16*(6), 1815-1824. https://doi.org/10.1007/s11368-016-1372-y

Coraggio, E., Han, D., Gronow, C., & Tryfonas, T. (2022). Water Quality Sampling Frequency Analysis of Surface Freshwater: A Case Study on Bristol Floating Harbour [Original Research]. *Frontiers in Sustainable Cities, 3*. https://doi.org/10.3389/frsc.2021.791595

Cotterill, V., Hamilton, D. P., Puddick, J., Suren, A., & Wood, S. A. (2019). Phycocyanin sensors as an early warning system for cyanobacteria blooms concentrations: a case study in the Rotorua lakes. *New Zealand Journal of Marine and Freshwater Research, 53*(4), 555-570. https://doi.org/10.1080/00288330.2019.1617322

Cottingham, K. L., Ewing, H. A., Greer, M. L., Carey, C. C., & Weathers, K. C. (2015). Cyanobacteria as biological drivers of lake nitrogen and phosphorus cycling. *Ecosphere, 6*(1), 1-19. https://doi.org/10.1890/ES14-00174.1

Cottingham, K. L., Weathers, K. C., Ewing, H. A., Greer, M. L., & Carey, C. C. (2021). Predicting the effects of climate change on freshwater cyanobacterial blooms requires consideration of the complete cyanobacterial life cycle. *Journal of Plankton Research, 43*(1), 10-19. https://doi.org/10.1093/plankt/fbaa059

Creed, I. F., McKnight, D. M., Pellerin, B. A., Green, M. B., Bergamaschi, B. A., Aiken, G. R., Burns, D. A., Findlay, S. E. G., Shanley, J. B., Striegl, R. G., Aulenbach, B. T., Clow, D. W., Laudon, H., McGlynn, B. L., McGuire, K. J., Smith, R. A., & Stackpoole, S. M. (2015). The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. *Canadian Journal of Fisheries and Aquatic Sciences, 72*(8), 1272-1285. https://doi.org/10.1139/cjfas-2014-0400

Crockford, L., O'Riordain, S., Taylor, D., Melland, A. R., Shortle, G., & Jordan, P. (2017). The application of high temporal resolution data in river catchment modelling and management strategies. *Environmental Monitoring and Assessment, 189*(9), 461. https://doi.org/10.1007/s10661-017-6174-1

Cruaud, P., Vigneron, A., Fradette, M.-S., Charette, S. J., Rodriguez, M. J., Dorea, C. C., & Culley, A. I. (2017). Open the SterivexTM casing: An easy and effective way to improve DNA extraction yields. *Limnology and Oceanography: Methods, 15*(12), 1015-1020. https://doi.org/10.1002/lom3.10221

Danhiez, F. P., Vantrepotte, V., Cauvin, A., Lebourg, E., & Loisel, H. (2017). Optical properties of chromophoric dissolved organic matter during a phytoplankton bloom. Implication for DOC estimates from CDOM absorption. *Limnology and Oceanography, 62*(4), 1409-1425. https://doi.org/10.1002/lno.10507

Daniel, A., Laës-Huon, A., Barus, C., Beaton, A. D., Blandfort, D., Guigues, N., Knockaert, M., Munaron, D., Salter, I., Woodward, E. M. S., Greenwood, N., & Achterberg, E. P. (2020). Toward a Harmonization for Using in situ Nutrient Sensors in the Marine Environment. *Frontiers in Marine Science, 6*(773). https://doi.org/10.3389/fmars.2019.00773

Dayarathne, K., Ishikawa, T., Kadeer, A., Yamaguchi, M., & Kawai-Yamada, M. (2024). The effect of light availability and light wavelength on growth, 2-MIB biosynthesis, and 2-MIB-related gene expression in Pseudanabaena foetida var. intermedia. Archives of Microbiology, 206(9), 367. https://doi.org/10.1007/s00203-024-04099-w

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology, 90*, 3566-3574. https://doi.org/10.1890/08-1823.1

De Senerpont Domis, L. N., Mooij, W. M., & Huisman, J. (2007). Climate-induced shifts in an experimental phytoplankton community: a mechanistic approach. In Gulati R.D., Lammens E., De Pauw N., & V. D. E. (Eds.), *Shallow lakes in a changing world* (Vol. 196, pp. 403-413). Springer. https://doi.org/10.1007/978-1-4020-6399-2_36

DEFRA. (2023). *Continuous Water Quality Monitoring Programme. Provisional technical guidance for sewerage undertakers on implementing s.82 of the Environment Act 2021*.

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences of the United States of America, 89*(12), 5685-5689. https://doi.org/10.1073/pnas.89.12.5685

Descy, J. P., Leprieur, F., Pirlot, S., Leporcq, B., Van Wichelen, J., Peretyatko, A., Teissier, S., Codd, G. A., Triest, L., Vyverman, W., & Wilmotte, A. (2016). Identifying the factors determining blooms of cyanobacteria in a set of shallow lakes. *Ecological Informatics, 34*, 129-138. https://doi.org/10.1016/j.ecoinf.2016.05.003

Devi, A., Chiu, Y.-T., Hsueh, H.-T., & Lin, T.-F. (2021). Quantitative PCR based detection system for cyanobacterial geosmin/2-methylisoborneol (2-MIB) events in drinking water sources: Current status and challenges. *Water Research, 188*, 116478. https://doi.org/10.1016/j.watres.2020.116478

Dodds, W. K., & Whiles, M. R. (2020). Chapter 14 - Nitrogen, Sulfur, Phosphorus, and Other Nutrients. In W. K. Dodds & M. R. Whiles (Eds.), *Freshwater Ecology (Third Edition)* (pp. 395-424). Academic Press. https://doi.org/10.1016/B978-0-12-813255-5.00014-4

Domingues, R. B., Barbosa, A. B., Sommer, U., & Galvão, H. M. (2011). Ammonium, nitrate and phytoplankton interactions in a freshwater tidal estuarine zone: potential effects of cultural eutrophication. *Aquatic Sciences, 73*(3), 331-343. https://doi.org/10.1007/s00027-011-0180-0

Donald, D. B., Bogard, M. J., Finlay, K., Bunting, L., & Leavitt, P. R. (2013). Phytoplankton-specific response to enrichment of phosphorus-rich surface waters with ammonium, nitrate, and urea. *PLoS One, 8*(1), e53277. https://doi.org/10.1371/journal.pone.0053277

Donald, D. B., Bogard, M. J., Finlay, K., & Leavitt, P. R. (2011). Comparative effects of urea, ammonium, and nitrate on phytoplankton abundance, community composition, and toxicity in hypereutrophic freshwaters. *Limnology and Oceanography, 56*(6), 2161-2175. https://doi.org/10.4319/lo.2011.56.6.2161

Dortch, Q. (1990). The interaction between ammonium and nitrate uptake in phytoplankton. *Marine Ecology Progress Series, 61*(1/2), 183-201.

Dunlap, C. R., Sklenar, K. S., & Blake, L. J. (2015). A costly endeavor: Addressing algae problems in a water supply. *Journal - American Water Works Association, 107*(5), E255-E262. https://doi.org/10.5942/jawwa.2015.107.0055

Durand, P., Breuer, L., Johnes, P. J., Billen, G., Butturini, A., Pinay, G., van Grinsven, H., Garnier, J., Rivett, M., Reay, D. S., Curtis, C., Siemens, J., Maberly, S., Kaste, O., Humborg, C., Loeb, R., de Klein, J., Hejzlar, J., Skoulikidis, N., Kortelainen, P., Lepisto, A., & Wright, R. (2011). Nitrogen processes in aquatic ecosystems. In M. A. Sutton, C. M. Howard, J. W. Erisman, G. Billen, A. Bleeker, P. Grennfelt, H. van Grinsven, & B. Grizzetti (Eds.), *The European Nitrogen Assessment* (pp. 126-146). Cambridge University Press. https://doi.org/9781107006126c07_p126-146.indd

Durrer, M., Zimmermann, U., & Jüttner, F. (1999). Dissolved and particle-bound geosmin in a mesotrophic lake (Lake Zurich): Spatial and seasonal distribution and the effect of grazers. *Water Research, 33*(17), 3628-3636. https://doi.org/10.1016/S0043-1354(99)00069-X

Dzialowski, A. R., Smith, V. H., Huggins, D. G., deNoyelles, F., Lim, N. C., Baker, D. S., & Beury, J. H. (2009). Development of predictive models for geosmin-related taste and odor in Kansas, USA, drinking water reservoirs. *Water Research, 43*(11), 2829-2840. https://doi.org/10.1016/j.watres.2009.04.001

Elser, J. J., Bracken, M. E. S., Cleland, E. E., Gruner, D. S., Harpole, W. S., Hillebrand, H., Ngai, J. T., Seabloom, E. W., Shurin, J. B., & Smith, J. E. (2007). Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial

ecosystems. *Ecology Letters, 10*(12), 1135-1142. https://doi.org/10.1111/j.1461-0248.2007.01113.x

Erratt, K. J., Creed, I. F., & Trick, C. G. (2020). Differential Drawdown of Ammonium, Nitrate, and Urea by Freshwater Chlorophytes and Cyanobacteria1. *Journal of Phycology, 56*(2), 458-468. https://doi.org/10.1111/jpy.12960

Espinosa, C., Abril, M., Guasch, H., Pou, N., Proia, L., Ricart, M., Ordeix, M., & Llenas, L. (2020). Water Flow and Light Availability Influence on Intracellular Geosmin Production in River Biofilms. *Frontiers in Microbiology, 10*, 3002. https://doi.org/10.3389/fmicb.2019.03002

Espinosa, C., Abril, M., Ponsa, S., Ricart, M., Vendrell-Puigmitja, L., Ordeix, M., Llenas, L., & Proia, L. (2021). Effects of the interaction between nutrient concentration and DIN:SRP ratio on geosmin production by freshwater biofilms. *Science of the Total Environment, 768*, 144473. https://doi.org/10.1016/j.scitotenv.2020.144473

Evans, D. J., & Johnes, P. J. (2004). Physico-chemical controls on phosphorus cycling in two lowland streams. Part 1 – the water column. *Science of the Total Environment, 329*(1), 145-163. https://doi.org/10.1016/j.scitotenv.2004.02.018

Fawley, M. W., & Fawley, K. P. (2004). A simple and rapid technique for the isolation of DNA from microalgae. *Journal of Phycology, 40*(1), 223-225. https://doi.org/10.1111/j.0022-3646.2004.03-081.x

Feng, W., Wu, F., He, Z., Song, F., Zhu, Y., Giesy, J. P., Wang, Y., Qin, N., Zhang, C., Chen, H., & Sun, F. (2018). Simulated bioavailability of phosphorus from aquatic macrophytes and phytoplankton by aqueous suspension and incubation with alkaline phosphatase. *Science of the Total Environment, 616-617*, 1431-1439. https://doi.org/10.1016/j.scitotenv.2017.10.172

Foy, R. H., & Kirk, M. (1995). Agriculture and Water Quality: A Regional Study. *Water and Environment Journal, 9*(3), 247-256. https://doi.org/10.1111/j.1747-6593.1995.tb00937.x

Franklin, H. M., Podduturi, R., Jørgensen, N. O. G., Roberts, D. T., Schlüter, L., & Burford, M. A. (2023). Potential sources and producers of 2-methylisoborneol and geosmin in a river supplying a drinking water treatment plant. *Chemical Engineering Journal Advances, 14*, 100455. https://doi.org/10.1016/j.ceja.2023.100455

Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics, 46*(Volume 46, 2015), 1-23. https://doi.org/10.1146/annurev-ecolsys-110411-160340

Gaget, V., Almuhtaram, H., Kibuye, F., Hobson, P., Zamyadi, A., Wert, E., & Brookes, J. D. (2022). Benthic cyanobacteria: A utility-centred field study. *Harmful Algae, 113*, 102185. https://doi.org/10.1016/j.hal.2022.102185

Gaget, V., Hobson, P., Keulen, A., Newton, K., Monis, P., Humpage, A. R., Weyrich, L. S., & Brookes, J. D. (2020). Toolbox for the sampling and monitoring of benthic cyanobacteria. *Water Research, 169*, 115222-115222. https://doi.org/10.1016/j.watres.2019.115222

Gaget, V., Humpage, A. R., Huang, Q., Monis, P., & Brookes, J. D. (2017). Benthic cyanobacteria: A source of cylindrospermopsin and microcystin in Australian drinking water reservoirs. *Water Research, 124*, 454-464. https://doi.org/10.1016/j.watres.2017.07.073

Gao, J., Zhu, J., Wang, M., & Dong, W. (2018). Dominance and Growth Factors of Pseudanabaena sp. in Drinking Water Source Reservoirs, Southern China. *Sustainability, 10*(11). https://doi.org/10.3390/su10113936

Giglio, S., Saint, C. P., & Monis, P. T. (2011). Expression of the geosmin synthase gene in the cyanobacterium anabaena circinalis AWQC318. *Journal of Phycology, 47*(6), 1338-1343. https://doi.org/10.1111/j.1529-8817.2011.01061.x

Glibert, P. M. (2020). From hogs to HABs: impacts of industrial farming in the US on nitrogen and phosphorus and greenhouse gas pollution. *Biogeochemistry, 150*(2), 139-180. https://doi.org/10.1007/s10533-020-00691-6

Glibert, P. M., & Berg, G. M. (2009). Nitrogen form, fate and phytoplankton composition. In W. M. K. V. S. Kennedy, J. E. Peterson, and W. C. Dennison (Ed.), *Experimental ecosystems and scale: Tools for understanding and managing coastal ecosystems* (pp. 183-189). Springer.

Glibert, P. M., Kelly, V., Alexander, J., Codispoti, L. A., Boicourt, W. C., Trice, T. M., & Michael, B. (2008). In situ nutrient monitoring: A tool for capturing nutrient variability and the antecedent conditions that support algal blooms. *Harmful Algae, 8*(1), 175-181. https://doi.org/10.1016/j.hal.2008.08.013

Glibert, P. M., Wilkerson, F. P., Dugdale, R. C., Raven, J. A., Dupont, C. L., Leavitt, P. R., Parker, A. E., Burkholder, J. M., & Kana, T. M. (2016). Pluses and minuses of ammonium and nitrate uptake and assimilation by phytoplankton and implications for productivity and community composition, with emphasis on nitrogen-enriched conditions. *Limnology and Oceanography, 61*(1), 165-197. https://doi.org/10.1002/lno.10203

Grand, M. M., Clinton-Bailey, G. S., Beaton, A. D., Schaap, A. M., Johengen, T. H., Tamburri, M. N., Connelly, D. P., Mowlem, M. C., & Achterberg, E. P. (2017). A Lab-On-Chip Phosphate Analyzer for Long-term In Situ Monitoring at Fixed Observatories: Optimization and Performance Evaluation in Estuarine and Oligotrophic Coastal Waters. *Frontiers in Marine Science, 4*, 255. https://doi.org/10.3389/fmars.2017.00255

Guttman, L., & van Rijn, J. (2012). Isolation of bacteria capable of growth with 2-methylisoborneol and geosmin as the sole carbon and energy sources. *Applied and Environmental Microbiology, 78*(2), 363-370. https://doi.org/10.1128/AEM.06333-11

Halliday, S. J., Skeffington, R. A., Bowes, M. J., Gozzard, E., Newman, J. R., Loewenthal, M., Palmer-Felgate, E. J., Jarvie, H. P., & Wade, A. J. (2014). The Water Quality of the River Enborne, UK: Observations from High-Frequency Monitoring in a Rural, Lowland River System. *Water, 6*(1), 150-180. https://doi.org/10.3390/w6010150

Halliday, S. J., Skeffington, R. A., Wade, A. J., Bowes, M. J., Gozzard, E., Newman, J. R., Loewenthal, M., Palmer-Felgate, E. J., & Jarvie, H. P. (2015). High-frequency water quality monitoring in an urban catchment: hydrochemical dynamics, primary production and implications for the Water Framework Directive. *Hydrological Processes, 29*(15), 3388-3407. https://doi.org/10.1002/hyp.10453

Halliday, S. J., Wade, A. J., Skeffington, R. A., Neal, C., Reynolds, B., Rowland, P., Neal, M., & Norris, D. (2012). An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales. *Science of the Total Environment, 434*, 186-200. https://doi.org/10.1016/j.scitotenv.2011.10.052

Hammond, N. W., Birgand, F., Carey, C. C., Bookout, B., Breef-Pilz, A., & Schreiber, M. E. (2023). High-frequency sensor data capture short-term variability in Fe and Mn concentrations due to hypolimnetic oxygenation and seasonal dynamics in a drinking water reservoir. *Water Research, 240*, 120084. https://doi.org/10.1016/j.watres.2023.120084

Hanson, D. (2023, September 2023). Designing an effective water quality monitoring programme. *Water Industry Journal*, (28), 50-51.

Harris, T. D., & Graham, J. L. (2017). Predicting cyanobacterial abundance, microcystin, and geosmin in a eutrophic drinking-water reservoir using a 14-year dataset. *Lake and Reservoir Management, 33*(1), 32-48. https://doi.org/10.1080/10402381.2016.1263694

Harris, T. D., Reinl, K. L., Azarderakhsh, M., Berger, S. A., Berman, M. C., Bizic, M., Bhattacharya, R., Burnet, S. H., Cianci-Gaskill, J. A., Domis, L. N. d. S., Elfferich, I., Ger, K. A., Grossart, H.-P. F., Ibelings, B. W., Ionescu, D., Kouhanestani, Z. M., Mauch, J., McElarney, Y. R., Nava, V., North, R. L., Ogashawara, I., Paule-Mercado, M. C. A., Soria-Píriz, S., Sun, X., Trout-Haney, J. V., Weyhenmeyer, G. A., Yokota, K., & Zhan, Q. (2024). What makes a cyanobacterial bloom disappear? A review of the abiotic and biotic cyanobacterial bloom loss factors. *Harmful Algae, 133*, 102599. https://doi.org/10.1016/j.hal.2024.102599

Harris, T. D., Smith, V. H., Graham, J. L., Waa, D. B. V. d., Tedesco, L. P., & Clercin, N. (2016). Combined effects of nitrogen to phosphorus and nitrate to ammonia ratios on cyanobacterial metabolite concentrations in eutrophic Midwestern USA reservoirs. *Inland Waters, 6*(2), 199-210. https://doi.org/10.5268/IW-6.2.938

Harris, T. D., Wilhelm, F. M., Graham, J. L., & Loftin, K. A. (2014). Experimental manipulation of TN:TP ratios suppress cyanobacterial biovolume and microcystin concentration in large-scale in situ mesocosms. *Lake and Reservoir Management, 30*(1), 72-83. https://doi.org/10.1080/10402381.2013.876131

Hatton-Ellis, T. W. (2016). Evidence Review of Lake Nitrate Vulnerable Zones in Wales. NRW Evidence Report No: 135. Bangor: *Natural Resources Wales*.

Heath, M. W., Wood, S. A., Brasell, K. A., Young, R. G., & Ryan, K. G. (2015). Development of Habitat Suitability Criteria and In-Stream Habitat Assessment for the Benthic Cyanobacteria Phormidium. *River Research and Applications, 31*(1), 98-108. https://doi.org/10.1002/rra.2722

Heathwaite, A. L., & Johnes, P. J. (1996). Contribution of nitrogen species and phosphorus fractions to stream water quality in agricultural catchments. *Hydrological Processes, 10*(7), 971-983. https://doi.org/10.1002/(SICI)1099-1085(199607)10:7<971::AID-HYP351>3.0.CO;2-N

Heathwaite, A. L., Johnes, P. J., & Peters, N. E. (1996). Trends in nutrients. *Hydrological Processes, 10*(2), 263-293. https://doi.org/10.1002/(SICI)1099-1085(199602)10:2<263::AID-HYP441>3.0.CO;2-K

Hecht, J. S., Zia, A., Clemins, P. J., Schroth, A. W., Winter, J. M., Oikonomou, P. D., & Rizzo, D. M. (2022). Modeling the sensitivity of cyanobacteria blooms to plausible changes in precipitation and air temperature variability. *Science of the Total Environment, 812*, 151586. https://doi.org/10.1016/j.scitotenv.2021.151586

Hoffman, D. K., McCarthy, M. J., Boedecker, A. R., Myers, J. A., & Newell, S. E. (2022). The role of internal nitrogen loading in supporting non-N-fixing harmful cyanobacterial blooms in the water column of a large eutrophic lake. *Limnology and Oceanography, 67*(9), 2028-2041. https://doi.org/10.1002/lno.12185

Hooper, A. S. (2023a). *Environmental triggers for geosmin and 2-MIB production in drinking water reservoirs* Cardiff University].

Hooper, A. S., Kille, P., Watson, S. E., Christofides, S. R., & Perkins, R. G. (2023b). The importance of nutrient ratios in determining elevations in geosmin synthase (geoA) and 2-MIB cyclase (mic) resulting in taste and odour events. *Water Research, 232*, 119693. https://doi.org/10.1016/j.watres.2023.119693

House, W. A. (2003). Geochemical cycling of phosphorus in rivers. *Applied Geochemistry, 18*(5), 739-748. https://doi.org/10.1016/S0883-2927(02)00158-0

Howard, C. S. (2020). *Taste and odor event dynamics of a midwestern freshwater reservoir*

Huang, H., Xu, X., Shi, C., Liu, X., & Wang, G. (2018a). Response of Taste and Odor Compounds to Elevated Cyanobacteria Biomass and Temperature. *Bulletin of Environmental Contamination and Toxicology, 101*(2), 272-278. https://doi.org/10.1007/s00128-018-2386-5

Huang, X., Huang, Z., Chen, X.-P., Zhang, D., Zhou, J., Wang, X., & Gao, N. (2018b). The predominant phytoplankton of Pseudoanabaena holding specific biosynthesis gene-derived occurrence of 2-MIB in a drinking water reservoir. *Environmental Science and Pollution Research, 25*(19), 19134-19142. https://doi.org/10.1007/s11356-018-2086-z

Huber, V., Wagner, C., Gerten, D., & Adrian, R. (2012). To bloom or not to bloom: contrasting responses of cyanobacteria to recent heat waves explained by critical thresholds of abiotic drivers. *Oecologia, 169*(1), 245-256. https://doi.org/10.1007/s00442-011-2186-7

Huisman, J., Sharples, J., Stroom, J. M., Visser, P. M., Kardinaal, W. E. A., Verspagen, J. M. H., & Sommeijer, B. (2004). Changes in turbulent mixing shift competition for light between phytoplankton species. *Ecology, 85*(11), 2960-2970. https://doi.org/10.1890/03-0763

Hupfer, M., & Lewandowski, J. (2008). Oxygen Controls the Phosphorus Release from Lake Sediments – a Long-Lasting Paradigm in Limnology. *International Review of Hydrobiology, 93*(4-5), 415-432. https://doi.org/10.1002/iroh.200711054

Hyenstrand, P., Burkert, U., Pettersson, A., & Blomqvist, P. (2000). Competition between the green alga Scenedesmus and the cyanobacterium Synechococcus under different modes of inorganic nitrogen supply. *Hydrobiologia, 435*(1), 91-98. https://doi.org/10.1023/A:1004008721373

Ibelings, B. W., Mur, L. R., & Walsby, A. E. (1991). Diurnal changes in buoyancy and vertical distribution in populations of Microcystis in two shallow lakes. *Journal of Plankton Research, 13*(2), 419-436. https://doi.org/10.1093/plankt/13.2.419

ICH Harmonised Tripartite Guideline (2005). *Validation of analytical procedures: text and methodology Q2 (R1)* International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use, Geneva, Switzerland.

Izaguirre, G., & Taylor, W. D. (1998). A pseudanabaena species from Castaic Lake, California, that produces 2-methylisoborneol. *Water Research, 32*(5), 1673-1677. https://doi.org/10.1016/S0043-1354(97)00379-5

Jähnichen, S., Jäschke, K., Wieland, F., Packroff, G., & Benndorf, J. (2011). Spatio-temporal distribution of cell-bound and dissolved geosmin in Wahnbach Reservoir: Causes and potential odour nuisances in raw water. *Water Research, 45*(16), 4973-4982. https://doi.org/10.1016/j.watres.2011.06.043

Jalili, F., Moradinejad, S., Zamyadi, A., Dorner, S., Sauvé, S., & Prévost, M. (2022). Evidence-Based Framework to Manage Cyanobacteria and Cyanotoxins in Water and Sludge from Drinking Water Treatment Plants. *Toxins, 14*(6). https://doi.org/10.3390/toxins14060410

Jensen, H. S., & Andersen, F. O. (1992). Importance of temperature, nitrate, and pH for phosphate release from aerobic sediments of four shallow, eutrophic lakes. *Limnology and Oceanography, 37*(3), 577-589. https://doi.org/10.4319/lo.1992.37.3.0577

Jeong, J.-Y., Lee, S.-H., Yun, M.-R., Oh, S.-E., Lee, K.-H., & Park, H.-D. (2021). 2-Methylisoborneol (2-MIB) Excretion by Pseudanabaena yagii under Low Temperature. *Microorganisms, 9*(12). https://doi.org/10.3390/microorganisms9122486

Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K. M., Andersen, H. E., Lauridsen, T. L., Liboriussen, L., Beklioglu, M., Özen, A., & Olesen, J. E. (2009). Climate Change Effects on Runoff, Catchment Phosphorus Loading and Lake Ecological State, and Potential Adaptations. *Journal of Environmental Quality, 38*(5), 1930-1941. https://doi.org/10.2134/jeq2008.0113

Jeppesen, E., Kronvang, B., Olesen, J. E., Audet, J., Søndergaard, M., Hoffmann, C. C., Andersen, H. E., Lauridsen, T. L., Liboriussen, L., Larsen, S. E., Beklioglu, M., Meerhoff, M., Özen, A., & Özkan, K. (2011). Climate change effects on nitrogen loading from cultivated catchments in Europe: implications for nitrogen retention, ecological state of lakes and adaptation. *Hydrobiologia, 663*(1), 1-21. https://doi.org/10.1007/s10750-010-0547-6

Jiang, Y., Cheng, B., Liu, M., & Nie, Y. (2016). Spatial and Temporal Variations of Taste and Odor Compounds in Surface Water, Overlying Water and Sediment of the Western Lake Chaohu, China. *Bulletin of Environmental Contamination and Toxicology, 96*(2), 186-191. https://doi.org/10.1007/s00128-015-1698-y

Johnes, P. J. (2007). Uncertainties in annual riverine phosphorus load estimation: Impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. *Journal of Hydrology, 332*(1-2), 241-258. https://doi.org/10.1016/j.jhydrol.2006.07.006

Johnes, P. J., & Heathwaite, A. L. (1992). A procedure for the simultaneous determination of total nitrogen and total phosphorus in freshwater samples using persulphate microwave digestion. *Water Research, 26*(10), 1281-1287. https://doi.org/10.1016/0043-1354(92)90122-K

Johnston, B. D., Finkelstein, K. M., Gifford, S. R., Stouder, M. D., Nystrom, E. A., Savoy, P., Rosen, J. J., & Jennings, M. B. (2024). *Evaluation of sensors for continuous monitoring of harmful algal blooms in the Finger Lakes region, New York, 2019 and 2020* (2024-5010). (Scientific Investigations Report, Issue. U. S. G. Survey

Jones, C. A., & Welch, E. B. (1990). Internal Phosphorus Loading Related to Mixing and Dilution in a Dendritic, Shallow Prairie Lake. *Research Journal of the Water Pollution Control Federation, 62*(7), 847-852.

Jordan, P., Arnscheidt, A., McGrogan, H., & McCormick, S. (2007). Characterising phosphorus transfers in rural catchments using a continuous bank-side analyser. *Hydrol. Earth Syst. Sci., 11*(1), 372-381. https://doi.org/10.5194/hess-11-372-2007

Jordan, P., & Cassidy, R. (2022). Perspectives on Water Quality Monitoring Approaches for Behavioral Change Research. *Frontiers in Water, 4*. https://doi.org/10.3389/frwa.2022.917595

Jørgensen, N. O. G., Podduturi, R., & Burford, M. A. (2016). Relations between abundance of potential geosmin- and 2-MIB-producing organisms and concentrations of these compounds in water from three Australian reservoirs. *Journal of Water Supply: Research and Technology - AQUA, 65*(6), 504-513. https://doi.org/10.2166/aqua.2016.001

Kang, M., Peng, S., Tian, Y., & Zhang, H. (2018). Effects of dissolved oxygen and nutrient loading on phosphorus fluxes at the sediment–water interface in the Hai River Estuary, China. *Marine Pollution Bulletin, 130*, 132-139. https://doi.org/10.1016/j.marpolbul.2018.03.029

Kehoe, M. J., Chun, K. P., & Baulch, H. M. (2015). Who Smells? Forecasting Taste and Odor in a Drinking Water Reservoir. *Environmental science & technology, 49*(18), 10984-10992. https://doi.org/10.1021/acs.est.5b00979

Khalil, B., & Ouarda, T. B. M. J. (2009). Statistical approaches used to assess and redesign surface water-quality-monitoring networks. *Journal of Environmental Monitoring, 11*(11), 1915-1929. https://doi.org/10.1039/B909521G

Kibuye, F. A., Almuhtaram, H., Zamyadi, A., Gaget, V., Owen, C., Hofmann, R., & Wert, E. C. (2021). Utility practices and perspectives on monitoring and source control of cyanobacterial blooms. *AWWA Water Science, 3*(6), e1264. https://doi.org/10.1002/aws2.1264

Kim, H., Jo, B. Y., & Kim, H. S. (2017). Effect of different concentrations and ratios of ammonium, nitrate, and phosphate on growth of the blue-green alga (cyanobacterium) Microcystis aeruginosa isolated from the Nakdong River, Korea. *Algae, 32*(4), 275-284. https://doi.org/10.4490/algae.2017.32.10.23

Kirchner, J. W., Feng, X., Neal, C., & Robson, A. J. (2004). The fine structure of water-quality dynamics: The (high-frequency) wave of the future. *Hydrological processes, 18*(7), 1353-1359.

Klausmeier, C. A., Litchman, E., Daufresne, T., & Levin, S. A. (2004). Optimal nitrogen-to-phosphorus stoichiometry of phytoplankton. *Nature, 429*(6988), 171-174. https://doi.org/10.1038/nature02454

Kosten, S., Huszar, V. L. M., Mazzeo, N., Scheffer, M., Sternberg, L. d. S. L., & Jeppesen, E. (2009). Lake and watershed characteristics rather than climate influence nutrient limitation in shallow lakes. *Ecological Applications, 19*(7), 1791-1804. https://doi.org/10.1890/08-0906.1

Kristensen, P., Whalley, C., Néry, F., Zal, N., & Christiansen, T. (2018). *European waters: assessment of status and pressures 2018* (1725-9177). (EEA Report 2018, Issue. E. E. Agency

Kutovaya, O. A., & Watson, S. B. (2014). Development and application of a molecular assay to detect and monitor geosmin-producing cyanobacteria and actinomycetes in the Great Lakes. *Journal of Great Lakes Research, 40*(2), 404-414. https://doi.org/10.1016/j.jglr.2014.03.016

Lee, J., Rai, P. K., Jeon, Y. J., Kim, K. H., & Kwon, E. E. (2017). The role of algae and cyanobacteria in the production and release of odorants in water. *Environmental Pollution, 227*, 252-262. https://doi.org/10.1016/j.envpol.2017.04.058

Lee, J. E., Park, R., Yu, M., Byeon, M., & Kang, T. (2023). qPCR-Based Monitoring of 2-Methylisoborneol/Geosmin-Producing Cyanobacteria in Drinking Water Reservoirs in South Korea. *Microorganisms, 11*(9). https://doi.org/10.3390/microorganisms11092332

Lee, L. (2020). *NADA: Nondetects and Data Analysis for Environmental Data*. In (Version R package version 1.6-1.1) https://CRAN.R-project.org/package=NADA

Lévesque, D., Cattaneo, A., Hudon, C., & Gagnon, P. (2012). Predicting the risk of proliferation of the benthic cyanobacterium Lyngbya wollei in the St. Lawrence River. *Canadian Journal of Fisheries and Aquatic Sciences, 69*(10), 1585-1595. https://doi.org/10.1139/f2012-087

Li, L., Wan, N., Gan, N. Q., Xia, B. D., & Song, L. R. (2007). Annual dynamics and origins of the odorous compounds in the pilot experimental area of Lake Dianchi, China. *Water Science and Technology, 55*(5), 43-50. https://doi.org/10.2166/wst.2007.160

Li, Z., Hobson, P., An, W., Burch, M. D., House, J., & Yang, M. (2012). Earthy odor compounds production and loss in three cyanobacterial cultures. *Water Research, 46*(16), 5165-5173. https://doi.org/10.1016/j.watres.2012.06.008

Liu, J., & Vyverman, W. (2015). Differences in nutrient uptake capacity of the benthic filamentous algae Cladophora sp., Klebsormidium sp. and Pseudanabaena sp. under varying N/P conditions. *Bioresource Technology, 179*, 234-242. https://doi.org/10.1016/j.biortech.2014.12.028

Liu, Y., Chu, Z., Jin, X., Zang, Q., Liang, C., & Guo, W. (2009). Geosmin production and distribution of Anabaena sp. *China Environmental Science, 29*(10), 1082-1085.

Lloyd, C. E. M., Freer, J. E., Collins, A. L., Johnes, P. J., & Jones, J. I. (2014). Methods for detecting change in hydrochemical time series in response to targeted pollutant mitigation in river catchments. *Journal of Hydrology, 514*, 297-312. https://doi.org/10.1016/j.jhydrol.2014.04.036

Lloyd, C. E. M., Freer, J. E., Johnes, P. J., & Collins, A. L. (2016). Technical Note: Testing an improved index for analysing storm discharge–concentration hysteresis. *Hydrol. Earth Syst. Sci., 20*(2), 625-632. https://doi.org/10.5194/hess-20-625-2016

Lloyd, C. E. M., Freer, J. E., Johnes, P. J., Coxon, G., & Collins, A. L. (2015). Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams. *Hydrological Processes, 30*(1), 135-152. https://doi.org/10.1002/hyp.10574

Lloyd, C. E. M., Johnes, P. J., Freer, J. E., Carswell, A. M., Jones, J. I., Stirling, M. W., Hodgkinson, R. A., Richmond, C., & Collins, A. L. (2019). Determining the sources of nutrient flux to water in headwater catchments: Examining the speciation balance to inform the targeting of mitigation measures. *Science of the Total Environment, 648*, 1179-1200. https://doi.org/10.1016/j.scitotenv.2018.08.190

Louati, I., Pascault, N., Debroas, D., Bernard, C., Humbert, J.-F., & Leloup, J. (2015). Structural Diversity of Bacterial Communities Associated with Bloom-Forming Freshwater Cyanobacteria Differs According to the Cyanobacterial Genus. *PLoS One, 10*(11), e0140614. https://doi.org/10.1371/journal.pone.0140614

Lucius, M. A., Johnston, K. E., Eichler, L. W., Farrell, J. L., Moriarty, V. W., & Relyea, R. A. (2020). Using machine learning to correct for nonphotochemical quenching in high-frequency, in vivo fluorometer data. *Limnology and Oceanography: Methods, 18*(9), 477-494. https://doi.org/10.1002/lom3.10378

Lürling, M., & Faassen, E. J. (2013). Dog poisonings associated with a Microcystis aeruginosa bloom in the Netherlands. *Toxins, 5*(3), 556-567. https://doi.org/10.3390/toxins5030556

Ma, S. N., Wang, H. J., Wang, H. Z., Li, Y., Liu, M., Liang, X. M., Yu, Q., Jeppesen, E., & Søndergaard, M. (2018). High ammonium loading can increase alkaline phosphatase activity and promote sediment phosphorus release: A two-month mesocosm experiment. *Water Research, 145*, 388-397. https://doi.org/10.1016/j.watres.2018.08.043

Ma, Z., Niu, Y., Xie, P., Chen, J., Tao, M., & Deng, X. (2013). Off-flavor compounds from decaying cyanobacterial blooms of Lake Taihu. *Journal of Environmental Sciences, 25*(3), 495-501. https://doi.org/10.1016/S1001-0742(12)60101-6

Maberly, S. C., Pitt, J.-A., Davies, P. S., & Carvalho, L. (2020). Nitrogen and phosphorus limitation and the management of small productive lakes. *Inland Waters, 10*(2), 159-172. https://doi.org/10.1080/20442041.2020.1714384

Macintosh, K. A., Jordan, P., Cassidy, R., Arnscheidt, J., & Ward, C. (2011). Low flow water quality in rivers; septic tank systems and high-resolution phosphorus signals. *Science of the Total Environment, 412-413*, 58-65. https://doi.org/10.1016/j.scitotenv.2011.10.012

Mackay, E. B., Feuchtmayr, H., De Ville, M. M., Thackeray, S. J., Callaghan, N., Marshall, M., Rhodes, G., Yates, C. A., Johnes, P. J., & Maberly, S. C. (2020). Dissolved organic nutrient uptake by riverine phytoplankton varies along a gradient of nutrient enrichment. *Science of the Total Environment, 722*, 137837. https://doi.org/10.1016/j.scitotenv.2020.137837

Mallin, M. A., & McIver, M. R. (2012). Pollutant impacts to Cape Hatteras National Seashore from urban runoff and septic leachate. *Marine Pollution Bulletin, 64*(7), 1356-1366. https://doi.org/10.1016/j.marpolbul.2012.04.025

Marsh, T. J., & Hannaford, J. (2008). *UK hydrometric register. A catalogue of river flow gauging stations and observation wells and boreholes in the United Kingdom together with summary hydrometric and spatial statistics*. Centre for Ecology & Hydrology.

McCarthy, M. J., Gardner, W. S., Lehmann, M. F., Guindon, A., & Bird, D. F. (2016). Benthic nitrogen regeneration, fixation, and denitrification in a temperate, eutrophic lake: Effects on the nitrogen budget and cyanobacteria blooms. *Limnology and Oceanography, 61*(4), 1406-1423. https://doi.org/10.1002/lno.10306

McCarthy, M. J., James, R. T., Chen, Y., East, T. L., & Gardner, W. S. (2009). Nutrient ratios and phytoplankton community structure in the large, shallow, eutrophic, subtropical Lakes Okeechobee (Florida, USA) and Taihu (China). *Limnology, 10*(3), 215-227. https://doi.org/10.1007/s10201-009-0277-5

Meybeck, M., Chapman, D. V., & Helmer, R. (1990). Global freshwater quality: A first assessment. In (pp. 306). Global Environmental Monitoring System/UNEP/WHO.

Miller, T. R., Beversdorf, L., Chaston, S. D., & McMahon, K. D. (2013). Spatiotemporal Molecular Analysis of Cyanobacteria Blooms Reveals Microcystis-Aphanizomenon Interactions. *PLoS One, 8*(9), e74933. https://doi.org/10.1371/journal.pone.0074933

Milovanović, I., Mišan, A., Simeunović, J., Kovač, D., Jambrec, D., & Mandić, A. (2015). Determination of Volatile Organic Compounds in Selected Strains of Cyanobacteria. *Journal of Chemistry, 2015*, 969542. https://doi.org/10.1155/2015/969542

Moorhouse, H. L., Read, D. S., McGowan, S., Wagner, M., Roberts, C., Armstrong, L. K., Nicholls, D. J. E., Wickham, H. D., Hutchins, M. G., & Bowes, M. J. (2018). Characterisation of a major phytoplankton bloom in the River Thames (UK) using flow cytometry and high performance liquid chromatography. *Science of the Total Environment, 624*, 366-376. https://doi.org/10.1016/j.scitotenv.2017.12.128

Moss, B., Kosten, S., Meerhoff, M., Battarbee, R. W., Jeppesen, E., Mazzeo, N., Havens, K., Lacerot, G., Liu, Z., De Meester, L., Paerl, H., & Scheffer, M. (2011). Allied attack: climate change and eutrophication. *Inland Waters, 1*(2), 101-105. https://doi.org/10.5268/IW-1.2.359

Moulton, T. L. (2018). *rMR: Importing Data from Loligo Systems Software, Calculating Metabolic Rates and Critical Tensions*. In (Version R package version 1.1.0) https://CRAN.R-project.org/package=rMR

Mowlem, M., Beaton, A., Pascal, R., Schaap, A., Loucaides, S., Monk, S., Morris, A., Cardwell, C. L., Fowell, S. E., Patey, M. D., & López-García, P. (2021). Industry Partnership: Lab on Chip Chemical Sensor Technology for Ocean Observing. *Frontiers in Marine Science, 8*, 1407. https://doi.org/10.3389/fmars.2021.697611

Muyzer, G. (1998). Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. *Molecular microbial ecology manual. 3.4.4*.

Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology, 59*(3), 695-700. https://doi.org/10.1128/aem.59.3.695-700.1993

Naes, H., Utkilen, H. C., & Post, A. F. (1989). Geosmin production in the cyanobacterium Oscillatoria brevis. *Archives of Microbiology, 151*(5), 407-410. https://doi.org/10.1007/BF00416598

National Centre for Atmospheric Science. (2023). *National Centre for Atmospheric Science informs Environment Agency report on UK drought*. National Environment Research Council. Retrieved 28th September from https://ncas.ac.uk/national-centre-for-atmospheric-science-informs-environment-agency-report-on-uk-drought/

Ndong, M., Bird, D., Nguyen-Quang, T., de Boutray, M.-L., Zamyadi, A., Vinçon-Leite, B., Lemaire, B. J., Prévost, M., & Dorner, S. (2014). Estimating the risk of cyanobacterial occurrence using an index integrating meteorological factors: Application to drinking water production. *Water Research, 56*, 98-108. https://doi.org/10.1016/j.watres.2014.02.023

Neal, C., Bowes, M., Jarvie, H. P., Scholefield, P., Leeks, G., Neal, M., Rowland, P., Wickham, H., Harman, S., Armstrong, L., Sleep, D., Lawlor, A., & Davies, C. E. (2012). Lowland river water quality: a new UK data resource for process and environmental management analysis. *Hydrological Processes, 26*(6), 949-960. https://doi.org/10.1002/hyp.8344

Nimick, D. A., Gammons, C. H., & Parker, S. R. (2011). Diel biogeochemical processes and their effect on the aqueous chemistry of streams: A review. *Chemical Geology, 283*(1), 3-17. https://doi.org/10.1016/j.chemgeo.2010.08.017

Ockenden, M. C., Deasy, C. E., Benskin, C. M. H., Beven, K. J., Burke, S., Collins, A. L., Evans, R., Falloon, P. D., Forber, K. J., Hiscock, K. M., Hollaway, M. J., Kahana, R., Macleod, C. J. A., Reaney, S. M., Snell, M. A., Villamizar, M. L., Wearing, C., Withers, P. J. A., Zhou, J. G., & Haygarth, P. M. (2016). Changing climate and nutrient transfers: Evidence from high temporal resolution concentration-flow dynamics in headwater catchments. *Science of the Total Environment, 548-549*, 325-339. https://doi.org/10.1016/j.scitotenv.2015.12.086

Oh, H. S., Lee, C. S., Srivastava, A., Oh, H. M., & Ahn, C. Y. (2017). Effects of environmental factors on cyanobacterial production of odorous compounds: Geosmin and 2-methylisoborneol. *Journal of Microbiology and Biotechnology, 27*(7), 1316-1323. https://doi.org/10.4014/jmb.1702.02069

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C., & Weedon, J. (2024). *vegan:*

Community Ecology Package. In (Version R package version 2.6-6.1) CRAN. https://CRAN.R-project.org/package=vegan

Olsen, B. K., Chislock, M., & Wilson, A. (2016a). Eutrophication mediates a common off-flavor compound, 2-methylisoborneol, in a drinking water reservoir. *Water Research, 92*, 228-234. https://doi.org/10.1016/j.watres.2016.01.058

Olsen, B. K., Chislock, M. F., Rebelein, A., & Wilson, A. E. (2016b). Nutrient enrichment and vertical mixing mediate 2-methylisoborneol and geosmin concentrations in a drinking water reservoir. *Water Supply, 17*(2), 500-507. https://doi.org/10.2166/ws.2016.159

Otten, T. G., Graham, J. L., Harris, T. D., & Dreher, T. W. (2016). Elucidation of taste and odor-producing bacteria and toxigenic cyanobacteria in a Midwestern drinking water supply reservoir by shotgun metagenomic analysis. *Applied and Environmental Microbiology, 82*(17), 5410-5420. https://doi.org/10.1128/AEM.01334-16

Outram, F. N., Lloyd, C. E. M., Jonczyk, J., Benskin, C. M. H., Grant, F., Perks, M. T., Deasy, C., Burke, S. P., Collins, A. L., Freer, J., Haygarth, P. M., Hiscock, K. M., Johnes, P. J., & Lovett, A. L. (2014). High-frequency monitoring of nitrogen and phosphorus response in three rural catchments to the end of the 2011–2012 drought in England. *Hydrology and Earth System Sciences, 18*(9), 3429-3448. https://doi.org/10.5194/hess-18-3429-2014

Owen, G. J., Perks, M. T., Benskin, C. M. H., Wilkinson, M. E., Jonczyk, J., & Quinn, P. F. (2012). Monitoring agricultural diffuse pollution through a dense monitoring network in the River Eden Demonstration Test Catchment, Cumbria, UK. *Area, 44*(4), 443-453. https://doi.org/10.1111/j.1475-4762.2012.01107.x

Ozaki, K., Ohta, A., Iwata, C., Horikawa, A., Tsuji, K., Ito, E., Ikai, Y., & Harada, K.-i. (2008). Lysis of cyanobacteria with volatile organic compounds. *Chemosphere, 71*(8), 1531-1538. https://doi.org/10.1016/j.chemosphere.2007.11.052

Paerl, H. W., & Fulton, R. S. (2006). Ecology of Harmful Cyanobacteria. In E. Granéli & J. T. Turner (Eds.), *Ecology of Harmful Algae* (pp. 95-109). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-32210-8_8

Paerl, H. W., Havens, K. E., Xu, H., Zhu, G., McCarthy, M. J., Newell, S. E., Scott, J. T., Hall, N. S., Otten, T. G., & Qin, B. (2020). Mitigating eutrophication and toxic cyanobacterial blooms in large lakes: The evolution of a dual nutrient (N and P) reduction paradigm. *Hydrobiologia, 847*(21), 4359-4375. https://doi.org/10.1007/s10750-019-04087-y

Paerl, H. W., & Huisman, J. (2009). Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports, 1*(1), 27-37. https://doi.org/10.1111/j.1758-2229.2008.00004.x

Paerl, H. W., & Otten, T. G. (2013). Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology, 65*(4), 995-1010. https://doi.org/10.1007/s00248-012-0159-y

Paerl, H. W., & Paul, V. J. (2012). Climate change: Links to global expansion of harmful cyanobacteria. *Water Research, 46*(5), 1349-1363. https://doi.org/10.1016/j.watres.2011.08.002

Paerl, R., Huang, H., & Ehrlich, L. (2022). *Investigating the Microbial Culprits of Taste/Odor Issues in City of Durham Drinking Water Reservoir Lake Michie and Algicidal Mitigation Tactics*. North Carolina Water Resources Research Institute.

Painter, K. J., Venkiteswaran, J. J., & Baulch, H. M. (2023). Blooms and flows: Effects of variable hydrology and management on reservoir water quality. *Ecosphere, 14*(3), e4472. https://doi.org/10.1002/ecs2.4472

Palmer-Felgate, E. J., Jarvie, H. P., Williams, R. J., Mortimer, R. J. G., Loewenthal, M., & Neal, C. (2008). Phosphorus dynamics and productivity in a sewage-impacted lowland chalk stream. *Journal of Hydrology, 351*(1), 87-97. https://doi.org/10.1016/j.jhydrol.2007.11.036

Parinet, J., Rodriguez, M. J., & Sérodes, J. B. (2013). Modelling geosmin concentrations in three sources of raw water in Quebec, Canada. *Environmental Monitoring and Assessment, 185*(1), 95-111. https://doi.org/10.1007/s10661-012-2536-x

Park, H. K., Kwon, M. A., Lee, H. J., Oh, J., Lee, S. H., & Kim, I. S. (2018). Molecular verification of bloom-forming aphanizomenon flos-aquae and their secondary metabolites in the Nakdong river. *International Journal of Environmental Research and Public Health, 15*(8), 1-14. https://doi.org/10.3390/ijerph15081739

Parmar, A., Singh, N. K., Pandey, A., Gnansounou, E., & Madamwar, D. (2011). Cyanobacteria and microalgae: A positive prospect for biofuels. *Bioresource Technology, 102*(22), 10163-10172. https://doi.org/10.1016/j.biortech.2011.08.030

Pattanaik, B., & Lindberg, P. (2015). Terpenoids and their biosynthesis in cyanobacteria. *Life, 5*(1), 269-293. https://doi.org/10.3390/life5010269

Pellerin, B. A., Downing, B. D., Kendall, C., Dahlgren, R. A., Kraus, T. E. C., Saraceno, J., Spencer, R. G. M., & Bergamaschi, B. A. (2009). Assessing the sources and magnitude of diurnal nitrate variability in the San Joaquin River (California) with an in situ optical nitrate sensor and dual nitrate isotopes. *Freshwater Biology, 54*(2), 376-387. https://doi.org/10.1111/j.1365-2427.2008.02111.x

Perkins, R. G., Slavin, E. I., Andrade, T. M. C., Blenkinsopp, C., Pearson, P., Froggatt, T., Godwin, G., Parslow, J., Hurley, S., Luckwell, R., & Wain, D. J. (2019). Managing taste and odour metabolite production in drinking water reservoirs: The importance of ammonium as a key nutrient trigger. *Journal of Environmental Management, 244*(March), 276-284. https://doi.org/10.1016/j.jenvman.2019.04.123

Perks, M. T., Owen, G. J., Benskin, C. M. H., Jonczyk, J., Deasy, C., Burke, S., Reaney, S. M., & Haygarth, P. M. (2015). Dominant mechanisms for the delivery of fine sediment and phosphorus to fluvial networks draining grassland dominated headwater catchments. *Science of the Total Environment, 523*, 178-190. https://doi.org/10.1016/j.scitotenv.2015.03.008

Perri, K. A., Bellinger, B. J., Ashworth, M. P., & Manning, S. R. (2024). Environmental Factors Impacting the Development of Toxic Cyanobacterial Proliferations in a Central Texas Reservoir. *Toxins, 16*(2). https://doi.org/10.3390/toxins16020091

Persaud, A. D., Paterson, A. M., Dillon, P. J., Winter, J. G., Palmer, M., & Somers, K. M. (2015). Forecasting cyanobacteria dominance in Canadian temperate lakes. *Journal of Environmental Management, 151*, 343-352. https://doi.org/10.1016/j.jenvman.2015.01.009

Peter, A., Köster, O., Schildknecht, A., & von Gunten, U. (2009). Occurrence of dissolved and particle-bound taste and odor compounds in Swiss lake waters. *Water Research, 43*(8), 2191-2200. https://doi.org/10.1016/j.watres.2009.02.016

Peterson, H. G., Hrudey, S. E., Cantin, I. A., Perley, T. R., & Kenefick, S. L. (1995). Physiological toxicity, cell membrane damage and the release of dissolved organic carbon and geosmin by Aphanizomenon flos-aquae after exposure to water treatment chemicals. *Water Research, 29*(6), 1515-1523. https://doi.org/10.1016/0043-1354(94)00300-V

Pham, M.-L., Askarzadmohassel, E., & Brandl, M. (2023). Growth of freshwater cyanobacterium Aphanizomenon sp. ULC602 in different growing and nutrient conditions [Original Research]. *Frontiers in Microbiology, 14*. https://doi.org/10.3389/fmicb.2023.1220818

Poulíčková, A., Hašler, P., Lysáková, M., & Spears, B. (2008). The ecology of freshwater epipelic algae: an update. *Phycologia, 47*(5), 437-450. https://doi.org/10.2216/07-59.1

Qi, C., Fang, J., Wang, G., Huang, H., Wang, Z., Si, Z., & Zhang, L. (2020a). Characterization of odorants in contrasting ecotypes of Lake Taihu: algae-dominated versus macrophyte-dominated zones. *Environmental Science and Pollution Research, 27*(34), 42221-42229. https://doi.org/10.1007/s11356-020-07896-0

Qi, C., Zhang, L., Fang, J., Lei, B., Tang, X., Huang, H., Wang, Z., Si, Z., & Wang, G. (2020b). Benthic cyanobacterial detritus mats in lacustrine sediment: Characterization and odorant producing potential. *Environmental Pollution, 256*, 113453-113453. https://doi.org/10.1016/j.envpol.2019.113453

Quilbé, R., Serreau, C., Wicherek, S., Bernard, C., Thomas, Y., & Oudinet, J.-P. (2005). Nutrient transfer by runoff from sewage sludge amended soil under simulated rainfall. *Environmental Monitoring and Assessment, 100*(1), 177-190. https://doi.org/10.1007/s10661-005-4773-8

Rand, J. M., Nanko, M. O., Lykkegaard, M. B., Wain, D., King, W., Bryant, L. D., & Hunter, A. (2022). The human factor: Weather bias in manual lake water quality monitoring. *Limnology and Oceanography: Methods, 20*(5), 288-303. https://doi.org/10.1002/lom3.10488

Rangel, M., Martins, J. C. G., Garcia, A. N., Conserva, G. A. A., Costa-Neves, A., Sant, Anna, C. L., & De Carvalho, L. R. (2014). Analysis of the Toxicity and Histopathology Induced by the Oral Administration of Pseudanabaena galeata and Geitlerinema splendidum (Cyanobacteria) Extracts to Mice. *Marine Drugs, 12*(1), 508-524. https://doi.org/10.3390/md12010508

Raven, J. A. (2010). Cyanotoxins: A poison that frees phosphate. *Current Biology, 20*(19), R850-R852. https://doi.org/10.1016/j.cub.2010.08.012

Recknagel, F., Adrian, R., Köhler, J., & Cao, H. (2016). Threshold quantification and short-term forecasting of Anabaena, Aphanizomenon and Microcystis in the polymictic eutrophic Lake Müggelsee (Germany) by inferential modelling using the hybrid evolutionary algorithm HEA. *Hydrobiologia, 778*(1), 61-74. https://doi.org/10.1007/s10750-015-2442-7

Redfield, A. C., Ketchum, B. H., & Richards, F. A. (1963). The influence of organisms on the composition of seawater. In *The sea* (Vol. 2, pp. 26-77).

Reinl, K. L., Harris, T. D., Elfferich, I., Coker, A., Zhan, Q., Domis, L. N. D. S., Morales-Williams, A. M., Bhattacharya, R., Grossart, H.-P., North, R. L., & Sweetman, J. N. (2022). The role of organic nutrients in structuring freshwater phytoplankton communities in a rapidly changing world. *Water Research*, 118573. https://doi.org/10.1016/j.watres.2022.118573

Reynolds, C. S. (2006). *The Ecology of Phytoplankton*. Cambridge University Press. https://doi.org/10.1017/CBO9780511542145

Reynolds, K. N., Loecke, T. D., Burgin, A. J., Davis, C. A., Riveros-Iregui, D., Thomas, S. A., St. Clair, M. A., & Ward, A. S. (2016). Optimizing Sampling Strategies for Riverine Nitrate Using High-Frequency Data in Agricultural Watersheds. *Environmental science & technology, 50*(12), 6406-6414. https://doi.org/10.1021/acs.est.5b05423

Richardson, J., Feuchtmayr, H., Miller, C., Hunter, P. D., Maberly, S. C., & Carvalho, L. (2019). Response of cyanobacteria and phytoplankton abundance to warming, extreme rainfall events and nutrient enrichment. *Global Change Biology, 25*(10), 3365-3380. https://doi.org/10.1111/gcb.14701

Rider, Z., Percich, A., Hiripitiyage, Y., Harris, T. D., Sturm, B. S. M., Wilson, A. E., Pollock, E. D., Beaver, J. R., & Husic, A. (2024). Drivers of cyanotoxin and taste-and-odor compound presence within the benthic algae of human-disturbed rivers. *Water Research, 253*, 121357. https://doi.org/10.1016/j.watres.2024.121357

Roberts, D. W. (2023). *labdsv: Ordination and Multivariate Analysis for Ecology.* In (Version R package version 2.1-0) https://CRAN.R-project.org/package=labdsv

Robson, A. J., & Reed, D. W. (1999). Statistical procedures for flood frequency estimation. In *Flood Estimation Handbook* (Vol. 3). Centre for Ecology & Hydrology, Wallingford.

Rode, M., Wade, A. J., Cohen, M. J., Hensley, R. T., Bowes, M. J., Kirchner, J. W., Arhonditsis, G. B., Jordan, P., Kronvang, B., Halliday, S. J., Skeffington, R. A., Rozemeijer, J. C., Aubert, A. H., Rinke, K., & Jomaa, S. (2016). Sensors in the Stream: The High-Frequency Wave of

the Present. *Environmental Science and Technology, 50*(19), 10297-10307. https://doi.org/10.1021/acs.est.6b02155

Rozemeijer, J., Jordan, P., Hooijboer, A., Kronvang, B., Glendell, M., Hensley, R., Rinke, K., Stutter, M., Bieroza, M., Turner, R., Mellander, P. E., Thorburn, P., Cassidy, R., Appels, J., Ouwerkerk, K., & Rode, M. (2025). Best practice in high-frequency water quality monitoring for improved management and assessment; a novel decision workflow. *Environmental Monitoring and Assessment*, 197(4), 353. https://doi.org/10.1007/s10661-025-13795-z

Rummens, K., De Meester, L., & Souffreau, C. (2018). Inoculation history affects community composition in experimental freshwater bacterioplankton communities. *Environmental Microbiology, 20*(3), 1120-1133. https://doi.org/10.1111/1462-2920.14053

Ryan, J. A., & Ulrich, J. M. (2011). xts: Extensible time series. *R package version 0.8-2*.

Saadoun, I. M. K., Schrader, K. K., & Blevins, W. T. (2001). Environmental and nutritional factors affecting geosmin synthesis by Anabaena sp. *Water Research, 35*(5), 1209-1218. https://doi.org/10.1016/S0043-1354(00)00381-X

Salvia-Castellví, M., François Iffly, J., Vander Borght, P., & Hoffmann, L. (2005). Dissolved and particulate nutrient export from rural catchments: A case study from Luxembourg. *Science of the Total Environment, 344*(1), 51-65. https://doi.org/10.1016/j.scitotenv.2005.02.005

Santschi, P., Höhener, P., Benoit, G., & Buchholtz-ten Brink, M. (1990). Chemical processes at the sediment-water interface. *Marine Chemistry, 30*, 269-315. https://doi.org/10.1016/0304-4203(90)90076-O

Schindler, D. W., Hecky, R. E., Findlay, D. L., Stainton, M. P., Parker, B. R., Paterson, M. J., Beaty, K. G., Lyng, M., & Kasian, S. E. M. (2008). Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proceedings of the National Academy of Sciences of the United States of America, 105*(32), 11254-11258. https://doi.org/10.1073/pnas.0805108105

Schmutz, S., & Sendzimir, J. (2018). *Riverine Ecosystem Management: Science for Governing Towards a Sustainable Future*. Springer Nature.

Scholefield, D., Le Goff, T., Braven, J., Ebdon, L., Long, T., & Butler, M. (2005). Concerted diurnal patterns in riverine nutrient concentrations and physical conditions. *Science of the Total Environment, 344*(1), 201-210. https://doi.org/10.1016/j.scitotenv.2005.02.014

Sefbom, J., Sassenhagen, I., Rengefors, K., & Godhe, A. (2015). Priority effects in a planktonic bloom-forming marine diatom. *Biology letters, 11*(5), 20150184.

Seitzinger, S. P. (1988). Denitrification in freshwater and coastal marine ecosystems: Ecological and geochemical significance. *Limnology and Oceanography, 33*(4part2), 702-724. https://doi.org/10.4319/lo.1988.33.4part2.0702

Shen, Q., Shimizu, K., Miao, H., Tsukino, S., Utsumi, M., Lei, Z., Zhang, Z., Nishimura, O., Asada, Y., Fujimoto, N., Takanashi, H., & Akiba, M. (2020). Effects of elevated nitrogen on the growth and geosmin productivity of Dolichospermum smithii. *Environmental Science and Pollution Research, 28*(1), 177-184. https://doi.org/10.1007/s11356-020-10429-4

Sinang, S. C., Reichwaldt, E. S., & Ghadouani, A. (2015). Local nutrient regimes determine site-specific environmental triggers of cyanobacterial and microcystin variability in urban lakes. *Hydrology and Earth System Sciences, 19*(5), 2179-2195. https://doi.org/10.5194/hess-19-2179-2015

Skeffington, R. A., Halliday, S. J., Wade, A. J., Bowes, M. J., & Loewenthal, M. (2015). Using high-frequency water quality data to assess sampling strategies for the EU Water Framework Directive. *Hydrology and Earth System Sciences, 19*(5), 2491-2504. https://doi.org/10.5194/hess-19-2491-2015

Smith, V. H. (1983). Low Nitrogen to Phosphorus Ratios Favor Dominance by Blue-Green Algae in Lake Phytoplankton. *Science, 221*(4611), 669-671. https://doi.org/10.1126/science.221.4611.669

Søndergaard, M., Jensen, J. P., & Jeppesen, E. (2003). Role of sediment and internal loading of phosphorus in shallow lakes. *Hydrobiologia, 506*(1), 135-145. https://doi.org/10.1023/B:HYDR.0000008611.12704.dd

Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods -- a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics, 23*, 1164-1167.

Su, M., Gaget, V., Giglio, S., Burch, M., An, W., & Yang, M. (2013). Establishment of quantitative PCR methods for the quantification of geosmin-producing potential and Anabaena sp. in freshwater systems. *Water Research, 47*(10), 3444-3454. https://doi.org/10.1016/j.watres.2013.03.043

Su, M., Jia, D., Yu, J., Vogt, R. D., Wang, J., An, W., & Yang, M. (2017). Reducing production of taste and odor by deep-living cyanobacteria in drinking water reservoirs by regulation of water level. *Science of the Total Environment, 574*, 1477-1483. https://doi.org/10.1016/j.scitotenv.2016.08.134

Trommer, G., Poxleitner, M., & Stibor, H. (2020). Responses of lake phytoplankton communities to changing inorganic nitrogen supply forms. *Aquatic Sciences, 82*(2), 22. https://doi.org/10.1007/s00027-020-0696-2

Tsao, H. W., Michinaka, A., Yen, H. K., Giglio, S., Hobson, P., Monis, P., & Lin, T. F. (2014). Monitoring of geosmin producing Anabaena circinalis using quantitative PCR. *Water Research, 49*, 416-425. https://doi.org/10.1016/j.watres.2013.10.028

Turner, B. L., Papházy, M. J., Haygarth, P. M., & Mckelvie, I. D. (2002). Inositol phosphates in the environment. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 357*(1420), 449-469.

Urbanek, S., & Horner, J. (2023). *Cairo: R Graphics Device using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11 and Win32) Output.* In (Version R package version 1.6-2) https://CRAN.R-project.org/package=Cairo

Üveges, V., Tapolczai, K., Krienitz, L., & Padisák, J. (2012). Photosynthetic characteristics and physiological plasticity of an Aphanizomenon flos-aquae (Cyanobacteria, Nostocaceae) winter bloom in a deep oligo-mesotrophic lake (Lake Stechlin, Germany). In Salmaso N., Naselli-Flores L., Cerasino L., Flaim G., Tolotti M., & P. J. (Eds.), *Phytoplankton responses to human impacts at different scales* (Vol. 221, pp. 263-272). Springer. https://doi.org/10.1007/978-94-007-5790-5_20

Vadeboncoeur, Y., Moore, M. V., Stewart, S. D., Chandra, S., Atkins, K. S., Baron, J. S., Bouma-Gregson, K., Brothers, S., Francoeur, S. N., Genzoli, L., Higgins, S. N., Hilt, S., Katona, L. R., Kelly, D., Oleksy, I. A., Ozersky, T., Power, M. E., Roberts, D., Smits, A. P., Timoshkin, O., Tromboni, F., Zanden, M. J. V., Volkova, E. A., Waters, S., Wood, S. A., & Yamamuro, M. (2021). Blue Waters, Green Bottoms: Benthic Filamentous Algal Blooms Are an Emerging Threat to Clear Lakes Worldwide. *BioScience, 71*(10), 1011-1027. https://doi.org/10.1093/biosci/biab049

van der Grinten, E., Janssen, A. P. H. M., de Mutsert, K., Barranguet, C., & Admiraal, W. (2005). Temperature- and Light-Dependent Performance of the Cyanobacterium Leptolyngbya Foveolarum and the Diatom Nitzschia Perminuta in Mixed Biofilms. *Hydrobiologia, 548*(1), 267-278. https://doi.org/10.1007/s10750-005-5324-6

Van Gremberghe, I., Vanormelingen, P., Van der Gucht, K., Souffreau, C., Vyverman, W., & De Meester, L. (2009). Priority effects in experimental populations of the cyanobacterium Microcystis. *Environmental Microbiology, 11*(10), 2564-2573. https://doi.org/10.1111/j.1462-2920.2009.01981.x

Vidal, M., López, A., Santoalla, M. C., & Valles, V. (2000). Factor analysis for the study of water resources contamination due to the use of livestock slurries as fertilizer. *Agricultural Water Management, 45*(1), 1-15. https://doi.org/10.1016/S0378-3774(99)00073-6

Vrede, T., Ballantyne, A., Mille-Lindblom, C., Algesten, G., Gudasz, C., Lindahl, S., & Brunberg, A. K. (2009). Effects of N:P loading ratios on phytoplankton community composition, primary production and N fixation in a eutrophic lake. *Freshwater Biology, 54*(2), 331-344. https://doi.org/10.1111/j.1365-2427.2008.02118.x

Wade, A. J., Palmer-Felgate, E. J., Halliday, S. J., Skeffington, R. A., Loewenthal, M., Jarvie, H. P., Bowes, M. J., Greenway, G. M., Haswell, S. J., Bell, I. M., Joly, E., Fallatah, A., Neal, C., Williams, R. J., Gozzard, E., & Newman, J. R. (2012). Hydrochemical processes in lowland rivers: insights from in situ, high-resolution monitoring. *Hydrol. Earth Syst. Sci., 16*(11), 4323-4342. https://doi.org/10.5194/hess-16-4323-2012

Wagner, C., & Adrian, R. (2009). Cyanobacteria dominance: Quantifying the effects of climate change. *Limnology and Oceanography, 54*(6part2), 2460-2468. https://doi.org/10.4319/lo.2009.54.6_part_2.2460

Walsby, A. E. (1994). Gas vesicles. *Microbiological Reviews, 58*(1), 94-144. https://doi.org/10.1146/annurev.pp.26.060175.002235

Wang, M., Xu, X., Wu, Z., Zhang, X., Sun, P., Wen, Y., Wang, Z., Lu, X., Zhang, W., Wang, X., & Tong, Y. (2019a). Seasonal Pattern of Nutrient Limitation in a Eutrophic Lake and Quantitative Analysis of the Impacts from Internal Nutrient Cycling. *Environmental science & technology, 53*(23), 13675-13686. https://doi.org/10.1021/acs.est.9b04266

Wang, Z., Huang, S., & Li, D. (2019b). Decomposition of cyanobacterial bloom contributes to the formation and distribution of iron-bound phosphorus (Fe-P): Insight for cycling mechanism of internal phosphorus loading. *Science of the Total Environment, 652*, 696-708. https://doi.org/10.1016/j.scitotenv.2018.10.260

Wang, Z., & Li, R. (2015). Effects of light and temperature on the odor production of 2-methylisoborneol-producing Pseudanabaena sp. and geosmin-producing Anabaena ucrainica (cyanobacteria). *Biochemical Systematics and Ecology, 58*(7), 219-226. https://doi.org/10.1016/j.bse.2014.12.013

Wang, Z., Song, G., Li, Y., Yu, G., Hou, X., Gan, Z., & Li, R. (2019c). The diversity, origin, and evolutionary analysis of geosmin synthase gene in cyanobacteria. *Science of the Total Environment, 689*(7), 789-796. https://doi.org/10.1016/j.scitotenv.2019.06.468

Wang, Z., Song, G., Shao, J., Tan, W., Li, Y., & Li, R. (2016). Establishment and field applications of real-time PCR methods for the quantification of potential MIB-producing cyanobacteria in aquatic systems. *Journal of Applied Phycology, 28*(1), 325-333. https://doi.org/10.1007/s10811-015-0529-1

Wang, Z., Xiao, P., Song, G., Li, Y., & Li, R. (2015). Isolation and characterization of a new reported cyanobacterium Leptolyngbya bijugata coproducing odorous geosmin and 2-methylisoborneol. *Environmental Science and Pollution Research, 22*(16), 12133-12140. https://doi.org/10.1007/s11356-015-4470-2

Watson, S. B. (2003). Cyanobacterial and eukaryotic algal odour compounds: signals or by-products? A review of their biological activity. *Phycologia, 42*(4), 332-350. https://doi.org/10.2216/i0031-8884-42-4-332.1

Watson, S. B., & Jüttner, F. (2019). Biological production of taste and odour compounds. In T.-F. Lin, S. Watson, A. Dietrich, & I. H. M. Suffet (Eds.), *Taste and Odour in Source and Drinking Water: Causes, Controls, and Consequences* (pp. 63-112). IWA Publishing. https://doi.org/10.2166/9781780406664_0063

Watson, S. B., Monis, P., Baker, P., & Giglio, S. (2016). Biochemistry and genetics of taste- and odor-producing cyanobacteria. *Harmful Algae, 54*, 112-127. https://doi.org/10.1016/j.hal.2015.11.008

Watson, S. B., & Ridal, J. (2004). Periphyton: a primary source of widespread and severe taste and odour. *Water Science and Technology, 49*(9), 33-39. https://doi.org/10.2166/wst.2004.0527

Watson, S. E., Taylor, C. H., Bell, V., Hooper, A. S., Bellamy, T. R., Kille, P., & Perkins, R. G. (2024). Utilising eDNA Methods and Interactive Data Dashboards for Managing Sustainable Drinking Water. *Sustainability, 16*(5). https://doi.org/10.3390/su16052043

Wentzky, V. C., Frassl, M. A., Rinke, K., & Boehrer, B. (2019). Metalimnetic oxygen minimum and the presence of Planktothrix rubescens in a low-nutrient drinking water reservoir. *Water Research, 148*, 208-218. https://doi.org/10.1016/j.watres.2018.10.047

Werner, C. M., Stuble, K. L., Groves, A. M., & Young, T. P. (2020). Year effects: Interannual variation as a driver of community assembly dynamics. *Ecology, 101*(9), e03104. https://doi.org/10.1002/ecy.3104

Wert, E. C., Korak, J. A., Trenholm, R. A., & Rosario-Ortiz, F. L. (2014). Effect of oxidant exposure on the release of intracellular microcystin, MIB, and geosmin from three cyanobacteria species. *Water Research, 52*, 251-259. https://doi.org/10.1016/j.watres.2013.11.001

Westerhoff, P., Rodriguez-Hernandez, M., Baker, L., & Sommerfeld, M. (2005). Seasonal occurrence and degradation of 2-methylisoborneol in water supply reservoirs. *Water Research, 39*(20), 4899-4912. https://doi.org/10.1016/j.watres.2005.06.038

Wetzel, R. G. (2001). *Limnology: Lake and River Ecosystems*. Academic Press.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. In (Version R package version 1.1.4) CRAN. https://CRAN.R-project.org/package=dplyr

Williams, M. R., King, K. W., Macrae, M. L., Ford, W., Van Esbroeck, C., Brunke, R. I., English, M. C., & Schiff, S. L. (2015). Uncertainty in nutrient loads from tile-drained landscapes: Effect of sampling frequency, calculation algorithm, and compositing strategy. *Journal of Hydrology, 530*, 306-316. https://doi.org/10.1016/j.jhydrol.2015.09.060

Williams, R. J., White, C., Harrow, M. L., & Neal, C. (2000). Temporal and small-scale spatial variations of dissolved oxygen in the Rivers Thames, Pang and Kennet, UK. *Science of the Total Environment, 251-252*, 497-510. https://doi.org/10.1016/S0048-9697(00)00401-0

Winston, B., Hausmann, S., Scott, J. T., & Morgan, R. (2014). The influence of rainfall on taste and odor production in a south-central USA reservoir. *Freshwater Science, 33*(3), 755-764. https://doi.org/10.1086/677176

Winter, J. G., Desellas, A. M., Fletcher, R., Heintsch, L., Morley, A., Nakamoto, L., & Utsumi, K. (2011). Algal blooms in Ontario, Canada: Increases in reports since 1994. *Lake and Reservoir Management, 27*(2), 107-114. https://doi.org/10.1080/07438141.2011.557765

Withers, P. J. A., & Jarvie, H. P. (2008). Delivery and cycling of phosphorus in rivers: A review. *Science of the Total Environment, 400*(1), 379-395. https://doi.org/10.1016/j.scitotenv.2008.08.002

Wood, S. A., Borges, H., Puddick, J., Biessy, L., Atalah, J., Hawes, I., Dietrich, D. R., & Hamilton, D. P. (2017). Contrasting cyanobacterial communities and microcystin concentrations in summers with extreme weather events: insights into potential effects of climate change. *Hydrobiologia, 785*(1), 71-89. https://doi.org/10.1007/s10750-016-2904-6

Wood, S. N. (2017). *Generalized additive models: an introduction with R, Second Edition*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association, 111*(516), 1548-1563. https://doi.org/10.1080/01621459.2016.1180986

Worrall, F., Gibson, H. S., & Burt, T. P. (2008). Production vs. solubility in controlling runoff of DOC from peat soils – The use of an event analysis. *Journal of Hydrology, 358*(1), 84-95. https://doi.org/10.1016/j.jhydrol.2008.05.037

Wu, J.-T., Ma, P.-I., & Chou, T.-L. (1991). Variation of geosmin content in Anabaena cells and its relation to nitrogen utilization. *Archives of Microbiology, 157*(1), 66-69. https://doi.org/10.1007/BF00245337

Wu, T., Zhu, G., Wang, Z., Zhu, M., & Xu, H. (2022). Seasonal dynamics of odor compounds concentration driven by phytoplankton succession in a subtropical drinking water reservoir, southeast China. *Journal of Hazardous Materials, 425*, 128056. https://doi.org/10.1016/j.jhazmat.2021.128056

Xiao, M., Wei, Y., Zheng, D., Shi, L., Huang, C., Wang, Y., & Zhu, G. (2024). Outbreak of 2-methylisoborneol in a drinking water reservoir attributed to the cyanobacterium Pseudanabaena sp. *Journal of Oceanology and Limnology*. https://doi.org/10.1007/s00343-024-4034-8

Xuwei, D., Min, Q., ren, R., Jiarui, L., Xiaoxue, S., Ping, X., & Jun, C. (2019). The relationships between odors and environmental factors at bloom and non-bloom area in Lake Taihu, China. *Chemosphere, 218*, 569-576. https://doi.org/10.1016/j.chemosphere.2018.11.121

Yamamoto, Y., & Nakahara, H. (2005). The formation and degradation of cyanobacterium Aphanizomenon flos-aquae blooms: the importance of pH, water temperature, and day length. *Limnology, 6*(1), 1-6. https://doi.org/10.1007/s10201-004-0138-1

Yamamoto, Y., & Nakahara, H. (2009). Life cycle of cyanobacterium Aphanizomenon flos-aquae. *Taiwania, 54*(2), 113-117.

Yang, J., Holbach, A., Wilhelms, A., Qin, Y., Zheng, B., Zou, H., Qin, B., Zhu, G., & Norra, S. (2019). Highly time-resolved analysis of seasonal water dynamics and algal kinetics based on in-situ multi-sensor-system monitoring data in Lake Taihu, China. *Science of the Total Environment, 660*, 329-339. https://doi.org/10.1016/j.scitotenv.2019.01.044

Yao, L., Huang, Y., Chen, L., & He, Y. (2022). Potential influence of overwintering benthic algae on water quality. *Journal of Environmental Sciences, 117*, 58-70. https://doi.org/10.1016/j.jes.2022.01.021

Yao, X., Ding, R., Zhou, Y., Wang, Z., Liu, Y., Fu, D., & Zhang, L. (2023). How internal nutrient loading forms in shallow lakes: Insights from benthic organic matter mineralization. *Water Research, 245*, 120544. https://doi.org/10.1016/j.watres.2023.120544

Yates, C. A., & Johnes, P. J. (2013). Nitrogen speciation and phosphorus fractionation dynamics in a lowland Chalk catchment. *Science of the Total Environment, 444*, 466-479. https://doi.org/10.1016/j.scitotenv.2012.12.002

Yates, C. A., Johnes, P. J., Brailsford, F. L., Evans, C. D., Evershed, R. P., Glanville, H. C., Jones, D. L., Lloyd, C. E. M., Marshall, M. R., & Owen, A. T. (2023). Determining patterns in the composition of dissolved organic matter in fresh waters according to land use and management. *Biogeochemistry, 164*(1), 143-162. https://doi.org/10.1007/s10533-022-00964-2

Yates, C. A., Johnes, P. J., Owen, A. T., Brailsford, F. L., Glanville, H. C., Evans, C. D., Marshall, M. R., Jones, D. L., Lloyd, C. E. M., Jickells, T., & Evershed, R. P. (2019a). Variation in dissolved organic matter (DOM) stoichiometry in U.K. freshwaters: Assessing the influence of land cover and soil C:N ratio on DOM composition. *Limnology and Oceanography, 64*(6), 2328-2340. https://doi.org/10.1002/lno.11186

Yates, C. A., Johnes, P. J., & Spencer, R. G. M. (2019b). Characterisation of treated effluent from four commonly employed wastewater treatment facilities: A UK case study. *Journal of Environmental Management, 232*, 919-927. https://doi.org/10.1016/j.jenvman.2018.12.006

Yen, H., Lin, T., Tseng, I., Tung, S., & Hsu, M. (2007). Correlating 2-MIB and microcystin concentrations with environmental parameters in two reservoirs in south Taiwan. *Water Science and Technology, 55*(5), 33-41. https://doi.org/10.2166/wst.2007.159

Yindong, T., Xiwen, X., Miao, Q., Jingjing, S., Yiyan, Z., Wei, Z., Mengzhu, W., Xuejun, W., & Yang, Z. (2021). Lake warming intensifies the seasonal pattern of internal nutrient cycling in the eutrophic lake and potential impacts on algal blooms. *Water Research, 188*, 116570. https://doi.org/10.1016/j.watres.2020.116570

Yue, Z., Chen, Y., Wu, Z., Cheng, X., Bao, Z., Deng, X., Shen, H., Liu, J., Xie, P., & Chen, J. (2024). Thermal stratification controls taste and odour compounds by regulating the phytoplankton community in a large subtropical water source reservoir (Xin'anjiang Reservoir). *Journal of Hazardous Materials, 466*, 133539. https://doi.org/10.1016/j.jhazmat.2024.133539

Zamyadi, A., Henderson, R. K., Stuetz, R., Newcombe, G., Newtown, K., & Gladman, B. (2016). Cyanobacterial management in full-scale water treatment and recycling processes: reactive dosing following intensive monitoring. *Environmental Science: Water Research & Technology, 2*(2), 362-375. https://doi.org/10.1039/C5EW00269A

Zeileis, A., & Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software, 14*(6), 1-27. https://doi.org/10.18637/jss.v014.i06

Zhang, M., Duan, H., Shi, X., Yu, Y., & Kong, F. (2012). Contributions of meteorology to the phenology of cyanobacterial blooms: Implications for future climate change. *Water Research, 46*(2), 442-452. https://doi.org/10.1016/j.watres.2011.11.013

Zhang, T., Li, L., Song, L., & Chen, W. (2009). Effects of temperature and light on the growth and geosmin production of Lyngbya kuetzingii (Cyanophyta). *Journal of Applied Phycology, 21*(3), 279-285. https://doi.org/10.1007/s10811-008-9363-z

Zhang, T., Zheng, L., Li, L., & Song, L. (2016). 2-Methylisoborneol production characteristics of Pseudanabaena sp. FACHB 1277 isolated from Xionghe Reservoir, China. *Journal of Applied Phycology, 28*(6), 3353-3362. https://doi.org/10.1007/s10811-016-0864-x

Zhu, M., Zhu, G., Zhao, L., Yao, X., Zhang, Y., Gao, G., & Qin, B. (2013). Influence of algal bloom degradation on nutrient release at the sediment–water interface in Lake Taihu, China. *Environmental Science and Pollution Research, 20*(3), 1803-1811. https://doi.org/10.1007/s11356-012-1084-9

Zimba, P. V., Dionigi, C. P., & Millie, D. F. (1999). Evaluating the relationship between photopigment synthesis and 2-methylisoborneol accumulation in cyanobacteria. *Journal of Phycology, 35*, 1422-1429. https://doi.org/10.1046/j.1529-8817.1999.3561422.x

List of Appendices

Appendix A – Supplementary material specific to Chapter 1

Appendix B – Supplementary material specific to Chapter 3

Appendix C – Supplementary material specific to Chapter 4

Appendix A

Table S.1. Summary of mesocosm and laboratory studies that investigated the effect of nutrient concentrations and ratios on geosmin and/or 2-MIB concentrations.

Experiment type	Time	Variable parameters	Manipulation	T&O metabolite	Result	Reference
Enclosures, 3000 L limnocorrals open to atmosphere, sealed at bottom. Polymyctic and mesotrophic drinking water reservoir located in the southeastern United States (Alabama) with 8 m depth	7 weeks (50 days) 28th October to 16th December 2013	Nutrients TN, TP (NO_3^- and PO_4^{3-}) & TN:TP ratios	The experimental layout consisted of three TN levels (300 ug/L (low), 1000 ug/L (medium), and 3000 ug/L (high)) with four TN:TP (2:1, 10:1, 33:1, and 90:1; by mass) manipulations. Nutrient additions were done with $NaNO_3^-$ and/or $NaH_2PO_4^{3-}$. Enclosures were fertilized weekly at a rate of 10% of the target TN and TP for each treatment to maintain TN and TN:TP over time	Geosmin and 2-MIB, whole water sample	They observed over a 700% increase in 2-MIB when both nitrogen and phosphorus were added, while adding only one had no significant effect. 2-MIB correlated more with diatom biovolume than with cyanobacteria. Diatoms increased in abundance throughout the experiment, while cyanobacteria declined in all treatments.	Olsen et al. (2016a)
Enclosures, 3800 L limnocorrals (diameter = 1.18 m, depth = 3.5 m), open to atmosphere, sealed at bottom. Polymyctic and mesotrophic drinking water reservoir	28 days (24th July 2014 to 21st August 2014)	Nutrients TN, TP (NO_3^- and PO_4^{3-}) & Vertical mixing	Four treatments were: (treatment A) control (ambient nutrients; no mixing), (treatment B) mixing only (ambient nutrients), (treatment C) nutrient addition only (no mixing), and (treatment D) mixing and nutrient addition. Target TN (1,000 µg/L) and TP (100 µg/L) was used in nutrient addition enclosures and spikes were	Geosmin and 2-MIB, whole water sample	2-MIB and geosmin increases occurred at different times during the experiment. Fertilization with phosphorus and nitrogen quickly boosted geosmin regardless of mixing, which was linked to cyanobacterial biovolume, but geosmin levels dropped below 20 ng/L by the end of the experiment.	Olsen et al. (2016b)

located in the southeastern United States (Alabama) with 8 m depth			done with $NaNO_3^-$ and $NaH_2PO_4^{3-}$.		In contrast, 2-MIB started low and peaked later, influenced by fertilization and mixing, and was also associated with cyanobacteria.	
Enclosures, 7.5 m deep, 1m^2 limnocorrals, open to atmosphere, sealed at bottom. Polymyctic and mesotrophic drinking water reservoir located in the southeastern United States (Alabama) with 8 m depth	3 weeks (26th July 2015 to 20th August 2015)	Depth and 3 fertilizer treatments (none, nitrate based and urea based)	Three fertilization treatments; (1) control (no nutrients), (2) nitrate-based, and (3) urea-based fertilizer) on chlorophyll *a* and phytoplankton biovolume as well as 2-MIB and geosmin, measured at four depths: 1.5, 3, 5, and 7 m across 3 weeks. "Nitrate" and "urea" treatment enclosures were fertilized with sodium phosphate to increase TP to 100 µgl−1 and with either sodium nitrate or urea to increase TN to 1000 µgl−1.	Geosmin and 2-MIB, whole water sample	The data suggest that the positive response of phytoplankton (e.g., cyanobacteria, such as Oscillatoria species) to the fertilization treatments is likely responsible for increased 2-MIB, while geosmin concentrations may be a function of actinobacteria-mediated decomposition in the hypolimnion.	Chislock et al. (2021)
18 microcosms under controlled conditions, using natural biofilm suspension form Ter river (NE, Spain) and flow through flumes.	21 days	Nutrient ratios (DIN:SRP) and nutrient concentrations (high and low)	Six treatments were set crossing three DIN:SRP ratios (A=4:1, B=16:1, C=64:1) with two nutrient concentrations (low and high).	Geosmin, intracellular and extracellular concentrations	They found that a low DIN:SRP ratio and a relatively high overall nutrient concentration (12 ug Ammonium-N, 110 ug Nitrate-N, 4 ug Phosphate-P)	Espinosa et al. (2021)

10 outdoor 3 m long flumes with a continuous water supply from the Ter river (Catalonia, NE Spain)	February–April, 2019	Light and water flow	Two light intensities were established: natural light (515 ± 57 µmol photons m^{-2} s^{-1}) and light reduced to 80% (107 ± 38 µmol photons m^{-2} s^{-1}), combined with five gradual water flows from 0.09 to 1.10 L/s (F1 = 0.09 L/s; F2 = 0.18 L/s; F3 = 0.36 L/s; F4 = 0.72 L/s; and F5 = 1.10 L/s).	Geosmin, whole water sample	Benthic cyanobacteria produced the highest geosmin concentrations in low light conditions and low water flow.	Espinosa et al. (2020)
Culture experiments with 3 liter flasks that contained 900 ml of the culture of geosmin-producing Anabaena sp. (isolated from Lake Ogletree 1991)	25 days	Light, temperature and nutrient ratios (NO_3^--N, NH_4^+-N and PO_4^{3-}-P)	For temperature studies (stable 17 µmol photons m^{-2} s^{-1}) they used 15, 20, 25 and 30 degrees C. For studies on light intensity effects, cultures (stable 20 degrees C) had 5, 17, 27 and 42 µmol photons m^{-2} s^{-1} incident light. Nutritional studies (stable 17 µmol photons m^{-2} s^{-1} and 20 degrees C) were spiked with varying concentrations of NH_4^+Cl-N (3, 26, 105, 328, 472), $NaNO_3^-$-N (2, 25, 124, 247 and 494 ug/L) or HPO_4^{3-}-P (2, 20, 59, 118, 235, 470, 941 ug/L)	Geosmin, whole water sample	In laboratory conditions with *Anabaena* sp., geosmin and chlorophyll *a* (Chl-a) concentrations increased with higher light intensity at 20°C. Beyond that, more light favored geosmin production over Chl-a, while higher temperatures promoted Chl-a instead of geosmin. Geosmin production was also linked to high Ammonium-N, low Nitrate-N, and low N:P ratios, which enhanced both growth and geosmin production.	Saadoun et al. (2001)
Culture experiments in 50 ml flasks containing 30 ml of culture with Lyngbya	42 days	Light and temperature	For the temperature treatment, the cultures were incubated at an illumination of 20 µmol photons m^{-2} s^{-1} under temperatures of 10, 25 and 35°C. For the light intensity treatment, the cultures were	Geosmin, whole water sample	They found that a low temperature together with a low light intensity stimulated the production of geosmin by *Lyngbya kuetzingii*, probably because Chl-a production	Zhang et al. (2009)

			incubated at 25°C under white-fluorescent lamps at light intensities of 0, 10, 20 and 75 µmol photons m^{-2} s^{-1}.		and metabolism were inhibited.	
kuetzingii (UTEX-1547)						
Culture experiments in 250 mL flasks containing a mixed culture of 2-MIB-producing Pseudanabaena sp. dqh15 and geosmin-producing Anabaena ucrainica CHAB 2155	16 days	Light and temperature	Cultures were incubated at 10 °C, 25°C, and 35 °C under 30 µmol photons m^{-2} s^{-1} (12:12 light/dark) for the temperature treatment. Another series of cultures was incubated at 10, 30, and 60 µmol photons m^{-2} s^{-1} under 25 °C for the light treatment.	Geosmin and 2-MIB, whole water sample	They found that growth-inhibiting temperatures and light intensities might stimulate cyanobacteria to produce geosmin or 2-MIB. The maximum geosmin production by *Anabaena ucrainica* was found at 10 °C and 60 µmol photons m^{-2} s^{-1} and the maximum 2-MIB production by *Pseudanabaena* sp. at 35 °C and 60 µmol photons m^{-2} s^{-1}.	Wang and Li (2015)
A 96-well plate PhotoBiobox system with 0.2 ml cultures of Anabaena sp. NIER, Anabaena sp. Chusori, and Anabaena sp. FACHB-1384	60 hours	Light and temperature	A whole range of temperatures and light intensities was tested (20-200 µmol photons m^{-2} s^{-1} and 16-28 degrees C)	Growth of cyanobacteria (Optical Density)	In this study, geosmin-producing Anabaena strains were better adapted to lower temperature (<20 °C) but were sensitive to high light intensity. However, 2-MIB-producing Anabaena grew well at a higher light intensity (>100 µmol photons m^{-2} s^{-1}) but was sensitive to low temperature.	Oh et al. (2017)

250 ml flasks with 160 ml cultures of Anabaena sp. NIER, Anabaena sp. Chusori	12 days	Nitrate concentrations	A range of nitrogen concentrations (Stable Light intensity of 50 μmol photons m^{-2} s^{-1} and temperature of 25o C) as follows: N 250 (original BG11 nitrate condition, 250 mg/l), N 2.5 (BG11 100× diluted nitrate condition, 2.5 mg/l), and N- (nitrate-free condition).	Geosmin, intracellular and extracellular concentrations	Despite the decrease in cell number, a high geosmin concentration was observed even after 8 days of cultivation in high nitrogen concentration. The geosmin amount per cell, however, was not affected by nitrate concentrations in batch cultivation	Oh et al. (2017)
Chemostat with 1 Liter flasks with 800 ml cultures of Anabaena sp. NIER and 2-MIB-producer Planktothrix sp. FACHB-1374	Until steady state was reached	Phosphorus limitation	Chemostat reactor dilution rates (0.10 d-1, 0.19 d-1, 0.25 d-1, 0.33 d-1) with stable 25 degrees C and 80 μmol photons m^{-2} s^{-1}	Geosmin and 2-MIB, intracellular and extracellular concentrations	The geosmin amount per cell decreased at lower dilution rates (higher P-limitation), but 2-MIB per cell was not affected by P-limitation	Oh et al. (2017)
500 ml flasks with 200 ml mixed culture of Dolichospermum smithii NIES-824	28 days	Nitrate concentrations	Elevated N loading with normal TN/TP ratio (13:1, by mass) by adding twice the amount of nitrate-N (Ca(NO$_3^-$)2·4H$_2$O and KNO$_3^-$) (Stable CT medium with TN/TP ratio of 8.1)	Geosmin, intracellular and extracellular concentrations	Elevated nitrogen levels increased Chl-a production but reduced geosmin synthesis in *D. smithii* NIES-824, suggesting a competitive relationship between Chl-a and geosmin production. Over 90% of geosmin remained intracellular during 28 days of cultivation. Geosmin synthase gene (*geo*A) expression was constant but decreased with higher	Shen et al. (2020)

					nitrogen concentrations during the exponential growth phase. The decline in *geo*A expression during the cell decline phase indicates a close link between *geo*A transcription, cell activity, and isoprenoid production.	

Appendix B
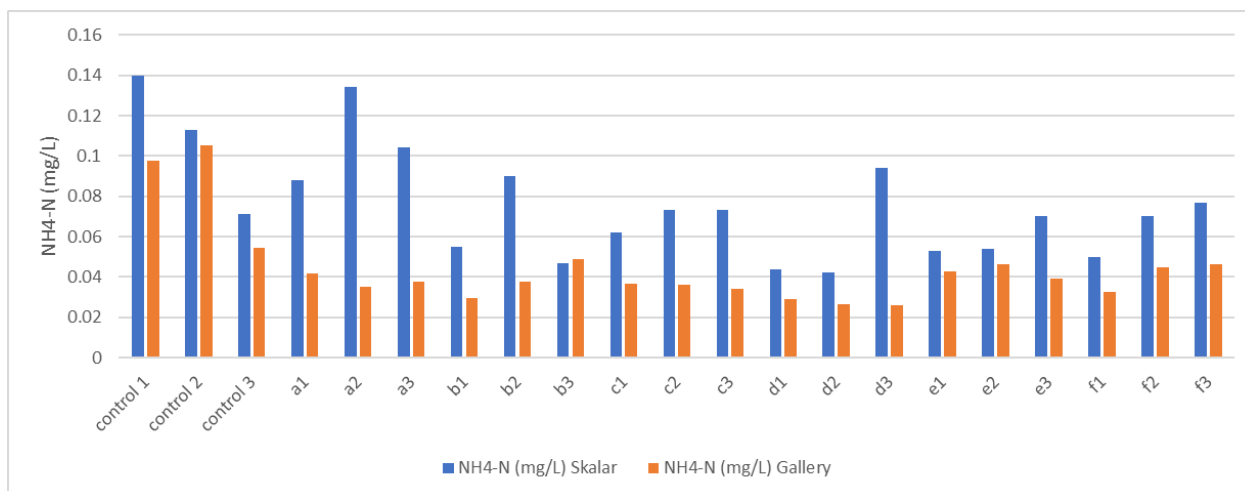
Figure S.1. Difference between Skalar San++ and Gallery^{TM} instrument for measuring NH_4^+-N (mg/L).

Table S.2. Comparison of different treatments of water samples in nutrient analysis

Device	Sample	NH_4^+-N (mg/L)	TOxN (mg/L)	NO_2^--N (mg/L)	NO_3^--N (mg/L)	SRP (mg/L)
Gallery	WR1 freezer	0.1328	0.1497	0.0129	0.1368	0.0183
Skalar	WR1 fridge	0.154	0.199	0.006	0.193	0
Skalar	WR1 freezer	0.123	0.217	0.005	0.212	0
Gallery	WR2 freezer	0.0441	0.8404	0.0028	0.8377	0.0092
Skalar	WR2 fridge	0.046	0.882	-0.004	0.886	0.003
Skalar	WR2 freezer	0.046	0.911	-0.005	0.916	0.001

Table S.3. Flasks Ctrl3, A2 and C1 area under the curve results for total carbon (mmol) in minutes for manual oxygen monitoring with a 04/07 start as well as a 08/07 start, and the continuous oxygen monitoring with a 07/07 start.

	Excel AUC DO manual; start: 04/07/2022. Total carbon (mmol)	Excel AUC DO manual; start: 08/07/2022. Total carbon (mmol)	Rstudio AUC DO continuous; start: 07/07/2022. Total carbon (mmol)
Ctrl3	6841.05	5510.82	5696.22
A2	9178.51	7849.65	7813.97
C1	8877.76	7612.18	7594.96

Table S.4. Average nutrient % change for each treatment between start and end of the experiment (Table A, B and C), average concentration of 2-MIB and geosmin (ng/L) produced (positive) or degraded (negative) between start and end of the experiment (Table D), 2-MIB and geosmin concentrations (ng/L) in start sample (sampled three times) and flasks at day 22 (Table E), as well as average nutrient uptake rate calculated in mg/day (Table F).

A	TN % change		TDN % change		PON % change		NH_4^+-N % change		DON % change		NO_3^--N % change	
Treatment	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ctrl	-17.81	16.43	-39.04	12.73	125.52	184.30	-3.80	35.26	-20.35	31.21	-78.92	11.72
a	-17.74	6.65	-47.44	16.21	348.27	145.27	-70.25	8.72	-27.56	19.73	-50.87	18.67
b	-28.18	2.46	-82.38	0.61	584.59	61.93	-75.67	7.61	-39.66	2.11	-96.75	1.15
c	-13.47	22.25	-80.74	0.51	639.26	319.13	-95.59	0.21	-34.55	23.29	-91.46	5.76
d	-18.54	5.80	-48.96	5.81	727.73	309.33	-71.98	1.92	-40.99	10.36	-50.22	6.64
e	-19.15	6.37	-65.90	5.57	609.37	384.18	-71.25	3.26	-47.71	8.92	-71.93	5.97
f	-20.68	14.61	-71.82	9.39	395.68	201.23	-94.82	0.75	-41.88	11.11	-67.22	23.42

B	TP % change		TDP % change		PP % change		SRP % change		DOP % change	
Treatment	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ctrl	3.80	45.63	-52.63	4.56	31.13	67.12	NA	NA	-39.05	3.30
a	-11.90	5.68	-63.59	23.17	635.26	234.27	-64.35	23.60	-50.37	16.39
b	-17.26	14.13	-90.35	8.32	559.65	106.23	-93.76	8.44	-48.91	8.46
c	-2.75	26.10	-91.53	9.37	624.55	174.19	-94.51	8.85	-57.03	12.94
d	-16.78	13.86	-95.65	0.04	349.03	104.70	-100.00	0.00	-61.03	4.50
e	-5.79	19.05	-88.79	1.26	89.93	38.32	-100.00	0.00	-55.62	7.51
f	-9.77	34.68	-87.60	1.24	45.23	72.94	-100.00	0.00	-48.15	0.00

C	DOC % change	
Treatment	Mean	SD
ctrl	0.54	15.96

a	20.15	35.40
b	48.40	17.83
c	29.35	9.46
d	55.48	20.61
e	4.06	18.08
f	47.74	42.73

D	2-MIB (ng/L)		Geosmin (ng/L)	
Treatment	Mean	SD	Mean	SD
ctrl	-20.0	17.0	2.8	3.5
a	164.7	95.4	3.7	4.3
b	251.4	213.9	12.8	16.9
c	64.7	99.0	2.9	5.0
d	521.4	551.8	3.5	4.9
e	17.4	41.0	2.9	6.7
f	6.7	41.7	38.8	58.1

E	Start 1	Start 2	Start 3	Ctrl1	Ctrl2	Ctrl3	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3	F1	F2	F3
2-MIB (ng/L)	50	43	43	16	15	45	310	120	200	250	530	110	82	28	220	310	1200	190	40	110	38	25	100	31
Geosmin (ng/L)	5	3.8	4.4	3.3	8.4	10	6.3	13	4.9	12	36	3.5	3.8	13	5.1	7.5	13	3.3	4.4	15	2.6	4.6	110	15

F	NH_4^+-N uptake (mg/day)		DON uptake (mg/day)		NO_3^--N uptake (mg/day)		SRP uptake (mg/day)		DOP uptake (mg/day)	
Treatment	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD

ctrl	6.83E-04	1.06E-03	4.49E-03	3.79E-03	8.86E-03	1.03E-03	NA	NA	2.40E-04	2.16E-05
a	3.91E-03	1.13E-03	1.21E-02	5.25E-03	1.93E-01	5.14E-02	5.78E-02	1.65E-02	2.73E-03	8.66E-04
b	4.96E-03	8.39E-04	8.71E-03	5.54E-04	6.06E-02	8.78E-04	1.58E-02	1.13E-03	7.78E-04	1.18E-04
c	3.01E-02	5.34E-04	7.74E-03	4.81E-03	2.95E-02	3.12E-03	1.42E-02	1.19E-03	8.64E-04	1.45E-04
d	2.88E-03	3.66E-04	2.01E-02	7.76E-03	2.00E-01	2.36E-02	5.40E-03	5.95E-05	5.83E-04	6.53E-05
e	4.37E-03	3.13E-04	1.06E-02	2.32E-03	4.73E-02	4.49E-03	8.01E-04	4.48E-05	3.75E-04	1.15E-04
f	2.95E-02	9.75E-04	8.77E-03	2.86E-03	2.27E-02	6.95E-03	7.40E-04	1.35E-04	2.95E-04	1.71E-08
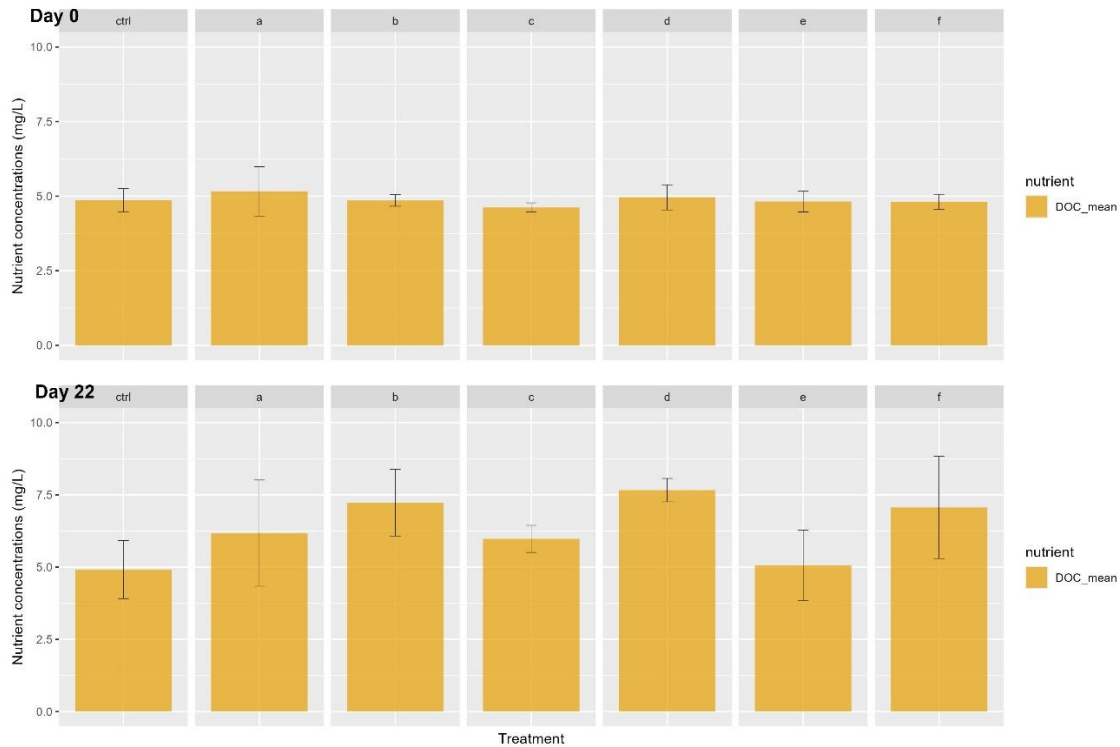
Figure S.2. Concentrations of DOC (mg/L) at the start (day 0) and the end of the experiment (day 22) averaged by treatment, with standard deviation indicated by error bars.

Table S.5. 16S rRNA indicator species result.

```
Multilevel pattern analysis
---------------------------

Association function: r.g
Significance level (alpha): 0.05

Total number of species: 487
Selected number of species: 59
Number of species associated to 1 group: 52
Number of species associated to 2 groups: 7

List of species associated to each combination:

Group Low  #sps.  48

stat p.value
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Verrucomicrobiales.f__Verrucomicrobiaceae.g__uncultured
0.766  0.0009 ***
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Pedosphaerales.f__Pedosphaeraceae.g__uncultured
0.765  0.0009 ***
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Rhizobiales.f__A0839.g__A0839
0.720  0.0020 **
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Rickettsiales.f__SM2D12.g__SM2D12
0.703  0.0002 ***
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Opitutales.f__Opitutaceae.g__Lacunisphaera
0.699  0.0038 **
d__Bacteria.p__Verrucomicrobiota.c__Chlamydiae.o__Chlamydiales.f__Parachlamydiaceae.g__Neochlamydia
0.696  0.0033 **
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Burkholderiales.f__Burkholderiaceae.g__Polynucleobacter
0.691  0.0036 **
d__Bacteria.p__WPS.2.c__WPS.2.o__WPS.2.f__WPS.2.g__WPS.2
0.683  0.0067 **
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Pedosphaerales.f__Pedosphaeraceae.__
0.675  0.0086 **
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Rhizobiales.f__Rhizobiaceae.g__Allorhizobium.Neorhizobium.P
ararhizobium.Rhizobium 0.666  0.0032 **
d__Bacteria.p__Bdellovibrionota.c__Bdellovibrionia.o__Bdellovibrionales.f__Bdellovibrionaceae.g__Bdellovibrio
0.658  0.0032 **
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Reyranellales.f__Reyranellaceae.g__Reyranella
0.647  0.0061 **
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Burkholderiales.__.__
0.646  0.0084 **
d__Bacteria.p__Armatimonadota.c__Armatimonadia.o__Armatimonadales.__.__
0.644  0.0054 **
```

```
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Pedosphaerales.f__Pedosphaeraceae.g__Pedosphaeraceae
0.628  0.0065  **
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__uncultured.f__uncultured.g__uncultured
0.618  0.0100  **
d__Bacteria.p__Actinobacteriota.c__Actinobacteria.o__Frankiales.f__Sporichthyaceae.g__hgcI_clade
0.613  0.0138  *
d__Bacteria.p__Acidobacteriota.c__Acidobacteriae.o__Acidobacteriae.f__Acidobacteriae.g__Paludibaculum
0.606  0.0107  *
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.__.__.__
0.602  0.0112  *
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Verrucomicrobiales.f__Verrucomicrobiaceae.__
0.600  0.0105  *
d__Bacteria.p__Cyanobacteria.c__Cyanobacteriia.o__Leptolyngbyales.f__Leptolyngbyaceae.g__LB3.76
0.596  0.0113  *
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Burkholderiales.f__T34.g__T34
0.583  0.0096  **
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Coxiellales.f__Coxiellaceae.g__Coxiella
0.578  0.0061  **
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Burkholderiales.f__Comamonadaceae.g__Rhizobacter
0.578  0.0138  *
d__Bacteria.p__Myxococcota.c__Polyangia.o__Polyangiales.f__Polyangiaceae.g__Pajaroellobacter
0.573  0.0179  *
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Chitinophagales.f__uncultured.g__uncultured
0.573  0.0215  *
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Steroidobacterales.f__Steroidobacteraceae.g__uncultured
0.569  0.0291  *
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Chitinophagales.f__Chitinophagaceae.g__Flavihumibacter
0.568  0.0226  *
d__Bacteria.p__Acidobacteriota.c__Blastocatellia.o__11.24.f__11.24.g__11.24
0.565  0.0262  *
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Chthoniobacterales.f__Terrimicrobiaceae.g__FukuN18_freshwat
er_group             0.565  0.0224  *
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__KI89A_clade.f__KI89A_clade.g__KI89A_clade
0.561  0.0298  *
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Verrucomicrobiales.f__Verrucomicrobiaceae.g__Prosthecobacte
r                    0.549  0.0235  *
d__Bacteria.p__Cyanobacteria.c__Cyanobacteriia.o__Leptolyngbyales.f__Leptolyngbyaceae.g__Phormidesmis_ANT.L52.6
0.544  0.0333  *
d__Bacteria.p__Patescibacteria.c__ABY1.o__Candidatus_Uhrbacteria.f__Candidatus_Uhrbacteria.g__Candidatus_Uhrbacteri
a                    0.542  0.0410  *
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Caulobacterales.f__Caulobacteraceae.g__Caulobacter
0.538  0.0289  *
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Chitinophagales.f__Saprospiraceae.g__uncultured
0.534  0.0292  *
```

```
d__Bacteria.p__Cyanobacteria.c__Cyanobacteriia.o__Pseudanabaenales.f__Pseudanabaenaceae.g__Pseudanabaena_PCC.7403
0.532  0.0157  *
d__Bacteria.p__Myxococcota.c__Myxococcia.o__Myxococcales.f__Myxococcaceae.g__P3OB.42
0.530  0.0293  *
d__Bacteria.p__Planctomycetota.c__Planctomycetes.o__Planctomycetales.f__Rubinisphaeraceae.g__SH.PL14
0.530  0.0349  *
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Burkholderiales.f__Nitrosomonadaceae.g__uncultured
0.523  0.0367  *
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Chitinophagales.f__Saprospiraceae.g__Lewinella
0.521  0.0408  *
d__Bacteria.p__Actinobacteriota.c__Actinobacteria.o__Frankiales.f__Sporichthyaceae.g__Sporichthyaceae
0.514  0.0432  *
d__Bacteria.p__Bacteroidota.c__SJA.28.o__SJA.28.f__SJA.28.g__SJA.28
0.502  0.0253  *
d__Bacteria.p__Dependentiae.c__Babeliae.o__Babeliales.f__UBA12409.g__UBA12409
0.493  0.0211  *
d__Bacteria.p__Verrucomicrobiota.c__Kiritimatiellae.o__WCHB1.41.f__WCHB1.41.g__WCHB1.41
0.490  0.0493  *
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Rhizobiales.f__Xanthobacteraceae.g__Xanthobacter
0.468  0.0453  *
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Cellvibrionales.f__Cellvibrionaceae.g__uncultured
0.452  0.0491  *
d__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Burkholderiales.f__Comamonadaceae.g__Limnohabitans
0.418  0.0494  *

 Group Medium  #sps.  1

stat p.value
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Holosporales.f__Holosporaceae.g__Candidatus_Paraholospora 0
.509  0.0474  *

 Group High  #sps.  3

stat p.value
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Chitinophagales.f__Chitinophagaceae.__                             0
.585  0.0280  *
d__Bacteria.p__Cyanobacteria.c__Cyanobacteriia.o__Pseudanabaenales.f__Pseudanabaenaceae.g__Pseudanabaena_PCC.7429 0
.556  0.0253  *
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Cytophagales.f__Cyclobacteriaceae.g__Algoriphagus                  0
.540  0.0036  **

 Group Low+Medium  #sps.  5

                                                                                                    stat p.valu
e
```

```
d__Bacteria.p__Verrucomicrobiota.c__Verrucomicrobiae.o__Pedosphaerales.f__Pedosphaeraceae.g__DEV114          0.590  0.016
0 *
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Sphingobacteriales.f__AKYH767.g__AKYH767                        0.561  0.026
2 *
d__Bacteria.p__Actinobacteriota.c__Actinobacteria.o__Frankiales.f__Sporichthyaceae.                           0.560  0.023
3 *
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Caulobacterales.f__Hyphomonadaceae.g__Hirschia 0.556  0.026
8 *
d__Bacteria.p__Bacteroidota.c__Bacteroidia.o__Chitinophagales.f__Chitinophagaceae.g__Terrimonas             0.522  0.044
6 *

 Group Medium+High  #sps.   2
                                                                                                                  sta
t p.value
d__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Caulobacterales.f__Hyphomonadaceae.g__UKL13.1         0.58
2  0.0209 *
d__Bacteria.p__Cyanobacteria.c__Cyanobacteriia.o__Leptolyngbyales.f__Leptolyngbyaceae.g__Leptolyngbya_SAG_2411 0.51
5  0.0482 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table S.6. rbcL indicator species result

```
Multilevel pattern analysis
---------------------------

Association function: r.g
Significance level (alpha): 0.05

Total number of species: 116
Selected number of species: 9
Number of species associated to 1 group: 8
Number of species associated to 2 groups: 1

List of species associated to each combination:

 Group Low  #sps.   6
                                                                               stat p.value
k__Eukaryota.p__Chlorophyta.c__Trebouxiophyceae.o__Chlorellales.f__Chlorellaceae.g__Gloeotila   0.549  0.0218 *
k__Eukaryota.p__Chlorophyta.c__Trebouxiophyceae.o__.f__.g__                                      0.528  0.0096 **
k__Eukaryota.p__Chlorophyta.c__Trebouxiophyceae.o__.f__.g__Choricystis                           0.508  0.0254 *
k__Eukaryota.p__Chlorophyta.c__Trebouxiophyceae.o__.f__.g__Botryococcus                          0.477  0.0447 *
k__Eukaryota.p__Chlorophyta.c__Chlorophyceae.__.__.__                                            0.468  0.0176 *
k__Eukaryota.p__Chlorophyta.c__Chlorophyceae.o__Sphaeropleales.f__Scenedesmaceae.g__Scenedesmus 0.448  0.0266 *

 Group Medium  #sps.   1
                                                                                 stat p.value
k__Eukaryota.p__Chlorophyta.c__Chlorophyceae.o__Sphaeropleales.f__Selenastraceae.g__Monoraphidium 0.628   0.005 **

 Group High  #sps.   1
                                                                                 stat p.value
k__Eukaryota.p__Chlorophyta.c__Chlorophyceae.o__Chlamydomonadales.f__Palmellopsidaceae.g__Asterococcus 0.48  0.0462 *

 Group Medium+High  #sps.   1
                                                                                 stat p.value
k__Eukaryota.p__Chlorophyta.c__Chlorophyceae.o__Sphaeropleales.f__Scenedesmaceae.g__Tetradesmus 0.541  0.0368 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix C
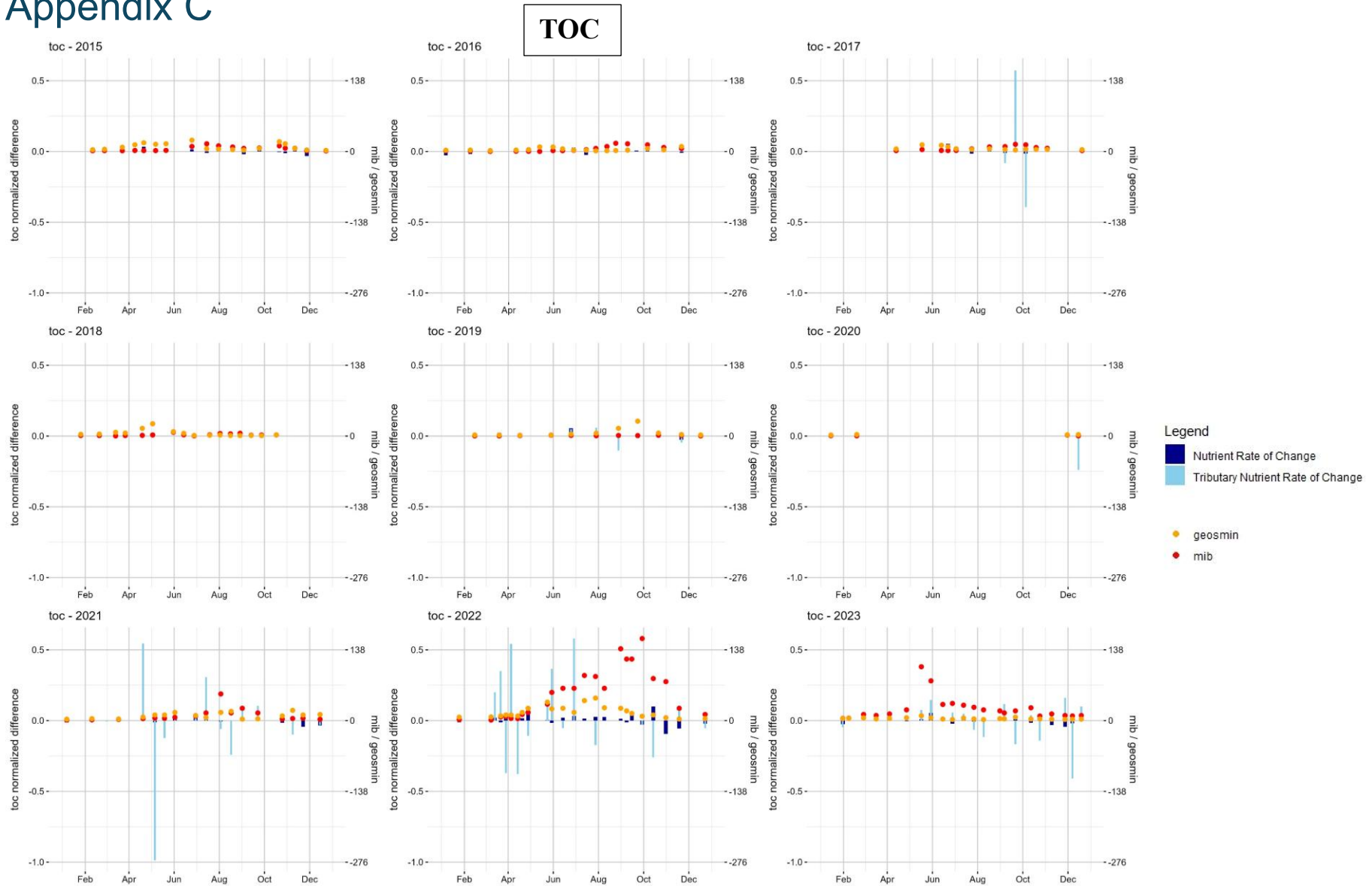
Figure S.3. TOC 2022 rate of change, compared with 2-MIB and geosmin concentrations (ng/L).
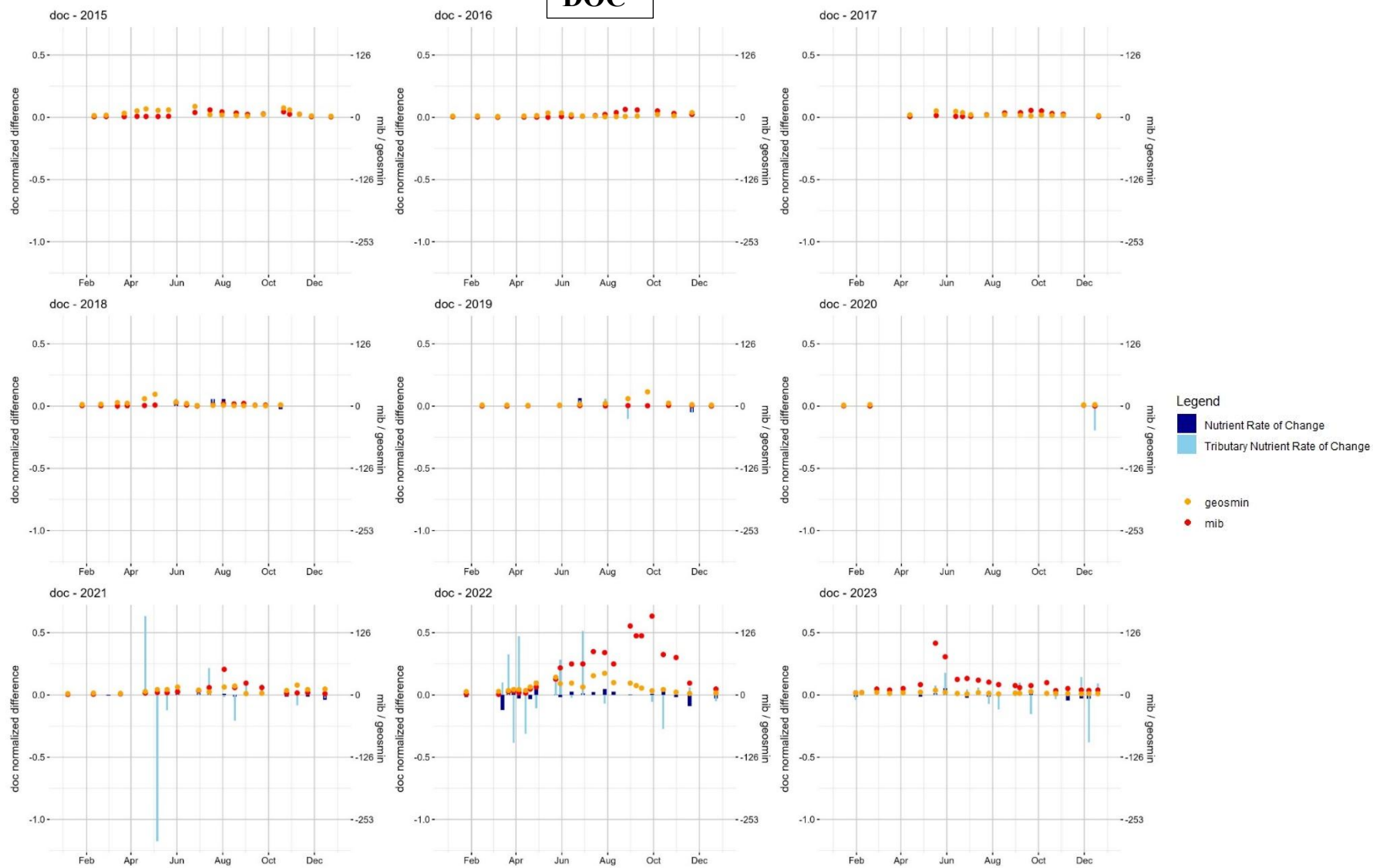
Figure S.4. DOC 2022 rate of change, compared with 2-MIB and geosmin concentrations (ng/L).

278
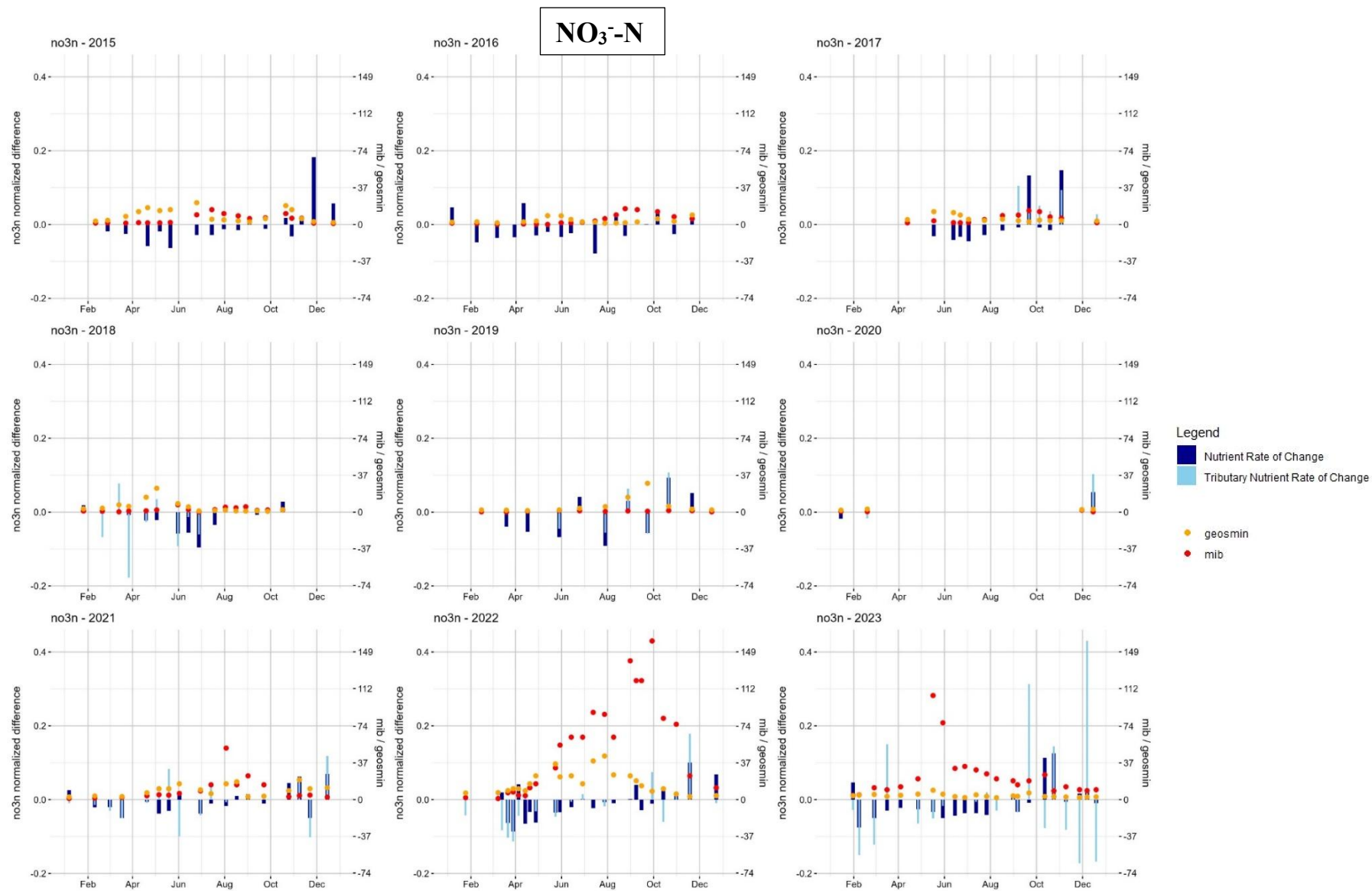
Figure S.5. NO₃⁻-N 2022 rate of change, compared with 2-MIB and geosmin concentrations (ng/L).

NH$_4^+$-N
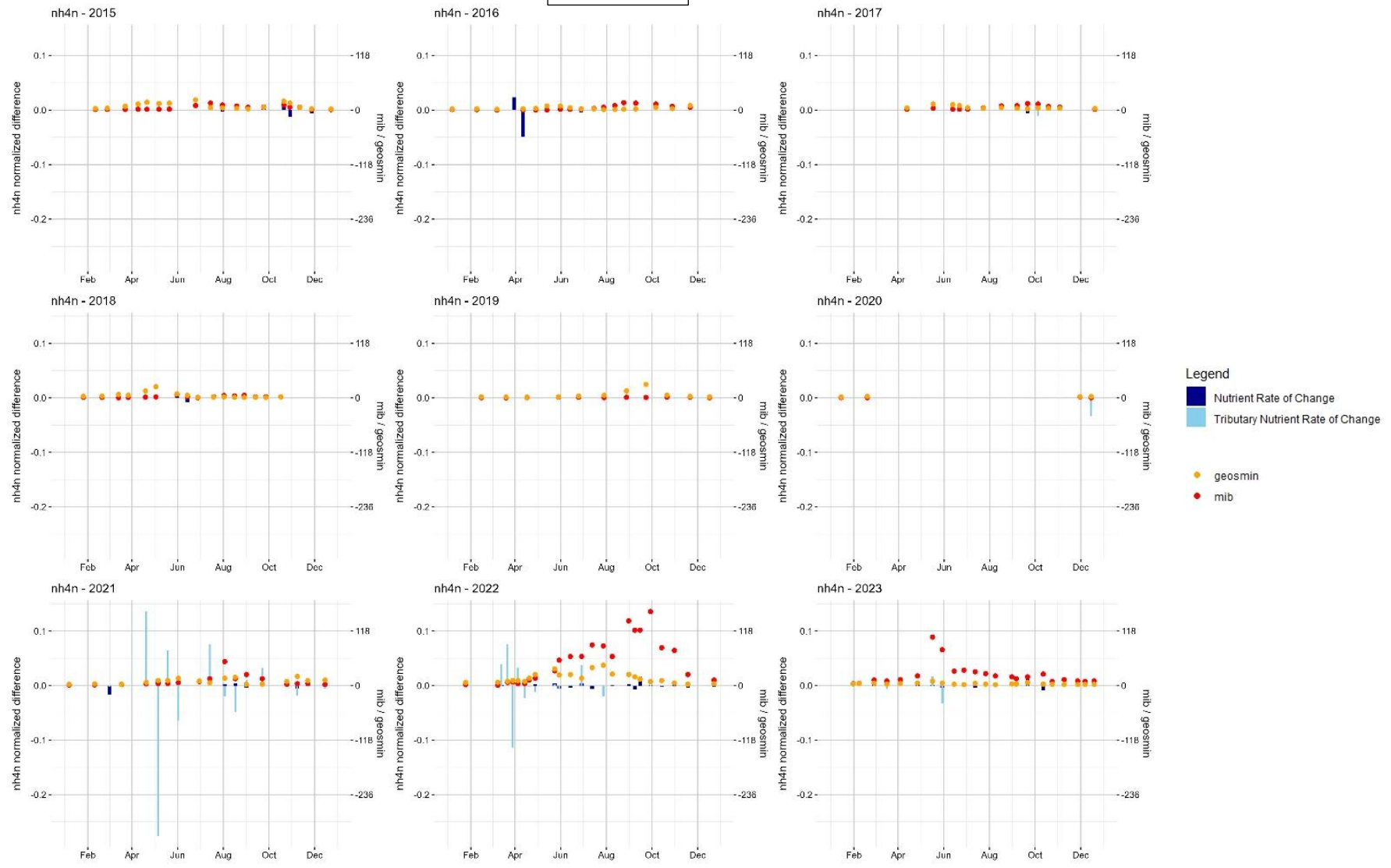
Figure S.6. NH$_4^+$-N 2022 rate of change, compared with 2-MIB and geosmin concentrations (ng/L).
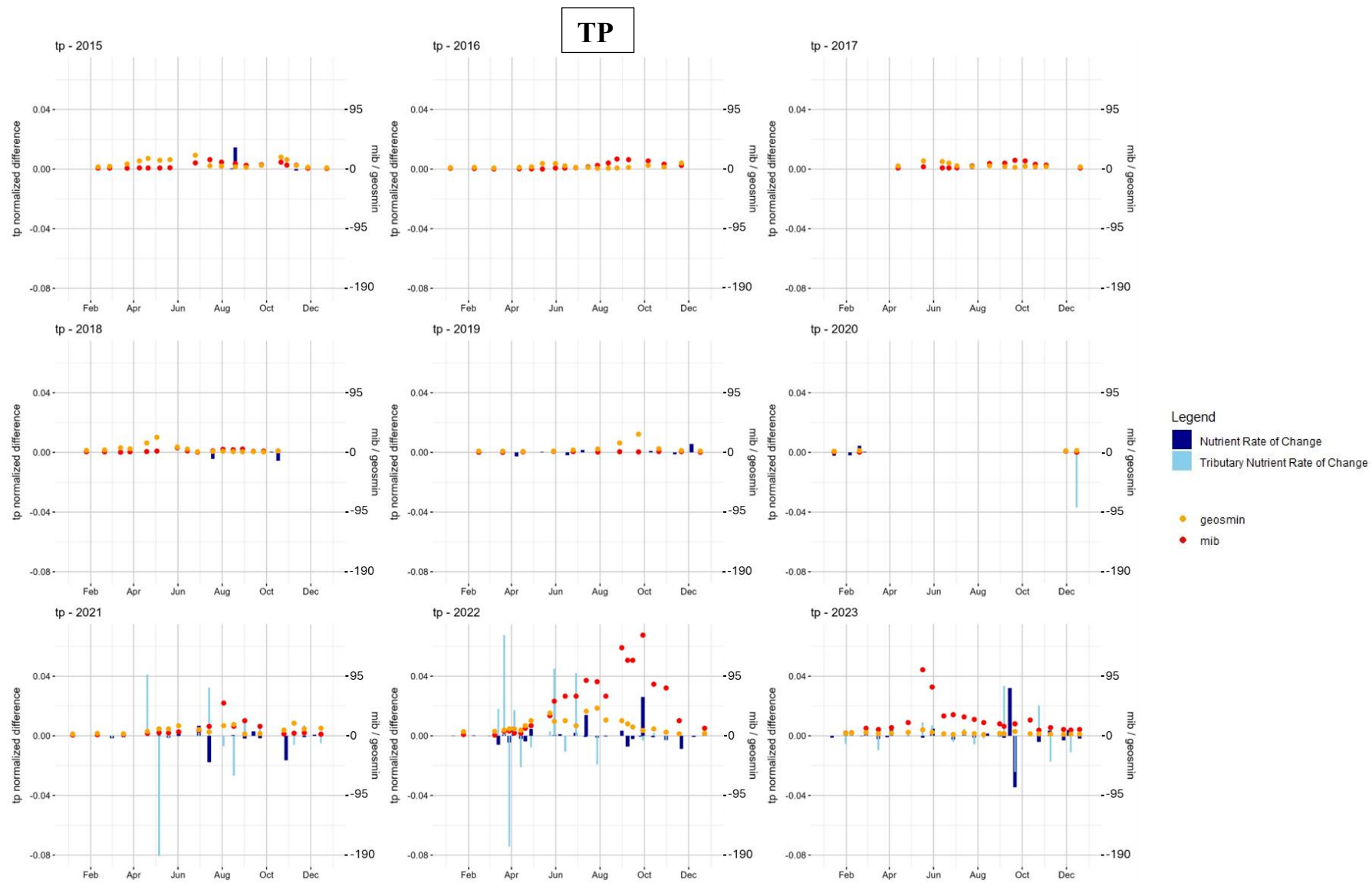
Figure S.7. TP 2022 rate of change, compared with 2-MIB and geosmin concentrations (ng/L).

281
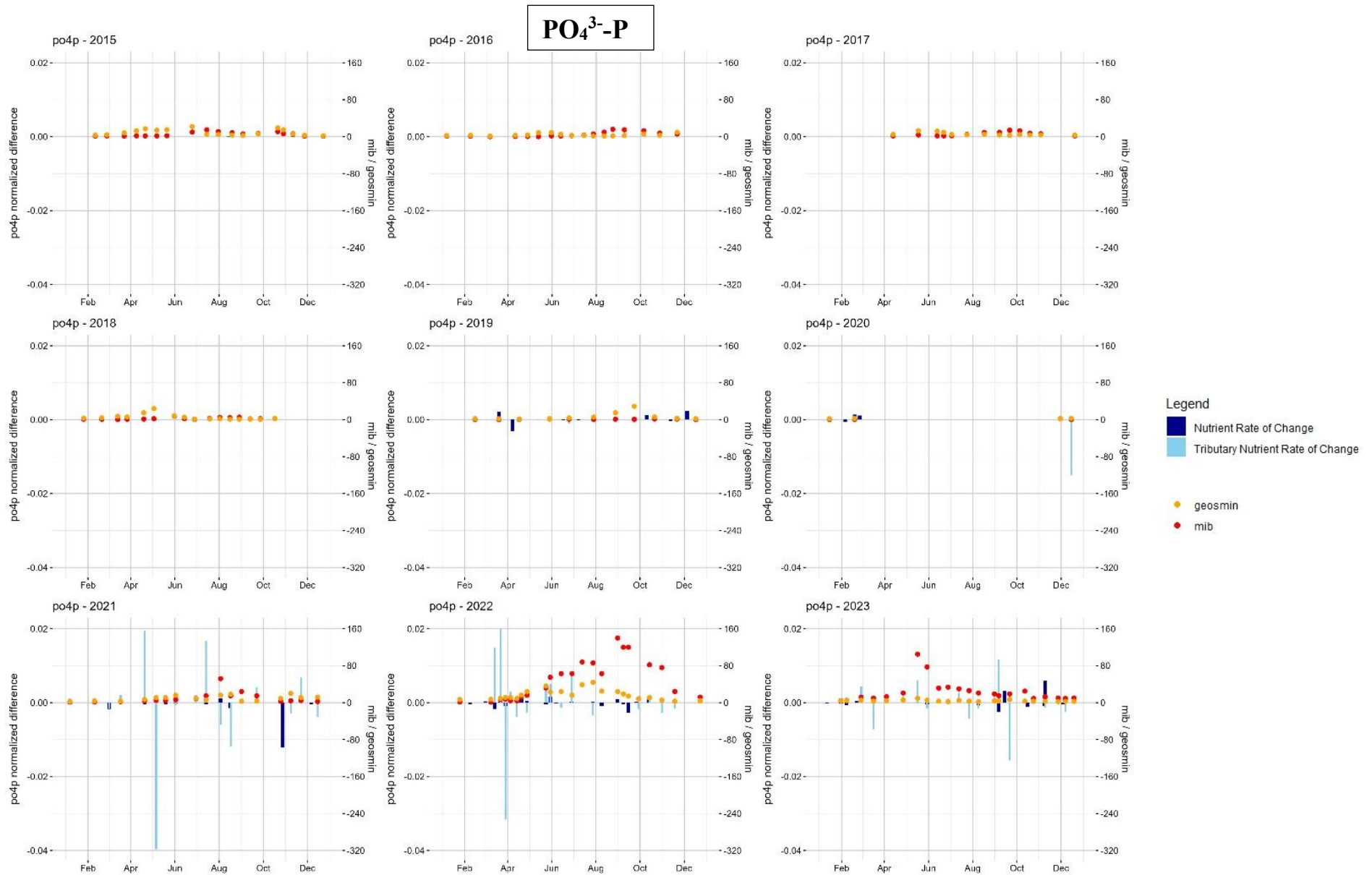
Figure S.8. PO$_4^{3-}$-P 2022 rate of change, compared with 2-MIB and geosmin concentrations (ng/L).

Table S.7. Correlations and simple GAMs on all data.

Predictor	Correlation 2-MIB	Correlation Geosmin
algaetot	0.18	0.15
alkalinity	0.21	0.02
aluminiumdis	-0.32	0.07
aluminiumtot	-0.13	-0.27
avg_temp_1week_prior	0.44	0.15
avg_temp_2weeks_prior	0.36	0.11
avg_temp_3weeks_prior	0.45	0.05
avg_temp_4weeks_prior	0.45	0.03
avg_tempdiff_1week_prior	0.15	0.24
avg_tempdiff_2weeks_prior	0.21	0.31
avg_tempdiff_3weeks_prior	0.22	0.17
avg_tempdiff_4weeks_prior	0.27	0.21
avg_winddir_diff_1week_prior	0.09	0.17
avg_winddir_diff_2weeks_prior	-0.07	0.21
avg_winddir_diff_3weeks_prior	0.13	0.16
avg_winddir_diff_4weeks_prior	0.10	0.03
avg_windspeed_1week_prior	-0.23	-0.24
avg_windspeed_2weeks_prior	-0.32	-0.14
avg_windspeed_3weeks_prior	-0.22	-0.04
avg_windspeed_4weeks_prior	-0.28	-0.14
bluegreenalgae	0.11	0.04
chla	0.16	0.03
chloride	0.55	0.21
cloudcover	-0.01	-0.02
cloudcover_lag_lag_1	-0.12	0.01
cloudcover_lag_lag_2	-0.04	-0.05
cloudcover_lag_lag_3	-0.11	-0.12
cloudcover_lag_lag_4	-0.15	-0.20
cloudcover_lag_lag_5	0.01	0.01
cloudcover_lag_lag_6	-0.12	0.07
cloudcover_lag_lag_7	-0.02	0.12
conductivity	0.22	0.21
dinsrp_ratio	-0.45	-0.18
dintp_ratio	-0.53	-0.28
doc	0.45	0.01
doc_trib	0.59	0.12
DOY	0.35	0.06
geosmin	0.36	1.00
greenalgae	0.18	0.24
iron	0.27	-0.11
irondis	0.42	0.24

manganese	0.42	0.05
manganesedis	0.51	0.32
max_solarenergy_1week_prior	0.27	0.31
max_solarenergy_2weeks_prior	0.35	0.32
max_solarenergy_3weeks_prior	0.40	0.30
max_solarenergy_4weeks_prior	0.42	0.28
max_solarradiation_1week_prior	0.27	0.31
max_solarradiation_2weeks_prior	0.35	0.32
max_solarradiation_3weeks_prior	0.40	0.30
max_solarradiation_4weeks_prior	0.42	0.29
max_temp_1week_prior	0.37	0.16
max_temp_2weeks_prior	0.41	0.16
max_temp_3weeks_prior	0.44	0.09
max_temp_4weeks_prior	0.42	0.07
max_tempdiff_1week_prior	0.08	0.13
max_tempdiff_2weeks_prior	0.15	0.30
max_tempdiff_3weeks_prior	0.24	0.18
max_tempdiff_4weeks_prior	0.21	0.14
max_winddir_diff_1week_prior	0.18	0.19
max_winddir_diff_2weeks_prior	-0.01	0.19
max_winddir_diff_3weeks_prior	0.17	0.14
max_winddir_diff_4weeks_prior	0.15	0.06
max_windgust_1week_prior	-0.33	-0.10
max_windgust_2weeks_prior	-0.44	-0.07
max_windgust_3weeks_prior	-0.30	0.07
max_windgust_4weeks_prior	-0.40	-0.07
max_windspeed_1week_prior	-0.27	-0.22
max_windspeed_2weeks_prior	-0.35	-0.17
max_windspeed_3weeks_prior	-0.25	0.03
max_windspeed_4weeks_prior	-0.32	-0.15
mib	1.00	0.36
min_temp_1week_prior	0.39	0.07
min_temp_2weeks_prior	0.48	0.02
min_temp_3weeks_prior	0.40	-0.03
min_temp_4weeks_prior	0.45	-0.02
min_windspeed_1week_prior	-0.09	-0.17
min_windspeed_2weeks_prior	-0.13	-0.07
min_windspeed_3weeks_prior	-0.17	-0.07
min_windspeed_4weeks_prior	-0.09	-0.10
nh4n	0.08	0.10
nh4n_trib	0.09	0.10
nh4no3_ratio	0.34	0.14
nh4no3_ratio_trib	0.13	0.10
no2n	-0.26	0.05

no2n_trib	0.39	0.09
no3n	-0.53	-0.14
no3n_trib	-0.33	-0.27
otheralgaetot	-0.04	-0.13
pH	0.18	0.06
po4p	-0.18	-0.19
po4p_trib	0.39	0.15
precip	0.05	-0.11
precip_lag_lag_1	-0.14	-0.16
precip_lag_lag_2	-0.09	-0.03
precip_lag_lag_3	-0.03	-0.19
precip_lag_lag_4	-0.07	-0.27
precip_lag_lag_5	-0.01	-0.13
precip_lag_lag_6	-0.16	-0.21
precip_lag_lag_7	-0.02	-0.06
silicate	-0.09	-0.14
solarenergy	0.18	0.29
solarenergy_lag_lag_1	0.19	0.28
solarenergy_lag_lag_2	0.26	0.27
solarenergy_lag_lag_3	0.23	0.31
solarenergy_lag_lag_4	0.22	0.37
solarenergy_lag_lag_5	0.18	0.28
solarenergy_lag_lag_6	0.31	0.21
solarenergy_lag_lag_7	0.23	0.25
solarradiation	0.16	0.28
solarradiation_lag_lag_1	0.17	0.28
solarradiation_lag_lag_2	0.25	0.27
solarradiation_lag_lag_3	0.22	0.30
solarradiation_lag_lag_4	0.22	0.36
solarradiation_lag_lag_5	0.19	0.29
solarradiation_lag_lag_6	0.30	0.21
solarradiation_lag_lag_7	0.22	0.25
sulphate	-0.56	-0.22
sum_precip_1week_prior	-0.10	-0.30
sum_precip_2weeks_prior	-0.20	-0.12
sum_precip_3weeks_prior	-0.11	-0.19
sum_precip_4weeks_prior	-0.16	-0.14
sum_solarenergy_1week_prior	0.26	0.32
sum_solarenergy_2weeks_prior	0.35	0.33
sum_solarenergy_3weeks_prior	0.40	0.30
sum_solarenergy_4weeks_prior	0.43	0.29
sum_solarradiation_1week_prior	0.26	0.32
sum_solarradiation_2weeks_prior	0.35	0.33
sum_solarradiation_3weeks_prior	0.40	0.30

sum_solarradiation_4weeks_prior	0.43	0.29
temp	0.35	0.13
temp_lag_lag_1	0.34	0.12
temp_lag_lag_2	0.39	0.17
temp_lag_lag_3	0.42	0.18
temp_lag_lag_4	0.40	0.12
temp_lag_lag_5	0.43	0.12
temp_lag_lag_6	0.44	0.10
temp_lag_lag_7	0.39	0.15
tempdiff	0.13	0.18
tempdiff_lag_lag_1	0.07	0.12
tempdiff_lag_lag_2	0.12	0.12
tempdiff_lag_lag_3	0.07	0.20
tempdiff_lag_lag_4	0.10	0.20
tempdiff_lag_lag_5	0.06	0.13
tempdiff_lag_lag_6	0.19	0.11
tempdiff_lag_lag_7	0.08	0.03
tempdiff_max_min_1week_prior	-0.05	0.05
tempdiff_max_min_2weeks_prior	-0.01	0.20
tempdiff_max_min_3weeks_prior	0.06	0.13
tempdiff_max_min_4weeks_prior	0.03	0.13
tempmax	0.37	0.16
tempmax_lag_lag_1	0.33	0.14
tempmax_lag_lag_2	0.39	0.18
tempmax_lag_lag_3	0.41	0.22
tempmax_lag_lag_4	0.38	0.16
tempmax_lag_lag_5	0.43	0.16
tempmax_lag_lag_6	0.44	0.13
tempmax_lag_lag_7	0.43	0.15
tempmin	0.27	0.01
tempmin_lag_lag_1	0.31	0.01
tempmin_lag_lag_2	0.30	0.12
tempmin_lag_lag_3	0.39	0.12
tempmin_lag_lag_4	0.36	0.07
tempmin_lag_lag_5	0.42	0.08
tempmin_lag_lag_6	0.42	0.07
tempmin_lag_lag_7	0.36	0.08
tn	-0.54	-0.15
tntp_ratio	-0.51	-0.32
toc	0.67	0.26
toc_trib	0.59	0.11
toxn	-0.54	-0.14
toxn_trib	-0.32	-0.27
toxnsrp_ratio	-0.44	-0.15

tp	0.13	0.11
tp_trib	0.39	0.15
turbidity	0.04	-0.27
winddir	-0.08	-0.04
winddir_diff	-0.09	-0.06
winddir_diff_lag_lag_1	-0.04	0.05
winddir_diff_lag_lag_2	0.00	0.10
winddir_diff_lag_lag_3	-0.01	0.01
winddir_diff_lag_lag_4	0.10	0.01
winddir_diff_lag_lag_5	0.01	0.13
winddir_diff_lag_lag_6	0.08	0.21
winddir_diff_lag_lag_7	-0.06	0.01
winddir_lag_lag_1	-0.04	-0.10
winddir_lag_lag_2	0.04	0.00
winddir_lag_lag_3	-0.01	-0.12
winddir_lag_lag_4	-0.04	-0.14
winddir_lag_lag_5	-0.09	-0.22
winddir_lag_lag_6	0.09	-0.03
winddir_lag_lag_7	0.03	-0.04
windgust	-0.26	-0.14
windgust_lag_lag_1	-0.15	-0.12
windgust_lag_lag_2	-0.26	-0.04
windgust_lag_lag_3	-0.20	-0.06
windgust_lag_lag_4	-0.23	-0.07
windgust_lag_lag_5	-0.15	-0.12
windgust_lag_lag_6	-0.30	-0.15
windgust_lag_lag_7	-0.26	-0.10
windspeed	-0.23	-0.20
windspeed_lag_lag_1	-0.08	-0.15
windspeed_lag_lag_2	-0.17	-0.14
windspeed_lag_lag_3	-0.06	-0.09
windspeed_lag_lag_4	-0.09	-0.12
windspeed_lag_lag_5	-0.07	-0.13
windspeed_lag_lag_6	-0.18	-0.16
windspeed_lag_lag_7	-0.10	-0.11
year	0.51	0.16

Table S.8. GAM all results Geosmin on all data.

Parameter	P Value	R2 Adj	Deviance Explained	Num Observations
DOY	0.002	0.119	0.139	141
precip	0.485	-0.003	0.004	141
tempmax	0.040	0.047	0.064	141
tempmin	0.534	-0.005	0.003	141
temp	0.089	0.038	0.056	141

windgust	0.120	0.011	0.017	141
windspeed	0.019	0.031	0.037	141
winddir	0.773	-0.007	0.001	141
cloudcover	0.342	-0.001	0.006	141
solarradiation	0.000	0.091	0.093	141
solarenergy	0.000	0.091	0.093	141
sum_solarenergy_2weeks_prior	0.000	0.093	0.102	141
min_temp_2weeks_prior	0.249	0.022	0.034	141
max_windgust_2weeks_prior	0.212	0.002	0.010	141
sum_precip_1week_prior	0.000	0.097	0.105	141
avg_windspeed_1week_prior	0.000	0.057	0.078	141
pH	0.001	0.122	0.153	136
turbidity	0.002	0.063	0.093	136
toc	0.005	0.054	0.064	117
doc	0.000	0.146	0.252	90
tn	0.005	0.071	0.094	140
nh4n	0.257	0.002	0.009	139
no3n	0.005	0.071	0.093	140
tp	0.272	0.002	0.013	92
po4p	0.034	0.022	0.042	93
manganesedis	0.010	0.084	0.087	76
manganese	0.589	-0.006	0.002	137
chla	0.644	-0.006	0.001	139
algaetot	0.707	-0.007	0.001	140
greenalgae	0.074	0.036	0.027	140
bluegreenalgae	0.338	-0.002	0.006	140
otheralgaetot	0.439	-0.002	0.004	138
tntp_ratio	0.002	0.130	0.168	92
nh4no3_ratio	0.121	0.009	0.017	139
geosmin	0.000	0.998	0.997	141
mib	0.000	0.136	0.126	141
toc_trib	0.377	-0.004	0.011	68
nh4n_trib	0.488	-0.007	0.005	85
no3n_trib	0.004	0.073	0.092	85
tp_trib	0.264	0.001	0.018	66
po4p_trib	0.223	0.001	0.020	66
nh4no3_ratio_trib	0.420	-0.006	0.007	85
year	0.000	0.329	0.341	141
conductivity	0.000	0.248	0.283	77

Table S.9. GAM all results 2-MIB on all data.

Parameter	P Value	R2 Adj	Deviance Explained	Num Observations
DOY	0.000	0.226	0.247	141

precip	0.415	-0.001	0.004	141
tempmax	0.000	0.109	0.131	141
tempmin	0.000	0.095	0.113	141
temp	0.000	0.126	0.149	141
windgust	0.002	0.065	0.063	141
windspeed	0.004	0.048	0.051	141
winddir	0.558	-0.004	0.002	141
cloudcover	0.644	-0.005	0.001	141
solarradiation	0.026	0.028	0.031	141
solarenergy	0.026	0.029	0.032	141
sum_solarenergy_2weeks_prior	0.000	0.117	0.109	141
min_temp_2weeks_prior	0.000	0.221	0.216	141
max_windgust_2weeks_prior	0.000	0.188	0.188	141
sum_precip_1week_prior	0.121	0.011	0.016	141
avg_windspeed_1week_prior	0.000	0.089	0.117	141
pH	0.038	0.042	0.059	136
turbidity	0.941	-0.007	0.000	136
toc	0.000	0.458	0.467	117
doc	0.000	0.417	0.421	90
tn	0.000	0.285	0.286	140
nh4n	0.384	-0.002	0.005	139
no3n	0.000	0.279	0.277	140
tp	0.254	0.005	0.013	92
po4p	0.001	0.125	0.178	93
manganesedis	0.000	0.219	0.177	76
manganese	0.000	0.297	0.294	137
chla	0.227	0.001	0.009	139
algaetot	0.254	0.000	0.008	140
greenalgae	0.108	0.026	0.019	140
bluegreenalgae	0.136	0.013	0.015	140
otheralgaetot	0.539	-0.005	0.002	138
tntp_ratio	0.000	0.286	0.252	92
nh4no3_ratio	0.002	0.118	0.102	139
geosmin	0.000	0.138	0.104	141
mib	0.000	0.972	0.944	141
toc_trib	0.000	0.454	0.478	68
nh4n_trib	0.805	-0.002	0.009	85
no3n_trib	0.007	0.136	0.138	85
tp_trib	0.000	0.395	0.427	66
po4p_trib	0.000	0.376	0.372	66
nh4no3_ratio_trib	0.352	0.033	0.036	85
year	0.000	0.521	0.535	141
conductivity	0.060	0.025	0.039	77
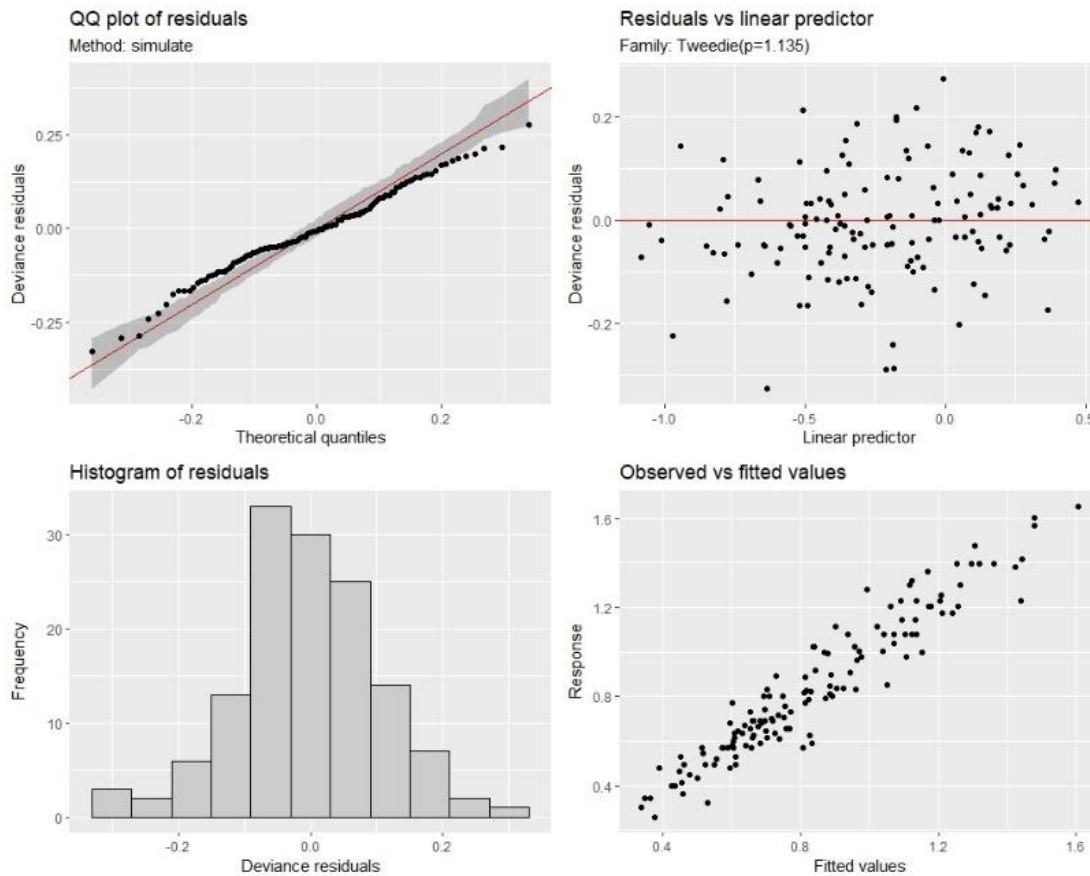
Figure S.9. Full GAM Geosmin residuals.

Table S.10. Full GAM Geosmin model parameters.

```
Family: Tweedie(p=1.132)
Link function: log

Formula:
log_geosmin ~ s(no3n) + s(sum_solarenergy_2weeks_prior) + te(year,
    DOY, bs = c("cr", "cc"), k = c(8, 12))

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.23581    0.01303  -18.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                  edf Ref.df      F p-value
s(no3n)                          1.00   1.00  1.300  0.2575
s(sum_solarenergy_2weeks_prior)  1.00   1.00  3.088  0.0826 .
te(year,DOY)                    49.28  59.86 10.879  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.849   Deviance explained = 90.4%
-ML = -26.175  Scale est. = 0.017948  n = 135
```

Table S.11. k.check output for full GAM Geosmin, k' is k-1 (k = smoothing parameter), edf should be below k' to prevent over or underfitting.

```
                              k'       edf  k-index  p-value
s(no3n)                        9  1.000009 1.005936   0.4700
s(sum_solarenergy_2weeks_prior) 9  1.000010 1.071505   0.7575
te(year,DOY)                  87 49.278364 1.227079   1.0000
```
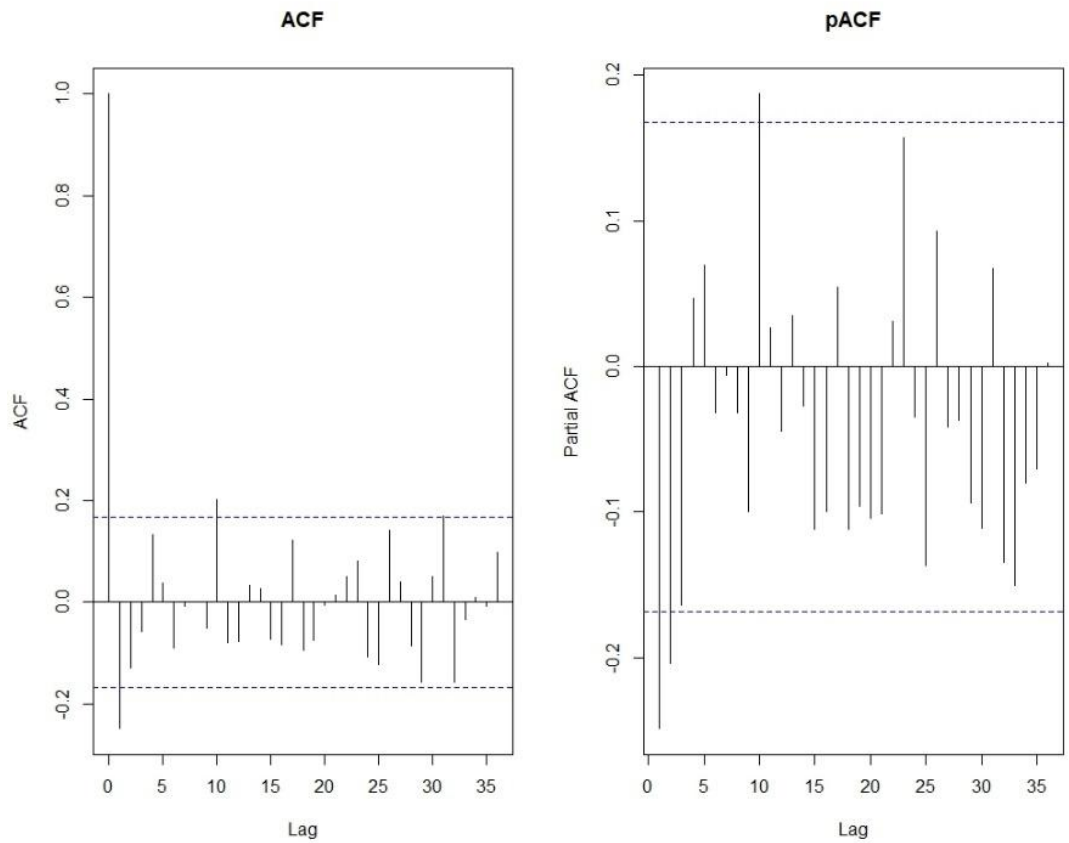

Figure S.10. Full GAM Geosmin autocorrelation (ACF) and partial autocorrelation (pACF).
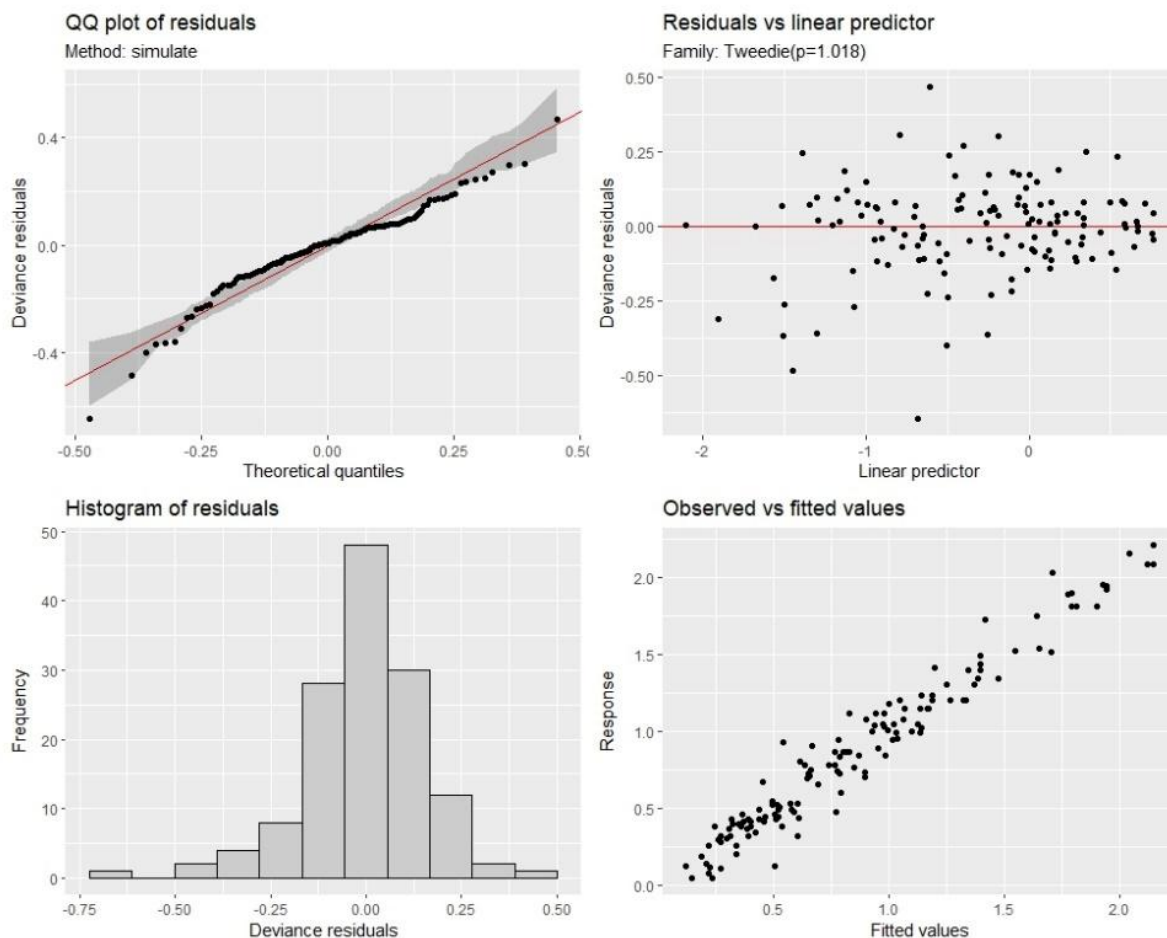
Figure S.11. Full GAM 2-MIB residuals.

Table S.12. Full GAM 2-MIB model parameters.

```
Family: Tweedie(p=1.019)
Link function: log

Formula:
log_mib ~ s(no3n) + s(max_windgust_2weeks_prior) + s(min_temp_2weeks_prior) +
    te(year, DOY, bs = c("cr", "cc"), k = c(8, 12))

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.31889    0.01851  -17.23   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                              edf Ref.df      F p-value
s(no3n)                     1.776  2.155  2.974  0.0414 *
s(max_windgust_2weeks_prior) 1.000  1.000  3.999  0.0484 *
s(min_temp_2weeks_prior)     1.000  1.000  0.800  0.3734
te(year,DOY)                34.873 44.186 18.225  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.929   Deviance explained = 92.7%
-ML = -21.089  Scale est. = 0.027917  n = 135
```

Table S.13. k.check output for full GAM 2-MIB, k' is k-1 (k = smoothing parameter), edf should be below k' to prevent over or underfitting.

```
                             k'      edf  k-index p-value
s(no3n)                       9  1.776183 1.103692  0.8650
s(max_windgust_2weeks_prior)  9  1.000004 1.052282  0.7225
s(min_temp_2weeks_prior)      9  1.000006 0.939841  0.1925
te(year,DOY)                 87 34.872552 1.136931  0.9750
```
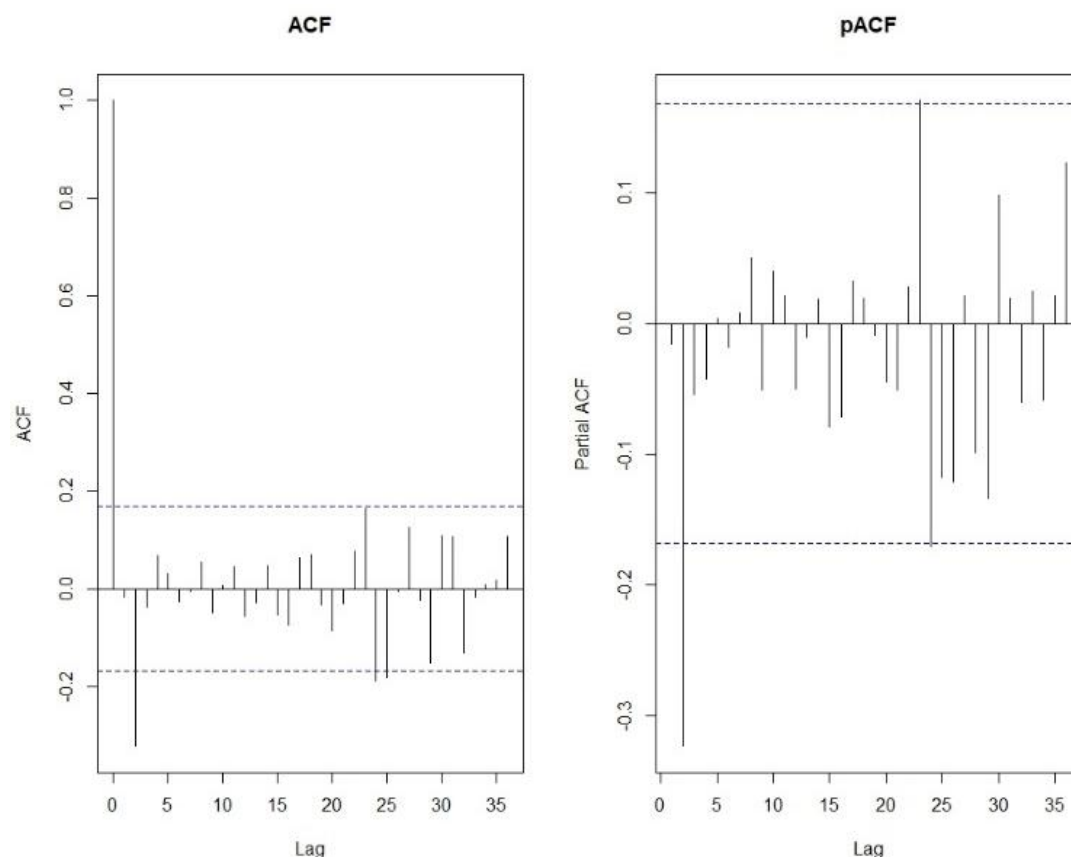

Figure S.12. Full GAM 2-MIB autocorrelation (ACF) and partial autocorrelation (pACF).

Table S.14. Correlations between environmental parameters and 2-MIB and geosmin, within the sensor deployment period (2022-2023).

Predictor	Correlation 2-MIB	Correlation Geosmin
atemp_max	0.57	0.38
atemp_max_max_1week_prior	0.69	0.43
atemp_max_max_2weeks_prior	0.72	0.46
atemp_max_max_3weeks_prior	0.78	0.56
atemp_max_max_4weeks_prior	0.74	0.48
atemp_mean	0.66	0.38
atemp_mean_avg_1week_prior	0.75	0.43
atemp_mean_avg_2weeks_prior	0.79	0.47
atemp_mean_avg_3weeks_prior	0.79	0.49
atemp_mean_avg_4weeks_prior	0.70	0.43
atemp_median	0.66	0.38

293

atemp_median_median_1week_prior	0.72	0.41
atemp_median_median_2weeks_prior	0.78	0.47
atemp_median_median_3weeks_prior	0.77	0.47
atemp_median_median_4weeks_prior	0.61	0.39
atemp_min	0.57	0.25
atemp_min_min_1week_prior	0.73	0.29
atemp_min_min_2weeks_prior	0.78	0.38
atemp_min_min_3weeks_prior	0.59	0.39
atemp_min_min_4weeks_prior	0.63	0.33
atemp_range	0.33	0.32
atemp_range_avg_1week_prior	0.48	0.31
atemp_range_avg_2weeks_prior	0.40	0.18
atemp_range_avg_3weeks_prior	0.60	0.31
atemp_range_avg_4weeks_prior	0.53	0.24
atemp_range_max_1week_prior	0.30	0.34
atemp_range_max_2weeks_prior	-0.07	-0.09
atemp_range_max_3weeks_prior	0.39	0.37
atemp_range_max_4weeks_prior	0.42	0.22
atemp_range_min_1week_prior	0.55	0.22
atemp_range_min_2weeks_prior	0.24	0.09
atemp_range_min_3weeks_prior	0.71	0.25
atemp_range_min_4weeks_prior	0.37	0.14
atemp_weekly_range_1week_prior	0.08	0.21
atemp_weekly_range_2weeks_prior	-0.01	0.06
atemp_weekly_range_3weeks_prior	0.47	0.41
atemp_weekly_range_4weeks_prior	0.43	0.28
chla_max	-0.13	0.64
chla_max_max_1week_prior	-0.18	0.64
chla_max_max_2weeks_prior	-0.15	0.57
chla_max_max_3weeks_prior	-0.35	0.61
chla_max_max_4weeks_prior	-0.32	0.14
chla_mean	-0.28	0.54
chla_mean_avg_1week_prior	-0.20	0.67
chla_mean_avg_2weeks_prior	-0.29	0.79
chla_mean_avg_3weeks_prior	-0.41	0.67
chla_mean_avg_4weeks_prior	-0.55	0.40
chla_median	-0.31	0.51
chla_median_median_1week_prior	-0.12	0.61
chla_median_median_2weeks_prior	-0.27	0.82
chla_median_median_3weeks_prior	-0.32	0.62
chla_median_median_4weeks_prior	-0.55	0.48
chla_min	-0.30	0.45
chla_min_min_1week_prior	-0.04	0.57
chla_min_min_2weeks_prior	-0.32	0.55
chla_min_min_3weeks_prior	-0.23	0.57
chla_min_min_4weeks_prior	-0.47	0.44

chla_range	0.09	0.67
chla_range_avg_1week_prior	-0.16	0.61
chla_range_avg_2weeks_prior	-0.26	0.61
chla_range_avg_3weeks_prior	-0.41	0.69
chla_range_avg_4weeks_prior	-0.43	0.15
chla_range_max_1week_prior	-0.17	0.62
chla_range_max_2weeks_prior	-0.15	0.52
chla_range_max_3weeks_prior	-0.38	0.63
chla_range_max_4weeks_prior	-0.30	0.09
chla_range_min_1week_prior	0.05	0.57
chla_range_min_2weeks_prior	-0.29	0.67
chla_range_min_3weeks_prior	-0.10	0.37
chla_range_min_4weeks_prior	-0.67	0.40
chla_weekly_range_1week_prior	-0.18	0.64
chla_weekly_range_2weeks_prior	-0.12	0.53
chla_weekly_range_3weeks_prior	-0.37	0.61
chla_weekly_range_4weeks_prior	-0.31	0.11
depth_max	-0.61	-0.47
depth_max_max_1week_prior	-0.70	-0.52
depth_max_max_2weeks_prior	-0.58	-0.44
depth_max_max_3weeks_prior	-0.65	-0.49
depth_max_max_4weeks_prior	-0.66	-0.48
depth_mean	-0.60	-0.46
depth_mean_avg_1week_prior	-0.62	-0.46
depth_mean_avg_2weeks_prior	-0.56	-0.41
depth_mean_avg_3weeks_prior	-0.59	-0.42
depth_mean_avg_4weeks_prior	-0.57	-0.41
depth_median	-0.60	-0.46
depth_median_median_1week_prior	-0.63	-0.47
depth_median_median_2weeks_prior	-0.56	-0.41
depth_median_median_3weeks_prior	-0.59	-0.42
depth_median_median_4weeks_prior	-0.56	-0.40
depth_min	-0.60	-0.46
depth_min_min_1week_prior	-0.47	-0.36
depth_min_min_2weeks_prior	-0.54	-0.38
depth_min_min_3weeks_prior	-0.53	-0.37
depth_min_min_4weeks_prior	-0.53	-0.37
depth_range	0.11	-0.10
depth_range_avg_1week_prior	-0.34	-0.28
depth_range_avg_2weeks_prior	-0.25	-0.34
depth_range_avg_3weeks_prior	-0.40	-0.40
depth_range_avg_4weeks_prior	-0.39	-0.31
depth_range_max_1week_prior	-0.36	-0.28
depth_range_max_2weeks_prior	-0.37	-0.36
depth_range_max_3weeks_prior	-0.42	-0.33
depth_range_max_4weeks_prior	-0.34	-0.23

depth_range_min_1week_prior	-0.15	-0.21
depth_range_min_2weeks_prior	-0.18	-0.31
depth_range_min_3weeks_prior	-0.16	-0.46
depth_range_min_4weeks_prior	-0.33	-0.31
depth_weekly_range_1week_prior	-0.37	-0.28
depth_weekly_range_2weeks_prior	-0.22	-0.33
depth_weekly_range_3weeks_prior	-0.41	-0.40
depth_weekly_range_4weeks_prior	-0.36	-0.33
do_max	-0.61	-0.19
do_max_max_1week_prior	-0.50	-0.10
do_max_max_2weeks_prior	-0.38	0.01
do_max_max_3weeks_prior	-0.25	0.18
do_max_max_4weeks_prior	-0.35	-0.09
do_mean	-0.67	-0.28
do_mean_avg_1week_prior	-0.79	-0.43
do_mean_avg_2weeks_prior	-0.68	-0.22
do_mean_avg_3weeks_prior	-0.68	-0.38
do_mean_avg_4weeks_prior	-0.66	-0.32
do_median	-0.67	-0.27
do_median_median_1week_prior	-0.80	-0.43
do_median_median_2weeks_prior	-0.70	-0.26
do_median_median_3weeks_prior	-0.70	-0.46
do_median_median_4weeks_prior	-0.66	-0.36
do_min	-0.68	-0.37
do_min_min_1week_prior	-0.83	-0.53
do_min_min_2weeks_prior	-0.73	-0.36
do_min_min_3weeks_prior	-0.80	-0.55
do_min_min_4weeks_prior	-0.64	-0.20
do_range	0.51	0.50
do_range_avg_1week_prior	0.72	0.57
do_range_avg_2weeks_prior	0.62	0.43
do_range_avg_3weeks_prior	0.82	0.62
do_range_avg_4weeks_prior	0.61	0.27
do_range_max_1week_prior	0.66	0.51
do_range_max_2weeks_prior	0.57	0.38
do_range_max_3weeks_prior	0.70	0.57
do_range_max_4weeks_prior	0.59	0.25
do_range_min_1week_prior	0.73	0.58
do_range_min_2weeks_prior	0.48	0.40
do_range_min_3weeks_prior	0.74	0.41
do_range_min_4weeks_prior	0.51	0.36
do_weekly_range_1week_prior	0.63	0.52
do_weekly_range_2weeks_prior	0.57	0.42
do_weekly_range_3weeks_prior	0.71	0.68
do_weekly_range_4weeks_prior	0.63	0.32
ec_max	0.45	0.45

ec_max_max_1week_prior	0.41	0.42
ec_max_max_2weeks_prior	0.38	0.39
ec_max_max_3weeks_prior	0.32	0.38
ec_max_max_4weeks_prior	0.28	0.34
ec_mean	0.44	0.48
ec_mean_avg_1week_prior	0.43	0.45
ec_mean_avg_2weeks_prior	0.37	0.39
ec_mean_avg_3weeks_prior	0.33	0.39
ec_mean_avg_4weeks_prior	0.28	0.36
ec_median	0.44	0.49
ec_median_median_1week_prior	0.44	0.47
ec_median_median_2weeks_prior	0.37	0.40
ec_median_median_3weeks_prior	0.33	0.41
ec_median_median_4weeks_prior	0.28	0.36
ec_min	0.41	0.49
ec_min_min_1week_prior	0.37	0.45
ec_min_min_2weeks_prior	0.34	0.39
ec_min_min_3weeks_prior	0.29	0.35
ec_min_min_4weeks_prior	0.30	0.37
ec_range	0.29	-0.25
ec_range_avg_1week_prior	0.20	-0.36
ec_range_avg_2weeks_prior	0.13	-0.40
ec_range_avg_3weeks_prior	0.31	-0.14
ec_range_avg_4weeks_prior	-0.08	-0.27
ec_range_max_1week_prior	0.11	-0.21
ec_range_max_2weeks_prior	0.07	-0.28
ec_range_max_3weeks_prior	0.29	0.04
ec_range_max_4weeks_prior	-0.17	-0.39
ec_range_min_1week_prior	0.21	-0.36
ec_range_min_2weeks_prior	-0.02	-0.24
ec_range_min_3weeks_prior	0.28	-0.23
ec_range_min_4weeks_prior	-0.01	0.01
ec_weekly_range_1week_prior	0.14	-0.14
ec_weekly_range_2weeks_prior	0.33	0.05
ec_weekly_range_3weeks_prior	0.23	0.24
ec_weekly_range_4weeks_prior	-0.02	-0.15
no3n_max	-0.94	-0.78
no3n_max_max_1week_prior	-0.94	-0.78
no3n_max_max_2weeks_prior	-0.96	-0.74
no3n_max_max_3weeks_prior	-0.96	-0.69
no3n_max_max_4weeks_prior	-0.96	-0.62
no3n_mean	-0.94	-0.77
no3n_mean_avg_1week_prior	-0.96	-0.75
no3n_mean_avg_2weeks_prior	-0.96	-0.71
no3n_mean_avg_3weeks_prior	-0.96	-0.67
no3n_mean_avg_4weeks_prior	-0.96	-0.62

no3n_median	-0.94	-0.77
no3n_median_median_1week_prior	-0.96	-0.75
no3n_median_median_2weeks_prior	-0.96	-0.72
no3n_median_median_3weeks_prior	-0.96	-0.67
no3n_median_median_4weeks_prior	-0.96	-0.61
no3n_min	-0.95	-0.78
no3n_min_min_1week_prior	-0.96	-0.76
no3n_min_min_2weeks_prior	-0.96	-0.69
no3n_min_min_3weeks_prior	-0.94	-0.64
no3n_min_min_4weeks_prior	-0.96	-0.61
no3n_range	-0.78	-0.60
no3n_range_avg_1week_prior	-0.84	-0.65
no3n_range_avg_2weeks_prior	-0.92	-0.64
no3n_range_avg_3weeks_prior	-0.92	-0.61
no3n_range_avg_4weeks_prior	-0.91	-0.56
no3n_range_max_1week_prior	-0.72	-0.65
no3n_range_max_2weeks_prior	-0.91	-0.67
no3n_range_max_3weeks_prior	-0.91	-0.64
no3n_range_max_4weeks_prior	-0.90	-0.57
no3n_range_min_1week_prior	-0.90	-0.73
no3n_range_min_2weeks_prior	-0.93	-0.67
no3n_range_min_3weeks_prior	-0.92	-0.60
no3n_range_min_4weeks_prior	-0.91	-0.56
no3n_weekly_range_1week_prior	-0.76	-0.62
no3n_weekly_range_2weeks_prior	-0.91	-0.73
no3n_weekly_range_3weeks_prior	-0.92	-0.66
no3n_weekly_range_4weeks_prior	-0.90	-0.59
orp_max	-0.41	-0.30
orp_max_max_1week_prior	-0.32	-0.15
orp_max_max_2weeks_prior	-0.41	-0.36
orp_max_max_3weeks_prior	-0.47	-0.28
orp_max_max_4weeks_prior	-0.48	-0.39
orp_mean	-0.45	-0.34
orp_mean_avg_1week_prior	-0.43	-0.27
orp_mean_avg_2weeks_prior	-0.44	-0.38
orp_mean_avg_3weeks_prior	-0.50	-0.33
orp_mean_avg_4weeks_prior	-0.52	-0.41
orp_median	-0.45	-0.34
orp_median_median_1week_prior	-0.42	-0.22
orp_median_median_2weeks_prior	-0.45	-0.38
orp_median_median_3weeks_prior	-0.47	-0.28
orp_median_median_4weeks_prior	-0.53	-0.42
orp_min	-0.46	-0.39
orp_min_min_1week_prior	-0.51	-0.44
orp_min_min_2weeks_prior	-0.46	-0.46
orp_min_min_3weeks_prior	-0.57	-0.44

orp_min_min_4weeks_prior	-0.54	-0.45
orp_range	0.53	0.53
orp_range_avg_1week_prior	0.66	0.55
orp_range_avg_2weeks_prior	0.66	0.45
orp_range_avg_3weeks_prior	0.76	0.45
orp_range_avg_4weeks_prior	0.75	0.38
orp_range_max_1week_prior	0.67	0.58
orp_range_max_2weeks_prior	0.62	0.44
orp_range_max_3weeks_prior	0.75	0.45
orp_range_max_4weeks_prior	0.76	0.34
orp_range_min_1week_prior	0.67	0.48
orp_range_min_2weeks_prior	0.46	0.31
orp_range_min_3weeks_prior	0.75	0.37
orp_range_min_4weeks_prior	0.53	0.36
orp_weekly_range_1week_prior	0.53	0.49
orp_weekly_range_2weeks_prior	0.45	0.36
orp_weekly_range_3weeks_prior	0.66	0.45
orp_weekly_range_4weeks_prior	0.68	0.36
ph_max	0.26	0.26
ph_max_max_1week_prior	0.38	0.39
ph_max_max_2weeks_prior	0.31	0.31
ph_max_max_3weeks_prior	0.46	0.43
ph_max_max_4weeks_prior	0.40	0.35
ph_mean	0.25	0.24
ph_mean_avg_1week_prior	0.24	0.18
ph_mean_avg_2weeks_prior	0.26	0.28
ph_mean_avg_3weeks_prior	0.35	0.29
ph_mean_avg_4weeks_prior	0.34	0.31
ph_median	0.24	0.24
ph_median_median_1week_prior	0.23	0.15
ph_median_median_2weeks_prior	0.26	0.26
ph_median_median_3weeks_prior	0.32	0.25
ph_median_median_4weeks_prior	0.34	0.31
ph_min	0.25	0.26
ph_min_min_1week_prior	0.06	0.03
ph_min_min_2weeks_prior	0.18	0.27
ph_min_min_3weeks_prior	0.16	0.12
ph_min_min_4weeks_prior	0.24	0.30
ph_range	0.17	0.17
ph_range_avg_1week_prior	0.41	0.36
ph_range_avg_2weeks_prior	0.37	0.28
ph_range_avg_3weeks_prior	0.55	0.37
ph_range_avg_4weeks_prior	0.49	0.27
ph_range_max_1week_prior	0.43	0.42
ph_range_max_2weeks_prior	0.37	0.25
ph_range_max_3weeks_prior	0.50	0.38

ph_range_max_4weeks_prior	0.46	0.21
ph_range_min_1week_prior	0.37	0.36
ph_range_min_2weeks_prior	0.33	0.26
ph_range_min_3weeks_prior	0.50	0.28
ph_range_min_4weeks_prior	0.43	0.28
ph_weekly_range_1week_prior	0.47	0.52
ph_weekly_range_2weeks_prior	0.34	0.24
ph_weekly_range_3weeks_prior	0.54	0.48
ph_weekly_range_4weeks_prior	0.46	0.29
turb_max	0.04	0.15
turb_max_max_1week_prior	-0.42	-0.17
turb_max_max_2weeks_prior	-0.45	0.25
turb_max_max_3weeks_prior	0.51	-0.94
turb_max_max_4weeks_prior	0.18	0.70
turb_mean	0.31	0.61
turb_mean_avg_1week_prior	-0.79	-0.60
turb_mean_avg_2weeks_prior	-0.10	-0.95
turb_mean_avg_3weeks_prior	0.63	-0.88
turb_mean_avg_4weeks_prior	0.96	-0.46
turb_median	0.28	0.62
turb_median_median_1week_prior	-0.94	-0.82
turb_median_median_2weeks_prior	-0.09	-0.91
turb_median_median_3weeks_prior	0.55	-0.92
turb_median_median_4weeks_prior	0.79	-0.66
turb_min	0.23	0.60
turb_min_min_1week_prior	-0.99	-0.92
turb_min_min_2weeks_prior	0.59	-0.84
turb_min_min_3weeks_prior	-0.81	-0.42
turb_min_min_4weeks_prior	0.55	-0.32
turb_range	0.03	0.09
turb_range_avg_1week_prior	-0.60	-0.37
turb_range_avg_2weeks_prior	-0.36	-0.41
turb_range_avg_3weeks_prior	0.66	-0.86
turb_range_avg_4weeks_prior	0.58	0.36
turb_range_max_1week_prior	-0.41	-0.15
turb_range_max_2weeks_prior	-0.47	0.29
turb_range_max_3weeks_prior	0.49	-0.95
turb_range_max_4weeks_prior	0.17	0.69
turb_range_min_1week_prior	-0.82	-0.64
turb_range_min_2weeks_prior	0.03	-0.97
turb_range_min_3weeks_prior	0.59	-0.90
turb_range_min_4weeks_prior	0.93	-0.69
turb_weekly_range_1week_prior	-0.37	-0.11
turb_weekly_range_2weeks_prior	-0.48	0.27
turb_weekly_range_3weeks_prior	0.50	-0.94
turb_weekly_range_4weeks_prior	0.17	0.69

wtemp_max	0.78	0.46
wtemp_max_max_1week_prior	0.79	0.46
wtemp_max_max_2weeks_prior	0.83	0.47
wtemp_max_max_3weeks_prior	0.82	0.48
wtemp_max_max_4weeks_prior	0.77	0.45
wtemp_mean	0.79	0.46
wtemp_mean_avg_1week_prior	0.80	0.46
wtemp_mean_avg_2weeks_prior	0.83	0.46
wtemp_mean_avg_3weeks_prior	0.82	0.46
wtemp_mean_avg_4weeks_prior	0.78	0.45
wtemp_median	0.79	0.46
wtemp_median_median_1week_prior	0.80	0.46
wtemp_median_median_2weeks_prior	0.83	0.46
wtemp_median_median_3weeks_prior	0.81	0.45
wtemp_median_median_4weeks_prior	0.77	0.45
wtemp_min	0.79	0.47
wtemp_min_min_1week_prior	0.80	0.46
wtemp_min_min_2weeks_prior	0.83	0.47
wtemp_min_min_3weeks_prior	0.82	0.45
wtemp_min_min_4weeks_prior	0.79	0.46
wtemp_range	0.18	0.09
wtemp_range_avg_1week_prior	0.37	0.26
wtemp_range_avg_2weeks_prior	0.54	0.37
wtemp_range_avg_3weeks_prior	0.58	0.37
wtemp_range_avg_4weeks_prior	0.60	0.41
wtemp_range_max_1week_prior	0.31	0.19
wtemp_range_max_2weeks_prior	0.43	0.29
wtemp_range_max_3weeks_prior	0.51	0.35
wtemp_range_max_4weeks_prior	0.47	0.36
wtemp_range_min_1week_prior	0.37	0.27
wtemp_range_min_2weeks_prior	0.53	0.30
wtemp_range_min_3weeks_prior	0.48	0.34
wtemp_range_min_4weeks_prior	0.51	0.24
wtemp_weekly_range_1week_prior	0.33	0.28
wtemp_weekly_range_2weeks_prior	0.32	0.18
wtemp_weekly_range_3weeks_prior	0.58	0.60
wtemp_weekly_range_4weeks_prior	0.17	0.13

Table S.15. GAM sensor period result geosmin, for data within the sensor deployment period (2022-2023).

Parameter	P Value	R2 Adj	Deviance Explained	Num Observations
DOY	0.004	0.201	0.286	45
precip	0.053	0.035	0.072	45
tempmax	0.032	0.147	0.206	45
tempmin	0.290	-0.001	0.024	45

temp	0.078	0.122	0.179	45
windgust	0.422	-0.012	0.013	45
windspeed	0.049	0.055	0.081	45
winddir	0.760	-0.021	0.002	45
cloudcover	0.532	0.016	0.049	45
solarradiation	0.000	0.194	0.236	45
solarenergy	0.000	0.197	0.237	45
sum_solarenergy_2weeks_prior	0.000	0.245	0.311	45
min_temp_2weeks_prior	0.039	0.050	0.085	45
max_windgust_2weeks_prior	0.960	-0.023	0.000	45
sum_precip_1week_prior	0.001	0.164	0.205	45
avg_windspeed_1week_prior	0.067	0.024	0.064	45
pH	0.035	0.131	0.213	42
turbidity	0.975	-0.025	0.000	42
toc	0.165	0.030	0.047	42
doc	0.140	0.040	0.055	41
tn	0.001	0.181	0.218	45
nh4n	0.602	-0.017	0.006	45
no3n	0.001	0.187	0.223	45
tp	0.092	0.051	0.065	45
po4p	0.220	0.000	0.029	45
manganesedis	0.014	0.141	0.141	42
manganese	0.006	0.189	0.176	42
chla	0.243	0.001	0.029	44
algaetot	0.763	-0.022	0.002	44
greenalgae	0.094	0.076	0.071	44
bluegreenalgae	0.315	-0.010	0.019	44
otheralgaetot	0.441	-0.010	0.014	42
tntp_ratio	0.012	0.187	0.237	45
nh4no3_ratio	0.426	-0.013	0.013	45
geosmin	0.000	1.000	1.000	45
mib	0.000	0.218	0.250	45
toc_trib	0.307	0.009	0.033	34
nh4n_trib	0.201	0.016	0.044	37
no3n_trib	0.003	0.181	0.215	37
tp_trib	0.116	0.038	0.066	37
nh4no3_ratio_trib	0.089	0.052	0.078	37
atemp_max_max_3weeks_prior	0.003	0.346	0.369	21
atemp_mean_avg_3weeks_prior	0.008	0.255	0.301	21
depth_max_max_1week_prior	0.004	0.234	0.312	22
do_range_avg_3weeks_prior	0.001	0.414	0.441	21
do_min_min_1week_prior	0.009	0.236	0.278	22
ec_max	0.012	0.148	0.234	23
ec_mean	0.008	0.176	0.261	23
no3n_max_max_2weeks_prior	0.012	0.539	0.651	12
no3n_mean_avg_2weeks_prior	0.025	0.477	0.580	12

no3n_range_avg_2weeks_prior	0.007	0.563	0.676	12
orp_range_avg_3weeks_prior	0.031	0.082	0.188	21
ph_range_avg_3weeks_prior	0.063	0.052	0.143	21
wtemp_mean_avg_2weeks_prior	0.011	0.168	0.252	22
ec_min	0.006	0.184	0.271	23
do_weekly_range_3weeks_prior	0.000	0.441	0.472	21
no3n_mean	0.001	0.617	0.764	12
no3n_max_max_1week_prior	0.001	0.676	0.804	11
no3n_weekly_range_2weeks_prior	0.002	0.583	0.725	12
depth_mean	0.007	0.176	0.265	23
depth_mean_avg_1week_prior	0.012	0.175	0.250	22
orp_range_avg_1week_prior	0.002	0.223	0.340	22
ph_weekly_range_1week_prior	0.008	0.219	0.285	22
ph_max_max_1week_prior	0.043	0.110	0.171	22
wtemp_weekly_range_3weeks_prior	0.003	0.399	0.377	21
orp_range_max_1week_prior	0.001	0.380	0.431	22
wtemp_max_max_3weeks_prior	0.008	0.241	0.295	21
wtemp_mean	0.008	0.168	0.254	23

Table S.16. GAM sensor period result 2-MIB, for data within the sensor deployment period (2022-2023).

Parameter	P Value	R2 Adj	Deviance Explained	Num Observations
DOY	0.000	0.583	0.631	45
precip	0.980	-0.023	0.000	45
tempmax	0.001	0.325	0.348	45
tempmin	0.000	0.409	0.421	45
temp	0.000	0.502	0.514	45
windgust	0.001	0.369	0.443	45
windspeed	0.022	0.078	0.111	45
winddir	0.227	0.070	0.097	45
cloudcover	0.676	-0.019	0.004	45
solarradiation	0.013	0.114	0.124	45
solarenergy	0.014	0.114	0.125	45
sum_solarenergy_2weeks_prior	0.000	0.261	0.273	45
max_windgust_2weeks_prior	0.052	0.054	0.078	45
sum_precip_1week_prior	0.872	-0.023	0.001	45
avg_windspeed_1week_prior	0.208	0.016	0.039	45
pH	0.593	-0.018	0.007	42
turbidity	0.091	0.045	0.060	42
toc	0.000	0.666	0.692	42
doc	0.000	0.543	0.525	41
tn	0.000	0.527	0.520	45
nh4n	0.007	0.125	0.144	45
no3n	0.000	0.522	0.515	45
tp	0.000	0.230	0.208	45

po4p	0.818	-0.022	0.001	45
manganesedis	0.000	0.606	0.605	42
manganese	0.000	0.444	0.437	42
chla	0.267	0.007	0.032	44
algaetot	0.944	-0.024	0.000	44
greenalgae	0.231	0.015	0.032	44
bluegreenalgae	0.809	-0.022	0.001	44
otheralgaetot	0.011	0.149	0.160	42
tntp_ratio	0.000	0.503	0.494	45
nh4no3_ratio	0.001	0.333	0.327	45
geosmin	0.001	0.185	0.185	45
mib	0.000	0.991	0.984	45
toc_trib	0.004	0.280	0.309	34
nh4n_trib	0.058	0.204	0.282	37
no3n_trib	0.065	0.100	0.110	37
tp_trib	0.141	0.027	0.054	37
nh4no3_ratio_trib	0.478	-0.012	0.013	37
atemp_max_max_3weeks_prior	0.000	0.699	0.800	21
atemp_mean_avg_3weeks_prior	0.000	0.628	0.686	21
depth_max_max_1week_prior	0.000	0.458	0.467	22
do_range_avg_3weeks_prior	0.000	0.731	0.766	21
do_min_min_1week_prior	0.000	0.668	0.720	22
ec_max	0.024	0.170	0.190	23
ec_mean	0.032	0.169	0.187	23
no3n_max_max_2weeks_prior	0.000	0.916	0.958	12
no3n_mean_avg_2weeks_prior	0.000	0.905	0.941	12
no3n_range_avg_2weeks_prior	0.000	0.939	0.977	12
orp_range_avg_3weeks_prior	0.001	0.560	0.608	21
ph_range_avg_3weeks_prior	0.018	0.322	0.387	21
wtemp_mean_avg_2weeks_prior	0.000	0.738	0.803	22
ec_min	0.038	0.135	0.158	23
do_weekly_range_3weeks_prior	0.001	0.547	0.606	21
no3n_mean	0.000	0.861	0.901	12
no3n_max_max_1week_prior	0.000	0.857	0.897	11
no3n_weekly_range_2weeks_prior	0.000	0.821	0.852	12
depth_mean	0.001	0.340	0.375	23
depth_mean_avg_1week_prior	0.004	0.354	0.376	22
orp_range_avg_1week_prior	0.005	0.269	0.306	22
ph_weekly_range_1week_prior	0.029	0.159	0.189	22
ph_max_max_1week_prior	0.069	0.089	0.130	22
wtemp_weekly_range_3weeks_prior	0.011	0.209	0.251	21
orp_range_max_1week_prior	0.001	0.376	0.398	22
wtemp_max_max_3weeks_prior	0.000	0.692	0.767	21
wtemp_mean	0.000	0.706	0.764	23
min_temp_2weeks_prior_REML	0.000	0.598	0.624	45