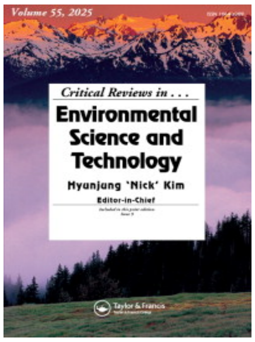


Critical Reviews in Environmental Science and Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/best20

Integrating sensor data and machine learning to advance the science and management of river carbon emissions

Lee E. Brown, Taylor Maavara, Jiangwei Zhang, Xiaohui Chen, Megan Klaar, Felicia Orah Moshe, Elad Ben-Zur, Shaked Stein, Richard Grayson, Laura Carter, Elad Levintal, Gideon Gal, Pazit Ziv, Frank Tarkowski, Devanshi Pathak, Kieran Khamis, José Barquín, Hemma Philamore, Misael Sebastián Gradilla-Hernández & Shai Arnon

To cite this article: Lee E. Brown, Taylor Maavara, Jiangwei Zhang, Xiaohui Chen, Megan Klaar, Felicia Orah Moshe, Elad Ben-Zur, Shaked Stein, Richard Grayson, Laura Carter, Elad Levintal, Gideon Gal, Pazit Ziv, Frank Tarkowski, Devanshi Pathak, Kieran Khamis, José Barquín, Hemma Philamore, Misael Sebastián Gradilla-Hernández & Shai Arnon (2025) Integrating sensor data and machine learning to advance the science and management of river carbon emissions, Critical Reviews in Environmental Science and Technology, 55:9, 600-623, DOI: 10.1080/10643389.2024.2429912

To link to this article: https://doi.org/10.1080/10643389.2024.2429912

Published online: 24 Nov 2024.

CRITICAL REVIEWS IN ENVIRONMENTAL SCIENCE AND TECHNOLOGY
2025, VOL. 55, NO. 9, 600–623
https://doi.org/10.1080/10643389.2024.2429912

Taylor & Francis
Taylor & Francis Group

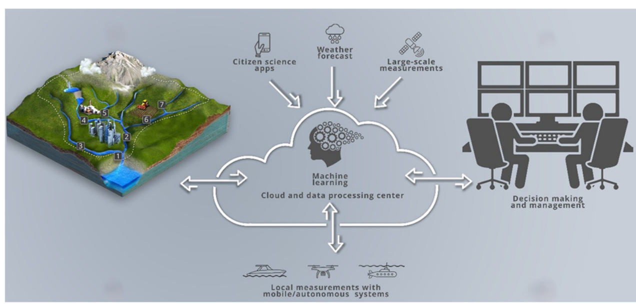
OPEN ACCESS Check for updates

Integrating sensor data and machine learning to advance the science and management of river carbon emissions

Lee E. Brown[a], Taylor Maavara[a], Jiangwei Zhang[b], Xiaohui Chen[b], Megan Klaar[a], Felicia Orah Moshe[c], Elad Ben-Zur[d], Shaked Stein[d], Richard Grayson[a], Laura Carter[a], Elad Levintal[e], Gideon Gal[d], Pazit Ziv[a], Frank Tarkowski[f], Devanshi Pathak[g], Kieran Khamis[h], José Barquín[i], Hemma Philamore[j], Misael Sebastián Gradilla-Hernández[k] and Shai Arnon[e]

[a]School of Geography and water@leeds, University of Leeds, Leeds, UK; [b]School of Civil Engineering, University of Leeds, Leeds, UK; [c]Department of Soil Conservation, Soil Erosion Research Station, Ministry of Agriculture, Bet Dagan, Israel; [d]Kinneret Limnological Laboratory, Israel Oceanographic & Limnological Research, Migdal, Israel; [e]Zuckerberg Institute for Water Research, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Israel; [f]Yorkshire Water Services Ltd, Bradford, UK; [g]Department of Aquatic Ecosystem Analysis and Management, Helmholtz Centre for Environmental Research-UFZ, Magdeburg, Germany; [h]School of Geography, Earth & Environmental Science, University of Birmingham, Birmingham, UK; [i]IHCantabria – Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander, Spain; [j]School of Engineering, Mathematics and Technology, University of Bristol, Bristol, UK; [k]Escuela de Ingenieria y Ciencias, Tecnologico de Monterrey, Nuevo Mexico, Mexico

ABSTRACT
Estimates of greenhouse gas emissions from river networks remain highly uncertain in many parts of the world, leading to gaps in global inventories and preventing effective management. In-situ sensor technology advances, coupled with mobile sensors on robotic sensor-deployment platforms, will allow more effective data acquisition to monitor carbon cycle processes influencing river CO_2 and CH_4 emissions. However, if countries are to respond effectively to global climate change threats, sensors must be installed more strategically to ensure that they can be used to directly evaluate a range of management responses across river networks. We evaluate how sensors and analytical advances can be integrated into networks that are adaptable to monitor a range of catchment processes and human modifications. The most promising data analytics that provide processing, modeling, and visualizing approaches for high-resolution river system data are assessed, illustrating how multi-sensor data coupled with machine learning solutions can improve both proactive (e.g. forecasting) and reactive (e.g. alerts) strategies to better manage river catchment carbon emissions.



Data measurement and integration can be used to advance assessments and management of river carbon dynamics and water quality.

KEYWORDS carbon dioxide; machine learning; methane; metabolism; sensors; water quality

HANDLING EDITORS Hyunjung (Nick) Kim

CONTACT Lee E. Brown l.brown@leeds.ac.uk School of Geography and water@leeds, University of Leeds, Leeds, UK.

Introduction

Despite the importance of river systems for water supply, and other ecosystem services such as regulation of nutrient cycles (e.g. nitrogen and phosphorus) and supporting fisheries, they are highly degraded ecosystems due to anthropogenic stressors such as modified flows, urbanization, agriculture and wastewater (Vörösmarty et al., 2010). By altering physical, chemical, and biological components of river systems, anthropogenic interventions play an important role influencing climate change through greenhouse gas (GHG) emissions. River systems globally contribute net estimated annual CO_2 emissions equivalent to 20-24% of fossil fuel emissions, 35-65% of the CH_4 emissions from all sources, and 4–5% of N_2O total emissions (Battin et al., 2023; Friedlingstein et al., 2022; Rosentreter et al., 2021). However, global estimates of river net GHG emissions remain highly uncertain, due to sparse data availability and inconsistent monitoring practices, perpetuating large gaps in international emissions inventories and preventing effective management (Rudee & Phillips, 2021). In this review we focus on CO_2 and CH_4 as gases that are most often quantified in river systems due to the availability of multiple sensors for measuring carbon cycle processes. Other GHGs with natural sources, notably N_2O and SF6, typically have to be measured with gas analyzers although dissolved gas sensors for the former are beginning to emerge (UNISENSE, 2024). The Paris Agreement (UNFCCC, 2018) signed at the UN Climate Change Conference of Parties (COP21) in 2015 recognized the crucial need to quantify GHG sources and sinks that have not yet been adequately quantified. More effective river catchment monitoring and management are needed urgently for countries to respond effectively to global climate change threats by better managing carbon emissions.

Quantifying aquatic carbon cycle processes is challenging. Processes such as photo-oxidation, metabolism (production, respiration), and methanogenesis can be estimated from dissolved gas measurements, organic matter degradation assays, GHG emissions (e.g. floating chambers), or from dissolved gas concentrations relative to atmospheric concentrations (Aho et al., 2021; Appling et al., 2018; Duc et al., 2013). However, most studies have collected short-duration datasets in situ, at small numbers of sites, with low temporal resolution. Estimates of photo-oxidation and decomposition with experimental manipulations are also typically resolved at weekly-to-monthly timescales. Even where daily-to-weekly sampling takes place, it often occurs at selected locations during daylight hours or misses important events such as flow peaks (Bieroza et al., 2023). Thus, we lack a clear understanding of how river stressors and management activities influence emission "hotspots" in space, and/or "hot moments" in time (W. Zhang et al. 2021), risking either over- or under-estimation of emissions. Recent reviews and opinion articles have broadly outlined a need for global river observation systems for river carbon monitoring (Battin et al., 2023; Dean & Battin, 2024) but lacked details on how these networks could be implemented. Here we evaluate how recent advances in autonomous (field deployable and wireless) sensor networks, and robotic mobile sensing platforms, can be harnessed to meet this requirement by combining high-frequency, continuous data at multiple locations, with machine learning (ML) models to improve carbon emission estimates and overall water management in river networks.

The emergence of sensor technologies for high-resolution space/time monitoring offers the potential to evaluate fundamental linkages between hydrological regime, physicochemical conditions, and nutrient dynamics to fill knowledge gaps in understanding processes related to carbon emissions. Links between river physical properties, network structure, and ecosystem carbon cycle parameters, including metabolism, have advanced notably with Cole et al. (2007) concept of "leaky pipes" for carbon loss along the land-ocean aquatic continuum, and the Pulse-Shunt Concept, which added transport vs reaction timescales related to flow (Raymond et al., 2016). Wollheim et al. (2018) proposed a similar River Network Saturation concept, describing how river networks become saturated with carbon at high flows, particularly in low-order streams, where terrestrial carbon is "pulsed" to river networks and "shunted" downstream because high flows restrict time for uptake reactions in quantifiable amounts. Thus, most annual downstream carbon export occurs during a small number of high-flow events (Raymond et al., 2016). At low flows, particularly in high-order rivers, carbon uptake fluxes and subsequent emissions are much higher as transport timescales are long and reactions can occur by photo-mineralization and co-metabolism on bio-aggregates (Battin et al., 2008). Continuous measurements of dissolved oxygen (DO) have enabled many of these advances in understanding river carbon cycling processes of primary production and respiration, but the spatial distribution of monitoring systems remains limited and globally unbalanced. For example, across the United States, the relatively widespread availability of sensor data maintained by the US Geological Survey (USGS) (Figure 1), has promoted an understanding of key drivers of river metabolism (Appling et al., 2018). A range of datasets are also collected in regional initiatives (e.g. Figure 1b, c) yet for large parts of the world, including much of the global south, we still have only patchy knowledge of the parameters needed to quantify carbon transformations and emissions, or data collected are not open access (Dean & Battin, 2024). Even in countries with advanced sensor networks, there are still large gaps spatially between sensor locations (Figure 1b), and high-order, poorly mixed rivers, which present challenges to developing representative datasets, unless multiple sensors are deployed across river cross sections.

Sensor network developments can improve our understanding of spatial and temporal carbon dynamics significantly (Segatto et al., 2023) but cost prevents monitoring all rivers. Coupling

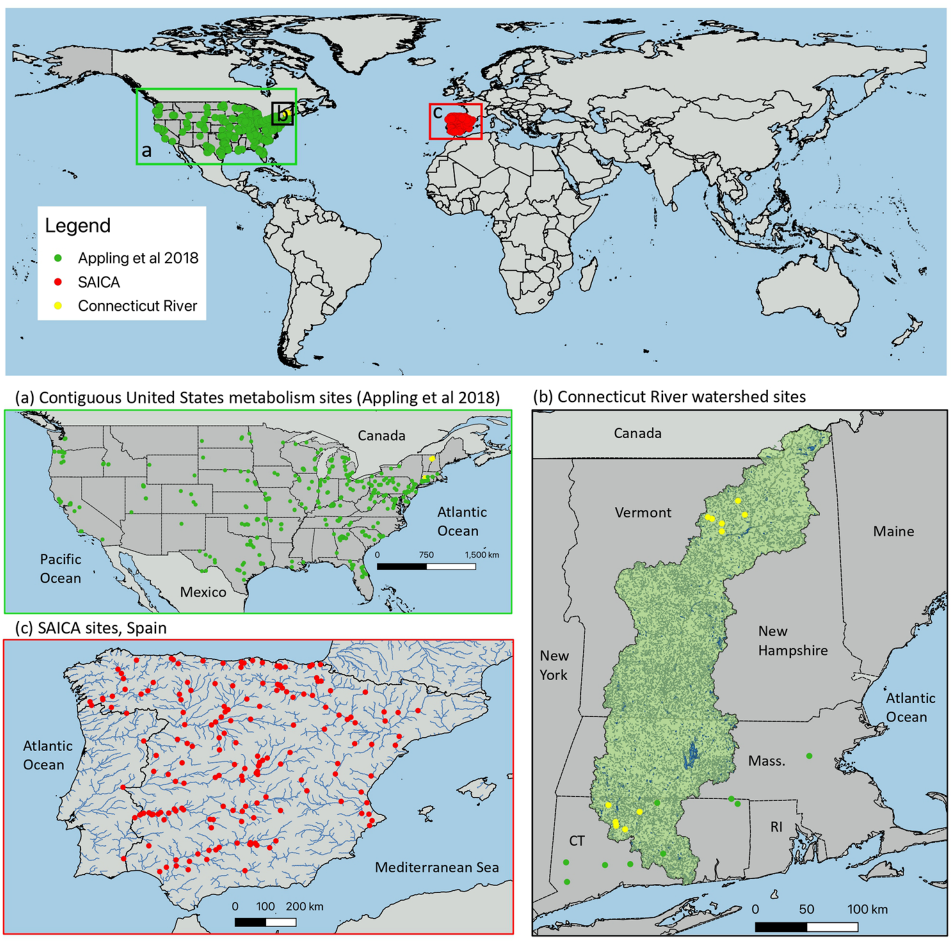


Figure 1. Distribution of river monitoring stations with sensors suitable for developing metabolism estimates and carbon emissions has a strong spatial bias. Examples of the most dense nationwide networks are: (a) USGS water quality monitoring sites in the continental USA used to predict river metabolism; (b) Spanish Environmental Department Water Quality Automatic Information System (SAICA); (c) a catchment-scale monitoring network in the Connecticut River, NE USA (Hosen et al., 2021).

sensor developments with advances in fixed sensor technology and data analytics, as well as mobile robotics and ML, will be vital to achieve spatially continuous data and interpolate spatially explicit datasets to derive whole catchment understanding (Khandelwal et al., 2023; O'Grady et al., 2021). By automating sensors using computer science advances and telemetry systems, it is becoming possible to monitor, in near real-time, how aquatic ecosystems are functioning. Additionally, the Internet of Things (IoT) offers significant potential in delivering up-to-date water quality data with a high level of precision and accuracy, enabling the detection of even minor fluctuations in water quality. IoT facilitates the connection of various instruments, including electronic devices and sensors, utilizing the communication infrastructure and cloud computing resources already in place (Amador-Castro et al., 2024). This offers the potential to validate existing carbon dynamic scientific models and develop the next generation of catchment-scale numerical predictive models. There is now a potential for a step-change in adaptive management, moving away from current low-resolution, relatively slow turnover data collection, with delayed analytics that impede effective decision-making, to faster and more accurate workflows, even at national scales. This will subsequently enable scientists to advance emission quantifications at national to global levels and develop intervention plans.

In this review, we evaluate the potential for using sensor data and machine learning to advance river carbon cycle processes and emission management, both responsively to stressor events and proactively to enhance the management of both water resource security and downstream river system services. Taking into consideration the key drivers of river carbon processes and emissions, we demonstrate how recent technological advances in the development and implementation of sensor networks for river catchment management can be harnessed to improve knowledge of aquatic processes. We examine how sensor and analytics advances offer new opportunities to develop strategic monitoring networks that can capture impacts resulting from a range of catchment processes and human modifications. We illustrate the benefits of incorporating emerging, affordable sensor technologies, and novel robotic sensor-deployment technologies, which allow for high-resolution monitoring, and explain how a variety of water quality parameters can be used to develop causal relationships between drivers and response variables. We then assess the most promising analytical approaches and methods for processing, modeling, and visualizing high-resolution river system data, demonstrating how novel applications of sensor networks coupled with artificial intelligence (AI) solutions could be developed.

Advancing river management with sensor networks and data analytics

Many traditional methods for monitoring river systems are resource-intensive and deliver results sometimes weeks-months after sampling and ecosystem changes occur (e.g. biological sampling followed by laboratory identification, then analysis/interpretation, or; "snapshot" sampling of water chemistry (Dean & Battin, 2024)). Manually operated sensors furthermore offer only a snapshot of temporal dynamics. Both delays and low-resolution data can result in less effective management responses, such as detecting pollution incidents, or optimizing water systems where tradeoffs between water supply and environmental needs are required. In contrast, there is increasing availability of affordable, robust, and high-resolution sensors, coupled with distributed data transfer systems (e.g. LoRaWAN - long range wide area network) and the array of data analytics solutions. If we are to truly revolutionize water resource management, river monitoring needs to embrace the collation of large, integrated datasets in complete packages rather than considering layered approaches (Dean & Battin, 2024) that re-iterate long-standing collection protocols. For instance, IoT devices can incorporate software sensors (such as those based on machine learning) for predicting a range of water quality parameters based on the information from physical sensors (Ba-Alawi et al., 2023), reducing monitoring costs. River ecosystem metabolism for example, which can be quantified routinely and continuously using optical measurements of DO, would be a core carbon cycle process measurement which has been found to respond consistently to environmental change with a high sensitivity, including detecting effects of river restoration practices (Ferreira et al., 2020), wastewater treatment upgrades (Arroita et al., 2019) and stressor events such as sedimentation (Aspray et al., 2017). When combined with a range of other sensor-based measurements, it offers significant potential for assessing the impacts of river system responses to human modification.

Surface and ground water quality monitoring is an established approach for determining river management objectives. There are numerous examples of statutory water quality monitoring programmes that target both below ground and surface water environments (e.g. EU Water Framework Directive), but monitoring to establish the status of the hyporheic zone is often limited or non-existent. This is an important gap in our knowledge as hyporheic zones are where surface and groundwater sources mix in the streambed creating strong physico-chemical gradients that influence a range of ecosystem functions, including nutrient cycling and biogeochemical reactions (Lewandowski et al., 2019). This is especially true when considering GHG exchange since the water column serves as the interface zone for exchange between the sedimentary environment (where many reactions occur) to the atmosphere. High-frequency measurements, such as temperature and DO, in the hyporheic zone are needed to integrate the heterogenous nature of the surface and subsurface environment (Klaar et al., 2020) and reaction rates (Shelley et al., 2017), thereby enabling quantification of biogeochemical processes in hyporheic zones including subsequent release of GHG from aquatic environments.

Reactive management responses to sensor data

Coupling telemetered sensor networks to data analytics solutions will be needed to enable the development of dynamic visualization dashboards, providing environmental managers with unprecedented insights into the real-time status of the whole river network. These sensor systems also present new opportunities for the democratization of catchment data with public-facing web-hosted applications. By engaging the public in the process, initiatives that improve insights into water quality by enhancing the level of detail and coverage in both space and time supplement the data generated by scientists and government organizations. To achieve this, it is crucial to establish and distribute appropriate and consistent protocols to the public (Amador-Castro et al., 2024). With additional potential for alerting citizen scientists, reliable information on water quality could be obtained quickly during events, increasing the spatiotemporal resolution and complementing the data produced by scientists and government institutions. As ML methods improve, the ability to upscale from localized data collection points in space, and to robustly infer system dynamics over time where data gaps exist (Segatto et al., 2021), offers potential for significant improvements in both reactive and proactive management (Figure 2). Predictions of future conditions (forecasting) will become possible, similar to recent advances applied to standing freshwaters (Lofton et al., 2023), enabling improved responses to future problems.

Sensor data can be used directly or aggregated to develop metrics for evaluating the ecological status of a river section. Such data will be used together with information on water quality and discharge to support decisions on water management, especially by elucidating links between water quality, ecosystem respiration, and carbon emission. For example, abstractors using river water for drinking water supply can identify contamination issues, such as high dissolved organic carbon (DOC) concentrations upstream, thus avoiding problems whereby disinfection byproducts make water unsuitable for human consumption (Valdivia-Garcia et al., 2019). Other examples include water utilities and hydropower companies that withdraw, store, and redistribute water around river systems facing management challenges related to altered water quality (Gillespie et al., 2015). Such approaches are already being tested, by diverting episodic events with elevated DOC in raw water sources away from water treatment works (Yorkshire Water, 2023). Although sensor networks can be costly to implement and maintain, wider operational cost-savings can be made by integrating forecasting into ML architecture to feed back to field sensors and samplers to collect higher resolution data which would otherwise need manual intervention. For

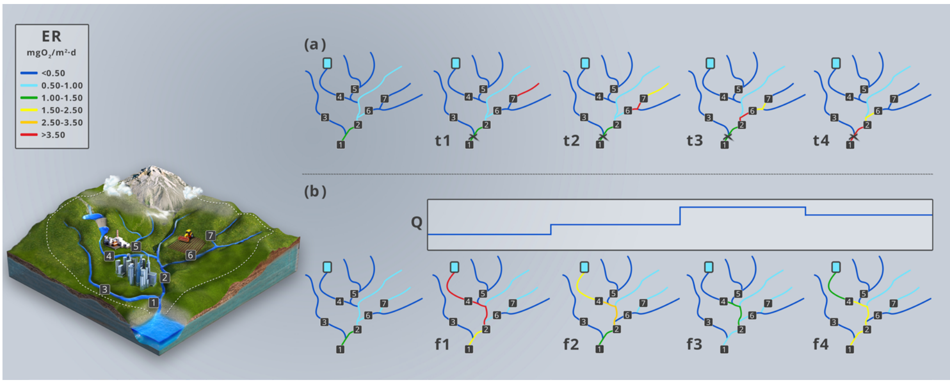


Figure 2. Distributed sensor networks can be developed to support water resource management: (a) *REACTIVE*; a river catchment with sensor locations denoted by numbers (1-7) spanning river channels. At t1, a stressor (e.g. organic pollution, sedimentation) appears (upstream of location 7) leading to enhanced ecosystem respiration (ER). Real-time analytics and visualization allow pollutant tracking through t2-t4, enabling water abstraction (denoted by x) to be deactivated at t3. (b) *PROACTIVE*; a river catchment with a large headwater reservoir. Hydrograph shows discharge (Q) scenarios f1-4. Low flow f1 elevates ER in the mainstem. With a regulatory target of ER 1-2.5, water release in f2 modifies only the segment below the reservoir. Excessive water release in f3 leads to overshoot of targets, allowing an optimal solution in f4 to tradeoff ecosystem recovery and water supply.

example, enhanced data collection during contamination events could be used to support regulator investigations, and during storms where runoff peaks are often missed, for enhanced understanding of water quality and carbon cycle dynamics.

Regulators and decisionmakers need access to high-quality data to develop, monitor, and enforce catchment management plans and legislation, and identify areas where persistent problems highlight the need for restoration, such as through payment for ecosystem service or nature-based solution initiatives. Additionally, managers of agricultural basins, which are recognized as a leading source of global water contamination (Liu et al., 2022) need evidence to manage and reduce the effects of sediment loads and adsorbed contaminants originating from soil erosion, and the use of agrochemicals (nutrients, herbicides, pesticides), all of which can lead to elevated GHG emissions from rivers (Xiao et al., 2021). By pinpointing river sections or sub-catchments suffering from stressors, prioritized and targeted management practices can meet multiple objectives to reduce emissions as part of the water-energy-food nexus in global resource systems.

Proactive management responses to sensor data

Proactive uses of sensor networks and analytics portals will benefit from long-term management planning through research and adaptive management. For example, experimental campaigns can be initiated to optimize management by modifying environmental flows from reservoirs (Figure 2b). At present, reservoir operators release water to support downstream ecosystems, aiming to maintain the quantity and quality of water, based on the taxonomic or behavioral response of targeted biological groups, such as fish and invertebrates (Gillespie et al., 2015). However, these flows can additionally modify downstream water quality, such as temperature, which is a strong control on carbon cycle processes (Yvon-Durocher et al., 2011). The water release may also alter emissions from previously dry sediments (Pérez-Calpe et al., 2022), and transfer dissolved GHG from in-reservoir processing (Shi et al., 2023) to modify downstream emissions (Guérin et al., 2006). With the ability of sensor networks to provide rapid insights into downstream river ecosystem responses to changes in outflow volume, reservoir managers could more effectively balance water supply requirements with minimizing downstream ecosystem damage and emissions.

GHGs are known to be emitted from all freshwaters and make up a large component of the carbon flux (Butman et al., 2016) but the lack of direct accounting for many of these systems, despite studies showing their important role in carbon budgets both naturally and when modified, can now be remedied with enhanced environmental data collection. This work is vital to inform improved Life Cycle Analysis (LCA), for example when evaluating GHG emissions from infrastructure projects such as hydropower systems (Hertwich, 2013). However, most LCA analyses currently ignore the role of the river itself, or are conducted only local to the development. Data on GHG emission from river networks need to be matched carefully with the spatio-temporal domain of the accounting methods, or vice-versa (Feng, 2005). Hence, investments in freshwater system monitoring infrastructure are needed from governments and businesses, otherwise strategies to reach net-zero are likely to be hampered (Chen et al., 2022). Following recent IPCC inventory refinements (IPCC, 2019), emissions from managed inland waters (e.g. farm ponds, reservoirs, and their outflows) now need to be quantified, which is adding some impetus to data collection, but for river systems it is also necessary to consider business operations effects "offsite" or "downstream," such as lengths of watercourses influenced by upstream contaminant inputs (Hu et al., 2018) or flow modifications (Shi et al., 2023). The integration of sensor networks with ML models will be a key step toward meeting this need for whole catchment understanding and improved management.

Developing sensor networks to enhance river catchment management

Drivers of river carbon cycle processes

Understanding the drivers of carbon cycling in rivers is needed to predict the effects of modification by stressors such as warming, land use changes, and flow regulation (Bernhardt et al., 2018). Unlike terrestrial systems, carbon processing in rivers may not be synchronized with subsequent emissions of the produced GHG due to the dynamic nature and spatiotemporal variability in their physicochemical and biological characteristics (Dodds et al., 2013). Continuous measurements of the key drivers of metabolism in rivers (e.g. hydrological conditions, light, temperature, organic matter availability, and nutrient concentrations) are needed therefore to compare and pinpoint their relative importance under different stressor regimes. Additionally, because metabolism and physicochemical drivers act at multiple spatial scales, from local (riparian vegetation, channel morphology) to regional (climate, topography), and vary along the river network (Alberts et al., 2017), their combined impacts can only be examined by in-situ sensor networks and remotely-sensed data products. For example, the extra dimensionality offered by mobile robots with on-board sensors offers a potential solution to measuring spatial variation in such parameters along the course of a river.

River hydrology plays a significant role in shaping metabolism in rivers due to its control on ecosystem structure and functioning (Hosen et al., 2019; Maavara et al., 2023; VON Schiller et al., 2008). In their study based on sensor data from 222 US rivers, Bernhardt et al. (2022) found light and flow stability (and their interaction) to be key controls on primary production and respiration. Studies in temperate rivers have shown that, in addition to the obvious seasonal drivers of Gross Primary Production (GPP) (i.e. light availability, including canopy shading, and temperature), GPP's dependence on flow should be considered in the context of river size (Hosen et al., 2019). In large rivers, GPP is maximized at low flows, but reduced in high flows due to short water residence times and high turbidity obscuring light availability (Pathak et al., 2022; Roberts et al., 2007), while there is little flow-related change on GPP fluxes in smaller rivers (Hosen et al., 2021; Maavara et al., 2023). The dependence of ecosystem respiration (ER) on flow is somewhat less straightforward; low flows may reduce benthic production due to riverbed drying, but ER can increase after flow resumes, fueled by the accumulation of terrestrial organic matter on the dry riverbed (Acuña et al., 2005). Maavara et al. (2023) showed that ER was generally maximized in a temperate forested watershed close to median flows when water residence times allowed ample time for carbon uptake to occur, with higher flows resulting in a deeper and wider water column allowing for more DOC availability and more uptake due to larger water column volume.

Although light availability is a key driver of primary production (Savoy & Harvey, 2021), it is not straightforward to model GPP as other factors impact river autotrophs such as turbidity, vegetation, nutrient availability/stoichiometry, and shading (Behrenfeld & Falkowski, 1997). Light and temperature models can be developed for whole river networks by calibrating ML approaches from local sensor data and scaling the findings using remote sensing products such as land use/cover classifications and digital elevation models for topographical information (Maavara et al., 2023; Segatto et al., 2021). Drones with attached sensors could be utilized to build and validate such models with high spatial resolution data. Light availability can also modify temperature (Nebgen & Herrman, 2019), which has a stronger control on ER compared to GPP, indicating a possibility of higher CO_2 emissions from rivers with climate warming (Demars et al., 2011). However, the impact of warming on emissions is still less predictable compared to estimates derived from metabolic theory (Battin et al., 2023). Incoming solar radiation can also mineralize DOC to inorganic forms and this must be considered alongside aquatic biological processes (Maavara et al., 2021). Indeed, recent research has shown that the magnitude of photo-mineralization during low flows as well as in winter often dramatically exceeds ER fluxes. Failing to consider year-round photo-mineralization fluxes may vastly underestimate the total magnitude of DOC uptake and CO_2 production (Maavara et al., 2023).

While GPP and ER are sensitive to nutrient loading, studies have yielded equivocal results regarding the impact of nutrient loading. Some studies have reported an increase in GPP and ER due to nitrogen and/or phosphorus loading (Kominoski et al., 2018), whereas others have suggested that nutrient concentrations may be only secondary drivers due to the effects of light and food web structure (Dodds & Cole, 2007). Conversely, some studies have found evidence for reverse causality, where metabolism variations strongly control riverine nutrient dynamics (Jarvie et al., 2018; Pathak et al., 2022). The development of sensor networks, and sensor-deployment technologies, designed specifically to monitor river ecosystem carbon dynamics will therefore enable the relative importance of multivariate drivers of metabolic processes and their feedback to be understood in far more detail. Moreover, bidirectional relationships between water quality and biogeochemical cycles could eventually be explained and predicted through ML using both water quality information from sensors alongside next generation sequencing data focused on microbial communities and their biogeochemical processes (Díaz-Torres et al., 2022; Fell et al., 2021) such as sulfur, nitrogen, phosphorus, and carbon metabolism.

Sensor networks for river system monitoring

Catchment management aimed at improving freshwater quality and reducing carbon emission is complicated due to multiple transport pathways that convey water and a wide range of contaminants into rivers (Khamis et al., 2018). These include point sources such as industrial and municipal wastes, and non-point contributions such as agriculture. Identifying hotspot areas (both sources and impacts) is a critical first step in developing adequate intervention measures to improve water quality. However, monitoring is needed to evaluate the effects of these intervention efforts, protect water quality, and meet regulations (Lofton et al., 2023). Sensor networks provide the potential to meet these aims, but operational water quality monitoring programs globally are commonly based on fixed sampling points with periodic manual collection of "grab" samples and subsequent laboratory analysis for targeted parameters. Due to limitations of personnel, equipment, and access this type of sampling can provide good spatial snapshots of river conditions at the time of sampling (Meyer et al., 2019) and manual water/gas samples are important to cross-validate sensor outputs, but is difficult to implement across entire catchments (Xing et al., 2013). These sampling approaches also largely miss sporadic extreme events, such as contaminant releases or stormflows (Charriau et al., 2016). In response to deficiencies in capturing event-based changes in river ecosystem properties, and with the emergence of more reliable sensor technology, high-frequency monitoring using field deployable sensors and actuators is increasing (Bieroza et al., 2023; Blaen et al., 2016). Autonomous and remotely operated robotic surface vehicles with on-board sensors have increased the achievable spatial resolution of field-deployed water quality sensors (Lee et al., 2023), and show great potential to improve detection of, and response to, short-term changes in river environments (Powers et al., 2018). For example, localization of a pollution hotspot could trigger reactive behaviors, such as increasing the resolution of data collection or tracking concentration gradients.

In-situ automated systems with multiple sensors that measure at high-frequency (typically 15–60 min resolution but can vary depending on the application) can be used to deliver near real-time data (Meyer et al., 2019; Singh et al., 2022). Various sensors can be deployed to quantify carbon cycling or to supply information on physicochemical drivers (Table 1). However, to advance catchment-scale carbon management, networks of these automated systems (i.e. sensor nodes (Figure 3)) are needed to pinpoint areas, such as those with high emissions, and to track event propagation through river basins (Zia et al., 2013). Further potential for enhancing the dimensionality of environmental data is emerging from the development of autonomous robotic platforms to deploy sensors in parts of river systems that are difficult to access. The integration of these approaches and datasets presents a challenge, but these networks offer significant potential for advances in real-time understanding and mitigation of risk for river users, managers, decision-makers, and regulators (Jankowski et al., 2021; O'Grady et al., 2021).

With the decreasing footprint and power consumption of contemporary environmental sensors, they can now typically be integrated into multiparameter sensor platforms (i.e. sondes) to provide new opportunities for quantifying an array of carbon cycle processes. Miniaturization and lower power requirements have also increased the portability of environmental sensors for deployment onboard mobile robots, leading to the emergence of commercially available, remotely operated

Table 1. Multiple parameters can be measured routinely with high-frequency sensors to advance understanding and management of river carbon cycling and emissions.

Sensor parameter	Relevance to freshwater C processes and emissions	Example studies (citations in [] are open-source sensor examples)
cDOM, tryptophan-like fluorescence, and absorption at 254 nm	Measures fractions of organic matter and correlates with DOC and Total organic Carbon (TOC) to understand carbon cycling	(Spencer et al., 2009)
Chlorophyll a and Phycocyanin	Represents processes of algae growth and primary production	(Chegoonian et al., 2022, Peipoch & Ensign, 2022)
CO_2/CH_4 gas flux chambers	Provides direct measurements of GHG emitted from water surfaces	(McClure et al., 2021, Y. Zheng et al., 2022) [Duc et al., 2013, Maher et al., 2019]
Dissolved CO_2 / CH_4	Dissolved GHG that can be potentially emitted	(Roberts et al., 2007, Crawford et al., 2017) [Butturini & Fonollosa, 2022]
Dissolved oxygen	Primary production produces oxygen whilst respiration consumes it.	(Aspray et al., 2017, Mejia et al., 2019, Jankowski et al., 2021) [Chan et al., 2021]
Electrical conductivity	Hydrological tracer within open system calculations of river metabolism	(Vieweg et al., 2016) [Méndez-Barroso et al., 2020]
Nitrate	Primary production can serve as an important nitrate sink	(Jarvie et al., 2018, Murray et al., 2020)
pH	Significant control on most biological processes, and influences bicarbonate buffering system, thus dissolved CO_2	(Hong et al., 2021, Klemme et al., 2022)
Turbidity	Elevated turbidity/sedimentation can block sunlight to primary producers, reducing primary production	(Honious et al., 2022) (Droujko et al., 2023)
Water level	Indicates changes in flow, which influence the rates of primary production and respiration	(Bernhardt et al., 2022) (Bresnahan et al., 2023)
Water temperature	Direct impact on gas solubility, metabolic rates (primary production, respiration, methanogenesis). Warming can shift the balance toward more CO_2 and CH_4 release rather than uptake into biomass	(Demars et al., 2011, Yvon-Durocher et al., 2011) (Hong et al., 2021)

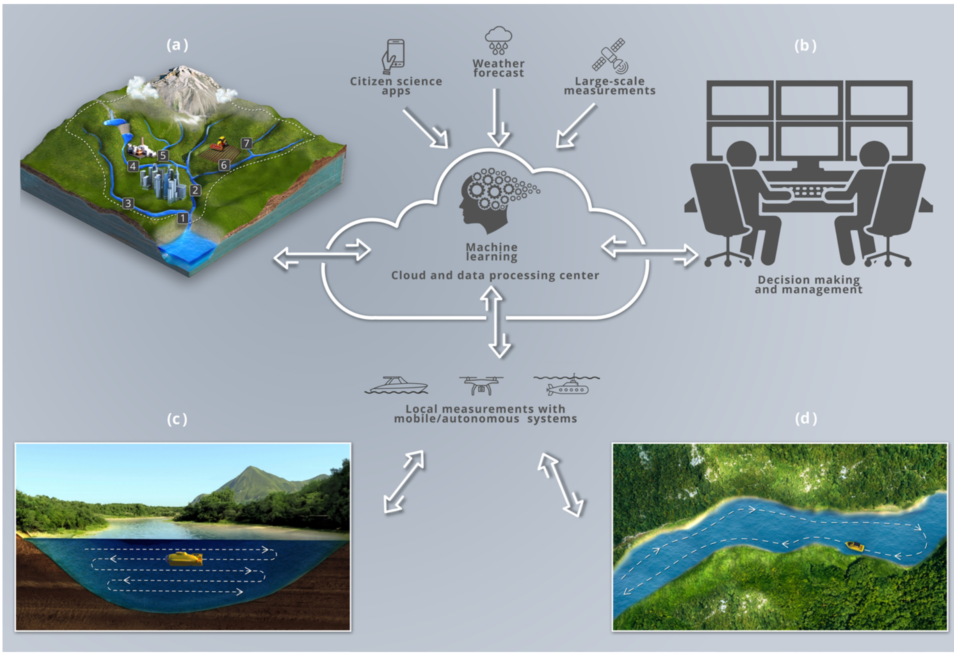


Figure 3. Data measurement and integration can be used to advance assessments and management of river carbon dynamics and water quality. (a) Sensor arrays along river catchments (locations 1-7) provide time-series of parameters (including carbon/ metabolism). Real-time data from the suite of sensors can then be exchanged and processed together with other automated information sources such as weather forecasts, satellite data, and local measurements (white arrows). ML tools can be incorporated to (b) alert humanized control centers for proposed actions, or take actions automatically using alarm rationalization (distinguishing between alarms and alerts). The dimensionality of data collection at these nodes (c, d) can be augmented by deploying mobile/autonomous systems to capture information from river cross sections, as well as from the reaches between fixed sensor nodes. Hierarchically nested structures of sensor arrays and other information sources can thus be used to advance optimization in water resource management.

and autonomous surface vehicles for environmental monitoring (HR Wallingford no date, YSI n.d.). However, probes for measuring dissolved CO_2, CH_4, and other proxies for organic matter (e.g. cDOM, UV254, etc.) remain poorly incorporated into carbon cycling estimations. Available sensors either require further development to improve their resolution and detection of multiple compounds such as emerging contaminants, or for dissolved gases atmospheric sensors must be deployed in bespoke water-tight, gas-permeable sleeves (Aho et al., 2021; Bernal et al., 2022) or direct measurements require combined chemical and optical measurements (Mendes et al., 2019). Despite the increasing number of sensors that measure parameters related to carbon, most river studies estimating whole-stream metabolism have used DO time series (Hoellein et al., 2013) but this method cannot resolve the change in respiration between day and night (Tromboni et al., 2022). Conversion of oxygen data to CO_2 production/uptake then relies on the use of respiratory quotients, with further work needed using concurrent O_2 and CO_2 measurements to understand sources of uncertainty, including organic matter composition and biological community influences (Bernal et al., 2022) as well as processes such as denitrification and sulfate oxidation that produce CO_2 without consuming DO.

Additional uncertainty must be minimized with appropriate corrections for reaeration of atmospheric-aquatic gas exchanges, using either tracer injections of inverse model fitting to sensor-derived dissolved gas time-series (Holtgrieve et al., 2016). Novel biosensors based on microbial-fuel cells (MFC), such as two-electrode bioelectrochemical systems that use microbial respiration to convert chemical energy to electricity (Cui et al., 2019), offer potential solutions to environmental sensors for aquatic respiration-related parameters. The MFC voltage or current response to aquatic respiration-related parameters (including DO, BOD, COD, and GHG) has been used as the basis for developing MFC-based biosensors (Wu et al., 2019), including commercial devices (e.g. HABS-2000 Online BOD Analyser). MFC-based sensors have additional benefits including low cost, environmental sustainability, the possibility of self-powered operation, portability, and reduced response times in the order of minutes.

Implementation of sensor networks requires significant investment and a long-term strategy that takes into account design, construction, maintenance, and data analysis approaches. Sensor networks are rarely constructed in a single effort due to their costs and are instead constructed incrementally. Actual costs are highly dependent on the variables being measured and on location (including remoteness). The major costs associated with sensor networks include the construction of the infrastructure for protecting the sensors, the sensor themselves, their maintenance and calibration (including personnel and sensor replacements), data transmission, and data analysis. Infrastructure and personnel cost are highly dependent on the location globally. Individual sensors have different costs; for example, while traditional sensors for measuring DO, EC, and pH cost approximately $1–5K USD each, a nitrate sensor can cost $30–50K USD. Including measurements that use autonomous systems (Figure 3) will further increase costs by at least several $1000 USD. Despite the high costs of sensor and analysis tools, there is a constant flow of new low-cost systems for measuring water quality in rivers that will likely reduce the cost of sensor networks by an order of magnitude in the near future.

Affordable sensor networks

In general, the high cost of using standard commercial sensors for water quality and carbon cycling presents a significant limiting factor for implementing high spatio-temporal sensor networks. "Affordable" sensors/devices have a price of at least one order of magnitude lower than an off-the-shelf commercial product though. Developments in affordable, open-source computing hardware, such as microprocessors (e.g. Arduino) and single-board microcomputers (e.g. Raspberry Pi) could bridge the gap between low-cost sensing and data logging, with wireless real-time data transmitting (Chan et al., 2021) enabling their use in a wide range of geographic locations and by a range of users, including citizen scientists, particularly as IoT networks become more widespread. While there are advantages to adopting citizen science methods for monitoring water quality, there are also several obstacles to overcome. Research indicates that the technology utilized for citizen science water quality monitoring should be cost-effective, easy to use, and capable of producing precise outcomes. In this context, IoT devices and sensors integrated with smartphones show potential as viable solutions (Amador-Castro et al., 2024).

Currently, there are low-cost reported solutions for aquatic measurements of multiple parameters directly relevant to understanding C cycling in aquatic systems (Table 1, column 3). In addition, there are multiple affordable datalogging devices and wireless communication options relying on local-, cellular-, and satellite- based solutions (Levintal et al., 2021). The field of affordable sensors is evolving rapidly, with new capabilities emerging constantly. For example, a compact multi-gas sensing platform for CO_2, CH_4, and N_2O sub-ppm measurements is under development (Wastine et al., 2022). Such a device can be used in automatic flux chambers to quantify, in real-time, emissions of all three main GHGs, from different locations within river catchments. Despite this, highly specialized equipment will increase monitoring costs, which is only feasible for limited initiatives. With respect to carbon cycling in rivers, there is still a need for affordable solutions for dissolved organic matter (DOM), nutrients, and dissolved gases other than O_2 and CO_2. The use of low-cost auto-samplers (Carvalho, 2020), portable spectrophotometers (Laganovska et al., 2020) or UV fluorescence spectroscopy (Yeshno et al., 2021) can potentially provide relevant solutions for autonomous water sampling and analysis, thus meeting the need for high-resolution monitoring without excessive costs. However, given the increasing availability of low-cost solutions for deployment by a range of users, these sensors must be developed, deployed, and maintained in line with robust protocols to ensure data accuracy.

In many cases, affordable sensors have not been designed for use in aquatic systems. Installing or developing a sensor station (node) may take longer to implement than standard commercial sensors and require different steps such as waterproofing, calibration, or processing of raw data, which can present barriers to non-technical users (Chan et al., 2021). There are also major challenges with incorporating multiple sensors, possibly with different outputs, into a single and stable working system such as a monitoring robot. Another barrier is psychological, as affordable sensors can sometimes be wrongly considered less appropriate for rigorous scientific research (Chan et al., 2021). Overcoming these challenges will lead to the development of low-cost sensor nodes, which will increase the affordability of deploying multiple nodes within an environment. This will also increase the spatial resolution, which is particularly valuable in areas where unpredictable extreme events are increasingly likely. Increasing accessibility of this technology, to economically developing nations will also be improved.

Mobile sensors and monitoring robots

Advances in mobile sensors and monitoring robots have increased the spatial resolution of environmental sensing, enabling capabilities such as detecting the location of pollution hotspots within water bodies (Powers et al., 2018). Mobile sensors can be divided into those designed to travel passively within water currents (Gardner et al., 2020; Marchant et al., 2015), and those that are actively mobilized using robotic technology. Unmanned robots are a promising solution for sensing in hard-to-reach locations, enabling spatially continuous data collection and monitoring over longer distances and time periods.

Unmanned/unpiloted aerial vehicles (UAV), also known as aerial drones, can monitor large areas and use multi-spectral imaging and on-board probes and samplers to measure parameters such as algal blooms, temperature, and light. Small-scale, uncrewed surface vehicles/vessels (USV), also known as unmanned surface vehicles/vessels can be equipped with on-board bathymetric, light detection and ranging (LiDAR), GPS, and flow sensors, which enable correlation of geographical and hydrological parameters with water quality sensing (e.g. temperature, suspended solids concentration and hydrocarbon concentration (Martinez Vargas et al., 2023)). In addition to carrying water quality sensors, samplers are also beginning to emerge with the ability to automate water sample collection and analysis on-board USVs, providing further potential for monitoring in-situ ecosystem processes (e.g. biochemical oxygen demand for respiration) (Fornai et al., 2012; Shabani et al., 2021). The availability of relatively low-cost autonomous underwater vehicles (AUV), USVs, and miniaturized environmental sensors has already led to the emergence of commercially available sensor-deployment robots (Lee et al., 2023). These devices are typically remotely controlled, with the state-of-the-art ability to navigate autonomously due to advances in AI and ML and particularly deep learning (DL) methods (Qiao et al., 2023).

Remaining obstacles to the use of robotic mobile sensing include legal constraints and physical limitations, such as achievable battery life. Autonomous navigation remains a challenge due to the high non-linearity and uncertainty of natural, and particularly, aquatic environments. As a result, AUVs are not yet well-developed for use in rivers, as the complex, low visibility environment is difficult to navigate, and many communication and localization technologies (e.g. GPS) cannot be used underwater. However, USVs and AUVs are well established in marine monitoring, showing the potential of this technology to be applied within rivers with further development.

Integrating high-resolution sensor data with analytics advances

Sensor networks present opportunities to develop a new understanding of fundamental environmental processes alongside applied management scenarios, by coupling high-resolution data sources with concurrent advances in statistical analysis and modeling. For example, 10 years ago, many studies using DO sensing technology typically comprised snapshots of metabolism on daily timescales at the river reach (~10–100 m) scale (Demars et al., 2015; Hoellein et al., 2013), sometimes with seasonal repeat sampling (VON Schiller et al., 2008). With the development of more reliable and robust sensors including automatic cleaning (e.g. wipers, pressurized air), DO time-series data have been collected to calculate continuous metabolism (i.e. GPP, ER, NEP) over periods of months-years (Pathak et al., 2022; Roberts et al., 2007) illustrating how some rivers can switch temporally from sources to sinks of GHGs. The spatial distribution of sampling networks has also seen a recent shift from reach-based assessments toward efforts to quantify metabolism for catchments and whole river systems (Rodríguez-Castillo et al., 2019; Segatto et al., 2021, 2023). Large volumes of high-resolution water quality data are becoming available from continental-global scale networks (Bernhardt et al., 2022). These ongoing increases in sensor data coverage offer the potential for significant improvements in pinpointing key drivers and constraints of aquatic ecosystem health, such as temperature, light, nutrients, and discharge (Bernhardt et al., 2018) enabling improved decision-making and more strategic intervention efforts.

Modeling approaches

River DOM processing is influenced by multiple dynamic drivers that often respond non-linearly to hydro-climatological events across catchments, such as floods, drought, and warming (Battin et al., 2023). In the past, modeling of DOM, carbon and nutrient reaction, and transport through river networks was hindered by the lack of high-resolution hydrology and hydrography data products at watershed, national, and global scales. As a result, models were typically limited to specific water body types (e.g. lakes only, river reaches/segments only) or grouped catchments where output could not be discretized in such a way to allow for spatiotemporal trends to be identified. The intersection of sensor technology, river models, and ML advances presents new opportunities for aquatic scientists and managers to develop digital representations of river systems (aka digital twins) to enhance aquatic science and management.

Increasing volumes of sensor data have enabled the expansion of metabolism estimation from the river reach scale to the network scale (Figure 4) using a range of model methods, including process-based (Segatto et al., 2020), empirical (Rodríguez-Castillo et al., 2019), ML (Segatto et al., 2021), or a combination (Maavara et al., 2023; Pathak et al., 2022). As sensor networks can gather data on the physical and chemical properties of rivers, such as temperature, light intensity, DO, and nutrient concentrations, these data are usually used as input in process-based metabolism models to estimate reach-scale processes (Appling et al., 2018; Demars et al., 2015).

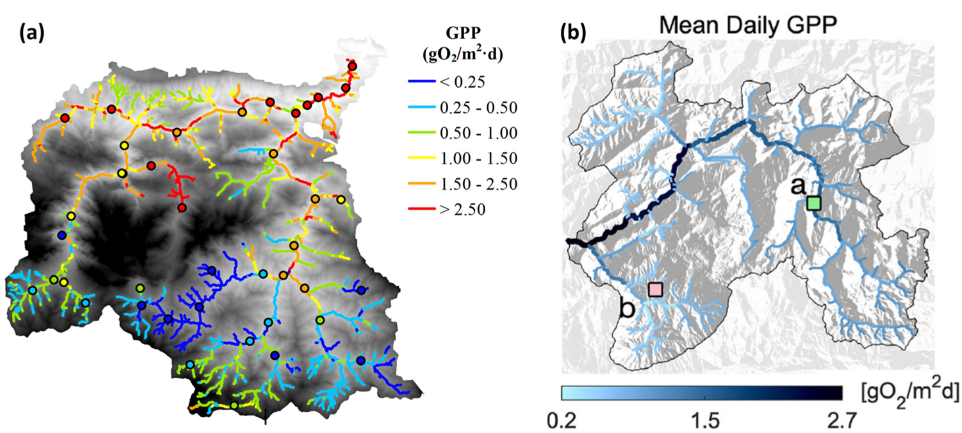


Figure 4. Catchment-scale river metabolism estimates can now be developed from distributed sensor networks: (a) GPP measured in the 1200 km^2 Deva-Cares catchment, northern Spain (Rodríguez-Castillo et al., 2019); (b) GPP measured in the 256 km^2 Ybbs river, Austria (Segatto et al., 2021). Network outputs such as these can be developed as visualization tools to aid catchment management decision-making, with dynamic updating in near-real time from linked sensors, telecommunication systems, and computational models.

Local metabolism rates can then be combined with information about the catchment environment to upscale to the river network scale, and as inputs for ML algorithms such as decision trees or neural networks.

In the past 5 years, high-resolution river network data products have become available at both national and global scales, including MERIT-Hydro and the associated GRADES dataset (Lin et al., 2019; Yamazaki et al., 2019) with 35 years of daily flow data from nearly 3 million river segments worldwide. These river network maps have enabled the further development of biogeochemical models that can be used alongside discrete location sensor data to quantify how nutrient and carbon sources, sinks, and transformations vary according to river size, flow, and season in large watershed networks. For example, Maavara et al. (2023) used the US National Hydrology Dataset (NHD Plus HR) product to develop a DOC model for the Connecticut River watershed, NE USA (Figure 1). This model was calibrated with GPP, terrestrial DOC loading, photo-mineralization, and respiration datasets, derived partly from a sensor network at 10 locations across 1st-8th order rivers. These continuous DO measurements at 15-min intervals were used to estimate GPP and ER using a Markov Chain Monte Carlo algorithm, which was then scaled to estimate GPP across the entire watershed during all flows and seasons, by calibrating a random forest ML model (Appling et al., 2018).

Efforts in improving process-based metabolism models have focused on expanding estimation to a more diverse set of river environments than previously possible, including estimation in river reaches with large discontinuities (e.g. flow and water quality regulation) or river reaches with significant transient storage (Pathak & Demars, 2023). Progress in this direction is valuable for reducing uncertainties in global estimates of freshwater carbon fluxes. Such process-based models could facilitate large-scale assessments of metabolism and its drivers across river environments, when combined with ML methods (Appling et al., 2018; Bernhardt et al., 2022). Several other physical properties currently overlooked in field studies may significantly impact metabolism and will need to be incorporated into future network models, for example sediment movement (Risse-Buhl et al., 2023; Schulz et al., 2023) and hyporheic/groundwater interactions (Galloway et al., 2019), which can have major impacts on ER.

Quantification of spatial and temporal dynamics of metabolism across river networks is important for estimating regional carbon emissions from rivers (Battin et al., 2023). However, only a few studies have focused on metabolism estimation at the river network scale (Figure 4). Rodríguez-Castillo et al. (2019) utilized the spatial stream network model to identify the factors that govern spatial variations in river metabolism within the Deva-Cares catchment in northern Spain, highlighting benthic biomass, river channel properties, and human activities as important controlling factors. Segatto et al. (2021, 2023) found that ER played a larger role in metabolic stability at the river network scale in the Ybbs River Austria, whereas GPP showed higher sensitivity to flow-induced disturbances and variations in light availability. Mejia et al. (2019) used the BAyesian Single-station Estimation (Grace et al., 2015) model to estimate metabolism over a year at ten sites across the Methow River network in Washington State, USA. Their findings indicated that metabolism timing may vary between sites within a river catchment due to the combined influence of local physicochemical conditions, despite having similar regional climates. Metabolism studies at the river network scale are admittedly data-intensive and these approaches need to be evaluated in river systems that are heavily polluted and where water quality often varies significantly over even short distances (Casillas-García et al., 2021). In these systems the implications may be that more dense networks of fixed and robot-mounted sensors are required, alongside additional predictor datasets such as point-source input locations and land use; however, such information is increasingly becoming available with advances in sensor technology, remote sensing products, and modeling techniques including ML. Mobile robots can be used to both increase the range and spatial resolution of the data on which models are trained and validate predictive models by increasing empirical field data collection.

Machine learning advances

AI is a field of study aimed at enabling machines to simulate human intelligence. Its origins can be traced back to the 1950s, when research primarily focused on automatic computers, self-improvement, and other related areas. ML is a subfield of AI, allowing machines to construct or improve computer programs automatically based on experience, rather than relying on explicit programming. Specifically, ML trains a regression or classification model through complex non-linear mapping with adjustable parameters, based on a training data set. Several recent river carbon cycle studies have used random forest (RF) ML algorithms; for example, Maavara et al. (2023) calibrated a RF that extrapolated GPP to almost 100,000 river reaches and lakes within the watershed using available predictor data such as flow, temperature, and canopy cover. Segatto et al. (2021, 2023) also improved metabolic upscaling by incorporating a temporal dimension into predictions of metabolic regimes by training RFs using long-term, sensor-based estimates of GPP and ER in the Ybbs River catchment in Austria, as well as catchment physical and climate properties. However, RFs typically require large datasets and their transferability to systems for which they have not been trained can be problematic. DL is an additional branch of ML, distinguished by multiple layers of neurons in neural network architecture, which provide a higher ability to represent complex functions than non-deep neural networks (J. Zhang et al. 2021). Owing to the availability of large datasets and advancements in computational power, DL garnered significant attention in the 2010s, particularly due to its breakthrough in image recognition and natural language processing. Additionally, DL has emerged as a promising tool for research in domains such as carbon cycling and hydrology.

Accurate quantification of carbon emissions from aquatic systems remains constrained by scientific uncertainties, high complexity of physical and chemical process linkages such as non-stationarity, dynamism, and non-linearity. As a result, prediction and forecasting with process-driven methods can be inaccurate; for rivers, water temperature, and discharge data currently provide the best opportunities for forecasting, whereas research on near-term biological/chemical predictions has advanced more quickly for lakes (McClure et al., 2021). DL has been suggested as a potential means to overcome uncertainty and nonlinearity in river sciences (Shen, 2018) and is now being applied in hydrologic predictions (water level, discharge (Xu et al., 2022)), regional rainfall-runoff linkages (J. Zhang et al. 2021) and water quality dynamics (H. Zheng et al., 2023). DL also has relevance in aquatic ecosystem prediction, including data mining and identifying outliers (Kim et al., 2022). With respect to water quality data, DL methods have been shown to offer potential to predict N and P concentrations from physical data that can be collected more easily with sensors (e.g. pH, turbidity, temperature, DO, conductivity) (Ba-Alawi et al., 2023). Moreover, DL can serve both as an auxiliary tool for process-driven methods, reducing computational loads in uncertainty analyses (Li et al., 2020) and as a component of process-driven models, describing a process difficult to characterize mathematically (Huang et al., 2022).

Physical models can now be embedded into DL models to improve performance and mitigate risks, by providing important supplementary information (Huang et al., 2022; Reichstein et al., 2019). Physics-informed neural network (PINN) models incorporate the residual of physics principles (e.g. governing equations) as a regulation in loss functions to enable learning by penalizing poor predictions (Tartakovsky et al., 2020). PINN is increasingly being applied in areas such as estimating water quantity and quality (Liang et al., 2019). Therefore, the development of physics-informed surrogate models that link DOM or DO concentrations and other water quality data with river flows could offer the potential for forecasting carbon emissions with greater accuracy and with improved consideration of uncertainty propagation.

Transfer learning (TL) developments offer additional potential for DL applications in water resource science and management. TL recognizes knowledge from a previous task and applies it to a new task (Pan & Yang, 2010). The previous task is usually an efficient ML model trained on large datasets, and then new tasks are related to the previous task but with smaller datasets.

TL methods in hydrology have focused mainly on data interpolation and prediction in areas where observed data are missing or unavailable. For example, Willard et al. (2021) showed how lake water temperature can be predicted in areas without monitoring, and Zhou (2020) developed real-time predictions of river water quality applied to situations where data were missing (e.g. broken sensors). Promising applications to river carbon cycle understanding and management could include learning between catchments that differ in data availability (e.g. Figure 1), enabling knowledge gained from the better-studied catchment(s) to advance understanding of the less-studied system(s).

Despite numerous successful DL applications in aquatic sciences, challenges and risks remain in applying these approaches for aquatic carbon management. Overarching issues for all ML applications include the potential for sensor and data processing security breaches (Richards et al., 2023) leading to risks for water security. A second issue concerns detection, as the accuracy of DL methods relies on the quantity of observational data. Insufficient data may prevent DL from achieving satisfactory precision (Cao et al., 2022); however, even in developed countries with well-established infrastructures, the cost of obtaining a substantial volume of high-precision environmental monitoring data such as that needed for river carbon cycle estimation could hinder the application of DL in some locations (Richards et al., 2023). Moreover, even water quality monitoring networks in developing countries are often limited by financial resources and technical capabilities and so must prioritize resource allocation. Third, DL methods work well only when training and test data are drawn from the same data feature space and distribution (Pan & Yang, 2010). This implies that DL methods must be specifically designed and tailored for context. Due to the influence of factors such as geometry and land cover, aquatic systems often differ between watersheds, meaning models from other study areas can lead to errors in prediction and risks for decision-making. However, by incorporating explicit mechanisms into the training process, DL models are beginning to emerge to overcome these issues, offering strong potential to advance further our understanding of river carbon cycling and emissions.

Conclusions

Global challenges associated with climate change adaptation and mitigation underpin the need for an accurate understanding of factors influencing carbon cycling at whole catchment scales. A primary barrier is the lack of water quality and emissions data, which are also key to improving water management by local-national business and regulatory organizations in river basins around the world, and to inform global policy developments by organizations such as the IPCC. The coincident need for data collection at higher resolution should be addressed by capitalizing on advances in distributed high-resolution sensor networks, combined with data analytic advances including ML methods. Combining high-resolution data and advanced analytics provide an opportunity to overcome challenges including resource limitation, access to remote areas, inconsistent monitoring practices, and/or data collection with insufficient spatial/temporal resolution. Benefits from expanding river carbon cycle process and emission understanding include closing knowledge gaps in international carbon accounting and facilitating more effective river catchment management.

While enhanced data collection and processing using in-situ sensors and advanced analytics can fill large gaps, logistical and financial constraints will still limit comprehensive sampling of complex spatial river networks at high-resolution. Therefore, advanced data analytics methods need to be developed concurrently to allow for scaling-up from point estimates in space to catchment scale, for filling in data gaps in time-series and for predicting water quality parameters for which robust and reliable sensors do not yet exist. DL methods have created significant opportunities and challenges in environmental research, although PINN and TL now provide a new basis to advance traditional DL methods. These methods are still in the research stage and significant investment will be needed to ensure confidence in water resource management applications. Nevertheless, rapid developments in data collection and analysis, with reducing costs, present unprecedented new potential for monitoring and improving the status of freshwater systems worldwide. Capitalizing on these technological advances quickly will be vital to address intersecting global crises in freshwater availability, water quality, biodiversity and climate change while maintaining the array of critical ecosystem services that freshwaters provide to humanity.

Author contributions

LEB, MK, EL, FOM, SS, EBZ, RG, LC, GG, FT, PZ and SA generated the idea and structure for the review at a workshop held in Israel in November 2022. LEB, TM, XC, JZ, MK, FOM, EL, DP, KK, SS, EBZ & SA developed the first detailed draft of the review. LEB, TM, JB and SA generated the figures and table. All authors reviewed and edited the manuscript text.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported primarily by funding from the Wohl Clean Growth Alliance and the British Council. Initial ideas were generated through work undertaken as part of the Euro-FLOW project by LEB, MJK, DP, PZ and JB, funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765553. TM is supported by a UK Natural Environment Research Council Independent Research Fellowship (NE/V014277/1). LC is supported by a UKRI Future Leaders Fellowship (MR/S032126/1). EBZ is supported by a grant from the Israeli Ministry of Science and Technology (#4755).

ORCID

Lee E. Brown http://orcid.org/0000-0002-2420-0088
Richard Grayson http://orcid.org/0000-0003-3637-3987
Gideon Gal http://orcid.org/0000-0002-6962-4001

Data availability

No data were generated in the production of this manuscript.

References

Acuña, V., Muñoz, I., Giorgi, A., Omella, M., Sabater, F., & Sabater, S. (2005). Drought and postdrought recovery cycles in an intermittent Mediterranean stream: Structural and functional aspects. *Journal of the North American Benthological Society*, *24*(4), 919–933. https://doi.org/10.1899/04-078.1

Aho, K. S., Hosen, J. D., Logozzo, L. A., McGillis, W. R., & Raymond, P. A. (2021). Highest rates of gross primary productivity maintained despite CO2 depletion in a temperate river network. *Limnology and Oceanography Letters*, *6*(4), 200–206. https://doi.org/10.1002/lol2.10195

Alberts, J. M., Beaulieu, J. J., & Buffam, I. (2017). Watershed land use and seasonal variation constrain the influence of riparian canopy cover on stream ecosystem metabolism. *Ecosystems (New York, N.Y.)*, *20*(3), 553–567. https://doi.org/10.1007/s10021-016-0040-9

Amador-Castro, F., González-López, M. E., Lopez-Gonzalez, G., Garcia-Gonzalez, A., Díaz-Torres, O., Carbajal-Espinosa, O., & Gradilla-Hernández, M. S. (2024). Internet of things and citizen science as alternative water quality monitoring approaches and the importance of effective water quality communication. *Journal of Environmental Management*, *352*, 119959. https://doi.org/10.1016/j.jenvman.2023.119959

Appling, A. P., Read, J. S., Winslow, L. A., Arroita, M., Bernhardt, E. S., Griffiths, N. A., Hall, R. O., Jr., Harvey, J. W., Heffernan, J. B., Stanley, E. H., Stets, E. G., & Yackulic, C. B. (2018). The metabolic regimes of 356 rivers in the United States. *Scientific Data*, *5*(1), 180292. https://doi.org/10.1038/sdata.2018.292

Arroita, M., Elosegi, A., & Hall, R. O.Jr, (2019). Twenty years of daily metabolism show riverine recovery following sewage abatement. *Limnology and Oceanography*, *64*(S1), S77–S92. https://doi.org/10.1002/lno.11053

Aspray, K. L., Holden, J., Ledger, M. E., Mainstone, C. P., & Brown, L. E. (2017). Organic sediment pulses impact rivers across multiple levels of ecological organization. *Ecohydrology*, *10*(6), e1855. https://doi.org/10.1002/eco.1855

Ba-Alawi, A. H., Heo, S., Aamer, H., Chang, R., Woo, T., Kim, M., & Yoo, C. (2023). Development of transparent high-frequency soft sensor of total nitrogen and total phosphorus concentrations in rivers using stacked convolutional auto-encoder and explainable AI. *Journal of Water Process Engineering*, *53*, 103661. https://doi.org/10.1016/j.jwpe.2023.103661

Battin, T. J., Kaplan, L. A., Findlay, S., Hopkinson, C. S., Marti, E., Packman, A. I., Newbold, J. D., & Sabater, F. (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nature Geoscience*, *1*(2), 95–100. https://doi.org/10.1038/ngeo101

Battin, T. J., Lauerwald, R., Bernhardt, E. S., Bertuzzo, E., Gener, L. G., Hall, R. O., Jr, Hotchkiss, E. R., Maavara, T., Pavelsky, T. M., Ran, L., Raymond, P., Rosentreter, J. A., & Regnier, P. (2023). River ecosystem metabolism and carbon biogeochemistry in a changing world. *Nature*, *613*(7944), 449–459. https://doi.org/10.1038/s41586-022-05500-8

Behrenfeld, M. J., & Falkowski, P. G. (1997). A consumer's guide to phytoplankton primary productivity models. *Limnology and Oceanography*, *42*(7), 1479–1491. https://doi.org/10.4319/lo.1997.42.7.1479

Bernal, S., Cohen, M. J., Ledesma, J. L. J., Kirk, L., Martí, E., & Lupon, A. (2022). Stream metabolism sources a large fraction of carbon dioxide to the atmosphere in two hydrologically contrasting headwater streams. *Limnology and Oceanography*, *67*(12), 2621–2634. https://doi.org/10.1002/lno.12226

Bernhardt, E. S., Heffernan, J. B., Grimm, N. B., Stanley, E. H., Harvey, J. W., Arroita, M., Appling, A. P., Cohen, M. J., McDowell, W. H., Hall, R. O., Jr., Read, J. S., Roberts, B. J., Stets, E. G., & Yackulic, C. B. (2018). The metabolic regimes of flowing waters. *Limnology and Oceanography*, *63*(S1), S99–S118. https://doi.org/10.1002/lno.10726

Bernhardt, E. S., Savoy, P., Vlah, M. J., Appling, A. P., Koenig, L. E., Hall, R. O., Jr., Arroita, M., Blaszczak, J. R., Carter, A. M., Cohen, M., Harvey, J. W., Heffernan, J. B., Helton, A. M., Hosen, J. D., Kirk, L., McDowell, W. H., Stanley, E. H., Yackulic, C. B., & Grimm, N. B. (2022). Light and flow regimes regulate the metabolism of rivers. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(8), e2121976119. https://doi.org/10.1073/pnas.2121976119

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

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. *The Science of the Total Environment*, *569–570*, 647–660. https://doi.org/10.1016/j.scitotenv.2016.06.116

Bresnahan, P., Briggs, E., Davis, B., Rodriguez, A. R., Edwards, L., Peach, C., Renner, N., Helling, H., & Merrifield, M. (2023). A low-cost, DIY ultrasonic water level sensor for education, citizen science, and research. *Oceanography*, *36*(1), 51–58. https://doi.org/10.5670/oceanog.2023.101

Butman, D., Stackpoole, S., Stets, E., McDonald, C. P., Clow, D. W., & Striegl, R. G. (2016). Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 58–63. https://doi.org/10.1073/pnas.1512651112

Butturini, A., & Fonollosa, J. (2022). Use of metal oxide semiconductor sensors to measure methane in aquatic ecosystems in the presence of cross-interfering compounds. *Limnology and Oceanography: Methods*, *20*(11), 710–720. https://doi.org/10.1002/lom3.10515

Cao, H., Xie, X., Shi, J., Jiang, G., & Wang, Y. (2022). Siamese network-based transfer learning model to predict geogenic contaminated groundwaters. *Environmental Science & Technology*, *56*(15), 11071–11079. https://doi.org/10.1021/acs.est.1c08682

Carvalho, M. C. (2020). Portable open-source autosampler for shallow waters. *HardwareX*, *8*, e00142. https://doi.org/10.1016/j.ohx.2020.e00142

Casillas-García, L. F., de Anda, J., Yebra-Montes, C., Shear, H., Díaz-Vázquez, D., & Gradilla-Hernández, M. S. (2021). Development of a specific water quality index for the protection of aquatic life of a highly polluted urban river. *Ecological Indicators*, *129*, 107899. https://doi.org/10.1016/j.ecolind.2021.107899

Chan, K., Schillereff, D. N., Baas, A. C. W., Chadwick, M. A., Main, B., Mulligan, M., O'Shea, F. T., Pearce, R., Smith, T. E. L., van Soesbergen, A., Tebbs, E., & Thompson, J. (2021). Low-cost electronic sensors for environmental research: Pitfalls and opportunities. *Progress in Physical Geography: Earth and Environment*, *45*(3), 305–338. https://doi.org/10.1177/0309133320956567

Charriau, A., Lissalde, S., Poulier, G., Mazzella, N., Buzier, R., & Guibaud, G. (2016). Overview of the Chemcatcher® for the passive sampling of various pollutants in aquatic environments Part A: Principles, calibration, preparation and analysis of the sampler. *Talanta*, *148*, 556–571. https://doi.org/10.1016/j.talanta.2015.06.064

Chegoonian, A. M., Zolfaghari, K., Leavitt, P. R., Baulch, H. M., & Duguay, C. R. (2022). Improvement of field fluorometry estimates of chlorophyll a concentration in a cyanobacteria-rich eutrophic lake. *Limnology and Oceanography: Methods*, *20*(4), 193–209. https://doi.org/10.1002/lom3.10480

Chen, L., Msigwa, G., Yang, M., Osman, A. I., Fawzy, S., Rooney, D. W., & Yap, P.-S. (2022). Strategies to achieve a carbon neutral society: A review. *Environmental Chemistry Letters*, *20*(4), 2277–2310. https://doi.org/10.1007/s10311-022-01435-8

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., Duarte, C. M., Kortelainen, P., Downing, J. A., Middelburg, J. J., & Melack, J. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems*, *10*(1), 172–185. https://doi.org/10.1007/s10021-006-9013-8

Crawford, J. T., Stanley, E. H., Dornblaser, M. M., & Striegl, R. G. (2017). CO2 time series patterns in contrasting headwater streams of North America. *Aquatic Sciences*, *79*(3), 473–486. https://doi.org/10.1007/s00027-016-0511-2

Cui, Y., Lai, B., & Tang, X. (2019). Microbial fuel cell-based biosensors. *Biosensors*, *9*(3), 92. https://doi.org/10.3390/bios9030092

Dean, J. F., & Battin, T. J. (2024). Future directions for river carbon biogeochemistry observations. *Nature Water*, *2*(3), 219–222. https://doi.org/10.1038/s44221-024-00207-8

Demars, B. O. L., Russell Manson, J., Ólafsson, J. S., Gíslason, G. M., Gudmundsdóttir, R., Woodward, G. U. Y., Reiss, J., Pichler, D. E., Rasmussen, J. J., & Friberg, N. (2011). Temperature and the metabolic balance of streams. *Freshwater Biology*, *56*(6), 1106–1121. https://doi.org/10.1111/j.1365-2427.2010.02554.x

Demars, B. O. L., Thompson, J., & Manson, J. R. (2015). Stream metabolism and the open diel oxygen method: Principles, practice, and perspectives. *Limnology and Oceanography: Methods*, *13*(7), 356–374. https://doi.org/10.1002/lom3.10030

Díaz-Torres, O., Lugo-Melchor, O. Y., de Anda, J., Orozco-Nunnelly, D. A., Gradilla-Hernández, M. S., & Senés-Guerrero, C. (2022). Characterizing a subtropical hypereutrophic lake: From physicochemical variables to shotgun metagenomic data. *Frontiers in Microbiology*, *13*, 1037626. https://doi.org/10.3389/fmicb.2022.1037626

Dodds, W. K., & Cole, J. J. (2007). Expanding the concept of trophic state in aquatic ecosystems: It's not just the autotrophs. *Aquatic Sciences*, *69*(4), 427–439. https://doi.org/10.1007/s00027-007-0922-1

Dodds, W. K., Veach, A. M., Ruffing, C. M., Larson, D. M., Fischer, J. L., & Costigan, K. H. (2013). Abiotic controls and temporal variability of river metabolism: Multiyear analyses of Mississippi and Chattahoochee River data. *Freshwater Science*, *32*(4), 1073–1087. https://doi.org/10.1899/13-018.1

Droujko, J., Kunz, F., Jr,., & Molnar, P. (2023). otz-T: 3D-printed open-source turbidity sensor with Arduino shield for suspended sediment monitoring. *HardwareX*, *13*, e00395. https://doi.org/10.1016/j.ohx.2023.e00395

Duc, N. T., Silverstein, S., Lundmark, L., Reyier, H., Crill, P., & Bastviken, D. (2013). Automated flux chamber for investigating gas flux at water–air interfaces. *Environmental Science & Technology*, *47*(2), 968–975. https://doi.org/10.1021/es303848x

Fell, S. C., Carrivick, J. L., Cauvy-Fraunié, S., Crespo-Pérez, V., Hood, E., Randall, K. C., Nicholass, K. J. M., Tiegs, S. D., Dumbrell, A. J., & Brown, L. E. (2021). Fungal decomposition of river organic matter accelerated by decreasing glacier cover. *Nature Climate Change*, *11*(4), 349–353. https://doi.org/10.1038/s41558-021-01004-x

Feng, H. (2005). The dynamics of carbon sequestration and alternative carbon accounting, with an application to the upper Mississippi River Basin. *Ecological Economics*, *54*(1), 23–35. https://doi.org/10.1016/j.ecolecon.2005.02.003

Ferreira, V., Elosegi, A., D. Tiegs, S., von Schiller, D., & Young, R. (2020). Organic matter decomposition and ecosystem metabolism as tools to assess the functional integrity of streams and rivers-A systematic review. *Water*, *12*(12), 3523. https://doi.org/10.3390/w12123523

Fornai, F., Bartaloni, F., Ferri, G., Manzi, A., Ciuchi, F., & Laschi, C. (2012). An autonomous water monitoring and sampling system for small-sized ASV operations. In 2012 Oceans (pp. 1–6). IEEE.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., … Zheng, B. (2022). Global carbon budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. https://doi.org/10.5194/essd-14-4811-2022

Galloway, J., Fox, A., Lewandowski, J., & Arnon, S. (2019). The effect of unsteady streamflow and stream-groundwater interactions on oxygen consumption in a sandy streambed. *Scientific Reports*, *9*(1), 19735. https://doi.org/10.1038/s41598-019-56289-y

Gardner, J. R., Ensign, S. H., Houser, J. N., & Doyle, M. W. (2020). Light exposure along particle flowpaths in large rivers. *Limnology and Oceanography*, *65*(1), 128–142. https://doi.org/10.1002/lno.11256

Gillespie, B. R., Desmet, S., Kay, P., Tillotson, M. R., & Brown, L. E. (2015). A critical analysis of regulated river ecosystem responses to managed environmental flows from reservoirs. *Freshwater Biology*, *60*(2), 410–425. https://doi.org/10.1111/fwb.12506

Grace, M. R., Giling, D. P., Hladyz, S., Caron, V., Thompson, R. M., & Mac Nally, R. (2015). Fast processing of diel oxygen curves: Estimating stream metabolism with BASE (BAyesian Single-station Estimation). *Limnology and Oceanography: Methods*, *13*, e10011.

Guérin, F., Abril, G., Richard, S., Burban, B., Reynouard, C., Seyler, P., & Delmas, R. (2006). Methane and carbon dioxide emissions from tropical reservoirs: Significance of downstream rivers. *Geophysical Research Letters*, *33*(21). https://doi.org/10.1029/2006GL027929

Hertwich, E. G. (2013). Addressing biogenic greenhouse gas emissions from hydropower in LCA. *Environmental Science & Technology*, *47*(17), 9604–9611. https://doi.org/10.1021/es401820p

Hoellein, T. J., Bruesewitz, D. A., & Richardson, D. C. (2013). Revisiting Odum (1956): A synthesis of aquatic ecosystem metabolism. *Limnology and Oceanography*, *58*(6), 2089–2100. https://doi.org/10.4319/lo.2013.58.6.2089

Holtgrieve, G. W., Schindler, D. E., & Jankowski, K. (2016). Comment on Demars et al. 2015, "Stream metabolism and the open diel oxygen method: Principles, practice, and perspectives. *Limnology and Oceanography: Methods*, *14*(2), 110–113. https://doi.org/10.1002/lom3.10075

Hong, W., Shamsuddin, N., Abas, E., Apong, R., Masri, Z., Suhaimi, H., Gödeke, S., & Noh, M. (2021). Water quality monitoring with Arduino based sensors. *Environments*, *8*(1), 6. https://doi.org/10.3390/environments8010006

Honious, S. A. S., Hale, R. L., Guilinger, J. J., Crosby, B. T., & Baxter, C. V. (2022). Turbidity structures the controls of ecosystem metabolism and associated metabolic process domains along a 75-km segment of a semiarid stream. *Ecosystems*, *25*(2), 422–440. https://doi.org/10.1007/s10021-021-00661-5

Hosen, J. D., Aho, K. S., Appling, A. P., Creech, E. C., Fair, J. H., Hall, R. O., Jr., Kyzivat, E. D., Lowenthal, R. S., Matt, S., Morrison, J., Saiers, J. E., Shanley, J. B., Weber, L. C., Yoon, B., & Raymond, P. A. (2019). Enhancement of primary production during drought in a temperate watershed is greater in larger rivers than headwater streams. *Limnology and Oceanography*, *64*(4), 1458–1472. https://doi.org/10.1002/lno.11127

Hosen, J. D., Aho, K. S., Fair, J. H., Kyzivat, E. D., Matt, S., Morrison, J., Stubbins, A., Weber, L. C., Yoon, B., & Raymond, P. A. (2021). Source switching maintains dissolved organic matter chemostasis across discharge levels in a large temperate river network. *Ecosystems*, *24*(2), 227–247. https://doi.org/10.1007/s10021-020-00514-7

HR Wallingford. (no date). The ARCboat range. https://equipit.hrwallingford.com/sites/default/files/2021-08/ARCboats-brochure.pdf. Retrieved August 23, 2024, from https://equipit.hrwallingford.com/arcboat/arcboat

Hu, B., Wang, D., Zhou, J., Meng, W., Li, C., Sun, Z., Guo, X., & Wang, Z. (2018). Greenhouse gases emission from the sewage draining rivers. *The Science of the Total Environment*, *612*, 1454–1462. https://doi.org/10.1016/j.scitotenv.2017.08.055

Huang, S., Xia, J., Wang, Y., Wang, W., Zeng, S., She, D., & Wang, G. (2022). Coupling machine learning into hydrodynamic models to improve river modeling with complex boundary conditions. *Water Resources Research*, *58*(10), e2022WR032183. https://doi.org/10.1029/2022WR032183

IPCC. (2019). Climate Change and Land: An IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. IPCC.

Jankowski, K. J., Mejia, F. H., Blaszczak, J. R., & Holtgrieve, G. W. (2021). Aquatic ecosystem metabolism as a tool in environmental management. *Wiley Interdisciplinary Reviews-Water*, *8*, e1521.

Jarvie, H. P., Sharpley, A. N., Kresse, T., Hays, P. D., Williams, R. J., King, S. M., & Berry, L. G. (2018). Coupling high-frequency stream metabolism and nutrient monitoring to explore biogeochemical controls on downstream nitrate delivery. *Environmental Science & Technology*, *52*(23), 13708–13717. https://doi.org/10.1021/acs.est.8b03074

Khamis, K., Bradley, C., & Hannah, D. M. (2018). Understanding dissolved organic matter dynamics in urban catchments: Insights from in situ fluorescence sensor technology. *Wiley Interdisciplinary Reviews-Water*, *5*, e1259.

Khandelwal, A., Castillo, T., & González-Pinzón, R. (2023). Development of The Navigator: A Lagrangian sensing system to characterize surface freshwater ecosystems. *Water Research*, *245*, 120577. https://doi.org/10.1016/j.watres.2023.120577

Kim, Y., Jeong, J., Park, H., Kwon, M., Cho, C., & Jeong, J. (2022). Development of a data-driven ensemble regressor and its applicability for identifying contextual and collective outliers in groundwater level time-series data. *Journal of Hydrology*, *612*, 128127. https://doi.org/10.1016/j.jhydrol.2022.128127

Klaar, M. J., Shelley, F. S., Hannah, D. M., & Krause, S. (2020). Instream wood increases riverbed temperature variability in a lowland sandy stream. *River Research and Applications*, *36*(8), 1529–1542. https://doi.org/10.1002/rra.3698

Klemme, A., Rixen, T., Müller-Dum, D., Müller, M., Notholt, J., & Warneke, T. (2022). CO2 emissions from peat-draining rivers regulated by water pH. *Biogeosciences*, *19*(11), 2855–2880. https://doi.org/10.5194/bg-19-2855-2022

Kominoski, J. S., Rosemond, A. D., Benstead, J. P., Gulis, V., & Manning, D. W. P. (2018). Experimental nitrogen and phosphorus additions increase rates of stream ecosystem respiration and carbon loss. *Limnology and Oceanography*, *63*(1), 22–36. https://doi.org/10.1002/lno.10610

Laganovska, K., Zolotarjovs, A., Vázquez, M., Mc Donnell, K., Liepins, J., Ben-Yoav, H., Karitans, V., & Smits, K. (2020). Portable low-cost open-source wireless spectrophotometer for fast and reliable measurements. *HardwareX*, *7*, e00108. https://doi.org/10.1016/j.ohx.2020.e00108

Lee, K.-H., Ali, S., Kim, Y., Lee, K., Kwon, S. Y., & Kam, J. (2023). High resolution mapping of nitrate loads of a reservoir using an uncrewed surface vehicle: Potential opportunities and challenges. *Water Resources Research*, *59*(11), e2023WR034665. https://doi.org/10.1029/2023WR034665

Levintal, E., Kang, K. L., Larson, L., Winkelman, E., Nackley, L., Weisbrod, N., Selker, J. S., & Udell, C. J. (2021). eGreenhouse: Robotically positioned, low-cost, open-source CO2 analyzer and sensor device for greenhouse applications. *HardwareX*, *9*, e00193. https://doi.org/10.1016/j.ohx.2021.e00193

Lewandowski, J., Arnon, S., Banks, E., Batelaan, O., Betterle, A., Broecker, T., Coll, C., Drummond, J. D., Gaona Garcia, J., Galloway, J., Gomez-Velez, J., Grabowski, R. C., Herzog, S. P., Hinkelmann, R., Höhne, A., Hollender, J., Horn, M. A., Jaeger, A., Krause, S., … Wu, L. (2019). Is the hyporheic zone relevant beyond the scientific community? *Water*, *11*(11), 2230. https://doi.org/10.3390/w11112230

Li, J., Lu, W., Wang, H., Fan, Y., & Chang, Z. (2020). Groundwater contamination source identification based on a hybrid particle swarm optimization-extreme learning machine. *Journal of Hydrology*, *584*, 124657. https://doi.org/10.1016/j.jhydrol.2020.124657

Liang, J., Li, W., Bradford, S. A., & Šimůnek, J. (2019). Physics-informed data-driven models to predict surface runoff water quantity and quality in agricultural fields. *Water*, *11*(2), 200. https://doi.org/10.3390/w11020200

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Liu, S., Kuhn, C., Amatulli, G., Aho, K., Butman, D. E., Allen, G. H., Lin, P., Pan, M., Yamazaki, D., Brinkerhoff, C., Gleason, C., Xia, X., & Raymond, P. A. (2022). The importance of hydrology in routing terrestrial carbon to the atmosphere via global streams and rivers. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(11), e2106322119. https://doi.org/10.1073/pnas.2106322119

Lofton, M. E., Howard, D. W., Thomas, R. Q., & Carey, C. C. (2023). Progress and opportunities in advancing near-term forecasting of freshwater quality. *Global Change Biology*, *29*(7), 1691–1714. https://doi.org/10.1111/gcb.16590

Maavara, T., Brinkerhoff, C., Hosen, J., Aho, K., Logozzo, L., Saiers, J., Stubbins, A., & Raymond, P. (2023). Watershed DOC uptake occurs mostly in lakes in the summer and in rivers in the winter. *Limnology and Oceanography*, *68*(3), 735–751. https://doi.org/10.1002/lno.12306

Maavara, T., Logozzo, L., Stubbins, A., Aho, K., Brinkerhoff, C., Hosen, J., & Raymond, P. (2021). Does photomineralization of dissolved organics matter in temperate rivers? *Journal of Geophysical Research: Biogeosciences*, *126*(7), e2021JG006402. https://doi.org/10.1029/2021JG006402

Maher, D. T., Drexl, M., Tait, D. R., Johnston, S. G., & Jeffrey, L. C. (2019). iAMES: An inexpensive, automated methane ebullition sensor. *Environmental Science & Technology*, *53*(11), 6420–6426. https://doi.org/10.1021/acs.est.9b01881

Marchant, R., Reading, D., Ridd, J., Campbell, S., & Ridd, P. (2015). A drifter for measuring water turbidity in rivers and coastal oceans. *Marine Pollution Bulletin*, *91*(1), 102–106. https://doi.org/10.1016/j.marpolbul.2014.12.021

Martinez Vargas, S., Vitale, A. J., Genchi, S. A., Nogueira, S. F., Arias, A. H., Perillo, G. M. E., Siben, A., & Delrieux, C. A. (2023). Monitoring multiple parameters in complex water scenarios using a low-cost open-source data acquisition platform. *HardwareX*, *16*, e00492.

McClure, R. P., Thomas, R. Q., Lofton, M. E., Woelmer, W. M., & Carey, C. C. (2021). Iterative forecasting improves near-term predictions of methane ebullition rates. *Frontiers in Environmental Science*, *9*, 756603. https://doi.org/10.3389/fenvs.2021.756603

Mejia, F. H., Fremier, A. K., Benjamin, J. R., Bellmore, J. R., Grimm, A. Z., Watson, G. A., & Newsom, M. (2019). Stream metabolism increases with drainage area and peaks asynchronously across a stream network. *Aquatic Sciences*, *81*(1), 9. https://doi.org/10.1007/s00027-018-0606-z

Mendes, J. P., Coelho, L., Kovacs, B., de Almeida, J. M. M. M., Pereira, C. M., Jorge, P. A. S., & Borges, M. T. (2019). Dissolved carbon dioxide sensing platform for freshwater and saline water applications: Characterization and validation in aquaculture environments. *Sensors*, *19*(24), 5513. https://doi.org/10.3390/s19245513

Méndez-Barroso, L. A., Rivas-Márquez, J. A., Sosa-Tinoco, I., & Robles-Morúa, A. (2020). Design and implementation of a low-cost multiparameter probe to evaluate the temporal variations of water quality conditions on an estuarine lagoon system. *Environmental Monitoring and Assessment*, *192*(11), 710. https://doi.org/10.1007/s10661-020-08677-5

Meyer, A. M., Klein, C., Fünfrocken, E., Kautenburger, R., & Beck, H. P. (2019). Real-time monitoring of water quality to identify pollution pathways in small and middle scale rivers. *The Science of the Total Environment*, *651*(Pt 2), 2323–2333.

Murray, E., Roche, P., Briet, M., Moore, B., Morrin, A., Diamond, D., & Paull, B. (2020). Fully automated, low-cost ion chromatography system for in-situ analysis of nitrite and nitrate in natural waters. *Talanta*, *216*, 120955. https://doi.org/10.1016/j.talanta.2020.120955

Nebgen, E. L., & Herrman, K. S. (2019). Effects of shading on stream ecosystem metabolism and water temperature in an agriculturally influenced stream in central Wisconsin, USA. *Ecological Engineering*, *126*, 16–24. https://doi.org/10.1016/j.ecoleng.2018.10.023

O'Grady, J., Zhang, D., O'Connor, N., & Regan, F. (2021). A comprehensive review of catchment water quality monitoring using a tiered framework of integrated sensing technologies. *The Science of the Total Environment*, *765*, 142766.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pathak, D., & Demars, B. O. L. (2023). Metabolism modeling in rivers with unsteady flow conditions and transient storage zones. *Journal of Geophysical Research: Biogeosciences*, *128*(3), e2022JG007245. https://doi.org/10.1029/2022JG007245

Pathak, D., Hutchins, M., Brown, L. E., Loewenthal, M., Scarlett, P., Armstrong, L., Nicholls, D., Bowes, M., Edwards, F., & Old, G. (2022). High-resolution water-quality and ecosystem-metabolism modeling in lowland rivers. *Limnology and Oceanography*, *67*(6), 1313–1327. https://doi.org/10.1002/lno.12079

Peipoch, M., & Ensign, S. H. (2022). Deciphering the origin of riverine phytoplankton using in situ chlorophyll sensors. *Limnology and Oceanography Letters*, *7*(2), 159–166. https://doi.org/10.1002/lol2.10240

Pérez-Calpe, A. V., de Guzman, I., Larrañaga, A., von Schiller, D., & Elosegi, A. (2022). Organic matter processing on dry riverbeds is more reactive to water diversion and pollution than on wet channels. *Frontiers in Environmental Science*, *9*, 817665. https://doi.org/10.3389/fenvs.2021.817665

Powers, C., Hanlon, R., & Schmale, D. G. (2018). Tracking of a fluorescent dye in a freshwater lake with an unmanned surface vehicle and an unmanned aircraft system. *Remote Sensing*, *10*(1), 81. https://doi.org/10.3390/rs10010081

Qiao, Y., Yin, J., Wang, W., Duarte, F., Yang, J., & Ratti, C. (2023). Survey of deep learning for autonomous surface vehicles in marine environments. *IEEE Transactions on Intelligent Transportation Systems*, *24*(4), 3678–3701. https://doi.org/10.1109/TITS.2023.3235911

Raymond, P. A., Saiers, J. E., & Sobczak, W. V. (2016). Hydrological and biogeochemical controls on watershed dissolved organic matter transport: Pulse-shunt concept. *Ecology*, *97*(1), 5–16. https://doi.org/10.1890/14-1684.1

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Richards, C. E., Tzachor, A., Avin, S., & Fenner, R. (2023). Rewards, risks and responsible deployment of artificial intelligence in water systems. *Nature Water*, *1*(5), 422–432. https://doi.org/10.1038/s44221-023-00069-6

Risse-Buhl, U., Arnon, S., Bar-Zeev, E., Oprei, A., Packman, A. I., Peralta-Maraver, I., Robertson, A., Teitelbaum, Y., & Mutz, M. (2023). Streambed migration frequency drives ecology and biogeochemistry across spatial scales. *WIREs Water*, *10*(3), e1632. https://doi.org/10.1002/wat2.1632

Roberts, B. J., Mulholland, P. J., & Hill, W. R. (2007). Multiple scales of temporal variability in ecosystem metabolism rates: Results from 2 years of continuous monitoring in a forested headwater stream. *Ecosystems*, *10*(4), 588–606. https://doi.org/10.1007/s10021-007-9059-2

Rodríguez-Castillo, T., Estévez, E., González-Ferreras, A. M., & Barquín, J. (2019). Estimating ecosystem metabolism to entire river networks. *Ecosystems*, *22*(4), 892–911. https://doi.org/10.1007/s10021-018-0311-8

Rosentreter, J. A., Borges, A. V., Deemer, B. R., Holgerson, M. A., Liu, S., Song, C., Melack, J., Raymond, P. A., Duarte, C. M., Allen, G. H., Olefeldt, D., Poulter, B., Battin, T. I., & Eyre, B. D. (2021). Half of global methane emissions come from highly variable aquatic ecosystem sources. *Nature Geoscience*, *14*(4), 225–230. +. https://doi.org/10.1038/s41561-021-00715-2

Rudee, A., & Phillips, J. (2021). *Why greenhouse gas inventories are important for natural and working lands—And how to fix them*. https://www.wri.org/technical-perspectives/greenhouse-gas-emissions-natural-working-lands

Savoy, P., & Harvey, J. W. (2021). Predicting light regime controls on primary productivity across CONUS river networks. *Geophysical Research Letters*, *48*(10), e2020GL092149. https://doi.org/10.1029/2020GL092149

Schulz, H., Teitelbaum, Y., Lewandowski, J., Singer, G. A., & Arnon, S. (2023). Moving bedforms control CO2 production and distribution in sandy river sediments. *Journal of Geophysical Research-Biogeosciences*, *128*, e2022JG007156.

Segatto, P. L., Battin, T. J., & Bertuzzo, E. (2020). Modeling the coupled dynamics of stream metabolism and microbial biomass. *Limnology and Oceanography*, *65*(7), 1573–1593. https://doi.org/10.1002/lno.11407

Segatto, P. L., Battin, T. J., & Bertuzzo, E. (2021). The metabolic regimes at the scale of an entire stream network unveiled through sensor data and machine learning. *Ecosystems (New York, N.Y.)*, *24*(7), 1792–1809. https://doi.org/10.1007/s10021-021-00618-8

Segatto, P. L., Battin, T. J., & Bertuzzo, E. (2023). A network-scale modeling framework for stream metabolism, ecosystem efficiency, and their response to climate change. *Water Resources Research*, *59*(3), e2022WR034062. https://doi.org/10.1029/2022WR034062

Shabani, F., Philamore, H., & Matsuno, F. (2021). An energy-autonomous chemical oxygen demand sensor using a microbial fuel cell and embedded machine learning. *IEEE Access*, *9*, 108689–108701. https://doi.org/10.1109/ACCESS.2021.3101496

Shelley, F., Klaar, M., Krause, S., & Trimmer, M. (2017). Enhanced hyporheic exchange flow around woody debris does not increase nitrate reduction in a sandy streambed. *Biogeochemistry*, *136*(3), 353–372. https://doi.org/10.1007/s10533-017-0401-2

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Shi, W., Maavara, T., Chen, Q., Zhang, J., Ni, J., & Tonina, D. (2023). Spatial patterns of diffusive greenhouse gas emissions from cascade hydropower reservoirs. *Journal of Hydrology*, *619*, 129343. https://doi.org/10.1016/j.jhydrol.2023.129343

Singh, S., Rai, S., Singh, P., & Mishra, V. K. (2022). Real-time water quality monitoring of River Ganga (India) using internet of things. *Ecological Informatics*, *71*, 101770. https://doi.org/10.1016/j.ecoinf.2022.101770

Spencer, R. G. M., Aiken, G. R., Butler, K. D., Dornblaser, M. M., Striegl, R. G., & Hernes, P. J. (2009). Utilizing chromophoric dissolved organic matter measurements to derive export and reactivity of dissolved organic carbon exported to the Arctic Ocean: A case study of the Yukon River, Alaska. *Geophysical Research Letters*, *36*(6). https://doi.org/10.1029/2008GL036831

Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., & Barajas-Solano, D. (2020). Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research*, *56*(5), e2019WR026731. https://doi.org/10.1029/2019WR026731

Tromboni, F., Hotchkiss, E. R., Schechner, A. E., Dodds, W. K., Poulson, S. R., & Chandra, S. (2022). High rates of daytime river metabolism are an underestimated component of carbon cycling. *Communications Earth & Environment*, *3*(1), 270. https://doi.org/10.1038/s43247-022-00607-2

UNFCCC. (2018). The Paris agreement.

UNISENSE. (2024). Nitrous oxide sensor user manual. https://unisense.com/wp-content/uploads/2024/01/2024.01-N2O-manual.pdf (accessed 15/11/2024)

Valdivia-Garcia, M., Weir, P., Graham, D. W., & Werner, D. (2019). Predicted impact of climate change on trihalomethanes formation in drinking water treatment. *Scientific Reports*, *9*(1), 9967. https://doi.org/10.1038/s41598-019-46238-0

Vieweg, M., Kurz, M. J., Trauth, N., Fleckenstein, J. H., Musolff, A., & Schmidt, C. (2016). Estimating time-variable aerobic respiration in the streambed by combining electrical conductivity and dissolved oxygen time series. *Journal of Geophysical Research: Biogeosciences*, *121*(8), 2199–2215. https://doi.org/10.1002/2016JG003345

VON Schiller, D., Martí, E., Riera, J. L., Ribot, M., Marks, J. C., & Sabater, F. (2008). Influence of land use on stream ecosystem function in a Mediterranean catchment. *Freshwater Biology*, *53*(12), 2600–2612. https://doi.org/10.1111/j.1365-2427.2008.02059.x

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555–561. https://doi.org/10.1038/nature09440

Wastine, B., Hummelgård, C., Bryzgalov, M., Rödjegård, H., Martin, H., & Schröder, S. (2022). Compact non-dispersive infrared multi-gas sensing platform for large scale deployment with sub-ppm resolution. *Atmosphere*, *13*(11), 1789. https://doi.org/10.3390/atmos13111789

Willard, J. D., Read, J. S., Appling, A. P., Oliver, S. K., Jia, X., & Kumar, V. (2021). Predicting water temperature dynamics of unmonitored lakes with meta-transfer learning. *Water Resources Research*, *57*(7), e2021WR029579. https://doi.org/10.1029/2021WR029579

Wollheim, W. M., Bernal, S., Burns, D. A., Czuba, J. A., Driscoll, C. T., Hansen, A. T., Hensley, R. T., Hosen, J. D., Inamdar, S., Kaushal, S. S., Koenig, L. E., Lu, Y. H., Marzadri, A., Raymond, P. A., Scott, D., Stewart, R. J., Vidon, P. G., & Wohl, E. (2018). River network saturation concept: Factors influencing the balance of biogeochemical supply and demand of river networks. *Biogeochemistry*, *141*(3), 503–521. https://doi.org/10.1007/s10533-018-0488-0

Wu, S.-S., Hernández, M., Deng, Y.-C., Han, C., Hong, X., Xu, J., Zhong, W.-H., & Deng, H. (2019). The voltage signals of microbial fuel cell-based sensors positively correlated with methane emission flux in paddy fields of China. *FEMS Microbiology Ecology*, *95*(3), fiz018. https://doi.org/10.1093/femsec/fiz018

Xiao, Q., Hu, Z., Hu, C., Islam, A. R. M. T., Bian, H., Chen, S., Liu, C., & Lee, X. (2021). A highly agricultural river network in Jurong Reservoir watershed as significant CO2 and CH4 sources. *The Science of the Total Environment*, *769*, 144558. https://doi.org/10.1016/j.scitotenv.2020.144558

Xing, Z., Chow, L., Rees, H., Meng, F., Li, S., Ernst, B., Benoy, G., Zha, T., & Hewitt, L. M. (2013). Influences of sampling methodologies on pesticide-residue detection in stream water. *Archives of Environmental Contamination and Toxicology*, *64*(2), 208–218. https://doi.org/10.1007/s00244-012-9833-9

Xu, W., Chen, J., & Zhang, X. J. (2022). Scale effects of the monthly streamflow prediction using a state-of-the-art deep learning model. *Water Resources Management*, *36*(10), 3609–3625. https://doi.org/10.1007/s11269-022-03216-y

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, *55*(6), 5053–5073. https://doi.org/10.1029/2019WR024873

Yeshno, E., Dahan, O., Bernstain, S., & Arnon, S. (2021). A novel analytical approach for the simultaneous measurement of nitrate and dissolved organic carbon in soil water. *Hydrology and Earth System Sciences*, *25*(4), 2159–2168. https://doi.org/10.5194/hess-25-2159-2021

Yorkshire Water. (2023). Yorkshire water invests in rural tech. Retrieved August 23, 2024, from https://www.yorkshirewater.com/news-media/news-articles/2023/yorkshire-water-invests-in-rural-tech/

YSI. (n.d.). rQPOD remote surface water vehicle. Retrieved August 23, 2024 from https://www.ysi.com/rqpod

Yvon-Durocher, G., Montoya, J. M., Woodward, G., Jones, J. I., & Trimmer, M. (2011). Warming increases the proportion of primary production emitted as methane from freshwater mesocosms. *Global Change Biology*, *17*(2), 1225–1234. https://doi.org/10.1111/j.1365-2486.2010.02289.x

Zhang, J., Chen, X., Khan, A., Zhang, Y-K., Kuang, X., Liang, X., Taccari, M. L., & Nuttall, J. (2021). Daily runoff forecasting by deep recursive neural network. *Journal of Hydrology*, *596*, 126067. (https://doi.org/10.1016/j.jhydrol.2021.126067

Zhang, W., Li, H., Xiao, Q., & Li, X. (2021). Urban rivers are hotspots of riverine greenhouse gas (N_2O, CH_4, CO_2) emissions in the mixed-landscape Chaohu lake basin. *Water Research*, *189*, 116624.

Zheng, H., Liu, Y., Wan, W., Zhao, J., & Xie, G. (2023). Large-scale prediction of stream water quality using an interpretable deep learning approach. *Journal of Environmental Management*, *331*, 117309. https://doi.org/10.1016/j.jenvman.2023.117309

Zheng, Y., Wu, S., Xiao, S., Yu, K., Fang, X., Xia, L., Wang, J., Liu, S., Freeman, C., & Zou, J. (2022). Global methane and nitrous oxide emissions from inland waters and estuaries. *Global Change Biology*, *28*(15), 4713–4725. https://doi.org/10.1111/gcb.16233

Zhou, Y. (2020). Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. *Journal of Hydrology*, *589*, 125164. https://doi.org/10.1016/j.jhydrol.2020.125164

Zia, H., Harris, N. R., Merrett, G. V., Rivers, M., & Coles, N. (2013). The impact of agricultural activities on water quality: A case for collaborative catchment-scale management using integrated wireless sensor networks. *Computers and Electronics in Agriculture*, *96*, 126–138. https://doi.org/10.1016/j.compag.2013.05.001