# Machine learning for long-term reconstruction and attribution of carbon and water-quality dynamics in inland waters

**PhD Thesis**

**Thesis supervisor:** Associate Prof. Dr. Zhongbing Chen

**Czech University of Life Sciences Prague**

**Faculty of Environmental Sciences**

**Department of Applied Ecology**

**Prague, 2026**

# Machine learning for long-term reconstruction and attribution of carbon and water-quality dynamics in inland waters

Feilong Shen

**Thesis**

This thesis is submitted in fulfillment of the requirements for the PhD degree at the Czech University of Life Sciences Prague, Faculty of Environmental Sciences.

Machine learning for long-term reconstruction and attribution of carbon and water-quality dynamics in inland waters.
Feilong Shen
Czech University of Life Sciences Prague (2026)

# Acknowledgement

I would like to begin by expressing my deepest gratitude to my supervisor, Associate Prof. Dr. Zhongbing Chen, of the Department of Applied Ecology, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, for his trust, patience, and unwavering support throughout my doctoral studies. His scientific insight, critical guidance, and encouragement shaped this dissertation and my development as an independent researcher.

I sincerely thank Prof. Ing. Jan Vymazal, CSc., for his support and contribution to the Department of Applied Ecology; Věra Nováková for her practical assistance with my research; and all departmental colleagues for their cooperation and help. I am especially grateful to Slavěna Zajícová and Jana Stýblová for their patient support with my doctoral courses, study-related administration, and other practical matters throughout my four years of doctoral study.

My sincere thanks go to Mengyang Xu of Xinyang Normal University, Di Tong of Zhejiang University, and Shasha Fan of Shanghai Normal University for their valuable help and encouragement in my research. Their support helped me overcome difficulties and sustain my enthusiasm for scientific work. I warmly thank Jifeng Li and Yingrun Chen at CZU, and Yu Zhu and Kai Zhu at Charles University, as well as the many friends who supported me during my four years in Prague. Their kindness and companionship made life away from home easier and filled this journey with lasting memories.

Above all, I thank my parents for decades of love, guidance, and encouragement. Their patience, sacrifices, and unwavering belief in me have supported every stage of my education. I am deeply grateful to my entire family, and I dedicate this dissertation to them.

The experimental work included in this PhD thesis was funded by the following research project:

IGA (Internal Grant Agency of the Faculty of Environmental Sciences, CZU. Project No. 2024B0005)

# Abstract

Inland waters play disproportionate roles in global carbon and nutrient cycles, yet their long-term dynamics remain poorly observed. Field networks provide direct measurements but are spatially sparse, temporally discontinuous, and particularly limited in small rivers that integrate catchment pressures. Satellite remote sensing and machine learning (ML) can extend observations beyond gauged sites, but feasibility depends on three conditions that are rarely evaluated together: whether the target is optically expressed, whether the water body is resolved by the sensor, and whether validation accounts for spatial dependence.

This dissertation aimed to reconstruct and attribute long-term changes in carbon and water quality using satellite observations and ML while defining the limits within which such inferences remain defensible. Harmonized surface reflectance from Landsat 5, 7, 8, and 9, Sentinel-2, Harmonized Landsat and Sentinel-2 (HLS), and the Aqua Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua) was processed in Google Earth Engine (GEE). Cloud screening, cross-sensor harmonization, object-level extraction, effective-pixel gating, and matching with extensive in situ records were implemented. Two contrasting systems were examined: eutrophic Lake Taihu, China, and 103 width-stratified agricultural river reaches in South Bohemia, Czechia.

Study 1 reconstructed dissolved organic carbon (DOC) concentration and storage in Lake Taihu from 1984 to 2023 using 200 Landsat–in situ matchups. Extreme gradient boosting (XGBoost) generalized best among four learners, reaching a validation coefficient of determination ($R^2$) of 0.73 on an independent dataset. Estimates were harmonized against MODIS-Aqua and combined with reconstructed water storage capacity. Lake-wide DOC concentration averaged approximately 4.72 mg $L^{-1}$ and followed a three-phase trajectory, while storage declined slightly but changed heterogeneously across the lake. Structural equation modelling identified meteorological forcing as the primary control, with human activity, bio-optical conditions, and vegetation acting as secondary or interacting influences.

Study 2 evaluated retrieval of ammonium nitrogen ($NH_4^+$-N), nitrate nitrogen ($NO_3^-$-N), and total phosphorus (TP) using 178,286 gated reach-object observations matched to 52,521 in situ measurements. Under leave-one-site-out spatial cross-validation (CV), $R^2$ values were −0.03, 0.05, and 0.11, respectively, whereas random CV inflated apparent skill by up to $\Delta R^2 = 0.57$. The principal constraint was geometric: the median narrow reach occupied only 16% of a 30 m pixel, leaving reflectance dominated by land and adjacency effects. Long-term records nevertheless showed declines of 80%, 48%, and 66%, followed by a plateau around 2007. This recovery was regional and width-uniform, and none of nine multi-scale riparian land-cover tests explained it under spatial validation.

Study 3 compared direct optical retrieval of chemical oxygen demand (COD) and five-day biochemical oxygen demand ($BOD_5$) with a random forest driven by satellite-

observable catchment attributes and point-source pressures. The modelled quantity was concentration rather than flux. Optical models failed at unvisited reaches, with Nash–Sutcliffe efficiencies of 0.02 for COD and 0.03 for $BOD_5$ and no benefit from native 10 m imagery. The catchment model achieved efficiencies of 0.29 and 0.40 and Kling–Gupta efficiencies of 0.33 and 0.47, respectively, on the $\log_{10}$ scale. $BOD_5$ prediction was strongest in the narrowest reaches. Reconstruction for 1985–2025 revealed an initial decline followed by a partial rebound, while the 2050 sensitivity analysis identified land-use management as a substantially stronger lever than warming within the tested scenarios.

Collectively, the three studies advance inland-water monitoring from algorithm-centred accuracy comparison towards a scale-aware framework of retrievability. They establish a double inseparability boundary: an optical boundary, where weak constituent signals cannot be separated from sub-pixel land contributions and adjacency contamination, and an attributional boundary, where spatial dependence and regional coherence forcing prevent basin-wide change from being decomposed into local causes. A near-zero spatially validated score is therefore not merely a poor modelling result but evidence that the requested information is unavailable at the chosen observational scale. Crucially, the dissertation also identifies a pathway beyond this limit: when direct observation of the water surface cannot support defensible retrieval, catchment context can recover part of the missing information. Its central contribution is therefore threefold: a long-term reconstruction where optical and spatial conditions permit, a quantitative diagnosis of where physical observability and statistical transferability fail, and a spatially validated catchment-based alternative for narrow rivers. This framework assigns satellite, catchment, and in situ observations to the scales at which each is informative, thereby strengthening the credibility of long-term reconstruction, attribution, and management of inland water change.

# Abstrakt

Vnitrozemské vody hrají nepoměrně významnou roli v globálních cyklech uhlíku a živin, jejich dlouhodobá dynamika však zůstává nedostatečně zachycena v prostoru i čase. Terénní monitorovací sítě poskytují přímá měření, jsou však prostorově řídké, časově nesouvislé a zvláště nedostatečné u malých vodních toků, které integrují tlaky působící v povodí. Družicový dálkový průzkum Země v kombinaci se strojovým učením (ML) může rozšířit pozorování za hranice měřených lokalit. Využitelnost tohoto přístupu však závisí na třech podmínkách, které jsou jen zřídka posuzovány současně: zda má cílová veličina opticky detekovatelný projev, zda je vodní útvar senzorem dostatečně prostorově rozlišen a zda validace zohledňuje prostorovou závislost dat.

Cílem této disertační práce bylo pomocí družicových pozorování a ML rekonstruovat dlouhodobou dynamiku uhlíku a ukazatelů jakosti vody a určit jejich řídicí faktory, přičemž byly zároveň vymezeny fyzikální a statistické meze, v nichž lze takové závěry považovat za vědecky obhajitelné. Harmonizovaná data povrchové odrazivosti z misí Landsat 5, 7, 8 a 9 a Sentinel-2, z produktů Harmonized Landsat and Sentinel-2 (HLS) a ze senzoru Moderate Resolution Imaging Spectroradiometer na družici Aqua (MODIS-Aqua) byla zpracována v prostředí Google Earth Engine (GEE). Zpracování zahrnovalo maskování oblačnosti, mezisenzorovou harmonizaci, extrakci objektů, filtraci podle počtu efektivních pixelů a párování s rozsáhlými záznamy in situ. Zkoumány byly dva kontrastní systémy: eutrofní jezero Taihu v Číně a 103 říčních úseků v zemědělsky využívaných povodích Jižních Čech, rozdělených podle šířky toku.

Studie 1 rekonstruovala koncentraci a zásobu rozpuštěného organického uhlíku (DOC) v jezeře Taihu v období 1984–2023 na základě 200 spárovaných pozorování Landsat–in situ. Ze čtyř hodnocených algoritmů vykázal model extreme gradient boosting (XGBoost) nejlepší schopnost generalizace a v nezávislém validačním souboru dosáhl koeficientu determinace $R^2 = 0{,}73$. Odhady z jednotlivých senzorů byly harmonizovány s MODIS-Aqua a spojeny s rekonstruovanou zásobou vody v jezeře. Průměrná koncentrace DOC v měřítku celého jezera činila přibližně 4,72 mg $L^{-1}$ a vykazovala třífázový časový průběh. Zásoba DOC během sledovaného období mírně poklesla, její změny však byly prostorově heterogenní. Modelování pomocí strukturních rovnic určilo meteorologické podmínky jako hlavní řídicí faktor, zatímco lidská činnost, biooptické podmínky a vegetace představovaly sekundární nebo vzájemně se ovlivňující faktory.

Studie 2 hodnotila možnost odvozovat koncentrace amonného dusíku ($NH_4^+$-N), dusičnanového dusíku ($NO_3^-$-N) a celkového fosforu (TP) na základě 178 286 objektových pozorování říčních úseků po aplikaci kvalitativních filtrů, spárovaných s 52 521 in situ měřeními. Při prostorové křížové validaci (CV) s vynecháním celé lokality dosáhly modely hodnot $R^2 = -0{,}03$, 0,05 a 0,11. Náhodně členěná CV naproti tomu nadhodnotila zdánlivou

predikční schopnost až o $\Delta R^2 = 0{,}57$. Zásadním omezením byla geometrie pozorování: říční úsek o mediánové šířce v nejužší třídě zabíral pouze 16 % plochy pixelu o rozlišení 30 m, a zaznamenaná odrazivost proto byla určována především okolní souší a efektem sousedství. Dlouhodobé záznamy přesto prokázaly pokles koncentrací o 80 %, 48 % a 66 % a následnou stabilizaci přibližně od roku 2007. Toto zlepšení mělo regionální charakter, bylo jednotné napříč šířkovými třídami a žádný z devíti víceměřítkových testů změn krajinného pokryvu v příbřežní zóně jej při prostorové validaci nevysvětlil.

Studie 3 porovnala přímý optický odhad chemické spotřeby kyslíku (COD) a pětidenní biochemické spotřeby kyslíku ($BOD_5$) s modelem náhodného lesa založeným na družicově pozorovatelných charakteristikách povodí a ukazatelích zatížení z bodových zdrojů. Modelovanou veličinou byla koncentrace, nikoli látkový tok. Optické modely při předpovědi v dosud nepozorovaných říčních úsecích selhaly: Nashův–Sutcliffeův koeficient účinnosti (NSE) dosáhl pouze 0,02 pro COD a 0,03 pro $BOD_5$ a nativní snímky s prostorovým rozlišením 10 m nepřinesly žádné zlepšení. Model využívající charakteristiky povodí naproti tomu dosáhl NSE 0,29 a 0,40 a Klingova–Guptova koeficientu účinnosti (KGE) 0,33 a 0,47 pro COD a $BOD_5$ na $\log_{10}$ transformované škále. Nejvyšší přesnost predikce $BOD_5$ byla dosažena v nejužších tocích. Rekonstrukce pro období 1985–2025 odhalila počáteční pokles následovaný částečným nárůstem, zatímco citlivostní analýza pro rok 2050 ukázala, že opatření v oblasti využití půdy měla v testovaných scénářích podstatně větší vliv než oteplování.

Souhrnně tyto tři studie posouvají monitorování vnitrozemských vod od pouhého porovnávání přesnosti algoritmů k měřítkově orientovanému rámci, který hodnotí, zda lze cílovou veličinu vůbec spolehlivě odvodit. Práce vymezuje dvojí hranici neoddělitelnosti: optickou hranici, za níž nelze slabý signál cílové složky oddělit od příspěvku souše v subpixelovém měřítku a od efektu sousedství, a atribuční hranici, za níž prostorová závislost a regionálně koherentní působení řídicích faktorů znemožňují rozložit změny v měřítku povodí na místní příčiny. Skóre blízké nule při prostorové validaci proto není pouze známkou nedostatečného modelu, ale dokládá, že požadovaná informace není v daném pozorovacím měřítku k dispozici. Zásadním výsledkem práce je rovněž nalezení cesty za tuto hranici: pokud odrazivost vodní hladiny neumožňuje vědecky obhajitelný přímý odhad, lze část chybějících informací získat z charakteristik povodí. Ústřední přínos práce je proto trojí: vytvoření dlouhodobé rekonstrukce tam, kde to optické a prostorové podmínky umožňují; kvantitativní určení mezí fyzikální pozorovatelnosti a statistické přenositelnosti; a vytvoření prostorově validované alternativy založené na charakteristikách povodí pro úzké vodní toky. Navržený rámec poskytuje vědecky obhajitelný základ pro kombinování družicových pozorování, informací o povodí a in situ měření v měřítkách, v nichž každý z těchto zdrojů skutečně nese využitelnou informaci, a tím zvyšuje věrohodnost dlouhodobých rekonstrukcí, atribuce příčin a řízení změn jakosti vnitrozemských vod.

# List of abbreviations

| | |
|---|---|
| AP | atmospheric pressure |
| AT | air temperature |
| AWEI | automated water extraction index |
| $BOD_5$ | five-day biochemical oxygen demand |
| BRDF | bidirectional reflectance distribution function |
| CDOM | colored dissolved organic matter |
| CES | $CO_2$ emissions and sequestration |
| Chl-a | chlorophyll-a |
| CHMI | Czech Hydrometeorological Institute |
| COD | chemical oxygen demand |
| CORINE | Coordination of Information on the Environment |
| CoV | coefficient of variation |
| CV | cross-validation |
| DOC | dissolved organic carbon |
| DOCc | dissolved organic carbon concentration |
| DOCs | dissolved organic carbon storage |
| EPSG | coordinate reference system registry code |
| ERA5 | fifth-generation ECMWF atmospheric reanalysis |
| ESA | European Space Agency |
| ETM+ | Enhanced Thematic Mapper Plus (Landsat 7) |

| | |
|---|---|
| FAI | floating algae index |
| GBT | gradient-boosted trees |
| GDP | gross domestic product |
| GEE | Google Earth Engine |
| HLS | Harmonized Landsat and Sentinel-2 |
| IF | Isolation Forest |
| IQR | interquartile range |
| KGE | Kling-Gupta efficiency |
| KGE′ | revised Kling-Gupta efficiency |
| LaSRC | land surface reflectance code |
| LEDAPS | Landsat Ecosystem Disturbance Adaptive Processing system |
| LOSO | leave-one-site-out |
| MAE | mean absolute error |
| MAPE | mean absolute percentage error |
| MBPNN | multilayer back-propagation neural network |
| MICE | multivariate imputation by chained equations |
| ML | machine learning |
| MNDWI | modified normalized difference water index |
| MODIS | moderate resolution imaging spectroradiometer |
| NDTI | normalized difference turbidity index |
| NDVI | normalized difference vegetation index |

| | |
|---|---|
| NDWI | normalized difference water index |
| $NH_4^+$-N | ammonium nitrogen |
| $NO_3^-$-N | nitrate nitrogen |
| NSE | Nash-Sutcliffe efficiency |
| NTL | nighttime light |
| OCSF | object-level cross-sensor framework |
| OLI | operational land imager (Landsat 8/9) |
| OLS | ordinary least squares |
| PBIAS | percent bias |
| PCA | principal component analysis |
| POP | population |
| QA | quality assurance |
| QC | quality control |
| $R^2$ | coefficient of determination |
| RF | random forest |
| RMSE | root mean square error |
| RPD | residual prediction deviation |
| $R_{rs}$ | remote-sensing reflectance |
| SD | standard deviation |
| SDD | secchi disk depth |
| SEM | structural equation modelling |

| | |
|---|---|
| SVM | support vector machine |
| TM | thematic mapper (Landsat 5) |
| TP | total phosphorus |
| TSI | trophic state index |
| USGS | United States Geological Survey |
| WD | water depth |
| WFD | water framework directive |
| WQP | water quality parameter |
| WSC | water storage capacity |
| WS | wind speed |
| XGBoost | extreme gradient boosting |

# Table of contents


| | | |
|---|---|---|
| **Chapter I** | General Introduction | 3 |
| **Chapter II** | Research Objectives and Questions | 13 |
| **Chapter III** | State of the Art | 19 |
| **Chapter IV** | Materials and Methods | 42 |
| **Chapter V** | Results and Discussion | 112 |
| **Chapter VI** | Conclusions | 189 |
| **Chapter VII** | Limitations and Future Perspectives | 203 |
| **References** | | 209 |
| **Curriculum vitae & List of publications** | | 227 |

# Chapter I

## General Introduction

# Content

1.1. Inland waters in the global carbon and nutrient cycles ..... 3
1.2. The monitoring gap in space and time ..... 5
1.3. Satellite remote sensing and machine learning ..... 6
1.4. Problem statement: retrievability, limits, and attribution ..... 9

# 1. General Introduction

## 1.1. Inland waters in the global carbon and nutrient cycles

Inland waters, including lakes, reservoirs, rivers, streams, and wetlands, occupy a relatively small fraction of the continental surface, yet they are active components of the global carbon and nutrient cycles. They receive material from terrestrial ecosystems, transform it within the aquatic continuum, exchange gases with the atmosphere, and transport or retain carbon and nutrients downstream. Lakes and reservoirs therefore function as biogeochemical reactors rather than passive conduits, with consequences for greenhouse-gas emissions, carbon burial, nutrient retention, and water quality (Cole et al., 2007; Downing, 2010; Tranvik et al., 2009).

Carbon enters inland waters in dissolved, particulate, and inorganic forms. Aquatic primary producers assimilate dissolved inorganic carbon into organic matter, whereas respiration, photochemical processing, and microbial decomposition return carbon to the water column and atmosphere as carbon dioxide and, under reducing conditions, methane. A fraction of organic and inorganic carbon is retained in sediments or exported through the river network (Regnier et al., 2022; Tranvik et al., 2018). For lakes, the global standing pool of dissolved organic carbon (DOC) in lake water has been estimated at approximately 0.729 Pg C (729 Tg), with most of that pool associated with large lakes; this value is a stock estimate, not an annual atmospheric flux (Toming et al., 2020).

DOC is chemically heterogeneous and is supplied by both terrestrial materials transported from catchments and organic matter produced within the water body. It supports heterotrophic metabolism, is altered by microbial and photochemical degradation, and can influence light penetration and the availability of nutrients and trace substances. Only part of the DOC pool is optically active: its chromophoric fraction, commonly described as colored dissolved organic matter (CDOM), absorbs and modifies light and can therefore provide a remote-sensing signal. The relationship

between DOC and CDOM is nevertheless variable because it depends on source composition, molecular structure, residence time, and decomposition history (Cole et al., 2007; Sobek et al., 2007; Tranvik et al., 2018). In eutrophic lakes, phytoplankton phenology and terrestrial inflow can further alter the DOC–CDOM relationship, which is one reason DOC retrieval is more difficult than retrieving a directly optical constituent (D. Liu et al., 2021).

Nutrients, ammonium nitrogen ($NH_4^+$-N), nitrate nitrogen ($NO_3^-$-N), and total phosphorus (TP), provide a second major pathway through which inland waters connect catchment processes to ecosystem change. Excess nitrogen and phosphorus increase primary production and can promote algal and cyanobacterial blooms, oxygen depletion, habitat degradation, and risks to drinking-water supply and recreation. Decades of experimental and comparative research show that reducing phosphorus is essential in many systems, but that effective control often requires coordinated management of both phosphorus and nitrogen (Carpenter et al., 1998; Conley et al., 2009; Smith and Schindler, 2009). The issue remains widespread and costly: one assessment estimated potential annual economic damages from eutrophication in United States freshwaters at approximately US$2.2 billion (Dodds et al., 2008). Satellite analyses also indicate that the risk of lacustrine algal blooms increased during the 2010s, while harmful algal blooms have become a globally significant inland-water hazard (Feng et al., 2024; Hou et al., 2022; Le Moal et al., 2019).

Organic loading links catchment wastewater and diffuse pollution to the oxygen regime of running waters. Chemical oxygen demand (COD) is an operational measure of chemically oxidizable material, whereas five-day biochemical oxygen demand ($BOD_5$) indexes the oxygen consumed by microbial degradation over a five-day incubation. Neither parameter is a direct compositional measurement of all organic carbon, but both are useful integrative indicators of oxygen demand and organic pollution. River organic pollution is also closely connected to nutrient cycling, self-purification, ecological condition, and human water security (Meybeck, 2003; Vorosmarty et al., 2010). In the European policy context, the Water Framework

Directive (WFD) links river-basin management to ecological and chemical status, including the support of physicochemical conditions such as oxygen balance and nutrient status, while the Urban Wastewater Treatment Directive establishes minimum treatment requirements for municipal wastewater (European Parliament and Council, 2024, 2000; Hering et al., 2010).

These carbon, nutrients, and organic-loading pathways are tightly coupled but are not interchangeable. DOC and CDOM describe a carbon pool with a variable optical expression; dissolved nutrients are generally not directly optical; and COD and $BOD_5$ represent aggregate oxygen-demand responses. Therefore, a general monitoring framework must distinguish between what can be observed directly in water reflectance, what can be inferred from correlated constituents, and what must be explained using hydrological and catchment information.

## 1.2. The monitoring gap in space and time

Freshwater quality is highly heterogeneous in both space and time. Concentrations can change over short distances along a river and over timescales ranging from hours to days in response to discharge, rainfall, mixing, biological activity, wastewater releases, and agricultural runoff. Conventional sampling and gauging remain essential because they provide direct measurements and process context. But the networks are often spatially sparse, temporally irregular, and unevenly distributed across small catchments. Global compilations such as the Global River Water Quality Archive illustrate both the value of historical observations and the heterogeneity of their spatial, temporal, and analytical coverage (Virro et al., 2021).

The gap is most consequential in low-order rivers and narrow headwaters. These channels receive diffuse agricultural inputs and small or intermittent point sources, integrate landscape processes before water reaches larger rivers, and can exert a disproportionate influence on downstream water quality. Therefore, land use is inseparable from river conditions, but the reaches most directly connected to local land management are commonly the least monitored (Allan, 2004). At the same time, a large

fraction of the global river network is narrower than the width needed to obtain a medium-resolution pixel dominated by water; the global river-width inventory of Allen and Pavelsky (2018) makes clear that many channels fall into this sub-pixel regime.

Long-term assessment adds a second difficulty. In situ programmes may contain gaps, changes in sampling frequency, changes in laboratory methods, and shifts in station location or catchment pressure. A trend inferred from a few stations can therefore confound regional change with hydrological sampling, station placement, or programme design. The problem is not simply the absence of observations. It is the mismatch between the spatial scale of management questions and the spatial and temporal support of the observations available to answer them (Sagan et al., 2020; Virro et al., 2021).

A monitoring strategy for inland waters must consequently provide both repeated temporal coverage and spatial context. For large lakes, this means resolving basin-scale gradients and seasonal dynamics. For rivers, it means retaining information across connected drainage systems while acknowledging that narrow channels, mixed pixels, and episodic pollution may remain below the sensor's resolution or the sampling programme's sampling interval. These requirements motivate the use of satellite Earth observation as a complement to, rather than a replacement for, field monitoring.

## 1.3. Satellite remote sensing and machine learning

Satellite Earth observation complements conventional water-quality monitoring by converting discrete field measurements into repeated, spatially explicit observations. It cannot replace laboratory analysis, but it can reveal spatial gradients, episodic events, and long-term changes that sparse station networks may miss. Advances in atmospheric correction, bio-optical modelling, spectral indices, and machine learning (ML) have expanded the use of remote sensing in inland waters (Palmer et al., 2015; Sagan et al., 2020; Topp et al., 2020). Satellites now support mapping of water extent, color, transparency, suspended sediment, chlorophyll-a (Chl-a), and algal blooms across areas

that field campaigns cannot efficiently cover (Dethier et al., 2022; Gardner et al., 2021; Pekel et al., 2016).

The Landsat archive is particularly valuable for long-term freshwater research. The record used here begins in 1984 and provides a medium-resolution basis for multi-decadal reconstruction (Maciel et al., 2023; Pekel et al., 2016). Measurements from the Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Operational Land Imager (OLI) are not spectrally identical, so uncorrected differences can introduce artificial discontinuities. Sentinel-2 adds finer resolution and denser recent coverage. The Harmonized Landsat Sentinel-2 (HLS) product places compatible observations on a common 30 m grid (Claverie et al., 2018), while Google Earth Engine (GEE) makes consistent archive-scale processing practical (Gorelick et al., 2017). HLS improves recent cross-sensor consistency but cannot replace the pre-Sentinel Landsat history.

Retrieval feasibility depends first on optical causation. Chl-a, suspended particles, and colored dissolved organic matter (CDOM) alter water-leaving reflectance through absorption and scattering. Their signals can support empirical, semi-analytical, or ML retrievals when the water surface is adequately resolved. Other variables are observed only through proxies. Dissolved organic carbon (DOC) may be related to blue-wavelength CDOM absorption, but this relationship varies among water bodies and seasons (Brezonik et al., 2015; Harkort and Duan, 2023). Dissolved nutrients and aggregate oxygen-demand indicators do not exhibit a unique reflectance signal at typical environmental concentrations. Their estimates instead depend on associations with phytoplankton, suspended sediment, water colour, land use, or point-source pressure (Odermatt et al., 2012; Sagan et al., 2020). Such relationships may be locally informative but do not constitute direct optical detection.

Observation geometry and preprocessing impose further limits. Because water reflectance is low, errors in atmospheric correction, cloud masking, or temporal matching can be large relative to the aquatic signal. In narrow rivers, mixed pixels combine water, banks, and riparian vegetation, while atmospheric scattering introduces

radiance from adjacent land (Bulgarelli and Zibordi, 2018; Kuhn et al., 2019). These effects intensify as the water-filled pixel fraction decreases. Finer spatial resolution can reduce mixing but cannot create a constituent-specific signal where none exists. Channel width and optical activity must thus be considered together (Allen and Pavelsky, 2018).

ML remains valuable within these physical limits. Random forests, gradient-boosted trees, support vector machines, and neural networks can represent nonlinear relationships among spectral bands, indices, and environmental covariates (Cao et al., 2020; Harkort and Duan, 2023; D. Liu et al., 2021; Pahlevan et al., 2022). Large matchup archives have also widened the calibration domain. AquaSat, for example, contains more than 600,000 pairings of Landsat reflectance and in situ water-quality observations (Ross et al., 2019). Dataset size alone does not ensure transferability, which also depends on the representation of optical water types, concentration ranges, seasons, sensors, and regions. A continental riverine TP model was developed only for Landsat-detectable rivers wider than approximately 30 m (Ramtel et al., 2024). Its success does not imply equivalent performance in channels only a few meters wide. ML cannot recover an absent optical signal or compensate fully for an unsupported prediction domain.

Validation design is thus part of the scientific question. Random cross-validation (CV) can include observations from neighboring sites or repeated observations from the same site in both the training and test sets. In spatially autocorrelated networks, this leaks local information into model evaluation, potentially overstating transferability (Karasiak et al., 2022; Ploton et al., 2020; Roberts et al., 2017). Spatially blocked CV and leave-site-out CV provide more defensible tests when prediction is intended for new locations. Their configuration must match the deployment target because block size, separation distance, sampling density, and environmental coverage affect the estimate (Milà et al., 2022a; Wadoux et al., 2021). Nearest-neighbor distance matching and area-of-applicability analysis can further identify predictions outside the calibration domain (Ludwig et al., 2023; Meyer and Pebesma, 2021).

This dissertation applies these principles within a common framework. Satellite archives are used where the target has a defensible optical or catchment-based pathway. Model claims are bounded by water-body size, sensor footprint, matchup quality, and validation design. Paired random and spatial CVs distinguish interpolation from transfer to unmonitored sites among familiar observations. This framework places successful reconstruction, failed optical retrieval, and catchment-based prediction within the same observational logic.

## 1.4. Problem statement: retrievability, limits, and attribution

This dissertation treats long-term water-quality mapping as an inference problem with three linked boundaries. The first is the retrievability boundary: whether the target variable contains a reproducible signal in the observed water reflectance. The second is the physical-resolution boundary: whether that signal survives atmospheric correction, mixed pixels, adjacency effects, changing water levels, and the sensor's spatial scale. The third is the attribution boundary: whether a mapped trend can be related to catchment management, hydrology, point sources, climate, or infrastructure rather than being interpreted as a purely descriptive pattern.

For DOC, the retrievability problem arises because only the CDOM-related fraction is directly visible and because CDOM–DOC relationships change with source and processing history. In eutrophic lakes, algae and suspended particles can further obscure or mimic the DOC-related signal. Existing work in Lake Taihu and at larger inland-water scales demonstrates the feasibility of satellite-based DOC estimation. It also shows why model performance and long-term extrapolation must be interpreted in relation to the calibration domain and optical environment (Harkort and Duan, 2023; D. Liu et al., 2021; Sobek et al., 2007).

For narrow rivers, the physical-resolution boundary is especially restrictive. A channel that is only a few meters wide may underfill a 10–30 m pixel, so the measured spectrum can include contributions from water, banks, vegetation, exposed sediment, and atmospheric or adjacency effects. Atmospheric-correction choices can materially

alter river reflectance and the retrieval of even optically active variables (Kuhn et al., 2019). Object-based aggregation and cross-sensor harmonization can improve the representation of reach-scale information, but they cannot create a pure-water signal where the channel is not resolved (Allen and Pavelsky, 2018; Blaschke, 2010a; Claverie et al., 2018).

Attribution requires a complementary catchment perspective. Water-quality trajectories arise from the interaction of land use, hydrology, climate, wastewater treatment, point-source pressure, and in-stream processing. Satellite observations can provide repeated measures of land-surface change and water appearance, but causal interpretation requires explicit temporal and spatial design, independent knowledge of management interventions, and validation that respects the dependence structure of the monitoring network (Allan, 2004; Gorelick et al., 2017; Vorosmarty et al., 2010). When the water signal is not sufficiently identifiable, satellite-observable catchment attributes may provide a more defensible route to prediction and scenario analysis than a forced optical inversion.

The studies that follow apply this framework across two complementary settings. The first uses the large, eutrophic Lake Taihu to reconstruct four decades of DOC concentration and storage from Landsat observations and an optimized Machine learning model. The second uses agricultural rivers in South Bohemia, Czech Republic, to test whether dissolved nutrients and oxygen-demand indicators can be retrieved in narrow channels under spatially honest validation, and to determine when catchment attributes provide a stronger basis for prediction. Together, these studies address three overarching objectives: (1) quantify long-term carbon, nutrient, and organic-loading dynamics from a multi-decadal Earth-observation archive; (2) identify the optical, spatial, and validation limits that govern transferability; and (3) separate descriptive reconstruction from defensible attribution and future-oriented management inference.

# Chapter II

## Research Objectives and Questions

# Content

2.1. Overall aim ..... 13
2.2. Specific objectives ..... 13
2.3. Research questions and hypotheses ..... 14

# 2. Research Objectives and Questions

## 2.1. Overall aim

The overall aim of this dissertation is to determine the extent to which ML, using multi-source satellite archives and ancillary environmental and catchment data, can reconstruct and attribute the long-term dynamics of carbon and water-quality variables in inland waters, and to identify the physical and statistical limits within which such inferences remain defensible. This aim is addressed through three first-author studies spanning a gradient from a large eutrophic lake to a network of narrow agricultural rivers. Together, these studies progress from successful retrieval and attribution, through rigorous identification of the limits of direct optical retrieval, to the evaluation of a catchment-based alternative where optical retrieval is unreliable.

A unifying methodological principle runs through the three studies: predictive skill is assessed honestly against the data's spatial structure. When the analytical unit is a spatially clustered network of river reaches, models are evaluated using spatial CV, which requires predictions at unseen sites. This ensures that reported skill reflects genuine, transferable information rather than interpolation among autocorrelated neighbors. This principle makes the positive and negative results across the dissertation directly comparable and ensures that conclusions about retrievability and attribution are conservative.

## 2.2. Specific objectives

The specific objectives, organized by the three studies, are as follows:

Study 1, DOC dynamics in a large eutrophic lake. The objectives were: (1) to reconstruct four decades (1984–2023) of DOC concentration and storage in Lake Taihu by integrating the Landsat archive with ML, selecting an optimal model among four candidate algorithms; (2) to quantify the spatiotemporal variability of DOC concentration and storage across the lake and its sub-regions; and (3) to identify and

rank the meteorological, anthropogenic, bio-optical, and vegetation drivers of DOC variability and to separate their direct and indirect pathways.

Study 2 examined the satellite retrieval and attribution of nutrients in narrow rivers. Its objectives were: (1) to determine whether dissolved nutrients ($NH_4^+$-N, $NO_3^-$-N, TP) could be retrieved across sites from a harmonized multi-sensor archive of narrow agricultural rivers and to quantify the sensitivity of retrieval performance to CV design; (2) to characterize long-term nutrient recovery and assess whether it could be attributed to local, reach-scale land-surface change; and (3) to formalize the identified limits as a double inseparability boundary and develop a transferable, spatially honest validation protocol for river water-quality monitoring.

Study 3 examined oxygen demand through catchment observation. Its objectives were: (1) to determine whether COD and $BOD_5$ could be retrieved optically across sites in narrow-river networks under spatial validation and, if not, to identify the underlying limitations; (2) to assess whether oxygen-demand concentrations could instead be predicted from satellite-observable catchment attributes and point-source pressure and to reconstruct their four-decade dynamics from 1985 to 2025; and (3) to evaluate the sensitivity of oxygen demand to plausible land-use and climate scenarios for 2050 and identify the factor offering the greatest management leverage.

Together, the three studies move from successful lake-scale retrieval to identifying retrieval limits in narrow rivers and, finally, to a catchment-based alternative, thereby addressing the overarching aim across contrasting inland-water systems.

## 2.3. Research questions and hypotheses

The three studies were guided by eight research questions (RQ1–RQ8), derived from the knowledge gaps identified in Sections 1.4 and 3.6. Each question is paired with a working hypothesis (H1–H8) stated so that it can be confirmed or rejected against an advance-defined criterion. The outcome of each hypothesis is stated explicitly in Chapter 6.

- **Study 1: Dissolved organic carbon in a large eutrophic lake**

RQ1. Can four decades of DOC concentration and storage in a large, optically complex eutrophic lake be reconstructed from the Landsat archive using ML?

H1. Non-linear ensemble learning retrieves DOC from Landsat surface reflectance with an accuracy sufficient for multi-decadal reconstruction (validation $R^2 \geq 0.7$) and outperforms neural-network and support-vector alternatives.

RQ2. Which meteorological, anthropogenic, bio-optical, and vegetation factors dominate four-decade DOC variability, and through which direct and indirect pathways do they act?

H2. Meteorological forcing is the first-order control, acting both directly and indirectly through bio-optical and vegetation pathways, rather than anthropogenic pressure acting alone.

- **Study 2. Dissolved nutrients in narrow rivers**

RQ3. Can dissolved nutrients ($NH_4^+$-N, $NO_3^-$-N, TP) in narrow rivers be retrieved at sites not seen during training?

H3. They cannot (leave-site-out $R^2 < 0.2$) because their optical activity is weak and indirect, and because narrow channels underfill the pixel.

RQ4. To what extent does random CV overstate apparent skill relative to spatial CV for spatially clustered river data?

H4. Random CV inflates apparent skill substantially, by $\Delta R^2 > 0.3$ within at least one river-width class, so that reported performance reflects interpolation among autocorrelated neighbours rather than transferable information.

RQ5. Is the long-term nutrient recovery of an agricultural river network a local, reach-scale signal attributable to land-cover change, or a regional signal that is uniform across channel widths?

H5. The recovery is regional and width-uniform, and reach-scale land-cover change explains a negligible share of the observed decline.

- **Study 3. Organic loading from catchment observation**

RQ6. Where optical retrieval of oxygen demand (COD, $BOD_5$) fails, can it be predicted from satellite-observable catchment properties and point-source pressure, and how does that skill vary with channel width?

H6. A catchment-attribute model predicts oxygen-demand concentrations across unseen sites with useful skill (leave-one-site-out (LOSO) Nash-Sutcliffe efficiency (NSE) > 0.25) and performs best in the narrow headwaters, where optical retrieval fails most severely.

RQ7. How has oxygen demand evolved over the four-decade, and how will it respond to land-use and climate scenarios through 2050? Is climate or catchment management the dominant lever?

H7. Catchment land-use management, rather than climate forcing, is the dominant lever on oxygen demand under plausible 2050 scenarios.

- **Synthesis**

RQ8. Taken together, what is within reach of satellite-plus-ML monitoring of inland-water quality, and what lies beyond the physical and statistical boundary imposed by sub-pixel mixing, adjacency contamination, weak optical activity, and spatial autocorrelation?

H8. Retrievability is a property not of the algorithm, but of the variable–system combination, set jointly by optical activity and by the water fraction of the pixel, and bounded by a double inseparability boundary, optical and attributional, that model flexibility cannot overcome.

# Chapter III

## State of the Art

# Content

3.1. The inland-water carbon continuum and its observational requirements ..... 19
3.2. Target variables and their optical accessibility ..... 23
3.3. Optical principles, satellite sensors and the long-term archive ..... 27
3.4. Retrieval methods and the evidence by variable ..... 31
3.5. Validation, long-term reconstruction and attribution ..... 35
3.6. Study systems and knowledge gaps ..... 37

# 3. State of the Art

This chapter sets out the background on which the three studies build: the inland-water carbon continuum and the observation it demands, the target variables and their very different accessibility to an optical sensor, the optical principles and the long-term satellite archive, the retrieval methods and the evidence for each variable, the cross-cutting questions of validation and attribution, and the two study systems with the gaps the studies close.

## 3.1. The inland-water carbon continuum and its observational requirements

Inland waters cover only a few per cent of the non-glacial land surface, yet they are neither passive nor quantitatively negligible in the global carbon cycle (Downing et al., 2006; Hoegh-Guldberg et al., 2019). The image of a passive pipe has been replaced by that of an aquatic continuum, a connected sequence of soils, groundwater, streams, rivers, ponds, lakes, reservoirs and estuaries through which terrestrially fixed carbon is progressively transformed before a residual fraction reaches the sea. Along it, carbon is repeatedly processed rather than merely transported: outgassed as carbon dioxide and methane, buried in sediments, and chemically and biologically altered, so that what enters the terrestrial margin differs markedly from what leaves at the coastal terminus. Synthesis estimates place carbon dioxide evasion from lakes alone at no less than 0.729 Pg C $y^{-1}$, comparable to the net oceanic carbon sink, and greenhouse-gas release from lakes and impoundments remain among the least constrained terms in regional carbon budgets (Cole et al., 2007; DelSontro et al., 2018; Regnier et al., 2022; Toming et al., 2020).

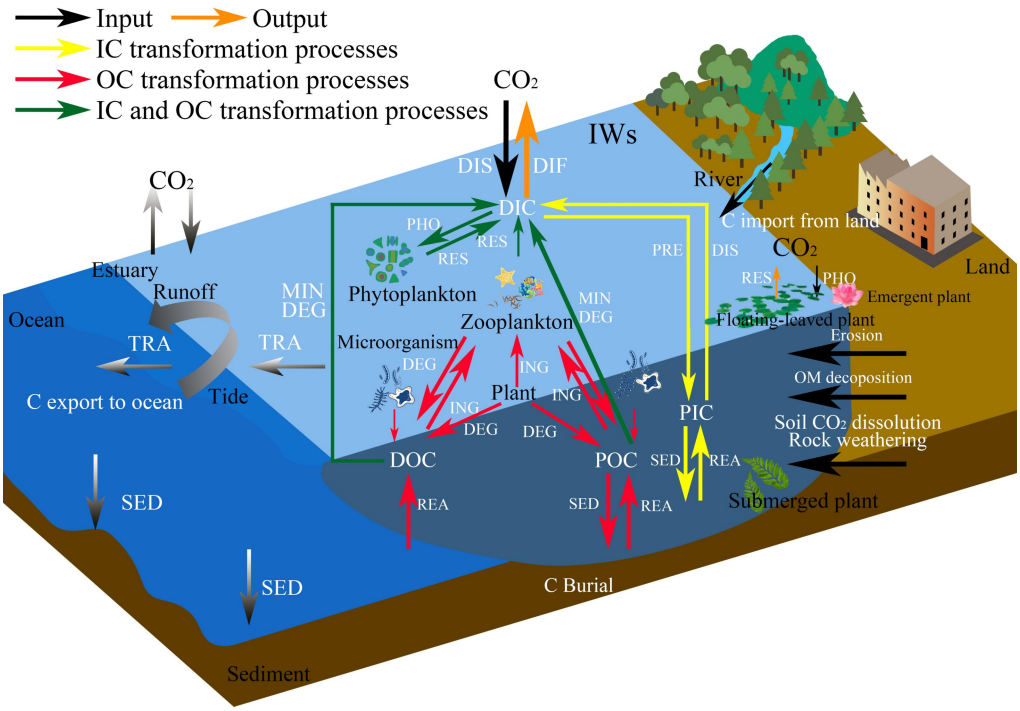



**Figure 3.1.** Carbon transport and transformation in inland waters. DIC, dissolved inorganic carbon; DOC, dissolved organic carbon; PIC, particulate inorganic carbon; POC, particulate organic carbon; OM, organic matter; PRE, precipitation; DIS, dissolution; PHO, photosynthesis; RES, respiration; SED, sedimentation; REA, release; MIN, mineralization; DEG, degradation; ING, ingestion; TRA, transport; DIF, diffusion. Thick arrows denote carbon processes in the estuary and thin arrows those in inland water bodies, including lakes, ponds, wetlands, rivers and reservoirs.

The processes summarized in Figure 3.1 act on two coupled pools. The inorganic pool exchanges carbon dioxide with the atmosphere by dissolution and diffusion, interconverts between dissolved and particulate forms, and delivers particulate inorganic carbon to the sediment, from which it can be released again (Gao and Yu, 2020). The organic pool is generated and reworked biologically, and particulate organic carbon is either buried or returned to the water column. Between the pools, aquatic plants and phytoplankton fix carbon dioxide, respiration returns it to the atmosphere and to the dissolved inorganic pool, and microbial mineralization together with

photodegradation converts organic to inorganic carbon (Bogard et al., 2017). Table 3.1 defines the terms used throughout this dissertation for these forms and fluxes. How the three fates, evasion, burial and downstream export, are partitioned depends on residence time, temperature, the lability of the organic matter, light and nutrient status: decomposition rates scale with retention time, so identical organic matter is processed to very different extents in a headwater stream and in a reservoir (Catalán et al., 2016a), and small, shallow systems are disproportionately reactive per unit area while large lakes and reservoirs integrate and store carbon over longer timescales (Cole et al., 2007; Regnier et al., 2022).

**Table 3.1.** Terminology used in this dissertation for carbon fluxes and processes in inland waters, adapted from the author's literature review of methodologies for assessing DOC.

| Term | Definition |
|---|---|
| Input | Organic and inorganic carbon of terrestrial origin transferred laterally from land to inland waters by wind, surface and subsurface runoff, or groundwater |
| Aquatic primary production | Fixation of atmospheric $CO_2$ by photosynthesis in aquatic primary producers |
| Burial | Storage of carbon, primarily organic but possibly inorganic, in sediments, removing it from the active carbon cycle on decadal to millennial timescales |
| Emissions (evasion, outgassing) | Release of $CO_2$ and $CH_4$ from inland waters to the atmosphere by diffusion, ebullition or degassing |
| Export | Organic and inorganic carbon delivered from inland waters downstream to the ocean |
| Carbon quality | Chemical composition of organic carbon, which governs properties such as chromophoricity (color) and lability |
| Diffusion | Emission of $CO_2$ and $CH_4$ at the air–water interface along a chemical gradient |
| Ebullition | Emission of $CH_4$ as bubbles produced in the sediment, transported through the water column and released into the atmosphere |
| Degassing | Emission of $CO_2$ and $CH_4$ by turbulent diffusion, for example during water release through turbines or spillways and with irrigation |
| Sedimentation | Breakdown of organic carbon followed by settling and deposition |

What enters the water is set at the land surface, where climate, vegetation, topography and landform control the carbon flux across the land-water interface and are overlain by mining, agriculture, animal husbandry, urbanization, industry and

impoundment (Chi et al., 2020). Figure 3.2 summarizes these mechanisms together with the observational hierarchy they require: in-situ sampling is the most accurate but confined to points, aerial platforms extend the scope at intermediate uncertainty, and only satellites deliver synoptic, repeatable and retrospective coverage at regional to global scale (Batina and Krtalić, 2023; Malerba et al., 2023). The temporal implication is equally important. A few years of monitoring cannot reliably distinguish directional change from natural variability, while many in situ records are too short or discontinuous to resolve long-term trajectories. The long-term satellite archive provides the temporal depth and spatial consistency needed to separate sustained trends from short-term variability at the scale required by the continuum perspective.

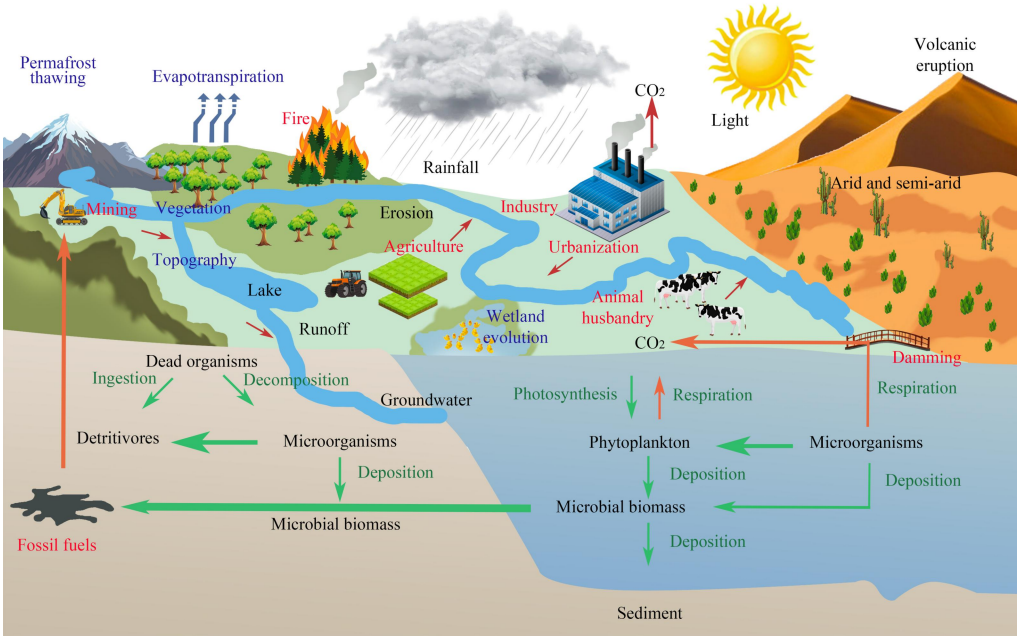



**Figure 3.2.** Control mechanisms for carbon transport and transformation in inland waters and the corresponding observational hierarchy. Red text marks anthropogenic processes and blue text processes related to climate change; thick red arrows represent carbon transport and emission, thin green arrows carbon input and biotransformation, and thin orange arrows carbon output and biotransformation. The left-hand panel shows the scope of monitoring and the platforms employed, from site or plot sampling through unmanned aerial systems and aircraft to satellites, together with the associated level of uncertainty.

## 3.2. Target variables and their optical accessibility

Carbon in inland waters divides into an organic and an inorganic pool, each with a dissolved and a particulate fraction, and only part of one of those four compartments carries an optical signal (Figure 3.3). DOC is operationally defined as the organic carbon passing a fine filter and constitutes roughly half of the total carbon pool in many lake waters (Cole et al., 2007). Within DOC, the chromophoric fraction, CDOM, is the light-absorbing component, while the remainder is divided into hydrophilic and hydrophobic components that leave no direct spectral signature. Optical accessibility is therefore a property not of the analyte but of the fraction of it that interacts with light.

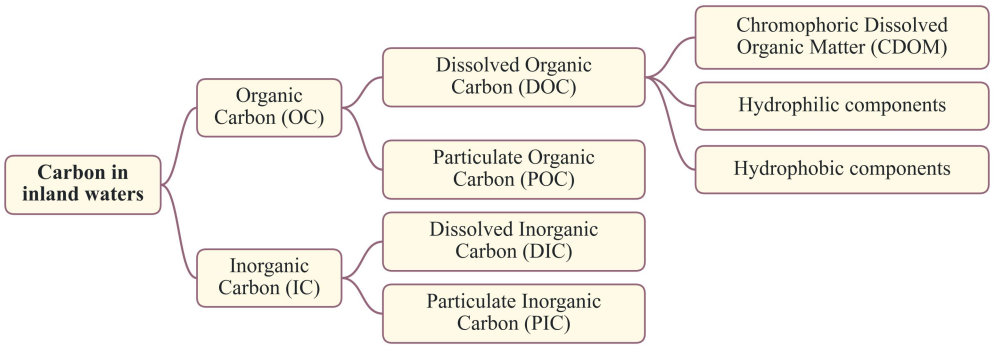



**Figure 3.3.** Classification of carbon in inland waters. Of the four compartments, only the chromophoric fraction of DOC interacts appreciably with light, which is why every optical estimate of DOC is an inference through a proxy rather than a direct measurement.

The DOC pool is sustained by two source classes whose proportions shape its optical behavior. Allochthonous DOC, leached from soils, peatlands and terrestrial vegetation, is typically aromatic, of high molecular weight and strongly colored; autochthonous DOC, produced by phytoplankton exudation, grazing and cell lysis, is comparatively aliphatic, less colored and more readily assimilated (Baines and Pace, 1991; Stedmon and Markager, 2005). Both are continuously reworked, as photochemical oxidation cleaves colored aromatic molecules into smaller and less colored compounds while microbial metabolism consumes the labile fractions, and

molecular-level work shows that persistence is governed by intrinsic molecular properties as much as by ambient conditions (Kellerman et al., 2015, 2014; Kothawala et al., 2014). DOC in turn governs the functioning of the water body, setting the depth of the photic zone, darkening the water, strengthening stratification and supplying the energy that sustains the heterotrophic food webs whose respiration drives the evasion described above; the widespread increases known as browning have been linked to recovery from acid deposition, changing land use, hydrological intensification and warming (Evans et al., 2017; Roulet and Moore, 2006).

In remote sensing, the key point is that DOC is only partially optically active. CDOM absorbs strongly in the ultraviolet and blue, with a near-exponential decline towards the red, and serves as a proxy for DOC (Brezonik et al., 2015; Kirk, 1994; Kutser, 2012). Because CDOM is a mixture of diverse origins, its quantity is defined operationally through the Napierian absorption coefficient at a reference wavelength, a (254) in drinking-water work, a (350) in limnology and a (440) in water-cooler remote sensing, where ocean-color sensors carry a channel near that wavelength (Lavonen et al., 2015; Massicotte et al., 2017; Zhu et al., 2014). Table 3.2 lists the parameters used to characterize CDOM abundance, composition and source. The relationship between CDOM absorption and DOC concentration is not universal: it varies in space and season with the source, and with the degree of photochemical and microbial alteration, and in temperate lakes it can fail outright (De Stefano et al., 2022; Griffin et al., 2018; Liu et al., 2020). In optically complex, eutrophic waters, the difficulty is compounded because phytoplankton pigments and suspended particles covary with and obscure the CDOM signal (Duan et al., 2014), which makes long-term DOC retrieval both valuable and technically demanding and motivates the use of flexible, non-linear estimators rather than simple band ratios.

**Table 3.2.** Parameters used to characterize the abundance, concentration, composition and source of CDOM, reproduced from the author's literature review.

| Parameter | Significance | Unit |
|---|---|---|
| a (254), a (280), a (350), a (375), a (440) | Absorption coefficients used to quantify CDOM abundance | $m^{-1}$ |
| DOC | DOC, used to quantify CDOM concentration | mg $L^{-1}$ |
| $SUVA_{254}$ | Ratio of a (254) to DOC, used to assess the aromatic degree of CDOM | $m^{-1}$ (mg $L^{-1}$)$^{-1}$ |
| S | Spectral slope, used to characterize CDOM composition and source | $\mu m^{-1}$, $nm^{-1}$ |
| $S^R$ | Spectral slope ratio, used to characterize CDOM composition and source | – |
| M | Relative molecular weight parameter | – |
| FI | Fluorescence index, used to trace CDOM composition and source | – |
| HIX | Humification index, used to assess the degree of humification | – |
| BIX | Biological index, used to quantify autochthonous CDOM | – |
| FPRI | Ratio of fluorescence peak C to peak T, used to track the relative contribution of autochthonous and allochthonous CDOM | – |
| $\delta^{13}C$, $\delta^{15}N$, $\delta^{2}H$, $\delta^{18}O$ | Stable isotopic composition of C, N, H and O, used to characterize CDOM composition and source | ‰ |
| O/C | Ratio of oxygen to carbon | – |
| H/C | Ratio of hydrogen to carbon | – |

The second family comprises the dissolved nutrients that drive eutrophication, of which the management-relevant forms are ammonium ($NH_4^+$-N), nitrate ($NO_3^-$-N) and TP; dual control of nitrogen and phosphorus is generally required to reverse enrichment (Carpenter et al., 1998; Conley et al., 2009; Smith and Schindler, 2009). Observing them across whole river networks is intrinsically difficult, because in flowing waters concentrations change sharply over short distances and respond within days to hydrology, land use and point sources (Tyler et al., 2016), and because continental analyses show trajectories reflecting interacting climatic, agricultural and policy forcing that scattered stations cannot resolve (Grizzetti et al., 2021). Crucially, dissolved nutrients are not optically active and can be inferred only through co-varying optically active constituents, so that even in favorable lake settings retrievals depend on turbidity surrogates and are markedly less reliable than retrievals of optically active variables (Ma et al., 2023; Mouw et al., 2015; Odermatt et al., 2012; Xiong et al., 2022).

The third family is organic loading, measured operationally as COD and $BOD_5$; it is a primary control on the oxygen regime and the ecological status of running waters and a core criterion under the European Union WFD (Hering et al., 2010; Meybeck, 2003; Vörösmarty et al., 2010). Like dissolved nutrients, it sits awkwardly within the optical paradigm because a chromophoric, particulate fraction interacts with light, whereas a substantial labile, colorless, and often point-source-derived fraction does not, yet it is strongly governed by catchment properties that are observable (Whitehead et al., 2009). Table 3.3 summarizes these differences, stating for each variable whether the analyte is optically active, the pathway by which a satellite estimate must be obtained, the principal constraint on that pathway, and the study in which the question is tested. The ordering it makes explicit is the one of the three studies' probes: a colored fraction gives DOC a genuine, if imperfect, optical handle, whereas dissolved nutrients and much of the organic loading offer none, so any apparent retrievability of the latter must be scrutinized as a possible artefact of covariation.

**Table 3.3.** Optical accessibility of the water-quality variables addressed in this dissertation, and the pathway by which each must be estimated from satellite reflectance.

| Variable | Optically active | Pathway to a satellite estimate | Principal constraint | Addressed in |
|---|---|---|---|---|
| Chl-a | Yes, pigment absorption is near 440 and 675 nm | Direct retrieval from band ratios, line-height indices or machine learning | Package effect; spectral overlap with CDOM and non-algal particles in Case 2 waters | Auxiliary variable, Studies 1–3 |
| Total suspended solids, turbidity | Yes, strong backscattering | Direct retrieval from red and near-infrared reflectance | Saturation at high load; bottom reflectance in shallow water | Auxiliary variable, Studies 1–3 |
| CDOM | Yes, exponential absorption in the ultraviolet and blue | Direct retrieval of the absorption coefficient at a reference wavelength | Separation from non-algal particle absorption is ill-posed | Study 1 |
| DOC | Only through its coloured fraction | Two-stage inference: reflectance to CDOM absorption, CDOM absorption to carbon | CDOM-to-DOC coupling varies with source; photobleaching removes colour but not carbon | Study 1 |

| Variable | Optically active | Pathway to a satellite estimate | Principal constraint | Addressed in |
|---|---|---|---|---|
| $NH_4^+$-N and $NO_3^-$-N | No | Proxy inference through Chl-a or turbidity | No measurable absorption or scattering between the visible and the shortwave infrared; the proxy does not transfer between settings | Study 2 |
| TP | Particulate fraction only | Proxy inference through turbidity or total suspended solids | Dissolved phosphate is optically silent; the particulate association is setting-specific | Study 2 |
| COD, $BOD_5$ | Chromophoric and particulate fraction only | Proxy inference through CDOM or turbidity, or prediction from catchment attributes | The colourless, point-source-derived fraction leaves no spectral signature | Study 3 |

## 3.3. Optical principles, satellite sensors and the long-term archive

The retrieval of any water-quality variable from a satellite image is constrained by radiative transfer along the path shown in Figure 3.4: only a small part of the radiance reaching the sensor has been in contact with the water, the remainder being sunlight scattered within the atmosphere, skylight reflected at the surface, and the specularly reflected solar beam. The fundamental distinction is between inherent optical properties, which depend only on the medium and its constituents, and apparent optical properties, which depend additionally on the geometry of the light field (Kirk, 1994). Of the inherent properties, the backscattering coefficient matters most for passive remote sensing, because only backscattered photons return through the surface to a downward-looking sensor. The chief apparent property is the remote-sensing reflectance (Rrs), which, to first order, is proportional to the ratio of backscattering to the sum of absorption and backscattering. That single proportionality is the physical reason that strongly absorbing, weakly scattering substances such as CDOM darken the water while strongly backscattering particles brighten it.

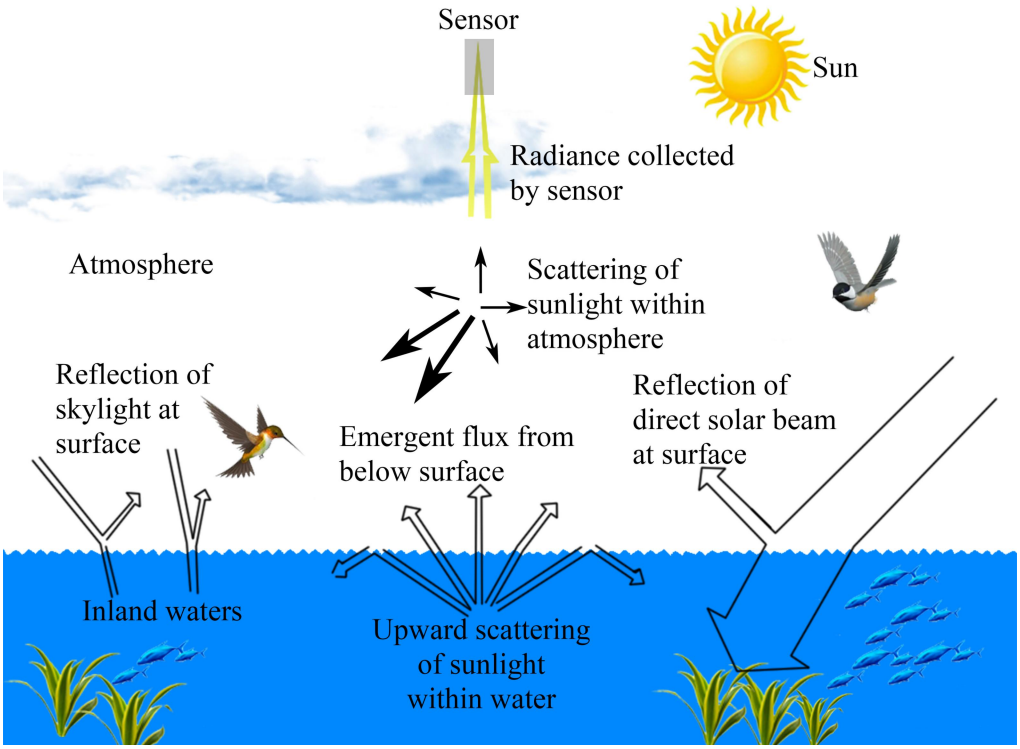



**Figure 3.4.** Spectral transfer between water constituents and a satellite sensor. Only the emergent flux from below the surface carries information about the water column; sunlight scattered within the atmosphere; skylight reflected at the surface and the reflected direct solar beam all contribute to the radiance collected by the sensor without having interacted with the constituents of the water.

Total absorption and backscattering are, to a good approximation, the additive sum of contributions from four constituent classes, and this additivity is the basis of all constituent retrieval. Pure water absorbs weakly in the blue and green and increasingly strongly towards the red and near-infrared, which is why the near-infrared signal of water is essentially nil. Phytoplankton absorb light through their pigments, chiefly Chl-a, with peaks near 440 and 675 nm, and the package effect makes that absorption non-linear with concentration. Non-algal particles absorb with a smooth exponential rise towards the blue and backscatter strongly, dominating the brightness of turbid inland waters, while CDOM absorbs with a near-exponential increase towards the ultraviolet

and blue but scatters negligibly, the steepness of that decrease being diagnostic of composition and source (Helms et al., 2008). Because CDOM and non-algal particle absorption share a similar blue-rising exponential shape, their separation from reflectance alone is intrinsically ill-posed, a recurrent obstacle to retrieving the coloured carbon fraction in turbid, eutrophic systems.

A second distinction concerns the optical classification of water bodies. In Case 1 waters, the optical properties co-vary with phytoplankton, so Chl-a largely parameterizes the optical state and simple band ratios succeed; inland and coastal waters are predominantly Case 2, in which sediment, CDOM and phytoplankton vary independently, a given reflectance spectrum can arise from many combinations of constituents, and algorithms calibrated for one water type fail in another (Mouw et al., 2015; Odermatt et al., 2012). The most consequential implication for this dissertation is the categorical separation of optically active from optically inactive constituents. Sediment, Chl-a and CDOM absorb, or backscatter appreciably within the range the sensor observes; dissolved inorganic nutrients do not, their few absorption features lying in the ultraviolet below the atmospheric cut-off and being far too weak at environmental concentrations to perturb the water-leaving signal, and the same holds for the colourless fraction of organic loading. Such constituents can be inferred only through statistical association with optically active constituents that happen to co-vary with them, an indirect and conditional pathway whose reliability must be established empirically rather than assumed.

The feasibility of long-term analysis rests on the archive summarized in Table 3.4. Landsat has provided 30 m imagery since 1984 through three successive instruments, creating a continuous long-term record uniquely suited to trend reconstruction (Wulder et al., 2019). Sentinel-2 has provided 10-20 m imagery since 2015, and the Harmonised Landsat and Sentinel-2 products place both families on a common grid with bandpass adjustment and bidirectional reflectance normalisation (Claverie et al., 2018). Coarser ocean-colour instruments are unsuited to small inland waters but remain valuable as temporally continuous references for cross-sensor harmonization. The hyperspectral

sensors resolve the spectral mixing that defeats multispectral instruments, yet their archives are short (Brando and Dekker, 2003), and airborne lidar adds structural detail at the cost of coverage (Malerba et al., 2023). Converting raw imagery into surface reflectance requires atmospheric correction (Roy et al., 2014a; Vermote et al., 2016), which is land-oriented and, over narrow channels, can leave residual aerosol effects and occasionally yield physically invalid negative reflectance, and the opening of these analysis-ready archives within cloud platforms has made planetary-scale decadal analysis tractable (Gorelick et al., 2017; Pekel et al., 2016).

**Table 3.4.** Satellite missions relevant to multi-decadal monitoring of inland-water quality.

| Mission and sensor | Operating period | Spatial resolution | Revisit | Relevance to inland-water monitoring |
|---|---|---|---|---|
| Landsat 5 TM | 1984–2013 | 30 m | 16 days | Origin of the long-term 30 m record and the only source for the 1980s and most of the 1990s |
| Landsat 7 ETM+ | 1999–2022 | 30 m | 16 days | Continuity through the 2000s; the 2003 scan-line-corrector failure leaves striped gaps |
| Landsat 8 and 9 OLI | 2013–present | 30 m | 16 days each, 8 days combined | Improved radiometric resolution and signal-to-noise ratio over dark water |
| Sentinel-2 MSI | 2015–present | 10, 20 and 60 m | 5 days with two satellites | Finer pixels partly resolve channels that Landsat cannot |
| Harmonized Landsat and Sentinel-2 | 2013–present | 30 m | 2–3 days | Bandpass-adjusted, BRDF-normalized surface reflectance from both families on a common grid |
| MODIS-Aqua | 2002–present | 250–1000 m | Daily | Too coarse for small water bodies, but a temporally continuous reference for cross-sensor harmonization |

Notes: TM, Thematic Mapper; ETM+, Enhanced Thematic Mapper Plus; OLI, Operational Land Imager; MSI, MultiSpectral Instrument; MODIS, Moderate Resolution Imaging Spectroradiometer; BRDF, bidirectional reflectance distribution function.

## 3.4. Retrieval methods and the evidence by variable

The literature on which these methods rest has grown rapidly and unevenly. A systematic search of the Web of Science for DOC, CDOM and their hydrophilic and hydrophobic fractions in inland waters, covering 1 January 1990 to 1 August 2023, returned 6,516 records. The co-occurrence of abstract terms and keywords in that corpus (Figure 3.5) shows a field organized around two intersecting groups of terms, one of water-body types and one of organic-carbon fractions, with remote sensing and ML forming a distinct and comparatively recent cluster. The distribution of the same keywords by year of publication shows the field expanding substantially since 2010, lakes dominating and peaking most recently, rivers attracting steadier but lower attention, and wetlands remaining markedly under-represented. The growth of the field is therefore real but unevenly distributed, and the smallest and most numerous systems are the least studied, which is precisely the gap the river studies address.

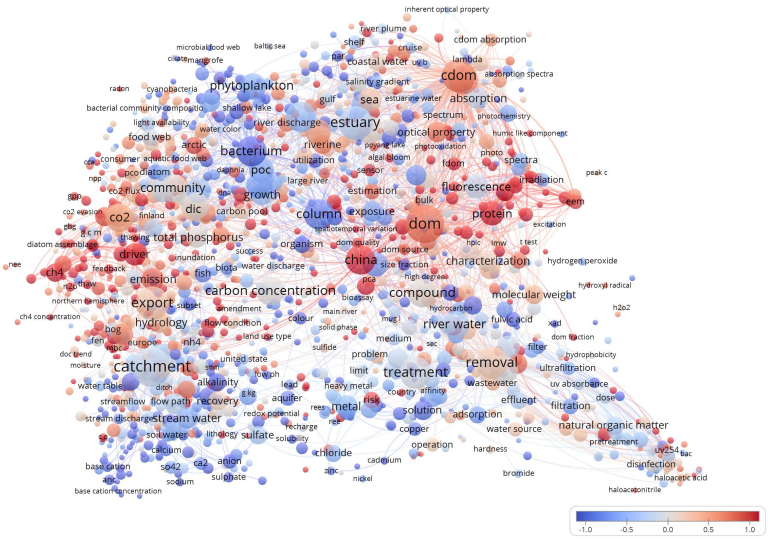



**Figure 3.5.** Co-occurrence of abstract terms and keywords in the literature on DOC in inland waters. Colour-coded clusters are term groupings determined by VOSviewer 1.6.19 from the relationships between items; the size of each circle is proportional to the frequency with which a term appears in abstracts and keywords, and the thickness of the connecting lines to the intensity of co-occurrence.

Algorithms fall into the three families compared in Table 3.5. Empirical algorithms relate concentration to reflectance or band ratios using statistical regression and are simple and transparent but are typically site- and season-specific (Matthews, 2011). Semi-analytical algorithms invert a physical model of the inherent optical properties and partition the result using assumed spectral shapes. The quasi-analytical algorithm with its CDOM refinements and the shallow-water bio-optical approach are in principle more transferable but require well-constrained optical models that are difficult to specify for inland waters (Li et al., 2018; Maritorena et al., 2002). ML algorithms learn flexible non-linear mapping without an explicit physical model, and approaches from regularized linear and kernel ridge regression through Gaussian processes to random forest (RF), support vector machine (SVM) and neural networks have all been applied to CDOM retrieval (Ruescas et al., 2018).

**Table 3.5.** Families of algorithms for retrieving water-quality variables from satellite reflectance.

| Family | Principle | Data requirement | Transferability | Principal limitation |
|---|---|---|---|---|
| Empirical | Statistical regression of concentration on reflectance or band ratios | Matchups from the target water body | Low; site- and season-specific | Degrades in optically complex waters and offers no physical basis for extrapolation |
| Semi-analytical | Inversion of a radiative-transfer approximation to absorption and backscattering, then partition by assumed spectral shapes | A specified optical model; few or no local matchups | Moderate to high in principle | Ill-posed wherever CDOM and non-algal particle absorption overlap, which is the norm in Case 2 waters |
| Machine learning | Flexible non-linear mapping learned from reflectance and derived features to concentration | Large matchup archives spanning the optical conditions of interest | Conditional; high inside the calibration domain, unverified outside it | Prone to overfitting and to optimistic skill estimates when validation ignores spatial structure |

ML has come to dominate inland-water retrieval because it accommodates the non-linear, multi-constituent character of optically complex waters and can exploit the large matchup archives now available (Cao et al., 2020; Ross et al., 2019). The problem

is supervised regression on tabular data, a structure that favours tree ensembles and kernel methods over convolutional architectures used in image classification. Random forests average predictions from an ensemble of decorrelated trees, reducing variance and improving predictive stability. Their behavior can be examined using post hoc interpretation methods. However, standard regression forests cannot predict beyond the range of response values observed during training. Gradient boosting builds its ensemble sequentially and often achieves high predictive accuracy on tabular spectral datasets. Careful tuning and regularization are nevertheless needed to limit overfitting. SVM remain competitive when training samples are limited. Neural networks do not consistently outperform well-tuned gradient boosting on the moderately sized, noisy satellite–in situ matchup datasets typical of water-quality remote sensing (Ge et al., 2024; Harkort and Duan, 2023; Kattenborn et al., 2021; D. Liu et al., 2021). Feature engineering often matters as much as the learner; the implementation and tuning used here are set out in Chapter IV.

The evidence available for each variable follows its optical accessibility. Because DOC is sensed only through its coloured fraction, every optical retrieval is a two-stage inference, from reflectance to CDOM absorption and from CDOM absorption to carbon (Fichot and Benner, 2011; Hestir et al., 2015). Band-ratio regression performs well in CDOM-dominated waters and has been successfully applied to northern and temperate lakes (Al-Kharusi et al., 2020; Kutser et al., 2005; Olmanson et al., 2016). However, these empirical relationships are often site- and season-specific. Semi-analytical inversion offers greater physical interpretability, but it can be ill-posed in optically complex inland waters. ML provides an alternative by capturing nonlinear relationships between spectral observations and DOC. Studies in large lakes and estuarine environments have demonstrated the potential of satellite observations for DOC estimation (Cao et al., 2018, 2020; Ge et al., 2024; Harkort and Duan, 2023; Kutser et al., 2015; Toming et al., 2020). Extending these approaches to multi-decadal trend analysis requires consistent processing and harmonization of the satellite archive, temporal validation, and uncertainty assessment. The deepest obstacle is nevertheless

biogeochemical: the coloured fraction per unit carbon shifts with source, and photochemical bleaching destroys color while leaving carbon in place, which places an upper bound on the accuracy attainable by any optical method and is aggravated in eutrophic, turbid waters by spectral interference (De Stefano et al., 2022; Duan et al., 2014; Griffin et al., 2018; Liu et al., 2020).

For dissolved nutrients, optical retrieval is less favorable because ammonium, nitrate, and dissolved phosphate lack diagnostic signals in satellite reflectance, making estimation inherently indirect. The chlorophyll-based pathway relies on the coupling between nutrient availability and algal biomass. This coupling can weaken because of light limitation, grazing, rapid flushing, or time lags in phytoplankton responses. Nutrient uptake into algal biomass can also decouple chlorophyll concentrations from concurrent dissolved nutrient concentrations. The turbidity-based pathway relies on the association of phosphorus with suspended mineral particles. It is generally more informative for particle-associated phosphorus than for dissolved phosphate or dissolved inorganic nitrogen (Ma et al., 2023; Xiong et al., 2022). These relationships are context-dependent, so models may fail when transferred to other settings. This limitation is compounded in narrow rivers, where optical proxies themselves can be poorly retrieved.

Optical remote sensing provides indirect information on organic pollution through associations with CDOM and suspended particulate matter. Retrieval reliability depends on the stability of these relationships. Point-source inputs may weaken them when changes in organic matter composition or oxygen demand are not reflected in the optical signal. Study 3 instead replaces explicit process-based catchment equations with ML mappings from observable catchment descriptors to in-stream concentrations, estimating organic pollution indicators from catchment characteristics rather than optical proxies (Vörösmarty et al., 2010; Whitehead et al., 2009).

## 3.5. Validation, long-term reconstruction and attribution

Several physical challenges bound what any inland-water retrieval can achieve, and the most severe arise from target size. A river a few metres wide cannot fill a 10-30 m pixel. The at-sensor signal thus mixes water with banks, riparian vegetation, and fields. A large share of the global river network lies below the width required for a clean water pixel, making this mixed-pixel problem pervasive rather than exceptional (Allen and Pavelsky, 2018). The adjacency effect compounds this problem. Photons reflected from bright land surrounding a dark water body are scattered into the sensor's view of the water. This contribution grows with the surrounding land fraction and remains especially difficult to correct when the channel is sub-pixel. Standard atmospheric correction alone may not adequately address this contamination (Bulgarelli and Zibordi, 2018; Kuhn et al., 2019; Vanhellemont and Ruddick, 2018). Object-based image analysis and multi-sensor harmonization are partial remedies only. Neither resolves whether the resulting data carry transferable information once the data's spatial structure is respected (Blaschke, 2010a; Maciel et al., 2023). Both limits are quantified directly in Studies 2 and 3.

Perhaps the most consequential methodological issue in this field is the design of model validation. ML studies conventionally report accuracy on random training-test splits, which assume that test samples are independent of training samples. River-monitoring networks, however, are clustered along channels, and water-quality fields are strongly spatially autocorrelated. Random CV can place autocorrelated neighbors in both training and test sets. This can overstate predictive skill at genuinely unmonitored locations (Karasiak et al., 2022; Ploton et al., 2020; Roberts et al., 2017). Spatially aware validation addresses this problem by separating training and test data at the site or spatial block level. Within each fold, all observations from a given site remain together in the same set. Leave-site-out and spatially blocked CV are advocated for evaluating predictions at unmonitored locations (Meyer and Pebesma, 2022; Milà et al., 2022a; Wadoux et al., 2021). This distinction can determine whether a variable appears retrievable at all. Throughout this dissertation, the gap in predictive

performance between random and spatial validation serves as a diagnostic of apparent skill associated with spatial leakage. The evaluation metrics adopted here, and their benchmarks, are set out in Section 4.3.6. Each reported metric must be interpreted in the context of the validation design that produces it.

Reconstructing multi-decadal records introduces further challenges, particularly because the three Landsat instruments differ in spectral response and rarely acquire synchronous observations. Artificial discontinuities may be mistaken for long-term environmental changes unless cross-sensor records are harmonized using bandpass adjustments or calibration against a temporally continuous external reference (Claverie et al., 2018; Zhao et al., 2024). Cloud cover, Landsat 7 scan-line-corrector failure and uneven acquisition create temporal gaps, while estimating carbon stocks requires time-varying water volumes that may also need reconstruction. Changes in the spatial composition of monitoring networks can also bias trend estimates unless explicitly accounted for. Attribution of reconstructed changes involves meteorological and hydrological forcing, eutrophication state, vegetation and human activity (Huang et al., 2017; Liu et al., 2020; Wang et al., 2024; Zhou et al., 2023). This remains challenging when spatially coherent changes reflect shared regional forcing, limiting the ability to distinguish local contributions from broader environmental influences using observational data.

Where optical retrieval is unreliable, the contributing catchment offers an alternative observational target, whose suitability depends on hydrological connectivity and transport processes. Small headwater catchments may have short flow paths and limited discharge for dilution, allowing nearby land-derived and effluent-derived inputs to influence in-stream concentrations, although attenuation depends on the constituent and flow conditions (Bishop et al., 2008). As catchment area increases, water sampled at a reach integrates inputs from a larger, more heterogeneous area over longer travel times. Locally observed catchment attributes may consequently explain less of the variability in concentration (Moatar et al., 2017). Some solutes exported from managed catchments exhibit chemostatic behavior, with concentrations remaining comparatively

stable across changes in discharge and potentially reflecting persistent catchment sources (Basu et al., 2010; Godsey et al., 2009; Musolff et al., 2015). These patterns motivate a testable hypothesis: concentrations in small channels that are difficult to resolve optically may be more predictable from catchment descriptors than those in larger rivers, provided that local source and transport relationships remain sufficiently consistent across sites. Study 3 evaluates this hypothesis using the same LOSO validation framework as in optical retrieval, enabling comparison of predictive performance under a common standard.

## 3.6. Study systems and knowledge gaps

The empirical work is grounded in two contrasting systems spanning the gradient from a large, well-resolved lake to a network of narrow rivers. Lake Taihu, in the lower Yangtze basin of eastern China, is the country's third-largest freshwater lake, with a surface area of about 2,347 $km^2$ and a mean depth of about 1.9 m. The lake is shallow, wind-exposed, and highly eutrophic, with frequent algal blooms and wind-driven sediment resuspension, creating dynamic and optically complex conditions. Its large extent relative to Landsat pixel size supports spatially resolved observations, while its long monitoring history provides in situ data for model calibration. Strong human pressures also make it an instructive case for long-term DOC reconstruction and driver attribution (Qin et al., 2010). Taihu is the setting of Study 1.

The South Bohemian River network in the southern Czech Republic provides the contrasting setting for Studies 2 and 3. The region is predominantly agricultural and forested, with drainage dominated by the Vltava, Lužnice and Otava systems and a historical landscape of fishponds. Monitored reaches are frequently only a few metres wide. The post-1989 transition involved reduced agricultural intensity and improved wastewater treatment, alongside long-term declines in nutrient concentrations and oxygen-demand indicators throughout the observation period. Most measured channels are no wider than a 30 m Landsat pixel. Of the 120 fixed monitoring stations, channel widths could be measured at 103 using national orthophotos with a spatial resolution

of 0.125 m; the remaining 17 were obscured by canopy. The measurable reaches have a median width of 9.2 m: 60 are narrower than 10 m, 30 are between 10 and 30 m, and 13 are wider than 30 m. Narrow reaches have a median width of approximately 5 m, which is below both the 30 m Landsat and 10 m Sentinel-2 Pixel sizes. These characteristics make the network a rigorous test bed for retrieval in narrow rivers, relevant to the widespread problem of unresolved channels.

The literature reviewed here establishes both the potential and the unresolved limitations of satellite remote sensing and ML for inland water-quality monitoring. A harmonized multi-sensor satellite archive, cloud-based processing, flexible ML models, and extensive in situ observations provide the foundation for long-term reconstruction. Three gaps remain. Firstly, robust long-term DOC products for large, optically complex eutrophic lakes remain scarce, leaving multi-decadal changes in DOC concentration and storage, as well as their drivers, insufficiently constrained. Secondly, dissolved nutrient retrieval in narrow rivers has rarely been assessed using spatially structured validation, leaving the extent of inflated performance under random CV poorly quantified. Thirdly, optical retrieval of oxygen-demand indicators and prediction from satellite-observable catchment attributes have rarely been evaluated using a common validation framework for the same target variables and sampling locations. Testing both approaches under matched validation conditions is necessary to determine whether catchment information provides transferable predictive skill when river-water optical signals are insufficient. Across these gaps, validation of spatially structured data remains inconsistent. This dissertation adopts spatial CV as a unifying evaluation principle, with the corresponding objectives and hypotheses presented in Chapter II.

# Chapter IV

## Materials and Methods

# Content

4.1. Study 1: Four decades of DOC dynamics in Lake Taihu ..... 42
4.1.1. Workflow of the Lake Taihu reconstruction ..... 43
4.1.2. Study area and field data ..... 45
4.1.3. Data preprocessing ..... 49
4.1.4. Feature engineering and selection ..... 55
4.1.5. Model construction and accuracy assessment ..... 61
4.1.6. Cross-sensor harmonization using MODIS-Aqua ..... 63
4.1.7. DOC storage quantification and WSC reconstruction ..... 66
4.1.8. Driver analysis: auxiliary datasets and attribution framework ..... 70
4.2. Study 2: Satellite retrieval and attribution of narrow-river nutrients in South Bohemia ..... 75
4.2.1. Workflow of the nutrient retrievability test ..... 75
4.2.2. Study area, monitoring network and width stratification ..... 76
4.2.3. In-situ nutrient records ..... 79
4.2.4. Object-level cross-sensor framework, quality gating and matchups ..... 81
4.2.5. Retrievability protocol: random versus spatial-block CV ..... 86
4.2.6. Model-free dependence and spatial-structure diagnostics ..... 88
4.2.7. Characterization of the long-term recovery ..... 89
4.2.8. Multi-scale riparian attribution under identical spatial rigour 89
4.3. Study 3: Modelling oxygen demand in narrow rivers from catchment characteristics ..... 90
4.3.1. Workflow and design principles of the oxygen-demand study ..... 91

4.3.2. Study area and width-stratified monitoring network ..... 95
4.3.3. In situ oxygen-demand concentrations ..... 97
4.3.4. Satellite archive and spatiotemporal matchups ..... 98
4.3.5. Two prediction pathways and their spatial validation ..... 101
4.3.6. Model evaluation ..... 106
4.3.7. Trends, reconstruction and 2050 scenario analysis ..... 108

# 4. Materials and Methods

This chapter describes the data sources, processing, and analytical methods of the three studies that form the core of this dissertation. Each study is presented in its own section, addressing in turn the reconstruction of DOC in a large eutrophic lake (Study 1), the retrieval and attribution of dissolved nutrients in narrow rivers (Study 2), and the catchment-driver prediction of organic loading in narrow rivers (Study 3).

## 4.1. Study 1: Four decades of DOC dynamics in Lake Taihu

Lakes occupy a small fraction of the land surface but play a disproportionately large role in carbon cycling (Regnier et al., 2022). Inland waters receive, transform, and store organic and inorganic carbon supplied by terrestrial ecosystems and river networks (Cole et al., 2007; Evans et al., 2017). A global modelling assessment of lakes larger than 0.1 $km^2$, including the Caspian Sea, estimated total DOC storage at 0.729 Pg C $yr^{-1}$ (Toming et al., 2020). Despite its biogeochemical importance, four-decade DOC dynamics in large inland lakes remain insufficiently constrained. Study 1 addresses this gap for Lake Taihu, China's third-largest freshwater lake, by combining Landsat observations from 1984 to 2023 with ML. The study reconstructs DOC concentration (DOCc) and estimates DOC storage (DOCs), examining their spatiotemporal variability and potential drivers. Its three objectives were: (1) to reconstruct four decades of DOCc using Landsat data and the best-performing ML model and estimate the corresponding DOCs; (2) to quantify spatiotemporal changes in DOCs; and (3) to identify potential drivers of DOCs variability. Together, these analyses provide a four-decade assessment of lake DOC concentration and storage, linking four-decade reconstruction with an evaluation of the environmental factors associated with their changes.

### 4.1.1. Workflow of the Lake Taihu reconstruction

Study 1 implemented an integrated workflow linking field observations, Landsat-based DOC retrieval, four-decade product generation, and interpretation. Six open-water zones were defined, and monthly DOCc data for 2007–2015 (n = 841) were compiled; other water-quality parameters (WQPs) were retained only for descriptive and diagnostic analyses. Field-data quality control (QC) using Isolation Forest (IF) outlier removal retained 818 observations, while missing values for transparency (TR) and water depth (WD) were imputed using multivariate imputation by chained equations (MICE) with predictive mean matching. Landsat–in situ matchups were constructed in GEE using a ±7-day window (Figure 4.1).

Aquatic Rrs was derived from atmospherically corrected Landsat surface reflectance, and quality assurance (QA)/QC was performed using the QA_PIXEL mask, which yielded 402 valid matchups. Interquartile range (IQR)-based filtering produced the modelling dataset of 200 samples. A feature pool comprising spectral bands, indices, and transformations was constructed, and Pearson correlation pre-filtering followed by RF importance retained 15 predictors. Four models were tuned using 10-fold CV on the training set (n = 140): multilayer back-propagation neural network (MBPNN), RF, SVM, and extreme gradient boosting (XGBoost). Performance was assessed on a held-out validation set (n = 60), and the best-performing model was selected for four-decade reconstruction.

The selected model was applied to the full Landsat archive (1984–2023) to reconstruct DOCc, with TM/ETM+/OLI estimates harmonized using MODIS-Aqua as an external reference. DOCs was estimated from harmonized DOCc and time-varying water storage capacity (WSC), for which the 1984–2006 record was hindcast using a fifth-generation ECMWF atmospheric reanalysis (ERA5)-driven XGBoost model. Finally, auxiliary driver datasets were harmonized and gap-filled, and their associations with DOCc and DOCs were assessed using structural equation modelling (SEM) and complementary diagnostics.

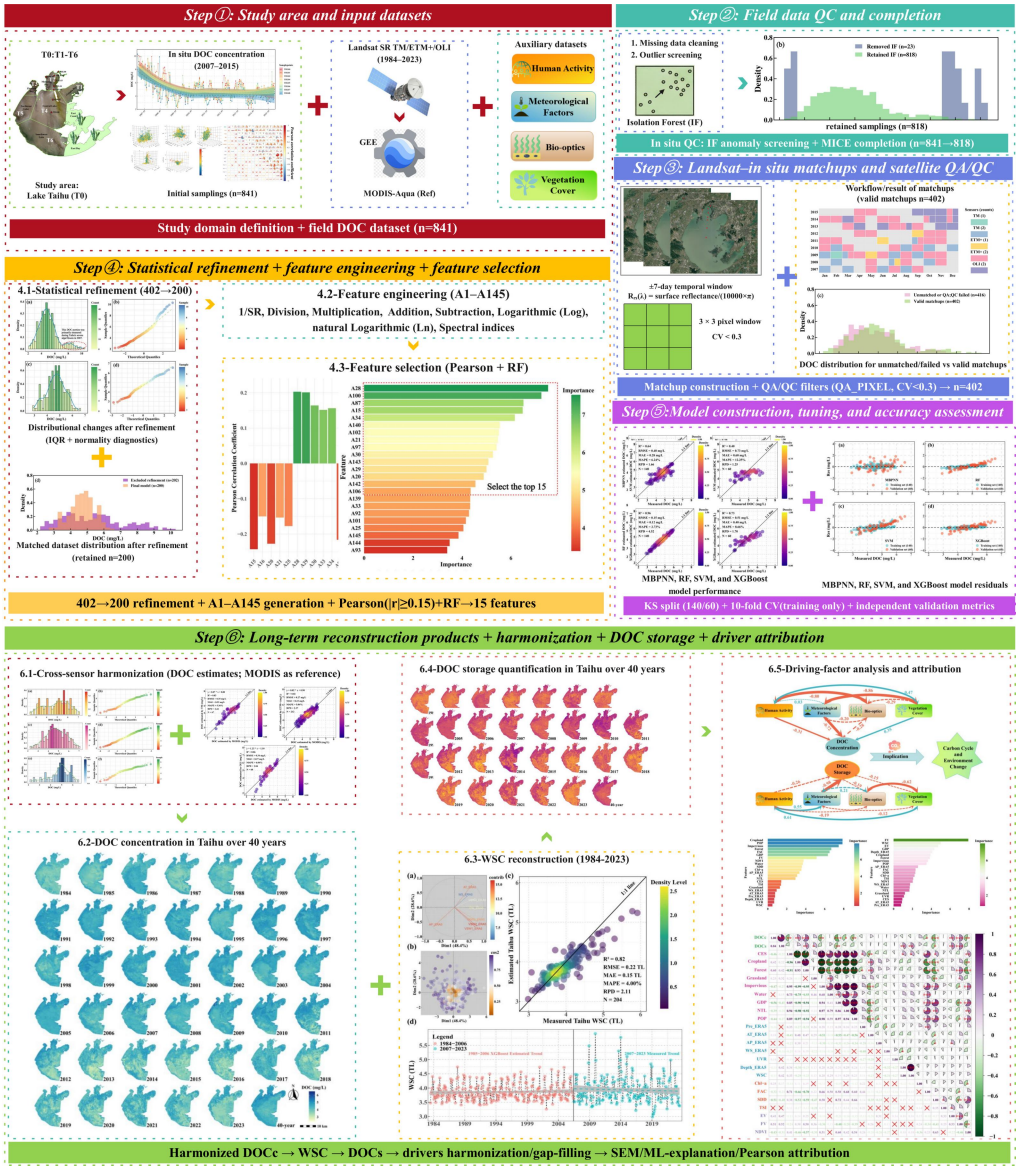



**Figure 4.1.** End-to-end methodological workflow for Landsat-based DOC retrieval, multi-sensor harmonization, four-decade reconstruction, DOC storage estimation, and driver attribution in Lake Taihu. The flowchart summarizes (i) study area and input datasets, (ii) field data quality control (QC) and completion, (iii) Landsat-in situ matchups and satellite quality assurance (QA)/QC, (iv) statistical refinement, feature engineering, and feature selection, (v) machine learning model construction tuning and

accuracy assessment, and (vi) generation of harmonized four-decade products (DOCc and DOCs) with post-retrieval driver analyses. Numbers in parentheses indicate the sample sizes retained after each filtering stage.

### 4.1.2. Study area and field data

Lake Taihu (approximately 30°56′–31°34′ N, 119°54′–120°36′ E) lies in the lower Yangtze River basin on the Jiangsu–Zhejiang border in eastern China. It is China's third-largest freshwater lake, with a surface area of 2,346.66 $km^2$ and a mean depth of 1.89 m (Wang et al., 2019). The lake is commonly divided into eight zones (Lai et al., 2024). Study 1 excluded East Bay (EB) and East Coast (EC) because dense aquatic vegetation alters water reflectance and can bias satellite-derived DOC retrieval (Qi et al., 2020). Analyses focused on six open-water zones, designated T1–T6: T1 (Gonghu Bay, GB), T2 (Meiliang Bay, MB), T3 (Zhushan Bay, ZB), T4 (Central Lake, CL), T5 (Northwest Lake, NL), and T6 (Southwest Lake, SL). T0 denotes the whole-lake domain used for lake-wide statistics (Figure 4.2).

A total of 841 field observations collected in Lake Taihu during 2007–2015 were compiled, including DOC and other WQPs. In Study 1, DOC was the only target variable used to calibrate and validate the Landsat-based retrieval models, and satellite-derived spectral bands, indices, and their transformations were the only predictors. The remaining in situ WQPs were not used as model predictors; instead, they supported descriptive statistics and correlation-based diagnostic interpretation, contextualising DOC variability and the discussion of reconstructed DOC patterns. The basic statistical characteristics of the WQPs over 2007–2015 are summarized in Table 4.1, and the temporal trends, multivariate interactions, and correlations of DOC with these WQPs are shown in Figure 4.3.

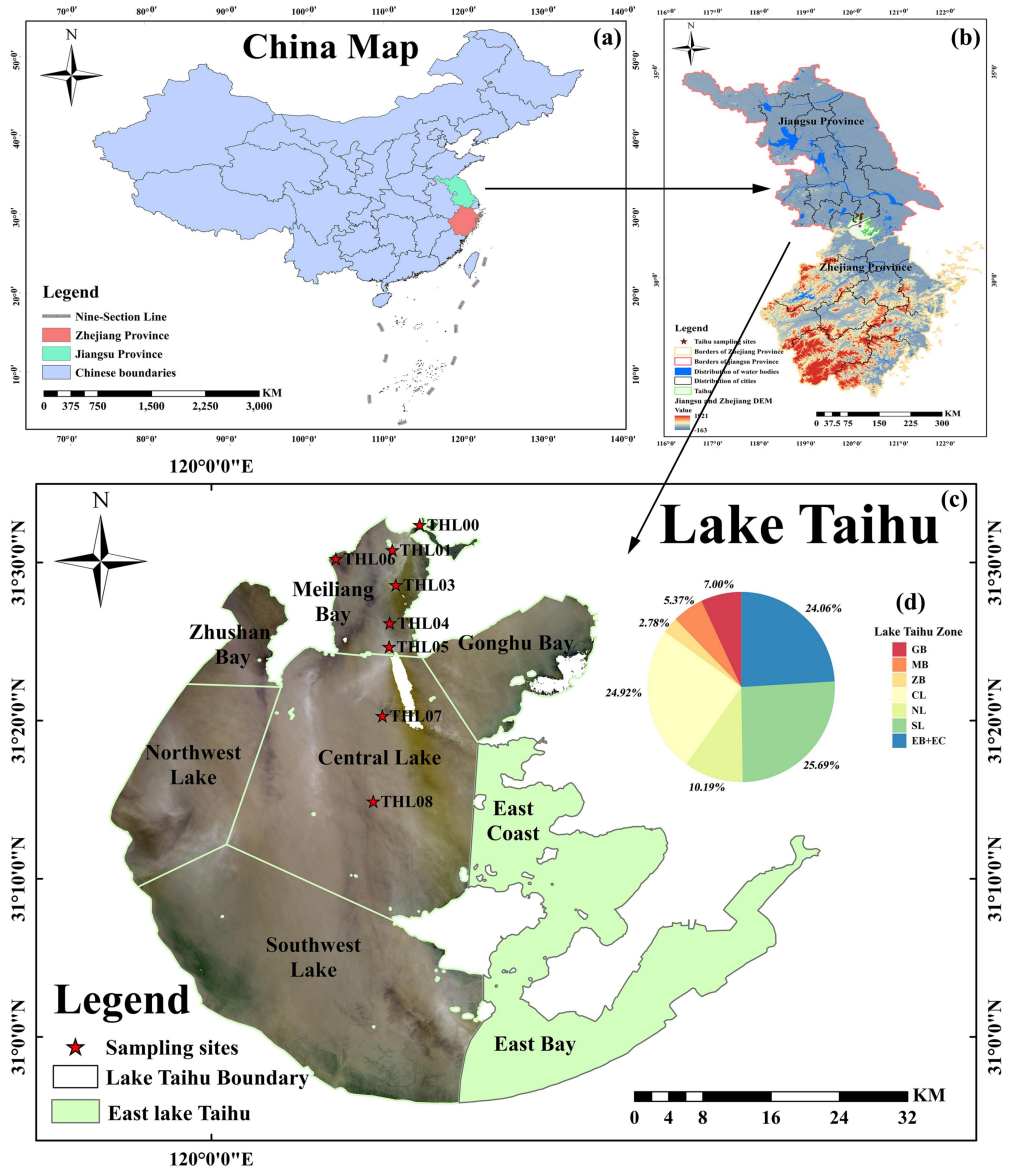



**Figure 4.2.** Study area, field sampling, and sub-lake zoning of Lake Taihu. (a) Location of Lake Taihu in China. (b) Location of Taihu in eastern China (Jiangsu-Zhejiang Provinces border). (c) In situ sampling sites for DOC and water-quality measurements (T1-T6); green areas indicate perennial aquatic-vegetation coverage. (d) Sub-lake zones and their areal proportions. For the sub-lake analysis, zones were labelled as T1-T6: T1 (Gonghu Bay, GB), T2 (Meiliang Bay, MB), T3 (Zhushan Bay, ZB), T4

(Central Lake, CL), T5 (Northwest Lake, NL), and T6 (Southwest Lake, SL). T0 represents the entire lake extent used for lake-wide statistics. East Bay (EB) and East Coast (EC) were excluded due to aquatic-vegetation effects on reflectance.

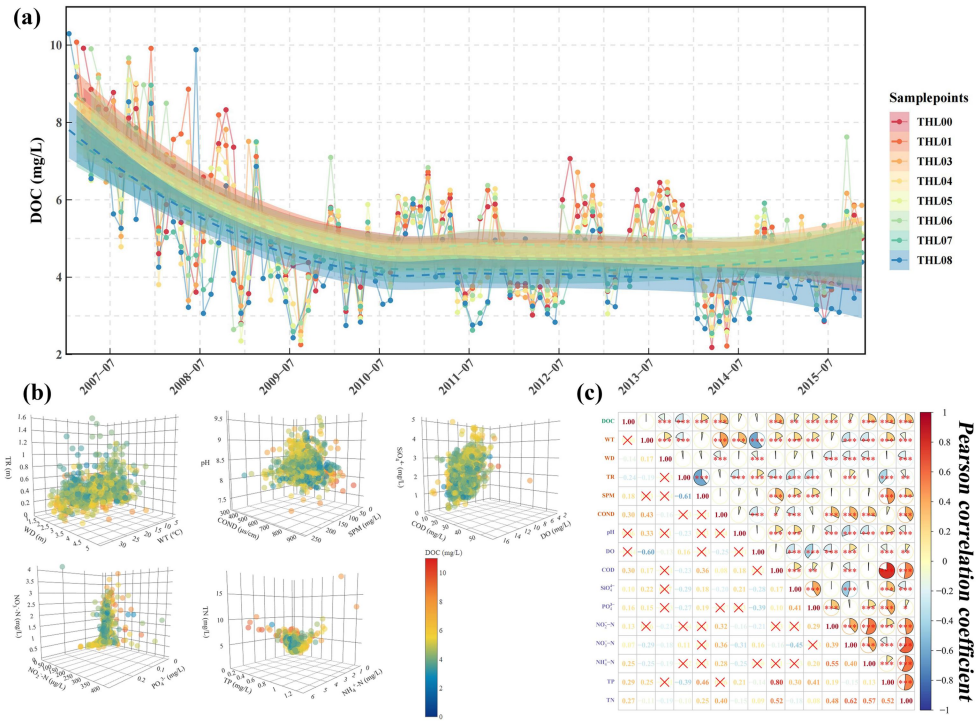



**Figure 4.3.** Temporal trends, multivariate interactions, and correlations of DOC with WQPs: (a) DOC trends across sampling points, (b) 3D bubble chart of DOC and key WQPs, (c) Pearson correlation matrix between DOC and key WQPs.

**Table 4.1.** Basic characteristics of WQPs in Taihu from 2007 to 2015.

| WQPs | Range | Mean | SD | Median | $Q_{25}$–$Q_{75}$ | N | Perid |
|---|---|---|---|---|---|---|---|
| DOC (mg/L) | 2.18–10.29 | 4.98 | 1.48 | 4.77 | 3.94–5.68 | 841 | 2007–2015 |
| WT (°C) | 1.15–33.46 | 17.58 | 8.92 | 19.14 | 9.02–25.01 | 841 | 2007–2015 |
| WD (m) | 1.00–5.00 | 2.51 | 0.52 | 2.50 | 2.20–2.80 | 841 | 2007–2015 |
| TR (m) | 0.00–1.60 | 0.38 | 0.22 | 0.35 | 0.23–0.50 | 841 | 2007–2015 |
| SPM (mg/L) | 2.68–246.93 | 45.81 | 34.83 | 36.30 | 22.00–57.40 | 841 | 2007–2015 |
| COND (μS/cm) | 295–915 | 495.72 | 108.40 | 480.00 | 410–565 | 841 | 2007–2015 |
| pH | 7.27–9.64 | 8.24 | 0.35 | 8.18 | 8.02–8.41 | 841 | 2007–2015 |
| DO (mg/L) | 2.3–16.29 | 9.19 | 2.13 | 9.14 | 7.62–10.75 | 841 | 2007–2015 |
| COD (mg/L) | 2.52–55.66 | 5.25 | 2.59 | 4.81 | 4.16–5.65 | 841 | 2007–2015 |
| $SiO_4^{4-}$ (mg/L) | 0.03–5.04 | 2.09 | 0.92 | 2.06 | 1.37–2.73 | 841 | 2007–2015 |
| $PO_4^{3-}$ (mg/L) | 0.00–0.29 | 0.02 | 0.02 | 0.01 | 0.00–0.02 | 841 | 2007–2015 |
| $NO_2^-$-N (μg/L) | 0.64–413.20 | 29.72 | 43.77 | 15.53 | 6.85–34.62 | 841 | 2007–2015 |
| $NO_3^-$-N (mg/L) | 0.04–4.06 | 0.80 | 0.74 | 0.63 | 0.17–1.20 | 841 | 2007–2015 |
| $NH_4^+$-N (mg/L) | 0.00–5.82 | 0.49 | 0.55 | 0.35 | 0.24–0.51 | 841 | 2007–2015 |
| TP (mg/L) | 0.02–1.25 | 0.13 | 0.10 | 0.10 | 0.07–0.15 | 841 | 2007–2015 |
| TN (mg/L) | 0.77–18.77 | 2.94 | 1.61 | 2.59 | 1.86–3.58 | 841 | 2007–2015 |

Notes: $Q_{25}$ and $Q_{75}$ are in the $1_{st}$ quartile and $3_{rd}$ quartile, respectively. This included the sample number of the WQPs and the study period.

### 4.1.3. Data preprocessing

The DOC dataset was preprocessed in three sequential steps to ensure data quality and to construct valid Landsat-in situ matchups for supervised model development. Firstly, the raw in situ dataset (n = 841, 2007–2015) was screened using IF anomaly detection (Liu et al., 2012, 2008). Missing symbols "/" were converted to "NA", and DOC was converted to numeric. The IF model was trained with 100 trees, a subsample size of 256, and an anomaly-score threshold of 0.65 to flag extreme anomalies, retaining 818 observations (n = 23 excluded). Missing values in TR and WD were filled using MICE with predictive mean matching, which does not affect DOC screening.

Secondly, the 818 retained observations were matched to Landsat scenes processed in GEE. A +/-7-day temporal window was paired with field sampling dates and satellite overpasses (Song et al., 2020). $R_{rs}(\lambda)$ (i.e., water-leaving reflectance used for water-quality applications) was derived from atmospherically corrected Landsat surface reflectance following Eq. S1 (i.e., $R_{rs}(\lambda)$=surface reflectance/(10000 × π)) and used as the aquatic reflectance input for DOC retrieval. (Z. Zhao et al., 2024b, 2024a). Pixels affected by cirrus, clouds, cloud shadows, and snow were removed using the QA_PIXEL bitmask, and local spatial homogeneity around each sampling location was assessed in a 3 x 3-pixel window (nine 30 m Landsat pixels, ~90 m x 90 m) using the coefficient of variation (CoV); only samples with CoV < 0.3 in all bands were retained (Xie et al., 2022; Yin et al., 2021). This procedure yielded 402 valid Landsat-in situ matchups, with an in-situ DOC range of 2.18–10.29 mg/L. Landsat bands from TM, ETM+, and OLI were harmonized as SR1–SR6 (corresponding to the BLUE, GREEN, RED, near-infrared (NIR), shortwave infrared 1 (SWIR1), and shortwave infrared 2 (SWIR2) bands), and the band–DOC relationships were examined. The 818-to-402 reduction primarily reflects matchup feasibility under the +/-7-day window and satellite QA/QC screening rather than outlier deletion, consistent with common practice in satellite-based water-quality retrieval (Chen et al., 2022; Harkort and Duan, 2023) (Figure 4.4).

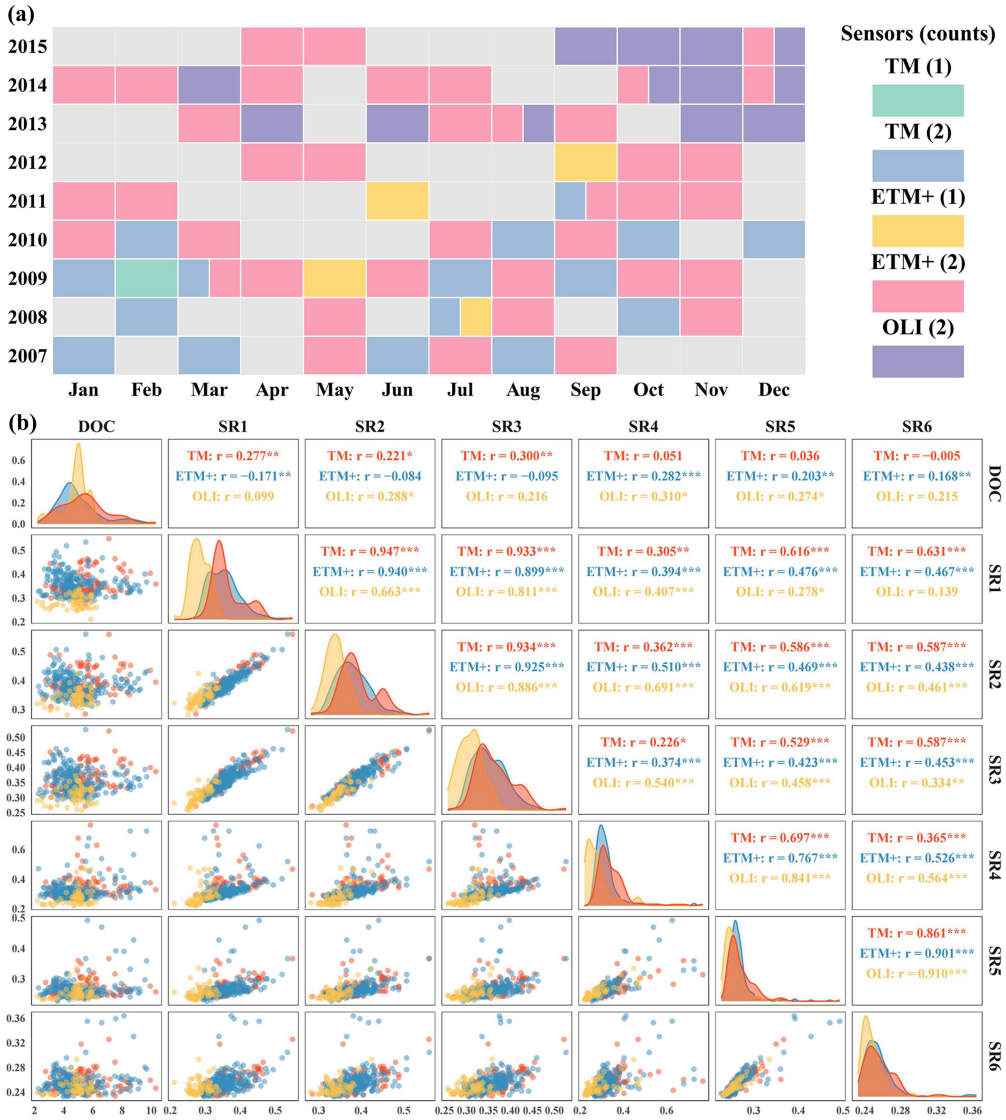



**Figure 4.4.** (a) Landsat sensor usage for DOC inversion in Taihu (2007–2015), with colored cells showing monthly availability and image counts from TM, ETM+, and OLI, (b) the scatterplot matrix analysis of DOC concentration and Landsat sensor reflectance bands (SR1–SR6) in Taihu. Note: The $R_{rs}(\lambda)$ of various bands from TM, ETM+, and OLI were uniformly designated as SR1–SR6, corresponding to BLUE, GREEN, RED, NIR, SWIR1, and SWIR2, respectively. In situ DOC range in the valid matchups (n = 402): 2.18–10.29 mg/L.

The sensor specifications of the Landsat missions used are summarized alongside MODIS in Table 4.2, and the monthly availability and per-sensor image counts are shown in Figure 4.4; the per-scene acquisition catalogues for TM, ETM+ and OLI are held in the supporting information of the corresponding publication. A consolidated summary of all satellite products, resolutions, QA/QC masks, and derived predictor families was also compiled (Table 4.3).

**Table 4.2.** Details of Landsat and MODIS.

| Landsat Sensors Type | GEE Collection | Band Details | Band Name | Centre Band (nm) | Resolution (m) |
|---|---|---|---|---|---|
| TM | LANDSAT/LT 05/C02/T1_L2 | Band 1 | BLUE | 485 | 30 |
| | | Band 2 | GREEN | 560 | 30 |
| | | Band 3 | RED | 660 | 30 |
| | | Band 4 | NIR | 835 | 30 |
| | | Band 5 | SWIR1 | 1650 | 30 |
| | | Band 7 | SWIR2 | 2223 | 30 |
| ETM+ | LANDSAT/LE 07/C02/T1_L2 | Band 1 | BLUE | 485 | 30 |
| | | Band 2 | GREEN | 560 | 30 |
| | | Band 3 | RED | 660 | 30 |
| | | Band 4 | NIR | 835 | 30 |
| | | Band 5 | SWIR1 | 1650 | 30 |
| | | Band 7 | SWIR2 | 2715 | 30 |
| OLI | LANDSAT/L C08/C02/T1_L 2 | Band 2 | BLUE | 482 | 30 |
| | | Band 3 | GREEN | 555 | 30 |
| | | Band 4 | RED | 655 | 30 |
| | | Band 5 | NIR | 865 | 30 |
| | | Band 6 | SWIR1 | 1609 | 30 |
| | | Band 7 | SWIR2 | 2201 | 30 |
| MODIS | MODIS/061/M YD09GA | Band 1 | RED | 645 | 500 |
| | | Band 2 | NIR | 858 | 500 |
| | | Band 3 | BLUE | 469 | 500 |
| | | Band 4 | GREEN | 555 | 500 |
| | | Band 5 | SWIR1 | 1240 | 500 |

| Landsat Sensors Type | GEE Collection | Band Details | Band Name | Centre Band (nm) | Resolution (m) |
|---|---|---|---|---|---|
| | | Band 6 | SWIR2 | 1640 | 500 |
| | | Band 7 | SWIR3 | 2130 | 500 |

Notes: In this study, the $R_{rs}(\lambda)$ of various bands from TM, ETM+, and OLI were uniformly designated as SR1–SR6, corresponding to BLUE, GREEN, RED, NIR, SWIR1, and SWIR2, respectively, to facilitate modelling and visualisation.

**Table 4.3.** Consolidated summary of all satellite products, resolutions, QA/QC masks, and derived predictor families.

| Satellite product | Landsat TM Collection2 Level-2 Surface Reflectance | Landsat ETM+ Collection2 Level-2 Surface Reflectance | Landsat OLI Collection2 Level-2 Surface Reflectance | MODIS Aqua Surface Reflectance |
|---|---|---|---|---|
| GEE collection ID | LANDSAT/LT05/ C02/T1_L2 | LANDSAT/LE07/ C02/T1_L2 | LANDSAT/LC08/ C02/T1_L2 | MODIS/061/MYD 09GA |
| Spatial resolution | 30 m | 30 m | 30 m | 500 m |
| Temporal sampling | 16-day nominal | 16-day nominal | 16-day nominal | Daily |
| Core variables | Surface reflectance bands; QA_PIXEL | Surface reflectance bands; QA_PIXEL; SLC-off | Surface reflectance bands; QA_PIXEL | Surface reflectance bands |
| QA/Mask used | QA_PIXEL; 3×3 CoV<0.3 | QA_PIXEL; 3×3 CoV<0.3; SLC-off handling | QA_PIXEL; 3×3 CoV<0.3 | Product QA in collection |
| Derived predictors generated | SR1–SR6; indices; transformations | SR1–SR6; indices; transformations | SR1–SR6; indices; transformations | Band reference and optional cross-sensor context |

Note: CoV denotes the coefficient of variation calculated within a 3 × 3-pixel neighbourhood.

Thirdly, statistical screening was applied to the matched dataset (n = 402). Outliers were identified using the IQR criterion (values < Q1 − 1.5 × IQR or > Q3 + 1.5 × IQR) (Aggarwal et al., 2019; Lotteraner et al., 2026; Uddin et al., 2024). The Shapiro–Wilk and two-sided Kolmogorov–Smirnov tests were used as distributional diagnostics at α=0.05, and screening retained 200 samples for model development. Figure 4.5 compares density-normalized DOC distributions across screening stages and between excluded and retained subsets. Figure 4.6 presents the normality assessments before and after outlier removal.

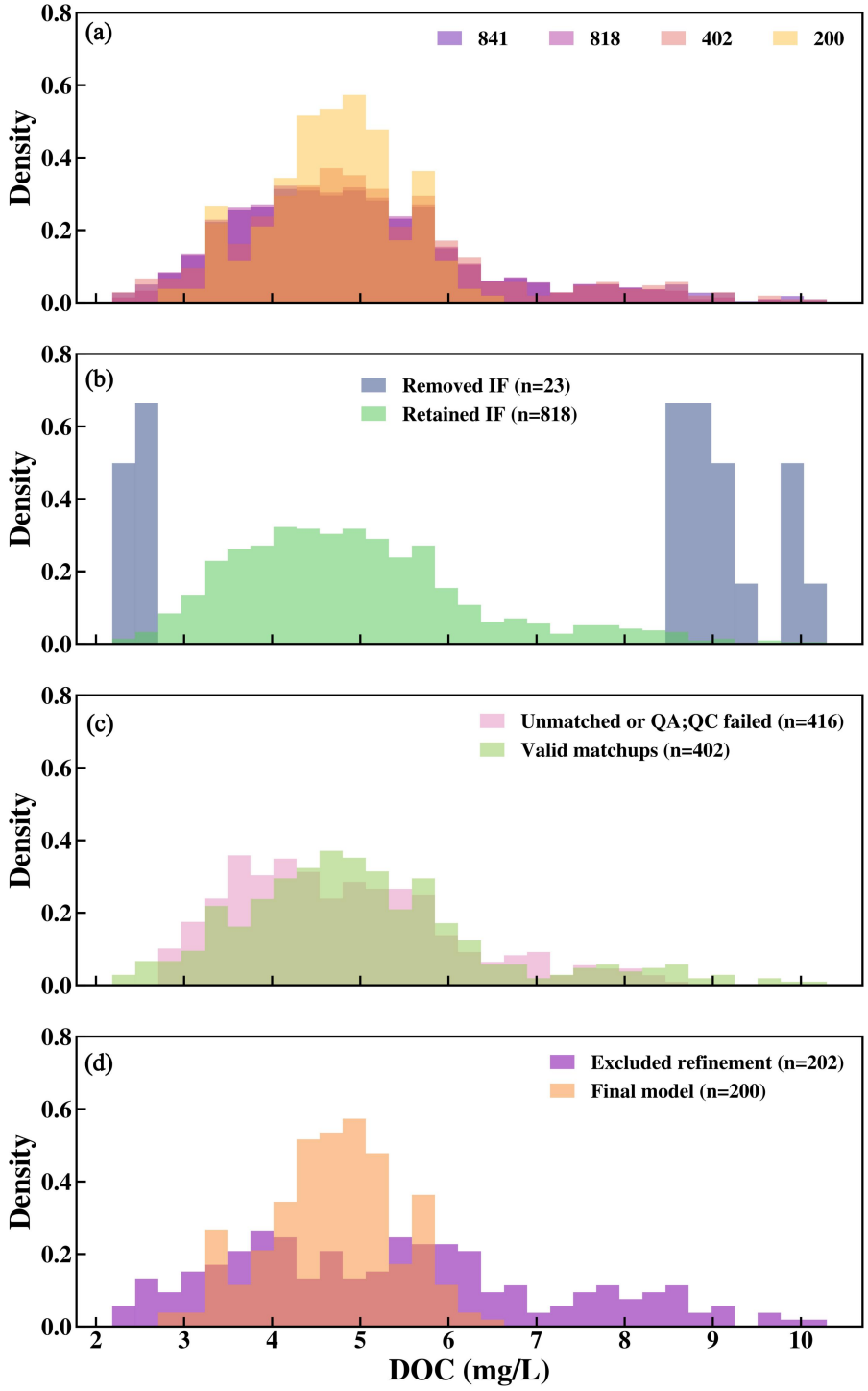

(a)
841
818
402
200
Density
(b)
Removed IF (n=23)
Retained IF (n=818)
Density
(c)
Unmatched or QA;QC failed (n=416)
Valid matchups (n=402)
Density
(d)
Excluded refinement (n=202)
Final model (n=200)
Density
0.0
0.2
0.4
0.6
0.8
2
3
4
5
6
7
8
9
10
DOC (mg/L)

**Figure 4.5.** Density-normalized dissolved organic carbon (DOC) distributions across screening steps and excluded-versus-retained subsets. (a) Stage-wise DOC distributions (n = 841/818/402/200); (b) Isolation Forest (IF) screening: removed vs retained (n = 23 vs 818); (c) Quality assurance/quality control (QA/QC) filtering: unmatched/failed vs valid matchups (n = 416 vs 402); (d) Interquartile range (IQR) refinement: excluded vs final model set (n = 202 vs 200).

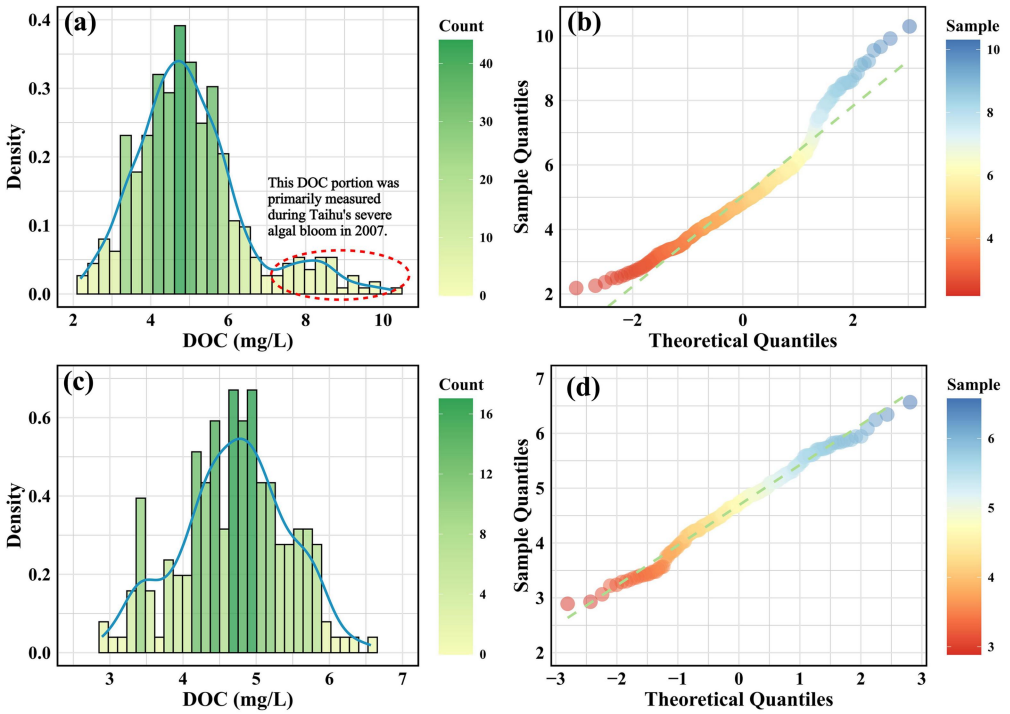



**Figure 4.6.** Statistical analysis of DOC concentrations in Taihu: distribution and normality assessment. (a) Histogram with density curve (2007–2015 data, 402 samples). Elevated DOC values (red dashed circle) are linked to algal blooms in 2007, (b) Q-Q plot for normality check (402 samples), (c) Histogram with density curve after outlier removal (IQR method, 2007–2015, 200 samples), (d) Q-Q plot for normality check after outlier removal (200 samples).

### 4.1.4. Feature engineering and selection

In optically complex eutrophic lakes such as Taihu, DOC-related signals mediated by blue-band CDOM absorption can be confounded by phytoplankton pigments and suspended particles, potentially limiting the robustness of band-only predictors (SR1–SR6). Band combinations, transformations and indices were added to represent spectral contrasts and improve robustness across optical conditions (Brezonik et al., 2015; Harkort and Duan, 2023; D. Liu et al., 2021). The index set included the normalized difference water index (NDWI), modified NDWI (MNDWI), normalized difference moisture index (NDMI), and floating algae index (FAI; Hu, 2009). The automated water extraction index (AWEI) variants for non-shadow (AWEI0) and shadow (AWEI1) conditions (Feyisa et al., 2014), together with the 2015 water index (WI2015; Fisher et al., 2016), were also included. These indices characterised water-related spectral contrasts and surface vegetation signals, with their relevance to DOC prediction evaluated through subsequent feature selection. Additional SR1–SR6 features comprised reciprocals, ratios, products, sums, differences and logarithmic transformations, including natural logarithms (Table 4.4); water indices derived for TM, ETM+ and OLI are listed in Table 4.5.

**Table 4.4.** Band transformations and spectral indices for DOC modelling.

| Original band | SR1 | SR2 | SR3 | SR4 | SR5 | SR6 |
|---|---|---|---|---|---|---|
| 1/SR | 1/SR1 | 1/SR2 | 1/SR3 | 1/SR4 | 1/SR5 | 1/SR6 |
| Division | SR1/SR2 | SR1/SR3 | SR1/SR4 | SR1/SR5 | SR1/SR6 | SR2/SR1 |
| | SR2/SR3 | SR2/SR4 | SR2/SR5 | SR2/SR6 | SR3/SR1 | SR3/SR2 |
| | SR3/SR4 | SR3/SR5 | SR3/SR6 | SR4/SR1 | SR4/SR2 | SR4/SR3 |
| | SR4/SR5 | SR4/SR6 | SR5/SR1 | SR5/SR2 | SR5/SR3 | SR5/SR4 |
| | SR5/SR6 | SR6/SR1 | SR6/SR2 | SR6/SR3 | SR6/SR4 | SR6/SR5 |
| Multiplication | SR1*SR1 | SR1*SR2 | SR1*SR3 | SR1*SR4 | SR1*SR5 | SR1*SR6 |
| | SR2*SR2 | SR2*SR3 | SR2*SR4 | SR2*SR5 | SR2*SR6 | SR3*SR3 |
| | SR3*SR4 | SR3*SR5 | SR3*SR6 | SR4*SR4 | SR4*SR5 | SR4*SR6 |
| | SR5*SR5 | SR5*SR6 | SR6*SR6 | | | |
| Addition | SR1+SR1 | SR1+SR2 | SR1+SR3 | SR1+SR4 | SR1+SR5 | SR1+SR6 |
| | SR2+SR2 | SR2+SR3 | SR2+SR4 | SR2+SR5 | SR2+SR6 | SR3+SR3 |
| | SR3+SR4 | SR3+SR5 | SR3+SR6 | SR4+SR4 | SR4+SR5 | SR4+SR6 |
| | SR5+SR5 | SR5+SR6 | SR6+SR6 | | | |
| Subtraction | SR1–SR2 | SR1–SR3 | SR1–SR4 | SR1–SR5 | SR1–SR6 | SR2–SR1 |
| | SR2–SR3 | SR2–SR4 | SR2–SR5 | SR2–SR6 | SR3–SR1 | SR3–SR2 |
| | SR3–SR4 | SR3–SR5 | SR3–SR6 | SR4–SR1 | SR4–SR2 | SR4–SR3 |
| | SR4–SR5 | SR4–SR6 | SR5–SR1 | SR5–SR2 | SR5–SR3 | SR5–SR4 |
| | SR5–SR6 | SR6–SR1 | SR6–SR2 | SR6–SR3 | SR6–SR4 | SR6–SR5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Logarithmic (Log) | Log(SR1) | Log(SR2) | Log(SR3) | Log(SR4) | Log(SR5) | Log(SR6) |
| | Log(1+SR1) | Log(1+SR2) | Log(1+SR3) | Log(1+SR4) | Log(1+SR5) | Log(1+SR6) |
| natural Logarithmic (Ln) | Ln(SR1) | Ln(SR2) | Ln(SR3) | Ln(SR4) | Ln(SR5) | Ln(SR6) |
| | Ln(1+SR1) | Ln(1+SR2) | Ln(1+SR3) | Ln(1+SR4) | Ln(1+SR5) | Ln(1+SR6) |
| Spectral indices | NDWI | MNDWI | NDMI | FAI | AWEI0 | AWEI1 |
| | $WI_{2015}$ | | | | | |

Note: When imported, some data column names contain special symbols. To simplify modelling, SR1–$WI_{2015}$ were sequentially renamed as A1–A145 during the process.

**Table 4.5.** Key water indicators for DOC estimation using TM, ETM+, and OLI reflectance.

| Water index | Equation | Reference |
|---|---|---|
| NDWI | $(\rho_{GREEN} - \rho_{NIR}) / (\rho_{GREEN} + \rho_{NIR})$ | (Gao, 1996) |
| MNDWI | $(\rho_{GREEN} - \rho_{SWIR\ 1}) / (\rho_{GREEN} + \rho_{SWIR\ 1})$ | (Duan et al., 2009) |
| NDMI | $(\rho_{NIR} - \rho_{SWIR\ 1}) / (\rho_{NIR} + \rho_{SWIR\ 1})$ | (Badapalli et al., 2024) |
| FAI | $FAI = \rho_{NIR} - [\rho_{RED} + (\rho_{SWIR\ 1} - \rho_{RED}) * (\frac{\lambda_{NIR} - \lambda_{RED}}{\lambda_{SWIR\ 1} - \lambda_{RED}})]$ | (Hu, 2009) |
| AWEI0 | $4 * (\rho_{GREEN} - \rho_{NIR}) - 0.25 * \rho_{NIR} - 2.75 * \rho_{SWIR\ 2}$ | (Feyisa et al., 2014) |
| AWEI1 | $\rho_{BLUE} + 2.5 * \rho_{GREEN} - 1.5 * (\rho_{NIR} + \rho_{SWIR\ 1}) - 0.25 * \rho_{SWIR\ 2}$ | (Feyisa et al., 2014) |
| $WI_{2015}$ | $1.7204 + 171 * \rho_{GREEN} + 3 * \rho_{RED} - 70 * \rho_{NIR} - 45 * \rho_{SWIR\ 1} - 71 * \rho_{SWIR\ 2}$ | (Fisher et al., 2016) |

Notes: Notably, the band range of OLI differs from that of TM and ETM+ when calculating FAI, requiring separate calculations.

A total of 145 features were generated from the original bands and derived indices and labelled A1–A145. Given the small sample size (n = 200), feature screening was applied to reduce dimensionality and the risk of overfitting. Firstly, constant-valued columns were excluded, and Pearson correlations between the remaining features and DOCc were calculated. A permissive threshold ($|r| \geq 0.15$) was used for preliminary screening, recognizing that informative predictors may exhibit only modest linear associations in optically complex waters (Brezonik et al., 2015). Secondly, an RF model comprising 100 decision trees ranked feature importance, and the 15 highest-ranked features were retained. These were dominated by NIR–visible and SWIR–visible contrasts, including ratios and differences involving SR4 or SR5 versus SR1–SR3, as well as MNDWI, FAI, and AWEI0. These features describe spectral variability associated with algal blooms, turbidity, and mixed-water pixels (D. Liu et al., 2021). The screening outcome is shown in Figure 4.7, and the resulting dataset comprised the 15 selected features, listed in Table 4.6, and DOC as the response variable.

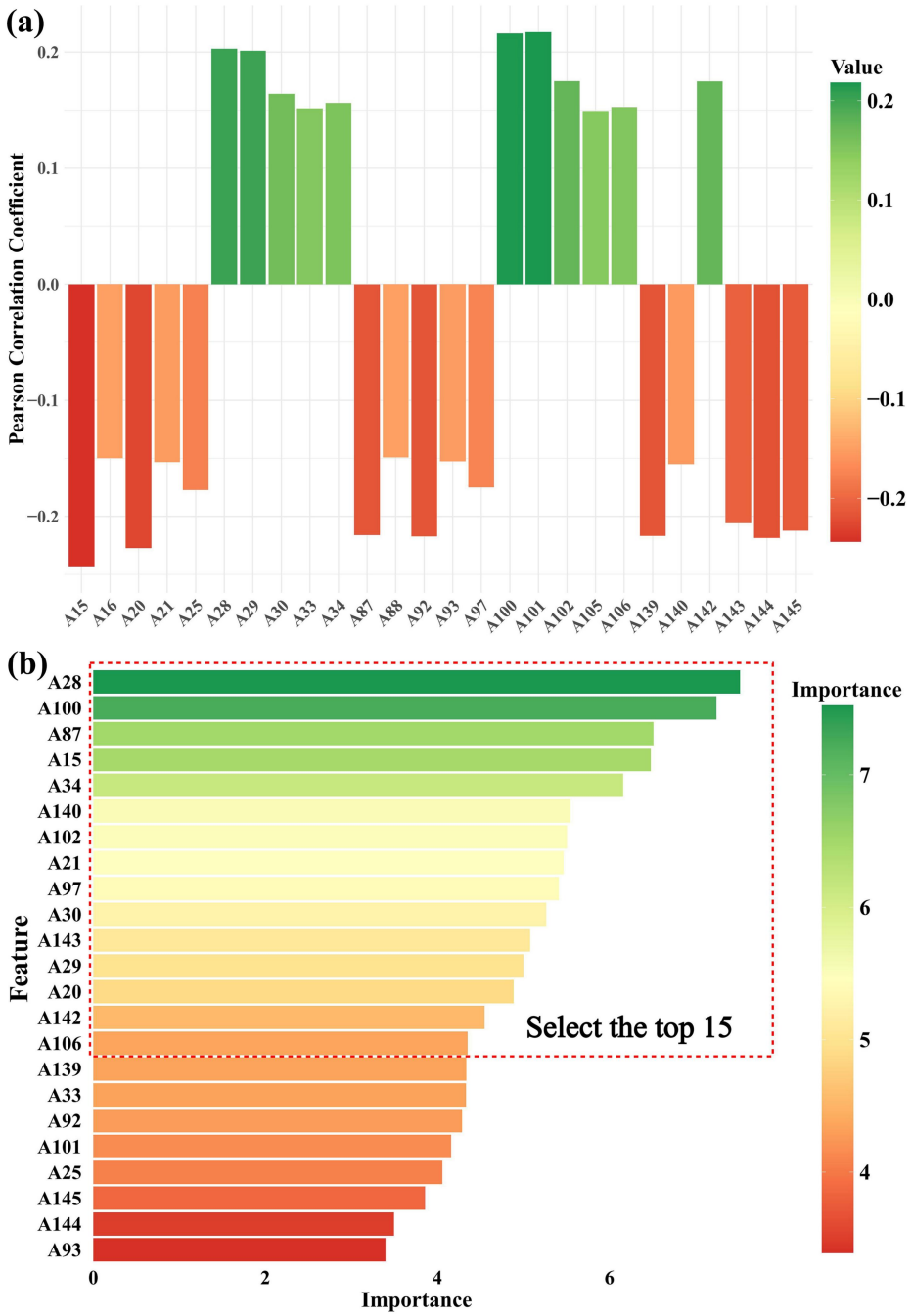



**Figure 4.7.** (a) Pearson correlation analysis identified features with significant linear relationships ($|r| \geq 0.15$) for dimensionality reduction, (b) The top 15 feature variables were selected based on their importance to enhance estimative power.

**Table 4.6.** Selected 15 modelling features via Pearson correlation and RF screening.

| Band abbreviation | Band combination | r | Importance |
|---|---|---|---|
| A15 | SR1/SR4 | -0.24 | 6.48 |
| A20 | SR2/SR4 | -0.23 | 4.89 |
| A21 | SR2/SR5 | -0.15 | 5.47 |
| A28 | SR4/SR1 | 0.20 | 7.52 |
| A29 | SR4/SR2 | 0.20 | 5.00 |
| A30 | SR4/SR3 | 0.16 | 5.26 |
| A34 | SR5/SR2 | 0.16 | 6.16 |
| A87 | SR1–SR4 | -0.22 | 6.51 |
| A97 | SR3–SR4 | -0.17 | 5.41 |
| A100 | SR4–SR1 | 0.22 | 7.24 |
| A102 | SR4–SR3 | 0.17 | 5.51 |
| A106 | SR5–SR2 | 0.15 | 4.35 |
| A140 | MNDWI | -0.16 | 5.55 |
| A142 | FAI | 0.17 | 4.55 |
| A143 | AWEI0 | -0.21 | 5.08 |

### 4.1.5. Model construction and accuracy assessment

The modelling dataset comprised 200 Landsat–in situ matchups with 15 input features. The Kennard–Stone algorithm divided the data into a training set (70%, n = 140) and a held-out validation set (30%, n = 60), promoting coverage of the feature space in the training subset (Shen et al., 2025c; Yang et al., 2021). Validation samples were excluded from predictive model fitting and hyperparameter tuning, and the same fixed set of 15 features was used in both subsets. Four ML models (MBPNN, RF, SVM, and XGBoost) were implemented in R and tuned using 10-fold CV within the training set. The tuned models were refitted on the full training set and compared on the held-out validation set, with the best-performing model selected for four-decade reconstruction. MBPNN used a regularized multilayer feed-forward network, while RF was implemented with the ranger package. SVM used a radial basis function (RBF)

kernel, with centering and scaling estimated within each training fold; XGBoost was implemented using xgboost and caret with numeric matrix inputs (Table 4.7).

Model performance was assessed using the coefficient of determination ($R^2$), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), standard deviation (SD), residual prediction deviation (RPD), and residuals (Res). Performance reported in the results was evaluated on the independent validation set (n = 60) unless explicitly stated otherwise. All preprocessing, training, validation, evaluation, and visualization were performed in R version 4.3.3.

**Table 4.7.** Optimal hyperparameters for machine learning algorithms.

| Algorithm | Best parameter combination | Value |
|---|---|---|
| MBPNN | Neurons | 20 |
| | Hidden layers | 2 |
| | Regularization strength | 0.001 |
| RF | Number of variables per split | 2 |
| | Split rule | Variance |
| | Minimum node size | 1 |
| | Number of trees | 1000 |
| | Maximum tree depth | 20 |
| SVM | Kernel | Radial basis function |
| | C | 3 |
| | epsilon | 0.1 |
| | Minimum loss reduction (Gamma) | 0.1 |
| XGBoost | Number of rounds | 100 |
| | Maximum tree depth | 6 |
| | Learning rate (eta) | 0.1 |
| | Gamma | 1 |
| | Column subsampling ratio | 0.7 |
| | Minimum child weight | 1 |
| | Subsample ratio | 0.9 |

### 4.1.6. Cross-sensor harmonization using MODIS-Aqua

Four-decade DOC reconstruction requires a temporally consistent Landsat record spanning TM, ETM+, and OLI. Limited synchronous acquisitions constrain direct inter-sensor comparisons (Shen et al., 2025a; Z. Zhao et al., 2024b). Study 1 harmonized DOC estimates rather than recalibrating surface reflectance, using daily MODIS-Aqua surface reflectance (500 m) from GEE as a temporally dense external reference to increase the availability of near-coincident observations. Eight stable calibration locations were selected within Taihu, and near-coincident Landsat and MODIS observations were extracted using water masking and QC criteria consistent with model development. For each location and date, DOC estimates averaged over 34 × 34-pixel Landsat windows (30 m resolution) and corresponding 2 × 2-pixel MODIS windows (500 m resolution) to reduce resolution mismatch (Z. Zhao et al., 2024a). Sensor-specific linear relationships with MODIS-based DOC estimates were derived to place ETM+ and OLI estimates on a TM-consistent DOC scale (Shen et al., 2025a). TM was selected as the reference sensor because it provided the best balance between agreement with MODIS and temporal coverage. Figure 4.8 presents sensor-specific DOC distributions and normality assessments, Figure 4.9 compares near-coincident Landsat- and MODIS-based DOC estimates, and Table 4.8 reports agreement statistics after harmonization.

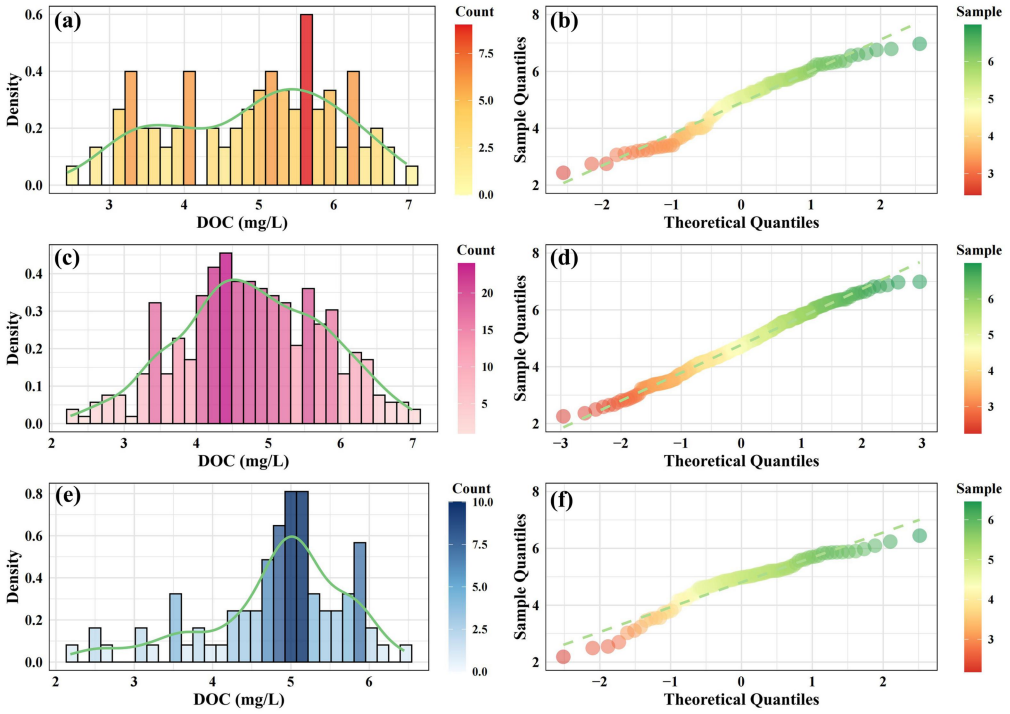



**Figure 4.8.** Distribution and normality assessment by sensor. (a,c,e) Histograms with density curve: TM (96), ETM+ (323), OLI (84); (b, d, f) Q-Q plots (TM, ETM+, OLI).

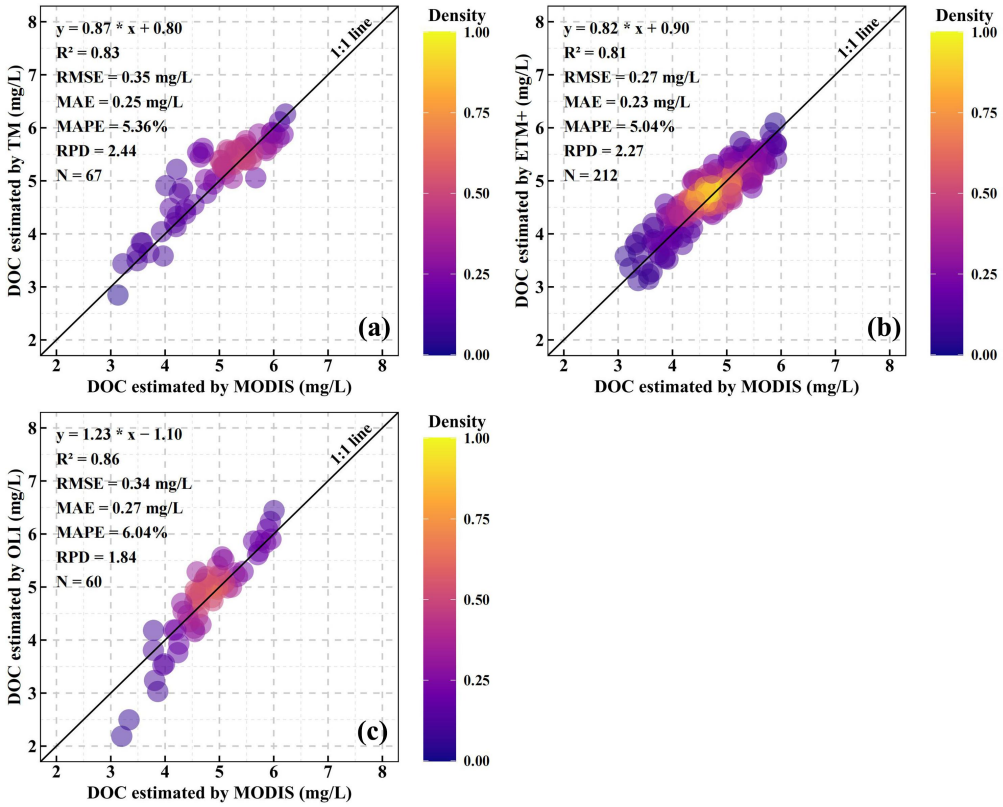



**Figure 4.9.** Comparison of DOC estimated from (a) TM, (b) ETM+, and (c) OLI with corresponding synchronous MODIS data.

**Table 4.8.** DOC estimation accuracy using XGBoost with Landsat and MODIS.

| Sensor | $R^2$ | RMSE (mg/L) | MAE (mg/L) | MAPE (%) | SD | RPD | Linear Formula |
|---|---|---|---|---|---|---|---|
| TM | 0.83 | 0.35 | 0.25 | 5.36 | 0.33 | 2.44 | y = 0.87 * x + 0.80 |
| ETM+ | 0.81 | 0.27 | 0.23 | 5.04 | 0.27 | 2.27 | y = 0.82 * x + 0.90 |
| OLI | 0.86 | 0.34 | 0.27 | 6.04 | 0.35 | 1.84 | y = 1.23 * x – 1.10 |

### 4.1.7. DOC storage quantification and WSC reconstruction

DOCc maps at 30 m spatial resolution were reconstructed from Landsat TM/ETM+/OLI observations for 1984–2023. Because satellite reflectance provides information only about the near-surface optically active layer, these maps represent near-surface concentrations rather than vertical profiles. The permanent-water extent was delineated using HydroLAKES (Messager et al., 2016), Global Surface Water Occurrence (Pekel et al., 2016), and four-decade Landsat imagery. Estimating DOCs also required time-varying WSC to account for substantial water-volume fluctuations associated with Taihu's shallow morphology and hydrological regulation (Qin et al., 2010). Monthly WSC for 1984–2006 was reconstructed using an ERA5-driven XGBoost model (Hersbach et al., 2020). Principal component analysis (PCA) summarized the dominant patterns of variation among ERA5 predictors: Dim1 was primarily associated with air temperature (AT) and multi-layer soil water content, whereas Dim2 was more closely associated with wind speed (WS). Observation-level $\cos^2$ values indicated the extent to which individual observations were represented in PCA space (Figure 4.10). Comparison with observed WSC for 2007–2023 yielded $R^2$ = 0.82, RMSE = 0.22 TL, MAE = 0.15 TL, MAPE = 4.00%, and RPD = 2.11 (N = 204), with data points clustered around the 1:1 line, particularly at WSC values of 3.5–4.5 TL. The monthly WSC hindcasts for 1984–2006 were combined with observations for 2007–2023 to produce a continuous series covering 1984–2023. Temporal variations in climatic, hydrological, and precipitation variables are shown in Figure 4.11, with summary statistics presented in Table 4.9.

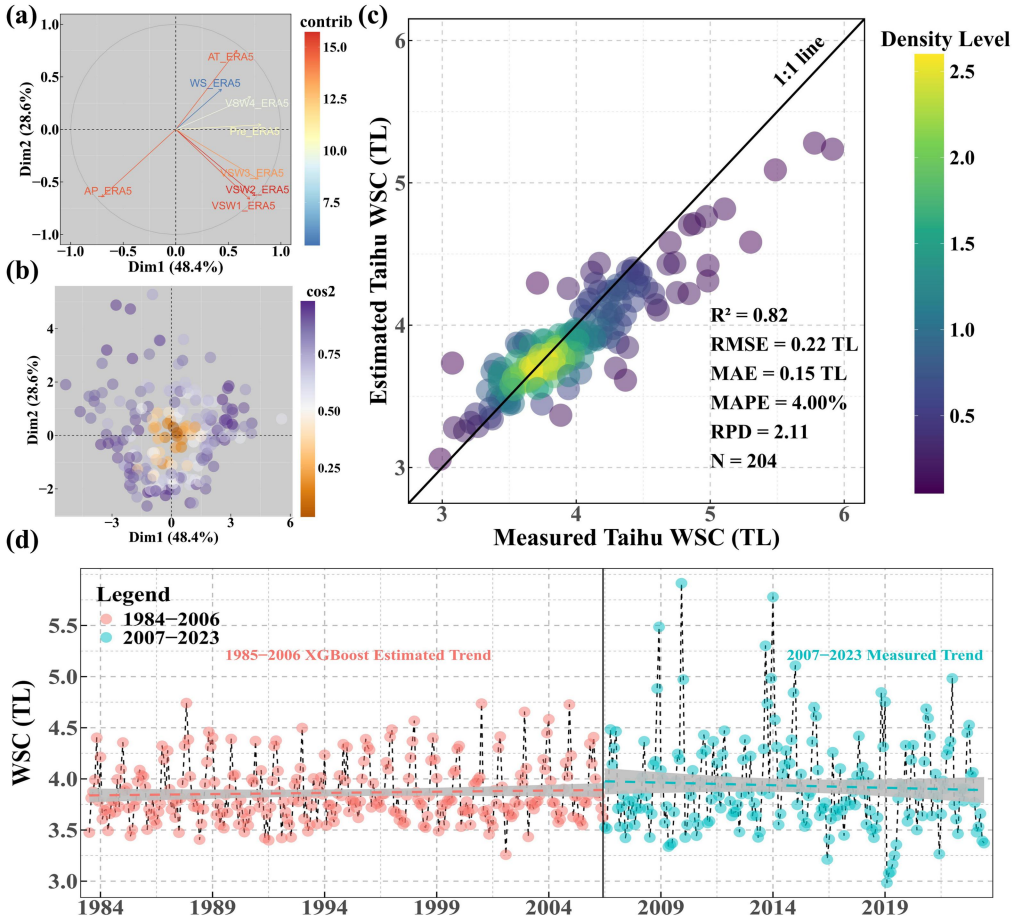



**Figure 4.10.** (a) and (b) Biplot and score plot of principal component analysis (PCA) of climate parameters of Lake Taihu based on WSC as the principal component, (c) scatter plot of measured WSC values, XGBoost model estimates, and model performance metrics, (d) time series of WSC in Taihu: model estimation and measured data (1984–2023).

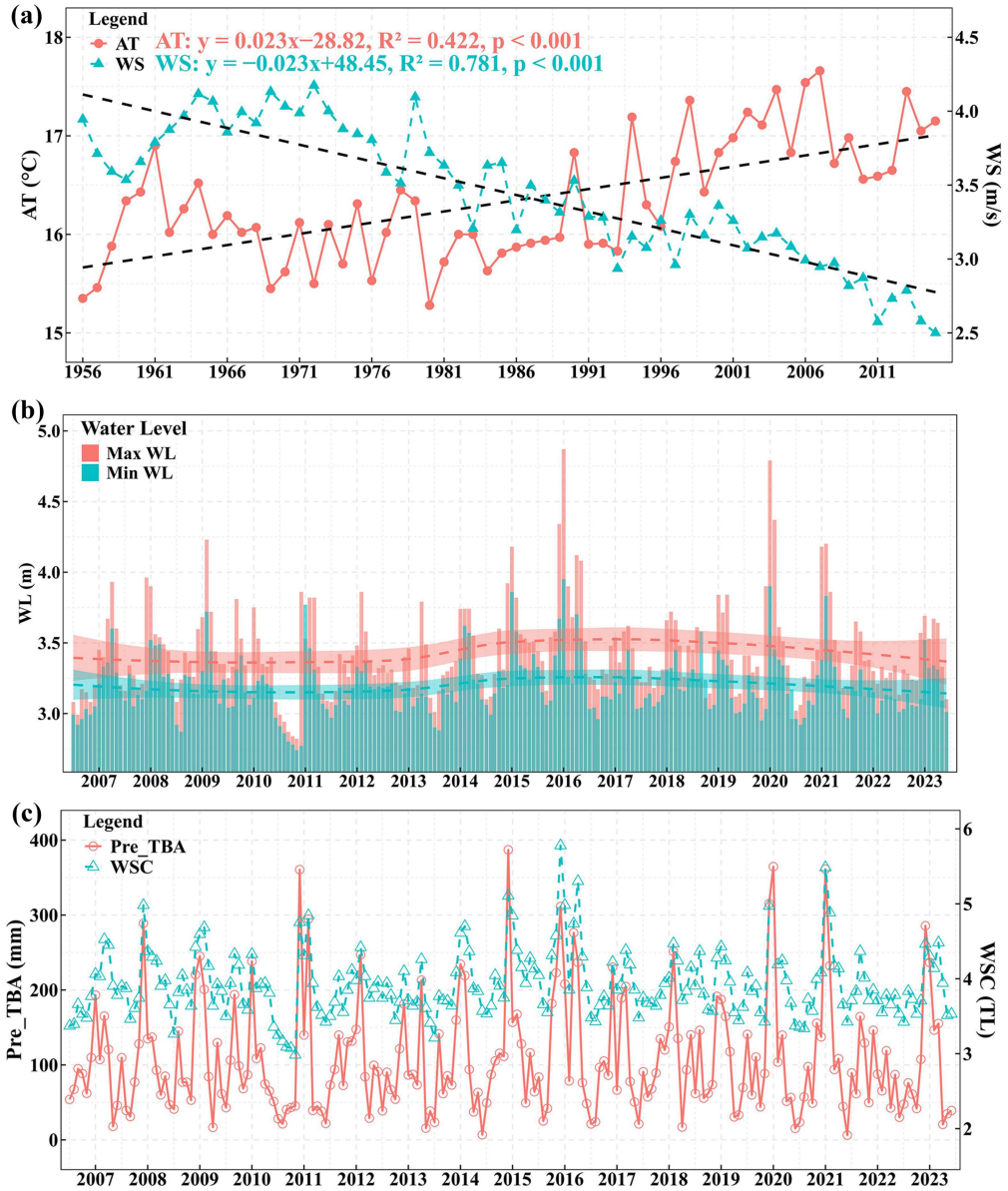



**Figure 4.11.** Temporal variations in climatic, hydrological, and precipitation parameters over Taihu (T1–T6): (a) AT and WS from 1956 to 2015, (b) Min and Max WL from 2007 to 2023, (c) monthly Pre_TBA and WSC from 2007 to 2023.

**Table 4.9.** Statistical summary of environmental parameters for Taihu.

| Environmental data | Range | Mean | SD | Median | $Q_{25}$–$Q_{75}$ | N | Measurement range | Perid |
|---|---|---|---|---|---|---|---|---|
| AT (°C) | 15.28–17.66 | 16.33 | 0.61 | 16.23 | 15.90–16.83 | 60 | Annual | 1956–2015 |
| WS (m/s) | 2.48–4.17 | 3.44 | 0.45 | 3.50 | 3.10–3.85 | 60 | Annual | 1956–2015 |
| Min WL (m) | 2.74–3.95 | 3.19 | 0.21 | 3.16 | 3.06–3.31 | 204 | Month | 2007–2023 |
| Max WL (m) | 2.82–4.87 | 3.43 | 0.32 | 3.36 | 3.23–3.58 | 204 | Month | 2007–2023 |
| Pre (mm) | 6.30–387.00 | 107.49 | 76.09 | 87.55 | 51.20–139.88 | 204 | Month | 2007–2023 |
| WSC (TL) | 2.98–5.91 | 3.93 | 0.46 | 3.85 | 3.64–4.17 | 204 | Month | 2007–2023 |

Notes: This table summarises the key environmental parameters used in the carbon inversion analysis of Taihu. The data for air temperature (AT) and wind speed (WS) were collected annually from 1956 to 2015 (Shi et al., 2018), while water levels (Min WL and Max WL), precipitation (Pre), and water storage capacity (WSC) were recorded monthly from 2007 to 2023. It was important to note the differences in measurement frequency (annual vs monthly) and the varied periods covered by each dataset. In addition, the WSC unit is a trillion litres (TL), which may differ from other public volume units. 100 million $m^3$ equals 100 billion litres, which is 0.1 TL.

Given the relatively small spatial extent of Taihu, a pixel-by-pixel approach was adopted to estimate DOCs (Song et al., 2018). Assuming vertical homogeneity of DOCc within each pixel, DOCs was computed by multiplying near-surface DOCc by the contemporaneous water-column height inferred from WSC and summing across all water pixels. To characterize four-decade spatial change, 40-year pixel-level differences in DOCc and DOCs were calculated as $\Delta DOC = DOC_{2023} - DOC_{1984}$. Lake-wide DOCs was quantified by multiplying annual near-surface DOCc maps by the contemporaneous lake-wide water-column height inferred from WSC and summing over all water pixels. Under these equations, DOCs is a year-specific multiplicative scaling of DOCc, which preserves relative spatial dispersion (SD/Mean) and facilitates direct comparison of within-lake heterogeneity metrics between DOCc and DOCs.

### 4.1.8. Driver analysis: auxiliary datasets and attribution framework

To interpret the four-decade variability of DOCc and DOCs (1984–2023), four groups of auxiliary drivers were assembled: (i) human activities, (ii) meteorological factors, (iii) bio-optical properties, and (iv) vegetation cover. These auxiliary drivers were not used as predictors in the reflectance-to-DOC retrieval model; they were applied only after DOC reconstruction to quantify potential controls and pathways. Human-activity indicators included $CO_2$ emissions and sequestration (CES), land-cover fractions (cropland, forest, grassland, impervious, water), gross domestic product (GDP), nighttime light (NTL), and population (POP), aggregated for the Lake Taihu Basin over Suzhou, Wuxi, Zhenjiang, Changzhou, Huzhou, Jiaxing, and Shanghai to represent anthropogenic pressures acting via watershed runoff, riverine inflows, and point-source inputs. Lake-scale drivers included ERA5-based precipitation (Pre_ERA5), air temperature (AT_ERA5), atmospheric pressure (AP_ERA5), and wind speed (WS_ERA5), as well as ultraviolet radiation (UVR), an ERA5-derived lake-scale annual water-depth proxy (Depth_ERA5, distinct from in situ WD), and WSC. Bio-optical indicators included Chl-a, floating algae cover (FAC), Secchi disk depth (SDD), and trophic state index (TSI). Vegetation indicators included emergent vegetation (EV), floating vegetation (FV), and Landsat-derived normalized difference

vegetation index (NDVI). The Lake Taihu Basin over which the human-activity indicators were aggregated is mapped in Figure 4.12, and the auxiliary datasets are summarized in Table 4.10.




**Figure 4.12.** Location of Lake Taihu Basin and Lake Taihu, including Suzhou, Wuxi, Changzhou (partially), Huzhou, Jiaxing, and the Qingpu District of Shanghai.

**Table 4.10.** Summary of auxiliary datasets.

| Datasets | Variables | Time scale | Spatial coverage | Unit | Sources |
|---|---|---|---|---|---|
| Human activity | $CO_2$ emissions and sequestration (CES) | 1997–2020 | Lake Taihu Basin | Mt C | (Chen et al., 2020) |
| | Cropland (Landcover) | 2000–2020 | Lake Taihu Basin | ha | (Xue et al., 2024) |
| | Forest (Landcover) | 2000–2020 | Lake Taihu Basin | ha | (Xue et al., 2024) |
| | Grassland (Landcover) | 2000–2020 | Lake Taihu Basin | ha | (Xue et al., 2024) |
| | Impervious (Landcover) | 2000–2020 | Lake Taihu Basin | ha | (Xue et al., 2024) |
| | Water (Landcover) | 2000–2020 | Lake Taihu Basin | ha | (Xue et al., 2024) |
| | Gross domestic product (GDP) | 1984–2023 | Lake Taihu Basin | Billion RMB | Resources and Environmental Science and Data Center (https://www.resdc.cn/) |
| | Nighttime light (NTL) | 1984–2020 | Lake Taihu Basin | / | (Zhang et al., 2024) |
| | Population (POP) | 2000–2020 | Lake Taihu Basin | People | (Xue et al., 2024) |
| Meteorological factors | Precipitation (Pre_ERA5) | 1984–2023 | Lake Taihu | mm | ERA5 Reanalysis Data in GEE |
| | Air temperature (AT_ERA5) | 1984–2023 | Lake Taihu | ℃ | ERA5 Reanalysis Data in GEE |
| | Atmospheric pressure (AP_ERA5) | 1984–2023 | Lake Taihu | Pa | ERA5 Reanalysis Data in GEE |
| | Wind speed (WS_ERA5) | 1984–2023 | Lake Taihu | m/s | ERA5 Reanalysis Data in GEE |
| | Ultraviolet radiation (UVR) | 1986–2021 | Lake Taihu | $kJ/m^2$ | Clear-sky erythemal UV dose (https://www.temis.nl/uvradiation/UVarchive/uvncfiles_msr2.php?) |

| | | | | | |
|---|---|---|---|---|---|
| | Water depth (Depth_ERA5) | 1984–2023 | Lake Taihu | M | Estimated using ERA5 Reanalysis Data |
| | Water storage capacity (WSC) | 1984–2023 | Lake Taihu | TL | Estimated using ERA5 Reanalysis Data |
| Bio-optics | Chl-a concentration | 1984–2019 | Lake Taihu | μg/L | (Xue et al., 2024) |
| | Floating algae cover (FAC) | 2003–2022 | Lake Taihu | % | (Xue et al., 2024) |
| | Secci disk depth (SDD) | 1986–2020 | Lake Taihu | M | (Xue et al., 2024) |
| | Trophic state index (TSI) | 1986–2020 | Lake Taihu | / | (Xue et al., 2024) |
| Vegetation cover | Emergent Vegetation (EV)-Aquaticvegetation | 2000–2020 | Lake Taihu | ha | (Xue et al., 2024) |
| | Floating Vegetation (FV)-Aquaticvegetation | 2000–2020 | Lake Taihu | ha | (Xue et al., 2024) |
| | Normalized difference vegetation index (NDVI) | 1984–2023 | Lake Taihu | / | Calculate NDVI values using Landsat data obtained from GEE |

Notes: Auxiliary datasets with occasional missing years were gap-filled using XGBoost regression (Cropland, Forest, Grassland, Impervious, Water, GDP, NTL, POP, Chl-a, FAC, SDD, TSI, EV, FV). CES gaps (2018–2020) were filled using ARIMA, and WSC (1984–2006) was reconstructed using the dedicated PCA+XGBoost hindcasting workflow. Depth_ERA5 refers to an ERA5-derived lake-scale annual water-depth proxy; WD refers to in situ water depth in the field WQPs (2007–2015).

All auxiliary variables were harmonized to the analysis time step of the DOC record (annual and zone-aggregated where applicable) using consistent spatial masks. Intermittent missing years within a dataset's native period were imputed using an XGBoost regression that leveraged time features, lagged terms of the target, and contemporaneous covariates. The model was trained exclusively on observed data, evaluated using blocked, time-aware CV, and then used to predict the missing entries (Shen et al., 2025b). CES gaps for 2018–2020 were filled separately using ARIMA. Imputation was not used to extrapolate variables beyond their original support period; analyses requiring multiple drivers were conducted over their common overlap period. The combined influence of the four driver groups on DOCc and DOCs was then quantified using three complementary approaches: (1) SEM to separate direct and indirect pathways among climate/hydrology, eutrophication proxies, vegetation, human activity, and DOC variability; (2) driver-based ML explanation models (trained to predict reconstructed DOCc/DOCs from auxiliary drivers) to obtain feature-importance rankings; and (3) Pearson correlation analysis with significance testing.

## 4.2. Study 2: Satellite retrieval and attribution of narrow-river nutrients in South Bohemia

Study 2 shifted the focus from a large, spatially well-resolved lake to a dense network of narrow agricultural rivers. Its aim was to assess the feasibility and limits of nutrient retrieval and attribution within a consistent, spatially explicit validation framework, rather than generating a nutrient product. Two questions were examined: whether nutrient concentrations could be predicted at previously unseen reaches and whether local, reach-scale land-surface change could explain the network's long-term nutrient trajectory. Model comparisons used matched observations and predictors, with shared fold assignments across validation designs. This framework helped assess whether limitations persisted across learners and distinguish constraints associated with the observations and spatial structure from algorithm-specific performance. The resulting limits were interpreted within the scope of the tested datasets, sensors, and methods.

### 4.2.1. Workflow of the nutrient retrievability test

The analysis was organized into three layers. The observation layer constructed a reach-object archive by harmonizing surface reflectance from seven satellite data streams, applying QA/QC, and matching satellite observations to in situ records. The inference layer evaluated nutrient retrievability, characterized spatial structure in nutrient concentrations, and assessed attribution of long-term changes. The synthesis layer integrated the retrieval and attribution findings into an overall diagnostic assessment. Spatial CV provided a common framework for evaluating predictive transferability to unmonitored locations (Figure 4.13).

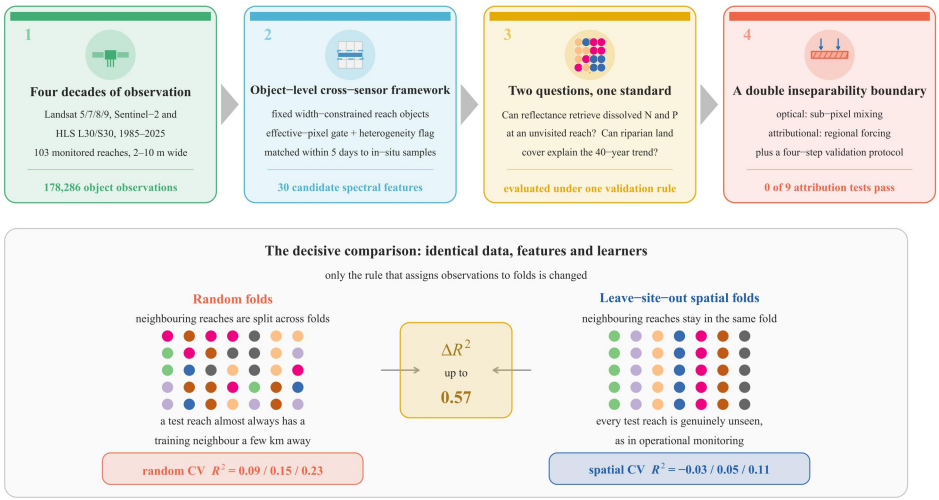



**Figure 4.13.** Analytical framework and the comparison on which the study turns. Top: the four stages of the analysis, from the long-term multi-sensor archive, through the object-level cross-sensor framework, to the two questions evaluated under a single validation standard, to the resulting double inseparability boundary; each card carries the corresponding headline quantity. Bottom: the decisive comparison. The same data, features, and learners are used throughout; only the rule that assigns observations to folds changes. Under random folds, neighboring reaches are split across folds, so a test reach almost always has a training neighbor a few kilometers away and reported skill reaches $R^2$ = 0.23. Under seven-fold spatial-block CV spatial folds, neighboring reaches stay in the same fold, and every test reach is genuinely unseen, as in operational monitoring, and skill falls to about zero; the gap reaches $\Delta R^2$ = 0.57. The dot lattices are schematic illustrations of the two-fold designs, not maps of the monitoring network.

### 4.2.2. Study area, monitoring network and width stratification

South Bohemia (Jihočeský kraj; approximately 48.6–49.6°N, 13.8–15.5°E) is drained by a dense river network dominated by the Vltava, Lužnice, and Otava systems. The region encompasses forested uplands, intensively cultivated lowlands, and a long-established fishpond landscape (Figure 4.14a–c). River water-quality observations

were obtained from the open database of the Czech Hydrometeorological Institute (CHMI). The monitoring network comprised 120 fixed stations, with a complete inventory provided in the supporting information of the corresponding publication. Channel widths could be measured at 103 stations using national orthophotos with a spatial resolution of 0.125 m, and these stations were retained for analysis. The remaining 17 stations were excluded because canopy cover completely obscured the channel. Representative orthophotos illustrate channel widths ranging from 1.89 to 84.61 m (Figure 4.14d–f).

The 103 retained reaches were divided into three width classes: W1 (<10 m; n = 60, 58.25%), W2 (10–30 m; n = 30, 29.13%), and W3 (>30 m; n = 13, 12.62%). Median widths were 4.9, 15.3 and 54.9 m for W1, W2 and W3, respectively, with an overall median of 9.2 m across the retained reaches (Table 4.11). Channel width affects the degree of land–water mixing within medium-resolution pixels and was therefore used as the primary stratification variable. The same width classes were applied throughout the retrieval test, trend analysis and attribution test. All spatial operations were performed using the coordinate reference system registry code (EPSG):5514.

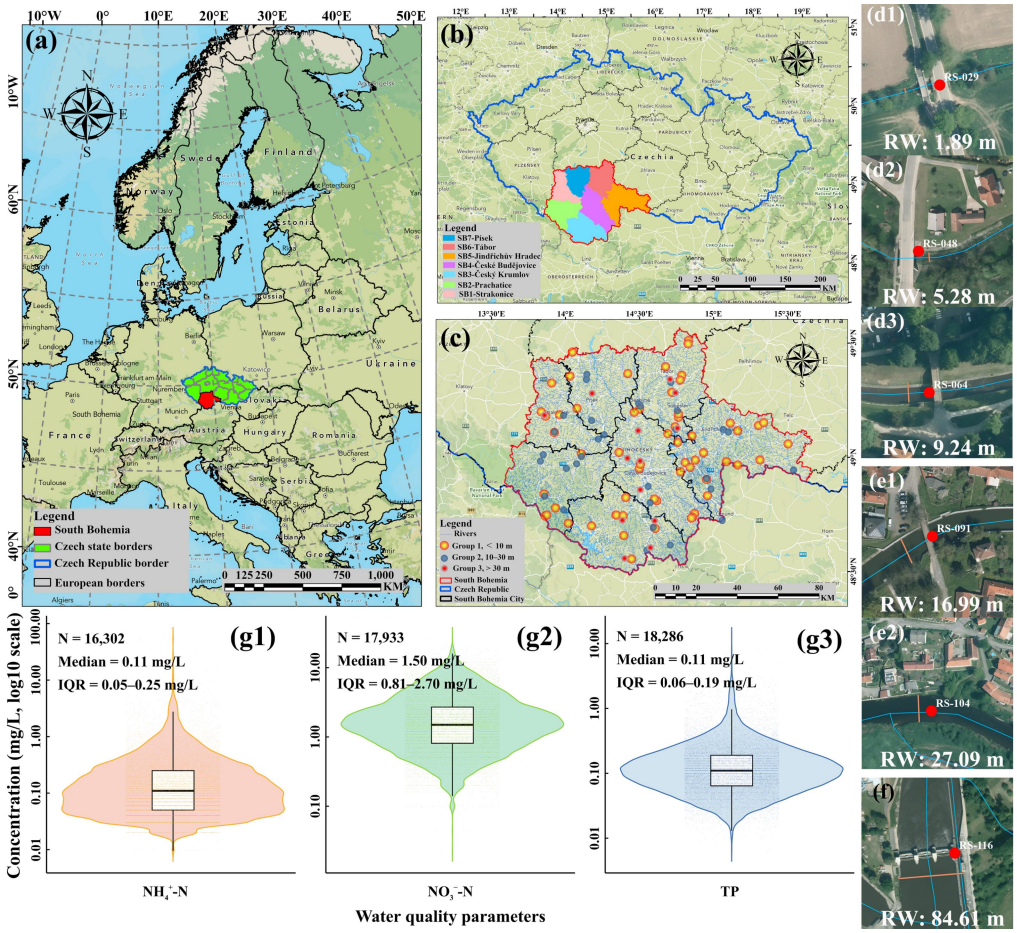



**Figure 4.14.** Study area, monitoring network, and nutrient concentration distributions. (a) Location of South Bohemia in Europe. (b) South Bohemia within the Czech Republic and its seven districts (SB1–SB7). (c) Monitoring stations classified into three river-width groups. (d1–d3) narrow reaches (< 10 m); (e1–e2) intermediate reaches (10–30 m); and (f) a wide reach (> 30 m). RW denotes orthophoto-derived river width. (g1–g3) Concentration distributions of $NH_4^+$-N, $NO_3^-$-N and TP, respectively, during 1990–2020. Violin plots with embedded box plots are presented on a $\log_{10}$ concentration scale, with sample sizes, medians and interquartile ranges indicated.

**Table 4.11.** Channel-width classes of the South Bohemian monitoring network and the object-level observation archive (1985–2025).

| Width class | Channel width | Median width (m) | Width / 30-m pixel | Reaches (n) | Reaches (%) | Object observations (n) | Observations (%) | Internally heterogeneous objects (%) |
|---|---|---|---|---|---|---|---|---|
| W1 | < 10 m | 4.9 | 0.16 | 60 | 58.25 | 96,283 | 54.00 | 5.97 |
| W2 | 10–30 m | 15.3 | 0.51 | 30 | 29.13 | 55,304 | 31.02 | 15.73 |
| W3 | > 30 m | 54.9 | 1.83 | 13 | 12.62 | 26,699 | 14.98 | 61.32 |
| All | — | 9.2 | 0.31 | 103 | 100.00 | 178,286 | 100.00 | 17.29 |

Notes: Objects are flagged as internally heterogeneous when the within-object coefficient of variation of green and near-infrared reflectance exceeds 0.30 in both bands.

### 4.2.3. In-situ nutrient records

Three eutrophication-related water-quality variables were analysed: $NH_4^+$-N, $NO_3^-$-N and TP. Following harmonisation for 1990–2020, the dataset comprised 16,302, 17,933 and 18,286 observations for $NH_4^+$-N, $NO_3^-$-N and TP, respectively, yielding 52,521 observations in total. Their corresponding median concentrations were 0.11, 1.50 and 0.11 mg $L^{-1}$. All three variables showed strongly right-skewed distributions within each width class (Figure 4.14g1-g3; Table 4.12). Concentrations were therefore $\log_{10}$-transformed for all subsequent analyses. Before trend estimation, observations at or below the analytical detection limits were screened to minimise potential floor effects. Measurements by CHMI followed standard analytical methods, which are summarised with supporting references in Table 4.13. The sampling design and temporal coverage are documented in Figure 4.15, which also shows the systematic difference in record length between narrow and wide reaches.

**Table 4.12.** Summary statistics of the eutrophication-related water-quality parameters ($NH_4^+$-N, $NO_3^-$-N and TP) for the South Bohemian monitoring dataset (1990–2020). Q1 and Q3 are the 25th and 75th percentiles; skewness is reported on the raw and $log_{10}$ scales.

| Parameter | $NH_4^+$-N | $NO_3^-$-N | TP |
|---|---|---|---|
| Time window | 1990–2020 | 1990–2020 | 1990–2020 |
| Date range | 1990-01-10 to 2020-12-16 | 1990-01-10 to 2020-12-16 | 1990-01-10 to 2020-12-16 |
| N | 16,302 | 17,933 | 18,286 |
| Q1 | 0.05 | 0.81 | 0.06 |
| Q3 | 0.25 | 2.70 | 0.19 |
| Min (mg $L^{-1}$) | 0.01 | 0.02 | 0.01 |
| Max (mg $L^{-1}$) | 56.00 | 27.00 | 13.00 |
| Median (mg $L^{-1}$) | 0.11 | 1.50 | 0.11 |
| Mean (mg $L^{-1}$) | 0.28 | 2.28 | 0.17 |
| SD (mg $L^{-1}$) | 1.07 | 2.27 | 0.28 |
| Skewness (raw) | 30.36 | 2.53 | 14.17 |
| Skewness ($log_{10}$) | 0.70 | -0.07 | 0.42 |

**Table 4.13.** In-situ analytical determination of $NH_4^+$-N, $NO_3^-$-N and TP by the Czech Hydrometeorological Institute (CHMI).

| Parameter | Analytical determination | Reference |
|---|---|---|
| $NH_4^+$-N | Spectrophotometry, indophenol-blue reaction | (APHA, 2017) |
| $NO_3^-$-N | Ion chromatography or spectrophotometry after reduction | (APHA, 2017) |
| TP | Persulfate digestion followed by the molybdenum-blue method | (Grasshoff et al., 2009; Murphy and Riley, 1962) |
| All parameters | Values at or below the analytical detection limit screened before trend estimation | (APHA, 2017) |

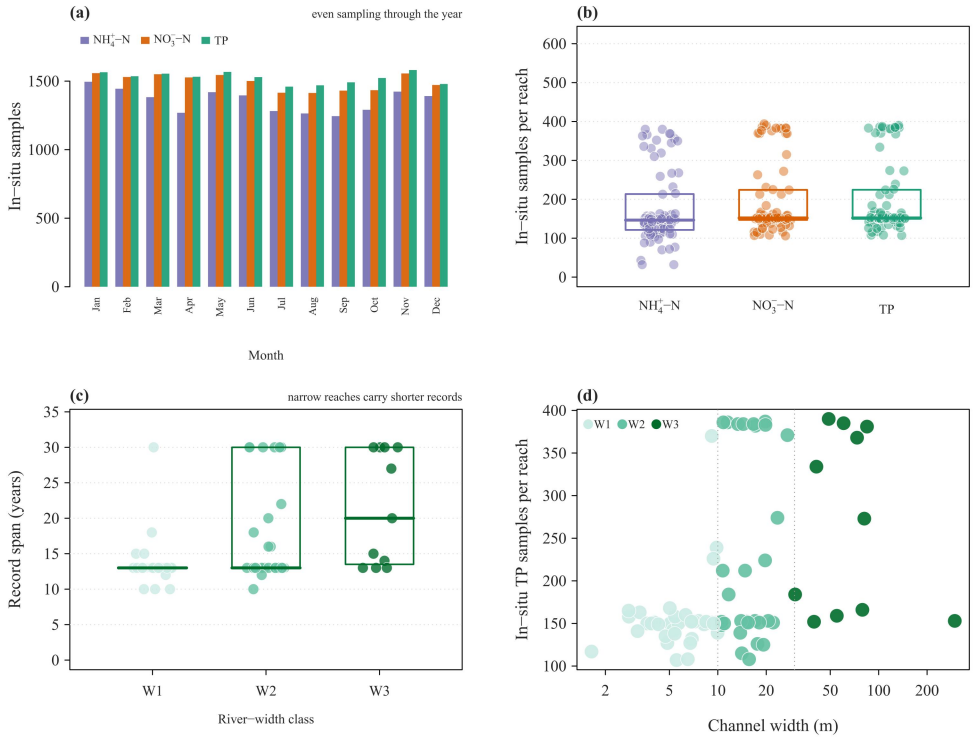



**Figure 4.15.** Design of the in-situ sampling program, 1990–2020. (a) In-situ samples per month and parameter. (b) Samples per reach for each parameter. (c) Record span by width class, showing that narrow reaches carry systematically shorter records, the confound controlled in the trend analysis reported with the results. (d) TP samples per reach against channel width, colored by width class.

### 4.2.4. Object-level cross-sensor framework, quality gating and matchups

In the object-level cross-sensor framework (OCSF), the analytical unit was a fixed, width-constrained river-reach object rather than an individual pixel. This design directly addressed the mixed-pixel problem in narrow rivers (Blaschke, 2010b; Blaschke et al., 2014). A reach centerline was digitized from 0.125-m orthophotos and buffered on both sides. The one-sided buffer distance was defined as the larger of half the interpreted channel width or a minimum of 2 m. Each object remained fixed throughout the analysis because surface-water extent observed at 30-m resolution can

vary substantially among scenes (Pekel et al., 2016). Redelineating objects for individual scenes could therefore confound changes in water quality with changes in object geometry.

Surface reflectance was obtained from seven harmonised data streams: Landsat 5, 7, 8, and 9 (Roy et al., 2014b; Wulder et al., 2019), Sentinel-2 (Drusch et al., 2012), and the HLS L30 and S30 products (Claverie et al., 2018), which have also been evaluated for aquatic applications (Maciel et al., 2023). Band definitions are provided in Table 4.14. The HLS products contain NASA-reprocessed Landsat 8/9 and Sentinel-2 observations and therefore share acquisition dates with the corresponding native streams. They were analyzed as separate processing streams to assess sensitivity to atmospheric correction and harmonization, rather than as independent satellite observations. Conclusions were evaluated separately within each stream, and cross-stream agreement was not interpreted as fully independent replication.

For each cloud-free object-date, summary statistics were calculated for the spectral bands, NDWI, MNDWI, NDVI, normalized difference turbidity index (NDTI), and selected band ratios, yielding 30 candidate predictors. Two complementary quality checks were applied to limit land contamination. Firstly, an effective-pixel gate retained an object-date only when the number of valid green-band pixels met a minimum threshold determined by pixel size and width class. Secondly, an object-date was flagged as internally heterogeneous when the within-object coefficients of variation for both green and near-infrared reflectance exceeded 0.30 (Kuhn et al., 2019; Vanhellemont and Ruddick, 2018). The corresponding thresholds and operators are listed in Table 4.15. Object-level reflectance was matched to in-situ observations acquired within ±5 days, and $\log_{10}$-transformed nutrient concentration was used as the response variable (Lehmann et al., 2023; Ross et al., 2019). Predictors were standardised within each training fold to prevent information leakage during feature scaling. Figure 4.16 quantifies the resulting geometric constraints, including why quality gating alone cannot recover an uncontaminated water signal from a subpixel channel.

**Table 4.14.** Band definitions and native spatial resolution for Sentinel-2 MSI, Landsat-5 TM, Landsat-7 ETM+ and Landsat-8 OLI/TIRS.

| Mission / Sensor | Band | Common name | Center (nm) | Bandwidth (nm) | Resolution (m) | Wavelength range (μm) |
|---|---|---|---|---|---|---|
| Sentinel2 (MSI) | B1 | Coastal aerosol | 443 | 20 | 60 | / |
| Sentinel2 (MSI) | B2 | Blue | 490 | 65 | 10 | / |
| Sentinel2 (MSI) | B3 | Green | 560 | 35 | 10 | / |
| Sentinel2 (MSI) | B4 | Red | 665 | 30 | 10 | / |
| Sentinel2 (MSI) | B5 | Rededge 1 | 705 | 15 | 20 | / |
| Sentinel2 (MSI) | B6 | Rededge 2 | 740 | 15 | 20 | / |
| Sentinel2 (MSI) | B7 | Rededge 3 | 783 | 20 | 20 | / |
| Sentinel2 (MSI) | B8 | NIR | 842 | 115 | 10 | / |
| Sentinel2 (MSI) | B8A | Narrow NIR | 865 | 20 | 20 | / |
| Sentinel2 (MSI) | B9 | Water vapour | 945 | 20 | 60 | / |
| Sentinel2 (MSI) | B10 | Cirrus | 1375 | 30 | 60 | / |
| Sentinel2 (MSI) | B11 | SWIR 1 | 1610 | 90 | 20 | / |
| Sentinel2 (MSI) | B12 | SWIR 2 | 2190 | 180 | 20 | / |
| Landsat5 (TM) | B1 | Blue | / | / | 30 | 0.45–0.52 |
| Landsat5 (TM) | B2 | Green | / | / | 30 | 0.52–0.60 |
| Landsat5 (TM) | B3 | Red | / | / | 30 | 0.63–0.69 |
| Landsat5 (TM) | B4 | NIR | / | / | 30 | 0.76–0.90 |
| Landsat5 (TM) | B5 | SWIR 1 | / | / | 30 | 1.55–1.75 |
| Landsat5 (TM) | B6 | TIR | / | / | 120 | 10.4–12.5 |
| Landsat5 (TM) | B7 | SWIR 2 | / | / | 30 | 2.08–2.35 |

| Mission / Sensor | Band | Common name | Center (nm) | Bandwidth (nm) | Resolution (m) | Wavelength range (μm) |
|---|---|---|---|---|---|---|
| Landsat7 (ETM+) | B1 | Blue | / | / | 30 | 0.45–0.52 |
| Landsat7 (ETM+) | B2 | Green | / | / | 30 | 0.52–0.60 |
| Landsat7 (ETM+) | B3 | Red | / | / | 30 | 0.63–0.69 |
| Landsat7 (ETM+) | B4 | NIR | / | / | 30 | 0.77–0.90 |
| Landsat7 (ETM+) | B5 | SWIR 1 | / | / | 30 | 1.55–1.75 |
| Landsat7 (ETM+) | B6 | TIR | / | / | 60 | 10.4–12.5 |
| Landsat7 (ETM+) | B7 | SWIR 2 | / | / | 30 | 2.09–2.35 |
| Landsat7 (ETM+) | B8 | Panchromatic | / | / | 15 | 0.52–0.90 |
| Landsat8 (OLI/TIRS) | B1 | Coastal aerosol | / | / | 30 | 0.433–0.453 |
| Landsat8 (OLI/TIRS) | B2 | Blue | / | / | 30 | 0.450–0.515 |
| Landsat8 (OLI/TIRS) | B3 | Green | / | / | 30 | 0.525–0.600 |
| Landsat8 (OLI/TIRS) | B4 | Red | / | / | 30 | 0.630–0.680 |
| Landsat8 (OLI/TIRS) | B5 | NIR | / | / | 30 | 0.845–0.885 |
| Landsat8 (OLI/TIRS) | B6 | SWIR 1 | / | / | 30 | 1.560–1.660 |
| Landsat8 (OLI/TIRS) | B7 | SWIR 2 | / | / | 30 | 2.100–2.300 |
| Landsat8 (OLI/TIRS) | B8 | Panchromatic | / | / | 15 | 0.500–0.680 |
| Landsat8 (OLI/TIRS) | B9 | Cirrus | / | / | 30 | 1.360–1.390 |
| Landsat8 (OLI/TIRS) | B10 | TIRS 1 | / | / | 100 | 10.60–11.19 |
| Landsat8 (OLI/TIRS) | B11 | TIRS 2 | / | / | 100 | 11.50–12.51 |

Notes: Landsat-9 (OLI-2) is spectrally equivalent to Landsat-8 (OLI) and follows the Landsat-8 band definitions above. The HLS L30 and S30 products are NASA harmonized surface reflectance resampled to a common 30-m grid, inheriting the band definitions of Landsat-8/9 OLI and Sentinel-2 MSI, respectively; they are therefore not listed separately. Band centre and bandwidth are given for Sentinel-2; wavelength range (as reported by USGS) is given for the Landsat sensors.

**Table 4.15.** Object definition, effective-pixel gate and internal-heterogeneity flag of the object-level cross-sensor framework.

| Component | Specification | Applies to | Setting |
|---|---|---|---|
| Reach object | Fixed polygon from a centreline digitized on 0.125-m orthophotos, buffered by the larger of half the interpreted width or 2 m | All streams | Fixed through time |
| Pixel masking | Cloud, shadow, snow and saturation masks applied before any summary | All streams | Product QA bands |
| Summary operators | Mean, median, standard deviation and interquartile range of each band, plus NDWI, MNDWI, NDVI, NDTI and band ratios | All object–dates | 30 candidate features |
| Effective-pixel gate | Minimum count of valid green-band pixels per object–date | Sentinel-2 (10 m), W1 | 3 pixels |
| Effective-pixel gate | Minimum count of valid green-band pixels per object–date | Sentinel-2 (10 m), W2–W3 | 5 pixels |
| Effective-pixel gate | Minimum count of valid green-band pixels per object–date | 30-m streams (Landsat 5/7/8/9, HLS L30 and S30), W1 | 2 pixels |
| Effective-pixel gate | Minimum count of valid green-band pixels per object–date | 30-m streams, W2–W3 | 3 pixels |
| Internal-heterogeneity flag | Within-object coefficient of variation of green and near-infrared reflectance | All streams | Flagged when both exceed 0.30 |
| Matchup rule | Nearest in-situ sample within a symmetric temporal window | All streams | ±5 days |
| Sensitivity | Alternative buffer distances, quality subsets and width-class subsets | Spatial-block CV verdict | Insensitive (Figure 5.10) |

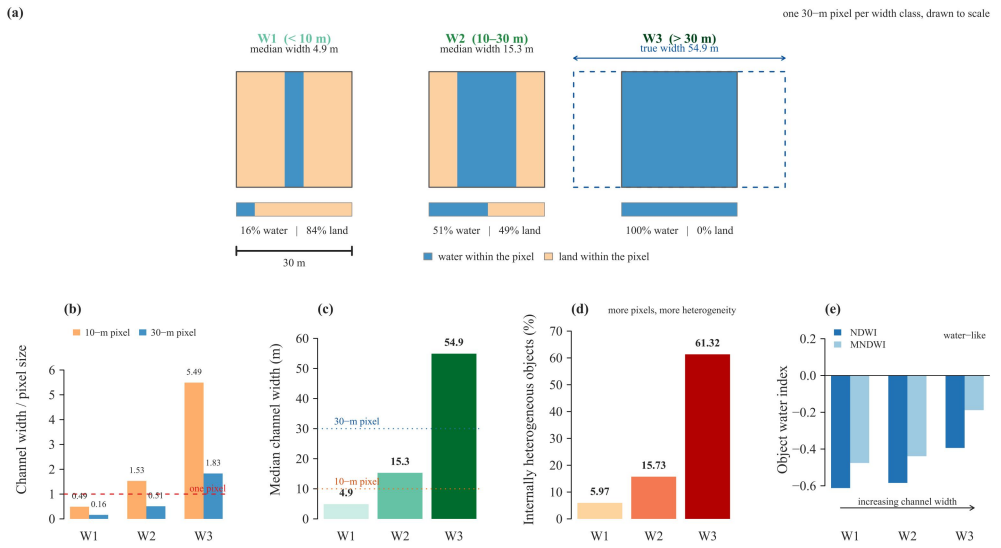



**Figure 4.16.** Sub-pixel geometry and object quality. (a) One 30-m pixel per width class, drawn to scale: the blue band is the median channel of that class, orange is land within the same pixel, and the bar beneath gives the resulting water and land fractions; for W3 the dashed outline marks the true channel width, which exceeds one pixel. (b) Median channel width as a fraction of a 10-m and a 30-m pixel. (c) Median channel width per class against the two-pixel sizes. (d) Fraction of objects flagged as internally heterogeneous. (e) Object-level water indices, which approach water-like values only in the widest class.

### 4.2.5. Retrievability protocol: random versus spatial-block CV

The protocol compared two seven-fold CV schemes using identical observations, predictors, and model classes. Three regressors represented nonlinear ensemble and regularized linear approaches: RF, gradient-boosted trees (GBT), and a regularized linear model (Kattenborn et al., 2021). The first scheme used conventional random fold assignments. The second used spatially blocked CV, in which monitoring sites were assigned to seven spatial clusters generated by k-means clustering of station coordinates. Each cluster served once as the test fold, keeping all observations from a given reach in the same fold and preventing reach-level overlap between training and

testing data (Milà et al., 2022a; Ploton et al., 2020; Roberts et al., 2017). These coordinate-based clusters were unrelated to the SB1–SB7 administrative districts.

Both schemes used seven folds, the same feature set, and common hyperparameter settings, so the comparison isolated the influence of fold assignment within the tested model configurations. Performance metrics were calculated from pooled out-of-fold predictions across all seven folds rather than average across fold-specific estimates. Potential inflation in apparent skill was quantified as $\Delta R^2 = R^2_{random} - R^2_{spatial}$, with positive values indicating higher apparent performance under random splitting (Karasiak et al., 2022). Both schemes were evaluated at four scopes: the full dataset and each of the three width classes. Robustness was additionally assessed across strata defined by sensor, season, matchup window, width class, and quality flag. The full modelling pipeline was refitted within each stratum rather than applying post hoc subsets to predictions from the full dataset. A 120-iteration reach-level bootstrap resampled entire reaches with replacement and retained the complete spatial CV pipeline for each replicate. The full protocol is specified in Table 4.16.

**Table 4.16.** Specification of the validation protocol applied identically to the retrieval and the attribution tests. The seven spatial folds are numerical clusters of station coordinates, mapped with the spatial-structure results, and are unrelated to the seven SB1–SB7 administrative districts.

| Element | Specification |
|---|---|
| Target | $Log_{10}$ concentration of $NH_4^+$-N, $NO_3^-$-N and TP |
| Predictors | 30 object-level reflectance features, standardized within each training fold |
| Learners | Random forest; gradient-boosted trees; regularized linear model |
| Scheme 1 | Conventional random k-fold CV |
| Scheme 2 | Spatial-block CV; folds are seven contiguous blocks from k-means clustering of station coordinates (Figure 5.16) |
| Fold count | Identical between the two schemes |
| Skill statistic | $R^2$ computed once over all pooled out-of-fold predictions (never averaged over folds) |
| Inflation | $\Delta R^2 = R^2_{random} - R^2_{spatial}$ |
| Scopes | Full sample; W1; W2; W3 |
| Stratified re-fits | Reflectance stream (7); season (4); matchup window (0–5 days); width class (3); quality gate (2) |
| Uncertainty | 120-iteration bootstrap resampling reaches with replacement, full spatial-CV pipeline rerun |

| Element | Specification |
|---|---|
| Mutual information | Binned estimator, computed across reaches and within reaches; significance from 999 permutations |
| Spatial diagnostics | Moran's I with k-nearest-neighbour weights (k = 5); empirical semivariogram (8-km bins); between- and within-reach variance partition; nearest-neighbour station distance |
| Attribution test | Identical folds; OLS retained as an in-sample baseline; coupling declared only if leave-block-out spatial $R^2 > 0.15$ |
| Software | Google Earth Engine (reflectance and land cover); R 4.6 with spdep (statistics) |

### 4.2.6. Model-free dependence and spatial-structure diagnostics

Because CV performance can depend on the learner, mutual information serves as a complementary model-agnostic diagnostic. Mutual information between each predictor and $\log_{10}$-transformed nutrient concentration was estimated from binned data across all reaches and within individual reaches, with significance assessed using 999 permutations. Higher within-reach than across-reach mutual information indicated more localized and potentially less transferable dependence. This analysis helped assess whether weak cross-validated skill reflected limitations of the regressors, although univariate mutual information based on binned data could not exclude multivariate feature interactions or dependence obscured by discretization.

The spatial structure of each nutrient field was characterized independently by the predictive models. Because inflation under random CV depends in part on the autocorrelation range relative to station spacing, four diagnostics were applied. Moran's I was calculated from reach-mean concentrations using k-nearest-neighbor weights (k = 5), with significance assessed by permutation. An empirical semivariogram of reach-mean concentrations was constructed using 8-km lag bins. Repeated observations were used to partition total variance into between-reach and within-reach components, while nearest-neighbor station distances were summarized to relate observed spatial structure to monitoring density (Dormann et al., 2007). Together, these measures described the persistence of site-level differences and whether network spacing adequately resolved spatial dependence.

### 4.2.7. Characterization of the long-term recovery

To distinguish temporal trends from changes in monitoring network composition, a fixed, long-span panel of 15 reaches was defined. Each reach was sampled in both the early (1993 or earlier) and late (2017 or later) periods. Parameter-specific median annual trajectories and net changes from the first to the last sampled year were calculated. Restricting the analysis to a consistent panel reduced potential bias from station turnover, which can create spurious trends when sites enter or leave the monitoring record. Reach-specific Theil–Sen slopes were also estimated for 76 reaches with sufficient temporal coverage, and their significance was evaluated using the Mann–Kendall test ($p < 0.05$). For all 103 reaches, temporal coverage was summarized by the number of observations per parameter, the first and last sampled years, and the number of sampled years. To assess width dependence, slopes were regressed against width class before and after adjustment for record span. This comparison tested whether the apparent width effect reflected the shorter and more recent records available for narrow reaches (Figure 4.15c).

### 4.2.8. Multi-scale riparian attribution under identical spatial rigour

Riparian buffers of 50, 150, and 500 m were delineated around reach objects and intersected with Coordination of Information on the Environment (CORINE) Land Cover data for 1990 and 2018. Land cover was reclassified into four categories: urban land, cropland, forest, wetland and semi-natural land. Changes in the areal proportion of each category were calculated at each buffer scale. Using three buffer distances reduced the risk that weak associations arose from a mismatch between the buffer and the relevant zone of influence. Relationships between reach-specific nutrient slopes and land-cover changes were evaluated using the protocol described in Section 4.2.5, combining leave-block-out spatial CV with Moran's I tests of residuals. Ordinary least squares (OLS) models fitted to the dataset yielded an optimistic in-sample baseline. A nutrient-scale combination was considered coupled only when its spatial-block cross-validated $R^2$ exceeded 0.15. Applying the spatial validation standard to retrieval and attribution enabled a comparison of their limitations. Reflectance and land-cover data were processed in GEE (Gorelick et al., 2017), and statistical analyses were performed in R 4.6 using the spdep package.

## 4.3. Study 3: Modelling oxygen demand in narrow rivers from catchment characteristics

Study 3 examined the same river network as Study 2 but focused on a different group of water-quality variables and a different observational hypothesis. COD and $BOD_5$ quantify chemically oxidizable and readily biodegradable material, respectively, and are widely used as indicators of organic pollution and oxygen demand. Under European water policy, oxygenation conditions are among the physicochemical parameters used to assess ecological status, while COD and $BOD_5$ are also regulated as wastewater discharge parameters. The study first tested whether these variables could be retrieved optically across narrow river reaches. Where the water-surface signal was insufficient, it then examined whether catchment characteristics could predict their concentrations. Both approaches were evaluated using LOSO spatial CV. This common validation standard enabled a direct comparison of the optical and catchment-based results with respect to their transferability to unmonitored reaches.

Throughout the study, the modelled, reconstructed, and projected response was concentration (mg $L^{-1}$), not flux. The CHMI monitoring network reports COD and $BOD_5$ concentrations but provides no co-located continuous discharge records for the investigated reaches. The observations, therefore, cannot be reliably converted into loads expressed as mass per unit time. This distinction is essential: the results describe water-quality conditions but do not quantify hydrological exports. A concentration-based analysis nevertheless matches the available monitoring data and the study objectives. It is also relevant to regulatory assessment, although wastewater-discharge limits and receiving-river concentrations represent different environmental contexts. The only exception is the wastewater register described in Section 4.3.5, which reports organic loads in units of mass per time and is therefore treated explicitly as a flux.

### 4.3.1. Workflow and design principles of the oxygen-demand study

Study 3 comprised six linked stages spanning data assembly, validation, reconstruction, and scenario analysis (Figure 4.17). Long-term in situ COD and $BOD_5$ records were first compiled and linked to a harmonized multisensor satellite archive (Stages 1–2). Direct optical retrieval of dichromate COD (COD-Cr, hereafter COD) and $BOD_5$ were then evaluated, while a parallel catchment-attribute model incorporated land-surface descriptors and point-source pressure indicators (Stage 3). Both models predicted concentrations rather than loads because co-located continuous discharge records were unavailable. Observed changes during 1985–2025 were subsequently characterized (Stage 4), and annual concentration fields were reconstructed across the river network (Stage 5). Finally, scenario experiments quantified the sensitivity of reconstructed concentrations to specified land-use and climate perturbations representing 2050 conditions (Stage 6). These experiments were treated as within-range sensitivity analyses rather than deterministic forecasts.

Two design principles ensured methodological consistency. Firstly, every predictive step used LOSO spatial CV, with all observations from the held-out reach excluded from training. Applying the same site-level validation to the optical and catchment models made their contrasting results directly comparable and limited inflation from repeated, spatially clustered observations. Secondly, the river-width gradient (W1–W3) was retained throughout to test how performance varied with observation scale and catchment integration. This was important because narrow headwaters experience severe pixel mixing but may retain stronger links between local catchment conditions and in-stream concentrations. Reflectance and catchment processing were conducted in GEE (Gorelick et al., 2017), and statistical analyses were performed in R 4.6.0. The optical test was interpreted in light of previous COD and $BOD_5$ retrieval studies, many of which did not explicitly separate monitoring sites during validation. Table 4.17 summarizes their algorithms, regions, sensors, study periods, and validation strategies, providing context for the present results.

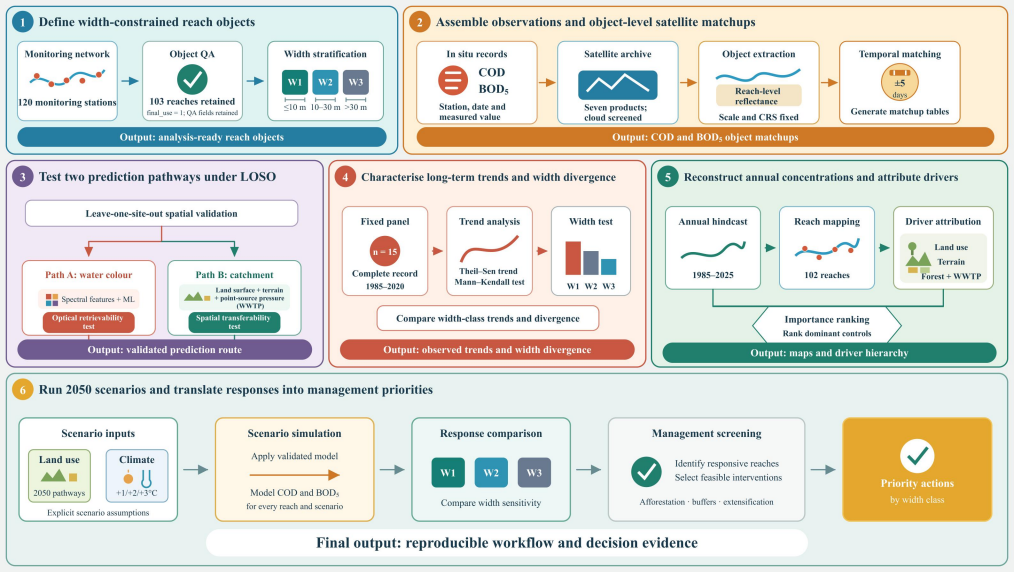



**Figure 4.17.** End-to-end workflow for monitoring oxygen-demand concentrations in narrow South Bohemian rivers (1985–2025). Six colour-coded stages link reach-object definition and width stratification (Stage 1), water-quality records and satellite matchups (Stage 2), dual-path modelling and the retrievability boundary (Stage 3), four-decade trends and width divergence (Stage 4), hindcast reconstruction and driver attribution (Stage 5), and scenario sensitivity and management priorities to 2050 (Stage 6). Each box shows its inputs, operations and output; river-width classes W1–W3 recur throughout.

**Table 4.17.** Prior remote-sensing studies of COD and $BOD_5$, summarizing the algorithm, study region, sensor, and period.

| Water quality parameters | Algorithm | Region | Instrument type | Estimated time scale | Year | Reference |
|---|---|---|---|---|---|---|
| COD | Quasi-Analytical Algorithm and Multivariable Linear Regression Models | China: Lake Taihu, Lake Qiandoahu, and the Yangtze River | Sentinel 3A/Ocean and Land Colour Instrument (OLCI) | 2010–2021 | 2023 | (Cai et al., 2023) |
| | Support vector regression (SVR) and one-dimensional convolutional neural networks (1D-CNN) | China: typical rivers flowing through Zhongshan | M600 Pro UAV (DJI, Shenzhen, China) equipped with the hyperspectral sensor, ZK-VNIR-FPG480 (Zhike, Beijing, China) | 2019.07–09; 2020.11-12 | 2022 | (Cai et al., 2022) |
| | Random forest (RF), Extreme gradient boosting (XGBoost), Categorical Boosting (CatBoost), and SVR | China: Dianshan Lake | Sentinel-2 | 2017–2023 | 2023 | (Dong et al., 2023) |

| Water quality parameters | Algorithm | Region | Instrument type | Estimated time scale | Year | Reference |
|---|---|---|---|---|---|---|
| | Convolutional long short-term memory (ConvLSTM) model | China: Dongping Lake, an impounded lake located in the Yellow River | Landsat 8 OLI | 2013–2018 | 2022 | (Zhang et al., 2022) |
| | XGBoost | China: Zhejiang Province | Sentinel-2 | 2019.01–2023.08 | 2024 | (Y. Zhao et al., 2024) |
| | RF | China: Poyang Lake | Sentinel-2 | 2021 | 2023 | (Zhao, 2023) |
| | Capsule Network and back-propagation neural network | China: Baiyangdian area | PSR-3500 Portable Hyperspectrometer | / | 2019 | (Deng et al., 2019) |
| $BOD_5$ | SVR, Radial-Basis Function (RBF) neural network, RF, XGBoost | China: Hong Kong SAR | Landsat 7 and Landsat 8 Collection 2 Level-2 | 2000.01–2020.12 | 2024 | (Zhou et al., 2024) |

### 4.3.2. Study area and width-stratified monitoring network

South Bohemia (Jihočeský kraj; 48.61–49.56°N, 13.78–15.46°E) covers approximately 10,058 km$^2$ in southern Czechia. Its dense drainage network is dominated by the Vltava, Lužnice, Otava, Malše, Nežárka, and Blanice river systems. These rivers cross forested uplands, intensively cultivated lowlands, and a long-established fish-pond landscape (Figure 4.18). Study 3 used the same monitoring network, the open CHMI database, and the channel-width screening procedure as Study 2 (Section 4.2.2). Of 120 fixed monitoring stations, 103 had measurable channels and were retained; 17 were excluded because dense canopy obscured the channel. The retained reaches were classified as W1 (< 10 m; n = 60; median = 4.9 m), W2 (10–30 m; n = 30; median = 15.3 m), and W3 (> 30 m; n = 13; median = 54.9 m), with an overall network median of 9.2 m. Width remained a fixed reach attribute throughout the study. The stratification was not intended to define hydraulically equivalent distinct river types. The classes represent increasing water-surface coverage within satellite pixels and support consistent assessment of land–water spectral mixing.

Channel widths were digitized on screen from the 2023–2024 Czech national orthophoto mosaic in EPSG:5514. The imagery has a ground sampling distance of 0.125 m, a reported planimetric accuracy of approximately 0.19–0.21 m, and a two-year update cycle. At each station, a bank-to-bank cross-section was drawn perpendicular to the locally mapped flow direction, with endpoints snapped to the visually identified banks. Where a pond, impoundment, bridge, or culvert prevented reliable measurement at the sampling location, the nearest unobstructed representative segment was measured, and the station was flagged. Widths across the 103 retained reaches ranged from 1.6 to 297 m, with an IQR of 4.0–16.2 m. Although narrow reaches were numerically dominant, they had fewer long monitoring records. This unequal temporal coverage was explicitly accounted for in subsequent trend and reconstruction analyses to avoid misinterpreting record length as a channel-width effect.

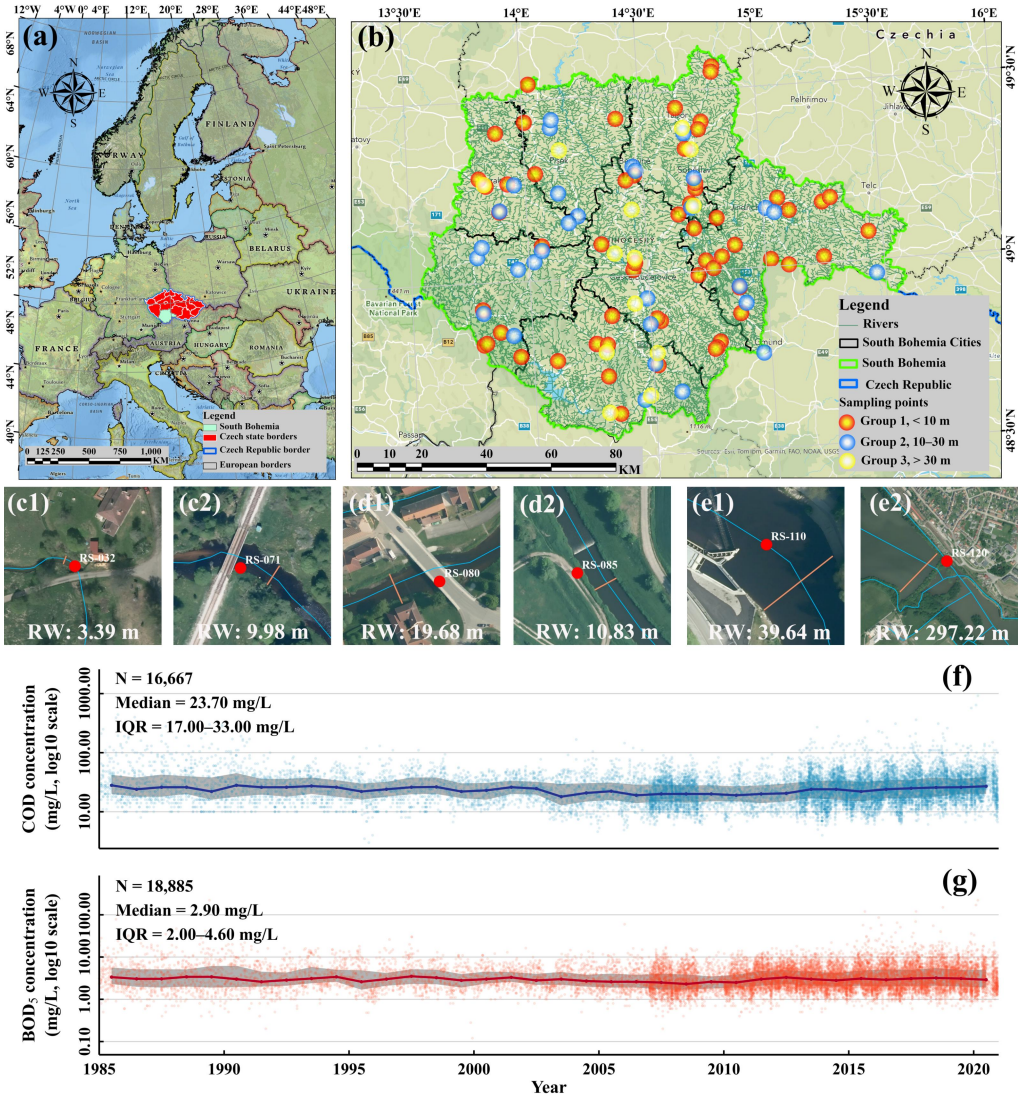



**Figure 4.18.** Study area, monitoring network, and organic-load setting in South Bohemia, Czech Republic. (a) Location within Europe; (b) the South Bohemian administrative districts and the 103 monitoring reaches classified into three river-width groups (≤10 m, 10–30 m, >30 m) from 0.125 m orthophotos; (c1, c2) representative narrow reaches (≤10 m), (d1, d2) intermediate reaches (10–30 m) and (e1, e2) a wide reach (>30 m); and (f, g) the 1985–2020 concentration distributions of COD and $BOD_5$ ($log_{10}$ scale, with yearly medians and interquartile bands).

### 4.3.3. In situ oxygen-demand concentrations

All available CHMI measurements of COD and $BOD_5$ collected from 2 January 1985 to 16 December 2020 were compiled. The dataset comprised 16,667 COD and 18,885 $BOD_5$ observations, with no missing or non-positive concentrations; median sample counts per reach were 115 and 149, respectively. Both distributions were strongly right-skewed (Figures 4.18f, g). Median concentrations (IQRs) were 23.7 mg $L^{-1}$ (17.0–33.0) for COD and 2.9 mg $L^{-1}$ (2.0–4.6) for $BOD_5$, while maxima reached 920 and 220 mg $L^{-1}$, respectively (Table 4.18). Concentrations were $log_{10}$-transformed before spectral, trend, and predictive analyses, including CV. This reduced skewness from 10.21 to 0.45 for COD and from 21.82 to 0.46 for $BOD_5$, reducing the influence of extreme observations. Descriptive statistics are reported on the original concentration scale.

**Table 4.18.** Summary statistics of the in-situ oxygen-demand concentrations for the South Bohemia monitoring dataset, 1985–2020.

| Statistic | COD | $BOD_5$ |
|---|---|---|
| Date range | 1985-01-02 to 2020-12-16 | 1985-01-02 to 2020-12-16 |
| N | 16,667 | 18,885 |
| Minimum (mg $L^{-1}$) | 3.00 | 0.12 |
| Q1 (mg $L^{-1}$) | 17.00 | 2.00 |
| Median (mg $L^{-1}$) | 23.70 | 2.90 |
| Mean (mg $L^{-1}$) | 28.02 | 3.87 |
| Q3 (mg $L^{-1}$) | 33.00 | 4.60 |
| Maximum (mg $L^{-1}$) | 920.00 | 220.00 |
| Standard deviation (mg $L^{-1}$) | 22.06 | 4.43 |
| Skewness (raw) | 10.21 | 21.82 |
| Skewness ($log_{10}$) | 0.45 | 0.46 |

### 4.3.4. Satellite archive and spatiotemporal matchups

All analyses used atmospherically corrected surface reflectance from five satellite missions represented by seven products (Table 4.19). Landsat 5 and 7 observations were obtained from the United States Geological Survey (USGS) Collection 2 Level-2 archive and processed using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS). Landsat 8 and 9 observations were processed using the Land Surface Reflectance Code (LaSRC) (Roy et al., 2014b; Vermote et al., 2016; Wulder et al., 2019). Sentinel-2A, Sentinel-2B, and Sentinel-2C were represented by native Level-2A surface-reflectance products (Drusch et al., 2012). The HLS archive included the Landsat-derived 30 m product (L30) and Sentinel-2-derived 30 m product (S30). Both products undergo LaSRC atmospheric correction and bidirectional reflectance distribution function (BRDF) normalization, while S30 is additionally adjusted to Landsat spectral bandpasses (Claverie et al., 2018). These processors were developed primarily for terrestrial applications and may retain aerosol and path-radiance errors over dark, narrow waters (Maciel et al., 2023). They were retained to permit consistent evaluation of operational products. A water-specific atmospheric correction might reduce radiometric errors but cannot resolve sub-pixel geometry or adjacency contamination.

The analytical unit was therefore a fixed, width-constrained river-reach object rather than an individual pixel (Kuhn et al., 2019; Vanhellemont and Ruddick, 2018). Where channel width approached or fell below pixel size, the recorded signal included contributions from banks, riparian vegetation, and adjacent fields. Atmospheric scattering and the sensor point-spread function can transfer these bright land signals into nominal water pixels, with contamination increasing as the water fraction decreases (Bulgarelli and Zibordi, 2018). Three safeguards reduced but could not eliminate this effect: restricting statistical summaries to the channel object, screening non-water contributions using the MNDWI and applying an effective-pixel quality gate.

For each object-date, the feature set included the mean, median, selected percentiles, and SD of each visible-to-shortwave-infrared band. It also included

MNDWI, NDVI, NDTI, NDWI and selected band ratios. Red-edge features were included for native Sentinel-2 and HLS S30. Object-dates with insufficient valid pixels were excluded by the effective-pixel gate. Objects exceeding sensor-specific coefficient-of-variation thresholds in both the green and near-infrared bands were flagged as internally heterogeneous, and this flag was used in the sensitivity analysis. The gated archive contained 178,286 object-level observations for 1985–2025. After observations containing negative reflectance values were excluded, 175,591 remained. Each satellite product was matched to the nearest in situ sample within a ±5-day window. The final modelling datasets comprised 17,200 COD and 17,658 $BOD_5$ feature–concentration pairs, with $log_{10}$-transformed concentration used as the response variable.

HLS observations are harmonized reprocessings of Landsat and Sentinel-2 acquisitions rather than independent overpasses. Native and HLS forms were both retained because the spatial-resolution comparison required native 10 m Sentinel-2 observations alongside their 30 m HLS equivalents. Among the 17,207 raw COD matchup rows, 9,471 (55%) occurred on reach-dates represented by multiple products. These observations corresponded to 11,897 unique reach-dates, leaving 5,310 additional product rows beyond the first observation (31%). For $BOD_5$, the corresponding values were 17,665 raw rows, 9,155 rows on overlapping reach-dates (52%), 12,543 unique reach-dates and 5,122 additional product rows (29%). Landsat 9 contributed to the full archive but not to the matchup datasets because in situ monitoring ended in 2020, before the mission launched in 2021. Consequently, six of the seven products contributed to model evaluation.

This product overlap creates a specific leakage risk under observation-level random splitting because native and harmonized versions of the same acquisition may be assigned to different folds. LOSO spatial CV assigns all observations from a given reach to the same fold, thereby preventing this form of same-reach duplication leakage. However, LOSO validation does not eliminate potential spatial dependence among neighbouring reaches.

**Table 4.19.** The seven surface-reflectance products harmonized in the object-level framework, with the number of gated reach-object observations each contributed to the 1985–2025 archive.

| Product | Mission | Platform units | Operator | Native resolution (m) | Key bands | Observations | Period in archive |
|---|---|---|---|---|---|---|---|
| HLS S30 | Sentinel-2 (harmonized) | S30 | NASA | 30 | harmonized Sentinel-2 | 49,009 | 2015–2025 |
| Sentinel-2 MSI L2A | Sentinel-2 | S2A, S2B, S2C | ESA / EU Copernicus | 10 / 20 / 60 | VIS–NIR–red-edge–SWIR | 45,295 | 2017–2025 |
| Landsat 5 TM | Landsat 5 | — | USGS / NASA | 30 | VIS–NIR–SWIR (+TIR) | 22,010 | 1985–2011 |
| Landsat 7 ETM+ | Landsat 7 | — | USGS / NASA | 30 | VIS–NIR–SWIR (+pan) | 21,884 | 1999–2022 |
| HLS L30 | Landsat (harmonized) | L30 | NASA | 30 | harmonized Landsat | 19,792 | 2013–2025 |
| Landsat 8 OLI | Landsat 8 | — | USGS / NASA | 30 | VIS–NIR–SWIR | 15,298 | 2013–2025 |
| Landsat 9 OLI-2 | Landsat 9 | — | USGS / NASA | 30 | VIS–NIR–SWIR | 4,998 | 2021–2025 |
| Total | five missions | nine platform labels | | | | 178,286 | 1985–2025 |

Notes: The archive contains nine distinct platform labels, which correspond to seven surface-reflectance products and five satellite missions, because the two HLS entries are harmonized re-processings of Landsat and Sentinel-2 acquisitions rather than independent missions, and the Sentinel-2 record combines three platform units. Landsat 7 acquisitions after 2003 are affected by the scan-line corrector failure. Observation counts are gated to reach-object observations before removal of negative-reflectance retrievals.

### 4.3.5. Two prediction pathways and their spatial validation

**Optical retrieval**. Direct optical retrieval was tested using RF regression implemented in the ranger package with 400 trees (Breiman, 2001). $Log_{10}$-transformed COD and $BOD_5$ concentrations were regressed against the object-level spectral features. Two validation schemes were applied to identical observations and predictors: conventional observation-level random 10-fold CV and LOSO CV. Under LOSO validation, all observations from the held-out reach were excluded from model training (Milà et al., 2022b; Roberts et al., 2017). The difference between random and LOSO performance was used to quantify inflation due to repeated-site information and product overlaps in random folds (Karasiak et al., 2022). It was not interpreted as a complete measure of spatial autocorrelation because dependence may remain among neighboring reaches. Models were evaluated for the complete dataset, within each width class, and separately for native 10 m Sentinel-2 and 30 m Landsat/HLS observations. Temporal tracking at held-out reaches was assessed using reach-specific Spearman correlations between LOSO predictions and observations. The proportion of reaches with correlations exceeding 0.3 was also calculated.

**Catchment-based prediction**. As an alternative to direct optical retrieval, annual concentrations were predicted using RF regression and spatially explicit catchment attributes derived from remote-sensing and ancillary geospatial datasets in GEE. Predictor sources and temporal coverage are listed in Table 4.20 (Muñoz-Sabater et al., 2021; Zanaga et al., 2022). Land-cover predictors comprised CORINE Land Cover fractions for 1990, 2000, 2006, 2012, and 2018, together with 10 m European Space Agency (ESA) WorldCover fractions. Terrain was represented by the Copernicus Digital Elevation Model GLO-30 and a derived slope index. Additional predictors included WorldPop population counts and ERA5-Land annual mean temperature and total precipitation for 1985–2025.

Point-source pressure was obtained from the European Environment Agency inventory compiled under the Urban Waste Water Treatment Directive. The inventory

contained 607 Czech wastewater-treatment plants, but only nine occurred within the assigned local hydrological units. Simple within-unit plant counts were therefore insufficient to represent point-source pressure. Instead, pressure was summarized for each reach using distance-weighted plant design capacity, distance-weighted discharged organic load, and distance to the nearest plant. The national wastewater connection rate reported by Eurostat served as the time-varying component. This series begins in 2000 and was extrapolated backwards to 1985, with all extrapolated years flagged.

The spatial support of the catchment predictors was defined explicitly because it determines their hydrological interpretation. Each reach was assigned to the fourth-order hydrological unit (povodí IV. řádu) of the Czech Digital Base of Water Management Data, which contains the monitoring station, via a point-in-polygon operation. These polygons represent local hydrological units rather than cumulative upstream drainage areas. Their median area was 12.2 $km^2$, with an IQR of 5.5–21.0 $km^2$ and an approximate range of 0.002–259 $km^2$. Median areas were similar among the W1, W2, and W3 reaches at 11.1, 11.7, and 13.5 $km^2$, respectively (Table 4.21). For headwater reaches, the assigned unit more closely approximates the contributing area. For wider downstream reaches, however, it represents only a non-cumulative portion of the area supplying the sampled water. The predictors should therefore be interpreted as descriptors of local catchment context. Station RS-100 was excluded from catchment modelling because it fell within an anomalous 0.0017 $km^2$ sliver polygon that was smaller than one CORINE pixel and lacked an ERA5-Land value. Consequently, catchment-based analyses used 102 reaches, whereas the optical and trend analyses retained all 103 reaches.

Four engineered descriptors were tested as a hydrological augmentation using the same LOSO folds. The first was annual precipitation multiplied by hydrological-unit area, representing incident precipitation volume rather than observed runoff or water yield. The second was a relative dilution-potential index, calculated by dividing this precipitation volume by the distance-weighted design capacity of potentially

contributing treatment plants. The remaining descriptors were reach-specific annual precipitation and temperature anomalies relative to their 1985–2025 means. Because area, precipitation, and temperature were already included in the baseline predictor set, this comparison evaluated the additional value of their interactions and temporal departures rather than introducing independent data sources.

Models were fitted at the site-year scale using annual median concentrations, comprising 1,392 COD and 1,578 $BOD_5$ site-years. Predictive performance was assessed using LOSO CV, a temporal hold-out trained through 2010 and evaluated on subsequent years, and width-stratified analyses. The marginal contribution of point-source predictors was assessed by refitting the models after removing these terms. All preprocessing, including predictor standardization, was estimated separately within each training fold to prevent information leakage.

**Table 4.20.** Satellite-observable catchment predictors and point-source-pressure variables used in the catchment-attribute model.

| Group | Variables | Source | Native resolution | Period / epochs |
|---|---|---|---|---|
| Land cover (CORINE) | Urban, agricultural, forest, wetland, water fractions | CORINE Land Cover | 100 m | 1990, 2000, 2006, 2012, 2018 |
| Land cover (WorldCover) | Class fractions (tree, cropland, built-up, water, …) | ESA WorldCover | 10 m | 2020–2021 |
| Terrain | Elevation; terrain-slope index | Copernicus GLO-30 DEM | 30 m | static |
| Population | Mean population count per 100 m pixel (× 100 ≈ persons km$^{-2}$) | WorldPop GP/100m/pop | 100 m | 2020 |
| Climate | Annual mean temperature, annual total precipitation | ERA5-Land monthly aggregates | ~9 km (extracted at 5 km) | 1985–2025 |
| Hydrological descriptors (tested; Table 5.10) | Annual catchment water yield; dilution-capacity index; precipitation and temperature anomalies | derived from ERA5-Land, DIBAVOD, UWWTD | — | 1985–2025 |
| Point source (static) | Distance-weighted design capacity (p.e.), discharged organic load, nearest-plant distance | EEA UWWTD (607 CZ plants) | point | 2021–2023 |
| Point source (temporal) | National population connected to waste-water treatment | Eurostat (env_ww_con) | national | 2000–2023 (back-extrapolated to 1985) |
| Catchment | Local fourth-order hydrological unit and its area (not cumulative upstream area) | Czech DIBAVOD | vector | static |
| Channel | Bank-to-bank width, width class | 0.125 m orthophoto | reach | 2023–2024 |

**Table 4.21.** Channel geometry, size of the local fourth-order contributing unit, and dilution capacity by width class.

| Width class | n | Median channel width (m) | Median unit area (km²) | Unit area IQR (km²) | Median annual precipitation (mm) | Median annual water yield ($10^6$ m³) | Median dilution capacity (m³ p.e.$^{-1}$) | Median cropland fraction | Median elevation (m) |
|---|---|---|---|---|---|---|---|---|---|
| W1 (≤ 10 m) | 60 | 4.9 | 11.1 | 5.6–20.6 | 803 | 9.4 | 577 | 0.29 | 486 |
| W2 (10–30 m) | 30 | 15.3 | 11.7 | 3.9–20.9 | 811 | 10.5 | 567 | 0.27 | 479 |
| W3 (> 30 m) | 13 | 54.9 | 13.5 | 9.1–24.5 | 798 | 10.6 | 469 | 0.20 | 441 |
| All | 103 | 9.2 | 12.2 | 5.5–21.0 | 807 | 10.4 | 554 | 0.26 | 476 |

Notes: The unit is the local DIBAVOD fourth-order hydrological unit that contains each station, not the cumulative upstream drainage area, which is why its size barely increases over an order of magnitude in channel width. Annual water yield is annual precipitation multiplied by unit area; dilution capacity is that yield divided by the distance-weighted design capacity of upstream treatment plants in population equivalents. Both are proxies used only in the augmentation experiment of Table 5.10.

### 4.3.6. Model evaluation

Because no single goodness-of-fit statistic fully captures correlation, variability, systematic bias, and residual error, model performance was evaluated using a complementary set of metrics widely applied in hydrological modelling. These included the Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970), the original and revised Kling–Gupta efficiencies (KGE and KGE′; Gupta et al., 2009; Kling et al., 2012), percent bias (PBIAS), the ratio of the root-mean-square error to the SD of the observations (RSR), and RPD. Let $o_i$ and $p_i$ denote the observed and predicted values for observation (i), respectively; $\bar{o}$ and $\bar{p}$ their arithmetic means; $s_o$ and $s_p$ their SDs; (r) their Pearson correlation coefficient; and (n) the number of observations. The RMSE was calculated as

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i-o_i)^2}$$

NSE was calculated as

$$\mathrm{NSE}=1-\frac{\sum_{i=1}^{n}(p_i-o_i)^2}{\sum_{i=1}^{n}(o_i-\bar{o})^2}$$

The original KGE was calculated as

$$\mathrm{KGE}=1-\sqrt{(r-1)^2+(\alpha-1)^2+(\beta-1)^2}$$

where α represents the variability ratio and β represents the bias ratio:

$$\alpha=\frac{s_p}{s_o}$$

$$\beta=\frac{\bar{p}}{\bar{o}}$$

The revised KGE′ replaced the standard-deviation ratio with the ratio of coefficients of variation:

$$\mathrm{KGE}'=1-\sqrt{(r-1)^2+(\gamma-1)^2+(\beta-1)^2}$$

where

$$\gamma = \frac{s_p/\bar{p}}{s_o/\bar{o}}$$

PBIAS was calculated as

$$\text{PBIAS} = 100 \times \frac{\sum_{i=1}^{n} (o_i - p_i)}{\sum_{i=1}^{n} o_i}$$

Under this convention, positive PBIAS values indicate underprediction, whereas negative values indicate overprediction. RSR and RPD were calculated as

$$\text{RSR} = \frac{\text{RMSE}}{s_o}$$

$$\text{RPD} = \frac{s_o}{\text{RMSE}} = \frac{1}{\text{RSR}}$$

Nash-Sutcliffe efficiency (NSE) is identical to the predictive coefficient of determination only when $R^2$ is defined as $1 - \text{SSE/SST}$ using the same evaluation dataset. It is not equivalent to the squared Pearson correlation coefficient. To avoid ambiguity, NSE and $R^2$ should not be used interchangeably unless this definition of $R^2$ is explicitly adopted. Similarly, RSR and RPD are reciprocal expressions of the same normalized prediction error and should not be treated as independent evidence of model performance.

Metrics were reported on the $\log_{10}$ scale used for model fitting and the original concentration scale after back-transformation. Log-scale metrics describe performance on the modelling scale, whereas original-scale metrics provide a physically interpretable assessment of concentration errors. PBIAS and the ratio components of KGE and KGE′ were interpreted on the original concentration scale.

Two considerations guided metric interpretation. Firstly, KGE does not share the same zero benchmark as NSE: predictions based on the observed mean yield a reference KGE of approximately −0.41 rather than zero (Knoben et al., 2019). Secondly, the frequently cited NSE > 0.50 criterion was developed for continuous, long-term monthly watershed simulations (Moriasi et al., 2007). It was not treated as a universal threshold for spatial prediction at previously unseen reaches.

### 4.3.7. Trends, reconstruction and 2050 scenario analysis

Trends were quantified following Section 4.2.7. Annual median concentrations were used to calculate reach-specific Theil–Sen slopes, and trend significance was assessed with the Mann–Kendall test at $p < 0.05$. The regional trajectory was derived from a fixed panel of 15 reaches with records spanning at least 30 years: one W1, nine W2, and five W3 reaches. Associations between trend slopes and width class were evaluated with record span included as a covariate. This adjustment reduced confounding from the shorter, more recent records available for many narrow reaches, but it was not interpreted as eliminating all sampling bias.

The final catchment model was applied to every year from 1985 to 2025. This period included 2021–2025, for which no in-situ COD or $BOD_5$ observations were available. The reconstruction was therefore entirely driver-based. Its interpretation was anchored to the LOSO and temporal hold-out validation results reported above. Reconstructed values were treated as spatially transferable annual estimates rather than observations or locally validated gap-filled time series.

For the 2050 sensitivity analysis, land-use scenarios retained 2018 land cover or multiplied cropland fractions by 1.15 (intensification) or 0.85 (extensification). The remaining land-cover fractions were renormalized to sum to one. These scenarios were combined with temperature increases of 1, 2, and 3°C, along with the prescribed proportional adjustments to precipitation. Point-source predictors were held at baseline values. Responses were expressed as percentage differences from the current-driver baseline and summarized by reach and width class. These results represent conditional model sensitivities, not forecasts. Scenario likelihood was not assessed; random forests cannot reliably extrapolate response relationships beyond the training domain, and the model captured persistent spatial contrasts more strongly than interannual variability within individual reaches.

# Chapter V

## Results and Discussion

# Content

5.1. Study 1: Four decades of dissolved organic carbon dynamics in Lake Taihu using comprehensive Landsat-5/7/8 observations and advanced machine learning models with investigation on driving factors ..... 112

5.1.1. Model performance and selection of XGBoost .....113

5.1.2. Spatiotemporal variability of DOC concentrations .....118

5.1.3. DOC storage changes in Taihu ..... 124

5.1.4. Driver attribution of DOC variability ..... 130

5.1.5. Carbon-cycle implications ..... 134

5.2. Study 2: Spatial cross-validation reveals a double inseparability boundary for satellite monitoring of narrow-river nutrients ..... 136

5.2.1. A long-term object archive, its sub-pixel geometry and the nutrient setting .....137

5.2.2. The optical inseparability boundary .....141

5.2.3. Validation design governs reported skill and reveals limited spatial transferability .....150

5.2.4. A regional long-term recovery with no detectable width dependence .....152

5.2.5. The attributional inseparability boundary ..... 153

5.2.6. The double inseparability boundary and the protocol it implies .....157

5.2.7. What remains feasible and the limits of the evidence .....159

5.3. Study 3: Catchment attributes outperform reflectance in predicting oxygen-demand concentrations of narrow headwater rivers: a spatially validated four-decade reconstruction ..... 161

5.3.1. The optical non-retrievability boundary for COD and $BOD_5$ .....162
5.3.2. Long-term trends and catchment-attribute prediction .....168
5.3.3. Hydrological basis and limits of catchment-attribute prediction ..... 175
5.3.4. Scenario sensitivity to 2050 ..... 178
5.3.5. Width-divergent trajectories, management implications and evidential limits .....181
5.4. Cross-study synthesis: A retrievability spectrum and the double inseparability boundary .....183

# 5. Results and Discussion

In accordance with the structure adopted throughout this dissertation, the results and their discussion are presented together for each study, followed by a cross-study synthesis.

## 5.1. Study 1: Four decades of dissolved organic carbon dynamics in Lake Taihu using comprehensive Landsat-5/7/8 observations and advanced machine learning models with investigation on driving factors

This section presents the principal findings of Study 1, focusing on the four-decade reconstruction and attribution of DOC dynamics in Lake Taihu using Landsat observations and ML models. This work has been published in the following peer-reviewed article:

**Shen, F.**, Li, J., Wang, T., Li, N., Yang, J., Xu, M., Yan, Y., He, L., Guo, Y., Chen, Z., 2026. Four decades of dissolved organic carbon dynamics in Lake Taihu using comprehensive Landsat-5/7/8 observations and advanced machine learning models with investigation on driving factors. Journal of Hydrology 676, 135696. https://doi.org/10.1016/j.jhydrol.2026.135696

### 5.1.1. Model performance and selection of XGBoost

Hyperparameters were tuned using 10-fold CV within the training set (n = 140). Training performance was reported for diagnostic purposes, while model selection was based primarily on performance on the held-out validation set (n = 60). On the training set, RF and XGBoost both achieved $R^2$ = 0.96, with RMSE values of 0.17 and 0.15 mg/L, respectively (Figure 5.1; Table 5.1). Compared with RF, XGBoost had slightly lower MAE (0.12 vs. 0.13 mg/L) and MAPE (2.73% vs. 2.91%) and a higher RPD (4.52 vs. 3.97). SVM showed the weakest training fit ($R^2$ = 0.64, RMSE = 0.40 mg/L), whereas MBPNN achieved intermediate performance ($R^2$ = 0.84, RMSE = 0.27 mg/L). On the validation set, XGBoost achieved $R^2$ = 0.73, RMSE = 0.51 mg/L, MAPE = 8.66%, and RPD = 1.70. RF had a slightly higher $R^2$ (0.78) but also higher RMSE (0.54 mg/L) and MAPE (9.09%), whereas MBPNN and SVM showed substantially weaker validation performance than training performance. Residuals ranged from −1.18 to 1.12 mg/L for XGBoost and from −1.19 to 1.40 mg/L for RF, indicating a narrower residual range for XGBoost (Figure 5.2). All models tended to overestimate DOCc at low concentrations (< 3.00 mg/L) and underestimate it at high concentrations (> 6.00 mg/L). Based on its lower validation errors and narrower residual range, XGBoost was selected for the four-decade reconstruction of DOC dynamics in Lake Taihu.

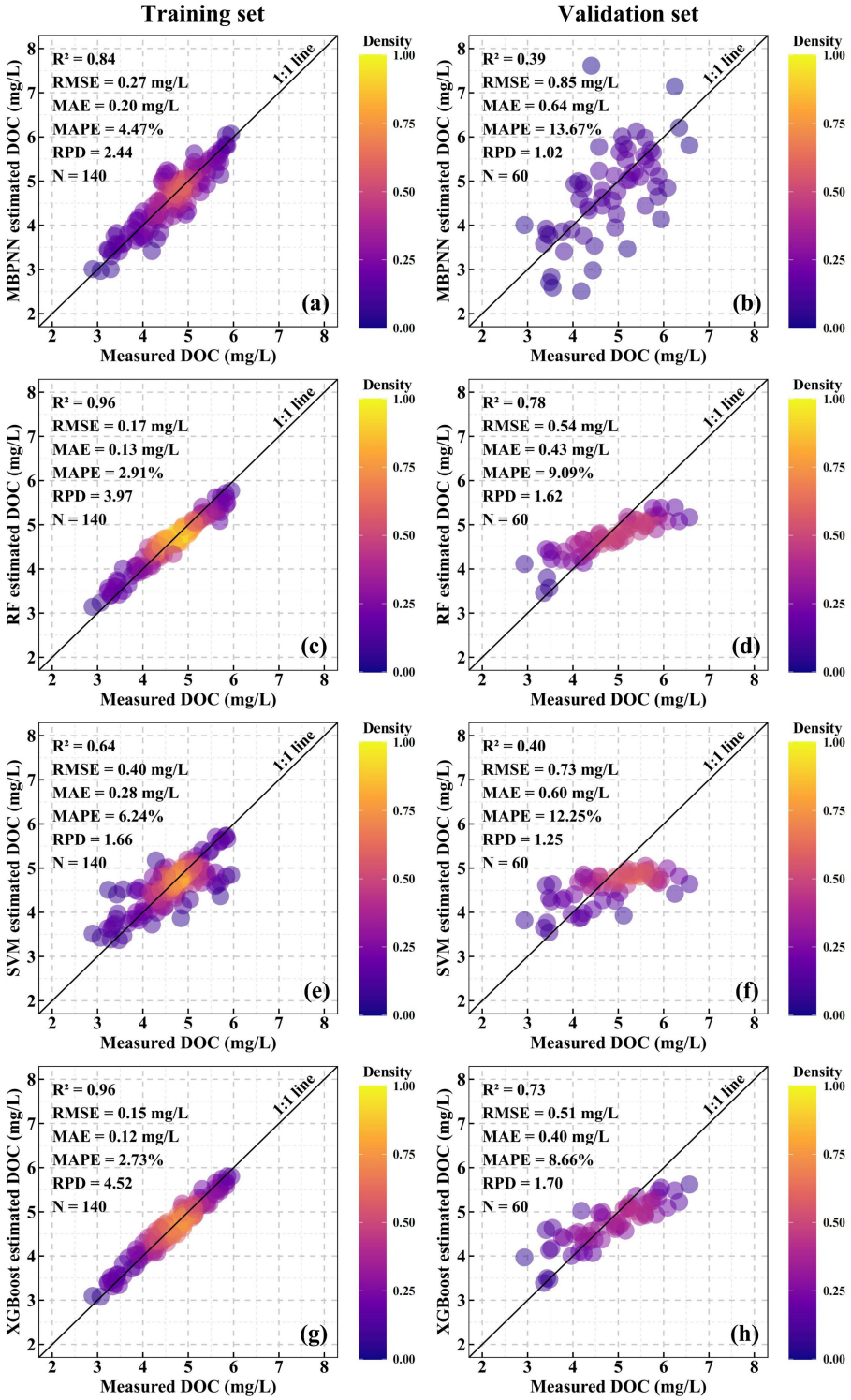

Training set
Validation set
MBPNN estimated DOC (mg/L)
RF estimated DOC (mg/L)
SVM estimated DOC (mg/L)
XGBoost estimated DOC (mg/L)
Measured DOC (mg/L)
Density
1:1 line
R² = 0.84
RMSE = 0.27 mg/L
MAE = 0.20 mg/L
MAPE = 4.47%
RPD = 2.44
N = 140
(a)
R² = 0.39
RMSE = 0.85 mg/L
MAE = 0.64 mg/L
MAPE = 13.67%
RPD = 1.02
N = 60
(b)
R² = 0.96
RMSE = 0.17 mg/L
MAE = 0.13 mg/L
MAPE = 2.91%
RPD = 3.97
N = 140
(c)
R² = 0.78
RMSE = 0.54 mg/L
MAE = 0.43 mg/L
MAPE = 9.09%
RPD = 1.62
N = 60
(d)
R² = 0.64
RMSE = 0.40 mg/L
MAE = 0.28 mg/L
MAPE = 6.24%
RPD = 1.66
N = 140
(e)
R² = 0.40
RMSE = 0.73 mg/L
MAE = 0.60 mg/L
MAPE = 12.25%
RPD = 1.25
N = 60
(f)
R² = 0.96
RMSE = 0.15 mg/L
MAE = 0.12 mg/L
MAPE = 2.73%
RPD = 4.52
N = 140
(g)
R² = 0.73
RMSE = 0.51 mg/L
MAE = 0.40 mg/L
MAPE = 8.66%
RPD = 1.70
N = 60
(h)

**Figure 5.1.** The scatterplots of ML estimation of Taihu DOC using Landsat $R_{rs}(\lambda)$. (a, b) MBPNN, (c, d) RF, (e, f) SVM, and (g, h) XGBoost. The training set had 140 samples, and the validation set had 60 samples.

**Table 5.1.** Performance evaluation of machine learning algorithms.

| Algorithm | Dataset | $R^2$ | RMSE (mg/L) | MAE (mg/L) | MAPE (%) | SD | RPD |
|---|---|---|---|---|---|---|---|
| MBPNN | Training | 0.84 | 0.27 | 0.20 | 4.47 | 0.27 | 2.44 |
| | Validation | 0.39 | 0.85 | 0.64 | 13.67 | 0.85 | 1.02 |
| RF | Training | 0.96 | 0.17 | 0.13 | 2.91 | 0.17 | 3.97 |
| | Validation | 0.78 | 0.54 | 0.43 | 9.09 | 0.53 | 1.62 |
| SVM | Training | 0.64 | 0.40 | 0.28 | 6.24 | 0.40 | 1.66 |
| | Validation | 0.40 | 0.73 | 0.60 | 12.25 | 0.69 | 1.25 |
| XGBoost | Training | 0.96 | 0.15 | 0.12 | 2.73 | 0.15 | 4.52 |
| | Validation | 0.73 | 0.51 | 0.40 | 8.66 | 0.51 | 1.70 |

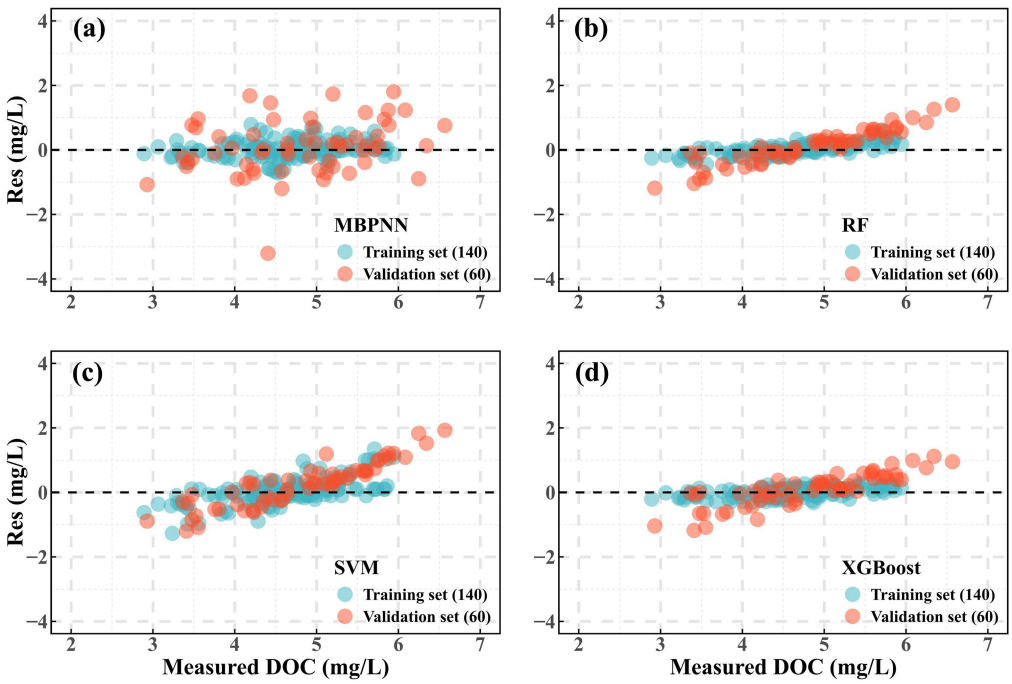



**Figure 5.2.** Residuals (Res) of measured and estimated DOC across four ML models. (a) MBPNN, (b) RF, (c) SVM, (d) XGBoost. Training set: 140, validation set: 60.

The selection of XGBoost was supported by its validation errors, while its capacity to represent nonlinear relationships is relevant to Taihu's optical complexity. Episodic algal blooms and wind-driven sediment resuspension can affect both DOC dynamics and optical conditions in the lake (Ding et al., 2018; Huang et al., 2017; Ye et al., 2015). Taihu's shallow, wind-exposed morphology promotes wave- and current-driven resuspension, rapidly altering suspended solids and turbidity (Ding et al., 2018; Li et al., 2017; Qin et al., 2004). These processes can enhance sediment–water exchange and redistribute river- and bloom-derived organic matter, contributing to short-term DOC fluctuations and spatial heterogeneity (Huang et al., 2017; Qin et al., 2004). Previous studies reported weak DOC–CDOM correlation ($r = 0.32$, $p < 0.01$) and limited explanatory power for blue–green band ratios (51.84% of DOC variability), illustrating the challenges posed by phytoplankton and mineral particles (D. Liu et al., 2021; Wang et al., 2018). XGBoost combines sequential tree fitting with explicit regularization to capture nonlinear relationships and interactions while limiting model complexity (Z. Zhang et al., 2024). Its validation RMSE (0.51 mg/L) and MAPE (8.66%) were slightly lower than those of RF (0.54 mg/L and 9.09%), although its $R^2$ was lower (0.73 vs. 0.78). These modest differences support its selection for this dataset but do not establish general superiority under weak DOC–CDOM coupling.

RF achieved similar validation performance but showed larger errors at high DOC concentrations and greater variability in 10-fold CV. Its residual range (−1.19 to 1.40 mg/L) was also wider than that of XGBoost (−1.18 to 1.12 mg/L). MBPNN and SVM exhibited larger training–validation discrepancies and less stable predictions at high concentrations, with training $R^2$ values of 0.84 and 0.64 and RMSE values of 0.27 and 0.40 mg/L, respectively. Rapid bloom dynamics relative to Landsat revisit intervals and within-scene heterogeneity may weaken satellite–field correspondence (Liu et al., 2024; H. Liu et al., 2021), while SVM performance also depends on kernel specification (Li et al., 2021). Larger residuals coincided with bloom onset or decay, sediment resuspension, and longer gaps between cloud-free scenes; XGBoost reduced, but did not eliminate, the concentration-dependent biases shared by all models. As a qualitative

consistency check, elevated DOC estimates in the 2009 black-bloom region were consistent with the MODIS anomaly detected on 11 May 2009 and the Landsat reconstruction reported by H. Liu et al. (2021). This agreement supports the plausibility of the spatial pattern but does not independently validate DOC concentrations under extreme optical conditions. MODIS-referenced harmonization was used to improve cross-sensor consistency for the four-decade reconstruction.

The tendency to overestimate low DOCc (< 3.00 mg/L) and underestimate high DOCc (> 6.00 mg/L) indicates compression of the predicted concentration range and warrants caution when interpreting reconstructed extremes. This pattern may partly reflect the limited representation of extreme concentrations following matchup selection (±7 days) and IQR screening, as well as interference with CDOM-related spectral signals during intense blooms or resuspension. Because DOC is inferred indirectly through its optically active fraction, increasing DOC need not produce a proportional change in reflectance. XGBoost attenuated but did not eliminate this bias, and the modest interannual amplitude of the lake-wide DOCc series should be interpreted with potential range compression in mind. Four-decade interpretation focused on zone- and lake-scale patterns rather than individual pixels or dates, recognizing that spatial aggregation can reduce random errors but cannot necessarily remove systematic bias. Convergent spatial patterns from XGBoost and RF provide evidence of consistency between the two ensemble methods, although their shared predictors and calibration data mean that common retrieval biases cannot be excluded.

## 5.1.2. Spatiotemporal variability of DOC concentrations

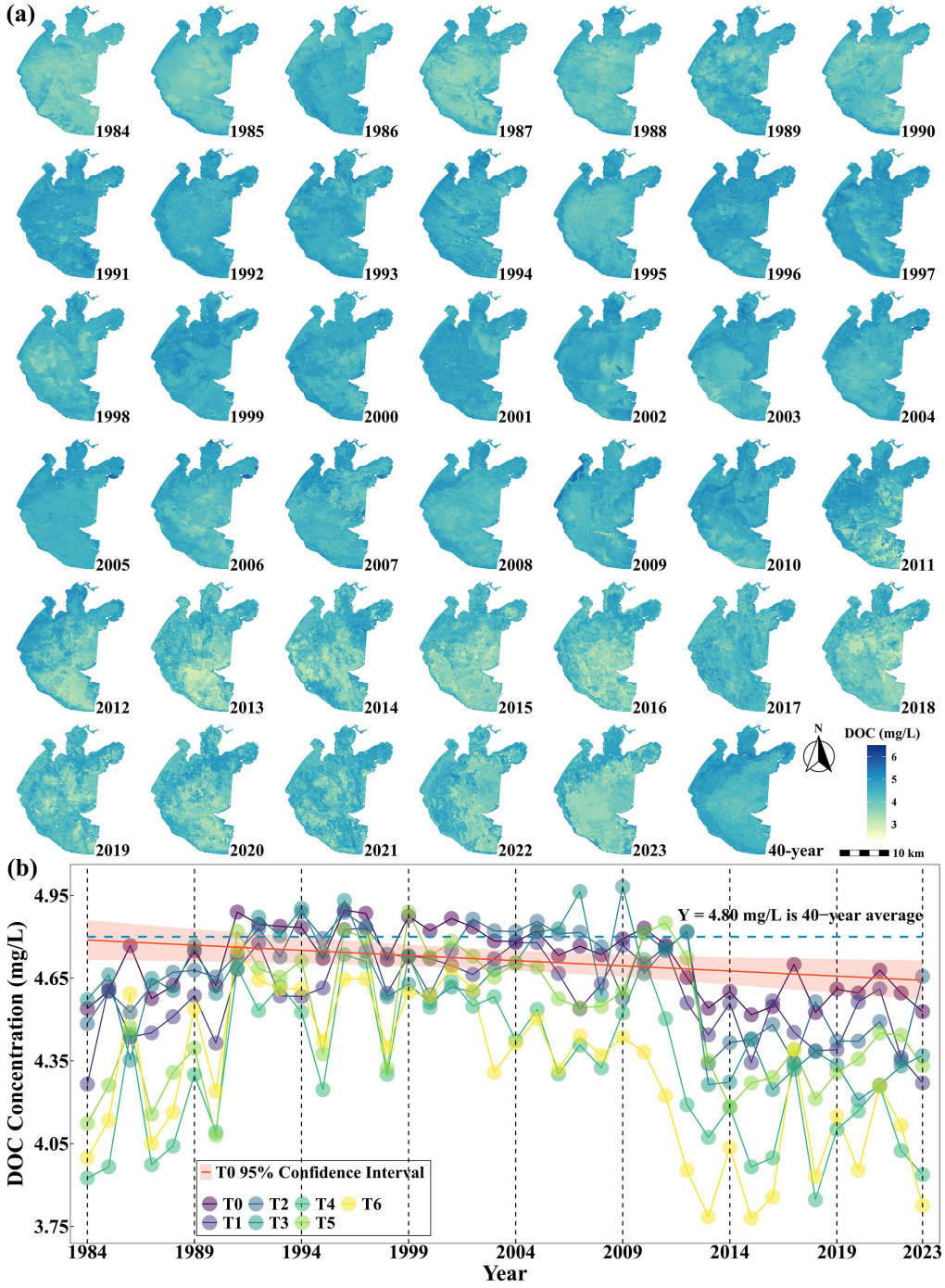


**Figure 5.3.** Spatiotemporal dynamics of DOC concentrations in Lake Taihu (1984−2023). (a) annual DOC distribution maps of Lake Taihu from 1984 to 2023, (b) temporal trends of DOC concentrations across different sub-regions (T0−T6) with the 40-year mean and 95% confidence interval.

From 1984 to 2023, DOCc in Taihu displayed pronounced spatial and temporal heterogeneity. To quantify within-lake heterogeneity explicitly, annual pixel-level CoV ($CoV_{pixel}$) and a scale-consistent block-based CoV ($CoV_{block}$; 3–10 km) were computed, together indicating persistent sub-lake gradients throughout the record (Figure 5.4; Table 5.3). On the whole-lake scale (T0), annual mean DOCc was 4.72 ± 0.12 mg/L (mean ± SD) across the 40-year record, representing a difference of 0.38 mg/L over the study period (Figure 5.3). The spatial change in DOCc between 1984 and 2023 ($\Delta DOC_c = DOC_{c2023} - DOC_{c1984}$) ranged from -2.46 to +2.50 mg/L across the lake. Among the six subregions, T1−T3 and T5 exhibited higher four-decade mean DOCc (4.51 ± 0.23–4.66 ± 0.17 mg/L), whereas T4 and T6 were lower, at approximately 4.31–4.32 mg/L (Table 5.2). Within-zone spatial dispersion was modest (Spatial CoV approximately 0.08–0.10), indicating persistent but relatively stable sub-lake contrasts.

**Table 5.2.** Zone-level descriptive statistics of annual Landsat-derived DOCc in Lake Taihu (T0−T6), 1984−2023.

| Zone | Min–Max (mg/L) | Mean ± SD (mg/L) | Median ($Q_{25}$–$Q_{75}$) (mg/L) | Spatial SD (mg/L) | Spatial CoV | N (years) | Period |
|---|---|---|---|---|---|---|---|
| T0 (Whole lake) | 4.52–4.90 | 4.72 ± 0.12 | 4.73 (4.60–4.82) | 0.46 | 0.10 | 40 | 1984–2023 |
| T1 (Gonghu Bay) | 4.27–4.83 | 4.59 ± 0.15 | 4.59 (4.45–4.71) | 0.38 | 0.08 | 40 | 1984–2023 |
| T2 (Meiliang Bay) | 4.34–4.89 | 4.66 ± 0.17 | 4.70 (4.52–4.81) | 0.37 | 0.08 | 40 | 1984–2023 |
| T3 (Zhushan Bay) | 4.21–4.98 | 4.60 ± 0.22 | 4.62 (4.38–4.76) | 0.35 | 0.08 | 40 | 1984–2023 |
| T4 (Central Lake) | 3.85–4.74 | 4.32 ± 0.27 | 4.31 (4.09–4.54) | 0.4 | 0.09 | 40 | 1984–2023 |
| T5 (Northwest Lake) | 4.08–4.89 | 4.51 ± 0.23 | 4.51 (4.32–4.72) | 0.38 | 0.09 | 40 | 1984–2023 |
| T6 (Southwest Lake) | 3.78–4.75 | 4.31 ± 0.28 | 4.38 (4.10–4.59) | 0.4 | 0.09 | 40 | 1984–2023 |

Notes: Statistics are summarized from the annual zone-mean DOCc time series (N = 40 years) derived from 30 m Landsat DOCc maps for 1984–2023. “Min–Max”, “Mean ± SD”, and "Median ($Q_{25}$–$Q_{75}$)" describe interannual variability of zone-mean DOCc, whereas “Spatial SD” and “Spatial CoV” quantify within-zone spatial heterogeneity from pixel-level DOCc fields. The East Bay and the East Coast were excluded due to aquatic vegetation effects on water reflectance.

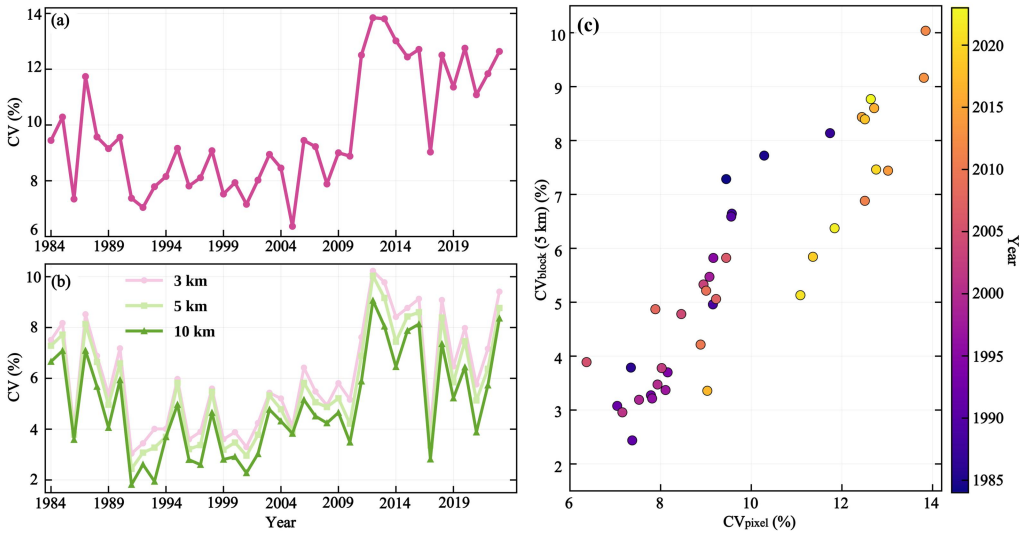



**Figure 5.4.** Within-lake spatial heterogeneity of annual Landsat-derived DOC products (1984–2023) using pixel-level and block-based coefficients of variation (CoV). (a) Annual pixel-level spatial CoV for DOC concentration maps (DOCc; mg/L), computed across all water pixels as $CoV_{pixel}$ = SD/Mean. (b) Block-based CoV ($CoV_{block}$) after aggregating annual DOCc maps to regular blocks (3 km, 5 km, 10 km); each block values the mean DOCc within the block, and CoV is computed across blocks. (c) Relationship between $CoV_{pixel}$ and the 5km $CoV_{block}$ across years (each point represents one year). For consistency with the original figure, "CV" in the axis labels denotes coefficient of variation (CoV), not cross-validation.

**Table 5.3.** Summary statistics of within-lake heterogeneity (DOCc; 1984–2023).

| Metric | Min | Max | Median |
|---|---|---|---|
| $CoV_{pixel}$ (%) | 6.37 | 13.85 | 9.16 |
| $CoV_{block}$ 3km (%) | 3.04 | 10.23 | 5.68 |
| $CoV_{block}$ 5km (%) | 2.44 | 10.04 | 5.27 |
| $CoV_{block}$ 10km (%) | 1.82 | 9.06 | 4.65 |
| $Mean_{pixel}$ | 4.03 | 4.74 | 4.44 |
| $SD_{pixel}$ | 0.29 | 0.59 | 0.41 |

The reconstructed DOCc record showed three broad temporal phases: an initial increase during 1984–2000, a decline after 2000, and a further decrease after 2013. During the first phase, annual zone means ranged from 4.83–4.98 mg/L in the higher-DOCc subregions (T1–T3 and T5) and from 4.74–4.75 mg/L in the lower-DOCc subregions (T4 and T6), with relatively stable concentrations during the 1990s. Lake-wide DOCc (T0) reached its maximum reconstructed annual mean of 4.90 mg/L in 1996, while annual means in T1 and T2 reached 4.83 and 4.89 mg/L, respectively, during the 1990s. After 2000, concentrations generally declined, particularly in T4–T6, although T5 fluctuated between 4.18 and 4.81 mg/L. DOCc decreased further across all zones after 2013, with annual means in the historically higher-DOCc subregions (T1–T3) often falling below 4.66 mg/L. T4 reached 3.85 mg/L in 2018, while T6 ended the record at 3.83 mg/L in 2023, compared with 4.00 mg/L in 1984. Overall, DOCc was generally lower after 2000, with T1–T3 retaining comparatively higher concentrations than T4–T6 (Figure 5.5).

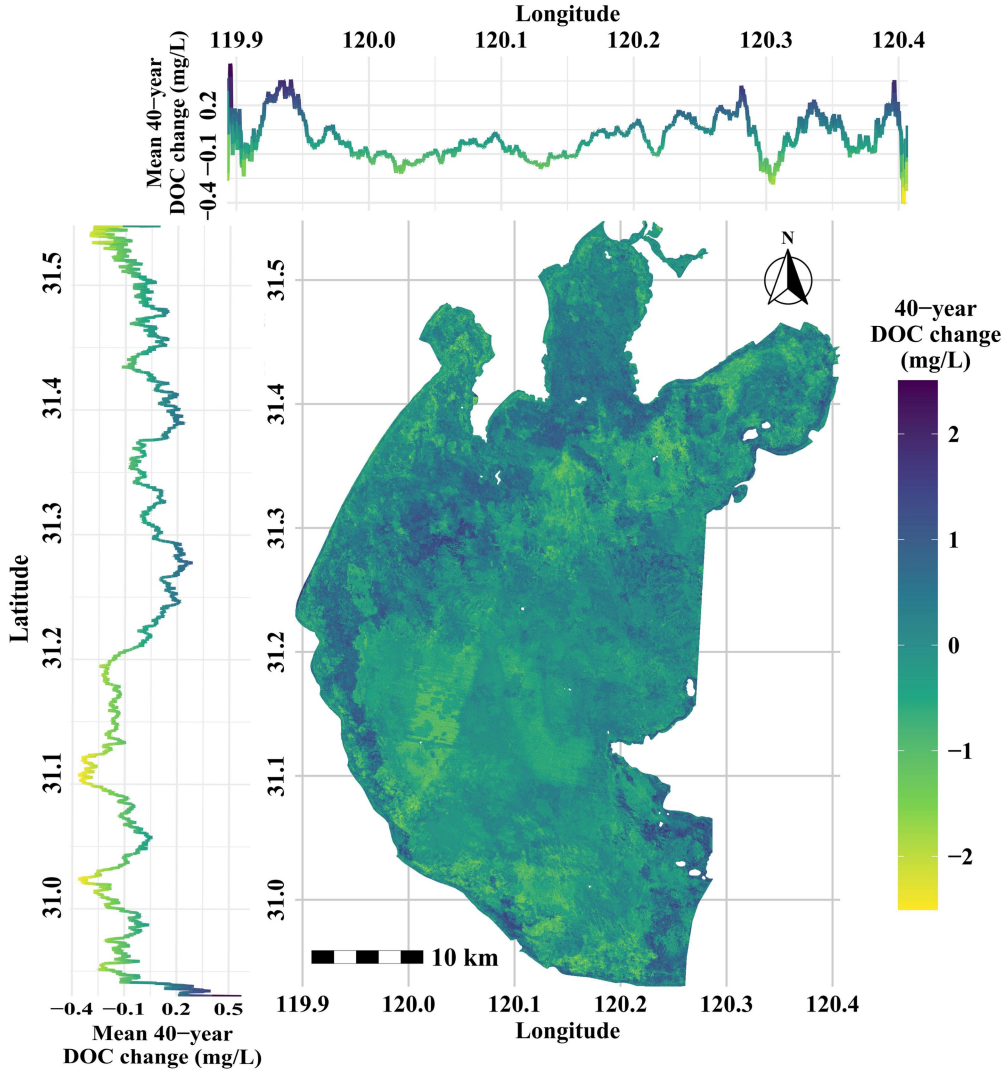



**Figure 5.5.** Spatial variations in DOC concentration changes (mg/L) in Taihu over 40 years (1984–2023), based on pixel-wise differences between 2023 and 1984. Negative values indicate decreases relative to 1984, and positive values indicate increases.

The spatial pattern of ΔDOCc (2023 minus 1984) showed marked heterogeneity, with pronounced decreases in the southern and central open-water zones and localized increases in some bays. These contrasts may reflect differences in underwater light availability, mixing, and exposure to source waters, which influence DOC transformation and retention. They are also consistent with spatial differences in the

relative importance of riverine loading in nearshore bays and in-lake processing in open waters.

The three-phase DOCc trajectory is broadly consistent with the trophic and management history of the Taihu Basin. The pre-2000 increases coincided with agricultural intensification, urbanization, and rising nutrient inputs associated with rapid economic development, which could enhance both terrestrial organic carbon delivery and bloom-derived production within the lake. The post-2000 decline, particularly after 2013, coincided with intensified pollution control, wastewater interception, and watershed restoration, including measures introduced in response to drinking-water crises. Improved point-source management may have reduced DOC delivery, whereas the effects of expanding impervious surfaces depend on changes in organic matter sources and runoff pathways. The stronger declines in T4–T6, compared with persistently higher DOCc in the northern bays (T1–T3), may reflect continued riverine inputs and bloom accumulation in the latter. These spatial differences are compatible with contrasting hydrodynamic conditions and carbon-processing pathways, but temporal coincidence alone cannot establish the contributions of individual management measures. Pixel- and block-level CoVs describe variability within the reconstructed record, but their persistence alone cannot demonstrate stable spatial gradients or establish robustness across sensors and acquisition conditions.

### 5.1.3. DOC storage changes in Taihu

Over the four-decade study period, pixel-level DOCs in Lake Taihu ranged from 5.12 to 9.17 kg C pixel$^{-1}$. At the whole-lake scale (T0), mean pixel-level DOCs were 6.57 ± 0.37 kg C pixel$^{-1}$. Spatial variability was modest, with a spatial SD of 0.64 kg C pixel$^{-1}$ and a spatial CoV of approximately 0.10 (Figure 5.6; Table 5.4). Higher DOCs values (>6.98 kg C pixel$^{-1}$) occurred mainly in northern and nearshore areas (Figure 5.7). Longitudinally, mean DOCs was 7.14 kg C pixel$^{-1}$ in the western band (119.9–120.0°E) and 7.21 kg C pixel$^{-1}$ in the eastern band (120.3–120.4°E). Both values exceeded the mean of approximately 6.65 kg C pixel$^{-1}$ in the central band (120.0–120.2°E). Latitudinally, the southern (31.0–31.2°N) and northern (31.4–31.5°N) bands

had mean DOCs values of 6.98–7.26 kg C pixel$^{-1}$. The central band (31.2–31.3°N) had a lower mean of approximately 6.65 kg C pixel$^{-1}$. Four-decade changes in pixel-level DOCs (ΔDOCs) ranged from −3.56 to +3.69 kg C pixel$^{-1}$, with a mean change of 0.01 ± 0.76 kg C pixel$^{-1}$.

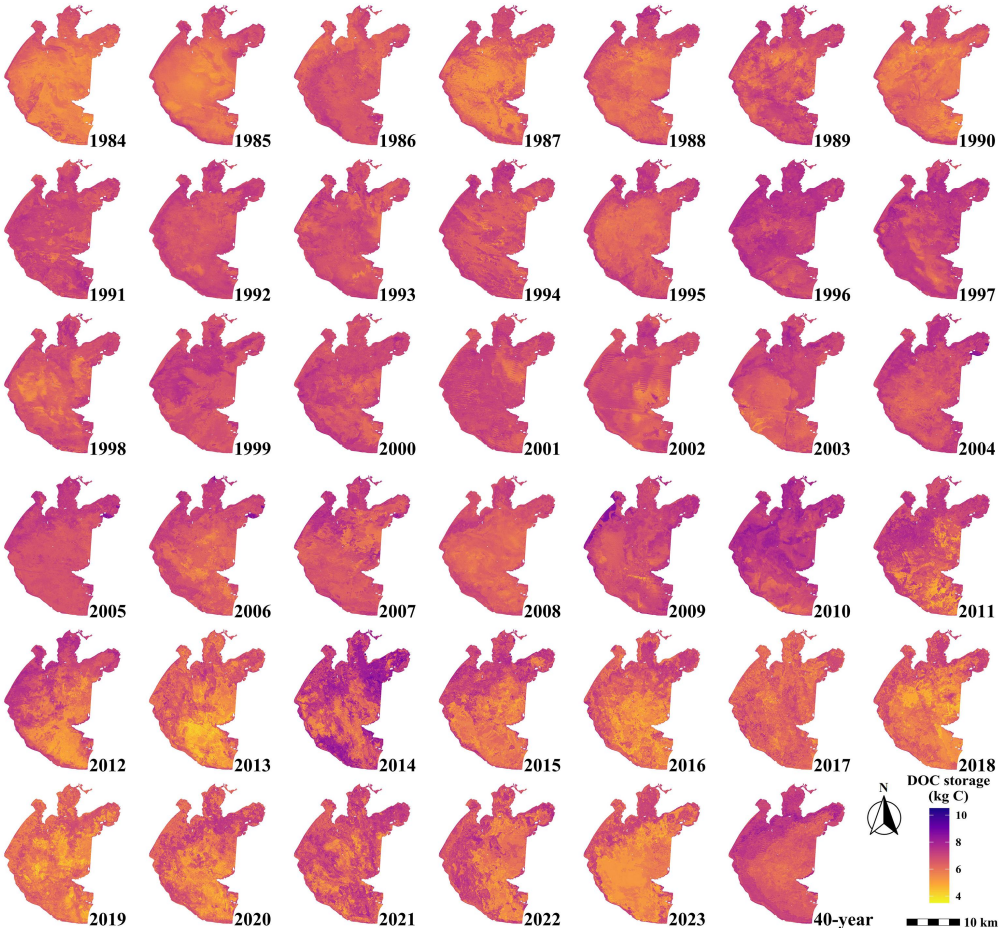



**Figure 5.6.** Dynamic changes in DOC storage in the T0 region over the past 40 years (1984–2023).

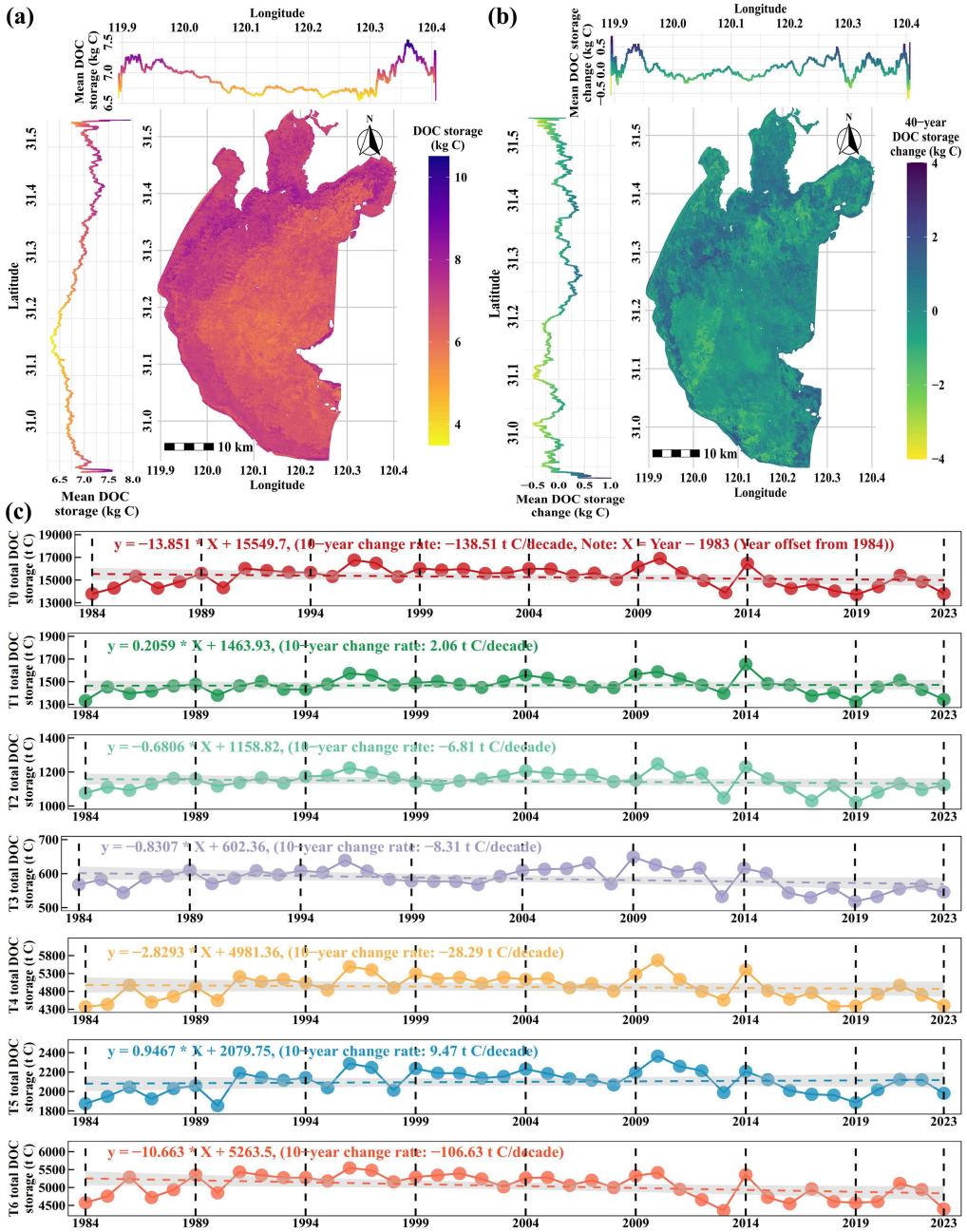



**Figure 5.7.** (a) Mean DOC storage (kg C) in Taihu (1984–2023), (b) DOC storage change (kg C) in Taihu (1984–2023), negative values indicate decreases relative to 1984, and positive values indicate increases. (c) Temporal trends of total DOC storage (t C) in Taihu subregions (T0–T6, 1984–2023).

**Table 5.4.** Summary statistics of within-lake heterogeneity (DOCs; 1984–2023).

| Metric | Min | Max | Median |
|---|---|---|---|
| $CoV_{pixel}$ (%) | 6.37 | 13.85 | 9.16 |
| $CoV_{block}$ 3km (%) | 3.04 | 10.23 | 5.68 |
| $CoV_{block}$ 5km (%) | 2.44 | 10.04 | 5.27 |
| $CoV_{block}$ 10km (%) | 1.82 | 9.06 | 4.65 |
| $Mean_{pixel}$ | 5.90 | 7.28 | 6.63 |
| $SD_{pixel}$ | 0.44 | 0.92 | 0.61 |

At the lake-wide scale (T0; n = 1,981,915 pixels), reconstructed total DOCs peaked at 16,906 t C in 2010 and reached 13,802 t C in 2023, close to the 1984 value of 13,788 t C. The mean annual total was 15,265.76 ± 859.02 t C, with a range of 13,710.15–16,905.96 t C. A linear fit to the annual totals yielded a slope of −138.5 t C/decade, indicating a slight decline in the fitted trend despite similar start- and end-year values. Among the subregions, T6 (Southwest Lake) showed the most negative slope (−106.6 t C/decade), with a maximum total of 5,414 t C and a final value of 4,402 t C. Between 1984 and 2023, totals increased slightly in T1 from 1,333 to 1,343 t C and in T5 from 1,874 to 1,982 t C. Their maximum was 1,654 t C in 2014 and 2,363 t C in 2010, respectively, while T5 had a positive fitted slope of 9.47 t C/decade. T2 increased from 1,076 t C in 1984 to 1,124 t C in 2023, whereas T3 and T4 showed small decreases. Overall, DOCs changes were spatially heterogeneous, with the strongest decline in T6 and comparatively smaller changes in the remaining zones (Table 5.5).

**Table 5.5.** Zone-level descriptive statistics of annual DOC storage (DOCs) in Lake Taihu (T0–T6), 1984–2023.

| Zone | Total Mean ± SD (t C) | Total Median (Q25–Q75) (t C) | Total Min–Max (t C) | Pixel Mean ± SD (kg C/pixel) | Spatial SD (kg C/pixel) | Spatial CoV | N (years) | Period |
|---|---|---|---|---|---|---|---|---|
| T0 (Whole lake) | 15,265.76 ± 859.02 | 15,394.64 (14,572.73–15,901.37) | 13,710.15–16,905.96 | 6.57 ± 0.37 | 0.64 | 0.1 | 40 | 1984–2023 |
| T1 (Gonghu Bay) | 1,468.15 ± 70.43 | 1,470.47 (1,432.09–1,502.70) | 1,321.59–1,653.56 | 6.85 ± 0.33 | 0.57 | 0.08 | 40 | 1984–2023 |
| T2 (Meiliang Bay) | 1,144.86 ± 50.84 | 1,149.77 (1,120.63–1,178.55) | 1,021.91–1,248.08 | 6.96 ± 0.31 | 0.55 | 0.08 | 40 | 1984–2023 |
| T3 (Zhushan Bay) | 585.33 ± 31.96 | 587.19 (566.82–609.02) | 518.38–649.12 | 6.87 ± 0.38 | 0.52 | 0.08 | 40 | 1984–2023 |
| T4 (Central Lake) | 4,923.36 ± 332.53 | 4,948.62 (4,686.82–5,151.78) | 4,364.00–5,667.12 | 6.46 ± 0.44 | 0.6 | 0.09 | 40 | 1984–2023 |
| T5 (Northwest Lake) | 2,099.16 ± 121.06 | 2,120.08 (2,012.75–2,189.44) | 1,854.16–2,363.19 | 6.73 ± 0.39 | 0.57 | 0.09 | 40 | 1984–2023 |
| T6 (Southwest Lake) | 5,044.90 ± 328.48 | 5,134.48 (4,749.50–5,299.61) | 4,355.97–5,545.59 | 6.43 ± 0.42 | 0.59 | 0.09 | 40 | 1984–2023 |

DOCs depends on both DOCc and the water-column height derived from annual WSC estimates. Its four-decade trajectory therefore reflects the gradual decline in near-surface DOCc and substantial interannual variations in lake water storage. The DOCs maximum in 2010 coincided with elevated WSC rather than peak DOCc, indicating a substantial hydrological contribution to the storage maximum. This highlights the importance of reconstructing WSC when assessing four-decade DOCs dynamics in a shallow, regulated lake. The ERA5-driven WSC reconstruction performed well in validation ($R^2 = 0.82$; RPD = 2.11), supporting its application to the period before 2007. However, reconstruction accuracy during this earlier period could not be directly verified because storage observations were unavailable. Validation points clustered around the 1:1 line within the typical storage range of 3.5–4.5 TL, indicating good agreement under commonly observed storage conditions.

The spatial distribution of DOCs broadly followed that of DOCc. Values were higher in the western and eastern longitudinal bands and the northern and southern latitudinal bands than in the central zone. Northern and nearshore areas exhibited relatively high storage per pixel, whereas central areas had lower values. Although DOCs also depends on the water-column height used in its calculation, attributing these spatial differences to local bathymetry requires spatially resolved water-depth information. Subregional trends also differed, with T6 showing the largest decline and T1 and T5 exhibiting modest increases. These contrasting trends indicate that local gains partly offset losses elsewhere in the lake. Monitoring DOCs at the subregional scale would therefore help identify changes that may be obscured by whole-lake averages and provide more spatially specific information for lake management.

### 5.1.4. Driver attribution of DOC variability

The SEM and auxiliary analyses linked reconstructed DOC dynamics to meteorological conditions, human activity, bio-optical state, and vegetation cover. Meteorological factors had the largest absolute direct path coefficients for DOCc and DOCs ($\beta = -0.72$ and $-0.98$, respectively). The corresponding coefficients were $-0.31$ and $-0.26$ for human activity, $-0.28$ and $-0.19$ for bio-optical conditions, and 0.35 and $-0.15$ for vegetation cover. Vegetation showed a positive association with DOCc and ranked second in absolute direct path magnitude, whereas its association with DOCs was negative. These variables were used only for post-reconstruction interpretation and diagnostic analyses and were not included in the DOC retrieval calibration (Figure 5.8).

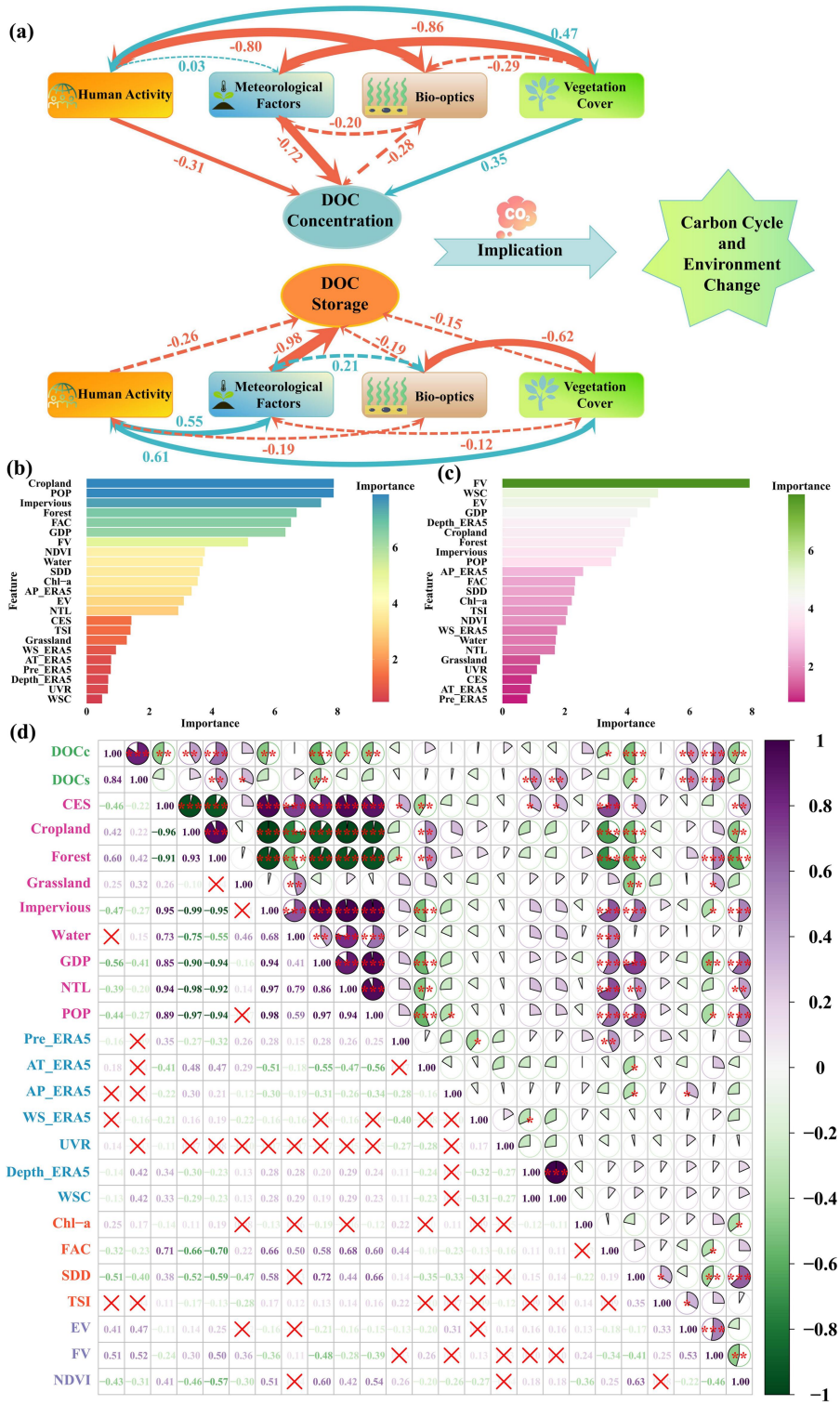

(a)
Human Activity
Meteorological Factors
Bio-optics
Vegetation Cover
DOC Concentration
DOC Storage
Implication
Carbon Cycle and Environment Change
(b)
(c)
Importance
Feature
(d)

**Figure 5.8.** (a) Structural equation modelling (SEM) diagram illustrating the direct and indirect impacts of human activities, meteorological factors, bio-optical properties, and vegetation cover on DOCc and DOCs in Taihu. Red and blue arrows represent negative and positive influences, respectively. All path coefficients are standardized (beta). (b) Importance of auxiliary data and DOCc, (c) Importance of auxiliary data and DOCs, (d) Pearson correlation matrix between DOC and key water quality parameters. $r_c$ and $r_s$ represent the Pearson correlation value between DOCc and DOCs with factors.

Meteorological conditions may affect DOC through several interacting pathways. Event-scale runoff can increase riverine and lateral DOC inputs, whereas sustained rainfall may dilute lake DOC when incoming water has lower concentrations than the receiving lake (Huang et al., 2017; Safieddine and Heald, 2017). Elevated AT can accelerate microbial mineralization, while wind-driven resuspension can mobilize and redistribute sediment-derived organic matter (Hu et al., 2016; Zhou et al., 2016). In the DOCc SEM, the meteorology–vegetation path was strongly negative ($\beta = -0.86$), but this association does not specifically demonstrate vegetation suppression by meteorological extremes. The larger meteorological coefficient for DOCs ($\beta = -0.98$) than for DOCc ($\beta = -0.72$) is compatible with meteorological influences on both concentration and water volume. However, meteorological dependence is also introduced through the ERA5-driven WSC reconstruction, so these coefficients do not independently quantify hydrological and biogeochemical contributions. The 2010 DOCs maximum, which coincided with elevated WSC rather than peak DOCc, illustrates the importance of distinguishing concentration from total storage.

Socioeconomic indicators were negatively associated with DOCc and DOCs: CES ($r_c = -0.46$, $r_s = -0.22$), the GDP proxy ($r_c = -0.56$, $r_s = -0.41$), and impervious cover ($r_c = -0.47$, $r_s = -0.27$). These relationships may reflect changes in carbon sources and hydrological connectivity, but do not establish that urbanization reduced DOC delivery (Wang et al., 2024; Zhou et al., 2023). Associations with cropland, forest, and grassland cover may reflect differences in terrestrial carbon supply; correlations with forest were positive for both DOCc and DOCs ($r_c = 0.60$, $r_s = 0.42$). NTL and POP were also

negatively associated with DOCc ($r_c$ = −0.39 and −0.44, respectively). Agriculture and aquaculture can alter nutrient supply and the balance between DOC production and mineralisation (D. Liu et al., 2021; Liu et al., 2020). Without direct measurements of industrial discharge or DOC fluxes, socioeconomic indicators remain proxies, and proposed pathways involving rapid mineralisation of labile carbon to $CO_2$ cannot be directly verified (Zhou et al., 2023).

Bio-optical associations were consistent with interactions among production, loss, and light-attenuation processes. Phytoplankton can release DOC through exudation and cell breakdown (Baines and Pace, 1991; Stedmon and Markager, 2005). Reported Chl-a–DOCc correlations were positive but modest ($r_c$ = 0.35 in the remote-sensing analysis and $r_c$ = 0.25 in this study; Liu et al., 2021). SDD was negatively associated with DOCc ($r_c$ = −0.51) and DOCs ($r_s$ = −0.40), potentially reflecting both organic-matter-related light attenuation and differences in photodegradation. Higher turbidity may limit photodegradation by reducing light penetration (Rodríguez-Murillo and Filella, 2015). The negative FAC–DOCc correlation ($r_c$ = −0.32) may reflect variation in the balance between production and consumption but does not uniquely identify enhanced microbial processing. Regional associations between eutrophication and DOC variability in eastern China provide broader context for these relationships (Liu et al., 2020).

Hydrological residence time, which cannot be inferred from WSC alone, affects both DOC export and in-lake processing. Longer residence may reduce flushing while allowing more mineralisation and photodegradation, so it does not necessarily increase DOC storage (D. Liu et al., 2021; Messager et al., 2016; Meyer-Jacob et al., 2019). Aquatic vegetation was positively associated with DOCc (FV: $r_c$ = 0.51; EV: $r_c$ = 0.41) and DOCs (FV: $r_s$ = 0.52; EV: $r_s$ = 0.47), consistent with potential plant-derived inputs and vegetation-mediated changes in water exchange (Zhang et al., 2023). Terrestrial NDVI was negatively associated with DOCc ($r_c$ = −0.43) and DOCs ($r_s$ = −0.31), but these correlations do not directly demonstrate greater soil-carbon retention or reduced DOC mobilization.

The opposing signs for aquatic vegetation and terrestrial NDVI highlight the need to distinguish in-lake and catchment vegetation rather than combine them into a single "greenness" indicator. These pairwise correlations also differ from the conditional relationships represented by SEM direct paths. Excluding the macrophyte-dominated EB and EC zones was intended to reduce vegetation-related optical interference but also limited direct inference of DOC within these habitats. Given weak DOC–CDOM coupling, relationships involving Chl-a, SDD, and FAC should be interpreted as bio-optical associations rather than direct measurements of DOC sources or losses. Where these indicators are derived from satellite reflectance, shared spectral information may also contribute to their correlations with reconstructed DOCc. The modest Chl-a correlation ($r_c$ = 0.25) indicates limited linear correspondence and does not, itself, identify the processes underlying DOC variability.

### 5.1.5. Carbon-cycle implications

The reconstructed lake-wide DOCs time series showed pronounced interannual variability and occasional fluctuations lasting several years. These patterns indicate that Taihu's dissolved carbon pool varied over timescales relevant to lake management. They may reflect changes in external loading, in-lake production, mineralization, and hydrological storage and exchange. Such variability is relevant to the lake carbon budget, although changes in DOCs alone cannot quantify atmospheric emissions or hydrological export. The maps also revealed persistent gradients between river-influenced nearshore sectors and open-water areas. These spatial differences may influence DOC processing through variations in light exposure. In turbid nearshore waters, reduced light penetration can constrain photochemical degradation, whereas clearer open-water conditions may allow greater photochemical processing (Z. Zhao et al., 2024a).

The reconstructions also captured episodic anomalies within the four-decade record. Elevated DOC estimates in the area affected by the 2009 black bloom were consistent with previous satellite evidence and a Landsat inversion for the same year (D. Liu et al., 2021). This agreement supports the detection of an event-related anomaly,

although unusual optical conditions may also affect retrieval accuracy. Such episodes may coincide with enhanced labile-carbon turnover and $CO_2$ evasion, but these flux responses were not measured here. Resolving episodic anomalies alongside four-decade changes can therefore help identify periods requiring targeted observations of carbon cycling. Positive associations of aquatic vegetation indicators (FV and EV) with DOCs, together with negative associations with NDVI during some periods, suggest that vegetation–DOC relationships varied among indicators and over time. These patterns may reflect differing contributions of primary production, exudation, and microbial consumption under changing optical and hydrodynamic conditions.

The three complementary attribution approaches, SEM, driver-based ML importance, and Pearson correlation, consistently identified meteorology as the factor most strongly associated with DOC variability. Human activity, bio-optical conditions, and vegetation showed additional associations with DOC dynamics. Agreement across these approaches supports the consistency of the findings, but they share underlying data and do not independently establish causality. SEM depends on an assumed causal structure, ML importance can be affected by correlated predictors, and Pearson correlation captures pairwise linear associations. Accordingly, the negative associations with CES, GDP, impervious cover, NTL, and POP were interpreted as contextual indicators of basin-wide change. Altered hydrological connectivity and mineralisation of labile inputs offer possible explanations, but these processes were not directly measured. The reconstructed dataset provides a multi-decadal baseline for tracking DOC storage and identifying persistent spatial patterns and episodic anomalies to guide targeted monitoring. Applying this framework to other optically complex lakes would require suitable satellite archives, local calibration and validation, and reliable water-storage estimates.

## 5.2. Study 2: Spatial cross-validation reveals a double inseparability boundary for satellite monitoring of narrow-river nutrients

This section presents the principal findings of Study 2, evaluating the spatial transferability and inherent limitations of satellite-based nutrient retrieval and land-cover attribution across narrow river networks using spatially explicit cross-validation. This work has been submitted for consideration as the following manuscript:

**Shen, F.**, Li, J., Fan, S., Luo, T., Guo, Y.*, Chen, Z.*. 2026. Spatial cross-validation reveals a double inseparability boundary for satellite monitoring of narrow-river nutrients. Manuscript submitted to ***Remote Sensing of Environment***.

### 5.2.1. A long-term object archive, its sub-pixel geometry and the nutrient setting

The quality-gated archive comprised 178,286 cloud-free, object-level observations across 103 river reaches from 1985 to 2025. Its temporal coverage reflected the succession of satellite generations. Landsat 5 anchored the early record from 1985 to 2011, while Landsat 7 became available in 1999 and Landsat 8 in 2013. Sentinel-2 and the HLS L30 and S30 products substantially increased observation density from 2015 onwards, and Landsat 9 contributed from 2021 (Figure 5.9a). Consequently, cumulative observation counts increased gradually during the first three decades and accelerated markedly after 2015. Within five years of becoming available, the combined Sentinel-2 and HLS processing streams surpassed the native Landsat streams in total archived observations (Figure 5.9b; Figure 5.10a, b). Usable observations were concentrated in the summer half-year (Figure 5.9c). All 103 reaches were represented, with a median of 1,677 observations per reach (Figure 5.9d). Totals for individual processing streams ranged from approximately 5.0 thousand observations for Landsat 9 to 49.0 thousand for HLS S30 (Figure 5.9e). Matching satellite observations to in-situ sampling dates yielded 16,607, 17,481, and 17,992 reflectance–concentration pairs for $NH_4^+$-N, $NO_3^-$-N and TP, respectively (Figure 5.9f). Across the three width classes, observations were distributed as 54% in W1, 31% in W2, and 15% in W3 (Figure 5.9g).

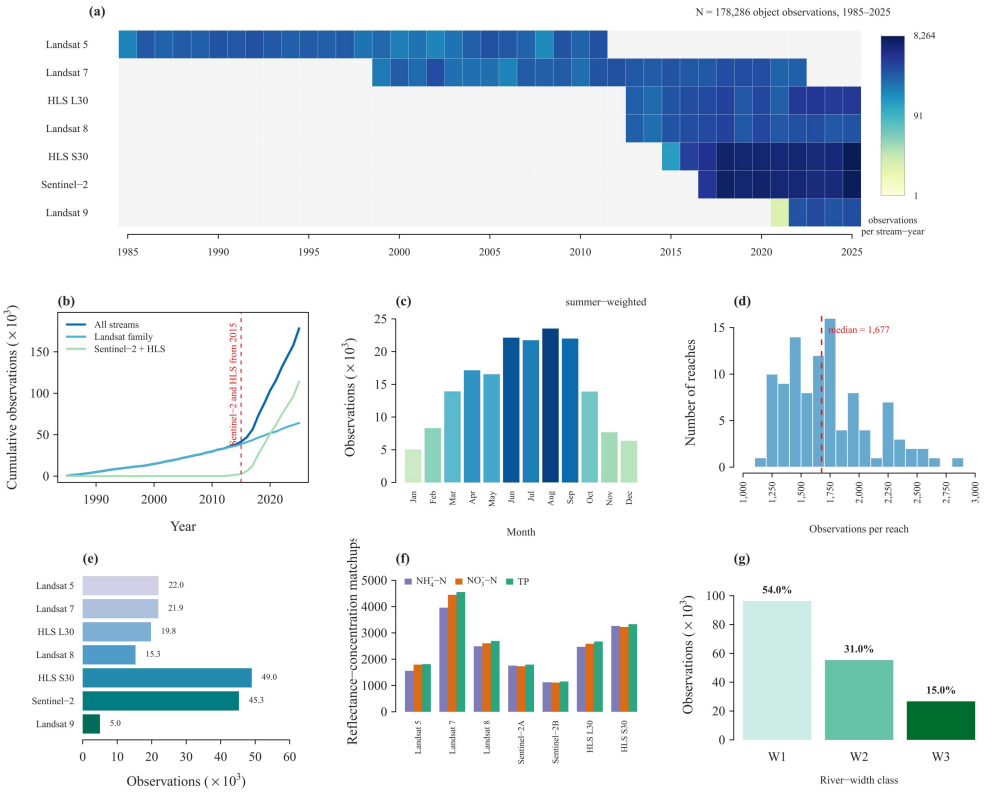



**Figure 5.9.** The long-term object-level archive (N = 178,286 observations over 103 reaches). (a) Observations per reflectance stream and year, showing the generational hand-over from Landsat 5 to Sentinel-2 and HLS. (b) Cumulative observations for all streams, the Landsat family, and Sentinel-2 plus HLS. (c) Monthly distribution of acquisitions. (d) Distribution of observations per reach. (e) Total observations per stream. (f) Reflectance-concentration matchups by stream and parameter. (g) Observations by river-width class.

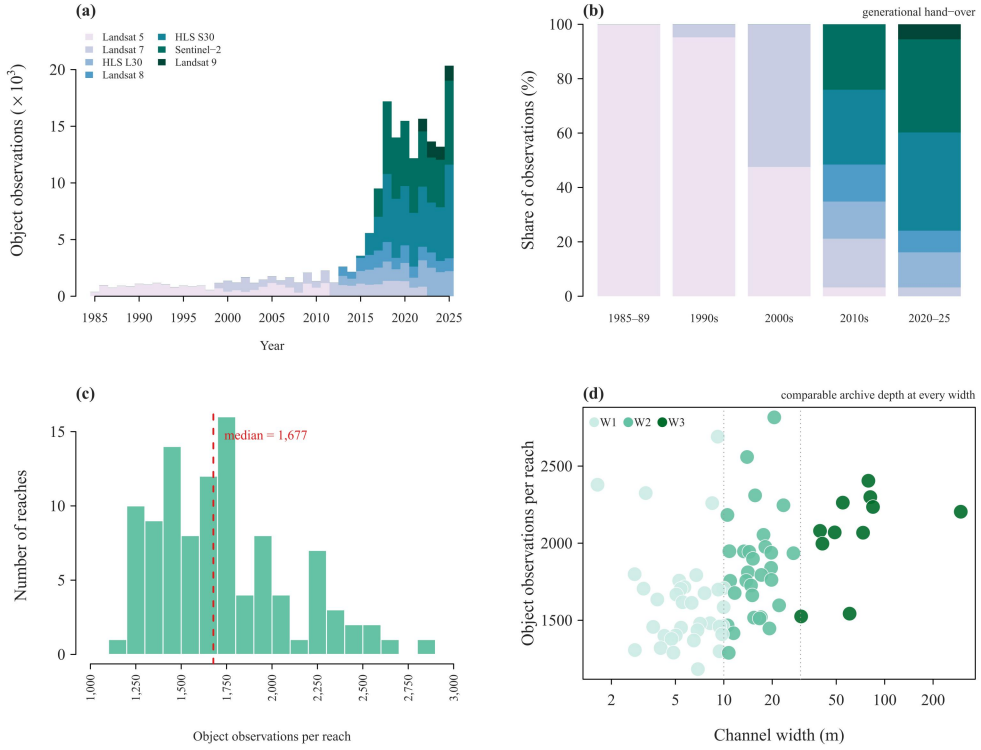



**Figure 5.10.** Detail of the object-level observation archive, 1985–2025 (N = 178,286 observations over 103 reaches). (a) Annual observations stacked by reflectance stream. (b) Share of observations contributed by each stream within five periods, showing the generational hand-over from Landsat 5 to Sentinel-2 and HLS. (c) Distribution of object observations per reach, with the median marked. (d) Object observations per reach against channel width, colored by width class; the dotted lines mark the 10-m and 30-m width thresholds.

The geometric constraint on nutrient retrieval was determined by the relationship between channel width and sensor pixel size (Figure 4.16). When drawn to scale against a 30-m pixel, the median W1 reach was only 4.9 m wide and occupied approximately 16% of the pixel width. The median W2 reach was 15.3 m wide and occupied 51%, whereas the median W3 reach, with a width of 54.9 m, extended beyond a single pixel (Figure 4.16a–c). Even at the 10-m resolution of Sentinel-2, the median W1 channel occupied only 0.49 of the pixel width. The proportion of object-dates classified as

internally heterogeneous increased from 5.97% in W1 to 15.73% in W2 and 61.32% in W3 (Figure 4.16d; Table 4.11). Narrow reaches contained relatively few valid pixels, whereas wider reaches contained sufficient pixels to reveal both within-channel variability and bank-edge effects. Object-level water indices also became progressively less negative as channel width increased. NDWI and MNDWI increased from −0.61 and −0.48 in W1 to −0.39 and −0.19 in W3, respectively (Figure 4.16e). This pattern indicated an increasing contribution from the water surface, although both indices remained negative even in the widest reaches.

The nutrient signals that the retrieval models were expected to recover also exhibited clear temporal structure (Figure 5.11). $NO_3^-$-N concentrations peaked in winter, a pattern consistent with lower biological uptake and greater leaching during this period. TP showed the opposite seasonal pattern, with a clear summer maximum, while $NH_4^+$-N followed a weaker cycle weighted towards winter (Figure 5.11a–d). Concentrations differed only modestly among the three width classes (Figure 5.11e), and all three parameters declined between the 1990s and the 2010s (Figure 5.11f). At the observation level, relationships among the parameters were weak to moderate. Spearman's ρ was 0.38 between $NH_4^+$-N and TP, 0.21 between $NO_3^-$-N and TP, and 0.15 between $NH_4^+$-N and $NO_3^-$-N (Figure 5.11g). The minimum reported concentrations reached apparent analytical floors of 0.0095, 0.0226 and 0.006 mg $L^{-1}$ for $NH_4^+$-N, $NO_3^-$-N and TP, respectively (Figure 5.11h). The next section evaluates whether these nutrient patterns could be retrieved at previously unseen reaches.

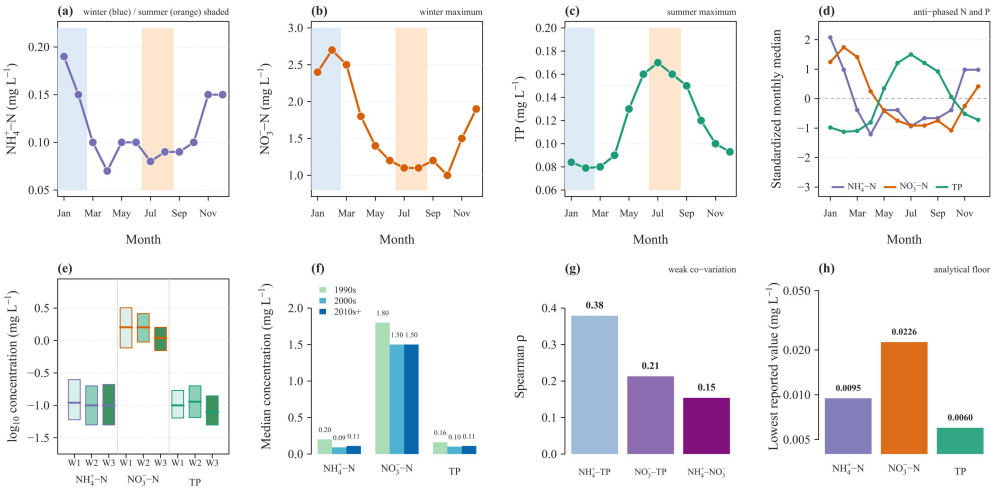



**Figure 5.11.** In-situ nutrient climatology, 1990–2020. (a–c) Monthly median concentrations of $NH_4^+$-N, $NO_3^-$-N and TP, with winter and summer shaded. (d) The same cycles standardized, showing the anti-phase between nitrate and phosphorus. (e) Concentration distributions by width class (median and IQR). (f) Decadal median concentrations. (g) Spearman correlations among the three parameters at the observation level. (h) Lowest reported value per parameter, the analytical floor.

### 5.2.2. The optical inseparability boundary

Optical relationships were weak across the full dataset. Among the 30 candidate features, the maximum absolute Spearman's ρ was 0.17 for $NH_4^+$-N, 0.23 for $NO_3^-$-N and 0.34 for TP (Figure 5.12a; Table 5.6). TP showed the strongest associations with near-infrared reflectance, NDVI, NDTI and the red-to-green ratio, consistent with indirect sensitivity to particulate-associated phosphorus. In contrast, $NH_4^+$-N and $NO_3^-$-N showed only weak relationships with visible reflectance. Except for the TP near-infrared relationship, no feature exceeded the predefined weak-signal threshold of $|\rho| = 0.30$. Furthermore, the strongest relationships remained highly scattered across concentration ranges spanning approximately two orders of magnitude (Figure 5.12b–d), indicating that the available spectral features provided insufficient information for robust nutrient retrieval in previously unseen reaches of the South Bohemian River.

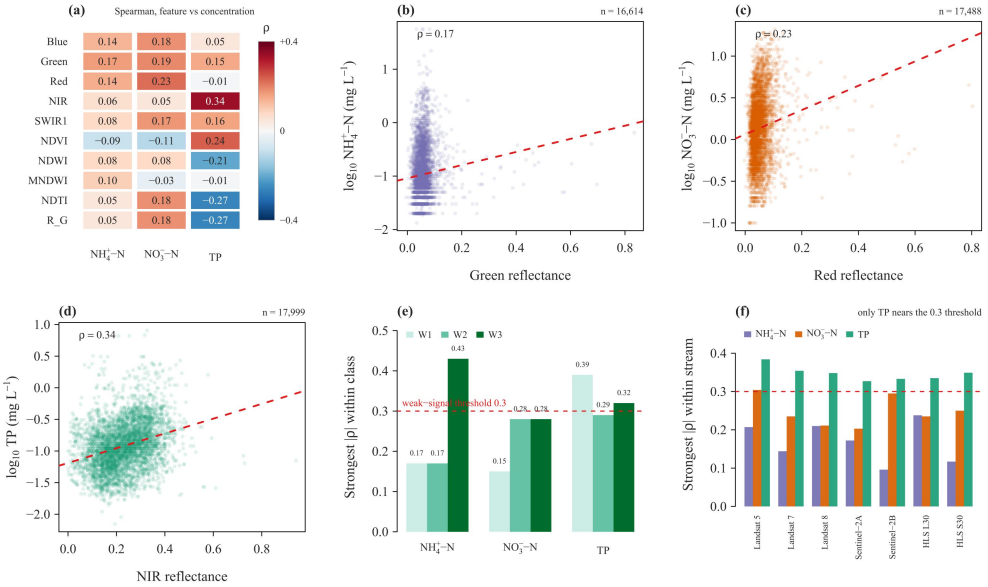



**Figure 5.12.** Spectral-nutrient relationships and the optical limit. (a) Spearman correlations between object-level features and $\log_{10}$ concentration over the full sample. (b-d) The strongest single relationship for each parameter: $NH_4^+$-N against green, $NO_3^-$-N against red and TP against near-infrared reflectance. (e) Strongest absolute rho within each width class. (f) Strongest absolute rho within each reflectance stream; the dashed line marks the 0.3 weak-signal threshold.

**Table 5.6.** Per-parameter in-situ dataset, matchup count, strongest optical signal, mutual information and full-sample cross-validated retrieval skill.

| Parameter | In-situ observations (n) | Median (mg $L^{-1}$) | Matchups (n) | max \|ρ\| | MI across reaches (nats) | MI within reach (nats) | Random-CV $R^2$ | Spatial-CV $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| $NH_4^+$-N | 16,302 | 0.11 | 16,607 | 0.17 | 0.08 | 0.12 | 0.09 | −0.03 | 0.12 |
| $NO_3^-$-N | 17,933 | 1.50 | 17,481 | 0.23 | 0.13 | 0.15 | 0.15 | 0.05 | 0.10 |
| TP | 18,286 | 0.11 | 17,992 | 0.34 | 0.13 | 0.16 | 0.23 | 0.11 | 0.12 |

Notes: Spatial CV is leave-site-out and $\Delta R^2$ = random − spatial; within the widest class (W3), the inflation rises to $\Delta R^2$ = 0.24, 0.41 and 0.57 for $NH_4^+$-N, $NO_3^-$-N and TP.

Stratification by channel width did not materially strengthen the optical signal. Within the width classes, only TP in W1 (|ρ| = 0.39) and $NH_4^+$-N in W3 (|ρ| = 0.43) exceeded the predefined threshold of 0.30 (Figure 5.12e; Figure 5.13). However, feature–nutrient correlations did not increase systematically with channel width or fractional water coverage. Across the reflectance streams, only TP exceeded the threshold, with maximum absolute correlations ranging from 0.33 to 0.38. The 10-m Sentinel-2 stream provided no clear advantage over the 30-m streams (Figure 5.12f). The absence of a consistent width-related improvement suggested that nutrient retrievability was not controlled solely by the proportion of water within a pixel.

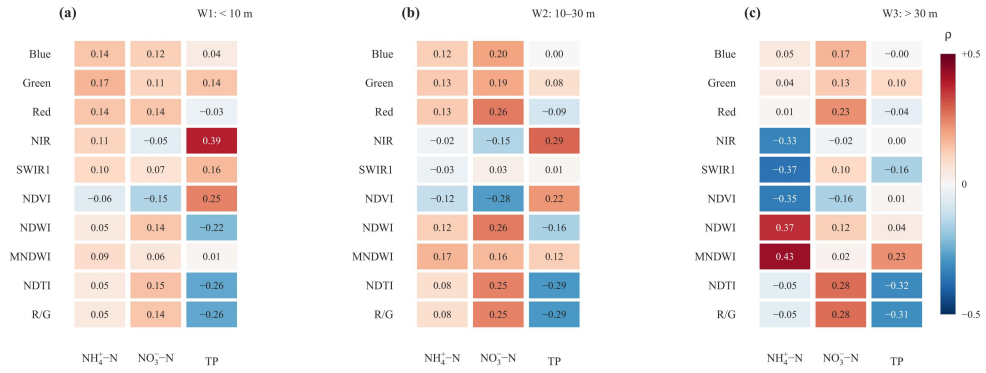



**Figure 5.13.** Feature-concentration correlations within each river-width class. Spearman correlations between object-level features and $\log_{10}$ concentration computed separately within (a) W1, (b) W2 and (c) W3. Correlations do not strengthen systematically as the channel fills the pixel, and only two of the thirty-class-wise maxima exceed the 0.3 weak-signal threshold. Figure 5.12a shows the equivalent matrix over the full sample.

Under spatially blocked CV, full-sample $R^2$ values were −0.03 for $NH_4^+$-N, 0.05 for $NO_3^-$-N, and 0.11 for TP. The median across all 36 parameter–model–scope combinations was −0.01, indicating that none of the three nutrient variables could be reliably retrieved at unseen reaches (Figure 5.14a, e; Table 5.6). This result was insensitive to the learner: for each parameter, the maximum difference in full-sample $R^2$ among the three regressors was 0.02 (Figure 5.15). In contrast, random CV using

identical data, predictors, and model settings yielded $R^2$ values of 0.09, 0.15, and 0.23, inflating $R^2$ by 0.10–0.12 for the full sample and by up to 0.57 within individual width classes. All 36 comparisons fell below the 1:1 line (Figure 5.14d).

Two diagnostics help explain this discrepancy. Firstly, inflation did not decline as channel width increased, providing no evidence that greater pixel purity improved transferability. Instead, mean inflation increased from approximately 0.15 in the narrowest class to parameter-specific values of 0.21, 0.39, and 0.56 in the widest class (Figure 5.14b). Because the widest class contained the fewest and most spatially clustered stations, this pattern was more consistent with monitoring-network geometry than with a simple width-dependent optical effect. Secondly, across-reach mutual information was low (0.08–0.13 nats) and consistently lower than within-reach values (Figure 5.14c), suggesting that the remaining dependence was predominantly local. This model-agnostic evidence indicates that the weak spatial performance was not specific to the tested regressors, although univariate, binned mutual information cannot exclude signals arising from multivariate interactions or weakened by discretization.

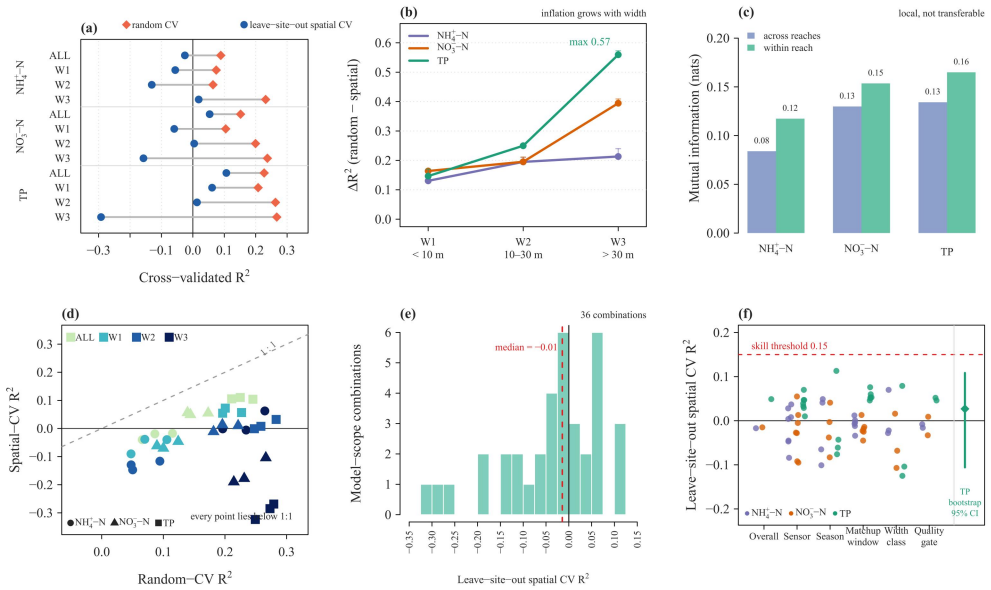


**Figure 5.14.** The optical inseparability boundary. (a) Comparison of $R^2$ under random and spatial-block CV, averaged over three learners, for the full sample and each width class. (b) The inflation $\Delta R^2$ grows with channel width; whiskers span the three learners. (c) Mutual information across reaches versus within reaches. (d) Spatial against random skill for all 36 model-scope combinations, coloured by fold scope and shaped by parameter; every point lies below the 1:1 line. (e) Distribution of spatial cross-validated $R^2$ across those combinations. (f) Spatial cross-validated $R^2$ recomputed within strata of sensor, season, matchup window, width class and quality gate, with a 120-iteration reach-resampling bootstrap for TP.

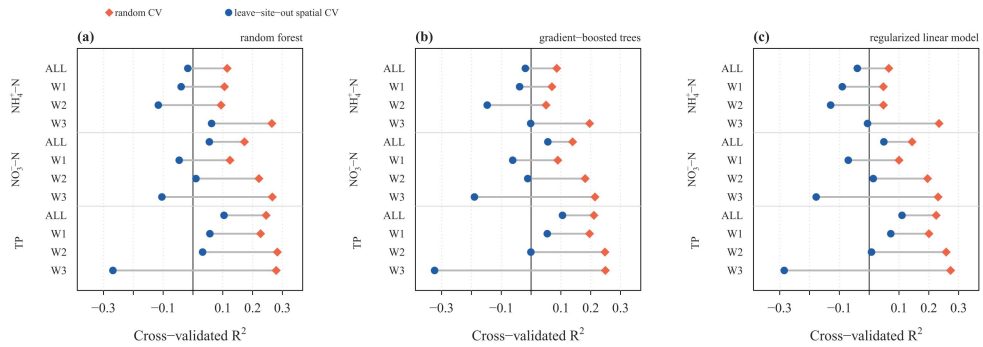



**Figure 5.15.** Cross-validated skill by learner. Comparison of $R^2$ under random and spatial-block CV for (a) random forest, (b) gradient-boosted trees and (c) the regularized linear model, at full sample and within each width class. The collapse from random to spatial validation, and its growth with channel width, are reproduced by all three learners, so the verdict is not an artefact of model choice. Figure 5.14a shows the average over the three.

Spatial structure in nutrient concentrations helps explain why random CV appeared skilful (Figure 5.16). Moran's I values for reach-mean $NH_4^+$-N, $NO_3^-$-N and TP concentrations were 0.35, 0.33, and 0.44, respectively, and all were significant at $p < 0.001$ (Figure 5.16a). The mapped fields and Moran scatterplot for TP showed the same positive spatial pattern (Figure 5.16e; Figure 5.17). The empirical semivariogram for $NO_3^-$-N increased from 0.054 at 4 km to 0.226 at 124 km without reaching a sill (Figure 5.16b). This pattern indicated spatial dependence or a broad-scale trend

extending beyond the sampled distances, although the correlation range could not be estimated within the study region. Between-reach differences accounted for 46%, 53%, and 59% of total variance, respectively (Figure 5.16c). The network was denser than this structure, with a median nearest-neighbor distance of 2.9 km (Figure 5.16d). Random folds could therefore place observations from the same or nearby reaches in both training and test sets. Dividing the stations into seven spatial blocks removed this shortcut and substantially reduced model performance (Figure 5.16f).

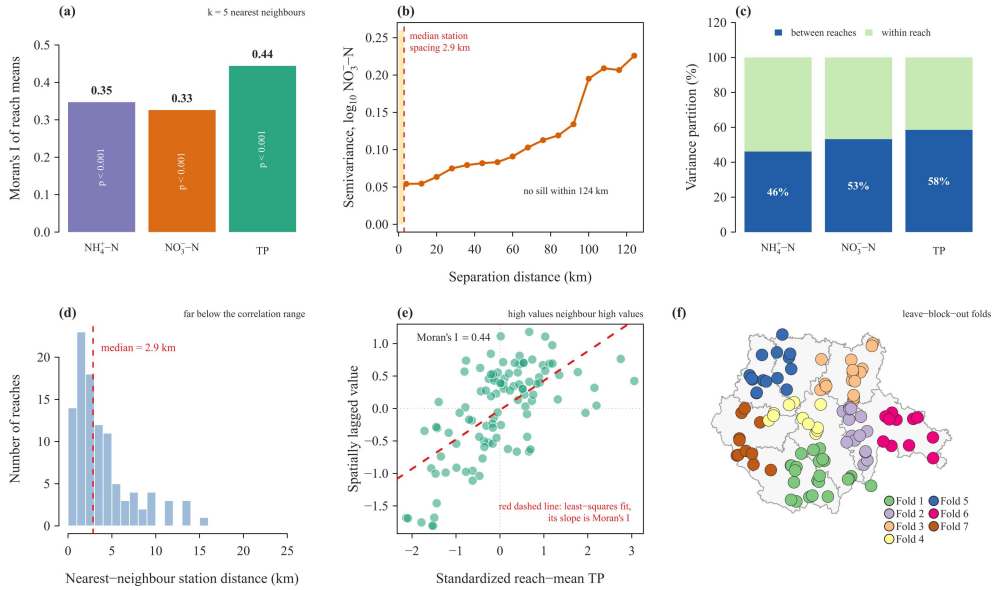



**Figure 5.16.** Spatial structure of the nutrient field and the design of the spatial folds. (a) Moran's I of reach mean concentrations with k-nearest-neighbor weights (k = 5). (b) Empirical semivariogram of reach-mean $\log_{10}$ $NO_3^-$-N; the shaded band marks the median station spacing. (c) Partition of variance between and within reaches. (d) Distribution of nearest-neighbour station distances. (e) Moran scatterplot for reach-mean TP; the red dashed line is the least-squares fit, whose slope is Moran's I. (f) The seven contiguous blocks that define the leave-block-out folds. These folds are obtained by k-means clustering of station coordinates and are numbered from Fold 1 to Fold 7; they are not the SB1-SB7 administrative districts in Figure 4.14b, whose boundaries are shown in grey for reference.

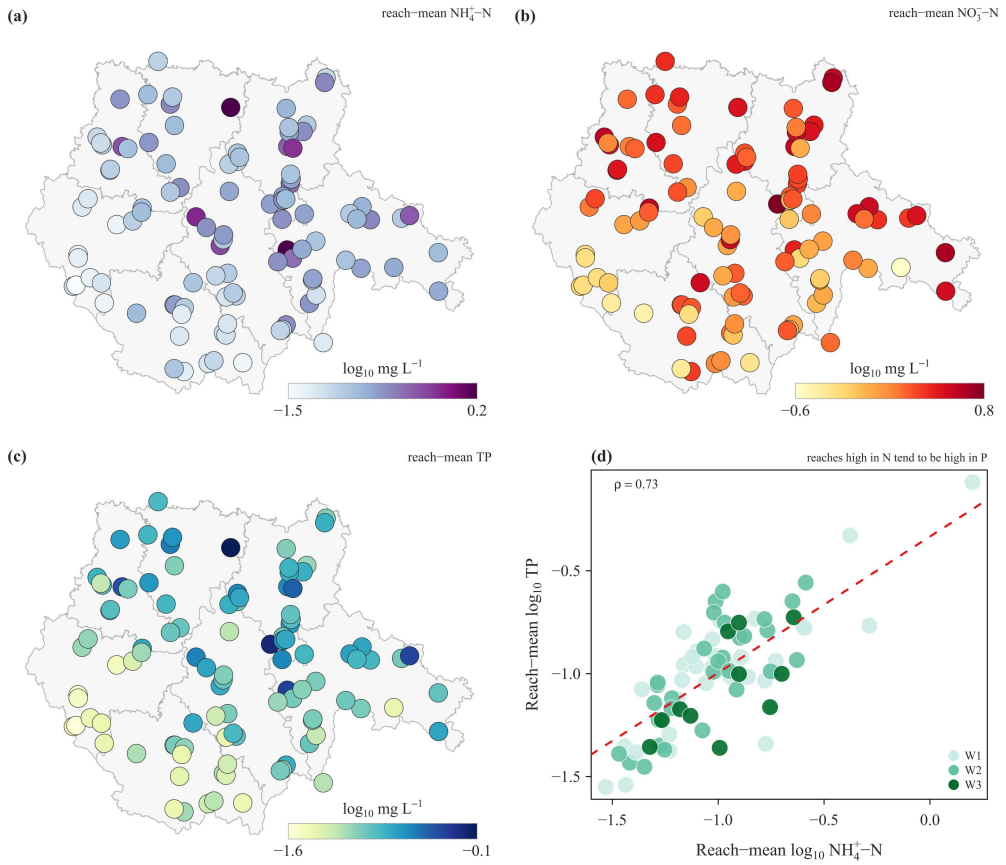



**Figure 5.17.** Reach-mean nutrient fields and their co-variation. (a–c) Reach-mean $\log_{10}$ concentrations of $NH_4^+$-N, $NO_3^-$-N and TP over 1990–2020, each on its own sequential color scale. (d) Reach-mean $\log_{10}$ TP against reach-mean $\log_{10}$ $NH_4^+$-N, coloured by width class; the quoted Spearman coefficient is a between-reach correlation and is therefore larger than the observation-level value reported in Figure 5.11g. Formal spatial-structure statistics for the fields in (a–c) are given in Figure 5.16.

The conclusion remained unchanged across all stratifications (Figures 5.14f and 5.18). Spatially cross-validated $R^2$ values ranged from −0.10 to 0.07 across data streams and from −0.10 to 0.11 across seasons. Tightening the matchup window to same-day observations changed $R^2$ by less than 0.04, while restricting the analysis to clean objects changed it by no more than 0.04. No stratum reached the 0.15 skill threshold. For TP, the bootstrap produced a median $R^2$ of 0.03 and a 95% interval of −0.11 to 0.11, which

included zero. Because the complete pipeline was refitted within each stratum rather than evaluated using post hoc subsets of predictions, these analyses provide stratum-specific robustness checks rather than summaries of a single model fit.

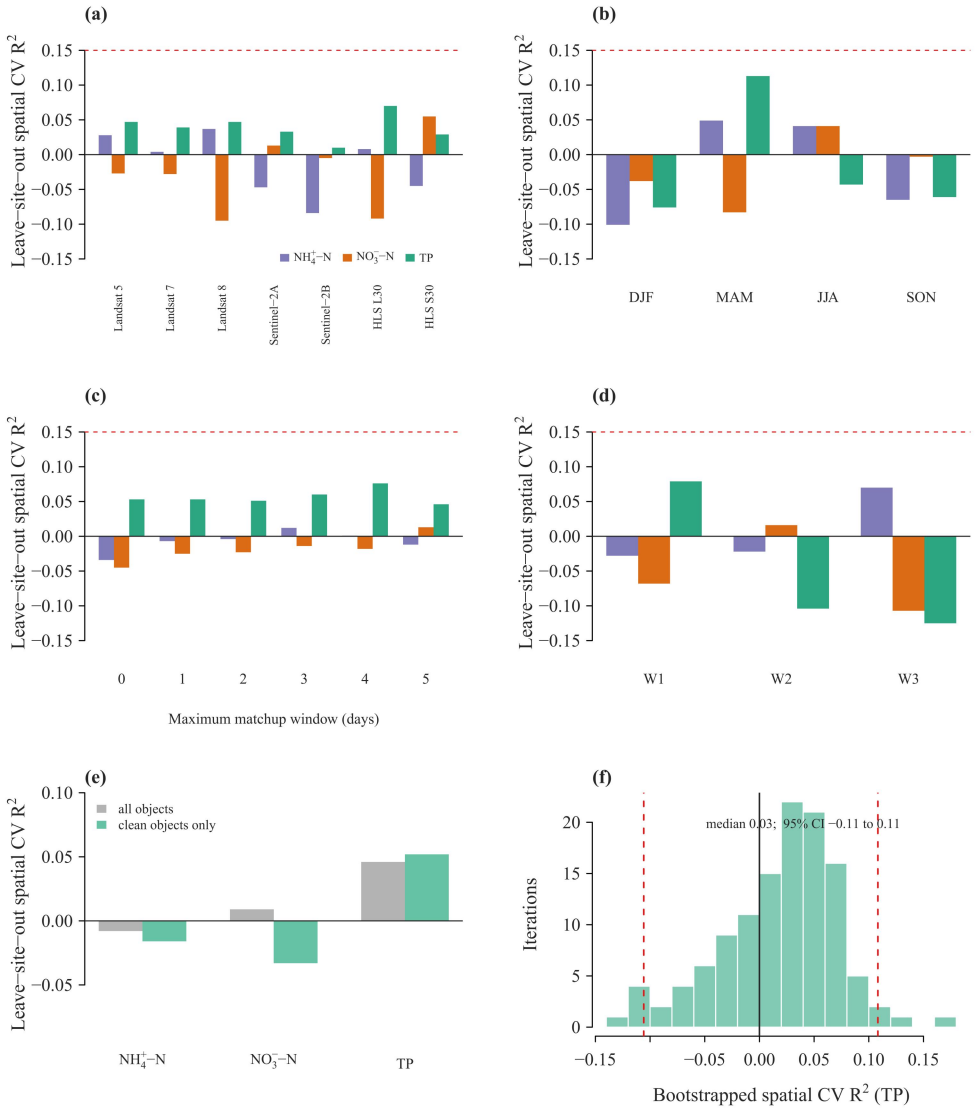



**Figure 5.18.** Full stratified robustness of the spatial-block CV verdict. $R^2$ under spatial-block CV, recomputed by refitting the entire pipeline within strata of (a) reflectance stream, (b) meteorological season, (c) maximum matchup window, (d) river-width class and (e) quality gate; the dashed line marks the 0.15 skill threshold. (f) Distribution

of the 120 bootstrap replicates of the spatial cross-validated $R^2$ for TP, obtained by resampling reaches with replacement, with the 2.5th and 97.5th percentiles marked. Figure 5.14f summarizes these strata on a single axis.

### 5.2.3. Validation design governs reported skill and reveals limited spatial transferability

The principal methodological finding of Study 2 was that validation design strongly affected the reported skill of dissolved-nutrient retrieval in narrow rivers. Using identical data, features, and learners, random CV produced $R^2$ values of up to 0.23, whereas LOSO CV yielded values close to zero. The largest difference reached $\Delta R^2 = 0.57$ across the evaluated model–scope combinations (Figure 5.14a, b). The sensors, atmospheric corrections, and models remained unchanged; only the assignment of observations to folds differed. Considered alone, the random-CV result could be interpreted as a modestly useful TP model. Spatial validation instead showed that predictive transferability to previously unseen reaches were not demonstrated. Therefore, the difference concerns the intended prediction domain rather than a minor aspect of the statistical procedure.

This finding is consistent with evidence from environmental mapping. Ploton et al. (2020) showed that conventional validation substantially overestimated forest-biomass prediction, while Meyer and Pebesma (2022) demonstrated the risks of prediction beyond a model's area of applicability. Karasiak et al. (2022) further linked optimistic performance estimates to spatial dependence between training and test samples. Conditions favouring this bias were pronounced in the present river network. Nutrient fields were spatially autocorrelated, with Moran's I reaching 0.44 (Figure 5.16a, e), and the semivariogram did not reach a sill within the sampled distances (Figure 5.16b). Meanwhile, the median distance between stations was only 2.9 km (Figure 5.16d), and individual reaches were sampled repeatedly. Random partitioning could therefore place observations from the same reach or correlated neighbouring

reaches in both training and test sets, producing interpolation-like performance rather than a reliable test of transfer to new locations.

Two additional diagnostics supported this interpretation. Firstly, the difference between random and spatial CV increased with channel width (Figure 5.14b). If mixed pixels alone explained the weak transferability, wider and more pixel-pure reaches would be expected to show improved spatial performance rather than a larger validation gap. The greatest inflation also occurred in the widest and most unevenly monitored class, indicating an influence of sampling geometry. Secondly, within-reach mutual information exceeded across-reach mutual information for every nutrient (Figure 5.14c). This pattern suggested that predictor–nutrient dependence was predominantly local, and that random CV benefited from reach-specific or neighbourhood information. Although this diagnostic does not identify the exact mechanism, it is consistent with the poor prediction of previously unseen reaches.

These findings have practical implications because ML retrievals are increasingly common in inland-water remote sensing, while spatially explicit validation remains less frequently applied (Sagan et al., 2020; Topp et al., 2020). For clustered and repeatedly sampled river data, random partitions may substantially overestimate transferable skill and could encourage premature replacement of field monitoring with satellite products. Two safeguards are therefore recommended: report random and spatially blocked CV together, and supplement predictive evaluation with spatial-structure and dependence diagnostics. Mutual information should be treated as supporting evidence rather than a pass–fail criterion because it does not capture all multivariate relationships. The results do not imply that all river-nutrient retrievals are unreliable. Physically meaningful retrieval remains possible for optically active targets and sufficiently wide, pixel-pure waters (Cao et al., 2020; Niroumand-Jadidi et al., 2021; Pahlevan et al., 2022). However, dissolved nutrient retrievals in narrow or spatially clustered river networks should be interpreted with caution until their transferability is demonstrated using spatial-block CV or other spatially blocked validation (Ludwig et al., 2023; Roberts et al., 2017; Wadoux et al., 2021).

### 5.2.4. A regional long-term recovery with no detectable width dependence

Within the fixed full-span panel, concentrations of all three nutrients declined markedly before stabilizing. From the first to the last sampled year, $NH_4^+$-N decreased by 80%, $NO_3^-$-N by 48%, and TP by 66%. Most of these reductions occurred during the 1990s, followed by an approximate plateau after 2007 (Figure 5.19a–e). Declines were widespread across the broader network: 57%, 82%, and 62% of the 76 well-sampled reaches exhibited negative trends in $NH_4^+$-N, $NO_3^-$-N and TP, respectively, while 38%, 51%, and 34% showed statistically significant declines (Figure 5.19f). Correspondingly, the distributions of reach-specific Theil–Sen slopes were centred below zero for all three nutrients (Figure 5.19h). Observation counts and record spans for individual reaches are provided in the supporting information of the corresponding publication.

The apparent tendency for wider rivers to exhibit stronger declines was not robust after accounting for record length. The unadjusted width effects yielded p-values of 0.023, 0.025, and 0.066 for $NH_4^+$-N, $NO_3^-$-N and TP, respectively; after controlling for record span, these increased to 0.68, 0.16, and 0.39 (Figure 5.19g). This attenuation is consistent with confounding by temporal coverage, as narrow reaches generally had shorter and more recent records (Figure 4.15c). The results therefore support a regional, temporally driven nutrient recovery without a detectable independent effect of channel width. This pattern is broadly consistent with continental analyses of agricultural rivers (Zhang et al., 2025, 2026) and with nutrient reductions reported across European rivers following decreased fertilizer use and improved wastewater treatment after 1989.

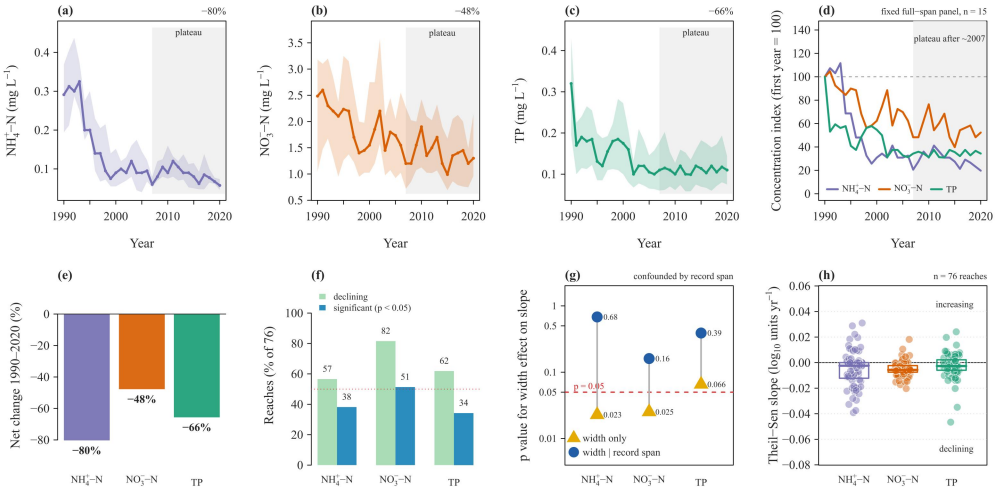



**Figure 5.19.** A regional, width-uniform long-term recovery. (a–c) Median annual trajectories of $NH_4^+$-N, $NO_3^-$-N and TP on the fixed full-span panel (n = 15), with interquartile bands and the post-2007 plateau shaded. (d) The same trajectories are indexed to the first year sampled. (e) Net change between the first and last year sampled. (f) Percentage of the 76 well-sampled reaches that declined, and that declined significantly. (g) The width effect on trend slope, before and after controlling for record span. (h) Distribution of per-reach Theil-Sen slopes.

### 5.2.5. The attributional inseparability boundary

Local land-surface change did not explain transferable spatial variation in nutrient recovery at any tested scale. Riparian land cover changed appreciably between 1990 and 2018. The mean magnitude of cropland change was greatest, reaching 6.0, 5.4, and 3.9 percentage points within the 50-, 150-, and 500-m buffers, respectively, followed by forest change at 3.0, 2.9, and 2.0 percentage points. Urban cover was the only class for which change increased with buffer radius (1.3, 1.8, and 2.0 points), whereas wetland and semi-natural cover changed least (2.3, 1.4, and 0.7 points; Figure 5.20a). Individual reaches lost up to 80 percentage points of cropland or gained up to 44 percentage points, producing a spatially heterogeneous pattern (Figures 5.20e, f, and 5.21).

These changes had little predictive value for nutrient recovery. Regressing reach-specific slopes against multiscale land-cover change produced weak in-sample ordinary least-squares fits ($R^2$ = 0.02–0.18), with the highest value obtained for TP within the 150-m buffer. Under leave-block-out spatial CV using the same folds as the retrieval analysis, $R^2$ values ranged from −97 to −1.3 across all parameters and scales, indicating performance below a mean-only benchmark (Figure 5.20b). None of the nine parameter–scale combinations reached the 0.15 threshold, and the results were consistent across buffer radii, making a simple scale mismatch unlikely within the tested range (Figure 5.20d; Table 5.7). TP residuals retained significant spatial structure at every scale (Moran's I, p = 0.001, 0.002, and 0.021). $NH_4^+$-N residuals approached significance at 50 m (p = 0.062), whereas $NO_3^-$-N residuals showed no significant spatial autocorrelation at any scale (p = 0.65–0.74; Figure 5.20c). These results suggest that TP recovery reflected spatially structured processes not represented by the local land-cover metrics.

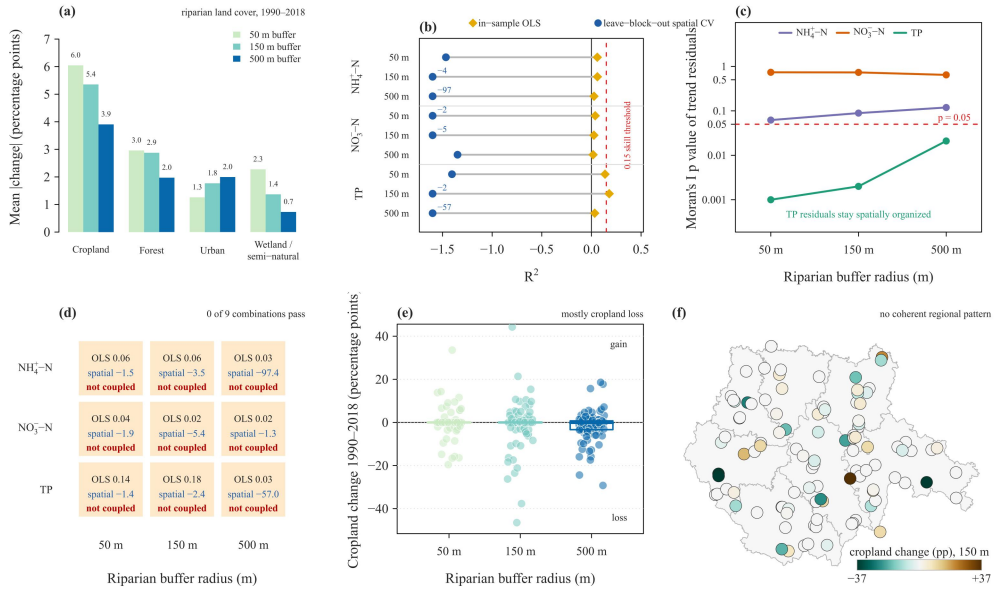



**Figure 5.20.** The attributional inseparability boundary. (a) Mean absolute change in riparian land cover between 1990 and 2018 by class and buffer radius. (b) In-sample OLS $R^2$ of per-reach trend slopes on multi-scale land-cover change compared with

leave-block-out spatial cross-validated $R^2$ for the same nine parameter-scale combinations; spatial values are truncated at −1.6 for display, and the three most extreme values are labelled. (c) Moran's I p value of the trend residuals. (d) Verdict matrix: none of the nine combinations reaches $R^2 > 0.15$. (e) Distribution of per-reach cropland change at the three buffer radii. (f) Spatial pattern of cropland changes at the 150-m buffer.

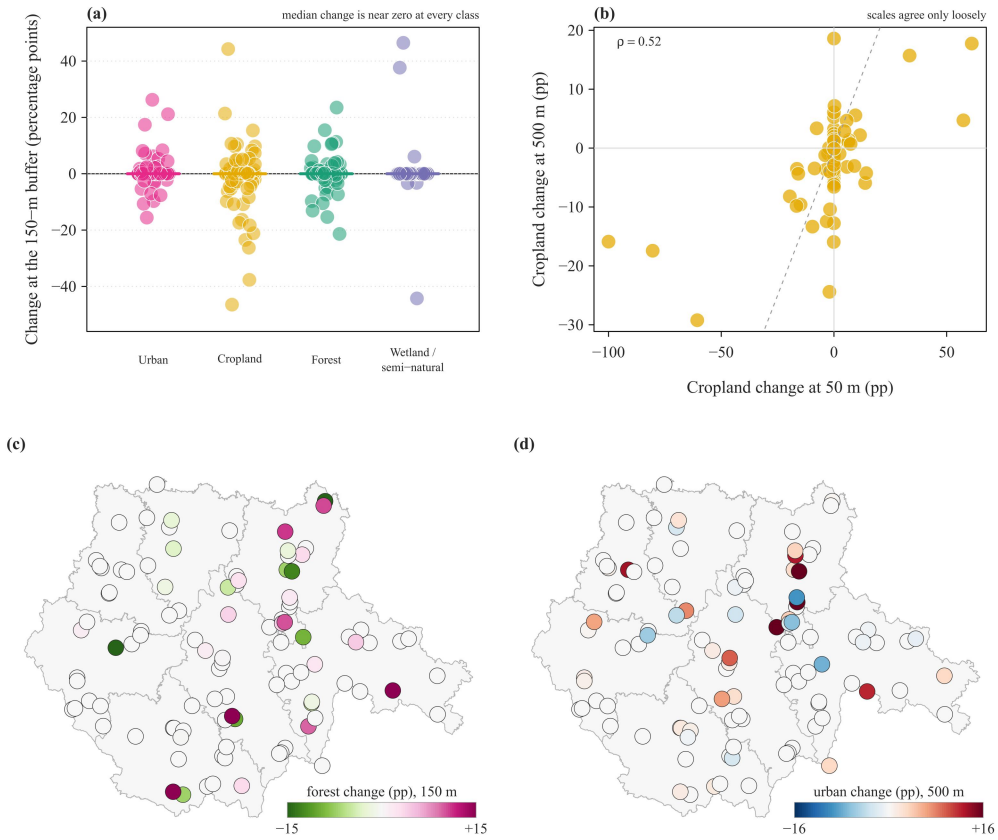



**Figure 5.21.** Detail of riparian land-cover change, 1990–2018. (a) Distribution of per-reach change at the 150-m buffer for the four Level-1 classes. (b) Cropland changes at the 500-m buffer against the same quantity at the 50-m buffer, with the 1:1 line. (c) Spatial pattern of forest change at 150 m. (d) Spatial pattern of urban change at 500 m. Figure 5.20a, e, f gives the class-mean magnitudes and the cropland pattern.

**Table 5.7.** Spatially cross-validated relationships between multi-scale riparian land-cover change and reach-specific nutrient trends.

| Parameter | Buffer (m) | OLS $R^2$ | Spatial-CV $R^2$ | Moran residual p | Verdict |
|---|---|---|---|---|---|
| $NH_4^+$-N | 50 | 0.06 | −1.47 | 0.062 | Not coupled |
| $NH_4^+$-N | 150 | 0.06 | −3.54 | 0.089 | Not coupled |
| $NH_4^+$-N | 500 | 0.03 | −97.42 | 0.119 | Not coupled |
| $NO_3^-$-N | 50 | 0.04 | −1.92 | 0.737 | Not coupled |
| $NO_3^-$-N | 150 | 0.03 | −5.36 | 0.733 | Not coupled |
| $NO_3^-$-N | 500 | 0.02 | −1.35 | 0.645 | Not coupled |
| TP | 50 | 0.14 | −1.41 | 0.001 | Not coupled |
| TP | 150 | 0.18 | −2.41 | 0.002 | Not coupled |
| TP | 500 | 0.03 | −56.96 | 0.021 | Not coupled |

Notes: None of the nine nutrient-scale combinations exceeded the predefined skill threshold of $R^2 > 0.15$. A negative spatial cross-validated $R^2$ indicates predictive performance worse than a mean-only benchmark for the held-out observations. The most extreme negative values indicate unstable out-of-block predictions due to the small number of reaches within each spatial block. "Moran residual p" denotes the permutation-based p-value for Moran's I calculated from the regression residuals using k-nearest-neighbor weights (k = 5).

Evidence from other regions supports this interpretation. Post-transition declines in nutrient levels in Central and Eastern European rivers have been linked to broad changes in climate and environmental management. Satellite observations have also documented large-scale changes in suspended sediment, organic pollution, and dissolved oxygen across river systems (Dethier et al., 2022; Sun et al., 2025; Tu et al., 2024; Yan et al., 2025; Zhang et al., 2025). Together, these studies show that important drivers of river water quality change operate at basin-to-continental scales and may not be represented by local land-cover metrics. This attributional boundary has direct implications for eutrophication management. The recovery observed here contains a strong regional component, but the present analysis cannot reliably separate it from local effects. Improvements at individual reaches should therefore not be attributed to local interventions based solely on reach-scale correlations. Some apparent local successes may instead reflect the shared basin-wide trajectory. Legacy nutrient stores may further delay or obscure water-quality responses to management for decades, increasing the risk of incorrect attribution (Van Meter et al., 2018). Robust evaluation

of intervention effectiveness therefore requires controlled paired-catchment designs or regionally explicit models. These approaches can distinguish common regional forcing from local intervention effects and spatially correlated background recovery.

### 5.2.6. The double inseparability boundary and the protocol it implies

Synthesizing the optical retrieval and driver-attribution results yields a single concept: the double inseparability boundary (Figure 5.22). The first is an optical boundary, whereby the studied nutrient concentrations could not be separated among reaches using medium-resolution reflectance. This limitation was primarily physical rather than learner-specific. At Landsat resolution, the median narrow channel occupied only 16% of a 30-m pixel (Figures 4.16a, b), leaving substantial contributions from banks, riparian vegetation, and adjacency effects. This constraint helps explain the weak across-reach spectral associations (Figure 5.12), the stronger within-reach than across-reach dependence (Figure 5.14c), and the apparent skill produced by spatial leakage under random CV. Plotting transferable skill against validation-induced inflation placed all three nutrient variables at or below the weak-signal threshold (Figure 5.22a). TP showed the strongest apparent optical association, consistent with its particle-associated fraction, whereas $NH_4^+$-N showed the weakest. This represents a narrow-river manifestation of the broader difficulty of retrieving non-optically active constituents from water reflectance (Mouw et al., 2015; Odermatt et al., 2012; Sagan et al., 2020).

The second, attributional boundary concerns the inability to resolve the real, substantial, and width-uniform regional recovery (Figure 5.19) into local reach-scale land-surface drivers. Multiscale riparian land-cover change explained no transferable across-reach variation under spatial CV, while TP residuals retained significant spatial structure (Figure 5.20). The two boundaries differ in mechanism: the first reflects a resolution-dependent observational constraint, whereas the second reflects a mismatch between local predictors and regionally coherent forcing. Their consequence is similar, as local satellite-derived information becomes insufficient to distinguish nutrient conditions or trends among reaches. Treating them jointly converts two negative results

into a testable diagnostic (Figure 5.22b). The same questions, whether a local pixel separates nutrient concentrations among reaches and whether local predictors separate their trends, can be evaluated in other monitoring networks using the four-step protocol developed here (Figure 5.22c). With many channels only 2–10 m wide, a median nearest-neighbor station spacing of 2.9 km, seven reflectance streams, and great post-1989 regional change, South Bohemia provides a demanding test case. Similar boundaries may occur in other agricultural headwater networks, but their occurrence and magnitude require explicit evaluation.

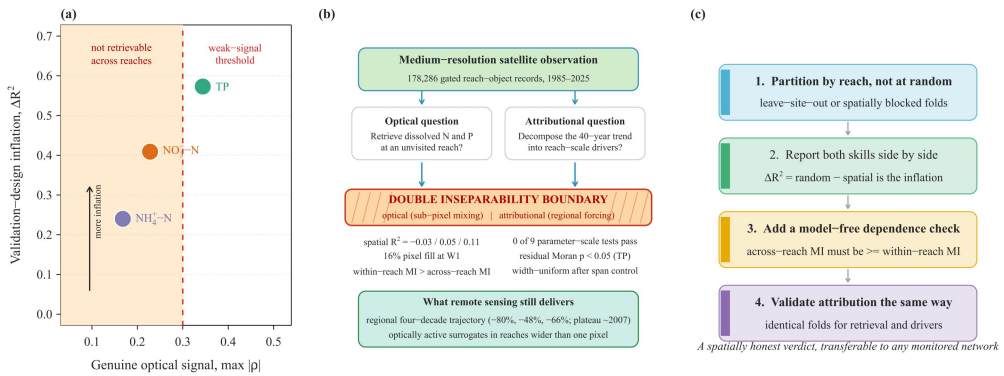



**Figure 5.22.** The double inseparability boundary and the validation protocol it implies. (a) Retrievability landscape: the genuine optical signal of each parameter, measured as the strongest feature correlation, against the inflation introduced by random validation. All three nutrients fall within the weak-signal region, where reported skill is dominated by the validation design. (b) The two boundaries, optical and attributional, that separate the questions a medium-resolution archive can answer from those it cannot, with the supporting evidence and what remains deliverable. (c) A four-step, transferable protocol for spatially honest verdicts in any monitored network.

### 5.2.7. What remains feasible and the limits of the evidence

Defining these boundaries clarifies the feasible scope of river water-quality monitoring. The long-term in-situ network resolves a coherent, region-wide nutrient decline after 1990 and a subsequent plateau, consistent with continental evidence of river recovery. These aggregated trajectories are valuable for evaluating large-scale nutrient policies, even where satellite imagery cannot predict dissolved nutrient concentrations at unvisited reaches. Remote sensing remains useful for optically active variables, including turbidity, chlorophyll, and bloom occurrence, particularly in wide or pixel-pure waters (Hou et al., 2022; Topp et al., 2021; Zhang et al., 2021).

For dissolved constituents that medium-resolution sensors cannot resolve in narrow channels, further tuning of ML algorithms using the same imagery is unlikely to overcome the physical information limit. Progress instead requires complementary measurements. Sub-metre imagery can reduce spatial mixing, while hyperspectral data may improve constituent discrimination where spatial resolution is adequate. Dense in situ sensor networks can capture local variability (Rode et al., 2016), while hybrid designs can combine spatially continuous remote sensing indicators with direct chemical measurements (Lehmann et al., 2023; Pi et al., 2022). The object-level, quality-gated, multi-sensor framework provides a consistent basis for testing retrievability before operational accuracy is claimed and for monitoring optically active variables in wider reaches. Satellite observations can therefore characterize regional environmental context and aggregated optical change. However, regulatory assessment or source attribution of dissolved nutrients in individual narrow reaches requires in situ measurements. Defensible monitoring architecture should assign each technology only to the variables and spatial scales for which it has demonstrated physical sensitivity and predictive validity. The double boundary thereby becomes a monitoring-design principle rather than simply a methodological limitation.

Three limitations qualify these conclusions. Firstly, the 100 m CORINE resolution is coarse relative to the median reach width of approximately 9 m and may miss fine-scale riparian changes. Consistent results across the 50, 150, and 500 m buffers reduce,

but do not eliminate, concern about this scale mismatch (Figure 5.20d). Secondly, the fixed full-span panel is small (n = 15) and has limited spatial variability. It supports a widespread regional decline but cannot identify its drivers or isolate local causal effects. Third, the in-situ records begin in 1990, after the 1989 transition. Therefore, the analysis characterizes a post-transition trajectory rather than a formal pre- and post-transition contrast. These limitations principally constrain trend attribution. By contrast, weak optical transferability was consistently demonstrated across 178,286 multi-decadal object observations, sensors, seasons, width classes, and sensitivity tests (Figure 5.14f).

## 5.3. Study 3: Catchment attributes outperform reflectance in predicting oxygen-demand concentrations of narrow headwater rivers: a spatially validated four-decade reconstruction

This section presents the principal findings of Study 3, which assesses whether satellite-observable catchment attributes can provide a spatially transferable alternative to direct reflectance-based retrieval for predicting COD and $BOD_5$ in narrow headwater rivers and for reconstructing their four-decade dynamics. This work has been submitted for consideration as the following manuscript:

**Shen, F.**, Li, J., Fan, S., Li, N., Li, Z., Xue, M.*, Chen, Z.*. 2026. Catchment attributes outperform reflectance in predicting oxygen-demand concentrations of narrow headwater rivers: a spatially validated four-decade reconstruction. Manuscript submitted to ***Journal of Hydrology***.

### 5.3.1. The optical non-retrievability boundary for COD and $BOD_5$

The multi-sensor reach-object archive comprised 178,286 cloud-free observations across 103 reaches during 1985–2025. Seven product streams derived from Landsat 5, 7, 8, and 9, Sentinel-2, and the HLS products contributed to the archive. Landsat 5 covered the period from the mid-1980s to the early 2010s, Landsat 7 entered the record in 1999, and Landsat 8 in 2013. Sentinel-2 and both harmonized product streams substantially increased observation density from 2015, while Landsat 9 contributed from 2021. Consequently, annual object counts increased from several hundred in the late 1980s to more than 20,000 by the mid-2020s, although the Landsat series maintained temporal coverage during the earlier decades (Figure 5.23c). Product-specific totals ranged from 4,998 observations for Landsat 9 to 49,009 for HLS S30 (Table 4.19). After observations containing negative reflectance values were excluded, 175,591 records remained.

Temporal matching with in situ sampling produced 17,200 COD and 17,658 $BOD_5$ reflectance–concentration pairs from six products (Figure 5.23e). Landsat 9 contributed no matchups because water-quality monitoring ended in 2020, before the mission was launched in 2021. Because native and harmonized products were matched separately, a single field sample could generate more than one product-specific matchup. The matchup totals therefore differed from the underlying in situ datasets, which contained 16,667 COD and 18,885 $BOD_5$ measurements collected between 1985 and 2020. Median matchup counts per reach were 115 for COD and 149 for $BOD_5$ (Figure 5.23b). Among the 103 reaches, 60 were classified as W1 (<10 m), 30 as W2 (10–30 m), and 13 as W3 (>30 m; Figure 5.23d). These classes contributed 8,151, 6,201, and 2,848 COD matchups and 8,477, 6,278, and 2,903 $BOD_5$ matchups, respectively. Fifteen reaches had records spanning at least 30 years, comprising one W1, nine W2, and five W3 reaches (Figure 5.23a, f).

Median concentrations were 23.7 mg $L^{-1}$ for COD and 2.9 mg $L^{-1}$ for $BOD_5$, whereas the corresponding means were 28.0 and 3.9 mg $L^{-1}$. IQRs were 17.0–33.0 mg $L^{-1}$ for COD and 2.0–4.6 mg $L^{-1}$ for $BOD_5$. Both variables were strongly right-skewed

on the original scale, with skewness values of 10.2 and 21.8, respectively. $Log_{10}$ transformation reduced these values to 0.45 and 0.46, supporting model development on the transformed scale (Figure 4.18f, g; Table 4.18).

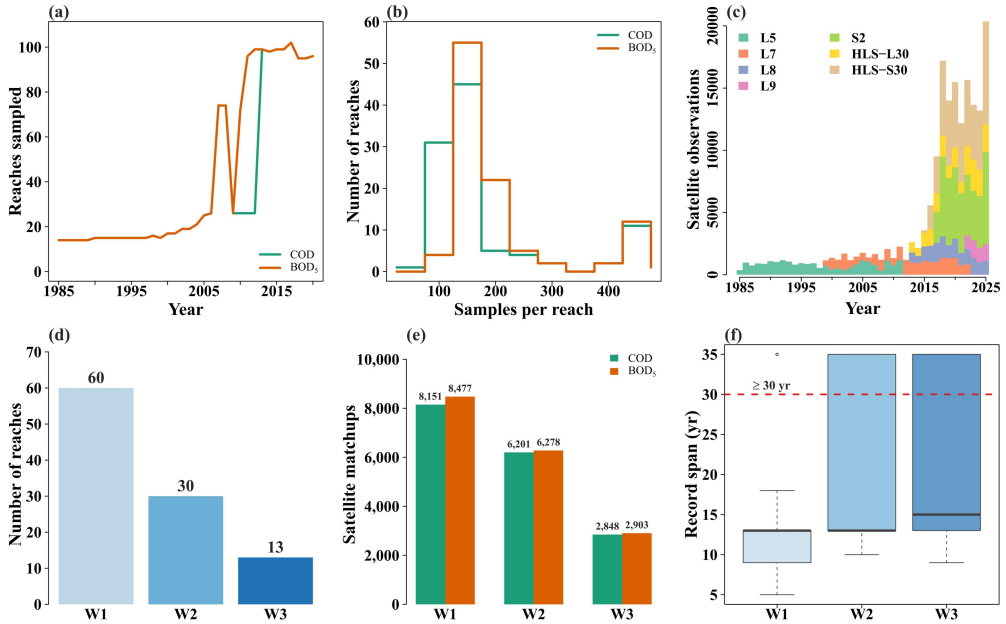



**Figure 5.23.** Data and monitoring overview. (a) Reaches sampled in situ per year; (b) samples per reach; (c) satellite observations by surface-reflectance product and year, for the seven products drawn from five missions (the two HLS products are shown separately from their native Landsat and Sentinel-2 counterparts); (d) river-width distribution (W1 = 60, W2 = 30, W3 = 13); (e) satellite–in-situ matchups by width class; (f) record-span distribution by width class.

Under LOSO validation, the spectral RF achieved Nash–Sutcliffe efficiencies of 0.02 for COD and 0.03 for $BOD_5$ across the full dataset, indicating negligible transferable skill at previously unseen reaches. In contrast, random ten-fold CV returned an NSE of 0.17 for both indicators, corresponding to an apparent skill inflation of 0.15 (Figure 5.24a). This contrast persisted across river-width classes. Random-validation NSEs for W1, W2, and W3 were 0.17, 0.11, and 0.19 for COD and 0.16, 0.14, and 0.25 for $BOD_5$, respectively. The corresponding LOSO values were 0.03,

−0.02, and −0.07 for COD and 0.04, 0.01, and −0.03 for $BOD_5$ (Figure 5.24b). Apparent skill inflation averaged approximately 0.13 in W1 and increased to 0.26 for COD and 0.28 for $BOD_5$ in W3.

Higher spatial resolution did not improve cross-reach transferability. Within W1, native 10 m Sentinel-2 observations returned NSE values of 0.01 for both COD and $BOD_5$, compared with 0.05 for the corresponding 30 m sensor subsets. Across all width classes, Sentinel-2 NSEs ranged from 0.01 to 0.04 for COD and from 0.01 to 0.08 for $BOD_5$, whereas the 30 m results ranged from −0.07 to 0.05 and from −0.04 to 0.05, respectively. No sensor-by-width combination exceeded an NSE of 0.08 (Figure 5.24c). Within-reach temporal correspondence was modest: median Spearman correlations between predicted and observed concentrations were 0.21 for COD and 0.23 for $BOD_5$. Correlations were highest in W1 (0.25–0.26), intermediate in W2 (0.17–0.20), and lowest in W3 (0.09–0.16), with only a minority of reaches exceeding 0.30 (Figure 5.24d). The network-wide median channel width was 9.2 m, while the W1 median was only 5.0 m, leaving most narrow reaches sub-pixel even at the 10 m Sentinel-2 resolution and substantially below the 30 m Landsat/HLS resolution (Figure 5.24e). Concentrations were also low and strongly right-skewed (Figure 5.24f). Collectively, these results show that spectral retrieval did not generalize across unmonitored reaches, although limited within-reach temporal information remained. The complete set of performance metrics is reported in Table 5.8.

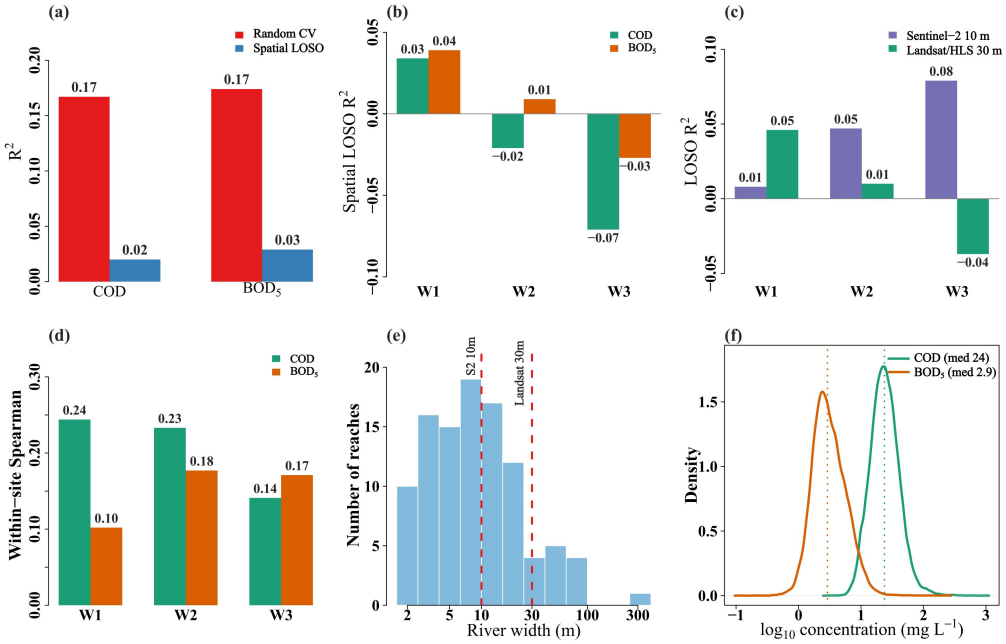



**Figure 5.24.** Optical retrieval skill for COD and $BOD_5$. (a) Skill under random and leave-one-site-out validation for the full sample; (b) leave-one-site-out values by width class; (c) values for native 10 m Sentinel-2 and for the 30 m Landsat/HLS subset; (d) median per-reach Spearman correlation between predicted and observed values; (e) channel-width distribution relative to the 10 m and 30 m pixel scales; (f) concentration distributions of COD and $BOD_5$ on a $\log_{10}$ scale.

**Table 5.8.** Optical retrieval skill for COD and $BOD_5$: random versus leave-one-site-out (LOSO) CV by width class, and the LOSO comparison between native 10 m Sentinel-2 and the 30 m Landsat/HLS subset.

| Parameter | Width | n | NSE (random) | NSE (LOSO) | Inflation | Median within-site ρ | NSE, Sentinel-2 10 m | NSE, Landsat/HLS 30 m |
|---|---|---|---|---|---|---|---|---|
| COD | All | 17,200 | 0.17 | 0.02 | 0.15 | 0.21 | 0.03 | 0.03 |
| COD | W1 | 8,151 | 0.17 | 0.03 | 0.13 | 0.25 | 0.01 | 0.05 |
| COD | W2 | 6,201 | 0.11 | −0.02 | 0.13 | 0.17 | −0.01 | −0.01 |
| COD | W3 | 2,848 | 0.19 | −0.07 | 0.26 | 0.09 | 0.04 | −0.07 |
| $BOD_5$ | All | 17,658 | 0.17 | 0.03 | 0.15 | 0.23 | 0.04 | 0.03 |
| $BOD_5$ | W1 | 8,477 | 0.16 | 0.04 | 0.12 | 0.26 | 0.01 | 0.05 |
| $BOD_5$ | W2 | 6,278 | 0.14 | 0.01 | 0.13 | 0.20 | 0.05 | 0.01 |
| $BOD_5$ | W3 | 2,903 | 0.25 | −0.03 | 0.28 | 0.16 | 0.08 | −0.04 |

Notes: Retrieval uses a random forest trained on object-level spectral features. NSE is computed on $\log_{10}$ concentrations and is equivalent to the out-of-sample coefficient of determination; inflation is the difference between the random and LOSO values. The final two columns repeat the LOSO test separately on the native 10 m Sentinel-2 subset and on the 30 m Landsat and HLS subset and are computed on those subsets rather than on the full sample, so they are not expected to reproduce the pooled LOSO column. All values are dimensionless.

The pattern of model failure indicates a predominantly physical rather than algorithmic limitation. The failure cannot be attributed solely to model choice, because a flexible RF trained on the complete spectral feature set still produced near-zero predictive skill at previously unobserved reaches. Higher spatial resolution was also insufficient: native 10 m Sentinel-2 observations performed no better than the 30 m products within the narrowest width class. These results identify channel geometry as the most plausible explanation. Given a network median width of 9.2 m and a W1 median of 5.0 m, the median narrow channel occupies only approximately half of a 10 m pixel in the cross-channel dimension and a much smaller proportion of a 30 m pixel. Object-level reflectance is therefore strongly influenced by riverbanks, riparian vegetation, surrounding fields, and adjacency effects, rather than primarily by water-leaving radiance. This limitation is particularly relevant to headwater networks, in which channels commonly remain narrower than medium-resolution pixels. Furthermore, the higher apparent skill under random CV and the stronger within-reach than across-reach dependence is consistent with models that exploit persistent site characteristics rather than learn transferable spectral relationships. Such site-specific information cannot support operational prediction at unmonitored reaches.

A compositional limitation compounds this geometric constraint. COD and $BOD_5$ are operational bulk measures that integrate constituents with contrasting optical properties. Chromophoric dissolved material and suspended particles absorb or scatter light and may therefore be detectable, whereas colorless dissolved and readily biodegradable materials can contribute substantially to oxygen demand without producing a distinct reflectance signal. The present study did not directly fractionate these components. Therefore, this mechanism should be interpreted as a physically plausible explanation rather than a demonstrated compositional pathway. At median concentrations of 23.7 mg $L^{-1}$ for COD and 2.9 mg $L^{-1}$ for $BOD_5$, the optically active fraction may have been too weak to overcome land contamination and sensor noise. Moreover, reductions in suspended and particulate material during water-quality recovery may further weaken the available optical proxy, even as dissolved oxygen

demand persists. Optical detectability may consequently decline as water quality improves, precisely when monitoring must remain sensitive to relatively small residual concentrations. This limitation distinguishes COD and $BOD_5$ from more optically responsive constituents, including chlorophyll, turbidity, and suspended sediment. Residual errors from land-oriented atmospheric correction over dark water may have added noise, but the consistent failure across sensors and resolutions suggests they were not the sole cause.

These findings delimit, rather than contradict, the potential of satellite monitoring for river water quality. Basin-scale Landsat analyses have reconstructed organic pollution histories in large rivers (Yan et al., 2025), dissolved oxygen has been estimated across lakes using remotely sensed proxies (Tu et al., 2024), and phosphorus has been retrieved where it covaries with optically active constituents (Xiong et al., 2022). Such applications generally benefit from water-dominated pixels, stronger optical proxies, or greater concentration contrasts. The narrow, relatively low-concentration ranges examined here provide none of these advantages. Therefore, the boundary identified in this study applies specifically to the direct retrieval of COD and $BOD_5$ from medium-resolution reflectance in narrow rivers, rather than to remote sensing of rivers.

### 5.3.2. Long-term trends and catchment-attribute prediction

On the fixed panel of 15 long-record reaches, both indicators exhibited non-monotonic temporal trajectories (Figures 5.25a, b). COD declined from approximately 26 mg $L^{-1}$ in the mid-1980s to 18–20 mg $L^{-1}$ during 2003–2010, representing a reduction of roughly 25%, before recovering to approximately 26 mg $L^{-1}$ by 2020. $BOD_5$ peaked around 1989, declined throughout the 1990s, and subsequently stabilised at approximately 2.6–2.7 mg $L^{-1}$. Reach-specific Theil–Sen slopes also varied systematically among width classes: W1 reaches showed predominantly positive trends, whereas W3 reaches showed predominantly negative trends for both indicators. This width-related contrast persisted after adjustment for differences in record span (Figures 5.25c, g, h), indicating that it was not solely attributable to the shorter and more recent

records of narrow reaches. Within individual reaches, the direction of change was generally consistent between COD and $BOD_5$. Because these regional trajectories were derived from a fixed panel, they were not confounded by changes in monitoring-network composition.

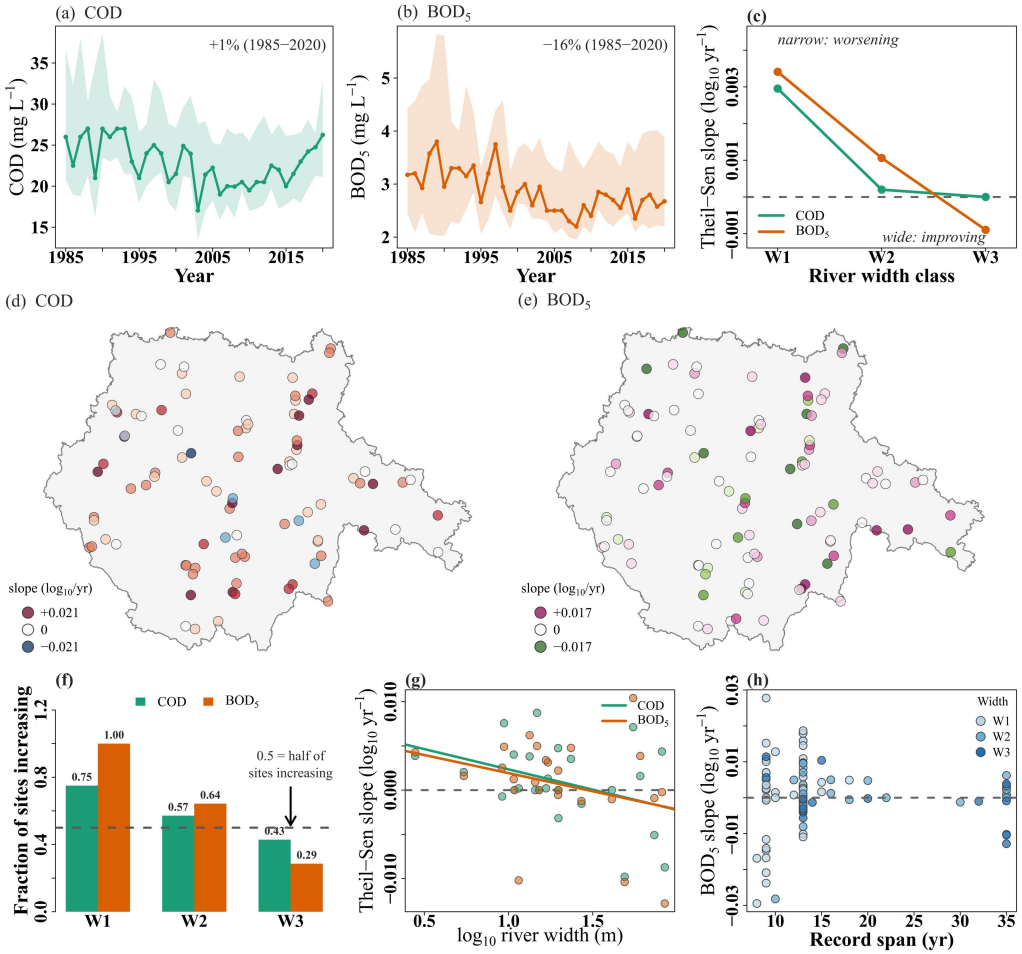



**Figure 5.25.** Long-term trends, 1985–2020. Regional trajectories of COD (a) and $BOD_5$ (b). (c) Theil–Sen slope ($log_{10}$ $yr^{-1}$) by width class. Maps of per-reach slopes for COD (d) and $BOD_5$ (e). (f) Fraction of sites increasing by width class. (g) Slope versus $log_{10}$ river width. (h) $BOD_5$ slope versus record span.

The catchment-based RF, using only catchment attributes and wastewater-related predictors, outperformed the spectral model at held-out reaches (Figure 5.26a–c). On the $\log_{10}$ modelling scale, $BOD_5$ achieved an NSE of 0.40, a KGE of 0.47, a KGE′ of 0.46, a PBIAS of +2.3%, and an RSR of 0.78, whereas COD achieved corresponding values of 0.29, 0.33, 0.33, +1.3%, and 0.84. After back-transformation, $BOD_5$ and COD yielded NSE values of 0.27 and 0.17, KGE values of 0.29 and 0.15, RMSE values of 1.7 and 12.0 mg $L^{-1}$, MAE values of 1.1 and 7.3 mg $L^{-1}$, and RPD values of 1.17 and 1.10, respectively. The lower original-scale efficiencies reflect the right-skewed concentration distributions (Table 5.9). Temporal hold-out validation, trained through 2010 and tested thereafter, produced NSE values of 0.47 for COD and 0.67 for $BOD_5$.

Width-specific performance was averaged over 11 repeated runs, with full ranges shown in Figure 5.26g. Mean $BOD_5$ NSE was 0.43 in W1 and 0.35 in both W2 and W3, indicating no performance disadvantage in the narrowest reaches. Mean COD NSE was 0.23, 0.35, and 0.26 in W1, W2, and W3, respectively. Run-to-run ranges reached 0.02 for the full dataset, and 0.04 within individual classes; consequently, the nearly identical W2 values for COD and $BOD_5$ (0.351 and 0.349) were not meaningfully different. Permutation importance ranked cropland fraction, elevation and forest cover above channel width, surface-water fraction, and wastewater predictors (Figure 5.26f). Site-mean concentrations were positively associated with cropland fraction for $BOD_5$ (Spearman's $\rho = 0.49$, $p < 0.001$) and COD ($\rho = 0.33$, $p < 0.001$; Figure 5.26i). Removing wastewater predictors reduced NSE to 0.33 for $BOD_5$ and 0.19 for COD and RPD to 1.12 and 1.05, respectively (Table 5.10), demonstrating their additional predictive value despite their lower importance rankings.

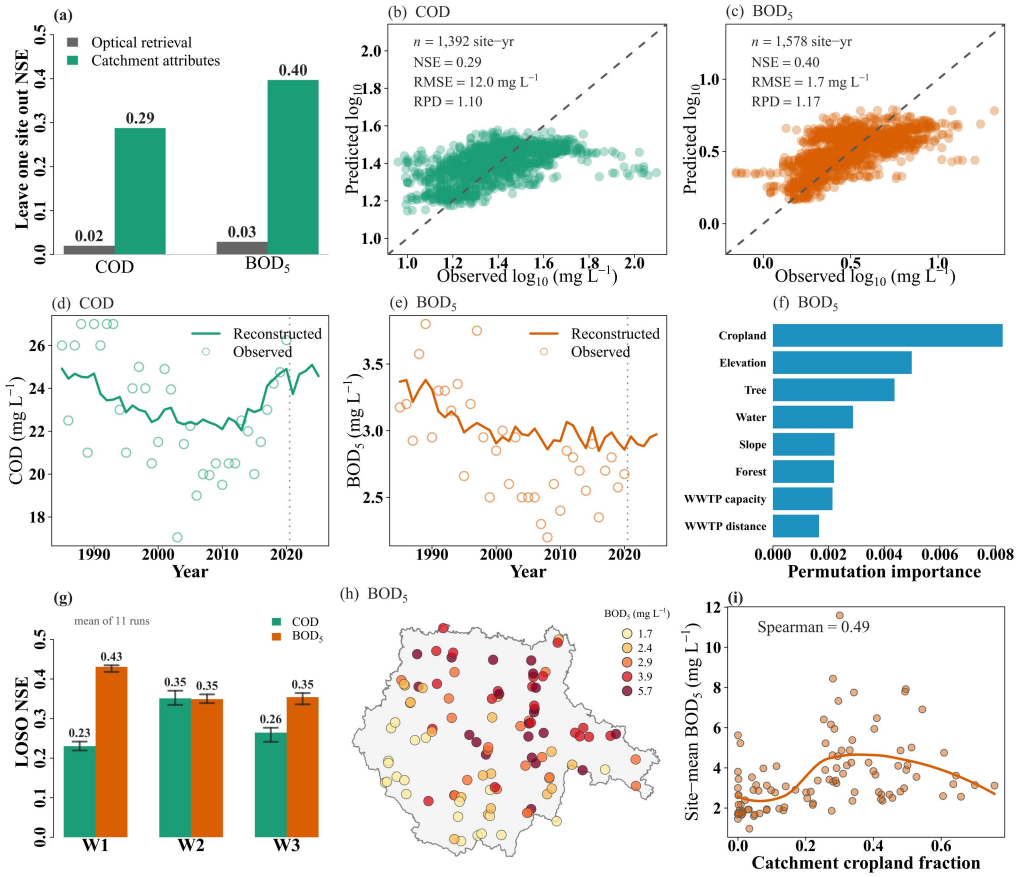



**Figure 5.26.** Catchment-attribute prediction and 1985–2025 reconstruction. (a) Leave-one-site-out skill for optical retrieval versus the catchment-attribute model. Predicted versus observed $\log_{10}$ concentrations for COD (b) and $BOD_5$ (c). Reconstructed and observed series for COD (d) and $BOD_5$ (e); the dotted line marks the start of the unmonitored 2021–2025 window. (f) Permutation importance of $BOD_5$ predictors. (g) Leave-one-site-out skill by width class, given as the mean over eleven repeat runs of the identical configuration, with whiskers spanning the full range across those runs; the between-class differences are of the same order as this run-to-run spread, so a single run must not be read as a ranking. (h) Map of the reconstructed mean $BOD_5$ over the unmonitored 2021–2025 window. (i) Observed site-mean $BOD_5$ versus catchment cropland fraction (Spearman = 0.49).

**Table 5.9.** Performance of the catchment-attribute random-forest models under leave-one-site-out CV, reported on the $\log_{10}$ scale used for fitting and on the untransformed concentration scale.

| Parameter | Subset | Scale | NSE | KGE | KGE′ | PBIAS (%) | RSR | RMSE | MAE | RPD |
|---|---|---|---|---|---|---|---|---|---|---|
| COD | All | $\log_{10}$ | 0.29 | 0.33 | 0.33 | +1.3 | 0.84 | 0.15 | 0.12 | 1.19 |
| COD | W1 | $\log_{10}$ | 0.23 | 0.28 | 0.28 | −0.5 | 0.88 | 0.16 | 0.12 | 1.14 |
| COD | W2 | $\log_{10}$ | 0.35 | 0.43 | 0.41 | +3.2 | 0.81 | 0.14 | 0.12 | 1.24 |
| COD | W3 | $\log_{10}$ | 0.26 | 0.29 | 0.29 | +1.0 | 0.86 | 0.16 | 0.10 | 1.17 |
| COD | All | concentration | 0.17 | 0.15 | 0.16 | −3.2 | 0.91 | 11.99 | 7.27 | 1.10 |
| $BOD_5$ | All | $\log_{10}$ | 0.40 | 0.47 | 0.46 | +2.3 | 0.78 | 0.16 | 0.13 | 1.29 |
| $BOD_5$ | W1 | $\log_{10}$ | 0.43 | 0.51 | 0.52 | −3.6 | 0.75 | 0.16 | 0.12 | 1.33 |
| $BOD_5$ | W2 | $\log_{10}$ | 0.35 | 0.47 | 0.43 | +9.0 | 0.81 | 0.16 | 0.13 | 1.24 |
| $BOD_5$ | W3 | $\log_{10}$ | 0.35 | 0.40 | 0.38 | +4.4 | 0.80 | 0.17 | 0.12 | 1.25 |
| $BOD_5$ | All | concentration | 0.27 | 0.29 | 0.31 | −5.6 | 0.85 | 1.70 | 1.08 | 1.17 |

Notes: Every value is the mean over eleven repeat runs of the identical configuration, because the between-class differences are of the same order as the run-to-run spread of the random forest; the full range across those runs is shown for NSE in Figure 5.26g. On that basis, the two W2 values (COD 0.351, $BOD_5$ 0.349) are indistinguishable, with COD exceeding $BOD_5$ in five of the eleven runs and falling below it in six. Models were evaluated on site-year annual medians (n = 1,392 for COD and 1,578 for $BOD_5$) across the 102 reaches with interpretable catchment predictors. RMSE and MAE are in $\log_{10}$ units on the $\log_{10}$ rows and in mg $L^{-1}$ on the concentration rows. Concentration-scale metrics are lower than log-scale metrics because the untransformed distributions are strongly right-skewed (Table 4.18), so a few extreme observations dominate the squared error. Under a temporal hold-out (train ≤ 2010, test > 2010), the full-sample NSE was 0.47 for COD and 0.67 for $BOD_5$. Within-reach year-to-year skill is far weaker: per-reach correlations between predicted and observed annual medians had medians of 0.54 for COD (66 reaches with ≥ 10 years) and 0.23 for $BOD_5$ (90 reaches), but per-reach NSE was negative at most reaches (medians −2.1 and −2.3), so the reconstruction resolves persistent differences between reaches rather than interannual variation within them.

Two additional tests constrained the interpretation of this predictive skill. Firstly, augmenting the predictor set with four derived hydrological descriptors, catchment water yield, a dilution-capacity index, and site-specific precipitation and temperature anomalies, changed cross-site NSE by no more than 0.01 in either direction (COD: 0.28 to 0.29; $BOD_5$: 0.40 to 0.39; Table 5.10). None of these variables ranked among the ten most important predictors. Site-mean concentrations were also unrelated to the dilution-capacity index (Spearman's rho = −0.02, p = 0.88 for COD; rho = −0.09, p = 0.36 for $BOD_5$) but were strongly associated with cropland fraction. These results are more consistent with catchment supply than with hydrological dilution as the dominant control on between-reach concentration differences across this network.

**Table 5.10.** Sensitivity of cross-site (leave-one-site-out) skill to the composition of the predictor set. Each pair of rows was run through identical folds with a fixed seed, so the difference is attributable solely to the predictors.

| Experiment | Configuration | COD NSE | COD KGE | $BOD_5$ NSE | $BOD_5$ KGE |
|---|---|---|---|---|---|
| Hydrological augmentation | Catchment attributes + point source (baseline) | 0.283 | 0.332 | 0.402 | 0.476 |
| Hydrological augmentation | + water yield, dilution index, precipitation and temperature anomalies | 0.294 | 0.348 | 0.393 | 0.471 |
| Point-source terms | Baseline (as above) | 0.283 | 0.332 | 0.402 | 0.476 |
| Point-source terms | Without the waste-water terms | 0.190 | — | 0.330 | — |
| Terrain-slope index | Baseline (separate run) | 0.296 | — | 0.404 | — |
| Terrain-slope index | Without the terrain-slope index | 0.285 | — | 0.404 | — |

Notes: All values are NSE and KGE on $\log_{10}$ concentrations. Repeat runs of the identical configuration differ due to random-forest stochasticity: across eleven seeds, the leave-one-site-out NSE varied by up to 0.02 for the full sample and by up to 0.04 within a single width class (Figure 5.26g), so differences of that magnitude are not interpreted as meaningful. Neither augmentation experiment changes any conclusion of the main text: the four hydrological descriptors alter skill by no more than 0.01, and none enters the ten most important predictors, and the terrain-slope index can be removed at no cost. Consistently, site-mean concentrations (mean $\log_{10}$ of all measurements per reach, the quantity plotted in Figure 5.26i) are uncorrelated with the dilution-capacity index (Spearman −0.02, p = 0.88 for COD; −0.09, p = 0.36 for $BOD_5$) but strongly correlated with catchment cropland fraction (+0.33 and +0.49, both $p < 0.001$), indicating a supply-controlled rather than dilution-controlled concentration field. By contrast, removing the waste-water terms costs 0.09 and 0.07 in NSE, which is well outside the run-to-run spread.

Secondly, the model captured spatial contrasts more effectively than temporal variability. Among reaches with at least 10 years of observations, median per-reach correlations between predicted and observed annual concentrations were 0.54 for COD (n = 66) and 0.23 for $BOD_5$ (n = 90). However, per-reach NSE was negative at most sites, with median values of −2.1 and −2.3, respectively. The reconstruction thus reproduced persistent differences among reaches but did not reliably capture year-to-year variability within individual reaches. Its cross-site predictive skill should not be interpreted as equivalent to temporal reconstruction skill at a given monitoring site.

Applying the catchment model to all 103 reaches for 1985–2025 generated annual reconstructed concentration fields, including model-based estimates for the unmonitored period of 2021–2025 (Figure 5.26d, e, h). Because no in situ measurements were available after 2020, these estimates were derived entirely from predictor data. Modelled COD and $BOD_5$ concentrations were generally highest in cropland-dominated lowland catchments and lowest in forested uplands, and this spatial gradient persisted throughout 2021–2025. The reconstructed regional trajectories exhibited weaker interannual variability than the observed long-record panel, consistent with the validation evidence that the model captured persistent spatial differences among reaches more effectively than year-to-year variation within individual reaches. Accordingly, the reconstruction is most defensibly interpreted as a sequence of spatially transferable annual estimates rather than as a gap-filled record of local temporal dynamics.

### 5.3.3. Hydrological basis and limits of catchment-attribute prediction

The catchment-attribute model demonstrates that information unavailable from the sub-pixel water surface can be partly recovered from the surrounding landscape. Under LOSO validation, its performance substantially exceeded that of the reflectance model, and $BOD_5$ skill was highest in W1, where optical retrieval failed most clearly. This inversion is hydrologically plausible. In small headwater catchments, local landscape attributes represent a larger proportion of the contributing area, source areas are generally closer to the channel, and land–river connectivity can preserve stable spatial signatures in stream chemistry. Larger rivers integrate more heterogeneous upstream sources and undergo longer, more varied transport and transformation pathways, thereby weakening their correspondence with local landscape descriptors (Abbott et al., 2018; Bieroza et al., 2024; Li et al., 2024). Thus, the small channel size that constrains optical retrieval can simultaneously increase the representativeness of local catchment attributes.

The spatial support of the predictors both reinforces and limits this interpretation. Each station was assigned to its corresponding fourth-order hydrological unit, which approximates the contributing area for many headwater reaches but represents only a non-cumulative, near-channel fraction of the drainage area that supplies wider rivers. The model therefore characterizes local catchment context rather than complete upstream influence. Consistent with this limitation, the width pattern was not universal. $BOD_5$ efficiency was higher in W1 than in W2 and W3 (0.43 versus 0.35 in both classes), whereas COD peaked in W2 (0.23, 0.35, and 0.26 for W1–W3). Moreover, reach-level errors were unrelated to hydrological-unit area for either COD (Spearman's $\rho = 0.08$, $p = 0.55$) or $BOD_5$ ($\rho = -0.11$, $p = 0.30$). The apparent width dependence should therefore be regarded as indicative rather than established. The difference between the indicators may reflect the greater sensitivity of $BOD_5$ to locally supplied biodegradable matter, whereas COD integrates a broader mixture of labile and refractory oxidizable materials that are affected by transport and processing (Catalán et

al., 2016b). However, because organic-matter composition and source contributions were not measured directly, this explanation remains a mechanistic hypothesis.

The hydrological augmentation tests further constrain interpretation. Adding catchment water yield, dilution capacity, and precipitation and temperature anomalies changed cross-site NSE by no more than 0.01, while site-mean concentrations were unrelated to the dilution-capacity index but positively associated with cropland fraction. These results indicate that the available landscape descriptors captured persistent spatial source patterns more effectively than the derived hydrological proxies. They do not demonstrate that discharge or dilution was unimportant. The water yields and dilution indices are simplified proxies and cannot represent reach-specific discharge, event connectivity, effluent timing, or transient in-stream processing. Moreover, annual medians suppress much of the seasonal and event-scale variability through which hydrological control is expressed. The observed pattern is nevertheless consistent with the relative temporal stability and chemostatic behavior reported for many solutes in managed catchments (Basu et al., 2010; Godsey et al., 2009; Musolff et al., 2015; van Vliet et al., 2023). Direct discharge measurements remain essential for converting concentrations into loads and resolving event-scale transport.

Cropland fraction, elevation, and forest cover provided the strongest predictive information, while wastewater variables made a smaller but measurable incremental contribution. Removing wastewater predictors reduced NSE from 0.40 to 0.33 for $BOD_5$ and from 0.29 to 0.19 for COD. Together with the positive concentration–cropland relationships, these results are consistent with the combined influence of diffuse agricultural pressure and point-source wastewater, both of which are well-established controls on agricultural headwater quality and regional organic pollution (Bieroza et al., 2024; Jones et al., 2023; Li et al., 2024). Nevertheless, permutation importance among correlated predictors does not quantify causal effects, and changes in predictive efficiency cannot be interpreted as source apportionment. The model identifies management-relevant covariates associated with spatial concentration

patterns, but it does not attribute observed concentrations to individual sources or interventions.

The broader implication is a shift in observational strategy: when the water surface cannot be resolved reliably, satellite and geospatial products can instead characterize the catchment supplying it. The required predictors are available from open pan-European or global datasets, although their temporal coverage and resolution differ. Transfer to other agricultural headwater networks should therefore be tested rather than assumed, using local observations, LOSO validation, and explicit assessment of whether new reaches fall within the model's area of applicability (Meyer and Pebesma, 2022, 2021). This supports a hybrid monitoring design in which spatially continuous catchment information identifies persistent gradients and guides sampling, while in situ measurements anchor local concentrations and temporal change (Palmer et al., 2015; Reichstein et al., 2019). Given the model's weaker representation of within-reach temporal variability, it is better suited to identifying persistently elevated or poorly represented reaches than to detecting short-term deterioration between surveys.

### 5.3.4. Scenario sensitivity to 2050

Relative to the current-driver baseline, the 2050 scenarios produced modest network-wide median changes but substantial variation among the 102 reaches (Figure 5.27). Intensification changed median COD and $BOD_5$ by −0.3% and +0.7%, respectively, whereas extensification changed them by −0.2% and −1.7% (Figures 5.27a, b). The +2 °C climate scenario increased median COD by 2.2% but $BOD_5$ by only 0.4%. When combined with climate change, intensification increased COD and $BOD_5$ by 2.7% and 0.3%, while extensification increased COD by 2.4% but reduced $BOD_5$ by 0.9%.

The $BOD_5$ response to land-use change varied among width classes. Intensification changed median $BOD_5$ by −0.5%, +6.0%, and +10.5% in W1, W2, and W3, respectively, compared with −4.0%, +1.1%, and +0.6% under extensification (Figure 5.27c; Table 5.11). Extensification reduced $BOD_5$ in 67% of W1 reaches, 45% of W2 reaches, and 38% of W3 reaches (Figure 5.27f). Although the overall association with cropland fraction was weak (Spearman's rho = 0.16), the largest reductions occurred in intermediate-cropland headwaters, while responses saturated at higher cropland fractions (Figure 5.27d). The mapped responses remained spatially heterogeneous under both land-use scenarios (Figures 5.27g–i).

Across the +1 to +3 °C scenarios, median $BOD_5$ changes remained below 1% in every width class (Figure 5.27e). Combining +2 °C warming with intensification changed $BOD_5$ by −0.4%, +6.0%, and +11.8% in W1–W3, compared with −4.6%, +0.7%, and +2.5% under extensification. Individual responses ranged from −41% to +48% for $BOD_5$ and from −29% to +37% for COD, although all width-class medians remained within −5% to +12% (Table 5.11). The principal result is the contrasting climate sensitivity of COD and land-use sensitivity of $BOD_5$, together with considerable spatial heterogeneity. These estimates represent within-range model sensitivities rather than deterministic forecasts for 2050 water quality.

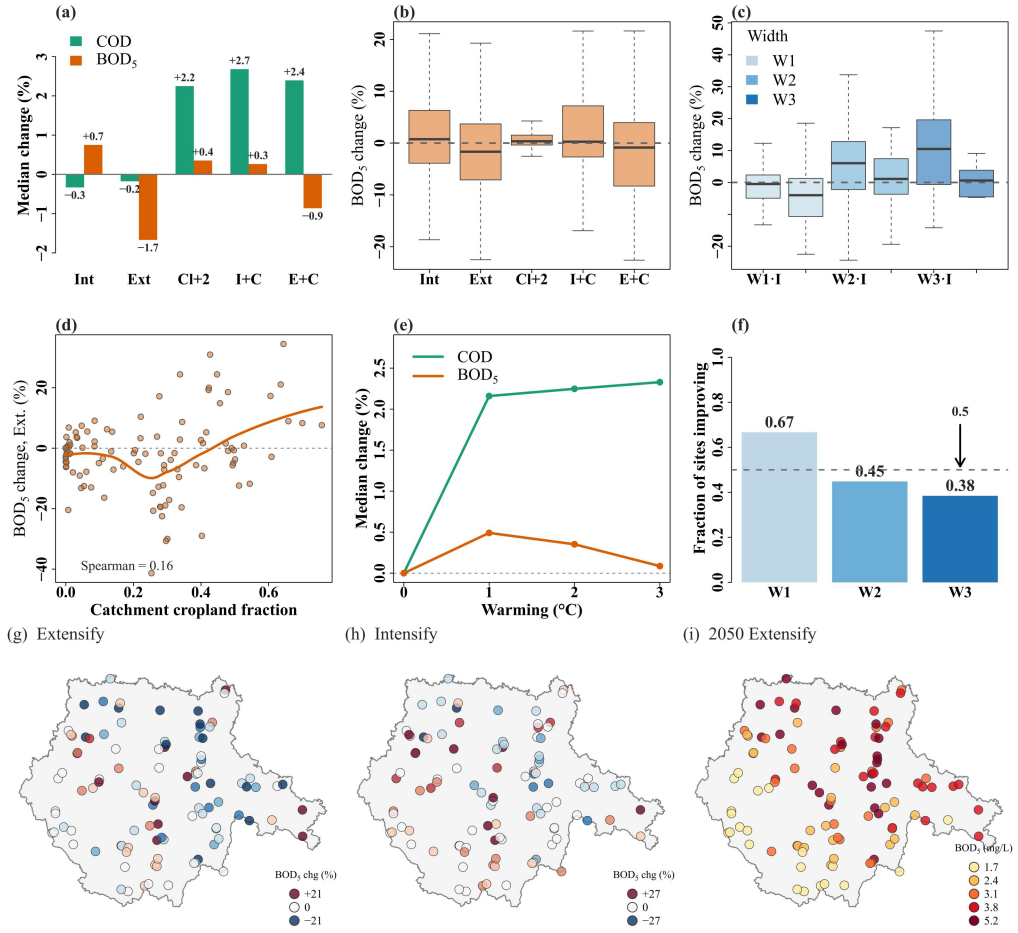



**Figure 5.27.** Sensitivity of COD and $BOD_5$ concentrations to 2050 driver scenarios. (a) Median percentage change for COD and $BOD_5$ across scenarios. (b) Distribution of $BOD_5$ changes across scenarios. (c) $BOD_5$ changes by width class under intensification. (d) Extensification $BOD_5$ change versus catchment cropland fraction (Spearman = 0.16). (e) Median change versus warming for COD and $BOD_5$. (f) Fraction of sites improving by width class under extensification. (g, h) Spatial pattern of $BOD_5$ changes under extensification (g) and intensification (h). (i) Modelled $BOD_5$ under the 2050 extensification scenario.

**Table 5.11.** 2050 scenario definitions and projected median percentage change in $BOD_5$ concentration by width class.

| Scenario | Cropland multiplier | Climate increment | W1 | W2 | W3 | Reaches reduced (W1/W2/W3, %) |
|---|---|---|---|---|---|---|
| Baseline (2018 land cover) | 1.00 | recent | 0.00 | 0.00 | 0.00 | — |
| Intensification | × 1.15 | — | −0.49 | +6.03 | +10.50 | — |
| Extensification | × 0.85 | — | −4.00 | +1.10 | +0.61 | 67 / 45 / 38 |
| Climate +1 °C | 1.00 | +1 °C | +0.45 | +0.42 | +0.85 | — |
| Climate +2 °C | 1.00 | +2 °C | +0.29 | +0.41 | +0.57 | — |
| Climate +3 °C | 1.00 | +3 °C | +0.49 | −0.04 | +0.08 | — |
| Intensification + climate | × 1.15 | +2 °C | −0.38 | +5.97 | +11.79 | — |
| Extensification + climate | × 0.85 | +2 °C | −4.64 | +0.67 | +2.52 | — |

Notes: Scenarios were constructed by multiplying the catchment cropland fraction by the specified factor and proportionally rescaling the remaining land-cover fractions so that all fractions summed to unity. The specified temperature increment was applied with a proportional change in precipitation, while the wastewater connection rate was held at recent levels. Percentage changes represent medians relative to the current-driver baseline across the 102 reaches with interpretable catchment predictors. Including the excluded reach altered only the W2 medians by no more than 1.3 percentage points. The final column reports the proportion of reaches showing reduced $BOD_5$ under extensification. Individual responses varied considerably: across all non-baseline scenarios, changes in $BOD_5$ ranged from −41% to +48% and in COD from −29% to +37%, with corresponding 5th–95th percentile ranges of −16% to +19% and −10% to +14%, respectively. The largest $BOD_5$ reductions under extensification occurred in headwater catchments with intermediate cropland cover, where the random forest response was steepest; at high cropland fractions, the fitted response flattened (Figure 5.26i and 5.27d). These results represent within-range scenario sensitivities and should not be interpreted as forecasts.

### 5.3.5. Width-divergent trajectories, management implications and evidential limits

The reconstructed regional trajectory and the reach-level trends indicate a non-monotonic change in oxygen demand. COD and $BOD_5$ declined during the 2000s but subsequently partially rebounded. The width-stratified results further suggest contrasting trajectories: concentrations increased in many narrow, agriculture-dominated headwaters, whereas wider rivers generally continued to improve. This divergence is consistent with differences in dominant pressure regimes. Improvements in wastewater treatment can reduce point-source organic inputs, while diffuse agricultural pressures remain strongly connected to small headwater streams (Bieroza et al., 2024; Jones et al., 2023; Li et al., 2024). However, the present observational analysis does not establish either mechanism causally. Because narrow reaches are both the most numerous and the least extensively monitored, their deterioration may be underrepresented in conventional monitoring networks despite its potential ecological importance.

The 2050 scenarios translate these spatial patterns into conditional management sensitivities. Under the prescribed cropland-extensification scenario, modelled $BOD_5$ decreased in approximately two-thirds of the narrow reaches. Conversely, agricultural intensification increased concentrations in several wider rivers that currently exhibit comparatively favorable conditions. Narrow, cropland-dominated catchments therefore represent priority locations for evaluating agricultural source control, riparian restoration, afforestation, and reduced cultivation intensity (Grizzetti et al., 2021; Gu et al., 2023). These findings do not quantify the effectiveness of individual interventions because the scenarios altered the cropland fraction rather than explicitly representing specific management practices. Similarly, the comparatively small temperature response is conditional on the tested climate range, fixed point-source pressures, and the predictor domain represented in the training data. It should not be interpreted as evidence that future climatic change will be unimportant for river water quality (Li et al., 2024; van Vliet et al., 2023). The broadly concordant width-dependent

patterns of COD and $BOD_5$ reduce, but do not eliminate, the possibility that the observed divergence reflects sampling structure.

Several limitations define the appropriate interpretation of these results. Firstly, the analysis concerns concentration rather than fluxes. Continuous discharge observations were unavailable across the 103 reaches; consequently, the mass export of oxidizable organic matter remains unresolved. Linking the water-quality records with discharge observations is therefore a priority. Such integration would permit estimation of organic-matter loads, analysis of concentration–discharge relationships, and examination of event-scale and seasonal dynamics that are obscured by annual medians.

Secondly, the predictor units represent local fourth-order hydrological areas rather than cumulative upstream drainage areas. They provide a reasonable approximation of the contributing area for headwaters but describe only near-channel context for larger rivers. Delineating complete upstream catchments from a hydrologically corrected digital elevation model would provide a stronger test of whether the observed width dependence persists after accounting for cumulative upstream pressures. Interpretation is also constrained by the unequal monitoring histories among width classes: the fixed long-record panel contains only one W1 reach, compared with nine W2 and five W3 reaches. The width-divergent trajectory should therefore be regarded as a network-level pattern requiring confirmation through additional long-term monitoring of narrow streams.

Thirdly, the model reproduces persistent differences among reaches more reliably than temporal variability within individual reaches. Negative reach-specific efficiencies at most sites show that the reconstruction should be used for spatial prioritization and assessment of broad regional change, rather than as a gap-filled local time series. The muted reconstructed variability may partly reflect static land-cover descriptors, incomplete treatment-plant commissioning histories, and the absence of wastewater-connection statistics before 2000. More temporally resolved wastewater,

agricultural management, and hydrological data would improve the attribution of year-to-year changes.

Additional uncertainty arises from the lower predictive precision for COD, possible delays caused by nutrient and organic-matter legacies in soils and sediments, and the limited capacity of random forests to represent relationships outside the environmental domain used for training (Powers et al., 2016; Van Meter et al., 2018). Moreover, the analysis covers a single region and holds point-source pressures constant in the scenario experiments. Independent validation in other hydroclimatic and agricultural settings, accompanied by formal assessment of the model's area of applicability, is therefore required before broader transfer (Meyer and Pebesma, 2022, 2021). At present, the results provide a defensible basis for identifying priority reaches and designing targeted monitoring, but not for assigning causal credit to individual interventions or forecasting site-specific concentrations in 2050.

## 5.4. Cross-study synthesis: A retrievability spectrum and the double inseparability boundary

In accordance with the structure adopted throughout this dissertation, the results and discussion have been presented together for each study. This section integrates evidence across the three studies. They trace progression from successful direct retrieval through its limits to a catchment-based alternative. Four determinants emerge: target optical activity, water-body size relative to the sensor footprint, validation design, and the hydrological relationship between a reach and its contributing landscape. The first three determine whether water quality can be retrieved from the water surface; the fourth determines whether catchment information provides a credible alternative when direct retrieval fails.

The first determinant is optical activity. In Study 1, DOCc was retrievable because CDOM covaried sufficiently with DOCc and produced a measurable reflectance signal. XGBoost predicted DOCc in Lake Taihu with a validation $R^2$ of 0.73. By contrast, $NH_4^+$-N and $NO_3^-$-N have no inherent optical signal. TP may be associated indirectly

with reflectance through particulate matter, phytoplankton, or turbidity, but the maximum |ρ| values were only 0.17, 0.23, and 0.34 for $NH_4^+$-N, $NO_3^-$-N, and TP, respectively. COD and $BOD_5$ occupy an intermediate position: chromophoric and particulate fractions affect reflectance, whereas dissolved compounds are often weakly expressed optically. Their spectral relationships are therefore indirect and composition dependent. Retrievability consequently followed each target's coupling to optically active material.

The second determinant is the water body's size relative to the sensor footprint. Lake Taihu contains abundant water-dominated pixels, allowing its signal to be separated from that of surrounding land. Most South Bohemian reaches, however, are comparable to or narrower than the available pixels. The median W1 channel occupies approximately 16% of a 30 m pixel and half of a 10 m pixel. Its reflectance is therefore strongly influenced by banks, riparian vegetation, and adjacent land. This geometry helps explain why optical retrieval in both river studies failed at unvisited reaches. The native 10 m Sentinel-2 imagery did not outperform the 30 m Landsat/HLS products in W1, showing that the improvement from 30 to 10 m remained insufficient. Nominal resolution must therefore be interpreted relative to channel width and effective water fraction.

The third determinant is validation design. In both river studies, random CV produced apparently positive skill, whereas LOSO CV revealed little transferable optical information. In Study 2, $\Delta R^2$ between random and spatial CV reached 0.57. In Study 3, the corresponding inflation was 0.15 for the full dataset, increasing to 0.26 for COD and 0.28 for $BOD_5$ in W3. Its increase with width is inconsistent with a purely mixed-pixel explanation, because greater water coverage should strengthen a genuine optical relationship. Instead, the pattern is consistent with models exploiting repeatedly sampled reaches, spatial autocorrelation, and overlapping native and HLS observations. The random-CV–LOSO-CV gap is therefore a diagnostic of validation-induced inflation rather than evidence of retrieval capability. Its recurrence across both target families is a central methodological finding.

These determinants define the boundary of double inseparability. The optical boundary is reached when the target signal cannot be reliably distinguished from land contamination and unrelated optical variability due to weak optical activity, sub-pixel coverage, and adjacency. The attributional boundary concerns a different problem: a spatially coherent regional trend cannot be decomposed reliably into local causes using the available predictors and spatial support. In Study 2, $NH_4^+$-N, $NO_3^-$-N and TP declined by 80%, 48%, and 66%, respectively, before reaching an approximate plateau after 2007. However, none of the nine parameter–buffer combinations produced transferable relationships between reach-level trends and local land-cover change under spatial-block CV. This result does not demonstrate that riparian land cover has no effect. Rather, its contribution could not be separated from regional forcing, given the available data, spatial units, and monitoring design. The boundaries differ in mechanism but share an inferential consequence: beyond either boundary, local conclusions lack transferable support. Both remain conditional on the observational system and scale tested here.

The fourth determinant is the hydrological relationship between a reach and its contributing landscape. Study 3 demonstrates a practical route beyond the optical boundary by shifting observation from the water surface to the landscape that supplies it. The catchment model achieved LOSO NSE values of 0.40 for $BOD_5$ and 0.29 for COD on the $\log_{10}$ scale, compared with 0.03 and 0.02 for the spectral models. Catchment information, therefore, provided substantially greater cross-site skill, although it did not reliably reproduce local temporal variation. $BOD_5$ performance was highest in W1, whereas COD performance peaked in W2; the headwater advantage is therefore indicator specific. Nevertheless, the pattern is hydrologically plausible. Local land-use and point-source pressures are more closely linked to stream chemistry in small headwaters, whereas larger rivers integrate inputs from broader, more heterogeneous upstream areas. The fourth-order units used here also closely approximate the contributing areas for headwaters than for downstream reaches. Thus, the small channel size that constrains optical observation can strengthen prediction

from the surrounding landscape. Catchment modelling does not replace in situ measurement, but it offers a more defensible basis for spatial screening where the water surface cannot be resolved.

Together, the studies define a retrievability spectrum rather than a binary division. At one end, an optically responsive constituent in a well-resolved water body can be retrieved directly, as demonstrated for DOCc in Lake Taihu. On the other hand, optically inactive constituents in sub-pixel channels cannot be retrieved with sufficient spatial transferability at the tested resolutions. Their regional trends can still be characterized from long-term in situ records, but local causes may remain inseparable from regional forcing, as demonstrated for the South Bohemian nutrients. COD and $BOD_5$ lie between these endpoints: neither was transferable from river reflectance, but part of their spatial variation was predictable from catchment attributes, supporting regional reconstruction and scenario sensitivity analysis. Position on this spectrum is a property of the target–system–sensor combination and the intended prediction task. ML complexity can use existing information more efficiently but cannot create a transferable signal that is absent from the observations.

The synthesis establishes a corresponding division of labor for inland-water monitoring. Direct optical retrieval is appropriate where the target is sufficiently coupled to reflectance, and the water body is adequately resolved. Catchment-based prediction is more defensible where targets are optically weak, or channels remain sub-pixel, although in situ measurements remain indispensable for regulatory assessment, flux estimation, and local source attribution. Regional trajectories may still support policy evaluation when a common force dominates, even when changes cannot be attributed to individual reaches or interventions. Validation must match the intended application: prediction at unvisited reaches requires LOSO CV or spatial-block CV, whereas future-year prediction requires temporal hold-out. Reporting random- and spatial-CV performance together makes potential inflation transparent. The retrievability spectrum and double inseparability boundary constitute the principal integrative contribution of this dissertation and provide the conceptual basis for the overall conclusions.

# Chapter VI

## Conclusions

# Content

6.1. Conclusions of Study 1 ..... 189
6.2. Conclusions of Study 2 ..... 190
6.3. Conclusions of Study 3 ..... 192
6.4. Summary ..... 194
6.4.1. Overview, aims, and methods ..... 194
6.4.2. Principal findings of the three studies ..... 195
6.4.3. The retrievability spectrum and the double inseparability boundary ..... 196
6.4.4. The decisive role of validation design ..... 197
6.4.5. Methodological contributions, reproducibility, and transferability ..... 199

# 6. Conclusions

This chapter presents the conclusions of each study and then draws together the overall summary of the dissertation.

## 6.1. Conclusions of Study 1

Study 1 developed and evaluated a framework combining four decades of Landsat observations (1984–2023) with machine learning to reconstruct DOC concentration (DOCc) and estimate DOC storage (DOCs) in Lake Taihu, a large, shallow, optically complex lake. Four models were compared following field-data quality control, satellite–in situ matchup construction, spectral feature engineering, and two-stage feature selection. XGBoost was selected based on its held-out validation performance ($R^2$ = 0.73, RMSE = 0.51 mg/L, MAPE = 8.66%, RPD = 1.70), with lower RMSE and MAPE and a narrower residual range than RF. MODIS-Aqua-referenced harmonization addressed cross-sensor differences, while observed water volumes and an ERA5-driven XGBoost hindcast supported depth-integrated DOCs estimation under the assumption of vertically uniform DOCc.

The reconstructions revealed spatial heterogeneity and distinct temporal phases. Whole-lake DOCc averaged 4.72 ± 0.12 mg/L, increasing through 2000, declining thereafter, and decreasing further after 2013. Northern bays (T1–T3) retained comparatively higher concentrations, while T4–T6 exhibited larger declines. Lake-wide DOCs peaked at 16,906 t C in 2010 and reached 13,802 t C in 2023, close to the 1984 value of 13,788 t C. The fitted four-decade slope was −138.5 t C/decade, with the strongest decline in the southwestern zone (T6) and modest increases in T1 and T5 between 1984 and 2023. The 2010 storage maximum coincided with elevated water volume rather than peak DOCc, highlighting the contribution of hydrological variability to storage dynamics.

SEM, ML importance, and correlation analyses identified meteorological conditions as a prominent correlate of reconstructed DOC variability, with the largest

absolute direct SEM coefficients for DOCc and DOCs (β = −0.72 and −0.98, respectively). Human-activity proxies, bio-optical conditions, and vegetation showed additional, potentially interacting associations. These analyses were complementary rather than independent, and the meteorological relationship with DOCs may partly reflect the ERA5-based water-volume reconstruction. Agreement between RF and XGBoost provided additional support for the spatial patterns, although shared inputs could retain common biases. Overall, Study 1 provides a spatially explicit, multi-decadal baseline for Taihu's dissolved-carbon pool, including a reconstructed DOC anomaly consistent with the 2009 black bloom. The workflow offers a basis for application to other lakes, subject to local calibration and independent validation. Future research should integrate depth-resolved in situ DOC profiles, hyperspectral observations, and hydrodynamic–biogeochemical information to improve source attribution and storage estimation. Comparing alternative water-volume reconstruction models and propagating uncertainties from DOC retrieval, volume estimation, and vertical representativeness would help establish the relative contributions of these factors to DOC uncertainty.

## 6.2. Conclusions of Study 2

Study 2 evaluated satellite-based monitoring of eutrophication-related nutrients in the narrow agricultural rivers of South Bohemia using long-term harmonized multisensor observations and monitoring data. Two principal limitations emerged. Firstly, nutrient concentrations could not be reliably retrieved in previously unseen reaches: spatially blocked CV yielded $R^2$ values of −0.03, 0.05, and 0.11 for $NH_4^+$-N, $NO_3^-$-N and TP, respectively, whereas random CV using identical data and models inflated $R^2$ by up to 0.57. This outcome was consistent with a physical resolution constraint. A typical narrow reach occupied only 16% of a 30-m pixel, leaving the reflectance signal strongly affected by banks, riparian vegetation, and adjacency effects. The conclusion remained unchanged across learners, data streams, seasons, matchup windows, width classes, quality criteria, and reach-level bootstrap resampling. Second, the monitoring record revealed substantial declines in $NH_4^+$-N, (80%), $NO_3^-$-N (48%),

and TP (66%), followed by a plateau after approximately 2007. After adjustment for record span, this recovery showed no clear dependence on river width. Local riparian land-cover change explained no transferable variation in recovery among reaches: none of the nine parameter–scale combinations exceeded the 0.15 skill threshold, and TP residuals retained significant spatial structure.

Together, these limitations define the boundary of double inseparability, comprising optical and attributional components. The optical boundary reflects the combined effects of the sensor footprint, mixed pixels, and the target variables' weak optical sensitivity. The attributional boundary reflects a mismatch between local predictors and the regional scale of the processes driving recovery. Although their mechanisms differ, both identify conditions under which local satellite-derived information is insufficient for transferable inference about narrow-river nutrient conditions or trends. Framing the two limitations jointly converts the negative results into a testable diagnostic. The protocol asks whether local reflectance can distinguish nutrient concentrations among reaches and whether local land-surface predictors can explain differences in their long-term trajectories.

These findings have two practical implications. Firstly, narrow-river retrieval models should be evaluated using spatially blocked CV, with random and spatial performance reported together and mutual information used as a complementary model-agnostic diagnostic. Secondly, nutrient chemistry that cannot be resolved by medium-resolution imagery requires alternative observations, including higher-spatial-resolution or channel-resolving hyperspectral imagery, dense in-situ measurements, and hybrid monitoring frameworks. The protocol developed here can be evaluated in other agricultural headwater networks, although its transferability should not be assumed. Importantly, the long-term in situ archive still resolved the regional recovery, while satellite observations remain valuable for optically active surrogates and sufficiently wide rivers. Study 2 thus does not discount satellite monitoring; it defines the conditions under which satellite-based nutrient inference can and cannot be trusted.

## 6.3. Conclusions of Study 3

Study 3 combined a long-term archive of harmonised multi-sensor imagery with 16,667 COD and 18,885 $BOD_5$ observations across 103 South Bohemian River reaches to determine how oxygen-demand concentrations in narrow rivers should be monitored. The central finding is that monitoring success depends more on the observational target than on spatial resolution. Under LOSO validation, spectral RF models achieved Nash–Sutcliffe efficiencies of only 0.02 for COD and 0.03 for $BOD_5$. Performance remained weak across all width classes and did not improve with 10 m Sentinel-2 imagery, whereas random CV produced apparently higher values of 0.11–0.25. The evidence points primarily to a physical observability constraint: median channel widths of 9.2 m across the network and 5.0 m in W1 leave most narrow reaches sub-pixel, while COD and $BOD_5$ have no unique optical signature. The reconstructed quantities are concentrations rather than loads because continuous discharge observations were unavailable; the products cannot quantify mass export.

Catchment attributes provided a more effective alternative. Under the same LOSO standard, RF driven by land use, terrain, climate, population, and wastewater pressure achieved efficiencies of 0.40 for $BOD_5$ and 0.29 for COD, representing an order-of-magnitude improvement over direct optical retrieval. $BOD_5$ performance was highest in W1 (0.43, compared with 0.35 in W2 and W3), whereas COD peaked in W2 (0.23, 0.35, and 0.26 in W1–W3). Because these differences were comparable to variation among repeated model runs, the width pattern is considered indicative rather than conclusive. Nevertheless, it is hydrologically plausible: small headwater catchments have shorter transport pathways and preserve a clearer relationship between catchment pressures and in-stream concentrations, whereas larger rivers integrate more heterogeneous source areas.

Cropland fraction was the leading predictor, and site-mean $BOD_5$ increased with cropland cover (Spearman's rho = 0.49). Wastewater variables also contributed substantially: removing them reduced efficiency from 0.40 to 0.33 for $BOD_5$ and from

0.29 to 0.19 for COD. Diffuse agricultural pressure and municipal wastewater thus jointly explained much of the transferable spatial signal. Adding catchment water yield, a dilution-capacity index, and local climate anomalies changed model performance by no more than 0.01, while the dilution-capacity index was unrelated to site-mean concentrations. These findings are more consistent with supply-controlled than dilution-controlled differences among reaches, although they do not imply that hydrological processes are unimportant.

The long-term analysis revealed non-monotonic change. Oxygen-demand concentrations declined through the 2000s before partly rebounding, while reach-specific trends diverged by channel width: narrow agricultural headwaters generally deteriorated, whereas wider rivers tended to improve. This contrast persisted after controlling for record span. However, the model reproduced persistent spatial differences more reliably than year-to-year variation. Median per-reach correlations were 0.54 for COD and 0.23 for $BOD_5$, but per-reach efficiency was negative at most sites. The 1985–2025 reconstructions, particularly the driver-only period of 2021–2025, should thus be interpreted as spatial products and descriptions of regional change rather than gap-filled local time series.

The scenario analysis identified distinct sensitivities of the two indicators. For $BOD_5$, cropland extensification reduced concentrations at 67% of W1 reaches, with a median change of −4.0%, whereas warming alone changed width-class medians by less than 1%. COD was more responsive to the climate perturbations, increasing by approximately 2.2% under the +2 °C scenario. Individual responses remained highly heterogeneous, and all scenarios represent within-range model sensitivities rather than deterministic forecasts. Monitoring of narrow agricultural rivers should accordingly combine catchment-based predictors with spatially honest validation, while retaining field observations for temporal assessment. Because the required predictors are available from open pan-European datasets, the framework is potentially transferable to other agricultural headwater networks, although its performance should be re-evaluated under local hydrological and monitoring conditions.

## 6.4. Summary

### 6.4.1. Overview, aims, and methods

This dissertation aimed to reconstruct and attribute long-term dynamics of carbon and water-quality variables in inland waters using machine learning and four decades of satellite observations, while defining when such reconstruction and attribution remain defensible. It was motivated by a mismatch between the importance of inland waters and the capacity to observe them. Conventional in situ monitoring is analytically authoritative but spatially sparse and temporally intermittent, especially where management interventions may have greatest leverage. The three studies span a gradient in systems and analytical outcomes: from a large, optically complex yet spatially well-resolved eutrophic lake, where retrieval was successful and enabled long-term reconstruction and driver attribution, to a network of narrow agricultural rivers, where the limits of optical retrieval were established, and a catchment-attribute alternative was evaluated.

All three studies share the observational foundation described in Chapter IV: harmonized, atmospherically corrected surface reflectance from the medium-resolution satellite archive, processed in a cloud-computing platform, and matched with extensive in situ datasets. Their analytical units and modelling strategies varied with the target properties and system geometry. Study 1 treated Lake Taihu as a spatially resolved water body and compared four ML algorithms using quality-controlled Landsat matchups before harmonizing sensor-specific estimates and reconstructing DOC concentration and storage. Studies 2 and 3 addressed the sub-pixel character of narrow rivers through an object-based, cross-sensor framework, an effective-pixel quality filter, and three river-width classes. Study 2 compared random seven-fold validation with seven-fold spatially blocked CV, whereas Study 3 compared random ten-fold validation with LOSO CV. This common commitment to spatially honest evaluation, despite the different fold definitions required by the two study designs, makes the

positive and negative results directly comparable and supports conservative interpretation.

### 6.4.2. Principal findings of the three studies

Study 1 demonstrated that DOCc and DOCs could be reconstructed over four decades in a large, optically complex eutrophic lake. XGBoost showed the strongest performance on the independent validation set. The reconstructed record revealed pronounced spatial and temporal heterogeneity, including a three-phase DOCc trajectory, persistently higher concentrations in the northern bays, and lower concentrations across the central and south-western zones. Lake-wide DOCs declined slightly over the study period, although the direction and magnitude of change varied spatially. Meteorological conditions emerged as the principal explanatory factors for both DOCc and DOCs, while human-activity proxies, bio-optical conditions, and aquatic vegetation exerted secondary or interacting influences.

Study 2 shifted from successful retrieval to the limits of physical and statistical retrieval. None of the three nutrient variables could be retrieved reliably across sites under spatially explicit CV, whereas random CV substantially inflated apparent skill, particularly in W3. The failure reflected the combined effects of sub-pixel channel geometry and weak target–reflectance coupling. Most narrow reaches occupied only a small fraction of a 30 m pixel, leaving their spectral signals dominated by banks and adjacent land, while $NH_4^+$-N and $NO_3^-$-N lacked direct optical signals, and TP was represented only indirectly through covariance with optically active constituents. Independently, the long-term nutrient decline was predominantly regional and width-uniform after controlling for record span. None of the nine parameter–buffer combinations produced transferable relationships between nutrient trends and riparian land-cover change under spatial-block CV, thereby defining both optical and attributional inseparability boundaries.

Study 3 encountered the same optical boundary for COD and $BOD_5$ but demonstrated a viable catchment-based alternative. Spectral models showed negligible

transferable skill under LOSO CV across all width classes, and native 10 m Sentinel-2 imagery did not outperform the 30 m products in W1. By contrast, the RF model driven by catchment attributes and wastewater indicators achieved substantially higher cross-site skill. $BOD_5$ prediction performed best in W1, where optical retrieval was least credible, whereas COD performance peaked in W2. The 1985–2025 reconstruction indicated an initial decline followed by a partial rebound, with deterioration concentrated in narrow agricultural headwaters while wider rivers generally continued to improve. The 2050 sensitivity analysis further showed that the prescribed land-use changes produced larger modelled responses than the tested warming increments. These results identify catchment management as a potentially important control on future oxygen demand, but remain conditional on the model structure, scenario definitions, and predictor range.

### 6.4.3. The retrievability spectrum and the double inseparability boundary

Taken together, the three studies show that the retrievability of an inland-water-quality variable is not a fixed property of either the target or the sensor alone. It emerges from the interaction of three factors: the optical activity of the target, the spatial scale of the water body relative to the sensor footprint, and the validation design used to assess predictive skill. Section 5.4 develops this retrievability spectrum in detail. Its defining feature is that these determinants are jointly necessary rather than fully compensatory. An optically active constituent may remain unretrievable when the channel occupies only a small fraction of a pixel, and the signal is dominated by adjacent land, while finer spatial resolution cannot make an optically inactive constituent directly observable. Study 1 occupies the favorable end of the spectrum, Study 2 the unfavorable end, and Study 3 an intermediate position in which oxygen-demand concentrations could not be retrieved reliably from reflectance but were partly predictable from catchment attributes.

Two implications follow. Firstly, increasing sensor resolution offers less leverage than is often assumed. A channel only a few meters wide remain sub-pixel even at 10 m resolution, while greater spatial detail cannot create a spectral signal that the target

does not possess. Secondly, within the South Bohemian network, optical retrievability was weakest in narrow agricultural headwaters, where diffuse pressures and opportunities for management intervention were greatest. The spectrum thus shifts the question from whether a variable can be mapped from space to the conditions under which a specific combination of target, water body, and sensor is defensible. It also identifies whether the water surface or the contributing catchment provides the more informative observational domain.

The river studies express two related limits, such as the boundary of double inseparability. The optical boundary is reached when the in-water signal cannot be separated reliably from mixed-pixel and land-adjacency effects. The attributional boundary is reached when a spatially coherent regional trend cannot be decomposed into local, reach-scale drivers using the available predictors. Study 3 provides a constructive response through the catchment-observation paradigm. Its success where optical retrieval failed is hydrologically coherent: the small catchment area relative to the receiving channel shortens transport pathways and strengthens the connection between land-based pressures and in-stream concentrations. The tendency for catchment models to perform better in narrow headwaters, while spectral models remain uniformly weak, is thus mechanistically plausible rather than contradictory. However, because the between-width differences in catchment-model performance were comparable to variation among repeated runs, they should be interpreted as supporting evidence rather than as a definitive width ranking.

### 6.4.4. The decisive role of validation design

The three studies demonstrate that model performance cannot be interpreted independently of the validation design used to estimate it. Random CV allows observations from the same reach to appear in both the training and test folds. In a repeatedly sampled and spatially autocorrelated network, a model may therefore reproduce persistent differences among known reaches without learning a relationship that transfers to an unvisited location. The overlap between native and HLS products introduces an additional risk because different representations of the same acquisition

may be divided between folds. Random CV is not intrinsically invalid, but it evaluates interpolation over observations drawn from an already-sampled network. It cannot, by itself, support claims of prediction at previously unmonitored reaches. LOSO CV provides a more appropriate test of cross-site transfer by withholding all observations from each test reach, although dependence among neighboring reaches may remain. Spatial-block CV imposes a stricter regional test by separating contiguous groups of sites, while temporal hold-out is required when the intended task is prediction into future years.

The contrast among these validation schemes was decisive in both river studies. In Study 2, random CV suggested a potentially useful nutrient-retrieval skill, whereas spatially explicit CV reduced $R^2$ to approximately zero, with inflation reaching $\Delta R^2 = 0.57$. Study 3 showed the same pattern for COD and $BOD_5$, and the random–LOSO difference increased to 0.26 and 0.28, respectively, in W3. The larger inflation in the widest and nominally most favorable channels is inconsistent with the expectation that increasing water coverage should improve genuinely transferable optical retrieval. Instead, it indicates that random-CV performance was strongly influenced by repeated site identity, spatial dependence, and sampling structure. Because W3 contained only 13 reaches, the precise magnitude of this inflation should be interpreted with caution; nevertheless, its recurrence across two independent studies and across different target variables demonstrates that the result was not confined to a single model or water-quality indicator.

Validation design also determined how the negative findings could be interpreted. The failure of the spectral models under LOSO and spatial-block CV does not establish that the target variables are universally unretrievable. It defines the limit of transferable retrieval for the variables, sensors, spatial resolutions, and channel geometries examined in this dissertation. Similarly, the failure of riparian land-cover models to predict nutrient trends under spatial-block CV does not show that land cover has no ecological effect. It demonstrates that the available land-cover changes could not separate local influences from the regionally coherent recovery. Importantly, the

catchment model in Study 3 achieved positive cross-site performance under the same LOSO framework that rejected the spectral models. Spatial validation therefore did not force all models to fail; it distinguished a predictor set containing transferable information from one dominated by local or site-specific structure.

The principal methodological conclusion is that validation must match the intended application. Claims concerning unvisited reaches require LOSO or spatial-block CV, whereas reconstruction beyond the monitoring period requires temporal hold-out. Random CV remains useful as a comparison because the difference between random and spatially explicit performance reveals the extent to which apparent skill depends on non-independent observations. Reporting these results together provides a transparent distinction between interpolation within a known monitoring network and prediction beyond it. Treating validation design as part of the scientific hypothesis, rather than as a routine modelling step, is therefore a central contribution of this dissertation.

### 6.4.5. Methodological contributions, reproducibility, and transferability

Beyond their substantive findings, the three studies provide several methodological contributions. The first is the use of spatial validation tailored to each monitoring design. Study 2 employed spatial-block CV, whereas Study 3 used LOSO CV. Each was compared with an otherwise identical random CV scheme. The difference between the two schemes quantified the apparent skill arising from spatial leakage. This approach elevated CV design from a routine modelling choice to a scientific diagnostic.

The second contribution is an object-based, cross-sensor framework with effective-pixel quality filtering. Reflectance was aggregated over fixed, width-constrained reach objects. Observations strongly affected by mixed pixels and land adjacency were then excluded. The framework integrated seven reflectance streams from Landsat, Sentinel-2, and HLS products into a consistent multi-decadal archive. The third contribution is the catchment-attribute strategy. It translated satellite-

observable catchment characteristics into models for spatial prediction, long-term reconstruction, and scenario-sensitivity analysis. Additional advances included harmonising sensor-specific DOC estimates against a temporally continuous external reference. ML was also used to hindcast WSC from atmospheric reanalysis. Monitoring-network turnover and record span were explicitly controlled during trend estimation.

The studies used openly accessible or fully documented data sources. Their procedures for data preparation, QA/QC, feature construction, model validation, and uncertainty analysis were specified as reproducible protocols. These features allow individual components to be inspected and reimplemented in other systems. Transferability nevertheless remains conditional. The Lake Taihu workflow is most applicable to large, spatially resolved water bodies with a detectable optical proxy and sufficient in situ calibration data. Estimating DOC storage also requires reliable hydrological information. The workflow is less suitable for narrow channels or optically inactive constituents, where changing the ML algorithm alone is unlikely to overcome observational constraints.

The negative river results provide a testable expectation for systems with channels narrower than the sensor footprint and targets lacking a direct optical signal. However, this expectation still requires local evaluation. The catchment-observation paradigm is applicable where consistent descriptors of land use, terrain, climate, population, and wastewater pressure are available. Representative in situ observations are also needed for spatial calibration. Model performance must be reassessed when catchment processes, predictor ranges, or management regimes differ. The most widely transferable contribution is the validation protocol. Paired random and spatial CV can be applied across regions, variables, and sensors. It quantifies skill inflation due to spatial dependence and tests whether a model can generalise to unmonitored locations.

# Chapter VII

## Limitations and Future Perspectives

# Content

7.1. Limitations of the studies ..... 203
7.2. Future perspectives ..... 206

# 7. Limitations and Future Perspectives

This chapter synthesizes the principal limitations of the three studies and outlines the future research directions arising from them. Consistent with the integrative structure of the dissertation, the limitations are organized thematically rather than by individual study. Each limitation is subsequently linked to the methodological or observational development required to address it.

## 7.1. Limitations of the studies

The first limitation concerns the optical, vertical, and temporal representativeness of the lake reconstruction. Satellite reflectance primarily characterizes the optically accessible surface layer, so the reconstructed DOCc fields represent near-surface conditions rather than depth-resolved profiles. DOCs was estimated by combining surface DOCc with WSC under an assumption of vertical homogeneity. Although frequent wind-driven mixing in shallow Lake Taihu probably limits persistent vertical gradients, episodic stratification, bloom accumulation, and sediment resuspension may temporarily decouple the surface from deeper water. DOCs should therefore be interpreted as a first-order, observation-constrained estimate rather than an exact inventory. The DOCc model was calibrated and validated using observations from 2007–2015 but applied to the 1984–2023 archive. Cross-sensor harmonization against MODIS-Aqua and residual diagnostics support consistency among sensors, but they cannot demonstrate that the DOCc–CDOM–reflectance relationship remained stationary outside the calibration period. Consequently, uncertainty is greater during rapid optical transitions and in the earliest Landsat years.

The second limitation concerns the physical observability of the river variables and the conditional performance of the catchment alternative. Under spatially explicit CV, the near-zero optical skill for the nutrient variables, COD and $BOD_5$, reflected the combined effects of sub-pixel channel coverage, land adjacency, and weak target–reflectance coupling. These results define an empirical boundary for the sensors,

resolutions, and channel geometries tested here; they do not establish universal non-retrievability under all observational systems. The Study 3 catchment model provided greater cross-site skill, but its accuracy remained moderate and depended on the quality and spatial support of the land-cover, climate, population, and wastewater predictors. The 2050 experiments are conditional sensitivity analyses rather than forecasts and do not represent future policy implementation, changes in point-source pressure, or responses beyond the model's training domain.

The third limitation concerns attribution and the spatial support of the ancillary data. In Study 1, GDP, NTL, and emission indicators served as proxies for human pressure rather than as direct measurements of discharge, carbon loading, or management. Their associations with DOCc and DOCs therefore provide contextual, not causal, evidence. Likewise, the failure of Study 2 to predict nutrient trends from riparian land-cover change does not demonstrate that local land cover has no effect on river chemistry. It shows that local effects could not be separated from the regional trajectory using the available data and spatial design. CORINE has a resolution of 100 m, compared with a median channel width of 9.2 m, and therefore characterizes the broader riparian setting rather than fine-scale changes immediately adjacent to narrow channels. Consistency across buffer scales indicates that buffer choice alone did not determine the outcome, although resolution and temporal-snapshot uncertainty remain. In Study 3, predictors were extracted from local fourth-order hydrological units rather than cumulative upstream drainage areas. These units contribute more closely to headwaters but represent only near-channel context for wider downstream reaches.

The fourth limitation concerns monitoring coverage and the quantities resolved by the river analyses. Record duration and completeness varied substantially among reaches, with long records least common in W1. The fixed panel used to characterize regional nutrient trajectories contained only 15 reaches, including one W1 reach, and therefore represented a restricted portion of the network. Moreover, nutrient observations began in 1990, after the 1989 socioeconomic transition; the analysis characterizes the post-transition trajectory but cannot provide a direct before-and-after

comparison. Study 3 modelled concentrations rather than fluxes because continuous discharge observations were not co-located with the water-quality records. The reconstructed COD and $BOD_5$ values therefore do not represent hydrological loads, and the mass export of oxidizable matter remains unquantified. The catchment model also resolved persistent spatial differences more effectively than within-reach temporal variability. Because reach-specific NSE values were negative at most sites, the reconstructed fields should be interpreted as spatial annual estimates and indicators of regional, multi-decadal change rather than as gap-filled local time series. Their muted temporal amplitudes may partly reflect static or temporally coarse predictors and incomplete historical wastewater information.

The final limitation concerns model uncertainty and external transferability. Alternative learners and configurations were not benchmarked across all analytical components; notably, the WSC hindcast relied on a single selected ML approach. More broadly, Study 1 examined one large, shallow eutrophic lake, whereas Studies 2 and 3 examined a single Central European River network dominated by narrow agricultural channels. The retrievability spectrum and double inseparability boundary are therefore empirical conclusions conditioned on the variables, sensors, environmental settings, and validation designs used here. Their conceptual basis may be transferable, but their quantitative thresholds and predictive performance should not be generalized to other lakes, catchments, or hydroclimatic regions without independent validation.

## 7.2. Future perspectives

The limitations identified in Section 7.1 define four priorities for future research. The first is to move from surface-based reconstruction towards depth-resolved observation of lake carbon. Future studies should combine depth-specific DOC sampling and excitation–emission matrix fluorescence analysis with vertical profiles of temperature, conductivity, and bio-optical properties. Moored sensors could resolve short-lived changes that routine sampling misses, particularly during stratification, bloom development, and sediment resuspension. Assimilating these observations into three-dimensional hydrodynamic–biogeochemical models would allow the assumption of vertical DOC homogeneity to be tested explicitly and would improve depth-integrated DOCs estimates. Repeated profiling across seasons and hydrological conditions would also clarify when surface DOCc reliably represents the full water column and when event-specific corrections are required.

The second priority is to exploit richer spectral and spatial observations while retaining physically realistic expectations of what they can resolve. Spaceborne hyperspectral data provide denser spectral sampling than the multispectral archive and may improve the estimation of CDOM absorption and spectral slope, as well as the separation of CDOM from phytoplankton and suspended sediment. DOCc would nevertheless remain an indirect retrieval, as the CDOM–DOC relationship varies across sources and seasons. Cross-calibration and data-fusion approaches could connect shorter hyperspectral records with the long-term Landsat archive without treating the sensors as interchangeable. For narrow rivers, very high-resolution satellite or airborne imagery, combined with water-specific atmospheric and adjacency corrections, should be evaluated relative to channel width and the effective water fraction. These observations may shift the optical boundary towards smaller channels, but they cannot guarantee retrieval where the target lacks a stable optical expression.

The third priority is to develop the catchment-observation approach into a more hydrologically complete framework. Its applicability should be tested for other partly

optical or optically inactive variables, including dissolved nutrients, using spatially explicit CV at unvisited reaches. Three extensions are particularly important. First, hydrologically representative discharge records should be coupled with water-quality observations to estimate fluxes, examine concentration–discharge relationships, and resolve seasonal and event-scale transport. Load estimation should explicitly account for sampling frequency and uncertainty rather than treating intermittent concentrations as continuous observations. Second, local fourth-order units should be replaced or supplemented by cumulative upstream catchments derived from a hydrologically corrected DEM and river network topology. This would allow upstream land use, point sources, drainage areas, and potential transport pathways to be represented consistently, particularly for wider rivers. Third, wastewater predictors should incorporate treatment-plant commissioning dates, changing connection rates, and temporally resolved discharge information. Agricultural-management histories, including fertilizer application and crop rotation, would provide complementary information on diffuse pressures. Combining these drivers with process-based catchment and in-stream water-quality models could improve mechanistic interpretation, while ML components could retain the spatial coverage and flexibility demonstrated in Study 3.

The fourth priority is methodological consolidation and external validation. The OCSF and the comparison between random and spatially explicit CV should be applied to additional regions, water-quality variables, and sensor configurations. Validation must remain aligned with the intended application: LOSO or spatial-block CV for transfer to unvisited reaches, temporal hold-out for reconstruction beyond the monitoring period, and independent regional validation for geographic transfer. Shared benchmark datasets containing channel geometry, satellite observations, and standardized in situ measurements would allow alternative models and preprocessing strategies to be compared under identical conditions. Future reconstructions should also propagate uncertainty from field measurements, sensor harmonization, predictor data, and model selection into the final spatial and temporal products.

Within this priority, transfer to new regions should be tested rather than assumed. Formal area-of-applicability diagnostics should determine whether new predictor combinations are represented within the training domain, and predictions outside that domain should be identified or withheld. Multi-region studies spanning different lake types, catchment sizes, optical conditions, and hydroclimatic settings would establish which components of the retrievability spectrum are general and which remain system-specific. In this way, the workflows developed here could contribute to long-term monitoring of inland-water carbon and water quality without extending model claims beyond the scales and environments supported by observation.

# References

Abbott, B.W., Gruau, G., Zarnetske, J.P., Moatar, F., Barbe, L., Thomas, Z., Fovet, O., Kolbe, T., Gu, S., Pierson-Wickmann, A.-C., Davy, P., Pinay, G., 2018. Unexpected spatial stability of water chemistry in headwater stream networks. Ecology Letters 21, 296–308. https://doi.org/10.1111/ele.12897

Aggarwal, V., Gupta, V., Singh, P., Sharma, K., Sharma, N., 2019. Detection of Spatial Outlier by Using Improved Z-Score Test, in: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). Presented at the 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 788–790. https://doi.org/10.1109/ICOEI.2019.8862582

Al-Kharusi, E.S., Tenenbaum, D.E., Abdi, A.M., Kutser, T., Karlsson, J., Bergström, A.-K., Berggren, M., 2020. Large-Scale Retrieval of Coloured Dissolved Organic Matter in Northern Lakes Using Sentinel-2 Data. Remote Sensing 12, 157. https://doi.org/10.3390/rs12010157

Allan, J.D., 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annual Review of Ecology, Evolution, and Systematics 35, 257–284. https://doi.org/10.1146/annurev.ecolsys.35.120202.110122

Allen, G.H., Pavelsky, T.M., 2018. Global extent of rivers and streams. Science 361, 585–588. https://doi.org/10.1126/science.aat0636

APHA, 2017. Standard methods for the examination of water and wastewater, 23rd Edition. American Public Health Association, AWWA, WPCF.

Badapalli, P.K., Nakkala, A.B., Gugulothu, S., Kottala, R.B., 2024. Dynamic Land Degradation Assessment: Integrating Machine Learning with Landsat 8 OLI/TIRS for Enhanced Spectral, Terrain, and Land Cover Indices. Earth Syst Environ. https://doi.org/10.1007/s41748-024-00442-9

Baines, S.B., Pace, M.L., 1991. The production of dissolved organic matter by phytoplankton and its importance to bacteria: Patterns across marine and freshwater systems. Limnology and Oceanography 36, 1078–1090. https://doi.org/10.4319/lo.1991.36.6.1078

Basu, N.B., Destouni, G., Jawitz, J.W., Thompson, S.E., Loukinova, N.V., Darracq, A., Zanardo, S., Yaeger, M., Sivapalan, M., Rinaldo, A., 2010. Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. Geophysical Research Letters 37.

Batina, A., Krtalić, A., 2023. A Review of Remote Sensing Applications for Determining Lake Water Quality. https://doi.org/10.20944/preprints202309.0489.v1

Bieroza, M., Hallberg, L., Livsey, J., Prischl, L.-A., Wynants, M., 2024. Recognizing agricultural headwaters as critical ecosystems. Environmental Science & Technology 58, 4852.

Bishop, K., Buffam, I., Erlandsson, M., Folster, J., Laudon, H., Seibert, J., Temnerud, J., 2008. Aqua Incognita: the unknown headwaters. Hydrological Processes 22, 1239.

Blaschke, T., 2010a. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Blaschke, T., 2010b. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic Object-Based Image Analysis – Towards a new paradigm. ISPRS Journal of Photogrammetry and Remote Sensing 87, 180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Bogard, M.J., Vachon, D., St.-Gelais, N.F., del Giorgio, P.A., 2017. Using oxygen stable isotopes to quantify ecosystem metabolism in northern lakes. Biogeochemistry 133, 347–364. https://doi.org/10.1007/s10533-017-0338-5

Brando, V.E., Dekker, A.G., 2003. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. IEEE Transactions on Geoscience and Remote Sensing 41, 1378–1387. https://doi.org/10.1109/TGRS.2003.812907

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brezonik, P.L., Olmanson, L.G., Finlay, J.C., Bauer, M.E., 2015. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. Remote Sensing of Environment, Special Issue: Remote Sensing of Inland Waters 157, 199–215. https://doi.org/10.1016/j.rse.2014.04.033

Bulgarelli, B., Zibordi, G., 2018. On the detectability of adjacency effects in ocean color remote sensing of mid-latitude coastal environments by SeaWiFS, MODIS-A, MERIS, OLCI, OLI and MSI. Remote sensing of environment 209, 423–438.

Cai, J., Meng, L., Liu, H., Chen, J., Xing, Q., 2022. Estimating Chemical Oxygen Demand in estuarine urban rivers using unmanned aerial vehicle hyperspectral images. Ecological Indicators 139, 108936. https://doi.org/10.1016/j.ecolind.2022.108936

Cai, X., Li, Y., Lei, S., Zeng, S., Zhao, Z., Lyu, H., Dong, X., Li, J., Wang, H., Xu, J., Zhu, Y., Wu, L., Cheng, X., 2023. A hybrid remote sensing approach for estimating chemical oxygen demand concentration in optically complex waters: A case study in inland lake waters in eastern China. Science of The Total Environment 856, 158869. https://doi.org/10.1016/j.scitotenv.2022.158869

Cao, F., Tzortziou, M., Hu, C., Mannino, A., Fichot, C.G., Del Vecchio, R., Najjar, R.G., Novak, M., 2018. Remote sensing retrievals of colored dissolved organic matter and dissolved organic carbon dynamics in North American estuaries and their margins. Remote Sensing of Environment 205, 151–165. https://doi.org/10.1016/j.rse.2017.11.014

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. Remote Sensing of Environment 248, 111974. https://doi.org/10.1016/j.rse.2020.111974

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., Smith, V.H., 1998. Nonpoint Pollution of Surface Waters with Phosphorus and Nitrogen. Ecological Applications 8, 559–568. https://doi.org/10.1890/1051-0761(1998)008%5B0559:NPOSWW%5D2.0.CO;2

Catalán, N., Marcé, R., Kothawala, D.N., Tranvik, L.J., 2016a. Organic carbon decomposition rates controlled by water retention time across inland waters. Nature Geoscience 9, 501–504. https://doi.org/10.1038/ngeo2720

Catalán, N., Marcé, R., Kothawala, D.N., Tranvik, L.J., 2016b. Organic carbon decomposition rates controlled by water retention time across inland waters. Nature Geosci 9, 501–504. https://doi.org/10.1038/ngeo2720

Chen, J., Chen, S., Fu, R., Li, D., Jiang, H., Wang, C., Peng, Y., Jia, K., Hicks, B.J., 2022. Remote Sensing Big Data for Water Environment Monitoring: Current Status, Challenges, and Future Prospects. Earth's Future 10, e2021EF002289. https://doi.org/10.1029/2021EF002289

Chen, J., Gao, M., Cheng, S., Hou, W., Song, M., Liu, X., Liu, Y., Shan, Y., 2020. County-level CO2 emissions and sequestration in China during 1997–2017. Sci Data 7, 391. https://doi.org/10.1038/s41597-020-00736-3

Chi, J., Nilsson, M.B., Laudon, H., Lindroth, A., Wallerman, J., Fransson, J.E.S., Kljun, N., Lundmark, T., Ottosson Löfvenius, M., Peichl, M., 2020. The Net Landscape Carbon Balance—Integrating terrestrial and aquatic carbon fluxes in a managed boreal forest landscape in Sweden. Global Change Biology 26, 2353–2367. https://doi.org/10.1111/gcb.14983

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002

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, 171–184. https://doi.org/10.1007/s10021-006-9013-8

Conley, D.J., Paerl, H.W., Howarth, R.W., Boesch, D.F., Seitzinger, S.P., Havens, K.E., Lancelot, C., Likens, G.E., 2009. Controlling Eutrophication: Nitrogen and Phosphorus. Science 323, 1014–1015. https://doi.org/10.1126/science.1167755

De Stefano, L.G., Valdivia, A.S., Gianello, D., Gerea, M., Reissig, M., García, P.E., García, R.D., Cárdenas, C.S., Diéguez, M.C., Queimaliños, C.P., Pérez, G.L., 2022. Using CDOM spectral shape information to improve the estimation of DOC concentration in inland waters: A case study of Andean Patagonian Lakes. Science of The Total Environment 824, 153752. https://doi.org/10.1016/j.scitotenv.2022.153752

DelSontro, T., Beaulieu, J.J., Downing, J.A., 2018. Greenhouse gas emissions from lakes and impoundments: Upscaling in the face of global change. Limnology and Oceanography Letters 3, 64–75. https://doi.org/10.1002/lol2.10073

Deng, C., Zhang, L., Cen, Y., 2019. Retrieval of Chemical Oxygen Demand through Modified Capsule Network Based on Hyperspectral Data. Applied Sciences 9, 4620. https://doi.org/10.3390/app9214620

Dethier, E.N., Renshaw, C.E., Magilligan, F.J., 2022. Rapid changes to global river suspended sediment flux by humans. Science 376, 1447–1452. https://doi.org/10.1126/science.abn7980

Ding, Y., Sun, L., Qin, B., Wu, T., Shen, X., Wang, Y., 2018. Characteristics of sediment resuspension in Lake Taihu, China: A wave flume study. Journal of Hydrology 561, 702–710. https://doi.org/10.1016/j.jhydrol.2018.04.026

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Schloesser, J.T., Thornbrugh, D.J., 2008. Eutrophication of U.S. Freshwaters: Analysis of Potential Economic Damages. Environ. Sci. Technol. 43, 12–19. https://doi.org/10.1021/es801217q

Dong, L., Gong, C., Huai, H., Wu, E., Lu, Z., Hu, Y., Li, L., Yang, Z., 2023. Retrieval of Water Quality Parameters in Dianshan Lake Based on Sentinel-2 MSI Imagery and Machine Learning: Algorithm Evaluation and Spatiotemporal Change Research. Remote Sensing 15, 5001. https://doi.org/10.3390/rs15205001

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R., 2007. Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: A Review. Ecography 30, 609–628.

Downing, J.A., 2010. Emerging global role of small lakes and ponds: little things mean a lot. Limnetica 29, 9–24. https://doi.org/10.23818/limn.29.02

Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, J.M., 2006. The global abundance and size distribution of lakes, ponds, and impoundments. Limnology and oceanography 51, 2388–2397.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment, The Sentinel Missions - New Opportunities for Science 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Duan, H., Ma, R., Loiselle, S.A., Shen, Q., Yin, H., Zhang, Y., 2014. Optical characterization of black water blooms in eutrophic waters. Sci. Total Environ. 482–483, 174–183. https://doi.org/10.1016/j.scitotenv.2014.02.113

Duan, H., Ma, R., Xu, X., Kong, F., Zhang, S., Kong, W., Hao, J., Shang, L., 2009. Two-Decade Reconstruction of Algal Blooms in China's Lake Taihu. Environ. Sci. Technol. 43, 3522–3528. https://doi.org/10.1021/es8031852

European Parliament and Council, 2024. Directive (EU) 2024/3019 of the European Parliament and of the Council of 27 November 2024 concerning urban wastewater treatment (recast) (Text with EEA relevance).

European Parliament and Council, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, OJ L.

Evans, C.D., Futter, M.N., Moldan, F., Valinia, S., Frogbrook, Z., Kothawala, D.N., 2017. Variability in organic carbon reactivity across lake residence time and trophic gradients. Nature Geosci 10, 832–835. https://doi.org/10.1038/ngeo3051

Feng, Y., Xiong, Y., Hall-Spencer, J.M., Liu, K., Beardall, J., Gao, K., Ge, J., Xu, J., Gao, G., 2024. Shift in algal blooms from micro- to macroalgae around China with increasing eutrophication and climate change. Global Change Biology 30, e17018. https://doi.org/10.1111/gcb.17018

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment 140, 23–35. https://doi.org/10.1016/j.rse.2013.08.029

Fichot, C.G., Benner, R., 2011. A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. Geophysical Research Letters 38. https://doi.org/10.1029/2010GL046152

Fisher, A., Flood, N., Danaher, T., 2016. Comparing Landsat water index methods for automated water classification in eastern Australia. Remote Sensing of Environment 175, 167–182. https://doi.org/10.1016/j.rse.2015.12.055

Gao, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gao, Y., Yu, G., 2020. Regional coupled CN-H2O cycle processes and associated driving mechanisms. Science China Earth Sciences 63, 1227–1236.

Gardner, J.R., Yang, X., Topp, S.N., Ross, M.R.V., Altenau, E.H., Pavelsky, T.M., 2021. The Color of Rivers. Geophysical Research Letters 48, e2020GL088946. https://doi.org/10.1029/2020GL088946

Ge, Y., Shen, F., Sklenička, P., Vymazal, J., Baxa, M., Chen, Z., 2024. Machine learning for cyanobacteria inversion via remote sensing and AlgaeTorch in the Třeboň fishponds, Czech Republic. Science of The Total Environment 947, 174504. https://doi.org/10.1016/j.scitotenv.2024.174504

Godsey, S.E., Kirchner, J.W., Clow, D.W., 2009. Concentration–discharge relationships reflect chemostatic characteristics of US catchments. Hydrological Processes: An International Journal 23, 1844–1864.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, Big Remotely Sensed Data: tools, applications and experiences 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Grasshoff, K., Kremling, K., Ehrhardt, M., 2009. Methods of Seawater Analysis. John Wiley & Sons.

Griffin, C.G., Finlay, J.C., Brezonik, P.L., Olmanson, L., Hozalski, R.M., 2018. Limitations on using CDOM as a proxy for DOC in temperate lakes. Water Research 144, 719–727. https://doi.org/10.1016/j.watres.2018.08.007

Grizzetti, B., Vigiak, O., Udias, A., Aloe, A., Zanni, M., Bouraoui, F., Pistocchi, A., Dorati, C., Friedland, R., De Roo, A., 2021. How EU policies could reduce nutrient pollution in European inland and coastal waters. Global Environmental Change 69, 102281.

Gu, B., Zhang, X., Lam, S.K., Yu, Y., van Grinsven, H.J.M., Zhang, S., Wang, X., Bodirsky, B.L., Wang, S., Duan, J., Ren, C., Bouwman, L., de Vries, W., Xu, J., Sutton, M.A., Chen,

D., 2023. Cost-effective mitigation of nitrogen pollution from global croplands. Nature 613, 77–84. https://doi.org/10.1038/s41586-022-05481-8

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology 377, 80–91.

Harkort, L., Duan, Z., 2023. Estimation of dissolved organic carbon from inland waters at a large scale using satellite data and machine learning methods. Water Research 229, 119478. https://doi.org/10.1016/j.watres.2022.119478

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J., Mopper, K., 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnology and Oceanography 53, 955–969. https://doi.org/10.4319/lo.2008.53.3.0955

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S., Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., de Bund, W. van, 2010. The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. Science of The Total Environment 408, 4007–4019. https://doi.org/10.1016/j.scitotenv.2010.05.031

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. Q.J.R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803

Hestir, E.L., Brando, V., Campbell, G., Dekker, A., Malthus, T., 2015. The relationship between dissolved organic matter absorption and dissolved organic carbon in reservoirs along a temperate to tropical gradient. Remote Sensing of Environment 156, 395–402. https://doi.org/10.1016/j.rse.2014.09.022

Hoegh-Guldberg, O., Northrop, E., Lubchenco, J., 2019. The ocean is key to achieving climate and societal goals. Science 365, 1372–1374. https://doi.org/10.1126/science.aaz4390

Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., Zheng, C., 2022. Global mapping reveals increase in lacustrine algal blooms over the past decade. Nat. Geosci. 15, 130–134. https://doi.org/10.1038/s41561-021-00887-x

Hu, B., Wang, P., Zhang, N., Wang, C., Ao, Y., 2016. Photoproduction of dissolved organic carbon and inorganic nutrients from resuspended lake sediments. Environ Sci Pollut Res 23, 22126–22135. https://doi.org/10.1007/s11356-016-7327-4

Hu, C., 2009. A novel ocean color index to detect floating algae in the global oceans. Remote Sensing of Environment 113, 2118–2129. https://doi.org/10.1016/j.rse.2009.05.012

Huang, C., Yunmei, L., Liu, G., Guo, Y., Yang, H., Zhu, A.-X., Song, T., Huang, T., Zhang, M., Shi, K., 2017. Tracing high time-resolution fluctuations in dissolved organic carbon

using satellite and buoy observations: Case study in Lake Taihu, China. Int. J. Appl. Earth Obs. Geoinformation 62, 174–182. https://doi.org/10.1016/j.jag.2017.06.009

Jones, E.R., Bierkens, M.F.P., van Puijenbroek, P.J.T.M., van Beek, L. (Rens) P.H., Wanders, N., Sutanudjaja, E.H., van Vliet, M.T.H., 2023. Sub-Saharan Africa will increasingly become the dominant hotspot of surface water pollution. Nat Water 1, 602–613. https://doi.org/10.1038/s44221-023-00105-5

Karasiak, N., Dejoux, J.-F., Monteil, C., Sheeren, D., 2022. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. Mach Learn 111, 2715–2740. https://doi.org/10.1007/s10994-021-05972-1

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

Kellerman, A.M., Dittmar, T., Kothawala, D.N., Tranvik, L.J., 2014. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. Nat Commun 5, 3804. https://doi.org/10.1038/ncomms4804

Kellerman, A.M., Kothawala, D.N., Dittmar, T., Tranvik, L.J., 2015. Persistence of dissolved organic matter in lakes related to its molecular characteristics. Nature Geosci 8, 454–457. https://doi.org/10.1038/ngeo2440

Kirk, J.T.O., 1994. Light and Photosynthesis in Aquatic Ecosystems. Cambridge University Press.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of hydrology 424, 264–277.

Knoben, W.J., Freer, J.E., Woods, R.A., 2019. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences 23, 4323–4331.

Kothawala, D.N., Stedmon, C.A., Müller, R.A., Weyhenmeyer, G.A., Köhler, S.J., Tranvik, L.J., 2014. Controls of dissolved organic matter quality: evidence from a large-scale boreal lake survey. Global Change Biology 20, 1101–1114. https://doi.org/10.1111/gcb.12488

Kuhn, C., de Matos Valerio, A., Ward, N., Loken, L., Sawakuchi, H.O., Kampel, M., Richey, J., Stadler, P., Crawford, J., Striegl, R., Vermote, E., Pahlevan, N., Butman, D., 2019. Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-*a* and turbidity. Remote Sensing of Environment 224, 104–118. https://doi.org/10.1016/j.rse.2019.01.023

Kutser, T., 2012. The possibility of using the Landsat image archive for monitoring long time trends in coloured dissolved organic matter concentration in lake waters. Remote Sensing of Environment 123, 334–338. https://doi.org/10.1016/j.rse.2012.04.004

Kutser, T., Pierson, D.C., Kallio, K.Y., Reinart, A., Sobek, S., 2005. Mapping lake CDOM by satellite remote sensing. Remote Sensing of Environment 94, 535–540. https://doi.org/10.1016/j.rse.2004.11.009

Kutser, T., Verpoorter, C., Paavel, B., Tranvik, L.J., 2015. Estimating lake carbon fractions from remote sensing data. Remote Sensing of Environment, Special Issue: Remote Sensing of Inland Waters 157, 138–146. https://doi.org/10.1016/j.rse.2014.05.020

Lai, L., Zhang, Y., Han, T., Zhang, M., Cao, Z., Liu, Z., Yang, Q., Chen, X., 2024. Satellite mapping reveals phytoplankton biomass's spatio-temporal dynamics and responses to environmental factors in a eutrophic inland lake. Journal of Environmental Management 360, 121134. https://doi.org/10.1016/j.jenvman.2024.121134

Lavonen, E.E., Kothawala, D.N., Tranvik, L.J., Gonsior, M., Schmitt-Kopplin, P., Köhler, S.J., 2015. Tracking changes in the optical properties and molecular composition of dissolved organic matter during drinking water production. Water Research 85, 286–294. https://doi.org/10.1016/j.watres.2015.08.024

Le Moal, M., Gascuel-Odoux, C., Ménesguen, A., Souchon, Y., Étrillard, C., Levain, A., Moatar, F., Pannard, A., Souchu, P., Lefebvre, A., Pinay, G., 2019. Eutrophication: A new wine in an old bottle? Science of The Total Environment 651, 1–11. https://doi.org/10.1016/j.scitotenv.2018.09.139

Lehmann, M.K., Gurlin, D., Pahlevan, N., Alikas, K., Conroy, T., Anstee, J., Balasubramanian, S.V., Barbosa, C.C., Binding, C., Bracher, A., 2023. GLORIA-A globally representative hyperspectral in situ dataset for optical sensing of water quality. Scientific data 10, 100.

Li, J., Yu, Q., Tian, Y.Q., Becker, B.L., Siqueira, P., Torbick, N., 2018. Spatio-temporal variations of CDOM in shallow inland waters from a semi-analytical inversion of Landsat-8. Remote Sensing of Environment 218, 189–200. https://doi.org/10.1016/j.rse.2018.09.014

Li, L., Knapp, J.L.A., Lintern, A., Ng, G.-H.C., Perdrial, J., Sullivan, P.L., Zhi, W., 2024. River water quality shaped by land–river connectivity in a changing climate. Nat. Clim. Chang. 14, 225–237. https://doi.org/10.1038/s41558-023-01923-x

Li, S., Song, K., Wang, S., Liu, G., Wen, Z., Shang, Y., Lyu, L., Chen, F., Xu, S., Tao, H., Du, Y., Fang, C., Mu, G., 2021. Quantification of chlorophyll-a in typical lakes across China using Sentinel-2 MSI imagery with machine learning algorithm. Science of The Total Environment 778, 146271. https://doi.org/10.1016/j.scitotenv.2021.146271

Li, Y., Tang, C., Wang, J., Acharya, K., Du, W., Gao, X., Luo, L., Li, H., Dai, S., Mercy, J., Yu, Z., Pan, B., 2017. Effect of wave-current interactions on sediment resuspension in large shallow Lake Taihu, China. Environ Sci Pollut Res 24, 4029–4039. https://doi.org/10.1007/s11356-016-8165-0

Liu, D., Du, Y., Yu, S., Luo, J., Duan, H., 2020. Human activities determine quantity and composition of dissolved organic matter in lakes along the Yangtze River. Water Research 168, 115132. https://doi.org/10.1016/j.watres.2019.115132

Liu, D., Shi, K., Chen, P., Yan, N., Ran, L., Kutser, T., Tyler, A.N., Spyrakos, E., Woolway, R.I., Zhang, Y., Duan, H., 2024. Substantial increase of organic carbon storage in Chinese lakes. Nat Commun 15, 8049. https://doi.org/10.1038/s41467-024-52387-2

Liu, D., Yu, S., Xiao, Q., Qi, T., Duan, H., 2021. Satellite estimation of dissolved organic carbon in eutrophic Lake Taihu, China. Remote Sensing of Environment 264, 112572. https://doi.org/10.1016/j.rse.2021.112572

Liu, F.T., Ting, K.M., Zhou, Z.-H., 2012. Isolation-Based Anomaly Detection. ACM Trans. Knowl. Discov. Data 6, 3:1-3:39. https://doi.org/10.1145/2133360.2133363

Liu, F.T., Ting, K.M., Zhou, Z.-H., 2008. Isolation Forest, in: 2008 Eighth IEEE International Conference on Data Mining. Presented at the 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17

Liu, H., Li, Q., Bai, Y., Yang, C., Wang, J., Zhou, Q., Hu, S., Shi, T., Liao, X., Wu, G., 2021. Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. Remote Sensing of Environment 256, 112316. https://doi.org/10.1016/j.rse.2021.112316

Lotteraner, L., Möller, T., Hofmann, T., 2026. The Importance of Being Thorough: How Data Analysis Choices Impact the Perceived Relationship between Pollutants and Predictors. Water Research 288, 124639. https://doi.org/10.1016/j.watres.2025.124639

Ludwig, M., Moreno-Martinez, A., Hölzel, N., Pebesma, E., Meyer, H., 2023. Assessing and improving the transferability of current global spatial prediction models. Global Ecology and Biogeography 32, 356–368. https://doi.org/10.1111/geb.13635

Ma, C., Zhang, H., Huo, S., Li, W., Liu, Y., Xiao, Z., Xu, Y., Wu, F., 2023. Improving the estimation of nitrogen and phosphorus concentrations in lakes and reservoirs using a stacked approach. Earth's Future 11, e2022EF003013.

Maciel, D.A., Pahlevan, N., Barbosa, C.C.F., Martins, V.S., Smith, B., O'Shea, R.E., Balasubramanian, S.V., Saranathan, A.M., Novo, E.M.L.M., 2023. Towards global long-term water transparency products from the Landsat archive. Remote Sensing of Environment 299, 113889. https://doi.org/10.1016/j.rse.2023.113889

Malerba, M.E., Duarte de Paula Costa, M., Friess, D.A., Schuster, L., Young, M.A., Lagomasino, D., Serrano, O., Hickey, S.M., York, P.H., Rasheed, M., Lefcheck, J.S., Radford, B., Atwood, T.B., Ierodiaconou, D., Macreadie, P., 2023. Remote sensing for cost-effective blue carbon accounting. Earth-Science Reviews 238. https://doi.org/10.1016/j.earscirev.2023.104337

Maritorena, S., Siegel, D.A., Peterson, A.R., 2002. Optimization of a semianalytical ocean color model for global-scale applications. Appl. Opt., AO 41, 2705–2714. https://doi.org/10.1364/AO.41.002705

Massicotte, P., Asmala, E., Stedmon, C., Markager, S., 2017. Global distribution of dissolved organic matter along the aquatic continuum: Across rivers, lakes and oceans. Science of The Total Environment 609, 180–191. https://doi.org/10.1016/j.scitotenv.2017.07.076

Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. International Journal of Remote Sensing 32, 6855–6899. https://doi.org/10.1080/01431161.2010.512947

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O., 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat Commun 7, 13603. https://doi.org/10.1038/ncomms13603

Meybeck, M., 2003. Global Analysis of River Systems: From Earth System Controls to Anthropocene Syndromes. Philosophical Transactions of the Royal Society B. Biological Sciences 358, 1935–55. https://doi.org/10.1098/rstb.2003.1379

Meyer, H., Pebesma, E., 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat Commun 13, 2208. https://doi.org/10.1038/s41467-022-29838-9

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods in Ecology and Evolution 12, 1620–1633.

Meyer-Jacob, C., Michelutti, N., Paterson, A.M., Cumming, B.F., Keller, W. (Bill), Smol, J.P., 2019. The browning and re-browning of lakes: Divergent lake-water organic carbon trends linked to acid deposition and climate change. Sci Rep 9, 16676. https://doi.org/10.1038/s41598-019-52912-0

Milà, C., Mateu, J., Pebesma, E., Meyer, H., 2022a. Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. Methods in Ecology and Evolution 13, 1304–1316. https://doi.org/10.1111/2041-210X.13851

Milà, C., Mateu, J., Pebesma, E., Meyer, H., 2022b. Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. Methods in Ecology and Evolution 13, 1304–1316. https://doi.org/10.1111/2041-210X.13851

Moatar, F., Abbott, B.W., Minaudo, C., Curie, F., Pinay, G., 2017. Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions. Water Resources Research 53, 1270–1287.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50, 885–900.

Mouw, C.B., Greb, S., Aurin, D., DiGiacomo, P.M., Lee, Z., Twardowski, M., Binding, C., Hu, C., Ma, R., Moore, T., Moses, W., Craig, S.E., 2015. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. Remote Sensing of Environment 160, 15–30. https://doi.org/10.1016/j.rse.2015.02.001

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.-N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth System Science Data 13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Murphy, J., Riley, J.P., 1962. A modified single solution method for the determination of phosphate in natural waters. Analytica Chimica Acta 27, 31–36. https://doi.org/10.1016/S0003-2670(00)88444-5

Musolff, A., Schmidt, C., Selle, B., Fleckenstein, J.H., 2015. Catchment controls on solute export. Advances in water resources 86, 133–146.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Niroumand-Jadidi, M., Bovolo, F., Bruzzone, L., Gege, P., 2021. Inter-Comparison of Methods for Chlorophyll-a Retrieval: Sentinel-2 Time-Series Analysis in Italian Lakes. Remote Sensing 13, 2381. https://doi.org/10.3390/rs13122381

Odermatt, D., Gitelson, A., Brando, V.E., Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. Remote Sensing of Environment 118, 116–126. https://doi.org/10.1016/j.rse.2011.11.013

Olmanson, L.G., Brezonik, P.L., Finlay, J.C., Bauer, M.E., 2016. Comparison of Landsat 8 and Landsat 7 for regional measurements of CDOM and water clarity in lakes. Remote Sensing of Environment, Landsat 8 Science Results 185, 119–128. https://doi.org/10.1016/j.rse.2016.01.007

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M.K., Loisel, H., Matsushita, B., Hà, N., Olmanson, L., Potvin, G., Simis, S.G.H., VanderWoude, A., Vantrepotte, V., Ruiz-Verdù, A., 2022. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sensing of Environment 270, 112860. https://doi.org/10.1016/j.rse.2021.112860

Palmer, S.C.J., Kutser, T., Hunter, P.D., 2015. Remote sensing of inland waters: Challenges, progress and future directions. Remote Sensing of Environment, Special Issue: Remote Sensing of Inland Waters 157, 1–8. https://doi.org/10.1016/j.rse.2014.09.021

Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422. https://doi.org/10.1038/nature20584

Pi, X., Luo, Q., Feng, L., Xu, Y., Tang, J., Liang, X., Ma, E., Cheng, R., Fensholt, R., Brandt, M., Cai, X., Gibson, L., Liu, J., Zheng, C., Li, W., Bryan, B.A., 2022. Mapping global lake dynamics reveals the emerging roles of small lakes. Nat Commun 13, 5777. https://doi.org/10.1038/s41467-022-33239-3

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat Commun 11, 4540. https://doi.org/10.1038/s41467-020-18321-y

Powers, S.M., Bruulsema, T.W., Burt, T.P., Chan, N.I., Elser, J.J., Haygarth, P.M., Howden, N.J., Jarvie, H.P., Lyu, Y., Peterson, H.M., 2016. Long-term accumulation and transport of anthropogenic phosphorus in three river basins. Nature Geoscience 9, 353–356.

Qi, T., Xiao, Q., Cao, Z., Shen, M., Ma, J., Liu, D., Duan, H., 2020. Satellite Estimation of Dissolved Carbon Dioxide Concentrations in China's Lake Taihu. Environ. Sci. Technol. 54, 13709–13718. https://doi.org/10.1021/acs.est.0c04044

Qin, B., Hu, W., Gao, G., Luo, L., Zhang, J., 2004. Dynamics of sediment resuspension and the conceptual schema of nutrient release in the large shallow Lake Taihu, China. Chin. Sci. Bull. 49, 54–64. https://doi.org/10.1007/BF02901743

Qin, B., Zhu, G., Gao, G., Zhang, Y., Li, W., Paerl, H.W., Carmichael, W.W., 2010. A Drinking Water Crisis in Lake Taihu, China: Linkage to Climatic Variability and Lake Management. Environmental Management 45, 105–112. https://doi.org/10.1007/s00267-009-9393-6

Ramtel, P., Feng, D., Gardner, J., 2024. Toward Large-Scale Riverine Phosphorus Estimation Using Remote Sensing and Machine Learning. Journal of Geophysical Research: Biogeosciences 129, e2024JG008121. https://doi.org/10.1029/2024JG008121

Regnier, P., Resplandy, L., Najjar, R.G., Ciais, P., 2022. The land-to-ocean loops of the global carbon cycle. Nature 603, 401–410. https://doi.org/10.1038/s41586-021-04339-9

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

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881

Rode, M., Wade, A.J., Cohen, M.J., Hensley, R.T., Bowes, M.J., Kirchner, J.W., Arhonditsis, G.B., Jordan, P., Kronvang, B., Halliday, S.J., 2016. Sensors in the stream: the high-frequency wave of the present. ACS Publications.

Rodríguez-Murillo, J.C., Filella, M., 2015. Temporal evolution of organic carbon concentrations in Swiss lakes: Trends of allochthonous and autochthonous organic carbon. Science of The Total Environment 520, 13–22. https://doi.org/10.1016/j.scitotenv.2015.02.085

Ross, M.R.V., Topp, S.N., Appling, A.P., Yang, X., Kuhn, C., Butman, D., Simard, M., Pavelsky, T.M., 2019. AquaSat: A Data Set to Enable Remote Sensing of Water Quality for Inland Waters. Water Resources Research 55, 10012–10025. https://doi.org/10.1029/2019WR024883

Roulet, N., Moore, T.R., 2006. Browning the waters. Nature 444, 283–284. https://doi.org/10.1038/444283a

Roy, D.P., Qin, Y., Kovalskyy, V., Vermote, E.F., Ju, J., Egorov, A., Hansen, M.C., Kommareddy, I., Yan, L., 2014a. Conterminous United States demonstration and characterization of MODIS-based Landsat ETM + atmospheric correction. Remote Sensing of Environment 140, 433–449. https://doi.org/10.1016/j.rse.2013.09.012

Roy, D.P., Wulder, M.A., Loveland, T.R., C.e., W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014b. Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment 145, 154–172. https://doi.org/10.1016/j.rse.2014.02.001

Ruescas, A.B., Hieronymi, M., Mateo-Garcia, G., Koponen, S., Kallio, K., Camps-Valls, G., 2018. Machine Learning Regression Approaches for Colored Dissolved Organic Matter (CDOM) Retrieval with S2-MSI and S3-OLCI Simulated Data. Remote Sensing 10, 786. https://doi.org/10.3390/rs10050786

Safieddine, S.A., Heald, C.L., 2017. A Global Assessment of Dissolved Organic Carbon in Precipitation. Geophysical Research Letters 44, 11,672-11,681. https://doi.org/10.1002/2017GL075270

Sagan, V., Peterson, K.T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B.A., Maalouf, S., Adams, C., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth-Science Reviews 205, 103187. https://doi.org/10.1016/j.earscirev.2020.103187

Shen, F., Li, J., Chen, Z., 2025a. Comment on "Remote sensing estimation of dissolved organic carbon concentrations in Chinese lakes based on Landsat images" by Zhilong Zhao, Kun Shi and Yunlin Zhang. Journal of Hydrology 662, 133990. https://doi.org/10.1016/j.jhydrol.2025.133990

Shen, F., Li, J., Chen, Z., 2025b. Correspondence on "High-Resolution Estimates of N2O Emissions from Inland Waters and Wetlands in China." Environ. Sci. Technol. 59, 12408–12409. https://doi.org/10.1021/acs.est.5c04391

Shen, F., Yang, J., Wang, T., Li, N., Que, S., Xu, M., Li, J., Chen, Z., 2025c. Smartphone-based soil color analysis and machine learning for free iron oxide prediction in subtropical B-horizon soils. Precision Agric 26, 77. https://doi.org/10.1007/s11119-025-10270-0

Shi, K., Zhang, Y., Zhu, G., Qin, B., Pan, D., 2018. Deteriorating water clarity in shallow waters: Evidence from long term MODIS and in-situ observations. Int. J. Appl. Earth Obs. Geoinformation 68, 287–297. https://doi.org/10.1016/j.jag.2017.12.015

Smith, V.H., Schindler, D.W., 2009. Eutrophication science: where do we go from here? Trends in Ecology & Evolution 24, 201–207. https://doi.org/10.1016/j.tree.2008.11.009

Sobek, S., Tranvik, L.J., Prairie, Y.T., Kortelainen, P., Cole, J.J., 2007. Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes. Limnology and Oceanography 52, 1208–1219. https://doi.org/10.4319/lo.2007.52.3.1208

Song, K., Liu, G., Wang, Q., Wen, Z., Lyu, L., Du, Y., Sha, L., Fang, C., 2020. Quantification of lake clarity in China using Landsat OLI imagery data. Remote Sensing of Environment 243, 111800. https://doi.org/10.1016/j.rse.2020.111800

Song, K., Wen, Z., Shang, Y., Yang, H., Lyu, L., Liu, G., Fang, C., Du, J., Zhao, Y., 2018. Quantification of dissolved organic carbon (DOC) storage in lakes and reservoirs of mainland China. Journal of Environmental Management 217, 391–402. https://doi.org/10.1016/j.jenvman.2018.03.121

Stedmon, C.A., Markager, S., 2005. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. Limnology and Oceanography 50, 686–697. https://doi.org/10.4319/lo.2005.50.2.0686

Sun, X., Tian, L., Fang, H., Walling, D.E., Huang, L., Park, E., Li, D., Zheng, C., Feng, L., 2025. Changes in global fluvial sediment concentrations and fluxes between 1985 and 2020. Nat Sustain 8, 142–151. https://doi.org/10.1038/s41893-024-01476-7

Toming, K., Kotta, J., Uuemaa, E., Sobek, S., Kutser, T., Tranvik, L.J., 2020. Predicting lake dissolved organic carbon at a global scale. Sci Rep 10, 8471. https://doi.org/10.1038/s41598-020-65010-3

Topp, S.N., Pavelsky, T.M., Jensen, D., Simard, M., Ross, M.R.V., 2020. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. Water 12, 169. https://doi.org/10.3390/w12010169

Topp, S.N., Pavelsky, T.M., Stanley, E.H., Yang, X., Griffin, C.G., Ross, M.R.V., 2021. Multi-decadal improvement in US Lake water clarity. Environ. Res. Lett. 16, 055025. https://doi.org/10.1088/1748-9326/abf002

Tranvik, L., Cole, J.J., Prairie, Y.T., 2018. The study of carbon in inland waters-from isolated ecosystems to players in the global carbon cycle. Limnology and Oceanography Letters 3, 41–48.

Tranvik, L.J., Downing, J.A., Cotner, J.B., Loiselle, S.A., Striegl, R.G., Ballatore, T.J., Dillon, P., Finlay, K., Fortino, K., Knoll, L.B., Kortelainen, P.L., Kutser, T., Larsen, S., Laurion, I., Leech, D.M., McCallister, S.L., McKnight, D.M., Melack, J.M., Overholt, E., Porter, J.A., Prairie, Y., Renwick, W.H., Roland, F., Sherman, B.S., Schindler, D.W., Sobek, S., Tremblay, A., Vanni, M.J., Verschoor, A.M., von Wachenfeldt, E., Weyhenmeyer, G.A., 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnol Oceanogr 54, 2298–2314. https://doi.org/10.4319/lo.2009.54.6_part_2.2298

Tu, Z., Zhang, Y., Shi, K., Gong, S., Zhao, Z., 2024. Landsat data reveal lake deoxygenation worldwide. Water Research 267, 122525. https://doi.org/10.1016/j.watres.2024.122525

Tyler, A.N., Hunter, P.D., Spyrakos, E., Groom, S., Constantinescu, A.M., Kitchen, J., 2016. Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. Science of The Total Environment 572, 1307–1321. https://doi.org/10.1016/j.scitotenv.2016.01.020

Uddin, M.G., Rahman, A., Rosa Taghikhah, F., Olbert, A.I., 2024. Data-driven evolution of water quality models: An in-depth investigation of innovative outlier detection approaches-A case study of Irish Water Quality Index (IEWQI) model. Water Research 255, 121499. https://doi.org/10.1016/j.watres.2024.121499

Van Meter, K.J., Van Cappellen, P., Basu, N.B., 2018. Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico. Science 360, 427–430.

van Vliet, M.T.H., Thorslund, J., Strokal, M., Hofstra, N., Flörke, M., Ehalt Macedo, H., Nkwasa, A., Tang, T., Kaushal, S.S., Kumar, R., van Griensven, A., Bouwman, L., Mosley, L.M., 2023. Global river water quality under climate change and hydroclimatic extremes. Nat Rev Earth Environ 4, 687–702. https://doi.org/10.1038/s43017-023-00472-3

Vanhellemont, Q., Ruddick, K., 2018. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. Remote Sensing of Environment 216, 586–597. https://doi.org/10.1016/j.rse.2018.07.015

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment, Landsat 8 Science Results 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Virro, H., Amatulli, G., Kmoch, A., Shen, L., Uuemaa, E., 2021. GRQA: Global River Water Quality Archive. Earth System Science Data 13, 5483–5507. https://doi.org/10.5194/essd-13-5483-2021

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., 2010. Global threats to human water security and river biodiversity. nature 467, 555–561.

Vorosmarty, 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, 555–61. https://doi.org/10.1038/nature09440

Wadoux, A.M.J.-C., Heuvelink, G.B.M., de Bruin, S., Brus, D.J., 2021. Spatial cross-validation is not the right way to evaluate map accuracy. Ecological Modelling 457, 109692. https://doi.org/10.1016/j.ecolmodel.2021.109692

Wang, G., Liu, J., Wang, Z., Xiang, Y., Heng, C.K., Li, X., 2024. Spatiotemporal evolution and interaction of water constraints and their socio-ecological drivers in the Taihu Lake Basin. Science of The Total Environment 949, 175155. https://doi.org/10.1016/j.scitotenv.2024.175155

Wang, S., Gao, Y., Li, Q., Gao, J., Zhai, S., Zhou, Y., Cheng, Y., 2019. Long-term and inter-monthly dynamics of aquatic vegetation and its relation with environmental factors in Taihu Lake, China. Science of The Total Environment 651, 367–380. https://doi.org/10.1016/j.scitotenv.2018.09.216

Wang, S., Li, J., Zhang, B., Spyrakos, E., Tyler, A.N., Shen, Q., Zhang, F., Kuster, T., Lehmann, M.K., Wu, Y., Peng, D., 2018. Trophic state assessment of global inland waters using a MODIS-derived Forel-Ule index. Remote Sensing of Environment 217, 444–460. https://doi.org/10.1016/j.rse.2018.08.026

Whitehead, P.G., Wilby, R.L., Battarbee, R.W., Kernan, M., Wade, A.J., 2009. A review of the potential impacts of climate change on surface water quality. Hydrological sciences journal 54, 101–123.

Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J.D., Hostert, P., Hughes, M.J., Huntington, J., Johnson, D.M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T.A., Schaaf, C., Schott, J.R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J.C., Wynne, R.H., Zhu, Z., 2019. Current status of Landsat program, science, and applications. Remote Sensing of Environment 225, 127–147. https://doi.org/10.1016/j.rse.2019.02.015

Xie, S., Liu, L., Zhang, X., Yang, J., 2022. Mapping the annual dynamics of land cover in Beijing from 2001 to 2020 using Landsat dense time series stack. ISPRS Journal of Photogrammetry and Remote Sensing 185, 201–218. https://doi.org/10.1016/j.isprsjprs.2022.01.014

Xiong, J., Lin, C., Cao, Z., Hu, M., Xue, K., Chen, X., Ma, R., 2022. Development of remote sensing algorithm for total phosphorus concentration in eutrophic lakes: Conventional or machine learning? Water Research 215, 118213. https://doi.org/10.1016/j.watres.2022.118213

Xue, K., Ma, R., Zhu, G., Hu, M., Cao, Z., Xiong, J., Zhang, Y., Xu, J., Huang, Z., Wu, Y., 2024. A comprehensive time-series dataset linked to cyanobacterial blooms in Lake Taihu. Sci Data 11, 1365. https://doi.org/10.1038/s41597-024-04224-w

Yan, N., Qiu, Z., Zhang, C., Yan, Y., Liu, D., 2025. Landsat monitoring reveals the history of river organic pollution across China during 1984–2023. Water Research 275, 123210. https://doi.org/10.1016/j.watres.2025.123210

Yang, J., Shen, F., Wang, T., Luo, M., Li, N., Que, S., 2021. Effect of smart phone cameras on color-based prediction of soil organic matter content. Geoderma 402, 115365. https://doi.org/10.1016/j.geoderma.2021.115365

Ye, L., Wu, X., Liu, B., Yan, D., Kong, F., 2015. Dynamics and sources of dissolved organic carbon during phytoplankton bloom in hypereutrophic Lake Taihu (China). Limnologica 54, 5–13. https://doi.org/10.1016/j.limno.2015.05.003

Yin, Z., Li, J., Liu, Y., Xie, Y., Zhang, F., Wang, S., Sun, X., Zhang, B., 2021. Water clarity changes in Lake Taihu over 36 years based on Landsat TM and OLI observations. Int. J. Appl. Earth Obs. Geoinformation 102. https://doi.org/10.1016/j.jag.2021.102457

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. https://doi.org/10.5281/zenodo.7254221

Zhang, H., Sun, H., Li, J., Li, Y., Zhang, L., Zhao, R., Hu, X., Ren, N., Tian, Y., 2025. Natural and anthropogenic imprints on seasonal river water quality trends across China. npj Clean Water 8, 49. https://doi.org/10.1038/s41545-025-00481-3

Zhang, H., Xue, B., Wang, G., Zhang, X., Zhang, Q., 2022. Deep Learning-Based Water Quality Retrieval in an Impounded Lake Using Landsat 8 Imagery: An Application in Dongping Lake. Remote Sensing 14, 4505. https://doi.org/10.3390/rs14184505

Zhang, L., Ren, Z., Chen, B., Gong, P., Xu, B., Fu, H., 2024. A Prolonged Artificial Nighttime-light Dataset of China (1984-2020). Sci Data 11, 414. https://doi.org/10.1038/s41597-024-03223-1

Zhang, X., Bai, J., Zhang, Z., Xie, T., Zhang, G., Liu, Y., Chen, G., Liu, Z., 2023. Plant invasion strengthens the linkages between dissolved organic matter composition and the microbial community in coastal wetland soils. CATENA 232, 107449. https://doi.org/10.1016/j.catena.2023.107449

Zhang, X., Zhang, H., Yin, D., Gu, B., Chen, L., 2026. Machine learning reveals disruptive nutrient pollution shifts in Chinese rivers to 2100. npj Clean Water 9, 39. https://doi.org/10.1038/s41545-026-00571-w

Zhang, Yibo, Zhang, Yunlin, Shi, K., Zhou, Y., Li, N., 2021. Remote sensing estimation of water clarity for various lakes in China. Water Research 192, 116844. https://doi.org/10.1016/j.watres.2021.116844

Zhang, Z., Zhang, H., Jin, Y., Guo, H., Tian, S., Huang, J.J., Zhu, X., 2024. Analysing the spatiotemporal variation and influencing factors of Lake Chaohu's CDOM over the past 40 years using machine learning. Ecohydrology 17, e2639.

Zhao, N., 2023. A New Method for Spatial Estimation of Water Quality Using an Optimal Virtual Sensor Network and In Situ Observations: A Case Study of Chemical Oxygen Demand. Sensors 23, 4739. https://doi.org/10.3390/s23104739

Zhao, Y., He, X., Pan, S., Bai, Y., Wang, D., Li, T., Gong, F., Zhang, X., 2024. Satellite retrievals of water quality for diverse inland waters from Sentinel-2 images: An example from Zhejiang Province, China. International Journal of Applied Earth Observation and Geoinformation 132, 104048. https://doi.org/10.1016/j.jag.2024.104048

Zhao, Z., Shi, K., Zhang, Y., 2024a. Remote sensing estimation of dissolved organic carbon concentrations in Chinese lakes based on Landsat images. Journal of Hydrology 638, 131466. https://doi.org/10.1016/j.jhydrol.2024.131466

Zhao, Z., Shi, K., Zhang, Y., Zhou, Y., Qin, B., 2024b. Increased dominance of terrestrial component in dissolved organic matter in Chinese lakes. Water Research 249, 121019. https://doi.org/10.1016/j.watres.2023.121019

Zhou, Y., He, B., Cao, X., Xiao, Y., Feng, Q., Yang, F., Xiao, F., Geng, X., Du, Y., 2024. Remotely sensed estimates of long-term biochemical oxygen demand over Hong Kong marine waters using machine learning enhanced by imbalanced label optimisation. Science of The Total Environment 943, 173748. https://doi.org/10.1016/j.scitotenv.2024.173748

Zhou, Y., Yu, X., Zhou, L., Zhang, Y., Xu, H., Zhu, M., Zhu, G., Jang, K.-S., Spencer, R.G.M., Jeppesen, E., Brookes, J.D., Kothawala, D.N., Wu, F., 2023. Rainstorms drive export of aromatic and concurrent bio-labile organic matter to a large eutrophic lake and its major tributaries. Water Research 229, 119448. https://doi.org/10.1016/j.watres.2022.119448

Zhou, Y., Zhou, J., Jeppesen, E., Zhang, Y., Qin, B., Shi, K., Tang, X., Han, X., 2016. Will enhanced turbulence in inland waters result in elevated production of autochthonous dissolved organic matter? Science of The Total Environment 543, 405–415. https://doi.org/10.1016/j.scitotenv.2015.11.051

Zhu, W., Yu, Q., Tian, Y.Q., Becker, B.L., Zheng, T., Carrick, H.J., 2014. An assessment of remote sensing algorithms for colored dissolved organic matter in complex freshwater environments. Remote Sensing of Environment 140, 766–778. https://doi.org/10.1016/j.rse.2013.10.015

# Curriculum vitae
# &
# List of publications

# Feilong Shen

**Date of birth:** 06.12.1996

**Place of birth:** Zhengzhou City, Henan Province, China

**Address:** Kamýcká 129, Suchdol, 165 00 Prague, Czech Republic

**E-mail:** shen@fzp.czu.cz; flydragon_shen@163.com

## Education

2022–2026 **Ph.D. candidate**

Applied and Landscape Ecology

Department of Applied Ecology, Faculty of Environmental Sciences

**Czech University of Life Sciences Prague, Czech Republic**

2019–2022 **Master's degree**

Resource Utilization and Plant Protection

Department of Geographic Information Science, College of Resources and Environment

**Huazhong Agricultural University, Wuhan, China**

2015–2019 **Bachelor's degree**

Agricultural Resources and Environment

Department of Soil Science, College of Resources and Environment

**Henan Agricultural University, Zhengzhou, China**

# Publications

**Shen, F.**, Li, J., Wang, T., Li, N., Yang, J., Xu, M., Yan, Y., He, L., Guo, Y.*, Chen, Z.*, 2026. Four decades of dissolved organic carbon dynamics in Lake Taihu using comprehensive Landsat-5/7/8 observations and advanced machine learning models with investigation on driving factors. Journal of Hydrology 676, 135696.

**Shen, F.**, Li, J., Fan, S., Luo, T., Guo, Y.*, Chen, Z.*. 2026. Spatial cross-validation reveals a double inseparability boundary for satellite monitoring of narrow-river nutrients. **Manuscript submitted to Remote Sensing of Environment**.

**Shen, F.**, Li, J., Fan, S., Li, N., Li, Z., Xue, M.*, Chen, Z.*. 2026. Catchment attributes outperform reflectance in predicting oxygen-demand concentrations of narrow headwater rivers: a spatially validated four-decade reconstruction. **Manuscript submitted to Journal of Hydrology**.

**Shen, F.**, Cao, Z., Zhang, C., Zhao, Y., Li, H., Liu, S., Jiang, G., Yang, J., Liu, F.*, 2026. City-scale assessment of green-roof carbon: high-frequency chamber fluxes and district inventories reveal microclimate-driven sinks and seasonal resilience. Building and Environment 294, 114431.

**Shen, F.**, Chen, Z.*, 2026. Comment on "Early warning of harmful cyanobacteria blooms based on high frequency in situ monitoring and intelligible machine learning modelling: The case study of Lake Müggelsee (Germany)" by Recknagel et al. (Water Research 287, 2025, 124514). Water Research 293, 125406.

**Shen, F.**, Yang, J., Wang, T.*, Li, N., Que, S., Xu, M., Li, J., Chen, Z., 2025. Smartphone-based soil color analysis and machine learning for free iron oxide prediction in subtropical B-horizon soils. Precision Agriculture 26, 77.

**Shen, F.**, Li, J., Chen, Z.*, 2025. Comment on "Remote sensing estimation of dissolved organic carbon concentrations in Chinese lakes based on Landsat images" by Zhilong Zhao, Kun Shi and Yunlin Zhang. Journal of Hydrology 662, 133990.

**Shen, F.**, Li, J., Chen, Z.*, 2025. Correspondence on "High-Resolution Estimates of N2O Emissions from Inland Waters and Wetlands in China." Environmental Science & Technology 59, 12408–12409.

Fan, S., Guo, X., Wang, X., Gao, Y., Xu, M., **Shen, F.***, Yang, X.*, 2026. Eggshell-derived calcium redistributes legacy phosphorus toward less labile operational pools while sustaining pak choi yield in greenhouse vegetable systems. Journal of Environmental Management 415, 130674.

Xu, M.*, Li, X., Fan, S., **Shen, F.***, Yao, C., Zhang, Z., Wang, J., Zhang, X., 2026. Improvement in vegetation ecological quality and its driving factors in China over 2001-2024. Environmental Research 297, 124151.

* Corresponding author.

# Grants and projects

**IGA Project No. 2024B0005.** Assessment of the changes in aquatic ecosystems in South Bohemia, Czech Republic. Internal Grant Agency of the Faculty of Environmental Sciences, Czech University of Life Sciences Prague. Role: Principal investigator.

# Participation in Conferences

2024.05 Centennial Celebration and Congress of the IUSS 2024, Florence, Italy. Oral presentation and poster presentation.

2022.11 Kostelecké inspirování 2022, Prague, Czech Republic. Poster presentation.

# Awards

2026.05 Dean's Award (3rd place), Faculty of Environmental Sciences, Czech University of Life Sciences Prague.

2025.10 Best Ph.D. Student Paper Award (4th place), Czech University of Life Sciences Prague.