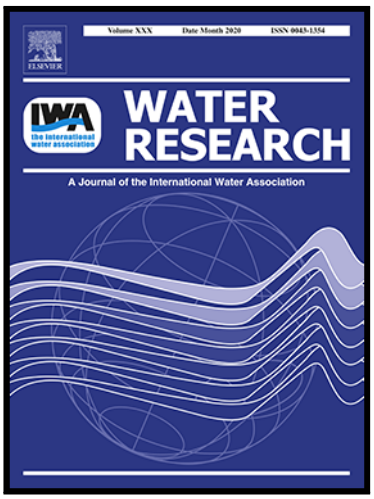


A comprehensive conceptual framework for signaling in-lake CO_2 through dissolved organic matter

Maofei Ni , Rui Liu , Weijun Luo , Junbing Pu , Shengjun Wu , Zhikang Wang , Jing Zhang , Xiaodan Wang , Yongmei Ma

PII: S0043-1354(24)01127-8
DOI: https://doi.org/10.1016/j.watres.2024.122228
Reference: WR 122228

To appear in: *Water Research*

Received date: 15 April 2024
Revised date: 4 August 2024
Accepted date: 6 August 2024

Please cite this article as: Maofei Ni , Rui Liu , Weijun Luo , Junbing Pu , Shengjun Wu , Zhikang Wang , Jing Zhang , Xiaodan Wang , Yongmei Ma , A comprehensive conceptual framework for signaling in-lake CO_2 through dissolved organic matter, *Water Research* (2024), doi: https://doi.org/10.1016/j.watres.2024.122228

Highlights:

- Lake CO_2 and DOM turnovers follow common biogeochemical pathways
- Subtropical lakes can be carbon sinks or sources due to varying CO_2 mechanisms
- DOM biodegradation governs temporal variations in CO_2 relative to terrestrial inputs
- Photo-mineralization of activated aromatic compounds fuels lake CO_2
- CO_2 levels are well modelled with 0.4%–2.1% overestimations *via* DOM signals

A comprehensive conceptual framework for signaling in-lake CO_2 through dissolved organic matter

Maofei Ni[1], Rui Liu[2,4], Weijun Luo[3], Junbing Pu[2,5], Shengjun Wu[6], Zhikang Wang[1], Jing Zhang[2]*, Xiaodan Wang[1], Yongmei Ma[6]**

1. College of Eco-Environmental Engineering, Guizhou Minzu University, Guiyang 550025, China.

2. School of Geography and Tourism, Chongqing Normal University, Chongqing 401331, China

3. State Key Laboratory of Environmental Geochemistry, Institute of Geochemistry, Chinese Academy of Sciences, Guiyang, 550081, China

4. The Key Laboratory of GIS Application Research, Chongqing Normal University, Chongqing 401331, China

5. Karst Research Team, Chongqing Key Laboratory of Carbon Cycle and Carbon Regulation of Mountain Ecosystem, School of Geography and Tourism, Chongqing Normal University, Chongqing 40133, China;

6. Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China.

*** Corresponding Author:**

Jing Zhang

Chongqing Normal University (CQNU)

Address: Chongqing Normal University, University Town, Shapingba District, Chongqing, P. R. China 401331

E-mail: zhangjing@cqnu.edu.cn

Yongmei Ma

Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences (CIGIT and CAS)

Address: 266, Fangzheng Avenue, Shuitu High-tech Park, Beibei, Chongqing 400714, China.

E-mail: mayongmei@cigit.ac.cn

Abstract

Organic carbon (C) and CO_2 pools are closely interactive in aquatic environments. While there are strong indications linking freshwater CO_2 to dissolved organic matter (DOM), the specific mechanisms underlying their common pathways remain unclear. Here, we present an extensive investigation from 20 subtropical lakes in China, establishing a comprehensive conceptual framework for identifying CO_2 drivers and retrieving CO_2 magnitude through co-trajectories of DOM evolution. Based on this framework, we show that lake CO_2 during wet period is constrained by a combination of biogeochemical processes, while photo-mineralization of activated aromatic compounds fuels CO_2 during dry period. We clearly determine that biological degradation of DOM governs temporal variations in CO_2 rather than terrestrial C inputs within the subtropical lakes. Specifically, our results identify a shared route for the uptake of atmospheric polycyclic aromatic compounds and CO_2 by lakes. Using machine learning, in-lake CO_2 levels are well modelled through DOM signaling regardless of varying CO_2 mechanisms. This study unravels the mechanistic underpinnings of causal links between lake CO_2 and DOM, with important implications for understanding obscure aquatic CO_2 drivers amidst the ongoing impacts of global climate change.

Keywords: Inland water CO_2, drivers and magnitude, dissolved organic matter, subtropical lakes, carbon cycling, biogeochemical pathways.

1. Introduction

Natural lakes act as dynamic reactors, emitters and sinks for aqueous CO_2 turnover, representing a substantial contribution to global carbon (C) cycle (Alin and Johnson, 2007; Pi et al., 2022). Most lakes (and reservoirs) show supersaturation in relation to atmospheric CO_2, with global CO_2 evasion estimated to range from 0.06 to 0.84 Pg C yr^{-1} (Raymond et al., 2013). The emission spectrum is admittedly broad, demonstrating that CO_2 drivers for current lakes are highly variable (Finlay et al., 2015; Perga et al., 2016; Weyhenmeyer et al., 2015). The prevailing consensus is that inland water CO_2 are fuelled by two main mechanisms: terrestrially-derived C inputs (Heathcote et al., 2015; Li et al., 2018) and actual degradation of organic C (Allesson et al., 2021; Begum et al., 2023). Terrestrial fluxes of CO_2 and dissolved organic C (DOC) released from soils into surface waters, for instance, typically account for CO_2 excess: the former may rapidly escape to the atmosphere (Campeau et al., 2019), while the latter will participate in aquatic degradation within lakes (Maberly et al., 2013). Biological respiration and photochemical mineralization of DOC result in the actual production of in-lake CO_2 (Dempsey et al., 2020). Consequently, it is important to recognize that the evolution of CO_2 and DOC likely shares common pathways. Yet to date, a conceptual framework for unraveling in-lake CO_2 through these co-trajectories with DOC is not well established, despite mounting evidence of their specific associations driven by geographical and biological processes (Hessen et al., 2017; Lapierre and Giorgio, 2015; Lapierre et al., 2013).

Dissolved organic matter (DOM), refers to the specific material form of DOC, involving the processes that govern organic C transport and transformation in natural waters (Lynch et al., 2019; Ni and Li, 2022). Lakes collect large quantities of allochthonous organic litters from surrounding shorelines, providing an important source, together with soil leaching, for DOM (Nakhavali et al., 2021; Wilkinson et al., 2013). Aquatic biology, meanwhile, functions as primary producers and contributors to autochthonous DOM pool in lakes (Hu et al., 2022). A large amount of natural DOM will undergo decomposition and eventually mineralize into CO_2 *via* photochemical and biological processes (Kellerman et al., 2015; Milstead et al., 2023). The DOM molecular composition, source and fate are interrelated within freshwater organic C pool and processing (Butturini et al., 2022). For example, in-lake lignin and soil fulvic acid originate commonly from terrestrial inputs (Cory et al., 2007; Yang et al., 2021), whereas carbohydrates, lipids and proteins are indicative of biological metabolism (Zhang et al., 2020). Microbial degradation and photobleaching alter DOM structure, and further modify molecular weight and recalcitrance (Hansen et al., 2016; Helms et al., 2014; Logozzo et al., 2021). All these characterizations can be simply extracted through DOM spectroscopy and molecular data analysis (Chen et al., 2019; Xu et al., 2020). From this perspective, DOM potentially serves as a signaling catalyst of organic C evolution.

Natural DOM can be established as a nexus, encompassing data on DOC origins and turnover, while also decodes the allocation of CO_2 *via* common DOC pathways

(Bodmer et al., 2016; Ni et al., 2023; Zhang et al., 2024). Specifically, the mechanistic underpinnings of the causal relationships between lake DOM and CO_2 involves: 1) the co-trajectories of lake CO_2 and organic C evolution (Begum et al., 2023; Lapierre et al., 2013); and 2) the inherent attributes of DOM that signal organic C source and fate (Kujawinski et al., 2009; Ni and Li, 2020). It is anticipated that dissolved CO_2 drivers can be identified through DOM signaling, as demonstrated by previous observations in aquatic environments (Kang et al., 2023; Ni et al., 2020b). However, these connections may be decoupled by spatiotemporal shifts of organic C pathways and mutual interferences among drivers (Winterdahl et al., 2016), making them not universally present. For instance, biologically derived connections between DOM and CO_2 could be decoupled due to inhibited respiration and strengthened geochemical connections e.g., terrestrial inputs and photochemical processes over space and time. Within this framework, the question arises whether there is a universal determination for retrieving both CO_2 levels and sources, in spite of dynamic CO_2 drivers and potential interferences. Furthermore, how can we figure out other CO_2 contributions that are unable to be gauged by explainable DOM variables.

In this study, we aim to identify and allocate in-lake CO_2 by linking with DOM signals, addressing the mechanistic underpinnings and ultimately providing a comprehensive conceptual framework for understanding these relationships. We collected 249 observations from 20 subtropical lakes in China, to quantify aqueous CO_2 levels, and analyse DOM spectroscopy and molecular information. Our

hypothesis postulated that the common pathways involved in CO_2 and DOM will vary at timescales attributing to biogeochemical shifts in subtropical lakes. In testing this hypothesis, we specifically: 1) examined the dynamics of aqueous CO_2 magnitude in lakes; 2) characterized DOM molecular composition, origin and fate; and 3) identified CO_2 drivers and modelled CO_2 levels through DOM signaling. This study proposes a DOM-based path for capturing varying CO_2 mechanisms in lakes, which can be universal and transferable for other natural waters.

2. Materials and methods

2.1. Study area and sampling

During August to April from 2022 to 2023, we sampled 134 sites from 20 lakes in subtropical China, geographically extending from 23º24′ to 31º32′N in latitude and from 98º57′ to 106º34′E in longitude (Fig. 1). Our sampling encompassed the bulk of representative subtropical lakes across Yunnan-Guizhou Plateau and Sichuan Basin, representing a broad spectrum of lake surface areas (<10–330 km^2). We realized that CO_2 and organic C biogeochemistry is potentially constrained by sampling locations. For instance, terrestrial C and nutrient levels are typically higher along the shorelines in comparison to offshore areas (Biddanda and Cotner, 2002). Consequently, our investigation involved sampling from both shorelines and more distant offshore areas to ensure spatial representativeness of lake samples (Fig. S1). Most lakes were revisited to incorporate temporal patterns of wet (May–October) and dry periods (December–April) in the subtropical region, resulting in a total of 249 observations

being collected. We acknowledge that geographical locations can affect regional climate and hydrology. Therefore, we carefully distinguish the wet and dry periods during the investigation. Specifically, it mostly includes mountainous lakes to minimize potential interferences from groundwaters and human activities. Being in the monsoonal climate zone, sampling locations exhibit mean annual temperature and precipitation varying within 10–20°C and 700–1300 mm yr^{-1}, respectively. The terrain is composed largely of mountains and hills, with elevations ranging from 260 to 2700 m. Detailed information on sampling locations see Supplement Materials (Appendix Table A1 and Fig. S1).

Before water sampling, gas was collected *in-situ* from the atmosphere using gas sampling bag at each lake. Surface water (~10 cm) was collected at each site, and 1 L water samples were filtered through glass microfiber filters (GF/F 47, 0.7 μm, Whatman). The filtrates were refrigerated in 100 mL high-density polyethylene (HDPE) plastic containers at 4°C. *In-situ* pH and water temperature (T) were measured using a portable CyberScan PCD 650 multi-parameter system (Eutech, USA), while wind speed (U_0) was determined with a Testo410-1 anemometer (Testo, Germany). Total alkalinity was titrated with Alkalinity Test MColortestTM (Merck, Germany). Subsequent UV absorbance and fluorescence spectroscopy were conducted for all water sample (n = 249). Given the high experimental cost, we mixed water samples (equal volumes) from each of the lakes as composite samples for Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS)

characterization to acquire comprehensive insights into molecular-level DOM. While this approach does not result in information loss regarding the DOM molecules, it presents a potential limitation in analysing the temporal variations of molecular DOM. Two small lakes were excluded due to insufficient filtrate volumes for FT-ICR MS, leaving a total of 18 composite samples. The subsequent solid-phase extraction (SPE) and FT-ICR MS analysis were commissioned to the China National Analytical Center, Guangzhou.

2.2. Laboratory analysis

Gas samples were determined for atmospheric CO_2 using gas chromatography-flame ionization detection (Fuli Analytical Instrument Co., Ltd., China). Water chemistry, UV absorbance and fluorescence spectroscopy were analysed for the filtered water samples. Lake DOC and dissolved inorganic carbon (DIC) concentrations were measured using a multi N/C 2100S analyzer (Analytik Jena, Germany) in triplicate. Chromophoric DOM was assessed using a double-beam scanning spectrophotometer (UV-5500 PC, Shanghai). The absorption spectra ranged from 200 to 700 nm with an interval of 1 nm. Fluorescence DOM was analysed using a RF-6000 Spectrophotometer (Shimadzu, Japan). Excitation-emission matrices (EEM) were scanned from 200 to 450 nm with a 5-nm interval for excitation wavelengths, and from 250 to 600 nm with a 1-nm interval for emission wavelengths.

Molecular composition of lake DOM was determined using FT-ICR MS (Bruker

Solari X, Germany). For SPE, water samples were acidified with hydrochloric acid to pH 2.0, and passed through a PPL column (200 mg, 3 mL). The PPL column was activated using 30 mL methanol, and salt was removed using 30-mL Milli-Q water and hydrochloric acid. The flow rate was kept at 1–2 mL min^{-1}, and samples were washed with 3 mL methanol (HPLC grade). Negative electrospray ionization (ESI) ion source was injected into a Bruker Solarix XR 7.0T FT-ICR MS system at a flow rate of 120 μL h^{-1}. The injection and extraction voltages were was set to 4.0 kV and -500 V for capillary inlets. The mass-to-charge ratios (m/z) were within the range of 100–1200, with an ion accumulation time of 0.02 or 0.06 s. During FT-ICR MS analysis, m/z = 369.119106 served as the internal reference ion for real-time mass calibration. It acquired 300 single transients/scans to improve the signal-to-noise ratio (S/N) of the target peaks. For high-quality FT-ICR MS, Suwannee River natural organic matter (SRNOM) was used as the standard for calibrating mass axis.

2.3. Lake CO_2 calculation and DOM characterization

We defined aqueous partial pressure of CO_2 (pCO_2) as CO_2 level in lakes, which was calculated from pH, water temperature, and total alkalinity or DIC using CO2SYS program (D.E. Pierrot et al., 2006). This program is developed on a basis of water chemistry or carbonate equilibria, enabling the calculation of DIC species concentrations, as well as pCO_2 or fugacity of CO_2 (fCO_2). The fCO_2 is about 0.3% lower than the pCO_2 under pressures of ~1 atm due to the non-ideality of CO_2. Carbonate equilibria are also constrained by the solubility coefficient of CO_2 (K_0),

first (K_1) and second dissociation (K_2) constants of carbonic acid in natural waters as follows:

$$fCO_2 = \frac{DIC}{K_0} \times \frac{[H^+]^2}{[H^+]^2 + K_1 \times [H^+] + K_1 \times K_2} \quad (1)$$

To better understand the potential CO_2 dynamics within the subtropical lakes, we computed the areal CO_2 efflux at distinct spatial and temporal scales. We estimated normalized gas transfer velocity (k_{600}, m d^{-1}) of CO_2 by relating it to wind speed, using previously reported empirical models (Li et al., 2016). It was converted into site-specific gas transfer velocity (k, m d^{-1}) with water temperatures and Schmidt numbers. Henry's constant of CO_2 (K_h, mmol m^{-3} μatm^{-1}) was corrected for *in-situ* temperature and pressure. The disparity between in-lake (pCO_{2water}, μatm) and atmospheric pCO_2 (pCO_{2air}, μatm) drives the transfer of CO_2. We noted relatively high CO_2 concentrations (434–496 ppm, with a mean of 463 ± 20 ppm) and thus pCO_{2air} in each lake (Fig. S2), potentially upscaling the baseline for assessing lake CO_2 outgassing. By measuring and compiling these parameters across sampling locations, we calculated water-air areal CO_2 efflux (mmol m^{-2} d^{-1}) using thin boundary layer (TBL) model as follows:

$$CO_2 \text{ efflux} = k \times K_h \times (pCO_{2water} - pCO_{2air}) \quad (2)$$

For detailed descriptions of estimating areal CO_2 efflux in the study subtropical lakes see Supplementary Information S1.

Chromophoric DOM was characterized by the non-normalized absorption coefficients of wavelengths at 254 (a_{254}, a proxy for aromatic abundance) (Ni et al.,

2020a), 350 (a_{350}, a proxy for lignin abundance) (Derrien et al., 2019) and 420 (a_{420}, an indication for photochemical mineralization of lake DOC) (Koehler et al., 2016) in Naperian units. Although these absorption coefficients are highly correlated, they have specific implications for DOM characterization, and the use of multiple coefficients is more conducive to accurate prediction of DOC concentration (Fichot and Benner, 2011). $SUVA_{254}$, a proxy for DOM aromaticity (D'Andrilli et al., 2022), was further calculated based on a_{254} and DOC concentration. Spectral slope coefficient ($S_{275\text{-}295}$), calculated from exponential fitting to the absorption spectrum of 275–295 nm, was used as a proxy for DOM microbial degradation, photobleaching and molecular weight (Helms et al., 2008). Spectral slope ratio (S_R), the ratio of $S_{275\text{-}295}$ to $S_{350\text{-}400}$ (computed similarly to $S_{275\text{-}295}$), had a negative correlation with DOM molecular weight. Fluorescence DOM was modelled using fluorescence regional integration across five EEM regions (Chen et al., 2003): region I (tyrosine-like), II (tryptophan-like) and IV (microbial exudates) represent biogenic DOM; region III (soil fulvic acid-like) links to terrestrial DOM; and region V expresses humic-like DOM. Fluorescence index (FI) and humification index (HIX) can indicate lake DOM origin and fate (Wilson and Xenopoulos, 2009). Specifically, FI was calculated from emission intensities of 450 to 500 nm at an excitation of 370 nm; while HIX was determined from total emission intensities of 435–480 nm to 300–345 nm at an excitation of 254 nm. We identified the main fluorophores with maximum fluorescence intensity (F_{max}) in the 20 subtropical lakes using parallel factor analysis (PARAFAC). The component models were assessed and validated using residual

analysis and split half analysis. For detailed descriptions of computing chromophoric and fluorescence DOM see Table S1.

Molecular composition of DOM was analysed using a molecular formula calculator based on criteria with elemental combinations of $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-1}S_{0-1}$, and assigned to the peaks with S/N > 4. For FT-ICR MS data, the modified aromaticity index (AI-mod), double bond equivalent (DBE) and nominal oxidation state of carbon (NOSC) were calculated for each assigned molecular as follows:

$$AI\_mod = \frac{1 + \mathrm{C} - \frac{1}{2}O - S - \frac{1}{2}(N + H)}{\mathrm{C} - \frac{1}{2}O - N - S} \quad (3)$$

$$DBE = 1 + \frac{1}{2}(2C - H + N) \quad (4)$$

$$NOSC = \frac{4\mathrm{C} + \mathrm{H} - 3\mathrm{N} - 2\mathrm{O} - 2\mathrm{S}}{\mathrm{C}} \quad (5)$$

Van Krevelen analysis was used to decipher molecular composition from 7 distinct regions based on the elemental ratios of H/C and O/C (Wang et al., 2023): carbohydrates (1.5 < H/C < 2.4; 0.71 < O/C < 1.2); amino-sugars (1.5 < H/C < 2.2; 0.52 < O/C < 0.71; N > 0); saturated compounds (1.5 < H/C < 2.2; 0 < O/C < 0.52; N > 0); tannin (0.5 < H/C < 1.5; 0.67 < O/C < 1.2); lignin (0.7 < H/C < 1.5; 0.1 < O/C < 0.67); unsaturated hydrocarbons (0.7 < H/C < 1.5; 0 < O/C < 0.1); condensed aromatic structures (0.2 < H/C < 0.7; 0 < O/C < 0.67). We included a visual map to aid in understanding the classification in Fig. S3. Furthermore, molecules were categorized into five compound groups based on AI-mod, H/C and O/C cutoffs (Hu et al., 2023; Linkhorst et al., 2017): combustion-derived polycyclic aromatic structures (CA, AI-mod > 0.66), soil-derived polyphenols and PCAs with aliphatic chains

(SP&PCAs, 0.66 > AI-mod > 0.50), soil-derived humic and highly unsaturated compounds (SH&UNs, AI-mod ≤ 0.50, H/C < 1.5), unsaturated aliphatic compounds (UAs, 2.0 ≥ H/C ≥ 1.5), and saturated fatty and sulfonic acids, carbohydrates (SF&SAC, H/C ≥ 2.0 or O/C ≥ 0.9).

2.4. Statistical analysis and data uncertainty

We tested normality and homogeneity of variance using Kolmogorov-Smirnov and Levene's test, respectively. Variables were log-transformed when it was statistically necessary to assume normality. Mann-Whitney *U* test was used to examine temporal variability of in-lake CO_2 and DOM datasets. One-way analysis of variance (ANOVA, with Turkey HSD post hoc) allowed us to assess statistical differences in the abundance or proportion of DOM molecular compositions. Linear regression and correlation analysis were assigned to indicate the homology of distinct DOM compositions, and show possible associations of pCO_2 with DOM variables. Given these linear relationships were likely impacted by mutual interactions between variables, we conducted an in-depth identification of how lake DOM signals CO_2 sources and forecasts CO_2 levels using machine learning. Specifically, random forest model was performed to identify relative importance of CO_2 drivers and predict CO_2 magnitude in lakes. It constructs multiple decision trees using random subsets of the training data, and synthesizes all predictions from each tree or node. This approach, handles high-dimensional and complex datasets of aqueous gas emissions (Rocher-Ros et al., 2023), demonstrating the ability to manage irrelevant variables without

causing overfitting. We defined the weights of relative importance (%) as the increased errors caused by permuting variables. The minimum leaf node size and number of trees were set to 2 and 500, respectively. The training and testing datasets have a ratio of approximately 2:1, containing 162 and 85 data points, respectively. The optimal results with the highest R^2 were selected to represent the regressions between observed and modelled pCO_2 across multiple modeling. In this study, mathematical statistics were performed using OriginLab OriginPro 2024 and SPSS statistical package (19.0), while random forest model was developed using MATLAB (2021b) algorithm.

Dissolved CO_2 concentrations calculated from water chemistry or carbonate equilibria may involve systematic errors caused by potential issues with pH and alkalinity measurements (Liu et al., 2020). For pH calibration, the breakpoint is often pH 5.4 in data handling: this includes measurements with pH > 5.4 to filter out implausible readings and instrument errors (Hotchkiss et al., 2015). Also, a low pH may indicate significant interference from organic acids to DIC species computing. In this study, we demonstrate 95% of pH measurements > 7.0, suggesting that CO_2 calculation errors resulting from pH could be relatively minimal (Fig. S4). For total alkalinity calibration, we realized that non-carbonate alkalinity sourced from organic and inorganic acids (e.g., borate, phosphate, silicate and hydrogen sulfate) can result in significant errors, particularly for low-concentration total alkalinity (<1000 μmol L^{-1}) (Liu et al., 2020). In this context, we quantified an average uncertainty of 19% for

total alkalinity using ancillary data e.g., DOC, dissolved phosphate and nitrogen. In fact, carbonate alkalinity and DIC in freshwaters are conceptually interchangeable terms in water chemistry calculations and thus CO2SYS program (D.E. Pierrot et al., 2006). Therefore, we constrained lake pCO_2 uncertainties using total alkalinity-based and directly measured DIC. We show that consistent with the evaluation from ancillary data, DIC and pCO_2 were overestimated by 19% and 23% from total alkalinity-based determination, respectively (Fig. S5). Therefore, we used measured DIC instead of total alkalinity in estimating lake pCO_2 in this study, in order to minimize the uncertainties. This converted the systematic errors into instrument errors of DIC (and DOC) measurements (< 2% in triplicate). UV absorptions, spectral slopes and S_R were analyzed in duplicates for 10% data, suggesting an uncertainty of < 2%. The inner filter effects of EEM data could be minimal (Ohno, 2002), with over 99% of the recorded absorbance at 254 nm being less than 0.3. This left us with one sample that was diluted with Milli-Q water to address the interference issue. Raman and Rayleigh scatter was collected for fluorescence data using interpolation (Bahram et al., 2006). Fluorescence intensities were normalized to Raman Unit (R.U.) using water Raman peak area (Lee et al., 2018).

3. Results and discussion

3.1. Subtropical lake CO_2 levels and dynamics

Our investigations encompassed mountainous lakes from the Sichuan Basin to the Yunnan-Guizhou Plateau in subtropical China (Fig. 1). Lake pCO_2 spanned four

orders of magnitude, ranging from <10 to 9175 µatm (17–3385 µatm, 95% confidence intervals) with a mean of 736 ± 1289 µatm. The diverse magnitudes indicate substantial fluctuations in CO_2 drivers within these lakes. Particularly, pCO_2 in the wet period had a broad range (Fig. 1a), averaging much higher at 1063 ± 1691 µatm than the dry period of 413 ± 534 µatm (Mann-Whitney *U* test, $p < 0.001$). This is likely modulated by temporal rainfall and soil flushing, therefore influencing lateral CO_2 inputs (Vachon et al., 2017). We found large spatial shifts in pCO_2 levels in the study subtropical lakes, especially high averages > 2000 µatm for locations with human disturbances e.g., Dianchi, Xihu and Qinglong (Fig. S6). It should be noted that ~70% of the samples were undersaturated with CO_2, exhibiting a high potential for C sequestration in subtropical lakes. Lake Chenghai, for instance, retained DIC concentrations of 5413–9228 µmol L^{-1} but yielded an unusually low pCO_2 of only 163 ± 76 µatm.

The modelled *k* values varied from 0.23 to 6.6 m d^{-1} (0.44–1.3 m d^{-1}, 95% confidence intervals) and averaged at 0.76 ± 0.53 m d^{-1} (Fig. S7), which is consistent with the global estimates (0.74 m d^{-1}) from similar wind speed models (Raymond et al., 2013). Combining paired measurements of pCO_2 and *k* values taken at each sampling location, we further calculated areal CO_2 efflux ranging from -66 to 231 mmol m^{-2} d^{-1} (-15–64 mmol m^{-2} d^{-1}, 95% confidence intervals) for these subtropical lakes (Fig. S8). The mean value, 5 ± 33 mmol m^{-2} d^{-1}, was significantly different from the median value of -6 mmol m^{-2} d^{-1}, for the CO_2 fluxes. This raises a conflicting

observation as these subtropical lakes overall represent a net CO_2 efflux because of some uniquely high values, despite most sampling locations functioning as CO_2 sinks. Site-specific extreme events, for example particular human disturbance (Luo et al., 2022; Raymond et al., 2008), may be responsible for this. Our areal CO_2 efflux is lower than the previously reported mean value (27 ± 66 mmol m^{-2} d^{-1}) for lakes and reservoirs across China (Ran et al., 2021), yet aligns with a recent real-time estimation (< 10 mmol m^{-2} d^{-1}) for lakes in Yunnan-Guizhou Plateau (Sun et al., 2023). It is important to note that we excluded non-carbonate alkalinity contributions by using detected DIC, and employed a relatively high baseline for atmospheric CO_2 based on *in-situ* measurements (Fig. S2). These calibrations avoided potential overestimations and resulted in relatively low magnitudes of lake CO_2 evasion estimated. Our findings suggest that the study lakes emitted CO_2 in the wet period (14 ± 42 mmol m^{-2} d^{-1}), but switched to uptake CO_2 in the dry period (-3 ± 15 mmol m^{-2} d^{-1}, Fig. S8). This implies that C turnover, along with CO_2 drivers, may differ substantially at timescales in these lake ecosystems.

3.2. Subtropical lake DOM characterizing

Lake DOC concentrations varied from 1.0 to 50.4 mg L^{-1} (1.9–23.0 mg L^{-1}, 95% confidence intervals), averaging ~2-fold higher in the wet period (11.4 ± 7.63 mg L^{-1}) than the dry season (5.9 ± 4.88 mg L^{-1}, $p < 0.001$) (Fig. 2a). Aromatic and lignin abundance, as well as photo-mineralization potentials of lake DOM, were greater in the wet period, as evidenced by absorption coefficients a_{254}, a_{350} and a_{420} ($p < 0.001$).

DOM aromaticity ($SUVA_{254}$), however, was higher in the dry period ($p < 0.001$, Fig. 2a). This demonstrates concentrating effect of aromatic compounds during periods of drought stress. Alternatively, biological degradation of DOM can also modulate its aromaticity over time (Ni and Li, 2020). We suggest that photobleaching and microbial degradation drove DOM decay, resulting in large accumulation of photo- and bio-metabolites, as well as alterations in DOM relative molecular weight in the dry period, as indicated by spectral slope coefficient $S_{275-295}$ (Fig. 2a). As a result, seasonality can alter DOM photochemical and metabolic pathways in lakes, governing complete (mineralization) or incomplete (biolysis and photobleaching) degradation of organic C. Shifting sunlight and DOM composition over seasons may straightforwardly account for these effects (Cory et al., 2014; Gonsior et al., 2013). It further combined with $S_{350-400}$ to compute S_R, demonstrating that DOM relative molecular weight was statistically similar across time periods ($p > 0.05$, Fig. 2a).

The PARAFAC modeling identified four fluorescent components in the subtropical lakes (Fig. 2b), which were closely associated with terrestrial (C1) and biogenic humic-like (C2), as well as tryptophan-like DOM (C3 and C4). The accumulation of C1 and C3 in particular, sourced from soil leaching and algal metabolism, was intensified by the drought conditions (Fig. S9). Specifically, the coexisting C3 and C4 during the wet period, two alternative tryptophan-like components, represented a changeable relationship in F_{max} (Fig. S10), which likely resulted from a transition state caused by biological degradation of DOM (Ni and Li,

2023). We realized that PARAFAC modeling may omit individualized components, and thus it employed fluorescence regional integrations for a comprehensive understanding of fluorescence DOM (Fig. 2c). Our results suggest that humic-like DOM (region V) had the highest proportion of 13%–81% across the five EEM regions (ANOVA, $p < 0.001$). Biogenic (region I + II + IV) and terrigenous DOM (region III) showed no significant temporal variations (Mann-Whitney U test, $p > 0.05$), contributing 12%–85% and 2%–17% to the EEM regions, respectively. This demonstrates a robust buffer capacity within the subtropical lakes, able to withstand seasonal fluctuations in DOM evolution. FI varied between 1.5–2.1 with a mean of 1.8 ± 0.10 (Fig. 2d), indicating that autochthonous and allochthonous sources can be equally significant for these fluorophores (implications of FI see Table S1 for more details). We found that soil fulvic acid-like component largely contributed to DOM humification in the lakes, as evidenced by the associations between region III and HIX (Fig. S11).

The DOM molecular formulae were highly consistent across the lakes, as shown by H/C and O/C distributions in van Krevelen Diagram (Fig. S12). Terrestrially derived lignin compounds (or carboxy-rich acyclic molecules) were predominated, contributing 77%–86% of molecular DOM (Fig. 3a). Amino-sugars and saturated compounds, in particular, also exhibited the significant proportions of 2%–4% and 8%-18%, respectively. These DOM molecules, can be preferentially biodegraded in aerobic environments with high H/C >1.5 (Sleighter et al., 2014;

Spencer et al., 2015). We thus suggest that terrestrial inputs and aquatic metabolism, as well as their potential interactions (Hotchkiss et al., 2015), may act as substantial pathways for organic C evolution in the lakes. We further classified DOM into five compound groups, demonstrating that SH&UNs were highly prevalent, with proportions of 71%–83% (Fig. 3b). This aligns with terrigenous signals from fluorescence DOM characterizing (Fig. 2). By contrast, DOM hydrogenation increased relative abundance of UAs (13%–26%), along with low O/C (< 0.5, Fig. S12) and NOSC (< 0, Fig. 3c), resulting in the accumulation of photo-labile compounds (McDonough et al., 2022) within the subtropical lakes. It is especially the case for lake Qinglong, of which DOM had higher AI-mod and DBE (Fig. 3c). Here, photo-produced aromatics may interact with DOM biodegradation, as proposed by a recent study (Hu et al., 2023). Therefore, it is anticipated that biotic processes also concatenate photochemical pathways of DOM, constraining organic C turnover in the study lakes.

3.3. Lake DOM signaling CO_2 drivers and magnitudes

Organic C shares common pathways with CO_2, establishing their causal links in aquatic environments. This supports previously documented associations between DOM and dissolved CO_2 (Li et al., 2024; Luo and Li, 2021). However, these relationships are typically particular at timescales. By linking pCO_2 to DOM variables, for instance, we found that a combination of biogeochemical drivers i.e., terrestrial inputs, photo-mineralization and biological metabolism, governed lake CO_2

during wet period. In contrast, photochemical processes likely modulated aquatic CO_2 during dry period (see Fig. S13). This was in agreement with temporal shifts of pCO_2 levels (Fig. 1), given that photo-mineralization of organic C corresponds only a small fraction of CO_2 evasion (Allesson et al., 2021) relative to more diverse sources. It should be noted that from the perspective of DOM molecules, aromaticity or aromatic structures were strongly associated with pCO_2, serving as an important indicator for lake CO_2 levels (Fig. S14).

Nevertheless, this effort to elucidate lake CO_2 cannot rank the drivers and predict CO_2 levels. Here, we performed machine learning to identify and allocate in-lake CO_2 (Fig. 4). The reliable results from linear regression analysis (Fig. S13) were complemented by random forest modeling, further showing that microbial degradation of DOM (with the relative high importance of biogenic signals i.e., region Ⅱ, Ⅳ and FI) dominantly explained lake pCO_2 among multifaceted CO_2 drivers in the wet period (Fig. 4a). Moreover, the modeling additionally identified significant importance of the terrigenous (a_{350}) and photochemical signals (a_{420} and an ancillary signal $S_{275\text{-}295}$) in the dry period (Fig. 4b), which reveals that terrestrial inputs, together with photolytic degradation of DOM, constrained lake pCO_2. This supports our hypothesis, suggesting the biogeochemical co-trajectories involved in lake CO_2 and DOM evolution can greatly differ over time. We highlight that these CO_2 mechanisms function not only independently but also interact within organic C processing. By examining the interactions of DOM variables, we can determine that

aquatic metabolism of terrestrial DOM and photo-mineralization of activated aromatic compounds emerged as two potential pathways for CO_2 production in the subtropical lakes (see Fig. S15).

On the basis of these biogeochemical co-trajectories, we clearly show robust correlations between observed and modelled pCO_2 across time periods (Fig. 4c). This indicates there is a generic way to retrieve CO_2 levels from DOM signaling, despite temporal variations in CO_2 mechanisms. Particularly, the modelled mean pCO_2 levels showed good agreement with the measured values, with overestimations of only 0.4% and 2.1% in the wet and dry periods, respectively (Fig. 4d). However, regarding high pCO_2 (> 2500 μatm), the modelled values had significant uncertainties. This may arise from the contributions not fully gauged by the "explainable DOM variables" (Please refer to Fig. 5 for the definition of this term). For instance, our study identified a shared pathway for the uptake of atmospheric polycyclic aromatic compounds and CO_2 by lakes from a molecular perspective (Fig. S14), which could not be simply omitted particularly in subtropical lakes. Photosynthetic consumption of pCO_2 also complicated predictions of lake CO_2 levels (Fig. S16). Particulate organic C typically accounts for 7%–38% of the total organic C pool (Ostapenia et al., 2009; Rouillard et al., 2011) and transforms into DOC through fragmentation (Kiuru et al., 2018), which potentially contributes 51–280 μatm to lake CO_2 levels on average. We note that China's subtropical lakes are susceptible to eutrophication and algal blooms, depending on temporal temperature changes (Yindong et al., 2021) and spatial

locations e.g., eutrophication-induced algal blooms often emerge along the shoreline rather than in the relative central areas (Stadig et al., 2020; Wang et al., 2018). This potentially shapes spatiotemporal patterns of CO_2 levels in these subtropical lake ecosystems (Fig. S17). Mounting evidence that human activities and groundwater can regulate aquatic DOM and CO_2 (Connolly et al., 2020; Duvert et al., 2018; Kang et al., 2023; McDonough et al., 2022). Yet, there is still a lack of the specific explainable DOM variables to correspond with these inputs.

In this study, we propose a DOM-based path for understanding CO_2 drivers and magnitudes in lakes (Fig. 5). The mechanistic basis relies on the ubiquitous biogeochemical co-evolution of aquatic CO_2 with DOM. It develops a comprehensive conceptual framework to unveil coupling relationships between diverse DOM indicators and aquatic CO_2, showing clearly how to differentiate and identify CO_2 sources. Furthermore, we suggest that machine learning can unravel the intrinsic connections between various drivers despite potential interferences, and thus establish a universal determination for retrieving CO_2 in most natural waters. However, it should be noted that the specific CO_2 drivers and associated DOM-based models may vary between distinct aquatic environments e.g., arid regions and saline lakes. Regardless of predictable levels and identifiable sources of lake CO_2 and organic C, this determination can provide significant implications for C turnover in aquatic ecosystems. We realize that biological degradation of organic C apparently governs lake CO_2 magnitude. For example, microbial metabolism acting as a determinant

drove much higher CO_2 levels during wet period rather than terrestrial inputs during dry period (Fig. 4). This understanding deviates from the prevailing view (Drake et al., 2018; Einarsdottir et al., 2017; Lapierre et al., 2013), yet aligns with the observations in shallow lakes (Bogard et al., 2019), suggesting that terrestrial C inputs may not be the fundamental hotspots of CO_2 emissions in subtropical lakes. Subtropical lakes sequester a significant amount of dissolved C but emit less CO_2 into the atmosphere (Raymond et al., 2013) due to biogenic constraints, demonstrating the significance in C capture and subsequent C neutrality. Our efforts to signal aqueous CO_2 through DOM, are thus anticipated to provide new insights into C cycling in subtropical lakes amidst the ongoing impacts of global climate change.

4. Conclusion

Natural DOM shares specific biogeochemical pathways with CO_2 in aquatic environments, establishing mechanistic underpinnings of their causal links. Based on these associations, we present a comprehensive conceptual framework to identify and allocate CO_2 through DOM signals in China's subtropical lakes. We demonstrate that lake CO_2 was constrained by terrestrial inputs, photo-mineralization and biological metabolism during the wet period, while photochemical processes largely fuelled CO_2 during the dry period. The interacting mechanisms involving aquatic metabolism of terrestrial DOM and photo-mineralization of activated aromatic compounds were specifically identified as key drivers of lake CO_2. Our findings highlight the crucial role of aquatic biology in regulating CO_2 dynamics relative to terrestrial C inputs

within subtropical lakes. Using machine learning, aqueous pCO_2 levels were well modelled through DOM signaling, achieving average overestimations of 0.4% and 2.1% in the wet and dry periods, respectively. This study proposes a generalizable approach for unraveling aquatic CO_2, which may provide important insights into C cycling from inland waters amidst ongoing climate change.

Declaration of Competing Interest

All authors agree this submission and the authors declare that there is no conflict of interests regarding the publication of this article

Acknowledgments

This study was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB40020200), National Natural Science Foundation of China (NSFC grant no. 42107091 and 42167050), National Special Support Plan for High-Level Talent to Junbing Pu (Young Talent Plan, 2022), Guizhou Provincial Science and Technology Projects (ZK[2024]061), Science and Technology Research Project of Chongqing Municipal Education Commission KJQN202200517), Natural Science Foundation of Chongqing (CSTB2022NSCQ-LZX0022, 2022NSCQ-MSX1046 and 2024NSCQ-MSX3061), and the Second Tibetan Plateau Scientific Expedition and Research Program (Sub-item: 2019QZKK0601-3). Special thanks are given to the editor, and two anonymous reviewers for their constructive comments and suggestions.

Author contributions: M.N. directed the study, carried out the field works and wrote the manuscript. J.Z. collected samples and analysed Digital Elevation Model data. Y.M. collected and analysed samples, and contributed to paper writing. R.L. supported field works and sample analysis. W.L. and J.P. contributed to data interpretation and paper writing. S.W., Z.W. and X.W. provided comments on the manuscript. All authors reviewed the manuscript.

References

Alin, S.R. and Johnson, T.C. 2007. Carbon cycling in large lakes of the world: A synthesis of production, burial, and lake-atmosphere exchange estimates. Global Biogeochem Cy 21(3), GB3002.

Allesson, L., Koehler, B., Thrane, J.-E., Andersen, T. and Hessen, D.O. 2021. The role of photomineralization for CO_2 emissions in boreal lakes along a gradient of dissolved organic matter. Limnology and Oceanography 66(1), 158-170.

Bahram, M., Bro, R., Stedmon, C. and Afkhami, A. 2006. Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. J. Chemom. 20(3-4), 99-105.

Begum, M.S., Park, J.H., Yang, L., Shin, K.H. and Hur, J. 2023. Optical and molecular indices of dissolved organic matter for estimating biodegradability and resulting carbon dioxide production in inland waters: A review. Water Res 228(Pt A), 119362.

Biddanda, B.A. and Cotner, J.B. 2002. Love Handles in Aquatic Ecosystems: The Role of Dissolved Organic Carbon Drawdown, Resuspended Sediments, and Terrigenous Inputs in the Carbon Balance of Lake Michigan. Ecosystems 5(5), 431-445.

Bodmer, P., Heinz, M., Pusch, M., Singer, G. and Premke, K. 2016. Carbon dynamics and their link to dissolved organic matter quality across contrasting stream ecosystems. Sci Total Environ 553, 574-586.

Bogard, M.J., Kuhn, C.D., Johnston, S.E., Striegl, R.G., Holtgrieve, G.W., Dornblaser, M.M., Spencer, R.G.M., Wickland, K.P. and Butman, D.E. 2019. Negligible cycling of terrestrial carbon in many lakes of the arid circumpolar landscape. Nature Geoscience 12(3), 180-185.

Butturini, A., Herzsprung, P., Lechtenfeld, O.J., Alcorlo, P., Benaiges-Fernandez, R., Berlanga, M., Boadella, J., Freixinos Campillo, Z., Gomez, R.M., Sanchez-Montoya, M.M., Urmeneta, J. and Romani, A.M. 2022. Origin, accumulation and fate of dissolved organic matter in an extreme hypersaline shallow lake. Water Research 221, 118727.

Campeau, A., Bishop, K., Amvrosiadi, N., Billett, M.F., Garnett, M.H., Laudon, H., Öquist, M.G. and Wallin, M.B. 2019. Current forest carbon fixation fuels stream CO_2 emissions. Nat Commun 10(1), 1876.

Chen, W., Teng, C.-Y., Qian, C. and Yu, H.-Q. 2019. Characterizing Properties and Environmental Behaviors of Dissolved Organic Matter Using Two-Dimensional Correlation Spectroscopic Analysis. Environ Sci Technol 53(9), 4683-4694.

Chen, W., Westerhoff, P., Leenheer, J.A. and Booksh, K. 2003. Fluorescence Excitation−Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter. Environ Sci Technol 37(24), 5701-5710.

Connolly, C.T., Cardenas, M.B., Burkart, G.A., Spencer, R.G.M. and McClelland, J.W. 2020. Groundwater as a major source of dissolved organic matter to Arctic coastal waters. Nat Commun 11(1), 1479.

Cory, R.M., McKnight, D.M., Chin, Y.-P., Miller, P. and Jaros, C.L. 2007. Chemical characteristics of fulvic acids from Arctic surface waters: Microbial contributions and

photochemical transformations. Journal of Geophysical Research: Biogeosciences 112(G4).

Cory, R.M., Ward, C.P., Crump, B.C. and Kling, G.W. 2014. Sunlight controls water column processing of carbon in arctic fresh waters. Science 345(6199), 925-928.

D.E. Pierrot, D.W.R. Wallace and E. Lewis 2006. MS Excel Program Developed for CO_2 System Calculations. ORNL/CDIAC-105a. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.

D'Andrilli, J., Silverman, V., Buckley, S. and Rosario-Ortiz, F.L. 2022. Inferring Ecosystem Function from Dissolved Organic Matter Optical Properties: A Critical Review. Environ Sci Technol 56(16), 11146-11161.

Dempsey, C.M., Brentrup, J.A., Magyan, S., Knoll, L.B., Swain, H.M., Gaiser, E.E., Morris, D.P., Ganger, M.T. and Williamson, C.E. 2020. The relative importance of photodegradation and biodegradation of terrestrially derived dissolved organic carbon across four lakes of differing trophic status. Biogeosciences 17(24), 6327-6340.

Derrien, M., Brogi, S.R. and Gonçalves-Araujo, R. 2019. Characterization of aquatic organic matter: Assessment, perspectives and research priorities. Water Research 163, 114908.

Drake, T.W., Raymond, P.A. and Spencer, R.G.M. 2018. Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. Limnology and Oceanography Letters 3(3), 132-142.

Duvert, C., Butman, D.E., Marx, A., Ribolzi, O. and Hutley, L.B. 2018. CO_2 evasion along streams driven by groundwater inputs and geomorphic controls. Nature Geoscience 11(11), 813-818.

Einarsdottir, K., Wallin, M.B. and Sobek, S. 2017. High terrestrial carbon load via groundwater to a boreal lake dominated by surface water inflow. Journal of Geophysical Research: Biogeosciences 122(1), 15-29.

Fichot, C.G. and Benner, R. 2011. A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. Geophys Res Lett 38(3), L03610.

Finlay, K., Vogt, R.J., Bogard, M.J., Wissel, B., Tutolo, B.M., Simpson, G.L. and Leavitt, P.R. 2015. Decrease in CO_2 efflux from northern hardwater lakes with increasing atmospheric warming. Nature 519(7542), 215-218.

Gonsior, M., Schmitt-Kopplin, P. and Bastviken, D. 2013. Depth-dependent molecular composition and photo-reactivity of dissolved organic matter in a boreal lake under winter and summer conditions. Biogeosciences 10(11), 6945-6956.

Hansen, A.M., Kraus, T.E.C., Pellerin, B.A., Fleck, J.A., Downing, B.D. and Bergamaschi, B.A. 2016. Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. Limnol Oceanogr 61(3), 1015-1032.

Heathcote, A.J., Anderson, N.J., Prairie, Y.T., Engstrom, D.R. and del Giorgio, P.A. 2015. Large increases in carbon burial in northern lakes during the Anthropocene. Nat Commun 6(1), 10016.

Helms, J.R., Mao, J., Stubbins, A., Schmidt-Rohr, K., Spencer, R.G.M., Hernes, P.J. and Mopper, K. 2014. Loss of optical and molecular indicators of terrigenous dissolved organic matter during long-term photobleaching. Aquat Sci 76(3), 353-373.

Helms, J.R., Stubbins, A., Ritchie, J.D., Minor, E.C., Kieber, D.J. and 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(3), 955-969.

Hessen, D.O., Håll, J.P., Thrane, J.E. and Andersen, T. 2017. Coupling dissolved organic carbon, CO_2 and productivity in boreal lakes. Freshwater Biology 62(5), 945-953.

Hotchkiss, E.R., Hall, R.O., Sponseller, R.A., Butman, D., Klaminder, J., Laudon, H., Rosvall, M. and Karlsson, J. 2015. Sources of and processes controlling CO_2 emissions change with the size of streams and rivers. Nature Geoscience 8(9), 696-699.

Hu, A., Choi, M., Tanentzap, A.J., Liu, J., Jang, K.-S., Lennon, J.T., Liu, Y., Soininen, J., Lu, X., Zhang, Y., Shen, J. and Wang, J. 2022. Ecological networks of dissolved organic matter and microorganisms under global change. Nat Commun 13(1), 3600.

Hu, J., Kang, L., Li, Z., Feng, X., Liang, C., Wu, Z., Zhou, W., Liu, X., Yang, Y. and Chen, L. 2023. Photo-produced aromatic compounds stimulate microbial degradation of dissolved organic carbon in thermokarst lakes. Nat Commun 14(1), 3681.

Kang, W., Hu, X., Feng, R., Wei, C. and Yu, F. 2023. DOM Associates with Greenhouse Gas Emissions in Chinese Rivers under Diverse Land Uses. Environ Sci Technol 57(40), 15004-15013.

Kellerman, A.M., Kothawala, D.N., Dittmar, T. and Tranvik, L.J. 2015. Persistence of dissolved organic matter in lakes related to its molecular characteristics. Nature Geoscience 8(6), 454-457.

Kiuru, P., Ojala, A., Mammarella, I., Heiskanen, J., Kämäräinen, M., Vesala, T. and Huttula, T. 2018. Effects of Climate Change on CO_2 Concentration and Efflux in a Humic Boreal Lake: A Modeling Study. Journal of Geophysical Research: Biogeosciences 123(7), 2212-2233.

Koehler, B., Broman, E. and Tranvik, L.J. 2016. Apparent quantum yield of photochemical dissolved organic carbon mineralization in lakes. Limnology and Oceanography 61(6), 2207-2221.

Kujawinski, E.B., Longnecker, K., Blough, N.V., Vecchio, R.D., Finlay, L., Kitner, J.B. and Giovannoni, S.J. 2009. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. Geochimica et Cosmochimica Acta 73(15), 4384-4399.

Lapierre, J.F. and Giorgio, P.A. 2015. Geographical and environmental drivers of regional differences in the lake pCO_2 versus DOC relationship across northern landscapes. Journal of Geophysical Research Biogeosciences 117(G3), 6841-6847.

Lapierre, J.F., Guillemette, F., Berggren, M. and del Giorgio, P.A. 2013. Increases in terrestrially derived carbon stimulate organic carbon processing and CO_2 emissions in boreal aquatic ecosystems. Nat Commun 4, 2972.

Lee, M.H., Osburn, C.L., Shin, K.H. and Hur, J. 2018. New insight into the applicability of spectroscopic indices for dissolved organic matter (DOM) source discrimination in aquatic systems affected by biogeochemical processes. Water Res. 147, 164-176.

Li, J., Liang, E., Deng, C., Li, B., Cai, H., Ma, R., Xu, Q., Liu, J. and Wang, T. 2024. Labile dissolved organic matter (DOM) and nitrogen inputs modified greenhouse gas dynamics: A source-to-estuary study of the Yangtze River. Water Research 253, 121318.

Li, S., Bush, R.T., Santos, I.R., Zhang, Q., Song, K., Mao, R., Wen, Z. and Lu, X.X. 2018. Large greenhouse gases emissions from China's lakes and reservoirs. Water Res. 147, 13-24.

Li, S., Bush, R.T., Ward, N.J., Sullivan, L.A. and Dong, F. 2016. Air–water CO_2 outgassing in the Lower Lakes (Alexandrina and Albert, Australia) following a millennium drought. Sci Total Environ 542, 453-468.

Linkhorst, A., Dittmar, T. and Waska, H. 2017. Molecular Fractionation of Dissolved Organic Matter in a Shallow Subterranean Estuary: The Role of the Iron Curtain. Environ Sci Technol 51(3), 1312-1320.

Liu, S., Butman, D.E. and Raymond, P.A. 2020. Evaluating CO_2 calculation error from organic alkalinity and pH measurement error in low ionic strength freshwaters. Limnology and Oceanography: Methods 18(10), 606-622.

Logozzo, L., Tzortziou, M., Neale, P. and Clark, J.B. 2021. Photochemical and Microbial Degradation of Chromophoric Dissolved Organic Matter Exported From Tidal Marshes. Journal of Geophysical Research: Biogeosciences 126(4), e2020JG005744.

Luo, J. and Li, S. 2021. Optical properties of dissolved organic matter in a monsoonal headwater stream, China: Insights for structure, source and riverine pCO_2. Journal of Cleaner Production 282, 124545.

Luo, J., Zhou, Q., Hu, X., Zeng, H., Deng, P., He, C. and Shi, Q. 2022. Lake Chemodiversity Driven by Natural and Anthropogenic Factors. Environ Sci Technol 56(9), 5910-5919.

Lynch, L.M., Sutfin, N.A., Fegel, T.S., Boot, C.M., Covino, T.P. and Wallenstein, M.D. 2019. River channel connectivity shifts metabolite composition and dissolved organic matter chemistry. Nat Commun 10(1), 459.

Maberly, S.C., Barker, P.A., Stott, A.W. and De Ville, M.M. 2013. Catchment productivity controls CO_2 emissions from lakes. Nature Climate Change 3(4), 391-394.

McDonough, L.K., Andersen, M.S., Behnke, M.I., Rutlidge, H., Oudone, P., Meredith, K., O'Carroll, D.M., Santos, I.R., Marjo, C.E., Spencer, R.G.M., McKenna, A.M. and Baker, A. 2022. A new conceptual framework for the transformation of groundwater dissolved organic matter. Nat Commun 13(1), 2153.

Milstead, R.P., Horvath, E.R. and Remucal, C.K. 2023. Dissolved Organic Matter Composition Determines Its Susceptibility to Complete and Partial Photooxidation within Lakes. Environ Sci Technol 57(32), 11876-11885.

Nakhavali, M., Lauerwald, R., Regnier, P., Guenet, B., Chadburn, S. and Friedlingstein, P. 2021. Leaching of dissolved organic carbon from mineral soils plays a significant role in the terrestrial carbon balance. Global Change Biol 27(5), 1083-1096.

Ni, M., jiang, S. and Li, S. 2020a. Spectroscopic indices trace spatiotemporal variability of dissolved organic matter in a river system with Karst characteristic. Journal of Hydrology 590, 125570.

Ni, M. and Li, S. 2023. Ultraviolet humic-like component contributes to riverine dissolved organic matter biodegradation. Journal of Environmental Sciences 124, 165-175.

Ni, M., Liu, R., Luo, W., Pu, J., Zhang, J. and Wang, X. 2023. Unexpected shifts of dissolved carbon biogeochemistry caused by anthropogenic disturbances in karst rivers. Water Research 247, 120744.

Ni, M.F. and Li, S.Y. 2020. Optical properties as tracers of riverine dissolved organic matter biodegradation in a headwater tributary of the Yangtze. Journal of Hydrology 582, 124497.

Ni, M.F. and Li, S.Y. 2022. Dynamics and internal links of dissolved carbon in a karst river

system: Implications for composition, origin and fate. Water Res 226, 119289.

Ni, M.F., Li, S.Y., Santos, I., Zhang, J. and Luo, J.C. 2020b. Linking riverine partial pressure of carbon dioxide to dissolved organic matter optical properties in a Dry-hot Valley Region. Sci Total Environ 704, 135353.

Ohno, T. 2002. Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. Environ Sci Technol 36(4), 742-746.

Ostapenia, A.P., Parparov, A. and Berman, T. 2009. Lability of organic carbon in lakes of different trophic status. Freshwater Biology 54(6), 1312-1323.

Perga, M.-E., Maberly, S.C., Jenny, J.-P., Alric, B., Pignol, C. and Naffrechoux, E. 2016. A century of human-driven changes in the carbon dioxide concentration of lakes. Global Biogeochem Cy 30(2), 93-104.

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. and Bryan, B.A. 2022. Mapping global lake dynamics reveals the emerging roles of small lakes. Nat Commun 13(1), 5777.

Ran, L., Butman, D.E., Battin, T.J., Yang, X., Tian, M., Duvert, C., Hartmann, J., Geeraert, N. and Liu, S. 2021. Substantial decrease in CO_2 emissions from Chinese inland waters due to global change. Nat Commun 12(1), 1730.

Raymond, P.A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Durr, H., Meybeck, M., Ciais, P. and Guth, P. 2013. Global carbon dioxide emissions from inland waters. Nature 503(7476), 355-359.

Raymond, P.A., Oh, N.-H., Turner, R.E. and Broussard, W. 2008. Anthropogenically enhanced fluxes of water and carbon from the Mississippi River. Nature 451(7177), 449-452.

Rocher-Ros, G., Stanley, E.H., Loken, L.C., Casson, N.J., Raymond, P.A., Liu, S., Amatulli, G. and Sponseller, R.A. 2023. Global methane emissions from rivers and streams. Nature 621(7979), 530-535.

Rouillard, A., Rosén, P., Douglas, M.S.V., Pienitz, R. and Smol, J.P. 2011. A model for inferring dissolved organic carbon (DOC) in lakewater from visible-near-infrared spectroscopy (VNIRS) measures in lake sediment. Journal of Paleolimnology 46(2), 187-202.

Sleighter, R.L., Cory, R.M., Kaplan, L.A., Abdulla, H.A.N. and Hatcher, P.G. 2014. A coupled geochemical and biogeochemical approach to characterize the bioreactivity of dissolved organic matter from a headwater stream. Journal of Geophysical Research: Biogeosciences 119(8), 1520-1537.

Spencer, R.G.M., Mann, P.J., Dittmar, T., Eglinton, T.I., McIntyre, C., Holmes, R.M., Zimov, N. and Stubbins, A. 2015. Detecting the signature of permafrost thaw in Arctic rivers. Geophys Res Lett 42(8), 2830-2835.

Stadig, M.H., Collingsworth, P.D., Lesht, B.M. and Höök, T.O. 2020. Spatially heterogeneous trends in nearshore and offshore chlorophyll a concentrations in lakes Michigan and Huron (1998–2013). Freshwater Biology 65(3), 366-378.

Sun, K., Jia, J., Wang, S. and Gao, Y. 2023. Real-time and dynamic estimation of CO_2 emissions from China's lakes and reservoirs. The Innovation Geoscience 1(3), 100031.

Vachon, D., Solomon, C.T. and del Giorgio, P.A. 2017. Reconstructing the seasonal dynamics and relative contribution of the major processes sustaining CO_2 emissions in northern lakes. Limnology and Oceanography 62(2), 706-722.

Wang, S.C., Liu, X., Liu, Y. and Wang, H. 2018. Contrasting patterns of macroinvertebrates inshore vs. offshore in a plateau eutrophic lake: Implications for lake management. Limnologica 70, 10-19.

Wang, Y., Li, N., Fu, Q., Cheng, Z., Song, Y., Yan, B., Chen, G., Hou, L.a. and Wang, S. 2023. Conversion and impact of dissolved organic matters in a heterogeneous catalytic peroxymonosulfate system for pollutant degradation. Water Research 241, 120166.

Weyhenmeyer, G.A., Kosten, S., Wallin, M.B., Tranvik, L.J., Jeppesen, E. and Roland, F. 2015. Significant fraction of CO_2 emissions from boreal lakes derived from hydrologic inorganic carbon inputs. Nature Geoscience 8(12), 933-936.

Wilkinson, G.M., Pace, M.L. and Cole, J.J. 2013. Terrestrial dominance of organic matter in north temperate lakes. Global Biogeochem Cy 27(1), 43-51.

Wilson, H.F. and Xenopoulos, M.A. 2009. Effects of agricultural land use on the composition of fluvial dissolved organic matter. Nature Geoscience 2(1), 37-41.

Winterdahl, M., Wallin, M.B., Karlsen, R.H., Laudon, H., Öquist, M. and Lyon, S.W. 2016. Decoupling of carbon dioxide and dissolved organic carbon in boreal headwater streams. Journal of Geophysical Research: Biogeosciences 121(10), 2630-2651.

Xu, W., Gao, Q., He, C., Shi, Q., Hou, Z.Q. and Zhao, H.-Z. 2020. Using ESI FT-ICR MS to Characterize Dissolved Organic Matter in Salt Lakes with Different Salinity. Environ Sci Technol 54(20), 12929-12937.

Yang, B., Ljung, K., Nielsen, A.B., Fahlgren, E. and Hammarlund, D. 2021. Impacts of long-term land use on terrestrial organic matter input to lakes based on lignin phenols in sediment records from a Swedish forest lake. Sci Total Environ 774, 145517.

Yindong, T., Xiwen, X., Miao, Q., Jingjing, S., Yiyan, Z., Wei, Z., Mengzhu, W., Xuejun, W. and Yang, Z. 2021. Lake warming intensifies the seasonal pattern of internal nutrient cycling in the eutrophic lake and potential impacts on algal blooms. Water Research 188, 116570.

Zhang, L., Fang, W., Li, X., Gao, G. and Jiang, J. 2020. Linking bacterial community shifts with changes in the dissolved organic matter pool in a eutrophic lake. Sci Total Environ 719, 137387.

Zhang, T., Zhou, L., Zhou, Y., Zhang, Y., Guo, J., Han, Y., Zhang, Y., Hu, L., Jang, K.-S., Spencer, R.G.M., Brookes, J.D., Dolfing, J. and Jeppesen, E. 2024. Terrestrial dissolved organic matter inputs accompanied by dissolved oxygen depletion and declining pH exacerbate CO_2 emissions from a major Chinese reservoir. Water Research 251, 121155.

Figure captions:

Fig. 1. Spatiotemporal distributions of aqueous pCO_2 across the study lakes in subtropical China. Dots show the mean pCO_2 in each lake during the wet period (a) and dry period (b). Blue lines describe drainage networks across the study area. Detailed information of sampling locations is provided in Fig. S1.

Fig. 2. Spectroscopic characteristics of DOM in the subtropical lakes. (a) Temporal pattern of DOC and chromophoric DOM in the lakes. The boxes with whiskers represent 25%–75% range within 1.5IQR. White and black lines show median and mean, respectively. Dots correspond to all data of UV parameters. (b) Three-dimensional view of primary fluorophores identified by PARAFAC modeling. (c) Fluorescence regional integration across five excitation-emission matrix regions. The waves display Kernel Smooth distributions of proportions for each region. (d) Distributions of FI and HIX across the time periods. The red line and grey area indicate linear regression and 95% confidence band, respectively. Dots correspond to all data of fluorescent parameters. The waves show Kernel Smooth distributions of the data.

Fig. 3. Molecular information of DOM across the subtropical lakes. (a) Proportions of DOM molecular compounds in each lake, (b) Proportions of compound groups include combustion-derived polycyclic aromatic structures (CA), soil-derived polyphenols and PCAs with aliphatic chains (SP&PCAs), soil-derived humic and highly unsaturated compounds (SH&UNs), unsaturated aliphatic compounds (UAs), and saturated fatty and sulfonic acids, carbohydrates (SF&SAC). (c) The modified

aromaticity index (AI-mod), double bond equivalent (DBE) and nominal oxidation state of carbon (NOSC) across the lakes.

Fig. 4. Lake DOM signaling CO_2 drivers and magnitudes using machine learning. (a) Random forest modeling assigning the relative importance of DOM variable in evaluating lake CO_2. The x axis with whisker shows the mean importance with standard deviation (s.d.). (b) The visual flows of relative importance from DOM variable to temporal pCO_2 in the lakes. The red and blue flows represent the most variables with weights > 0.2. (c) The relationships between observed and modelled pCO_2 across the time periods. The waves show Kernel Smooth distributions of the data. (d) Comparison of observed and modelled pCO_2 in the lakes. The boxes with whiskers represent 25th–75th percentiles with s.d. White and black lines show median and mean, respectively. Dots correspond to all data of observed and modelled pCO_2.

Fig. 5 A comprehensive conceptual framework for signaling in-lake CO_2 through DOM. We suggest that lake organic C or DOM follows the shared biogeochemical pathways with lake CO_2. It includes the common drivers 1) e.g., photochemical drivers, aquatic metabolism, terrestrial inputs and atmospheric uptake that are directly signalled by explainable DOM variables; and 2) e.g., groundwater inputs, human activities and photosynthesis that co-vary with lake CO_2 although explainable variables signaling these relationships are lacking. The term “explainable DOM variables” refers to DOM variables that can specifically indicate certain sources or pathways regarding biogenic, terrigenous, and photochemical processes. By contrast, ancillary signals may interact with various pathways but are not specifically indicative

of any single explainable process. The causal relationships can be visualized through linear regression analysis, correlation analysis or machine learning. This helps to identify CO_2 drivers, rank contributions to lake CO_2, and particularly discover interactive trajectories in modulating lake CO_2 (e.g., aquatic metabolism of terrestrial DOM, and photo-mineralization of activated aromatic compounds). By compiling and analysing these common pathways, it allows us to predict lake CO_2 levels using machine learning. We highlight that this conceptual framework could be generic and transferable for other natural waters.

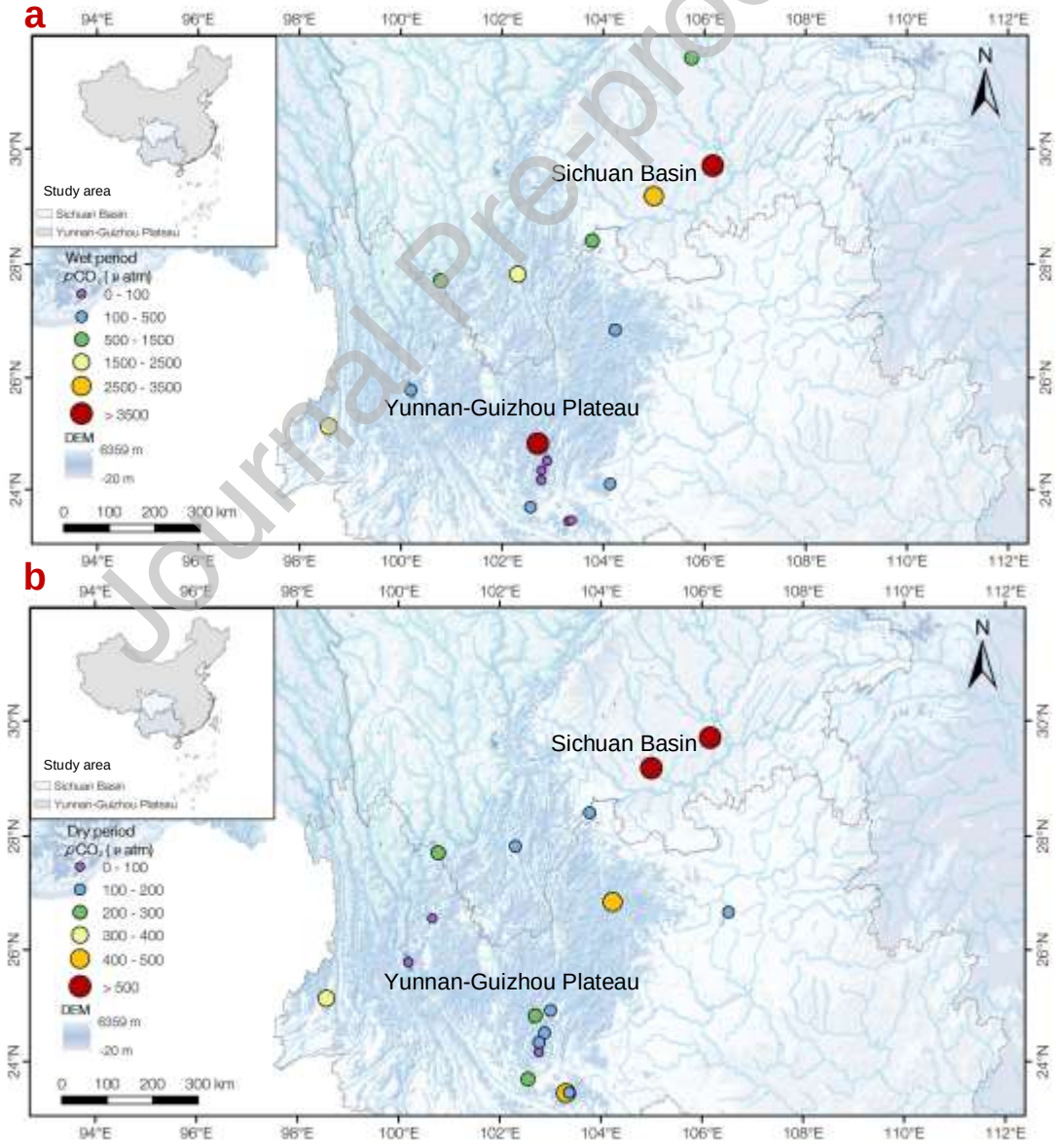


Fig. 1. Spatiotemporal distributions of aqueous pCO_2 across the study lakes in subtropical China. Dots show the mean pCO_2 in each lake during the wet period (a) and dry period (b). Blue and black lines describe drainage networks and boundaries across the study area, respectively. Detailed information of sampling locations is provided in Fig. S1.

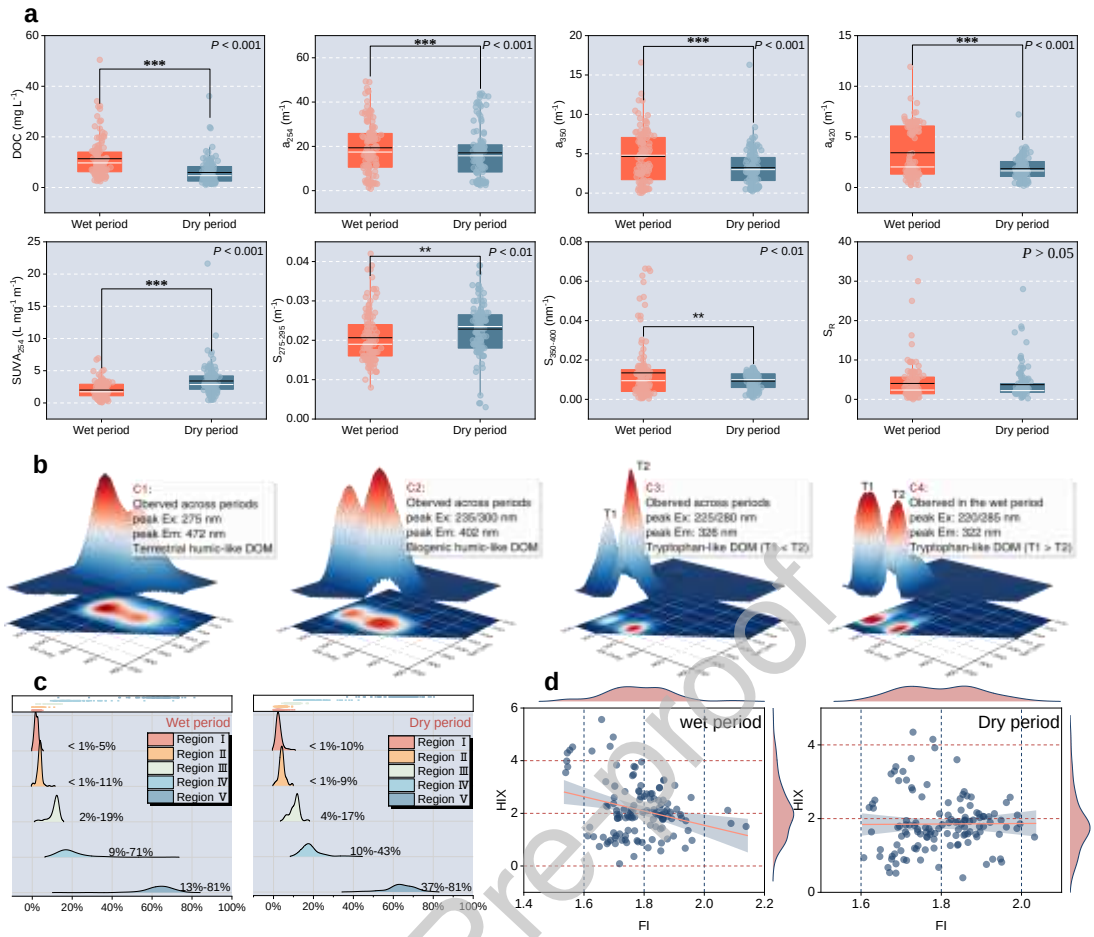


Fig. 2. Spectroscopic characteristics of DOM in the subtropical lakes. (a) Temporal pattern of DOC and chromophoric DOM in the lakes. The boxes with whiskers represent 25%–75% range within 1.5IQR. White and black lines show median and mean, respectively. Dots correspond to all data of UV parameters. (b) Three-dimensional view of primary fluorophores identified by PARAFAC modeling. (c) Fluorescence regional integration across five excitation-emission matrix regions. The waves display Kernel Smooth distributions of proportions for each region. (d) Distributions of FI and HIX across the time periods. The red line and grey area indicate linear regression and 95% confidence band, respectively. Dots correspond to all data of fluorescent parameters. The waves show Kernel Smooth distributions of the data.

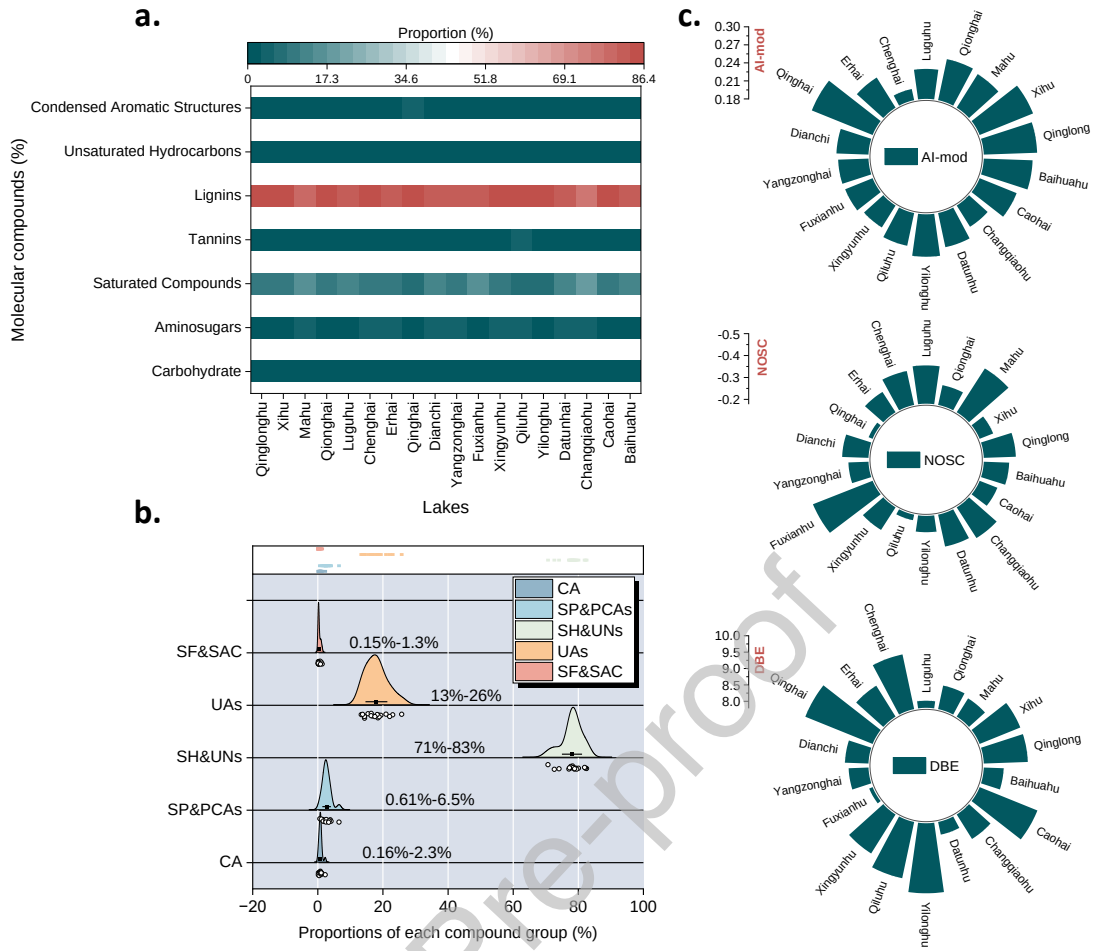


Fig. 3. Molecular information of DOM across the subtropical lakes. (a) Proportions of DOM molecular compounds in each lake, (b) Proportions of compound groups include combustion-derived polycyclic aromatic structures (CA), soil-derived polyphenols and PCAs with aliphatic chains (SP&PCAs), soil-derived humic and highly unsaturated compounds (SH&UNs), unsaturated aliphatic compounds (UAs), and saturated fatty and sulfonic acids, carbohydrates (SF&SAC). (c) The modified aromaticity index (AI-mod), double bond equivalent (DBE) and nominal oxidation state of carbon (NOSC) across the lakes.

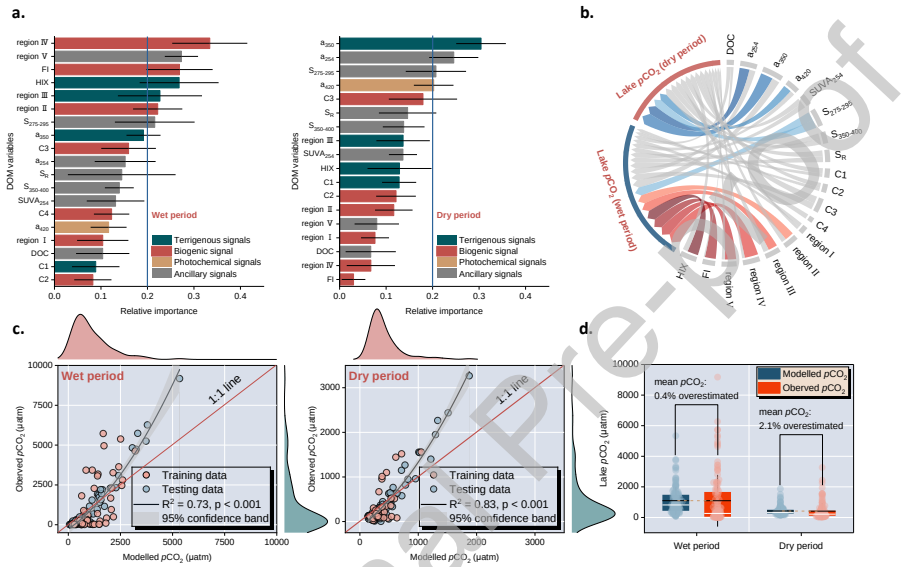


Fig. 4. Lake DOM signaling CO_2 drivers and magnitudes using machine learning. (a) Random forest modeling assigning the relative importance of DOM variable in evaluating lake CO_2. The x axis with whisker shows the mean importance with standard deviation (s.d.). (b) The visual flows of relative importance from DOM variable to temporal pCO_2 in the lakes. The red and blue flows represent the most variables with weights > 0.2. (c) The relationships between observed and modelled pCO_2 across the time periods. The waves show Kernel Smooth distributions

of the data. (d) Comparison of observed and modelled pCO_2 in the lakes. The boxes with whiskers represent 25th–75th percentiles with s.d. White and black lines show median and mean, respectively. Dots correspond to all data of observed and modelled pCO_2.

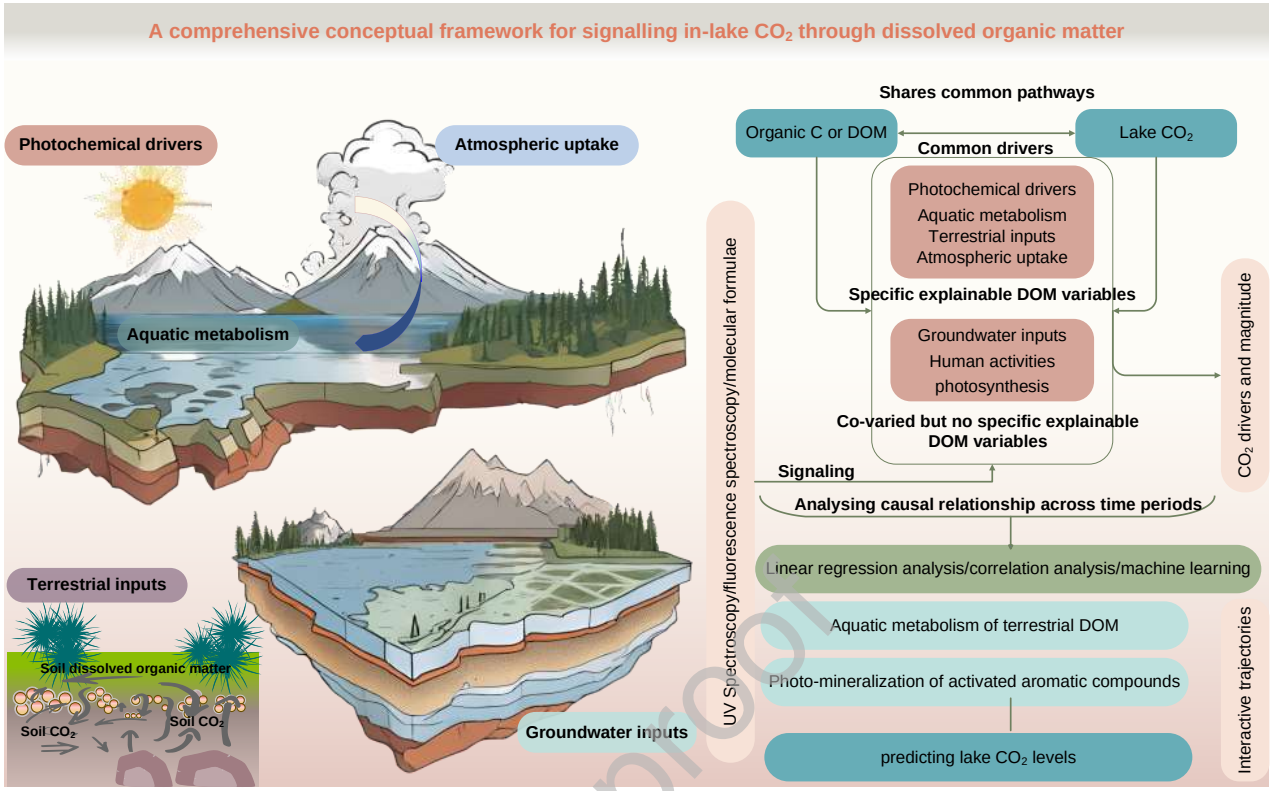
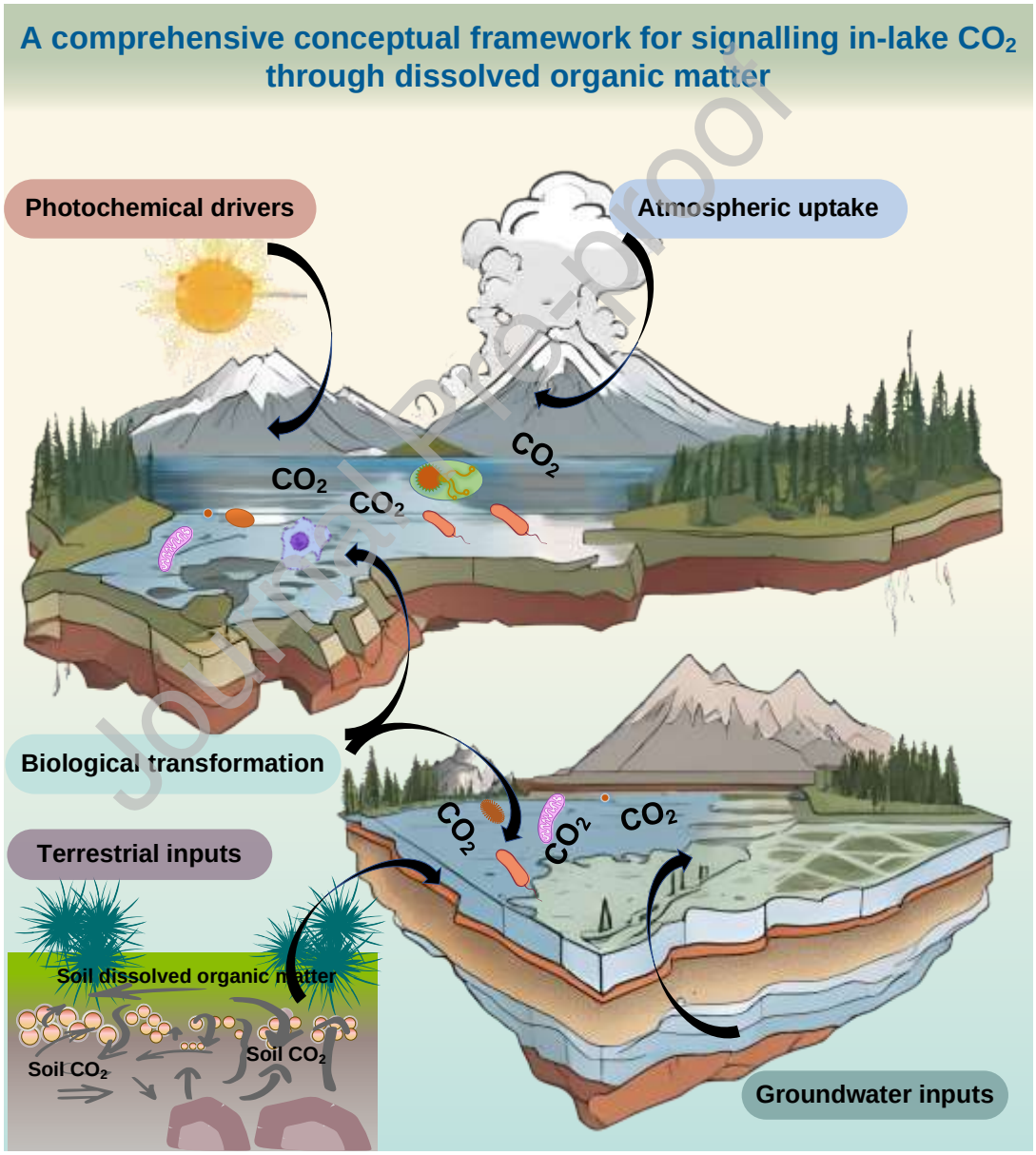


Fig. 5. A comprehensive conceptual framework for signaling in-lake CO_2 through DOM. We suggest that lake organic C or DOM follows the shared biogeochemical pathways with lake CO_2. It includes the common drivers 1) e.g., photochemical drivers, aquatic metabolism, terrestrial inputs and atmospheric uptake that are directly signalled by explainable DOM variables; and 2) e.g., groundwater inputs, human activities and photosynthesis that co-vary with lake CO_2 although explainable variables signaling these relationships are lacking. The term "explainable DOM variables" refers to DOM variables that can specifically indicate certain sources or pathways regarding biogenic, terrigenous, and photochemical processes. By contrast, ancillary signals may interact with various pathways but are not specifically indicative of any single explainable process. The causal relationships can be visualized through linear regression analysis, correlation analysis or machine learning. This helps to identify CO_2 drivers, rank contributions to lake CO_2, and particularly discover

interactive trajectories in modulating lake CO_2 (e.g., aquatic metabolism of terrestrial DOM, and photo-mineralization of activated aromatic compounds). By compiling and analysing these common pathways, it allows us to predict lake CO_2 levels using machine learning. We highlight that this conceptual framework could be generic and transferable for other natural waters.

Graphical Abstract



Declaration of Interests

All authors agree this submission and the authors declare that there is no conflict of interests regarding the publication of this article