Reviewer 1

Review of the paper "Multidecadal trends in CO2 evasion and aquatic metabolism in a large temperate river" by An Truong Nguyen et al.

This paper is focused on a timely and relevant question, which is to better under how fluvial ecosystems regulate the global C cycle. The data set, with more than three decades of data, is unique not only because of its length but also because there are very few high temporal resolution data of this quality in large rivers. I sincerely congratulate the authors for their vision and perseverance to put together this impressive data set. Moreover, the paper reports interesting results illustrating that rivers can act either as sources or sinks of carbon, and that this pattern can change seasonally but also at large time scales depending on the nutrient status of the ecosystem. This finding has important implications for understanding how fluvial networks work and their contribution to global C fluxes under present and future anthropogenic pressures. Overall, I think this research will be of interest to the audience of Biogeoscience, though the paper requires major changes to improve clarity and streamline data analysis and the interpretation of the results before publication. Below, I provide some general and specific comments and suggestions, which I hope will be of help to the authors when crafting the revised version of the paper.

Response: We would like to thank the reviewer for these extensive comments and suggestions. We have addressed each of the general and specific comments below in this blue color font. Please note that the Line numbers indicated in this letter is based on the track changes pdf version.

General comments

Long-term trends in groundwater CO2 inputs. One of my main concern is about long-term changes in groundwater CO2 inputs. As mentioned by the authors, it seems that the observed long-term decrease in FCO2 is mostly associated with a decrease of about 50% in groundwater inputs between the phytoplankton dominated and the macrophytes dominated periods. How reasonable this is? At the very end of the discussion, the authors suggest that there has been a generalized decrease in groundwater CO2 fluxes in the Loire catchment. Yet, it is not clear whether the long-term trend in discharge data support this explanation. How reasonable is to think that CO2 concentrations in groundwater have change if there have not been large changes in groundwater levels, neither in weathering rates. Overall, this flux is highly uncertain, and difficult to constrain with independent data.

Response: We agree that these external fluxes are uncertain and difficult to constrain with current data. First, we want to clarify that the observed decrease in our calculated external CO2 inputs is substantial. The mean annual external CO2 input decreased from 1008 ± 551 in the 1990-2000 period to 472 ± 129 gC m^{-2} yr^{-1} in the 2011-2021 period (Table 1), a reduction that exceeds the inter-annual variability and propagated uncertainty of our estimates. Therefore, we interpret this as a significant long-term shift that requires explanation.

To address question of "How reasonable is this?", we frame the argument as follows: a decrease in external CO2 flux must be driven by either (1a) a decrease in groundwater discharge (Qgw), (1b) a decrease in groundwater CO2 concentration (Cgw), or (1c) a combination of both.

Regarding groundwater discharge **(Qgw):** While our data show a modest long-term decline in river discharge (~13% over 32 years, Figure S8), we now present new evidence from a representative local borehole at Montifault (20 km from our site). After removing pumping effects with the EROS model (Thiéry, 2018), the data shows a clear decreasing trend in the piezometric level of the nappe since 2003 (New Figure S_groundwater_level). This evidence for decreasing groundwater levels, also noted at a regional scale (Binet et al., 2022; Baulon et al., 2022), supports a reduction in groundwater discharge (Qgw) contributing to the river. However, the modest scale of these hydrological changes suggests they are insufficient to be the sole driver of the >50% decrease in the calculated external CO_2 flux.

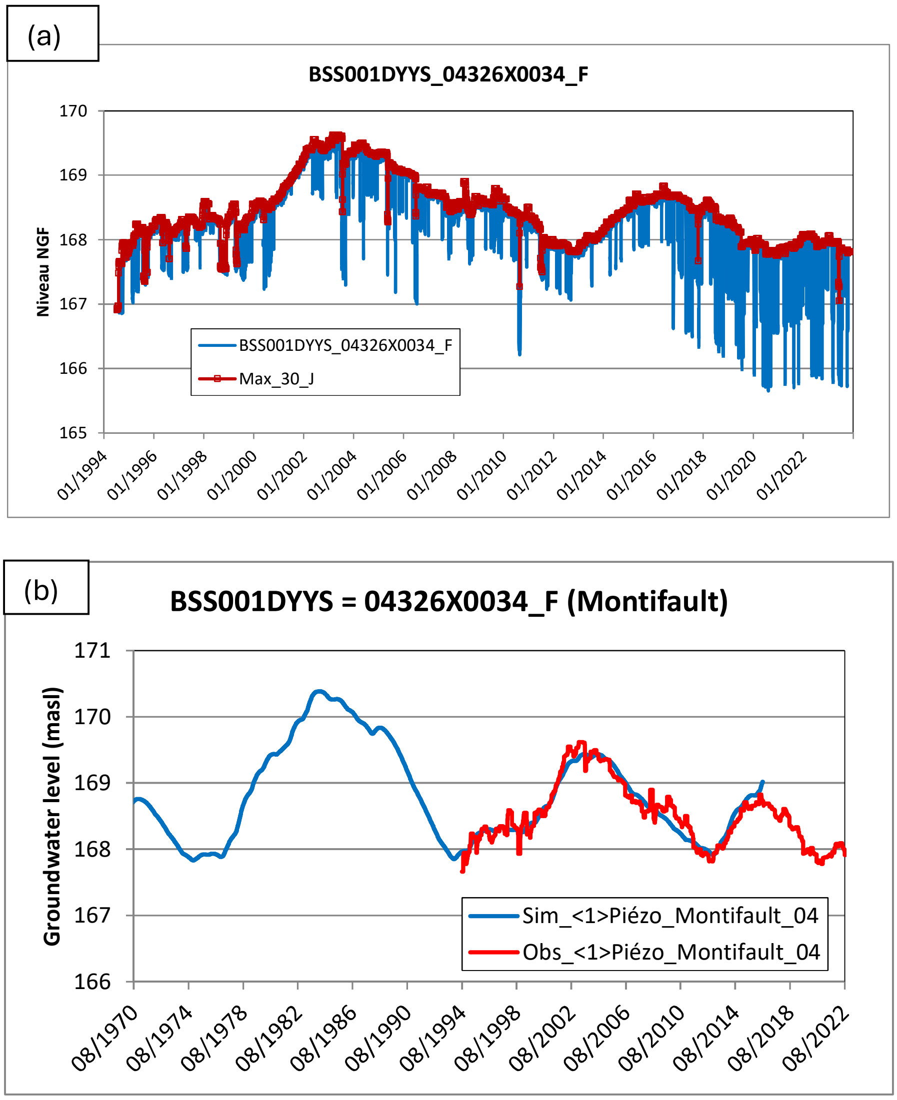


Figure S_groundwater_level: Decreasing trend in the groundwater level in Montifault (20 km from our site). (a) raw data, (b) after removing pumping effects with the EROS model (Thiéry, 2018)

Reference: Thiéry, D.: Logiciel ÉROS version 7.1. Guide d'utilisation, in: Rapport BRGM/RP-67704-FR, 175 pp., http://infoterre.brgm. fr/rapports/RP-67704-FR.pdf (last acccess: 12 October 2022), 2018.

Regarding groundwater CO2 concentration (Cgw): The discrepancy points towards a significant decrease in groundwater CO2 concentration as a key driver. To investigate this, we analyzed 30-year records of both pH and Total Alkalinity (TA) from several local groundwater monitoring stations near Dampierre (New Figure S_groundwater_quality). This analysis reveals a long-term increasing trend in groundwater pH of ~0.1-0.2 units while groundwater TA remained relatively stable, mirroring trends in the surface water (Figure 2a). At stable alkalinity, a pH increase of 0.1 units corresponds to a ~20-25% decrease in pCO2, lending more direct support to our hypothesis.

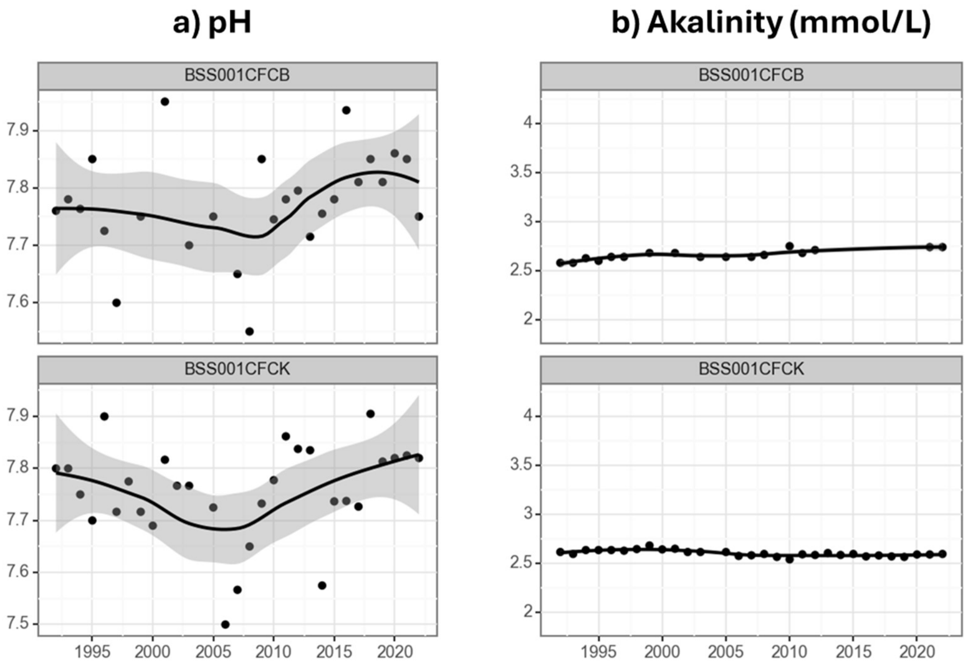


Figure S_groundwater_quality. pH and alkalinity in groundwater monitoring stations in the vicinity (5 km radius) of the Dampierre study site (1990-2021). Points represent individual measurements, and solid lines are LOESS smoothers with 95% confidence intervals (shaded areas). Source: https://hubeau.eaufrance.fr/page/api-qualite-nappes

We updated the discussion L555 – L579: "Further evidence from a representative local borehole at Montifault (20 km from our site) shows a clearly decreasing trend in the piezometric level since 2003 (Appendix D, Figure D3)[….] This analysis revealed a long-term increasing trend in groundwater pH, particularly after 2008, while groundwater TA remained relatively stable (Figure D4) […] Together, the evidence for both reduced groundwater discharge and lower groundwater pCO2 provides a robust explanation for the observed multi-decadal decline in external CO2 sources to the Loire River."

Re-oligotrophication. This phenomenon becomes crucial for understanding the temporal patterns in stream metabolic activity and CO2 sources, yet the magnitude of change of nutrients and DOM in the study river over time is barely mentioned. Even if this shift in water chemistry has been explained in a previous paper, some more quantitative information will help to better framed the discussion and interpretation of the results of this paper.

Response: We will add specific details on nutrient and chlorophyll-a changes, referencing previous work on the Loire River. We will also recall these changes in the Discussion when interpreting shifts in metabolic activity and CO2 sources. We will update the main manuscript

L77-86: "The Loire River (France) was one of the most eutrophic rivers in Europe at that time with total phosphorus (TP) concentrations frequently exceeded 0.2 mg P L-1 and chlorophyll-a concentrations often surpassed 100 µg L-1, with summer peaks reaching over 200 µg L-1 (Minaudo et al., 2015; Moatar & Meybeck, 2005). Despite potential autotrophic activity, the CO2 dynamics during these periods remains poorly documented due to the lack of comprehensive CO2 data, leaving a gap in our understanding of whether the river predominantly acted as a CO2 source or sink.

Following efforts to reduce nutrient inputs between the early 1990s and the mid-2000s, TP concentrations declined by approximately 50-70%, and mean summer chlorophyll-a concentrations decreased to <30 µg L-1 (Minaudo et al., 2015)."

Terminology. The authors use many different concepts to describe whether their system is dominated by macrophytes or phytoplankton, whether it is in an oligotrophic or eutrophic state, and finally classify the system behavior in four trophflux states as a function of CO2 fluxes and metabolic activity, which is the cornerstone of the results. For instance, "macrophyte-dominated" and "oligotrophic" regimes as well as "phytoplankton-dominated" and "eutrophic" regimes are used at the beginning. Also, the authors refer to "regime", "states", or "periods" non-consistently when referring to either "trophic conditions" or to metabolic activity. Overall, my suggestion is to simplify a bit this terminology and make sure to refer always in the same terms to the same concepts. For instance, only use either macrophyte- vs phytoplankton-dominate OR oligotrophic- vs eutrophic- regimes, and be consistent referring to either "states", "regimes", or "periods".

Response: We agree with consistent terminology to use only macrophyte- vs phytoplankton, so remove eutrophic/ oligotrophic

L21: " a shift from a phytoplankton-dominated regime to a macrophyte-dominated regime in ca. 2005"

Metabolic stoichiometry to convert O2 to CO2 moles. More details on these calculations are needed. An important aspect is whether conversions were similar for the phytoplankton- and the macrophyte-dominated periods, and to discuss the uncertainty associated with these calculations.

Response: In our metabolism calculations, gross primary production (GPP) and ecosystem respiration (ER) were initially estimated in O2 units by the streamMetabolizer model. We converted these to carbon units (mmol C m2 d, as reported in the manuscript) using a molar O2:C ratio of 1:1. For simplicity, we did not vary the photosynthetic or respiratory quotient between the phytoplankton-dominated and macrophyte-dominated regime.

This approach is supported by recent study of Diamond et al. (2025; some of us are co-authors). Indeed, a detailed analysis by Diamond et al. (2025) on this dataset found a significant difference in the Ecosystem Quotient (EQ: apparent mol O2 produced per mol DIC consumed), with a median EQ of ~1.3 during the phytoplankton-dominated period and ~1.0 under macrophyte dominance. They noted this ~30% change in EQ largely explains the observed change in O2-based GPP between the two regimes, implying that GPP in carbon units was likely more stable over time than our 1:1 conversion would suggest. However, this measured difference in EQ is primarily relevant during autotrophic periods. The vast majority of the internal CO_2 production in our system occurs during heterotrophic periods (i.e., winter), where the Respiratory Quotient (RQ) is the key stoichiometric parameter. There is not sufficient evidence to suggest that RQ varied systematically over time during these dominant heterotrophic periods. Therefore, while using a 1:1 ratio likely underestimates the magnitude of carbon fixation (autotrophic sink) during the phytoplankton-dominated summers, we posit it remains a reasonable first-order approximation for the heterotrophic periods that dominate the internal CO_2 source budget. Based on this evidence, we consider the fixed ratio a conservative and justified assumption for estimating multi-decadal trends.

L198-205: "GPP and ER were then converted to carbon units (g C m^{-2} d^{-1}) using a fixed molar O_2:C ratio of 1:1. This assumption is widely used in river metabolism studies and reflects the stoichiometry of aerobic metabolism (Trentman et al., 2023). Although photosynthetic and respiratory quotients (PQ and RQ) can vary with autotrophic community composition, recent long-term analysis of the Loire River by Diamond et al. (2025) showed that such variability does not lead to cumulative bias in net ecosystem production or CO_2 budgets when integrated over decadal timescales. Therefore, we adopt this approach as a reasonable and conservative approximation for estimating long-term carbon dynamics, while acknowledging it as a source of short-term uncertainty.

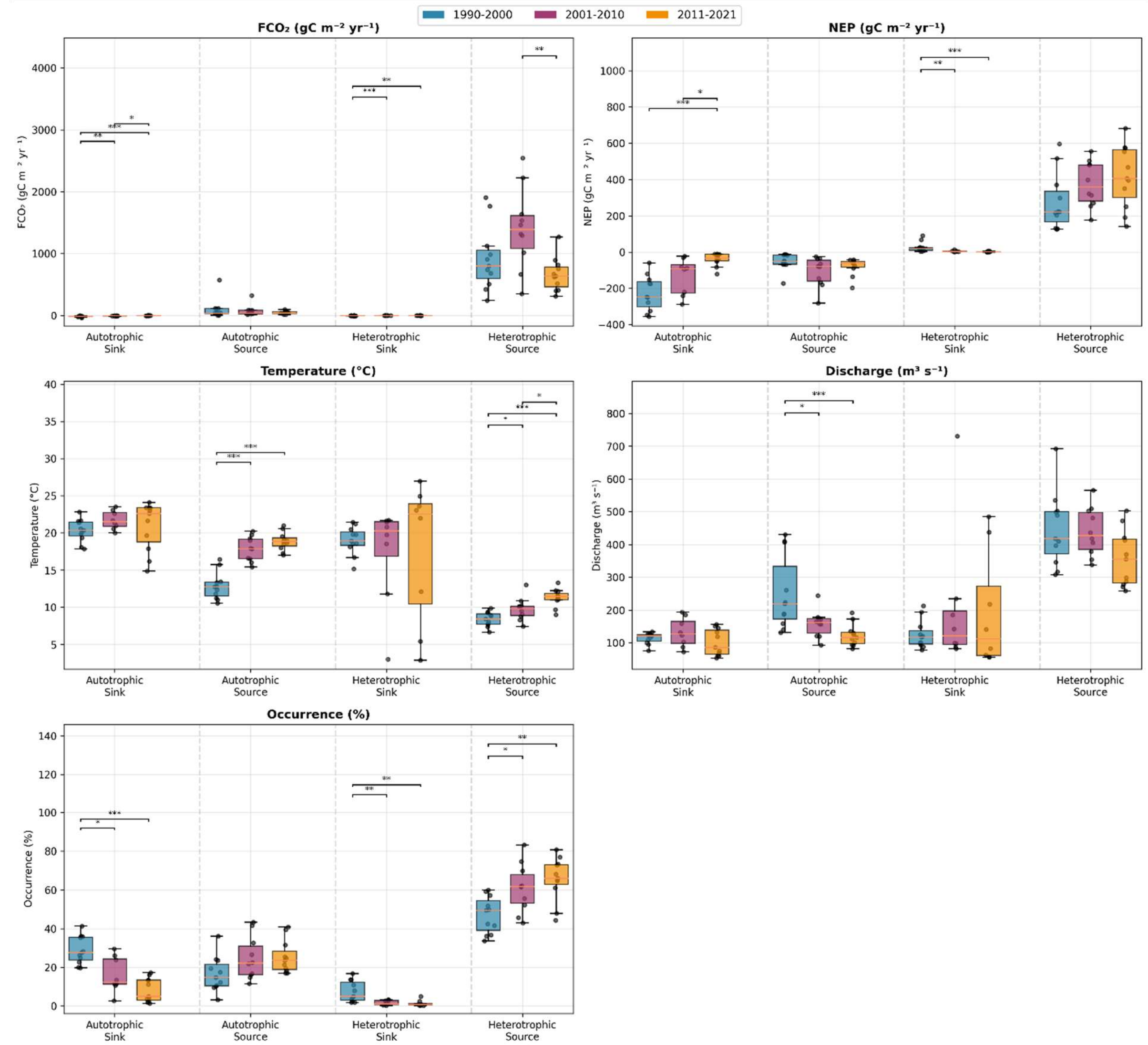
Reference: Diamond, J. S., Nguyen, A. T., Abril, G., Bertuzzo, E., Chanudet, V., Lamouroux, R., & Moatar, F. (2025). Inorganic carbon dynamics and their relation to autotrophic community regime shift over three decades in a large, alkaline river. Limnology and Oceanography.

Comment: Change point analysis and statistical analysis. Is this analysis important enough to keep it in the main manuscript? At the end of the day, the authors are splitting the data set per decades. While I agree that the changepoint analysis somehow supports to split the data like by decades, I wonder whether it might be enough to add this analysis in the supplementary materials. On the other hand, the results would be better supported if the authors use statistical tests to explore whether differences among periods (and/or states) for the different variables are statistically significant. This would help to more clearly distinguish the most remarkable changes, and avoid qualitative statements.

Response: We believe the change point analysis is valuable for the main manuscript because it provides an objective basis for dividing the 32-year dataset into periods that reflect statistically identified shifts in key variables (FCO_2, NEP, Alkalinity, pH, etc.), rather than relying on arbitrary decadal splits or solely on previously published ecological shift years which might not perfectly align with the specific biogeochemical fluxes we are analyzing.

Regarding statistical testing, we agree with the reviewer and have now implemented statistical comparisons among periods within each trophic state. However, note that our manuscript also presents long-term trend analyses (Theil-Sen slopes) for annual values of key variables within each trophlux state over the entire 32-year duration (Figure 3). This addresses the gradual changes over time, not only for 3 decades.

We conducted non-parametric Kruskal-Wallis tests to assess overall differences between decades within each trophic state for key variables (FCO_2, NEP, discharge, temperature, and occurrence frequency). For cases where significant differences were detected ($p<0.05$), we performed post-hoc Mann-Whitney U tests with Bonferroni correction to identify which specific decade pairs differed significantly.



L812: Figure C2. Statistical comparison of key variables across decades (1990-2000, 2001-2010, 2011-2021) within each trophic state. Boxplots show median (horizontal line), interquartile range (box), and distribution of annual values (black dots). Colored boxes represent different decades: blue (1990-2000), purple (2001-2010), and orange (2011-2021). Statistical significance of differences between decades was assessed using Kruskal-Wallis tests, with significant pairwise differences (Mann-Whitney U test) indicated by horizontal bars with asterisks (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

Internal vs external sources of CO_2. While I understand the point of the authors, this is an ***oversimplification of CO2 sources***. For instance, by referring to "external CO_2 sources" the authors imply there is no other internal sources than aerobic metabolism producing CO_2 in the study system. How reasonable is to assume that there is **no anaerobic metabolism**? The authors should include some rational about this assumption, or else refer to "Other sources" rather than to "External sources". Regarding "internal sources", I wonder ***whether diel signals of dissolved oxygen fully capture the metabolism associated with photoautotrophs***. In Table 1, the authors report negative values for "external CO_2 inputs" which seems unrealistic. A potential explanation could be a systematic underestimation of the photoautotrophic activity by either phytoplankton or macrophytes, which might be more evident during this state, though may be happening also during other states. On the other hand, how feasible is that "external inputs" vary so much among states within a given decade? The authors should better discuss and, if possible, constrain, this factor to the best of their knowledge.

Response: We acknowledge that our terminology and framework are simplified but it is typical in many previous studies. In our analysis, "internal CO_2 source" specifically refers to CO_2 produced by net ecosystem respiration. We want to state clearly that we do not assume there is no anaerobic metabolism. The diel oxygen method for calculating ecosystem respiration (ER) implicitly accounts for most anaerobic metabolism. This is because the reduced secondary metabolites produced during anaerobic processes (e.g., NH_4^+, Mn^{2+}, Fe^{2+}, sulfides) are typically re-oxidized in other parts of the ecosystem, a process that consumes oxygen and is therefore captured within the integrated ER term. The only significant anaerobic pathways not accounted for by this oxygen consumption are denitrification (which produces N_2 gas) and methanogenesis. In a large, generally well-oxygenated river like the Loire, these pathways are expected to be a minor component of the overall carbon and oxygen budget compared to aerobic respiration. We will add a concise explanation of this to the Methods section to clarify that our ER term represents a robust measure of total ecosystem oxygen demand.

Regarding whether diel O_2 signals fully capture the metabolism associated with all photoautotrophs, our open-water diel O_2 method (*streamMetabolizer*) quantifies net changes in DO in the water column. This integrates the metabolic activity of phytoplankton suspended within the water and the portion of benthic photoautotroph (e.g., submerged macrophytes, benthic algae) metabolism that results in O_2 exchange with the overlying water. We agree there is a potential issue if macrophytes or benthic algal create very localized O_2 supersaturation that do not fully mix with the water on a diel timescale, then our method might slightly underestimate GPP. However, given the size and flow of the Loire as a large temperate river, we expect the water to be sufficiently mixed that most macrophyte oxygen production is recorded.

Regarding negative external CO_2 in Table 1, particularly for the Heterotrophic-Sink state (which we note is rare, occurring 1-7% of days annually): Our manuscript already provides a physical explanation for this phenomenon. We state, 'The heterotrophic-sink state implies that despite the net conversion of biomass into water column CO_2, there is still a CO_2 undersaturation relative to the atmosphere, likely due to prior autotrophic uptake. We expect the heterotrophic-sink state to be a temporary occurrence, reflecting temporal lags...'. This reflects the river acting as a strong net sink for CO_2 from all sources combined due to this biologically-driven undersaturation, not an active consumption by external sources like groundwater.

We add the Appendix section discuss Heterotrophic-Sink state:

L786-787: The Heterotrophic-Sink state (NEP < 0, FCO_2 < 0) represents a condition where the river is a net CO_2 sink from the atmosphere despite ongoing net ecosystem respiration. As discussed in the main text, we primarily attribute this transient state to significant CO_2 undersaturation in the water column resulting from intense prior autotrophic uptake. A recent analysis by Diamond et al. (2025) using this same dataset provides strong evidence for this mechanism. They found that heterotrophic-sink events were temporary, lasting an average of 4.7 ± 4.1 days, and that 76% of these events were immediately preceded by an autotrophic-sink state. This confirms our interpretation that the heterotrophic-sink state is not a stable condition but rather a short-lived transitionary phase as the river shifts from being an autotrophic-sink to either an autotrophic-source or a heterotrophic-source state. An additional potential contributing factor is the possible underestimation of GPP by the diel oxygen method. If oxygen produced within dense benthic macrophyte beds does not fully mix into the water column on a diel timescale, the reach-integrated GPP would be underestimated. This would result in a calculated NEP that is more negative than the true value, which could contribute to the Heterotrophic-Sink state.

We agree with the reviewer's suggestion that a systematic underestimation of photoautotrophic activity (GPP) could also contribute to these calculated negative values is a pertinent consideration. If true GPP were indeed higher, particularly in periods preceding or during the Heterotrophic-Sink state, the calculated NEP would be more positive (or less negative). This would, in turn, make the calculated 'External CO_2' less negative or potentially positive. While our diel O_2 method captures bulk water metabolism, and we expect considerable mixing in the Loire, some underestimation of GPP from dense macrophyte beds (due to direct O_2 loss or localized consumption without full mixing) is a possibility.

Concerning the variability of calculated 'external inputs' among states within a decade: These are mean annual values for days falling into specific trophlux states. Different trophlux states are demonstrably associated with different mean hydroclimatic conditions (e.g., temperature, discharge, as shown in Table 1).

Sources of uncertainty. The authors need to better consider in their calculations the different sources of uncertainty. The supplementary materials tackle some of these sources of uncertainty, but some of this rationale needs to be moved to the main text, and other additional sources of uncertainty such as those associated with respiration and photosynthetic coefficients, anaerobic respiration, k600 in large streams (note that Raymond equations are useful for small streams with complete water column mixing, which is not the case of large rivers), and GPP not captured by DO signals in the water column (which may happen for macrophytes) should also be considered.

Response: Several sources of uncertainty mentioned have been addressed in detail in our responses to other comments and will be incorporated into the revised manuscript:

O_2:C stoichiometry: we clarify our 1:1 O_2:C assumption in the Methods

L189-196: GPP and ER were then converted to carbon units (g C m-2 d-1) using a fixed molar O_2:C ratio of 1:1. This assumption is widely used in river metabolism studies and reflects the stoichiometry of aerobic metabolism (Trentman et al., 2023). Although photosynthetic and respiratory quotients (PQ and RQ) can vary with autotrophic community composition, recent long-term analysis of the Loire River by Diamond et al. (2025) showed that such variability does not lead to cumulative bias in net ecosystem production or CO_2 budgets when integrated over decadal timescales. Therefore, we adopt this approach as a reasonable and conservative approximation for estimating long-term carbon dynamics, while acknowledging it as a source of short-term uncertainty

k600 in large streams: we expand the Methods section to discuss the comparison of our streamMetabolizer-derived k600 with Raymond et al. (2012) equations, acknowledge the challenges of k600 estimation in large, explain why discrepancies between methods.

L723 – 747: B2. Validation of the Gas exchange coefficient (k600)

The k600 values estimated by the StreamMetabolizer model were compared with seven k600 calculated from seven fitted equations proposed by Raymond et al., (2012b) for streams and small rivers. Both k600 estimates exhibited similar seasonal fluctuations, with the lowest values occurring in summer and the highest in winter. The comparison revealed that the mean absolute percentage error (MAPE) between the StreamMetabolizer estimates and the mean k600 from the seven fitted equations ranged from 36% to 62%. Specifically, the Raymond et al., (2012) k600 estimates tended to be higher in summer and lower in winter compared to those estimated by the StreamMetabolizer model. Such discrepancies can arise because streamMetabolizer co-estimates K600 with GPP and ER by fitting observed DO dynamics, making its estimate sensitive to the strength of the biological signal, whereas empirical equations rely solely on hydraulic proxies for turbulence. However, the k600 values derived from StreamMetabolizer fall within the same order of magnitude as those from the seven fitted equations (Figure B3). To maintain internal consistency between the metabolic and FCO2 calculations, the k600 estimates from streamMetabolizer were used for all subsequent flux calculations.

Figure B3. Comparison of the gas exchange coefficient (k600) estimated by streamMetabolizer (black line) with the mean and range (blue shaded area) of values derived from seven empirical equations from Raymond et al. (2012b).

Figure B4. Compare k600 between the mean of seven equations Raymond et al. 2012 and StreamMetabolizer for 1990-2021. Colors indicate discharge quantiles (Q1-Q4, legend above). R2 and RMSE shown per discharge quantile range.

GPP not captured by DO signals (macrophytes): we add a discussion on the potential for GPP underestimation in dense macrophyte beds due to mechanisms like direct O2 loss via ebullition or localized O2 dynamics not fully mixing with the bulk water.

L775 – 779: An additional potential contributing factor is the possible underestimation of GPP by the diel oxygen method. If oxygen produced within dense benthic macrophyte beds does not fully mix into the water column on a diel timescale, the reach-integrated GPP would be underestimated. This would result in a calculated NEP that is more negative than the true value, which could contribute to the Heterotrophic-Sink state.

Contribution of internal sources to total CO2. Overall, I wonder whether it makes sense to report -NEP/FCO2 in all cases since the implications of the mass balance are quite different depending on whether the stream is acting as a source or a sink of CO2. In particular: (1) No doubt about what -NEP/FCO2 implies for the heterotrophic-CO2 source state; (2) For the autotrophic-CO2 source state, the contribution of the stream to FCO2 is actually 0%, and photoautrotrophs could contribute to reduce “external CO2 inputs” by xx % (i.e. -NEP/external CO2 rather than -NEP/FCO2); (3) For the heterotrophic-CO2 sink state, it has no sense to me that groundwater is not contributing CO2, unless the stream is losing water, and in this case, there might be either an unaccounted pool fixing CO2 from the water column, or GPP is systematically underestimated for whatever reason; (4) For the autotrophic-CO2 sink state, the internal source could contribute to balance out 100% of the “external CO2 sources”, and contribute to fix additional CO2 from the atmosphere (i.e. not sure whether -NEP/FCO2 is really meaningful in this case). From a mass balance perspective, Figure 5 (in the discussion) makes much more sense than Table 1, and my feeling is that the manuscript would be more easy to follow if the results focus on the mass balances.

Response: Actually, the footnotes * and ** in Table already try to explain this, but it may still remain complex for readers. As also discussed in our response to your General Comment regarding the interpretation of 'External CO2 ' and negative values, we will revise the presentation in Table 1 and suggest reader follow the mass balance in Figure 5 for the negative ratio cases.

L369-370:

Variable	Period	CO_2 source	CO_2 sink	All states

		Heterotrophic	Autotrophic	Heterotrophic	Autotrophic	
Occurrence (% of days)	1990-2000	47.3 ± 9.4	16.7 ± 9.2	7.3 ± 5.4	28.7 ± 7.0	100
	2001-2010	61.2 ± 12.7	25.3 ± 11.2	1.7 ± 1.1	15.5 ± 9.0	100
	2011-2021	65.8 ± 11.3	26.2 ± 8.6	1.1 ± 1.5	7.3 ± 5.7	100
FCO_2 budget (% of annual flux by each state)	1990-2000	94.6 ± 13.8	9.2 ± 11.4	-0.8 ± 1.4	-3.0 ± 4.2	100
	2001-2010	92.4 ± 11.3	8.2 ± 11.2	-0.1 ± 0.0	-0.6 ± 0.5	100
	2011-2021	91.8 ± 5.6	8.7 ± 5.8	-0.1 ± 0.2	-0.4 ± 0.3	100
–NEP ($gC\ m^{-2}y^{-1}$)	1990-2000	277 ± 158	-54 ± 46	25.5 ± 28.0	-225 ± 97	23 ± 222
	2001-2010	376 ± 127	-111 ± 83	3.8 ± 3.0	-131 ± 100	162 ± 234
	2011-2021	417 ± 173	-82 ± 48	2.0 ± 2.4	-36 ± 35	300 ± 232
FCO_2 ($gC\ m^{-2}y^{-1}$)	1990-2000	954 ± 514	102 ± 148	-4.4 ± 4.5	-21 ± 12	1031 ± 531
	2001-2010	1453 ± 666	88 ± 104	-0.6 ± 0.8	-7.8 ± 5.4	1534 ± 620
	2011-2021	717 ± 274	59 ± 28	-0.9 ± 1.8	-2.6 ± 2.1	773 ± 272
External CO_2 ($gC\ m^{-2}y^{-1}$)	1990-2000	677 ± 477	157 ± 92	-29 ± 32	204 ± 93	1008 ± 551
	2001-2010	1077 ± 595	199 ± 172	-4 ± 1	123 ± 95	1372 ± 528
	2011-2021	299 ± 140	140 ± 55	-2 ± 3	34 ± 31	472 ± 129
–NEP/FCO_2 (%)	1990-2000	37 ± 27	NA*	NA*	NA**	5 ± 29
	2001-2010	28 ± 9	NA*	NA*	NA**	7 ± 15
	2011-2021	57 ± 10	NA*	NA*	NA**	34 ± 27
Temperature (°C)	1990-2000	8 ± 4	12 ± 4	18 ± 3	20 ± 3	13 ± 6
	2001-2010	10 ± 5	18 ± 5	18 ± 7	21 ± 3	13 ± 7
	2011-2021	11 ± 5	18 ± 5	10 ± 9	21 ± 3	14 ± 6
Discharge ($m^3\ s^{-1}$)	1990-2000	454 ± 326	276 ± 225	115 ± 60	111 ± 56	301 ± 290
	2001-2010	436 ± 323	160 ± 110	137 ± 69	115 ± 59	323 ± 296
	2011-2021	362 ± 278	120 ± 87	205 ± 136	80 ± 63	277 ± 259

* Not applicable. In these cases, the –NEP/FCO_2 ratio does not represent a simple contribution of internal respiration to CO_2 evasion and is therefore not reported. The mass balance for these states is presented in Figure 5.

** Not Applicable. Internal metabolism consumes both external CO_2 and the CO_2 supplied from the atmosphere through the gas exchange at the air-water interface.

The external CO_2 ($gC\ m^{-2}y^{-1}$) = FCO_2 + NEP

Regarding the Heterotrophic-Sink state: We maintain that this state is physically plausible and does not imply an absence of groundwater CO2 contribution. This transient state is explained by the carbonate system's buffer capacity. A significant percentage of Heterotrophic-Sink occurrences directly follow Autotrophic-Sink periods (Diamond et al., 2025). The intense biological CO2 drawdown during the preceding autotrophic phase creates a large CO2 deficit. Due to the buffering delay, the river can remain a net sink for atmospheric CO2 even as it briefly switches to net heterotrophy, before net respiration is sufficient to overcome the deficit and turn the river into a CO2 source.

L783-796: B4. Methodological considerations for the Heterotrophic-Sink State

Specific comments

Comments from reviewer 1	Responses from Nguyen et al.,

Abstract
 	
L26-27. For (ii), not clear from this sentence whether the contribution of aquatic metabolism was higher or lower during the eutrophic or the oligotrophic trophic regime.
 	
For (iii), better highlight the predominant role of the river as a source of CO_2.
 	
For (iv), might be more informative to highlight the seasonality of FCO_2, and how this was modulated by the trophic regime rather than referring to the hysteresis patterns.
	L23-44: We demonstrate that: i) annual FCO_2 varied an order of magnitude among years (range = 200–2600 g C m2 yr-1) with a long-term decrease trend, mainly linked to decreased groundwater contribution; ii) the mean annual contribution of aquatic metabolism to total FCO_2 was 40%, increasing from 37 ± 27% in phytoplankton-dominated period to 57 ± 10% in macrophyte-dominated period; iii) while the river predominantly acted as a CO_2 source, it occasionally functioned as a CO_2 sink ($FCO_2 < 0$) during summer, though this sink behavior constituted a minor component (-0.6%) of the FCO_2 budget; and iv) FCO_2 exhibited strong seasonality linked to discharge, exhibiting hysteresis where FCO_2 levels at equivalent discharge were 1.5 to 2 times higher during the rising limb (autumn) compared to the falling limb (spring). The magnitude of this hysteresis diminished in the later macrophyte-dominated period, indicating a changing seasonal discharge control.

L31."dynamic within and across years" as a function of what? The amount of nutrients?	L45-48: This study makes clear that river FCO_2—and its source—is dynamic within and across years, driven by hydro-climatic variations and biological activity. Catchment-scale hydrogeological changes can be a more dominant driver of long-term riverine CO_2 evasion than in-stream ecological regime shifts, controlling the balance between internal and external CO_2 production.
Introduction L38. Better say "is assumed to come", since there are already several published studies challenging this assumption.	L53: Most CO_2 flux (FCO_2) is often assumed to come from "external" sources
L54-56. In which way these variables controlled by discharge (inputs of carbon and nutrients) would influence metabolic activity and the balance between GPP and ER?	L68-81: The seasonal hydrology plays an important role in determining the magnitude and timing of CO_2 emissions from rivers, as changes in flow rates affect the transport of nutrient, organic carbon from surrounding land as input for stream metabolism and the exchange of CO_2 between the river and the atmosphere (Cole et al., 2007; Hotchkiss et al., 2015)
L72-74. Clarify what do you mean by "positive NEP yields local organic matter increases". Do you mean in the form of algal biomass? Increases in particulate and dissolved organic matter? Explain why "this autotrophic state" is most	L97-105: In this "autotrophic" state (GPP > ER), positive NEP means a net production of organic matter within the reach, leading to increases in biomass (e.g., algal, macrophyte) and potentially contributing to particulate and dissolved

common in larger rivers, and clarify why “this is typically missed by FCO2 sampling campaigns”.	organic matter pools. Such autotrophic periods, sometimes leading to CO2 undersaturation, can be more prevalent or sustained in larger rivers due to factors like greater water residence times, increased light availability across wider channels, and potentially a greater buffering capacity against rapid changes in external CO2 inputs compared to smaller streams (Hotchkiss et al., 2015). However, these periods, especially if transient or occurring outside of typical low-flow summer conditions, can be missed FCO2 sampling campaigns that are often infrequent or biased towards specific seasons, thus underestimating their occurrence and impact.
L75. Indicate if this finding is general to all large rivers, or if it was observed in specific rivers.	L99-103: Such autotrophic periods, sometimes leading to CO2 undersaturation, can be more prevalent or sustained in larger rivers due to factors like greater water residence times, increased light availability across wider channels, and potentially a greater buffering capacity against rapid changes in external CO2 inputs compared to smaller streams (Hotchkiss et al., 2015). It is recommend checking Hotchkiss et al., 2015 for the full list of rivers from that study.
L81. For contextualization purposes, provide some information on how nutrient concentrations changed between the eutrophic and oligotrophic regimes and this phenomenon happen. It is also important to recall this in the discussion, to better interpret the large changes in both groundwater CO2 inputs and metabolic activity within the river.	L114-124: The Loire River (France) was one of the most eutrophic rivers in Europe at that time with total phosphorus (TP) concentrations frequently exceeding 0.2 mg P L-1 and chlorophyll-a concentrations often surpassing 100 µg L-1, with summer peaks reaching over 200 µg L-1 (Minaudo et al., 2015; Moatar & Meybeck, 2005). Despite potential autotrophic activity, the CO2 dynamics during these periods remain poorly documented due to the lack of comprehensive CO2 data, leaving a gap in our understanding of whether the river predominantly acted as a CO2 source or sink. Following efforts to reduce nutrient inputs between the early 1990s and the mid-2000s, TP concentrations declined by approximately 50-70%, and mean summer chlorophyll-a concentrations decreased to <30 µg L-1 (Minaudo et al., 2015).
L85. By the growing season do you mean spring and summer?	L125-127: during the spring–summer growing season (Diamond et al., 2022).
L83-87. These two sentences can be merged and shorten.	L125-127: The ecosystem transition was followed by a delayed shift in the river’s metabolic regime around 2012–2014, with GPP declining and NEP decreasing by roughly 10% during the spring–summer growing season

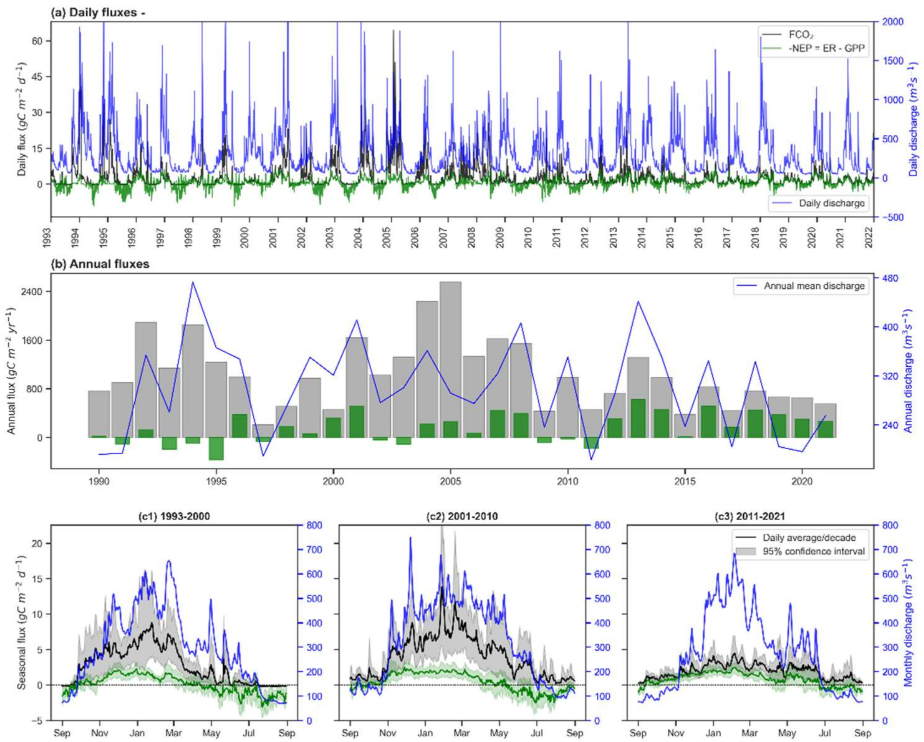
	(Diamond et al., 2022). decreasing by roughly 10% during the spring–summer growing season (Diamond et al., 2022)
L88. Please, could you provide other examples of re-oligotrophication in developed countries? This shift towards lower nutrient concentration in large rivers is not so evident giving the modest improvements in water chemistry observed in the last decades in Europe.	There are a few well-documented examples of nutrient load reductions leading to ecological shifts in major rivers within developed countries. Our manuscript cites Ibáñez et al. (2023) for a global review of re-oligotrophication and its effects, which further supports this broader context. 1) The Rhine River was with orthophosphate concentrations declining from >0.4 mg/L (early 1970s) to <0.1 mg/L (1998), accompanied by broader ecological rehabilitation efforts. 2) Recent work on the Upper Mississippi River (USA) documented substantial macrophyte community recovery (1998–2020), including increased species diversity and shifts from free-floating to submerged plants. 3) The River Thames (UK) exemplifies this trend, with phosphorus loads reduced by approximately 80% over 40 years due to improved wastewater treatment and agricultural management, though nutrient concentrations remain above ecological limiting levels and rising temperatures continue to promote algal blooms. Jarvie, H. P., Worrall, F., Burt, T. P., & Howden, N. J. K. (2025). A 150-year river water quality record shows reductions in phosphorus loads but not in algal growth potential. Communications Earth & Environment, 6(1), 62 Ibáñez, C., et al. (2022). Ecosystem-level effects of re-oligotrophication and N:P imbalances in rivers and estuaries on a global scale. Global Change Biology, 29(2), 261–282 Larson, D., Jones, M., Weigel, B., Gray, B., & Ovaskainen, O. (2024). River re-oligotrophication and hydrologic changes abruptly contributed to macrophyte community shifts and recovery (https://aslo.secure-platform.com/2024/gallery/rounds/16/details/11174)

L91-93. Note that the hypothesis is lacking the reasoning behind. Why did you expect an increase in the contribution of FCO2 from aquatic metabolism if, according to the earlier paragraph, you actually observed a decrease in NEP with reoligotrophication?	The reasoning was that reduced GPP would mean less CO2 uptake, potentially leading to higher pCO2 and thus higher FCO2. If ER remained similar or increased, then –NEP would increase, further boosting FCO2. The phrasing "internal source contribution would increase" refers to the ratio –NEP/FCO2. If –NEP increases (more net respiration) and FCO2 also increases or changes less drastically, the ratio can increase.
L94. Note that in the earlier paragraphs you refer to "regime" when talking about the trophic conditions, but to "states or periods" when referring to stream metabolic activity. To help the reader, better be consistent with the terminology throughout.	As answered in your earlier comment. We have systematically revised the paper to adopt a clear and consistent terminological framework. Our approach is as follows: - Regime: to describe the long-term, multi-year ecological condition of the river. So, we then refer to the 'phytoplankton-dominated regime'. - State: This term is used for short-term, daily conditions. So, 'metabolic state' (i.e., 'autotrophic state' or 'heterotrophic state') and the four specific 'trophlux states' which combine metabolic state with CO2 flux. - Period: This term is used neutrally to refer to the specific timeframes identified by our change-point analysis (e.g., 'the 1990–2000 period'). L88-89: The eutrophic **state** was common in large rivers throughout the 1980s and 1990s, characterized by high nutrient concentrations and high chlorophyll-a, leading to a net autotrophic **state**. L109-112: Specifically, we predicted that these increases would manifest coincidentally with the shifts in phytoplankton to macrophyte-dominated **regime** in 2005 and stream metabolism **regime** in 2012.
L96. How do you expect discharge to influence FCO2 in the first term? Would FCO2 increase or decrease with discharge? Why will Q influence FCO2 differently in the macrophyte dominated period across seasons? And why this seasonal influence will not emerge in the phytoplankton dominated period?	Generally, FCO2 is expected to increase with discharge due to increased turbulence (higher kCO2) and potentially increased delivery of CO2-rich terrestrial/groundwater (higher pCO2). We will include it in introduction: L134-163: Finally, discharge influences FCO2 through multiple mechanisms, including gas transfer velocity, delivery of external CO2, and inhibition of in-stream primary production. Phytoplankton GPP can be sensitive to discharge (e.g., washout, turbidity), while macrophyte GPP may be less directly flow-dependent (Diamond et al., 2022). We therefore predicted that the overall control

	of discharge (Q) on FCO_2 would change with the shift to macrophyte dominance. Specifically, we anticipated that the seasonal hysteresis patterns observed in the FCO_2-Q relationship (where FCO_2 differs between rising and falling limbs of the hydrograph at similar Q) would be altered, potentially becoming less pronounced or showing a different shape if macrophyte GPP imparts a more stable baseline of CO_2 uptake across varying flow conditions compared to phytoplankton.
L112. And during winters?	L178-181: During summer, low flows (<150 m^3 s^{-1}), the study site is typically shallow (around 1 m deep) and wide (330 m); during winter, with higher discharges (e.g., >500 m^3 s^{-1}), depths can increase significantly, typically ranging from 2 to 3 m.
L130. How did you transform stream metabolic rates from O_2 to C units? Which stoichiometric ratios did you use (i.e., photosynthetic and respiration coefficients). Did you consider whether these coefficients vary between the phytoplankton- and the macrophyte-dominated regimes?	L212-219: GPP and ER were then converted to carbon units (g C m^{-2} d^{-1}) using a fixed molar O_2:C ratio of 1:1. This assumption is widely used in river metabolism studies and reflects the stoichiometry of aerobic metabolism (Trentman et al., 2023). Although photosynthetic and respiratory quotients (PQ and RQ) can vary with autotrophic community composition, recent long-term analysis of the Loire River by Diamond et al., (2025) showed that such variability does not lead to cumulative bias in net ecosystem production or CO_2 budgets when integrated over decadal timescales. Therefore, we adopt this approach as a reasonable and conservative approximation for estimating long-term carbon dynamics, while acknowledging it as a source of short-term uncertainty.
L137. Only 12% of discards is a big success! Why did you choose to fill the gaps? Please, indicate in the main text whether main results and conclusions hold if not filling the gaps.	We chose to fill gaps to have a continuous daily time series for FCO_2 and NEP calculations, which is required for accumulative annual budgets and consistent change point/trend analysis. More detail in the appendix: B1. Handling of Metabolism Model Outputs L781 – 796: In general, this correction did not substantially alter the annual GPP or ER calculations. Replacing negative GPP with the 75th percentile increased annual GPP by an average of 1.3% (ranging from 0.1% to 5.3%), while setting negative GPP to zero resulted in a smaller increase, ranging from 0.04% to 3.4% (Figure B1). Similarly, the annual ER calculations across different treatments for unrealistic ER values show no significant differences, with an average flux

	variation of around 1%, except in 1995, where the difference reaches 15% (Figure B2). Figure B1. Comparison of annual GPP estimates based on different approach for handling negative GPP values: retaining negative GPP, setting negative GPP to zero, and replacing negative GPP with the 75th percentile of estimated GPP from the streamMetabolizer model. Figure B2. Comparison of annual ER estimates based on different approach for handling negative ER values: retaining negative ER, setting negative ER to zero, and replacing negative ER with the 75th percentile of estimated ER from the streamMetabolizer model.
L139. I think Supplementary Section S2 is missing or, if it comes latter, Supplementary sections should be reordered to be cited in order in the main text.	Done, we organized all the supplement and put the key information into the appendices of the main manuscript.
L150-151. What do you mean by "river CO2 state compared over 32 years"? Do you mean that there were no long-term trends in concentration? That there were no statistically significant differences in average CO2 concentration between the two trophic regimes? Best used past tense.	L268-272: The uncertainty of estimated TA leads to ±11% uncertainty in pCO2 estimation, however there **was** no significant difference in the river's CO2 state over 32 years
L156. Delete "with" after "multiplying".	L276: The k600 was calculated by multiplying river depth with K600
L157. Add "the" between "with" and "seven".	L811: The k600 values estimated by the StreamMetabolizer model were compared with the seven k600 calculated from seven fitted equations proposed by Raymond et al., (2012b)
L161-165. To be formal, define all the terms included in these equation (depth and T).	L288: CO2,water is aqueous CO2 (mmol m-3) estimated by pyCO2SYS, and CO2,air is CO2 in equilibrium with the atmosphere using global monthly atmospheric CO2 from 1990–2021 from National Oceanic and Atmospheric Administration Global Monitoring Laboratory (https://gml.noaa.gov/ccgg). K600 is the gas exchange rate coefficient normalized to a Schmidt number of 600 (d-1) from streamMetabolizer, ScCO2 is the Schmidt number for CO2 (unitless), T is the water temperature in degrees Celsius (oC).

L166.This subtitle is sort of funky. Something like "The trophlux categories" would be enough.	"Trophlux" is a term we introduced, combining "trophic" and "flux", but we prefer to introduce it after explaining autotrophic/heterotrophic and source/sink states
L167-169. Well, for NEP these states have been defined for decades. I think Odum reserves credit here.	L333: "Following Odum (1956), if NEP is positive (GPP > ER), the river reach is considered net autotrophic, while if NEP is negative (GPP < ER), it is considered net heterotrophic."
L178. Could also the heterotrophic-sink state also occur during high discharges because of high gas exchange with the atmosphere?	In our 32-year dataset, we observed heterotrophic-sink states only during brief transition periods (1-14 days, typically June-August) when the system was shifting from autotrophic-sink to heterotrophic-sink under low discharge conditions, not during high flow events. We attribute the heterotrophic-sink state to a delay for heterotrophy compensate for the CO_2 depletion due to buffer capacity of the carbonate system in the alkaline waters. The physical constraints of mass balance make sustained heterotrophic-sink conditions during high discharge highly improbable. Besides, the gas exchange intensity will not change the direction of the CO_2 flux, only its intensity.
L191. Explain briefly what is a seasonal decomposed time series.	L355-358: This decomposition, performed using the statsmodels Python package (Seabold & Perktold, 2010), separates a time series into trend, seasonal, and residual components, allowing us to specifically identify change points in the characteristics of the seasonal cycle (e.g., amplitude changes) independently of the long-term trend (Appendix C, Section C1).
L198-201. For the reader to follow this prediction the expectations need to be better explained in the last paragraph of the introduction.	The revised introduction in your comment L96 addressed this
L201-204. Please, provide some hints of how the two metrics used to characterize the hysteresis loops are expected to change between the phytoplankton vs macrophyte-dominated regimes.	L371-373: We evaluated the hysteresis loops by the direction of hysteresis (clockwise or counterclockwise) and the magnitude (i.e., difference of FCO_2 at the same discharge but in the rising and falling flow stage). We anticipated that the shift to a macrophyte-dominated period, with potentially more stable GPP across flow conditions, might lead to a reduction in the magnitude (i.e., a "flattening") of these hysteresis loops for FCO_2 and NEP.
Results	L382-387: The pronounced seasonal variability drove successive transitions among different trophlux states. Typically, the river would shift from being heterotrophic and a CO_2 source in winter, towards becoming more autotrophic in spring and potentially a CO_2 sink in summer during low flows, before returning

L210. Explain better this successive seasonal transition between autotrophic/heterotrophic of NEP, sink/source of FCO2.	to heterotrophic source conditions in autumn with rising flows. This general seasonal pattern, involving changes in both NEP (autotrophic/heterotrophic balance) and FCO2 (sink/source status), recurred each year, though its specific timing and intensity varied (Figure 1c).
L218-221. Would be helpful to include discharge in Figure 1.	To do this without cluttering the main figure, we have created a new figure in the Supplementary Information that plots fluxes mean discharge together. L273-275: The corresponding hydro-climatic context, including mean annual discharge for each year, is provided in the Supplementary Information (Figure S3) for a detailed comparison  Figure S3: Comparison of carbon fluxes (-NEP (green) and FCO2 flux (grey)) with river discharge (blue) across daily, annual, and seasonal timescales.

L222. Change "-" by "to" between "-383" and "584" to smooth the reading (suggestion holds for the whole results section). Note that the units are expressed differently in the main text and in Figure 1 (y vs yr)	Done: from -383 to 584 g C m-2 yr-1
L224. Add "net" before "source".	Done: the Loire River was a net source of CO2 to the atmosphere
Figure 2. Honestly, pCO2, pH, alkalinity, and k600 could be moved to the supplementary.	We feel these parameters are central to understand the timing of shifts we are discussing. Moving them to SI would detach the visual evidence for the identified change points from the main narrative. We prefer to keep Figure 2 in the main text but ensure the discussion of these change points is focused and directly relevant to the main hypotheses.
L244-249. Better apply a statistical test for comparing average values for each period and variable.	L907: Figure C2. Statistical comparison of key variables across decades within each trophlux. Boxplots show median (horizontal line), interquartile range (box), and distribution of annual values (black dots). Colored boxes represent different decades: blue (1990-2000), purple (2001-2010), and orange (2011-2021). Statistical significance of differences between decades was assessed using Kruskal-Wallis tests, with significant pairwise differences (Mann-Whitney U test) indicated by horizontal bars with asterisks (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).
L250. For the whole section and throughout the manuscript, would be more helpful to the reader if you refer to "heterotrophic-CO2 source" rather than to "heterotrophic-source". Also, the text would flow better if you refer consistently to the 4 trophlux states throughout.	We explained this in 2.3.3. Categorizing NEP-FCO2 states by autotrophic/heterotrophic and source/sink states
L255. Clarify whether this is the annual range or an average value for each decade, and if the later, provide s.e.	L437: The joint occurrence of the heterotrophic-source state thus ranged from 47.3 ± 9.4% in 1990-2000 to 66.8 ± 11.3% in 2011-2021 (Table 1),
L256. "coinciding with low water temperature and high discharge". This result is quite rough. Please use a two-way ANOVA test or similar to support this statement. Clarify whether the 90% refers to all data or to each decade.	Done, coinciding with low water temperature and high discharge (Figure C2). L907: Figure C2. Statistical comparison of key variables across decades within each trophlux. Boxplots show median (horizontal line), interquartile range (box), and distribution of annual values (black dots). Colored boxes represent different

	decades: blue (1990-2000), purple (2001-2010), and orange (2011-2021). Statistical significance of differences between decades was assessed using Kruskal-Wallis tests, with significant pairwise differences (Mann-Whitney U test) indicated by horizontal bars with asterisks (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).
L257-259. Bis as my earlier comment. If these are averages per decades, add the s.e., and clarify to which decade refer each number.	L442-445: Within this heterotrophic-source state, the contribution of internal sources (–NEP/FCO2) to total CO2 emissions varied across the decades: it was 37 ± 27% in 1990-2000, 28 ± 9% in 2001-2010, and increased to 57 ± 10% in 2011-2021 (Table 1). This implies that external CO2 sources accounted for the remaining proportion in each period.
L260-261. This sentence is confusing. Moreover, two of the remaining three trophlux states act as sinks rather than sources of CO2, so why one would expect them to contribute to FCO2?	We mentioned all states since we need them for the net FCO2 calculation. L446 – 447: The remaining three trophlux states (autotrophic-source, autotrophic-sink, and heterotrophic-sink) had a combined net impact of less than 10% to total FCO2 despite their regular occurrence (e.g., up to 50% of time) during the 1990-2000 decade
L261-263. Just refer to the occurrence of the autotrophic-CO2 sink state to follow the same logic throughout.	L448 – 449: The autotrophic-sink state, driven by high GPP, typically occurred for 1–3 months during the summer growing season
L263. Are these average values? Then add s.e.	L450-451: The autotrophic-sink reduced annual FCO2 by -3.0 ± 4.2% during 1990–2000 and by -0.4 ± 0.3% in 2011–2021
L267. Say more clearly that this state represented a small sink of CO2.	L454-455: The heterotrophic-sink state occurred rarely (1–7% of time) and had a small influence on the annual FCO2 budget (reducing it by 0.1% to 0.8% across the decades).
Table 1. The caption should better explain the variables to make sure that the table is self-explanatory. All the "footnotes" included at the bottom of the table include important information to interpret the data and thus should be included in M&M (some of this info would help to answer some of the below questions). It would be helpful to provide statistical tests among periods and states, and rewrite the results of this subsection in light of the result of these statistical tests. This Table arises some issues on how the calculated variables should be interpreted. It would be nice to add some text in the M&M helping the reader to interpret these	Most of these questions are related to previous responses. We briefly answer here: 1. We revise the Methods to state explicitly that Heterotrophic-Sink is an accounting outcome under specific transient conditions and does not imply a physical negative external CO2 source. 2. You are correct, the negative ratio is mathematically correct but confusing, we remove this metric.

values. Another possibility is to focus on the mass balance calculations (now in the discussion). More specifically:

1. How can external CO2 sources be negative? This suggests that there is some other unaccounted process fixing CO2, and also uncertainties associated with your calculations, which should be better constrained. Overall is unrealistic to think that groundwater is not supplying, but consuming CO2.
2. How can be the contribution of internal processes to total CO2 evasion (i.e. -NEP/CO2) a negative value as reported for both the autotrophic-CO2 source? In this case, I would say that the contribution of internal processes to CO2 evasion is 0%, and that photoautotrophic organisms are contributing to fix more CO2 than supplied by groundwater.
3. How can the contribution of internal processes to total CO2 evasion be higher than 100% as reported for the autotrophic-CO2 sink state? In this case, are photoautotrophs fixing CO2 from the atmosphere to fulfill their photosynthetic requirements?.

3. We used the mass balance diagram in Figure 5 to clearly illustrate that NEP is the largest flux, supported by both FCO2 and external inputs in the autotrophic-CO2 sink state.

Besides, in Table 2, the external CO2 is not negative, what is negative is the difference in the external flux between autumn and spring, ie this variable is higher in spring than in autumn. This is why we called it as hysteresis in the Table as below.

	Flux*	1990–2000	2001–2010	2011–2021
River metabolism ($g\ C\ m^{-2}\ d^{-1}$)	$-NEP_{300}$ spring	-0.28 ± 0.41	-0.69 ± 0.11	0.83 ± 0.18
	$-NEP_{300}$ autumn	0.66 ± 0.15	1.43 ± 0.13	1.5 ± 0.06
	$-NEP_{300}$ hysteresis	0.94 ± 0.43	2.12 ± 0.17	0.68 ± 0.18
Total CO_2 fluxes ($g\ C\ m^{-2}\ d^{-1}$)	$FCO_{2\ 300}$ spring	1.55 ± 0.44	2.65 ± 0.15	2.42 ± 0.12
	$FCO_{2\ 300}$ autumn	4.77 ± 0.14	4.15 ± 0.25	2.34 ± 0.08
	$FCO_{2\ 300}$ hysteresis	3.22 ± 0.47	1.5 ± 0.29	-0.09 ± 0.14
External CO_2 flux ($g\ C\ m^{-2}\ d^{-1}$)	($FCO_{2\ 300}$+ NEP_{300}) spring	1.83 ± 0.2	3.34 ± 0.05	1.59 ± 0.14
	($FCO_{2\ 300}$+NEP_{300}) autumn	4.12 ± 0.05	2.72 ± 0.12	0.83 ± 0.05
	($FCO_{2\ 300}$+NEP_{300}) hysteresis	2.29 ± 0.2*	-0.62 ± 0.13**	-0.76 ± 0.15
$-NEP/FCO_2$ (%)	$-NEP/FCO_{2\ 300}$ spring	-26 ± 33%	-26 ± 5%	34 ± 6%
	$-NEP/FCO_{2\ 300}$ autumn	14 ± 3%	34 ± 1%	64 ± 1%
	$-NEP/FCO_{2\ 300}$ hysteresis	40 ± 34%	61 ± 5%	30 ± 6%

* The hysteresis flux is equal to the difference of flux in autumn and spring, i.e., 2.29 = 4.12–1.83, where the 300 subscript refers to the fact that these measurements are averages from mean discharge at 300 $m^3\ s^{-1}$.
** Negative hysteresis flux indicates lower flux in autumn (rising waters stage), i.e., -0.62 = 2.72–3.34

L283. Better fully write down what -13%/12 years mean, since this is an important result. The same in line 289.	This is a total reduction of approximately 13% over the 32-year study period
L292. Better write "with either annual discharge or annual temperature".	L508: The decrease in +NEP was not significantly correlated with either annual discharge ($R2 = 0.03$) or annual temperature ($R2 = 0.00$) (Table D2).
L292. What about the other two states?	L510: Regarding the other two trophlux states, the autotrophic-source state showed a significant increasing trend in occurrence, similar to the heterotrophic-source state (Figure 3a). This state was also associated with increasing annual

	temperatures and decreasing annual discharge over time (Figure 3c,e). However, unlike the heterotrophic-source state, there was no significant long-term trend in FCO2. In contrast, the heterotrophic-sink state existed rarely over the study period and showed no significant long-term trends due to its very infrequent occurrence.
L276-281. So, if there is a 62% reduction in external inputs, but only a 13% reduction in discharge, what could explain the reduction in groundwater CO2 concentration over time?	We discussed this in the response in the previous comments. Basically, the decline in external CO2 likely reflects a decrease in CO2 concentration in groundwater inflows or less connectivity.
Figure 3. Why did you use Theil-Sen slopes rather than regular linear slopes? Mention this briefly in M&M. Add "only for the heterotrophic-CO2 source state" in f.	We indeed explained it in Methods that we used Theil-Sen slope from Mann-Kendall because it's robust to outliers and not assuming normal distribution, referencing pyMannKendal
L301. The M&M methods should better explain how the seasonality is imbedded in the rising and falling stages of the discharge.	Actually, the Methods (Lines 198-204, first manuscript) describe evaluating hysteresis against discharge across periods. The rising stage is autumn, falling is spring. We will clarify more: L369-373: We evaluated the hysteresis loops by the direction of hysteresis (clockwise or counterclockwise) and the magnitude (i.e., difference of FCO2 at the same discharge but in the rising and falling flow stage). We anticipated that the shift to a macrophyte-dominated period, with potentially more stable GPP across flow conditions, might lead to a reduction in the magnitude (i.e., a "flattening") of these hysteresis loops for FCO2 and NEP.
L305- "and from CO2 sink to source". Really? It seems the stream was acting as a source almost all year in Figure 4, except for some particular days.	We double-checked Figure 4: indeed, in the early years (1990–2000) there were brief periods (particularly July–August) where FCO2 dipped below zero (river acting as a sink). That is why we described a seasonal sink phase in summer
L309. "The contribution of external sources largely mirrored these patterns" This sentence needs some extra clarification.	L536-539: The contribution of external CO2 sources (FCO2 + NEP) also exhibited a clockwise hysteresis loop with discharge, with higher external contributions during the rising limb (autumn/winter) and lower contributions during the falling limb (spring) at equivalent discharge rates, generally mirroring the patterns observed for FCO2 and –NEP (Figure 4, bottom row).
Figure 4. Why are you showing temperature in the color ramp?	We included temperature as the color gradient in Figure 4 because temperature strongly influences both biological activity and gas exchange. Plotting

	temperature along the hysteresis loops allows the reader to see the seasonal progression (e.g., moving from cool conditions to warm and back) associated with the rising vs. falling limb. This additional context enriches the interpretation of the loops, as it aligns with seasonal metabolic drivers.
L316. Add “the” between “in” and “three”.	Done L547: "...varied across the three decades..."
L317-328. Is this analysis of the hysteresis at 300 m3/s really a fundamental result of the paper? I understand the authors are doing this analysis to showcase the seasonal patterns exhibited by the variables studied. But this is already shown in Figure 4. My suggestion would be to withdraw these text (and Table) from the results and just select some of these numbers to illustrate the magnitude of these seasonal changes in the discussion.	Q = 300 m3/s is the yearly average discharge in Loire river. Figure 4 shows the overall loops, but extracting precise differences at a common discharge point from the figure is difficult. We believe Table 2 provides valuable quantitative support for the changes in hysteresis described. It's a concise table. We prefer to keep it to substantiate the claims about changing hysteresis magnitude.
L319. I don’t think the lack of slope makes the relationship more linear or more predictable than in the previous two decades.	We rephrase to be more precise L551-555: The near-zero FCO2 hysteresis magnitude at 300 m3 s-1 in 2011–2021 (Table 2) indicates that FCO2 values at this discharge were very similar during both the rising (autumn) and falling (spring) limbs of the hydrograph. This reduction in the hysteresis loop area suggests that the relationship between FCO2 and discharge became less dependent on the seasonal progression of the hydrograph in recent years, making FCO2 at a given moderate discharge more predictable regardless of season.
Discussion L330-342. In this first paragraph, it might be good to put some numbers to this “re-oligotrophication process”.	Addressed in General Comment
L336. How do the authors explain such a decrease in external CO2 sources? Is because decrease in groundwater discharge, CO2 concentrations, or both? This comes very late in the discussion, but may be good to provide some hint here.	Addressed in General Comment

L331. Add “contribution of” before “internal source”.	Done L569: increase in the contribution of internal source of FCO2
L340-342. The reason why the authors expected to observed a weaker discharge control on FCO2 when macrophytes dominated is not clear in the introduction, so it is difficult to follow the rationale here. Not clear either what the authors mean by “weakened discharge-external CO2 source”.	We mean that the relationship between river discharge and the magnitude of external CO2 inputs became less strong in the later period. If external source hysteresis flattened (Fig 4), it means external CO2 inputs became less variable between seasons at similar discharges.
L347-349. Can you be more specific on how climate or environmental changes influence the occurrence of trophlux transitions?	L592-596: The transitions among these trophlux states are influenced by seasonal climatic drivers (temperature, solar radiation, discharge patterns, which themselves are subject to long-term climate change) and broader environmental changes like nutrient loading (e.g., re-oligotrophication) or external factors such as groundwater, making such long-term analyses critical for systems undergoing similar pressures.
L354. Be consistent with terminology throughout. Delete “metabolic”, the trophlux state does depend on the metabolic regime by definition (bis in line 356). Not clear what these ranges in parentheses are referring to, is this a range for the three decades? Perhaps it would be easier to provide an average value of the annual occurrence for the 32 years.	L597 – 601: By contrast, the mean annual occurrence of the autotrophic-sink state depended strongly on the trophic conditions of the river, decreasing from 28.7 ± 7.0% of days in the eutrophic 1990-2000 period to 7.3 ± 5.7% in the oligotrophic 2011-2021 period (Table 1).
L359-363. Has this oligotrophication process being accompanied by changes in DOM? Why a decrease in nutrient availability has influenced GPP more than ER?	Diamond et al. (2022) discuss these GPP/ER dynamics for the Loire. In our revised MS, we will add a brief explanation. L600-612: This shift in trophlux state dominance reflects the ecosystem's response to re-oligotrophication (Diamond et al., 2022). Reduced nutrient availability, particularly phosphorus, directly restrained GPP more substantially than ER because ER is supported by both autochthonous organic matter (linked to GPP) and allochthonous inputs from the catchment.
L370. Actually, more than a data point! You could say, for instance, “This study sheds new light” or “it’s a relevant contribution”….	L623: This work provides a significant long-term perspective to our understanding of the contribution of internal and external sources to FCO2 in large rivers

L372. Not sure "rigorously" is the best adjective in this case, something like "quantify at high temporal resolution" may be more appropriate.	L625: capacity to quantify the relative strength
L379. "the temporal evolution of discharge is equally important to its magnitude"? Clarify, please.	L631-633: While the magnitude of discharge is a known control on CO_2 dynamics, our results highlight that the seasonal timing and the rising or falling limb of the hydrograph are equally crucial for determining CO_2 fluxes and sources, due to hysteresis effects.
L403. Change "had an annual CO_2 sink" by "was acting as a CO_2 sink"	L665: During the 1990–2000 decade, the Loire River was acting as a CO_2 sink for almost half the years due to high rates of GPP.
L410. Why large river autotrophs benefit from being less affected by external CO_2 sources?	L671-674: In addition, autotrophy can exert a stronger control on CO_2 dynamics in larger rivers compared to small streams. This is due to factors such as increased light penetration across wider channels supporting higher areal GPP, and a larger water volume where internal metabolic signals may be less rapidly overwhelmed by the proportional influence of external CO_2 inputs from groundwater or riparian zones (Hotchkiss et al., 2015).
L411. Well, if it occurred in 2005, it was not a "long-term shift", but an "abrupt shift".	The shift from phytoplankton to macrophyte dominance was a process that occurred around 2005, not instantaneously in that single year. Minaudo et al. (2015) describe this transition. "Long-term shift" here refers to the change in the dominant primary producer type over the multi-decadal study.
L415. Provide in parenthesis the rate of annual increase in temperature and of decrease in discharge to get a sense of the magnitude of these changes without the need to dive on to the supplementary materials.	L680-688: Moreover, while annual water temperature increased (+0.18°C per decade, +5.7°C over 32 years) and annual river discharge decreased (-4% per decade, -13% over 32 years) (Figure 3, Table D1), these hydroclimatic trends did not show a direct, strong correlation with the timing or magnitude of these decadal metabolic shifts (Figure D2), suggesting that their manifestation on the magnitude of NEP is insignificant.
L417. Therefore…oligotrophication implies an improvement of water quality but a decrease in the capacity of the river to act as a CO_2 sink. This seems like an important take home message.	L475-477: The occurrence of this state has gradually declined over the three decades (Table 1), highlighting an important environmental trade-off where improved water quality from re-oligotrophication has diminished the river's capacity to act as a CO_2 sink.
L422. Which "linkage" and "the variation" of what? Do you mean, that a decrease in FCO_2 over time cannot be attributed to changes in groundwater inputs because discharge showed no clear decreases over time?	"This linkage" refers to the idea that reduced discharge leads to reduced lateral CO_2 transport. The sentence means that while reduced discharge seasonally might correlate with lower external CO_2 (Fig D1 shows Q vs External), the inter-annual trend of decreasing discharge doesn't strongly explain the inter-

	annual trend of decreasing external CO_2 (Fig D2 shows weak R^2 for annual Q vs annual External). This was already revised for General Comment 1.
L423-425. According to figure S7 and S8 discharge explains from 19-26% of external CO_2 sources.	L502-504: In addition, while the discharge can explain ca. 37-48% of the decay in FCO_2, it only explains ca. 19-26% of the variation in external CO_2 source magnitude (Figures D1 and D2)
L439. "low frequency variation" of what?	L522: "Trend analysis of groundwater table levels in France over the past 30 years shows low-frequency variations, specifically multi-annual (~7 years) and decadal (~17 years) cycles in groundwater level (Baulon et al., 2022)."
L439-450. Overall, I found this part of the discussion quite speculative. Afterall, why groundwater CO_2 fluxes have decreased so in the last decades? How the observed multi-annual low frequency variations relate with the results presented? How can these groundwater inputs be constrained in future studies?	As detailed in our response to your General Comment (and further supported by our new analysis of local groundwater level and groundwater pH and alkalinity trends presented in New Figure D3 & D4), our revised discussion now more robustly addresses the potential reasons for a decrease in groundwater CO_2 fluxes.
L462-464. Not sure what the authors mean in this sentence. Could you rewrite?	L553 – 557: While our study infers changes in external CO_2 sources, a full understanding of carbon transfer at the groundwater-river interface requires more DIC data from groundwaters and riparian zones (Deirmendjian & Abril, 2018; Duvert et al., 2018). Future work should focus on obtaining such data to better quantify these fluxes.
L455-468. Could you be more specific about what do you mean with "new exploration" and "extrapolation on river networks"?	L556-561: We also suggest that developing approaches to extrapolate findings from such detailed site studies across diverse river networks, integrating them with spatial hydrological and ecological data, will be key to better understanding and predicting how global changes will influence the balance between internal and external CO_2 production at broader scales, ultimately refining estimates of the role of rivers in the global carbon budget
Figure S7. Mention the color ramp in the caption.	L769: Figure D1. Relationship of daily fluxes and daily discharge or daily water temperature. Points are colored by year, as indicated by the color bar. L773: Figure D2. Relationship of annual fluxes and annual discharge or annual water temperature. Points are colored by year, as indicated by the color bar
Figure S9. What means NGF? Check units for annual external CO_2 (g C m^{-2} y^{-1})	NGF stands for "Nivellement Général de la France," which is the official vertical datum for France. I change figure from the groundwater level in overall France into a station near to study site.

	L776: Figure D3. Decreasing trend in the groundwater level in Montifault (20 km from the study site). (a) raw data, (b) after removing pumping effects with the EROS model (Data source: (Thiéry, 2018)). NGF stands for "Nivellement Général de la France," which is the official vertical datum for France.

Reviewer 2

General comments

Nguyen et al. report on long-term CO2 and metabolism data in a large temperate river. They show long-term shifts in the autotrophic/heterotrophic balance of the river, which they link to management changes that occurred since the early 2000s (i.e. lower nutrient inputs leading to lower GPP and a more heterotrophic river). Their data also suggest strong seasonal variations in both metabolism and CO2 emissions. Interestingly, while the river becomes more heterotrophic over time, hence with increased internal CO2 production, CO2 emissions tend to decrease in parallel. The authors attribute this decline in emissions to a decrease in external CO2 inputs at the catchment scale. They also show extreme year-to-year variability in both river metabolism and CO2 emissions, a finding that confirms the limitations of single-year studies and the need for long-term data where feasible.

This is a potentially great study based on a rare long-term dataset of paired O2 and (indirect) CO2 measurements. The study offers insights into the links between river metabolism and CO2 emissions, an emerging research area that is receiving some attention in smaller streams but remains unexplored in larger rivers – even less so over such extended timeframes. The findings should be of broad interest to the community, as improving our understanding of the temporal variations (diel, seasonal and interannual) in the source/sink status of rivers is a priority. However, several aspects of the paper require some improvement before it can be published.

Response: We thank Reviewer 2 for constructive feedback and recognition of the study's potential. We have addressed the general and specific comments below.

First, the Methods section lacks details that can help readers assess the robustness of the datasets and the validity of the methods. I see some details are in the SI, but some information should appear in the main text. Details regarding any QAQC of the pH, alkalinity and oxygen datasets are crucial, as the entire study relies on these parameters. Importantly, the authors mention somewhere that older membrane sensors were replaced with optical sensors in 2008. This date coincides with a clear step increase in pH values and a resulting decrease in pCO2 values (as expected) and CO2 emissions (Figure 2), which brings up a crucial question: how much of the observed long-term decrease in CO2 emissions is a result of this sensor change? How confident are the authors about the continuity of the pH measurements across the entire time-series?

Response:

We agree that demonstrating the robustness and continuity of the dataset, particularly across the 2008 sensor change, is essential. We have clarified and expanded the QA/QC description and will move a summary of this information to in the appendix of manuscript and more detail in the supplement.

To directly address the concern about the 2008 sensor change, we performed a cross-validation of the entire pH time series against independent data. The detailed QA/QC framework and this validation are described below:

Appendix L565- 595:

A1. Data Acquisition and QA/QC Framework

The dataset combines high-frequency continuous monitoring data from Électricité de France (EDF) with lower-frequency grab sample data from both EDF and the Loire-Brittany Water Agency (AELB).

The continuous measurement system of EDF is a floating platform with a temperature sensor and sensors for pH (range 0–14 pH unit), DO (range 0–20 mg L^{-1}), and conductivity (range 0–1000 µS cm^{-1}) (Campbell l ®). The surface water at 20 cm depth is pumped (ca. 0.5 L s^{-1}) through the system and measurements are recorded every 5 seconds, with average values saved every hour. It should be noted that data was collected both upstream and downstream of the nuclear power plant, with the upstream station located at the entrance of the dam and the downstream station located approximately 2-5km downstream of the dam. The data used for data analysis in this study was upstream station because of its data completeness. Grab sampling data was collected by EDF and Loire-Brittany Water Agency (AELB), including pH, conductivity, and alkalinity from 1990-2021, with frequency ranging from daily to monthly. Grab sampling data exists only in the upstream of the nuclear power plant. While AELB provided data for the period of 1990 to 2003 for these parameters, EDF supplied data from 2007 to 2021.

EDF's environmental monitoring operates under mandatory oversight from ASN (Autorité de Sûreté Nucléaire), France's nuclear safety authority, with comprehensive QA/QC protocols documented for two main periods. Prior to 2008, measurements were made with membrane sensors and validated using a multi-level quality control methodology that included routine calibrations and automated checks (Moatar et al., 2001), achieving an accuracy of approximately ±0.3 pH units and ±8% for DO. In 2008, the sensor technology was upgraded to optical sensors, and since 2009, all procedures have complied with ISO 17025:2005 standards, requiring documented calibration with certified reference standards and regular external audits.

We cross-validate the pH data from EDF with independent Water Agency grab samples (www.naiades.eaufrance.fr). The cross-validation confirms measurement continuity across the sensor transition, showing improved post-2008 agreement (pH residuals: -0.17±0.35 to -0.10±0.32 pH units) while preserving statistically significant long-term trends (**Figure A1**). Post-2008 optical sensor indicates less differences with the grab samples, but both continuous sensor data from EDF and grab samples data from water agency are well consistent for the long term trend. This demonstrates that the sensor upgrade enhanced data quality rather than creating systematic artifacts, with the step change in 2008 contributing to but not solely explaining the multi-decadal trends.

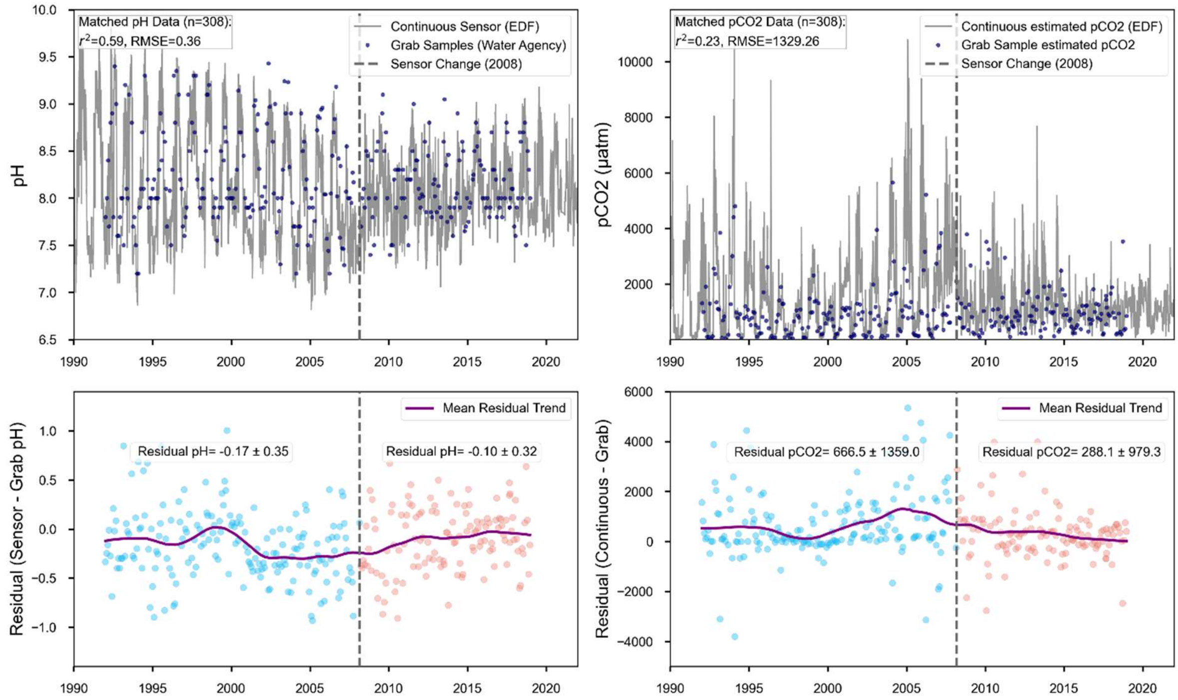


Figure A1. Comparison of continuous sensor (EDF) vs. independent grab samples from water agency for pH and estimated pCO_2 in 1990-2021. Bottom panels show residuals (Sensor - Grab) with points colored pre/post sensor change in 2008, a LOWESS trend, and mean residual ± SD.

More detail about QA/QC in the Supplement information:

1) Historical QA/QC Framework (1990-2008): Moatar et al. (2001) Methodology

During 1990-2008, DO was measured with electrochemical membrane sensors and pH with combination electrodes; these were calibrated at scheduled intervals (using two-point pH buffer solutions and air-saturation for DO) and checked against laboratory benchmarks. The measurement accuracy achieved (including routine calibration and laboratory cross-comparisons) was about ±0.3 pH units for pH and ±8% for DO (in mg O2 L-1). Sensors were inspected, cleaned (to prevent biofouling or sediment clogging), and had consumables (e.g. DO membranes) replaced as needed in accordance with the manufacturer's recommendations and agency protocols.

The monitoring system implemented the quality control methodology developed by Moatar et al. (2001), including: (1) Multi-level automated validation with specific range checks (pH: 6.0-10.0, DO: 0-20 mg L-1), persistence testing (>48 hours constant values), and rate-of-change detection (±0.5 pH units, ±3 mg L-1 DO hourly limits); (2) Cross-station validation comparing upstream and downstream measurements with acceptance criteria of ±0.2 pH units and ±1 mg L-1 DO; (3) Systematic drift correction with linear correction applied when bias exceeded 0.1 pH units over 7-day periods; (4) Expert review integration incorporating discharge patterns to eliminate false anomaly detection; and (5) Performance monitoring maintaining >95% data recovery with documented calibration procedures.

In our study, we specifically employed the open-source pyhydroqc toolkit (Jones et al., 2022) to automate anomaly detection (range, persistence, spike, and drift checks) and then visually inspected flagged periods in conjunction with discharge and the downstream station as an extra safeguard. Through this process, approximately 10% of the hourly pH and O2 data (mostly in the early 1990s) were removed. Short gaps (<6 h) were linearly interpolated, while longer gaps were filled using data from the paired station or smoothed via a seasonal Kalman filter, as detailed below sections.

2) Modern QC Framework (2008-present)

Since 2009, all monitoring laboratories must comply with ISO 17025:2005 international standards, requiring: (1) Documented calibration procedures using certified reference standards traceable to national standards (LNE - Laboratoire National de Métrologie et d'Essais); (2) Staff competency verification through annual testing and training programs; (3) Method validation with comprehensive quality management systems; (4) Regular external audits by COFRAC (Comité Français d'Accréditation) every 15 months; and (5) Inter-laboratory comparison testing organized by ASN to ensure measurement quality.

The 2008 sensor upgrade from membrane to optical technology occurred within this consistent regulatory framework, with both sensor generations operating under identical mandatory ASN oversight requirements.

On another note, the metabolism modelling using streamMetabolizer is described too briefly, and there are no details on the conversion of metabolism fluxes (expressed in O2 units) into CO2 fluxes (unless I have missed it). The indirect estimation of k600 is another area that needs to be further scrutinised – where and when do the two methods (empirical models versus metabolism) yield the largest discrepancies, and why might that be?

Response: The reference Diamond et al. (2021) contains the detailed model structure, priors, and fitting used for the Loire. We can briefly reiterate key aspects.

"Briefly, the Bayesian model used a state-space formulation to estimate daily GPP, ER, and K600 from from hourly DO. Priors K600 was constrained as a function of daily discharge and depth using empirical relationships (Raymond et al., 2012) to improve identifiability, as detailed in Diamond et al. (2021). Model convergence and fit were assessed using standard Bayesian diagnostic tools (e.g., R-hat statistics, visual inspection of observed vs. modeled DO)."

Regarding O2 to C conversion: "GPP and ER are estimated by streamMetabolizer in oxygen units (g O2 m-2 d-1) and subsequently converted to carbon units (g C m-2 d-1) for calculating NEP and for comparison with FCO2. In this study, we assumed a 1:1 molar ratio for O2:C for both GPP (photosynthetic quotient, PQ = 1) and ER (respiratory quotient, RQ = 1), and applied uniformly across all years and trophic regimes.

Regarding k600: Supplementary Figure S5 compares streamMetabolizer K600 (converted to k600 by multiplying by depth) with the mean and range of seven Raymond et al. (2012) equations. The Raymond et al. (2012) k600 estimates tended to be higher in summer and lower in winter compared to those estimated by the StreamMetabolizer model. Reasons for discrepancies:

- Raymond equations are empirical, based on hydraulic variables (Q, depth, slope, velocity). They do not consider the DO data like streamMetabolizer
- In summer (low flow, lower turbulence), Raymond eqs might overestimate if turbulence is low. streamMetabolizer K600 might be lower if diel DO is strong and suggests less reaeration is needed to explain DO patterns.

- In winter (high flow, high turbulence), Raymond might give high k600. If metabolism is very low, streamMetabolizer K600 might also be high to balance any residual DO variation, or it could be poorly constrained. We will add a brief synthesis of this to the main Methods.

L158-165: GPP and ER were then converted to carbon units (g C m-2 d-1) using a fixed molar O2:C ratio of 1:1. This assumption is widely used in river metabolism studies and reflects the stoichiometry of aerobic metabolism (Trentman et al., 2023). Although photosynthetic and respiratory quotients (PQ and RQ) can vary with autotrophic community composition, recent long-term analysis of the Loire River by Diamond et al., (2025) showed that such variability does not lead to cumulative bias in net ecosystem production or CO2 budgets when integrated over decadal timescales. Therefore, we adopt this approach as a reasonable and conservative approximation for estimating long-term carbon dynamics, while acknowledging it as a source of short-term uncertainty.

L169-171: The K600 values derived from the model were validated against established empirical equations and found to be of the same order of magnitude, as detailed in Appendix B, Section B2.

L676 – 695: B2. Validation of the Gas exchange coefficient (k600)

The k600 values estimated by the StreamMetabolizer model were compared with the seven k600 calculated from seven fitted equations proposed by Raymond et al., (2012b) for streams and small rivers. Both k600 estimates exhibited similar seasonal fluctuations, with the lowest values occurring in summer and the highest in winter. The comparison revealed that the mean absolute percentage error (MAPE) between the StreamMetabolizer estimates and the mean k600 from the seven fitted equations ranged from 36% to 62%. Specifically, the Raymond et al., (2012) k600 estimates tended to be higher in summer and lower in winter compared to those estimated by the StreamMetabolizer model. Such discrepancies can arise because streamMetabolizer co-estimates K600 with GPP and ER by fitting observed DO dynamics, making its estimate sensitive to the strength of the biological signal, whereas empirical equations rely solely on hydraulic proxies for turbulence. However, the k600 values derived from StreamMetabolizer fall within the same order of magnitude as those from the seven fitted equations (Figure B3). To maintain internal consistency between the metabolic and FCO2 calculations, the k600 estimates from streamMetabolizer were used for all subsequent flux calculations.

Figure B3. Comparison of the gas exchange coefficient (k600) estimated by streamMetabolizer (black line) with the mean and range (blue shaded area) of values derived from seven empirical equations from Raymond et al. (2012b)

Second, some of the interpretations, particularly regarding the potential drivers of the observed long-term decrease in CO2 emissions, need to be expanded. The authors mention a decrease in external CO2 sources, but could this trend instead reflect carbonate buffering processes, i.e. the conversion of some of the CO2 into alkalinity? As the authors do not seem to have collected any pCO2 data in shallow groundwater, could the above hypothesis be tested by updating their mass balance in Figure 5 with an additional term representing the downstream export of CO2 and/or DIC? Alternatively, have the authors considered any concomitant land use change that could explain some of the decline in CO2 inputs? Or could the decline be, at least in part, an artefact related to the sensor change in 2008?

Response:

About carbonate buffering, our data (Figure 2a) show that alkalinity in the Loire at the study site has been relatively stable over the 32 years, with no significant long-term increasing trend that would support a major shift towards CO2 sequestration as HCO3-.

About adding downstream export of CO2/DIC in mass balance (Figure 5) would make it a full DIC budget for the river reach, which is a different scope. We do not have downstream DIC concentration data at the same temporal resolution to robustly calculate DIC export over 32 years. So we prefer to keep Figure 5 as it is, but we have much detail to discuss the external CO2 now in this revised version.

L522 – 534: Further evidence from a representative local borehole at Montifault (20 km from our site) shows a clearly decreasing trend in the piezometric level since 2003 (Appendix D, Figure D3). This regional decrease in groundwater levels supports a reduction in groundwater discharge to the river.

To further explore the potential drivers behind the inferred long-term decrease in external CO2 inputs, we analyzed available data of pH and TA during 1990-2021 from several groundwater monitoring stations situated in aquifers hydraulically connected to our study site. This analysis revealed a long-term increasing trend in groundwater pH, particularly after 2008, while groundwater TA remained relatively stable (Figure D4). At stable alkalinity, an increase in pH directly corresponds to a decrease in pCO2. This inferred decline in groundwater pCO2 provides a strong, complementary mechanism explaining the observed reduction in external CO2 inputs to the river. Given the relative stability of in-river TA over the study period (Figure 2), it is likely that these shifts in the carbonate system are driven by changes in CO2 supply rather than major changes in catchment-scale weathering rates.

On another note, I would invite the authors to discuss how representative their single measurement site is of the whole river system. Are the findings scalable to the entire river system? Some of the authors have worked on spatial variations in metabolism across river networks, so this should be a relatively easy addition.

Response: Our site is in the middle Loire, the freshwater zone. Dynamics upstream (smaller, steeper tributaries) and further downstream (closer to estuary, larger, slower) will differ. However, the re-oligotrophication and macrophyte shift were observed over large sections of this part of the river (Minaudo 2015) or in other river systems (Ibáñez et al., 2022). So, the temporal trends driven by these broad-scale changes are likely representative for similar large, temperate river sections that underwent such changes. However, the absolute magnitudes of FCO2, NEP, and their balance will vary spatially. The mechanisms and patterns (e.g., importance of trophlux state transitions, hysteresis, impact of trophic shifts on CO2 sources) are likely relevant and scalable concepts for understanding other large river sections.

Specific comments from reviewer 2	**Responses from Nguyen et al.,**
Abstract L30. Remove "was" in "the degree of which was depended"	We completely update this sentence: L29-31: FCO2 exhibited strong seasonality linked to discharge, exhibiting hysteresis where FCO2 levels at equivalent discharge were 1.5 to 2 times higher during the rising limb (autumn) compared to the falling limb (spring)
Introduction L37-38. This statement should be updated with the more recent estimates in Liu et al. (2022). https://www.pnas.org/doi/abs/10.1073/pnas.2106322119	Thanks. L40-43: Streams and rivers are a major component of inland water CO2 evasion, with the most recent estimates for this flux at 2.0 ± 0.2 Pg C yr−1 (Liu et al., 2022). Earlier foundational work suggested this riverine flux accounted for approximately 60% of all inland waters (Raymond et al., 2013).
L47-49. While I agree that seasonal and interannual variations remain under-studied, this statement omits recent studies that report paired CO2 and DO measurements for several years. Perhaps tone this statement down a bit.	L55 – 58: Consequently, while progress has been made, a comprehensive understanding of the full spectrum of temporal variability (seasonal to multi-decadal) in FCO2 and its sources remains limited, particularly for large river systems. Recent work by Young et al. (2025) highlighted this complexity by documenting strong seasonal variability driven by hydrological events, temperature fluctuations, and biological productivity in a temperate river.
L77. "most eutrophic river"	L99: "one of the most eutrophic rivers in Europe "
L86. "NEP decreased by approximately 10%"	L102: with GPP declining and NEP decreasing by roughly 10% during the spring–summer growing season (Diamond et al., 2022).
Methods L117-119. What sensors were used for pH and DO measurements? What was their measurement range, accuracy, frequency of cleaning and calibration? Some of this information is in the SI but much of it should be moved to the main text.	We put most of these information into the appendices of the manuscript which should be easier to check without download the SI document. L569: A1. Data acquisition and QA/QC framework The dataset combines high-frequency continuous monitoring data from Électricité de France (EDF) with lower-frequency grab sample data from both EDF and the Loire-Brittany Water Agency (AELB). The continuous measurement system of EDF is a floating platform with a temperature sensor and sensors for pH (range 0–14 pH unit), DO (range 0–20 mg L-1), and conductivity (range 0–1000 µS cm-1) (Campbell l ®). The surface water at 20 cm depth is

	pumped (ca. 0.5 L s-1) through the system and measurements are recorded every 5 seconds, with average values saved every hour. It should be noted that data was collected both upstream and downstream of the nuclear power plant, with the upstream station located at the entrance of the dam and the downstream station located approximately 2-5km downstream of the dam. The data used for data analysis in this study was upstream station because of its data completeness. Grab sampling data was collected by EDF and Loire-Brittany Water Agency (AELB), including pH, conductivity, and TA from 1990-2021, with frequency ranging from daily to monthly. Grab sampling data exists only in the upstream of the nuclear power plant. While AELB provided data for the period of 1990 to 2003 for these parameters, EDF supplied data from 2007 to 2021. EDF's environmental monitoring operates under mandatory oversight from ASN (Autorité de Sûreté Nucléaire), France's nuclear safety authority, with QA/QC protocols documented for two main periods. Prior to 2008, measurements were made with membrane sensors and validated using a multi-level quality control methodology that included routine calibrations and automated checks (Moatar et al., 2001), achieving an accuracy of approximately ±0.3 pH units and ±8% for DO. In 2008, the sensor technology was upgraded to optical sensors, and since 2009, all procedures have complied with ISO 17025:2005 standards, requiring documented calibration with certified reference standards and regular external audits.
L121-122. What is the uncertainty of indirectly estimating alkalinity on pCO2 estimates?	L705-722: B3. Uncertainty Analysis Estimating FCO2 and NEP using models such as PyCO2SYS and streamMetabolizer often involves large uncertainties, particularly when considering the propagation of errors in all model input data and the summing/multiplying of these uncertainties in calculating fluxes (Battin et al., 2023; Kirk & Cohen, 2023). Estimating FCO2 and NEP involves uncertainties from multiple sources. While a full error propagation was beyond the scope of this study, we assessed the impact of the largest source of input uncertainty: the reconstruction of daily TA which was based on daily conductivity. The error in TA could potentially affect conclusions regarding the temporal distribution of CO2 sink/source states throughout the year, as well as comparisons with NEP. The average error in the reconstructed TA was ±190 µmol/L. Propagating this uncertainty through the pyCO2SYS model resulted in an uncertainty of ±11% in the final pCO2 estimates. As shown in Table B1, this level of uncertainty did not alter the main conclusions of the study. The annual distribution of trophic states remains consistent, with a maximum deviation of only 3% fluxes. Moreover, the dominance of the CO2 source–heterotrophic state throughout the year remains

almost unchanged, with less than a 1% difference under any range of TA uncertainty, though the magnitude of FCO2 could vary up to 602 to 841 gC m-2 y-1.

Table B1. Impact of TA uncertainty on the occurrence and FCO2 of each trophlux state.

	Period	CO_2 source						CO_2 sink					
		Heterotrophic			Autotrophic			Heterotrophic			Autotrophic		
		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
% of days	1990-2000	47	47	48	16	17	18	8	7	7	30	29	28
	2001-2010	60	61	61	24	25	27	2	2	2	14	16	13
	2011-2021	65	66	66	25	26	28	1	1	1	9	7	6
FCO_2 (gC m^{-2} y^{-1}.)	1990-2000	831	954	1100	87	103	119	-5	-4	-4	-22	-21	-20
	2001-2010	1267	1454	1669	75	88	102	-1.3	-0.6	-1	-7.4	-7.8	-6.6
	2011-2021	602	717	841	49	59	71	-1.5	-0.9	-1.2	-3.4	-2.6	-2.1

L131-132. “To avoid unrealistic estimates of K600, values were constrained…”

L666: we replaced unrealistic estimates

L130-141. This section on metabolic modelling needs additional methodological details. Please specify the priors used in the model, how model performance was assessed, etc.

L161 – 169: GPP, ER and K600. These estimates are supported by the streamMetabolizer, a R package (Appling et al., 2018). The model setup for the Loire River was described by Diamond et al., (2021) and Diamond et al., (2025).

A detail information of the modelling method from Diamond et al., 2025 paper is extracted below.

	4 DETAILED METABOLISM MODELLING INFORMATION Metabolism modelling was identical to that of Diamond et al. (2021) but is presented again here for clarity. We used a single-station open channel method to estimate stream metabolism (Odum 1956). We estimated daily fluxes for GPP (g O_2 m^{-2} d^{-1}), ER (g O_2 m^{-2} d^{-1}), and the gas exchange constant (K_{600} [d^{-1}]) using inverse fitting of DO dynamics with a state space approach (i.e., including process and observation error) and Bayesian inference with Markov chain Monte Carlo sampling (Appling et al. 2018). To constrain K_{600}, we pooled its estimates based on discharge (Figure S2), corresponding to the *b_Kb_oipi_tr_plrckm.stan* model in the *streamMetabolizer* R package (version 0.10.9) (Appling et al. 2018). The hyperprior, i.e., the prior distribution of the mean of K_{600} for all days, was lognormally distributed—Lognormal(1.1, 0.1)—which was based on O'Connor and Dobbins equations (O'Connor and Dobbins 1958) and floating dome measurements (G. Abril, unpublished). Priors for daily GPP and ER were normally distributed—N(8, 6) and N(-7.1, 7.1), respectively—where GPP was based on literature ranges from prior unpublished investigations and ER was based on developer recommendations. Four Markov chains were run in parallel on four cores, with 1000 warmup
L153. Remove "the" before FCO2.	Done L190
L156. Remove one of the occurrences of "with"	Done L193: The k600 was calculated by multiplying river depth with
L157-159. It would help to show a scatter plot comparing the k600 values from the Raymond models with those from streamMetabolizer, so readers can see how well they match. Also, please explain why the two estimates might be different at high and low flow. As importantly, please clarify which quotients were used	We add the scatter plot and update the discussion for the differences between 2 methods. L687-695: Specifically, the Raymond et al., (2012) k600 estimates tended to be higher in summer (low discharge) and lower in winter (high discharge) compared to those estimated by the StreamMetabolizer model (Figure B4). Such discrepancies can arise because streamMetabolizer co-estimates K600 with GPP and ER by fitting observed DO dynamics, making its estimate

to convert O2 fluxes to CO2 fluxes, and explain how those values were chosen.	sensitive to the strength of the biological signal, whereas empirical equations rely solely on hydraulic proxies for turbulence. At high flow (winter), streamMetabolizer may underestimate k600 because the strong turbulence and deep water can weaken the biological signal (i.e., the daily change in dissolved oxygen), which the model relies on for its estimations. The Raymond et al. (2012) hydraulic equations, by contrast, are driven by high velocity and are less sensitive to this biological signal dampening. At low flow (summer), streamMetabolizer may estimate a higher k600 because the biological signal is very strong and clear in the shallow, warm, and productive water. The model may attribute a larger portion of the observed oxygen change to gas exchange to best fit the data. However, the k600 values derived from StreamMetabolizer fall within the same order of magnitude as those from the seven fitted equations (Figure B3). To maintain internal consistency between the metabolic and FCO2 calculations, the k600 estimates from streamMetabolizer were used for all subsequent flux calculations.

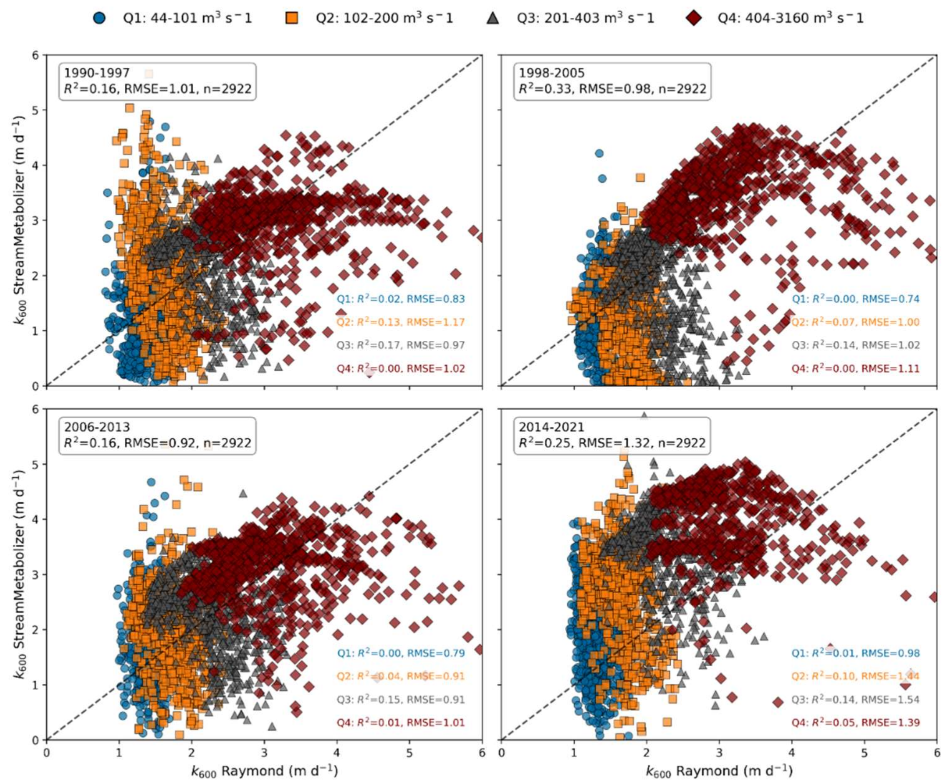
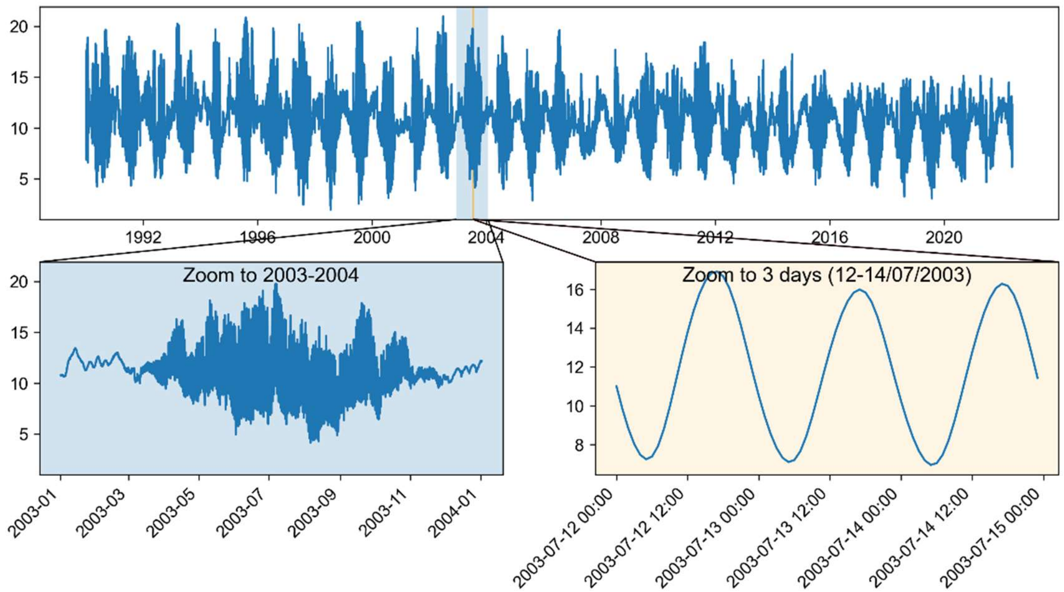


	Figure B4: Compare k600 between mean of 7 equations Raymond 2012 and StreamMetabolizer for 1990-2021. Colors indicate discharge quantiles (Q1-Q4, legend above). R2 and RMSE shown per discharge quantile range. L162-170: GPP and ER were then converted to carbon units (g C m-2 d-1) using a fixed molar O2:C ratio of 1:1. This assumption is widely used in river metabolism studies and reflects the stoichiometry of aerobic metabolism (Trentman et al., 2023). Although photosynthetic and respiratory quotients (PQ and RQ) can vary with autotrophic community composition, recent long-term analysis of the Loire River by Diamond et al., (2025) showed that such variability does not lead to cumulative bias in net ecosystem production or CO2 budgets when integrated over decadal timescales. Therefore, we adopt this approach as a reasonable and conservative approximation for estimating long-term carbon dynamics, while acknowledging it as a source of short-term uncertainty
Results L206. It would be useful to show (at least in the SI) a plot of the long-term DO time-series.	We add a figure to the SI showing the long-term (32-year) time series of hourly DO and pH concentrations. 

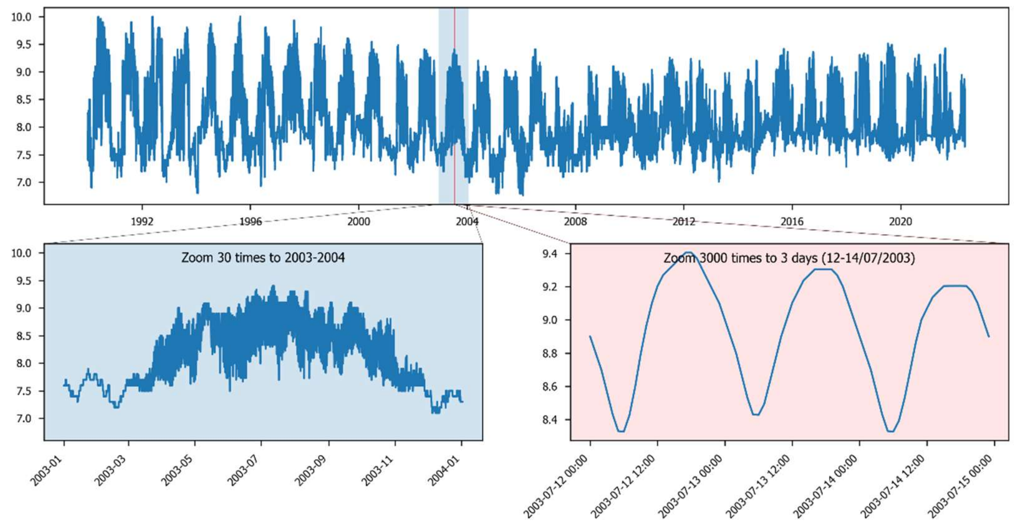
<table>
<tr><td></td><td>Figure S2a: Hourly DO (mgO2/L) in Dampierre station in 1990-2021


Figure S2b: Hourly pH in Dampierre station in 1990-2021</td></tr>
<tr><td>L210. “autotrophic/heterotrophic for NEP, sink/source for FCO2”</td><td>L261: NEP (autotrophic/heterotrophic) and FCO2 (sink/source),</td></tr>
<tr><td>L236. “while there was”</td><td>Done L287</td></tr>
<tr><td>L229. Figure 2. It seems that most of the decrease in pCO2 is driven by an increase in pH rather than a change in alkalinity. How confident are the authors that this is not related to the sensor replacement that occurred in 2008?</td><td>We addressed this in General Comment
More detail in Appendix A: Dataset integrity and validation</td></tr>
</table>

Discussion L336-337. What explanations can the authors put forward to explain such decreases in external CO_2 sources? Are they discharge-related? Do they relate to biogenic or geogenic sources? Can the decline be the result of carbonate buffering, i.e. some of the CO_2 is converted to alkalinity follwoing increases in pH?	These questions are at the core of the discussion on external CO_2 sources. We update the discussion section of external CO_2 and add new Appendix D: Analysis of Hydroclimatic and External Drivers on FCO_2. Discharge-related: Just partly, but inter-annual Q trend only explains ~19% of external CO_2 trend Biogenic/geogenic sources: External CO_2 comes from soil respiration (biogenic), groundwater (can be biogenic from overlying soils or geogenic from carbonate/silicate weathering reactions that also produce CO_2 or consume it and affect DIC). A decrease could mean less soil CO_2 production/transport, or changes in groundwater chemistry/pathways. Carbonate buffering: Addressed in General Comment. Stable alkalinity suggests this is not an accelerating in-river sink mechanism explaining the FCO_2 trend. The Discussion section these, focusing on groundwater levels and groundwater pH, alkalinity changes. L527: Further evidence from a representative local borehole at Montifault (20 km from our site) shows a clearly decreasing trend in the piezometric level since 2003 (Appendix D, Figure D3). This regional decrease in groundwater levels supports a reduction in groundwater discharge to the river. To further explore the potential drivers behind the inferred long-term decrease in external CO_2 inputs, we analyzed available data of pH and TA during 1990-2021 from several groundwater monitoring stations situated in aquifers hydraulically connected to our study site. This analysis revealed a long-term increasing trend in groundwater pH, particularly after 2008, while groundwater TA remained relatively stable (Figure D4). At stable alkalinity, an increase in pH directly corresponds to a decrease in pCO_2. This inferred decline in groundwater pCO_2 provides a strong, complementary mechanism explaining the observed reduction in external CO_2 inputs to the river. Given the relative stability of in-river TA over the study period (Figure 2), it is likely that these shifts in the carbonate system are driven by changes in CO_2 supply rather than major changes in catchment-scale weathering rates. […] Together, the evidence for both reduced groundwater discharge and lower groundwater pCO_2 provides a robust explanation for the observed multi-decadal decline in external CO_2 sources to the Loire River.
L374-377. Solano et al. (2023) present a figure summarising the range of percent contributions of NEP to FCO_2 across the literature, which could provide useful context here.	L452: More recently, Solano et al. (2023) highlighted the vast range of this contribution, with values spanning from less than 10% to over 100%. Our results in the Loire River, which also show strong seasonal variability in this metric, align with this broader context. This reinforces that

https://aslopubs.onlinelibrary.wiley.com/doi/full/10.1002/lno.12334	within a single large river system, the contribution of internal metabolism to FCO2 can be highly dynamic, covering much of the spectrum reported across diverse global systems
L418. OK, some of my earlier questions are addressed in this section. I still think further discussion is needed, particularly regarding the potential influence of the sensor change in 2008, and the role of carbonate buffering. On this second point, I suggest integrating downstream export into the budgets presented in Figure 5.	We added an explicit discussion of the sensor change in 2008 in the Appendix A. In Appendix A1 and shown in Figure A1 of our manuscript, this analysis confirms that while the post-2008 optical sensors show improved agreement with grab samples, the statistically significant long-term trends are preserved across the sensor transition Your suggestion to integrate it into Figure 5 is robust. However, the primary goal of our Figure 5 is to visually partition the fluxes that directly produce the observed atmospheric evasion (FCO2), which is the focus of our study, not the full reach-scale DIC budget. Fortunately, the complex inorganic carbon dynamics in the Loire are precisely the focus of the companion paper by Diamond et al. (2025), which uses the same dataset.
L440. Add "scales" after multi-annual and decadal.	Done, but cycles maybe more suitable L530: multi-annual (~7 years) and decadal (~17 years) cycles