nature portfolio

Peer Review File

Beavers can convert stream corridors to persistent carbon sinks

Corresponding Author: Dr Lukas Hallberg

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Communications Earth & Environment.

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Dr Hallberg,

Your manuscript titled "Beavers can convert stream corridors to persistent carbon sinks" has now been seen by 3 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised substantial concerns that must be addressed. In light of these comments, extensive revisions will be required before we can further consider the manuscript for publication. We would, however, be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. Additionally, the following editorial thresholds should be addressed in the revised manuscript:

* Clearly present sufficient novel insights into the role of beavers in converting carbon sinks.
* Clearly discuss the methodological details so that they support all the conclusions.

If additional work allows you to either incorporate or refute these criticisms, we will be happy to look at a substantially revised manuscript. If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

When resubmitting, please provide a point-by-point response to the reviewers' comments. Please submit your responses as a separate file, distinct from your cover letter where you can add responses to the Editors' comments that you do not want to be made available to the reviewers. Word files are preferred. We recommend that any figures, tables or graphs that are included in the response to reviewers are also included in the main article or Supplementary Information.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and any completed checklist:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact us if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Somaparna Ghosh, PhD
Associate Editor,
Communications Earth & Environment
Consulting Editor,
Communications Sustainability

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

For Manuscripts that fall into the following fields:
• Behavioural and social science
• Ecological, evolutionary & environmental sciences
• Life sciences
An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript
You can download the form here:
https://www.nature.com/documents/nr-reporting-summary.pdf

For your information, you can find some guidance regarding format requirements summarized on the following checklist: (https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf) and formatting guide (https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have collected data to assess the impact of beaver dams on carbon storage and fluxes. The data was collected from one beaver-impacted wetland over a short timeframe (one year) and includes data on stream discharge, dissolved and gaseous carbon fluxes from surface waters, sediments, and vegetation. The authors conclude that the beaver dam was a large carbon sink driven by DIC infiltration into the subsurface. They quantify the impact of the dam on cumulative carbon storage over decades assuming complete infilling of the wetland by sediment and wood and compare their estimates to a scenario without beaver dams.

I applaud the authors on collecting a broad dataset that includes many of the important stores and fluxes that can be used to consider the impacts of beaver expansion on carbon stores and fluxes. Unfortunately, I do not believe that the methods and particularly the discharge data are sufficient to support the carbon balance or conclusions. The authors' calculations of DIC infiltration – calculated to be the largest C flux - are fully dependent on the difference in discharge between the upstream and downstream sites, yet several aspects of the discharge measurement are problematic, relying on insufficient data and unclear methods. If such a large infiltration flux were present, it could also be measured using seepage meters, piezometer arrays, or additional tracer additions. Elevated DIC in the subsurface could be assessed with coring. Any/all of these additions would be really useful support for the authors' conclusions if additional discharge data cannot be acquired. Beyond discharge, other methods are not sufficiently described.

Major Points:
1) The difference in discharge is used to infer a very large infiltration flux and subsurface DIC retention, which greatly exceeds the other carbon fluxes. To feel confident in this interpretation, the discharge data needs to be accurate. Unfortunately, the discharge data relies on extrapolation and incomplete, complex methods that are not justified or fully explained:

a. Discharge at the upstream and downstream locations are regularly above 0.5 m3/s, with maximums closer to 3 m3/s (Figure S1), yet there is not a single measurement at the downstream location at a discharge above 0.25 m3/s (Figure S6). Therefore the downstream discharge is often based on an extrapolation of the rating curve rather than data and thus has exceedingly large uncertainty bounds that could easily swamp the infiltration term.

b. The rating curves used to relate stream stage to discharge were defined using simple logarithmic models. Rating curves are seldom well-described by a simple function except in cases where channel cross sections are perfectly uniform, a case limited to a control by a weir or flume rather than natural banks. That both the upstream and downstream cross sections fit this ideal seems quite unlikely, unless weirs were used, which is not mentioned in the text. The uncertainty in discharge associated with this approach needs to be considered.
c. The methods do not mention any barometric correction of the pressure data, which would lead to inaccurate water depth measurements. The sensor range is also not mentioned, which impacts the precision of these instruments. Given the small range or water levels measured (~ 0.2m), this uncertainty is quite important and potentially impactful on the calculated discharge
d. The rating curves include a term for the electrical conductivity of the water, which is an unusual approach. Clearly water chemistry does not control discharge, so why was this term included?
e. Why were slug additions used rather than more traditional methods of measuring discharge? Slug additions are appropriate in some situations, but can also result in major errors if one isn't careful to consider mixing lengths, storage zones, etc. How did you ensure/check accuracy in the method?
f. Due to challenges relating the added chloride to measured electrical conductivity, the authors chose instead to relate their data to other data from a nearby site collected by 'local authorities'. No source was cited or data made available.
g. Discharge at the downstream site was calculated in a different way than the upstream site (equation 2), using an equation that suggests that downstream discharge is dependent on upstream discharge and upstream tracer mass added. This does not make intuitive sense, and this equation can be reduced to equation 1. It's unclear why this equation was formulated in this way or how it is helpful.
h. The downstream discharge is "validated" using Manning's equation with a surface roughness of 0.02. It is unclear what variables were used to model discharge using the Manning's equation – ie. How did you calculate slope? Wetted perimeter? Furthermore, in this case, choosing a roughness value serves to calibrate (rather than validate) the model, and any value could have been selected to match the other estimate. Is this value reasonable based on the actual roughness of the channel? And is this a reasonable value across the entire range of discharge, from 0 to 3 m3/s? Roughness commonly varies with stage.

2) Other methods are not clear
a. How was water surface area determined?
b. Where/how did you acquire the variables to calculate ET?
c. Cumulative C storage estimates - Why did you choose 13 years as the appropriate length to calculate cumulative C storage given that you report an estimated 'active lifespan' of ~ 33 years? And how was 33 years determined? Is this the span of the beaver or the pond? From my experience beaver ponds are seldom a constant size as new dams are created to inundate new areas and access more resources. Old dams are not maintained and eventually breach. How might these processes impact your estimates?

3) Improper methods
a. Figure 2 Multiple t-tests are discouraged due to the potential for type 1 errors. It would be more appropriate to compare the groups with ANOVA followed by post-hoc tests.
b. Biological assimilation (equation S5) is calculated based solely on upstream and downstream concentrations and upstream discharge. Masses are not used and the infiltration flux is not included. A full mass balance is needed to properly constrain biological assimilation. This decision cascades through the following equations used to determine C loads and fluxes(6 through 9).
c. Throughout, units are missing from the equations.

Other comments by line number:

Figure 1 – ET line is not continuous, changes thickness, highest in October and still quite high in January? This is odd. Shouldn't ET be highest in the summer months?

254 – please report mean discharge, that would be more meaningful and more quantitative than '2nd order stream'.
259 – could you outline the beaver wetland, dam, etc.? It's location is not at all clear from Figure S4c.
279 – does it make sense to maintain a constant EC conversion across tests? This implies that water quality is constant.
281 – Where can the data from the local authorities be found? Do they provide uncertain bounds associated with their measurements, rating curve, etc.?
282 – Downstream discharge was inferred? Why was downstream discharge not measured directly? This equation is odd, given that all of the upstream parameter cancel out (given equation 1), and so this could be reduced to equation 1
287 – Why a roughness of 0.02? Is this reasonable? What is the cross section like? Any roughness value could be used, thereby calibrating the output and resulting in a close fit.
320 – is there precedent for including EC in the rating curve? I haven't seen this done before and it is unclear why / how discharge would be dependent on EC. The solutes that act together to control EC often vary with discharge, but likely not all in the same direction and magnitude.

Supplement –
17 - QS is not defined.
22- 'halos' is an odd word choice
Fig S4 – the location of the beaver dam is not clear
191 – These equations are missing units, so it's unclear why you need a unit conversion. This equation is missing QDS.
Table S4 – RMSE is not terribly useful in a log form, what is the magnitude of uncertainty in relevant units of carbon. Nash-

Sutcliffe efficiency is not presented in this table (though it appears in the description).

Reviewer #2 (Remarks to the Author):

Manuscript summary
The manuscript by Hallberg et al. provides a comprehensive carbon budget of a temperate stream reach affected by beaver damming. The budget is based on a mass balance approach using one year of seasonally resolved field measurements of carbon fluxes and stocks across the aquatic, terrestrial and atmospheric environment, complemented by hydrological and morphometrical surveys and literature data. The authors conclude that the stream corridor acts like a carbon sink, mainly due to downwelling of DIC rich water, which more than compensated soil CO2 emissions as the main source term. Based on a hypothetical calculation, the authors argue that this sink function was much larger than if the stream corridor was not influenced by beavers. The authors upscale their results to Switzerland, finding that C burial in beaver dams have the potential to offset a notable amount of national C emissions.

Overall assessment
Streams and wetlands play important roles in the global carbon cycle, emitting and burying carbon at high rates, respectively. By damming, beavers make stream reaches resemble wetlands, but the effects on the carbon budget remain unknown. This knowledge gap partly stems from the difficulty of integrating the diverse carbon stocks and fluxes in these systems. The present manuscript makes an ambitious and valuable contribution toward addressing this knowledge gap. The paper is based on a comprehensive dataset, with an impressive diversity of measurements and estimates of carbon budget components. This complexity of measurements is mostly cautiously described and elegantly condensed in text and figures. Overall, the manuscript is well written and structured to attract a broad audience in Geoscience through clear and accessible language. The figures are generally of high quality and the literature appears to be covered appropriately, but I have a few suggestions below on literature that I think should also be considered, if space allows. Importantly, I have some concerns about the study design as well as some calculations and assumptions behind the carbon budget, as specified below. In summary, with some major revisions, the manuscript likely provides a significant contribution to freshwater research, with immediate and broad implications for stream ecology, biogeochemistry, and hydrology.
Note that I focused on assessing the aquatic components of the carbon budget, as my knowledge on terrestrial components is rather limited.

General comments
(1) A core analysis of the study involves the evaluation of the beaver impact on the stream carbon budget, relative to a counterfactual scenario for conditions without beaver activity. Ideally, a before-after control-impact design should be chosen for addressing such a research question to disentangle intervention effects from natural variability. Since such an approach would be very costly, I understand that the present study uses a more simplified approach. However, the limitations and assumptions of this approach should be explained more clearly. It remains somewhat unclear to me whether the study utilizes a before-after approach or a control-impact (space-for-time substitution) approach. It seems to me like as different approaches were used to assess different carbon budget components, but this is not fully clear to me from the text. For example, L. 194 ("based on known conditions prior to beaver occupation") and Fig. 3A ("floodplain sediments pre-beaver introduction") implies to me that some (sediment?) sampling was done at the beaver-impacted site prior to intervention, whereas L. 271 ("The field data used in this study was collected between January 2022 and January 2023, with additional sediment depth measurements and sampling in March 2023 and April 2024") suggests to me that no such pre-intervention sampling was done. Please clarify the exact approaches chosen to assess beaver impacts for the different carbon budget components (Table S1) and acknowledge underlying limitations/assumptions.

(3) I have two concerns about the calculations of the long-term carbon storage potential:
(i) To estimate the long-term carbon storage potential of the beaver pond, the authors assume that the lifespan is solely determined by the sediment infilling time (33 years)(L. 174). The authors argue that infilling "lead[s] to a stable state in long-term C storage" (L. 179) and "persistent carbon sinks" (manuscript title). Yet, dam breakage may be expected to happen before the dam is filled, often within the first couple of years after dam building (Wróbel & Krysztofiak-Kaniewska 2024, https://doi.org/10.1007/s10452-024-10130-5 and studies cited therein). Not accounting for the likelihood of dam failure would overestimate the longer-term carbon storage potential of beaver dams. This should be discussed or incorporated in the calculations, perhaps by expanding the discussion on the stability of C deposits in L. 245.
(ii) The authors acknowledge in L. 246 that the fate of (particulate) inorganic carbon remains uncertain following infilling and meadow succession. I think this discussion should be expanded by acknowledging that perhaps even larger uncertainties exist regarding the fate of dissolved inorganic carbon. A major fraction of the beaver-induced C sink is made up by DIC infiltration and it is unknown what happens with this DIC in groundwater (Klaus 2023, https://doi.org/10.1038/s43247-023-00885-4). Part of the DIC in groundwater could potentially be discharged into surface waters or diffused through soils and then re-emitted as CO2 to the atmosphere at various time scales (Duvert et al. 2018, https://doi.org/10.1038/s41561-018-0245-y), which would weaken the C sink function of the beaver dam.

(4) In the extrapolation, beaver-driven C burial is compared to Switzerland's annual carbon emissions. Why was only one component (burial) of the carbon budget used in this comparison, rather than utilizing the more comprehensive net carbon balance of the beaver-affected stream, corrected for the net carbon balance of the stream free from beaver activity?

(5) I get the impression that the authors did a careful job at including the most important components of the stream carbon budget. Yet, I wonder how they reason about the following additional components not accounted for in the budget:

(i) I am somewhat surprised to see low CH_4 emissions from the beaver pond. CH_4 emissions from such systems are typically high and a large proportion of these emissions is through ebullition (Ford & Naiman 1987, https://doi.org/10.1139/z88-076; Weyenmeyer 1999, https://doi.org/10.1029/1999GB900047; Gazolla Gatti et al. 2018, https://doi.org/10.1016/j.scitotenv.2018.07.042; Smufer et al. 2023, https://doi.org/10.1007/s10021-023-00835-3). As discussed in L. 234-235, most previous studies were from more organic rich systems, which could explain lower diffusive emissions in the studied Swiss system. Still, ebullition could potentially play an additional role that remains unassessed in the present study. I see that carbon content in the sediment was higher in the beaver pond relative to control sites (Fig. 3a), generally suggesting increased potential for CH_4 ebullition. The manuscript would benefit from a convincing mechanistic explanation for why CH_4 ebullition would be negligible at their study sites, or alternatively, an acknowledgement of uncertainties from ignoring CH_4 ebullition.
(ii) POC export (e.g. downstream export, or atmospheric export through emerging aquatic insects commonly favoured by beaver activity, see e.g. Schloemer et al. 2025, https://doi.org/10.1111/fwb.70046)
(iii) Upstream- or groundwater CH_4 inputs.

(6) Since measurements were done in one year and then extrapolated to the lifespan of the beaver pond, I wonder how much the carbon budget could vary across years? From Fig. 2a, it looks like that January starts at much lower levels than December ends, suggesting potential for considerable interannual variation, at least during winter.

(7) Was there any significant temporal autocorrelation in the data? If so, was it accounted for in the statistical analyses (t-test, ANOVA, multiple linear regression)? Not accounting for autocorrelation my bias test outcomes.

Specific comments
L. 27: I would suggest to replace “complete” by “comprehensive”, because it is near impossible to account for all components of a carbon budget.
L. 69-71: Does this sentence refer to Fig. 1? If so, add a reference to it.
L. 77-79: “The estimated average loss rate of 0.13 m day-1 is high for wetland systems, but physically plausible given the permeable gravel substrate and the elevated hydraulic gradients created by the impoundment” requires support by the literature.
L. 97-100: I cannot quite follow how the proportions for DIC, POC, biomass accumulation and DOC were derived from the data shown in Table 1. Can this be clarified to the less observant reader?
L. 106: It might help readers not familiar with northern hemisphere mid-latitude seasonality to specify what is meant by “spring”, “summer, “autumn” etc., here and throughout manuscript.
Table 1: Consider rearranging the table to reduce excessive blank space, particularly in the lower right corner. For example, the output terms could be placed to the right of the storage terms.
L. 146: consider replacing “driven” by “primarily driven” to reflect that there are other terms that also contribute (but in a minor way)
L. 155-156: The carbon concentration in sediments depends on bulk density, so I wonder to what extent the increase in concentration (%) would reflect an increase in carbon accumulation or total inventory (g C/m2) relative to being a result of compaction? While I don’t question the explanations for increased C concentration listed in L. 157-159, I still wonder if the authors have data on bulk density to evaluate potential compaction effects? (L. 229 suggests bulk density data exists).
Figure 4: It is difficult for me to distinguish the different colours in the arrows. Increasing arrow size or choosing more contrasting colours may solve this problem.
Figure 4: Clarify the uncertainty measure given for Net C of the beaver-impacted stream reflect.
Figure 4: Can the authors also provide an uncertainty of the Net C for the counterfactual scenario without beaver? Uncertainties are discussed nicely in Text S7 and Figure 4 would benefit if these uncertainties were reflected in the figure.
L. 225: To support the idea that gravels could enhance subsurface carbon losses, one could cite the study by Boodoo et al. (2019, https://doi.org/10.1002/hyp.13450), who demonstrated that streamwater downwelling account for a large contribution of the CO_2 budget in a gravel bar of a pre-alpine temperate stream.
L. 226: I might have missed it, but where is the stabilized sediment TOC storage data (98 t ha-1) shown (Table / Figure)?
L. 227-229: I think the reason for the differences in TOC storage and GHG fluxes between the Swiss system and previously studied systems elsewhere could be clarified a bit more by rephrasing L. 227-233, highlighting more clearly that the difference in TOC storage and GHG emissions could be are result of the Swiss study system having generally lower organic carbon content (%).
L. 241-243: I get a bit confused by the focus on TOC accumulation in the context of beaver-mediated C sinks, while elsewhere (e.g. L. 30-31), the focus is on subsurface removal of inorganic carbon.
L. 256: How could the historic channelization and clay lining of the study stream have affected the study results and how representative is the channelization and clay lining of the study stream for other streams in Switzerland that the results are extrapolated to?
L. 272: Were V-notches installed in the stream to assure robust discharge rating curves? IF not, how well-defined and stable were the stream banks were water level was monitored?
L. 289: Were DOC samples be stored dark?
L. 288-292: How would you rate the risk of the water samples having changed in alkalinity (e.g. through CO_2 degassing or metabolic activity in the samples) between sampling and analysis?
L. 292: Which method was used to convert alkalinity to DIC? Would alkalinity contributions from organic acids be an issue? If so, alkalinity could be corrected for these contributions using DOC measurements (Lydersen et al. 2004, https://doi.org/10.1016/j.scitotenv.2003.12.005)
L. 299: Data interpretation would be aided by more details on the flux chambers used. What are the dimensions; were they ventilated; how did soil chambers differ from chambers deployed in open water; were the open water chambers drifting or static; did the soil chambers rest on air-tight collars?

L. 310: What criteria were used when deciding on the depth of the soil core samples / why were different depths chosen for the wetland and forest sites?
L. 353: Can you provide a reference to motivate the choice of 5% mass subtraction?
Supplementary Material L. 52: On what basis was assumed that deadwood volume was the same above and below the water surface?
Supplementary Material L. 159-169: How were the factors (e.g. net surface radiation, heat fluxes) included in the Priestley-Taylor energy balance model measured?
Supplementary Material L. 187: Why was the DOC assimilation pathway excluded? It is widely accepted that DOC can be assimilated in streams (e.g. Creed et al. 2015, http://dx.doi.org/10.1139/cjfas-2014-0400). DIC production and DOC assimilation can be linked through mineralization. Hence, if the authors account for DIC production, the exclusion of DOC assimilation should be motivated.
Supplementary Material L. 217-219: Where does the estimate for Ra_above come from / how was it derived?
Supplementary Material L. 217-220: “aligning with forest flux tower data across temperate Europe” should be supported by a reference.
Supplementary Material L. 222: I question the reasoning that the stream can be assumed net heterotrophic when stream CO2 emissions exceed stream GPP. Even net-autotrophic streams could show this pattern when stream CO2 emissions are largely sourced from groundwater CO2 inputs rather than respiration in the stream (Duvert et al. 2018, https://doi.org/10.1038/s41561-018-0245-y).
Supplementary Material L. 232: “chamber bias, limited diel coverage” could be supported by literature, e.g. Vingiani et al. 2021, https://doi.org/10.5194/bg-18-1223-2021 and Attermeyer et al. 2021, https://doi.org/10.1038/s43247-021-00192-w).
Supplementary Material L. 235: In small streams, turbulence and hence gas exchange is mainly driven by channel bed friction rather than wind action (Alin et al. 2011, https://doi.org/10.1029/2010JG001398).
Supplementary Material, Fig. S3: What does “Channel US (100m)” mean? I cannot find this location in the map in Fig. S4.
Supplementary Material, Fig. S6: What does “upstream reference” mean? I cannot find this location in the map in Fig. S4.
Supplementary Material, Table S4: NSE = Nash-Sutcliffe efficiency is indicated in the table header, but I cannot find this metric in the table.

Technical corrections
L. 37 “climate mitigations” should be “climate change mitigation”
Supplementary Material L. 45: I am not familiar with the term “photosynthetic activity radiation”. Do you mean “photosynthetically active radiation”?
Supplementary Material, L. 300: “show” -> “shows”

Reviewer #3 (Remarks to the Author):

Please find the attachment.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Dr Hallberg,

Your manuscript titled "Beavers can convert stream corridors to persistent carbon sinks" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment, provide you satisfactorily address all the reviewers' comments along with the following points:

* Clearly discuss the uncertainty analysis for the C budget.
* We noticed that you added a new co-author "Sarah Thurnheer (#7)" and the order of the authors' position have been changed. Please clearly write the justifiable reason of this change in your rebuttal letter.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Portfolio, please visit https://www.nature.com/commsenv/open-access

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
Link Redacted
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Somaparna Ghosh, PhD
Associate Editor,
Communications Earth & Environment
Consulting Editor,
Communications Sustainability

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

Halberg et al construct a carbon balance for a beaver-influenced stream in Switzerland. Beavers are typically thought of as ‘ecosystem engineers’ that build dams and modify hydrologic fluxes. This work goes a long way in demonstrating the potential impact of beaver dams on carbon cycling and storage as well. This is particularly relevant given beaver range expansion in Europe. Accomplishing this balance is no mean feat given the challenges of quantifying both lateral and vertical fluxes in complex terrain of partially-flooded beaver-impacted streams. The authors have collected a broad suite of data to quantify beaver-induced changes in the C flux and storage.

The revised manuscript adds important new details on the methods and results, and I generally found that the reviewer's rebuttals and edits were thoughtful and thorough. My only remaining question is regarding uncertainty quantification. I think it is important to report uncertainty associated with this approach and appreciate that the authors have done so. But it would also help to be more transparent regarding how uncertainty was quantified and propagated.

Uncertainty quantification:
I'm confused about how uncertainty was determined for the C budget. I would expect the uncertainty to be additive for the categories of Table 1, but it is not. For example, Total inputs in Table 1 were determined by summing DIC, DOC, POC, and CO_2 fluxes, but adding the uncertainties of these would result in a higher value than what is shown. How was uncertainty quantified? I don't see methods regarding uncertainty.

Related to this, what is the rationale for removing high flow days from the C flux calculation? High discharge days could be particularly important to solute and particulate transport components (both lateral and vertical) of the flux, so I'm not sure that simply removing those days from the analyses accurately captures the overall uncertainty associated with periods that exceed the stream rating curve.

Even when interpolating within a rating curve, there is always some uncertainty, and looking at the rating curves (figure s6), there is quite a bit of scatter in measurements at the upstream gauge. Was uncertainty in the rating quantified and considered in water and C fluxes?

The method for calculating biological removal has changed, now using the average discharge between the two gages. It's unclear to me why the average is used to calculate the mass, rather than using each gage's calculated discharge. Given substantial differences between upstream and downstream discharge, using an average adds another degree of imprecision / uncertainty.

Generally, uncertainty in aqueous mass fluxes should be multiplicative, considering both uncertainty in the concentration and discharge. Was this done?

Related to the total long-term sequestration, the authors report a maximum value (assuming full in-filling). I appreciate the new text highlighting this choice and explaining the rationale. That this represents a maximum should also be clarified in the abstract. And at line 194 (in track-changed doc) you could note this is the estimated maximum active lifespan of the dam

Reviewer #2 (Remarks to the Author):

I congratulate the authors for a convincing rebuttal letter and thorough revisions. I have reviewed an earlier version of the manuscript and have only few minor comments left on the revised version, as outlined below.

In the reply to comment R2.28 the authors write that the "water chambers […] were open and drifting" and that "No chambers were ventilated since it was a closed measurement system". Please clarify what is meant by "open". To me "open" contradicts "closed". Also, I am not aware of open floating chambers for flux measurements in aquatic systems and would appreciate more details or a reference that describes the method. Finally, if chambers were drifting, how can their location be indicated as a point in Fig. S4b)? How long was the drifting transect and how consistent was it at different discharge conditions? Depending on discharge, chamber travel times may vary. This can impact flux estimates in two ways: if the transect is kept constant, the time over which concentration changes inside the chambers can be measured would vary which may bias rate of change estimates. Alternatively, if the travel time is kept constant, the length of the transect may differ, potentially integrating variable stream reaches. How did the authors deal with this trade-off?

I noticed that a co-author was added to the author list. For reasons of integrity, I would appreciate if the authors could briefly motivate this addition.

Line-by-Line comments (line number refers to marked up version).

Main text
L. 102 Update the uncertainty estimate for annual C sink, to conform with the change in the abstract (L. 33).
L. 126: is it correct that the uncertainty for DIC reductions remained the same, despite change in the mean? Are all numbers throughout the manuscript updated correctly?
L. 191: are you sure that DIC only sequesters as calcium carbonates and no other carbonate form? If not, consider just writing "carbonates".
L. 268: "a pathway can account for" -> "a pathway that can account for"
L. 282: "resulting in a substantially lower sediment C storages" -> "resulting in substantially lower sediment C storages"

Supplementary Material
L. 14: "which were also reflected baseflow" – something is missing here
L. 45: "photosynthetic active radiation" -> "photosynthetically active radiation"
Text S5: Clarify units to the symbols introduced here.

L. 217: “Time Discharge” Do you mean “Time, Discharge”?
L. 319: For consistency, also explain “Channel DS”.

Reviewer #3 (Remarks to the Author):

I thank the authors for the thoughtful revision of the manuscript and for addressing all reviewer comments. My comments and concerns have been adequately addressed.

Authors' responses - revision 1
Dear Editor and Reviewers,

We would like to thank you for taking the time reviewing our manuscript and providing valuable comments on how to improve our manuscript entitled "Beavers can convert stream corridors to persistent carbon sinks".

The manuscript has been carefully revised (changes highlighted in red font color as track changes) and we have addressed all comments with detailed responses further below in this document.

We submit this revised version for your consideration for publication in *Communications Earth & Environment*.

Reviewer 1 comments	**Authors' responses** (Line numbers refer to revised version with track changes)
R1.1) The difference in discharge is used to infer a very large infiltration flux and subsurface DIC retention, which greatly exceeds the other carbon fluxes. To feel confident in this interpretation, the discharge data needs to be accurate. Unfortunately, the discharge data relies on extrapolation and incomplete, complex methods that are not justified or fully explained: a. Discharge at the upstream and downstream locations are regularly above 0.5 m3/s, with maximums closer to 3 m3/s (Figure S1), yet there is not a single measurement at the downstream location at a discharge above 0.25 m3/s (Figure S6). Therefore the downstream discharge is often based on an extrapolation of the rating curve rather than data and thus has exceedingly large uncertainty bounds that could easily swamp the infiltration term.	We have clarified the uncertainty in Q estimates due to rating curve extrapolation, wherein 11% (upstream) and 14% (downstream) of cumulative Q are above the measured range with salt slugs. **This is now reported on lines 79-81**. To test the impact of Q extrapolation on DIC load changes along the reach, we excluded all days (n = 36) where estimated downstream Q exceeded the measured maximum Q. This resulted in a net DIC removal of 130 t C yr-1, compared to the 122 t C yr-1 when including Q extrapolation. Although this is a slightly lower value, we maintain that DIC removal via infiltration is substantial also during lower Q and that extrapolation of higher Q does not strongly impact the DIC sink behaviour. **This is now reported on lines 127-130**.
R1.2) The rating curves used to relate stream stage to discharge were defined using simple logarithmic models. Rating curves are seldom well-described by a simple function except in cases where channel cross sections are perfectly uniform, a case limited to a control by a weir or flume rather than natural banks. That both the upstream and downstream cross sections fit this ideal seems quite unlikely, unless weirs were used, which is not mentioned in the text. The uncertainty in discharge associated with this approach needs to be considered.	To generate a power law relation between depth and discharge, we do not need to assume any particular geometry, it is purely empirical and emerges as a relation for most slightly irregular channels. If depth roughly scales with cross sectional area, then it is a very well known approximation from Q=vh and that a power law relation can emerge, but with the exact coefficients of this relation varying substantially from site to site. In highly variable cross sections, or especially in those with considerable overbank flow and/or backwater effects, we agree that a single function can fail to capture the relation between depth and discharge. However, our cross sections have had considerable management intervention to maintain their trapezoidal shape, and no overbank flow at these cross sections was observed during the monitoring period (noting that we were able to monitor the highest flow event on site during the year of measurement).
R1.3) The methods do not mention any barometric correction of the pressure data, which would lead to inaccurate water depth measurements. The sensor range is also not mentioned, which impacts the precision of	Water pressure measurements were corrected for barometric pressure using a pressure sensor placed on land (not submerged during higher flows). **This is now reported on line 321**.

these instruments. Given the small range or water levels measured (~ 0.2m), this uncertainty is quite important and potentially impactful on the calculated discharge	The typical water level error of HOBO U20L-01 is ± 0.01 m and have a negligible impact on measured water stages, ranging between 0.25 – 1.16 m (upstream) 0.11 – 0.75 m (downstream). **The error is now reported on lines 320-321**.
R1.4) The rating curves include a term for the electrical conductivity of the water, which is an unusual approach. Clearly water chemistry does not control discharge, so why was this term included?	In response to this concern, we have decided to not use the EC intercept in the rating curves, and instead use the traditional approach of $Q = a(h-h0)^b$ in the revised manuscript. **This is now clarified on lines 391-398 and shown in SI Fig. S6**. The new Q estimates (also including the revised calculation of Q measured with salt slugs in response to comment R1.17) resulted in slightly different Q at US and DS, which **we have propagated into recalculations of water balance (e.g., Fig. 1; Fig. S1; SI Text S1), and dissolved C loads (Table 1; Fig. 2; Fig. 4; Fig. S2).** The newly derived Q only resulted in slight changes in the water balance, which strengthened the infiltration response and thus reductions in DIC (previous = 104 t C yr-1, revised = 122 t C yr-1). In the data repository, the files "Discharge and stage", "Dissolved carbon loads", and "Water balance fortnightly" have been updated to reflect the revised results (doi: 10.6084/m9.figshare.29500514).
R1.5) Why were slug additions used rather than more traditional methods of measuring discharge? Slug additions are appropriate in some situations, but can also result in major errors if one isn't careful to consider mixing lengths, storage zones, etc. How did you ensure/check accuracy in the method?	Tracer based methods for discharge measurement are very common and is itself a traditional method. In this case, salt slug additions were ideal because of the short reach lengths (<150 m) and rapid travel times (<5 min) at upstream and downstream locations. Mixing lengths were determined on the basis of the recommended 25 times stream widths or greater (e.g., Day & Day 1977; https://www.jstor.org/stable/43944411), the upstream site has a shorter mixing length but greater turbulence due to enhanced roughness. **We have clarified mixing lengths on lines 317-318**.
R1.6) Due to challenges relating the added chloride to measured electrical conductivity, the authors chose instead to relate their data to other data from a nearby site collected by 'local authorities'. No source was cited or data made available.	The upstream reference Q data was measured at the station called 511 Mederbach-Marthalen (47.6209 °N, 8.6425 °E), monitored by Zurich Civil Engineering Office (data source: https://www.zh.ch/de/umwelt-tiere/wasser-gewaesser/messdaten/abfluss-wasserstand.html). The Q is estimated using an established rating curve for a well-defined section (photograph of station provided on the website). **Information about the station is now provided on lines 388-390**. We also made the Q from the upstream reference available in the updated file "Discharge and stage" in the data repository.
R1.7) Discharge at the downstream site was calculated in a different way than the upstream site (equation 2), using an equation that suggests that downstream discharge is dependent on upstream discharge and upstream tracer mass added. This does not make	We appreciate this comment and understand the potential confusion. Since we no longer use a constant EC conversion coefficient, equation 2 has been removed. **Our revised method is described in SI Text S5 on lines 140-158**. See also our response to comment R1.17.

intuitive sense, and this equation can be reduced to equation 1. It's unclear why this equation was formulated in this way or how it is helpful.	
R1.8) The downstream discharge is "validated" using Manning's equation with a surface roughness of 0.02. It is unclear what variables were used to model discharge using the Manning's equation – ie. How did you calculate slope? Wetted perimeter? Furthermore, in this case, choosing a roughness value serves to calibrate (rather than validate) the model, and any value could have been selected to match the other estimate. Is this value reasonable based on the actual roughness of the channel? And is this a reasonable value across the entire range of discharge, from 0 to 3 m3/s? Roughness commonly varies with stage.	To avoid uncertainty in choosing an appropriate roughness value, we removed the estimate for downstream Q with Manning's equation.
R1.9) Other methods are not clear a. How was water surface area determined?	This is thoroughly explained in SI Text S6: Water balance analysis - Evapotranspiration modelling and water storage change (SI lines 188 - 198): "To resolve the ETwetland term, acting on the inundated surface, and determine dVwetland, we estimated daily water surface area and volume using drone-based RGB imagery collected during six surveys in 2022. Water extent for each survey was mapped with supervised classification, using ArcGIS Pro (3.3.1) toolboxes Train Support Vector Machine Classifier and Classify Raster. Then, water extent at daily resolution across up-, mid-, and downstream wetland sections were estimated from relationships between water stage (at DS) and observed water surface area, fitted using linear or logarithmic regressions (Supplementary Fig. 7). To avoid water surface overestimation, linear relationships were fitted for the observed range of water surface area. Above this range, another linear regression was fitted between the maximum observed water surface area and total surface area. For the upstream wetland section, daily water surface area was fitted with a logarithmic function as this section only showed limited inundation throughout 2022."
R1.10) Where/how did you acquire the variables to calculate ET?	The data sources for ET calculation **are now clarified in SI (lines 182 – 187)**: "Data for calculating ET were both obtained from measurements in the study area (water temperature, air temperature and barometric pressure) and from MeteoSwiss (relative humidity and global radiation) using data from a nearby weather station in Marthalen (47.6244 ° N, 8.6533 ° E; https://www.meteoswiss.admin.ch/services-and-publications/applications/measurement-values-and-measuring-networks.html#param=messnetz-partner&table=false).
R1.11) Cumulative C storage estimates - Why did you choose 13 years as the appropriate length to calculate cumulative C storage given that you report an estimated 'active lifespan' of ~ 33 years? And how was 33 years determined? Is this the span of the beaver or the pond?	The cumulative C storages in table 1 refer to the time period from beaver introduction (early 2010) until the monitoring period of this study (2022).

From my experience beaver ponds are seldom a constant size as new dams are created to inundate new areas and access more resources. Old dams are not maintained and eventually breach. How might these processes impact your estimates?	The active life span of 33 years is the predicted period of beaver activity in the reach that will still occur after 2022, **clarified on lines 194-196**. This estimate is based on our sediment deposition modelling, the wetland is estimated to approach complete infilling after approximately 33 years – see Supplementary Text S4, lines 111-128. Upon infilling, we assume that beaver activity will cease because of lack of wetland inundation, resulting in a less suitable beaver habitat. We acknowledge that the predicted lifespan of 33 years is indeed an approximation, based on the successive infilling that we have observed for over a decade at this site. It is certainly true that beaver dam breaches could alter or stop this process, thus our estimates are contingent on the current deposition conditions remaining intact. However, we note at this site that the main channel has been dammed multiple times, without significant breaches or falling into disrepair during 13 years. Moreover, the low gradient nature of the site means that even if the dams were no longer maintained, maintaining the large pond and wetland can remain for a considerable time period. It is difficult for new dams to be created that alter the size of this system, as only a single dam is needed to cause the inundation in this low gradient system, and under these conditions it is difficult to add many more dams to significant effect (Larsen et al. 2021; https://doi.org/10.1016/j.earscirev.2021.103623). **Consequences of dam failure are now discussed on lines 281-288, and cumulative C storages across a range of beaver wetland lifespans are shown in the new SI Table S2**.
R1.12) Improper methods a. Figure 2 Multiple t-tests are discouraged due to the potential for type 1 errors. It would be more appropriate to compare the groups with ANOVA followed by post-hoc tests.	We have updated Fig. 2 according to your suggestion.
R1.13) Biological assimilation (equation S5) is calculated based solely on upstream and downstream concentrations and upstream discharge. Masses are not used and the infiltration flux is not included. A full mass balance is needed to properly constrain biological assimilation. This decision cascades through the following equations used to determine C loads and fluxes(6 through 9).	We thank the reviewer for raising this point. We have clarified the term as biogeochemical removal and production, replacing biological assimilation/production. We have also modified Q location used for calculating changes in biogeochemical change loads, now using the average between QUS and QDS, instead of only QUS. **See our revised text in SI, Text S7, lines 209-212**. Fig. S2b have also been updated to reflect the corrected calculations of biogeochemical sink/source of DIC and DOC loads, with only small differences compared to the previous version. In the case of DIC, this approach is based on the fact that concentration changes between upstream and downstream can only occur due to biological production/removal and chemical precipitation/dissolution (or evapotranspiration but this is negligible as shown in the water balance). Thus, concentration changes can be directly linked to biogeochemical processes and are unaffected by infiltration (this affects load but not concentration). Then, load changes

	associated with biogeochemical processing along the reach are inferred by multiplying the concentration changes with the average of QUS and QDS, to account for differences in input/output Q.
R1.14) Throughout, units are missing from the equations.	We have now specified units in all equations in Main text and SI.
R1.15) Figure 1 – ET line is not continuous, changes thickness, highest in October and still quite high in January? This is odd. Shouldn't ET be highest in the summer months?	ET is continuous but very low in comparison to the other water balance components. ET peaks in June and is substantially lower in October and January, see raw data in "Water balance fortnightly" in repo, which is the data shown in Fig. 1.
R1.16) 254 – please report mean discharge, that would be more meaningful and more quantitative than '2nd order stream'.	Median Q is reported in SI Text S1, lines 13-14. Median Q is a result and not a study site characterisation.
R1.17) 279 – does it make sense to maintain a constant EC conversion across tests? This implies that water quality is constant.	Given the uncertainty in deriving a constant EC conversion coefficient based on further upstream reference Q, we chose to convert EC measurement into NaCl concentrations using a different approach, utilising the background ionic strength in stream water and ionic strength of added NaCl to convert measured EC into NaCl concentrations, following the method of McCleskey et al 2025 (https://doi.org/10.1029/2024WR037771). **Our new method is described in detail in SI Text S5, lines 140-158**. The resulting Q estimates from the ionic strength method are highly consistent with previous estimates using the EC conversion coefficient = 0.5 (US RMSE = $1.02*10^{-3}$; DS RMSE = $1.55*10^{-3}$).
R1.18) 281 – Where can the data from the local authorities be found? Do they provide uncertain bounds associated with their measurements, rating curve, etc.?	This is addressed in our response to comment R1.6.
R1.19) 282 – Downstream discharge was inferred? Why was downstream discharge not measured directly? This equation is odd, given that all of the upstream parameter cancel out (given equation 1), and so this could be reduced to equation 1	This is addressed in our response to comment R1.17.
R1.20) 287 – Why a roughness of 0.02? Is this reasonable? What is the cross section like? Any roughness value could be used, thereby calibrating the output and resulting in a close fit.	To avoid uncertainty in choosing an appropriate roughness value, we removed the estimate for downstream Q using Manning's equation.
R1.21) 320 – is there precedent for including EC in the rating curve? I haven't seen this done before and it is unclear why / how discharge would be dependent on EC. The solutes that act together to control EC often vary with discharge, but likely not all in the same direction and magnitude.	This is addressed in our response to comment R1.4.
R1.22) Supplement – 17 - QS is not defined.	We have now added the definition of QS in SI, line 18.
R1.23) 22- 'halos' is an odd word choice	"Halos" have been removed from the sentence.
R1.24) Fig S4 – the location of the beaver dam is not clear	We have now modified Fig. S4 to clarify the location of the beaver wetland, also including locations of the Marthalen weather station and Upstream Q reference.

R1.25) 191 – These equations are missing units, so it's unclear why you need a unit conversion. This equation is missing QDS.	Units have now been added to the equation. The average between QUS and QDS is now used, see our response to comment R1.13.
R1.26) Table S4 – RMSE is not terribly useful in a log form, what is the magnitude of uncertainty in relevant units of carbon. Nash-Sutcliffe efficiency is not presented in this table (though it appears in the description).	In Table S5, we have updated the calculation of RMSE and R^2 for non-logged values, showing uncertainty in kg C day^{-1}. We removed Nash-Sutcliffe efficiency from the table text and goodness-of-fit for SS since this parameter was not included in results (poor fit with WRTDS modelling).

Reviewer 2 comments	**Authors' response** (Line numbers refer to revised version with track changes)
R2.1) A core analysis of the study involves the evaluation of the beaver impact on the stream carbon budget, relative to a counterfactual scenario for conditions without beaver activity. Ideally, a before-after control-impact design should be chosen for addressing such a research question to disentangle intervention effects from natural variability. Since such an approach would be very costly, I understand that the present study uses a more simplified approach. However, the limitations and assumptions of this approach should be explained more clearly. It remains somewhat unclear to me whether the study utilizes a before-after approach or a control-impact (space-for-time substitution) approach. It seems to me like as different approaches were used to assess different carbon budget components, but this is not fully clear to me from the text. For example, L. 194 ("based on known conditions prior to beaver occupation") and Fig. 3A ("floodplain sediments pre-beaver introduction") implies to me that some (sediment?) sampling was done at the beaver-impacted site prior to intervention, whereas L. 271 ("The field data used in this study was collected between January 2022 and January 2023, with additional sediment depth measurements and sampling in March 2023 and April 2024") suggests to me that no such pre-intervention sampling was done. Please clarify the exact approaches chosen to assess beaver impacts for the different carbon budget components (Table S1) and acknowledge underlying limitations/assumptions.	We thank the reviewer for raising this important point. The study does not rely on a before–after or control–impact (BACI) type design, as no pre-beaver or matched-control data were available or appropriate. Instead, we adopt a process-based counterfactual approach, combining empirical measurements from the present beaver-influenced system with physiologically and hydrologically constrained estimates of the same reach in its pre-beaver state. This method explicitly uses known boundary conditions (channel confinement, vegetation composition, pre-2010 topography and soils) to simulate the "no-beaver" scenario in a mass-conservative and internally consistent framework (see SI Text S8, lines 228-275). We emphasise that while before–after or control–impact designs are valuable, they are not inherently superior for long-term ecosystem-engineering processes, which involve irreversible geomorphic and ecological transitions that rarely permit true replication or baseline data. Both designs also carry their own limitations: before–after comparisons can confound temporal trends (e.g. climate or land-use changes), whereas control–impact comparisons can misrepresent hydrological or geomorphological equivalence. Our approach therefore represents a complementary third framework, conceptually similar to a counterfactual mass-balance model, which has been used successfully in recent whole-ecosystem carbon assessments (e.g., Knose et al 2025, https://doi.org/10.1021/acsestwater.4c00592; Magliocca et al., 2022, https://doi.org/10.1080/1747423X.2023.2173325). **We have clarified this logic and its assumptions in the revised Main text (lines 216-218) and SI (Text S8 lines 229-232).** No sediments were sampled before beaver introduction, sediment cores were separated into two horizons: post-beaver (non-consolidated) and pre-beaver (consolidated) sediments based on

	penetration resistance. The two horizons differed clearly in both consolidation and texture (from visual observations). **This is now clarified on lines 368-372.**
R2.2) I have two concerns about the calculations of the long-term carbon storage potential: (i) To estimate the long-term carbon storage potential of the beaver pond, the authors assume that the lifespan is solely determined by the sediment infilling time (33 years)(L. 174). The authors argue that infilling "lead[s] to a stable state in long-term C storage" (L. 179) and "persistent carbon sinks" (manuscript title). Yet, dam breakage may be expected to happen before the dam is filled, often within the first couple of years after dam building (Wróbel & Krysztofiak-Kaniewska 2024, https://doi.org/10.1007/s10452-024-10130-5 and studies cited therein). Not accounting for the likelihood of dam failure would overestimate the longer-term carbon storage potential of beaver dams. This should be discussed or incorporated in the calculations, perhaps by expanding the discussion on the stability of C deposits in L. 245.	We acknowledge that the predicted lifespan of ~33 years is an approximation, based on the successive infilling that we have observed for over a decade at this site. We cannot rule out that dam failure could alter the predicted lifespan and the C storage persistence within the wetland, but based on the stable dam conditions over 13 years, we have chosen to base our future predictions on intact dam integrity. **We have now addressed the consequences of a potential dam failure in Main text (lines 281-288) and show cumulative C storages across a range of beaver wetland lifespans in the revised SI Table S2**. See also our response to comment R1.11.
R2.3) The authors acknowledge in L. 246 that the fate of (particulate) inorganic carbon remains uncertain following infilling and meadow succession. I think this discussion should be expanded by acknowledging that perhaps even larger uncertainties exist regarding the fate of dissolved inorganic carbon. A major fraction of the beaver-induced C sink is made up by DIC infiltration and it is unknown what happens with this DIC in groundwater (Klaus 2023, https://doi.org/10.1038/s43247-023-00885-4). Part of the DIC in groundwater could potentially be discharged into surface waters or diffused through soils and then re-emitted as CO_2 to the atmosphere at various time scales (Duvert et al. 2018, https://doi.org/10.1038/s41561-018-0245-y), which would weaken the C sink function of the beaver dam.	We have expanded the discussion about the fate of DIC infiltration according to your suggestions on lines 189-192.
R2.4) In the extrapolation, beaver-driven C burial is compared to Switzerland's annual carbon emissions. Why was only one component (burial) of the carbon budget used in this comparison, rather than utilizing the more comprehensive net carbon balance of the beaver-affected stream, corrected for the net carbon balance of the stream free from beaver activity?	We appreciate this comment and understand the potential confusion. The comparison to Switzerland's annual carbon emissions (lines 232–234) specifically refers only to the long-term carbon burial rate (which is sediment + deadwood accumulation), not to the more instantaneous (but complete) annual C balance of the beaver wetland, covered in the first half of the paper. This distinction is deliberate because: 1) The annual mass balance (98 ± 34 t C yr^{-1}) captures transient flux exchanges within the reach (including subsurface DIC storage/infiltration with an uncertain fate, and short-lived biomass turnover). 2) The long-term burial rate (≈ 46 t C yr^{-1}, equivalent to 4.7 t C ha^{-1} yr^{-1}) represents net sequestration of carbon that is effectively locked

	away over much longer (decadal to centennial) timescales, comparable in persistence to national emission inventories. Hence, the upscaling to Switzerland's annual emissions was based on a per-area sequestration flux of permanent C storage, conceptually aligned with the flux of CO_2 emissions to the atmosphere. Including short-term internal cycling would not be appropriate for such a comparison because those fluxes do not represent enduring atmospheric removals. **We have clarified this rationale in the revised manuscript (lines 238-240)** by explicitly distinguishing between annual dynamic C balance and long-term burial-derived sequestration rate.
R2.5) I get the impression that the authors did a careful job at including the most important components of the stream carbon budget. Yet, I wonder how they reason about the following additional components not accounted for in the budget: (i) I am somewhat surprised to see low CH4 emissions from the beaver pond. CH4 emissions from such systems are typically high and a large proportion of these emissions is through ebullition (Ford & Naiman 1987, https://doi.org/10.1139/z88-076; Weyenmeyer 1999, https://doi.org/10.1029/1999GB900047; Gazolla Gatti et al. 2018, https://doi.org/10.1016/j.scitotenv.2018.07.042; Smufer et al. 2023, https://doi.org/10.1007/s10021-023-00835-3). As discussed in L. 234-235, most previous studies were from more organic rich systems, which could explain lower diffusive emissions in the studied Swiss system. Still, ebullition could potentially play an additional role that remains unassessed in the present study. I see that carbon content in the sediment was higher in the beaver pond relative to control sites (Fig. 3a), generally suggesting increased potential for CH4 ebullition. The manuscript would benefit from a convincing mechanistic explanation for why CH4 ebullition would be negligible at their study sites, or alternatively, an acknowledgement of uncertainties from ignoring CH4 ebullition. (ii) POC export (e.g. downstream export, or atmospheric export through emerging aquatic insects commonly favoured by beaver activity, see e.g. Schloemer et al. 2025, https://doi.org/10.1111/fwb.70046) (iii) Upstream- or groundwater CH4 inputs.	We appreciate the reviewer's thoughtful comments and the opportunity to clarify the treatment of methane (CH_4) and particulate organic carbon (POC) in our budget. We hope we can address these in turn below: (i) CH_4 emissions. The cited studies indeed report higher CH_4 fluxes in some beaver systems; however, most were conducted in organic-rich or peat-forming wetlands and focused on gas exchange processes rather than total carbon budgets. Our study distinctly integrates CH_4 within a complete C mass balance, expressed as carbon fluxes (CH_4–C), allowing a direct comparison across all C pathways. On that basis, our measured CH_4–C fluxes fall within the mid to lower range of reported beaver-wetland values, but are not anomalously low. Incorporating ebullition explicitly is important, and we note that most of the cited studies either do not quantify it or infer it indirectly. In one beaver wetland where ebullition was measured directly, this pathway contributed up to ~70 % of total CH_4 emissions. Even if we applied this upper-end correction to our system, the resulting CH_4–C flux would still contribute < 0.5 % of the total C balance, consistent with our conclusion that CH_4 plays a minor role in the overall C mass budget. Mechanistically, we attribute the modest CH_4 fluxes to the mineral-rich, non-peat substrate and the presence of alternative electron acceptors (N, Fe, and SO_4^{2-}), as confirmed by unpublished water quality data from the site. These conditions favour denitrification, iron reduction, and sulfate reduction pathways, which outcompete methanogenesis under moderately reducing conditions. Hence, potential CH_4 production is naturally constrained compared to the highly organic, strongly anaerobic pond sediments typical of published studies from North American and boreal beaver systems. **We have**

	clarified the mechanistic explanation for low CH4 fluxes and acknowledged remaining uncertainties associated with unmeasured ebullition fluxes (lines 268-273). We also monitored upstream CH_4 concentrations (Fig. S3b), indicating that potential groundwater or upstream CH_4 inputs are minor contributors to total fluxes and are effectively accounted for in the reach-scale balance. (ii) POC export and insect-mediated fluxes. We recognise that particulate transport is among the more uncertain components of the budget**, now clarified on lines 130-133**. However, our sediment trapping efficiency and reach geometry imply that downstream POC export is likely low relative to deposition, and aerial carbon export through emerging insects, while ecologically relevant, represents a negligible fraction of the mass balance at the reach scale. We hope these points and their uncertainties are **now clarified in the Main text (lines 268-273 and 130-133)**. Overall, the inclusion of CH_4, DOC, DIC, POC, and gaseous CO_2–C terms ensures that the dominant carbon pathways are captured quantitatively, and that omitting small, poorly constrained terms such as CH_4 ebullition does not affect the interpretation of the reach as a strong net C sink.
R2.6) Since measurements were done in one year and then extrapolated to the lifespan of the beaver pond, I wonder how much the carbon budget could vary across years? From Fig. 2a, it looks like that January starts at much lower levels than December ends, suggesting potential for considerable interannual variation, at least during winter.	We thank the reviewer for raising this point and agree that annual carbon fluxes are expected to vary to some extent interannually with hydrology, temperature, and vegetation dynamics. However, in this system the within-year (seasonal) variation in hydrology and surface extent is likely to be much greater than the interannual variation, and this seasonal cycle is explicitly captured in our measurements. However, it is most important to clarify that our long-term extrapolation does not extend the full annual C budget linearly over time. Rather, it applies only to the sequestration components, namely sediment and deadwood carbon burial, which accumulate progressively and can be more confidently constrained from observed sedimentation rates, bulk densities, and carbon contents (Supplementary Text 4). These storage terms integrate over many years and are thus less sensitive to short-term variability. The annual C budget presented here (98 ± 34 t C yr^{-1}) represents one complete hydrological year and was chosen precisely because it captures the dominant seasonal fluctuations in wetland inundation, gas exchange, and dissolved carbon

	transport. Interannual variation would modulate these fluxes to some extent but is unlikely to exceed the magnitude of the seasonal variability we already resolve. The long-term extrapolation therefore reflects the cumulative, irreversible sequestration trajectory of the wetland, rather than an assumption of static annual behaviour. **We have further clarified this distinction on lines 61-62, 214-215**.
R2.7) Was there any significant temporal autocorrelation in the data? If so, was it accounted for in the statistical analyses (t-test, ANOVA, multiple linear regression)? Not accounting for autocorrelation my bias test outcomes.	Temporal autocorrelation is only relevant for the multiple regression, predicting CO2 fluxes from air temperature and exposed surface area. **Results of Durbin-Watson tests are now included in SI Table S6**, indicating that no autocorrelation is present. The t-tests and ANOVAs reported (e.g. Fig. 3a–b; SI Fig. S3) were used solely to compare independent groups of spatially distinct samples (e.g. sediment types, flux interfaces), not sequential observations over time. The time-resolved C fluxes shown in Fig. 2 were analysed descriptively within a mechanistic mass-balance framework, not subjected to hypothesis testing or regression analyses. Therefore, no temporal autocorrelation corrections were required.
R2.8) L. 27: I would suggest to replace “complete” by “comprehensive”, because it is near impossible to account for all components of a carbon budget.	We have changed the phrasing accordingly.
R2.9) L. 69-71: Does this sentence refer to Fig. 1? If so, add a reference to it.	Yes, Fig. 1b is now referenced in the sentence.
R2.10) L. 77-79: “The estimated average loss rate of 0.13 m day-1 is high for wetland systems, but physically plausible given the permeable gravel substrate and the elevated hydraulic gradients created by the impoundment” requires support by the literature.	This sentence is now referenced (Russo et al., 2012; https://doi.org/10.1016/j.jhydrol.2012.02.022 and Boodoo et al., 2019; https://doi.org/10.1002/hyp.13450).
R2.11) L. 97-100: I cannot quite follow how the proportions for DIC, POC, biomass accumulation and DOC were derived from the data shown in Table 1. Can this be clarified to the less observant reader?	We chose to simplify the calculation of these proportions and now only compare the relative contribution of DIC, DOC and POC to total fluvial C retention (input DIC, DOC, POC – output DIC, DOC, POC). **This is now clarified on lines 104-106.**
R2.12) L. 106: It might help readers not familiar with northern hemisphere mid-latitude seasonality to specify what is meant by “spring”, “summer, “autumn” etc., here and throughout manuscript.	We thank the reviewer for the suggestion. Terms such as “spring,” “summer,” “autumn,” and “winter” are used in their northern hemisphere sense (spring = March–May; summer = June–August; autumn = September–November; winter = December–February). To aid clarity for international readers, **we have added this definition at first use on lines 112-113**.
R2.13) Table 1: Consider rearranging the table to reduce excessive blank space, particularly in the lower right corner. For example, the output terms could be placed to the right of the storage terms.	We have rearranged Table 1 to reduce negative space.

R2.14) L. 146: consider replacing “driven” by “primarily driven” to reflect that there are other terms that also contribute (but in a minor way)	We have changed the phrasing accordingly.
R2.15) L. 155-156: The carbon concentration in sediments depends on bulk density, so I wonder to what extent the increase in concentration (%) would reflect an increase in carbon accumulation or total inventory (g C/m2) relative to being a result of compaction? While I don’t question the explanations for increased C concentration listed in L. 157-159, I still wonder if the authors have data on bulk density to evaluate potential compaction effects? (L. 229 suggests bulk density data exists).	We confirm that bulk density measurements were obtained and explicitly incorporated in all sediment carbon stock and accumulation calculations. Carbon concentrations (%) were used in Fig. 3a to compare compositional differences between sediment types, while total carbon inventories (g C m^{-2} and t C) were computed from measured bulk density and sediment depth data (SI Text 4). These bulk density values therefore directly determine the sediment C accumulation and total storage estimates presented in Table 1 and Fig. 3c. The comparison of C % across interfaces (lines 170–175) was intended only to highlight compositional enrichment, not to imply increased storage independently of density. **We have clarified this distinction on line 173**.
R2.16) Figure 4: It is difficult for me to distinguish the different colours in the arrows. Increasing arrow size or choosing more contrasting colours may solve this problem.	We have revised Fig. 4 with higher contrast in arrow colours.
R2.17) Figure 4: Clarify the uncertainty measure given for Net C of the beaver-impacted stream reflect.	The uncertainty measure is one standard deviation, which is now clarified in Table 1 (line 148) and Fig. 4 (line 246).
R2.18) Figure 4: Can the authors also provide an uncertainty of the Net C for the counterfactual scenario without beaver? Uncertainties are discussed nicely in Text S7 and Figure 4 would benefit if these uncertainties were reflected in the figure.	We have now also included standard deviations for the counterfactual scenario and revised Table 4 and Table S1. The net C balance was also revised to 0.5 t C yr-1, correcting a calculation error.
R2.19) L. 225: To support the idea that gravels could enhance subsurface carbon losses, one could cite the study by Boodoo et al. (2019, https://doi.org/10.1002/hyp.13450), who demonstrated that streamwater downwelling account for a large contribution of the CO2 budget in a gravel bar of a pre-alpine temperate stream.	We thank the reviewer for the suggestion and have included this reference to the discussion on the fate of DIC infiltration on line 189.
R2.20) L. 226: I might have missed it, but where is the stabilized sediment TOC storage data (98 t ha-1) shown (Table / Figure)?	The stabilized TOC storage expressed per ha is not explicitly shown in any figure or table, but the total wetland storage in t C is shown in Fig. 3 as the PC store at the stable state timestep.
R2.21) L. 227-229: I think the reason for the differences in TOC storage and GHG fluxes between the Swiss system and previously studied systems elsewhere could be clarified a bit more by rephrasing L. 227-233, highlighting more clearly that the difference in TOC storage and GHG emissions could be are result of the Swiss study system having generally lower organic carbon content (%).	We thank the reviewer for this suggestion and we have now clarified the lower OC content as a potential explanation for both lower TOC storage and C emissions observed in the studied system (lines 271-273).
R2.22) L. 241-243: I get a bit confused by the focus on TOC accumulation in the context of beaver-mediated C	As discussed in the reply to comment R2.4, we focus on both the annual dynamics wherein DIC is an important C sink on short-term, but also decadal

sinks, while elsewhere (e.g. L. 30-31), the focus is on subsurface removal of inorganic carbon.	C sequestration in which TOC accumulation is a more important C sink. Further, previous studies have mainly quantified TOC accumulation (and not TIC) and this is therefore an obvious term for comparison with literature.
R2.23) L. 256: How could the historic channelization and clay lining of the study stream have affected the study results and how representative is the channelization and clay lining of the study stream for other streams in Switzerland that the results are extrapolated to?	We thank the reviewer for raising this point. The historic channelisation and partial clay lining of the study reach are indeed features of its land-use legacy; however, they do not form the basis of our extrapolation. The long-term projection and national upscaling focus on the hydrological and biogeochemical mechanisms enabled by beaver pond formation, namely enhanced surface–subsurface exchange, sediment trapping, and long-term carbon burial. These mechanisms have been individually found to arise wherever beavers re-establish ponded, low-gradient floodplain conditions (Larsen et al 2021), and are not specific to the Marthalen reach. While the legacy modifications may influence the absolute magnitude of infiltration at this particular site (by e.g. constraining vertical percolation in places), the overall pattern of discharge attenuation and large subsurface water losses in the overall budget is consistent with observations from other beaver-influenced systems across the world (see Chaubey & Ward, 2006, doi:10.3808/jei.200600073; Graham et al., 2022, https://doi.org/10.1002/hyp.14735; Puttock et al., 2017, https://doi.org/10.1016/j.scitotenv.2016.10.122). **We have clarified in SI Text (lines 286-289)** that our extrapolation is based on these general hydrological processes, not on the historic geomorphic configuration of this individual reach.
R2.24) L. 272: Were V-notches installed in the stream to assure robust discharge rating curves? IF not, how well-defined and stable were the stream banks were water level was monitored?	V-notch weirs were not installed, as beyond very small channels, they are unsuitable for most natural streams. The channel reaches are 7-10 m wide, with variable flows that would submerge or destabilise such structures. As detailed in our methods and supplementary information, we established stage–discharge rating curves from repeated NaCl tracer-dilution gauging under a wide range of flow conditions. This approach provides robust discharge estimates without altering channel hydraulics. Channel geometry and bank stability were assessed through repeated drone surveys and field inspections during the monitoring period, and no significant morphological change was observed at either gauging location. **We have added a short statement (lines 396-398) clarifying these points**.
R2.25) L. 289: Were DOC samples be stored dark?	DOC samples were stored refrigerated in amber bottles and then transported same day to the lab at EAWAG, and subsequently analysed the following day. **This is now clarified on line 337**.

R2.26) L. 288-292: How would you rate the risk of the water samples having changed in alkalinity (e.g. through CO2 degassing or metabolic activity in the samples) between sampling and analysis?	This risk is negligible since samples were filtered (0.22 µm), stored cool, airtight with no headspace, and analysed the following day (lines 336-339).
R2.27) L. 292: Which method was used to convert alkalinity to DIC? Would alkalinity contributions from organic acids be an issue? If so, alkalinity could be corrected for these contributions using DOC measurements (Lydersen et al. 2004, https://doi.org/10.1016/j.scitotenv.2003.12.005)	Determination of DIC is now reported on lines 339-342. Channel water pH (6.65 – 8.59) was close to or above neutral, which in combination with low to moderate DOC concentrations (2.96 – 9.01 mg/L) suggests that organic acid suppression of the acid neutralizing capacity were negligible in this system. This contrasts boreal systems with higher inputs of DOC rich in carboxylic acids (e.g., Buffam et al., 2007, https://doi.org/10.1029/2006JG000218). **We have clarified this on lines 342-344.**
R2.28) L. 299: Data interpretation would be aided by more details on the flux chambers used. What are the dimensions; were they ventilated; how did soil chambers differ from chambers deployed in open water; were the open water chambers drifting or static; did the soil chambers rest on air-tight collars?	Soil chambers (volume = 15.2 L, area = 4.3 dm2) had static collars permanently installed for the year, made air-tight with the chamber lid, water chambers (11.8 L, 7.5 dm2) were open and drifting, and tree chambers (0.5 L, 1.1 dm2) were static and air-tight. No chambers were ventilated since it was a closed measurement system. **This is now clarified on lines 354-357**.
R2.29) L. 310: What criteria were used when deciding on the depth of the soil core samples / why were different depths chosen for the wetland and forest sites?	In the wetland, the sampling depths were chosen to capture both pre-beaver (loose, unconsolidated) and post-beaver horizons (consolidated). In the forest sites, we sampled to capture the O and A horizons, where possible. **This is now clarified on lines 368-372**.
R2.30) L. 353: Can you provide a reference to motivate the choice of 5% mass subtraction?	We have chosen to remove the assumed 5% contribution of HCO3- to GPP and instead attribute all GPP to CO2 uptake. Larger, net autotrophic rivers have been shown to support substantial fractions of HCO3- uptake due to CO2 undersaturation (Aho et al., 2021; https://doi.org/10.1002/lol2.10195). Since we do not have any data on CO2 saturation or the carbonate buffering system, and because the studied stream is most likely net heterotrophic, we cannot motivate the 5% contribution from HCO3-. **The C budget have been revised accordingly (Table 1).**
R2.31) Supplementary Material L. 52: On what basis was assumed that deadwood volume was the same above and below the water surface?	This was an approximation based on visual observations of felled deadwood in the pond.
R2.32) Supplementary Material L. 159-169: How were the factors (e.g. net surface radiation, heat fluxes) included in the Priestley-Taylor energy balance model measured?	This is clarified in our response to comment R1.10.
R2.33) Supplementary Material L. 187: Why was the DOC assimilation pathway excluded? It is widely accepted that DOC can be assimilated in streams (e.g. Creed et al. 2015, http://dx.doi.org/10.1139/cjfas-2014-0400). DIC production and DOC assimilation can be linked through mineralization. Hence, if the authors	We thank the reviewer for this observation. This was a mistaken formulation because DOC production is in fact included (see Fig. S2). **We have therefore removed this sentence in SI (lines 210-211)**.

account for DIC production, the exclusion of DOC assimilation should be motivated.	
R2.34) Supplementary Material L. 217-219: Where does the estimate for Ra_above come from / how was it derived?	We thank the reviewer for pointing out the need for clarification. The aboveground autotrophic respiration term (Ra_above) was derived from standard partitioning relationships between total autotrophic respiration (Ra), gross primary production (GPP), and carbon-use efficiency (CUE) in temperate deciduous forests. After estimating Ra_root from soil CO_2 flux partitioning (Ra_root/Rh_soil = 0.66), we added an aboveground component consistent with literature-reported ratios of Ra_above ≈ 25–35 % of total Ra . This approach yields a total Ra = 6.2 t C ha^{-1} yr^{-1}, corresponding to Ra/GPP ≈ 0.42 and CUE ≈ 0.58, which fall within common ranges for mixed temperate broadleaf forests. **We have clarified this terminology in the revised SI Text 8 (lines 243–251)**.
R2.35) Supplementary Material L. 217-220: "aligning with forest flux tower data across temperate Europe" should be supported by a reference.	This statement is now supported by Luyssaert et al., 2007 (https://doi.org/10.1111/j.1365-2486.2007.01439.x).
R2.36) Supplementary Material L. 222: I question the reasoning that the stream can be assumed net heterotrophic when stream CO2 emissions exceed stream GPP. Even net-autotrophic streams could show this pattern when stream CO2 emissions are largely sourced from groundwater CO2 inputs rather than respiration in the stream (Duvert et al. 2018, https://doi.org/10.1038/s41561-018-0245-y).	We appreciate the reviewer's clarification and agree that high CO_2 evasion relative to aquatic GPP does not, in isolation, prove that the stream is biogeochemically net heterotrophic, since groundwater or soil CO_2 inputs can also sustain elevated emissions as the reviewer points out. **We have therefore re-phrased "net heterotrophic" to "a net source of CO_2" (SI line 252)**.
R2.37) Supplementary Material L. 232: "chamber bias, limited diel coverage" could be supported by literature, e.g. Vingiani et al. 2021, https://doi.org/10.5194/bg-18-1223-2021 and Attermeyer et al. 2021, https://doi.org/10.1038/s43247-021-00192-w).	Thank you for providing appropriate references, they have now been included to support the statements in SI (line 262).
R2.38) Supplementary Material L. 235: In small streams, turbulence and hence gas exchange is mainly driven by channel bed friction rather than wind action (Alin et al. 2011, https://doi.org/10.1029/2010JG001398).	We have revised this sentence by removing the mentioning of wind-driven turbulence, (SI line 265).
R2.39) Supplementary Material, Fig. S3: What does "Channel US (100m)" mean? I cannot find this location in the map in Fig. S4.	This location refers to the US location, where Q, stage, water sampling and GHG were measured. "(100 m)" refers to its upstream distance to the open wetland. For GHG, additional measurements were made 250 m upstream of the wetland. To clarify and keep the terms consistent with the map in Fig. S4, **we now denote the locations as "Channel US" and "Channel US_{250}", and further explain that Channel US is located 100 m upstream of the open wetland in SI Fig. S3 (lines 319-321).**
R2.40) Supplementary Material, Fig. S6: What does "upstream reference" mean? I cannot find this location in the map in Fig. S4.	Upstream reference refers to Q monitored 1 km upstream by the Zurich Civil Engineering Office. See our response to comment R1.6 for further details.

R2.41) Supplementary Material, Table S4: NSE = Nash-Sutcliffe efficiency is indicated in the table header, but I cannot find this metric in the table.	The mentioning of NSE has now been removed. It was not used for evaluating the fit.
R2.42) L. 37 "climate mitigations" should be "climate change mitigation"	The phrasing is now changed accordingly.
R2.43) Supplementary Material L. 45: I am not familiar with the term "photosynthetic activity radiation". Do you mean "photosynthetically active radiation"?	Yes, we mean photosynthetically active radiation, the phrasing is now changed accordingly.
R2.44) Supplementary Material, L. 300: "show" -> "shows"	The phrasing is now changed accordingly.

Reviewer 3 comments	**Authors' response** (Line numbers refer to revised version with track changes)
R3.1) On line 49, why would beaver activity enhance light availability? I would think the opposite, meaning that the enhanced riparian vegetation associated with beaver activity would create more shading.	Before beaver introduction, the study site was a forested floodplain, with mixed stands of deciduous and coniferous trees. When the reach was dammed by beavers, the stands were felled, resulting in a large increase in light availability. The succession of riparian vegetation has not reached equivalent canopy closure compared to preceding conditions with trees. **This is thoroughly described in the Main text (lines 305-311).**
R3.2) Fig 1. What is channel, US, MS, and DS specifically? What is channel the largest area? A site map in the main paper (not supplemental) would be useful so we could see what these actually are.	These are the upstream, midstream, downstream, and channel areal sections, used for estimating water surface area. **These sections are now shown as an inset in Fig. S4**. Due to the limited allowed number of figures and tables in the main text, we have chosen to highlight results rather than study site location.
R3.3) Line 155. It took me a bit to realize this analysis was based on a space for time substitution, I'd make that point clearly.	We appreciate that the reviewer raised this point and have clarified that the sediment comparisons rely on a space-for-time substitution (line 172).
R3.4) Line 272. The rating curves are extrapolated well beyond the bounds of your data. I would strongly advise against doing that. Also why do you use EC in the rating curve equation?	See our rationale for rating curve extrapolation, and sensitivity analysis in our response to comment R1.1. We revised the rating curves and chose to only include the traditional form: $Q = a(h-h_0)^b$. See our response to comment R1.4 for further details.
R3.5) Line 320. I don't follow the use of EC. Can you explain how this helps constrain the relationship?	See our response to comment R1.4 for further details.
R3.6) Line 327. Do you have a weather station? Where did you get the data to do this?	All weather data was collected from the nearby weather station in Marthalen (47.6244 ° N, 8.6533 ° E), sourced from Meteoswiss (https://www.meteoswiss.admin.ch/services-and-publications/applications/measurement-values-and-measuring-networks.html#param=messnetz-partner&table=false). **This is now clarified in SI (lines 182-187).**

Authors' responses - revision 2

Dear Editor and Reviewers,

We would like to thank you for taking the time and effort reviewing our manuscript and providing valuable comments on how to improve our manuscript entitled "Beavers can convert stream corridors to persistent carbon sinks".

The manuscript has been carefully revised (changes highlighted in red font color as track changes) and we have addressed all comments with detailed responses further below in this document.

We would also like to clarify the addition of Sarah Thurnheer to the author list. She was an integral part of the measurement and monitoring team, and also made valuable analysis and conceptual contributions. However, she was on extended leave at the time of finalising the manuscript and the initial submission process, and couldn't be contacted to confirm authorship and so was instead included in the acknowledgements. Following the initial submission, she returned from leave and was happy to confirm authorship. Her position in the author list reflects her contribution and no other changes in the order of authors were made.

We submit this revised version for your consideration for publication in *Communications Earth & Environment*.

Reviewer 1 comments	**Authors' responses** (Line numbers refer to revised version with track changes)
R1.1) Uncertainty quantification: I'm confused about how uncertainty was determined for the C budget. I would expect the uncertainty to be additive for the categories of Table 1, but it is not. For example, Total inputs in Table 1 were determined by summing DIC, DOC, POC, and CO2 fluxes, but adding the uncertainties of these would result in a higher value than what is shown. How was uncertainty quantified? I don't see methods regarding uncertainty.	We thank the reviewer for raising this point and welcome the opportunity to clarify how uncertainty was treated in the carbon budget. The assumption that uncertainties in Table 1 should be summed arithmetically is not generally valid for aggregated mass balances. Many flux components (e.g. DIC, DOC, POC, and CO_2) depend on the same discharge estimates, meaning that their uncertainties are not independent and their covariance is non-zero. In addition, the overall carbon balance includes both additive and subtractive terms (inputs, outputs, and storage changes), such that uncertainties cannot be combined by simple linear addition. Finally, temporal integration from daily to fortnightly and annual scales reduces random variability in several components. For these reasons, uncertainties in aggregated fluxes were propagated using variance-based error propagation, accounting explicitly for shared discharge terms and covariance among flux components, rather than by linear summation of individual uncertainties. Under this framework, it is statistically expected that the uncertainty of a summed quantity can be smaller than the arithmetic sum of the individual component uncertainties. We acknowledge that this logic was not described explicitly in the original Methods section, as it follows standard statistical practice for correlated variables, but we have now clarified this approach in the revised manuscript (lines 405-406) to avoid potential confusion.

R1.2) Related to this, what is the rationale for removing high flow days from the C flux calculation? High discharge days could be particularly important to solute and particulate transport components (both lateral and vertical) of the flux, so I'm not sure that simply removing those days from the analyses accurately captures the overall uncertainty associated with periods that exceed the stream rating curve.	We thank the reviewer for this comment and would like to clarify that high-flow days were not removed from the primary carbon flux calculations. All reported carbon budgets (e.g. Table 1 and associated figures) include the full hydrological year, including high-discharge periods. The exclusion of high-flow days was performed ***only as a sensitivity analysis***, with the specific aim of evaluating the influence of discharge estimates that required extrapolation beyond the measured range of the rating curve. This analysis was designed to test the robustness of the inferred carbon fluxes to uncertainty associated with rating-curve extrapolation, rather than to define the adopted methodology. Importantly, this sensitivity test demonstrates that even when days with extrapolated discharge estimates are excluded, the resulting annual carbon fluxes differ only modestly from those obtained using the full dataset. This confirms that the main conclusions of the study are not driven by high-flow extrapolation and that uncertainty associated with extreme discharge events does not dominate the annual carbon budget.
R1.3) Even when interpolating within a rating curve, there is always some uncertainty, and looking at the rating curves (figure s6), there is quite a bit of scatter in measurements at the upstream gauge. Was uncertainty in the rating quantified and considered in water and C fluxes?	We thank the reviewer for raising this point. Uncertainty in the discharge rating curves was explicitly quantified using leave-one-out cross-validation, with the coefficient of determination (R^2) reported in the original version of Supplementary Fig. S6, and explained in the caption. In the revised manuscript, we have additionally included root mean squared error (RMSE) estimates for the relationship between measured discharge and stage at both monitoring locations to further quantify rating-curve uncertainty. The scattered grey symbols shown in Supplementary Fig. S6 correspond to discharge measurements from an independent monitoring station located approximately 1 km upstream of the wetland. These data are shown for contextual comparison only and were not used in the load or flux calculations presented in the manuscript. Uncertainty in discharge estimates was therefore accounted for through the quantified rating-curve performance and was implicitly incorporated into the subsequent load modelling and carbon flux calculations. We also provide greater details on how discharge uncertainty propagates through the load estimation framework in our response to comment R1.5.
R1.4) The method for calculating biological removal has changed, now using the average discharge between the two gages. It's unclear to me why the average is used to calculate the mass, rather than using each gage's calculated discharge. Given substantial differences between upstream and downstream discharge, using an average adds another degree of imprecision / uncertainty.	We thank the reviewer for this comment and appreciate the opportunity to clarify the rationale for using the average discharge between the two gauges. The purpose of this calculation is to infer biogeochemical removal or production within the reach, based on concentration changes between upstream and downstream locations. This term is not intended to represent total load changes, which also include hydrological losses such as infiltration, but rather to isolate in-reach biogeochemical processing using concentration differences as a proxy.

	To express these concentration changes in terms of mass, we multiply them by the average discharge between upstream and downstream locations, which provides the most appropriate representation of the effective flow traversing the reach under conditions of net water loss. Using upstream discharge alone would systematically overestimate biogeochemical processing by scaling concentration changes with water that does not exit the reach, whereas using downstream discharge alone would underestimate processing by ignoring inflowing water that contributes to in-reach reactions. Using the reach-averaged discharge therefore reduces bias and does not introduce an additional degree of imprecision. Differences between upstream and downstream discharge are explicitly accounted for elsewhere in the mass balance through the hydrological and infiltration terms. The numerical impact of this refinement is small, but it improves the physical consistency of the biogeochemical partitioning.
R1.5) Generally, uncertainty in aqueous mass fluxes should be multiplicative, considering both uncertainty in the concentration and discharge. Was this done?	We agree with the reviewer that, in principle, uncertainty in aqueous mass fluxes arises from both concentration and discharge uncertainty. In the present study, this was accounted for through the structure of the load modelling and subsequent aggregation. First, dissolved carbon loads were estimated using WRTDS, which explicitly propagates uncertainty in concentration conditional on discharge, rather than treating concentration and discharge uncertainties as independent quantities. Second, uncertainty in discharge is shared across multiple flux components, as the same discharge estimates are used consistently throughout the mass balance. Treating discharge uncertainty as an independent term for each flux component would therefore result in double-counting the same source of uncertainty. For this reason, we did not explicitly re-multiply independent σQ and σC terms for each individual flux. Instead, discharge uncertainty was quantified at the rating-curve level and its influence on integrated fluxes was evaluated through sensitivity analyses. This approach accounts for multiplicative uncertainty while avoiding artificial inflation of uncertainty arising from redundant propagation of shared discharge errors.
R1.6) Related to the total long-term sequestration, the authors report a maximum value (assuming full in-filling). I appreciate the new text highlighting this choice and explaining the rationale. That this represents a maximum should also be clarified in the abstract. And at line 194 (in track-changed doc) you could note this is the estimated maximum active lifespan of the dam	We thank the reviewer for this suggestion, we have clarified that this value represents an upper estimate (rather than a true maximum) in abstract (lines 36-37) and discussion (line 193).

Reviewer 2 comments	**Authors' response** (Line numbers refer to revised version with track changes)
R2.1) In the reply to comment R2.28 the authors write that the "water chambers […] were open and drifting"	We thank the reviewer for highlighting this ambiguity. The floating chambers used in this study

and that "No chambers were ventilated since it was a closed measurement system". Please clarify what is meant by "open". To me "open" contradicts "closed". Also, I am not aware of open floating chambers for flux measurements in aquatic systems and would appreciate more details or a reference that describes the method. Finally, if chambers were drifting, how can their location be indicated as a point in Fig. S4b)? How long was the drifting transect and how consistent was it at different discharge conditions? Depending on discharge, chamber travel times may vary. This can impact flux estimates in two ways: if the transect is kept constant, the time over which concentration changes inside the chambers can be measured would vary which may bias rate of change estimates. Alternatively, if the travel time is kept constant, the length of the transect may differ, potentially integrating variable stream reaches. How did the authors deal with this trade-off?	were open at the bottom (water–air interface) but sealed to the atmosphere above, which is standard practice for floating flux chambers in aquatic systems. During measurements, the chamber–analyser assembly operated as a closed gas measurement system, meaning that the chamber headspace was isolated from the atmosphere and connected in a closed loop to the gas trace analyser. No chamber ventilation was applied. System integrity was continuously verified in real time by the analyser, with deviations from a linear flux increase immediately indicating any leakage or loss of closure. The chambers were floating but not freely drifting. During the short measurement intervals, chambers were lightly tethered to prevent downstream drifting while still allowing them to move vertically and laterally with the water surface and respond naturally to local surface flow and turbulence. This ensured that chambers sampled the same locations indicated in Fig. S4b, while avoiding artificial suppression of air–water gas exchange due to rigid fixation. As a result, chamber position, measurement duration, and sampled reach length were consistent across discharge conditions. Chamber deployment did not involve drifting transects, and therefore no trade-off between travel distance and measurement time arose. **We have clarified the chamber setup and terminology accordingly in the revised manuscript (line 334).**
I noticed that a co-author was added to the author list. For reasons of integrity, I would appreciate if the authors could briefly motivate this addition.	We thank the reviewer for raising this point. In accordance with the journal's authorship guidelines, the addition of a co-author was fully disclosed and justified in the cover letter submitted with Revision 1. The added author made a substantive contribution to the revised manuscript that meets the journal's criteria for authorship. For transparency, we have now also included an additional justification for this in the introductory section of our response to reviewers and editors (see above), to clarify the reasons behind this.
L. 102 Update the uncertainty estimate for annual C sink, to conform with the change in the abstract (L. 33).	We thank the reviewer for noticing this, we have corrected it accordingly.
L. 126: is it correct that the uncertainty for DIC reductions remained the same, despite change in the mean? Are all numbers throughout the manuscript updated correctly?	We thank the reviewer for spotting this. The slightly different uncertainty for the revised modelling of DIC and DOC were indeed not updated. We have corrected this in Table 1 and Fig. 4, and where loads are mentioned (lines 33, 101, 125, 185). We have also now double-checked all reported uncertainties to make sure that all reported values are correct.
L. 191: are you sure that DIC only sequesters as calcium carbonates and no other carbonate form? If not, consider just writing "carbonates".	We thank the reviewer for this suggestion. In stream environments, DIC predominately precipitates into calcium carbonates but we have re-

	phrased it as "carbonates" which also includes e.g., MgCO3 precipitation which may form under far rarer geochemical conditions.
L. 268: "a pathway can account for" -> "a pathway that can account for"	Revised accordingly.
L. 282: "resulting in a substantially lower sediment C storages" -> "resulting in substantially lower sediment C storages"	Revised accordingly.
L. 14: "which were also reflected baseflow" – something is missing here	Revised as "which were also reflected in baseflow"
L. 45: "photosynthetic active radiation" -> "photosynthetically active radiation"	Revised accordingly.
Text S5: Clarify units to the symbols introduced here.	We have now clarified all units in Text S5.
L. 217: "Time Discharge" Do you mean "Time, Discharge"?	Yes, this is now revised accordingly.
L. 319: For consistency, also explain "Channel DS".	Revised accordingly.

Review of Hallberg et al. Beaver can convert stream corridors to persistent carbon sinks

This is a very well written paper and I have very minor suggestions.

On line 49, why would beaver activity enhance light availability? I would think the opposite, meaning that the enhanced riparian vegetation associated with beaver activity would create more shading.

Fig 1. What is channel, US, MS, and DS specifically? What is channel the largest area? A site map in the main paper (not supplemental) would be useful so we could see what these actually are.

Line 155. It took me a bit to realize this analysis was based on a space for time substitution, I'd make that point clearly.

Line 272. The rating curves are extrapolated well beyond the bounds of your data. I would strongly advise against doing that. Also why do you use EC in the rating curve equation?

Line 320. I don't follow the use of EC. Can you explain how this helps constrain the relationship?

Line 327. Do you have a weather station? Where did you get the data to do this?