Response to Wil Wollheim

WW.1*: This study addresses concentration vs. discharge relationships in streams and rivers. The authors hypothesize that instream uptake will result in "bent" logC vs. logQ relationships, because net instream removal will cause lower concentrations than expected given loadings at low flows. They apply a new metric (curvature) to quantify this effect using both field data and modeling results from 13 different river networks that range in size and characteristics. They also do an extensive analysis across parameter space to understand which network characteristics have most influence on curvature and network scale removal. They find that in stream uptake can indeed lead to more bent logC-logQ relationships (because the curvature parameter is more negative), and that channel hydraulics (the width and depth vs. Q relationships) have the strongest influence. Uptake velocity (the biological parameter) seems to have less influence, which was surprising. They suggest that the curvature parameter could be used to quantify network scale removal using only the C vs. Q information, adding a potentially useful tool to understand network scale dynamics.*

I think this is overall an interesting analysis and potentially a very useful approach for quantifying network scale uptake. There are a few things to consider further, emphasize or discuss, and a couple of things that would increase the understandability.

We thank the reviewer Wil Wollheim for their useful comments and their interest in our work. We address all the reviewer comments (italic) one by one below with responses in normal font.

WW.2: *The result depends strongly on the assumption that the relationship between C and Q for loading from the landscape remains linear across seasons (i.e. the parameter b is constant). One of the difficulties getting at broad scale aquatic function is isolating landscape inputs and aquatic processes (inherent in any river network scale analysis). The constant "b" assumption is what allows inference that the bent C vs. Q relationship results from network-scale nutrient retention. Given that this analysis uses C and Q measured across seasons (as opposed to individual storm events), with seasonality correlated with flow conditions, how likely is that? That is, the loading C vs. Q relationship will differ between summer and winter, with the former tending to have lower C (e.g. due to higher riparian uptake). Would that also result in bent curves? I think this is an important consideration, worthy of some discussion.*

We thank the reviewer for this comment. It is correct that the conclusions presented in this paper depend on the assumption of a linear land to stream loading vs Q with a slope that remains constant throughout the seasons. If the loading C vs. Q would differ seasonally, this could indeed result in bent loading curves which would make it hard to attribute observed bending at the catchment outlet to in-stream processes. There are however indications that for land to stream NO3- loading, b can be constant (Basu et al., 2011). This also connects to the idea of nitrate being mainly transport limited and not source limited especially in catchments with agriculture. There is a clear indication of this in our well studied test catchment Selke on the basis of high-frequency nitrate concentration analysis: Storm event-CQ slopes rarely changed over the seasons with similar mobilization patterns in summer and winter (Winter et. 2021). We will incorporate those new findings into the text. Nevertheless, we cannot exclude alternative loading patterns when detecting bent C-Q relationships in observed data (esp. in low-nitrate environments with potential source limitations), which we will mention and discuss explicitly in the revised manuscript. In our work we explore if in-stream removal can result in bent log(C)-log(Q) relationships at the catchment outlet and find that it can. We therefore argue that the assumption of a constant 'b' is feasible for use in the explorative modelling approach presented in this paper but surely deserves attention in future research.

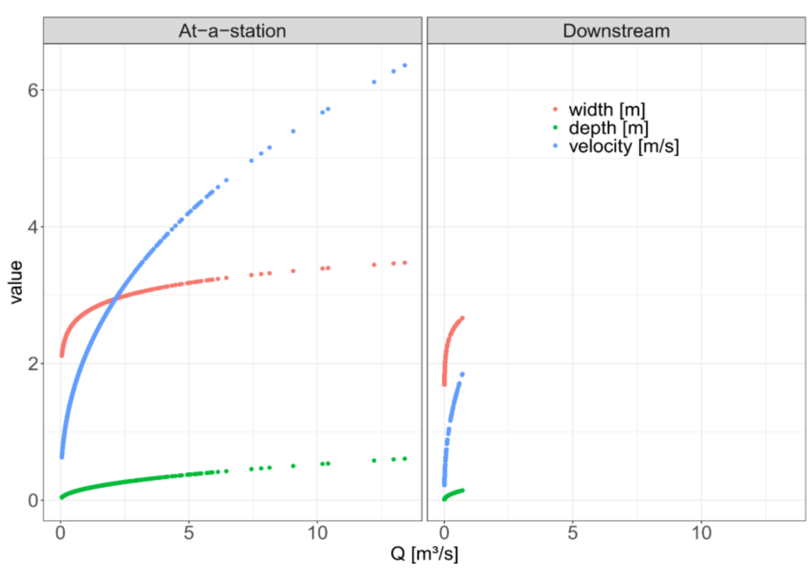
WW.3: *I agree with one of the findings that the hydraulic dimensions are among the dominant factors when considering network scale removal. However, in this analysis (if I understand right), a single hydraulic equation is used for width (and depth), i.e. w = Kw * Q ^ aw, and a single aw is applied over both space and time. However, the hydraulics of rivers are such that the change in width with changing flow at any given site (due to storms) differs from the change in mean flow in the downstream direction (at-a-site vs. downstream hydraulic relationships). Typically the at-a-site change in w is much lower (~0.1) than in the downstream direction (~0.5) (See Knighton 1998. Fluvial forms and processes: a new perspective). It appears the best calibrated fit for one of the watersheds (Table C1) was 0.09, closer to the typical at-a-site relationship. This will greatly affect the pattern of removal within the network (small vs. large rivers) as well as with changing flow. Note that the constant (Kw) is the width (m) when Q = 1m3/s. So if you have a low aw, that means large rivers stay relatively narrower and small rivers stay relatively wide (since width doesn't change much). The calibrated aw is closer to the at-a-site change (where increasing flow is accommodated mostly by changes in velocity) than the downstream change (where increasing flow is accommodated mostly by changes in width). This may explain why uptake velocity is relatively unimportant (which I was surprised by), and also why water velocity comes out as so important. It would be worth confirming whether the modeled widths match observations, and* *reporting the mean width of small headwater rivers (<5km2) and larger rivers (> ~400km2) to evaluate if they are reasonable.*

We thank the reviewer for these remarks. Width (w) and depth (d) have distinct parameters for their hydraulic equations, i.e. *w = Kw * Q ^ aw* (Eq.2.1) and *d = Kd * Q ^ ad* (Eq.2.2), respectively. Note that both channel hydraulic parameters; w and d, depend on discharge (Q) and therefore they vary depending on flow conditions. However the reviewer is right that each equation is applied over both space and time. This will be made clear in the revised manuscript when these equations are first introduced.

The effect of a low and high aw leading to constant or varying stream widths respectively the reviewer describes is also illustrated in the conceptual Fig. B5. This figure is currently briefly referenced in the manuscript, however based on the reviewer's comments we would add a sentence explaining this more in the revised manuscript.

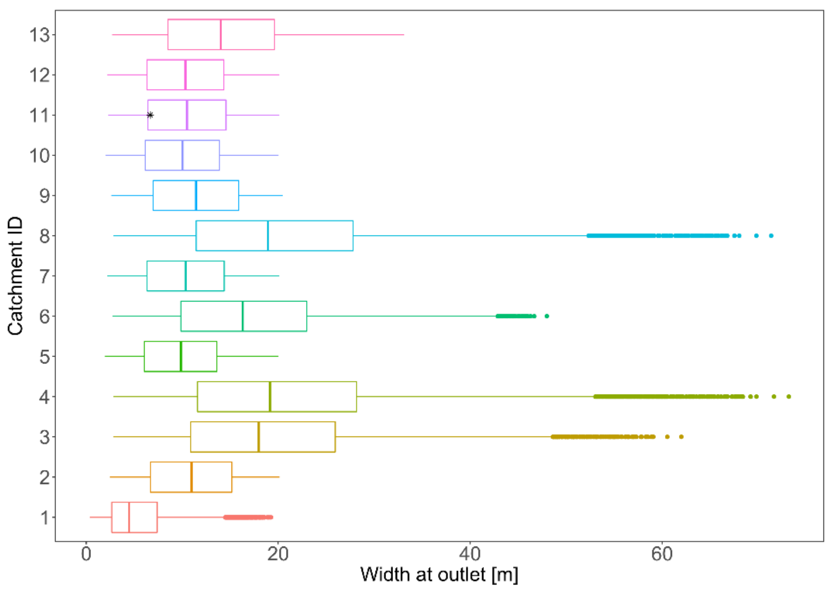
In the 'At-a-station' panel of Fig. R2 below we evaluate the changes in velocity, depth and width with Q for a grid cell in the middle of the Selke network (point B in Fig. 3 in the manuscript). The 'Downstream' panel - that considers all the network grid cells - shows the channel characteristics width, depth and velocity for a time t, with a Q of 0.70 $m^3 s^{-1}$ at the outlet. The values for the parameters aw, Kw, ad and Kd are the same for both scenarios (Table C1).

Figure R2 shows a larger variability of Q in the at-a-station panel with higher Q's that are indeed accommodated mostly by the increasing velocity. As the *Curvature* metric captures the shape of an 'At-a-station' log(C) vs log(Q) relationship that is driven by the Q variability, it might explain why the channel



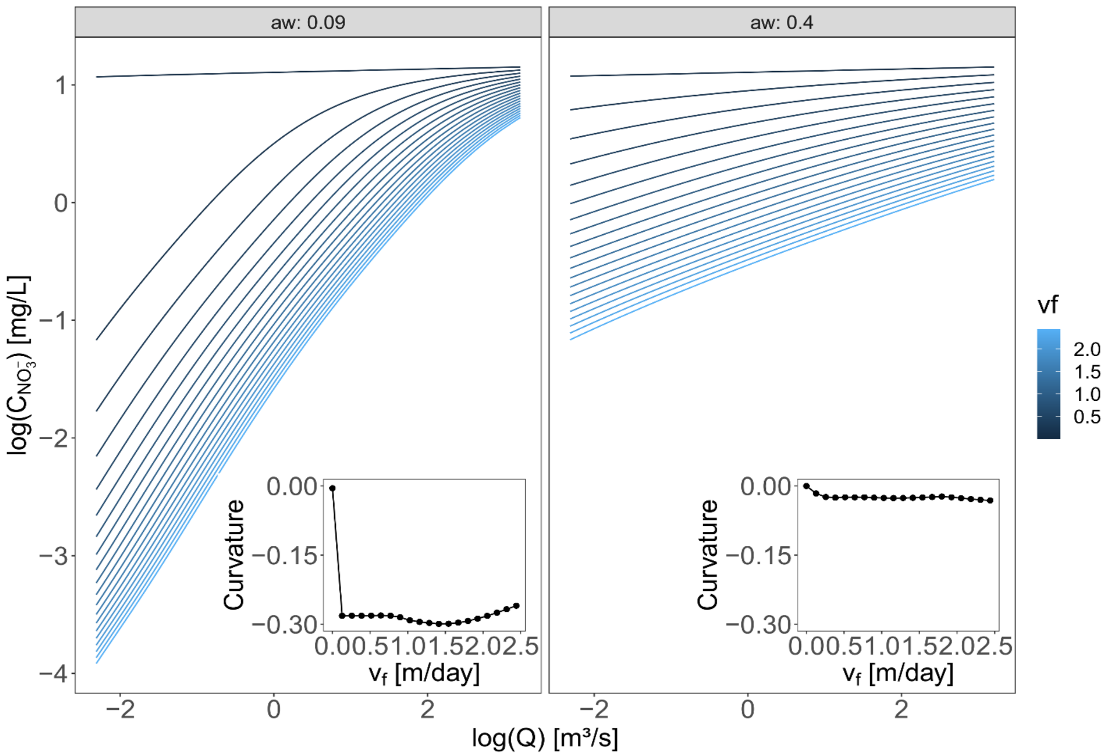
characteristics come out as more important for shaping this signal, compared to the uptake velocity. Therefore, we will add this consideration to the section discussing the PAWN sensitivity analysis.

In the figure below (R3) we show the median widths at the catchment outlet for each of the >11000 model parameter combinations used in the Monte Carlo simulation. These parameter combinations were chosen randomly within some set physical boundaries (l.233-237; Table 1) and will therefore cover realistic as well as unrealistic stream channel widths at the catchment outlet as can be seen in Fig. R3. The width for the selke Meisdorf was reported by Rode et al., 2016 and is indicated in the boxplot with an asterisk. This point falls well within the simulated width range for this catchment. Following this comment, we will add a sentence in the revised manuscript stating that the modelled widths are to a large degree reasonable. The figure with the width at the outlet will be added to the supporting information.



WW.4: *It is interesting and a bit surprising that vf had a relatively small impact. The authors state that if vf = 0, there is no bending (conservative) – and of course I agree. But it seems that a low vf would then result in only slight bending, which will only increase as vf increases. Does this pattern not occur? The choice of vf in the paper is appropriate for denitrification, but it is on the low side total N uptake (assimilation) which could be 5-10x higher than for denitrification (e.g. Mulholland et al. 2008 found denitrification was ~15% of gross nitrate uptake). Net assimilation may also be important in watersheds at certain times, particularly during lower flow summers (storing N over medium time scales, or transforming to PON or DON). Might this ever be a factor in the watershed considered. Could the Monte Carlo analysis address this possibility by using a higher Vf to determine at what point vf dominates the bending?*

We thank the reviewer for their comment. In Fig. R4 below we show C-Q relationships for increasing uptake velocities vf resulting from network model simulations of the Selke Meisdorf catchment. As an example, two different aw are displayed with the other parameter values as stated in Table C1. We see that although concentrations gradually decrease for increasing vf, *Curvature* remains rather constant for these example simulations apart from values close to zero for very small vf. Because this figure can help to understand the dynamics between vf, aw and *Curvature* we will add it in the revised manuscript.



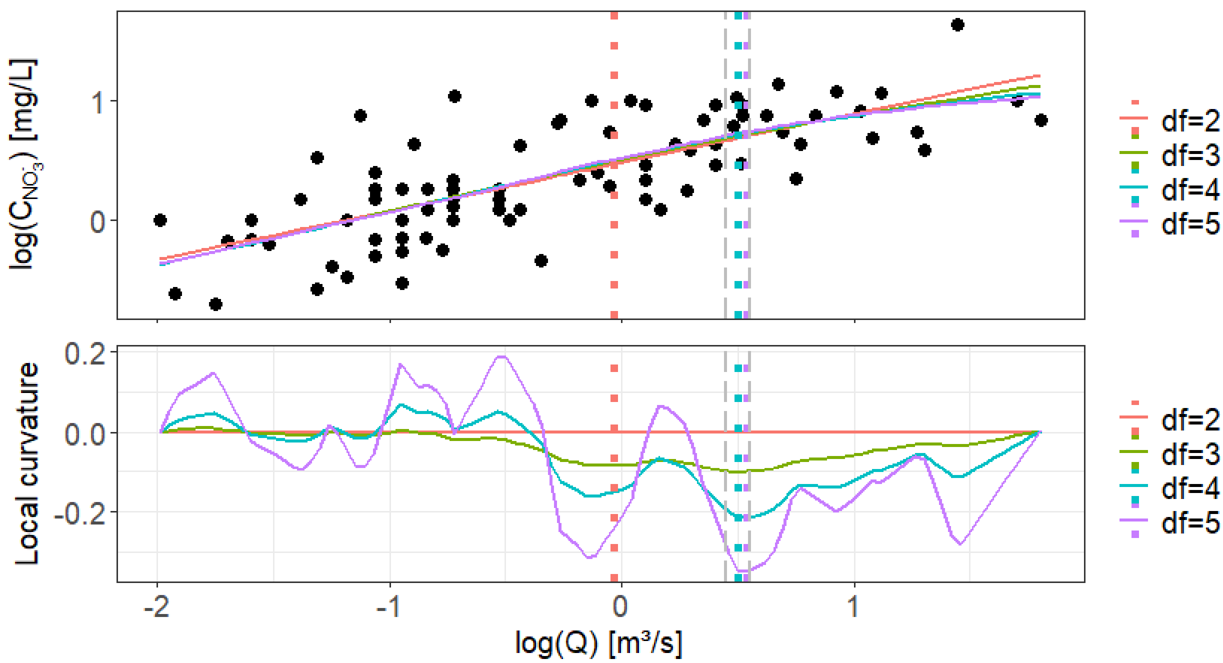
We assigned the values for vf based on a database compiled by Marcé et al., 2018. Here vf was collected from 83 published studies for >260 rivers (1-3rd order mainly). The studies used addition experiments that were typically conducted under base flows or low flow conditions and calculated vf based on the nutrient spiraling equations accounting for biotic (assimilatory and dissimilatory) and abiotic uptake (Stream solute workshop, 1990). However the range of 10^{-4} to 0.25 m/day for vf we finally selected based on a subset of this dataset is indeed low. The values for the median vf, 1st and 3rd quartile respectively 1.30, 0.47 and 5.76 m/day, taking into account the entire database. These values might exceed the ‘real’ in-situ vf as they were mostly obtained from nutrient additions (Hensley et al., 2014; Mulholland and Tank, 2002). As an example Fig. R4 above shows how a wider vf range (between 10^{-4} to 2.5 m/day) does not affect *Curvature* much. The higher vf values clearly lead to lower concentrations that are sometimes below the limit of quantification in real-world data (which we don't observe at the Selke and the other test catchments). This leads us to conclude that in the Monte Carlo analysis we can focus on the rather lower range of uptake where most changes in bending happens. We would therefore include Fig. R4 in the revised manuscript.

WW.5: *Given that these C vs. Q patterns are based on samples collected over the year there is also the confounding effect of temperature on biological activity. Denitrification is often represented with Q10 = 2, so winter (cold temperature) reactivity could be much lower. I know this was not part of the analysis, but given the use of C collected over seasons, it seems important to factor in somehow, at least in the discussion. The temperature effect, correlated with Q, would cause a more rapid shift to saturation with increasing flow (since most of flow change is likely seasonally driven, given the sampling regime). Should discuss whether this factor is potentially important, why or why not?*

This is an interesting point, thank you. We agree that this factor can be potentially important and surely will also interact with the strong seasonality in discharge and flow velocity. We will add a brief discussion (Section 3.4) on this topic in the revised manuscript. Additionally, we will alter the section in the methods where we state our basic assumptions on the stationarity on geometry and reaction parameters in time and space to be more clear."

WW.6: *I also had a question about how "bentness" (=curvature) is represented in Figure B1, discussed, and demonstrated. It would help me a lot (and I assume other readers) if some of the empirical patterns of log C vs. log Q were shown. Examples for different values of the curvature parameter (end members, the median, and 0) would be helpful. Especially since one of the conclusions is about the utility of these low frequency empirical data sets (L641) and given that much of the recent literature has used high frequency data to get at C vs. Q relationships.*

We thank the reviewer for this comment and suggest to replace Fig. B1 by the Fig. R3 below, where the iterative fitting of the Selke data is shown in the upper panel and the corresponding local curvature in the lower panel. The value of the *Curvature* metric results from the region of the largest instantaneous change.



WW.7: *Also, I would consider some of the wording regarding "less curvature". I initially assumed that meant straighter. But in fact, "less curvature" meant a more negative curvature parameter, which is actually more bent. It took me a while to get straight.*

We thank the reviewer for pointing this out and agree that the current wording with 'small' and 'large' *Curvature* might be confusing. We therefore decided to use the terms 'low' and 'high' *Curvature* to describe respective more and less bent simulated log(C)-log(Q) relationships. This wording will be applied consistently throughout the text.

WW.8: *In conceptual figure B1, I think that the bentness as I understand it should show a straight line at high flow parallel to the curvature equal 0 line, but bending down as flows decline. If the dynamic is saturation, it should approach the slope set by the loading function. Would it make sense to modify Figure B1 to reflect that (if indeed correct)? I also think some empirical patterns, showing what the curvature parameters is, would also help increase the intuitiveness of the results. A demonstration of how curvature is fit would be good in the appendix (to make section 2.1 easier to understand).*

Thank you for this comment. A demonstration of the fitting of Curvature is shown in the response of comment WW.6. Also, we agree that adding the slope of the loading function would improve the manuscript and would therefore propose altering Fig. R4 to mention that vf = 0 is also the land to stream loading linear log(C)-log(Q) slope.

WW.9: *I appreciated the test of the model predictions against observations in the Selke watershed. The correspondence looks excellent! But I did not quite understand how the seasonality of concentration emerges give the low removal proportions (I assume this is network scale removal by the entire network), and the fact the loading C vs. Q relationship is flat (b = 0.014). It seems that loading is fairly constant and removal in Figure 2a is very small (<5% at all times). So what causes the large drop during summer? I would add another line that represents the export assuming conservative mixing (Vf = 0). Also, in Figure 3, add the observed C vs. Q relationship.*

We thank the reviewer for these comments. Here we show the cumulative removal percentage when taking into account all the incoming and removed loads in the network and show the corresponding evolution of this network wide percentage of removal over time. Part of the seasonality of the concentration is driven by the streamflow as the input load L = a*Q^b+1, so in the case of b= 0.014, higher Q will still result in higher incoming loads everywhere in the river network. That the removal overall in the catchment is fairly small, does not mean that there are no locations and times in the network were removal percentages are high (see headwaters in Fig. 3). We agree that it would be helpful to add the conservative mixing scenario (vf=0) to Fig. 2 and will adjust this in the revised manuscript. This will also allow to differentiate effect of uptake compared to seasonality induced by loading patterns on the concentrations observed at the outlet for this Selke Meisdorf example.

For the comment regarding Fig.3 we would like to refer the reviewer to our response on WW. 34.

WW.10: *What is driving the runoff (water transfer from land to water) variability over time in each watershed?*

Runoff is driven by the variability in meteorological forcings (e.g., P, T, ...); which afterwards is modulated by land-surface properties (e.g., terrain, soil, vegetation, and geological attributes). In the context of this study, the water land to stream transfer over time is dictated by the discharge time series at the catchment outlet. The observed discharge daily discharge variability at the outlet is distributed to the individual stream sections according to their upstream area with the assumption that the discharge [mm/d] on each day is spatially homogeneous. We will mention this explicitly in l.181 of the methods section in the revised manuscript and in Sect 3.1 when discussing the results of the Selke validation example.

WW.11: *While the conclusions provide clear and useful summaries, I found the final conclusion seemed underwhelming. I think more of the implications of these findings could be emphasized, and why they would be useful. Tie back to the big picture of C vs. Q, role of network removal, and management.*

Thank you for these useful suggestions. We will revise the final conclusion emphasizing the implications of our findings in context of C-Q relationships and from the view-points of network removal and management aspects.

WW.12: *Line 116: should read "log" C-Q*

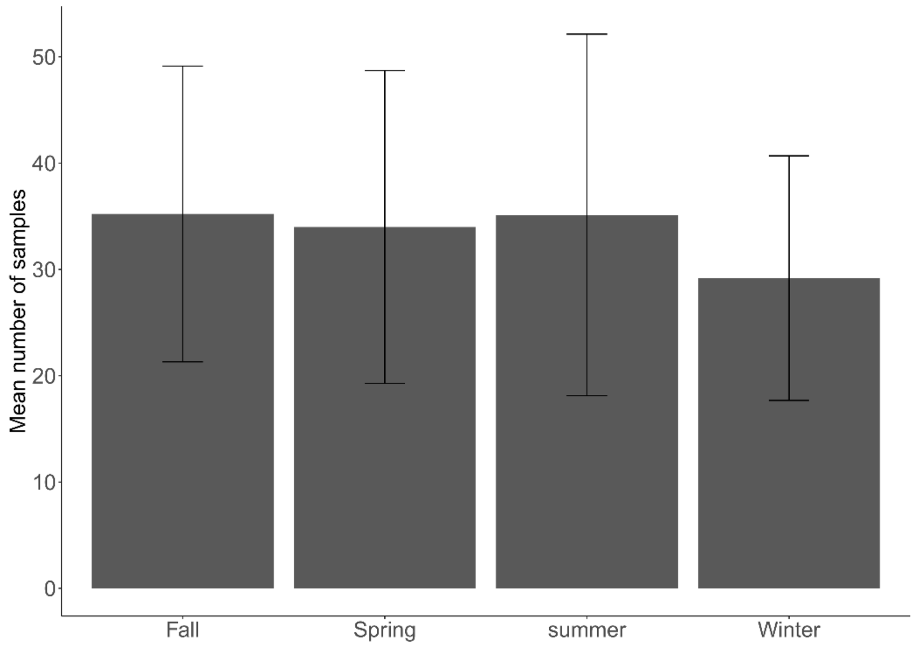
We thank the reviewer for this remark. We will adjust the notation to read log(C)-log(Q) throughout the lines 116-123.

WW.13: *L 135. Where does the value "402" come from?*

The maximum number of coupled C and Q samples within one station is 402. This will be specified in the revised manuscript.

WW.14: *L137. Meaning that at least 10% of the observations come from every season? Still, less sampled seasons could be underrepresented. What seasons were most samples collected?*

We will add a new figure to the supplementary information of the revised manuscript (Fig. R5; see also below) displaying the mean number of observations and the standard deviation for each season. In the fall, spring and summer there were on average 35 samples collected per station while in the winter the average number of collected samples was 30. We argue there was no underrepresentation of a given season, which we will mention when presenting the French data.



WW.15: *L182. What does this parameter definition mean?*

The ratio of a_d to a_w corresponds to a parameter r [-]∈ R^+which prescribes the cross section geometry relation such that a triangular channel cross section is represented by r = 1, a parabolic channel cross section by r = 2 and channel cross sections with progressively flatter bottoms and steeper banks by increasing values of r (Dingman, 2007). The width-discharge relation in Eq. (2.1) is conceptually illustrated in Fig. B6 for two sets of a_w and K_w (l.185-188). To make it clearer, we will change the order of these sentences so that they come directly after the definition of the parameters in l. 182.

WW.16: *L196. Why does the equation have "b+1" rather than just b?*

Because the model is mass balance based we calculate with load (L) rather than concentration (C). As $L = CQ$ and $C = cQ^b \leftrightarrow L = cQ^{b+1}$

WW.17: *L238. Explain what PAWN stands for when first introduced.*

PAWN is derived from the authors names (Pianosi and Wagener) - who introduced this method - and as such it does not have any meaning. Thus, we do not report what PAWN stands for, since it is not relevant to the analyses.

WW.18: *Table 1. Kw is not unitless, it has units of the dimension. (it is equivalent to the width at 1 m3/s or whatever units of Q you use). Same with Kd.*

We thank the reviewer for pointing this out. The units of K_w and K_d will be adjusted to $[L^{1-3*a_w}.T^{-a_w}]$ and $[L^{1-3*a_d}.T^{-a_d}]$ respectively in Table 1 (Dingman, 2007).

WW.19: *Table 2. Please add the watershed scale runoff (mm/d) to this table. It will allow comparison of how the different watersheds function. Q at the outlet is then just that times the watershed area. Is median Q the median of all river reaches, or the median at the mouth over time?*

We thank the reviewer for this comment. The watershed scale runoff will be added to Table 2. The median Q is taken as the median discharge at the basin mouth over time. This will be specified in l. 263.

WW.20: *Table 3. Why such small ranges for some of these parameter but not others?*

We distributed the non-missing simulation data over 20 percentiles and selected the percentiles corresponding to low, medium and high values (according to literature). Thus each class can have a different range; however for one variable the number of 'simulation data points' in each class is the same. This will be clarified in the revised manuscript.

WW.21: *Figure B4. Define the variables*

We thank the reviewer for this comment. We will add the definition of each variable to the caption of Fig. B5 : "Here, the flow length through a grid cell i is l_i [L], w_i [L] and d_i [L] are the respective width and average depth of the reach and P_i [L] is the corresponding stream channel wetted perimeter. The uptake velocity is denoted as v_f. The local discharge Q_i [L^3 T^{-1}] consists of upstream incoming discharge Q_{i-1} [L^3 T^{-1}] and land to stream runoff Q_{ls} [L^3 T^{-1}]. Similarly, the local load L [M T^{-1}] consists of upstream incoming load $L_{in.up}$ [M T^{-1}] and the land to stream load $L_{in.ls}$ [M T^{-1}], where $L_{in.}$ = $L_{in.up}$ +$L_{in.ls}$. Finally, the local load removed is denoted as $L_{r,i}$ [M T^{-1}]"

WW.22: *Table C1. The parameters for the Selke catchment suggests that inputs of NO3 are relatively chemostatic (fairly low "b"). This would lead to C vs. Q flattening out at high flows. It may be helpful to include a "conservative tracer" scenario to each of the catchments, which will be based on the C vs. Q of loading from the landscape. The divergence (always lower), will indicate bentness. Consider representing Figure B1 in this way.*

We thank the reviewer for this helpful comment. Comparing the divergence between the conservative tracer scenario and the resulting log(C)-log(Q) curve would indeed be another way to indicate 'bentness' and the effect of instream uptake. In this paper however 'bentness' is quantified with the *Curvature* metric as no conservative tracer scenario is needed to interpret it. Nevertheless, we agree that indicating the conservative tracer scenario is useful in this explorative approach at least as an example in the Selke Meisdorf case. We refer the reviewer to the response to WW.9 for more details.

WW.23: *L295. Explain what KSmax means in words and whether high values are better or worse.*

This information was indeed missing here. We will add a sentence in the revised manuscript so the section would read: "In this study, we applied Eq. (7) using $n_i = 10$ conditioning intervals for each input parameter and used the maximum KS value, KS_{max} , as a summary statistic, which is appropriate for screening non-influential input parameters. KS_{max} ranges between 0 to 1 and the higher the value the higher the influence of the parameter on the output. In particular, a value of 0 indicates that the parameter does not have any effect on the model output."

WW.24: *L312. I am not sure that the catchment wide Da adds much to the overall analysis, and could be dropped.*

We thank the reviewer. The catchment wide Da was included to check if the simulated values distribute around 1 and help the reader to understand if our scenarios rather create overall more reaction or more transport driven cases. This was mainly motivated by the surprisingly low impact of the vf and the prominent role of velocities on the uptake and bending. We thus wanted to explore if all our catchments are just transport driven which is not the case. We prefer to keep the Da number here but will be explaining our intention with Da more elaborately in the revised text.

WW.25: *L356-358 and Figure 3. The comparison of % removed and absolute amount removed within each grid cell is interesting and useful, but not the complete story. There are many more medium and large river grid cells than headwater grid cells along any nutrient loads flow path. So cumulative removal by larger rivers likely approaches or maybe even surpassed that of cumulative removal by the headwaters, particularly at high flows (see Wollheim et al. 2006 and 2018). Consider adding that metric as well.*

Interesting point. We will add an inset to Fig. 3 that shows with boxplots the cumulative removal and cumulative incoming load for each of the grid cells within a certain Strahler stream order. We will also calculate the total cumulative N removal for each stream order and describe our findings in the text of the revised manuscript.

WW.26:*L384. Wouldn't median over represent low flow periods, rather than total fluxes (since most flows are low, storm flows relatively infrequent).*

That is a good point. With the median we focus not on the total removed load but on how frequent is a certain removal efficiency. This is what we already state in the preceding lines l.380-384. The alternative (removal based on total, cumulative fluxes) would heavily weight single large discharge events. We will add and elaborate the text on these aspects in the revised manuscript.

WW.27: *L416. It is not clear in the table of watershed characteristics why C1 and C10 have so much higher Lr.perc than the others. What causes the large variability among watersheds? Cumulative percent removal should always increase with watershed size. Are you reporting the median within a watershed? I think cumulative removal would be a better metric.*

We thank the reviewer for this comment. It is indeed true that we report the median removal in each river network (see response to comment WW.26). We already discuss some possible reasons for the higher efficiency in C1 and C10 in l.47-l.483 in the manuscript: "The percentage load removed, $L_{r.perc}$, is notably lower catchments with high Q – like 3, 4 and 8 (Table 4) which follows the narrative in Sect. 3.1 that uptake efficiency decreases with increasing Q because of increasing loads to the system (Wollheim et al., 2018; Mulholland et al., 2008) that also result in less efficient uptake within the reactive surface area (Peterson et al., 2001; Hensley et al., 2014). The high $L_{r.perc}$ in small catchments 1 and 10 could then be attributed to their low Q, however why the small catchment 5 does not have similar uptake performance is less clear." Nevertheless we agree it would be interesting to add the cumulate removal as well as the 50th percentile to Table 4 in the revised manuscript. Also we will revise this section to include the discussion of the effect of the runoff [mm/day], added in Table 2 (WW. 19).

WW.28: *L458. I have a hard time understanding why catchments results are distinct, when all the parameters are the same. L461 says local loading and uptake differed, but what basis, since all the parameters are the same! Some of the other explanations in this paragraph are similarly unclear. It seems the model predictions can be summarized to see if the statements are true.*

We would like to thank the reviewer for these comments. We used a fixed set of 11107 parameter combinations (l.232) in each of the study catchments. During a model simulation, one of these parameter combinations was applied and the parameters are kept constant in space and in time for simplicity. However, all catchments do have an individual network structure and individual discharge conditions that largely explain the spatial differences between the catchments. For example, the channel hydraulic variables (w and d; Eq. 2) can vary significantly depending on the discharge values in each study catchment (Q). We will clarify this in the methods section and restate it in the discussion.

WW.29: *L470. Is Q higher is some catchments because they are stormier (runoff vs. Q focus). Q integrates watershed size and storminess.*

See WW.19, we will add runoff [mm/d] to Table 2. Both, absolute Q and Q variability are mainly the response to the climatic drivers precipitation and evapotranspiration. In Germany, the climatic drivers follow an East to West gradient and depend on the altitude. The storminess is captured by the CV of Q that is reported in Table 2 and not correlated to absolute [m3/s] and specific [mm/d] discharge. We would refer to Table 2 in the text at this point.

WW.30: *L473. Is the runoff the same in the small catchments as the large?*

We refer the reviewer to the response on comment WW.29.

WW.31: *L476. Important point! What about flow regime (frequency of different runoff events over time). Are they similar among catchments?*

This is nicely captured by CV (Q) which integrates the frequency of runoff events and the differences in recession constant (so the catchments “flashiness” in response to rainfall) (Botter et al., 2013). We will add some statements on the Q and CV (Q) differences in the catchment description to make that clear from the beginning on.

WW.32: *L511. Replace “Curvature” with “Curvature Parameter” because less curvature is more bent.*

We are not sure we understand the reviewer’s comment here.

WW.33: *L641. I think to make this conclusion, you need to include more empirical relationships.*

We thank the reviewer for pointing this out. In the revised manuscript we will stress that these conclusions hold true for the data set and parameter range that we used for our analysis. As we mentioned in the paper outlook, enlarging this approach to more catchments and gather more empirical evidence to explore this further would have to be done in future. We will stress in l.641of the revised manuscript that our results suggest this.

WW.34: *Figure 3. Add the observed C vs. Q (fitted relationship, with their R2) as a model test to this figure. Important to know how close predictions come to observation*

We thank the reviewer for this comment. In Fig. 2 simulated and observed NO_3^- concentrations are shown at the Selke Meisdorf station with the goodness-of-fit metrics NSE and pbias. Because adding those fits in Fig. 3 as well would make the figure harder to read and repetitive we would not follow your suggestion here. Nevertheless, in the revised manuscript we will refer to Fig. 2 at this point.

WW.35: *Figure 5. Nice summary of all the correlations, with color coding.*

We thank the reviewer for this comment.

WW.36: *Figure 7. I found this figure to be impossible to interpret. I think more explanation in caption needed. What are the histograms? What are the decision values? Why do variables show up multiple times? Not sure how useful the Cart analysis is based on the discussion here.*

Thank you. We will extend the captions in this Figure and some a guideline in the Figure to explain the CART concept better. We show the CART analysis because it is a visual guide through the multivariate space. Simple correlations do not capture parameter interactions and we therefore argue that CART is a valid tool here. The variables that appear in the internal nodes of the tree can be interpreted as being influential with respect to the dependent variables considered (here Lr, Da and vf). Variables can show up multiple times in the tree, revealing interactions between variables for different values of that variable. CART has been applied before in the context of sensitivity analysis, e.g. in Almeida et al. (2017) to identify the controls of landslides and in Singh et al. (2014) to identify the controls of runoff. In the manuscript, we will clarify the objectives of the CART analysis and link them to the previous analyses (PAWN, correlation).

REFERENCES

Almeida, S., Ann Holcombe, E., Pianosi, F. and Wagener, T.: Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change, Nat. Hazards Earth Syst. Sci., 17(2), 225–241, doi:10.5194/nhess-17-225-2017, 2017.

Singh, R., Wagener, T., Crane, R., Mann, M. E. and Ning, L.: A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds: Application to a watershed in Pennsylvania, USA, Water Resour. Res., 50(4), 3409– 3427, doi:10.1002/2013WR014988, 2014.

Botter, G., Basso, S., Rodriguez-Iturbe, I., & Rinaldo, A. (2013). Resilience of river flow regimes. Proceedings of the National Academy of Sciences of the United States of America, 110(32), 12925-12930. doi:10.1073/pnas.1311920110