

Revision notes of the manuscript

Kindly note that in the revision notes, comments from the reviewers are marked with “**Comment**”, while our responses are marked with “**Response**”.

Responses to the comments from reviewers:

Reviewer 1

Comment: I enjoyed reading "Seasonal variation and release of soluble reactive phosphorus in an agricultural upland headwater in central Germany" by Rode et al. The authors investigated potential delivery flow paths for P during various baseflow conditions: groundwater (GW) discharge, hyporheic exchange through stream sediments, and soil drainage. This case study is valuable for understanding how P in agricultural catchments is hydrologically delivered to streams and buffered along the way.

I think the manuscript needs at least moderate revision to make the story more effective and clearer. Additionally, I am not so convinced that the GW P delivered to the stream is "geogenic" nor that sediment porewaters are not adding any SRP to the stream -- these points need more nuance.

Response: We thank reviewer #1 for the positive comment. Regarding the concerns, please see our detailed point-to-point responses below.

Comment: Main criticisms

Inconsistencies in methods, results, and conclusions lead to discounting sediment porewaters:

Comment: The radon data are used to support conclusions about GW gains within the stream but much remains unclear. What were the radon concentrations in the GW wells? How was the degassing actually determined? Further, if stream turbulence is the key component behind degassing (L198), then why is the degassing of Rn so much greater in the very slow flow of the summer campaign compared to the winter one?

Response: The radon groundwater endmember that was determined from samples taken from three groundwater wells adjacent to the stream and two subsurface agricultural drains amounted to 23.2 kBq/m³. This value will be added to the text in sect. 3.3.

The way how the degassing was determined is comprehensively described in Schubert et al. (2020), sect. 2.2.2. In order to keep our manuscript as concise as possible, we prefer to not repeat this lengthy discussion/explanation but to rather cite the Schubert et al. paper. A related remark has been added to the text in sect. 2.2.1 of our manuscript.

Indeed, stream turbulence is a key component for degassing, as strong turbulence increases the concentration gradient at the water/air interface. Still, stream width and depth are quite influential as well, if the stream is wide and shallow. In particular, the width/depth ratio has major impact. A high ratio allows a strong gas exchange between water and air. As mentioned by the reviewer in another comment (L313) “the water level” was “very shallow for the September 2020 event”. With a stream width of about 30 cm and a depth of only 2 - 3 cm, the large water/air interface per water volume below dominated the degassing intensity. In addition, the very low flow velocity of the stream water (around 0.07 m s⁻¹) gave the radon time to escape. Even though turbulence played no major role in September, we cannot expect laminar flow of the water (due to the bed roughness of the stream). Thus, even though turbulence was not visible or obvious, the low water depth and the comparably large width of the stream resulted in an intense contact of the water with the water/air interface and hence in strong degassing. Therefore, (and as mentioned in sect. 3.3) the degassing of radon is much more intense in the summer campaign.

Comment: Given the issues noted here by the authors, it may help to consult the review by Raymond et al. 2012 to set an initial estimate of Rn degassing for this system based on basic stream hydrology information.

Response: As mentioned by the reviewer, Raymond et al. (2012) is a comprehensive review article. The paper discusses gas transfer velocities in streams and small rivers based on a metadata analysis of 563 gas tracer release experiments. On the other hand, the focus of the study that we discuss in our manuscript was not on gas transfer from surface water bodies. Hence we don't want to discuss the matter in a supererogatory extend and prefer to rather cite related publications. In the revised manuscript, we will mention Raymond et al. (2012) (sect. 2.2.1), since it discusses the issue of degassing comprehensively.

Comment: My impression of the Rn data is that flow in the summer campaigns was so slow that the gas transfer velocity for Rn was perhaps dominated by gas diffusion. The authors seem to corroborate this point on L313-315.

Response: Indeed, radon loss by gas diffusion played a significant role in the summer campaign. This influence was allowed for by the FINIFLUX model accordingly. The model itself is described in detail in Frei, S. and B.S. Gilfedder (2015).

Comment: Further, the water level must have indeed been very shallow for the September 2020 event (L313). For this event, using the reported velocity of ~0.07 m/s, discharge of 0.51 L/s, assuming the smallest stream width of 0.33 m (derived from 10 m / 30 as mentioned on L174-175), and further assuming a simple rectangular stream channel, I calculate a stream depth of 2.2 cm! (Relaxing the assumptions above may yield an even shallower depth.)

Response: The stream depth varied over distance, but about 2.5 cm gives a rough idea of the situation in September, indeed. We allowed for this rather challenging situation by adjusting the associated FINIFLUX parameters accordingly, reasonably and carefully. The FINIFLUX model itself is described comprehensively in Frei, S. and B.S. Gilfedder (2015).

Comment: Given that the stream was moving this slow and the depth was this shallow, I really doubt that sediment porewaters have such apparently little connection to SRP in the water column. Note that the sampled sediment porewaters 7 cm deep (L225) are relevant, but it's likely the uppermost 1 to 2 cm of the sediment that dominate P exchange between sediments and the water column. I suspect that SRP in the porewaters may possibly be lower towards the top of the benthos due to the oxygen gradient (see Palmer-Felgate et al. 2010 for example) but this still may not completely stop a significant flux of SRP to the overlying water column. It's been observed elsewhere that SRP can increase in summer months and that it's likely tied to redox status of benthic sediments (Smolders et al. 2017 cited in text; L75-79) -- why might this argument not apply in this case study?

Response: We found very high porewater concentrations in a depth of 7 cm in the downstream sampling point (20 times the concentrations in the stream water). We agree that the uppermost sediment layer will be dominant for exchange and likely have lower SRP concentrations. Here we do not have uppermost porewater concentration data proving this. However, the main evidence that this shallow stream does not take up significant amounts of P from the porewater is the observed water and SRP flux in the last 600 m of the stream. Data indicate that neither water nor SRP is gained here but that largest part of the flux comes from the upstream stream section. The downstream porewater show clear indication of reductive mobilization of iron and DP (manuscript line 355-356). Here we would make this more clear and add the Smolders et al. 2017 reference.

Note that the limited influence of shallow porewater in the upstream, strongly gaining, section is based on the different P speciation in porewater compared to the one observed in the stream water downstream of that section. Rather the gaining groundwater sampled upstream was similar to stream water. So we have three arguments here to underline the limited influence of the streambed SRP source: 1. Distinct water and SRP flux gaining in the upstream section but not in the downstream section. 2. Hydrochemical similarity of stream water and gaining groundwater. 3. Hydrochemical

dissimilarity of sediment pore water and stream water. We will improve our argumentation in section 3.4 to better pinpoint that.

Comment: Although it's possible that the more oxic sediments at the very surface should have a stronger sorption capacity than the deeper sediments, there's also the possibility of colloidal P -- generated from below -- bypassing the sorption sites and increasing the SRP signal (Gottselig et al. 2017). (I'm assuming filtered water samples [filter size not stated, 0.45 microns?], were still not small enough to prevent these colloids.) There seems to be enough Fe and DOC in the waters to support this.

Response: Our argumentation is not that the possible oxic boundary layer between sediment and water prevents P transport. We argue that the chemical composition (ratio SRP/DP, nitrate, ammonia) and the age of the DOC are very different in the pore water and in the stream. If there was efficient transport of solutes from the pore water to the sediment, then the chemical composition and age of the DOC in the stream would have to approximate the composition and age in the pore water - but this was not the case (lines 330-339, 346-350). It may well be that (colloidal-bound) P from below bypassed the sorption sites and was measured as SRP in the stream. In this case, however, it was not diffusion (from the boundary layer) but advective transport from deeper layers. This would have carried not only phosphate but also nitrate, ammonium and modern DOC into the stream and, if significant, would have approximated the chemical composition and age of the DOC. But there was no indication of this. We therefore assume that there was transport of P from the sediment by diffusion and advective transport through microchannels, but this was insignificant in extent. The pore size of filters was 0.22 μm . We will include this information.

Comment: It is inappropriate to treat SRP or DP as conservative endmembers mixing in this system, as text on e.g., L331-335 suggest is the basis for arguing that sediment porewaters are insignificant. A more compelling argument is needed. (Further, the data cited on these lines is from 7 cm deep so the DP there may be even less relevant.)

Response: We understand that SRP or DP alone cannot be considered conservative endmembers because, for example, the concentration of SRP can change by uptake in algae or bacteria or by precipitation with Fe. As stated above, however, we do not consider the concentration but the ratio of SRP to dissolved organic P (DOP) as being significant. We have referred to this as the fraction of SRP in dissolved P (DP), but would express this in the revised manuscript as the ratio SRP/DOP, where we define DOP as DP minus SRP. Thus, at the upper station, where the significant inputs of water and solutes occur, we would encounter an SRP/DOP ratio of 0.078 in the sediment and 0.935 in the stream. The assumption that turnover in the stream reverses the ratios of sediment-borne P would require dramatic conversion of DOP to SRP, which is unrealistic in such a short travel time in the small stream reach. We therefore believe that the SRP/DOP ratios are a good argument that supports our argumentation.

Comment: Only geogenic P in GW?

I find it surprising that GW P in this predominantly agricultural catchment (receiving regular P applications (L104)) is considered geogenic (paragraph starting L401). It seems that these soils have reasonably high P (L120-122). It is even stated on L125 that fluctuations in soil P "suggests a transport of soluble P compounds to deeper layers". [Note that this sentiment reverses later in the paper on L416-418.] So can we really assert here that there's no agricultural P reaching GW and, later on the flowpath, the stream?

Response: This is a good point. In the cited reference only the top soil layer was investigated, which means only the first top 2 cm. The main objective of the former study was to evaluate the transfer of soluble P from the top soil to surface runoff. Keeping the relatively high P concentrations in the top layer after fertilizer application in mind it is possible that soluble P compounds can be transferred to the underlying layers. On the other hand the loamy soils show a distinct reduction in TP and in C_{org} - concentration with increasing soil depth. In a typical agricultural soil of the study catchment, TP concentrations were elevated in the Ap horizon (778 mg/kg) but showed a dramatic drop to a mean

value of 193 mg/kg in the B and C horizon below the plough pan in 40 cm depth, indicating a very high P sorption capacity of the soil. This indicates that P does not leach from the Ap horizon to deeper layers in the investigated soil (Kistner 2007). This is consistent with our data of tile drain runoff, which indicate lower DP concentrations than both surface water and groundwater. Because investigated tile drains only cover a small but wet area in the valley bottom of the study catchment other areas of the catchment may show different DP concentration in percolation water of the soils. On the other hand this is not very likely because high groundwater levels, as they are in the area where tile drains exist, may accelerate the exchange between soils and groundwater while long seepage distances in more remote areas from the valley bottom may increase sorption processes and increase DP retention (Wriedt et al. 2019). Although soil P leaching is not very likely in the study catchment we cannot completely rule out any loss from agricultural soils to groundwater. We will consider this in the revised manuscript. See also our response to the next comment.

Comment: It seems that true natural reference points for GW SRP (i.e. with no history of agriculture) would be needed to support this -- is this available in the Wriedt et al. 2019 reference?

Response: The study of Wriedt et al. (2019) reveal mean DP values in groundwater of hard rock areas of 0.03 mgPL⁻¹ with a range of 0.01 to 0.10 mgPL⁻¹. He did not find significant differences between land use types (arable land, grassland and forest) but higher outlier values occur more often under arable land compared to forest. In our study area mean DP groundwater concentrations were with 0.05 mgPL⁻¹ slightly above the mean background concentration and therefore agricultural impact cannot completely be ruled out although our soil P and DP groundwater and tile drain data do not confirm DP leaching to deeper soil layers and into groundwater. We will consider this in the revision of the manuscript.

Comment: I take issue here because it is increasingly acknowledged that many of our agricultural catchments are dealing with P "legacies" which will take decades or longer to remediate even with drastic action. Wrongly attributing P sources can lead to even weaker action and longer times for surface waters to recover, if at all.

Response: We fully agree with this arguing. Although in Germany the P balance of agricultural land is closed since more than 20 years TP is mostly elevated in the top soil of agricultural land and often twice as high as natural background conditions (Schachtschabel et al. 1992, Pöthig et al. 2010). This legacy P is subject to transport via surface runoff and soil erosion and under certain conditions also via leaching through the soil and tile drains into surface waters (e.g. Rowe et al. 2016, McCrackin et al. 2018). Reduction measures of these legacy P loss to surface waters should therefore be targeted to the dominant transport pathways. As referenced in our introduction many catchments with certain soil and groundwater conditions may be susceptible for transport of legacy P also via subsurface flow. In our study we argue that under such specific catchment conditions with loamy soils and their high P absorption capacity soils P leaching may be less important compared to other transport pathways like surface runoff. We will consider the issue of legacy P in our discussion.

Comment: The discussion on L415-417 seems to ignore that N and P behave very differently in their transport; seeing faster movement of nitrate than for P is not surprising and doesn't support the statement on L416-417. In fact, the Dupas et al. 2017 paper cited in text argues this very point (and is summed up in their Figure 7).

Response: We agree with this comment because our reasoning is possibly not detailed enough. Indeed, compared to N leaching P leaching from soil is mostly low. This finding does not allow to relate this transport pathway to other P transport pathways. Multiple processes can change the availability of N and P in the groundwater and the further transport to the stream and N:P ratios may change via the transport from soil to streams. Therefore the input from the soil is likely not a sufficient indicator for assessing the transport of these two compounds from the subsurface to the stream. Some misunderstanding is possibly caused by the fact that we did not clearly distinguish between soil processes and transport processes from soil to the stream. We will limit our rationale to the leaching of DP from soil. Soil P profile data do not support significant transport of P from the Ap

soil horizon to deeper soil layers, thus SRP transport via seepage water can be assumed to be limited. We will consider this and revised the manuscript accordingly.

Comment: I actually concur with what's said on L426-428: SRP is probably well buffered in the catchment but that means that those sorption sites may slowly leak P to the system, maintaining ecologically relevant SRP concentrations in the stream for a long time.

Response: We agree with this comment. Although SRP leaching is very limited from the loamy soils in the study catchment we cannot completely rule out slowly leaking of P. Available data on SRP transport within the soil column do not suggest considerable SRP leaching from the soil but this data are limited to a single soil profile. Although the silt part of the soils dominate throughout the catchment hydromorphic properties may super-impose locally. Further long term studies on P leaching in soils are missing. Therefore possible long term leaking of P cannot be entirely ruled out and may slightly vary within the catchment due to varying soil properties.

Other general comments:

Stream description:

Comment: The catchment is described quite thoroughly, but what about the stream itself? Slope, width & depth at time of sampling, morphology type, substrate characteristics, light availability, etc. would be quite pertinent to this study.

Response: The stream itself has a length of 1747 m a slope of 2 %, with a mean discharge of 0.33 m³ s⁻¹ (5.5 l s⁻¹). The stream has a mean width of 0.4 m and a depth of 0.05 m. The substrate consists of fine and mid-granular sand fraction. The light availability is very high, because of only a small forest area of 3 % of the catchment size, which stands mainly not along the stream.

Comment: Illustrating on Figure 1 or adding a new subplot with focus on the stream itself to show locations of all the measurements (tracers, sediments, etc.) would be great context.

Response: We will add a new figure including all measurement sites along the stream and its surrounding.

Comment: Suggest sticking to the red/blue for summer/winter throughout (e.g. avoid the change in Figure 7). Please try to fix the x-axes in the figures with stream distance (m) so that they're similar/more comparable across figures. The aspect ratio for Fig 6 is too wide. The changes in figure styles and dimensions when 'distance from upstream' is on the x-axis makes it more difficult to follow the story across figures.

Response: We will adjust the figures to be consistent in color schemes and the use of the x-axis.

Comment: Consider combining Fig 2 and 3 to save space. Figure 4 could also be combined with 2 and 3 by showing the observed groundwater level for the study period overlaid upon the prior 10 year average -- just an idea.

Response: We will combine Figure 2 and 3. Including also Fig. 4 with a fourth information level would make the figure too complex. Therefore we would like to leave Figure 4 as it is but we could put this figure in the supplement and refer to it in the text as follows: "The groundwater levels [m below surface] in the sampling period ranged between 0.5 and 1.1 m. Relating to the 3 sampling campaigns groundwater levels were at 0.65 m (Jan. 2019), 1.0 m (Sep. 2019) and 0.85 m (Sep. 2020) below surface."

Comment: Double y axes in general: My preference is to avoid them if possible. It clouds the comparisons between seasons and visual points of data do not immediately map to values on the y-axis (until I re-read the caption to know which is which)

Response: We can add the x-axes to all sub-plots and put (a)-(d) into the sub-plots.

Comment: In some cases, they're entirely unnecessary, such as in Figure 7. (Also, why not just plot both series on log₁₀? Having double axes and different transformations is doubly confusing -- this applies to both radon figures.)

If double axes have to be used, I think a simple but very helpful improvement would be to color-code the y-axes text/title with the same red/blue color scheme for the points.

Response: We agree with the reviewer's preference of avoiding double y-axes. At the same time, we have to mention that her/his opinion is a rather subjective perception. We love double y-axes! Still, we color-coded the y-axes in Fig. 7, as suggested.

Comment: With all the data available in some of the figures, the summary stats in Tables 1 and 3 aren't necessary and can be removed to reduce clutter; just make references to the stats of interest within text.

Response: We agree with this and will revise the manuscript accordingly

Comment: I'm not sure the 3rd subplot in Figure 6 (with net change in SRP flux) is necessary -- it's simple enough to examine the pattern in the actual SRP flux data in the bottom subplot, in my opinion.

While the log-axes in other figures are more clearly denoted (e.g. Figure 8), it's not very obvious in Figure 9 -- making figure styles more consistent would help here.

Response: Fig. 6 agreed - we can take out 6c without a loss of information. We will adding ticks to 9a similar to Figs. 7 and 8 to make more clear that axes are logarithmic.

Comment: Writing and writing style:

I had forgotten all of Section 3 was results and discussion, as there wasn't too much discussion until much later. (The last few paragraphs felt like a sudden 'dump' of discussion.) Perhaps this paper is better suited to separate R & D sections?

Response: We disagree here. Yes, we state pure results in terms of concentration and fluxes. But later on (3.4 & 3.5) we do both together. Separating the results from discussion here would need a lot of repetition. However, we can make all this more clear in the chapter title: 1. Observed discharge and stream SRP concentrations. 2. Observed longitudinal water and SRP fluxes. With words like "assessing" and "integrating" in the last chapters we make pretty clear that we also discuss the results.

Comment: Please streamline the introduction to build its focus up to the study objectives. I think introducing the main sources in points (a) to (d) on L50-54 and elaborating each is a fair way to structure the intro, but it seems that structure was forgotten after the intro had covered point 'c' (sediments, L71-79); L80 onwards breaks with that structure leaving me wondering where the intro was headed. Further, if the intro is going to focus on "SRP mobilisation ... in various headwater compartments" (L50) as the first paragraph indicated, then lead sentences in paragraphs such as on L63 should more clearly follow that thread. I.e., 'SRP mobilisation' should be the common theme throughout and clearly connected.

Response: We will streamline the point d) accordingly and follow the suggestion on "SRP mobilization" and will revise the manuscript accordingly

Comment: Avoid excessive passive voice throughout. E.g., L336-337 could be rewritten to avoid the "were found" and so make the sentence clearer: "At the outlet, concentrations Fe, DOC, dissolved P, and NH₄ were greater in the sediment pore water than in the stream (Table 2),..."

Response: good point, we will revise the manuscript accordingly

Comments: Proofreading for clearer English would improve the paper. E.g., I'd suggest revising the sentence on L273-274 to: "The spatial pattern of SRP flux largely followed the pattern observed for discharge." This keeps the emphasis on the main topic of this paragraph (SRP flux) rather than on

'spatial pattern of discharge' in the original sentence. Another example sentence that could be made clearer/more effective is L56-59.

Response: We will revise the manuscript accordingly

Comments: Please break up the paragraph on L401-439 (139 lines!). Further, consider incorporating the discussion with the broader literature done here more evenly throughout Section 3.

Response: We will consider this where appropriate

Comments: L177-178: I'm confused by this sentence. What is the 'time lapse' issue here and why should that matter if you're moving upstream?

Response: That was not the best term used here. We mean time interval of 30 min between injections and will change that accordingly.

Comments: Section 2.2.1: was there any potential for significant sub-daily variation in discharge or was it stable?

Response: During all the field campaigns the discharge was stable

Comments: L199-201: it is stated here that there are multiple ways to estimate the radon degassing from the stream -- which the authors described as a "crucial parameter" just above. So... how was this actually determined here in the present study?

Response: Schubert et al. (2020), sect. 2.2.2, discusses comprehensively how the degassing rate was determined. In order to keep our manuscript as concise as possible, we prefer not to repeat this lengthy discussion but rather to cite the Schubert et al. paper. A related remark will be added to the text of our manuscript in sect. 2.2.1.

Comments: Section 2.2.2: was there any reason for not measuring Rn in the September 2019 campaign? Also, why were sampling locations different from the salt tracer points?

Response: Redoing the Rn measurements was logistically not feasible in Sep 2019. Yes, the sampling locations were similar between both Rn measurements and tracer tests. We will make this more clear in the text.

Comments: L344-350: I don't have any expertise with radiocarbon dating but is it really the case that DOC in the stream is millennia old? Could there be more discussion with literature in this section? Additionally, there's no mention here about the stream metabolism involved. If there's enough light, much of the DOC is likely autochthonous for summer baseflow, especially considering the high nutrient availability.

Response: Yes, the DOC in streams can indeed be thousands of years old, especially if it is groundwater. This does not mean that the water is that old, but the old DOC may have been dissolved and mobilized by recent hydrolysis of old soil POC. In other streams in this area we have found DOC that had a radiocarbon age of up to 3000 years. We will discuss this briefly and include a reference.

The degradation of the DOC also depends on the light irradiation, but is largely determined by the residence time. In incubation experiments using stream DOC, low degradation rates between <0.5% and 2.3% per day were measured (Kamjunke et al. 2022). Thus, a significant change of ¹⁴C values by selective decomposition of young or old DOC is unlikely. We will discuss this briefly in the revised manuscript.

Comments: L354: Can more concrete evidence be given instead of stating "...probably transported by preferential flow paths."

Response: See also reply above. There is no additional evidence. The gaining groundwater sampled in the deep streambed has the same signature as the stream water further downstream. We do not see evidence that the shallow pore water (having a different signature) is significantly contributing to the stream flow chemistry. We therefore propose that the groundwater gains bypass the shallow

pore water. This is likely done by preferential flow paths. Schmidt et al. (2006) is a good example of this behavior showing by temperature-based flux measurement that a fine-grained streambed with low hydraulic conductivity favors preferential flow paths so that largest part of the incoming groundwater flux is channeled to a few distinct locations. We will revise this section and give this literature example as an additional argument.

Comments: L402: where in the paper are the 'oxidised groundwater conditions' established? Could this be included in Table 2?

Response: Right. This was not made clear before but can be seen in Tab 2. (arguing with the low Fe and NH₄ concentrations). We will add this (Line 328 in the original manuscript) to chapter 3.4. We will also make a reference to Table 2 in this point of the manuscript.

Comments: L432: Please avoid the "source/sink" dichotomy for P and sediment sorption. It ignores the transient nature of sediment P and how, really, sorptive materials in catchments only buffer P.

Response: We will consider this in the revised manuscript

Comments: Technical comments

Please superscript all atomic mass numbers: e.g. ²²²Rn and ¹⁴C; fix subscripts too ("P_{sat}" on L59, "NO₃" on L69)

L30: 'SRP losses' here is kind of ambiguous in terms of the direction of the flux.

L31: 'SRP-fraction for dissolved P' is unclear.

L59: 'soil P sorption saturation' on its own doesn't mean greater P mobility unless you refer to soils with high P sorption saturation (typically, the saturation is expressed as some sort of degree).

Additionally, is "P_{sat}" ever used again in this manuscript? If not, no need for a new acronym.

L69: "NO₃" should be defined.

L75: don't use "molybdate reactive P" if 'SRP' is used everywhere else

L89-90: suggest moving the list of pathways out of the parentheses as they're quite important

L90: "locate" instead of "localise"?

L92: the "gaining- and losing water fluxes" needs to be reworked, avoid the dangling hyphen if no hyphen is used for 'losing'

L112-114: this sentence doesn't seem to state anything clearly -- what's the message here?

L122: move the Kistner et al. 2013 reference up one sentence (to align with the "Previous research has..."). Also, no need to give a new acronym for DPS if that's not used again.

L150: I think there's a zero instead of "O" in the "NO₃" here.

L150: The Dupas et al. 2017 reference here is out of place; save this for the discussion.

L213-215: Were the probes calibrated on the day of measurement? Please cite a reference for the 'standard methods' for the P analyses and give some note of accuracy and/or method detection limit.

Response: We will consider all technical notes in the revised manuscript

Comments: L217: how was dissolved iron measured? detection limit?

Response: Dissolved Fe concentrations were measured by ICP-OES (Perkin Elmer 7300 DV). The limit for determination was 0.01 mg L⁻¹.

Comments: L268: 'proportion' not 'proportionate'

L270: "neutral" is an odd term to use here, perhaps switch for "had little net change in discharge"

Response: We will consider all technical notes in the revised manuscript

Comment: L275: perhaps switch out 'gained' for 'contributed'?

Figure 7: please also add to the caption or indicate on the figure what the dashed line means (L286). And shouldn't this apply to Figure 8 too? Also, please keep the time labels consistent with other plots (i.e. January 2019 and September 2020 instead of 'winter' / 'summer').

Response: Actually, the water was "gained" by the stream (the groundwater "contributed" to the stream flow). The meaning of the dashed line is explained in the text above Fig. 7. Still, we added the

information to the figure caption, too. The information "January 2019" and "September 2020" has been added to the figure caption.

Comment: L310, how reliable are the values given here for groundwater given the text on L296? Is some value representing uncertainty (e.g. confidence interval) possible here?

Response: We state that the recorded data and the data evaluation based on the FINIFLUX model suggests that about 60 – 65 % of all water entering the stream is contributed by groundwater. The uncertainty of the result is mainly a function of the assumed radon degassing rate. Still, the related FINIFLUX parameters have been chosen carefully and are reasonable. A discussion of the FINIFLUX model including error propagation would be beyond the scope of the study. The model is described in detail in Frei, S. and B.S. Gilfedder (2015).

Comment: Section 3.4: edit the title to include "in summer baseflow (September 2020)" as that's crucial context in this whole section.

L336: make NH₄⁺ more consistent throughout text (replace "NH₄"); consider also including the valence for "NO₃-"

Response: We will consider all technical notes in the revised manuscript

Comment: Table 2: why is the Fe in the outlet stream sample "n.d." (unsure whether this is "not determined" or "not detected")? Or was it below detection? If the below detection, what was the MDL?

Response: n.d. actually means not determined, but I have mixed this with a non-analyzed value from another sample. The correct value is 0.027 mg L⁻¹, which we enter in the table. Thanks for the hint.

Comment: Table 2: please add pH, dissolved oxygen (or some indication of redox status), conductivity, and temperature here as that would be very helpful context (and seems to have been measured according to 2.2.3).

Response: We will add these compounds within the table

Comment: L432: should this be 'House 2003'? This reference is missing

References -- suggest checking all:

Kleinman et al. 2009; van Dael 2020 reference is duplicated

also some refs are CAPITALIZED

LLFG 2021 reference (L104) is missing

Response: We will check and correct this references

References cited

Gottselig N, Amelung W, Kirchner JW, et al (2017) Elemental composition of natural nanoparticles and fine colloids in European forest stream waters and their role as phosphorus carriers. *Global Biogeochem Cycles* 31:1592–1607. <https://doi.org/10.1002/2017GB005657>

Palmer-Felgate EJ, Mortimer RJG, Krom MD, Jarvie HP (2010) Impact of point-source pollution on phosphorus and nitrogen cycling in stream-bed sediments. *Environ Sci Technol* 44:908–914. <https://doi.org/10.1021/es902706r>

Raymond PA, Zappa CJ, Butman D, et al (2012) Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers. *Limnology and Oceanography: Fluids and Environments* 2:41–53.

Additional References:

Kistner, I., 2007: Anwendung des Modells ANIMO zur Simulation des gelösten Phosphors im Oberflächenabfluss auf der Feldskala und der Phosphorverfügbarkeit im Oberboden auf der Einzugsgebietsskala (in German). Dissertation, Martin Luther University Halle, 182 p.

Kamjunke N, Beckers, LM, Herzprung P (2022) Lagrangian profiles of riverine autotrophy, organic matter transformation, and micropollutants at extreme drought. *Science of the Total Environment* 828:154243. <https://doi.org/10.1016/j.scitotenv.2022.154243>

Scheffer, F. 1992: *Lehrbuch der Bodenkunde/ Scheffer Schachtschabel*. 13. Auflage, Stuttgart: Enke Verlag, 491 p.

Pöthig, R., Behrendt, H., Opitz, D., Furrer, G., 2010. A universal method to assess the potential of phosphorus loss from soil to aquatic ecosystems. *Environ. Sci. Pollut. Res.* 17, 497–504.

Schmidt, C., Bayer-Raich, M., Schirmer, M., (2006): Characterization of spatial heterogeneity of groundwater-stream water interactions using multiple depth streambed temperature measurements at the reach scale. *Hydrol. Earth Syst. Sci.* 10 (6), 849 - 859