

## 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:** This is Referee #1 again, apologies for my slow review. I think the manuscript is still in need of some major revisions before consideration for publication. I previously pointed out the inaccuracy and danger in (over)interpreting these data to conclude that P in this stream is “geogenic” – I disagree on this point even more now – and the authors seemed to concur somewhat in their response... yet the abstract and discussion remains unchanged on this point.

The ultimate source of P measured in the stream is only, at most, a secondary point of interest for this paper: the actual objective was on differentiating between proximal P inputs to the stream (riparian wetlands, stream sediments, and deeper groundwater) during low flow. This study was not designed to investigate whether “intensive arable land use within the catchment was the cause” [L453] – the frame of reference in this study is much further down the flowpath, closer to the stream. Further, I’ll discuss below why the discussion on the source of P in deeper groundwater for this catchment is highly uncertain. So, I insist this argument for geogenic P as the primary P source in the stream be removed entirely since (1) it’s unsubstantiated, (2) it’s not the point of the study, and (3) it’s ultimately harmful for the management of P in catchments, very much like the one in this study, that face P legacies.

**Response:** We thank the reviewer for the comment and agree that the major aim of this paper is to elucidate proximal input pathways of SRP in the stream. In the study we can show that groundwater inputs are the dominant source of instream SRP at low flow conditions. As indicated in our former response we agree that specifying the sources of SRP in the groundwater is still associated with uncertainties. We therefore made changes in the abstract (see also below). Although the identification of sources of SRP in the groundwater was not the main objective of this study, we should provide the reader a brief explanation/discussion from which sources the SRP within groundwater may stem but, at the same time, acknowledge the uncertainties. We refer here to results from earlier studies. Therefore we would like to keep the discussion on sources of SRP in the groundwater. Nevertheless we discuss this issue more carefully and prevent stipulations in the discussion section. See also our comments below.

**Comment:** Related to the above comment, the introduction gives the impression that “land-to-water” SRP fluxes (L103) and a connection to land use (L93-96) is a focus of the study. But this study was not designed to address this. The methods used here cannot address this land use question because they cannot differentiate between ultimate P sources once they’ve already reached an intermediate zone – the “potential source zones” on L237 – such as groundwater. So the text should be focused to the actual parameters of the study.

**Response:** We agree with the referee to make the focus of the manuscript more clear and modified Line 103 and used “proximal SRP” instead of “land-to-water SRP”. Further we specified the line 93-96

more clearly “(transfer from deep groundwater,...)” and “to localise the major source areas of SRP along the stream”.

**Comment:** Another major concern is the lack of a cohesive “story” in the paper. For example, the Conclusions is not coordinated with other sections in the paper (particularly the Abstract), doesn’t really follow from the discussion, and has this sudden appearance of soil erosion as a major topic (why is this just now showing up?). The results + discussion (Section 3) doesn’t have clear points anchoring it (see also my comment advocating for separate results and discussion). The mention of “riparian wetlands” in the abstract and introduction gets virtually no follow-up for the rest of the paper. Et cetera. Overall, the paper needs much more global cohesion to be effective.

**Response:** Thanks for this comment and we revised the manuscript accordingly by separating results and discussion and considering the riparian wetlands in all manuscript sections appropriately.

**Comment:** Relatedly, I strongly suggest separating the results and discussion sections as it suits this particular paper much better. I think Reviewer #2 also pointed this out. For most of Section 3, the reader is waiting for the story of the paper to come together (since we’re told this is also “Discussion”), which really only starts to take form in the latter half of 3.4 and then 3.5. Since sections 3.1 – 3.3 are mostly results, 3.4 is half/half, and 3.5 is all discussion anyways, it is not a stretch to simply condense results into its own section (along with some editing for conciseness), then have a well-rounded Discussion that pieces together all the evidence to tell the story. This edit would not require “a lot of repetition” of results. In fact I think the text would either be the same length or even shorter. The story of the paper might then be clearer and more effective.

**Response:** We agree to consider the riparian wetlands in more detail in the result and discussion section. Moreover we now separated the result and discussion section, see the revised manuscript.

**Comment:** The paper has a major focus on groundwater (GW) being a dominant source of P to the stream (e.g. paragraph on L440) and at times the text seems to suggest it’s virtually the only source of P or the only process that matters (e.g. the text excludes/discounts other potential sources but doesn’t provide/support additional, potential sources). However, just based on simple mass balance, this doesn’t add up. Taking a GW discharge value of 0.75 L/s (Fig 6) and a GW SRP value of ~40 to 50 ug/L, what gives a GW P load of 0.03 to 0.04 mg P/s. But this is considerably lower than the winter total stream SRP load (~0.055 mg P/s) yet considerably greater than both the summer total stream SRP loads (~0.02 mg P/s). I think the authors should address this and discuss reasons for the disparity for both winter and summer conditions. In my view: other sources were mobilised in the winter event and in summer there may be in-stream/hyporheic processes that diminish/buffer the GW P flux.

**Response:** We do not argue that groundwater is the only SRP source in the Schäfertal catchment but the major SRP source during low flow. Groundwater SRP flux in the winter campaign can be approximated using the observed groundwater concentrations and the Radon-based inflow of groundwater or drainage water. The observed SRP flux at the outlet is within the uncertainty of the Radon analysis and the variability of groundwater concentrations. Therefore we do not expect considerable other sources than groundwater during the observed conditions but surely need to acknowledge the uncertainties in our calculations. We added this argumentation to the result section 4.1. Note that in summer 2019 discharge was even lower than 2020 (which also suggests a clear dominance of GW to total discharge) but SRP concentrations at the outlet were different in 2019 and 2020. This suggests variation in groundwater SRP concentration of the same contributing area in time or variation in groundwater SRP concentration in different contributing groundwater source areas. The older SRP groundwater data (Figure7) also suggest that SRP GW concentrations are not constant and may vary in space and time. Moreover we agree that during summer low flow conditions also

other in-stream/hyporheic processes may potentially lower SRP stream concentration, especially in 2020, but SRP fluxes in the downstream reach, where no substantial exchange between groundwater and stream water was observed, did not significantly change. This is in line with findings of Bernot et al. (2008) who observed only low SRP uptake in small agricultural streams with comparable low SRP concentrations. We consider this in the revised discussion section.

### Specific comments

**Comment:** Again, language throughout could be greatly improved. I've pointed out only some examples (not exhaustive) below in technical comments. This applies not only to grammar but also to conceptual and technical language. I also have several specific comments below around technical aspects of the study which I think need refinement; these edits should also greatly aid the story of the paper.

A key point throughout this study is the connection of (deeper) groundwater to the stream under low flow conditions, supplying virtually all the flow, which coincides with generally the highest SRP concentrations. While I agree on this water source being prevalent at low flows, I'm not sure this study establishes this as the sole source for any given low flow period, as is illustrated with the two September campaigns. We only see detailed data for September 2020, which seems demonstrative of the groundwater P hypothesis, but notably September 2019 had much greater stream SRP concentrations (up to ~60-70 ppb). Unless deeper groundwater also increased SRP by 20-30 ppb, which I wouldn't expect as usually groundwater P is less dynamic, this high stream SRP leaves us wondering what the story is for September 2019 since groundwater SRP is likely lower. (Is there groundwater SRP data for this campaign too? I'm not seeing it.)

**Response:** This is a good point but again groundwater SRP concentrations are not constant and may vary in space and time. This can be seen in Figure 7. The boxplot shows that 75% of the SRP groundwater concentrations ranged between 0.03 and 0.07 mgPL<sup>-1</sup>. This range still explains the measured data at the outlet for both campaigns. Note also the log-log-linear relationship between concentration and discharge in the three measurement campaigns shown in Fig. 7a. SRP concentration clearly show an increase with decreasing discharge. Assuming that the groundwater fraction in total stream discharge increases with decreasing discharge this appears plausible for us. Again: We do not explicitly exclude other sources of SRP but argue that groundwater is the dominant one. We did not measure SRP in the groundwater wells in 2019.

**Comment:** Further on the groundwater SRP: these groundwater SRP values are really high in general. I've read through the Wriedt et al. 2019 reference (mein Deutsch ist schlecht aber ich weiß genug) and I'd assume that the Schäfertal catchment would classify as part of their "Bergregion" (being part of the Harz mountains/Berge) – the Bergregion generally sees groundwater "phosphate" [note they analysed P via molybdenum-blue, i.e., it's SRP] concentrations of roughly 50 ppb as PO<sub>4</sub> or less (hence, around 16 ppb or less as P) depending on depth, etc.. The Schäfertal catchment in this current study, however, seems to have much greater groundwater SRP concentrations than this natural reference of 16 ppb SRP (despite its oxidated status), with Figure 7 suggesting a median around 50 ppb P (more than 5- fold greater than the natural reference). Having worked in catchments with comparable geologies (greywacke), I would be floored by 50 ppb SRP being "natural" or "geogenic". This is all to say: I sincerely doubt the dominant source of P in the study stream as being "geogenic". I'm not suggesting that the text explores this point – as it's not the point of the study – but rather that the authors reckon with this and remove unnecessary/unsupported text from the Discussion, then refocus the text on how the proximal sources of P coupled with hydrological flowpaths determines in-stream SRP for their three sampling campaigns.

**Response:** Comparing the comments of the reviewer with our own statements in our discussion in section 3.5 and the conclusions on the contribution of agriculture and geogenic sources to groundwater SRP we think these viewpoints are not very different - we do not rule out agricultural

impact on groundwater SRP concentration. But we agree to change the wording in the abstract which suggested that geogenic sources are the overwhelming contributor of SRP to groundwater (see revised manuscript).

**Comment:** Relatedly, the authors argue for the catchment soils being highly P sorptive and thus not conducive to P leaching to groundwater (L467-468). A change in soil organic matter content and total P with depth is not necessarily an argument for limited soil P leaching (L467) since P sorption capacity is not a function of total P and is only somewhat related to organic matter. We'd need more direct observations of P sorption capacity, such as the degree of P saturation (DPS). In fact, the Kistner et al. (2013) in-text reference gives topsoil (0-5 cm) DPS in the Schäfertal of roughly 31% -- this is fairly high and saturated to the point that the soil is conducive to P loss in either surface or subsurface runoff (Fischer et al. 2017; Kleinman 2017). Further P loss in the subsurface soil would depend more on the soil's sorption chemistry, which is unavailable. So I am not convinced that soil P (particularly that coming from the historical fertiliser use in this predominantly agricultural catchment) here isn't reaching groundwater.

**Response:** There still seems to be some misunderstanding. What we wanted to point out is that the soils in the study watershed are highly sorptive and there are several indications which support this arguing, for example, if soils had limited sorption capacities, we would not expect such low SRP concentrations in drain water. This does not mean that no SRP from the soil column can reach the groundwater, and we did not state this in the discussion. The mean DPS value of roughly 31% was calculated according to van der Zee and Van Riemsdijk (1988). These values are in general lower than those calculated in Fischer et al. (2017) and Pöthig et al. (2010) and not directly comparable. Although WSP of around 13 mg P/kg indicate well P supplied soils for plant growth risk of leaching into groundwater was estimated low because of large distance between arable soils and groundwater head (Kistner 2007). Moreover half a year after P fertilizer application water soluble P content of the top soil was only 0.3% higher than those of unfertilized fields (Kistner 2007). This indicates high P sorption capacity of the soil which leads to a quick adsorption of readily available P compounds from the applied P fertilizer. Here we only argue that P leaching is likely low compared to other soil and aquifer properties. We still think that discussing these aspects is valuable for understanding SRP transport in such loamy lower mountain range catchments.

**Comment:** While it's ok (in my humble opinion) to use the term "soluble reactive P", it does seem silly to use that term but then call the unreactive P in the same filtered water sample as "dissolved organic P" (I'm assuming that's what "DOP" stands for, as it wasn't defined). Additionally, I'd be cautious about calling the unreactive P in a filtered sample "organic" – it likely is predominantly organic P but won't necessarily be 100% organic P (simply "unreactive", as in "molybdate-unreactive", is more accurate).

**Response:** While it is common using the molybdate-reactive P (soluble reactive P), it would be quite unusual referring to the remaining P fraction as "unreactive". We included into the method section: "We operationally define the difference of TDP and SRP as dissolved organic phosphorus (DOP), although this fraction may contain some inorganic phosphorus."

**Comment:** L234: Dissolved P (DP) was not defined in the methods (L266). Is this total dissolved P? If so, I suggest including "total" in the name (i.e. TDP).

**Response:** We now defined dissolved P (DP) as total dissolved P (TDP).

**Comment:** A focus of the paper is on baseflow conditions and the winter (January 2019) campaign is referred to as baseflow (L312). I think this needs to be more careful, as I'm not sure this event is really 'baseflow'. Discharge during the January campaign looks to me like a receding limb following a rain on snow event (Fig 2) – not exactly normal baseflow. Groundwater contribution as % of total

discharge would be somewhat lower than (groundwater meaning what your groundwater endmember represents). Instead, much more Q is likely coming from runoff and shallower stores (soil matrix). [In contrast, the summer September campaign is thoroughly deep groundwater, which the model results also support. This makes sense to me hydrologically and strikes me more as 'baseflow' as I understand it.] Perhaps the authors just need to clarify what is meant by 'baseflow'.

**Response:** We define baseflow as stream discharge, which is not containing direct runoff from a previous runoff event. It is the portion of the streamflow that is not directly generated from excess rainfall during a rain event, i.e., the streamflow that would exist without the contribution of direct runoff from the rainfall. This is a hydrological definition. Discharge in January campaign was relatively high and the reviewer is possibly right in assuming that some interflow, i.e., flow coming from shallower stores (e.g., soil matrix) from soils is also contributing to the discharge.

**Comment:** L301-302: So, the 'groundwater' endmember includes not only groundwater wells but also agricultural drains, and apparently these water sources have fairly similar Rn concentrations. This means it's difficult, based on the Rn data, to differentiate between these water sources to the stream. Is it possible at some points that the drains are delivering P to the stream? At very low baseflow, couldn't this P be remobilised into the water column? At the very least, it's confusing how the different subsurface sources of water sometimes get lumped into "groundwater".

**Response:** Generally, radon concentrations in water decrease (1) by decay and (2) by degassing. In an aquifer (i.e., in groundwater) decay is balanced by radon production and degassing does not occur. In the agricultural drains decay is NOT balanced by radon production but the residence time of the water within the drains is too short to allow notable decay. Degassing does not occur in the drains because the radon concentration in the air-filled headspace of the drain is in partition equilibrium with the drain water, or in some cases water-filled. Hence groundwater and drain water show the same radon concentration, which makes it reasonable that these two subsurface sources of water "get lumped into" a radon groundwater end-member. At the same time the reviewer is right in saying that with the radon method it is not possible to distinguish groundwater from the water produced by the drains, as they are essentially the same water, despite perhaps coming from different areas of the aquifer. We now use "lumped radon groundwater and tile drain water endmember" instead of "groundwater endmember" in connection with the Rn based calculation of subsurface inflows when a contribution of tile drain water cannot be excluded (winter campaign).

**Comment:** L308-312: This section should be deleted. First, it doesn't make sense to compare the Rn concentrations across different conditions anyways. The Rn tracer will degas back to atmosphere very differently across conditions. The summer low flow likely allowed tons of Rn degassing despite how much the GW contributes to Q. This Rn concentration comparison just has no place here. Second, the text here seems to imply that discharge in January was mostly GW while in September it was less so – the rest of the text argues the opposite. Absolute GW discharge could've been higher in January (although the data/model results suggest the GW Q's were pretty similar, which seems about right to me), but it made up proportionally \*less\* of the total Q.

**Response:** FIRST: We don't see the reviewers point why "it doesn't make sense to compare the Rn concentrations across different conditions". The differing degassing rates are explicitly accounted for by the FINIFLUX model. We don't simplify degassing by assuming a loss of "tons of Rn" but quantified the loss based on a specific degassing coefficient considered in FINIFLUX. The sensitivity of the modelled groundwater discharge rates on this applied degassing coefficient was assessed in detail (as discussed in the mns.). Hence, from our perspective the Rn concentration comparison makes perfect sense here.

SECOND: The message, which is clearly conveyed throughout the section is that river discharge in summer (September) was mostly GW while in winter (January) "normal baseflow" contributed about 60 – 65% of GW to the river discharge. This is the case because, on the one hand, GW discharge rates

were comparable (0.6 l/s vs. 0.75 l/s) but, on the other hand, river discharge rates differed and were considerably lower during summer. This shows that while in winter multiply flow paths and water sources contribute to total stream discharge, during draught conditions groundwater almost exclusively makes up stream flow.

**Comment:** L318-327: The sensitivity analysis here is nice to have. The text here demonstrates why I previously recommended giving an a priori value for the degassing coefficient based on the Raymond et al. 2012 review (see their empirical equations). I think this would bolster the discussion on this point. This is important because, for example, the degassing may be much greater than that parameterized for the winter event which implies that there must be greater 'groundwater' input (including the agricultural drains) in the first ~450 m which is also where the greatest increase in SRP flux occurs. I suspect the degassing may be too low because of the poor model fit in the downstream sections. Additionally, could a sensitivity analysis be performed for the summer campaign too?

**Response:** Comparing the various degassing equations is a good approach to better understand the uncertainties associated with the degassing coefficient ( $k$ ) when calculating radon mass-balances, especially for small streams. The approach was executed (and discussed in detail) in the original study by Schubert et al., (2020) applying the commonly used equations by O'Connor and Dobbins (1958) and by Negulescu and Rojanski (1968). These two equations are often used in radon studies to capture the possible ranges in radon gas exchange (see e.g., Unland et al., 2013 or Atkinson et al., 2013). Unfortunately, neither of these equations captured the down-stream radon decrease, which is why we also attempted a propane injection and a fitting method, assuming downstream was either neutral or losing as suggested by the water balance (see e.g., Cartwright et al., 2014). In the current paper we want to abstain from repeating the facts published in Schubert et al., (2020) but have made the best estimate of  $k$  from the different methods discussed there. Then conducted a sensitivity analysis for both summer and winter campaigns based on 25 % of this value, which seems reasonable since the empirical equations were not able to capture the radon loss in this very small stream (most degassing equations are made for larger rivers where sewage input and associated oxygen reduction was a problem historically). We added related text to the manuscript (see manuscript with track changes) . Still, generally the reviewer is right in saying that the higher the degassing, the more groundwater inflow is needed to compensate the mass-loss to the atmosphere.

**Comment:** L342: It's not necessarily that there was "intense radon degassing" during this low flow – degassing would be much more intense due to turbulence (a key influence of radon loss) during higher flow, no? I think it's just that stream velocity was very slow so the time available for degassing was much greater.

**Response:** Indeed, the long time period available for degassing increased radon loss by degassing between two measurement sites. However, the low flow conditions also lead to a very shallow water level in the creek (only cm). And a very shallow water level is favorable for degassing because it results in a very high ratio water surface / water volume. The water column depth is a very important parameter, as seen by its inclusion in almost all empirical degassing equations. Generally, the depth is likely a significant uncertainty in the calculation of the degassing in particular for very shallow water levels, as it is difficult to measure and extrapolate it over the whole stream length.

**Comment:** Section 3.4: SRP:DOP means little, has no biogeochemical basis, and doesn't suit as a tracer; just focus on the SRP and DOP masses themselves. You can make the same point but with much more clarity by just focusing on DOP concentration – the ratio is just a distraction.

**Response:** SRP and DOP (as defined) are both fractions of phosphorus. They differ (biogeochemically) in their origin (SRP can be released by zooplankton grazers, DOP can be released by anaerobic organic matter decomposition). The ratio of SRP/DOP is specific for different sources of P. The ratio can therefore be used to identify the source of phosphorus in the stream by comparing

the SRP/DOP ratio in the stream with the ratios in potential sources. A ratio tells much more about possible sources than concentrations of either SRP or DP, as concentrations can change rapidly by e.g. dilution, while ratios remain stable. The same principle is applied with other elements, e.g. the ratio of chloride to sulfate to characterize different groundwater bodies. We therefore think using the ratio of SRP to DOP is more convincing than concentrations of either SRP or DOP.

**Comment:** Section 3.4: This section repeatedly refers to porewaters from 7 cm deep as “shallow”. Considering that stream itself is only 1 to 3 cm deep and that the substrate is fairly fine (L130), 7 cm below the substrate is actually very deep in the hyporheic zone, if it can be considered hyporheic at all (and remember: sediment porewater’s connection to the water column is mostly via hyporheic exchange).

Contribution to total hyporheic exchange tails off with depth enormously so it’s likely the waters down here move extremely slowly and represent a tiny minority of total hyporheic exchange flow. Boano et al. 2014 discusses this in detail. Any discussion of sediment should make it clear this is relatively \*deep\* sediment, and so sentences like L390-391 should be given better context. [Even with the reductive dissolution of Fe oxide bound P, it isn’t clear what effect this has on the water column because this porewater must travel upwards through the hyporheic zone and likely comes across the zone of DO penetration and therefore potential for re-sorption.]

**Response:** We suggest to keep the description of the porewater from 7 cm deep as “shallow” because thickness of stream sediments was more than 20 cm although it is difficult to assess whether also deeper parts of the sediment are contributing to exchange with the stream water. Moreover one should keep in mind that water depth is highly variable in time and not limited to the range of 1-3 cm. We added this arguing also in the manuscript.

### Technical comments

[Again, these aren’t exhaustive. Please proofread carefully throughout.]

L23: “...often with the highest concentrations”

done

L27: “of SRP fluxes”

done

L28: revise to “... and sampled for SRP, iron, and 14C-DOC...”

done

L33: “and was thus...” – this logical connection doesn’t make sense here because the abstract doesn’t establish for the reader key points such as (1) SRP is high in groundwater and (2) summer low flow is dominated by GW (deeper GW, rather than other shallower sources). Please improve the connection.

done

L35-36: Remove “Examination of”; replace “confirms” with a verb more along the lines of “corroborates” or similar – temper the tone with scientific caution.

done

L37: I’m not sure what’s being argued here with the bit on ‘seepage from agricultural phosphorous’. [Note that P doesn’t end with “ous”!] What does this mean to the reader at this point in the abstract?

done

L53: Stream sediments don’t just provide P via reductive dissolution, there’s also desorption (among other processes).

done

L56: I would advise against strongly scientific certain language like “undoubtedly”. Could just revise this sentence to say “Subsurface transport is often dominated by preferential flow...”

done

L58: “therefore”

done

L59: How does a factor “such as soil P sorption saturation (Psat)” “greatly increase P mobility through soils” ? Do you mean, soils with a high P saturation?

Yes, you are right. We corrected accordingly

L69: Nitrate is an anion: NO<sub>3</sub>-

done

L74: I don't think “monomeric” is the term you mean here (I don't think phosphate counts as a monomer...). Do you mean “orthophosphate”, i.e., phosphate in its various hydrated states (HPO<sub>4</sub><sup>2-</sup>, H<sub>2</sub>PO<sub>4</sub><sup>1-</sup>, etc.)?

**Response:** Because the notion monomeric is possibly not clear, we took it out.

**Comment:** L75: “Molybdate reactive P redox-mediated release” – this isn't clear as written; please proofread throughout. Also, I pointed out in the previous review that it does not make sense to use ‘molybdate reactive P’ in one section but ‘soluble reactive P’ or other variants elsewhere. The ‘reactive’ implies ‘molybdate reactive’. It's fine to me to use ‘molybdate reactive P’ in the manuscript but if so, use it throughout to avoid confusion. (The literature is full of varying definitions of the same analyte; let's not muddy the waters more.)

**Response:** We agree with the reviewer and use SRP consistently within the whole manuscript.

L93: Rather than ‘level’, just use ‘concentration’ for consistency.

done

L129: 1747 meters?

This is correct

L130: Delete ‘fraction’

done

L131: Revise the forest area part for grammar.

done

L132: TP is not defined.

done

L136: “DSP” – was this supposed to be ‘DPS’?

done

**Comment:** L158: “runoff” seems like the wrong term here; consider “...under these drought conditions, water from deeper, older storage contributed more to stream discharge, particularly after a wet season [or in comparison to a wet season?]”

**Response:** runoff is also often used for the sum of all components of water entering the stream but sometimes also restricted to water from the surface of an area of land, therefore we now use the term water. Furthermore “particularly after a wet season” is correct.

**Comment:** L163: Edit to: “Baseflow data show clear seasonal variations, with NO<sub>3</sub> peaking in winter while DOC and SRP peak in summer.” [is the Dupas et al. 2017 reference specific to your study stream? Does it contain this data? Unless the answer to both is ‘yes’, then remove the reference]

**Response:** done; yes, Dupas et al. 2017b is referring to the study stream. Thanks for that because we missed to add the second manuscript of Dupas et al. from 2017 and we now distinguish between Dupas et al. 2017a and Dupas et al. 2017b.

**Comment:** Should the title of section 2.2.1 be broader than just “lateral inflows”? E.g. 2.2.1.1. doesn't really cover “lateral” inflows, but stream discharge itself. Maybe “Stream hydrological measurements” in general, or similar.

**Response:** good point because 2.2.1 is related to water quantity and 2.2.2 is related to water quality.



Our idea was to indicate the importance of lateral inflow and its spatial distribution along the stream but of course we also consider vertical inflows through the stream bottom. We select “Water inflows to the stream” to still consider the spatial variation of inflows to the stream.

**Comment:** L212-213: This sentence adds little here. Instead, could you add here or perhaps in the results what gas transfer velocity could be expected (independent of your model) for your study stream based on the relationships provided by Raymond et al. 2012? I.e. give a value or range of plausible values. This would provide confidence in the parameter value obtained via your FINIFLUX model calibration.

**Response:** Generally, all published empirical equations that aim at quantifying  $k$  (and thus, the ones presented by Raymond et al.) are derived by fitting degassing data from a wide range of different “model streams”. However, none of these model streams are as small as the Schäfertal stream. That makes the Raymond equations only suitable to only a limited extent for parameterizing radon degassing in our case. However, as requested by the reviewer we added the resulting information to the manuscript. We made considerable changes in section 3.3 of the manuscript.

**Comment:** Section 2.2.1.2. This section needs to be more clearly organized and probably should feature some paragraph breaks. The point raised on L209 (“A crucial parameter... is the rate of radon degassing...”) just gets brought up but is left unresolved.

**Response:** Section 2.2.1.2. (L209): We tried to give the section a clearer structure by adding some paragraph breaks. However, we don’t agree that the rate of radon degassing “*just gets brought up but is left unresolved*” in the text. Radon degassing is sufficiently discussed in the section 3.3 of the manuscript or the related literature is cited.

L217: “traces” – “tracer”

done

L226: Item (iv) is unclear to me.

**Response:** Both item (iv) and item (v) relate to  $^{222}\text{Rn}$  activities; i.e., (iv)  $^{222}\text{Rn}$  in stream water specific for the sub-section and (v)  $^{222}\text{Rn}$  in the overall groundwater.

L235: “soluble reactive P (SRP)”

done

L239: switching between “Fe” and “iron”...

**Response:** we used now Fe throughout the text with the exception of “iron–reductive conditions”

L244: remove the comma after “both”

done

L253: italicize the “ $g$ ” for the gravitational constant

done

Figure 5: The axis on the right side has a “0.0” label – this is inconsistent with the log scale. Edit this axis to be more like that in Figure 6B.

done

**Comment:** L339-340: This sentence confused me because I knew the January 2019 total discharge was much greater (5 to 6 L/s) until I realised this sentence referred to the water gained within the study reach – perhaps edit to make this clearer.

**Response:** We clarified this sentence in section 3.3.

L341: ‘stream depth’ instead of ‘water level’

done

**Comment:** L342: It's not necessarily that there was "intense radon degassing" during this low flow – degassing would be much more intense due to turbulence (a key influence of radon loss) during higher flow, no? I think it's just that stream velocity was very slow so the time available for degassing was much greater.

**Response:** Indeed, the long time period available for degassing increased radon loss by degassing between two measurement sites. However, the low flow conditions also lead to a very shallow water level in the creek (only cm). And a very shallow water level is favorable for degassing because it results in a very high ratio water surface / water volume. The water column depth is a very important parameter, as seen by its inclusion in almost all empirical degassing equations. Generally, the depth is likely a significant uncertainty in the calculation of the degassing in particular for very shallow water levels, as it is difficult to measure and extrapolate it over the whole stream length.

**Comment:** L356: Why isn't electric conductivity included in Table 1? This is actually the closest thing to a 'conservative tracer' and can help compare the different waters. Table 1: It's not clear what the distinction is between the two groundwater entries are (perhaps explain in caption?). Also, maybe it's just the formatting, but should the second GW entry have "n=6" as the station?

**Response:** Electric conductivity was measured in surface water only but not in groundwater and in sediment pore water. Thus we do not report it in Table 1 but we give the values which were not very different in upstream and downstream station in the text. n=6 was added to the table.

**Comment:** L388: Before this point (e.g. L232), I was already wondering why nitrate wasn't in Table 1 but sentences like this especially highlight this – can nitrate be added?

**Response:** Yes, we added the nitrate values but without groundwater wells

**Comment:** Figure 7: Clarify the sources for the historical data here (they're not from this study, correct?) in the caption.

**Response:** we added the information on the data to the Figure 7.

#### **Additional references**

Boano, Fulvio, Judson W Harvey, A. Marion, Aaron I Packman, R. Revelli, L. Ridolfi, and A. Wörman. "Hyporheic Flow and Transport Processes: Mechanisms, Models, and Biogeochemical Implications." *Reviews of Geophysics* 52 (2014): 603–79.  
<https://doi.org/10.1002/2012RG000417>.Received.

Fischer, P., R. Pöthig, and M. Venohr. "The Degree of Phosphorus Saturation of Agricultural Soils in Germany: Current and Future Risk of Diffuse P Loss and Implications for Soil P Management in Europe." *Science of The Total Environment* 599–600 (December 1, 2017): 1130–39.  
<https://doi.org/10.1016/j.scitotenv.2017.03.143>.

Kleinman, Peter J. A. "The Persistent Environmental Relevance of Soil Phosphorus Sorption Saturation." *Current Pollution Reports* 3, no. 2 (June 1, 2017): 141–50.  
<https://doi.org/10.1007/s40726-017-0058-4>.

## Reviewer 2

**Comment:** I have looked through the responses to Reviewers 1 and 2 and the authors have done a thorough job in addressing the previous two reviewers' comments and concerns.

This manuscript addresses an important subject: understanding and apportioning P sources under baseflow periods of greatest ecological sensitivity. Many papers have assumed that elevated P concentrations under baseflow reflect point source contributions. By combining salt dilution testing and novel  $^{222}\text{Rn}$  and  $^{14}\text{C}$ -DOC measurements, with more routine water quality analyses, the authors have been able to demonstrate that groundwater is a dominant source of SRP to an upland headwater stream. I enjoyed reading the manuscript and I recommend the paper for publication, subject to some small revisions:

**Response:** many thanks for this general positive comment

**Comment:** Line 180: "gross gains, gross losses and net change"— please specify "in flow" or "water flux"

**Response:** This refers to water flow and was added to this sentence.

**Comment:** Line 234 – It would be helpful if the authors could specify what they measure as DP. I assume that this is, in fact, a total dissolved P (TDP) fraction, using the same digestion step as for TP, but for a filtered sample? It would be helpful to at least provide some very concise information here about analytical methods, so that the reader can clearly differentiate these two dissolved/soluble P fractions.

**Response:** We now defined dissolved P (DP) as total dissolved P (TDP).

Line 355 – Please change "steam" to "stream"; also conductivity units should be  $\text{mS cm}^{-1}$ ?

done

Line 424 – "PO<sub>4</sub><sup>+</sup>" should be "PO<sub>4</sub><sup>3-</sup>"

done

Line 439 – I suggest changing this to "driven by within-river biogeochemical processes"

**Response:** we choose "in-stream biogeochemical processes" because the term "river" would be misleading for the small creek.

**Comment:** Line 444-445 – As far as I am aware, the Jarvie et al 2008 paper deals with one UK rural catchment where elevated baseflow P concentrations were attributed to agricultural point sources and septic tank sources, not groundwater. The statement, as currently written, incorrectly implies that groundwater is a dominant source of P in UK rural catchments. This is misleading; however, in \*some\* locations groundwater can be a source of ecologically-significant P concentrations. The authors may find the following reference useful: Holman et al (2008): "Phosphorus in Groundwater— An Overlooked Contributor to Eutrophication?" *Hydrological Processes* 22(26): 5121 – 5127.

**Response:** we agree with the reviewer and modified this sentence accordingly to prevent misunderstanding and also added the valuable reference of Holman et al. 2008.

Line 455 Please change "Nitrates" to "Nitrate"

done

Lines 500-503: "Currently enriched P in agricultural soils.....McCrackin et al 201)". This the first time in the manuscript that these issues about agricultural soils as long-term legacy P sources are raised. This is not a conclusion of this study and I recommend that this section of the Conclusions be moved to the Discussion.

done

Minor comment: throughout the manuscript, subscripts and superscripts were often omitted (mg L<sup>-1</sup>, NO<sub>3</sub>, NH<sub>4</sub>, etc.)

done