

Review for "Seasonal variation and release of soluble reactive phosphorus in an agricultural upland headwater in central Germany" by Rode et al. (hess-2022-126; submitted to HESS 2022-4-8; major revision 2022-9-1)

### General comments

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.

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.

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.

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.

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 P/L, that 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.

### Specific comments

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.)

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.

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.

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).

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).

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 then (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’.

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”.

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.

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?

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.

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.]

### Technical comments

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

L23: "...often with the highest concentrations"

L27: "of SRP fluxes"

L28: revise to "... and sampled for SRP, iron, and <sup>14</sup>C-DOC..."

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.

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

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?

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

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..."

L58: "therefore"

L59: How does a factor "such as soil P sorption saturation ( $P_{sat}$ )" "greatly increase P mobility through soils"? Do you mean, soils with a *high* P saturation?

L69: Nitrate is an anion:  $NO_3^-$

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 ( $HPO_4^{2-}$ ,  $H_2PO_4^{1-}$ , etc.)?

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.)

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

L129: 1747 meters?

L130: Delete 'fraction'

L131: Revise the forest area part for grammar.

L132: TP is not defined.

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

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?].”

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]

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.

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.

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.

L217: “traces” – “tracer”

L226: Item (iv) is unclear to me.

L235: “soluble reactive P (SRP)”

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

L244: remove the comma after “both”

L253: italicize the “g” for the gravitational constant

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.

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.

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

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.

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?

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?

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

#### **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>.