We use the following format for comments and responses:

R1C1 = Reviewer 1, comment 1. R1AR1 = Reviewer 1, Author Response 1.

General comments

R1C1 - The manuscript by McLagan *et al.* is well-written and describes a well-constructed study. The experiments detailed in this manuscript are well documented, frequently sampled, and analyzed to answer important research questions for Hg contamination, transport, and sorption/desorption. The resulting data is valuable and provides noteworthy contributions for future Hg research to build on. The high sorptive capacity of Hg in these sediments is an exciting find. While data on Hg sorption and desorption and presented retardation and distribution coefficients for Hg at contaminated concentrations in environments with low DOM may be the main selling points of the study, the data on Hg isotope fractionation associated with processes of sorption and desorption and evidence of Hg reduction in oxygenated soils are also noteworthy take-aways. The manuscript is recommended for publication in SOIL with minor suggested edits.

R1AR1 – We thank the reviewer for their time and overall positive support of the manuscript.

R1C2 - A general suggestion for improvement relates to the possible comparisons between the experiments and the real-world sediments sampled for this study. It is detailed that 74% of the material was discarded during the sieving process (Ln 130-131). In addition, there is a mention of an "optimal density" for the column packing (Ln 147), and the Authors went through several packing methods in the preliminary experimental phase (S1) where the density appears to have been a major factor in the consideration. This raises the following questions: 1) what the target density is, and how does it relate to the real world, e.g., what system the experiments now can be said to most closely mimic, and 2) what the effect of sieving off ¾ of the total weight could have for translating the findings to real-world systems. Discussing this would be a welcome inclusion, as would a site description of where the samples were collected.

R1AR2 – We would address the reviewers concern about the removal of coarser (large size fraction) materials by highlighting the importance of surface interactions during sorption of analyte from liquid-to-solid phase. The "specific surface area", which relates to the surface area (SA)-to-volume (V) ratio of the materials is critical; fine materials have much greater SA:V ratios; and hence, fine materials dominate sorption. Coarser materials dominate mass but play only a minor role in analyte sorption; thus, they are removed from column experiments. [References: e.g. Zhu et al., 1997; Cucarella & Renman, 2009]. In addition, large materials create more heterogeneous and less reproducible columns, which also supports their removal

(Lewis & Sjöstrom, 2010). Hence, this size fraction selection (<2 mm) followed general practices in the column experiment literature.

In terms of column density, we followed guidelines presented by Lewis & Sjöstrom (2010) for these aquifer materials which have been defined as sandy-gravel. The following sentence has been added to the methods:

"According to Lewis and Sjöstrom (2010), the average bulk densities range from 1.2 – 2.0 g/cm3 for sands and 1.6 – 2.0 for gravel. Thus, we deem the achieved bulk density of the columns to be appropriate for these materials, particularly as densities of the removed coarse materials are higher (solid densities are estimated at 2.65 g cm3; Lewis and Sjöstrom (2010))."

R1C3 - It is also suggested that the text be edited in a few instances to improve the conciseness and structure. This includes the beginning of the introduction, which starts very broadly with little direct relevance to the study, and the section between Ln 462-475, which can be shortened to a few sentences to keep the spotlight on the major findings in the manuscript.

R1AR3 – [comment from leading author] I thank the reviewer for their suggestion (and this is similar to a suggestion made by reviewer 2). However, I deem this to be more related to preference of writing style. I prefer to introduce my papers with something different that provides unique information that some readers may learn and benefit from rather than repeating the same information that is stated in most every paper on a particular topic. For instance, I would ask how many papers has the reviewer read on mercury that have included some form of "mercury is a persistent, bioaccumulative, and toxic contaminant..." (or some other very well noted anecdote on mercury) within the first few sentences? Here, I have introduced information on why mercury is such a unique metal (its quantum/electronic structure), information that is not well known even to many mercury scientists.

Moreover, I would argue that this has much "direct relevance to the study" as the behaviour of mercury in such saturated environments is influenced greatly by the quantum/electronic structure. For instance, the volatility of elemental mercury, which we link as a potential loss process of Hg from groundwater (and surface soils and water) is directly linked to elemental mercury not forming metal-metal bonds. It is my stylistic preference that the introduction remain in this form; I do feel strongly that individual writing styles still very much have a space and a place within scientific writing.

With reference to Ln 462-475. This was discussed at length between co-authors as is not a straightforward concept. As such, this paragraph was very carefully worded to ensure all readers could follow the explanation. We are very encouraged that the reviewer understands this behaviour, but it may not be as straightforward for other

readers. We wish not to shorten the discussion to ensure that all readers can follow it clearly.

Specific Comments

R1C4 - Ln 172-174: The tested HgCl2 concentrations are presented as estimates of original concentrations during the years of kyanization activity despite being three magnitudes below this concentration level. It is clear that this is a rough estimate and that lower concentrations of stock solutions were not feasible due to experimental constraints. However, can any data or calculations be referenced to support estimating a 1000x concentration loss in the porewater over 5-6 decades?

R1AR4 - The original concentration of the solution applied during the kyanisation process was 0.66% or 6600 ppm HgCl2 solution. Contamination at the site was from spills of this solution, which contains the very soluble HgCl2. Thus, we know solution ~50x stronger the EXP2 and >100x stronger than EXP1 were entering the top of the soil profile and this site operated from 1904-1965 with 10-20 T of Hg being lost to the soils and aquifer. These data are stated in the study and other referenced work examining this site. Without question, experimental constraints (time) was a major factor in selecting these experimental stock solution THg concentrations, but considering the history of the site we do deem them to be applicable to this site and others. This is supported by the fact that Miretzky et al., 2005 (the only other study Hg column experiments) applied similar concentration to a similar column experiment design. This paragraph has been updated as follows to better describe these considerations:

"Stock solutions were 46.1 ± 0.1 mg L-1 in EXP1 (n = 6) and 144 ± 6 mg L-1 in EXP2 (n = 12) and were selected for (i) time considerations (see Figure S1.6) and (ii) as these values balanced HgCl2 concentration applied during industrial activities (6600 ppm; spillages of this solution to the top of the soil profile) and as estimates of the original concentrations of HgCl2 contaminated solution entering the soil-groundwater system considering recently measured groundwater concentrations up to 164 ± 75.4 µg L-1 are still observed 55 years after closure cessation of the industrial activities at the site the solid-phase materials were extracted (McLagan et al., 2022)."

R1C5 - Ln 181, Table S2.2: Based on the 24 and 48-hour analyses, a decision was made to let the columns equilibrate for seven days before starting the experiment. The analysis shows that more than two days are needed to achieve equilibration. Is there any data to back up that seven days should be enough?

R1AR5 – Equilibrating the columns with the uncontaminated solution for 7 days was deemed a conservatively long period of time to ensure the equilibration between solution and solid phase based on the 24 and 48 hour assessments. We have provided all data both related to the preliminary experiments and the actual

experiments. This includes a very detailed summary of the preliminary experiments to ensure readers could follow our experimental design process and more easily replicate if desired. We do not have data confirming complete equilibration at 7 days (it would have been included if it was available). However, with the extended equilibration period, all co-authors deemed this would not have a significant impact on the results, an estimate that we stand by now.

R1C6 - Ln 242: How can the Authors know that no Hg0 loss occurs when analysis is performed on wet samples? A reference would be appreciated.

R1AR6 – This is a good pick up by the reviewer and an oversight on our part. This will be changed to: "All analyses were performed on wet samples to minimise any potential losses of Hg(0)." Of course this relates to the fact that if samples were dried (oven, air, or freeze drying) much greater Hg(0) losses can be expected due the volatility of Hg(0), which is in turn linked to its quantum/electronic structure.

R1C7 - Ln 320-330, Figure 2: With the eluate concentration only reaching 91% of the stock solution concentration, the Authors state that EXP1 likely did not reach equilibrium. Despite this, the Freundlich model fit to EXP1 indicates that the maximum totHg eluate concentration was reached when ca 5.7 L solution had passed through, after which the increase in Hg in the eluate ceased. If EXP1 did not reach equilibrium, would a maximum eluate concentration be passed? Or, is there reason to doubt the appropriateness of the model for the sorption phase in EXP1?

R1AR7 – The model is fitted to the data, if we were able to extend the equilibration period further, the model would be different, but it is highly unlikely a different model (non-Freundlich model: i.e., Langmuir) would have fit the data better. Other data fitting models were applied to both the sorption and desorption components of both experiments, but in both cases Freundlich and exponential decay functions, respectively, were easily the best fits. Moreover, Freundlich functions are common in describing column experiment breakthrough curves (e.g., Pan et al., 2005).

R1C8 - Ln 557: How can it be known that Hg0 exceeded solubility after the 25% breakthrough if Hg0 was only qualitatively measured? If this is a speculation, that should be indicated more clearly.

R1AR8 – This is an excellent point. We will use the term "semi-quantitative" in place of qualitative to assess the liquid phase speciation analysis and make this change throughout. While these data have a high uncertainty, Hg(0) is the species that we have the most confidence in, particularly in the case of these experiments. Hg(0) is measured simply by purging the sample with nitrogen gas directly into the CVAAS (without any addition of BrCl). In these experiments, the CVAAS was in the adjacent room to the room where the column experiments were conducted. All Hg(0) analyses were performed immediately (<30 mins) from sample collection. Thus, we have a higher confidence in these Hg(0) results than we had from the McLagan et

al., 2022 paper that required fresh (unstabilised) samples to be transported from the site (across Germany) and up to 5 days could pass resulting in potential loss of Hg(0). The following sentence will be added to the methods:

"Confidence in liquid phase Hg(0) concentrations is higher than for other species, as these result from purging untreated samples of Hg(0) with nitrogen gas directly into the CVAAS. All Hg(0) samples were analysed within 30 mins of sample collection."

R1C9 - Ln 571-575: While I support the conclusions, the calculations for the Hg0 production in the contaminated sediments hinge on the assumption that the experiments will mimic what is happening in the contaminated site sediments. Several factors can be envisioned that vary between these two settings and which may play a role: The grain size composition (see general comment above), oxygen abundance, and changes in, e.g., temperature and light conditions. A discussion about the implications of these differences would be a welcome inclusion.

R1AR9 – Justification for the removal of the coarse fraction has been made. McLagan et al. (2022) and Bollen et al. (2008), both show this is a consistently oxic aquifer. In terms of light, the experiments were performed in a basement lab and lights turned off whenever staff were not in attendance. There is minor potential for photochemical reductions on materials at the surface of the columns as the columns were transparent (opaque columns were considered but dismissed due to the inability to monitor the columns for breakup and/or preferential flow paths). Temperatures in the aquifer were consistent between 12 and 14 deg C, while temperatures in the lab were around 20 degC. Again, there could be some minor influence from the 6-8 degC on reductions.

To be clear, column experiments do provide a more representative reconstruction of real world conditions than batch experiments, which our group have already conducted. Nonetheless, they are still not the actual conditions, and we would like to highlight here that these differences between lab experiments (batch and columns) and real world conditions are very well known and heavily discussed within the literature over many decade (perhaps close to a century) of research on movements of analytes in soils and aquifer (and briefly in the introduction to the manuscript). Therefore, we do not feel another in-depth discussion on these limitations is warranted.

One final point here is that our research group is one of the only groups to focus on the study of mercury dynamics in aquifers in "the real world" (along with the group of Dr. Carl Lamborg). We have collected, analysed and published extensive field data from groundwater wells and soil cores from this and other sites. Yet these data themselves are not entirely conclusive (as is the case with all environmental field studies). Thus, it is critical that we use other tools such as column experiments to complement field studies and generate more holistic, total systems assays to improve our understanding of mercury biogeochemistry in soil-groundwater

systems, which is the ultimate goal of this work (and these field studies); an objective that the reviewer themselves in their opening paragraph have stated that we have achieved here.

R1C10 - Furthermore, the calculations for the back-of-the-envelope calculation could be included in the supplementary information to clarify how this estimation was done.

R1AR10 – Figure S9.6 and Table S9.1 will be added to the SI showing the details (including the percentage of the integrated area of the peaks; 0.11% Hg(0)) for Hg(0) and Hg(II) peaks from the mean PTD curves for all columns from EXP2. The text referring to the *back-of-the-envelope* calculations has been updated as follows:

"If we conservatively assume conservative values fora mean depth of contamination of 2 m (aquifer ≈3-4 m depth; Bollen et al., 2008; McLagan et al., 2022), mean THg concentration of 2 mg kg-1 (solid phase THg concentration of 2-162 mg kg along the contaminated aquifer; Bollen et al., 2008), the fraction of Hg(0) produced per day is 0.01 – 0.001% of the THg (based off 0.1% Hg(0) peak integration of total peak area of mean PTD curve from EXP2; see Section S9) based off data from Bollen et al., 2008; McLagan et al., 2022), and the same bulk density and flow rates as in our experiments, we can produce a back-of-the-envelope estimate of the mass of Hg(0) produced and potentially lost from the aquifer to overlying soils."

R1C11 - S1, Background Investigations section: It is stated that the equipment was tested for DOC to investigate the origin of the discoloration. What was the result of these tests?

R1AR11 - "DOC measurements from deionised water leaches (after initial rinsing all equipment with dionised water) were below limits of detection." This sentence will be added to Section S1.

Technical Corrections

R1C12 - Ln 116: Remove the "from".

R1AR12 – This will be changed to "source from an area adjacent to a former…"

R1C13 - Ln 155: "Names" should be "named."

R1AR13 – will be changed.

R1C14 - Ln 246: The protocol summary makes more sense if the relative rather than the absolute volume of BrCl in the modified aqua regia is specified.

R1AR14 – The part in parenthesis will be changed to "(8 mL HCl, 3 mL HNO3, and 1mL BrCl)".

R1C15 - Ln 338: Insert "to" in "species used to generate stock solution".

R1AR15 – will be changed.

R1C16 - Figure 2: The dashed lines are faint, and the color contrasts between the Freundlich and exponential decay functions are hard to discern. It is suggested that the figure be edited to make it easier to interpret.

R1AR16 – Figure 2 quality and clarity will be updated to reflect this request. Similarly, at the request of the associated editor, the quality/resolution of all figures has been improved.

R1C17 - Ln 364: "EXP" should be "EXP1".

R1AR17 – will be changed.

R1C18 - Ln 394: "recovers" should be "recovery".

R1AR18 – will be changed.

R1C19 - Ln 407: A period sign is missing.

R1AR19 – will be changed.

R1AR20 - Ln 428, Ln 476: The use of "that" or "this" in the first sentence of the paragraph, referring to the end of the previous one, impacts readability.

R1AR20 – updates to these sentences will be made accordingly.

R1C21 - Ln 434, 495: "Overtime" should be "over time".

R1AR21 – will be changed.

R1C22 - Ln 464-464, figure 3: It is suggested to label the panels in Figure 3 a) and b) and make the corresponding references in the text.

R1AR22 – Panel Labels will be added.

R1C23 - Ln 567: "...that fraction..." should be "...that the fraction...".

R1AR23 – will be changed.

R1C24 = References: The Miretzky 2005 reference is not listed.

R1AR24 – will be added.

R1C25 - S1: There are references to Appendices A 3 and A 4, but no appendix is included apart from the supplementary information. It appears that the S1:3 and S1:4 figures, respectively, are what is meant to be referenced.

R1AR25 – will be changed.

R1C26 - Fig S1.5: While not critical for a SI figure, the figure could be improved by shortening the X-axis (to 100 instead of 180 minutes).

R1AR26 – This will be updated to reflect the suggestion.

R1C27 - S1: The "results of the preliminary test" section: The takeaway, aided by figure S1.6, is that the concentration was ramped to reach a concentration where the eluate concentration was high enough for the experiment. This is, however, not clear when reading the text and should be clarified.

R1AR27 - The figure caption states: "Three different initial solutions were tested. In the first section (white), a solution with 1.7 ± 0.2 mg L-1 was used, in the second section (blue) 7.5 ± 0.8 mg L-1 and in the last section (violet) 49.3 ± 4.4 mg L-1."

The text states: "The pre-experiment was divided into three sections, where the initial solution had different concentrations (1.7 ± 0.2 mg L-1, 7.5 ± 0.8 mg L-1, and 49.3 ± 4.4 mg L-1). At this stock solution concentration, the increase in the liquid phase THg concentration was very little after 10 days. Hence it was deemed too low, and the stock solution increased. At 7.5 ± 0.8 mg L-1, the concentration in the liquid phase reached 32.8% of the maximum possible concentration of 7.5 ± 0.8 mg L-1 of the initial solution after four more days, despite 9 L of 1.7 ± 0.2 mg L-1 already being added to the columns. Thus, the stock solution concentration was increased again to 49.3 ± 4.4 mg L-1."

We are unsure that we can be much clearer on this, especially when the reviewer themselves are stating that they understand?

R1C28 - S2: Table S2.2 header "Elements" should be specified to clarify that wavelengths are listed below.

R1AR28 – will be changed.

R1C29 - S2: "Table S1.5" should be "Table S2.5".

R1AR29 – will be changed.

REFERENCES:

Lewis, J., & Sjöstrom, J. (2010). Optimizing the experimental design of soil columns in saturated and unsaturated transport experiments. *Journal of contaminant hydrology*, *115*(1-4), 1-13.

Pan, B. C., Meng, F. W., Chen, X. Q., Pan, B. J., Li, X. T., Zhang, W. M., ... & Sun, Y. (2005). Application of an effective method in predicting breakthrough curves of fixed-bed adsorption onto resin adsorbent. *Journal of Hazardous Materials*, *124*(1-3), 74-80.

Cucarella, V., & Renman, G. (2009). Phosphorus sorption capacity of filter materials used for on-site wastewater treatment determined in batch experiments–a comparative study. *Journal of environmental quality*, *38*(2), 381-392.

Zhu, T., Jenssen, P. D., Maehlum, T., & Krogstad, T. (1997). Phosphorus sorption and chemical characteristics of lightweight aggregates (LWA)-potential filter media in treatment wetlands. *Water Science and Technology*, *35*(5), 103-108.

Bollen, A., Wenke, A., & Biester, H. (2008). Mercury speciation analyses in HgCl2-contaminated soils and groundwater—implications for risk assessment and remediation strategies. *Water Research*, *42*(1-2), 91-100.

McLagan, D. S., Schwab, L., Wiederhold, J. G., Chen, L., Pietrucha, J., Kraemer, S. M., & Biester, H. (2022). Demystifying mercury geochemistry in contaminated soil–groundwater systems with complementary mercury stable isotope, concentration, and speciation analyses. *Environmental Science: Processes & Impacts*, *24*(9), 1406-1429.