Dear editors and reviewers,

In this letter, we will give a point-by-point response (in blue) to the different remarks raised by the two reviewers, followed by a list of changes with line numbers.

RC 1 – Response

De Donker et al. present a novel and ambitious inverse methodology to infer maps of erosion rates from mineralogical data derived from XRD analysis of suspended sediments. Their method defines a model matrix, which uses bedrock mineralogy data also from XRD analysis and a geological map to characterise the catchment, and then uses a non-linear inversion to infer erosion rates based on the mineralogical data of the suspended sediments. They provide a thorough analysis of the inversion method, including testing two different inverse schemes, as well as testing a wide range of various parameter values to explore their limitations. Having done this, they then apply their method to the Gornergletscher catchment, presenting an erosion rate map for the area. Generally I found the manuscript to be well written, and will likely be suitable for publication following some further exploration/explanation of the limitations of the inversion.

General comment: We thank the reviewer for the positive and constructive feedback, for the attentive reading of our manuscript and for the helpful suggestions.

Firstly, I think the authors do an excellent job of acknowledging and testing many of the limitations of the inversion, and so I commend them for this. However, I found some of the presentation of the scatter plots in figures 7-10 to be slightly unintuitive. It would be really helpful, perhaps in a supplement, or alongside these figures, that the **resulting erosion rate maps of the synthetic tests** are shown. Of course, it would be nice to see all of them, but perhaps 3 for each parameter tested would aid the reader in understanding how different the inversion result can be. For example, what do the three erosion rate maps look like for the three different geological map inputs? Perhaps the results are very similar, but at the moment, it is difficult to infer how much variation is possible for the model results based on different inversion parameters being changed.

Reply: We have now provided the resulting posterior erosion maps in Appendix B and refer to them in the main text (l. 285). This allows the reader to see example outputs for each parameter scenario, thank you for the suggestion.

On this note, I perhaps would also like to see a more complex synthetic test set up than the one presented. Something akin to a checkerboard test might be ideal, or a scenario where there are more than one peak in erosion rate. From what I can tell, the present synthetic test has high erosion rates across two similar lithologies

(Stockhorn-Turftgrat-Gornergrat and ZSF ophiolites?), and low rates elsewhere. This is quite a simple set up. Would the inversion scheme be able to identify **two different peaks of erosion** that are spatially discrete within these two units? What about three peaks spread out across the catchment? Hence, I would like to see a slightly **more complex synthetic erosion rate map tested**. Having said this, I did like the testing of two different inversion schemes on the synthetic data to decide which one is more suitable – nice analysis.

Reply: We have added a test using a true erosion map with three Gaussian peaks in Appendix A and refer to it in the main text (l. 281–282). This demonstrates the limitations of the lithological setup clearly. We did not implement a checkerboard pattern, as it would conflict with the smooth erosion-domain assumption.

Finally, I felt that the XRD data needed explaining a little better and perhaps slightly more exploration of the associated errors. For instance, the **number of bedrock XRD analyses** is not stated. I wondered whether if only one per lithology were analysed, how different two samples from the same lithology could be, and how much error this could introduce? If only one is used per lithology, is the assumption that each mapped lithology is homogenous fair? From what I can tell from the unit descriptions they can be quite variable. How is this variability accounted for? Hence, I would like to firstly see **greater detail given for the acquisition of XRD data and mineralogy data**, and perhaps some exploration of inversion results akin the analysis presented in figure 8 for an **error introduced when the measured bedrock mineralogy is different from the true bedrock mineralogy**. Here you could synthesise a sediment mineralogy based on one set of bedrock mineralogies, and then randomly change each mineralogy in the A matrix by some value of the 'error'. I am not sure whether this is reasonable or not however, as I am not an expert in XRD, and I am not sure how the bedrock XRD data was collected.

Reply: We have added error bars on the normalized XRD peak-area data for lithological units where multiple samples were available (Figure 11). These show that, while variability exists between samples, the overall pattern remains consistent. Using forward-inverse tests with intersample XRD signal averages in the forward model and single-sample XRD signals in the inverse model, propagating this additional uncertainty into the posterior solution via the data covariance, we obtain posterior erosion maps that closely reproduce the true erosion pattern, with only a moderate increase in misfit (as described in l. 368-370). We also included additional details on the source sampling strategy in Section 3.2 (l. 364-370) to clarify how variability and potential error are accounted for.

One final comment, the introduction is duplicated. One needs to be removed, and the section numbers redone.

Reply: Thank you for the attentive reading, the duplicate introduction has been removed, and section numbering has been updated accordingly.

I hope the authors find these help to improve the manuscript. I also outline a few line comments below.

Line 61 – Should be an in-text citation.

Reply: Replaced with an in-text citation.

Line 270ish Gorner-gletscher then gornergletcher, needs to be consistent throughout the manuscript.

Reply: All instances of “Gorner Glacier” and “Gorner-gletscher” have been standardized to “Gornergletscher.”

Figure 7. Originial geology – should be original.

Reply: Corrected “Originial” to “Original.”

We thank the reviewer again for the constructive suggestions, which have helped improve the clarity and completeness of the manuscript.

RC 2 – Response

Dear editor, dear authors,

I really enjoyed reading this article which presents a very original and potentially promising fingerprinting approach that involves cost-effective XRD analysis on fine sediment.

The manuscript is very well-written and presented (although the introduction is repeated), with a very clear argument and beautiful figures that are efficiently used to support the argument. I like the presentation of the methods, the validation that uses both synthetic and natural datasets, and the discussion of the opportunities and limitations.

General comment: We thank the reviewer for the positive and constructive feedback, and for the careful reading of our manuscript. We also thank the reviewer for the suggestions on additional references and discussion points.

My main comments are about the limitations. Firstly, I would like to know more about the **grain size** of the sediment sampled (suspended sediment), and how it compares to the ground bedrock samples. I assume a significant part of these grains will be in the silt/clay fraction, where minerals formed as a result of **chemical weathering** may abound. The presence of such minerals in the sampled suspended sediment may cause errors, as they wouldn't feature in the crushed bedrock. This may be an argument that supports the use of this method in places with low chemical weathering intensity (as in this study – mountainous / polar regions), but may prescribe its use in other areas (tropical)? Can the choice of target minerals help avoid this problem?

Reply: We have added a more detailed description of the grain size of suspended sediment in Section 2.4. As measured by laser diffraction (Malvern Mastersizer), the sediment is dominated by silt-sized particles (median d50 ≈ 20–30 μm), with only a small clay fraction (<2 μm) of ~6–7%. This indicates that most of the sediment lies within the 2–63 μm "glacial flour" fraction typical of glacially produced sediment. We have added a discussion in Section 4 on the implications of this for the application of the method in other environments, where careful selection of target minerals and a focus on well-mixed fluvial sediments may help mitigate potential biases.

Secondly, I feel that there could be a bias associated with the way sediment grains of various sizes may be preferentially produced by the weathering/erosion of some particular rock types, and/or the action of specific processes. I note that this probably applies to all fingerprinting methods, but could be particularly relevant to this study that stresses the need for strong mineralogical differences between units for the method to produce satisfying results, and the mention of glacial processes. This work illustrates nicely that this can cause enormous problems, with the low zircon concentration in the ophiolite units mentioned as a potential cause for the highly unstable results of the inversion of the zircon data (i.e., erosion rates > 8,000 mm/yr). **Could the weathering/erosion of granite produce more grains in the sand fraction, while the weathering/erosion of serpentinite and slates/schist produce more grains in the silt/clay fraction?** This has been observed in the abrasion of pebbles of different lithologies, so could potentially apply here? Additionally, glacial processes may produce more clay-size sediment (glacial flour) compared to other processes elsewhere in the catchment. This is why I feel the reader needs to know more about the grain size of the sediment analysed; these potential issues could also be discussed further in the discussion, if the authors felt it would be useful.

Reply: We agree that preferential production of sediment grains by lithology and process is an important consideration. We have clarified in the manuscript (Section 2.4 and Section 4) that glacial processes produce a significant proportion of fine silt-sized particles ("glacial flour"), and that this is consistent across lithologies in our catchment. We also discuss that differences in grain-size production between lithologies could influence the representativeness of suspended sediment and that this limitation is relevant to all sediment fingerprinting approaches. In our dataset, the focus on primary rock-forming minerals in selected tracer minerals and the limited amount of chemical weathering reduce the potential bias, but we acknowledge that in other catchments or under different erosion regimes, grain-size effects could be more pronounced.

Finally, some minor comments:

- the following paper may be relevant: Sediment tracing in the upper Hunter catchment using elemental and mineralogical compositions: Implications for catchment-scale suspended sediment (dis)connectivity and management. Kirstie Fryirs, Damian Gore. Geomorphology 193 (2013) 112–121. http://dx.doi.org/10.1016/j.geomorph.2013.04.010.

  Reply: We have added the suggested reference: Fryirs and Gore (2013, Geomorphology, 193, 112–121) in the introduction discussing different existing sediment tracing approaches. Indeed, this work is relevant as it also applies XRD-based fingerprinting to suspended sediments. Our approach differs in several key aspects: we use binned "raw" XRD data rather than proprietary software outputs (Fryirs and Gore use the "PANalytical HighScore Plus" software), which reduces costs and simplifies data processing. Furthermore, whereas Fryirs & Gore focus on source contributions, their resulting map remains spatially discrete (showing the cluster locations) and they do not produce spatially explicit erosion rate maps, which is the main output of our inversion methodology.

- Some methods are criticised because "they link sediment directly to its origin, without modelling storage or transport", but I don't think the new approach models storage or transport either?

  Reply: We clarified that storage and transport effects represent the main limitations of our method, as is the case for many fingerprinting approaches (l. 29).

- The abstract could be more specific about the specific findings of the work

  Reply: Two parts have been added in the abstract to highlight the findings.

I hope you find these comments helpful.

Mikael Attal

We thank the reviewer again for the helpful comments, which have allowed us to improve both the clarity and completeness of the manuscript.

Line by line changes

l.1: Updated author address

l. 9-11: Highlighted some findings in abstract

l.15: Removed duplicate of introduction

l.27-28: Added ", which represents the biggest limitation to the method proposed in this article."

l.69: Changed to in-text citation

l.208-228: Added "An important limitation of many sediment fingerprinting approaches is the dependence of tracer concentrations on grain size and post-depositional processes (D'Haen et al., 2012; Laceby et al., 2017; Collins et al., 2017). These issues are commonly referred to as grain-size effects, tracer fractionation, and source fertility problems (Garzanti et al., 2009). Specific concerns include a) the production of new minerals through chemical weathering in the finest fraction of sediments and b) lithology-dependent grain-size production, while suspended sediment samples predominantly capture fine particles. Three lines of evidence suggest that these effects are limited in our research setting. First, the grain size of suspended sediment in the Gorner River was measured using laser diffraction (Malvern Mastersizer). The suspended sediment is dominated by silt-sized particles, with median grain sizes of 20–30 μm (d50 = 18–31 μm) and only a small clay fraction (<2 μm) of 6–7%. Most of the sediment therefore lies within the 2–63 μm silt fraction typical of glacially produced suspended sediment ("glacial flour"). These grain sizes indicate that the sediment is largely produced by physical comminution beneath the glacier rather than by chemical weathering. Second, the mineralogical analysis focuses exclusively on primary rock-forming minerals (e.g., quartz, feldspars, pyroxenes, amphiboles, micas, serpentine minerals, garnet). Minerals that are typical products of chemical weathering (in this climatic setting: vermiculite, smectites or mixed layers in small amounts) were not included in the selected tracer minerals. The presence of chemically weathered minerals in the clay fraction is therefore unlikely to bias the fingerprinting results. Third, we tested whether mineralogical signals vary with grain size by comparing pump-sampler samples with depth-integrated samples collected at the same site. Pump samplers collect sediment at a fixed height within the water column and therefore tend to sample a narrower grain-size range, whereas depth-integrated samples capture a broader grain-size distribution. If mineral abundances were strongly grain-size dependent, we would expect large differences between these two sampling methods. However, the mineralogical signals are very similar between the pump and depth-integrated samples, suggesting that grain-size dependent mineralogical fractionation is limited in this dataset."

l.284-285: Added: “The quasi-Newton scheme was also tested using a more complex synthetic erosion rate map; the results of this experiment are presented in Appendix A.”

l.290: Added: “For the detailed erosion rate map results, we refer to Appendix B.”

l.363-370: Added: “For the Monte Rosa (granite) unit, four samples have been collected (two sand samples and two rock samples that were crushed before XRD analysis); for the Zermatt Saas Fee ophiolites (metabasites, eclogites) unit, one sand and one rock sample have been collected; for all other units, a single sample is used. The error bars in Figure 11 represent the inter-sample variability for these two units. Forward–inverse tests indicate that this variability remains sufficiently low to allow recovery of the true erosion pattern when this uncertainty is propagated via the data covariance matrix (Cd) as tracer-specific standard deviations. In this case, the misfit sum(| etrue − epost |) ranges between 260 and 330 mm/y when using single-sample fingerprints in the inversion and inter-sample averages in the forward model, compared to 210 mm/y when inter-sample averages are used consistently in both forward and inverse models.

l.449-453: Added: “Finally, mineral-based sediment fingerprinting is most robust when applied to well-mixed fine sediments, such as suspended sediments in river systems or glacial marine deposits (Andrews et al., 2023). In contrast, environments characterised by strong chemical weathering or poorly mixed coarse deposits (e.g., moraines) may exhibit grain-size dependent mineral fractionation or the formation of secondary minerals, which could bias interpretations (Iverson et al., 1996; Haldorsen, 1981; Von Eynatten et al., 2012; Caracciolo et al., 2012; Crompton et al., 2019).”

l.470-495: Added appendix A and B to include reviewer suggestions: a forward-inverse test with three erosion peaks in the true erosion model and displaying the resulting maps for the scenario tests.

Figure 11: Replaced figure to include errorbars, normalised instead of "raw" XRD fingerprints

We thank the editors and reviewers for their time and constructive feedback, which has significantly improved the manuscript.

Best wishes,

Fien De Doncker and co-authors