Puhl and co-authors present a new time-series of surface suspended sediment data (2015-2018) from a series of nested basins in the Koshi River basin (Nepal) affected by co- and post-seismic landsliding following the 2015 Gorkha earthquake. They supplement these data (sediment grain size, concentration) with grain size data from mass-wasting deposits associated with the 2015 event, and carry out end-member modelling analysis to determine whether and where a grain size signal from the earthquake is evident in post-event fluvial suspended sediment loads. Their results suggest that any signal relating to co- and post-seismic landsliding in the coarse tail of the suspended sediment load was relatively short-lived, and was largely diluted by the mountain front. The authors interpret this result to mean that fine sediment evacuation has either been relatively effective, or that material remains stored on hillslopes/low-order channels. Overall, this is an interesting dataset and I thought the use of the end-member modelling was a nice way to try and pull apart some of these spatial trends and think about how fine sediment is being routed through this heavily perturbed landscape. I think there is scope to sharpen up the layout of the paper, and in particular clarify several aspects of the methodology/analysis. Figures are generally of good quality, although I have added a few suggestions below that may help improve clarity. The text is generally well written, although in areas the core message could be sharpened or rewritten to help guide the reader through. I also felt like several of the figures in the supplemental information document were really useful in supporting the authors' interpretations, so would ask them to consider moving them into the main body of the text if they felt appropriate as well. I have three larger comments/areas, and a more extensive list of minor comments/suggestions that I hope are useful to the authors moving forwards.

<u>Larger comments (these are noted in more detail below):</u>

1) The elephant in the room here is the nature of the suspended sediment sampling. If the authors could be a bit more upfront in the manuscript about the potential errors/uncertainty that the surface nature of their sampling might have, it would be very useful (rather than just a concluding comment in the discussion). Even some kind of sensitivity analyses would add significant value.
2) The other more major aspect I have several comments on below is in regard to the level of detail on some of the sampling/methodological aspects, which I think would aid the reader (e.g., descriptions of the mass-wasting deposits, or even a few photos to illustrate their nature).
3) The layout/structure of the methods and results sections could be more closely aligned with a clearer distinction between directly measured and secondary/derived datasets. For example, the methods section flows from field/primary data on grain size to EMMA analysis using these grain size data. It would be intuitive to see the results reported in a similar fashion to how they are presented in the methods, where new primary data are clearly presented and

any subsequent analyses using these data then follows. The first section 4.1 using the EMMA results would perhaps benefit from contextualisation using the grain size data first.

Detailed comments by section:

Introduction

L31. True, in supply-limited settings anyway. This is then mentioned in the second paragraph, which does feel a little repetitive so would consider revisiting this opening section and streamlining.

L35. The 'type' of sediment supply process – perhaps the 'nature' would be a better way to describe, as type suggests some kind of classification.

L83. You also end up discussing the GLOF event quite a lot, so wonder if worth mentioning here as well? Something about a unique opportunity to consider back-to-back events?

Section 2.1/2.2 – could these be merged into a single section?

L123. By landslide rates, are you considering numbers/frequency/total areas or volumes? If any further info such as landslide size/area are available, it would be helpful to add to Table 1 as well.

Methods.

The methods were not totally clear (& therefore, reproducible) so some expansion would be helpful.

- How were water surface samples collected? Was this just a bottle dunk? Were they taken at the channel edge or thalweg? Were any repeat samples taken on a single day to compare differences in values?
- Which landslide deposits were sampled (i.e., provide lat/lon), and how were they sampled? Was material just scraped from the surface, or was any from depth? Was the deposit predominantly fine grained, or is there a coarse mode that wasn't sampled? How many/how much sample(s) per landslide? Was there any evidence of variability in the fine grain fraction over the surface (i.e., how representative was that sample)? Were they in different lithologies/climatic zones?
- L172. How specifically were subsamples produced in order to preserve characteristics?

- L184. Does this maximum values also correspond to typical values observed in the suspended record (i.e., are coarser sizes than 2.5 mm being detected)? Some greater justification as to why this truncation is reasonable beyond the apparatus limits would be useful.
- L189-191. This statement needs to be caveated and rephrased, with further details added. Is this calculated using a daily Q? Some comment on the daily fluctuation in Q is needed if so, and how that might influence any calculations (it may not, but useful to state if the case). The biggest thing here is that these are not calculations of total sediment load – they only represent suspended load, and also rely on the assumption that vertical mixing is sufficient that any surface sample is representative of the depth-integrated concentration. These are then presented in Fig. 4 without any consideration of potential error or uncertainty. I would suggest a more detailed discussion on potential limitations/caveats of this approach, and possible effects on any derived/estimated values (later comments on this).
- Section 3.2. It would be useful to start this paragraph explaining why this approach was chosen over others (i.e., how does it suit the types of data available) to address the overall aim of the manuscript.
- Section 3.2. I appreciate that the details of how the approach works is published elsewhere, but I found the explanation here quite difficult to follow, and it might benefit from a simple schematic to explain the basics of this approach to the non-expert, before a clear series of statements on how it was applied to these new data and what the key outputs were.
- 'in other words' (L203, 217) and 'fundamentally' (L218) – feels like some of the sentences are being re-explained.

Results

Please see my earlier comments on the structure of the manuscript. I see the appeal of presenting the EMMA results first, but the results from this would be better contextualised if the patterns/magnitudes of grain size in the measured data were first presented for each system.

- L237. First mention of 'EM' naming protocol – define here (not in L242).
- Fig.2. The identification of the modal values are helpful here! Otherwise please add a label to the y-axis and be consistent with how μm is reported (micrometer/ μm seems to be used interchangeably throughout). It would be helpful to use the same figure formatting as used in Fig. 3 though, which contains slightly more detail (e.g., % of each EM) and stretching out of the x-axis is more helpful with identifying individual EMs if the plots are stacked vertically as in Fig. 3.

- The caption on Fig. 3 also states that 2015 landslide deposits were sieved down to 600 microns – how much material was removed (i.e., >600 microns), and how might that influence any interpretations/results if the coarse tail has been removed? Does this potentially interfere with the EMMA?
- L256. It would be really cool to see some of the fluvial EMs stacked with the landslide/DF ones to show how well they overall, but appreciate that would be an unnecessarily messy figure which probably wouldn't be worth making…
- Section 4.2. The first few sentences here are grammatically incorrect. Please rephrase.
- Section 4.2. On L275ish, the 'peak sediment flux' is estimated at 23-45 tonnes per second? Is this correct (this seems huge?)? I would also take care with how this is presented and phrased. This is just for the suspended sediment. It also isn't clear what temporal resolution Q data were used. Would it be more accurately to call it an instantaneous flux? Note that these were described as sediment loads in the methods – load and flux are not the same, so please rectify language for consistency.
- L276ish. It would be useful to show the measured sediment concentrations somewhere. I think this would help separate out the increase in calculated sediment flux that is driven by an increase in just Q versus one which is also driven by an increase in suspended sediment concentration. I appreciate you can work this out if you compare Fig. 4A subpanels but would be easier if the data were just shown somewhere given that they are measured.
- Section 4.3. Ok, I can see why you are presenting the EMMA results first, so you can compare to the grain size data here, but I would still argue that it makes more sense to present the new empirical data separately at the beginning. This also answers my earlier comment around whether you can plot the EMMA outputs against the measured data. If you are going to do a comparison like this, it would be useful to state why you are using the specific EM that you do to provide the reader with a quick reminder of the key findings from the EMMA.
- L291. Please avoid relative phrases about grain size such as 'predominantly coarse grained'. Use absolute references/values instead. Please check throughout the results section for other instances of this.
- L293. No clear pattern in what? State. These first few sentences of 4.3.1 are completely relative so please rewrite this with something supported with quantitative examples/values.
- L294-299. Interesting! Why do you think the grain sizes transported in high/low flow seasons is so comparable?
- Fig. 5. Given the gaps in some of the data during the low flow windows, would it not make sense to mask these periods out, or not link the data points together

with a solid straight line, rather than it look like a grain size trend? Alternatively used a dashed line to connect up points in the middle panel?

- Section 4.4. Perhaps include a couple of sentences in the methods to explain how/why you explored seasonal patterns so this section doesn't feel completely out of the blue? Alternatively switch the first two paragraphs up here so it first explains *why* you decided to do it, rather than just stating what you did.
- L385. This concentration is based on the EMMA though, right? You also directly measured concentration, but I don't think this is shown anywhere. The language needs to be really careful about what these plots are showing.
- Out of curiosity, what were the differences in total ISM Q for each ISM season? Also for the non-ISM seasons (low flow). This might be useful to include somewhere to get a sense of how 'wet' and 'dry' seasons were through the study period.
- Fig. 9. Typo on x-axis for 'concentratio'

<u>Discussion</u>

- L418. The distribution as a whole was coarser, or just key statistics?
- L457. Is the absence of suspended sediment not a larger limiting factor?
- L458. But bedload and suspended load export timescales are very likely to differ, so this contrast shouldn't be a massive surprise? Some argument here perhaps reflecting on the grain sizes of the landslides that were sampled might be indicative here. Were they predominantly fine or coarse (i.e., > than the sizes you sampled?) grained?
- L500. Perhaps support with a suitable reference?
- L502….due to these down slope sorting effects?
- L551. I'm not sure I'd push as far to that it 'reveals the post-seismic response', but it offers valuable insights and new testable hypotheses. For example, how could future work identify whether the absence of signal is due to rapid evacuation of a relatively modest fine sediment supply, or just exhaustion of the most easily accessible deposits?
- L590. Both of the cited papers here do show some degree of vertical grain size coarsening/change, so I don't think this can be dismissed quite so easily. But this isn't just grain size this applies to, it is also the concentration gradient that needs to be addressed. In both instances (grain size, concentration), I don't think this can really be ignored and some more upfront caveating is needed earlier on in the paper, as well as the discussion. Having it here feels like a bit of an afterthought, given the potential it could have on calculations. For example, some of the near-bed suspended sediment concentrations in Morin et al. (2018) are >double the surface ones. Similarly, D90 values can vary by > an order of magnitude (see Fig. 5 in Morin et al. 2018). It might be that Q also has an impact

here – perhaps these effects are more/less obvious under different flow conditions?

Conclusion

- L617. It implies that fine sediment was, it doesn't say anything about coarse sediment. Without knowing more about the nature of the mass wasting inputs, this is tricky to evaluate.

Supplemental materials

Figure S1. Ah, here they are. Please could you modify the shape/size of the symbol based on landslide/fluvial/debris flow? It would make them easier to differentiate.

Figure S2. This is really helpful/interesting figure, so was surprised to see it buried away here. I think it really strengthens your interpretation at the end of the manuscript, so wonder whether it should be in the main body? Same goes for Figure S3.

Figure S8/10. Again, this is a really key figure for me when working through the interpretation of the data which shows the change in concentration rather than flux (which is shown in the main body of the paper). Potential to include this in the results?

Figure S11. Please could you add labels for which station is which? I actually think this is probably a cleaner way to present the data (relative to what is in the main text) and avoids erroneous interpretation by filling in gaps between samples. I'd consider reporting like this in the main text for Fig. 5/6/7/8 top panel?