To the editorial team at *Earth Surface Dynamics*,

I have completed my review of the manuscript "egusphere-2023-1278: Decadal-scale decay of landslide-derived fluvial suspended sediment after Typhoon Morakot" by Ruetenik et al. [https://egusphere.copernicus.org/preprints/2023/egusphere-2023-1278/]. The manuscript uses decades-long river suspended sediment and discharge measurements from a series of stations around Taiwan to investigate the impact of the 2009 Typhoon Morakot its resultant landslides on river suspended sediment delivery over time. The authors show that among a subset of focus stations after Typhoon Morakot, rivers generally had large increases in the rating-curve parameter $\tilde{a}$ (meaning that more sediment is delivered for a given discharge) and smaller decreases in the rating-curve parameter b (meaning that the amount of sediment delivered becomes less dependent on the magnitude of discharge), both compared to pre-Morakot averages. The peak for $\tilde{a}$ occurred within ~1 year, while the decrease in b peaked after a couple of years, and both values decayed back to pre-Morakot levels, with the most-affected regions decaying within a couple of decades.

Overall, after a close review, I have found the manuscript to be well-written, clear, and with few typographical or grammatical errors. It is clear that the submission has been treated with care. The work addresses relevant scientific questions to the readership of ESurf, presents novel and reasonable results, and has a method that seems technically sound and reproducible. While the manuscript requires little in the way of major changes, I have some minor comments for the editor's and authors' consideration. I have also included line-level comments. While I think some of my comments bear addressing, others are merely suggestions or observations.

I thank the authors for a well-written manuscript. I have enjoyed reading it, and I hope that my comments prove helpful.

Best regards,

Harrison Martin
Postdoctoral Scholar Research Associate
Division of Geological and Planetary Sciences
Caltech

Minor:

- The fundamental data of the study are centered around 87 river measurement stations in Taiwan. At different points in the manuscript, either the entire dataset, a subset of 24 stations split into North and South halves, or other subsets/splits are measured and reported-upon. At times, I found myself unsure of which stations were being analyzed and reported upon. I have noted in the line-level comments for some specific spots where I was unsure. Overall, I'd encourage more clarity about which stations are used for which analyses, and justifications for why.
- Many results are focused around the 24 focus stations. I'd like to see something explaining why these 24 were selected, evidence that they are fairly representative of the whole dataset, how they were split into North and South, and quantitative presentations of their associated parameters. In particular, part of the results hinges implicitly on the idea that the 24 stations and associated basins are comparable to one another, and vary only in that the northern 12 experienced less precipitation and less landsliding than the south. A series

of box-and-whisker plots, or table values with standard deviations, should be able to easily clear this up either at the end of Methods or start of Results. This would also help with justifying the split between North and South focus stations, and allow the landslide intensity data to be presented earlier (as they currently do not arrive until the final paragraphs/figure of the results).

- Some of the changes, such as the year-over-year changes in rating-curve parameters and Qs observed immediately after Morakot, would be more impactful to the reader if they were presented alongside some context of the significance compared to year-over-year changes before Morakot. Currently they are compared to pre-Morakot mean or median values, which by nature are central values without any context about variability. It would be nice to know if the magnitude of year-over-year changes is larger than, or within the range of, previous non-Typhoon "noise".
- The regressions (and interpretations) between rating curve and suspended sediment changes are done against landsliding intensity, as per the paper title. They are not, to the best of my knowledge, done against precipitation intensity, which is another reasonable mechanism that could be driving changes and for which data are available. It would be enlightening to regress changes against both to see if landsliding is the mechanism driving the change, or if both are correlated because they are both driven by precipitation. This is particularly interesting because there are a few Northern focus stations that apparently received precipitation but not landsliding, and yet reacted more similarly to the Southern stations than to the other Northern stations. Additionally, for most stations, the volume of excess sediment transported by rivers post-Typhoon (via suspended sediment alone) exceeds the volume of landslided materials resulting from the Typhoon. This suggests that non-landslided materials might also have made up a substantial part of the response, and this should be acknowledged and discussed. Regressing changes against rainfall intensity should help with this discussion.

Lines:

8-10: This is a bit tricky since it's an abstract and necessarily without elaboration, but I'd suggest offering a brief introduction to the coefficients here. This can even take the form of a brief phrase in a sentence, along the lines of " […] the discharge-normalized rating curve coefficient $\tilde{a}$ was higher […] the first year after Morakot (2010), indicating heightened sediment concentrations for given discharges." And similarly for *b*, saying it indicates a greater dependence of sediment concentration on discharge, or an increased sensitivity.

10: This could easily just be me, but seeing "Morakot (2010)" made me think it was a citation.

15-16: I think that "Values of a˜ tended to decline faster in basins with more intense landsliding" from 15 is repetitive with "Shortly after Morakot, changes in $\tilde{a}$ and *b* tended to be larger in basins with more intense landsliding"?

22-27: An effective summary and an interesting result. I appreciate the clarification that just because the increase in sediment transport disappears on a short timescale, that does not mean that the sediment deposited by landslides disappears on short timescales. Not sure that it needs to be in the manuscript, but it does make me wonder where sediment transport rates and sediment availability become detached in these sort of systems. Perhaps it winnows away all of the fine material and becomes transport-limited over short timescales, or maybe much of the landslide material is at elevations inaccessible to the river?

37: To the best of my understanding, Yanites et al. (2010) dealt with bedload transport, not suspended sediment. Additionally, unless you're referring to a different finding, according to Figure 2D the longest timescale for removal of (bedload) material is ~600 years. I'd suggest a different citation for this idea.

44-47: Interesting that such a big event only tripled the average annual pre-landslide sediment load. My intuition was that the increase would have been more significant.

85: Based on the manuscript's abstract and introduction so far, I get the impression that the manuscript is specifically focused on analyzing fluvial <u>suspended</u> sediment fluxes, so it might be worth adding that to this line.

106-110: This method of splitting up the data between north and south is framed as distinguishing between areas with lower and higher precipitation/landslide intensity, respectively. That's probably a good representation and valid, especially looking at Figure 1. It could be worth quantifying and reporting this here anyways to make this point quantitatively by presenting, for example, average precipitations per group). There may also be other factors beyond precipitation that varied systematically from south to north; Yanites et al. (2018: DOI: 10.1002/esp.4353) focused only on the southern tip, but found or discussed systematic latitude trends in maximum elevation, slope, local relief, channel steepness, erosion/exhumation rates, and channel width.

Figure 1: I find it a bit challenging to identify all of the small red dots because of their size, having the same color as the large red dots, and sometimes being apparently overlain by other dots. Perhaps enlarging them and using a third color could help. Additionally, I could see some readers wishing the colored contours using a perceptually uniform colorbar so as to not artificially introduce breaks, as jet tends to do [https://colorcet.com/].

126-127: Here is a nice, concise introduction to the meaning of the parameters; some version of this also mentioned above would be helpful.

131: I came back to this line after reaching section 2.3 and feeling that landsliding intensity had not been emphasized earlier in the methods. While it's mentioned in the Introduction and shown in Figure 1, I think it's worth making sure that you clarify that you are measuring the influence of Morakot-<u>induced landsliding</u>, and not Morakot itself (ex. precipitation). This fits better with the title of the paper and introduction. I'd also look at mentions in section 2.2 and the caption of Figure 3 to ensure you properly emphasize that you are estimating the effects of landsliding (due to Typhoon Morakot) on suspended sediment discharges.

134: Consider removing "applying".

135: I think that readers who have not constructed rating curves may not understand what centering means in an intuitive sense. Is it akin to horizontally or vertically shifting values in some log-space? It looks like it might be normalization, or comparison of values as deviations from a central value. A few words would be appreciated, especially since it seems Cohn et al. (1992) do not provide much in the way of explanation themselves.

143: Ahh! It seems this line contains the graphical explanation I was looking for above.

151: To each year's C and Q data <u>for each station</u>, right?

161: “sediment load” -> “suspended sediment load”

166: Where/how did you get or measure A for each gauging station? It isn’t mentioned as part of the WRA dataset on lines 95-96. I’m assuming a DEM and some routing toolkit, but it bears mentioning.

169-170: A good disclaimer. I think it would be worth a half-sentence about how far off you might expect the suspended sediment load to be from the total sediment load. Is suspended load 10% of the total load? 50%, 90%?

Eqn (5): Should E’s subscript be 1 instead of 3, since it is associated with downstream gauging station 1?

179-180: You say here that you apply Eqn’s 1-5 to each gauging station’s measurements each year, to yield estimates of parameters at each gauging station. However, in sections 3.1 and the first two paragraphs of 3.2, it seems that you only present the results (and statistics) for estimates of $\tilde{a}$, b, and Q_s for the 24 focus stations. Then, for the second half of 3.2, you say you calculated Q_s for all 87 stations and use them to calculate and present E for all all 87 stations. I’m unclear as to the selection criteria for the focus stations and why only 24 were presented and used for summary statistics, if it appears that values were calculated for all 87. Do we believe that the focus stations are representative of all stations?

188: The tense elsewhere has been past, but is present here. This may be intentional but it’s worth giving a look-over to ensure consistency.

188-193: The first sentence is a good topic, but the rest of this paragraph seems a bit out-of-place. It might work if you add what’s missing after the first sentence: a connection that says we are aiming to quantify the effects of Morakot by using changes in rating curves before and after a given storm. Then the following lines justify that it’s reasonable to anticipate seeing these changes in rating curves; in fact, we should expect to.

194-215: This is a nice and clear explanation of why average Qs rates pre- and post-storm aren’t sufficient to answer the research question. Only note is that you may be able to combine the final two sentences (213-215) into the paragraph above (208-212).

222-223: Marc et al. (2018) is cited twice in the same phrase.

222-224: “in which it is assumed” as used here applies to “corrections”, and not the catalogue estimates themselves. Is this intentional? I’d also define c and p here; you can still mention the values later in the paragraph, though if you did not do the estimations yourself (as you say on line 228) you might not need to mention the explicit values of p and c here.

229: Insert “for each station”

232: I think section 3 should include a summary of the differences between the North and South sections in terms of the “input” conditions, to communicate to the reader that they are comparable in all manners (drainage areas, typical discharges) except for the treatment condition (precipitation and landsliding intensity). Much of the following results relies on them being distinct in terms of precipitation and landsliding intensity, and presumably not in other ways. As such, why not include a figure with a couple of box-and-whisker or violin plots showing that there is a quantitative difference between the two groups in terms of precipitation and landslide intensity and how significant the difference is or isn't? That would make statements like the first sentence of Figure 4 caption easier to justify and write, since it will be in readers' heads already.

234-235: refer to Figure 4

242-244: Is there a figure showing this anywhere? Could be simple and useful to communicate the differences between the two zones. Either as two box-and-whisker plots, or with station # on the x-axis and $\tilde{a}/\tilde{a}_{pre}$ on the y.

252: This is entirely up to personal style, but if one wanted to avoid rhetorical or prompting questions and reserve them for research questions or prompting discussion sections, one way to re-write this could be "While this could account for the observation [...] at some stations, we consider this unlikely."

252-257: This is a good argument, but I'm not sure it's necessary unless you expect readers to bring it up unprompted. Could be shortened if you wanted, though I always respect including disclaimers. On the other hand, if it were very important to prove this, you could do a similar calendar split on each of the pre-Morakot years and see how common or uncommon such a within-year jump in $\tilde{a}$ is.

Figure 4 caption: The final sentence is a result, and is not shown in this figure, so I do not believe that this idea should first appear here unless it is accompanied by a reference to an associated supplemental figure. Is it shown or discussed elsewhere? Additionally, I'm not sure that the word "confirming" in the phrase "[...] confirming that basins with minimal landsliding experienced smaller changes in rating curves" is the most appropriate, as I'm not sure we've seen quantitative evidence yet that the northern and southern stations differ systematically in landsliding intensity beyond that shown visually in Figure 2. It almost looks like it might correlate similarly well to precipitation intensity.

Section 3.1: This might be personal taste as well, but I'd consider including a table of summary results/statistics between the north and south areas. The table could present many of the average pre-Morakot and post-Morkot 2009/2010 values for the different parameters of interest ($\tilde{a}$, b, Q_s). That would replace much of the writing, but I suppose that might make this section quite brief, and might not work for comparisons that rely on looking at values in 2011 (such as b in the south), so might not be the effort.

258-270: I think it might be helpful to put the magnitude of these changes into context for readers who are not as familiar with Taiwan or the parameters. It was not as big of an issue with values of $\tilde{a}$, where changes were on a factor of ~2-8x. Here, all we can see is that the changes at the southern stations from pre-Morakot to post-Morakot 2009-2010 were much smaller than the changes from pre-Morakot to 2011, and the northern changes from pre-Morakot to post-Morakot 2010 are somewhere in the middle. What we lack is an understanding of the relative scale between the largest changes we see here compared and the sort of changes over comparable timescales that would be expected from normal annual variations without a typhoon. An easy way to do this is to compare the magnitude of these changes to, say, the standard deviation of the values pre-Morakot or the average magnitude of changes between years. Essentially, when we see an average year-over-year drop from 0.59 to 0.29, is that an exceptionally strong signal, or is that within the range of noise we could see in pre-Morakot measurements and thus indistinguishable from random chance? It should be possible to do some sort of basic statistical test to see which of the changes assigned to Morakot are significant vs insignificant, given the pre-Morakot history in annual rating curve parameters.

281-283: Above, you compared post-Morakot values of rating curve parameters to the mean pre-Morakot values, but here you compare post-Morakot values of Q_s to the median pre-Morakot values. Any particular reason for the different method?

283-284: I think I know what you mean, but saying “rapid drop-off in Q_s after Morakot” almost sounds like it declined after Morakot. I think it’d be more accurate to say that the drop-offs occurred after the post-Morakot peaks. Additionally, here (as well as the caption for Figure 4) a reference is made to changes in annual discharge values. Are the timeseries of discharge, or even just annual averages per-station or for north vs south, shown anywhere?

287-279: I’m not sure if the “all but one” station mentioned in the first line is station S12 mentioned in the second line, or if these two lines are related to one another.

Figure 5: Interestingly, N1-3 have increases in suspended sediment discharge that look like those of the southern stations. I wondered if N1-3 are the south-most of the Northern stations. Based on Figure 2, it looks like these stations are southern, but not as much as N11/12. It also looks like these stations did experience elevated precipitation compared to the other Northern stations, but N2 and N3 experienced no landslides. This leads one to wonder if some of the increased Q_s observed is due to non-landslide associated mechanisms that also correlate to precipitation. It would be interesting to regress Q_s separately against landslide intensity and against precipitation.

Figure 6: How are these basins delineated? Is there a DEM involved? Is it from the WRA dataset?

295-307: I think that these results should be in a section of their own, and not part of “3.2 Suspended sediment discharge Q_s”. Additionally, though it was mentioned in the methods, it bears repeating here that this is basin-averaged erosion or deposition rates associated with suspended sediment, specifically (as opposed to total actual erosion which presumably would be higher as it included bedload, etc.).

296: remove “and”.

297: Is this station better-known as S# or N#? I suppose not, if it’s not one of the focus stations.

298: How does this compare to “normal”, non-Typhoon annual basin-average erosion rates for this type of setting?

Figure 7: Missing the word “of” in “By contrast, at seven of the twelve northern stations [...]”

325-337: I like this presentation, and I also think it’s a reasonable result to find that basins that over-deliver sediment have lower landslide intensities than those that under-deliver. I suppose this suggests that the range of Q_s amounts delivered by rivers exists over a narrower range than that of landslide intensities, and that the former is only moderately sensitive to the latter. Have you tried this same regression against rainfall intensity (which also has units of L^3/L^2)? That might give some insight as to where some of the non-suspended sediment is coming from. It’s also interesting that in most cases, ΔQ_s is greater than the volume of landslided material considering the former doesn’t include bedload transport, and I imagine a lot of the landslide material is coarser than would be expected from suspended sediment. This suggests to me that either landslides somehow “prime” landscapes to deliver excess sediment, or much of this excess suspended sediment you observe is not, in fact, material that was captured in the 2018 landslide dataset. Do you have a few words about what the character of this material might be, or the process that mobilizes it?

Figure 8: The caption says what is shown by each point, but not what each point is. Is each one of the 24 focus stations?

Figure 9: I'd recommend adding a horizontal line showing the pre-Morakot value for reference. Also, are these the same data as are shown in Figure 4? If so, any particular reason/justification for using $\log(\tilde{a})$ there (and later in section 4.2) and $\ln(\tilde{a})$ now?

377-380: The question and content asked and described here are presented earlier in the paragraph above, on lines 368-373. I'd suggest rearranging or combining.

Figure 10: I am thinking about how to compare the values shown here in b) to the take-away from lines 365-359, where the minimum decay timescale was ~4 years and the maximum was ~ 9 years. It seems that many here are >9 years.

383-391: This argument is reasonable, and I think it makes sense. It could probably be validated by looking at modern satellite images. However, I am not sure how to rectify it with the data shown in Figure 8, which shows that many/most of the stations have already delivered more excess sediment (purely via suspended load) than the total landslided volume. The volume of delivery would presumably be even larger if bedload were included. I think this should be acknowledged and/or discussed somewhere.

482: I believe that this is the first time that the characteristic decay time of 3-255 years has been presented, at least in words. Should be presented before the conclusions section. I assume that this is the total range among all stations, and the 4-9 years presented before is among a subset of stations (focus, or high-landslide-intensity and also decaying, etc.).

483-484: Do you have any intuition or understanding of why rates of $\tilde{a}$ should respond faster to Typhoon Morakot than rates of b? Even wild speculation of some physical mechanism or process could be enlightening and provide opportunities for future authors to test against field data or models.

489-491: The idea that the typhoon's influence should disappear entirely within a few decades works well with characteristic decay times of 4-9 years, but perhaps not with decay times up to 255 years.