Authors' Response to Reviews of

Controls on fluvial grain sizes in post-glacial landscapes

immediate

,

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

1. Cover Letter

Dear Editor,
We appreciate the opportunity to revise our manuscript. In response to the invaluable feedback from all reviewers, we have made changes to enhance the overall quality and clarity of our paper. We respond to the general comments and then more specifically to the line comments.
Best wishes,
Anya, Mikael, Simon and Fiona

2. Reviewer #1

This paper presents a broad analysis of grain size in streams in postglacial settings which builds on previous studies in non-glaciated landscapes that have been conducted using random forest and Spearman correlations (Snelder, 2011 and others). These methods are employed by the authors in a novel landscape setting and with an alternative approach use to data capture - a citizen science survey. The paper frames a clear hypothesis and is well structured and clearly written. I consider the paper will provide a valuable contribution to the field following some revision. I have three main points which I think should be addressed to enhance the work for publication.

As outlined in the main points below, I feel the paper would benefit from further discussion of the limitations of the data captured by the citizen science survey approach, including the distribution of sites and the representativeness of the images of sediment characteristics at the reach scale. Additionally, (and if supported by the data), I think more could be made of the sites where multiple data points are available for a single stream, particularly to illustrate the model for sediment delivery established for post-glacial settings.

A more substantial point relates to the erodibility metric used, and I believe further consideration of the formulation of this metric may be required to ensure it adequately reflects the geological variability in Scotland. Although I suggest the metric needs revision, I don't believe changes will alter the main conclusion of the paper. I look forward to seeing the final work in print.

Best wishes,

Katie Whitbread

2.1. Main points

RC: ***1. Data distribution and representativeness:***

RC: ***The sites are quite clustered and some discussion of potential impact of clustering bias on the outcomes is needed. To what extent does the distribution of sites cover a reasonable range of values of your key parameters (erodibility, steepness etc.)? Are very big, or very small channels undersampled? Could you include some plots in the supplementary information?***

RC: ***Note that the distribution of data in Snelder et al. (2011) is a well-spread sampling pattern at the national scale which is pretty much ideal, so it's important to consider the effect of clustering in relation to the lack of correlation in your study.***

RC: ***In terms of representativeness of the sample, grain size may vary locally within reaches, e.g. between riffles and pools, inner channels and channel bars, even between adjacent bars. The Snelder (2011) study describes a process of visual assessment of the geomorphic components of a reach and areal proportions of different grain-size deposits, followed by selection of representative points for sampling. Are there ways you can assess the representativeness of the sample (in the context of the reach morphology)?***

RC: ***The paper would benefit from further consideration of the distribution and representativeness of the data in the discussion as part of a more extended treatment of the limitations of the crowd-sourcing approach.***

AR: Thank you for these comments. Yes, overall our sample sites are more clustered than in the study of Snelder et al. (2011). This clustering likely reflects the preferences and accessibility of citizen scientists, in contrast to the more structured sampling design used by the French National Agency. We have added statistical distributions of the environmental variables to the Supplementary Data. The sampled sites span a reasonable range of environmental conditions. Some distributions are less uniform; for example, sample elevation shows a positively skewed distribution with fewer high-elevation samples, which reflects the topography of Scotland and potential preferences of citizen scientists. We emphasise that correlations remain weak between grain size and environmental metrics within the sampled ranges.

AR: We have added the following text to the final paragraph of the Discussion:It is important to note that our sample sites are more spatially clustered than in the study of Snelder et al. (2011). This clustering likely reflects the preferences and accessibility of citizen scientists, in contrast to the more structured sampling design used by the French National Agency for Water and Aquatic Environments. Compared to more structured sampling approaches, clustering has both benefits, such as enabling the exploration of detailed trends in selected river basins, and also limitations, including a more skewed range of sampled environmental variables. Moreover, we highlight that many of the sampled sites were taken by citizens, who may not have selected gravel bars that were representative of the river reach, and as such represents a limitation to our methodology. However, as discussed in the Methods section, we examined the context photograph associated with each survey upload to only select gravel bars that were considered to be representative of the river reach.

RC: ***2. Analysis of single-stream sample groups***

RC: ***The treatment of sites where multiple samples are available for the same stream is very limited. It would be interesting if more could be made of this data – for example a fuller treatment of the data for the River Feshie shown in figure 6 (b) to assess whether there are localised spatial trends and points where disruption occurs due to sediment inputs. Is there potential to include e.g. long profiles where data and key local features of influence could be plotted data could be projected on to a long profile? This could help to illustrate the spatial model in figure 7 and bolster the conclusion by demonstrating it operating through the local dynamics in the study area.***

AR: Thank you for these comments, these are really good ideas. We have added a more detailed analysis of the Feshie River, which is a basin where we have a high density of grain size measurements and a detailed understanding of geomorphic processes following our recent study Towers et al. (2025). Our results show that the Feshie River demonstrates large variations in grain size, with no apparent downstream fining trends, and flow competence. To show this, we have added a long profile plot (Figure 6c) of the Feshie with grain sizes and key geomorphic domains marked (e.g., paraglacial and glacial sediment stores, U-shaped valley). We suggest that the lack of trends may arise from the glacial modification of topography. This is supported by our recent paper Towers et al. (2025), which suggests that paraglacial sediment stores contribute large amounts of sediment to the modern Feshie river.

RC: ***3. The erodibility metric***

RC: ***Metamorphic grade is encompassed by the lithology described on BGS maps (e.g. a gneiss is high metamorphic grade by definition, and a metamorphosed sandstone is either a "metasandstone" or a "wacke"). I'm therefore not convinced of the relevance of combining separate metrics reflecting lithology and metamorphic grade when using the BGS bedrock map. The metamorphic grade seems redundant. Also, what is the estimate of rock strength (LL) based on? Has a strength dataset been used?***

RC: ***The classification shown in the erodibility map in supplementary Figure S2 is different to what I would expect, and I think this points to the need for revision of the metric.***

RC: ***• The moderate erodibility estimated for the Ordovician-Silurian age metasandstones ("wackes") across the Southern Uplands appears similar to the Devonian-aged Caithness Flagstone (an unmetamorphosed lacustrine siltstone/fine sandstone sequence) in the far northeast, and to the Coal Measures of Carboniferous age in the Midland Valley. I'd expect the Southern Uplands 'wackes' to have lower erodibility.***

RC: ***• The highest erodibility units (yellow) seem to be associated with small outcrops of Early to Middle Devonian conglomerates around the fringes of the Moray Firth and with units that occur along the Highland Boundary Fault zone and Southern Uplands Fault. It is not clear why these specific Devonian conglomerates should be more erodible than other Devonian units including conglomerates, sandstone and lacustrine siltstone and mudstone which are estimated with medium erodibility.***

RC: ***• I am unsure what would be giving rise to the high erodibility values along the Highland Boundary Fault – perhaps the Highland Border Complex, but these strata are metamorphosed.***

RC: ***• I would expect the highest erodibility in strata in the unmetamorphosed Permian sandstones which occur in basins within the Southern Uplands and small exposures of Triassic/Jurassic/Permian strata around the Moray Firth Coast - but currently these seem to have intermediate erodibility values.***

RC: ***Finally – this approach to characterising rock erodibility was developed to configure models of river incision into rock rather than as a control on grain-size distributions supplied to channels. It would be worth noting, in the discussion at least, that grain-sizes supplied to channels are significantly influenced by the nature of discontinuities (joints and faults) – see overview in Sklar (2024). Whilst in general terms, lithology/rock strength influences fracturing and is therefore somewhat included in your method, fracture density can be very locally variable (e.g. Neely and DiBiase, 2020; Whitbread et al., 2024) and the degree to which this may influence the supply of material to your sites isn't known. There has been a fair amount of recent work on grain-size distributions on hillslopes and I think it is important to note the emerging literature in this area and consider it in relation to your conclusions.***

AR: The erodibility index was used to provide a general, landscape-scale estimate of relative rock erodibility across Scotland. The index distinguishes between first-order contrasts in erodibility between major rock types — for example, indicating that granite is more resistant to erosion than coal. We agree that the index represents a simplified approach and has limitations, which we have now addressed in the Discussion. Despite these uncertainties, there are studies that use similar methods to investigate bedrock erodibility and grain size (Abeshu et al., 2021; Snelder et al., 2011), and we still think that it is valuable approach, in particular for a global analysis. We also tried an alternative, simpler approach, where we explored the influence of the proportion of igneous, metamorphic and sedimentary rocks in the upstream catchment on grain size, under the assumption that more resistant rocks (igneous and metamorphic) fracture less for a given stress history. Using this method, we similarly found no significant correlations between upstream rock type and grain size.

AR: We have added text explaining the link between erodibility and grain size as we realise this was not properly explained in the manuscript.

AR: In the Methods: 'Broadly speaking, more resistant rocks produce clasts that fracture less and have lower abrasion rates (Attal and Lavé, 2009; Sklar, 2024). This means that coarser grain sizes would be expected in basins with more resistant lithologies, assuming all else equal.

AR: We have also added text outlining the source data used to estimate rock strength, L_L, which is a parameter in the erodibility calculation.

AR: In the Methods: 'Our constraints on rock strength, L_L, come from the Campforts et al. (2020) study (see Supp. Table 2 in Campforts et al. (2020)). We have added the spreadsheet associated with the erodibility index to the Supp. data (Table S3). '

AR: Thank you for the final suggestion regarding the incorporation of the bedrock erodibility index to the Discussion. We have added the following text:

AR: 'Moreover, many of the environmental variables are defined at the landscape scale and are therefore relatively coarse; for example, the bedrock erodibility index may not capture local variations in erodibility and thus grain size (e.g., fracture spacing can vary substantially due to fault-related deformation at the local scale (Neely and DiBiase, 2020; Whitbread et al., 2024)).'

2.2. Specific comments:

RC: ***Figure 1: Increasing the size of the maps and photos would be helpful – these are very small in the preprint. Add labels showing key locations mentioned in the text would also be useful (see comment below about image 3). If addressing the point above about the use of multiple points on key streams, could you highlight grouped data for key streams on this plot? Or perhaps show those more clearly on Figure 3). Use of a hillshade terrain model as the base map may also help illustrate the range of topographic settings associated with the data distribution.***

AR: We have re-arranged and increased the size of the maps and photos. We have also added labels to show the key catchments referred to in this study.

RC: ***Figure 2: the segmentation of the photograph in c) is very difficult to see – increasing the photo size and upping the contrast of the lines would be helpful.***

AR: We have increased the size of Figure 2c. Changing the thickness of the red lines would require modifying the PC code, with which we are not familiar to that extent.

RC: ***Figure 3: Labelling is included on the map, but not very clearly – increasing the image size and using lines to show the association of the label and relevant points more clearly would be helpful.***

AR: We have increased the size of the map and the river labels. We have also added lines connecting each river to its corresponding location.

RC: ***Figure 7: Could you include an illustration of this model from one or more of your catchments where you have multiple points on the same stream? See notes on the text below. (But if you leave the figure as is, can you increase the size of the photos as they are too small to see clearly.)***

AR: To better illustrate our model, we have added a long profile plot of the Feshie river showing grain size and flow competence (see Figure 6c). The Feshie demonstrates no relationship between grain size and flow competence, and in particular has reaches with exceptionally coarse grain sizes with a low flow competence in relation to the rest of the catchment. This scenario corresponds to Scenario (e) in our post-glacial model shown in Figure 7.

AR: We have increased the size of the two photographs and also fixed a river name typo in (e)

RC: ***Supplementary figures S1 and S2: Increase the size of the maps***

AR: Increased

RC: ***Line 186 / Table 1: What scale and version was the bedrock geological map used? E.g. was it 1:50,000 Bedrock v8.***

AR: The scale and resolution has been added to the table: 1:50,000 scale, version 8.

RC: ***Line 198 / Table 1: What scale and version was the superficial geological map used?***

AR: The scale and resolution has been added to the table: 1:50,000 scale, version 8.

RC: ***Line 242 / Equation 6: The inclusion of Q and W in equation 6 means that it isn't just rearranged from equation 5. Could you explain the inclusion or cite a relevant paper?***

AR: Thank you for highlighting this. We just missed one step, that is, an explanation that the unit discharge is the discharge divided by the channel width. Importantly, we made a mistake by putting $0.15g^{0.5}$ in the numerator instead of the denominator, which probably led to more confusion. Equation (6) has now been corrected, with a statement that 'Q/W is the ratio of discharge to channel width (or unit discharge)'. Equation 6 is now:

$$D = [\frac{QS^M}{W * 0.15g^{0.5}}]^{2/3} \quad (1)$$

RC: ***Line 264: There is very cursory treatment of the data for rivers with multiple samples. How many rivers had multiple samples? How many samples and over what range of drainage areas? Could you show the variation of D84 and potentially other variables on a long profile of the stream? This may be particularly useful if it helps illustrate your model in Figure 7.***

AR: We have the highest density of samples along the Feshie River and, as such, have added a long profile plot of the Feshie to Figure 6c, including the key geomorphic landforms that influence sediment supply and transport. We have also added the river names that we are referring to in Line 264 (Feshie, Dee, and Tay).

RC: ***Line 291 / Figure 6 (b): See main point 2 above – is there more in this data for the Feshie? It seems like there are perhaps two clusters of data, one with a trend and one without. Are there parts of the catchment with localised trends in flow competence, vs. parts disrupted by sediment inputs? Could you use local relationships in the Feshie or other streams to illustrate your model in Figure 7?***

AR: Thanks for this comment. We have coloured the markers on Figure 6b according to the tributary to better aid the analysis. For each river in the Feshie basin, we do not observe any apparent relationship with flow competence. Negating some points, it could be said that the Chaoil tributary displays a positive relationship between grain size and flow competence. However, field observations, satellite images and BGS superficial maps show that the Chaoil tributary is incising directly into glacial till implying that it is not an example of a region not disrupted by glacially-conditioned sediment inputs.

RC: ***Lines 345-352: The discussion of the citizen science approach should also address limitations of the approach and consider how these could be mitigated in future studies. As noted above, the clustered distribution of sampling sites is a key limitation – I think you should include some discussion of the potential impact of this on the outcome of your study.***

AR: Thank you for this comment. We have added the following text to the final Discussion paragraph.

AR: 'It is important to note that our sample sites are more spatially clustered than in the study of Snelder et al. (2011). This clustering likely reflects the preferences and accessibility of citizen scientists, in contrast to the more structured sampling design used by the French National Agency for Water and Aquatic Environments. Compared to more structured sampling approaches, clustering has both benefits, such as enabling the exploration of detailed trends in selected river basins, and also limitations, including a more skewed range of sampled environmental variables. Moreover, we highlight that many of the sampled sites were taken by citizens, who may not have selected gravel bars that were representative of the river reach, and as such represents a limitation to our methodology. However, as discussed in the Methods section, we examined the context photograph associated with each survey upload to only select gravel bars that were considered to be representative of the river reach. '

References:

Neely, A.B. and DiBiase, R.A., 2020. Drainage area, bedrock fracture spacing, and weathering controls on landscape-scale patterns in surface sediment grain size. Journal of Geophysical Research: Earth Surface, 125(10), p.e2020JF005560.

Sklar, L.S., 2024. Grain Size in Landscapes. Annual Review of Earth and Planetary Sciences, 52.

Whitbread, K., Thomas, C. and Finlayson, A., 2024. The influence of bedrock faulting and fracturing on sediment availability and Quaternary slope systems, Talla, Southern Uplands, Scotland, UK. Proceedings of the Geologists' Association, 135(1), pp.61-77.

Citation: https://doi.org/10.5194/egusphere-2024-3084-RC1

3. Reviewer #2

Review of "Controls on fluvial grain sizes in post-glacial landscapes", submitted by Towers et al et Earth Surface Dynamics.

This manuscripts presents a very interesting and well-designed study on the relationships between grain sizes in rivers and environmental parameters, in the post-glacial landscapes of Scotland. The paper is well written and supported by nice data, however, I think it still requires some work before publication. In fact, I identify several points that could easily be a bit more developed to better root the study and the results. Please find my main comments below with some more minor points. I don't see any issue in addressing my comments and none of them should modify the conclusions of this study so I look forward to reading the published version.

3.1. Specific comments

Introduction (and where relevant)

RC: ***I would have expect reference to the recent review of Sklar "Grain size in landscapes"***

Sklar, 2024, Annual Reviews https://doi.org/10.1146/annurev-earth-052623-075856

AR: Good point. We have added this reference to line 28.

RC: ***l. 30 "primarily by abrasion": deposition and sorting are also first-order parameters in size reduction during transport. This should be addressed in the Introduction. As grain size fining will modify the size of the sediments, I was also expecting something about the rate of fining and how the position along the river could affect the size. I think this should be mentioned here or in the Methods section as this is, I think, an important question: how the location of your samples with respect to the river network can affect your results?***

AR: Good point, we agree with this comment. We have added the following text: Once a sediment grain enters a river network, it reduces in size primarily through size-selective transport and abrasion (Sternberg, 1875). The distributions of fluvial grain sizes have therefore been correlated to the longitudinal flow distance along a channel (e.g., Gomez et al., 2001; Moussavi-Harami et al., 2004; Rice and Church, 1998; Sklar et al., 2006).

RC: ***l. 36-48 in this paragraph, there is a mix of grain production (l. 39) and transport (l. 41), and in grain size and flux (l. 47). This should be clarify by adding a few sentences and/or separating sentences to avoid mixing of concepts.***

AR: We have removed the sentence discussing sediment flux as this was confusing and introduced new concepts outwith the main scope of this paragraph. We have added a sentence citing a study which explored the influence of rock type on fluvial grain size. The new sentence reads as follows. 'For example, Lai et al. (2021) showed that more resistant rock types, such as volcanic rocks, were associated with coarser bedload grain sizes in comparison to less resistance rocks, such as sedimentary mudstone and flysch lithologies, in the Coastal Range, Taiwan.'

RC: ***I really appreciate the final paragraphs, they state in a very clear way the motivation and purposes of the study.***

AR: Thanks!

Methods

RC: ***Data collection is a mix of automatically and manually segmented pictures with the addition of a few manual measurements. I totally understand the method, however, this part needs some clarification. In fact, it is well known (among other, Kellerhals and Bray, 1971 or Bunte and Abt, 2001) that grain-size distributions are not equivalent when the sediments are sampled by grid (Wolman counts) or by area. How did you account for this effect in your data set?***

AR: This was a major issue that we completely overlooked - we are extremely grateful for this comment. Because PebbleCountsAuto had been tested in multiple locations and studies, we wrongly assumed that it would give an estimate of the true distribution, that is, the distribution that one obtains using grid-by-number or volume-by-weight methods. But it doesn't, so we had to devise a correction procedure, similar to that devised and tested by Kellerhals and Bray in 1971. The results were not perfect, and we found that the 'true' grain size (manually derived from grid-by-number) was systematically bracketed by the uncorrected and corrected PebbleCounts values. We therefore decided to run our whole analysis using the average of these two values (which ends up closest to the measured value) to complement the analysis on the uncorrected data which now features in Supplementary Materials. The "Comparison between grain sizing tools" section becomes a "Comparability of and comparison between grain sizing tools" section, with an updated Figure 2 and the following content describing the issue and correction:

AR: "A first consideration is that grain size distributions obtained via PebbleCountsAuto are not directly comparable to those obtained via applying a grid on a photo or measuring grains at set intervals along a transect in the field. The latter methods belong to the grid-by-number set of procedures, which have been shown to provide the most accurate representation of the true grain size distribution, with the results theoretically directly comparable to those obtained using volume-by-weight methods, providing the sediment is isotropic (Kellerhals and Bray, 1971). PebbleCountsAuto is an area-by-number procedure, whereby the algorithm attempts to measure all grains visible on the surface and then produces a distribution based on the number of grains; this type of procedure has been shown to be systematically biased towards the finest sediment and to therefore underestimate percentiles (Kellerhals and Bray, 1971). Kellerhals and Bray (1971) proposed, tested and validated a conversion procedure to correct for this bias: a conversion factor D^2 needs to be applied to the area-by-number data, where D is the diameter of the sediment grain considered, approximated in our case by the grain's intermediate axis.

AR: To correct the PebbleCountsAuto dataset, we proceeded as follows: for each dataset produced by PebbleCountsAuto, a frequency of 1 was given to each measured grain, which is what the software would have done when computing the grain size distribution. This frequency was then multiplied by D^2 for each grain, giving a new value. The fraction that each grain represents in the grain size distribution was then computed as the ratio of this new weighed frequency divided by the sum of all new weighed frequencies for the entire measured population. These corrected data were used to produce cumulative grain size distributions and extract relevant percentiles in the following."

AR: The final paragraph in this section is now as follows:

AR: "The application of the correction factor described earlier in this section is a limited success: as expected, it leads to an increase in the PebbleCountsAuto grain sizes but also to a systematic overestimation of the grain sizes, and by various amounts (Figure 2). We believe that this is likely due to the fact that PebbleCountsAuto is not a perfect area-by-number method, in that it does not measure all the grains on the surface but only a subset (Figure 2c). Differences in grain shapes can also influence the effectiveness of the correction which assumes all grains are similarly shaped and uses the D^2 multiplier to make the 1-D measurements three-dimensional (Kellerhals and Bray, 1971). In the absence of a clear relationship between the amount of overestimation and metrics such as grain size or the fraction of grains not measured, and because the actual grain size measured manually is systematically bracketed by the corrected and uncorrected PebbleCountsAuto values, we decided to retain the mean of the latter two values for the following analysis. We find that this mean value is on average within 7 % and 8 % of the manually measured value for d50 and d84, respectively, in comparison to the uncorrected grain sizes being within 27 % and 21 %, and the corrected grain sizes being within 42 % and 37 % of the measured value for d50 and d84, respectively. While this issue represents a significant limitation of our approach, we also believe this approach is the most pragmatic, as it is necessary and provides a corrected measurement that is the closest to the actual measured value. We ran the following analysis using both the corrected-uncorrected mean and the uncorrected PebbleCountsAuto grain sizes. The results presented in the main text are those obtained using the corrected-uncorrected mean grain sizes; those obtained using the uncorrected PebbleCountsAuto grain sizes are presented in the Supplementary Data. Given the lower errors associated with the d84 percentile, we focus our analysis on this percentile."

AR: Thank you again for this comment, we are extremely grateful!

RC: ***In addition, it seems that you used a grid for the manual counts on pictures with an imposed number of nodes, rather than a fix distance. How would this affect the distributions?***

AR: This should not affect the results if the distribution of the grains across the surface is uniform: the grid just gives the reader a pointer to select pebbles in an unbiased manner. The results would be affected if the distribution were not uniform (e.g., patches of coarser or finer sediment) or if reducing the size of the grid led to grains being counted multiple times. We cannot rule out these effects of course, but considering the variety of gravel bar sizes, grain size distributions and environments that were studied, we had no choice but to adapt the size of the grid so that we could quantify the grain size of the sediment in an unbiased manner over an area which we deemed representative of the gravel bar.

RC: ***PebbleCounts does not segment all the grains so that the distribution is not by area neither by grid. Did you check that the results from the automatic approach are consistent with the manual one?***

AR: We believe that this is one of the main reasons why the application of the correction factor didn't work well, despite being proven successful in the study by Kellerhals and Bray. PebbleCounts is neither area nor grid, and we find that the correct grain size lies between the two (uncorrected-by-area and corrected-by-number). This is one of the reasons we decided to use the average value between these two, as explained above.

RC: ***Finally, due to the overlapping of the grains, samples from pictures tend to underestimate the diameters. If imbrication or grain shape are very different from one site to another, one could expect some differences related to the method and not the grains themselves. This is very difficult to quantify but it should at least be acknowledged.***

AR: Thank you for this comment. We have added a paragraph at the end of Section 2.2.1 highlighting this.

AR: The paragraph reads as follows: "In addition, pebble imbrication or orientation on a gravel bar (e.g., intermediate axis not perfectly visible in the horizontal plane) may also lead to an underestimation of grain sizes in photo counting compared to distributions obtained from tape measure lines (Attal and Lavé, 2006). The degree of underestimation in grain sizes associated with photo counting is likely to vary between localities and represents a limitation of our method."

RC: ***The selection, definitions and calculations of the environmental variables should be further developed and illustrated. In the current form, it is a bit difficult to understand the choices, the calculations. The way erodibility is defined could be a bit more developed too, as it is not that common. I would appreciate a bit***

more context and values (what do the landscapes look like, do we have some diversity or are they very similar, etc).

AR: We have expanded the definition and reasoning behind our use of the erodibility index. We have discussed the general science behind the environmental variables presented in this section in the Introduction. We have added a sentence referring the reader back to the Introduction. We also refer the reviewer to the end of the Introduction for descriptions of post-glacial Scottish landscapes (U-shaped valleys, glacial and paraglacial deposits).

RC: ***Random forest regressor and Spearman's correlation might not be familiar to all readers, please explain them a bit more the Methods.***

AR: Thanks for this comment. We have given an extensive description of a random forest regression in Section 2.4., which is located at the end of the Methods. We have added two sentences describing Spearmnan's correlation test: 'We also perform a Spearman's rank correlation test between grain size and each environmental variable. Spearman's rank correlation measures the strength and direction of the relationship between two variables.'

RC: ***l. 210 "all grain sizes are available for transport": this is a very strong hypothesis. How true and universal can it be? If not true, how could it alter your interpretations?***

AR: We have addressed the 'global' application of the flow competence metric in the Results section. We refer the reviewer to the last paragraph of the Results.

AR: "...We acknowledge that actual trends may be obscured in the noise due to the dataset amalgamating data from a very wide range of geological and geomorphological settings across Scotland. We therefore isolate the Feshie River basin which is very dynamic (Towers et al., 2025), with evidence of frequent bedload transport (Matthews et al., 2024) and for which we have 18 data points. "

RC: ***l. 214 supply-limited is often defined as a lack of sediments with no reference to a specific size. This could be mentioned as your definition might not be the most commun one.***

AR: Good point. We have made this more clear by adding the following sentence to the paragraph: "We emphasise that, while the terms supply- and transport-limited are sometimes used in the literature with reference to sediment availability, our reference is to grain size (Attal et al., 2015)."

RC: ***Equation 6 is not just a simple rearrangement of Equation 5. Please add the intermediate steps and write explicitly the final equation mentioned l. 244.***

AR: Thank you for highlighting this - the other reviewer did too. We just missed one step, that is, an explanation that the unit discharge is the discharge divided by the channel width. Importantly, we made a mistake by putting $0.15g^{0.5}$ in the numerator instead of the denominator, which probably led to more confusion. Equation (6) has now been corrected, with a statement that 'Q/W is the ratio of discharge to channel width (or unit discharge)'. Equation 6 is now:

$$D = [\frac{QS^M}{W * 0.15g^{0.5}}]^{2/3} \quad (2)$$

RC: ***Reading the abstract, I was really interested in the citizen science approach, yet, it is barely addressed in the Methods. How did you set up the survey? How many participants? What is the quality of the data? I think this is innovative and it should be better explained so that other groups can use the same approach.***

AR: The citizen science approach is described in the Section called Grain size data collection. We used ESRI's

Survey123 platform to collect the data. We did not collect data on user identity (due to GDPR) so we do not have data on participant numbers. The main issue that led to binning of data was poor documentation of location (e.g., location recorded by phone was not on a river). We did not record the amount of binned data which in hindsight would have been useful. We have also added more information to the final paragraph of the Discussion discussing opportunities and limitations with the approach (see later below comments).

Results

RC: ***Figure 3 is a bit difficult to read because it is quite small. Please consider increasing the size. For comparison, it could be great to add a scale in mm on panel b (or to add a panel c with the distribution of D84 in mm).***

AR: We have increased the size of the map and the river labels. We have also added lines connecting the river with corresponding location. A second scale has been added to the x axis showing grain size in mm.

RC: ***Figure 4 is barely used. In addition to the correlations with D84, you could explain the other correlations (for example, slope and aridity) and built on it either in the Results or in the Discussion.***

AR: This is an interesting point however we don't consider the relationships between the other variables to be within the scope of this study. The main purpose of Figure 4 is to show the relationship between grain size and each environmental variable.

RC: ***l. 289 it is not clear to me how you can conclude that the river has the potential but not the sediments. I think this is an important aspect of this study that could be further developed (see also my comment on line 210).***

AR: As discussed, if sediment grain size is transport-limited, i.e., controlled by flow competence, a power relationship can be expected between the grain size D and the quantity ω_m. However, we do not see any clear relationship on our flow comp plot. In Line 289 (of the original manuscript), we state that along reaches with a high flow competence, most grains are smaller than the median grain size. We suggest that along these reaches the river has the ability to transport larger sediment than present, which we term supply-limited.

Discussion

RC: ***l. 297 substrate cover is not mentioned previously and not explored in the Results. Please consider removing this point or addressing it in the whole manuscript.***

AR: We have removed the point concerning substrate cover, thanks for pointing that out.

RC: ***Here again, I was a bit disappointed by the way the citizen science survey is discussed. I guess there is more to say about it as it is not a classic method. For example, how long did it take to collect the data ? Did you stop or not the survey, when and why? How time consuming it is to deal with the variety of data with respect to "classic" field work?***

AR: We ran the survey for approx. 1 year. We have added this information to the methods. If we had had more time then we would have ran the survey for longer. The most time-consuming part of the workflow was measuring pebble sizes from photographs, especially when pebbles had to be manually measured (and probably similar to the time it takes to measure grain size in the field through traditional field measurements).

AR: We have also added more detail to the discussion about the limitations of the survey. The following text has been added.

AR: It is important to note that our sample sites are more spatially clustered than in the study of Snelder et al.

(2011). This clustering likely reflects the preferences and accessibility of citizen scientists, in contrast to the more structured sampling design used by the French National Agency for Water and Aquatic Environments. Compared to more structured sampling approaches, clustering has both benefits, such as enabling the exploration of detailed trends in selected river basins, and also limitations, including a more skewed range of sampled environmental variables. Moreover, we highlight that many of the sampled sites were taken by citizens, who may not have selected gravel bars that were representative of the river reach, and as such represents a limitation to our methodology. However, as discussed in the Methods section, we examined the context photograph associated with each survey upload to only select gravel bars that were considered to be representative of the river reach. '

RC: ***I would have appreciate some discussion about grain size variability along a stream, or at the same position if you have such data. A more focused analysis from rivers with similar variables could also be interesting, ie, the whole data set shows no trend but maybe at smaller scales, there are some. Did you explore this?***

AR: This comment aligns with comments from Reviewer 1. We have added a more detailed analysis of the Feshie River, which is a basin where we have a high density of grain size measurements and a detailed understanding of geomorphic processes following a recent study (Towers et al., 2025). Our results show that the Feshie River demonstrates large variations in grain size, with no apparent downstream fining trends, and flow competence. To show this, we have added a long profile plot (Figure 6c) of the Feshie with grain sizes and key geomorphic domains marked (e.g., paraglacial and glacial sediment stores, U-shaped valley). We suggest that the lack of trends may arise from the glacial modification of topography. Moreover, our recent paper Towers et al. (2025) suggests that paraglacial sediment stores contribute a large amount of sediment to the modern Feshie river.

RC: ***There are a few typos (parenthesis with references, l, 239-240)***

References

Abeshu, G. W., Li, H.-Y., Zhu, Z., Tan, Z., and Leung, L. R. (2021). Median bed-material sediment particle size across rivers in the contiguous U.S. *Earth System Science Data Discussions*, pages 1–22.

Attal, M. and Lavé, J. (2006). Changes of bedload characteristics along the Marsyandi River (central Nepal): Implications for understanding hillslope sediment supply, sediment load evolution along fluvial networks, and denudation in active orogenic belts. *Special Paper of the Geological Society of America*, 398(January):143–171.

Attal, M. and Lavé, J. (2009). Pebble abrasion during fluvial transport: Experimental results and implications for the evolution of the sediment load along rivers. *Journal of Geophysical Research: Earth Surface*, 114(F4).

Attal, M., Mudd, S. M., Hurst, M. D., Weinman, B., Yoo, K., and Naylor, M. (2015). Impact of change in erosion rate and landscape steepness on hillslope and fluvial sediments grain size in the Feather River basin (Sierra Nevada, California). *Earth Surface Dynamics*, 3(1):201–222.

Campforts, B., Vanacker, V., Herman, F. d., Vanmaercke, M., Schwanghart, W., Tenorio, G. E., Willems, P., and Govers, G. (2020). Parameterization of river incision models requires accounting for environmental heterogeneity: Insights from the tropical Andes. *Earth Surface Dynamics*, 8(2):447–470.

Gomez, B., Rosser, B. J., Peacock, D. H., Hicks, D. M., and Palmer, J. A. (2001). Downstream fining in a rapidly aggrading gravel bed river. *Water Resources Research*, 37(6):1813–1823.

Kellerhals, R. and Bray, D. I. (1971). Sampling Procedures for Coarse Fluvial Sediments. *Journal of the Hydraulics Division*, 97(8):1165–1180.

Lai, L. S.-H., Roering, J. J., Finnegan, N. J., Dorsey, R. J., and Yen, J.-Y. (2021). Coarse sediment supply sets the slope of bedrock channels in rapidly uplifting terrain: Field and topographic evidence from eastern taiwan. *Earth Surface Processes and Landforms*, 46(13):2671–2689.

Matthews, B., Naylor, M., Sinclair, H., Black, A., Williams, R., Cuthill, C., Gervais, M., Dietze, M., and Smith, A. (2024). Sounding out the river: Seismic and hydroacoustic monitoring of bedload transport. *Earth Surface Processes and Landforms*.

Moussavi-Harami, R., Mahboubi, A., and Khanehbad, M. (2004). Analysis of controls on downstream fining along three gravel-bed rivers in the Band-e-Golestan drainage basin NE Iran. *Geomorphology*, 61(1):143–153.

Neely, A. B. and DiBiase, R. A. (2020). Drainage area, bedrock fracture spacing, and weathering controls on landscape-scale patterns in surface sediment grain size. *Earth and Space Science Open Archive*, pages 1–22.

Rice, S. and Church, M. (1998). Grain size along two gravel-bed rivers: Statistical variation, spatial pattern and sedimentary links. *Earth Surface Processes and Landforms*, 23(4):345–363.

Sklar, L. S. (2024). Grain Size in Landscapes. *Downloaded from www.annualreviews.org. Guest (guest*.

Sklar, L. S., Dietrich, W. E., Foufoula-Georgiou, E., Lashermes, B., and Bellugi, D. (2006). Do gravel bed river size distributions record channel network structure? *Water Resources Research*, 42(SUPPL.):1–22.

Snelder, T. H., Lamouroux, N., and Pella, H. (2011). Empirical modelling of large scale patterns in river bed surface grain size. *Geomorphology*, 127(3-4):189–197.

Sternberg (1875). Untersuchungen uber langen-und Querprofil geschiebefuhrender flusse. *Zeitshrift fur Bauwesen*, 25:483–506.

Towers, A. H., Mudd, S. M., Attal, M., Clubb, F. J., Binnie, S. A., Dunai, T. J., and Haghipour, N. (2025). Denudation rates and holocene sediment storage dynamics inferred from in situ 14c concentrations in the feshie basin, scotland. *Earth Surface Processes and Landforms*, 50(4):e70043.

Whitbread, K., Thomas, C., and Finlayson, A. (2024). The influence of bedrock faulting and fracturing on sediment availability and quaternary slope systems, talla, southern uplands, scotland, uk. *Proceedings of the Geologists' Association*, 135(1):61–77.