We thank the AE for his constructive comments. Our replies are given in *italics*.

AE comments
The Authors have made a good set of revisions to the reviewers suggestions. I have one point I feel should be addressed if possible under minor revisions. To me there seemed some ambiguity in the response to point 2 of reviewer 2 "R2: With the new two theories, the authors can apply two theories to the results, and try to explain the scatter more specifically and carefully, rather than just saying there is a better trend or the theory probably works better with no proof."

Authors Response: We actually think that we do what the reviewer suggests here: we discuss the variation in the data using available models from the literature and physics-based reasoning. We are unsure which two new theories the reviewer is referring to here.
*We interpret the comment of the reviewer to pertain to the treatment of the data, and therefore to a different aspect of the paper then the statement in the conclusion quoted by the AE. Let us explain. In sections 4.1 and 4.2, we discuss the data obtained in our experiments, in particular how geotechnical properties control the erodibility as measured in the mill experiments. There, we compare the data to theory. In contrast to the statement of the reviewer that we use "two new theories", we discuss the data in context of the state-of-the-art model from the literature (originally put forward by Sklar and Dietrich, 2004), and then extend this theory further using physics-based reasoning (section 4.2.2). In a final step, we argue that even the extended theory is insufficient to explain the variation in the data. In response, we suggest that the dominant missing control lies in the grain size and provide a qualitative argument on how this control should work and scale.*
*As stated in the rebuttal, it is unclear to us which "two new theories" the reviewer referred to here – we discuss an "old" theory, and extensions of it.*
*We do not think that reviewer refers to the upscaling arguments here (section 4.3). First, these are also not "new theories". Second, they cannot be applied to "*the results […] to explain the scatter more specifically and carefully*", as suggested by the reviewer. See below for further details.*
*We still think we do what the reviewer suggests – i.e., "*apply two theories to the results, and try to explain the scatter*". We (i) empirically compare mill erosion and geotechnical data (section 3.1), we discuss some general observations (section 4.1), concluding that a purely empirical approach does not yield any advances, and that the data need to be looked at in the context of theory (line 293-295). We do the latter by comparing the data to the existing state-of-the-art theory (section 4.2.1), and then extend it by using physics-based reasoning (section 4.2.2). Finally, we suggest that the theory is incomplete and suggest directions for further investigation (section 4.2.3).*
*As a final remark, we disagree strongly with the remark that we just claim that "*there is a better trend or the theory probably works better with no proof*". As explained above we evaluate the data within the context of theory to which we provide novel insights. We believe our analysis and discussion yields novel and significant advances.*

However in the conclusions: "We provide two theoretical frameworks where the relative erodibilities measured in our mills scale linearly to field situations, based on (i) a revised stream power incision model, and 525 (ii) on a sediment-flux-dependent incision model including the tools and cover effects. As such, the relative erodibilities measured in the laboratory can be applied to scale erosion rates over long time scales."

*This statement from the conclusion pertains to section 4.3 ("Application of the laboratory experiments to natural rivers"). There, we develop arguments for the scaling of erodibility in the theoretical frameworks of the stream power incision model and sediment-flux-dependent incision. We argue that in both frameworks, erodibility scales linearly for erosion rates. However, the two model frameworks lead to different predictions for the dependence of the channel long profile on lithology. The idea of the chapter is to provide a broader context of the applicability of our results. Please note a couple points on this:*

- *Neither of the arguments can be considered as a "new theory" – although we provide some new points of view (especially for the stream power model), we evaluate existing theories for this particular application. In our point of view, our points here have not been explicitly made in the literature before, but are implicit in published theory (e.g., Turowski, 2021, as cited in the paper).*
- *The upscaling arguments are completely independent both of the relationship between erodibility and geotechnical properties, and of the particular structure of the data.*
- *As such, it is not necessary to fully understand how geotechnical properties determine erodibility to apply them. The relative mill-measured relative erodibilities can be applied to erosion rates in natural rivers.*

*The latter is the main point that we want to make in the statement of the conclusion, quoted above, is that the measured erodibility values can be directly applied to natural rivers.*

I was not entirely convinced by the rebuttal offered by the authors here - there seemed to be a bit of contradiction in the response and the statement in the conclusion and I would invite the authors to have another look at their response to this issue - and perhaps either re-frame their response or provide some text in the paper that explains or makes this clearer? If (as stated in the response) you think you are actually doing what the reviewer suggested - perhaps that section could be re-examined and made clearer?

*We have implemented some revisions to clarify these issues:*
*We changed the final sentence of the introduction to (line 65)*
*"In addition, we discuss the broader application of the observed erodibilities, in particular their upscaling of the results to natural rivers within the two currently competing theoretical frameworks of the stream power model and sediment-flux-dependent bedrock erosion."*
*We added explanatory sentences to the opening paragraph of section 4.3:*
*"In this section, we put the observations on erodibility from our laboratory-scale experiments into a broader perspective." (line 418) and "We note that the arguments presented in this section are independent of the precise geotechnical controls on erodibility and the structure of the mill data." (line 425).*
*In the conclusion (line 521), we changed the statement quoted by the AE to*
*"We suggest that the relative erodibilities measured in our mills scale linearly to field situations, based on theoretical reasoning using both (i) a revised stream power incision model, and (ii) on a sediment-flux-dependent incision model including the tools and cover effects."*