Response to comments on **gc-2021-22-R1 "** *GC Insights: Identifying conditions that sculpted bedforms – Human insights to build an effective AI*"

Comments were again kindly provided by two reviewers (RC1, RC2), and we are glad that they both felt the changes in Revision #1 greatly improved the manuscript. Please find below our point-by-point response to this second round of comments. Comments are in grey, and responses in black. A manuscript with changes tracked is also provided. We have also made some minor changes to remain within the word limit.

We hope that the editor will find the manuscript suitable for publication after these minor corrections.

All the best

John
(on behalf of the authors)

RC1 (Anonymous)

I would like to thank the authors for the detailed point-by-point response and revision, which was useful to better understand the work. The clarity of the article has substantially improved, and the addition of the supplementary material is valuable.
> Thank you.

I have two remaining comments, one technical and one more conceptual:

- In Fig 1(e), what is the meaning of the diagonal line linking the datasets? My understanding is that these are from different experiments, and are only ordered indicatively. Is there some trend that can be meaningfully arrived at here, given the qualitative differences between the experiments? If not, the line may be inappropriate. (There is some indication of 'non-linearly increasing velocity', but it perhaps if the authors want to draw out a relationship, that should be explored more.)
> The reviewer correctly identifies that the data are from four different experiments. These are ordered in terms of shear velocity, which increases monotonically yet non-linearly from Expt. 1 to Expt. 4. As such the ordering is useful, yet we accept that adding a trend line is not strictly statistically justified. It was added for visual and illustrative purposes only, and we have therefore removed it as upon reflection the improvement in the methods is visible without it.

- Underlying the discussion is, I think, the idea that ML techniques might work well where non-experts have been demonstrated to make good decisions based on particular visual data. This is a really interesting idea, but I wonder if it could a) be made more explicit, and b) if this comes from somewhere in the literature or is the supposition of the authors. Either way, drawing out this as a testable proposition, even briefly, would strengthen the paper, I think, as it has implications beyond just this domain.
> We thank the reviewer for focussing us on this idea, and requesting that we clarify our presentation of it. It is a speculation of the authors, based on our observation of the results of the human participants, and we have endeavoured to draw this out briefly as a testable proposition. This interesting idea is now in the final paragraph of the paper, giving it potential to be recognised outside the immediate domain of this study.

In general, the idea of establishing the sufficiency of a particular set of data inputs (e.g. shear velocity), to be able to determine an outcome, seems to be an interesting and valuable contribution.
> Thank you.

Reviewer #2 (Ward)

The revised version of this paper is better-organized and focused, with a more consistent terminology, and has addressed most of the review comments from the previous round. I think it can be published following minor revisions.
> Thank you.

It still appears as if there are some small gaps in the information provided from paragraph to paragraph. This is much improved from the first version, and I appreciate the difficulty imposed by the extra-short article format. I hope the suggestions below can be accommodated without too much added length.
> Please see detailed responses below.

Abstract - The abstract needs rewriting. It currently begins with a result, followed by a lengthy and somewhat confusing sentence that jumps topics from survey participants to machine learning, with a final inference about pre-processing. It is not clear how this inference follows from the beginning of the sentence. I suggest starting with something like the final sentence, modified: "Preliminary investigations with an ANN illustrate that a geoscience comm activity…", and split the middle sentence to clarify each clause.
> We have re-written the abstract, starting by placing the sentence that the reviewer suggested at the beginning.

39-40 Should the question of whether expert knowledge is useful be added here? As Q1a perhaps? It leads to one of the more interesting outcomes, which is reported in the abstract and section 3, but is never introduced as part of the study in the current version.
> The inclusion of both experts and non-experts is now made explicit at the start of Section 2 (Methods). Unfortunately, like the question of single bedforms vs sequences of bedforms, there is no space available to explicitly introduce this explicitly as a separate question.

56-57 re: time and distance profiles. While the presentation of this disparity is much improved, this will still be a bit confusing to the uninitiated.

Suggestion: could you simply elucidate, e.g. "…referred to as distance-height profiles, although the fluvial experimental data were collected as time series of bedform height passing below a sensor with time as a proxy for distance"? Pressed for space, you could find a few words at the beginning of the sentence and just say "For simplicity, these varied data…"
> As suggested, the start of the sentence has been shortened to save words, and we state that time is used as a proxy for distance in the fluvial data. We also now point to reader to the caption of Fig, 1 where there is a brief explanation of how the data are collected that corresponds well to the reviewer's suggestion (i.e. "*Distance-height profiles (strictly speaking time-series) like those given unannotated to participants, i.e. one from each experiment 1-3, all scaled to the same dimensions. Horizontal axis is time because in the flume tank a stationary sensor recorded height as bedforms passed beneath it*.")

Note from a physical standpoint I am not convinced the height-time and height distance profiles are directly equivalent, because bedform sizes vary in these profiles, and don't different sized bedforms migrate at different rates? However I don't think it will affect the outcomes presented here.
> From a physical standpoint, we agree. Different bedforms do indeed migrate at different rates, but this does not affect the conclusions presented here.

85-89 There seems to be a gap here, the paragraph starts talking about the baseline ANN experiment (which didn't work so well) and then does not signal that the remaining results are from the ANN with shape-fitting. Or is the difference the ANN with multiple bedforms? Suggest clarifying which results correspond to which ANN experiments. This will make the following discussion (101-107) more impactful.

> We have included a sentence explicitly clarifying that all experiments after the baseline one are based upon shape-fitting inputs to the ANN.

Again I appreciate the difficulty of the short format, the authors have done an admirable job of condensing a couple of potentially complex topics, so I hope it is helpful to point out where an outside reader will find a leap or two being made.
> Thank you.

Dylan Ward