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

Comments were kindly provided by two reviewers (RC1, RC2). Both reviewers highlighted the challenge of clearly communicating within this concise format, and we have made numerous modifications to address this.

Please find below our response to the comments. Comments are in grey, and responses in black. Although a fully revised manuscript is not yet prepared, we use 'changed' to indicate some simpler modifications where it was easiest to communicate by simply actioning the comment (see provisionally revised manuscript at the end of this pdf).

RC1

This paper explores an interesting idea. I particularly liked the idea of comparing and complementing AI-based and human decision-making, and the insights concerning the relevance of expertise are interesting.

As a reader with a background in machine learning and AI, rather than geoscience, I have a number of comments and questions concerning the current version of the manuscript:

- It is not clear to me what precisely the input to the ANN was, or what the task was (classification of time series of photos, I think, but this is left rather implicit). Was the input a sequence of images in each case? How many were included in the training / testing sets? How was the algorithm's performance evaluated? In general, there is not sufficient information in the paper to understand exactly what was done in the neural network part of the study. I would like to see complete detail, and/or reference to an implementation available for scrutiny/study.

> In order to provide sufficient information to reproduce the work, and yet comply with the length-limit of the *GC Insights* format, the input data and pseudo-code for the ANN are now provided as Supporting material. The algorithms performance was ultimately evaluated using out-of-sample prediction of experiment number (Fig. 1e) using 40% of the data withheld from training/validation. Performance during ANN training was using RMSE, with details of this now in the pseudo-code provided.

- Why the particular choice of ANN? Why those numbers of neurons and in that configuration? Were other options considered?

> Based on our prior experience, and the limited data available, we selected a small network (i.e. limited number of layers and nodes). After this, there was some *ad hoc* refinement, leading to the configuration used. We do not claim that this is optimal, and now explicitly state (e.g. in the Abstract) that this is a preliminary analysis. This is in line with the guidance for a *GC Insights* article that work must "*be well-founded and methodologically robust, based on evidence or analysis that can be openly inspected, but does not have to be comprehensively explored*".

- It seems that the results of the ANN study are very preliminary, and no conclusions can yet be drawn. (The manuscript uses phrases like "that training an ANN on these data alone should be productive".) Is that right? What are the conclusions drawn from the ANN work to date, or is it still at the speculative stage? The paper was unclear, to me, about this.

> The results of the ANN study are indeed preliminary; their purpose is to demonstrate that avenues suggested by the survey might be fruitful. The abstract and text have been re-phrased to be clearer that this is the intention. To use the example cited, we have clarified in paragraph 1 of the Discussion that the 'should' comes from the survey and in paragraph 2 stated that the ANN work provides initial numerical support for the idea. So, conclusions from the ANN are

1. It is possible to build an ANN with some predictive power (i.e. one successful example is sufficient to demonstrate the principle).
2. For this, sequences of bedforms produce better results than individual ones
3. And 'help' in the form of pre-computing morphological parameters mitigates the issue of data volumes that are available in this geoscience context.

- The manuscript claims that "Thus, insights from the participants have contributed to building an effective AI to reliably infer flow conditions from bedform morphology". I didn't see justification for this in the paper. The argumentation and explanation of the ANN results (and the lack of clarity around which are speculations and which are results) make it difficult to see where this comes from.

> This was a typographic error. We intended to say 'will contribute' rather than 'have contributed', because demonstrating the 'have contributed' needs a successful result that is following in a full paper after building an AI incorporating the insights gained here and fully applying it. This sentence has been removed in the rewrite of the discussion, but we will endeavour to avoid similar in the revised manuscript.

I think the submission could make for an interesting contribution, if the method, results, and contributions were more clearly stated.

> Throughout, we have re-written the manuscript with the aim of more clearly describing our work. In particular, pseudo-code and data are now provided in Supplementary Material to clarify the method. The ANN results are now in the Results section, instead of being presented as a follow-up, and we have re-phrased to clarify the narrative of the paper and thus its contribution.

RC2

The paper is a concisely written report of a survey of a mix of geoscientists and non- geoscientists to determine whether bedforms from different processes and environments could be classified based on shape alone. The results are interpreted in the context of building an artificial neural network model for the same task. The survey results indicate that non-geoscientists perform as well as specialists at this task, as it is primarily a shape matching exercise that does not require specialist knowledge. Unsurprisingly, identifying environment from an individual bedform was more difficult for respondents than when a series of bedforms were provided. The paper then applies an ANN model to the task with qualitatively similar results. The involvement of the non-geoscientists in the survey group is the “geoscience communication” aspect of the paper.

> This is a fair summary. Although the results may be unsurprising, they were and continue to be of use to the PhD student and inter-disciplinary supervisory team in focussing the possible routes forward in designing an ANN. Illustratively, the use of derived parameters such as height (*H*) and width (*W*) is a 'natural' approach for a geomorphologist (e.g. lead author), but not to machine learning specialists, even in a data limited scenario. As such we argue that there is value in exploring and confirming what might appear obvious from some viewpoints.

The shorter-than-short article format seems to have resulted in a few leaps being made that would benefit from further explanation. I will highlight these in my technical comments below.

There are also a few scattered errata

> Thank you. We have endeavoured to find and correct these.

...., but overall the text is clear.

> Thank you.

Abstract, elsewhere throughout - the way the profiles are described is highly inconsistent. Here and elsewhere they are called distance-depth profiles, presumably referring to bathymetric depth, but they are plotted as height vs distance, ie, topographic profiles, and presented as such to the survey respondents. The text later switches to talking about the shear velocity experiment profiles as elevation time series, the explanation for which is not in the main text but buried in the figure caption. While the timeseries could be recast as spatial series related through the migration rate of the bedforms, this step seems to have been omitted, so they are really timeseries presented as topography to the survey respondents and lumped in with the other experiments in the paper abstract and introduction as topography.

> We agree that within the context of a concise paper, for a general geoscience communication audience (i.e. including non-geoscientists), a consistent approach to describing the profiles without using jargon (e.g. topography, time-series) is best. This simplification is now introduced explicitly, at the same point in the

manuscript as the data are introduced. Of course this hides the detail, some of which is retained in the figure caption (i.e. glacial and aeolian are from topographic DEMs, fluvial are time series of forms passing under a sensor, and marine data are time series of acoustic measurements taken from a moving boat).

Line 31 - “this has been attempted for experimental parameters” is vague, what experiments? What parameters? ML has been applied to lots and lots of things...
> Froehlich (2020) used machine learning to attempt to predict fluvial flow regime from the standard experimental parameters (e.g. Froude number) in numerous flume tank experiments and natural streams recorded in the literature. This is the closest ML application to what we are attempting, but the reviewer identifies that it is too distant to include without further explanation. This previous work is now incorporated in a more general statement in the sentence before, and the confusing statement removed.

46 - “Scale readily distinguishes...” I get that the purpose was to see if there was scale- independent data in the shapes that would allow discrimination of these bedforms, but if the ultimate purpose is to build a model to do so, wouldn’t scale be included as it’s such reliable indicator?
> An exciting possibility we are exploring is the creation of a scale-independent ANN model, which might have the capability to be used across environments as from this it might be possible to gain insights into any unifying elements or characteristics of physical processes (e.g. to comment on Martian dunes from Earth analogues). For this, our current view is that scale is best omitted else the ANN would in effect be internally creating a separate model for each environment without transferability.

46 - Also, it appears that some sense of relative scale WAS in fact provided on the survey. The tick mark spacing on the frame of each profile seems to have been scaled along with the profiles when the extracted bedforms were rescaled. It is therefore easy to tell from the height of the bedform relative to the tick mark spacing that certain profiles came from a specific “training” profile. It would be hard to determine whether survey respondents registered this or not post hoc, and to me it calls the results into question. Why were the tick marks included on the “scale free” profiles at all?
> The reviewer is correct that, in hindsight, it would probably have been better to omit tick-marks entirely from the survey. However, we are confident the results we focus upon are sound as we asked questions in the survey to elicit whether or not such effects/biases were present. Our view for this primarily comes from two questions. First, 'a scale' (or similar) was the dominant response in a number of questions during the survey (e.g. Q3.36) paraphrased as "*What contextual information would have helped*?". Thus, if participants used tick-marks as a reliable scale indicator, it was a subconscious process. Also, when asked if anything made them feel confident of otherwise about their decisions (e.g. Q3.35), no participant mentioned using the tick-marks in the way the reviewer describes.
> The survey questions and responses are provided in the accompanying material for inspection, however to be concise and in line with the scope of the *GC Insights* format (i.e. "*does not need to be fully explored*") various aspects of the contextual information there (e.g. data resolution, roughness) are not explicitly considered in the main text. Consistent data quality, however, is one reason for our focus on the fluvial experiments in terms of our analysis and conclusion.
> We also focus on conclusions that are insensitive to such possible weaknesses in the research method e.g. '*short sequences are better than individuals*'.

57 - the parenthetical “(experiment number)” is initially confusing, replace perhaps with “(coded by experiment number) “or something like that.
> Text modified in line with the reviewer's suggestion.

58 - here is where the text starts talking about time series instead of topography with no explanation unless you happen to read the figure caption first.
> Thank you for this guidance on first uses of terminology, we now use 'distance-height profiles' throughout unless unavoidable for technical correctness.

63-65 - “For Q1... expected of guesswork” this sentence is a bit overloaded with info and convoluted to read.

65-67 - “For Q2...” this sentence is also convoluted.
> We note the reviewer's point, however the GC Insights format dictates concise wording. The sentences may need re-reading, but we believe they are clear, saving words for descriptive and less technical parts of the article. We have removed the technical point in brackets relating to the 16% expectation to avoid confusion.

81 - more explanation needed, as far as I know glacier bed topography and aeolian ripples are most certainly NOT modeled by the same physical equations... unless you mean not “modeled” but “described empirically"... there are process and form similarities across many environments, but perhaps this statement is oversimplified?
> Thank you for reminding us that bedforms (e.g. aeolian, glacial) are often modelled with different equations, and seen as different, but they are similar enough that they *can be* described (quantitatively, qualitatively, using equations) similarly. To soften our statement, therefore, the text has been modified to change 'are often' to 'can be'. The references used all take a view where the environments are modelled similarly e.g Fowler (2002) in "*Evolution equations for dunes and drumlins*" uses the same Exner equations for both.

90 - this is hard to follow. There is a leap between the previous mention of preprocessing to the description of the preprocessing method, which is embedded in a topic sentence.
> Pre-processing is now introduced in the Methods section.

83-95 - This section of the discussion justifies the preprocessing and introduces the ANN, is it really discussion? Seems more like methods.
> Thank you for this prompt. We agree. The initial structure is a legacy of the way in which the work unfolded. The original mini-project reported here was only the survey, and we added the preliminary follow-up ANN work later, which in reality aided our discussion and use of the survey results. The ANN is now introduced in the methods section, with technical detail now supplied in accompanying pseudo-code, and results of the ANN are now in the results section.
How the preprocessing method follows from the survey results needs more explanation.
> Please see modified discussion.
As for the potential pitfall of training the ANN on the raw data, was this attempted? With what result? It is merely asserted that it might be a problem.
> Thank you. Our results on running the ANN on the raw time-series data are now reported, allowing us to more clearly present this argument.
As for the fitting of flat topped-cones, there are different ways to fit things, which may depend on scale of interest etc.... More info about this step is needed. (Actually some detail is provided back in the methods section - but it’s not clear that the SWT/frustum algorithm described there is the same as the “fitting flat topped cones” described here. Also, SWT is never defined as an acronym.)
> Thank you for this comment. We entirely agree that fitting the cones, or indeed other methods of deriving shape parameters, will produce somewhat different results. The method used is scale-independent (i.e. fitting is after scaling into a unit box), and explicitly accounts for data gaps and density variations. Full details are given in the references cited (Hillier, 2006; Hillier, 2008), sufficient to reproduce the method, and identical parameters (e.g. for the SWT filter) are used here as in those papers.
> The pre-processing is now introduced in Methods, and the term 'frustum' no longer used to clarify that flat-topped cones are being fitted by the SWT/shape-fitting algorithms.
> SWT acronym is now defined at first use.

95 - The aside here about taking the statistical moments of the signal seems to obviate the other preprocessing. It’s also not clear whether it means that the ANN is more successful when any of these moments are used individually or if they are used together. The next sentence, the conclusion that insights from the participants led to a more successful model, does not seem to follow from this sentence about statistics.
> For clarity and focus this sentence about statistical moments has been removed. The reviewer rightly identifies that this sentence is somewhat tangential, and in a concise paper it is necessary to focus on key illustrative results of work done.
> There are issues with the use of these four statistical moments together. For example, much of the skill comes from mean height, which is arbitrary with respect to bedform processes and so very unlikely to remain

useful even when transferring to different fluvial experiments. Unpacking this is beyond the scope of a *GC Insights* article. It will be covered in the geomorphology / machine learning paper to follow.

99 - “in the future”
> Text modified.

So, I’m not really sure what to make of this study. It seems like a good exercise overall, to understand whether expert knowledge might help build a better ANN classifier, but most of the insights could have been gleaned just from trying the ANN in the first place.
> We fully accept the comment about just trying various ANNs, but after trying various routes, and building and training numerous ANNs with relatively limited success, the survey helped up to focus the design process (i.e. provided insights to help design an effective ANN). Perhaps the main insights are (i) it is possible and (ii) there is value in 'helping' the ANN in the context of limited data, which is the opposite of machine learning communities default desire to train ANNs on raw data so as to impose as few *a priori* assumptions as possible. The text has been adapted to more clearly communicate this.
Coupled with some of the issues of scale info on the survey, the unclear link between the survey results and how they informed building the ANN, and the condensed article format leading to omitted context, it’s hard to recommend publication without substantial revision.
> Thank you for your constructive feedback, we have revised the text to clarify how the survey results informed the building of the ANN, and focussed the text to clarify its purpose.

Identifying conditions that sculpted bedforms: Human insights to build an effective AI

John K. Hillier[1], Chris Unsworth[2], Luke De Clerk[3], Sergey Savel'ev[3]

[1]Geography and Environment, Loughborough University, Loughborough, LE1 3TU, UK
[2]School of Ocean Sciences, Bangor University, Bangor, LL59 5AB, UK.
[3]Dept. Physics, Loughborough University, Loughborough, LE1 3TU, UK.

Correspondence to: John K Hillier (j.hillier@lboro.ac.uk)

Abstract. In a survey, 42 participants could visually recognise which flow conditions created bedforms (e.g. sand dunes, riverbed ripples) from distance-height profiles, particularly if multiple bedforms are present. Furthermore, an interpreter's geoscience expertise does not help whilst only one set of 'training' examples are needed, together indicating that an effective machine learning or 'AI' algorithm might be trained successfully from limited data alone especially if it is 'helped' (e.g. by pre-processing to fit flat-topped cones). Preliminary investigations with an artificial neural network confirm these insights.

1 Introduction

Environmental flows shape the surface they flow over. The variety of features produced (e.g. sand ripples on a beach), known as bedforms, reflect and preserve characteristics (e.g. speed, depth) of the flowing ice, water or air (Venditti, 2012; Bullard et al., 2011; Storrar and Stokes, 2007). The relationships between bedform morphology and flow are contested where observation is extremely difficult, such as under ice-sheets (e.g. Clark et al., 2018; Hillier et al., 2018; Rose, 1987; King et al., 2009), and best understood for unidirectional water flow over sand in a laboratory setting, mimicking a river (e.g. Fig. 1a). Even in this idealised fluvial setting, it is difficult to construct a 1-to-1 link between bedform type (e.g. ripples or dunes) and specific flow conditions (Venditti, 2012; Froehlich, 2020). Illustratively, ripples have a higher aspect ratio (H/L, for height (H) and length (L)) than dunes (e.g. Allen, 1968); yet the observational ranges overlap (Venditti, 2012; Yalin, 1972), creating uncertainty when attempting to link morphology with hydraulic conditions. Many variables related to hydraulics and/or the physics of sediment movement have been proposed to remove the overlap in bedform stability diagrams such as Fig. 1a (see Venditti, 2012). Only recently, has a distinct and non-overlapping zonation of bedform type and flow structure been developed using a quantity called shear velocity (Duran Vinet et al., 2019). Inverting this result may help realise the aspiration of developing a means to reliably infer flow conditions from bedform morphology (e.g. Duran Vinet et al., 2019; Venditti, 2012; Myrow et al., 2018), which is often the only option (e.g. sedimentary structures preserving the geological past, Mars)(e.g. Ohata et al., 2017; Edgett and Lancaster, 1993).

Machine learning or 'AI' algorithms, such as Artificial Neural Networks (ANNs) have great potential in geomorphology (e.g. Sofia et al., 2016; Froehlich, 2020; Valentine and Kalnins, 2016; Shumack et al., 2020) and offer an opportunity to examine this problem as they do not assume simple (e.g. linear or 1-to-1) relationships between inputs and predicted variables (Wang et al., 2009; Faruk, 2010). There may be unexploited quantitative morphological subtleties to categorise bedforms, or even to accurately position them on stability diagrams. This work examines the scope for using ANNs to distinguish the flow conditions in which relict bedforms originated by asking if the ability exists in non-artificial (human) intelligence for two particulars:

Q1 - Is it possible to identify the environment (e.g. river, desert) of a bedform's genesis from its shape?

Q2 - In the fluvial environment, is it possible to distinguish flow conditions?

2 Method, Data & Ethics

An online survey was conducted, initially at the '*Non-equilibrium flows and landforms*' workshop (19^{th} May 2021), with participation expanded using authors' close contacts (friends, colleagues, and family). For Q1, participants attributed distance-height profiles across 34 individual bedforms, and 13 bedform sequences (≥3 bedforms) to one of four environments (*fluvial* [river], *glacial*, *marine*, *aeolian* [desert]). For Q2, participants ranked three profiles according to flow strength (shear velocity), thrice for individual forms, and thrice for bedform sequences. Examples were provided to isolate visual shape analysis from prior knowledge (Fig. 1b), black and white profiles were used to exclude contextual clues (e.g. dataset characteristics, other features in the landscape), and the order of options (e.g. B, A, C) was shuffled for each participant to prevent bias. Scale (e.g. metres) readily distinguishes environment without using bedform shape, so it was not given.

Ethical approval was given by the Ethics Review Sub-Committee at Loughborough University.

Aeolian data are ASTER (v2) across linear and transverse dune types from the Namib desert (Bullard et al., 2011), glacial are from near Lough Gara in Ireland (Hillier and Smith, 2008), fluvial are from four laboratory experiments (Expts. 1-4) of non-linearly increasing shear velocity (Unsworth, 2015), and marine data are from the Irish Sea. For consistency, simplicity and accessibility, these varied data are referred to as distance-height profiles, with 'time' equated to 'distance' where strictly necessary. For the survey, representative examples of individual bedforms and sequences were manually selected from these datasets. Pre-processing to estimate bedform's height (*H*) and width (*W*) - see Fig. 1c & d - used the Spatial Wavelet Transform (SWT) algorithm (Hillier, 2008) and fitting of flat-topped cones (Hillier, 2006).

ANN analysis to follow up the survey used a Multi-Layered Perceptron (MLP) with four hidden layers with 28, 56, 56 and 28 nodes, each with a ReLU activation function. In a baseline analysis, input to predict the fluvial flow regime (coded by experiment number) was non-overlapping profile segments 160 seconds long. After this, to 'help' the ANN bedform shapes (H, W) were input once each per analysis, either (i) individually or (ii) as groups of five bedforms in increasing size order, selected at random without replacement. Weights and biases were updated using the Adam Optimiser of PyTorch using a loss function that calculates the Mean Squared Error (MSE), all within a feedforward back-propagation approach.

3 Results

Of the 42 survey participants 25 self-identified as geoscientists, and 16 did not. For Q1, participants correctly identified the one of four environments (e.g. fluvial, aeolian) in which individual features originated 32% of the time, slightly if significantly (2-tailed t-test, $p \ll 0.01$) better than the 25% expected of guesswork. This rises to 51% for bedform sequences. For Q2, participants ranked entirely correctly 3 flow strengths (Expts. 1-3) for 46% of individual features, and 60% of sequences, much better than the 16% expected of guesswork ($p \ll 0.01$).

In none of the questions did geoscientists perform better than non-geoscientists, with mean percentages of correct answers being indistinguishable (2-tailed t-test, $p > 0.05$). The overall sentiment is encapsulated by one comment:

> "*I felt this was a geometrical exercise of recognising same patterns at different scales. I did not feel that my experience as an "expert" in bedforms really made any difference from, say, my son taking the test.*"

Several participants commented that their ability to distinguish environments might be to do with characteristics of the data (e.g. smoothness due to data resolution), not bedform shape. This is a potential pit-fall of training an ANN, avoided here by only analysing the fluvial data.

In the baseline ANN analysis, flow regime was predicted poorly ($r^2 = 0.03$). Visually, individual bedforms' shapes (H, W) overlap between experiments but their trends, and averages over a number of points, are distinctly different (Fig. 1d). This maps directly to results of the ANN (Fig. 1e). Individual forms are weakly predicted (light grey, $r^2 = 0.11$), but sub-sets of 5 bedforms more strongly so (grey, $r^2 = 0.56$), particularly if very small bedforms present in all experiments ($H < 0.5$ cm) are excluded (dark grey, $r^2 = 0.80$).

4 Discussion

Morphologies from differing environments (e.g. glacial, fluvial) can be viewed as similar, indicators of analogous processes at work (e.g. Shaw, 1983), and modelled with identical equations (e.g. Fowler, 2002; Duran Vinet et al., 2019) or statistics (e.g. Hillier et al., 2016; Einstein, 1937). Despite well-recognised similarities in appearance, however, the survey results clearly demonstrate a level of ability to distinguish flow conditions from distance-height data of the bed, and imply that an ANN should perform better if utilising sequences of bedforms rather than evaluating individual forms in isolation. Interestingly, geoscientists' *a priori* and contextual knowledge added little, indicating that training an ANN on distance-height data alone should be productive. Furthermore, one training dataset sufficed for the survey's participants, prompting the idea that the requirement to train ANNs performing pure pattern recognition on 1000s of datasets (e.g. Bishop, 1996) might be mitigated by 'helping' the ANN by pre-processing to identify key aspects of a geomorphological pattern (e.g. *H* and *W*). This would be useful as examples in nature are often limited in number.

Preliminary analysis with an ANN (Fig. 1) supports these speculations. It demonstrates that an AI with some level of predictive efficacy can be built, and this improves by using bedform sequences (Fig. 1e). Simple shape parameters that would have been readily understood by all participants (*H* and *W*) were selected to help the ANN, although relevant geo-statistical properties (e.g. Malinverno, 1988; Powell et al., 2016; Singh et al., 2011) might have been used instead. The increase in predictive skill from r^2 of 0.03 to 0.80 with pre-processing 'help' demonstrates, in principle, the utility of this approach when building an effective AI for geomorphology. This study was on equilibrium conditions but in the future ANNs may be key to disentangling forms and flow for transitional, non-equilibrium conditions (e.g. see Myrow et al., 2018).

Supporting material

The survey form and anonymised answers are provided for completeness and transparency. For the ANN, input data with associated README file and pseudo-code are provided to make the work reproducible.

Acknowledgements

The “Non - Equilibrium Turbulence and landforms workshop” was organised by Tim Marjoribanks, Chris Keylock, Christopher A. Unsworth, Daniel R. Parsons and Jonny Higham, with support from the British Society of Geomorphology. We thank Matt Baddock for preparing the aeolian data, and the 42 anonymous participants for completing the survey.

References

Allen, J. R. L.: Their Relation to Patterns of Water and Sediment Motion, North Holland Publishing Company, Amsterdam, 433 pp., 1968.

Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford University Press, 502 pp., 1996.

Bullard, J., White, K., and Livingstone, I.: Morphometric analysis of aeolian bedforms in the Namib Sand Sea using ASTER data, 36, 1534–1549, 2011.

Clark, C. D., Ely, J. C., Spagnolo, M., Hahn, U., Hughes, A., and Stokes, C. R.: Spatial orgainzation of drumlins, 43, 499–513, https://doi.org/10.1002/esp.4192, 2018.

Duran Vinet, O. D., Adreiotti, B., Claudin, P., and Winter, C.: A unified model of ripples and dunes in water and planetary environments, 12, 345–350, https://doi.org/10.1038/s41561-019-0336-4, 2019.

Edgett, K. S. and Lancaster, N.: Volcaniclastic aeolian dunes: terrestrial examples and application to martian sands, 25, 271–297, https://doi.org/10.1006/jare.1993.1061, 1993.

Einstein, H. A.: Bedload tranport as a probability problem, in: Sedimentation, edited by: Shen, W. H., Colorado State University, Fort Collins, C1–C105, 1937.

Faruk, D. Ö.: A hybrid neural network and ARIMA model for water quality time series prediction, 23, 586–594, https://doi.org/10.1016/j.engappai.2009.09.015, 2010.

Fowler, A. C.: Evolution equations for dunes and drumlins, 96, 377–387, 2002.

Froehlich, D. C.: Neural Network Prediction of Alluvial Stream Bedforms, 146, 04020084, https://doi.org/10.1061/(ASCE)HY.1943-7900.0001831, 2020.

Hillier, J. K.: Pacific seamount volcanism in space and time, 168, 877–889, https://doi.org/10.1111/j.1365-246X.2006.03250.x, 2006.

Hillier, J. K.: Seamount detection and isolation with a modified wavelet transform, 20, https://doi.org/10.1111/j.1365-2117.2008.00382.x, 2008.

Hillier, J. K. and Smith, M.: Residual relief separation: digital elevation model enhancement for geomorphological mapping, 33, 2266–2276, https://doi.org/10.1002/esp, 2008.

Hillier, J. K., Kougioumtzoglou, I. A., Stokes, C. R., Smith, M. J., and Clark, C. D.: Exploring explanations of subglacial bedform sizes using statistical models, 11, e0159489, https://doi.org/10.1371/journal.pone.0159489, 2016.

Hillier, J. K., Benediktsson, Í. Ö., Dowling, T. P. F., and Schomacker, A.: Production and preservation of the smallest drumlins, https://doi.org/10.1080/11035897.2018.1457714, 2018.

King, E. C., Hindmarsh, R. C. A., and Stokes, C. R.: Formation of mega-scale glacial lineations observed beneath a west Antarctic ice stream, 2, 585–596, 2009.

Lewis, D. W. and McConchie, D. M.: Practical sedimentology, 2nd ed., Springer, US, 212 pp., 1994.

Malinverno, A.: Statistical Studies Of Seafloor Geomorphology, PhD Thesis, Lamont, 1988.

Myrow, P. M., Jerolmack, D. J., and Perron, J. T.: Bedform Disequilibrium, 88, 1096–1113, https://doi.org/10.2110/jsr.2018.55, 2018.

Ohata, K., Naruse, H., Yokokawa, M., and Viparelli, E.: New Bedform Phase Diagrams and Discriminant Functions for Formative Conditions of Bedforms in Open-Channel Flows, 122, 2139–2158, https://doi.org/10.1002/2017JF004290, 2017.

Powell, D. M., Ockelford, A., Rice, S. P., Hillier, J. K., Nguyen, T., Reid, I., Tate, N. J., and Ackerley, D.: Structural properties of mobile armors formed at different flow strengths in gravel-bed rivers, 121, https://doi.org/10.1002/2015JF003794, 2016.

Rose, J.: Drumlins as part of a glacier bedform continuum, in: Drumlin Symposium, edited by: Menzies, J. and Rose, J., Balkema, Rotterdam, 103–116, 1987.

Shaw, J.: Drumlin formation related to inverted melt-water erosional marks, 29, 461–479, 1983.

Shumack, S., Hesse, P., and Farebrother, W.: Deep learning for dune pattern mapping with the AW3D30 global surface model, 45, 2417–2431, https://doi.org/10.1002/esp.4888, 2020.

Singh, A., Lanzoni, S., and Wilcock, P. R.: Multiscale statistical characterization of migrating bed forms in gravel and sand rivers, 47, W12526, https://doi.org/10.1029/2010WR010122, 2011.

Sofia, G., Hillier, J. K., and Conway, S. J.: Frontiers in Geomorphometry and Earth Surface Dynamics: Possibilities, limitations and perspectives, 4, https://doi.org/10.5194/esurf-4-721-2016, 2016.

Southard, J. B. and Boguchwal, L. A.: Bed configurations in steady unidirectional water flow part 2. Synthesis of flume data, 60, 658–679, 1990.

Storrar, R. and Stokes, C. R.: A glacial geomorphological map of Victoria Island, Canadian Arctic., 3, 191–201, https://doi.org/10.1080/jom.2007.9710838, 2007.

Unsworth, C.: River Dunes in Unsteady Conditions, PhD, University of Hull, 2015.

Valentine, A. and Kalnins, L.: An introduction to learning algorithms and potential applications in geomorphometry and Earth surface dynamics, 4, 445–460, https://doi.org/10.5194/esurf-4-445-2016, 2016.

Venditti, J. G.: Bedforms in Sand-Bedded Rivers, in: Treatise on Geomorphology, edited by: Shroder, J. and Wohl, E., 137–162, 2012.

Wang, W., Chau, K., Cheng, C., and Qui, L.: A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series, 374, 294–306, https://doi.org/10.1016/j.jhydrol.2009.06.019, 2009.

Yalin, M. S.: Mechanics of Sediment Transport., Pergamon, Oxford., 1972.

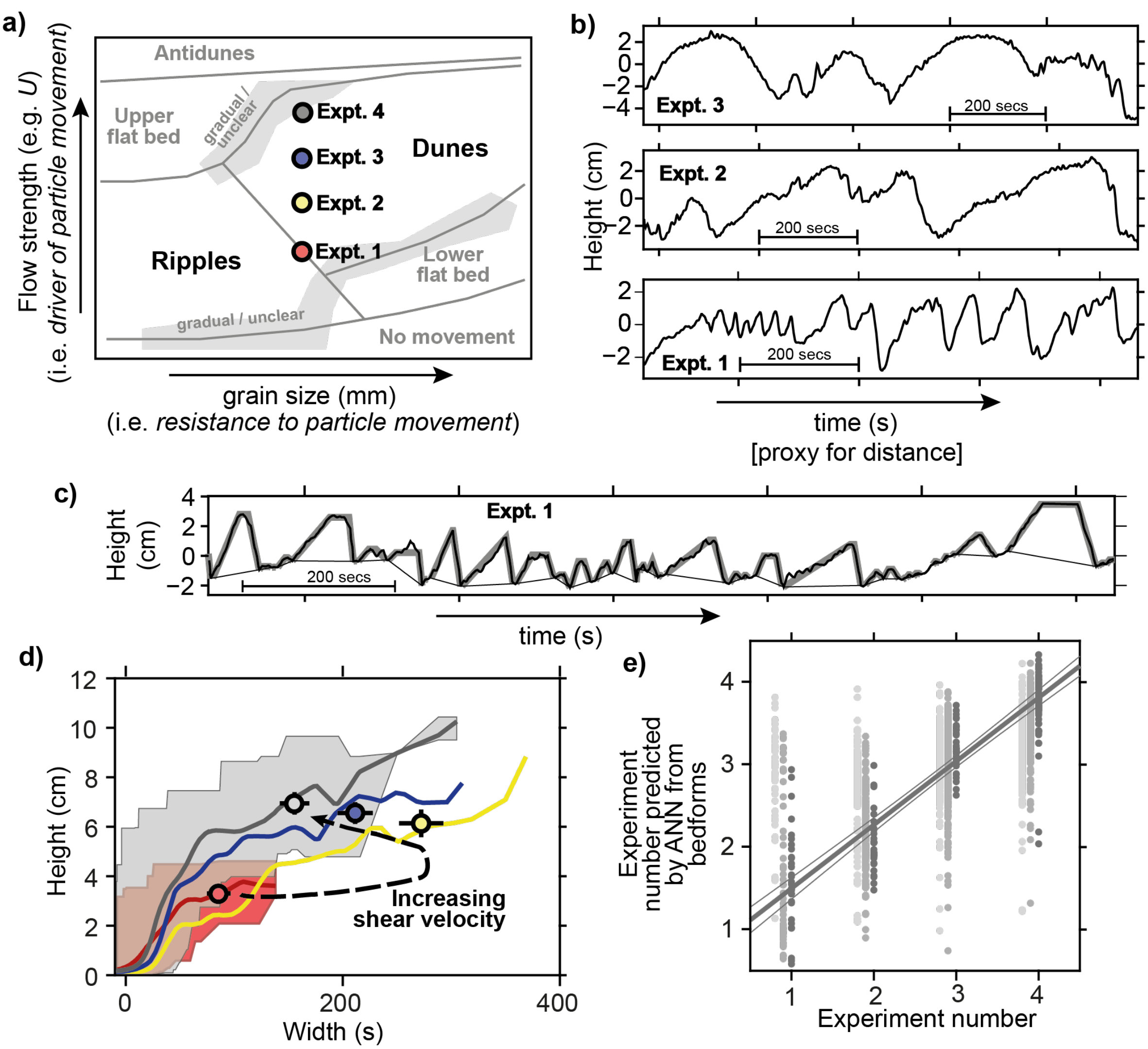

Figure 1: (a) Illustrative bedform stability 'phase diagram' for unidirectional fluvial (i.e. river) bedforms, synthesized from multiple sources (Ohata et al., 2017; Lewis and McConchie, 1994; Southard and Boguchwal, 1990). Main types considered here (i.e. ripples and dunes) are highlighted. Experiments 1-4 are positioned indicatively. (b) 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. (c) Example of how H and W are determined. Measured heights (thick black line) are processed using the SWT algorithm to identify bedforms, drawing a line beneath them (thin black line) then approximated as flat-topped cones (grey lines). SWT parameters as Hillier (2008). (d) Height-width relationships for the 4 experiments, with colours as in (a): lines are sliding means with W (Gaussian weights, width 60), shaded areas are full ranges for Expts. 1 & 4, and dots are the means (±2σ) of upper quartile of the data when the small bedforms (i.e. $H < 0.5$ cm) are excluded. (e) Comparison of actual experiment number for out of sample prediction by the ANN using H and W: individual bedforms (light grey), subsets of 5 bedforms with (grey) and excluding (dark grey) small bedforms.