GC Insights: 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. Insights from a geoscience communication activity, verified using preliminary investigations with an artificial neural network, illustrate that observation of humans' abilities can help design an effective Artificial Intelligence or 'AI'. Even given only one set of 'training' examples, survey participants could visually recognise which flow conditions created bedforms (e.g. sand dunes, riverbed ripples) from their shapes, but an interpreter's geoscience expertise does not help. Together, these observations were interpreted as indicating that a machine learning algorithm might be trained successfully from limited data, particularly if it is 'helped' by pre-processing bedforms into a simple shape familiar from childhood play.

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. Hillier et al., 2018; King et al., 2009), and best understood for unidirectional water flow over sand in a laboratory setting, mimicking a river. 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 (height/length) than dunes (e.g. Allen, 1968); yet the observational ranges overlap (Venditti, 2012), 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. Only recently has a distinct and non-overlapping zonation of bedform type and flow-sediment condition 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, which is often the only option for inferring past environmental conditions on Earth (Leary and Ganti, 2020) or Mars (Ohata et al., 2017; Edgett and Lancaster, 1993).

Machine learning or 'AI' algorithms, such as Artificial Neural Networks (ANNs) have great potential in geomorphology (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). Unexploited morphological subtleties may exist by which 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 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 (19th May 2021), expanded to participants without geomorphological expertise 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 (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. Distance-height profiles of these data were created, although for the fluvial measurements 'time' was used as a proxy for 'distance' (see Fig. 1b). For the survey, representative examples of individual bedforms and sequences were manually selected from these datasets. Pre-processing to estimate bedforms' 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 pseudo-sequences - 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, all within a feedforward back-propagation algorithm.

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 pitfall of training an ANN, avoided here by only analysing the fluvial data.

In the baseline ANN analysis, flow regime was predicted poorly ($r^2 = 0.03$). Fitting a simplified geometry (H, W) to bedforms improves results dramatically, particularly if pseudo-sequences of bedforms are used (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, $r^2 = 0.80$). This is consistent with a visual assessment (Fig. 1d) where individual morphologies overlap between experiments but their trends, and averages over a number of bedforms, are distinctly different.

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 similarities in appearance, the survey results clearly demonstrate a level of ability to distinguish flow conditions from distance-height data of the bed and, unsurprisingly, 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 all required visual cues lie within the distance-height profiles. Furthermore, one training dataset sufficed for the survey's participants, a stark contrast to the 1000s of datasets required to train ANNs performing pure pattern recognition (e.g. Bishop, 1996), suggesting that participants drew on significant previous learning (e.g. identification of basic idealized shapes). Together, these observations prompt the testable idea that an effective ANN might be efficiently trained by ‘helping’ it via pre-processing profiles into simple shape parameters that would have been readily understood by all participants (H, W).

Preliminary analysis with an ANN supports our speculations. It demonstrates that an AI with predictive efficacy can be built using limited data, improved by using bedform sequences (Fig. 1e). The increase in predictive skill to 0.80 with pre-processing 'help' demonstrates, in principle, the utility of this approach when building an effective AI for geomorphology that avoids the crippling need for 1000s of datasets when examples in nature are often limited in number. Speculatively, it follows that machine learning techniques might work well and be trained efficiently wherever non-experts make good decisions based on images. This study was on equilibrium conditions, but illustrates that ANNs may be key to linking forms and flow for transitional, non-equilibrium conditions (e.g. Myrow et al., 2018).

Code & data Availability

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

Supplement link

To be added by Copernicus.

Author contributions

JH conceived and led the work. JH and CU designed and ran data collection at the workshop. LDC, JH and CU undertook analyses and data preparation. All authors contributed to the analytical design, writing the manuscript, and development of the understanding and ideas presented.

Competing interests

Hillier is an executive editor of *Geoscience Communication*. The peer-review process was guided by an independent editors (handling and executive), and the authors have no other competing interests to declare

Acknowledgements

We thank Dylan Ward and an anonymous reviewer for their thoughtful and challenging comments which significantly improved the manuscript. 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.

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.

Leary, K. C. P. and Ganti, V.: Preserved Fluvial Cross Strata Record Bedform Disequilibrium Dynamics, 47, 1–11, 2020.

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

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.

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.

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.

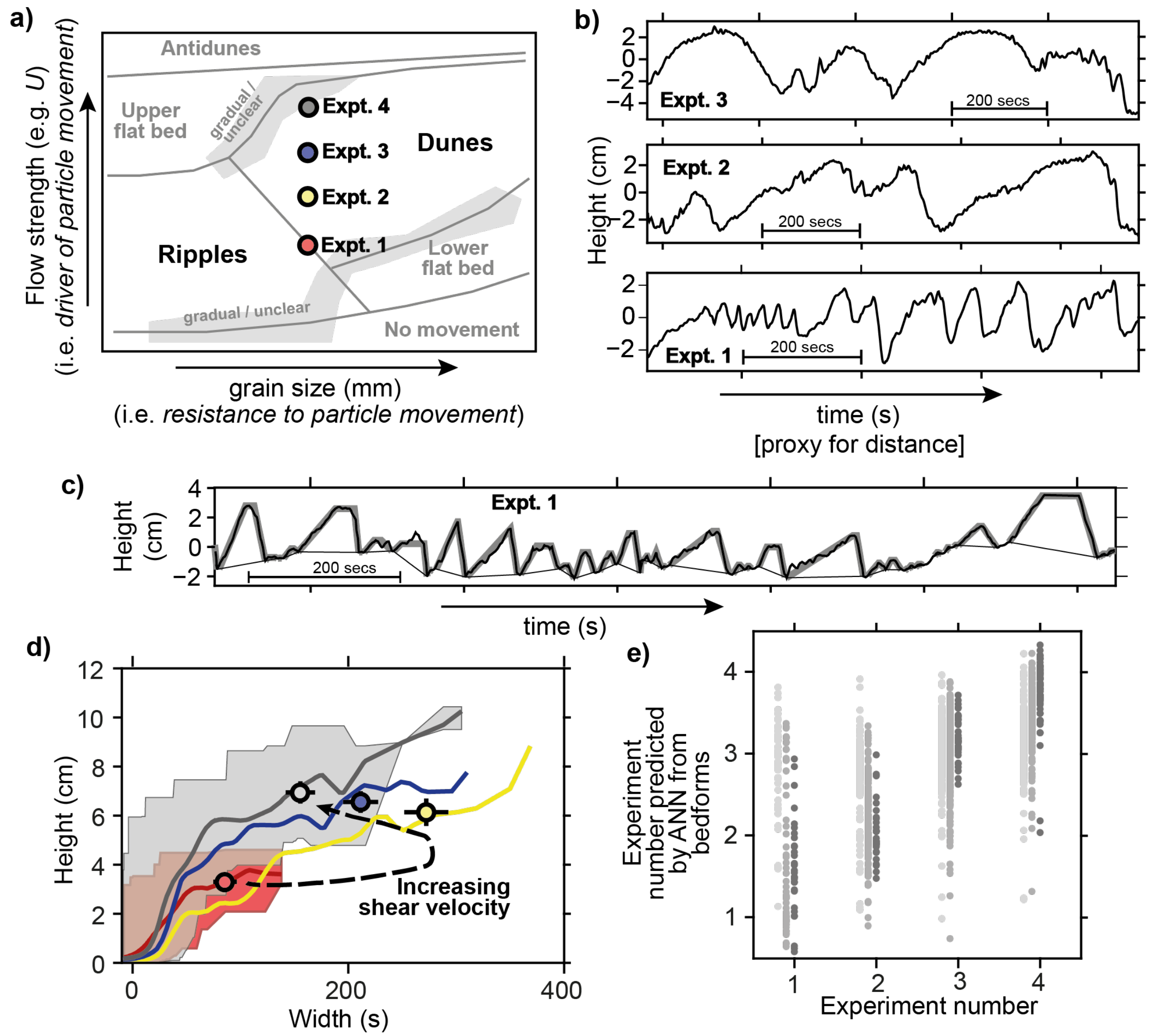

Figure 1: (a) Illustrative bedform stability 'phase diagram' for unidirectional fluvial (i.e. river) bedforms, synthesized from multiple sources (Ohata et al., 2017; Lewis and McConchie, 1994; Southard and Boguchwal, 1990). Main types considered here (i.e. ripples and dunes) are highlighted. Experiments 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.