Review of: Hosseiny *et al.* Development of a machine learning model for river bedload https://doi.org/10.5194/esurf-2022-23 -- Basil Gomez

It has recently been claimed that the search for a formula or model that can "be broadly applied" is a fallacious pursuit, as not all rivers function in the same transport regime (Gomez and Soar, 2022). This is because, although bedload transport efficiency and the size of sediment in motion are adjusted to the environmentally controlled rate at which sediment is supplied to a river system (Gomez, 2022), the time-variations in transport rates observed at a reference section are the product of transitory, in-channel changes in sediment availability; related, for example, to the passage of bedforms, scour and fill, the formation/breakup of armor, or the injection/exhaustion of sediment derived from proximal sources. In only one transport regime does the narrow range of inherent variability in the transport rate at a given flow magnitude indicate that it might be possible to compute temporally representative transport rates for a range of flows. It has also been suggested the sophisticated passive and active monitoring technologies, that routinely generate long-term, continuous high-resolution datasets, have transformed bedload research from a data-scarce (theory-driven) into a data-rich (data-driven) science (Gomez *et al.*, 2022). Through the use of modern methodologies designed to tease out trends, reveal structure in bivariate (and multivariate) datasets, these 'big data can help elucidate nonlinear and complex behavior, reveal structure and trends that were previously obscured, and link bedload transport dynamics to the underlying drivers across different timescales. A procedure (quantile LOWESS) it is advocated be applied to such data is also a machine learning task (Gomez and Soar, 2022).

Machine learning may indeed afford "a compelling approach to leverage the growing wealth of bedload transport observations towards the development of a data driven predictive model", but I fundamentally disagree with the authors' approach to predicting (or testing methods for predicting) bedload transport rates. Machine learning techniques clearly have the potential to help isolate patterns (that must necessarily be explained in physical terms) if the data utilized embrace the specificity of rivers in the universe of fluvial systems with respect, for example, to the transport regimes or efficiency metapopulations Gomez and Soar (2022) and Gomez (2022) identified. However, many of the rivers in the collection of data the authors draw on are referred to in the three aforementioned papers, and it should be abundantly clear that applying any machine learning technique to such a profusion of data is unlikely to yield results that "lay[s] the foundations for efficient and accurate predictions of river bedload". This is because, as Gomez and Church (1989) demonstrated, not all bedload transport data are created equal; some characterize 'equilibrium conditions', while other data, for example, represent the condition of partial transport where not all size fractions are transport in the same proportion as they are present on the bed.

In light of the above it would seem, at the very least, appropriate to specify the transport conditions to which a particular formula pertain, because most formulae attempt to predict the maximum rate of bedload transport for a given set of conditions and use data that match these criteria in any test of their performance. Moreover, there are also obvious errors in the data set the authors use that could easily be resolved with reference to the original works. As an example, I offer my (Gomez, 1983) data that are incorrectly recorded as having a uniform D_{50} of 11m when, in fact, reference to the original paper shows a D_{50} of from ~3 mm to ~1.3 mm was specified for each measurement (transect – for reference, the size of the H-S sampler is also incorrectly specified and only an unspecified portion of the data are utilized). Moreover, the conditions these (my) data describe are clearly not a test of any bedload transport formulae or methodology for prediction! There is also nonconformity in the conditions the data seek to describe, inasmuch as some of the essential variables (the energy slope is the most common example) are reported as constant values, whereas in actuality they vary. The median grain size of the bedload is another value which is, in fact (see Gomez, 2022), rarely specified reported for each rate and is often a surrogate value (such as the median diameter of the local bed material, from which the bedload is assumed to have been derived). For all the above reasons, it would seem unlikely that any machine learning process would provide intelligible or meaningful results when applied indiscriminately to the growing wealth of bedload transport observations.

I also have a concern about the data screening, which apparently involved arbitrarily removing extreme (the upper and lower 5th percentiles) data, because "outliers generally degrade the overall performance of the resulting model", and the indiscriminate use of four bedload transport models to compare to and build intuition for the predictions of the ANN model. The authors "compared bedload flux measurements to predictions from these four bedload transport models and the trained ANN model". Do they really expect, for example, the Wilcock and Crowe or Recking models to be applicable to Mountain Creek, a sand-bed stream with an assumed bedload D_{50} of 0.9 mm?

In summary, I am sensitive to the fact that the authors may have been unaware of, or chose to ignore, papers that were published as their manuscript was being prepared for publication. However, I cannot support the arbitrary and indiscriminate use of data and the application of bedload transport models in their study. There are a number of very large (comprising hundreds to tens of thousands of measurements, see Gomez *et al.*, 2022 and Gomez and Soar, 2022 for examples), high quality bedload transport data sets to which the bedload transport models they utilize are eminently applicable that would also serve as a rigorous test of the ANN model they have developed. My recommendation is that the paper be withdrawn as I do not consider it to be suitable for publication in its present form. The ANN model, which represents the core of the paper, may have utility, but it needs more meticulous testing using the 'big data' to which it is inherently more applicable.

Gomez, B., 1983, Temporal variations in bedload transport rates: the effect of progressive bed armouring, *Earth Surface Processes and Landforms*, **8**, 41 54.

Gomez, B., 2022, The efficiency of the river machine, *Geomorphology*, **410**, 108271, doi:10.1016/j.geomorph.2022.108271.

Gomez, B. and Church, M., 1989, An assessment of bedload sediment transport formulae for gravel bed rivers, *Water Resources Research*, **25**, 1161-1186.

Gomez, B. and Soar, P.J., 2022, Bedload transport: beyond intractability, *Royal Society Open Science*, **9**: 211932, doi: 10.1098/rsos.211932.

Gomez, B., Soar, P.J. and Downs, P.W., 2022, Good vibrations: Big data impact bedload research, *Earth Surface Processes and Landforms*, **47**, 129–142.