Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2019-81-RC2, 2019

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Ensemble models from machine learning: an example of wave runup and coastal dune erosion" *by* Tomas Beuzen et al.

Anonymous Referee #2

Received and published: 27 May 2019

Review of: Ensemble models from machine learning: an example of wave runup and coastal dune erosion

By Thomas Beuzen, Evan B. Goldstein, and Kristen D. Splinter

I have reviewed the above manuscript and find that it will be acceptable for publication in Natural Hazards and Earth System Sciences following only a very few very minor revisions. The paper "Ensemble models from machine learning: an example of wave runup and coastal dune erosion" uses a machine learning technique, Gaussian Process Regression, to develop a probabilistic wave runup model able to be implemented in an ensemble approach. The wave runup model is then applied to a deterministic dune erosion model to demonstrate the power of hybrid approaches over typical deterministic approaches. This topic is of considerable importance, is definitely appropriate

for this journal, and will be useful to a broad audience. I have only a couple of general and specific comments that might aid in improving the manuscript and these are offered below. This manuscript is extremely well written and therefore I am not submitting an annotated version of the manuscript.

General Comments: The authors make the bold (and most likely correct) statement that the development of a perfect deterministic parameterization of wave runup using only the typical inputs of beach slope, wave height, and wave period is improbable. They then go on to develop a GP runup model that has higher skill than the most typical deterministic runup model used today (Stockdon et al., 2006). However, to build this new model they still use the same three easily obtainable inputs. While perfectly reasonable for this paper's demonstration purposes, I am left wondering whether or not GP could be used to build an even better runup model if other input forcing dimensions were included? Figure 4 appears to have some structure in it, with low values of R2 overpredicted and high values underpredicted. Can we learn something from this? Even a few suggestions and/or speculations from the authors would be welcome about machine learning directions for developing even better runup models. In developing the input Hs and Tp time series for both the creation of the runup model and for the ultimate test against the dune erosion event, it is mentioned that SWAN is used to transform all conditions into the nearshore before being linear back shoaled. Did the authors really run 100s to 1000s of individual SWAN simulations? This effort seems like it must have had a high computational cost? Since the paper emphasizes the efficiency of the GP runup model some more detail of this step in the process is warranted. Have the authors considered developing simple look up tables, or better yet, a GP model of SWAN to simplify this stage of the process? The decision to use MDA for developing the training data seems sound. However, a list, or discussion of other possible space filling algorithms might be useful for readers embarking on their own GP applications. I commend the authors for their relatively parsimonious and clear explanation of GP theory in section 2.1. However, I suspect that this treatment will still be an occasionally opaque to some readers (including this reviewer). My only

suggestion here is to continue to work on describing machine learning approaches such as GP in as clear of terms as possible. This paper does this as well as I have seen.

Specific Comments: Line 81-82: Maybe add to this growing body of literature by including: Parker, K., P. Ruggiero, K. Serafin, and D. Hill. 2019. "Emulation as an Approach for Rapid Estuarine Modeling." Coastal Engineering 150: 79–93. https://doi.org/10.1016/j.coastaleng.2019.03.004. Line 235-236: I thank the author for identifying which toolkit they used in developing the runup GP model. However, it might be helpful for a broad group of readers if the authors listed other potential toolkits that could also have been used – say for example in Matlab, or R?

Line 544-566: The statement about 10,000 samples taking less than one second on a standard desktop computer is repetitive at this point.

Thanks very much for the opportunity to review this very exciting manuscript.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-81, 2019.