

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/nhess-2022-221-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2022-221

Anonymous Referee #1

Referee comment on "A predictive equation for wave setup using genetic programming"
by Charline Dalinghaus et al., Nat. Hazards Earth Syst. Sci. Discuss.,
<https://doi.org/10.5194/nhess-2022-221-RC1>, 2022

The paper discusses finding equations for wave setup using machine learning (ML) algorithms, and the main contribution is in applying ML algorithms to the geophysical problems. In general, the paper is clearly presented and well organized. However, I have two main concerns about this paper.

The first question is on the results by ML algorithms. One of the main contributions of this paper is the equation (14), but the physical implication behind it is unclear. As authors understand, complicated equations driven by the ML algorithms will give well-fitting results. At the same time, the equations are meaningful if they are physically interpretable. There are three terms in eq. (14), and the last term reversely relates the setup height (M) with the grain size (D_{50}). In line 297-298, authors mentioned "*This second order effect could tentatively be related to beach permeability, which increases with sediment size and results in a lower setup.*" However, to my knowledge, the permeability is related to the distribution of the grain size, not the average of the grain size.

The second one is on the sample size and data availability. The sample size of 491 cases is relatively small to apply ML algorithms. And it seems that more data are available from the provided link (<https://coastalhub.science/data>). It would be better to mention the reasons to use Stockdon and Holman 2011 data only. Moreover, I could not find the grain size (D_{50}) from Stockdon and Holman, 2011 (<https://pubs.usgs.gov/ds/602/>) or (<https://coastalhub.science/data>). Authors need to provide a complete data set, and how they acquired the grain size.

Here are some minor comments:

L114 "has open" -> "has opened"

L311 "Although presenting extremely promising results" -> ambiguous

L330 "being a data-driven technique, it will only get more accurate as more data becomes available." Not just more data but high quality data are necessary.

L332 "able to represent"->"representing"