Dear Dr. Ribas,

Thank you for your thoughtful review of our manuscript. We have carefully reviewed your comments and have revised the manuscript accordingly. Please see below our responses to each one of your comments.

This article presents a new application of genetic algorithms to obtain empirical formulas for the maximum wave setup. The authors train and test a GP model with 9 different data sets of wave setup measurements. The pieces used by the GP model to build the equations are 6 variables and 6 operators, which are among those used in previous setup formulas. Then, they compare the obtained predictors with other 7 existing formulas for maximum wave setup. They obtain an extremely simple predictor that give wave setup with the same accuracy than the best of the previous existing formulas and a more complex expression that outperforms them.

Obtaining more accurate formulas for the maximum wave setup is important to increase our capacity to predict flooding since wave setup can contribute significantly to inundation episodes. This is especially crucial in the present framework of climate change and the urgent need of quantifying its potential effects. Beyond the fact that the two new obtained formulas can be directly used, the article also shows the potential of a new methodology (genetic algorithms) to capture the trends and provide accurate formulas for wave setup. This is very interesting since it can be applied to obtain better empirical formulas as soon as new better-quality data is available. The article is well written and their approach and results are of great interest for the coastal research community.

My overall impression is very positive but I still have a few comments and suggestions that might improve the manuscript: there are a few extra analyses that I missed in this study (specific comment 1), some parts of their methodology and results should be described in more detail (specific comment 2) and some parts of the text could be synthesized (specific comment 3). Below there is a longer description of these comments, together with a list of possible technical corrections. Overall, I recommend publication after these minor issues have been considered by the authors.

We appreciate that you found our manuscript interesting and thank you for the positive analysis. Below we address each specific comment in detail.

1) Potential extra analysis (from more to less important) that could be added:

- To my opinion, it would be extremely interesting to apply the two new formulas together with the other 7 to a new data set not included in the training and test sets. This would allow quantifying if the new formulas maintain their better predictive capacity beyond the data that was used to train them.

This is a valuable suggestion, thank you. Unfortunately, it is still challenging to find freely available data that include not only wave characteristics and beach morphology but also measured wave setup. Moreover, it was beyond the scope of this work to produce new data. Hopefully, soon more data will become available. As suggested, we now recommend in the Discussion that future research should test the formulas using an entirely new data set not included in the training and test sets. See below.

"An avenue for future research includes the validation of the GP predictors (Eqs. (10) and (11)) by applying them to datasets not included in the training. This will further assess the predictive capability of the new formulas and the importance of each term in the equations."

- If available, it might be enlightening to show beach profiles corresponding to the 9 data sets used in this study. Maybe the sets that show less correlation (Duck 94 and SandyDuck) have a specificity. In particular, I would say that Duck beach has a mixture of fine and coarse sand and it often displays a double slope profile, with a larger slope at the foreshore corresponding to a coarser portion and a smaller slope in the rest of the profile linked to the finer portion.

Thank you for the suggestion. However, we do not have access to the cross-shore transects but only to the mean foreshore slope. Stockdon et al. (2006)'s paper presents the profile of each beach, but due to Copyright reasons, we are not reproducing it in this paper. Based on the available information, it is not clear what is the specificity that differentiates Duck 94 (and SandyDuck) from the others. A possible explanation is now added to the Discussion section as follows.

"Martins et al. (2022) suggest that it might be difficult to differentiate between swash and wave motions near the shoreline in the field, particularly for steeper foreshores. The considerable influence of the swash circulation within a cusp field described by Stockdon et al. (2006) during the Duck 94 experiment could be an explanation for the lowest correlation of the measured data with the GP predictors results. In essence, accounting for all the effects and processes that may be important for wave setup remains an arduous challenge."

Added reference: Martins, K., Bertin, X., Mengual, B., Pezerat, M., Lavaud, L., Guérin, T., and Zhang, Y. J.: Wave-induced mean currents and setup over barred and steep sandy beaches, Ocean Modelling, 179, 102 110, https://doi.org/10.1016/j.ocemod.2022.102110, 2022.

- A variable that is important for nearshore processes that do not appear in the study is wave direction with respect to the shore normal. Do the authors have the wave direction corresponding to the different points in their data set? At least, they could plot the mean direction during the experiments and discuss if the sets that show less/more correlation show a specific direction.

Unfortunately, we do not have wave direction data. However, thank you for the valuable suggestion. We incorporated this comment in the Discussion section as a suggestion for future works. See below.

"More importantly, the novel inclusion of D_{50} as a second-order effect may indicate that we still have very limited information to describe an entire beach (e.g., not considering the presence of multiple bar systems, or even accounting for wave direction)."

- I am curious to know what would happen if Eq. (14) is applied using only the two first terms (so without the term related with grain size). Maybe this is out of the scope of this study but what would happen if this is plotted in Figure 5, too? That is, I wonder how sensitive is equation (14) to the last term? The text related to Figure 5 about the role of D50 (lines 240-244) might be expanded by including such analysis, if the authors find this interesting.

Thank you for the suggestion. We ran the test, and the results are presented in Figures a and b below. Without the last term, the equation does not represent well the measured data, underpredicting wave setup. Looking at the metrics, the coefficient of determination (R^2) essentially indicates that the equation predicts as well as blind guesses made around the average observed data. In this sense, this predictor needs the third term. We did not include this analysis in the paper because removing a term from the machine learning (ML) result changes the entire solution. The GP method does not evaluate the physical meaning of parameters but only how well the model predicts. It is our work to choose the best and physically meaningful equation, as pointed out in the first paragraph of the Discussion section.

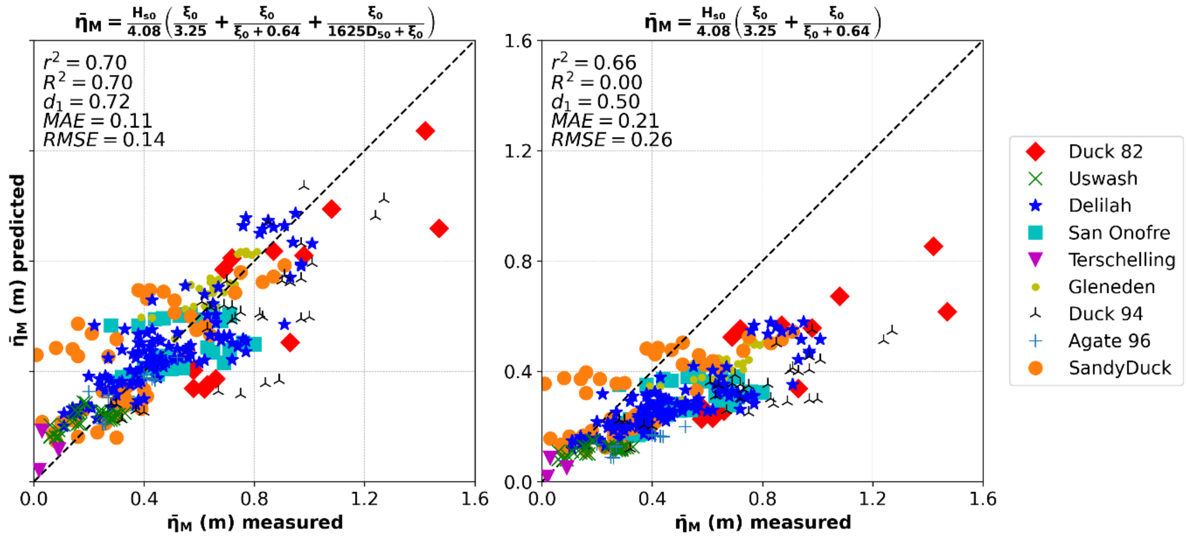

Figure a - Measured versus predicted maximum wave setup (ηM) using the testing data for Eq. (10) (former Eq. (13)) (left panel) and Eq. (11) (former Eq. (14)) (right panel). Different markers/colours refer to different field experiments, as referenced in the legend.

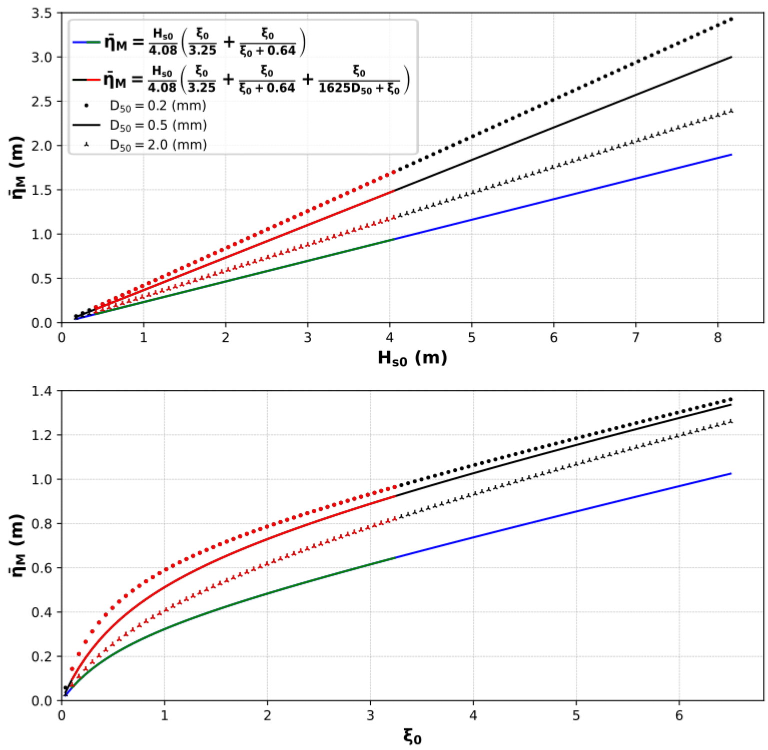

Figure b - General behaviour of the maximum wave setup (η_M) predictors presented and Eq. (11) (former Eq. (14)) with and without the third term, as a function of deep-water significant wave height (H_{s0}), Iribarren number (ξ_0) and median sediment diameter (D_{50}). D_{50} is represented by its minimum (0.2 mm), mean (0.5 mm) and maximum (2.0 mm) values in the dataset. Data within the measured range are depicted with red/green points. Black/blue points represent an extrapolated range for H_{s0} and ξ_0.

2) The following issues should be clarified in the text:

- Section 1: How are surf zones slopes and foreshore slopes defined? (e.g. which water depths?).

According to Stockdon et al. (2006), surf zone slope is the slope between the shoreline (cross-shore position of η_M), and the cross-shore location of wave breaking, and foreshore slope is the average foreshore slope with respect to the still water level ± twice the standard deviation of the continuous water level. We agree that the information used in our work should have been better described. We now clarified the foreshore slope in the Data section as follows.

"The dataset contains measurements of: maximum setup (η_M), foreshore beach slope (β_f) - average foreshore slope with respect to still water level ± twice the standard deviation of the continuous water level, and associated offshore wave characteristics (H_{s0} – significant wave height, and T_p – peak period) from 10 field experiments on sandy beaches resulting in a total of 491 measurements."

- Line 124: Iribarren number in the paper is computed with the foreshore slope, right? This is not what was used in the previous study where it first appear and thereby in the definition of line 58. Please,

Thank you for pointing out this confusion. The Irribaren number presented by Holman and Sallenger Jr. (1985) is also computed using the foreshore slope. This has been corrected and the following sentence in the text now becomes clear, as pointed out in the technical corrections (lines 58-60). See below.

"Later, Holman and Sallenger Jr (1985) found a more accurate correlation than the one presented by Guza and Thornton (1981) by relating the setup with the surf similarity parameter (Iribarren number: $\xi_0 = \beta_f /(H_{s0}/L_0)^{0.5}$, where β_f is the foreshore slope, H_{s0}/L_0 is the wave steepness and L_0 the offshore wavelength). However, when isolating low tide data, no significant trend was found with ξ_0, indicating the probable setup dependency on the entire surf zone's bathymetry and not only on the foreshore slope."

- Section 2: The authors should justify the choice of the variables in their GP model. The 6 used variables sound completely reasonable to me given the previous predictors published in the literature but I think the paper would benefit from a sentence arguing this.

The only freely available dataset containing setup, wave characteristics, and beach morphology parameters was from Stockdon et al. (2006). We included a sentence at the beginning of the Data section clarifying this (see below) and also indicated in the second paragraph of the Conclusion the need of additional variables.

"To make setup predictions through a data-driven model, it is necessary to have the input and output data to train it. The input data is related to physical processes that induce the output, wave setup. In this work, a dataset meeting these requirements, representing a large variety of beach and wave conditions compiled by Stockdon et al. (2006) has been used to develop a predictor of wave setup."

Moreover, why don't they also include beta_s (surf zone slope)?

Unfortunately, we do not have datasets containing surf zone slope and wave setup (as well as other variables). In the conclusion, we have mentioned the necessity of additional parameters and data to improve wave setup predictions.

Related to this, a comment about the introduction. Somewhere at the beginning of page 3, it would be nice to summarize what are the variables that have been used in the existing predictors (that are presented in the following paragraphs). The variables keep popping up each time a new predictor is *presented and I think that anticipating potential variables at the beginning of page 3 would improve the text.*

We thank you for the suggestion, but we have decided not to include this list of variables previously in the text. We believe this prevent duplication and it is easier for the reader to learn the meaning of each variable as they advance in the text other than needing to return to the beginning each time a new variable appears.

- Section 2: Also, the choice of operators in the GP model should be justified. In particular, what does it mean x^x? Why is this operator chosen instead of x^y or x^const?

Thank you for pointing out this confusion. We agree that x^x was not the best representation for a power function and x^y (with y being a constant) is in fact clearer. In any case, we have decided to delete it from the table and add it only in the text since it was not used to find the current predictors in the end, only in the search. See the corrected text below.

"We have run the model with different setups such as different mathematical operators (addition, subtraction, multiplication, division, square root, power, log, absolute, inverse, sine, cosine, tangent), population sizes (2,000 - 500,000), generations (20 - 10,000), tournament sizes (10 - 1,000), parsimony coefficients (0.0005 - 0.01) and genetic operations proportions."

Also, why the parameter values are limited from -5 to +5?

We chose this range of constants after testing values outside this range that resulted in poor predictions.

- Line 174: Clarify what is a parsimony coefficient and how does it work.

We have clarified the meaning and the use of the parsimony coefficient in the text as follows.

"A parsimony coefficient is used to penalize long equations, avoiding bloat (longer equations with no significant improvement in fitness). It is used during a tournament to deduct from the fitness result of the longer equation among two competitors that present identical results, being the longer one discarded."

- Section 3: How the GP model arrives at the factor 1625 in Eq. (14)? What is written in the Methods section is that constant range is from -5 to +5 (e.g. Table 2).

The original mathematical expression resulting from the GP model was the one below. However, we did simplify it to the one found in the paper (originally Eq. 14, now Eq. 11). We have clarified this information in the text.

$$\eta_M = \frac{H0}{4.08} * \left(\frac{\xi 0}{3.25} + \frac{\frac{\xi 0}{3.25}}{\frac{\xi 0}{3.25} + 0.197} + \frac{\frac{\xi 0}{3.25}}{\frac{D50}{0.002} + \frac{\xi 0}{3.25}} \right)$$

- Section 3: How robust is the result of the GP model? How the final formulas are chosen? Every time it is run, even with the same input parameter values of Table 2, it provides different formulas like Eq. (13) and (14)? Or it manages to always converge to these two selected ones?

Every time we run the model with the exact same setup, the result is the same. If the input parameters are changed, the algorithm usually converges to a different solution although at times the same form as Eqs. (10) and (11) (former Eqs. (13) and (14), respectively) can be obtained. We have clarified this information in the footnote of Table 2 (see below). Also, as indicated in the Methodology Section, it is the role of the operator to choose which formula is the most appropriate and this has been done on the basis of physical interpretation and robustness of the predictions beyond the training range.

"The choice of the values above was driven by extensive testing and sensitivity analyses performed as part of this work. The final model setup for each equation varies slightly. Details of both final codes are available and can be accessed through the Code and data availability section."

- In line 233, the authors write that Duck94 showed less correlation than the rest. However, watching in detail at Fig. 4, my impression is that SandyDuck also stands out, at least in Eq. (13) plot.

Although some SandyDuck data stand out in part of its data points, as seen in Figure 4, the remaining data is well represented by the new predictors. The same does not happen with Duck 94, since the new predictors underestimate almost all data points. When plotting both datasets only (and the metrics – Figure c), it is easy to visualize these differences. The metrics presented show that the SandyDuck dataset performs better than Duck 94. We have clarified this information in the text as follows.

"Among all field experiments, Duck 94 (intermediate to reflective with large wave conditions) was the beach that showed less correlation with our models when comparing individual metrics."

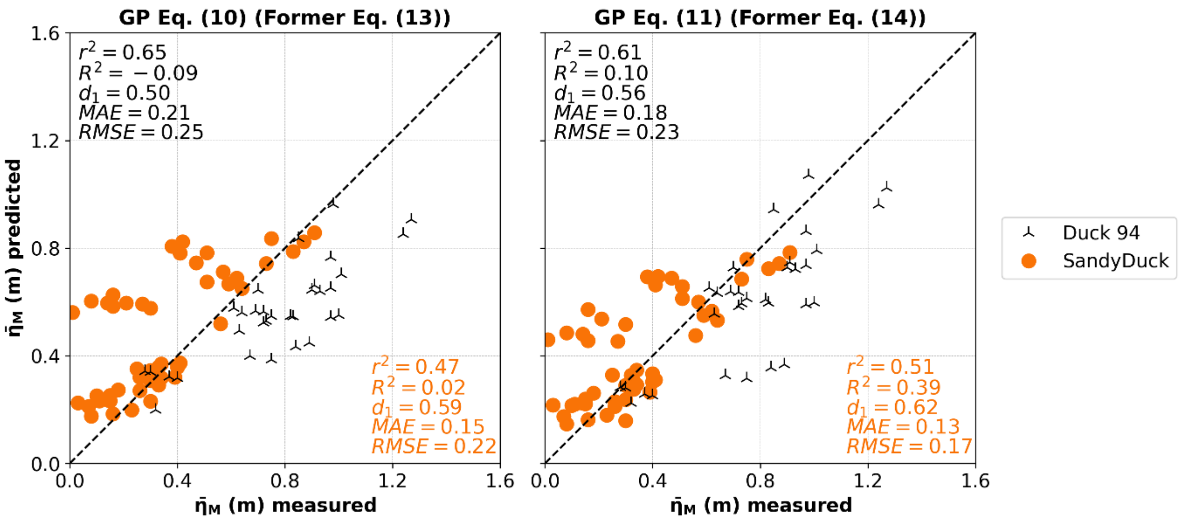

Figure c - Measured versus predicted maximum wave setup (η_M) using the testing data for Eq. (10) (former Eq. (13)) (left panel) and Eq. (11) (former Eq. (14)) (right panel). Different markers/colours refer to different field experiments, as referenced in the legend.

- Line 254: The authors mention that the best of the previous models (Ji et al., 2018) has the disadvantage of having one coefficient more than Eq. (13). What do they mean?

Thank you for pointing out this mistake. This statement is not correct. Both equations have the same number of variables and coefficients in the formula. We have deleted this sentence from the text.

- Line 270-271: Ji et al. (2018) formula also presents a good fit for dissipative and reflective conditions, right? This should be acknowledged.?

Yes, this is true. We have acknowledged this in the Results section as follows.

"Overall, our ML-driven approach achieved better results with Eq. (11) outperforming all other predictors. Similarly, Eq. (10) exhibits good results, the same ones as Ji et al. (2018)'s equation, which also performs well on dissipative and reflective beach conditions."

3) The text is in general well written and to the point but, to my opinion, the following parts could be synthesized:

- Delete lines 18-24 in the Intro. This contains standard knowledge, included in any book on nearshore processes, it is not needed in research articles, in my opinion.

We agree and have deleted those lines.

- Delete or synthesize lines 40-45 in the Intro. Is this adding something, given that most of the formulas below are purely empirical? At least, the formula could be deleted, maintaining only a summary of the main findings of Bowen et al. (1968). Alternatively, when introducing the work of Battjes (1974), the previous result of Bowen et al. (1968) could be acknowledged.

We agree and have synthesised those lines. One of the main findings of Bowen et al. (1968) was that the theory, based on the concept of radiation stress, underpredicts measured wave setup values. We believe this information is important, therefore we have maintained it.

- Lines 250-258 could be more to the point. For example, I think it is unnecessary to write all the numbers of Table 3 (they can be seen on the Table).

We agree with the suggestion and have deleted the metrics presented in the text.

Technical corrections

We thank you for the careful reading. All the suggestions in technical corrections have been accepted and corrected in the updated version of the paper. The only exception was the one related to figure 5 "*Figure 5: My impression is that showing the results of the two equations in the same panel would be more illustrative, since they could be more easily compared.*". We have tested it, but when both equations are plotted together, they mostly overlap, presenting a similar behaviour and being difficult to distinguish one from another. The idea of this figure is not to compare the equations but to show that both have coherent behaviour with the real environment, including using data that the model was not trained or tested on (i.e., extrapolated data).

Once again, we thank the reviewer very much for the time and careful consideration of our study. These have greatly improved the quality of our manuscript.

Kind regards,

The Authors.

References:

Battjes, J. A.: Computation of set-up, longshore currents, run-up and overtopping due to wind-generated waves, Ph.D. thesis, Delft University of Technology, http://resolver.tudelft.nl/uuid:e126e043-a858-4e58-b4c7-8a7bc5be1a44, 1974.

Bowen, A., Inman, D., and Simmons, V.: Wave ‘set-down’and set-up, Journal of Geophysical Research, 73, 2569–2577, https://doi.org/10.1029/JB073i008p02569, 1968.

Holman, R. A. and Sallenger Jr, A.: Setup and swash on a natural beach, Journal of Geophysical Research: Oceans, 90, 945–953, https:380//doi.org/10.1029/JC090iC01p00945, 1985.

Ji, C., Zhang, Q., and Wu, Y.: An empirical formula for maximum wave setup based on a coupled wave-current model, Ocean Engineering, 147, 215–226, https://doi.org/10.1016/j.oceaneng.2017.10.021, 2018.

Martins, K., Bertin, X., Mengual, B., Pezerat, M., Lavaud, L., Guérin, T., and Zhang, Y. J.: Wave-induced mean currents and setup over barred and steep sandy beaches, Ocean Modelling, 179, 102 110, https://doi.org/10.1016/j.ocemod.2022.102110, 2022.

Stockdon, H. F., Holman, R. A., Howd, P. A., and Sallenger Jr, A. H.: Empirical parameterization of setup, swash, and runup, Coastal Engineering, 53, 573–588, https://doi.org/10.1016/j.coastaleng.2005.12.005, 2006.