The manuscript by Athanasiou et al. proposes a surrogate model of a time-consuming numerical model (XBeach) for the prediction of coastal erosion under extreme conditions at the regional scale (the Netherlands' coast). The main benefit of the approach is speeding up the prediction process without significantly decreasing the accuracy of erosion estimates. This is particularly useful in the application of early warning systems. Similar methodologies have been proposed before, but the novelty here is the application on a regional scale and the inclusion of the beach profile as an input. The latter makes the model capable of reproducing coastal erosion for a wide range of typological coast profiles, enabling accurate and fast predictions regardless of the initial state of the beach. The framework is presented as a flexible and transferable tool, so similar applications to other processes should be viable.

The manuscript is well presented, it reads well, and figures are of scientific rigor. The proposed methodology is thoroughly explained and easy to interpret, facilitating its future application to similar issues. There are some instances where the quality of the manuscript would benefit from clarifications. For example, a few (but important) details of the statistical model used here were not explained in detail. Correcting this will not require great effort, hence I suggest accepting the manuscript after minor revisions.

Please find below a more detailed review, providing line numbers where some corrections or clarifications might help improve the quality of manuscript. I hope you find my suggestions useful.

Kind regards.

Line 72. I would expect a sentence here highlighting the novelty and relevance of the proposed model as compared to the ones mentioned just before.

Line 125. What was the criteria for determining highly dynamic areas?

Line 142. There could be Hs extreme events coinciding with a non-extreme SSL (and vice versa). What if, for instance, SSL (even though it does not exceed a threshold) is high enough to cause erosion together with an extreme Hs. Wouldn't it be more appropriate to include extreme Hs AND/OR extreme SSL?

Line 145. Remove (2016)

Line 172. I wonder if a GPD is appropriate to describe the marginal of D. D is not defined based on threshold exceedances here, and although a GPD can be used to describe other extremes (such as annual maxima), I am not convinced this variable may have a heavy-tail behaviour.

Line 176. A Gaussian copula is mentioned here, so I assume it is a Multivariate Gaussian copula what is used here to model the dependence structure between the four hydrodynamic variables. This copula is chosen based on earlier work by Li et al. (2014b) as it was deemed suitable at the Ijmuiden-06 station. This may seem a reasonable choice, as long as the data used here does not differ substantially from Li et al. (2014b). But I wonder if the

Multivariate Gaussian copula would also be appropriate for the other stations. I can see a validation at the Euro platform station (Figure 3), but not for the other stations.

Line 180. 100,000 synthetic events are sampled based on Monte-Carlo. How does this number translate into length of data? What is the rate of events per year?

Line 185. Related to the previous comment in line 176, it is stated here that dependency statistics were smaller than 5%, but I have the feeling this only refers to the data presented in Figure 3. Could you also report metrics for the other stations?

Figure 3. The resolution of this figure should be increased. Also, how do the copula-based most extreme simulations presented here compare with estimates of most extreme historic events not included in the copula analysis as observations, such as the 1953 one? This could also give an indication of how realistic the most extreme synthetic events are, especially for simulations far more extreme than the observed ones. Perhaps a return period/value comparison between extreme historical events not included as observations (e.g., 1953) and synthetic events from the copula analysis would be insightful and reinforce your message about the suitability of the statistical model.

Line 275. You may mention this later, but how was this division done? 50/50? Was this a k-fold validation? How did you determine a suitable calibration/dataset division to ensure that was a good way of selecting the architecture of the ANNs?

Figure 10. Encouraging that the model performs best for the most complete pre- and post-storm profiles available (2019), but it is also true that this event was not particularly erosive. Would the model perform as well for more erosive events if we had complete pre-post/storm profiles as in 2019? I seem to remember the winter season of 2013-14 was particularly extreme (waves and surge) in northern Europe (this especially applied to the UK, but I imagine the Netherlands was also impacted by these series of storms). Are there records of complete pre- post-storm transects for that particularly extreme winter? It could be more insightful to show how the model performs for more erosive events while being validated with complete transects (if they are available).

Figure 12. This is a nice and interesting figure. I wonder what's the effect of altering the number of TCPs included in the training process.

Line 552. There could be also problems in tropical-storm prone areas. Especially the fitting of marginal distributions and the copula approach, given the rarity of those events.