Notes to authors:

Dear Authors

I think your manuscript has been greatly improved, and it is clear you have ingested the paper of Thompson et al., (2019) well, and this focusing of the results onto sediment sizes for which you have measurements of their mobility has clarified the manuscript. I also appreciate breaking down your predictive ability for sediment motion into individual classes as well as the entire suite of grain sizes you have data for – the distinction of variability within and between sediment classes is well taken. I still think you need to switch to using a threshold for suspension, rather than initiation of motion, because at present (and particularly for your coarser sediments) the thresholds are quite different for initiation of motion and suspension. Hopefully you can include both the initiation of motion and suspension in your analysis, and consequently a (hopefully!) improved predictive ability when the threshold for suspension is used. I also have some concerns about how well the storm in November 2015 is modelled – primarily as your validation year doesn't appear to have a storm of such magnitude. You need to demonstrate that the model is operating in a physically realistic way under these higher-energy conditions (perhaps by comparing offshore forcings between the validation year and 2015).

Lastly, (and not for this manuscript!) I hope you will be able to expand on this "natural laboratory" with further field measurements and modelling, particularly in incorporating the grain size distributions, and potentially EPS measurements to (hopefully) further improve the correlation between predicted and observed threshold of motion – information of that kind is vital for understanding what is needed to accurately model the threshold of motion in real environments.

Abstract.

Please differentiate between critical threshold of motion, and suspension. These are two separate parameters.

The final sentence: "Based on the numerical model data, the critical shear stresses from the newly proposed model, τcr (d50, ρB , ChlA(t)), were rarely exceeded based on only wave-induced motions in most of the model grid, but could, nonetheless, be exceeded to up around 10% of the times in smaller areas.." needs reworking, it is a confused sentence.

- "were rarely exceeded based on only wave-induced motions in most of the model grid" this is one result, and valid on its' own (with added quantification).

- "but could, nonetheless, be exceeded to up around 10% of the times in smaller areas..", what are these smaller areas? What's different about them relative top the rest of the model? Is this result expected or unexpected? What interesting thing do we learn from this?

Introduction.

The point about models typically using the median grain size to define a threshold of motion as a limitation to many models is well taken, but perhaps laboured. Plenty of models use multiple grain sizes, or distributions, if the detail is available.

You are still not differentiating between threshold of motion and suspension, these are well known to be different.

Methods

It doesn't appear that there's any infragravity waves (20-300 second periods), but these have been found to be important for near shore sediment transport in the Baltic sea (see referances in Weisse *et al.*, 2021). Can you explain why your site doesn't appear to have any? Long period waves are often crucial for sediment transport in coastal environments.

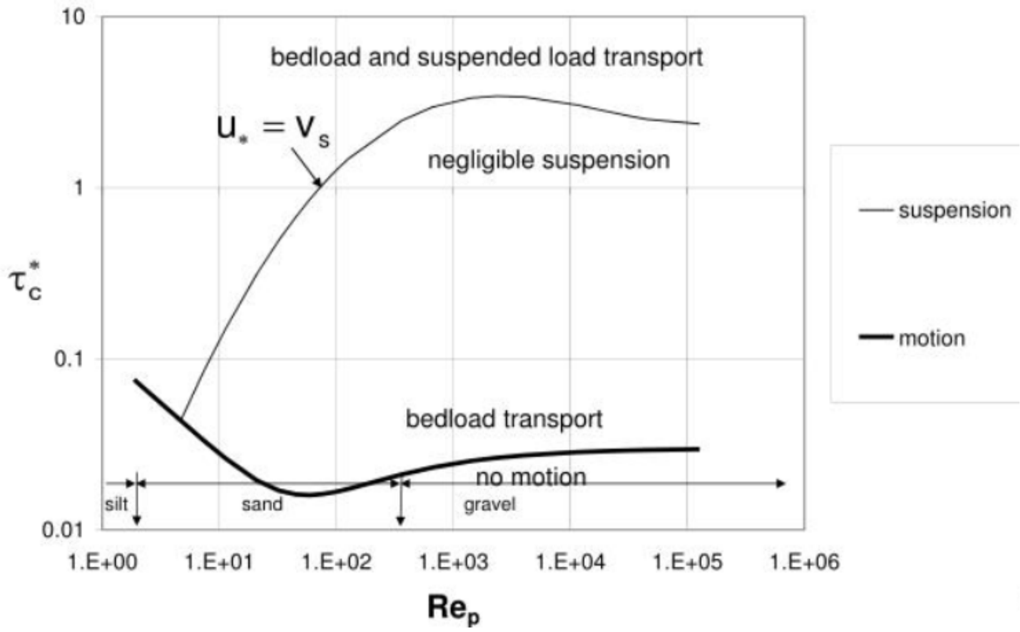
Figure 3 is concerning. The wave period from the SWAN model and wave buoy does not compare well, and suggests that wave height is poorly represented. Wave heights need to be shown in figure 3 as well. I would also include wave energy to help the reader see why any periods of poor correlation can be discounted (i.e. Low energy, not important).

There is also a concern that the period of model validation does not appear to be used in the rest of the paper. The 2017 validation period is unused in section 5, due to the lack of wave data for validation during the 2014 and 2015 measurement campaign. As such Figures 6-7 (modelling period undefined) and Figures 8-9 (a storm event on 28 November 2015) are presumably all from the same 2015 model run.

I have two issues here:

1) Results in section 5 – which model runs, and which time periods are shown need to more clearly labelled for each figure.
2) All things being equal, the different model validation period to those shown in the results section *should* be fine. My concern however is that the range of bottom orbital velocities shown in the November 2015 storms is far higher than any during the validation period; 0.8 m/s (Figure 8) in 2015 vs a maximum of 0.1 m/s in 2017 (Figure 3). Given the absence of observations for this period or set of conditions, additional justification is needed to demonstrate that the model is operating in a physically realistic way under these higher-energy conditions. A comparison of offshore wave climate between years and a discussion of whether the validated conditions adequately represent the extremes simulated would be useful.

I think you have to switch to using theta_sus over the current threshold of motion. These two thresholds are not the same especially for gravels and sands – see figure below.



At the moment your use of the Soulsby-Whitehouse equation is giving you the initiation of motion, not suspension. You need to also use the van rijn (or an equivalent equation) for the threshold of suspension: https://www.leovanrijn-sediment.com/papers/Threshsandmud2020.pdf

$\theta^*_{sus} = \frac{0.30}{1+D^*} + 0.1\left(1 - \exp^{-0.05D^*}\right)$,

You could add this threshold into your current work, and see how much more of the variability in your results is explained by using a threshold for suspension over a threshold for motion.

Figure 5. There seems to be far greater variance for EROMES data than for the predicted (Width of scatter far higher). Can you suggest why this is?

Lines 280 to 292 – needs references

Lines 293- "the three parameter linear model could explain most of the variation". Please give us some numbers!

Line 297 onwards: "and we surmise that a single variable (ChlA) might not be enough the capture the biological effects well enough to sufficiently model the variations in critical stress on muddy seabeds". You could certainly expand here, what about EPS?

Line 306- "1) Our in situ data could not provide representative values for all of the classes."

-A Value of....what exactly

Line 317- "as horizontal wave motion does not attenuate significantly when the wavelength is at least 20 times the water depth."

-Reference needed

Thompson, C.E.L. *et al.* (2019) 'Benthic controls of resuspension in UK shelf seas: Implications for resuspension frequency', *Continental Shelf Research*, 185, pp. 3–15. Available at: https://doi.org/10.1016/j.csr.2017.12.005.

Weisse, R. *et al.* (2021) 'Sea level dynamics and coastal erosion in the Baltic Sea region', *Earth System Dynamics*, 12(3), pp. 871–898. Available at: https://doi.org/10.5194/esd-12-871-2021.