

Review of “Predictive mapping of organic carbon stocks and accumulation rates in surficial sediments of the Canadian continental margin” by Epstein et al.

General comments (section evaluating overall quality of the preprint)

Overall, this manuscript is in great condition, and I recommend publication. There are a couple of things about this manuscript I do not agree with but in all, I would say minor edits are needed. This publication builds off previous efforts to predict seafloor properties including organic carbon content, mass accumulation rate, and mud content. These results are then incorporated into basic calculations to estimate density and accumulation rates. This manuscript focuses primarily the Canadian EEZ and the final results are relevant to a wide range of interests. The methods this paper uses are sound and robust however there are several issues I have which I outline briefly here.

Increasing spatial variability does not decrease uncertainty. There are several times the author makes this point. As an example, you can assume the median and/or mean of an observational dataset and arrive at a residual value (obs-pred) similar to that of the prediction. Further, statements made regarding the term “in-situ” data should be removed. None of this data was truly *in-situ* data as this data was pulled shipboard and processed ex-situ (i.e., on-deck). Overall, the text is very well written. However, the length and level of detail in the manuscript is at times very overwhelming (e.g., methods). The manuscript could benefit from being shortened to add to clarity for the reader. Certain details presented in the manuscript are more appropriate for the supplemental material (e.g., modules used). A major issue I would like to see addressed in a revision is the MAR prediction. It does not make sense to me the way this observed data was generated (more details in the specific comments below). Further, more specific comments about the manuscript I would like to see addressed below.

#### Specific comments

Line 29: Consider rearranging this sentence for clarity to say density and accumulation for mud content, sediment dry bulk density, and organic carbon content. Otherwise it reads differently, I at first questioned the difference between organic carbon content, organic carbon density, and organic carbon accumulation.

Line 38: See general comments regarding the term of *in-situ*

Figure 1. Why the red and grey parts? After reading the manuscript I did not have a full understanding of what this is.

Quite a few predictors are up to several orders of magnitude upsampled to be at the resolution of prediction. By upsampling you are not adding new information (variance) therefore your prediction will not be any more accurate and/or uncertain. Consider this in relation to general comments regarding spatial uncertainty and variability.

Line 197: Why do annulus and square windows? And why these different annulus widths? Are these selections related to processes or arbitrary? Do you find that the square windows cause artifacts on the final predictions?

Lines 201-203: You do this to several sets of grids. How different are these grids after this process? This is a lot of manipulation; how do you think this affects the final prediction? Particularly for those predictions that place emphasis on these as predictor grids?

Line 207: It would be beneficial to use same terms (shelteredness, exposition) in the text as in the Table 1.

Line 296: What is the lowest vertical cell? How often was this close enough to the sea bottom? Enough to be accurate to constrain what the value at the seafloor would actually be?

General note about methods/predictor grid processing: The paper would benefit from figure(s) outlining what was done perhaps within the supplemental material. In some cases, there was similar processing for various sets of grids so reusing the same flow diagram and referencing it might provide clarity. The level of detail given throughout the manuscript is great, however it is cumbersome to understand at times.

Line 273: Why three different ocean circulation models? State explicitly. Additionally, when you put all these grids together do you see artifacts?

General comment about grids for Section 2.3.4: After performing all this processing, are these grids any better than the global circulation model estimates? I understand that the grids are better resolution when taken on time slice or over particular areas (you mention explicitly nearshore areas) however, is some of that fidelity for other regions stripped by the processing techniques?

Section 2.4: This is grain size not composition.

Line 434: What about Hayes et al., 2021 (<https://doi.org/10.1029/2020GB006769>) and the CASCADES Martens et al., 2021 (<https://doi.org/10.5194/essd-13-2561-2021>) dataset?

Section 2.5.2: I am confused some by this section. Why the upper 30 cm? Why not select the upper  $n$  cm that are controlled by the data depth distribution? 30 cm can account for long periods of geologic time in some cases (lower sed rates). Further, by approximating a mean decay function are you starting to incorporate effects of degradation? What uncertainty are you introducing by this entire processing? This seems a lot of unnecessary data processing for observed data.

Section 2.7: I do not agree with using this data as observed data. Why not just upsample the prediction? The observed data is not real and thus should not be used in this manner. No prediction even if the predictors are fantastic will give you the correct result if the observed data is not solid. Further, why sample it spatially randomly and not at the same locations you have observed data for OC? I think you are arbitrarily making your predictions better by randomly sampling as you are covering the feature (predictor) space more uniformly. We find the same phenomenon occurs when performing this on synthetic data. The synthetic data will always outperform as it covers the feature space more uniformly and predictors are inherently tied to spatial phenomenon. I would suggest if you (although I do not agree with it) are to use this MAR data like this then you should sample it spatial and with as many samples as other datasets (e.g., OC) that have been sampled. This will represent the natural (but unfortunate) bias that occurs in marine sediment sampling.

Line 530: This is not completely spatial explicit if you are adding  $x$  and  $y$  coordinates as predictor variables?

Line 607: Why change the  $mtry$  and  $min\_n$  number and not the number of trees?

Line 626: How did you merge these datasets? Were there edge effects?

What about using confidence intervals on non- normally distributed data?

Section 2.9: Why are you calculating dry bulk density in several different ways? Why are you calculating porosity with three different equations? This section lacks clarity and I left feeling confused about what was done.

Did you do anything to avoid collinearity between the predictor grids that were selected?

Figure 4 caption: Indicate more clearly which parts of these is ALE (distros drawn by line) and which is a rug plot (bottom)

On what grounds (citations?) were certain predictors selected as *a priori* for input to the various predictions? (e.g., Fig 3)

Line 788: is this a typo? RMSE of 0.09 arcsin{%OC} also what about line 808 "RMSE of 0.206 log<sub>10</sub>{g cm<sup>-2</sup>yr<sup>-1</sup>}" Why say it like this?