Earth Syst. Sci. Data Discuss., referee comment RC2
https://doi.org/10.5194/essd-2021-23-RC2, 2021

Comment on essd-2021-23

Anonymous Referee #2

Referee comment on "Modelling seabed sediment physical properties and organic matter content in the Firth of Clyde" by Matthew C. Pace et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-23-RC2, 2021

Overview

The authors provide a machine-learning based approach to estimate surficial sediment properties in the Firth of Clyde.

These types of systems are key loci for OC burial, and their spatial heterogeneity and lack of data availability impedes a better understanding of these systems. Therefore, this approach is very welcome.

Broadly, the article itself is appropriate for the publication of the dataset. Furthermore, data set provided via the doi in the abstract is almost complete, only lacking the Rmd file (see comments below). Considering the article and dataset: as detailed in the comments below, it needs to clarified if the RF was trained on open-shelf datasets (i.e. the doi Seabed_properties.csv). Because the mineralogy between the open-shelf and the firth of Clyde are rather different, the model appears not to work optimally for clay-rich sediments. This needs to be addressed as a key model limitation.

The presentation quality and language are excellent, there are only minor suggestions w.r.t. improvements of Figures.

There are some key elements which need to be improved upon, after which I think this paper would be suitable for publication in ESSD. Once published, it would contribute to the scientific community as well as ecosystem mapping services and nature conservation efforts.

Firstly, concerning the Random Forest approach:

- **Random Forests are well-known for overfitting data**. Meaning that the model may well work the dataset they were trained on, but not for other areas. Presently, this is not sufficiently addressed in the manuscript (e.g. in the discussion). One of the key elements for publication in ESSD is that results are scalable and applicable in other systems. The authors primarily cite other studies that also use RF and use this as grounds that this is approach is appropriate. For a robust statistical argument however, one would run another model to compare it to the RF one. If you can prove RF gives better predictions, this would build your case.

To support the statements concerning varying degrees of variability, it would be helpful to see semi variograms (in the SI).

- **Spatial-cross validation:**

As this data is indeed spatially not independent, it is appropriate that the authors did spatial cross-validation. However, what is not clear is how the assumption of blocks of $1/8^{th}$ by 1/4 th of a degree (please also translate this to km in the method section) holds. Now, the only support for this assumption is that it was also used in another study (Wilson et al 2018).

Why was this chosen? Why is this value optimal, considering it is constant and variability is higher closer to land? How many "blocks" does this generate for your region? Is the sample density homogeneous enough to ensure enough datapoints per block? How do the

results change when a different size is chosen? How did you determine this size was optimal?

Why is better than a simple alternative such as k-means clustering? This needs to be answered.

- **Presently authors are not (yet) in accordance with the ESSD code publishing policy.** Authors can make an Rmd file with the key steps of their R code modelling process using the ESSD template. This way they would ensure repeatability and increase the chances of scalability in case other users want to apply a similar approach to their regions. At present, as a reviewer, readers have no sight on (i) relative importance explanatory factors and (ii) effectiveness of spatial cross-validation blocks (iii) how missing data was dealt with (did the authors use proximity weights?)

This should be included.

- **Applicability of the model:**

The evaluation of the RF-based prediction shows that there are some issues. Again, RFs are known for overfitting, but the issues need to be addressed in the discussion section of the paper. Crucially, as the authors describe and show (e.g. Figure 2), that significant chunks of Clyde sediments are nearly 100% clay. Clay in turn, is seen as a key predictor of OC content. Additionally, permeability predictions are generated from the mud content percentage (Figure 12). So, an accurate prediction of minerology in clay-rich and clay-only zones are important. In Figure 7 it is shown that for samples which are completely mud (no sand) the model is not predicting well at all. Similarly, in Figure 8, fine grained, i.e. muddy, samples (D50 < ~0.02 mm) are not predicted accurately. It is pertinent that this is addressed. It is not entirely clear yet if the mineralogy models were based on the North Sea open-shelf datasets (see detailed points, Seabed_properties.csv). If they are, this could explain why the predictions are poor for the Clyde. For gravel the results are so very poor (r2 = 0.08) that the model does not seem appropriate. It could be useful to have a map (in the SI) which shows the residuals between the measured and modelled values on a map, not just the x-y plot.

- **The timeline for data collection, and conclusions built thereon:**

As a reviewer I appreciate this type of data is sparse, and that it was an effort to collect the existing data.

However, much of the surficial sediment data - which is key for the conclusions – has been collected asynchronously (Table 1). For example, from the RF, clay content is a key predictor for OC content. This is makes sense considering organo-mineral interactions. However (Table 1) the clay content was measured between 1969-1980 whilst the OC and ON was determined 2005-2006. The authors conclude that e.g. trawling has not been adversely impacting OC content, but is that reasonable if the data that is built has been taken 25-35 years apart? I image fishing methods and intensity have changed in that time period. In the McIntyre (2012) survey cited, it is mentioned that trawling started in the 1960's and peaked in the 1980, altering the ecosystem and benthic foodwebs. So trawling peaked after the data on mineralogy, on which many conclusions are built, was collected. Does that impact the discussion e.g. in lines 424 on changes since the 1980?

- **Minor comments:**

Line 80: to further support this statement include this authors more recent paper (Luisetti et al., 2020) and the paper highlighting the economic value by (Avelar et al., 2017)

Introduction, support for RF: as mentioned above, you need to provide a reason other than that is has been used by others to prove it is a good/the best approach.

Wilson et al 2018 reference: in the reference list there is only a Wilson et al., 2017 – are

they the same one?

Figure 2: Gravel: in the paper it is said there is not a lot of gravel data. Yet the map, most data is marked at 0. Is it a true 0 or is it N/A? Considering the other data, N/A seems likely. If so, it's the latter, data should be gray, the current color scheme is misleading.

Section 2.2.3: Median Grain Size. this section needs some more detail. There is the identically named section 2.4.3. I suspect some missing information is there. Please make one clear section to avoid confusion. Which variables are calculated on the bases of which other, and which geographic zone? It is not completely clear now. Are you assuming that what works for the open shelf also applies for the Firth and the lochs? If so, please clarify and explain.

Section 2.2.4 idem: it is not clear if the authors assume the relationships that hold for the north sea in general can be translated without correction to the Firth (which must have different flow and fishing exposure). Section needs additional details to clarify this.

Line 255 Provide more details (if necessary, in SI) w.r.t. FVCOM: how were the variables mapped to the grid? Did you assume a weighed gaussian around each measurement, or did you just map the value to the nearest point? Or are the grid nodes identical to the measurements? Any assumption must be clarified and explained. Also, w.r.t. the # of grid points, is there potentially a typo? In Figure 8 it says n = 3244 samples, in the text (line 254) it says 39449 grid nodes. That's an order of magnitude higher. Also include these details in the Rmd.

Lines 255-259: also in accordance with the ESSD guidelines: provide in Rmd

Line 266: provide details to bootstrapping (frequency)

Line 312: regarding trawling there has been the recent paper by (Sala et al., 2021) which could also be good to cite

Line 314: RFE. Provide more details as to which variations of RFE were used and why. This may be done automatically when you provide the Rmd.

Line 325: provide reference for density assumption

Line 333: provide a bit more detail regarding the AUC, what are the reference values (0-1) and can this thus seen as a robust model? Cite the package used

Figure 7: Prediction accuracy for the mineralogy: predictions are rather poor for clay-rich sites, and some details are missing. Is the data shown in Figure 7 the out-of-bag dataset of the RF?

Figure 8 please add log scale increments with ggplot

Lines 343: authors state results may be applied to other settings, but lack proof of this. Statements on mud-rich sediments such also be nuanced in accordance with model robustness results.

Lines 449: as mentioned, are the inferences on trawling consistent with the various timepoints of data collection and the exploitation as described by Mcintyre (2012)? If yes, please clarify. If not, adjust. Stating that more research is needed to answer this

Section 4.1: connect limitations to RF model performance.

Line 474: doi does not work. The doi in the abstract does work.

- References:

Avelar, S., van der Voort, T. S. and Eglinton, T. I.: Relevance of carbon stocks of marine

sediments for national greenhouse gas inventories of maritime nations, Carbon Balance Manag., 12(1), doi:10.1186/s13021-017-0077-x, 2017.

Luisetti, T., Ferrini, S., Grilli, G., Jickells, T. D., Kennedy, H., Kröger, S., Lorenzoni, I., Milligan, B., van der Molen, J., Parker, R., Pryce, T., Turner, R. K. and Tyllianakis, E.: Climate action requires new accounting guidance and governance frameworks to manage carbon in shelf seas, Nat. Commun., 11(1), 1–10, doi:10.1038/s41467-020-18242-w, 2020.

Sala, E., Mayorga, J., Bradley, D., Cabral, R. B., Atwood, T. B., Auber, A., Cheung, W., Costello, C., Ferretti, F., Friedlander, A. M., Gaines, S. D., Garilao, C., Goodell, W., Halpern, B. S., Hinson, A., Kaschner, K., Kesner-Reyes, K., Leprieur, F., McGowan, J., Morgan, L. E., Mouillot, D., Palacios-Abrantes, J., Possingham, H. P., Rechberger, K. D., Worm, B. and Lubchenco, J.: Protecting the global ocean for biodiversity, food and climate, Nature, (December 2019), doi:10.1038/s41586-021-03371-z, 2021.