

Earth Syst. Sci. Data Discuss., author comment AC1
<https://doi.org/10.5194/essd-2021-23-AC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-23**

Matthew C. Pace et al.

Author 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-AC1>, 2021

On behalf of all my co-authors on this paper, I thank the two anonymous referees for their insightful and thorough comments. I also thank Craig Smeaton for his community comment that helps contextualises our work within the landscape of ongoing seabed mapping efforts.

The referees' comments touched on several important related points, and I have therefore collated these comments and described how they will be addressed within the revised manuscript.

Anonymous Reviewer 1

AR1 - *Explanatory environmental variables were used to predict rock presence, sediment composition, particulate organic carbon, and particulate organic nitrogen. While these have been submitted to a formal variable selection process as recommended in the literature, it would be good to know why certain predictor variables were chosen in the first place. Usually, such selections are based on a general understanding of the modelled system, experience from previous studies, and data availability. It would be beneficial to briefly outline, which predictor variables were initially chosen and why.*

MP - The selection of initial explanatory variables was largely based on a general understanding of the geomorphology of the Firth of Clyde and other coastal marine systems, as well as the availability of full coverage data to facilitate the generation of high-resolution quantitative maps. Some text outlining the selection of these explanatory variables has been added to section 2.4 with an explanation of additional predictors for organic carbon and nitrogen given in section 2.4.6.

AR1 - *To minimise the impact of spatial autocorrelation on performance estimates, a spatial cross-validation was run with data binned in blocks of 0.125° latitude by 0.25° longitude. It is encouraging to see that spatial autocorrelation is increasingly accounted for in marine modelling studies; however, it would be necessary to explain why the above-mentioned block size was chosen. Usually, the block size should be determined experimentally, e.g., by estimating the spatial autocorrelation range from an empirical variogram. The *r* package *blockCV* (Valavi et al., 2019) provides a tool to determine block sizes and might be used here to establish a suitable separation distance.*

MP - The block size was selected to balance the need for a suitably large number of spatial blocks and an adequate number of data points within each block. The internal sampling carried out by the Random Forests algorithm makes *a priori* assessment of spatial autocorrelation challenging (see discussion on cross-validation in Diesing et al., 2017). Hence, we show that the chosen size was suitable for each of the models by generating an empirical semi-variogram for residuals calculated for the full model (see Supplementary Material). It is clear that the selected block size exceeds the range of the semivariogram, following Roberts et al. (2017), and that, for some models, Random Forests internal bootstrapping in the full model sufficiently removes spatial dependencies within the data.

AR1 - *There is very limited information on the random forest models that were built apart from general information. It would be desirable to include basic information such as the (hyper-)parameters that were chosen and how/why. If the authors feel this would unnecessarily increase the length of the manuscript, the information could be provided as a supplement. Even better would be the provision of the R code, which will increase openness and transparency. Additionally, in the case of the rock model, it would be necessary to explain which threshold value was chosen to convert probabilities into presence/absence predictions. As no information was given, I suspect the "default" of 0.5 was selected. It has, however, been demonstrated that this threshold value is not always optimal and a range of alternative threshold criteria exist (Freeman and Moisen, 2008a, 2008b). I suggest exploring alternative thresholds, assuming the default has been used.*

MP - We fully agree with the recommendation to provide the R code used in this work to document the methods used and increase transparency and repeatability. Accordingly, we provide a compiled R Markdown report as Supplementary Material to the revised manuscript. This contains the code used to fit statistical models to data, code to generate predictive seabed maps and additional detail of model validation methods, diagnostics, and outputs. In this way, the markdown document is a means of addressing methodological queries without increasing the length of the manuscript.

Random Forests hyperparameters were tuned to optimise model performance. Increasing the number of trees 'grown' provided more stable model estimates, whereas modulating the number of variables sampled at each node had limited effect on model outputs. Due to the considerable length of the supplementary material, this investigation is not shown in the Markdown report. Nevertheless, we use the tuned number of trees in the models presented in the R Markdown report.

The threshold parameter is not readily tuneable for Random Forests models built using classification trees in the *ranger* package. However, we do not anticipate that alternative thresholds will translate to marked improvements in seabed rock map quality due to limitations in the availability of empirical data.

AR1 - *Model performance is reported as area under the curve for rock, r-squared for sediment composition, particulate organic carbon, and particulate organic nitrogen and the explained variance in the case of the median grain-size and permeability. Could the authors explain whether r-squared and explained variance have different definitions here? Additionally, it would be necessary to provide more detail on why certain performance indicators were chosen and others not. For example, in the case of the rock model, which is a binary, presence-absence model, different performance metrics could be used to estimate model calibration and discrimination for continuous and binary outputs (Lawson et al., 2014). The area under the curve measures discrimination of the continuous*

predictions, but when looking at binary predictions, sensitivity and specificity might be chosen. Calibration might be estimated with the root mean squared error in the case of continuous predictions and mean accuracy in the case of binary predictions.

MP - Within the revised manuscript, R-squared and explained variance have identical definitions, following the “traditional” R² formulation used within the caret package: . This is amended from the original manuscript, where some models were assessed using the “corr” formulation used within the caret package: the squared correlation between the observed and predicted response variable. This latter formulation was less robust (Kvalseth, 1985) and we therefore apply a more conservative metric. Some additional text is added to the main data paper to clarify these choices, and the implementation is shown in the supplementary R code.

Following the Reviewer’s recommendations, we expand the number of metrics used to assess the predictive performance of the seabed rock presence-absence model. These include the sensitivity, selectivity, overall accuracy and mean square error of the model. For ease of interpretation, we replace AUC in the manuscript with sensitivity, specificity and overall accuracy as these provide information on the overall quality of predictions as well as true prediction rate of rock presence.

AR1 - *While sediment composition data were transformed prior to modelling, it is not clear to me whether a transformation was applied to the content of particulate organic carbon and nitrogen. As these parameters are reported as proportions, it would be advisable to apply an arcsine transformation (Sokal and Rohlf, 1981).*

MP – Predictions from Random Forests models are constrained by the range of response data available, thus predictions may not be extrapolated beyond empirical measurements. Nevertheless, transformations may improve Random Forests model performance when the response data are highly skewed. An arcsine transformation did not markedly improve particulate organic carbon and nitrogen model performance – shown in the supplementary R Markdown document – and, hence, no transformation was applied.

AR1 - *Permeability was predicted with median grain size and mud content in two different models; however, porosity was predicted with median grain size only. This is a little surprising, as previous studies found close relationships between porosity and mud content (e.g. Jenkins, 2005; Silburn et al., 2017) and such a prediction would be “one step closer to measured data”, as the authors put it. Was there a specific reason why this relationship was not considered? If not, it might be advisable to investigate whether a similar relationship could be found based on the authors’ dataset.*

MP - The original application of these data products was as validated inputs to a coastal and shelf sea ecosystem model. A relationship between porosity and mud content was not required for this work and therefore not parameterised. Moreover, pairwise measurements of porosity and median grain size were more abundant than porosity and mud content measurements in our assembled dataset. We follow the reviewer’s suggestion and provide a parameterisation for this relationship in the main data paper as it may be useful to other researchers. However, we retain the median grainsize relationship as the basis for the map of predicted sediment porosity in the Firth of Clyde as it is informed by a larger volume of empirical data.

AR1 - *I would also suggest using a different colour palette than the red-white-blue palette used in most figures. Such a palette would suggest a pattern diverging from a central value rather than a continuous increase. See for example Crameri et al. (2020) for advice on choosing suitable colour palettes.*

MP - We follow the reviewer's recommendation and select a sequential colour-blind-friendly palette for use on most figures using the viridis package in R.

AR1 - *Finally, I noticed that the data doi is not working (yet). This will have to be fixed. I was, however, able to obtain the datasets from <https://pureportal.strath.ac.uk/>. Datasets are provided in csv and netCDF format. I assume this is because the predictions were made on an unstructured grid. If possible, I would suggest to additionally supply outputs as georeferenced tiff files, as this is a common format in the science community working with geographic information systems.*

MP - The DOI link has been amended in the revised manuscript. Work is ongoing to supply the data products as georeferenced tiff files and to provide a function to interpolate the data from an unstructured to a structured grid.

Anonymous Reviewer 2

AR2 - 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 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).

MP - The reviewer makes the important observation that Random Forests models perform poorly when predictions are extrapolated to different systems, in part due to the constraint of predictions to the value range of empirical measurements. Accordingly, throughout this manuscript, Random Forests models are fitted exclusively to data collected across the Firth of Clyde from a range of subtidal settings. Moreover, the basis of this manuscript is the presentation of a suite of mapped data products addressing present important gaps in the Firth of Clyde, with random forests serving as the statistical machinery to generate full-coverage predictions from point measurements within the same area.

For each of the mapped variables presented, an initial assessment was undertaken to select the modelling framework, contrasting the performance of linear models, additive models and machine learning tools. For most variables, Random Forests models outperformed other statistical models, with the notable exception of median grain size, where a generalised additive model demonstrated better predictive accuracy. These initial investigations were not presented for two reasons. Firstly, this was beyond the scope of the paper and do not provide the reader with a deeper understanding of the development and limitations of the presented data products. Secondly, inclusion of these initial investigations in either the main paper or supplementary information would add

considerable additional length and complexity to these texts.

To provide clarity to the reader, we revise the manuscript to make reference to the initial assessment of Random Forests performance relative to other model approaches.

Semivariograms are provided for full Random Forests models in Supplementary Information – see response to Anonymous Reviewer 1.

AR2 - 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 1/8th 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.

MP – We agree with the suggestion that some additional detail on the spatial cross-validation methods is required, and, as described in the response to Anonymous Reviewer 1, we accordingly provide additional clarifying text to the main data paper and technical details within the provided supplementary information. We add some justification on selection of spatial blocks to the supplementary material and show that the block size selected was appropriate for the datasets modelled. We would like to clarify here that Wilson et al. (2018) utilise similar methods (see Roberts et al. 2017) but mapped seabed properties over larger and coarser spatial scales and hence used a block size with 1 by 1 degree resolution.

Following the reviewer's recommendation, we translate the dimensions of the data blocks to kilometres. However, this is an approximation due to latitudinal differences in degrees and distance.

AR2 - 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.

MP - We agree with this suggestion and provide an R markdown report to document the code used and provide additional methodological detail, particularly the application of spatial cross-validation and relative explanatory variable importance, for each of the maps generated. However, it was prohibitively complex to adapt this code (comprising almost 3000 lines of code) into the ESSD template, and we hope that the format presented is

acceptable to the journal.

AR2 - 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 mineralogy 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 ($D_{50} < \sim 0.02$ 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 ($r^2 = 0.08$) 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.*

MP – Establishing the credibility of the presented data products and ensuring high predictive accuracy despite limited available data is crucial to this paper. The reviewer correctly notes how sediment geotechnical properties, and organic carbon and nitrogen content were mapped onto previously generated predictive maps of sediment grain size fractions, with clay content playing a key role. We diagrammatically illustrate the dependencies between data, models and maps in the revised manuscript with the inclusion of Figure 5.

Figure 7 (Figure 8 in the revised manuscript) represents a validation test of generated sediment grain size fraction map rather than the fitted Random Forests model. We accomplished this by collecting fresh measurements of sediment mud, sand and gravel, and contrasting these against values predicted for these locations from the generated seabed maps (described in lines 281-283). This represents a robust test of utility of these maps as predictions are also impacted by the mapping resolution and local-scale sediment grain size spatial heterogeneity. However, the noted mismatch may also be a limitation of the sample processing methodology adopted for the collection of independent validation data. These data were derived from subsamples collected from each grab (section 2.3), whereas the BGS data that largely drives mud content predictions were derived from the stacked-sieve processing of the whole grab sample (Deegan et al., 1973).

Conversely, mismatches between observed and predicted median grain size for muddy sediments ($D_{50} < 0.02$ mm) is an irreducibly limitation of the model used. The GAM predicts the whole sediment median grain size based on mud, sand and gravel content, implicitly assuming a single grain size value associated with each fraction that contributes to the whole sediment median grain size as a function of the proportions of the remaining size fractions. In the absence of sand and gravel information, such as for clay samples, the model can only predict the mean median grain size that is associated with 100% mud content within the data set. However, this is not expected to significantly impact derived maps of sediment geotechnical properties as relationships with median grain size < 0.02 mm are typically flat (Figures 10 and 11).

We amend the discussion of the data paper to address these issues. We add a figure mapping the residuals in the supplementary material.

AR2 - 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 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 imagine 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?

MP - We agree that any inferences of a relationship between two variables must be made on contemporaneous measurements. For each of our presented data products, models were fitted to empirical mineralogy data collected from single or replicate samples rather than estimated from predicted maps. The fitted model was then used to predict the spatial distribution of sediment OC and ON in the Clyde using the previously generated map of sediment mud content. It was therefore on the basis of the fitted model rather than predicted seabed map that we inferred a lack of trawling impact (derived from the annual average swept-area ratio between 2009-2016) on whole sediment organic carbon content.

AR2 - 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)*

MP - These papers have been cited in the revised manuscript.

AR2 - Introduction, support for RF: *as mentioned above, you need to provide a reason other than that it has been used by others to prove it is a good/the best approach.*

MP - In the revised manuscript, we refer to initial assessments where Random Forests models outperformed other modelling techniques (section 2.4).

AR2 - Wilson et al 2018 reference: *in the reference list there is only a Wilson et al., 2017 - are they the same one?*

MP - This is a typo in the references that has been amended in the revised manuscript.

AR2 - 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.*

MP - The plotted measurements here are compositional data that must sum to 100%.

Hence, it is appropriate to consider gravel data as true zeroes as these data lie upon the same simplex as the measurements where gravel was recorded. It was only Marine Scotland data where gravel content was not measured (see Table 1).

AR2 - *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.*

MP – Section 2.2.3 refers to the compilation of data, whereas the identically named section 2.4.3 refers to the mapping methods utilised. We clarify in section 2.2.3 that sparse data for the Firth of Clyde were supplemented by data from surrounding coastal seas to fit a statistical model predicting the whole-sediment median grain size from fractions of mud, sand and gravel. In section 2.4.3, the form of this model is presented and we support the assumption that dependencies among these measures of grain size are robust across geographic regions with references to literature.

AR2 - *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.*

MP – We clarify in section 2.2.4 that North Sea permeability and porosity data was used to supplement fresh Firth of Clyde measurements with the purpose of deriving expressions that extend to finer grained Clyde sediments. We explicitly state in section 2.4.4 the assumption that for higher-energy and coarser-grained sediments, North Sea relationships may be applicable to the Clyde.

AR2 - *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.*

MP – A total of 39449 grid points in an unstructured arrangement consistent with the FVCOM for the Firth of Clyde were used to map the seabed properties presented in this paper. The 3244 data points in Figure 8 refer to the number of empirical measurements of median grain size and sediment size fractions used to fit median grain size statistical models (these measurements collected from the Clyde and surrounding UK coastal waters. Figure 3 shows the distribution of non-Clyde data).

AR2 - *Lines 255-259: also in accordance with the ESSD guidelines: provide in Rmd*

MP – We provide the code used to map explanatory variables to the unstructured grid in the R Markdown document.

AR2 - Line 266: *provide details to bootstrapping (frequency)*

MP - We amend the text to "sampling" and provide the number of iterations

AR2 - Line 312: *regarding trawling there has been the recent paper by (Sala et al., 2021) which could also be good to cite*

MP - We cite Sala et al. (2021) in the revised manuscript.

AR2 - 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.*

MP - All functions and code related to RFE are provided in the Rmd supplied as supplementary material.

AR2 - Line 325: *provide reference for density assumption*

MP - Citation for the source of this assumption added

AR2 - 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*

MP - This is superseded in the revised manuscript with the use of alternative performance metrics. Additional text is added interpreting model performance results and suggesting that the model is fit-for-purpose.

AR2 - 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?*

MP - Additional text in figure caption to clarify that the observation data shown were independent, recent measurements that not used in the model fitting process. This is a robust test of not only the model performance but the predictive capacity of the generated sediment map.

AR2 - Figure 8 *please add log scale increments with ggplot*

MP - log-scale increments added to this figure.

AR2 - 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.

MP – It is unclear which line the reviewer refers to. On line 434 in the original manuscript we argue that the fitted permeability-grain size relationships may have wider application given the paucity of empirical seabed sediment permeability data. These functions were fitted to data from the Firth of Clyde and the North Sea, and hence cover a wide range of natural marine sediment types.

AR2 - *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*

MP - Analysis was carried out on sediment grain size and organic C/N measurements from samples collected between 2005 – and 2017. Trawling effort has been consistently high over this period and the analysis therefore shows that trawl intensity gradients do not relate to observed patterns of sediment organic carbon and nitrogen content. However, these observations do not account for any historical redistribution of sediment and organic matter with the onset of trawling in the 1960s and additional research is required to investigate this. The text has been amended to better reflect this.

AR2 - *Section 4.1: connect limitations to RF model performance.*

MP – The limitations section has been expanded to encompass Random Forests models performance and limitations – particularly the inability of models to generate predicted values beyond the range of the observation data used to fit the model.

AR2 - *Line 474: doi does not work. The doi in the abstract does work.*

MP - This has been amended in the revised manuscript.