Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-353-RC1, 2018

Biogeosciences Discussions
Open Access
EGU

Interactive comment on "Quantitative mapping and predictive modelling of Mn-nodules' distribution from hydroacoustic and optical AUV data linked by Random Forests machine learning" *by* Iason-Zois Gazis et al.

Anonymous Referee #1

Received and published: 17 August 2018

General comments

This is a well written paper on the combined use of AUV imagery and acoustic surveys for the assessment of manganese nodules, which shows clear scientific and industrial relevance. However, it shows some similarities to Alevizos et al 2018 (similar approaches, but different locations). Both the size of the area covered and the number of images, highlight the use of AUVs and the importance of automated approaches for environmental assessment. The authors have made a point of being very transparent about their approach, and the statistical details provided in the appendices provide

Printer-friendly version

Discussion paper

extra confidence in the results presented (more of the statistical explanations of the results could be moved to the appendix, e.g. assessment of normality in 4.1).

Specific comments

(1) There is reference to a particular MBES depth data processing that guarantees removal of artifacts and improvements of georeferencing, but no reference or descriptions are given. Particularly in deep waters, inaccurate AUV positioning will be an issue, especially when trying to related photographs to 3m resolution bathymetric grids.

(2) As mentioned in line 420-425, choice of scale is important in deriving terrain metrics and a quantitative justification for the choice of chosen scales should be provided.

(3) The calibration of the model section is not as clear as it could be. Lines 223-225 need to clearly state that the default values were used for the assessment of training/testing sample size only. Lines 238 says that after training sample size was determined were mtry and ntree examined, but Line 241 mentions that for each case of different training sample size, ntree and mtry, the model was run ten times. The latter sentence should be split to clarify that each training sample size was not also tested for each different numbers of ntree and mtry. Similarly, Line 243, if I understood correctly, Appendix A only presents the averages for the 10 training size runs, and not the ntree and mtry runs. The wording here also needs to be clarified.

(4) I am not convinced that the approach taken can be used to determine the optimal training sample size proportion. More data is likely to yield better models, but by decreasing number of testing data points, one can also expect MSR to keep decreasing (as was shown here). A much more interesting question would be how many samples are needed to obtain accurate predictions.

(5) RF models also provide a measure of uncertainty, it would be interesting the provide uncertainty maps for the associated predictions and discuss potential spatial trends if any.

(6) Autocorrelation in Mn nodule distribution was discussed, but whether model residuals showed any spatial autoccorelation was nor assessed, nor were the effects of this autocorrelation on model assessment discussed.

(7) The discussion is very much focused on the model and although the exploratory nature of machine learning algorithm is mentioned, a little more discussion of the causation mechanisms (or potential hypothesis) would be valuable.

Technical corrections

Line 38. I would suggest changing sea bottom for seafloor

Lines 39-45 I would suggest specifically introducing the term backscatter, as I believe that to be one of the main data product used for to show Mn trends in regional surveys

Line 61 Reference style

Lines 81-83 Awkward, please rephrase

Line114 In the marine environment,

Line 131 remove scale)

Line 133 is deeper and has less relief

Line 189, while

Line 222 a threshold of 0.95 for correlation of variable seems very high

Line 256 in the study area,

Line 259 First sentence seems repetitive

Line 260 I do not think that the word ‘alternation’ here is the right one

Line 263 to 0.18%

Line 265 Awkward, please rephrase

Printer-friendly version

Discussion paper

Line 266 change approved by to supported

Line 270 measurements

Line 274 like Kriging

Line 275 area, and it is an important step

Line 276 and the produced bathymetric derivatives

Line 290 after a distance of 400m

Line 345 I would suggest to avoid finishing a sentence with too

Line 356 For our data,

Figure 12 b) for which mtry and c) for which ntree?

Line 385 Table 5 MAE, MSE and RMSE were not introduced previously, only MSR was mentioned in to method section

Line 413 The analysis of RF variable importance

Line 414 specific depth ranges

Line 415-417 I would suggest removing this sentence as it is not necessary

Line 418 All of them also contribute in a nonlinear way

Line 444 the study area, equal to

Lines 458-459 Awkward, please rephrase

Lines 468-471 Awkward, please rephrase

Line 471 Conversely,

Line 475 clues as to why

Line 491 Along these lines, several authors have included

Line 513 ‘a priori’

Line 516 as well as their size

Line 518 interest). Finally,

Line 561 the remaining derivatives

Lines 565-566 training and testing records

Line 570 Should be 4.1

Throughout, ground truth vs ground-truth, hydo-acoustic vs hydroacoustic, circa vs ca., space vs no-space between value and unit

---

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-353, 2018.

Printer-friendly version

Discussion paper