

Biogeosciences Discuss., author comment AC4
<https://doi.org/10.5194/bg-2021-189-AC4>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC4

Brandon J. McNabb and Philippe D. Tortell

Author comment on "Improved prediction of dimethyl sulfide (DMS) distributions in the northeast subarctic Pacific using machine-learning algorithms" by Brandon J. McNabb and Philippe D. Tortell, Biogeosciences Discuss., <https://doi.org/10.5194/bg-2021-189-AC4>, 2021

I appreciated the analysis of impact of gridding resolution on the results. However, I wonder about the impact of binning the DMS data monthly regardless of year. Looking at the data in Figure 4, there is significant patchiness which I can only imagine is temporally and spatially variable. Given the power of the machine learning algorithms, why not use the full complexity of the dataset and pair the DMS observations with the closest (spatially and temporally) measurement of the predictor data sources?

We considered this, but there were a couple of factors that influenced our decision to use a monthly resolution. Training these models on daily resolution data could introduce bias due to autocorrelation among observations from the same cruise (L426-427, also discussed in Wang et al. 2020) and monthly-binned data allows us to reduce this source of uncertainty. In addition, daily resolutions have poor spatial coverage due to cloud cover creating gaps in the satellite's field of view. Monthly averaged predictors thus require less interpolation to match observations in space/time, allowing uncertainty in the models' training accuracy to be reduced.

Two machine learning algorithms were used in this study but there wasn't a robust analysis of which one was better and why. Should future studies use one over the other? Does one need to try multiple methods? Such a discussion would be a valuable addition.

We feel that the results presented largely show a strong agreement between the two methods (discussed in Sec. 3.2), as illustrated by the similar predictive accuracy (Fig. 3), spatial distributions (Fig. 4), and coherent predictor correlations (Fig. 6,7). Although there are some areas where the models deviate spatially, these are also areas with poor observational coverage (L271-273), which makes it difficult to ascertain whether one model's estimates are superior to the other. Future studies will likely benefit from applying both approaches to other regions of interest for DMS, where differences in algorithm performance may become more apparent.

Minor comments:

- The methods are very sparse. More information on the machine learning algorithms should be included (e.g. was this done with a package? If so which one?) This is in the 'code availability' statement to some extent but should be included in the methods along

with a brief description of the algorithms and differences between the two.

We have expanded Sec. 2.2 (L138-145) to briefly describe the two algorithms used. We have also added a line noting the specific package/functions used (L153-154).

- Only 20% of the data was held back for testing. It seems that it would be better to have a 50/50 split to provide a sufficiently large dataset for testing to confirm the robustness of the results.

The major limitation of these machine learning algorithms is that their performance is sensitive to the size of the training dataset. As a result, the typical approach is to feed a larger fraction of the data (ex. 70%, see Weber et al. (2019); Roshan & DeVries (2017)) into the training process to allow for appropriate learning of the underlying patterns. In contrast to these global studies, we have chosen a slightly more restrictive train:test split of 80:20 to compensate for the reduced sample size and smaller geographic extent associated with a single region.

- Are there any issue with correlations between the predictor variables? For example, many are derived from MODIS and so should have inherent correlations (ie not independent measurements).

There is likely some inherent covariance between predictors, to a degree, given their distributions are dependent on similar processes (for example: circulation patterns, or nutrient depletion via biological production). We note, however, that we have taken steps to reduce any covariance that may confound the models' results, such as including only a single biological predictor (see Sec. 2.6) and iteratively testing the addition of each new predictor on the RFR and ANN performance during development (for example, the extinction coefficient, K_d , was removed as it decreased R^2 due to covariance with other predictors).

- Figure 1: It seems a bit surprising that the R^2 value decreases so dramatically with resolution but the DMS flux barely changes. Is this just due to the large spatial variability in the flux?

Yes, this is due to spatial variability.

Line 37 missing an 'a' → by a suite of environmental ...

Thank you, this has been corrected.

Line 152: typo? Should it be modified from?

Thank you, this was a typo. In response to another reviewer's comments, this section has now been rephrased using a new k parameterization.

Eq 4: are the coefficients provided anywhere?

There are no coefficients for Eq. 4 (SRD) which is used within the VS07 model.

Figure 1: caption refers to green lines/symbols instead of black

Thank you, this has been corrected.

Line 261: it would be helpful to provide the fractional area represented by the study region. For example, if it accounts for only <1% but accounts for 4-8% that is more impactful.

We have removed this line in response to another reviewer's comment. In short, we have reevaluated our calculations to report the summertime regional-averaged fluxes only as $T_g S$, as assumptions in the conversion to an annual flux estimate were likely erroneous.

Line 461: it should be "approach for modeling"

Thank you, this has been corrected.