Biogeosciences Discuss., referee comment RC2
https://doi.org/10.5194/bg-2021-189-RC2, 2021

Comment on bg-2021-189

Anonymous Referee #2

Referee 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-RC2, 2021

This manuscript describes two machine learning techniques used to model DMS distributions in the NE Pacific Ocean. The authors find their techniques perform better than currently used statistical models in mapping DMS seawater concentrations. In addition, they are able to use ancillary parameters, such as PAR, nitrate, and sea surface height to derive predictive relationships with DMS. The authors also compute DMS air-sea fluxes using the modeled DMS concentrations.

The manuscript is well written and is a useful addition to our toolkit/knowledge about DMS distributions. Machine learning techniques are becoming more readily utilized and these techniques should be applied to surface ocean distributions of important trace gases, especially those that are hard to predict. I recommend that this manuscript be published in BG after the following minor comments are addressed.

Specific comments:

Lines 98-100 – Did the authors check for any additional data not included in the PMEL database?

Paragraph starting on line 102 – Were any of these satellite derived parameters ground-truthed against in situ data during any of the cruises?

Lines 150-152 – The Nightingale et al. (2000) parameterization of k is not really appropriate for DMS. It is becoming more and more clear that the k wind speed-based parameterization for DMS should be linear (Blomquist et al., 2017; Bell et al., 2013;

Zavarsky et al., 2018).

Line 224 – typo, should be ANN not AAN

Section 3.4 – It seems highly likely (and I believe the authors allude to this too in the discussion starting on line 388 paragraph) that the correlations found (especially with SSH) are indirect. The real driver of DMS distributions is likely nutrients and type of microbes present. If the SSH represents eddies carrying the relevant nutrients, SSH is not really a universal parameter that can be used to describe DMS distributions everywhere. It would be good to see how that works in other regions without much eddy activity. Were phosphate and bacteria looked into? It seems that bacterial counts and types are important, but difficult to account with things like satellite data. Is it possible to compile that info from the in-situ measurements – or is that too low resolution for the techniques?

Discussion (and Intro) – Why was this region chosen instead of one with more data coverage? Or why not try two different regions and compare findings? The area and number of data points (compared to the region that is mapped) seems small (i.e., Figure 4).

Section 4.1 (and methods section starting on line 187) – Why don't the two iron limitation proxies resemble each other at all? Also, the use of SSN is not really a unique identifier (e.g., effect of nutrients and photochemistry, as stated in the paragraph starting on line 388). How can this be practically handled when using SSN as a predicter? And if the relationships cannot be understood, why is it used?

References

Bell et al., 2013, https://doi.org/10.5194/acp-13-11073-2013

Blomquist et al., 2017, https://doi.org/10.1002/2017JC013181

Zavarsky et al., 2018, https://doi.org/10.1029/2017JD028071