Wang et al. Global ocean dimethyl sulfide climatology estimated from observations and an artificial neural network

Review of revised manuscript by Martí Galí
General comments
The manuscript has seen some substantial improvements compared to the first version. I appreciated the more comprehensive comparison between DMS fields generated by the ANN and those generated with simpler empirical algorithms, including remote sensing algorithms. The comparison is informative as to the uncertainties in different global DMS products.
However, I was disappointed to find that several inconsistencies and inaccuracies remain in the text. These issues have to be addressed and clarified.

We tried out best to make the paper clear and easy to follow. The changes are highlighted in the revised manuscript.

In the previous review I prompted the authors to provide a more up-to-date view of DMS(P) cycling processes. However, this did not require adding long discussions, just improving the quality of the information and the accuracy of literature references. Now, in section 3.3, the authors discuss why some environmental variables got selected as ANN predictors in relation to available knowledge. But they do not show any results about the sign of the effects of each predictor variable used in the ANN, or whether the sign changes from one region to another. Thus, the discussion becomes quite pointless (because our understanding does not increase), and could be more succinct. The distinction between “controlling factors” and “predictors” should also be clearer throughout, as this paper does not shed light on the controlling factors.

We have added sensitivity results to demonstrate how environmental parameters impact the DMS concentration distribution. Meanwhile, we merged part of Sec. 3.3 into Sec. 3.7 Sensitivity test, and shortened the discussion to make it clear and succinct.

l.387-440
Section 3.3 screens key parameter combinations that have the highest prediction skill. To demonstrate how these parameters influence the predicted distribution and sea-to-air flux of DMS, we ran a series of sensitivity tests. In each test, we increase/decrease one environmental parameter at a time. Fig. 9 shows annual mean differences between perturbed models and the control model. These sensitivity tests show regional differences in the sign of the perturbations anomalies. This non-linear behavior of the ANN model is not possible with a simple linear model.

For the temperature sensitivity test, we uniformly increase SST by 2 ˚C for the whole ocean (Fig. 9a). Compared to the control case, DMS concentrations are lower in most of the low and middle latitude oceans and higher in high latitude oceans, especially in the Southern Ocean, the Bering Sea, and the high latitude N. Atlantic Ocean. In contrast, the linear regression model shows no correlation between SST and DMS. SST alone with date and location parameters has very low prediction ability (ranked 244th over 255 models). When combined with other parameters, SST helps to improve the model performance. For example, the combination of SST and MLD ranks 2nd among all models.

For the mixed layer depth sensitivity test, we decrease MLD by 10% to mimic the stronger stratification in a warming world (Fig. 9c). DMS concentrations increase in most of the ocean, in line with the linear regression result. In the PAR sensitivity test, we uniformly increased PAR by 10% with the expectation that light exposure will increase in the future because of MLD shoaling (Fig. 9e). DMS concentrations increase with increased PAR, in agreement with the linear regression result and also with the physiological role of DMS. First, high radiation negatively influences the bacterial population/activity, which decreases DMS consumption (Galí et al., 2013a, b, c; Royer et al., 2016). Second, high radiation promotes DMS production by inducing oxidative stress within algal cells (Toole et al., 2006; Sunda et al., 2002; Royer et al., 2016).

For the salinity sensitivity test, we uniformly decrease surface ocean salinity by 1 psu. Similar to the temperature sensitivity result, the changes of DMS concentration show regional variations. DMS concentrations increase in most of the Southern Ocean, the high latitude N. Atlantic Ocean, and the Arctic Ocean, whereas DMS levels decrease in the eastern North Pacific Ocean, the Indian Ocean, and South Atlantic Ocean (Fig. 9b). The linear regression model also shows that there is no significant correlation between DMS and salinity. As in the case for temperature, salinity works synergistically with the other environmental parameters to predict the DMS concentration.

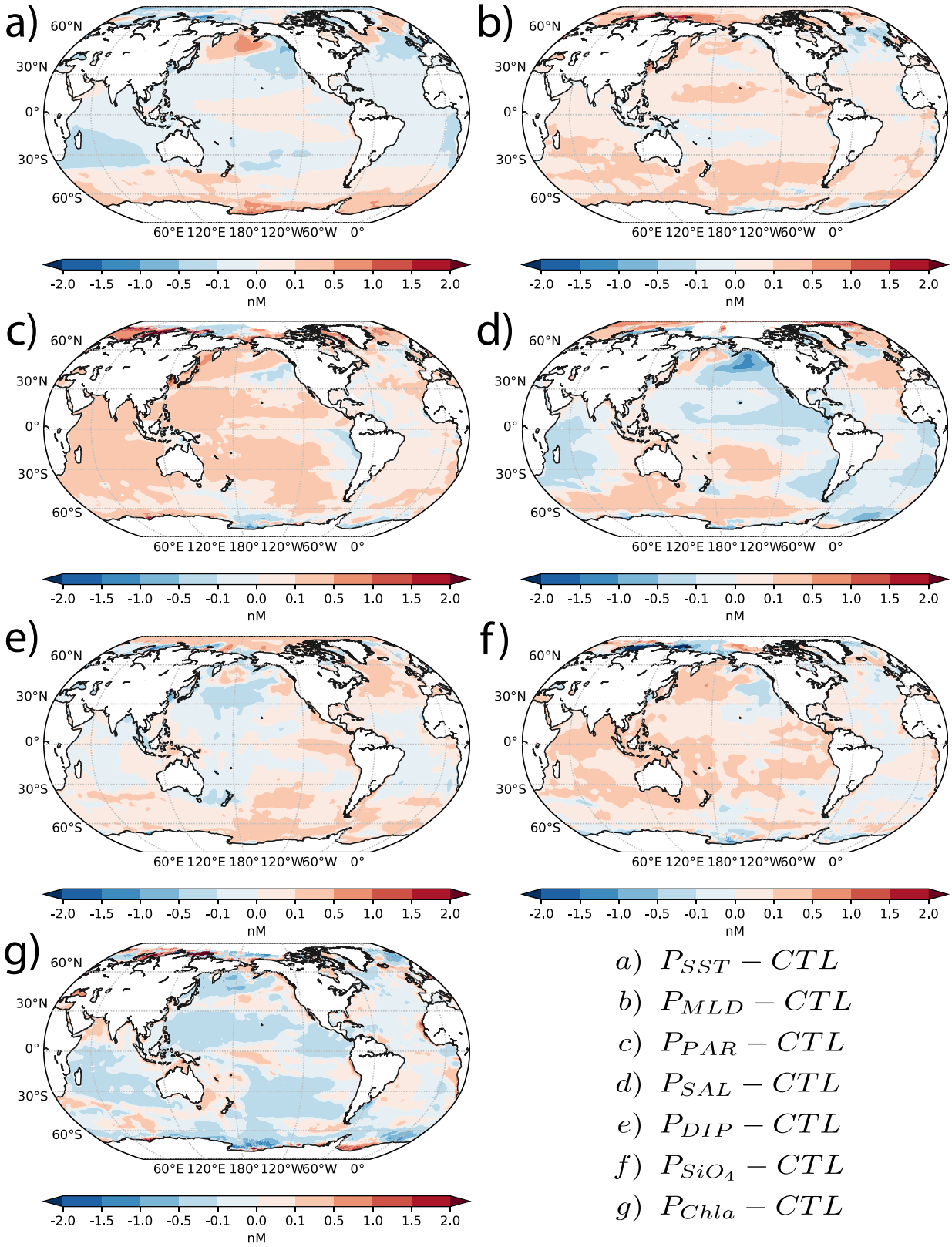
Figs. 9e and f show the sensitivity tests for DIP and SiO_4, respectively. For these tests, we decrease DIP and SiO_4 concentrations by 10% with the expectation that increasing ocean stratification due to global warming will decrease the nutrient supply from the deep ocean. In certain regions, the two nutrient perturbations have nearly opposite effects. For example, DMS concentrations drop slightly in the western Pacific and Indian Ocean for the DIP perturbation experiment, whereas the concentrations have almost opposite patterns in those regions for the SiO_4 perturbation experiment. In the eastern Pacific Ocean, the Southern Ocean, and high latitude N. and S. Atlantic oceans, reduced DIP concentration triggers an increase of DMS concentrations, which might be related to nutrient stress, which can increase DMSP production by low DMSP producers (e.g. diatoms) (McParland and Levine, 2019). The increase of DMS concentration for the SiO_4 perturbation is potentially due to a regime shift away from diatoms, which are low DMSP producers, to other more prolific DMSP producers.

Fig. 9g shows the sensitivity test for Chl a. In the test, we decreased Chl a concentration by 10% to mimic the decreased primary production caused by ocean stratification and nutrient depletion. Overall, the most apparent changes are in the subtropical gyres, where DMS concentrations are lower than the control run. DMS concentrations increase in some marginal seas and coastal oceans such as the Arabian Sea and eastern coast of Australia. Previous studies of the relationship between DMS and Chl a have produced contradictory results. Strong correlations have been reported in basin scale studies (e.g. Yang et al., 1999). On the other hand, there are numerous studies that observed no correlation between DMS and Chl a (e.g. Dacey et al., 1998; Kettle et al., 1999; Toole and Siegel, 2004). The inconsistent relationships indicate the complexity of the reduced sulfur cycle.

On a global scale, the increase of temperature does not significantly change sea-to-air flux (15.96 Tg S yr^{-1} compared to 15.89 Tg S yr^{-1} for the control run based on GM12) because the elevated DMS concentrations in the high latitude oceans are compensated by the reduced concentrations in the low latitude oceans. Similar to the case for the temperature perturbation, the salinity perturbation has a small effect on sea-to-air flux of DMS (15.88 compared to 15.89 Tg S yr^{-1}). The overall increases of DMS concentration in the MLD, PAR, and SiO_4 perturbation tests lead to increases of DMS sea-to-air flux of 0.56, 0.96, and 0.91 Tg S yr^{-1}, respectively. The Chl a perturbation model is the only one that shows a slight decrease in the sea-to-air flux of DMS (15.59 Tg S yr^{-1} compared to 15.89 Tg S yr^{-1}).

Of course, the ocean is a very complex system and changes in these environmental parameters will be correlated. For example, the projected temperature increase will lead to a stronger surface ocean stratification that will result in shoaling of MLD and reduced nutrient supplies from the deep ocean, which together will decrease primary production in the ocean. Based on our model results, if these effects work

jointly, the DMS sea-to-air flux will increase more than each of the individual perturbations. Assuming that larger DMS sea-to-air fluxes induce greater cloud albedo, then we might expect the changes in DMS to represent a negative climate feedback.



a) $P_{SST} - CTL$

b) $P_{MLD} - CTL$

c) $P_{PAR} - CTL$

d) $P_{SAL} - CTL$

e) $P_{DIP} - CTL$

f) $P_{SiO_4} - CTL$

g) $P_{Chla} - CTL$

Fig. 9 Differences of annul mean DMS concentration between perturbation models and the control model. Specific figure indexes are listed in the figure, where Pxxx represents a perturbed model and the subscript xxx indicates which parameter is changed. C T L is the control model that is the average of our top 10 model results (Fig. 5)s

As in the previous version of the manuscript, the authors make a strong case against the binning of in situ DMS measurements, and present a new figure (Fig. 2) to support their choice. However, I still find that the "binning issue" is treated in an inconsistent manner. For example, despite advocating the importance of using non-binned data, the authors do not use time-resolved satellite Chl as predictor although such data are available for most of the DMS measurements used to train the ANN. Instead, the authors match non-binned DMS data to a heavily time-binned global climatology of satellite Chl from SeaWiFS, which does not even cover the entire ocean color satellite era. Finally, they produce global 1x1 monthly DMS climatologies by applying the ANN, previously trained on non-binned DMS data, to climatological fields of the predictor variables. To be clear, I am not asking the authors to perform strict matchups between DMS and satellite data, which I think is beyond the scope of their study (note, however, that a DMS-satellite match-up database is available for measurements older than 2012: https://doi.org/10.5281/zenodo.2205131). In my view the paper needs clearer and more consistent reasoning and methodology regarding the utility of binned data to produce climatologies, because (1) the whole point of making climatologies is precisely collapsing temporal and spatial variability, and (2) the trained ANN is not implemented to produce non-climatological fields in this study.

As we explained in our response to the first reviews, binning the data will reduce data variance. This point has also been demonstrated by Derevianko et al. (2009). For the DMS dataset, sampling location and time are important predictors (Section 3.2 and 3.3). Binning and averaging the data lead to loss of these valuable information. Specific to the ANN method, binning the data leads to overfitting, which is shown in Fig. 2.

We argue that there are differences between how a model is trained and how a model is applied. When you train a model, it is important to make the model be able to generalize. That way, you can apply the model to either a coarse or a fine resolution filed. For example, we trained the network model using the raw data and later we are able to apply the model to the very fine resolution NAAMES fields (0.0417° ×0.0417°) (Bell et al. in preparation). We also published our models and code, therefore, any interested reader can reproduce the results and can also apply the model to their own data, regardless of data resolution.

Regarding Chl a, we compared PMEL in situ Chl a to the satellite derived Chl a. Chl a data from these two sources are well correlated (R^2 = 0.64 on logarithm scale). Chl a is not a strong DMS predictor. The ANN model does a fairly good job even without it. Moreover, we do not have time series MLD and nutrient data to strictly match DMS measurements. To match DMS with only Chl a would not be meaningful, because the other parameters are all climatological.

In connection with comment above: no quantitative assessment of overfitting is provided in section 3.4 nor figure 2. Just a qualitative assessment based on “how much the training and validation errors overlap”. Although the inclusion of figure 2 is welcome, a quantitative assessment would provide stronger support to the authors’ choices (ie training the ANN using non-binned data).

The quantitative assessments have been added.
l. 301-302:
For example, the averaged RMSE on natural logarithm scale for the 10 best ANN models is 0.608 for the validating dataset and 0.600 for the training dataset when using the un-binned data, whereas the RMSE is 0.655 (validating) and 0.635 (training) for the model constructed using the binned data (See Fig. 2b).

Specific comments
Line numbers refer to the document bg-2020-72-AC4-supplement.pdf
Abstract
The reasoning in the second sentence is backwards, in my opinion. In addition, I think this paper deals with the “predictors”, not with the “controlling factors”.

The sentence is revised to,

Knowledge of the global-scale distribution, seasonal variability, and sea-to-air flux of DMS is needed in order to improve understanding of atmospheric sulfur, aerosol/cloud dynamics and albedo.

We have changed “controlling factors” to “predictors”.

Introduction
L45: Some phytoplankton do not produce significant amounts of DMS, just its precursor DMSP. Please clarify.

Corrected. The new sentence is as follows,
l.47-48
In these models, phytoplankton are divided into different groups based on their ability to produce DMSP, the precursor of DMS.
L44-51: This entire paragraph is misleading. Reading it, one may deduce that the inability of biogeochemical models to capture global DMS distributions stems from the excessive simplification of phytoplankton diversity, lumping together very different taxa in a few phytoplankton functional types. Le Clainche et al. (2010) propose otherwise: it is the excessive dependence of DMS on phytoplankton dynamics in sulfur cycling parameterizations what results in poor model skill. And they explicitly suggest that modulation of community DMS production yields by environmental stressors (light, nutrients) must be included in model parameterizations. For example, by Vallina et al. (2008) showed the importance of adding light stress effects on DMS production in the Sargasso Sea using a model that had a single phytoplankton compartment. I refer the authors to my previous review for papers reporting experimental evidence of nutrient

and light stress. Finally, note that citation of Le Clainche's paper does not support the statements made in this paragraph.

We have rephrased the last two sentences to make our statement clearer.
l.53-55
Despite this level of modeling detail, there are still large discrepancies between the model simulations and in situ measurements (Tesdal et al., 2016). Le Clainche et al. (2010) suggested that environmental conditions should be included in future model development because DMS cycling depends strongly on phytoplankton dynamics.

L57: "Interpolation from neighboring provinces". I suggest adding "weighted" before interpolation. Actually, what Lana et al. (2011) did in provinces with insufficient monthly data substituting their seasonal cycle by that of the "biogeochemically closest" province, weighted by the "local province" average.
Corrected. Thanks.

L59-69: The message of this paragraph is unclear and the newly added sentences disrupt the flow. In addition, I think the authors should refer to "predictors" instead of "controlling factors". The ANN approach as implemented here does not reveal the controlling factors.
We have rephrased the paragraph for clarity. The revised section is as follows,
l. 69-74
In this study, we explore the relationships between DMS and environmental parameters using a machine learning method. Such relationships are hard to detect using traditional linear regression methods, because environmental parameters do not directly influence DMS concentration. They control the distribution of marine algae that determines the distribution of DMSP (a precursor of DMS) and its conversion to DMS (Kiene et al., 2000; Simó, 2001). The objective of this paper is to discover the relationships between DMS and environmental variables, with the goal of constructing a novel monthly-resolved DMS climatology.

We have added sensitivity tests that reveals how each environmental parameter can influence DMS distribution, and thus we believe the use of controlling factors are appropriate.

Methods
L84-85: With the current writing, it is unclear whether measurements were matched to the 1997- 2010 SeaWiFS climatology or to the multiyear time series. In the latter case, the period 2011- present would not be covered by matching satellite data. In the former case, one may wonder if the 1997-2010 climatology is representative of the 2011-present period in all ocean areas. Please clarify and provide a solid argumentation for not matching DMS data to satellite data at the best available resolution.
Note: I found the answer later, in L232. As suspected, DMS data were matched to climatological Chl. Note also that global Chl climatologies from multiple sensors spanning 1997-2020 are available (e.g. GlobColour, ESA CCI).

We have made the Chl *a* data source clear in Sec. 2.1. The sentence is in line 89-90.

L84: What reprocessing? Access date? Same applies to L94.
Information added. Thank you.

L114: This contradicts L91, where the authors say that “For consistency, we use only Chl-a data retrieved from SeaWiFS in the following multilinear and network models”.

There is no contradiction.

L114 describes how we run linear regression model. In the linear regression model, we use both in situ and satellite Chl a data.
To make it clear, we have added the following text in the manuscript.

l.118-119
In a third step, to keep Chl a data sources consistent as described previously, we use satellite Chl a;

L117: The statement that “there is no available climatological [DMSP] dataset to fill the missing values” is not entirely true. There is no climatological DMSP dataset based on objective interpolation of in situ data. But there is a global sea-surface DMSPt climatology based on the remote sensing algorithm of Galí et al. 2015, available here: https://doi.org/10.5281/zenodo.2558511.
Thank you for pointing this out. We have rephrased the sentence.

l. 122-123
DMSPt is not included, because there is no observation based climatological dataset to fill the missing values.

The DMSP database that the reviewer referred to is a valuable resource, but we do not intend to include that in our model, so that we avoid building our model based on another model’s results.

L118 and elsewhere: Please replace SiO by SiO_4.
Thank you, SiO has been replaced with SiO_4.

L154-162: It is unclear what the authors did. I could not understand whether data had finally been shuffled or not, and why.
We have rephrased the corresponding sentence to make it clear as follows,

l.160-163
The data was split into the above sets manually rather than automatically. This is because data collected from the same cruise are highly intercorrelated. The common practice of shuffling and randomly splitting the data produces an over-fitted model because the validating data can be predicted using near-neighbor values. This kind of apparent skill does not generalize to regions with large data gaps, which we need for constructing a robust climatology.

Section 2.4.2: I don't see the point of training the ANN using non-binned data to, afterwards, compute the DMS climatology from 1x1 gridded climatologies of the predictor data. If training the model on non-binned data was so advantageous, I would expect the authors to first compute global gridded DMS time series at the highest possible resolution, and only afterwards collapse the multiyear fields into a climatology. The advantage of training the ANN with non-binned DMS data would be demonstrated if a climatology produced using the minimal possible amount of climatological predictor fields outperformed another one computed directly from climatological predictor fields.
Please see our reply to general comment #2;

Equation 7: this equation produces negative DMS flux at WS < 1.33. Although such low wind speeds are infrequent in a climatology, did the authors add a correction to avoid potential negative values?
Yes, negative $K_{w,660}$ was set to zero in our algorithm. We have made it clear in the text.

Results and Discussion
L224: I suggest replacing "it is relatively easy to parameterize in a biogeochemical model" by something like "it is a priori more amenable to simple parameterizations".
The sentence has been modified. Thanks.

L248: "the model roughly predicts the level of DMS concentrations" sounds a bit vague. Can you please report skill metrics like RMSE or bias in addition to R2?
R^2 are added. Thanks.

L257: [the ANN] "also incorporates diurnal and seasonal signals present in the data". Please add something like "but see below", or place this sentence after discussing why time of day does not improve ANN predictions. In fact this sentence is a bit contradictory with the finding described two paragraphs later.
We changed the wording as follows,

l.257-258
The ANN model can also incorporate sampling time and coordinate signals present in the data (see below).

L264: "As shown in Fig. 2... sampling location and date alone can explain 44% of the validation data variance". Figure 2 does not show R2. Please rephrase.
We have removed "As shown in Fig. 2,", because we realized that adding R^2 values in Fig. 2 is too messy because there are too many R^2 values.

L265: The discussion that starts here is interesting but needs a little revision. The analysis of Vallina & Simó 2007 did not address diel cycles, so I do not agree with this sentence: "Given the strong correlation between solar radiation and DMS concentration reported by Vallina and Simó (2007), one would expect that adding sampling time would improve the model performance".

The occurrence of predictable diel cycles was assessed by Royer et al. 2015 (Small-scale variabilityscale variability patterns of DMS and phytoplankton in surface waters of the tropical and subtropical Atlantic, Indian, and Pacific Oceans) using continuous underway DMS data collected across the global oligotrophic oceans. They concluded there was no such a universal diel cycle, for the reasons pointed out in the manuscript (different possible outcomes of the balance between DMS sources and sinks over the die cycle).
We agree that Vallina and Simó 2007 is not appropriate to support DMS diurnal cycle, we thus have deleted the reference and change the wording. The revised text is as follows,

l.266-167
Time of day can be another possible predictor if DMS concentration varies diurnally. However, adding time of day to the model increases RMSE slightly (Fig. 2a).

L281: I do not think the cryoprotectant role has anything to do with global DMS patterns, considering that water temperature in most oceanic regions never decreases to freezing temperatures. Please remove or elaborate a better explanation. In addition, what is the sign of the SST influence on DMS?
We have removed the corresponding sentence.

We have added sensitivity tests, DMS and SST do not change unidirectionally (please see our earlier reply to the general comments).

On the contrary, I agree with the sentence that concludes the paragraph. SST and MLD have known for decades to define and capture a large deal of the biogeographic patterns of the ocean (e.g., Fay & McKinley, 2014, ESSD).
Thanks.

L290: “rate” should be “rate constant” (= specific rate). Not the same as rate!
Done, thank you.

L295: “loss” should be “transport” because sometimes turbulent vertical transport can result in net inputs to the upper mixed layer.
Done, Thanks.

L351: Please remove “runoff”. I would use expressions like “higher freshwater content” or “salinity stratification”. Freshwater may come from ice melt, continental runoff, etc... And in addition, much of the stratification in high northern latitudes is due to encounter between fresher Pacific-derived waters and more salty Atlantic waters.
Done, Thanks.

L361: It is risky to use these maxima to illustrate your point. If anything, I would rather use quantiles. Very high Chl from SeaWiFS may result from algorithm artifacts in

CDOM-ladden waters, whereas Chl data from PMEL were not quality-controlled and some were collected in estuarine areas (Galí et al., 2015, RSE).

In the manuscript, we stated that in situ Chl a has greater variance compared to satellite Chl a, whereas, the reviewer indicated the opposite. In addition, coastal data are filtered out according to the reviewer's suggestion in the first round of revision.

We anyway changed the expression and used percentile to demonstrate our idea.

l. 325-327
Extreme conditions are smoothed out in climatological data, e.g. in the DMS database the 99 percentiles of in situ Chl-a concentration is 12.58 mg/m^3, whereas it is only 6.85 mg/m^3 in the SeaWiFS climatology.

L371-374: In the marginal ice zone, DMS can reach high concentrations during-after ice breakup without any need for sea-ice-algae release. Simply because the biomass of high-DMS(P) phytoplankton (Phaeocystis) in the water column can be very high. See the compilation by Levasseur (2013; Nat Geo). Out of the seasonal ice zone, the argument on the cryoprotection function does not make sense, since subpolar waters where coccolithophores bloom in both hemispheres do not reach freezing temperatures, and for example Emiliania huxleyi blooms typically happen late in summer at tempertaures >10C. Please rephrase.

We have changed the wording to emphasize on the main cause of high DMS in high latitude. We have also put cryoprotection and release of ice algae in the context with ice edge zone.
The revised text is as follows.

l.338-344
The high DMS concentration during the summertime at high latitudes is believed to accompany blooms of coccolithophores and Phaeocystis, which are strong DMSP producers (Neukermans et al., 2018; Wang et al., 2015). The shoaling mixed layer depth during the summer provides favorable conditions, i.e. stable and warm, with adequate irradiation for coccolithophores and Phaeocystis growth (Galí et al., 2019). Additionally, high DMS concentrations at ice edge zones have also been observed. These high concentrations are due to the release of ice algae that are prolific DMSP producers (Stefels et al., 2012; Webb et al., 2019). As an important cryoprotectant and osmolyte, DMSP helps ice algae to cope with the low temperature and high salinity conditions (Thomas and Dieckmann, 2002).

L396: In the text you report a global DMS emission of 15.9 Tg S yr-1using the GM12 parameterization. However, when I look at Fig. 8, I see a mean monthly DMS emission of about 1.5 Tg S yr-1 (and, for sure, greater than 1.4). Multiplying 1.5x12, we get an annual emission of 18 Tg S yr-1. Therefore, I wonder how do you arrive at the value of 15.9 Tg S yr^{-1}.
Thank you for pointing this out. We accidently plotted the flux based on N00, whereas used GM12 in the caption. We have updated the figure to make the figure and caption consistent.

L430-434: Agree with this point.
Thanks.

Technical corrections
L84: Check verb tense in this sentence and concordance with the previous one.
Done, are --> were

L94: in the paragraph above the resolution was 9.2 km, now it is 93 km. Please check.
We have double checked the resolution; it is 9.2 km. Thank you for catching it.

L173: coordination or coordinate?
Corrected, thank you.

L348: please add “concentrations” after “higher”.
The word has been added.

L426: principal, not “principle”.
Corrected. Thanks.