[i]Review of the manuscript submitted to Biogeosciences Discussions: "Global ocean dimethyl sulfide climatology estimated from observations and an artificial neural network", by Wei-Lei Wang, Guisheng Song, François Primeau, Eric S. Saltzman, Thomas G. Bell, and J. Keith Moore.

General comments

The manuscript by Wang and coauthors proposes an interesting methodological development: the use of artificial neural networks (ANN) to produce a global gridded climatology of dimethylsulfide (DMS) concentration at the sea-surface. This is a relevant topic because DMS emission drives aerosol formation in the remote marine atmosphere, with subsequent effects on aerosol and cloud radiative forcing. Measurements of sea-surface DMS concentration are too sparse to be directly usable in studies of atmospheric chemistry and global sulfur biogeochemistry. Therefore, a number of techniques have been used in the past to produce global gridded DMS fields: from objective interpolation (on which the standard climatological product is based; Lana et al., 2011) to empirical remote sensing algorithms or prognostic ocean biogeochemistry models.

Thank you for the positive words.

The "artificial intelligence" approach proposed by Wang and coauthors is a necessary step to improve existing global DMS products. The article is generally well written and I appreciated the succinct style. However, in its current form the study suffers from a number of important shortcomings:

1. Repetition of results that have been presented more in depth in previous papers. The results of the correlation analysis between DMS(P) and environmental variables are of little interest as they are extremely similar to those reported in previous papers, where they were analyzed more in depth and with a more solid theoretical underpinning. The stepwise multilinear regression (sMLR), which is used mainly to contrast its limited predictive power against the greater predictive power of the ANN, is only partially described.

Response: Our goal with the sMLR model is indeed to gradually demonstrate that ANN is better than traditional linear/multilinear models. In the revised version, we added more in-depth discussion of each model results, and added more details about the multilinear model. We also tried to minimize any repetition of previous findings.

2. Failure to perform appropriate quality control of the raw DMS, DMSP and chlorophyll (Chl) data, for example following procedures detailed by Galí et al. 2015 for the same global DMS database used by Wang et al. The main flaw is the use of in situ fluorometric Chl measurements and satellite-retrieved Chl as if they were equivalent –they are not.

Response: This is a good point. Thank you for pointing this out.
In the revised version, we followed the guideline introduced by Galí et al. 2015 to conduct quality control. Specifically, we removed DMS data with concentration less than 0.1 nM and greater than 100 nM, we also removed data with salinity less than 30 psu, so that we focus our

study on the open ocean. We removed DMSPt data that are less than 1 nM. Other than that, we did not do any binning and averaging to preserve the original data variance. Finally, there are 10404 pairs of DMS-Chl-*a* and 4061 pairs DMS-DMSPt, which is substantially more than what was reported by Galí et al. 2018 (with 3637 DMS-DMSPt and 8141 DMS-Chl pairs). This is because, the PMEL database has been greatly expanded, it now has over 80K DMS data points.

For Chl a data, we have added the following discussion (l.84 – 1.92).
"SeaWiFS Chl-*a* data (Level 3-binned, spatial resolution of 9.2 km) from December 1997 to March 2010 were matched to DMS data according to coordinates and sampling date. We compared PMEL in situ Chl *a* to SeaWiFS Chl *a*, which are well correlated on logarithmic scale (R^2 = 0.64) with a slope of 0.67 and an intercept of -0.06, $[\log(Chl_{SeaWiFS}) = 0.67\log(Chl_{in-situ}) - 0.01]$, which means that on logarithmic scale SeaWiFS Chl-*a* concentrations are on average ~30% lower than those of in situ Chl-*a* concentrations. This is possibly because SeaWiFS Chl *a* is calibrated based on HPLC determined Chl *a* (Morel et al., 2007), which on average is ~40% lower than that determined using Fluorometric method (Sathyendranath et al., 2009). Unfortunately, there is no flag in the database showing how Chl *a* was determined. For consistency, we use only Chl-*a* data retrieved from SeaWiFS in the following multilinear and network models."

3. Inaccurate reasoning regarding the utility of data binning for the purpose of calculating monthly climatologies. What is the value of using raw (non-binned) measurements if (1) most of them are matched to climatological fields of the predictor variables? and (2) the final purpose is calculating a monthly climatology at coarser spatial resolution, which by definition aims at smoothing out interannual and small-scale variability? For example, to what extent the introduction of more than 10,000 new measurements, taken at high resolution in a relatively small region that was already quite well documented (temperate NW Atlantic), adds relevant information when it comes to computing monthly climatologies? Wouldn't it be more appropriate to bin all measurements beforehand to the coarsest resolution at which predictor variables are available, and then train the ANN? The authors should treat these issues more accurately and provide evidence for the advantages (if any) of using raw DMS data (including high-resolution transects), e.g. comparing statistics for ANN trained on raw vs. binned data. This being said, I agree that capturing the weight of extremes, ie the non-normal statistical distribution of DMS, is important (Galí et al., 2018).

Response: Since our initial submission, the PMEL database dramatically expanded. Now there are a total of 86,785 valid DMS measurements (concentration greater than 0.1 nM and less than 100 nM according to your instructions), that is 71% larger than the number of data we initially used (51,161). For the expanded data set, ~93% of DMS are accompanied with in situ SST measurements, ~81% are accompanied with in situ salinity measurements. More importantly, each data point has their unique location and sampling time signatures. As shown in the following figure, sampling time (date) and location information is a strong DMS predictor, which together can decrease DMS root mean square error to 0.64 (on natural logarithm scale). Adding other climatological predictors can further improve the model performance.

The NAAMES dataset has 6,786 valid data points, which are ~7% of the total data points (93,571 = 86785+6786). All data are accompanied with in situ Chl *a*, SST, and SAL measurements. For parameters without in situ measurements, high resolution data are used to match DMS measurements, 0.0417° ×0.0417° for PAR, 0.5° ×0.5° for MLD, and 1° ×1° for NO3, which ensures most of DMS have a set of unique predictors. As shown in Table 1, merging NAAMES data with PMEL data does not significantly change the statistic.

Moreover, binning the data will reduce data variance, which has been demonstrated by Derevianko et al. (2009). The objective of this study is to train an ANN with as much data as possible, and let the ANN do any fitting. The statement "the final purpose is calculating a monthly climatology at coarser spatial resolution" is only partially true. The model can be applied to coarse resolution predictor fields, but also to very fine resolution predictor fields. For example, we have applied the network to fine resolution NAAMES fields for comparison with in situ DMS measurements (Bell et al., in preparation).

Lastly, binning data will necessarily result in loss of information. A great amount of information is associated with sampling location and date as shown in the following figure (Fig. 2a in MS). By binning the data into monthly 1°× 1° grid, the number of data points decreases significantly from 82,996 to only 9,018; sampling date feature (365) will be average to 12 months, and coordination combinations will be averaged from 87,332 × 87,332 to 180° × 360°, which represents a substantial loss of information. For ANN models, using less data points can lead to overfitting (See Fig. 2b).

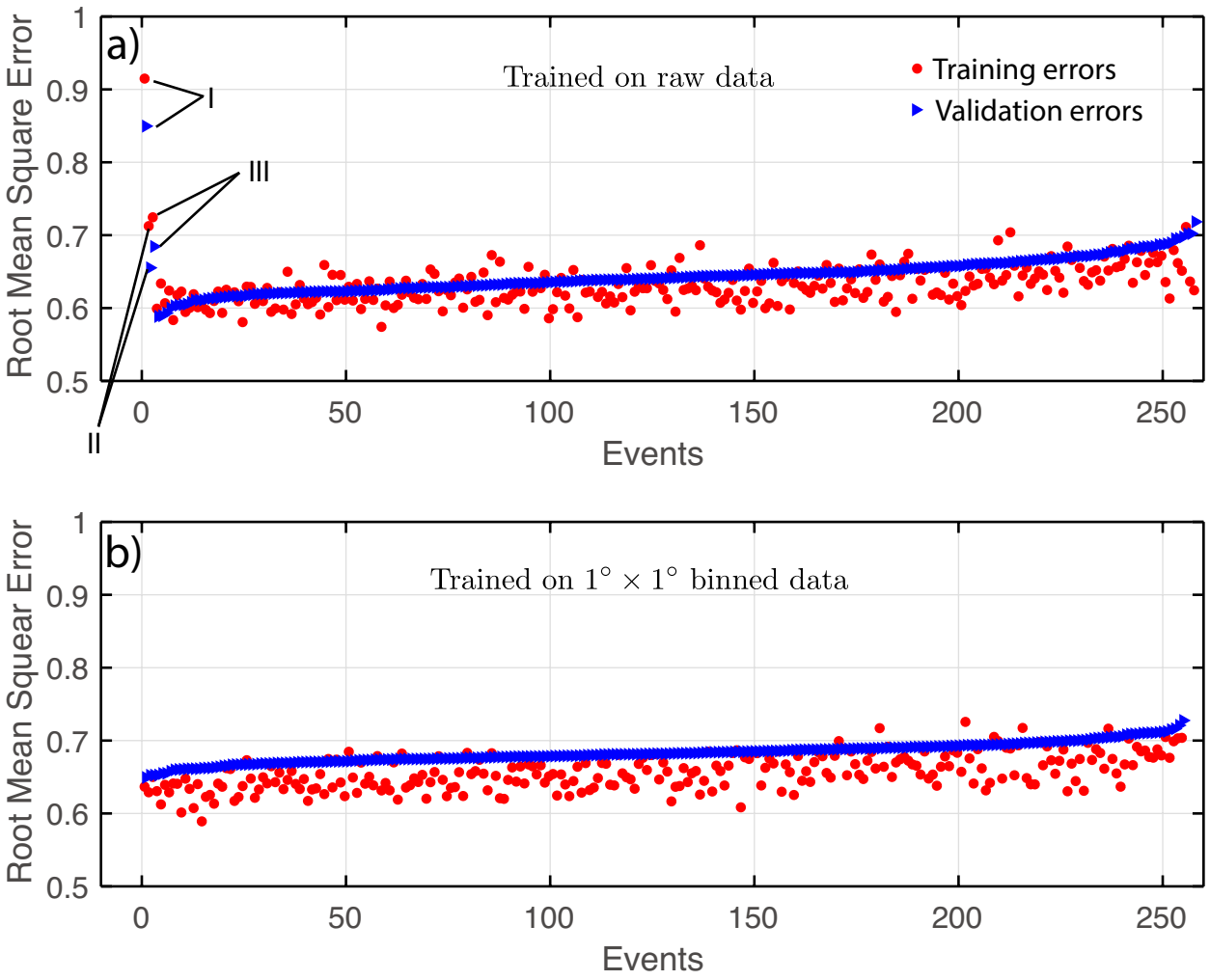

Fig. 2 Parameter sensitivity tests on raw and binned data. (a) Root mean square error on logarithmic scale for the model trained using raw data; (b) Root mean square error on logarithmic scale for the model trained using binned data. The time and location parameters are tested separately without combining with environmental parameters as shown in the upper panel, (I) with only location parameters; (II) with location and day of year parameters; and (III) with location, day of year, and time of day parameters. The model with three location parameters (I) has a root mean square error on natural logarithmic scale of ~0.83, which decreases to ~0.65 by adding sampling day of year parameters (II), however, increases to ~0.67 by adding sampling time parameters (III). We, therefore, do not include sampling time parameters in the following tests. We tested every possible combination of the eight parameters (PAR, MLD, SST, SAL, Chl a, DIP, DIN, and SiO), which in total are 255 tests.

4. Limited discussion of the advantages of the ANN approach, especially in regions that are challenging for prognostic and empirical models. For example, the ANN method does not outperform the gridded climatology (Lana et al., 2011) in the subarctic Northeast

Pacific in August and September, when DMS concentrations are much higher than what one would expect based on global-scale relationships. If the ANN does not outperform the (admittedly limited) objective interpolation approach in a region that contains data, how can we trust ANN predictions in regions with no DMS measurements? An analysis comparing seasonal DMS patterns across different biogeochemical regimes (eg ANN vs. objective interpolation and remote sensing algorithms) would be very welcome and would strengthen the arguments for adopting the ANN as a standard method to compute climatologies.

Response: We have added extensive discussion about the ANN approach on page 9 (l. 259 - l.274).

We also have added discussion emphasizing on comparison with previous models (P.12 l. 342 – l.364 and Figs. 3 and 4).

5. Misuse or inappropriate citation of some key references (e.g. Simó & Dachs 2002, Toole and Siegel 2004) and, more generally, omission of relevant references from the past 10-20 years. I think the view of marine DMS(P) cycling presented in this article is a bit outdated, especially regarding (1) upper-ocean DMS budgets and DMS turnover times due to biological processes, which ultimately control concentrations, and (2) the role of heterotrophic organisms and processes, which decouple DMS from phytoplankton abundance and taxonomy in much of the global ocean.

Response: Thank you for the suggestions. We have updated our reference and revised the biogeochemical description of DMS/DMSP.

In addition to the points above, I have to admit that approaches such as ANN leave me, as a reader, with the feeling I did not learn much about the global distribution of DMS and its controlling factors. This would not necessarily be a criticism as long as the choices the authors made to configure and tune the ANN were sufficiently justified, but I missed some depth of information in this regard. The ANN itself remains a mysterious black box to me and, even if the overall results look reasonable (with exceptions, as highlighted above), I am unable to appreciate whether Wang and coauthors made an optimal implementation of the ANN.

This is an important point. It motivated us to do more tests to help open the “black box” as discussed in the revised MS. (l.259 – l. 264).

“From traditional linear or multilinear models, one can easily figure out which parameter is a strong predictor and how a predictor influence the state variable (e.g. the correlation between DMSP and DMS). An ANN model is much more complex, it adjusts weights of each node that connect inputs and outputs, therefore, the relationship between inputs and outputs is subtle. That’s why ANN models are generally referred to as a "Black Box". In this study, we design experiments that help open this "Black Box" and reveal parameters that drive surface ocean DMS distributions.”

And please also read the subsequent discussion (l.264 – l.317).

I prompt the authors to address the formal, conceptual and technical criticisms made above. I honestly hope these constructive criticisms will improve the study and, more broadly, pave the way towards sensible implementation of AI techniques to compute DMS climatologies (which will likely become routine in the near future).

Thank you for your constructive comments.
I hope you will find that the paper has been greatly improved based on comments of you and two other anonymous reviewers.

Finally, note that in the specific comments below I will frequently refer to my own papers, simply because some of them are very relevant for the present study and, in some cases, the only ones available. Of course, the authors are free to decide what citations they incorporate. For all these reasons, and for the sake of transparency, I decided to sign the review.
We enjoyed reading your papers.

Specific comments
Abstract
Please reshape taking into account the main criticisms, especially concerning the amount of variance captured when using raw or binned data (general point 3). Raw DMS data variance could be biased towards high-resolution data representative of small-scale variability if no homogenization of the spatial-temporal scales covered by the measurements is applied (the PMEL database consists mostly of coarse resolution data).
We have taken most of your advice, and the manuscript has been thoroughly revised. More tests have been conducted, and our results do not significantly change, which means our method is robust. The abstract has been edited accordingly.

Introduction

L19: OK, but the approach proposed here does not reveal the factors controlling DMS variability. Rephrasing suggested.

We have added more tests to figure out how the environmental changes can influence surface ocean DMS distribution. With this revision, we believe that the sentence here is appropriate.

L24: If the authors insist on the mechanistic point of view (not sure is the right line of thought in this paper), I suggest adding “process rate measurements” here. They hamper predictive models even more strongly than limited observations of DMS concentration.

The term “process rate measurements” has been added.
Meanwhile, we did more experiments to exploit the mechanisms.

L30: To the best of my knowledge, the term “summer paradox” was coined by Simó & Pedrós-Alió 1999 (Nature), so I suggest crediting them for it.

Corresponding reference has been added. Thank you.

L49: Relevant citations here are Le Clainche et al. 2010 (GBC) (S cycling model inter-comparison seeking to understand the processes responsible for the summer paradox) and Tesdal et al. 2016 (Env Chem) (the most extensive comparison among gridded climatologies, empirical and prognostic models published so far, to my knowledge).

Corresponding references have been added. Thank you.

L61: DMS is produced by some marine algae, some bacteria, and mostly as a result of food web interactions (Kiene et al. 2000; Simó et al. 2001, Stefels et al. 2007, Curson et al. 2011, Moran et al. 2012, etc.). Please nuance and refine.

Corresponding references have been added, and now the text is as follows,

“The precursor of DMS, DMSP, is mainly produced by marine algae (e.g. Kiene et al., 2000; Curson et al., 65 2011), and a small fraction of DMSP is transformed to DMS by marine algae and/or bacteria lyases (Simó, 2001; Stefels et al., 2007; Curson et al., 2011; Moran et al., 2012), and mostly as a result of food web interactions (Kiene et al., 2000; Simó, 2001).”

Methods

L73: Over 10,000 measurements came from NAAMES alone. Not binning the data might give too much weight to particular conditions sampled during NAAMES, at least in the multilinear regression.

The valid NAAMES data points are 6,939, which accounts for ~7% of the total data with expanded PMEL database. For the NAAMES data, we matched the DMS observation with super high-resolution satellite products, which ensures most of the data points have unique predictor combinations. We have added more results and discussion to argue why binning the data is not a good choice.

L75: The DMSPt, Chla, SST and SSS data in the PMEL database require some quality control. This was documented by Galí et al. 2015. Quality controlled datasets with stringent satellite match-up criteria (ie minimizing the use of climatological coarse resolution data), as well as a piece of code used to clean data, are publicly available on github: https://github.com/mgali/DMS-SAT_DATA_DEV_VAL.

Thank you for the useful tips, we followed your instruction to clean up the data. Specifically, we removed DMS concentration higher than 100 nM and lower than 0.1 nM, DMSPt concentration lower than 1 nM. We also removed data with salinity lower than 30 psu to focus on open ocean.

L75: Fluorometric Chl is on average around 40% higher than HPLC Chl (Sathyendranath et al., 2009), and the proportion sometimes varies quite a bit. Satellite Chl is validated against HPLC measurements (e.g. Morel et al., 2007).
Yes, Good point.
For Chl a data, we have added the following discussion (l.84 – 1.92).

"SeaWiFS Chl-*a* data (Level 3-binned, spatial resolution of 9.2 km) from December 1997 to March 2010 were matched to DMS data according to coordinates and sampling date. We compared PMEL in situ Chl *a* to SeaWiFS Chl *a*, which are well correlated on logarithmic scale (R^2 = 0.64) with a slope of 0.67 and an intercept of -0.06, $[\log(Chl_{SeaWiFS}) = 0.67\log(Chl_{in-situ}) - 0.01]$, which means that on logarithmic scale SeaWiFS Chl-*a* concentrations are on average ~30% lower than those of in situ Chl-*a* concentrations. This is possibly because SeaWiFS Chl *a* is calibrated based on HPLC determined Chl *a* (Morel et al., 2007), which on average is ~40% lower than that determined using Fluorometric method (Sathyendranath et al., 2009). Unfortunately, there is no flag in the database showing how Chl *a* was determined. For consistency, we use only Chl-*a* data retrieved from SeaWiFS in the following multilinear and network models."

L77: Less than 3.5 years do not make a good Chl climatology in many ocean areas in my experience. Data products covering much longer periods are available on NASA's ocean colour website. Please update datasets, and specify also what reprocessing was used.
Good point. We updated our climatology so that the current climatology used is from 1997-2010. The new climatology is Level 3-binned (L3BIN with spatial resolution of 9.2 km) from SeaWiFS.
The change of Chl a climatology does not significantly change our results.

L79: The more recent climatology of Holte et al. 2017 seems to outperform that of Schmidtko et al. 2013 in areas of deep winter convection (subpolar North Atlantic) or where deep mixing prevails (Southern Ocean circumpolar current). In some cases the differences are important. Please consider using the Holte et al. 2017 MLD climatology.
Good point. We updated our MLD climatology and used MIMMOC one in the revised model.
The change of MLD climatology does not significantly change our results.

L81: Please specify the nutrient datasets used.
Good point. We added descriptions about the nutrient data sources (l.98 – l.101).

L105-107: I see some contradiction here. Data extremes typically arising from nonlinear dynamics are often smoothed out when averaging data. Your predicted variable (DMS) retains full variability but predictor fields do not, because apart from SST they largely originate from monthly climatologies. How can meaningful nonlinear relationships be identified?
Good point, but it is partially true. Except for SST that has the most in situ observational data (81069 for PMEL data), in situ salinity was reported for ~74% of the DMS data. More importantly, every data point has their unique time-space signatures (5 parameters in the

model). We used high resolution Chl *a* (0.418), PAR (0.418), and MLD (0.5×0.5) climatologies for the PMEL data. For the NAMMS data, the Chl *a*, SST, and PAR have even higher resolution (0.0417×0.0417). The high-resolution data and unique time-space ensure that each data has a unique signature. See also our response to your general question No. 3.

L128-130: Are these parameters default ones, or tuned manually to achieve reasonable fits in this particular study?

These parameters are called hyper-parameters in machine learning language. We manually tuned the parameters to prevent over-/under- fitting the data (l.159 – l.163).

L137: Inclusion of time of day is interesting, although diel variability was not been mentioned earlier in the manuscript. Are hourly predictions useful for computing climatologies? Although DMS can oscillate on diel time scales (Galí et al., 2013; Royer et al., 2016) diel cycles do not seem to follow a fixed pattern, at least in low-latitude high-resolution datasets (e.g. Royer et al., 2015).
Good point, we retested the diurnal parameters (two time parameters). Adding them slightly worsen the performance of the model (Fig. 2a in the MS). We have added the following discussion (l.265 – l.274),

"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. However, it increases RMSE slightly (Fig. 2a). Galí et al. (2013c) studied diel cycle at the Mediterranean Sea and Sargasso Sea. Among their four experiments (three in the Mediterranean Sea and one in the Sargasso Sea) regular diel variation was observed at only one experiment in the Mediterranean Sea at summer season, with highest DMS values observed at midnight and lowest values at midday. In all the other experiments, diel variations for both DMS and DMSPt pools were small. Gross community DMS production during the daytime was two to three times higher than that in the nighttime, but the high DMS production was compensated by greater photochemical and microbial consumption (Galí et al., 2013c). The balance between DMS production and consumption appears to dampens DMS diel variation. This may explain why adding time parameters does not improve the ANN model's predictive ability."

Results
L170 and 177: Note Galí et al. (2018) reported an R2 of 0.42 (r of 0.65) between DMS and DMSPt using the same datasets with stricter quality control. Similarly for DMS vs. Chl: R2 of 0.20 (r of 0.45).

We followed your instructions to clean up the data. Our new R^2 value for DMS and in situ Chl *a* is 0.21 (n = 10,404 compared to 8141 in Gali et al. (2018)), for DMS and DMSPt is 0.41 (n = 4060, compared to 3637 in Gali et al. (2018)). Both pairs have larger data than those by Gali et al. (2018) for two reasons, 1) the expansion of the online database; 2) no averaging being done.

L172-176: These sentences look a bit contradictory and may need further elaboration (or else, can be removed). Is it straightforward or not to predict DMS from DMSPt? Are measurements sufficient or not? Regarding DMSP prediction, another relevant study is that by McParland & Levine (2018). Regarding the relationship between DMSPt and DMS in the global PMEL database, Galí et al. (2018) is a relevant reference.

We have removed the corresponding sentence. Meanwhile, we have added more discussion as follows (l.226 – l.229).

"McParland and Levine (2019) developed a mechanistic model that related intracellular DMSP concentration to environmental stress, and coupled the model with MIT ecosystem model (DARWIN) to estimate global ocean DMSP distribution. Galí et al. (2015) first applied a remote sensing algorithm to obtain a DMSP climatology, from which they predict DMS climatology through an empirical relationship with PAR (Galí et al., 2018)."

L179: The weaker relationship between DMS and Chl in the entire dataset probably results from the higher proportion of oligotrophic low latitude data (where DMS is anticorrelated to Chl over the seasonal cycle) compared to in situ Chl-DMS data pairs. The difference between in situ fluorometric Chl and satellite Chl may also play a role. Finally, note that the global PMEL DMS database is biased towards productive conditions (Galí et al. 2018; figure 7) which influences global DMS-Chl correlations. In summary, the correlation between DMS and Chl in global datasets is not really meaningful as it depends strongly on how evenly represented are the different ocean biomes.
Thank you for pointing this out.
We made the following corrections (l.314 – l.317).

"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 biogeochemical reduced sulfur cycle. As suggested by Simó (2001), not only can phytoplankton biomass, taxonomy, and activity influence DMS production, but so does food-web structure and dynamics. The inconsistent relationship may also explain the low ranking of Chl a in the models."

L187-190: This is incorrect. Dilution is not the main explanation for the negative relationship between MLD and DMS (as originally proposed by Aranami and Tsunogai, 2004). The main explanation is the different balance between biological DMS sources and sinks, as explained by Galí & Simó (2015). In the handful of studies that have made DMS budgets including the vertical mixing term, vertical DMS transport has never been found to dominate DMS budgets in the MLD over relevant (~daily) time scales. Check for example Bailey et al. 2008 (DSR), Herrmann et al. 2012 (CSR), Galí et al. 2013 (GBC), Royer et al. 2016 (Sci Rep), etc. DMS turnover in the surface layer due to vertical transport is generally an order of magnitude slower than biological turnover or biological + photochemical turnover, at least. Please correct.
Thank you for pointing this out.
We have clarified the reasoning in the text and also as follows (l.289 – l.296).

"MLD is another important predictor. High DMS concentrations in the open ocean have been detected when the water column is most stratified (Simó and Pedrós-Alió, 1999). The authors proposed that a stratified (high light) environment nourishes strong DMSP producers, or that phytoplankton cellular DMSP quota increases in such an environment. High conversion rates from DMSP to DMS in stratified waters is another reason for high DMS concentrations when MLD is shallow. Meanwhile, the biological DMS consumption rate decreases in oligotrophic oceans (Galí and Simó, 2015). A dilution model was also proposed to explain the anti-correlation between DMS concentration and MLD (Aranami and Tsunogai, 2004). The authors proposed that mixed layer deepening entrains water with little or no DMS into surface waters and dilutes surface DMS concentrations, but recent studies have shown that DMS loss rate via vertical mixing is orders of magnitude lower than production/consumption rates (e.g. Galí et al., 2013c; Royer et al., 2016)."

L191-199: The strongest evidence for light-driven DMS production in natural plankton assemblages comes from recent work by myself and colleagues (Galí et al. 2013a, b, c; Royer et al. 2016). Evidence for light-driven DMS production in Toole et al. (2006) (otherwise, a great piece of work!) is indirect as that study focused on DMS removal processes.
We have updated the reference. Thank you.

Section 3.2: see general comment 3 on data binning.
Please refer to our response to your comments #2.

L201-204: It is unclear here if the authors made the appropriate comparisons with Simó & Dachs 2002 and Vallina & Simó 2007 empirical models. Was DMS compared with surface PAR or with the solar radiation dose in the mixed layer as done by Vallina and Simó 2007? Similarly, did the authors correctly apply the double algorithm used by Simó and Dachs 2002, where different equations are used depending on the value of Chl/MLD? Or just computed a single regression of DMS against Chl/MLD?
Yes, we used exactly the same model, and have made this clear in the text as follows (l.319 – l.329),

"Simó and Dachs (2002) obtained high R^2 values between DMS concentration and the ratio of Chl a to MLD (Chl/MLD) when Chl/MLD is greater than or equal to 0.02, and between DMS concentration and ln(MLD) when Chl/MLD is less than 0.02. We tried exactly the same model on raw PMEL data with in situ Chl-a measurements and climatological MLD, and found that both correlations between DMS and Chl/MLD (n = 4,921, R^2 =~ 0 .1) and between DMS and ln(MLD) (n = 5,978, R^2 =~ 0) are statistically insignificant. To reduce interannual variability, we binned in situ Chl a and DMS into monthly 1° × 1° grid, and retested the above model. We found that the correlations are still statistically insignificant. (R^2=~ 0)

Vallina and Simó (2007) reported an R2 of 0.95 (n=14) between DMS concentration and SRD. We applied the same linear regressions on both raw data and monthly 1° × 1° data, and found

no significant correlations between DMS and SRD as calculated according to Vallina and Simó (2007):,

$$SRD = SI \cdot \frac{1}{Kd490 \cdot MLD}(1 - \exp(-Kd490 \cdot MLD),$$

where SI is solar insolation (W m-2); Monthly SI data are from a CESM simulation (Wang et al., 2019)."

Section 3.3: Methods section mentions 8 initial variables (PAR, MLD, Chl a, SSS, SST, DIN, DIP, and SiO), but, what predictor variables were included in the final multilinear regression model (MLR)? Does the R2 of the MLR refer to linear of log space?

The Choice of parameter combinations are based on BIC criterion.
Yes, R^2 value is for natural log space. We have made this clear in the revised paper.

Section 3.4: Since this subsection describes the main technical innovation of this paper, a deeper explanation of why the ANN gives these results would be very welcome. See general comment 4.

Yes, we have extensively discussed the ANN model on pages 9.

Section 3.5: For the authors' information, global DMS fields produced with the remote sensing algorithm of Galí et al. 2018, as well as the algorithms of Simó & Dachs 2002 and Vallina & Simó 2007, are available in this repository: https://doi.org/10.5281/zenodo.2558511. Corresponding Matlab and R codes are available on a linked github repository: https://github.com/mgali/DMS-SAT_ALGORITHM.

Good resources. We have plotted the data together with our new prediction (Fig.3 and Fig. 4).

L235: These references are not appropriate here. Please cite studies that actually documented DMS(P) dynamics in subpolar or polar blooms of coccolithophores or Phaeocystis.

Good point, we updated the discussion as follows (l.371 – l.376),

"The summertime high DMS concentration at high latitudes is believed to be linked 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). High DMS concentrations at high latitudes have also been observed to accompany blooms of coccolithophores and Phaeocystis, which are strong DMSP producers (Neukermans et al., 2018; Wang et al., 2015). The shoaling of mixed layer depth in summer provides favorable conditions, i.e. stable and warm, with adequate irradiation for coccolithophores and Phaeocystis growth (Galí et al., 2019)"

L247-248: I do not find reasonable that DMS decreases below 0.1 nM in a subtropical gyre in winter. By examining the maps in Fig. 3 I would say ANN DMS is mostly between 0.1 and 0.5 nM, which still looks a bit low but more realistic according to my experience. DMS concentrations lower than 0.1 nM are extremely rare in both the PMEL database and in global estimates made with empirical algorithms (see Galí et al. 2018 figure 7).

In the revised MS, we rerun the model, and now there is not region with DMS concentration lower than 0.1 nM.

L250-254: Please check Galí and Simó 2015 (GBC) for a mechanistic explanation of the summer paradox.
We have changed the wording as follows (l.389 – l.393),

"Fig. 6 compares monthly mean Chl-a concentrations to DMS concentrations in N. and S. hemisphere gyres. The concentrations are normalized to the range of 0 to 1. It is clear that Chl a and DMS are anti-correlated, DMS concentration peaks at summer season when Ch-a concentration is generally low. This phenomenon is previously termed as "summer DMS paradox" (Simó and Pedrós-Alió, 1999). This pattern is more apparent in the S. hemisphere gyres, because the terrestrial influence is smaller in the S. hemisphere than in the N. hemisphere."
In general, I suggest making a figure showing the climatological seasonal cycles in different ocean biomes or regions, to better support the description made in the text.

Good point.
We plotted monthly mean Chl-*a* and DMS concentrations for N. and S. hemisphere gyres as shown in the following figures. It is clear that DMS and Chl *a* are anti-correlated in the gyres. DMS peaks in the summer when Chl *a* is at annual minimum.

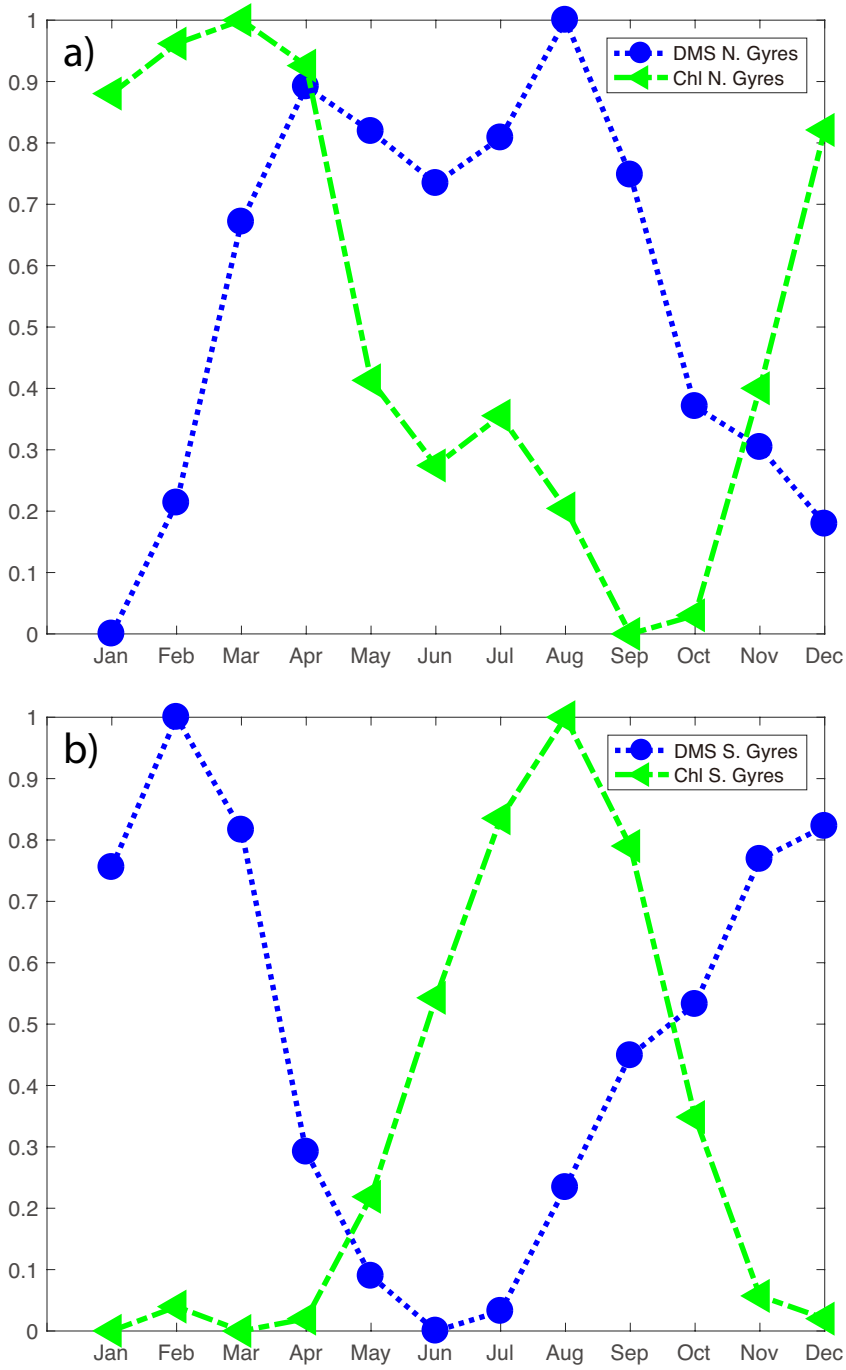

Figure 8. Distributions of monthly mean DMS and Chl-*a* concentrations for N. and S. hemisphere gyres. The gyres are defined as regions between 30◦ and equator where annually mean DIP concentration is below 0.2 µM. Monthly mean concentrations are normalized to the range of 0 to 1.

L273-293: Here I strongly suggest citing Tesdal et al. 2016.
Reference has been cited, and more discussion has been added.

Figures
Figure 3 and 5: I suggest using a color scale with different colors to help readers appreciate concentration patterns.

Thank you for this suggestion, we have updated the colormaps for both figures.

Figure 6: I strongly recommend splitting results into northern and southern hemisphere given the strong seasonality of DMS (also wind speed and SST), which results in opposed seasonal patterns.

Yes, another good suggestion, we updated the figure accordingly.

Minor corrections
L131: What does “epochs” mean in this context? Please use synonym for readers that are not expert in ANN or similar techniques.

one **epoch** = one forward pass and one backward pass of *all* the training examples

In terms of artificial neural networks, an epoch refers to one cycle through the full training dataset. Usually, training a neural network takes more than a few epochs. In other words, if we feed a neural network the training data for more than one epoch in different patterns, we hope for a better generalization when given a new "unseen" input (test data).

We have added more explaination in the revised MS (l.167 – l.169),
An epoch consists of one forward pass and one backward pass of all the training examples.

L218: The “tracer-tracer” term is not needed here (quite specific to bgc modelling).

Thank you. We have removed the term.

Reviewer references (only if not cited by the authors)
Aranami, K., & Tsunogai, S. (2004). Seasonal and regional comparison of oceanic and atmospheric dimethylsulfide in the northern North Pacific: Dilution effects on its concentration during winter. Journal of Geophysical Research: Atmospheres,109(D12).
Bailey, K. E., Toole, D. A., Blomquist, B., Najjar, R. G., Huebert, B., Kieber, D. J., ... & Del Valle, D. A. (2008). Dimethylsulfide production in Sargasso Sea eddies. Deep Sea Research Part II: Topical Studies in Oceanography,55(10-13), 1491-1504.
Curson, A. R., Todd, J. D., Sullivan, M. J., & Johnston, A. W. (2011). Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. Nature Reviews Microbiology, 9(12), 849-859.
Galí, M., Ruiz-González, C., Lefort, T., Gasol, J. M., Cardelús, C., Romera-Castillo, C., & Simó, R. (2013). Spectral irradiance dependence of sunlight effects on plankton dimethylsulfide production. Limnology and oceanography, 58(2), 489-504.

Galí, M., Simó, R., Vila-Costa, M., Ruiz-González, C., Gasol, J. M., & Matrai, P. (2013). Diel patterns of oceanic dimethylsulfide (DMS) cycling: Microbial and physical drivers. Global Biogeochemical Cycles, 27(3), 620-636.
Galí, M., Simó, R., Pérez, G., Ruiz Gonzalez, C., Sarmento, H., Royer, S. J., ... & Gasol, J. M. (2013). Differential response of planktonic primary, bacterial, and dimethylsulfide production rates to static vs. dynamic light exposure in upper mixed-layer summer sea waters.
Galí, M., & Simó, R. (2015). A meta-analysis of oceanic DMS and DMSP cycling processes: Disentangling the summer paradox. Global Biogeochemical Cycles,29(4), 496-515.
Herrmann, M., Najjar, R. G., Neeley, A. R., Vila-Costa, M., Dacey, J. W., DiTullio, G. R., ... & Vernet, M. (2012). Diagnostic modeling of dimethylsulfide production in coastal water west of the Antarctic Peninsula. Continental Shelf Research,32, 96-109.
Holte, J., Talley, L. D., Gilson, J., & Roemmich, D. (2017). An Argo mixed layer climatology and database. Geophysical Research Letters, 44(11), 5618-5626.
Kiene, R. P., Linn, L. J., & Bruton, J. A. (2000). New and important roles for DMSP in marine microbial communities. Journal of Sea Research, 43(3-4), 209-224.
Le Clainche, Y., Vézina, A., Levasseur, M., Cropp, R. A., Gunson, J. R., Vallina, S. M., ... & Bopp, L. (2010). A first appraisal of prognostic ocean DMS models and prospects for their use in climate models. Global biogeochemical cycles, 24(3).
Moran, M. A., Reisch, C. R., Kiene, R. P., & Whitman, W. B. (2012). Genomic insights into bacterial DMSP transformations. Annual review of marine science, 4, 523-542.
Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., & Franz, B. A. (2007). Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. Remote Sensing of Environment,111(1), 69-88.
Royer, S. J., Mahajan, A. S., Galí, M., Saltzman, E., & Simó, R. (2015). Small-scale variability patterns of DMS and phytoplankton in surface waters of the tropical and subtropical Atlantic, Indian, and Pacific Oceans. Geophysical Research Letters, 42(2), 475-483.
Royer, S. J., Galí, M., Mahajan, A. S., Ross, O. N., Pérez, G. L., Saltzman, E. S., & Simó, R. (2016). A high-resolution time-depth view of dimethylsulphide cycling in the surface sea. Scientific reports,6, 32325.
Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., ... & Platt, T. (2009). Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. Marine Ecology Progress Series,383,73-84.
Simó, R., & Pedrós-Alió, C. (1999). Role of vertical mixing in controlling the oceanic production of dimethyl sulphide. Nature,402(6760), 396-399.
Tesdal, J. E., Christian, J. R., Monahan, A. H., & von Salzen, K. (2016). Evaluation of diverse approaches for estimating sea-surface DMS concentration and air–sea exchange at global scale. Environmental Chemistry,13(2), 390-412.