Global ocean dimethyl sulfide climatology estimated from observations and an artificial neural network

Wei-Lei Wang[1], Guisheng Song[2], François Primeau[1], Eric Saltzman[1,3], Thomas Bell[1,4], and J. Keith Moore[1]

Corresponding author: Wei-Lei Wang, Department of Earth System Science, University of California at Irvine, Irvine, California, USA. (weilei.wang@gmail.com)

[1]Department of Earth System Science, University of California at Irvine, Irvine, California, USA

[2]School of Marine Science and Technology, Tianjin University, Tianjin, 300072, China

[3]Department of Chemistry, University of California at Irvine, Irvine, California, USA

[4]Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, UK

Abstract.

Marine dimethyl sulfide (DMS) is important to climate due to the ability of DMS to alter Earth's radiation budget. However, a knowledge of the global-scale distribution, seasonal variability, and sea-to-air flux of DMS is needed in order to understand the factors controlling surface ocean DMS and its impact on climate. Here we examine the use of an artificial neural network (ANN) to extrapolate available DMS measurements to the global ocean and produce a global climatology with monthly temporal resolution. A global database of 57,810 ship-based DMS measurements in surface waters was used along with a suite of environmental parameters consisting of lat-lon coordinates, time-of-day, time-of-year, solar radiation, mixed layer depth, sea surface temperature, salinity, nitrate, phosphate, silicate, and oxygen. Linear regressions of DMS against the environmental parameters show that on a global scale mixed layer depth and solar radiation are the strongest predictors of DMS, however, they capture 14% and 12% of the raw DMS data variance, respectively. The multi-linear regression can capture more (29%) of the raw data variance, but strongly underestimates high DMS concentrations. In contrast, the ANN captures ∼61% of the raw data variance in our database. Like prior climatologies our results show a strong seasonal cycle in DMS concentration and sea-to-air flux. The highest concentrations (fluxes) occur in the high-latitude oceans during the summer. We estimate a lower global sea-to-air DMS flux (17.90±0.34 Tg S yr^{-1}) than the prior estimate based on

a map interpolation method when the same gas transfer velocity parameterization is used.

Key points:

- A trained artificial neural network model captures DMS variance much better than linear/multilinear models do;
- Surface water DMS concentration has strong seasonal variability with high concentration in summer and low in winter;
- Annually, the ocean emits 17.17 to 17.90 Tg S to the atmosphere, which is lower than previous estimates.

1. Introduction

Dimethylsulfide emitted from the surface ocean is the major precursor for aerosol sulfate in the marine atmosphere. These aerosols play a significant role in the climate system both directly, through aerosol radiative effects and indirectly, through their role as cloud condensation nuclei and influence on cloud radiative properties [*Andreae and Rosenfeld*, 2008]. Assessing the impact of DMS on global climate requires an understanding of the DMS distribution and the factors controlling variability on a variety of spatial and temporal scales. Dimethylsulfide is produced in surface waters, mainly via enzymatic cleavage of the biogenic compound dimethyl sulfoniopropionate (DMSP; [e.g. *Stefels et al.*, 2007]). The abundance of DMS in surface waters is a function of numerous factors controlling production, loss rates and pathways of both DMSP and DMS [*Simó*, 2001; *Toole and Siegel*, 2004]. Developing mechanistic and predictive models of surface ocean DMS is challenging due to limitations of the existing observational database.

Given the biogenic origin of DMS, early efforts focused on the relationship between DMS and Chl a (a proxy for biomass). Positive correlations between DMS and Chl a have been reported on basin scales [e.g. *Andreae and Barnard*, 1984; *Yang et al.*, 1999]. However, this positive correlation disappears when more data are used. *Kettle et al.* [1999] found no significant relationship between DMS and Chl *a* based on the global DMS data set available at the time. The weak relationship may be caused by the so-called "summer DMS paradox", which describes a phenomenon where a maximum DMS concentration is commonly detected in low latitude waters when phytoplankton biomass is low [*Vallina et al.*, 2008; *Toole and Siegel*, 2004]. *Kettle et al.* [1999] also tested linear regression models

on a compilation of data, including sea surface salinity and temperature, nitrate, silicate, phosphate, and Chl *a*, and concluded that no simple algorithm based on linear regression could be used to create monthly DMS fields, indicating that more complex mechanisms control surface DMS concentrations.

Simó and Dachs [2002] achieved a strong relationship between heavily binned and averaged DMS data and mixed layer depth (MLD). *Vallina and Simó* [2007] found a linear relationship between DMS concentration and solar radiation dose (SRD) in the coastal northwest Mediterranean. They conducted a global scale study by dividing the ocean into 10° latitude by 20° longitude boxes and correlating SRD and the box averaged DMS concentration. A strong linear relationship was detected in this filtered dataset. *Derevianko et al.* [2009] reexamined the relationship between SRD/MLD and DMS concentration by using 1° by 1° bins, and found that only a small fraction (14%) of the DMS variance was captured by a linear model based on SRD or MLD. These authors also pointed out that the previously identified strong relationship between MLD/SRD and DMS "result from the reduction in the total variance in the data due to binning" [*Derevianko et al.*, 2009].

Prognostic models have also been used to obtain climatological DMS distributions. In these models, phytoplankton are divided into different groups based on their ability to produce DMS. For example, diatoms produce less DMS than *Coccolithophores* and *Phaeocystis* [e.g. *Bopp et al.*, 2003; *Vogt et al.*, 2010; *Gypens et al.*, 2014]. *Elliott* [2009] implicitly incorporated *Phaeocystis* in a model by assuming that DMS yields are simply related to temperature. The work of *Wang et al.* [2015] explicitly incorporated *Phaeocystis* into the Biogeochemical Elemental Cycling (BEC) model and included DMSP production from each phytoplankton group, along with DMS leakage pathways from algal cells, (grazing,

lysis, and exudation). Despite this level of modeling detail, there remain large discrepancies between the model simulations and in-situ measurements [*Vogt et al.*, 2010; *Wang et al.*, 2015].

The DMS climatologies used in most climate models were obtained by extrapolating observed DMS to the global ocean using objective analysis schemes [*Kettle et al.*, 1999; *Lana et al.*, 2011]. In those climatologies, observational data were first binned and averaged into 1° by 1° grid squares, which were then grouped into 57 static biogeographic provinces according to *Longhurst* [1998]. Many provinces lacked adequate data to create a reliable climatology (Fig. A1). In those situations, temporal interpolations were used to fill the blanks, and to create a first-guess map. Distance-weighted interpolation was used to fill the remaining gaps. Major gaps remain in the observational data base for wintertime in the high latitudes of both hemispheres.

Machine learning is being increasingly used in oceanography and geosicence studies [*Bergen et al.*, 2019]. For example, *Roshan and DeVries* [2017] applied an artificial neural network (ANN) to extrapolate observed dissolved organic carbon (DOC) to the global ocean. *Rafter et al.* [2019] used an ensemble of neural networks to study oceanic δ^{15}N distribution. ANNs have also been used to study DMS on regional scales [e.g. *Humphries et al.*, 2012]. The popularity of machine learning partially stems from one of its inherent advantages: it can detect non-linear relationships that traditional linear regression models are unable to capture. Since DMS is produced by marine algae, we expect that there exists a functional relationship between parameters controlling the growth of phytoplankton or species distribution. Therefore, the objective of the present paper is to explore possi-

ble relationships between DMS and environmental variables, with the goal of creating a monthly-resolved DMS climatology.

The paper is organized as follows. We begin by exploring the relationships between DMS concentration and various environmental parameters taken one at a time using linear regression. We then do a stepwise multilinear regression to create a reference model to which we compare our neural network model results. Lastly, we train an ANN using DMS measurements and environmental parameters. With the trained networks, we extrapolate the sparse measurements globally to obtain gridded fields of monthly DMS distributions and sea-to-air DMS fluxes.

2. Materials and Methods

2.1. Data sources and cleaning

Surface ocean DMS data were obtained from the Global Surface Seawater DMS database and from the North Atlantic Aerosol and Marine Ecosystems experiment [NAAMES, [*Behrenfeld et al.*, 2019] (Table A1) . In total, 58,100 measurements were used, compared to 47,313 used by *Lana et al.* [2011]. The Global Surface Seawater DMS database also includes some ancillary in-situ data, such as DMSP, Chl *a*, sea surface temperature (SST), and salinity (SSS). Such data were used if available. If not, monthly climatology data from other sources (Table A1) are used to fill the gaps. SeaWIFS Chl *a* data from December 1997 to March 2001 were averaged to create monthly climatologies. Photosynthetically available radiation (PAR) from January 2003 to March 2017 were averaged to get monthly PAR climatologies. In high latitude oceans, mixed layer depth climatologies from [*de Boyer Montégut et al.*, 2004; *Keerthi et al.*, 2013] contain missing data, which are filled with a monthly climatology from a simulation with the ocean component of the

Community Earth System Model (CESM) forced with a repeating thirty year cycle (1980-2009) of NCEP reanalysis datasets [*Wang et al.*, 2019]. Sea ice cover output was similarly averaged into a monthly climatology and was used for estimating air-sea gas exchange. Nutrient data (nitrate, phosphate, and silicate) were also included in the neural network analyses, since they can exert influence on phytoplankton distribution and thus influence DMS production [*Wang et al.*, 2015; *Archer et al.*, 2009]. The ancillary data are then matched with DMS data according to the time of year.

2.2. Linear regressions

Linear regression models are conducted on two sets of data to diagnose the predictive skill of each ancillary variable. As a first step, we restrict the regression model to data sites where both DMS and the predictor variable are simultaneously available. This selection process yields a total of 10,823 pairs for Chl *a* and DMS, 4,559 pairs of total DMSP (DMSPt) and DMS, 36,814 pairs of sea-surface salinity and DMS, and 43,142 pairs of sea-surface temperature and DMS, respectively. In a second step, the unmeasured predictors (i.e. MLD, PAR, Nitrate (DIN), Phosphate (DIP), and Silicate (SiO_4^{4-}), SST, SSA, and Chl *a*) are filled in using monthly climatology data from the previously cited sources. DMSPt is not included, because there is no available climatological dataset to fill the missing values.

To reduce the dynamic range, we log-transform the DMS, DMSPt, Chl *a*, MLD, DIP, DIN, and silicate data. We do not log-transform SST to avoid losing data with temperature below (equal to) zero. Data with salinity less than 10 are removed to eliminate near shore and estuary measurements to focus on the open ocean. The corresponding predictors are then standardized to their z-score, $Z \equiv (C - \overline{C})/\sigma$, where C is predictor's

concentration; $\overline{C}$ is the mean of the variables; and σ is standard deviation of the variables. Matlab's `polyfit` function is applied to each pair to fit a first degree polynomial, i.e. a linear regression.

2.3. Multilinear regression

We begin by applying a step-wise multi-linear regression model to the environmental data using Matlab's `stepwiselm` function. We consider a total of eight potential DMS predictors: PAR, MLD, Chl *a*, SSS, SST, DIN, DIP, and SiO. Bayesian Information Criterion (BIC) of 0.01 is used as a criterion for accepting or rejecting a predictor, which means that predictors are removed if they induce a BIC increase of more than 0.01.

2.4. Artificial Neural Network (ANN)

To assess the possibility that a non-linear model might provide better prediction, we train artificial neural networks (ANNs) using the `Keras` deep learning toolbox in Python. The ANN requires that the predictor fields be available for every DMS data point so we fill missing values in the environmental dataset with climatological data. We eliminate DMS measurements that are under CESM ice coverage, leaving us with 49,798 DMS measurements with a complete set of predictors. DMS concentration along with the eight environmental predictors (PAR, MLD, Chl *a*, SSS, DIN, DIP, and SiO) are log-transformed. The predictors' dynamic ranges are then constrained to the [0,1] interval using a minmax normalization, i.e. $C_{norm} \equiv (C - C_{min})/(C_{max} - C_{min})$, where C_{min} and C_{max} are the minimum and maximum values in the data C, respectively.

The in-situ sampling times (months and hours) were converted to periodic functions using sine and cosine functions to address the data continuity issue, such that in a diurnal

or seasonal cycle the start (0 hr or January) and the end (24 hr or December) of a cycle share the same properties, but would be numerically different. The coordinate space notations have a similar issue in the longitudinal direction. The conversions are conducted according to *Gade* [2010] and *Gregor et al.* [2017] as follows:

$$\begin{bmatrix} \mathrm{H1} \\ \mathrm{H2} \end{bmatrix} = \begin{bmatrix} \cos(\mathrm{hour}\frac{2\pi}{24}) \\ \sin(\mathrm{hour}\frac{2\pi}{24}) \end{bmatrix}, \tag{1}$$

$$\begin{bmatrix} \mathrm{M1} \\ \mathrm{M2} \end{bmatrix} = \begin{bmatrix} \cos(\mathrm{month}\frac{2\pi}{12}) \\ \sin(\mathrm{month}\frac{2\pi}{12}) \end{bmatrix}, \tag{2}$$

$$\begin{bmatrix} \mathrm{L1} \\ \mathrm{L2} \\ \mathrm{L3} \end{bmatrix} = \begin{bmatrix} \sin(\mathrm{lat}\frac{\pi}{180}) \\ \sin(\mathrm{lon}\frac{\pi}{180})\cos(\mathrm{lat}\frac{\pi}{180}) \\ \text{-}\cos(\mathrm{lon}\frac{\pi}{180})\cos(\mathrm{lat}\frac{\pi}{180}) \end{bmatrix}. \tag{3}$$

The dataset is then separated into training and external validating sets. Data from the fifteen two-degree-latitude bands (69°N−71°N, 59°N−61°N, 49°N−51°N, 39°N−41°N, 29°N−31°N, 19°N−21°N, 9°N−11°N, 1°N−1°S, 9°S−11°S, 19°S−21°S, 29°S−31°S, 39°S−41°S, 49°S−51°S, 59°S−61°S, 69°S−71°S) are left out for external validation (10,836 points). The remaining data (38,962 points) are used to train the neural network. The network has one input layer with 15 input nodes corresponding to the 15 predictors (8 environmental predictors and 7 time and coordination signatures), two dense hidden layers with 128 nodes each, and one output layer with one node corresponding to the predicted DMS concentration.

To avoid overfitting, we add two dropout layers with a dropout ratio of 30% after each hidden layer. We also apply a L2 kernel regularizer for each hidden layer with the regulation parameter value set to 0.001. When the network is trained, the training data set is internally split to two units, 80% for training and 20% for internal validation. We

monitor the internal validation mean squared error, and stop the training when there is no error deduction in 50 epochs. Only the best model with the lowest validation mean squared error is saved. We tested different network setups, the current setting achieves goodness of fit, but avoids overfitting. We train an assemble of networks based on ten random states (random seed of 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024 generated using the Python `tensorflow` package).

To obtain a monthly and annual DMS climatology, we interpolate the corresponding state variables (PAR, MLD, Chl *a*, SSS, SST, DIN, DIP, and SiO) onto a 1° by 1° grid. Coordinates and target months are transformed accordingly. We then apply the trained networks to obtain hourly DMS concentrations, which are averaged into daily and monthly means. Monthly results from 10 random states are then used to produce the final monthly climatology and to analyze uncertainties.

2.5. Sea-to-air flux

Air-sea gas transfer is estimated using the following bulk formula,

$$F = K_w(C_w - C_a/H), \tag{4}$$

where F is sea-to-air gas exchange flux, C_a and C_w are bulk air and bulk water gas concentrations, and K_w (cm/hr) is the overall gas transfer velocity, expressed in water side units [*Liss*, 1974]. K_w reflects the combined resistance to gas transfer on both sides of the interface, as follows:

$$1/K_w = 1/k_w + 1/(Hk_a)), \tag{5}$$

where H is the dimensionless (gas/liquid) Henry's law constant and k_a and k_w are gas transfer velocities in air and seawater. The surface ocean is strongly supersaturated in

DMS with respect to the overlying atmosphere, with $C_w \gg C_a$, simplifying the flux equation 4 to

$$F = K_w C_w, \tag{6}$$

For this study we used two parameterizations for K_w. The *Goddijn-Murphy et al.* [2012] parameterization (hereafter GM12) is based on regressions between satellite based wind-speed observations with shipboard in-situ measurements of DMS gas transfer velocities using eddy covariance. The GM12 parameterization for K_w normalized to a S_c number of 660 is

$$K_{w,660} = 2.1U_{10} - 2.8, \tag{7}$$

where U_{10} is wind speed (m/s) at 10 m above sea surface. We also utilized the *Nightingale et al.* [2000] (hereafter N00), which is based on shipboard ^{3}He/SF$_6$ dual tracer experiments. Their parameterization for water side only DMS gas transfer velocity at a Schmidt number of 660 ($\kappa_{w,660}$) is calculated as follows,

$$k_{w,660} = (0.222U_{10}^2 + 0.333U_{10})(Sc_{DMS}/600)^{-0.5}, \tag{8}$$

where Sc_{DMS} is calculated as a function of temperature after *Saltzman et al.* [1993]. A total transfer velocity is obtained from N00 as follows,

$$K_{w,660} = k_{w,660}(1 - \gamma_a), \tag{9}$$

where γ_a is atmospheric gradient fraction given by $\gamma_a = 1/(1+k_a/\alpha k_{w,660})$ [*McGillis et al.*, 2000]. Air side DMS transfer velocity is given as $k_a = 659U_{10}(M_{DMS}/M_{H_2O})^{-0.5}$, where M_{DMS} and M_{H_2O} are the molecular weights of DMS and water, respectively [*McGillis et al.*, 2000].

DMS fluxes were calculated using a satellite-based wind speed climatology (Table A1) and surface ocean DMS concentrations from the ANN results. Ice fraction data are from the CESM simulation monthly climatology. DMS fluxes from ice-covered regions are set to zero, although DMS concentration in or below sea ice is not necessarily zero.

3. Results and discussion

3.1. Linear regressions

The linear regression coefficients and R^2 values are summarized in Table 1. For the test using in-situ measurements, DMS and DMSPt show the strongest positive correlation with a R^2 value of 0.41. It is not surprising to find the strong relationship between total DMSP (DMSPt) and DMS, since DMS derives from the enzymatic cleavage of DMSP [*Stefels*, 2000; *Stefels et al.*, 2007]. However, it is difficult to infer global scale climatological DMS distributions from total DMSP measurements, because there are approximately 10-fold fewer measurements of total DMSP compared to DMS. On the other hand, since DMSP is directly produced by phytoplankton and has no sea-to-air gas exchange, it is relatively easy to parameterize in a biogeochemical model [*Galí et al.*, 2015] The strong relationship between DMS and DMSP point toward a potential way to model marine seawater DMS.

The second strongest predictor is in-situ Chl *a* ($R^2 = 0.19$). When missing Chl *a* data are filled in using the satellite-based climatological values, the correlation between Chl *a* and DMS becomes weaker ($R^2 = 0.07$). The weak relationship may be caused by greater variance in the larger dataset (41,146 vs 9,942). The positive correlation between Chl *a* and DMS can be explained by the fact that the precursor of DMS, namely DMSP, is biogenic. Such a relationship has been reported in basin scale studies [e.g. *Yang et al.*, 1999]. On the other hand, negative correlations between DMS and Chl *a*

have also been detected in coastal waters of the Mediterranean and in the Sargasso Sea [*Toole and Siegel*, 2004]. The low and inconsistent R^2 values 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, so do food-web structure and dynamics.

When tested against climatological data with gaps filled-in, MLD has the strongest negative correlation with DMS (R^2 = 0.14, n = 53,391), with a slope of -0.36. It has been reported that in the open ocean high DMS concentrations have been detected when the water column is most stratified [*Simó and Pedrós-Alió*, 1999]. When the mixed layer deepens, water with no or little DMS is entrained into the surface waters and dilutes surface DMS concentrations [*Derevianko et al.*, 2009]. Additionally, stratified wasters get more light which is good for phytoplankton growth and can stimulate DMS production due to enhanced oxidative stress.

Climatological PAR is the second strongest predictor (R^2 = 0.12, n = 54,683) of raw DMS data with a positive correlation. Physiologically, the correlation between PAR and DMS can be explained by two reasons. First, high radiation negatively influences the bacterial population and accompanying DMS consumption [*Toole et al.*, 2006]. Second, high radiation promotes DMS production by exerting oxidative stress on algal cells [*Toole et al.*, 2006; *Sunda et al.*, 2002]. Strong correlation between monthly binned and averaged solar radiation dose (SRD) and DMS concentration has been reported (R^2 = 0.94) at the Blanes Bay Microbial Observatory located in the coast of northwest Mediterranean [*Vallina and Simó*, 2007]. Correlations between DMS and the other parameters (SST, SSS, DIP, DIN, and SiO) are generally weak ($R^2 < 0.001$), indicating that although these

environmental parameters can influence biomass distribution in the ocean, their direct influence on DMS concentration is subtle.

3.2. Binned data versus raw data

Simó and Dachs [2002] obtained a high R^2 value between DMS concentration and the ratio of Chl *a* and MLD (Chl/MLD). *Vallina and Simó* [2007] reported an R^2 of 0.95 (n=14) between DMS concentration and SRD. We applied the same linear regressions on the raw data (not binned), and found that the correlation between solar radiation and DMS is weak, and the correlation between DMS and Chl/MLD is statistically insignificant (n = 9,942, $R^2 =\sim 0$). Compared to *Simó and Dachs* [2002] and *Vallina and Simó* [2007], we have significantly more data points. For example, in this study, there is a total of 9,942 DMS measurements accompanied with simultaneous Chl *a* measurements versus 2,385 data points used in *Simó and Dachs* [2002], and 54,683 (DMS, MLD) pairs in this study versus 26,400 in *Vallina and Simó* [2007]. Another noticeable difference between the current study and previous analyses is that both *Simó and Dachs* [2002] and *Vallina and Simó* [2007] binned the data into large longitude and latitude grids. By doing so, the raw data variance is greatly reduced. Additionally, both studies applied data cleaning processes to exclude "outliers". For example, *Simó and Dachs* deliberately got rid of "A few groups of data that did not fit the algorithm equation". In our analyses, we use all available raw data.

3.3. Multilinear regression

Fig. 1a compares a multilinear regression model prediction to observations ($R^2 = 0.28$). Generally, the multilinear model captures more of the variance than the univariate linear

regression model. However, it is apparent that the multilinear regression model significantly underestimates high DMS concentrations. The generally low correlation hinders the possibility of reliably extrapolating the model to the global ocean.

3.4. ANN

Fig. 1b displays the tracer-tracer correlation between DMS observations and ANN predictions. Compared to simple linear and multilinear regression models, ANN captures much more of the observed DMS variance ($R^2 = 0.61$, n= 49,798). Compared to previous extrapolations [*Kettle et al.*, 1999; *Lana et al.*, 2011], the ability of the ANN to build a nonlinear relationship between DMS and environmental predictors allows it to capture more of the variance. It also incorporates diurnal and seasonal signals present in the data. As a result, the extrapolation obtained from the ANN considers the relationship with geographical neighbors and also with temporal relationships.

3.5. DMS distributions

Fig. 3 displays monthly DMS concentration distributions predicted by the ANN. Generally, DMS concentrations in polar regions show strong seasonality. The highest DMS concentrations are in summer when light and temperature are ideal for primary production. For example, in austral summer, the Southern Ocean circumpolar regions display the highest DMS concentration (>10 nM). DMS concentration in the Scotia Sea and west coast of South America (30°S - 60°S) display the highest DMS concentration, which gradually decreases and falls below 0.5 nM in the following months when primary production is limited by light or low temperature. In boreal summer, DMS concentration in the Bering Sea and Greenland Sea can exceed 20 nM.

The summertime high DMS concentration at high latitudes is consistent with the hypothesis that phytoplankton use DMSP as a cryoprotectant [*Karsten et al.*, 1992]. It is found that the same phytoplankton (Antarctic macroalga) contains higher DMSP concentration in the polar regions than in the temperate regions [*Karsten et al.*, 1990]. High DMS concentrations at high latitudes have been observed to accompany blooms of *Coccolithophores* and *Phaeocystis*, which are prolific DMSP producers [*Iida et al.*, 2002; *Toole and Siegel*, 2004]. At the same time, the shoaling of mixed layer depth in summer can also help to maintain high surface DMS concentration.

Another interesting region is the Pacific equatorial upwelling region. Large-scale upwelling brings nutrient-rich waters to the surface, which nourish highly productive phytoplankton communities. Overall, the seasonality in the equatorial Pacific is weaker than that in polar regions, but there is still a clear seasonal pattern. In the period from December to February, the tongue with high DMS concentration ($\sim$3 nM) extends to the west Pacific Ocean reaching the east coast of Australia and the Philippine Sea. The tongue gradually retreats eastward in the following months. From September to November, the tongue is constrained to the eastern Pacific and DMS concentration falls to its lowest values ($<$2.0 nM). High DMS concentrations in the west Pacific ocean from November to February are also predicted by *Lana et al.* [2011].

The subtropical gyres show consistently low DMS concentrations and weak seasonal cycles throughout the year. In the southern hemisphere gyres, DMS concentrations are highest during austral summer, when the ocean is strongly stratified and local primary production is low. There are hot spots where DMS concentration exceeds 3 nM in December and February. DMS concentrations are generally low (falling below 1 nM) during

austral spring and winter seasons. In the period from April to September, DMS concentrations in the S. Atlantic gyre fall below 0.1 nM. In the northern hemisphere gyres, DMS concentrations are high during the boreal summer season. The concurrence of high DMS concentration with low phytoplankton biomass (Chl *a*) in summer is described as the "summer DMS paradox" [*Simó and Pedrós-Alió*, 1999]. A few mechanisms have been proposed to collectively explain the decoupling of DMS concentration and phytoplankton biomass: 1) gyres in summertime are dominated by phytoplankton species that produce relatively more DMSP, 2) ultraviolet radiation suppresses bacterial DMS consumption, 3) direct DMS production by phytoplankton relax UVR-induced oxidative stress, 4) solar-radiation stimulates DMS exudation.

The Indian Ocean as a whole shows quite interesting seasonality. DMS concentration is high from December to February, and gradually decreases in the following months, and reaches minimum in September. Our prediction in the Indian subtropical gyre (ISSG) is quite consistent with the result reported by *Lana et al.* [2011]. However, our highest concentration is lower than that reported by *Lana et al.* [2011] (4 nM compared to 6 nM). We predict a similar seasonal cycle in the equatorial Indian Ocean (MONS) compared to *Lana et al.* [2011]: DMS concentrations are higher in boreal summer than winter. The so-called "summer DMS paradox" is once again observed in these regions.

Monthly mean concentrations are plotted along with results from *Lana et al.* [2011] (Fig. 2). Our predictions display similar seasonal patterns but overall lower average concentrations compared to *Lana et al.* [2011]. The differences are the largest in the months of January, October, November, and December. To find out which parts of the ocean contribute to the difference, monthly difference maps are plotted in Fig. 4. Generally,

the differences are small where measurement data are present. For example, the Atlantic ocean was extensively surveyed in October (overall difference < 0.5 nM), whereas, the Indian and Pacific oceans have sparse data coverage and the largest differences. This is especially true in the months of January, October to December in the Indian ocean, and in the months of January, February, October to December in the western Pacific ocean. The results are not surprising, because our model matches observational data well, and *Lana et al.* [2011] model has the lowest bias when measurement data are available. However, when observational data are absent, our model builds on internal relationship between DMS and environmental parameters, and *Lana et al.* [2011] relies only on temporal and spatial relationships, which can be very weak due to the sparsity of the observational data (Fig. 4 and Fig. A1).

3.6. Sea-to-air flux

In this study, we computed monthly sea-to-air DMS fluxes using both the GM12 and N00 gas transfer velocity parameterizations (Fig. 5 and Fig. 6). These yield global DMS annual fluxes of 17.17±0.32 Tg S yr^{-1} and 17.90±0.34 Tg S yr^{-1}, respectively. The uncertainties (errorbars) are calculated according to DMS distributions from 10 ANN models based on different random seeds. These fluxes are on the lower end of *Lana et al.* [2011] (15.0 to 40.4 Tg S yr^{-1}) climatology, but exhibit similar patterns of geographic and seasonal variability (Table 2).

On an annual, zonal average basis, the DMS flux is highest in the tropical oceans and decreases towards the poles. However, our zonally averaged fluxes are generally lower than those of *Kettle et al.* [1999] and *Lana et al.* [2011]. To find out the possible reason(s) for our low flux estimates, we recalculated DMS fluxes based on the N00 flux parameterization

and DMS concentrations from *Lana et al.* [2011]. The newly calculated flux is 23.48 Tg S yr^{-1}, lower than the originally reported value of 28.1 Tg S yr^{-1} [*Lana et al.*, 2011], and indicates that the climatological fields of wind speeds, SST, and SSS used to compute gas transfer velocity contribute ∼45% of the flux difference. The difference between the DMS concentration fields from this study versus *Lana et al.* [2011] contribute the remaining ∼55% of the difference.

Geographically, in the high-latitude northern hemisphere, sea-to-air DMS fluxes are low in boreal winter, even though wind speeds are high. The DMS flux tends to increase in the proceeding months and reaches a maximum in boreal summer, despite the lower wind speeds (Fig. A2). The inverse relationship between wind speed and DMS flux indicates that the high DMS flux is mainly driven by high seawater DMS concentrations. Large sea-to-air DMS fluxes at high latitudes in austral summer are driven jointly by high DMS concentrations and high wind speeds (Fig. 5 and Fig. A2). Although DMS concentration is low in austral winter, sea-to-air DMS fluxes are intermediate owing to year round high wind speed in this region. The eastern tropical Pacific ocean displays a year-round intermediate sea-to-air DMS flux. This is mainly driven by the high DMS concentration in this region, since the wind speeds here are generally low (Fig. 5 and Fig. A2).

Fig. 6 displays integrated monthly global DMS fluxes based on the two piston velocity parameterizations. It is noticeable that DMS fluxes based on the GM12 parameterization are lower than those based on the N00 parameterization. Furthermore, DMS fluxes are highest in the winter months (Dec., Jan., and Feb.) and March; there is a another but smaller peak in the months of July and August because of northern hemisphere flux peaks. The boreal winter (austral summer) high fluxes are due to the flux from the southern

hemisphere oceans, which also have a large surface area. Although DMS fluxes are high in the high-latitude northern hemisphere summer, the sea surface area is much smaller than the southern hemisphere oceans. High DMS fluxes in the southern hemisphere have profound impact to the Earth's climate because there are less terrestrial and anthropogenic aerosol inputs compared to the northern hemisphere.

4. Conclusions

The artificial neural network (ANN) used in this study has some advantages compared to the prior methods used to develop DMS climatologies. Most importantly, the ANN utilizes available measurements to fill regions without DMS observations, using non-linear relationships trained in more data rich regions/seasons. By contrast, objective interpolation methods are spatial/temporal averages of sparse data with no underlying basis in environmental variability. As a result, the ANN approach captures significantly more of the raw data variance than simple linear/multilinear models. Simple models achieve comparable fits only after heavily binning the DMS observations [e.g. *Simó and Dachs*, 2002; *Galí et al.*, 2015; *Vallina and Simó*, 2007; *Galí et al.*, 2018]. The ANN is computationally more expensive than the linear/multilinear models, but considerably less expensive than prognostic biogeochemical models [e.g. *Vogt et al.*, 2010; *Wang and Moore*, 2011; *Wang et al.*, 2015]. The principle weakness of the ANN approach is that it does not easily provide scientific insight into the relationships between the parameters. Some insight could be gained by running sensitivity tests in which the response to perturbation of a single parameter is diagnosed.

The ANN approach is a useful tool for developing trace gas climatologies. It may also be useful as a means of assessing the sensitivity of DMS to past/future changes in climate by coupling the ANN to prognostic biogeochemical models. Caution is warranted in the interpretation of such efforts because there is as yet no basis for assessing whether the relationships obtained by training on contemporary measurements apply to the past or will hold the future. Such relationships could be investigated using paleoceanographic and ice core data [*Osman et al.*, 2019].

The annual sea-to-air DMS flux calculated in this study is slightly (∼23%) lower than the objective interpolation method of *Lana et al.* [2011] using the same sea-to-air gas exchange models. DMS concentrations from this study are similar to *Lana et al.* [2011] where measurements are abundant, so we infer that the difference is likely caused by positive bias in the objective interpolation method for data-sparse regions/seasons.

Acknowledgments. We thank the observational DMS community for making their measurements publicly available. J.K.M., F.P., and W.L.W. are supported by DOE Earth System Modeling program (DE-SC0016539). G.S is supported by the Natural Key Research and Development Program of China (2017YFC1404403). E.S.S. and T.G.B. are supported by the NASA North Atlantic Aerosols and Marine Ecosystems Study (NAAMES), which was funded through the NASA Earth Venture Suborbital program. (NNX#15AF31G). The data for DMS concentrations and sea-to-air flux are available at DOI: 10.5281/zenodo.3631875. The sources of ancillary data are listed in Table A1.

Competing Interests:

The authors declare that they have no competing financial interests.

References

Andreae, M., and D. Rosenfeld (2008), Aerosol–cloud–precipitation interactions. part 1. the nature and sources of cloud-active aerosols, *Earth-Science Reviews*, *89*(1-2), 13–41.

Andreae, M. O., and W. R. Barnard (1984), The marine chemistry of dimethylsulfide, *Marine Chemistry*, *14*(3), 267–279, doi:10.1016/0304-4203(84)90047-1.

Archer, S. D., D. G. Cummings, C. A. Llewellyn, and J. R. Fishwick (2009), Phytoplankton taxa, irradiance and nutrient availability determine the seasonal cycle of dmsp in temperate shelf seas, *Marine Ecology Progress Series*, *394*, 111–124.

Behrenfeld, M. J., R. H. Moore, C. A. Hostetler, J. Graff, P. Gaube, L. M. Russell, G. Chen, S. C. Doney, S. Giovannoni, H. Liu, et al. (2019), The north atlantic aerosol and marine ecosystem study (naames): Science motive and mission overview, *Frontiers in Marine Science*, *6*.

Bergen, K. J., P. A. Johnson, M. V. De Hoop, and G. C. Beroza (2019), Machine learning for data-driven discovery in solid Earth geoscience, *Science*, *363*(1299), doi:10.1126/science.aau0323.

Bopp, L., O. Aumont, S. Belviso, and P. Monfray (2003), Potential impact of climate change on marine dimethyl sulfide emissions, *Tellus, Series B: Chemical and Physical Meteorology*, *55*(1), 11–22, doi:10.1034/j.1600-0889.2003.042.x.

de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, *JJ. Geophys. Res.*, *109*(12), 1–20, doi:10.1029/2004JC002378.

Derevianko, G. J., C. Deutsch, and A. Hall (2009), On the relationship between ocean DMS and solar radiation, *Geophysical Research Letters*, *36*(17), 2–5, doi:

10.1029/2009GL039412.

Elliott, S. (2009), Dependence of DMS global sea-air flux distribution on transfer velocity and concentration field type, *J. Geophys. Res.*, *114*, G02,001, doi: 10.1029/2008JG000710.

Frouin, R., J. McPherson, K. Ueyoshi, and B. A. Franz (2012), A time series of photosynthetically available radiation at the ocean surface from seawifs and modis data, in *Remote Sensing of the Marine Environment II*, vol. 8525, p. 852519, International Society for Optics and Photonics.

Gade, K. (2010), A non-singular horizontal position representation, *Journal of Navigation*, *63*(3), 395–417.

Galí, M., E. Devred, M. Levasseur, S.-J. Royer, and M. Babin (2015), A remote sensing algorithm for planktonic dimethylsulfoniopropionate (dmsp) and an analysis of global patterns, *Remote Sensing of Environment*, *171*, 171–184.

Galí, M., M. Levasseur, E. Devred, R. Simó, and M. Babin (2018), Sea-surface dimethylsulfide (DMS) concentration from satellite data at global and regional scales, *Biogeosciences*, *15*(11), 3497–3519, doi:10.5194/bg-15-3497-2018.

Garcia, H. E., R. A. Locarnini, T. P. Boyer, J. I. Antonov, O. K. Baranova, M. M. Zweng, J. R. Reagan, and D. R. Johnson (2013), World Ocean Atlas 2013, Volume 4 : Dissolved Inorganic Nutrients (phosphate, nitrate, silicate), *NOAA Atlas NESDIS 76*, *4*(September), 25, doi:10.1182/blood-2011-06-357442.

Goddijn-Murphy, L., D. K. Woolf, and C. Marandino (2012), Space-based retrievals of air-sea gas transfer velocities using altimeters: Calibration for dimethyl sulfide, *J GEOPHYS RES: OCEANS*, *117*(C8).

Gregor, L., S. Kok, and P. M. Monteiro (2017), Empirical methods for the estimation of Southern Ocean CO2: Support vector and random forest regression, *Biogeosciences*, *14*(23), 5551–5569, doi:10.5194/bg-14-5551-2017.

Gypens, N., A. V. Borges, G. Speeckaert, and C. Lancelot (2014), The dimethylsulfide cycle in the eutrophied Southern North Sea: A model study integrating phytoplankton and bacterial processes, *PLoS ONE*, *9*(1), doi:10.1371/journal.pone.0085862.

Humphries, G. R., C. J. Deal, S. Elliott, and F. Huettmann (2012), Spatial predictions of sea surface dimethylsulfide concentrations in the high arctic, *Biogeochemistry*, *110*(1-3), 287–301.

Iida, T., S. Saitoh, T. Miyamura, M. Toratani, H. Fukushima, and N. Shiga (2002), Temporal and spatial variability of coccolithophore blooms in the eastern bering sea, 1998-2001, *Progress in Oceanography*, *55*(1-2), 165–175.

Karsten, U., C. Wiencke, and G. Kirst (1990), The ß-dimethylsulphoniopropionate (DMSP) content of macroalgae from antarctica and southern chile, *Bot Mar*, *33*, 143–146.

Karsten, U., C. Wiencke, and G. O. Kirst (1992), Dimethylsulphoniopropionate (DMSP) accumulation in green macioalgae from polar to temperate regions: interactive effects of light versus salinity and light versus temperature, *Polar Biology*, *12*(6-7), 603–607, doi:10.1007/BF00236983.

Keerthi, M. G., M. Lengaigne, J. Vialard, C. de Boyer Montégut, and P. M. Muraleedharan (2013), Interannual variability of the Tropical Indian Ocean mixed layer depth, *Climate Dynamics*, *40*, 743–759, doi:10.1007/s00382-012-1295-2.

Kettle, A., and M. Andreae (2000), Flux of dimethylsulfide from the oceans : A comparison of updated data sets and flux models, *Journal of Geophysical Research*, *105*, 26,793–26,808, doi:10.1029/2000JD900252.

Kettle, A. J., M. O. Andreae, D. Amouroux, T. W. Andreae, T. S. Bates, H. Berresheim, H. Bingemer, R. Boniforti, M. A. Curran, G. R. DiTullio, G. Helas, G. B. Jones, M. D. Keller, R. P. Kiene, C. Leek, M. Levasseur, G. Malin, M. Maspero, P. Matrai, A. R. McTaggart, N. Mihalopoulos, B. C. Nguyen, A. Novo, J. P. Putaud, S. Rapsomanikis, G. Roberts, G. Schebeske, S. Sharma, R. Simó, R. Staubes, S. Turner, and G. Uher (1999), A global database of sea surface dimethylsulfide (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month, *Global Biogeochem. Cycles*, *13*(2), 399–444, doi:10.1029/1999GB900004.

Lana, A., T. G. Bell, R. Simó, S. M. Vallina, J. Ballabrera-Poy, A. J. Kettle, J. Dachs, L. Bopp, E. S. Saltzman, J. Stefels, J. E. Johnson, and P. S. Liss (2011), An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean, *Global Biogeochem. Cycles*, *25*(1), doi:10.1029/2010GB003850.

Liss, P. S. (1974), Flux of gases across the air-sea interface, *Nature*, *247*, 181–184.

Longhurst, A. R. (1998), *Ecological geography of the sea*, Elsevier.

McGillis, W., J. Dacey, N. Frew, E. Bock, and R. Nelson (2000), Water-air flux of dimethylsulfide, *J GEOPHYS RES: OCEANS*, *105*(C1), 1187–1193.

NASA (2012), *SeaWinds on QuickSCAT Level 3 surface wind speed for climate model comparison*, Ver. 1. PO.DAAC, CA, USA, doi:https://doi.org/10.5067/QSSWS-CMIP1.

NASA (2018), Goddard space flight center, ocean ecology laboratory, ocean biology processing group, sea-viewing wide field-of-view sensor (seawifs) chlorophyll data, doi:

data/10.5067/ORBVIEW-2/SEAWIFS/L3M/CHL/2018.

Nightingale, P. D., G. Malin, C. S. Law, A. J. Watson, P. S. Liss, M. I. Liddicoat, J. Boutin, and R. C. Upstill-Goddard (2000), In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers, *Global Biogeochem. Cycles*, *14*(1), 373–387.

Osman, M. B., S. B. Das, L. D. Trusel, M. J. Evans, H. Fischer, M. M. Grieman, S. Kipfstuhl, J. R. McConnell, and E. S. Saltzman (2019), Industrial-era decline in subarctic atlantic productivity, *Nature*, *569*(7757), 551.

Rafter, P., A. Bagnell, D. Marconi, and T. Devries (2019), Global trends in marine nitrate N isotopes from observations and a neural network-based climatology, *Biogeosciences*, *16*(13), 2617–2633, doi:10.5194/bg-16-2617-2019.

Roshan, S., and T. DeVries (2017), Efficient dissolved organic carbon production and export in the oligotrophic ocean, *Nature communications*, *8*(1), 2036.

Saltzman, E., D. King, K. Holmen, and C. Leck (1993), Experimental determination of the diffusion coefficient of dimethylsulfide in water, *J GEOPHYS RES-OCEANS*, *98*(C9), 16,481–16,486.

Simó, R. (2001), Production of atmospheric sulfur by oceanic plankton: Biogeochemical, ecological and evolutionary links, *Trends in Ecology and Evolution*, *16*(6), 287–294, doi:10.1016/S0169-5347(01)02152-8.

Simó, R., and J. Dachs (2002), Global ocean emission of dimethylsulfide predicted from biogeophysical data, *Global Biogeochem. Cycles*, *16*(4), 26–1.

Simó, R., and C. Pedrós-Alió (1999), Short-term variability in the open ocean cycle of dimethylsulfide, *Global Biogeochem. Cycles*, *13*(4), 1173–1181.

Stefels, J. (2000), Physiological aspects of the production and conversion of DMSP in marine algae and higher plants, *Journal of Sea Research*, *43*(3-4), 183–197, doi: 10.1016/S1385-1101(00)00030-7.

Stefels, J., M. Steinke, S. Turner, G. Malin, and S. Belviso (2007), Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling, *Biogeochemistry*, *83*, 245–275.

Sunda, W. G., D. Kieber, and R. P. Kiene (2002), An antioxidant function of DMSP and DMS in marine algae, *Nature*, *418*, 317–320, doi:10.1038/nature00851.

Toole, D., D. Slezak, R. Kiene, D. Kieber, and D. Siegel (2006), Effects of solar radiation on dimethylsulfide cycling in the western atlantic ocean, *Deep Sea Research Part I: Oceanographic Research Papers*, *53*(1), 136–153.

Toole, D. A., and D. A. Siegel (2004), Light-driven cycling of dimethylsulfide (dms) in the sargasso sea: Closing the loop, *Geophysical Research Letters*, *31*(9).

Vallina, S. M., and R. Simó (2007), Strong Relationship Between DMS and the Solar Radiation Dose over, *Science*, *315*, 506–509, doi:10.1126/science.1133680.

Vallina, S. M., R. Simó, T. R. Anderson, A. Gabric, R. Cropp, and J. M. Pacheco (2008), A dynamic model of oceanic sulfur (DMOS) applied to the Sargasso Sea: Simulating the dimethylsulfide (DMS) summer paradox, *J. Geophys. Res.*, *113*(1), G01,009, doi: 10.1029/2007JG000415.

Vogt, M., S. M. Vallina, E. T. Buitenhuis, L. Bopp, and C. Le Quéré (2010), Simulating dimethylsulphide seasonality with the Dynamic Green Ocean Model PlankTOM5, *J. Geophys. Res.*, *115*(6), C06,021, doi:10.1029/2009JC005529.

Wang, S., and J. K. Moore (2011), Incorporating Phaeocystis into a Southern Ocean ecosystem model, *J. Geophys. Res.*, *116*(1), C01,019, doi:10.1029/2009JC005817.

Wang, S., S. Elliott, M. Maltrud, and P. Cameron-Smith (2015), Influence of explicit phaeocystis parameterizations on the global distribution of marine dimethyl sulfide, *Journal of Geophysical Research: Biogeosciences*, *120*(11), 2158–2177, doi: 10.1002/2015JG003017.

Wang, W.-L., J. K. Moore, A. C. Martiny, and F. W. Primeau (2019), Convergent estimates of marine nitrogen fixation, *Nature*, *566*(7743), 205–211.

Yang, G. P., X. T. Liu, L. Li, and Z. B. Zhang (1999), Biogeochemistry of dimethylsulfide in the South China Sea, *Journal of Marine Research*, *57*(1), 189–211, doi: 10.1357/002224099765038616.

Table 1. Results of linear regression model.

Parameter	R^2	Slope	No.
DMSPt	0.41	0.82	4,657
Chl a[1]	0.19	0.46	9,942
MLD	0.14	-0.36	53,391
PAR	0.12	0.34	54,683
Chl a[2]	0.07	0.26	41,146
SST	$\sim$0	-0.01	49,391
SSS	$\sim$0	-0.07	49,007
DIP	$\sim$0	0.03	50,350
DIN	$\sim$0	0.02	50,077
SiO	$\sim$0	0.04	50,304

[1] Original data.

[2] Data with climatology fill-in.

Table 2. Annually-averaged zonal mean DMS flux (Tg S/yr) for this study (W20), *Lana et al.* [2011] (L11), and *Kettle and Andreae* [2000](K00). L11 and K00 are computed with the *Nightingale et al.* [2000] parameterization of the piston velocity[N00]. Flux in this study is calculated using both the *Nightingale et al.* [2000], [N00], and *Goddijn-Murphy et al.* [2012],[GM12], parameterizations. Uncertainties are estimated based on models from 10 different random seeds. Errorbars correspond to $\pm 1\sigma$.

Latitude	K00[N00]	L10[N00]	W20[N00]	W20[GM12]
90°-80°N	0.0	0.0	0.00±0.00	0.00±0.00
80°-70°N	0.2	0.1	0.05±0.01	0.05±0.01
70°-60°N	0.2	0.2	0.11±0.01	0.11±0.01
60°-50°N	0.7	0.9	0.36±0.02	0.34±0.02
50°-40°N	1.4	1.5	0.60±0.02	0.56±0.02
40°-30°N	1.5	1.5	0.92±0.03	0.88±0.03
30°-20°N	1.4	1.4	1.07±0.04	1.09±0.04
20°-10°N	2.4	2.6	1.83±0.08	1.77±0.08
10°-0°N	2.3	2.6	1.87±0.08	1.93±0.10
0°-10°S	1.9	2.2	2.03±0.11	2.11±0.12
10°-20°S	2.3	3.5	2.14±0.15	2.09±0.15
20°-30°S	1.9	3.0	1.63±0.11	1.61±0.11
30°-40°S	1.9	2.7	1.98±0.11	1.87±0.10
40°-50°S	2.4	2.8	2.12±0.16	1.79±0.14
50°-60°S	2.0	2.1	1.10±0.11	0.89±0.09
60°-70°S	1.2	0.9	0.09±0.01	0.08±0.01
70°-80°S	0.2	0.1	0.00±0.00	0.00±0.00
80°-90°S	0.0	0.0	0.00±0.00	0.00±0.00
Total	24.0	28.1	17.90±0.34	17.17±0.32

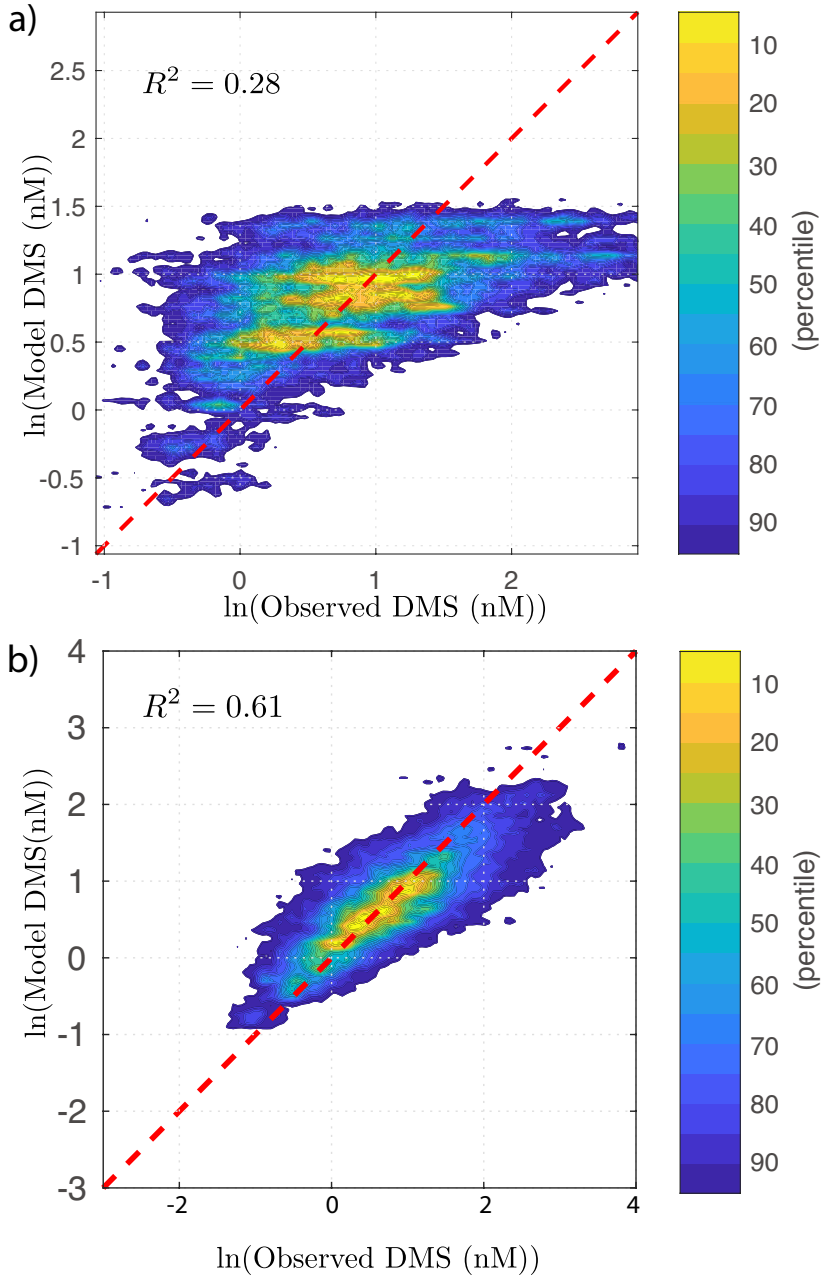

Figure 1. Model versus observation plots on logarithmic scale: (a) multilinear regression model; (b) artificial neural network model. The color indicates the fraction of the joint distribution explained as a percentile that falls within a region of concentration space.

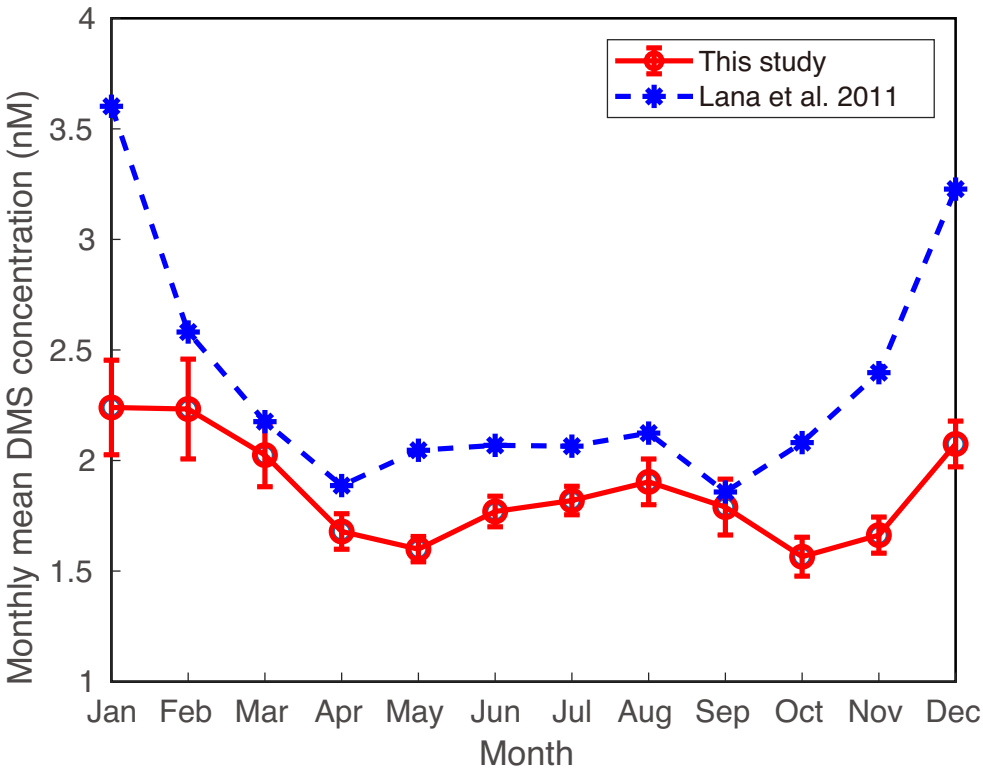

Figure 2. Comparisons of monthly mean DMS concentrations between this study and the study of *Lana et al.* [2011]. Uncertainties are estimated based on models from 10 different random seeds. Errorbars correspond to $\pm 1\sigma$.

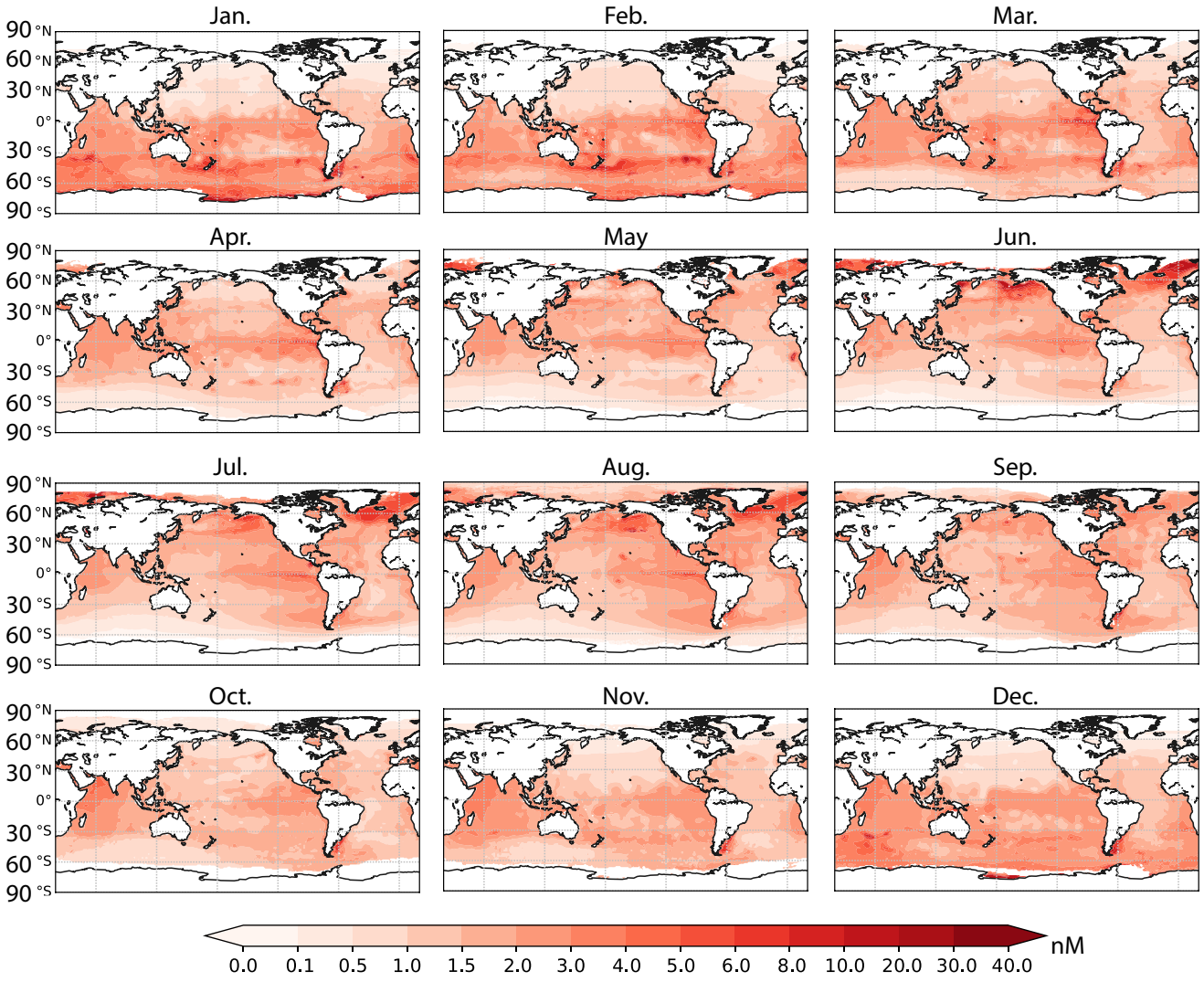

Figure 3. Monthly DMS concentration (nmol L^{-1}) estimated based on artificial neural network.

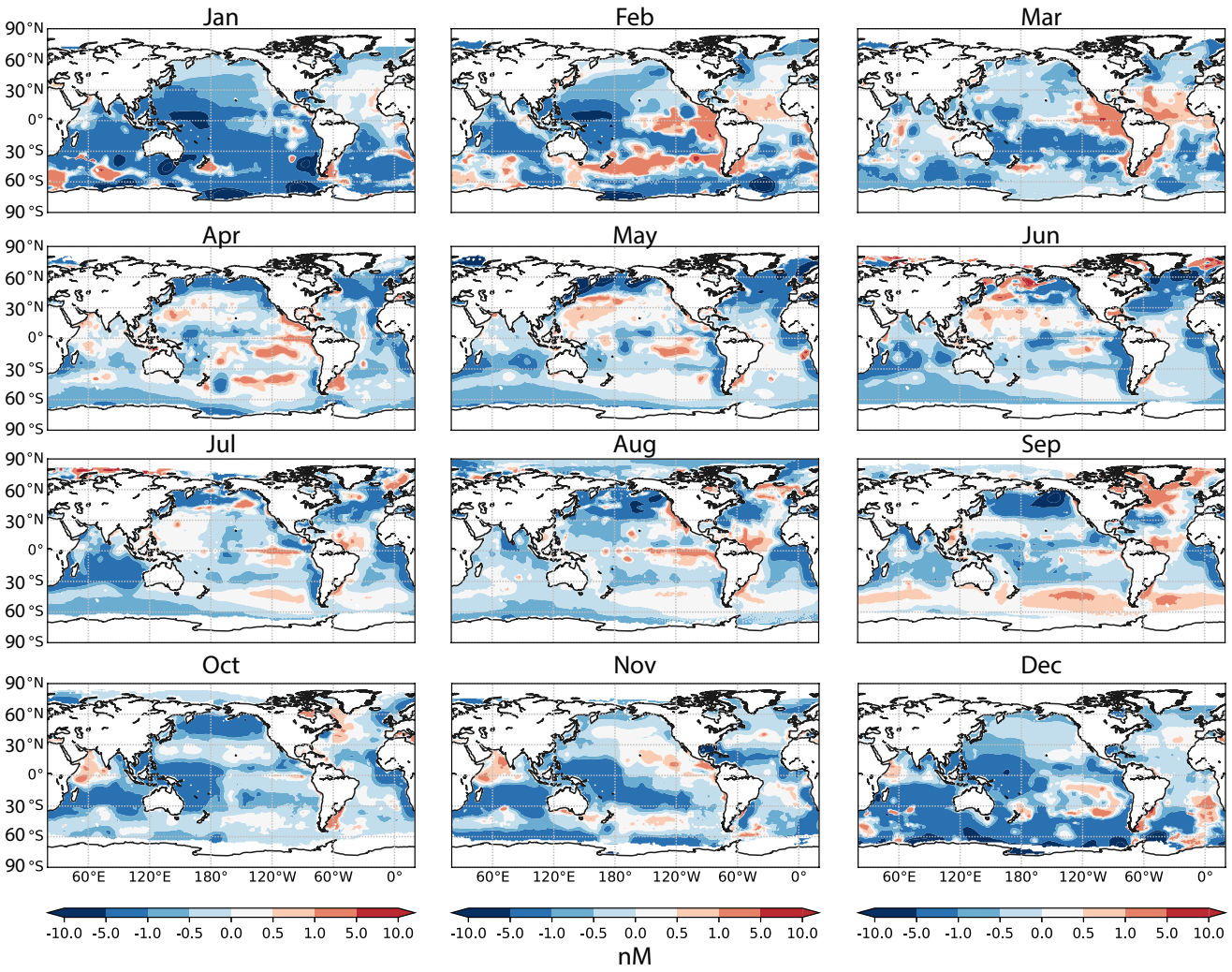

Figure 4. DMS concentration differences (nmol L^{-1}) between concentrations from the current study and those from *Lana et al.* [2011].

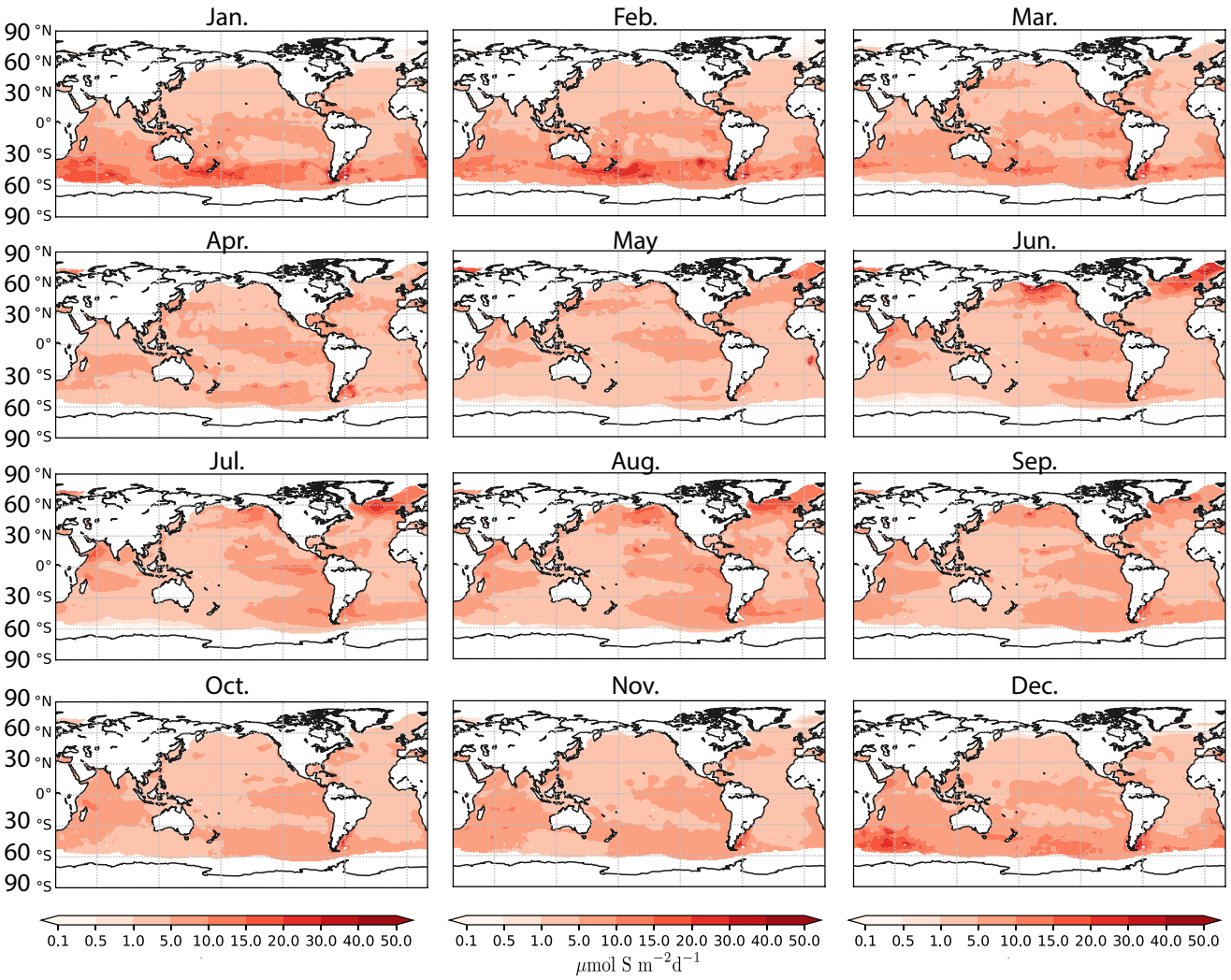

Figure 5. Monthly DMS flux (μmol S m^{-2} day^{-1}) calculated based on *Goddijn-Murphy et al.* [2012] parameterization. The figures are arranged according to seasons, first row is for northern hemisphere winter, second for spring, third for summer, and last row for autumn.

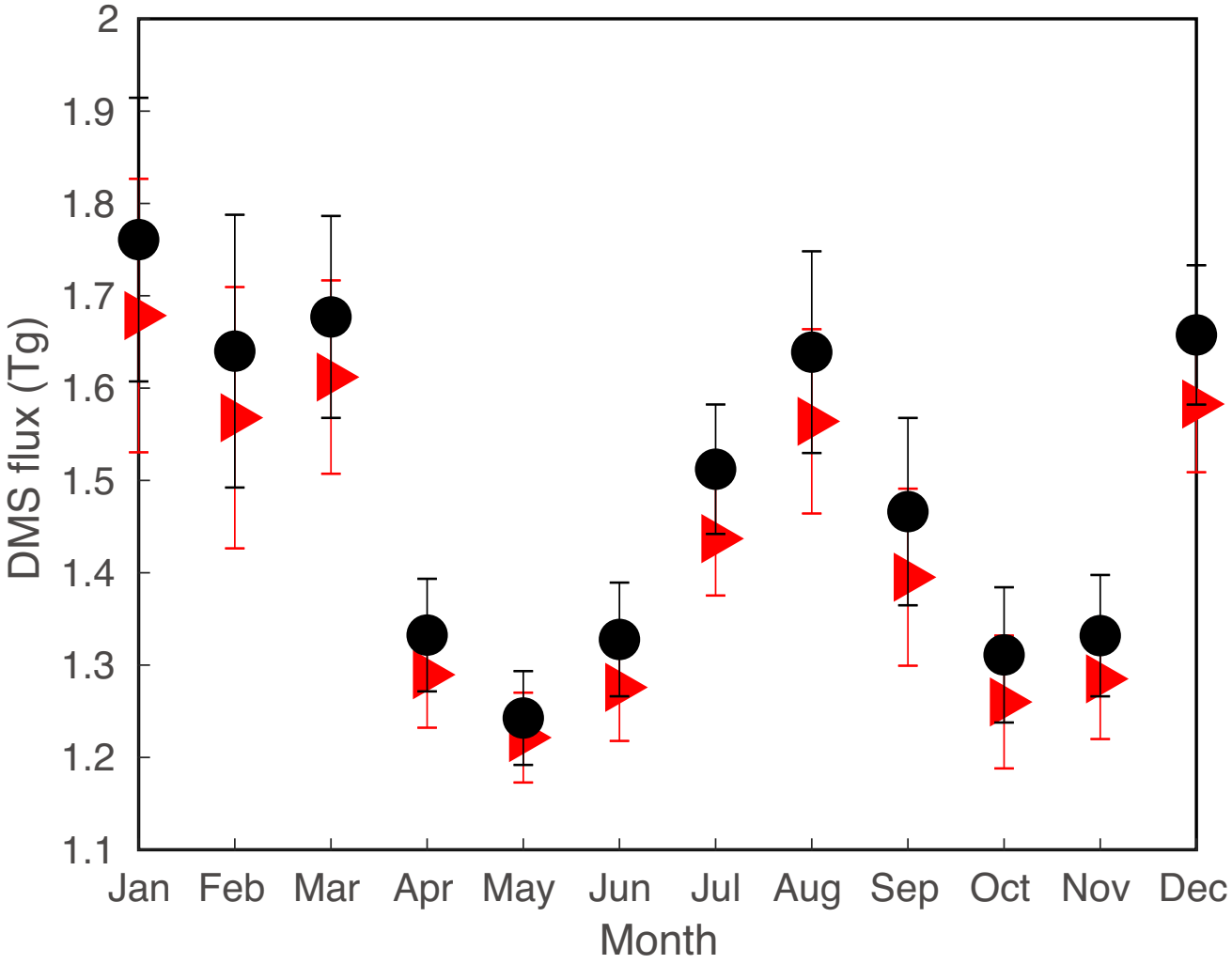

Figure 6. Area and month integrated DMS sea-to-air flux (Tg S month^{-1}). Red triangles represent flux calculated based on *Goddijn-Murphy et al.* [2012] parameterization, and black circles are flux based on *Nightingale et al.* [2000]. Uncertainties are estimated based on models from 10 different random seeds. Errorbars correspond to $\pm 1\sigma$.

5. Appendix Table and Figures

Table A1. DMS and ancillary data sources.

Variables	Sources	units	References
DMS[1]	http://saga.pmel.noaa.gov/dms/	nM	[*Kettle et al.*, 1999]
DMS[2]	NAAMES	nM	[*Behrenfeld et al.*, 2019]
Chl	https://oceandata.sci.gsfc.nasa.gov/SeaWiFS/	μg L^{-1}	[*NASA*, 2018]
MLD	http://www.ifremer.fr/cerweb/deboyer/mld/Data.php	m	[*de Boyer Montégut et al.*, 2004]
PAR	https://oceancolor.gsfc.nasa.gov/atbd/par/	Einsteins m^{-2} d^{-1}	[*Frouin et al.*, 2012]
WSP	https://podaac.jpl.nasa.gov/dataset	m s^{-1}	[*NASA*, 2012]
SST	WOA2013	C	[*Garcia et al.*, 2013]
SSS	WOA2013	psu	[*Garcia et al.*, 2013]
DIP	WOA2013	μM	[*Garcia et al.*, 2013]
DIN	WOA2013	μM	[*Garcia et al.*, 2013]
SiO	WOA2013	μM	[*Garcia et al.*, 2013]
ICE	CESM model	-	[*Wang et al.*, 2019]

[1] Original online data.
[2] New data from the North Atlantic Aerosol and Marine Ecosystems experiment .

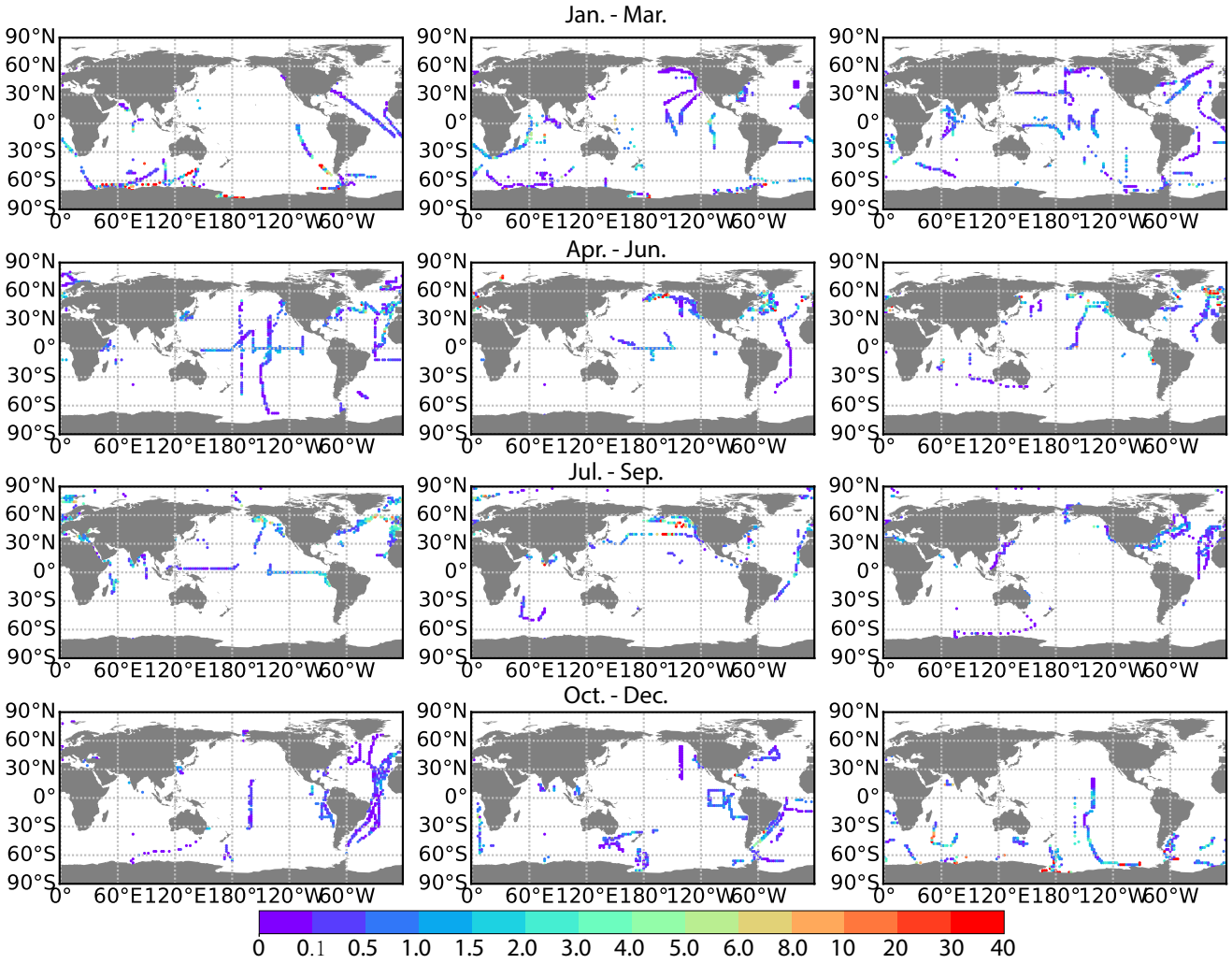

Figure A1. Distribution of DMS observations partitioned into each month. The color indicates DMS concentration (nM).

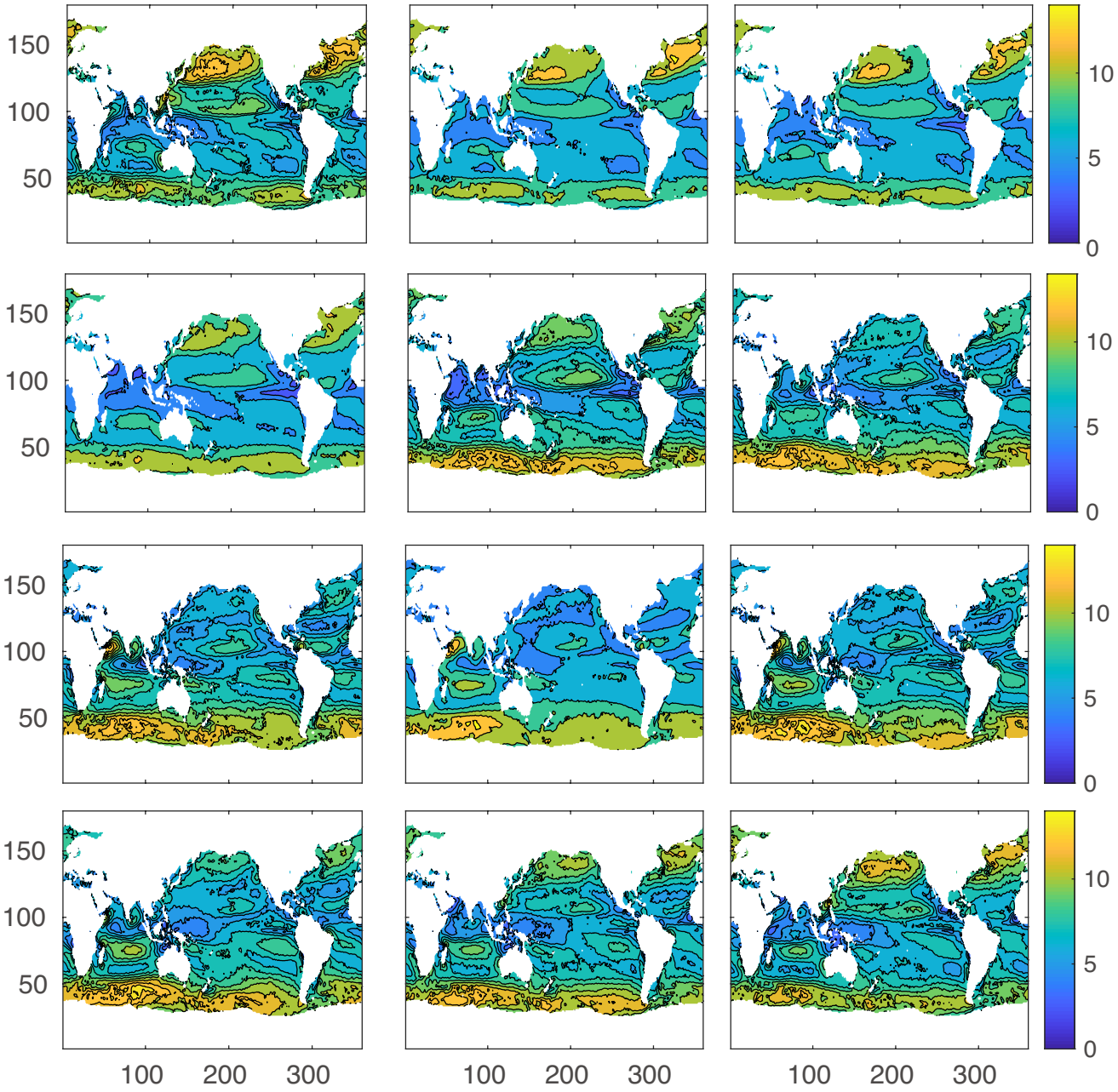

Figure A2. Climatological wind speed (m/s).