1 A 20-year (1998-2017) global sea surface dimethyl sulfide gridded dataset with daily resolution

3 Shengqian Zhou[1], Ying Chen[1,2,3], Shan Huang[4,5], Xianda Gong[6,7], Guipeng Yang[8,9,10], Honghai Zhang[8,9,10],
4 Hartmut Herrmann[5], Alfred Wiedensohler[5], Laurent Poulain[5], Yan Zhang[1,2], Fanghui Wang[1], Zongjun
5 Xu[1], Ke Yan[1]

6 [1]Shanghai Key Laboratory of Atmospheric Particle Pollution Prevention, Department of Environmental Science & Engineering,
7 Fudan University, Shanghai, 200438, China
8 [2]Institute of Eco-Chongming (IEC), National Observations and Research Station for Wetland Ecosystems of the Yangtze
9 Estuary, Shanghai, 200062, China
10 [3]Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China
11 [4]Institute for Environmental and Climate Research, Jinan University, Guangzhou, 511443, China
12 [5]Leibniz Institute for Tropospheric Research, Leipzig, 04318, Germany
13 [6]Research Center for Industries of the Future, Westlake University, Hangzhou, 310030, China
14 [7]Key Laboratory of Coastal Environment and Resources of Zhejiang Province, School of Engineering, Westlake University,
15 Hangzhou, 310030, China
16 [8]Frontiers Science Center for Deep Ocean Multispheres and Earth System, and Key Laboratory of Marine Chemistry Theory
17 and Technology, Ministry of Education, Ocean University of China, Qingdao, 266100, China
18 [9]Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and
19 Technology, Qingdao, 266071, China
20 [10]College of Chemistry and Chemical Engineering, Ocean University of China, Qingdao, 266100, China

21 *Correspondence to*: Ying Chen (yingchen@fudan.edu.cn)

22 **Abstract.** The oceanic emission of dimethyl sulfide (DMS) plays a vital role in the Earth's climate system and is a significant

23 source of uncertainty in aerosol radiative forcing. Currently, the widely used monthly climatology of sea surface DMS

24 concentration cannot meet the requirement for accurately simulating DMS-derived aerosols by chemical transport models.

25 Thus, there is an urgent need to construct a global sea surface DMS dataset with high time resolution spanning multiple years.

26 Here we develop an artificial neural network ensemble model based on 9 environmental factors, which demonstrate high

27 accuracy and generalization in predicting DMS concentrations. Subsequently, a global sea surface DMS concentration and

28 flux dataset (1°×1°) with daily resolution covering the period from 1998 to 2017 is established. According to this dataset, the

29 global annual average concentration was ~1.82 nM, and the annual total emission was ~17.9 TgS yr^{-1}, with ~60% originating

30 from the southern hemisphere. While overall seasonal variations are consistent with previous DMS climatologies, notable

31 differences exist in regional-scale spatial distributions. The new dataset enables further investigation of daily and decadal

32 variations. During 1998–2017, the global annual average concentration exhibited a slight decrease, while total emissions

33 showed no significant trend. Benefiting from the incorporation of daily and interannual variation information, the DMS flux

34 from our dataset showed a much stronger correlation with observed atmospheric methanesulfonic acid concentration compared

35 to those from previous monthly climatologies. As a result, it can serve as an improved emission inventory of oceanic DMS

36   and has the potential to enhance the simulation of DMS-derived aerosols and associated radiative effects. The new DMS
37   gridded products are available at https://zenodo.org/record/10279659 (Zhou et al., 2023).

1 Introduction

39   Dimethyl sulfide (DMS) mostly produced by ocean biota accounts for more than half of natural sulfur emissions and
40   contributes substantially to sulfur dioxide in the troposphere (Sheng et al., 2015; Andreae, 1990) which can be oxidized to
41   sulfuric acid and form sulfate aerosols (Barnes et al., 2006; Hoffmann et al., 2016). Sulfate aerosols play an important role in
42   climate systems by both scattering solar radiation and altering cloud condensation nuclei (CCN) and albedo (Masson-Delmotte
43   et al., 2021). Recent studies have proven that CCN over remote oceans and polar regions are primarily composed of non-sea-
44   salt sulfate (nss-SO_4^{2-}) (Quinn et al., 2017; Park et al., 2021). Given 70% coverage of the Earth's surface by the ocean and the
45   weak influence of anthropogenic SO_2 over open oceans, marine biogenic DMS can be the most important source of nss-SO_4^{2-}
46   and regulates regional and global climate (McCoy et al., 2015). Accordingly, DMS has been suggested to be the key substance
47   in the postulated feedback of marine phytoplankton to climate warming (the "CLAW" hypothesis) (Charlson et al., 1987).
48   However, there are still many controversies surrounding the CLAW hypothesis (Quinn and Bates, 2011). To address these
49   controversies, we need to accurately simulate the global climate effects of DMS-derived aerosols. This requires further
50   exploration of complex atmospheric chemistry and climate processes (Hoffmann et al., 2016; Novak et al., 2021), as well as
51   high-fidelity and high-resolution sea surface DMS concentrations and emission fluxes as emission inventories. According to
52   the uncertainty estimation by Carslaw et al. (2013), the uncertainty in DMS emission flux is the second largest contributor to
53   the overall uncertainty in the radiative forcing of natural aerosols. Therefore, improving the understanding of the
54   spatiotemporal variations of DMS in global oceans is currently an important task.

55   There are complex production and consumption mechanisms of DMS in the upper ocean, which makes it difficult to well
56   capture the dynamics and distributions of sea surface DMS across different oceanic regions. Dimethylsulfoniopropionate
57   (DMSP), the major precursor of DMS, is synthesized mainly by phytoplankton in the photic zone and plays a variety of
58   physiological functions in algal cells (Stefels, 2000; Sunda et al., 2002; McParland and Levine, 2018). The DMSP yield differs
59   dramatically among algal species (Stefels et al., 2007; Keller et al., 1989), and DMS can be produced through DMSP
60   intracellular and extracellular cleavage by both algae and bacteria (Alcolombri et al., 2015; Zhang et al., 2019). Therefore, the
61   oceanic DMS produced via multiple pathways can be affected by many biotic and abiotic factors, including temperature,
62   salinity, solar radiation, mixed layer depth, nutrients, oxygen, acidity, etc. (Simó and Pedrós-Alió, 1999b; Vallina and Simó,
63   2007; Stefels, 2000; Zindler et al., 2014; Six et al., 2013; Omori et al., 2015; Stefels et al., 2007). In addition, seawater DMS
64   has multiple removal pathways (bacterial consumption, photodegradation, sea-to-air ventilation, etc.), further complicating the
65   DMS cycling (Stefels et al., 2007). Therefore, although previous studies have developed several empirical algorithms (Simó
66   and Dachs, 2002; Belviso et al., 2004b; Vallina and Simó, 2007) and process-embedded prognostic models (Kloster et al.,

67   2006; Vogt et al., 2010; Belviso et al., 2011; Wang et al., 2015) based on relevant variables (mixed layer depth, chlorophyll *a*,

68   nutrients, radiation, phytoplankton group, etc.) to estimate the distribution of DMS, their results showed significantly different

69   distribution patterns and obvious inconsistency with observations in many regions (Tesdal et al., 2016; Belviso et al., 2004a).

70   Recently, Galí et al. (2018) developed a new empirical algorithm following a parameterization of DMSP (Galí et al., 2015).

71   The estimated DMS field exhibited a generally higher consistency with observations than those derived from the previous two

72   algorithms SD02 (Simó and Dachs, 2002) and VS07 (Vallina and Simó, 2007), but this method did not consider the influences

73   of nutrients and there are still great biases in certain regions (e.g., near the Antarctic).

74   Since Lovelock et al. (1972) first discovered the ubiquitous presence of DMS in seawater, numerous observations of sea

75   surface DMS have been conducted worldwide, accumulating a large amount of observational data to date. Based on these

76   worldwide measurements, monthly climatology of global DMS can be generated through interpolation and extrapolation

77   (Hulswar et al., 2022; Kettle et al., 1999; Lana et al., 2011). The latest version incorporated 873,539 raw observations (48,898

78   after data filtration and unification for climatology development), and the estimated global annual mean concentration and

79   total flux are 2.26 nM and 27.1 TgS yr^{-1}, respectively (Hulswar et al., 2022). However, despite the large data volume, there

80   are still considerable spatial and temporal disparities, which may lead to large uncertainties for the regions or periods with

81   sparse observations. In addition, the observational data from different years were equally combined and the interannual

82   variations cannot be addressed by this approach.

83   In recent years, the application of data-driven approaches like machine learning to Earth system science has drawn more and

84   more attention. Compared with traditional approaches and process models, machine learning explores larger function space

85   and captures more hidden information from the big data, hence it often provides a better simulation and prediction performance

86   (Reichstein et al., 2019; Zheng et al., 2020; Bergen et al., 2019). A recent study has demonstrated that artificial neural network

87   (ANN) can capture much more (~66%) of the raw data variance than multilinear regression (~39%), and a global monthly

88   climatology of DMS has been developed based on ANN (Wang et al., 2020). The machine learning techniques have also been

89   used to simulate the distribution of DMS in the Arctic (Galí et al., 2021), North Atlantic Ocean (Bell et al., 2021; Mansour et

90   al., 2023), and East Asia (Zhao et al., 2022).

91   However, to our best knowledge, there is no global-scale sea surface gridded DMS dataset with both high time resolution

92   (daily) and long-term coverage (> 10 years), which is urgently needed for modeling the atmospheric processes and climatic

93   implications of oceanic DMS. Since the sea surface concentration and sea-to-air emission flux of DMS can vary greatly from

94   day to day (Simó and Pedrós-Alió, 1999a) and the emitted DMS makes effects on the atmosphere in a time scale of several

95   hours to days, just using monthly climatology of DMS as the emission inventory cannot capture the details and may lead to

96   large modeling biases compared with observed concentrations of atmospheric DMS or its oxidation products (Chen et al., 2018;

97   Fung et al., 2022).

98   Here, we build a 20-year (1998 – 2017) global sea surface DMS gridded dataset (1°×1°) with a daily resolution based on a

99   data-driven machine learning approach (ANN ensemble). This product can improve our understanding of the spatiotemporal

100   variations of oceanic DMS. More importantly, it can serve as an updated emission inventory of marine biogenic DMS for

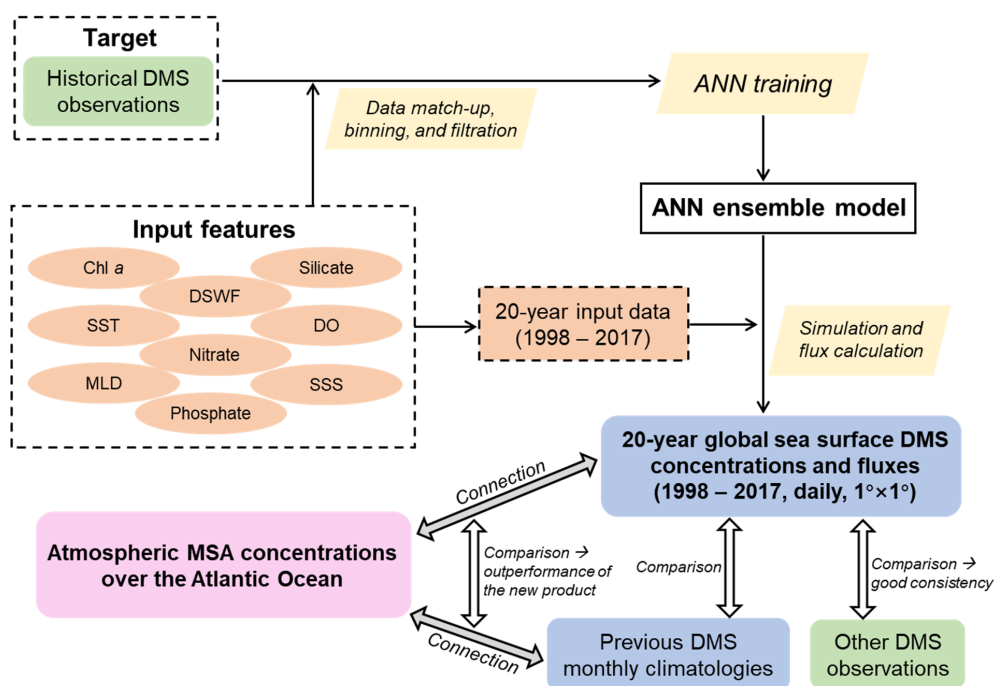
101   chemical transport models, which is very beneficial for improving the simulation of atmospheric processes of DMS and lower

102   down the uncertainties in marine aerosol's climate effects. The paper consists of the following main parts which are

103   demonstrated in Fig. 1: (1) the development of the machine learning model based on global DMS measurements and 9 ancillary

104   environmental variables; (2) the derived spatial and temporal distributions of DMS and comparisons with previous estimates;

105   (3) an example showing the superiority of our newly developed DMS field in linking atmospheric biogenic sulfur; and (4) the

106   uncertainties and limitations of our approach and the resulting data product.

107

108   **Figure 1.** Flowchart of this study, including the construction of the new DMS gridded dataset and subsequent analyses of the

109   validity and superiority of this product.

2 Methodology

2.1 Input datasets

The in-situ DMS measurement data used for training the machine learning model was obtained from the Global Surface Seawater DMS (GSSD) database (Kettle et al., 1999). The GSSD database contains a total of 87,801 DMS measurements obtained from 266 cruises and fixed-site observation campaigns from 11 March 1972 to 27 August 2017 (https://saga.pmel.noaa.gov/dms/, last access: 1 April 2020). The spatial distribution of theses in-situ observational data is shown in Fig. S1, which covers all major regions of the global ocean.

We selected 9 environmental variables related to DMS biogeochemical processes as input features, including chlorophyll *a* (Chl *a*), sea surface temperature (SST), mixed layer depth (MLD), nitrate, phosphate, silicate, dissolved oxygen (DO), downward short-wave radiation flux (DSWF), and sea surface salinity (SSS). The data sources and relevant information of 9 input variables and DMS are listed in Table 1. Chl *a* was obtained from both in-situ observations accompanied by DMS data and satellite remote sensing products (SeaWiFS and Aqua-MODIS, 0.083°×0.083°). Daily SST data (0.25°×0.25°) was from the NOAA OI SST V2 high-resolution blended reanalysis dataset (Huang et al., 2021). Daily MLD, DSWF, and SSS were from the modeling outputs of NASA's "Estimating the Circulation and Climate of the Ocean" (ECCO) consortium, Version 4 Release 4 (V4r4) (Forget et al., 2015). The LLC90 (Lat-Lon-Cap 90) grid is a native grid used for ECCO data, which has 5 faces containing 13 regional tiles covering the global ocean, with spatial resolution varying from 22 km to 110 km (Forget et al., 2015). As for the concentrations of nitrate, phosphate, silicate, and DO, the monthly climatologies (1°×1°) of World Ocean Atlas 2018 (WOA18) based on the historical observational data from the World Ocean Database (WOD) were used (Garcia et al., 2019b, a). The surface wind speed (WS) and sea ice fraction (SI) data are also needed in the calculation of sea-to-air flux. Here we utilized the daily 10-meter WS data from ECCO V4r4 and the daily SI data from NOAA OI SST V2. Since there are multiple different spatial grids and temporal resolutions among all datasets, the data match-up has been conducted as described in the next section.

132

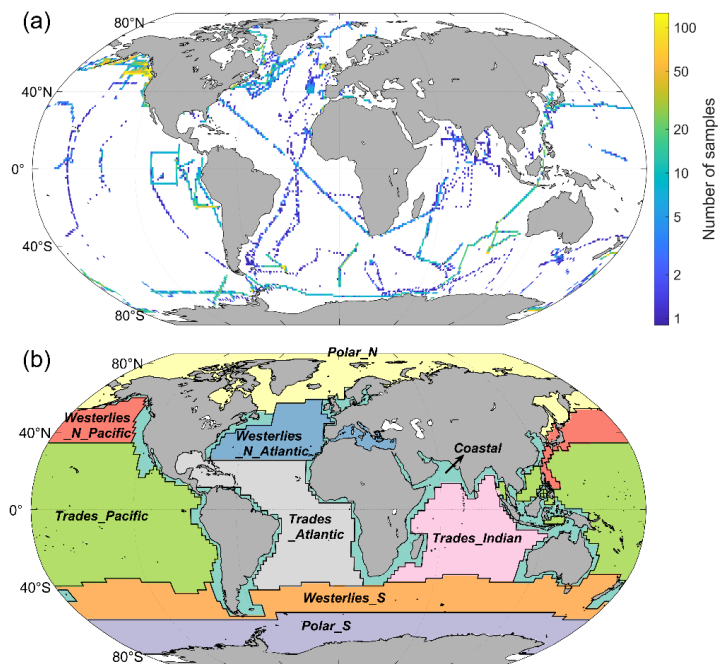



Figure 2. (a) The distribution of 34,118 DMS observational data after matchup, filtration, and binning for constructing the

ANN model. The grid size is 1°×1°. (b) Nine oceanic regions separated based on Longhurst's biomes (Longhurst, 1998).

135

Table 1. The data sources and related information of variables used for model development, DMS simulation, and flux calculation

Variable	Data source	URL	Temporal resolution	Temporal coverage	Spatial grid
DMS	GSSD database	https://saga.pmel.noaa.gov/dms/	In-situ	Mar. 1972 – Aug. 2017	-
Chl *a*	GSSD database	https://saga.pmel.noaa.gov/dms/	In-situ	Oct. 1980 – Aug. 2017	-
	SeaWiFS	https://oceandata.sci.gsfc.nasa.gov/	Daily, 8-day, monthly	Sep. 1997 – Dec. 2011	0.083°×0.083°
	Aqua-MODIS			Jul. 2002 – present	
SST	NOAA OI SST V2	https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html	Daily	Sep. 1981 – present	0.25°×0.25°
MLD	NASA ECCO V4r4	https://data.nas.nasa.gov/ecco/data.php?dir=/eccodata/llc_90/ECCOv4/Release4	Daily	Jan. 1992 – Dec. 2017	LLC90
DSWF					
SSS					
Nitrate	WOA18	https://www.nodc.noaa.gov/OC5/woa18/woa18data.html	Monthly climatology	-	1°×1°
Phosphate					
Silicate					
DO					
WS	NASA ECCO V4r4	https://data.nas.nasa.gov/ecco/data.php?dir=/eccodata/llc_90/ECCOv4/Release4	Daily	Jan. 1992 – Dec. 2017	LLC90
SI	NOAA OI SST V2	https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html	Daily	Sep. 1981 – present	0.25°×0.25°

2.2 Data preprocessing for model development

The data extraction and match-up were performed according to the sampling location and time in the GSSD database, as well as the time range and grid distribution of each variable. For satellite-retrieved Chl *a* concentration, the data of the grids covering

142 DMS sampling points were extracted and matched. If one grid data is missing, the average value of the 5×5 grids nearby will

143 be extracted. The SeaWiFS and Aqua-MODIS data during their overlapping periods were averaged. For other variables, only

144 values in the grids matching the DMS sampling locations were extracted.

145 There are in-situ Chl a measurements accompanied by some of the GSSD data, which were used along with satellite-retrieved

146 Chl a. The in-situ Chl a have many low-precision values (defined as < 0.1 mg m^{-3}, and the number of significant digits is 1),

147 and we removed those values. For a certain in-situ observation campaign, if the number of low-precision values is larger than

148 10 and accounts for more than half, all in-situ Chl a data from this campaign will be removed. In addition, the in-situ Chl a

149 data in the GSSD database are measured by two different methods, i.e., Turner fluorometry and high-performance liquid

150 chromatography (HPLC). In order to improve mutual consistency, a conversion between the data from these two methods was

151 applied and then the in-situ Chl a concentrations were adjusted to match up with satellite Chl a following the functions

152 described in Galí et al. (2015). After that, the statistical outliers for all $\log_{10}$(Chl a) (outside the range of average ± 3 standard

153 deviations) were eliminated. The $\log_{10}$ transformation was applied to make the data distribution close to normal distribution.

154 When finally selecting the $\log_{10}$(Chl a) corresponding to each DMS data, the precedence order of the various datasets is as

155 follows: in-situ data, daily satellite data, 8-day averaged satellite data, and monthly averaged satellite data. The comparison

156 between in-situ and different satellite-retrieved Chl a data is shown in Fig. S2. There is a good consistency between in-situ

157 and daily satellite Chl a data (R^2 > 0.5, RMSE < 0.4), indicating it is rational to combine the in-situ and satellite Chl a data

158 together.

159 DMS and extracted MLD and three nutrients (nitrate, phosphate, silicate) were also performed $\log_{10}$ transformation. Together

160 with SST, DO, DSWF, and SSS, their statistical outliers were excluded as mentioned above. After data filtration, a total of

161 54,191 DMS measurements can be matched by all variables. To avoid multi-data points gathering in a certain time (the same

162 day) and a small region (< 0.05°×0.05°), these data points were averaged, and finally the number of binned samples used for

163 model development is 34,118. The spatial distribution of these binned samples is shown in Fig. 2a.

164 **2.3 Artificial neural network training and validation**

165 The 34,118 binned samples after the previously mentioned data preprocessing were used to develop the artificial neural

166 network (ANN) model. The target feature is $\log_{10}$(DMS), and the input features are $\log_{10}$(Chl a), SST, $\log_{10}$(MLD),

167 $\log_{10}$(nitrate), $\log_{10}$(phosphate), $\log_{10}$(silicate), DO, DSWF, and SSS. The data of all variables were standardized before

168 training. We randomly selected 5% of the samples (n = 1706) to be entirely excluded from training, as a subset for global

169 validation and overfitting test. Then, the remaining samples (n = 32,412) were randomly split into three parts, that is, 70% for

170 training, 15% for validation, and 15% for testing. Our feed-forward back-propagation (BP) neural network contains one hidden

171 layer with 20 nodes, and the training algorithm is the Levenberg-Marquardt algorithm. Rooted mean squared error (RMSE) is

172 chosen as the indicator for performance evaluation during training, and the upper limit of the number of iterations in each

173 training session is 1000. The training processes were carried out with Neural Network Toolbox on Matlab R2014b. We have
174 trained the ANN 100 times and each training session started independently with a new random partition of those 32,411
175 samples. The average output of 100 trained ANNs showed an obviously higher consistency with the target than individual
176 ANN. But as the number of training sessions ($N_{training}$) is larger than 80, the simulation performance tends to be stable (Fig.
177 S3). Therefore, we used the average of 100 ANN outputs as our final model output. This kind of multiple-training approach is
178 also called "ANN ensemble" or "Monte Carlo cross-validation", which has been widely used to improve the model
179 generalization and performance (Sigmund et al., 2020; Holder et al., 2022) as well as get a better model evaluation (Dubitzky
180 et al., 2007).

181 To evaluate whether there is a spatial preference for DMS simulation, the comparisons of simulated and observed DMS
182 concentrations in different types of oceanic regions were conducted. We divided the global ocean into 9 regions based on
183 Longhurst's biomes (Longhurst, 1998). There are 6 biomes in Longhurst's definition, including Coastal, Polar_N, Polar_S,
184 Westerlies_N, Westerlies_S, and Trades (the .shp file of Longhurst's biomes and provinces was downloaded from
185 https://www.marineregions.org/downloads.php#longhurst). We further divided Westerlies_N into Westerlies_N_Pacific and
186 Westerlies_N_Atlantic, and divided Trades into Trades_Pacific, Trades_Indian, and Trades_Atlantic by different basins as
187 shown in Fig. 2b.

2.4 Deriving the 20-year global DMS distributions

2.4.1 Simulation of sea surface DMS concentrations

190 First, we constructed the daily gridded dataset of input variables with a spatial resolution of 1°×1° from 1998 to 2017 using
191 the data sources listed in Table 1 (except in-situ Chl a data). For those datasets with a higher spatial resolution than 1°×1°,
192 values in each 1°×1° grid were averaged. As for satellite Chl a data, the priority level was the same as mentioned in Section
193 2.2, and SeaWiFS and Aqua-MODIS datasets were combined. As for nutrients and DO, only monthly climatologies were
194 provided in WOA18 and the same value was used for all the days of each month. Hence, these variables lack information on
195 inter-annual and day-to-day variations, but the spatial and monthly variations are well captured. The potential uncertainties in
196 predicted DMS fields associated with this issue are discussed in Section 4. Then, the obtained gridded dataset was fed into the
197 ANN ensemble model, and the 20-year global distribution of sea surface DMS concentration with the daily resolution was
198 simulated.

2.4.2 Calculation of sea-to-air fluxes

200 The sea-to-air fluxes of DMS were calculated on the basis of simulated surface DMS concentrations following equation (1):

201 $$DMS\ flux = Kt \times (DMS_w - \frac{DMS_a}{H}) \tag{1}$$

202   Here DMS_w and DMS_a are DMS concentrations in surface seawater and air, respectively. H is Henry's law constant of DMS.

203   Since $\frac{DMS_a}{H}$ is usually $\ll DMS_w$, this term was omitted in the calculation. Kt is the total transfer velocity considering the sea

204   ice coverage fraction (SI):

205    $Kt = k_t \times (1 - SI)$ (2)

206   k_t is the total transfer velocity without considering sea ice which is calculated by equation (3):

207    $\frac{1}{k_t} = \frac{1}{k_w} + \frac{1}{k_a \times H}$ (3)

208   Here k_w and k_a are the water-side transfer velocity and air-side transfer velocity, respectively. We used the same approach as

209   Galí et al. (2019) to obtain k_w, k_a, and H for DMS, where the effect of wind speed was considered for k_a, and the influences of

210   SST and SSS were considered for H. The calculations of k_a and H followed the parameterizations put forward by Johnson

211   (2010). As for k_w calculation, we adopted the bubble scheme (Woolf, 1997), which divided the sea-to-air mass transfer process

212   into turbulence- and bubble-mediated gas exchange. The calculated k_w based on the bubble scheme is lower than that of

213   Nightingale's scheme (Nightingale et al., 2000) under the condition of high wind speed and has a smaller deviation from the

214   actual value (Beale et al., 2014; Galí et al., 2019). Before calculation, WS and SI data were also binned by 1°×1° grid. By

215   using WS and SI together with SST and SSS datasets, we obtained the daily gridded Kt and then calculated the sea-to-air DMS

216   fluxes (daily, 1998–2017) by multiplying simulated DMS concentrations by Kt values.

217   **3 Results**

218   **3.1 Model performance**

219   As shown in Fig. 3a, the newly-developed ANN ensemble model captures a substantial part of data variance globally ($\log_{10}$

220   space $R^2 = 0.624$ and RMSE = 0.267). 92.3% of ANN simulated concentration values fall within 1/3 to 3 times of corresponding

221   true values. The simulation performance for 5% of the dataset not used in training ($R^2 = 0.604$, RMSE = 0.274, and 91.6% of

222   data within the range of 1/3 to 3 times of observations) is very close to that for the training dataset (Fig. 3b), suggesting no

223   obvious overfitting. The ANN model exhibits much better performance compared to previous empirical and process-based

224   models ($R^2 = 0.01\sim0.14$) (Tesdal et al., 2016) as well as the satellite-based algorithm ($R^2 = 0.50$) (Galí et al., 2018). The ANN

225   model developed by Wang et al. (2020) showed a slightly higher performance ($R^2 = 0.66$, RMSE = 0.264 for training dataset)

226   than our model, likely due to their more complex ANN configuration (two hidden layers with 128 nodes each) and the including

227   of sample location and time into input features. However, the more complex model will significantly increase the

228   computational cost, and the incorporation of location and time information may weaken the physical interpretability. On the

229   other hand, the performance improvement is very limited. Therefore, we keep the simpler model configuration. It should be

230   noted the ANN model may not reproduce the extreme DMS concentrations (> 20 nM or < 0.2 nM) well, but the fraction of

231   these values is less than 3% and has little impact on the overall performance.

232 Our ANN model also shows good performance in each of the 9 separated oceanic regions without harmful spatial preference.

233 As illustrated in Fig. 3c, the $\log_{10}$ space RMSEs are all below 0.3 except for the Coastal region (RMSE = 0.304). The percentage

234 of simulated values within the 1:3 and 3:1 lines ranges from 88.7% to 99.4%. Although the R^2 values in 3 Trades regions are

235 lower than 0.5, which is related to the relatively small variation range of the DMS, the RMSEs in these regions are quite low

236 and not inferior to those of other regions.

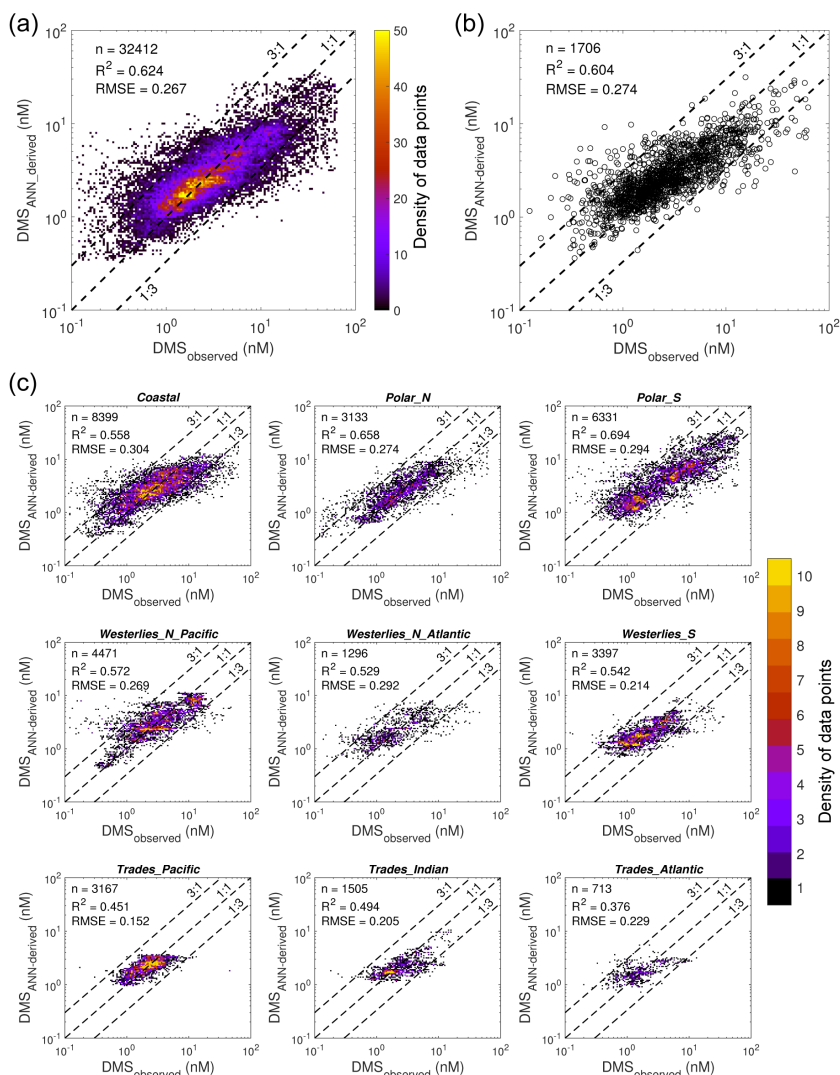



Figure 3. Comparisons between ANN-simulated and observed DMS concentrations in the GSSD database. (a) Scatter density for all simulated versus observed DMS concentration data used in ANN training. (b) Comparison between the simulation results and the 5% of GSSD observational data not used for training. (c) Scatter density for simulated versus observed DMS concentration data in each of the 9 regions. The number of data (n), the $\log_{10}$ space R^2 and the root mean squared error (RMSE) are also displayed.

243

244   It is worth noting that there may be intrinsic connections between the 5% excluded DMS subset and trained dataset both from
245   the GSSD database, because the data from the same cruise or fixed-site campaign have certain continuity. To further evaluate
246   the reliability of the ANN model, we compared the simulated DMS concentrations with the observational data not archived in
247   the GSSD database, which are obtained from 35 cruises in Northeast Pacific, West Pacific, and North Atlantic (number of data
248   = 6,478). These data include (1) off-line sampling and measurement data of 31 cruises of *Line P Program* in Northeast Pacific
249   (Steiner   et   al.,   2011)   (9   February   2007   –   26   August   2017,   number   of   data   =   177,
250   https://www.waterproperties.ca/linep/index.php, last access: 23 November 2020), (2) underway measurements during *SONNE*
251   *cruise 202/2 (TRANSBROM)* in West Pacific (Zindler et al., 2013) (9 – 23 October 2009, number of data = 115,
252   https://doi.org/10.1594/PANGAEA.805613, last access: 23 November 2020), (3) underway measurements during 3 cruises of
253   the *North Atlantic Aerosols and Marine Ecosystems Study (NAAMES)* (Behrenfeld et al., 2019; Bell et al., 2021) (11 – 30
254   November 2015, 14 May – 4 June 2016, 6 – 24 September 2017, number of data = 6,186, https://seabass.gsfc.nasa.gov/naames,
255   last access: 27 November, 2020). Before the comparison, the data measured within a 1°×1° grid and at the same day were
256   binned by arithmetic average.

257   The comparisons between these observed DMS concentrations and ANN simulation are shown in Fig. 4. As for the *Line P*
258   *Program*, it should be noted that there are 7 cruises included in the GSSD database, but those data were obtained by underway
259   measurements, different from the off-line data used here. Hence, these cruises were retained and marked in Fig. 4a but
260   eliminated in subsequent statistical analysis (Fig. 4b-c). It can be seen that the simulation well captures the seasonal variation,
261   which is generally August > June > February. In addition, the simulation can also partially reproduce the changes between
262   different stations, though the performance is not as good as the overall comparison between different cruises. As for underway
263   measurements, although a certain degree of overestimation in the West Pacific and underestimation in North Atlantic was
264   observed, most of the simulated values (88.4%) are within the range of 1/3 to 3 times of observations. If all of the above data
265   are binned by each cruise, the simulations will demonstrate high consistency with observations as shown in Fig. 4c ($R^2 = 0.784$,
266   RMSE = 0.227). These results further evidence the validity of our ANN model in simulating the concentrations of sea surface
267   DMS.

268

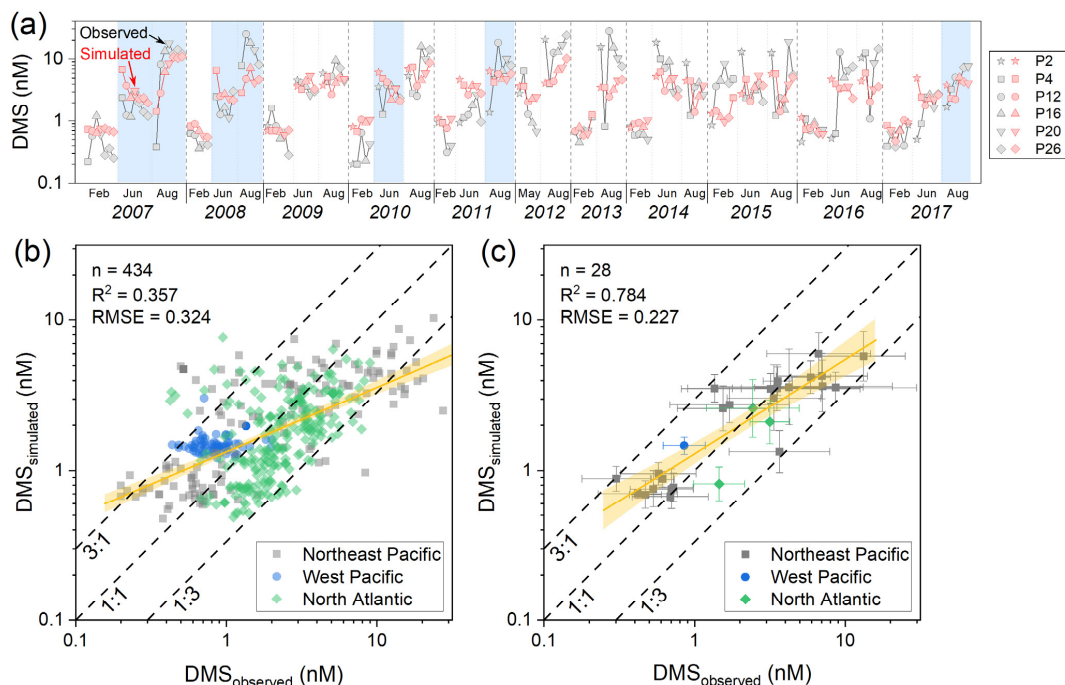



Figure 4. Comparisons between the ANN model simulation results and DMS observational data not archived in the GSSD database. (a) Time series of simulation results and DMS observational data obtained from *Line P Program*. The different markers represent different stations of *Line P*. The blue shades cover the data obtained from the cruises included in the GSSD database but with a different method. (b) Scatter plot of simulated versus observed DMS concentrations. (c) The same as panel b but for averaged data of each cruise. The yellow lines and shaded bands are linear fittings and corresponding 95% confidence intervals for $\log_{10}$ space data. The R^2 and RMSE displayed in the figure also correspond to $\log_{10}$ space data.

3.2 DMS distribution

3.2.1 Spatial and seasonal variations

The monthly climatology of ANN-simulated DMS concentrations in the global sea surface from 1998 to 2017 is shown in Fig. 5. Overall, the DMS concentrations in mid-high latitude regions exhibit a seasonal variation of high in summer and low in winter, which is consistent with the results of many previous observational studies. In the northern hemisphere, high DMS concentrations (>2.5 nM) in summer mainly occur in two regions. One is the North Pacific (40°–60° N) extending from the coast of Japan and Russia to the coast of Alaska and Canada, where the concentration generally peaks in August (Fig. 6). The

282 other is the subarctic North Atlantic (50°–80° N), where a significant increase of DMS concentration firstly occurs in the 45°–
283 50° N in May and then gradually shifts to the north of 50° N in July, consistent with the spatiotemporal evolution of solar
284 radiation intensity. The peak concentration date at the same latitude in the North Atlantic is generally earlier than in the North
285 Pacific (Fig. 6). In the southern hemisphere, there is a clear DMS-rich area near 40° S (where the Subtropical Convergence
286 lies) in summer, forming a ring-shaped high-concentration band nearly parallel to the latitude. The highest seasonal mean
287 concentration (December–February) at a specific latitude is 3.58 nM, taking place at 41.5° S (Fig. 8). Southward from this
288 region, there is a low-DMS area in the 51°–61° S, where the average concentration is below 2.0 nM in all seasons. However,
289 in the coastal waters of Antarctica (south of 60° S), significantly high concentrations also appear in summer, reaching over 4.0
290 nM, even higher than the region near 40° S (Fig. 4 and 8). In addition to the above regions, some typical upwelling zones also
291 exhibit relatively higher DMS concentrations, such as the eastern Pacific and the Southeast Atlantic. The former is at low
292 latitudes with no obvious seasonal variation, while the latter exhibits higher concentrations from October to February. The
293 high nutrient concentrations in upwelling areas could support higher primary productivity, leading to more intense biological
294 activities including the production of biogenic sulfur.

295

296

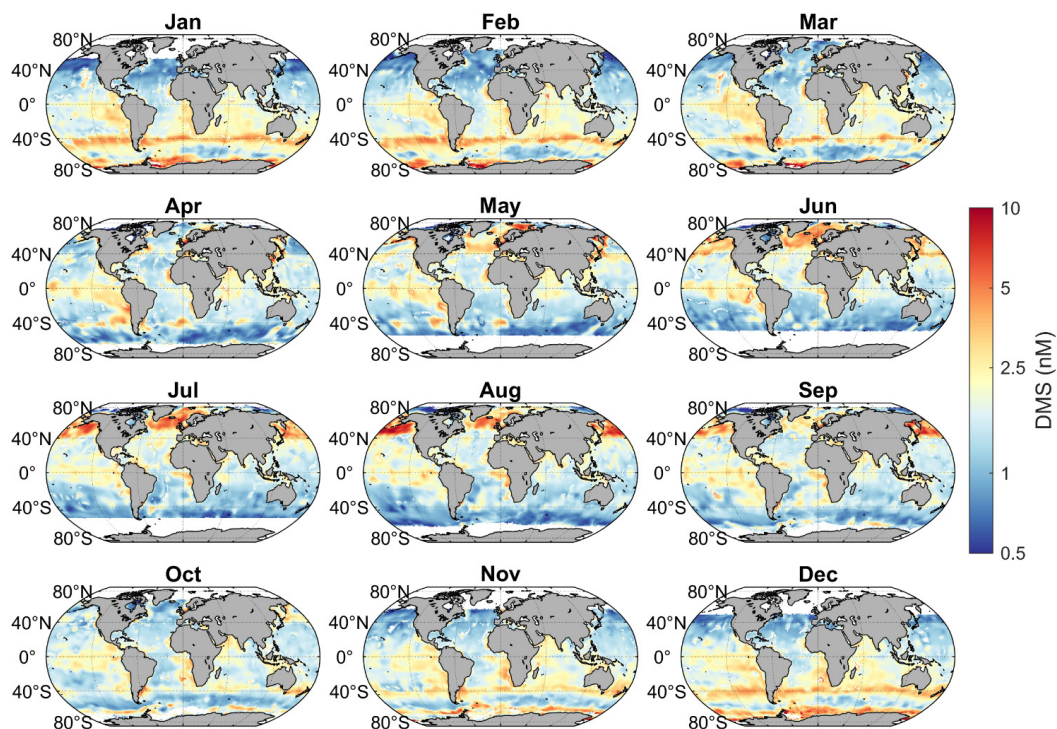



Figure 5. Monthly climatology of global sea surface DMS concentration during 1998 to 2017.

298

299

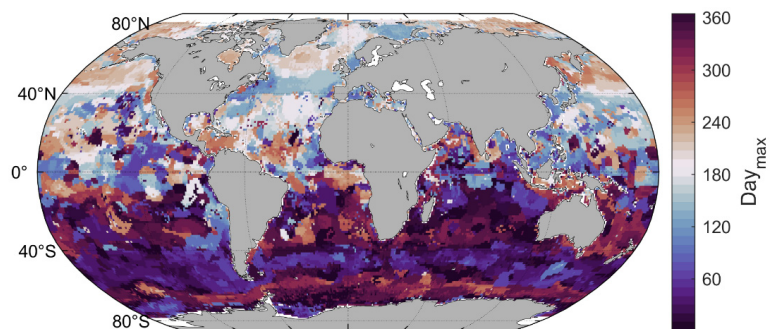



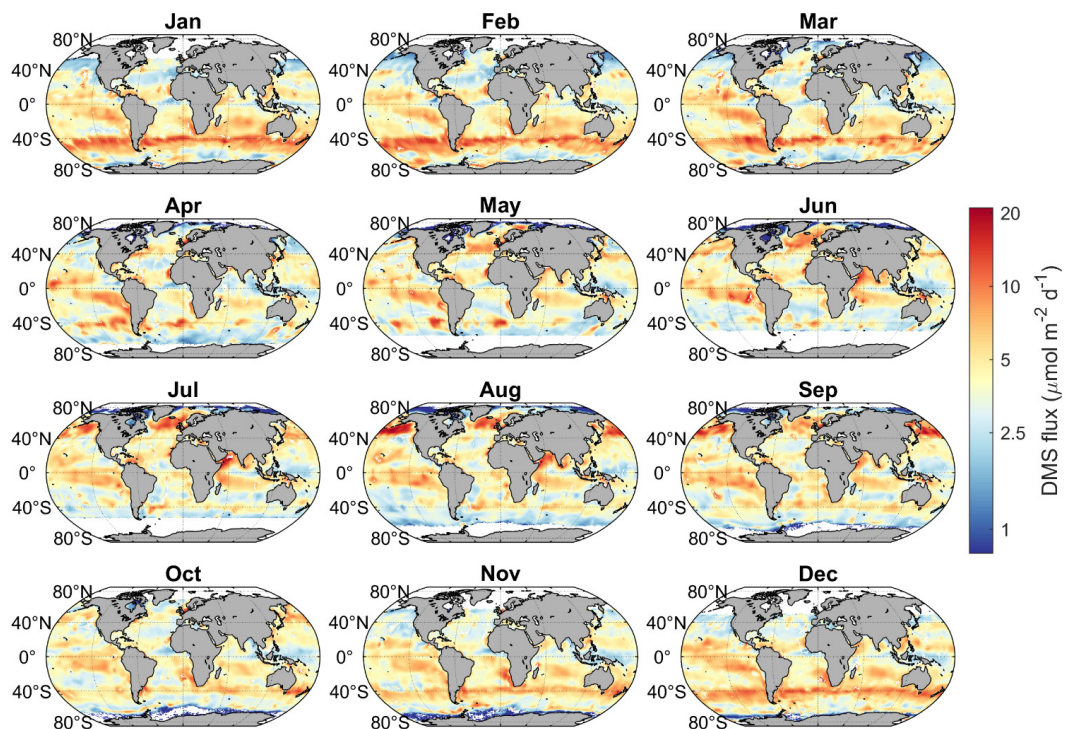
Figure 6. The day of the year with the highest sea surface DMS concentration for each grid.

301

302   The spatiotemporal variation of DMS emission flux is generally consistent with that of concentration. As shown in Fig. 7,
303   DMS fluxes are also significantly higher in summer in most mid-high latitude regions, and the high-flux regions generally
304   overlap with the hot spots of DMS concentration. This indicates that the distribution of sea surface DMS concentration is the
305   main factor controlling the distribution pattern of DMS emissions at the global scale, and the effect of transfer velocity is
306   secondary. However, in certain specific regions, the variations of DMS flux and concentration may be inconsistent. For
307   example, in the Arabian Sea and the central Indian Ocean, the high transfer velocities (Fig. S4) caused by higher wind speeds
308   during the period from June to September can significantly uplift the emission fluxes, although the concentrations are relatively
309   lower than in other months. In the polar regions, especially along the coast of Antarctica, although the DMS concentration is
310   high in summer, the sea ice coverage will greatly hinder the release of DMS, thus the emission flux remains at a lower level.

311   As shown in Fig. 8, the higher wind speeds in autumn and winter at mid to high latitudes result in higher total transfer velocities,
312   leading to smaller summer-to-winter ratios of DMS emission flux compared to that of DMS concentration. In low latitudes,
313   the existence of the trade wind zones in both hemispheres further differentiates two high-flux bands compared to concentration,
314   located between 5° to 20°. The emission fluxes in the equatorial region between these two trade zones are significantly lower.
315   Although the latitudinal distributions of mean DMS emission fluxes in the southern and northern hemispheres are almost
316   symmetrical, the huge difference in ocean area between the two hemispheres results in a significantly higher total emission
317   from the southern hemisphere. Since anthropogenic SO_2 emissions are mainly concentrated in the northern hemisphere,
318   oceanic DMS plays a much more important role in the southern hemisphere, especially over the regions south of 40° S where
319   the DMS emission is high and the perturbation of anthropogenic pollution is low.

320   According to our newly built DMS gridded dataset, the global area-weighted annual mean concentration of DMS at the sea
321   surface from 1998 to 2017 was ~1.82 nM (1.80–1.84 nM), which is at an intermediate level among the values (1.6 to 2.4 nM)
322   obtained by various methods in previous studies (Tesdal et al., 2016). The global annual mean DMS emission to the atmosphere
323   was 17.9 TgS yr^{-1} (17.7–18.4 TgS yr^{-1}), with 10.7 TgS yr^{-1} (59.4%) from the southern hemisphere and 7.3 TgS yr^{-1} (40.6%)
324   from the northern hemisphere. It is worth noting that there is no valid satellite Chl a data in high latitudes in winter, and the
325   Chl a data missing also exists in other regions, thus the simulated DMS concentrations are missing under these circumstances.
326   This part of missing data was filled in the calculation of global mean DMS concentration and total emission. The specific
327   method is shown in Appendix B. Without filling in the missing data, the global annual mean concentration and total emission
328   are 1.86 nM (1.84–1.88 nM) and 16.8 TgS yr^{-1} (16.6–17.2 TgS yr^{-1}), respectively.

329

331 **Figure 7.** Monthly climatology of global DMS sea-to-air flux from 1998 to 2017.

Earth System
Science
Data

334

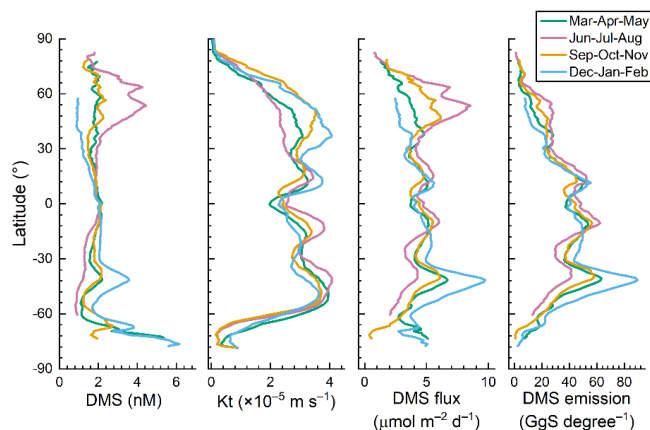



Figure 8. The latitudinal distributions of sea surface DMS concentration, total transfer velocity (Kt), sea-to-air flux, and total

emission in different seasons during 1998–2017. Only the latitudes in a season with no more than 75% of data missing are

shown in the figure.

3.2.2 Comparisons with other global DMS climatologies

Here we compare the distributions of DMS concentration derived from our ANN simulation (referred to as Z23) with four

previously constructed climatologies, including (1) L11: the widely used second version of interpolation/extrapolation-based

climatology established by Lana et al. (2011), (2) H22: an updated version of L11 incorporating much more DMS

measurements and using dynamic biogeochemical provinces (Hulswar et al., 2022), (3) G18: the DMS concentration field

estimated by a two-step remote sensing algorithm (Galí et al., 2018), and (4) W20: the previous DMS climatology simulated

by ANN (Wang et al., 2020).

In general, all datasets can reflect the basic pattern of high DMS concentration in summer and low concentration in winter, but

there are significant differences in their specific distributions. Due to the limitation of the method used, DMS_{L11} exhibits

relatively lower spatial heterogeneity (i.e., higher patchiness), which may not well capture the detailed spatial variability on a

regional scale. Compared with DMS_{L11}, DMS_{Z23} is significantly lower at high latitudes in summer and in the South Indian

Ocean and Southwest Pacific Ocean from December to February (Fig. 9a). Particularly in the southern polar region (Polar_S),

DMS_{L11} can reach > 10 nM in summer, which is 1–3 times higher than DMS_{Z23} (Fig. 9e). However, DMS_{Z23} around the

Antarctic in March remains the similar level as in summer, and it is significantly higher than DMS_{L11} as well as the other three

climatologies. DMS_{H22} shows fewer differences with DMS_{Z23} in northern high latitudes, the South Indian Ocean, and the

Southwest Pacific Ocean, but the summertime concentrations in most of Polar_S region are also > 2 nM higher than DMS_{Z23}

354   (Fig. 9b). The global annual mean DMS concentrations in L11 and H22 are 2.43 nM and 2.26 nM, about 33.5% and 24.2%

355   higher than Z23, respectively.

356   G18 exhibits the lowest global annual mean concentration (1.63 nM) among these climatologies, about 10.4% lower than Z23.

357   There are more than 60% of global oceans with higher DMS_{Z23} than DMS_{G18} in all seasons, which is especially obvious in

358   autumn and winter for both hemispheres (Fig. 9c). In addition, DMS_{Z23} is > 2 nM (> 100%) higher than DMS_{G18} in North

359   Pacific during boreal summer and near the Antarctic during austral summer and autumn. However, there are extremely high

360   DMS concentrations (> 5 nM) in some coastal seas (such as the coast of Patagonia and Peru, the southwestern coast of Africa,

361   the western coast of North America, and the Sahara Desert) based on the G18 estimate, and this characteristic was not

362   completely reproduced by other DMS fields. W20 exhibits the highest consistency with Z23 in spatiotemporal distribution

363   patterns as well as the lowest difference in global annual mean concentration (1.74 nM, only 4.4% lower than Z23). But in

364   specific regions, there are also obvious disagreements. For example, the summertime DMS_{Z23} is > 1 nM (> 40%) lower than

365   DMS_{W20} in more than half of the Arctic area, but in the southeastern Pacific and South Atlantic during austral summer and

366   autumn DMS_{Z23} is significantly higher than DMS_{W20} (Fig. 9e). Moreover, only DMS_{Z23} forms a nearly complete high-

367   concentration annular band at ~ 40° S during austral summer.

368

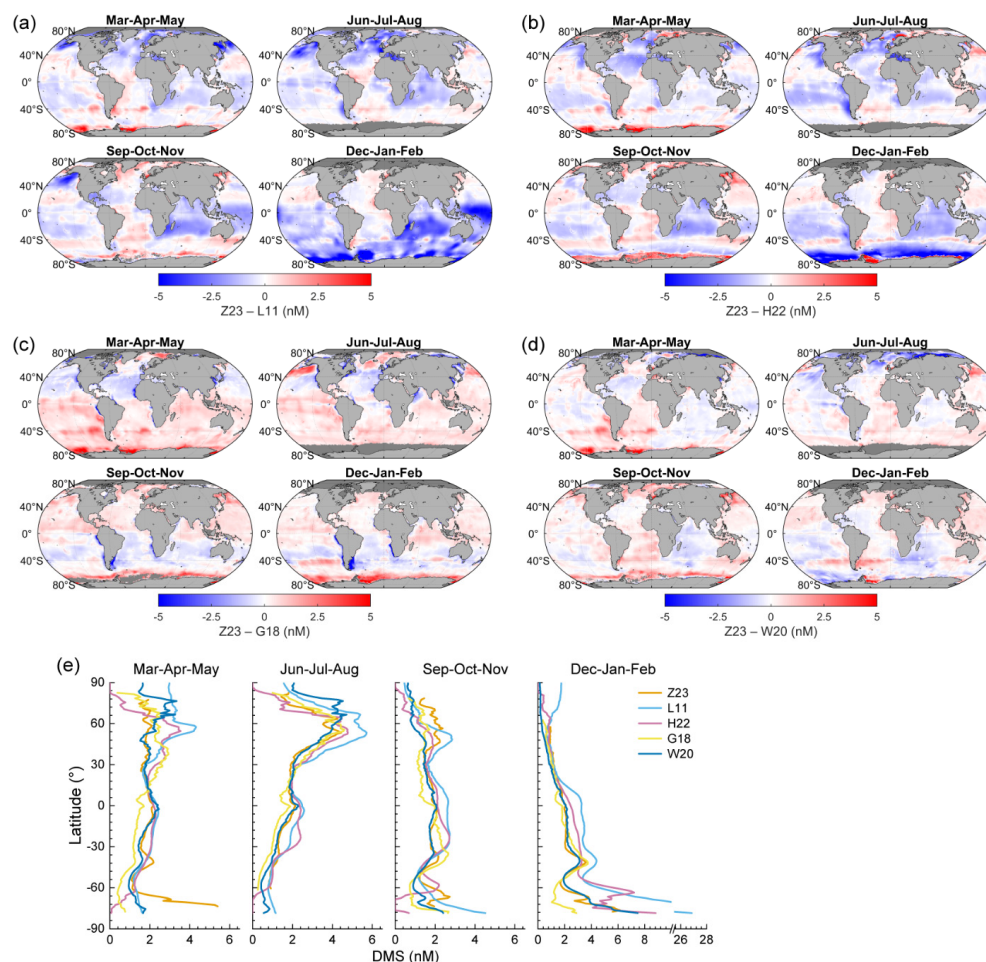



Figure 9. (a–d) The spatial distributions of DMS concentration differences between Z23 and four previously estimated fields
in different seasons: (a) L11, (b) H22, (c) G18, and (d) W20. Dark gray regions in the ocean represent data missing in at least
one field. (e) Comparisons between the latitudinal distributions of Z23 and four previous DMS fields in different seasons.

3.2.3 Decadal changes

One of the advantages of our ANN-derived DMS dataset is that it contains information on interannual variations. Here we
present the decadal trends of DMS concentration, Kt, and emission flux during 1998–2017 at global and regional scales.

376 Overall, the interannual variability of DMS concentration in most global oceanic regions is relatively small. 90% of the global
377 oceanic area exhibited a difference of less than 1 nM between the maximum and minimum annual average concentrations
378 during this 20-year period. As for the relative variation range ((maximum – minimum)/mean) of the annual average
379 concentration, it was lower than 0.2 in 66% of the global oceanic area, mainly distributed between the latitude 40° S and 40°
380 N. The annual mean concentration in the polar regions varied greatly (Fig. 10a). In particular, the Labrador Sea, Baffin Bay,
381 Barents Sea, Kara Sea, East Siberian Sea, and Chukchi Sea in the Arctic showed a significant increasing trend (> 0.03 nM yr^{-1}, $P < 0.05$). In contrast, some areas around the Antarctic showed a significant decreasing trend, such as the inner Weddell Sea
382
383 and Ross Sea (> 0.05 nM yr^{-1}), but the annual means in marginal seas of the Antarctic between 0°–180° E increased. In addition,
384 there was a significant increasing trend at ~50° S, but the rate of increase was relatively slow (< 0.02 nM yr^{-1}). In the low-mid
385 latitude regions where the variation range of DMS was small, the overall trend was decrease, and the large area of these regions
386 resulted in a mild decrease in global average concentration, with a rate of ~0.0012 nM yr^{-1} (Fig. 10d). The highest concentration
387 (1.84 nM) occurred in 2000, while the lowest concentration (1.80 nM) occurred in 2017. Due to the increasing SST and WS
388 in most mid-low latitude regions (Fig. S5), the Kt of DMS also showed an overall increasing trend, especially in the eastern
389 equatorial Pacific (Fig. 10b). The increase in Kt can offset the decrease in DMS concentration to some extent, resulting in no
390 significant trend in global DMS fluxes during this 20-year period (Fig. 10d).

391 In the Arctic region, which is the most sensitive to climate warming (Screen et al., 2012; Serreze and Barry, 2011), the sea ice
392 coverage has significantly decreased over the past 2 decades, particularly in the Barents Sea and Kara Sea and further north (>
393 1% yr^{-1} for annual mean SI, Fig. S5). The retreat of summertime sea ice leads to an expansion of open-sea surface, further
394 promoting the increase in DMS emission, which is consistent with the results of Galí et al. (2019). As shown in Fig. 11, the
395 annual DMS emission flux in the Polar_N region increased from 0.48 TgS yr^{-1} before 2002 to 0.58 TgS yr^{-1} in 2016. If we
396 just consider the fluxes during May–August north of 70° N, as in Galí et al. (2019), the lowest value was 37 GgS in 2003 and
397 the highest was 64 GgS in 2016. In contrast, the sea ice fraction in the Southern Ocean has significantly increased (Fig. S5),
398 leading to a significant decrease in the Kt and DMS emission flux (Fig. 10b-c). The highest annual total emission flux in the
399 Polar_S region occurred in 1999 (0.85 TgS) and the lowest occurred in 2013 (0.63 TgS), which decreased by ~26%. For other
400 oceanic regions, the average DMS concentrations in the Coastal, Westerlies_N_Pacific, and all 3 Trades regions exhibit
401 decreasing trends over the past 20 years, while the concentration in Westerlies_S has increased ($P < 0.05$, Fig. 11). As for the
402 DMS flux, only the Westerlies_N_Atlantic and Trades_Pacific regions showed a slight increase, while there was no significant
403 trend in other low-mid latitude regions.

404

405

406

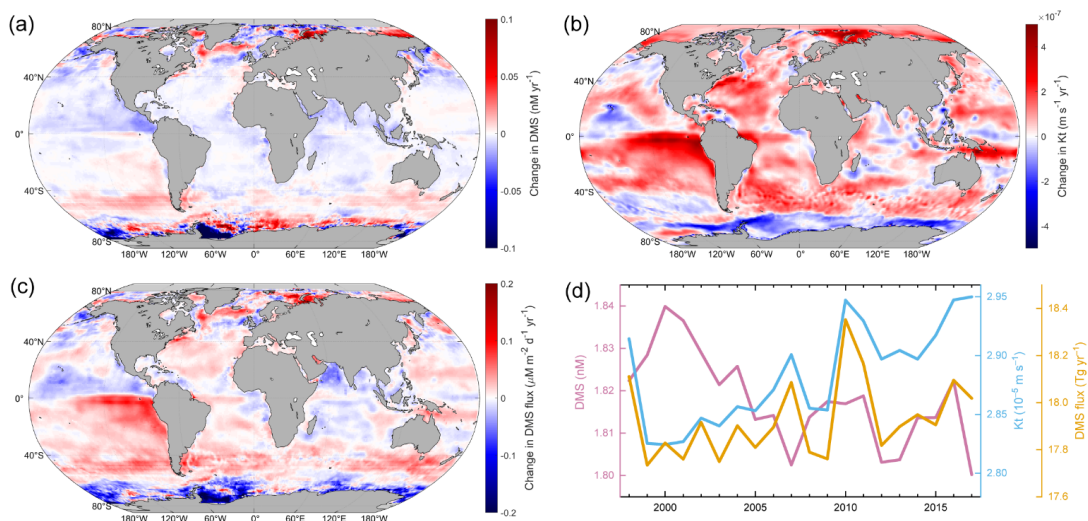



Figure 10. (a–c) The spatial distributions of changes in (a) DMS concentration, (b) Kt, and (c) DMS emission flux from 1998 to 2017. The linear regression slopes for the annual means are taken as the changing rates here. (d) The temporal changes of global annual mean DMS concentration, Kt, and total emission flux from 1998 to 2017.

410

411

Figure 11. The temporal changes of annual mean DMS concentration, Kt, and total emission flux in different regions during
1998 to 2017.

3.3 Connection with atmospheric biogenic sulfur

One of the main aims of developing this daily gridded DMS dataset (Z23) covering multiple years is to provide a better
emission inventory of marine biogenic DMS and effectively improve the modeling performance for atmospheric sulfur
chemistry, especially for simulating sulfur aerosols. In order to investigate whether our newly constructed DMS dataset can
reach this goal, we employed a backward trajectory-based method to establish the connection between sea surface DMS
emissions and DMS's oxidation products in the atmosphere. The correlation was compared to those calculated by applying
previously reported DMS climatologies (L11, H22, G18, and W20).

Here we use the observed concentrations of particulate methanesulfonic acid (MSA) over the Atlantic Ocean as a reference.
MSA is one of the major end-products of DMS in the atmosphere and is solely from the oxidation of marine biogenic DMS
over remote oceans (Saltzman et al., 1983; Savoie et al., 2002; Osman et al., 2019). Therefore, there is likely to be a dependence

425 of the variation of MSA concentration on the DMS emission fluxes. During four transection cruises in the Atlantic conducted

426 by *R/V Polarstern* (20 April – 20 May 2011, 28 October – 1 December 2011, 10 April – 15 May 2012, and 27 October – 27

427 November 2012), the MSA concentrations in submicron aerosols were measured online using a High-Resolution Time-of-

428 Flight Aerosol Mass Spectrometer with the time resolution of 20 min. We applied 1-h averaged data to match the air mass

429 backward trajectories calculated every hour. The ship tracks are shown in Fig. S6, and detailed information about the cruises

430 and measurement method was given by Huang et al. (2016). The 72-hour air mass backward trajectories reaching the ship

431 position were calculated every hour by the HYSPLIT model with the start height of 100 m (Stein et al., 2015), and then the air

432 mass exposure to DMS emission (AEDMS), i.e., the weighted average of DMS emission flux along the trajectory path was

433 calculated following the approach of Zhou et al. (2021). We used 5 different DMS gridded datasets, including Z23, L11, H22,

434 G18, and W20. For Z23, the calculated daily DMS fluxes were used. For the other 4 monthly climatologies, we applied the

435 daily Kt data from Z23 to calculate the DMS fluxes to exclude the confounding influences by different Kt parameterizations.

436 In this calculation, the same concentration was assigned to all days in a month without interpolation. The detailed calculation

437 procedures for AEDMS are illustrated in Appendix C.

438 MSA concentrations were significantly higher in late spring than those in autumn for both North and South Atlantic Oceans

439 (Fig. 12a). For example, during the boreal spring cruise in 2011, the average MSA concentration over the North Atlantic (0.068

440 μg m^{-3}, north of 25° N) was about an order of magnitude higher than the average concentration over the South Atlantic (0.006

441 μg m^{-3}, south of 5° S). In contrast, during the boreal autumn cruise in 2011, the average concentration over the South Atlantic

442 (0.034 μg m^{-3}, south of 5° S) was ~5 times higher than that over the North Atlantic (0.006 μg m^{-3}, north of 25° N). In addition

443 to this major seasonal pattern, there was also a minor MSA concentration peak between 5°–15° N in both seasons. The spatial

444 and seasonal variations of AEDMS based on the Z23 dataset (referred to as AEDMS_Z23) largely coincided with these MSA

445 concentration patterns (Fig. 12a). It should be noted that the AEDMS/MSA ratio between 5°–15° N was significantly higher

446 than those in other high-MSA regions, which may result from the DMS simulation biases near the coast of West Africa or the

447 lower DMS-to-MSA conversion yields related with air temperature and oxidant species (Barnes et al., 2006; Bates et al., 1992).

448 There were also several AEDMS peaks in North Atlantic during November 2012, inconsistent with the continuously low MSA

449 concentrations. Given the high precipitation rates along the trajectory, a strong wet scavenging process might significantly

450 reduce aerosol concentrations (Wood et al., 2017).

451 The AEDMS derived from other DMS concentration fields showed similar variations as AEDMS_Z23 (Fig. 12a). It was not

452 surprising since all DMS concentration fields exhibit similar large-scale spatiotemporal patterns, and the same air mass

453 transport path and Kt were applied in different AEDMS calculations. However, because those DMS monthly climatologies

454 have low temporal resolutions and lack interannual changes, the resulting AEDMS was less effective in capturing the

455 variability in fine scale or among different years. Here we focus on the high-MSA periods to elaborate on this issue, which

456 corresponds to the latitudes north of 25° N in boreal spring (S1 and S2 in Fig. 12a) and south of 5° N in boreal autumn (A1

457 and A2 in Fig. 12a). As shown in Fig. 12b–c, hourly MSA concentrations exhibited significantly stronger correlations with
458 AEDMS_Z23 than with other AEDMS time series in both seasons, which means AEDMS_Z23 can explain much more (0.78–
459 2.71 times higher) variance of MSA concentration. It was even more obvious for 6-h averaged data, as AEDMS_Z23 can
460 explain ~70% of the variance during S1 and S2, while others can only explain ~30% (Fig. S7). These remarkably stronger
461 connections between Z23 and atmospheric DMS-derived aerosols mainly benefited from the combined effects of higher time
462 resolution and accuracy of DMS concentrations as well as the incorporation of interannual variations. For example, the ratio
463 of average MSA concentration during S1 to that during S2 (S1-to-S2 ratio) was 1.89, and the A2-to-A1 ratio was 1.58.
464 AEDMS_Z23 exhibited a slightly lower but still significant interannual variation degree, where the S1-to-S2 ratio and A2-to-
465 A1 ratio were 1.59 and 1.34, respectively. However, this interannual variation cannot be reproduced by other datasets, where
466 the S1-to-S2 ratio and A2-to-A1 ratio were in the range of 1.09–1.32 and 1.05–1.11, respectively. The results further prove the
467 fidelity of our newly developed DMS gridded data product, which has the prospect of greatly improving the modeling
468 performance for atmospheric DMS processes compared with previously reported climatologies. Of course, the AEDMS
469 method used here is a highly simplified approach without considering the complex DMS chemistry in the atmosphere, and the
470 intercomparisons based on chemical transport models can be used in the future to obtain a more straightforward conclusion.

471

472

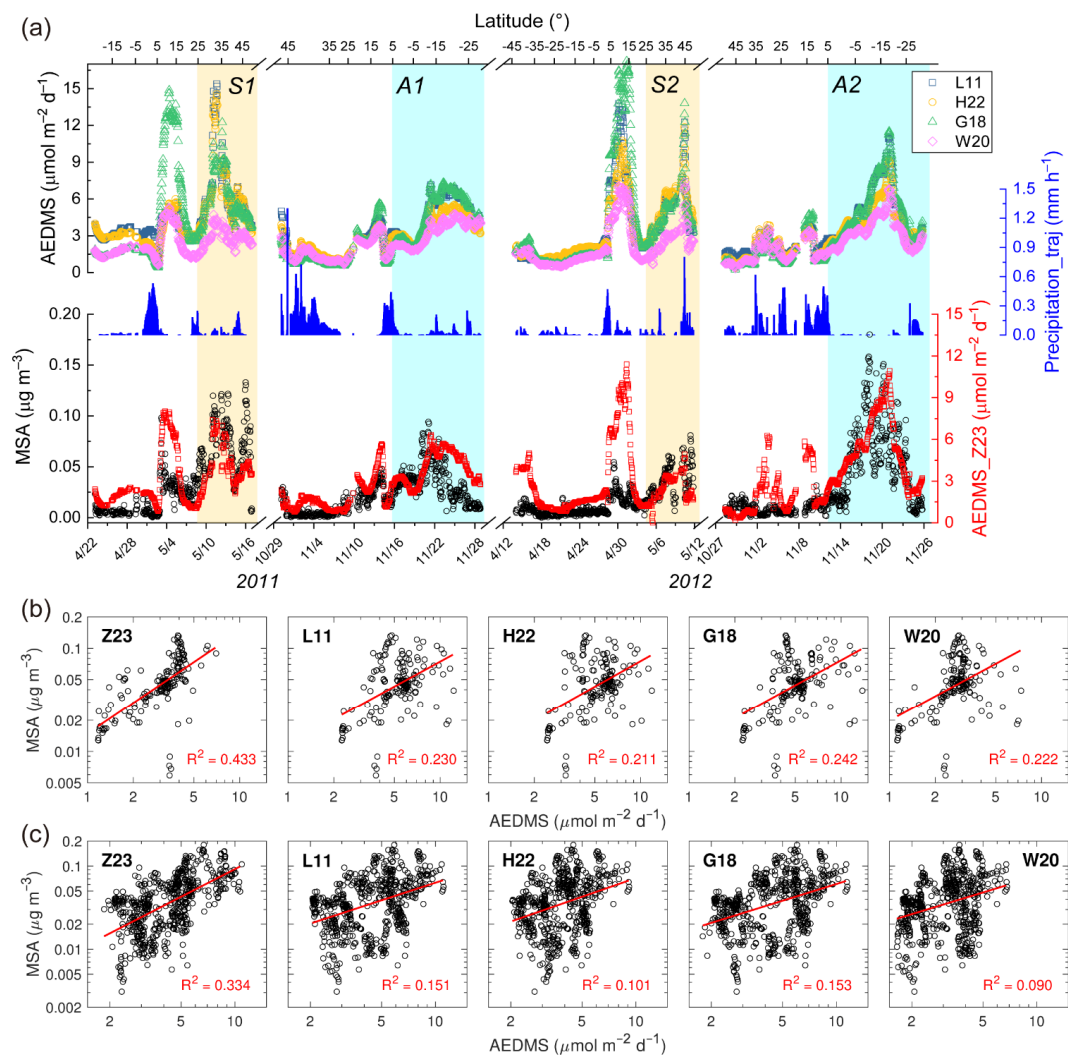



Figure 12. (a) Time series of observed MSA concentration, AEDMS calculated based on different DMS concentration datasets,

and average precipitation along the backward trajectory (Precipitation_traj) during four Atlantic cruises in 2011–2012. (b–c)

Correlations between hourly MSA concentration and AEDMS based on different DMS concentration datasets (b) in the region

north of 25° N in boreal spring (the shaded area S1 and S2 in panel a) and (c) in the region south of 5° N in boreal autumn (the

477   shaded area A1 and A2 in panel a). Data points during the periods with air mass time fraction within the boundary layer less

478   than 90% were removed.

479   **4 Uncertainties and limitations**

480   Although our ANN ensemble model and derived DMS dataset demonstrate certain advantages compared to previous studies,

481   as discussed in Section 3.3, there are still notable uncertainties and limitations, which result in the ~38% uncaptured variance

482   (Figure 3a) and non-negligible simulation biases. First, the input factor datasets from different sources and the observed sea

483   surface DMS concentrations inherently possess certain uncertainties, which can introduce biases in the ANN learning process.

484   Second, ANN models may not fully capture all intricate data features, and the outcomes obtained from each training can exhibit

485   certain randomness. In this study, the average standard deviation of simulated $\log_{10}$DMS values from 100 neural networks was

486   0.256, and the 5%-95% range of the coefficient of variation for DMS concentration was 0.36–1.96, with a mean of 1.06. If the

487   detailed uncertainties associated with each data source are known, future investigations could employ Monte Carlo methods

488   to estimate the uncertainties of final results arising from the aforementioned two factors (Abdar et al., 2021; Moradkhani et al.,

489   2012). Third, while the DMS observational data covers all major oceanic basins, certain regions such as the Southeast Pacific

490   remain underrepresented. Advances in online measurement technologies offer promising opportunities for acquiring more

491   extensive and convenient observational data (Hulswar et al., 2022). In the future, more observations are needed for these

492   underrepresented regions, and the model can be retrained and updated accordingly.

493   The ideal scenario for this study would entail having all input datasets covering a 20-year period with a daily resolution.

494   However, this is not the case for the nutrients and DO datasets utilized here, and only a portion of the Chl *a* datasets are

495   available at a daily resolution, potentially impacting the model performance to some extent. Nevertheless, machine learning

496   techniques can capture the interactions among various input factors. Two key factors, SST and MLD, influence the surface

497   concentration of these species, and we provided the multi-year daily datasets of SST and MLD. Therefore, during the training

498   process, the model can learn some of the daily and interannual variation information regarding nutrients and DO, as well as

499   their impact on DMS concentration, thereby partially compensating for the model performance. Of course, in the future, the

500   utilization of accurate daily resolution data instead of monthly climatologies may yield improved simulation results.

501   Additionally, apart from the 9 variables incorporated in this study, other environmental parameters, such as trace metal

502   elements (Li et al., 2021), can also influence the DMS concentration. Not incorporating these factors may introduce potential

503   biases. Thus, further field measurements of these factors are necessary to comprehend their spatiotemporal distributions, which

504   are likely to enhance the model's ability to simulate sea surface DMS concentrations.

505   When using our newly developed DMS dataset, two issues should be aware of. First, there is a notable amount of missing data

506   during winter in polar regions due to the unavailability of satellite Chl *a*, which restricts the investigation of subsequent DMS

507   atmospheric processes related to DMS. Nevertheless, given the low phytoplankton biomass and extensive sea ice coverage

508 during winter, DMS emissions are typically at the lowest level of the year, thus the data missing has a relatively small impact.
509 Second, the ANN ensemble model exhibits limited capacity in accruately reproducing extremely high and low concentrations
510 of DMS (Fig. 3a). Therefore, in some nearshore areas with intensive biological activity, this dataset may potentially
511 underestimate the actual DMS concentrations.

512 **5 Code and data availability**

513 The generated gridded datasets of DMS concentration, total transfer velocity, and flux have been deposited at
514 https://zenodo.org/record/10279659 (Zhou et al., 2023) and can be downloaded publicly. The ANN model code and the Matlab
515 scripts for data analysis are available from https://zenodo.org/record/8077751 (Zhou, 2023).

516 **6 Conclusion**

517 Based on the global sea surface DMS observations and associated data of 9 relevant environmental variables, an ANN
518 ensemble model was trained. The ANN model effectively captures the variability of DMS concentrations and demonstrates
519 good simulation accuracy. Leveraging this ANN model, a global sea surface DMS gridded dataset with a daily resolution
520 spanning 20 years (1998–2017) was constructed. The global annual average concentration was ~1.82 nM, falling within the
521 range of previous estimates, and the annual total emission was ~17.9 TgS yr^{-1}. High DMS concentrations and fluxes took place
522 in summer in North Pacific (40°–60° N), North Atlantic (50°–80° N), the annular band of ~40° S, and the Southern Ocean.
523 With this newly developed dataset, the day-to-day changes and interannual variations can be investigated. The global annual
524 average concentration shows a mild decreasing trend (~0.0012 nM yr^{-1}), while the total emission remains stable. There were
525 more significant decadal changes in certain regions. Specifically, the annual DMS emission in the Arctic and Antarctic regions
526 showed opposite trends.

527 To further validate the robustness and advantages of our new dataset, an airmass trajectory-based approach was applied to link
528 the DMS flux and atmospheric MSA concentration. Compared to previous monthly climatologies, the airmass exposure to
529 DMS calculated using our new dataset explains a greater amount of variance in atmospheric MSA concentration over the
530 Atlantic Ocean. Therefore, despite the presence of uncertainties and limitations, the new dataset serves as an improved DMS
531 emission inventory for atmospheric models and holds the potential to significantly enhance the simulation of DMS-induced
532 aerosols and their associated climatic effects.

533

534 **Appendix A: Acronyms**
535 AEDMS   Air mass exposure to DMS emission
536 ANN   Artificial neural network
537 BLH   Boundary layer height

538	CCN	Cloud condensation nuclei
539	Chl *a*	Chlorophyll *a*
540	DMS	Dimethyl sulfide
541	DMSP	Dimethylsulfoniopropionate
542	DO	Dissolved oxygen
543	DSWF	Downward short-wave radiation flux
544	ECCO	Estimating the Circulation and Climate of the Ocean
545	GSSD database	Global Surface Seawater DMS database
546	Kt	Total transfer velocity
547	MLD	Mixed layer depth
548	MSA	Methanesulfonic acid
549	MSE	Mean squared error
550	NAAMES	North Atlantic Aerosols and Marine Ecosystems Study
551	RMSE	Rooted mean squared error
552	SI	Sea ice fraction
553	SST	Sea surface temperature
554	SSS	Sea surface salinity
555	WOA18	World Ocean Atlas 2018
556	WOD	World Ocean Database
557	WS	Wind speed

558

Appendix B: Filling the missing DMS concentration data

559

560 In the calculation of global mean DMS concentration and total emission flux, the missing data was complemented based on

561 the adjacent valid values. First, for a certain global DMS gridded dataset, we calculated the proportion of missing data for each

562 latitude and identified the absolute minimum latitudes with missing proportions above 0.5 for the southern and northern

563 hemispheres, denoted as *Lat1* and *Lat2*, respectively (assuming Lat1 is for the southern hemisphere with negative value, and

564 *Lat2* is for the northern hemisphere with positive value). Next, for latitudes outside the range of *Lat1* and *Lat2* (including *Lat1*

565 and *Lat2*), the missing data were replaced by the average of all valid values with latitude < *Lat1* + 10° or > *Lat2* – 10°. For the

566 latitudes between Lat1 and Lat2, the missing data at a certain grid was replaced by the average value of all adjacent grid points

567 within 5° differences in both longitude and latitude. This operation was iterated until all such missing values were filled.

568

Appendix C: The calculation of airmass exposure to DMS emission (AEDMS)

569

570 Here the AEDMS index followed the similar calculation of the air mass exposure to Chl *a* (AEC) in previous studies (Arnold

571 et al., 2010; Park et al., 2018; Zhou et al., 2021). We adopted the same approach presented in Zhou et al. (2021), only replacing

572 the Chl *a* concentration by DMS flux, as shown in the following equation (A2):

573 $$AEDMS = \frac{\sum_{i=0}^{72} DMS\,flux_i \cdot e^{-\frac{t_i}{72}} \cdot \frac{600}{BLH}}{n} \tag{A2}$$

574 Here *i* represents the *i*-th trajectory point of the 72-hour backward trajectory (0-th for the receptor point). $DMS\,flux_i$ represents

575 the mean DMS flux within a radius of 20 km at the location of *i*-th trajectory point. $DMS\,flux_i$ is set to zero if the point locates

576 on land or the air mass pressure is below 850 hPa (usually in the free troposphere with little influence of surface emission). t_i

577 is the tracking time of the trajectory point (unit: hour) and $e^{-\frac{t_i}{72}}$ is the weighting factor to assign higher values for regions

578 closer to the receptor point. *n* is the total number of trajectory points with valid DMS flux (including 0). To better connect with

579 the atmospheric concentrations in the marine boundary layer, the normalization by boundary layer height (BLH) is added by

580 the $\frac{600}{BLH}$ term. The BLH below 50 m is replaced by 50 m.

Author contributions.

582 SZ and YC designed the research. SZ, FW, ZX, and KY collected the data and did the data preprocessing. SZ implemented

583 the model development and performed the simulation with assistance from GY, HZ, and YZ. SH, HH, AW, and LP provided

584 the measurement data of atmospheric MSA over the Atlantic Ocean. SZ conducted the data analysis and visualization with

585 advice from YC and XG. SZ and YC wrote the manuscript with inputs from all authors.

Competing interests.

587 The authors declare that they have no conflict of interest.

Acknowledgements.

589 We greatly thank National Oceanic and Atmospheric Administration's Pacific Marine Environmental Laboratory for

590 maintaining the Global Surface Seawater DMS Database. We acknowledge Dr. Chenzhao Li for sharing the code of global

591 sensitivity analysis and Dr. Martin Johnson for sharing the code of DMS transfer velocity calculation. We also thank Dr. Rich

592 Pawlowicz for developing and sharing the M_Map toolbox for Matlab (https://www.eoas.ubc.ca/~rich/map.html), which was

593 used in the mapping of this study. XG was supported by the Research Center for Industries of the Future (RCIF) at Westlake

594 University and Westlake University Education Foundation.

Financial support.

This work is jointly supported by Natural Science Foundation of Shanghai (22ZR1403800), National Key Research and Development Program of China (2016YFA0601304), and National Natural Science Foundation of China (41775145).

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Information Fusion, 76, 243-297, 10.1016/j.inffus.2021.05.008, 2021.

Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D. S., and Vardi, A.: Identification of the algal dimethyl sulfide–releasing enzyme: a missing link in the marine sulfur cycle, Science, 348, 1466-1469, 2015.

Andreae, M. O.: Ocean-Atmosphere Interactions in the Global Biogeochemical Sulfur Cycle, Mar. Chem., 30, 1-29, Doi 10.1016/0304-4203(90)90059-L, 1990.

Arnold, S. R., Spracklen, D. V., Gebhardt, S., Custer, T., Williams, J., Peeken, I., and Alvain, S.: Relationships between atmospheric organic compounds and air-mass exposure to marine biology, Environ. Chem., 7, 232-241, 10.1071/en09144, 2010.

Barnes, I., Hjorth, J., and Mihalopoulos, N.: Dimethyl sulfide and dimethyl sulfoxide and their oxidation in the atmosphere, Chem. Rev., 106, 940-975, 10.1021/cr020529+, 2006.

Bates, T. S., Calhoun, J. A., and Quinn, P. K.: Variations in the Methanesulfonate to Sulfate Molar Ratio in Submicrometer Marine Aerosol-Particles over the South-Pacific Ocean, J. Geophys. Res.-Atmos., 97, 9859-9865, 10.1029/92JD00411, 1992.

Beale, R., Johnson, M., Liss, P. S., and Nightingale, P. D.: Air–Sea Exchange of Marine Trace Gases, in: Treatise on Geochemistry (Second Edition), edited by: Holland, H. D., and Turekian, K. K., 2, Elsevier, Oxford, 53-92, 2014.

Behrenfeld, M. J., Moore, R. H., Hostetler, C. A., Graff, J., Gaube, P., Russell, L. M., Chen, G., Doney, S. C., Giovannoni, S., Liu, H., Proctor, C., Bolaños, L. M., Baetge, N., Davie-Martin, C., Westberry, T. K., Bates, T. S., Bell, T. G., Bidle, K. D., Boss, E. S., Brooks, S. D., Cairns, B., Carlson, C., Halsey, K., Harvey, E. L., Hu, C., Karp-Boss, L., Kleb, M., Menden-Deuer, S., Morison, F., Quinn, P. K., Scarino, A. J., Anderson, B., Chowdhary, J., Crosbie, E., Ferrare, R., Hair, J. W., Hu, Y., Janz, S., Redemann, J., Saltzman, E., Shook, M., Siegel, D. A., Wisthaler, A., Martin, M. Y., and Ziemba, L.: The North Atlantic Aerosol and Marine Ecosystem Study (NAAMES): Science Motive and Mission Overview, Front. Mar. Sci., 6, 10.3389/fmars.2019.00122, 2019.

Bell, T. G., Porter, J. G., Wang, W.-L., Lawler, M. J., Boss, E., Behrenfeld, M. J., and Saltzman, E. S.: Predictability of Seawater DMS During the North Atlantic Aerosol and Marine Ecosystem Study (NAAMES), Front. Mar. Sci., 7, 596763, 10.3389/fmars.2020.596763, 2021.

626   Belviso, S., Bopp, L., Moulin, C., Orr, J. C., Anderson, T. R., Aumont, O., Chu, S., Elliott, S., Maltrud, M. E., and Simó, R.:
627   Comparison of global climatological maps of sea surface dimethyl sulfide, Glob. Biogeochem. Cycles, 18,
628   10.1029/2003gb002193, 2004a.

629   Belviso, S., Moulin, C., Bopp, L., and Stefels, J.: Assessment of a global climatology of oceanic dimethylsulfide (DMS)
630   concentrations based on SeaWiFS imagery (1998-2001), Canadian Journal of Fisheries and Aquatic Sciences, 61, 804-816,
631   10.1139/f04-001, 2004b.

632   Belviso, S., Masotti, I., Tagliabue, A., Bopp, L., Brockmann, P., Fichot, C., Caniaux, G., Prieur, L., Ras, J., Uitz, J., Loisel,
633   H., Dessailly, D., Alvain, S., Kasamatsu, N., and Fukuchi, M.: DMS dynamics in the most oligotrophic subtropical zones of
634   the global ocean, Biogeochemistry, 110, 215-241, 10.1007/s10533-011-9648-1, 2011.

635   Bergen, K. J., Johnson, P. A., de Hoop, M. V., and Beroza, G. C.: Machine learning for data-driven discovery in solid Earth
636   geoscience, Science, 363, eaau0323, 10.1126/science.aau0323, 2019.

637   Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V.,
638   Woodhouse, M. T., Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing,
639   Nature, 503, 67-71, 10.1038/nature12674, 2013.

640   Charlson, R. J., Lovelock, J. E., Andreaei, M. O., and Warren, S. G.: Oceanic phytoplankton, atmospheric sulphur, cloud
641   albedo and climate, Nature, 326, 655-661, 10.1038/326655a0, 1987.

642   Chen, Q., Sherwen, T., Evans, M., and Alexander, B.: DMS oxidation and sulfur aerosol formation in the marine troposphere:
643   a focus on reactive halogen and multiphase chemistry, Atmos. Chem. Phys., 18, 13617-13637, 10.5194/acp-18-13617-2018,
644   2018.

645   Dubitzky, W., Granzow, M., and Berrar, D. P.: Fundamentals of data mining in genomics and proteomics, Springer Science
646   & Business Media, 2007.

647   Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M., and Wunsch, C.: ECCO version 4: An integrated framework
648   for non-linear inverse modeling and global ocean state estimation, Geosci. Model Dev., 8, 3071-3104, 2015.

649   Fung, K. M., Heald, C. L., Kroll, J. H., Wang, S., Jo, D. S., Gettelman, A., Lu, Z., Liu, X., Zaveri, R. A., Apel, E. C., Blake,
650   D. R., Jimenez, J. L., Campuzano-Jost, P., Veres, P. R., Bates, T. S., Shilling, J. E., and Zawadowicz, M.: Exploring dimethyl
651   sulfide (DMS) oxidation and implications for global aerosol radiative forcing, Atmos. Chem. Phys., 22, 1549-1573,
652   10.5194/acp-22-1549-2022, 2022.

653   Galí, M., Devred, E., Levasseur, M., Royer, S.-J., and Babin, M.: A remote sensing algorithm for planktonic
654   dimethylsulfoniopropionate (DMSP) and an analysis of global patterns, Remote Sensing of Environment, 171, 171-184,
655   10.1016/j.rse.2015.10.012, 2015.

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

658   Galí, M., Devred, E., Babin, M., and Levasseur, M.: Decadal increase in Arctic dimethylsulfide emission, P. Natl. Acad. Sci.
659   USA, 116, 19311-19317, 10.1073/pnas.1904378116, 2019.

660 Galí, M., Lizotte, M., Kieber, D. J., Randelhoff, A., Hussherr, R., Xue, L., Dinasquet, J., Babin, M., Rehm, E., and Levasseur,
661 M.: DMS emissions from the Arctic marginal ice zone, Elementa: Science of the Anthropocene, 9,
662 10.1525/elementa.2020.00113, 2021.

663 Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, M., Zweng, M., Mishonov, A., Baranova, O., and Seidov,
664 D.: World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Dissolved Oxygen Saturation.
665 A. Mishonov Technical Ed.; NOAA Atlas NESDIS 83, 38 pp., 2019a.

666 Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, M., Zweng, M., Mishonov, A., Baranova, O., and Seidov,
667 D.: World ocean atlas 2018. Vol. 4: Dissolved inorganic nutrients (phosphate, nitrate and nitrate+ nitrite, silicate). A. Mishonov
668 Technical Ed.; NOAA Atlas NESDIS 84, 35 pp., 2019b.

669 Hoffmann, E. H., Tilgner, A., Schroedner, R., Bräuer, P., Wolke, R., and Herrmann, H.: An advanced modeling study on the
670 impacts and atmospheric implications of multiphase dimethyl sulfide chemistry, P. Natl. Acad. Sci. USA, 113, 11776-11781,
671 10.1073/pnas.1606320113, 2016.

672 Holder, C., Gnanadesikan, A., and Aude-Pradal, M.: Using neural network ensembles to separate ocean biogeochemical and
673 physical drivers of phytoplankton biogeography in Earth system models, Geosci. Model Dev., 15, 1595-1617, 10.5194/gmd-
674 15-1595-2022, 2022.

675 Huang, B., Liu, C., Freeman, E., Graham, G., Smith, T., and Zhang, H.-M.: Assessment and Intercomparison of NOAA Daily
676 Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1, Journal of Climate, 34, 7421-7441, 10.1175/jcli-d-
677 21-0001.1, 2021.

678 Huang, S., Poulain, L., van Pinxteren, D., van Pinxteren, M., Wu, Z., Herrmann, H., and Wiedensohler, A.: Latitudinal and
679 Seasonal Distribution of Particulate MSA over the Atlantic using a Validated Quantification Method with HR-ToF-AMS,
680 Environ. Sci. Technol., 51, 418-426, 10.1021/acs.est.6b03186, 2016.

681 Hulswar, S., Simó, R., Galí, M., Bell, T. G., Lana, A., Inamdar, S., Halloran, P. R., Manville, G., and Mahajan, A. S.: Third
682 revision of the global surface seawater dimethyl sulfide climatology (DMS-Rev3), Earth System Science Data, 14, 2963-2987,
683 10.5194/essd-14-2963-2022, 2022.

684 Johnson, M. T.: A numerical scheme to calculate temperature and salinity dependent air-water transfer velocities for any gas,
685 Ocean Sci., 6, 913-932, 10.5194/os-6-913-2010, 2010.

686 Keller, M. D., Bellows, W. K., and Guillard, R. R.: Dimethyl sulfide production in marine phytoplankton, in: Biogenic Sulfur
687 in the Environment, edited by: Saltzman, E. S., and Cooper, W. J., ACS Publications, 1989.

688 Kettle, A. J., Andreae, M. O., Amouroux, D., Andreae, T. W., Bates, T. S., Berresheim, H., Bingemer, H., Boniforti, R., Curran,
689 M. A. J., DiTullio, G. R., Helas, G., Jones, G. B., Keller, M. D., Kiene, R. P., Leck, C., Levasseur, M., Malin, G., Maspero,
690 M., Matrai, P., McTaggart, A. R., Mihalopoulos, N., Nguyen, B. C., Novo, A., Putaud, J. P., Rapsomanikis, S., Roberts, G.,
691 Schebeske, G., Sharma, S., Simo, R., Staubes, R., Turner, S., and Uher, G.: A global database of sea surface dimethylsulfide
692 (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month, Glob.
693 Biogeochem. Cycles, 13, 399-444, 10.1029/1999gb900004, 1999.

694 Kloster, S., Feichter, J., Maier-Reimer, E., Six, K. D., Stier, P., and Wetzel, P.: DMS cycle in the marine ocean-atmosphere

695 system–a global model study, Biogeosciences, 3, 29-51, 2006.

696 Lana, A., Bell, T. G., Simó, R., Vallina, S. M., Ballabrera-Poy, J., Kettle, A. J., Dachs, J., Bopp, L., Saltzman, E. S., Stefels,

697 J., Johnson, J. E., and Liss, P. S.: An updated climatology of surface dimethylsulfide concentrations and emission fluxes in the

698 global ocean, Glob. Biogeochem. Cycles, 25, GB1004, 10.1029/2010gb003850, 2011.

699 Li, H., Zhou, S., Zhu, Y., Zhang, R., Wang, F., Bao, Y., and Chen, Y.: Atmospheric Deposition Promotes Relative Abundances

700 of High-Dimethylsulfoniopropionate Producers in the Western North Pacific, Geophys. Res. Lett., 48, e2020GL092077,

701 10.1029/2020GL092077, 2021.

702 Longhurst, A. R.: Ecological Geography of the Sea, Academic Press, 1998.

703 Lovelock, J. E., Maggs, R. J., and Rasmussen, R. A.: Atmospheric Dimethyl Sulphide and the Natural Sulphur Cycle, Nature,

704 237, 452-453, 10.1038/237452a0, 1972.

705 Mansour, K., Decesari, S., Ceburnis, D., Ovadnevaite, J., and Rinaldi, M.: Machine learning for prediction of daily sea surface

706 dimethylsulfide concentration and emission flux over the North Atlantic Ocean (1998-2021), Sci. Total. Environ., 871, 162123,

707 10.1016/j.scitotenv.2023.162123, 2023.

708 Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.

709 I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, R.

710 e.: IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment

711 Report of the Intergovernmental Panel on Climate Change, 2021.

712 McCoy, D. T., Burrows, S. M., Wood, R., Grosvenor, D. P., Elliott, S. M., Ma, P. L., Rasch, P. J., and Hartmann, D. L.: Natural

713 aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo, Science Advances, 1, e1500157,

714 10.1126/sciadv.1500157, 2015.

715 McParland, E. L., and Levine, N. M.: The role of differential DMSP production and community composition in predicting

716 variability of global surface DMSP concentrations, Limnol. Oceanogr., 64, 757-773, 10.1002/lno.11076, 2018.

717 Moradkhani, H., DeChant, C. M., and Sorooshian, S.: Evolution of ensemble data assimilation for uncertainty quantification

718 using the particle filter-Markov chain Monte Carlo method, Water Resources Research, 48, 10.1029/2012wr012144, 2012.

719 Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., Boutin, J., and Upstill-Goddard, R. C.:

720 In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers, Glob. Biogeochem.

721 Cycles, 14, 373-387, 10.1029/1999gb900091, 2000.

722 Novak, G. A., Fite, C. H., Holmes, C. D., Veres, P. R., Neuman, J. A., Faloona, I., Thornton, J. A., Wolfe, G. M., Vermeuel,

723 M. P., Jernigan, C. M., Peischl, J., Ryerson, T. B., Thompson, C. R., Bourgeois, I., Warneke, C., Gkatzelis, G. I., Coggon, M.

724 M., Sekimoto, K., Bui, T. P., Dean-Day, J., Diskin, G. S., DiGangi, J. P., Nowak, J. B., Moore, R. H., Wiggins, E. B., Winstead,

725 E. L., Robinson, C., Thornhill, K. L., Sanchez, K. J., Hall, S. R., Ullmann, K., Dollner, M., Weinzierl, B., Blake, D. R., and

726 Bertram, T. H.: Rapid cloud removal of dimethyl sulfide oxidation products limits SO2 and cloud condensation nuclei

727 production in the marine atmosphere, P. Natl. Acad. Sci. USA, 118, e2110472118, 10.1073/pnas.2110472118, 2021.

728 Omori, Y., Tanimoto, H., Inomata, S., Wada, S., Thume, K., and Pohnert, G.: Enhancement of dimethylsulfide production by
729 anoxic stress in natural seawater, Geophys. Res. Lett., 42, 4047-4053, 10.1002/2015gl063546, 2015.

730 Osman, M. B., Das, S. B., Trusel, L. D., Evans, M. J., Fischer, H., Grieman, M. M., Kipfstuhl, S., McConnell, J. R., and
731 Saltzman, E. S.: Industrial-era decline in subarctic Atlantic productivity, Nature, 569, 551-555, 10.1038/s41586-019-1181-8,
732 2019.

733 Park, K.-T., Lee, K., Kim, T.-W., Yoon, Y. J., Jang, E.-H., Jang, S., Lee, B.-Y., and Hermansen, O.: Atmospheric DMS in the
734 Arctic Ocean and Its Relation to Phytoplankton Biomass, Glob. Biogeochem. Cycles, 32, 351-359, 10.1002/2017gb005805,
735 2018.

736 Park, K. T., Yoon, Y. J., Lee, K., Tunved, P., Krejci, R., Ström, J., Jang, E., Kang, H. J., Jang, S., Park, J., Lee, B. Y., Traversi,
737 R., Becagli, S., and Hermansen, O.: Dimethyl Sulfide-Induced Increase in Cloud Condensation Nuclei in the Arctic
738 Atmosphere, Glob. Biogeochem. Cycles, 35, e2021GB006969, 10.1029/2021gb006969, 2021.

739 Quinn, P. K., and Bates, T. S.: The case against climate regulation via oceanic phytoplankton sulphur emissions, Nature, 480,
740 51-56, 10.1038/nature10580, 2011.

741 Quinn, P. K., Coffman, D. J., Johnson, J. E., Upchurch, L. M., and Bates, T. S.: Small fraction of marine cloud condensation
742 nuclei made up of sea spray aerosol, Nat. Geosci., 10, 674-679, 10.1038/ngeo3003, 2017.

743 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process
744 understanding for data-driven Earth system science, Nature, 566, 195-204, 10.1038/s41586-019-0912-1, 2019.

745 Saltzman, E. S., Savoie, D. L., Zika, R. G., and Prospero, J. M.: Methane sulfonic acid in the marine atmosphere, J. Geophys.
746 Res., 88, 10897, 10.1029/JC088iC15p10897, 1983.

747 Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., Forget, M. H., Maass, H., and Platt, T.: Carbon-to-
748 chlorophyll ratio and growth rate of phytoplankton in the sea, Marine Ecology Progress Series, 383, 73-84,
749 10.3354/meps07998, 2009.

750 Savoie, D. L., Arimoto, R., Keene, W. C., Prospero, J. M., Duce, R. A., and Galloway, J. N.: Marine biogenic and
751 anthropogenic contributions to non-sea-salt sulfate in the marine boundary layer over the North Atlantic Ocean, J. Geophys.
752 Res., 107, 4356, 10.1029/2001jd000970, 2002.

753 Screen, J. A., Deser, C., and Simmonds, I.: Local and remote controls on observed Arctic warming, Geophys. Res. Lett., 39,
754 L10709, 10.1029/2012gl051598, 2012.

755 Serreze, M. C., and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Global and planetary
756 change, 77, 85-96, 2011.

757 Sheng, J.-X., Weisenstein, D. K., Luo, B.-P., Rozanov, E., Stenke, A., Anet, J., Bingemer, H., and Peter, T.: Global atmospheric
758 sulfur budget under volcanically quiescent conditions: Aerosol-chemistry-climate model predictions and validation, J.
759 Geophys. Res.-Atmos., 120, 256-276, 10.1002/2014jd021985, 2015.

760 Sigmund, G., Gharasoo, M., Hüffer, T., and Hofmann, T.: Deep Learning Neural Network Approach for Predicting the
761 Sorption of Ionizable and Polar Organic Pollutants to a Wide Range of Carbonaceous Materials, Environ. Sci. Technol., 54,
762 4583-4591, 10.1021/acs.est.9b06287, 2020.

763 Simó, R., and Pedrós-Alió, C.: Short-term variability in the open ocean cycle of dimethylsulfide, Glob. Biogeochem. Cycles,
764 13, 1173-1181, 10.1029/1999gb900081, 1999a.

765 Simó, R., and Pedrós-Alió, C.: Role of vertical mixing in controlling the oceanic production of dimethyl sulphide, Nature, 402,
766 396-399, 10.1038/46516, 1999b.

767 Simó, R., and Dachs, J.: Global ocean emission of dimethylsulfide predicted from biogeophysical data, Glob. Biogeochem.
768 Cycles, 16, 1078, 10.1029/2001gb001829, 2002.

769 Six, K. D., Kloster, S., Ilyina, T., Archer, S. D., Zhang, K., and Maier-Reimer, E.: Global warming amplified by reduced
770 sulphur fluxes as a result of ocean acidification, Nat. Clim. Change, 3, 975-978, 10.1038/nclimate1981, 2013.

771 Stefels, J.: Physiological aspects of the production and conversion of DMSP in marine algae and higher plants, J. Sea. Res.,
772 43, 183-197, 2000.

773 Stefels, J., Steinke, M., Turner, S., Malin, G., and Belviso, S.: Environmental constraints on the production and removal of the
774 climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling, Biogeochemistry, 83, 245-275,
775 10.1007/s10533-007-9091-5, 2007.

776 Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT Atmospheric
777 Transport and Dispersion Modeling System, B. Am. Meteorol. Soc., 96, 2059-2077, 10.1175/bams-d-14-00110.1, 2015.

778 Steiner, N. S., Robert, M., Arychuk, M., Levasseur, M. L., Merzouk, A., Peña, M. A., Richardson, W. A., and Tortell, P. D.:
779 Evaluating DMS measurements and model results in the Northeast subarctic Pacific from 1996–2010, Biogeochemistry, 110,
780 269-285, 10.1007/s10533-011-9669-9, 2011.

781 Sunda, W., Kieber, D., Kiene, R., and Huntsman, S.: An antioxidant function for DMSP and DMS in marine algae, Nature,
782 418, 317-320, 2002.

783 Tesdal, J.-E., Christian, J. R., Monahan, A. H., and Salzen, K. v.: Evaluation of diverse approaches for estimating sea-surface
784 DMS concentration and air–sea exchange at global scale, Environ. Chem., 13, 390-412, 10.1071/EN14255, 2016.

785 Vallina, S. M., and Simó, R.: Strong relationship between DMS and the solar radiation dose over the global surface ocean,
786 Science, 315, 506-508, 10.1126/science.1133680, 2007.

787 Vogt, M., Vallina, S. M., Buitenhuis, E. T., Bopp, L., and Le Quéré, C.: Simulating dimethylsulphide seasonality with the
788 Dynamic Green Ocean Model PlankTOM5, J. Geophys. Res., 115, 10.1029/2009jc005529, 2010.

789 Wang, S., Elliott, S., Maltrud, M., and Cameron-Smith, P.: Influence of explicit Phaeocystis parameterizations on the global
790 distribution of marine dimethyl sulfide, J. Geophy. Res.-Biogeosci., 120, 2158-2177, 10.1002/2015jg003017, 2015.

791 Wang, W.-L., Song, G., Primeau, F., Saltzman, E. S., Bell, T. G., and Moore, J. K.: Global ocean dimethyl sulfide climatology
792 estimated from observations and an artificial neural network, Biogeosciences, 17, 5335-5354, 10.5194/bg-17-5335-2020, 2020.

793 Wood, R., Stemmler, J. D., Rémillard, J., and Jefferson, A.: Low-CCN concentration air masses over the eastern North Atlantic:

794 Seasonality, meteorology, and drivers, J. Geophys. Res.-Atmos., 122, 1203-1223, 10.1002/2016jd025557, 2017.

795 Woolf, D. K.: Bubbles and their role in gas exchange, in: The Sea Surface and Global Change, edited by: Liss, P. S., and Duce,

796 R. A., Cambridge University Press, Cambridge, 173-206, 1997.

797 Zhang, X. H., Liu, J., Liu, J., Yang, G., Xue, C. X., Curson, A. R. J., and Todd, J. D.: Biogenic production of DMSP and its

798 degradation to DMS-their roles in the global sulfur cycle, Sci. China Life Sci., 62, 1296-1319, 10.1007/s11427-018-9524-y,

799 2019.

800 Zhao, J., Ma, W., Bilsback, K. R., Pierce, J. R., Zhou, S., Chen, Y., Yang, G., and Zhang, Y.: Simulating the radiative forcing

801 of oceanic dimethylsulfide (DMS) in Asia based on machine learning estimates, Atmos. Chem. Phys., 22, 9583-9600,

802 10.5194/acp-22-9583-2022, 2022.

803 Zheng, G., Li, X., Zhang, R. H., and Liu, B.: Purely satellite data-driven deep learning forecast of complicated tropical

804 instability waves, Science Advances, 6, eaba1482, 10.1126/sciadv.aba1482, 2020.

805 Zhou, S., Chen, Y., Paytan, A., Li, H., Wang, F., Zhu, Y., Yang, T., Zhang, Y., and Zhang, R.: Non-Marine Sources Contribute

806 to Aerosol Methanesulfonate Over Coastal Seas, J. Geophys. Res.-Atmos., 126, e2021JD034960, 10.1029/2021jd034960,

807 2021.

808 Zhou, S., Chen, Y., Huang, S., Gong, X., Yang, G., Zhang, H., Herrmann, H., Wiedensohler, A., Poulain, L., Zhang, Y., Wang,

809 F., Xu, Z., and Yan, K.: A 20-year (1998-2017) global sea surface dimethyl sulfide gridded dataset with daily resolution, v2.0,

810 Zenodo [Data set], https://doi.org/10.5281/zenodo.10279659, 2023.

811 Zhou, S.: An artificial neural network ensemble model for sea surface DMS simulation, v1.0, Zenodo [Data set],

812 https://doi.org/10.5281/zenodo.8077751, 2023

813 Zindler, C., Bracher, A., Marandino, C. A., Taylor, B., Torrecilla, E., Kock, A., and Bange, H. W.: Sulphur compounds,

814 methane, and phytoplankton: interactions along a north–south transit in the western Pacific Ocean, Biogeosciences, 10, 3297-

815 3311, 10.5194/bg-10-3297-2013, 2013.

816 Zindler, C., Marandino, C. A., Bange, H. W., Schütte, F., and Saltzman, E. S.: Nutrient availability determines dimethyl sulfide

817 and isoprene distribution in the eastern Atlantic Ocean, Geophys. Res. Lett., 41, 3181-3188, 10.1002/2014gl059547, 2014.

818