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

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], Hartmut Herrmann[5], Alfred Wiedensohler[5], Laurent Poulain[5], Yan Zhang[1,2], Fanghui Wang[1], Zongjun Xu[1], Ke Yan[1]

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

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

Abstract. The oceanic emission of dimethyl sulfide (DMS) plays a vital role in the Earth's climate system and constitutes a substantial source of uncertainty in evaluating aerosol radiative forcing. Currently, the widely used monthly climatology of sea surface DMS concentration falls short of meeting the requirement for accurately simulating DMS-derived aerosols by chemical transport models. Hence, there is an urgent need for a high-resolution, multi-year global sea surface DMS dataset. Here we develop an artificial neural network ensemble model using 9 environmental factors as input features, which well captures the variabilities of DMS concentration across different oceanic regions. Subsequently, a global sea surface DMS concentration and flux dataset (1°×1°) with daily resolution spanning from 1998 to 2017 is established. According to this dataset, the global annual average concentration was ~1.71 nM, and the annual total emission was ~17.2 TgS yr^{-1}, with ~60% originating from the southern hemisphere. While overall seasonal variations are consistent with previous DMS climatologies, notable differences exist in regional-scale spatial distributions. The new dataset enables further investigations into daily and decadal variations. Throughout the period 1998–2017, the global annual average concentration exhibited a slight decrease, while total emissions showed no significant trend. The DMS flux from our dataset showed a stronger correlation with observed atmospheric methanesulfonic acid concentration compared to those from previous monthly climatologies. Therefore, it can serve as an improved emission inventory of oceanic DMS and has the potential to enhance the simulation of DMS-derived

aerosols and associated radiative effects. The new DMS gridded products are available at https://doi.org/10.5281/zenodo.11879900 (Zhou et al., 2024).

1 Introduction

Dimethyl sulfide (DMS), primarily produced by ocean biota, accounts for more than half of natural sulfur emissions and significantly contributes to sulfur dioxide in the troposphere (Sheng et al., 2015; Andreae, 1990), which can be oxidized to sulfuric acid and form sulfate aerosols (Barnes et al., 2006; Hoffmann et al., 2016). Sulfate aerosols play an important role in climate systems by scattering solar radiation, changing cloud condensation nuclei (CCN) population, and altering cloud properties (Masson-Delmotte et al., 2021). Recent studies have proven that CCN over the remote ocean and polar regions are primarily composed of non-sea-salt sulfate (nss-SO_4^{2-}) (Quinn et al., 2017; Park et al., 2021). Given the weak influence of anthropogenic SO_2 over open oceans, marine biogenic DMS emerges as a crucial source of nss-SO_4^{2-}, regulating oceanic climate (McCoy et al., 2015). Accordingly, DMS has been suggested to be the key substance in the postulated feedback loop of marine phytoplankton to climate warming (the "CLAW" hypothesis) (Charlson et al., 1987), albeit facing several controversies (Quinn and Bates, 2011). To accurately simulate the climate effects of DMS-derived aerosols, high-fidelity and high-resolution data on sea surface DMS concentrations and emission fluxes are required, along with further exploration of complex atmospheric chemical and physical processes (Hoffmann et al., 2016; Novak et al., 2021). It has been indicated that the uncertainty in DMS emission flux is the second largest contributor to the overall uncertainty associated with natural aerosols in evaluating the aerosol indirect radiative forcing (Carslaw et al., 2013). Therefore, understanding the spatiotemporal variations of DMS in global oceans is currently an important task.

There are complex production and consumption mechanisms of DMS in the upper ocean, which makes it difficult to well capture the dynamics and distributions of sea surface DMS across different regions. Dimethylsulfoniopropionate (DMSP), the major precursor of DMS, is synthesized mainly by phytoplankton in the photic zone and plays a variety of physiological functions in algal cells (Stefels, 2000; Sunda et al., 2002; McParland and Levine, 2018). The DMSP yield varies significantly among algal species (Stefels et al., 2007; Keller et al., 1989), and DMS can be produced through DMSP intracellular and extracellular cleavage by both algae and bacteria (Alcolombri et al., 2015; Zhang et al., 2019). Therefore, the oceanic DMS produced via multiple pathways can be affected by many biotic and abiotic factors, including temperature, salinity, solar radiation, mixed layer depth, nutrients, oxygen, acidity, etc. (Simó and Pedrós-Alió, 1999a; Vallina and Simó, 2007; Stefels, 2000; Zindler et al., 2014; Six et al., 2013; Omori et al., 2015; Stefels et al., 2007). In addition, seawater DMS undergoes various removal pathways (bacterial consumption, photodegradation, sea-to-air ventilation, etc.), further complicating its cycling (Stefels et al., 2007; Galí and Simó, 2015; Hopkins et al., 2023). Therefore, although previous studies have developed several empirical algorithms (Simó and Dachs, 2002; Belviso et al., 2004b; Vallina and Simó, 2007) and process-embedded prognostic models (Kloster et al., 2006; Vogt et al., 2010; Belviso et al., 2011; Wang et al., 2015) based on relevant variables

(mixed layer depth, chlorophyll *a*, nutrients, radiation, phytoplankton group, etc.) to estimate the distribution of DMS, their results showed significantly different patterns and inconsistency with observations in many regions (Tesdal et al., 2016; Belviso et al., 2004a). Recently, Galí et al. (2018) developed a new empirical algorithm following a parameterization of DMSP (Galí et al., 2015). The estimated DMS field exhibited a generally higher consistency with observations than those derived from previous algorithms SD02 (Simó and Dachs, 2002) and VS07 (Vallina and Simó, 2007), but this method did not consider the influences of nutrients and still exhibited substantial biases in certain regions (e.g., near the Antarctic).

Since Lovelock et al. (1972) first discovered the ubiquitous presence of DMS in seawater, numerous observations of sea surface DMS have been conducted worldwide, yielding a substantial volume of observational data to date. Based on these worldwide measurements, monthly climatology of global DMS can be generated through interpolation and extrapolation (Hulswar et al., 2022; Kettle et al., 1999; Lana et al., 2011). The latest version incorporated 873,539 raw observations (48,898 after data filtration and unification for climatology development), and the estimated global annual mean concentration and total flux are 2.26 nM and 27.1 TgS yr^{-1}, respectively (Hulswar et al., 2022). However, despite the abundance of data, significant spatial and temporal disparities persist, potentially introducing large uncertainties in regions or periods with sparse observations. Furthermore, the observational data from different years within a particular month were combined together for interpolation and extrapolation, and the interannual variations cannot be investigated by this approach.

In recent years, the application of data-driven approaches like machine learning to Earth system science has drawn more and more attention. Compared with traditional approaches, machine learning explores larger function space and captures more hidden information from the big data, hence it often provides a better prediction performance (Reichstein et al., 2019; Zheng et al., 2020; Bergen et al., 2019). For instance, a recent study demonstrated that artificial neural network (ANN) can capture much more (~66%) of the raw data variance than multilinear regression (~39%), and a global monthly climatology of sea surface DMS concentration has been developed based on ANN model (Wang et al., 2020). The machine learning techniques have also been used to simulate the distribution of DMS in the Arctic (Humphries et al., 2012; Qu et al., 2016), North Atlantic Ocean (Bell et al., 2021; Mansour et al., 2023), Northeast Pacific Ocean (McNabb and Tortell, 2022), Southern Ocean (McNabb and Tortell, 2023), and East Asia (Zhao et al., 2022).

However, to our best knowledge, there is currently no global-scale sea surface gridded DMS dataset with both high time resolution (daily) and long-term coverage (> 10 years). Such a dataset is urgently needed for modeling the atmospheric processes and climatic implications of oceanic DMS. The sea surface concentration and sea-to-air emission flux of DMS can vary greatly from day to day (Simó and Pedrós-Alió, 1999b), and the emitted DMS exerts effects on the atmosphere over time scales of several hours to days. Relying solely on monthly climatology of DMS as the emission inventory may fail to capture important details and could lead to large modeling biases compared to observed concentrations of atmospheric DMS or its oxidation products (Chen et al., 2018; Fung et al., 2022).

Here, we build a 20-year (1998 – 2017) global sea surface DMS gridded dataset (1°×1°) with a daily resolution based on a data-driven machine learning approach (ANN ensemble). This product can improve our understanding of the spatiotemporal variations of oceanic DMS. More importantly, it can serve as an updated emission inventory of marine biogenic DMS for chemical transport models, which is beneficial for enhancing the simulation of atmospheric processes of DMS and reducing the uncertainties in marine aerosol's climate effects. The paper consists of four main parts as depicted in Fig. 1: (1) the development of machine learning model based on global DMS measurements and 9 ancillary environmental variables; (2) the derived spatial and temporal distributions of DMS and comparisons with previous estimates; (3) an example showing the superiority of our newly developed DMS field through its correlation with atmospheric biogenic sulfur; and (4) the uncertainties and limitations inherent in our approach and the resulting data product.

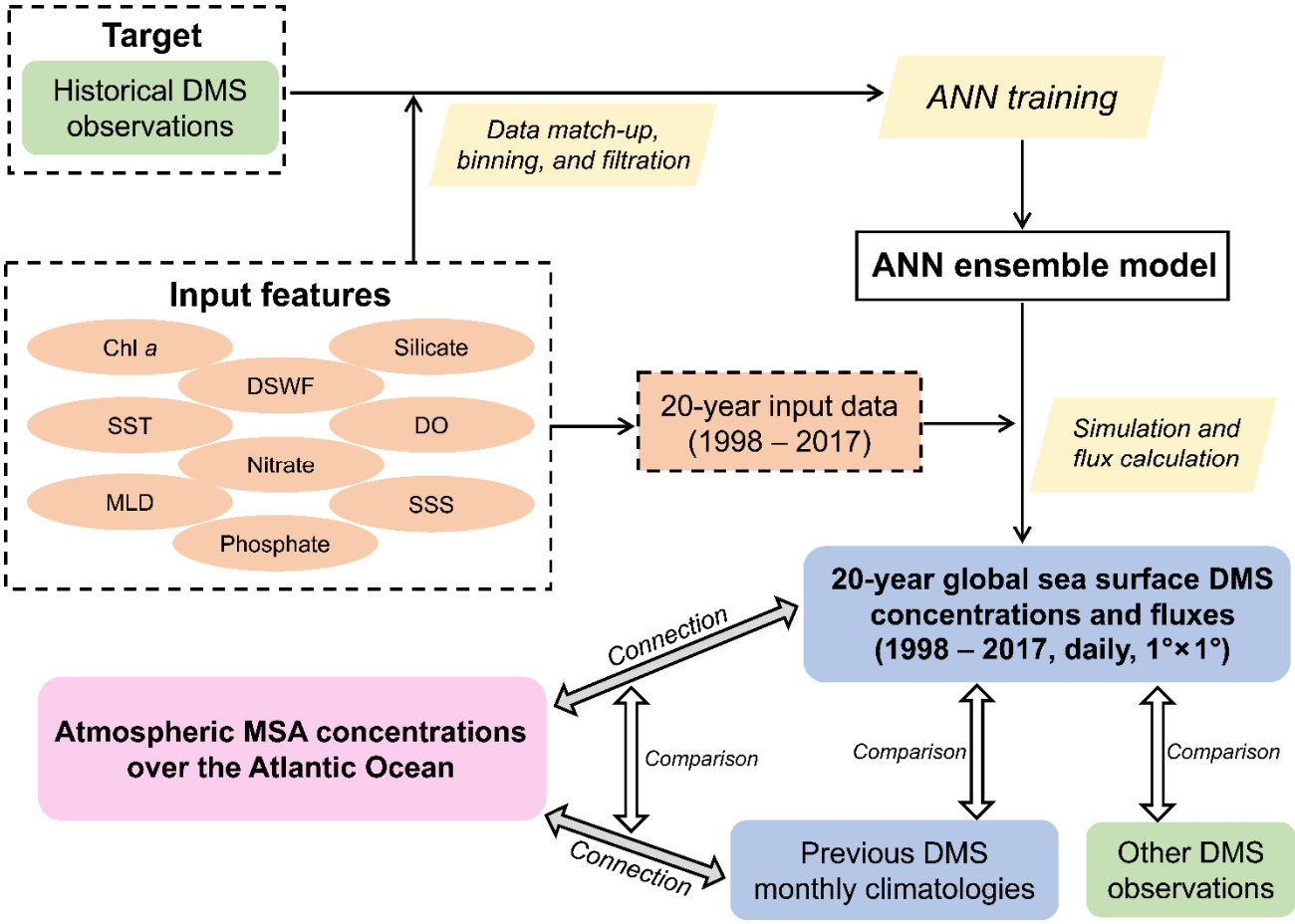

Figure 1. Flowchart of this study, including the development of ANN ensemble model, construction of new DMS gridded dataset, and subsequent evaluations of this product.

2 Methodology

2.1 Input datasets

The in-situ DMS measurement data used for training the machine learning model primarily sourced from the Global Surface Seawater DMS (GSSD) database (Kettle et al., 1999). The GSSD database contains a total of 87,801 DMS measurements collected across 266 cruise and fixed-site observation campaigns from 11 March 1972 to 27 August 2017 (https://saga.pmel.noaa.gov/dms/, last access: 1 April 2020). Hulswar et al. (2022) consolidated other DMS measurements not included in the GSSD database to establish an updated DMS climatology. Here we incorporated these additional data predating 2017, originating from 8 campaigns (number of samples = 6,711). The spatial distribution of theses combined 94,512 in-situ observational data is shown in Fig. S1, which covers all major regions of the global ocean.

We selected 9 environmental variables relevant 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 these 9 input variables and DMS are listed in Table 1. Chl *a* data were obtained from both in-situ observations, co-located with DMS data, and satellite remote sensing products (Copernicus-GlobColour, Level-4, daily, 0.042°×0.042°). The Copernicus-GlobColour Level-4 dataset integrates multiple upstream sensors including SeaWiFS, MODIS-Aqua & Terra, MERIS, VIIRS-SNPP & JPSS1, and OLCI-S3A & S3B, with an interpolation procedure applied to fill missing data (Garnesson et al., 2019). Daily SST data (0.25°×0.25°) were 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 sea surface concentrations of nitrate, phosphate, silicate, and DO were from the CMEMS global biogeochemical multi-year hindcast dataset (daily, 0.25°×0.25°). The surface wind speed (WS) and sea ice fraction (SI) data are needed in the calculation of sea-to-air flux (details are provided in Section 2.4.2). 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 among all datasets, the data match-up has been conducted as described in the next section.

Table 1. The data sources and relavant 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	-
	Other campaigns integrated in Hulswar et al. (2022)	https://data.mendeley.com/datasets/hyn62spny2/1	In-situ	Feb. 2000 – Jun. 2016	-
Chl *a*	GSSD database	https://saga.pmel.noaa.gov/dms/	In-situ	Oct. 1980 – Aug. 2017	-
	Copernicus-GlobColour Level-4	https://data.marine.copernicus.eu/product/OCEANCOLOUR_GLO_BGC_L4_MY_009_104/description	Daily	Sep. 1997 – present	0.042°×0.042°
	CMEMS global biogeochemical multi-year hindcast (only used for the simulation of DMS concentration in polar regions when satellite Chl *a* is unavailable)	https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_BGC_001_029/description	Daily	Jan. 1993 – present	0.25°×0.25°
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 (22 – 110 km)
DSWF					
SSS					
Nitrate	CMEMS global biogeochemical multi-year hindcast	https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_BGC_001_029/description	Daily	Jan. 1993 – present	0.25°×0.25°
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 (22 – 110 km)
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 based on the sampling location and time associated with each DMS measurement record, as well as the temporal range and grid distribution of each variable. For satellite-retrieved Chl *a*, the data of the grids covering DMS sampling locations were extracted. If the data of the corresponding grid is missing, the average

value of the 5×5 grids nearby was calculated and used. For other variables, only values in the grids matching the DMS sampling locations were extracted.

There are in-situ Chl *a* measurements co-located with certain GSSD data. They were also used along with satellite-retrieved Chl *a*. In-situ Chl *a* measurements with low precision (defined as < 0.1 mg m^{-3}, and the number of significant digits is 1) were removed. For a specific in-situ observation campaign, if the number of low-precision values is larger than 10 and accounts for more than half, all in-situ Chl *a* data from this campaign were excluded. In addition, the in-situ Chl *a* data in the GSSD database were measured by two different methods: Turner fluorometry and high-performance liquid chromatography (HPLC). In order to improve mutual consistency, a conversion between the data from these two methods was applied and then the in-situ Chl *a* concentrations were adjusted to match up with satellite Chl *a* following the functions described in Galí et al. (2015). After that, the statistical outliers for all $\log_{10}$(Chl *a*) (outside the range of average ± 3 standard deviations) were eliminated. The comparison between in-situ and satellite-retrieved Chl *a* data is shown in Fig. S2. A strong consistency between in-situ and daily satellite Chl *a* data ($R^2 > 0.5$, RMSE < 0.4) suggests the rationale for integrating these datasets. The $\log_{10}$ transformation was applied to make the data distribution close to normal distribution. When finally selecting the $\log_{10}$(Chl *a*) corresponding to each DMS data, in-situ data were prioritized where available; otherwise, the satellite-retrieved data were used.

DMS and extracted MLD and three nutrients (nitrate, phosphate, silicate) were also performed $\log_{10}$ transformation. The statistical outliers of each variable were excluded as mentioned above. After data filtration, a total of 633,361 samples with valid data for all variables were obtained. To avoid data aggregation bias stemming from multiple data points gathered within a narrow temporal and spatial range (i.e., the same day and within a region smaller than 0.05°×0.05°), these data points were averaged. Consequently, 41,157 binned samples were utilized for subsequent model development, with their spatial distribution depicted in Fig. 2a.

We divided the global ocean into 9 regions based on Longhurst's biomes (Longhurst, 1998). There are 6 biomes in Longhurst's definition, including Coastal, Polar_N, Polar_S, Westerlies_N, Westerlies_S, and Trades (the .shp file of Longhurst's biomes and provinces was downloaded from https://www.marineregions.org/downloads.php#longhurst). We further divided Westerlies_N into Westerlies_N_Pacific and Westerlies_N_Atlantic, and divided Trades into Trades_Pacific, Trades_Indian, and Trades_Atlantic by different oceanic basins, as shown in Fig. 2b. It is noteworthy that there are 11,237 samples in the Coastal region, constituting 27.3% of the entire sample set, despite the Coastal biome accounting for only 9.7% of the global ocean area. Given the distinct seawater physiochemical and biological conditions in coastal seas compared to other regions, the disproportionately higher density of samples within the Coastal biome might cause the model to overly prioritize this region. To mitigate this data imbalance and ensure the model captures broader patterns in open oceans, we adjusted the data distribution during model training and validation processes. Specifically, for each training session, a portion of coastal samples is randomly removed, ensuring the proportion of coastal samples in the total sample set (denoted as $F_{coastal}$) matches its area proportion.

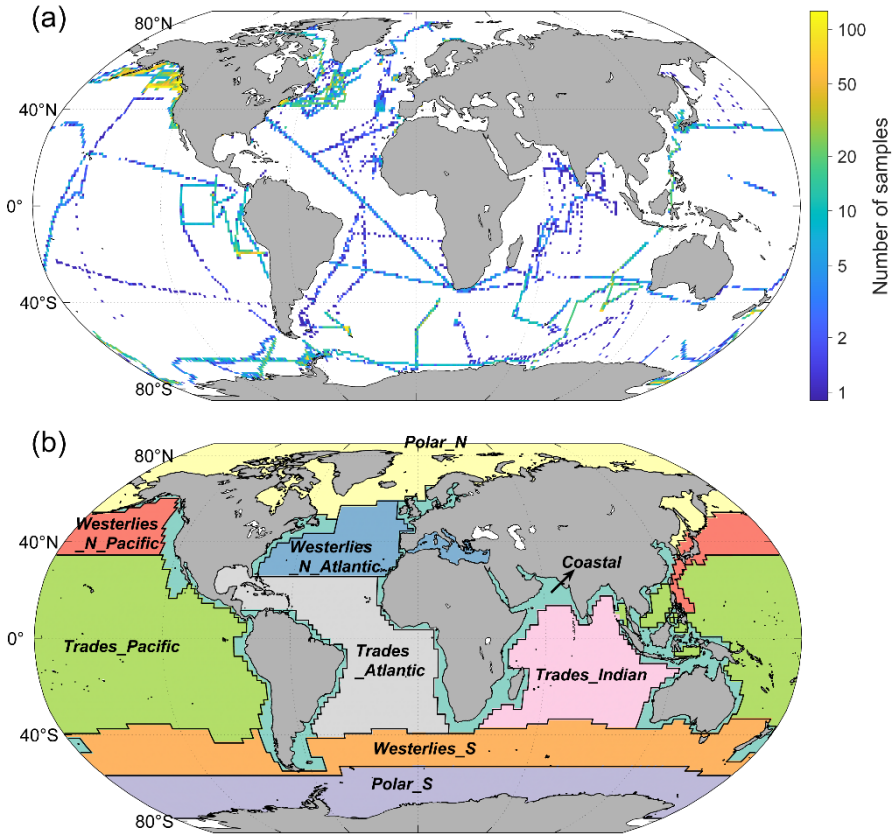

Figure 2. (a) The distribution of 41,157 DMS observational data after matchup, filtration, and binning for constructing the ANN model. The grid size is 1°×1°. (b) Nine oceanic regions separated based on Longhurst's biomes (Longhurst, 1998).

2.3 Artificial neural network training and validation

The 41,157 binned samples after the previously mentioned data preprocessing were used to develop the artificial neural network (ANN) model. The target feature is $\log_{10}$(DMS), and the input features are $\log_{10}$(Chl *a*), SST, $\log_{10}$(MLD), $\log_{10}$(nitrate), $\log_{10}$(phosphate), $\log_{10}$(silicate), DO, DSWF, and SSS. The data of all variables were standardized before training.

We randomly selected 10% of the samples (n = 4,116) to be entirely excluded from training, as a testing subset for global validation and overfitting test. Specifically, 401 samples were randomly selected from Coastal biome, and 3,715 samples were selected from other biomes to compose the testing subset, matching the proportion of coastal area in global oceans (9.7%). Then, the remaining samples (n = 37,041) were utilized for training and cross validation, with a constraint of $F_{coastal}$ equal to 9.7% in each training session as mentioned above.

Our feedforward fully connected neural network comprises two hidden layers, with 15 nodes in each layer. The activation functions for the first and second layers are ReLU and tanh, respectively. We applied L2 regularization (lambda = 1E-4) to counteract overfitting. The loss function is mean square error (MSE). Training stops if the validation loss is greater than or equal to the minimum validation loss computed so far 20 times in a row. The training processes were carried out with Statistics and Machine Learning Toolbox on Matlab 2022b. We repeated the data split (for training and validation sets) and training processes for 100 times and obtained 100 neural networks. The average prediction results of multiple ANNs shows a much higher consistency with the observations than a single ANN (Fig. S3). As the number of ANNs ($N_{training}$) increases, the accuracy of model predictions initially improves and then stabilizes. We adopted the average output of 20 ANNs as the final output, balancing performance and computational costs effectively. This kind of multiple-training approach, often termed "ANN ensemble" or "Monte Carlo cross-validation", has been widely used to improve the model generalization and performance (Sigmund et al., 2020; Holder et al., 2022) as well as get a better model evaluation (Dubitzky et al., 2007).

2.4 Deriving the 20-year global DMS distributions

2.4.1 Simulation of sea surface DMS concentrations

First, we constructed the daily gridded dataset of input variables with a spatial resolution of 1°×1° from 1998 to 2017 based on the data sources listed in Table 1 (except in-situ Chl *a* data). Datasets with a higher spatial resolution than 1°×1° were binned into 1°×1°. In polar regions, the satellite Chl *a* data are missing during winter, and the Chl *a* data from CMEMS global biogeochemical multi-year hindcast were used to fill the missing values. Then, the obtained gridded dataset was fed into the ANN ensemble model, and the 20-year global distribution of sea surface DMS concentration with daily resolution was simulated.

2.4.2 Calculation of sea-to-air fluxes

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

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

Here DMS_w and DMS_a are DMS concentrations in surface seawater and air, respectively. H is Henry's law constant of DMS. 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 ice coverage fraction (SI):

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

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

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

Here k_w and k_a are the water-side transfer velocity and air-side transfer velocity, respectively. We used the same approach as 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 SST and SSS were considered for H. The calculations of k_a and H followed the parameterizations of Johnson (2010). As for k_w calculation, we adopted the bubble scheme (Woolf, 1997), which divided the sea-to-air mass transfer process into turbulence- and bubble-mediated gas exchange. The calculated k_w based on the bubble scheme is lower than that of Nightingale's scheme (Nightingale et al., 2000) under conditions of high wind speed, exhibiting a smaller deviation from the measurements (Beale et al., 2014; Galí et al., 2019). Before calculation, WS and SI data were also binned by 1°×1° grid. By using WS and SI together with SST and SSS datasets, we obtained the daily gridded *Kt* and then calculated the sea-to-air DMS fluxes (daily, 1998–2017) by multiplying simulated DMS concentrations by *Kt* values.

3 Results

3.1 Model performance

As shown in Fig. 3a, the newly developed ANN ensemble model captures a substantial part of data variance globally ($\log_{10}$ space $R^2 = 0.651$ and RMSE = 0.262). 92.8% of ANN simulated concentration values fall within 1/3 to 3 times of corresponding true values. The performance for testing set ($R^2 = 0.640$, RMSE = 0.267, and 92.7% of data within the range of 1/3 to 3 times of observations) is very close to that for the training set (Fig. 3b), suggesting no obvious overfitting. The ANN model exhibits better performance compared to previous empirical and process-based models ($R^2 = 0.01$~0.14) (Tesdal et al., 2016) as well as the satellite-based algorithm ($R^2 = 0.50$) (Galí et al., 2018). The ANN model developed by Wang et al. (2020) showed a similar performance ($R^2 = 0.66$, RMSE = 0.264 for training set) to our model, despite their more complex ANN configuration (two hidden layers with 128 nodes each) and the inclusion of sample location and time into input features. However, the more complex model will significantly increase the computational cost, and the incorporation of location and time information may weaken the physical interpretability.

The performance of the model was evaluated across each of the nine oceanic regions. As illustrated in Fig. 3c and 4, the $\log_{10}$ space RMSEs are all below 0.32 (equivalent to a concentration ratio of 2.09 in linear space), except for the Coastal region (training: RMSE = 0.322, $R^2 = 0.479$; testing: RMSE = 0.332, $R^2 = 0.480$). Since the Coastal region comprises only 9.7% of the global oceanic area, the comparatively lower performance in this area has minimal impact on the overall ability to predict the spatiotemporal distributions of DMS on a global scale. Despite the R^2 values in Trades_Pacific and Trades_Atlantic being lower than 0.5, which is related to the relatively narrow variation range of DMS concentration, the RMSEs in these regions remain quite low and comparable to those of other regions. In general, our ANN ensemble model demonstrates a satisfactory capacity to reproduce variations in DMS concentrations across diverse oceanic regions.

However, it is noteworthy that our model tends to underestimate extremely high DMS concentrations and overestimate extremely low concentrations. Overall, the linear regressions between ANN-predicted and observed DMS concentrations yield slopes significantly lower than unity across all regions (Fig. 3c and 4), and there are significantly positive correlations between prediction residuals (observation – prediction) and observed $\log_{10}$(DMS) (Fig. S5 and S6). From a data perspective, this may be partly due to the insufficient number of samples with extreme DMS concentrations (known as underrepresentation), making it difficult to adequately capture the relevant information during training process. To test this point, we adopted a weighted resampling strategy to bolster the number of samples in the minority class before training, which has been widely used in machine learning to deal with the data imbalance issue (Haibo et al., 2008; Yu and Zhou, 2021; Chawla et al., 2002). The basic idea is to set a higher probability of being sampled for the minority class with extreme DMS concentrations, and the details are illustrated in Fig. S7 and explained in Appendix B. The results indicate that the weighted resampling scheme cannot fully alleviate the model bias. Although it does elevate the overall prediction-versus-observation slopes from ~0.59 to ~0.63, this improvement is marginal (Fig. S8 and S9). In several regions like Westerlies_S and Trades biomes, the slopes are even lower than original values. Furthermore, the data become more scattered after implementing the weighted resampling, resulting in increased RMSE and decreased R^2. Therefore, there are other potential issues causing the model bias, which are discussed in Section 4. The original model, trained without weighted resampling, was adopted for subsequent analysis and the construction of the gridded DMS dataset.

Primarily owing to the underestimation of high DMS concentrations, a negative mean bias (MB) in DMS concentration is evident across all regions, ranging from -0.18 to -2.02 nM (Table 2). The normalized mean bias (NMB, the ratio between mean bias and mean observed concentration) ranges from -8.7% to -32.2%. The most significant NMB emerges in Coastal and Trades_India regions, while NMB remains within -25% for other regions. The global MB and NMB are -1.05 nM and -22.1%, respectively. It is worth noting that these biases are compared against historical DMS observations, which were conducted within a very limited geographical area and time periods. Thus, they cannot be interpreted as the actual mean modelling bias for the entire region. On the other hand, the negative biases at high end of the concentrations are partially cancelled out by the positive biases at low end during the averaging over the entire region. The bias at a specific grid could be much larger. Nevertheless, those extreme DMS concentrations (> 15 nM or < 0.3 nM), exhibiting the most significant modeling bias, represent only a minority of the entire sample set (6.9%). Our model adeptly reproduces the majority of observations with moderate DMS concentrations across all regions, with the percentage of predicted values falling within 1/3 to 3 times of observations ranging from 87.0% to 98.8%.

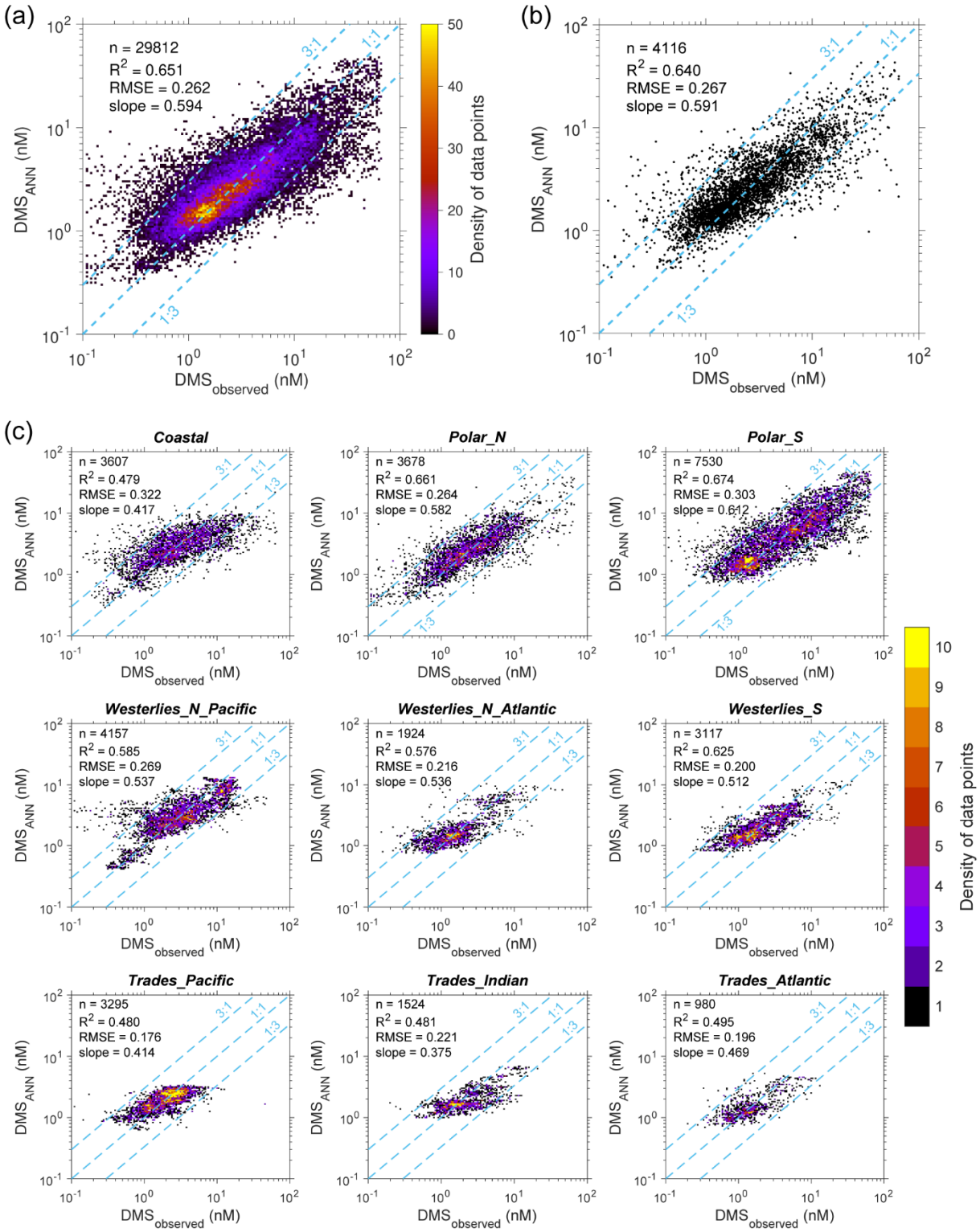
(a)
n = 29812
R2 = 0.651
RMSE = 0.262
slope = 0.594
3:1
1:1
1:3
DMSANN (nM)
DMSobserved (nM)
Density of data points
(b)
n = 4116
R2 = 0.640
RMSE = 0.267
slope = 0.591
(c)
Coastal
n = 3607
R2 = 0.479
RMSE = 0.322
slope = 0.417
Polar_N
n = 3678
R2 = 0.661
RMSE = 0.264
slope = 0.582
Polar_S
n = 7530
R2 = 0.674
RMSE = 0.303
slope = 0.612
Westerlies_N_Pacific
n = 4157
R2 = 0.585
RMSE = 0.269
slope = 0.537
Westerlies_N_Atlantic
n = 1924
R2 = 0.576
RMSE = 0.216
slope = 0.536
Westerlies_S
n = 3117
R2 = 0.625
RMSE = 0.200
slope = 0.512
Trades_Pacific
n = 3295
R2 = 0.480
RMSE = 0.176
slope = 0.414
Trades_Indian
n = 1524
R2 = 0.481
RMSE = 0.221
slope = 0.375
Trades_Atlantic
n = 980
R2 = 0.495
RMSE = 0.196
slope = 0.469

Figure 3. Comparisons between ANN-simulated and observed DMS concentrations. (a) Scatter density for simulated versus observed DMS concentrations of the samples used in ANN training. (b) Comparison between the simulated versus observed DMS concentrations of testing set. (c) Comparison between the simulated versus observed DMS concentrations of the samples used in ANN training across 9 regions. The number of data points (n), $\log_{10}$ space R^2, root mean square error (RMSE), and linear regression slope are also displayed.

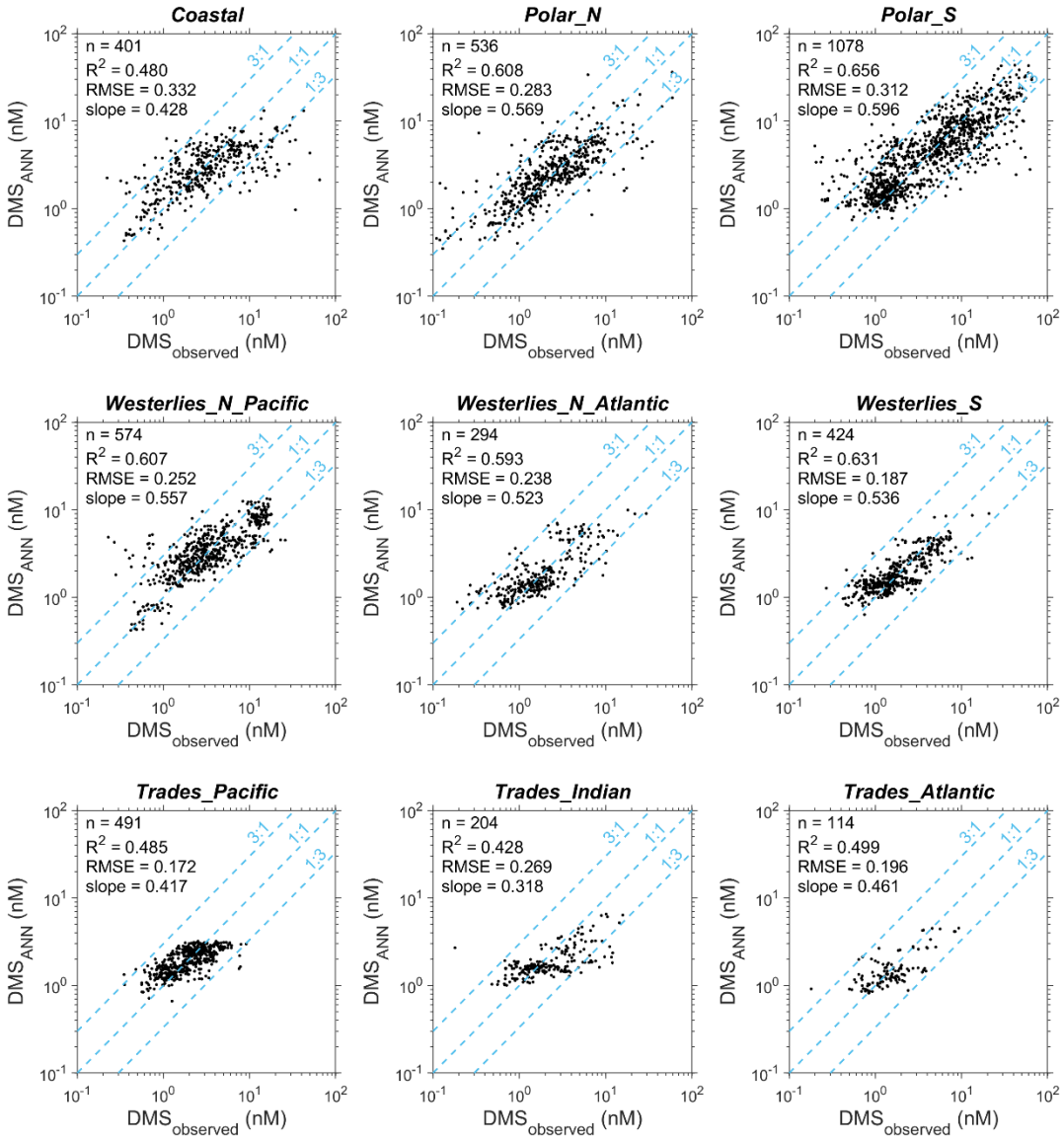

Figure 4. Comparisons between the simulated versus observed DMS concentrations of the testing set across 9 regions.

Table 2. The mean bias and normalized mean bias of the ANN-predicted DMS concentrations against observations across different regions.

Region	Mean bias (nM)	Normalized mean bias
Coastal	-1.55	-32.2%
Polar_N	-0.90	-21.4%
Polar_S	-2.02	-24.1%
Westerlies_N_Pacific	-0.91	-18.8%
Westerlies_N_Atlantic	-0.24	-10.4%
Westerlies_S	-0.36	-14.1%
Trades_Pacific	-0.19	-8.7%
Trades_Indian	-0.73	-26.7%
Trades_Atlantic	-0.18	-10.1%
Global	-1.05	-22.1%

It is worth noting that there may be intrinsic connections between the 10% excluded testing subset and the training set, because the data from the same cruise or fixed-site campaign have certain continuity. To further evaluate the reliability of the ANN model, we compared the simulated DMS concentrations with the observational data from fully independent campaigns, which are obtained from 33 cruises in Northeast Pacific, West Pacific, and North Atlantic (number of data = 6,478). These data include (1) discrete sampling and measurement during 31 cruises of *Line P Program* in Northeast Pacific (Steiner et al., 2011) (9 February 2007 – 26 August 2017, number of data = 177, https://www.waterproperties.ca/linep/index.php, last access: 23 November 2020), (2) underway measurements during *SONNE cruise 202/2 (TRANSBROM)* in West Pacific (Zindler et al., 2013) (9 – 23 October 2009, number of data = 115, https://doi.org/10.1594/PANGAEA.805613, last access: 23 November 2020), (3) underway measurements during the third *North Atlantic Aerosols and Marine Ecosystems Study (NAAMES)* campaign (Behrenfeld et al., 2019; Bell et al., 2021) (6 – 24 September 2017, number of data = 1, 025, https://seabass.gsfc.nasa.gov/naames, last access: 27 November, 2020). Before the comparison, the data measured within a 0.05°×0.05° grid and at the same day were binned by arithmetic average.

The comparisons between these observed DMS concentrations and ANN simulation are shown in Fig. 5. Regarding the *Line P Program*, it should be noted that there are 7 cruises included in the GSSD database, but those data were obtained by underway measurements, different from the discrete sampling (Niskin bottle) data used here. Hence, these cruises were retained and marked in Fig. 5a but excluded in subsequent statistical analysis (Fig. 5b-c). It can be seen that the model effectively captures the seasonal variation in Northeast Pacific, which is generally August > June > February (Fig. 5a). However, the small-scale spatial variations can only be partially reproduced by the model in certain campaigns, such as those in June and August of 2007, June of 2009, August of 2012, and August of 2016. Notably, the model generally underestimates high DMS

concentrations during summer, particularly those exceeding 10 nM, consistent with earlier discussions. Aggregating data from all campaigns across three regions, the $\log_{10}$ space RMSE of simulated DMS concentrations against observations is 0.274, marginally higher than the training set. Most simulated values (93.0%) are within the range of 1/3 to 3 times of observations. The results further evidence that there is no significant overfitting in our model. When data from each campaign are binned, simulations demonstrate high consistency with observations, as depicted in Fig. 5c (RMSE = 0.249, R^2 = 0.758). In summary, although our ANN ensemble model may not precisely reproduce small-scale variations and extreme values in specific regions and periods, it reasonably captures overall large-scale variations.

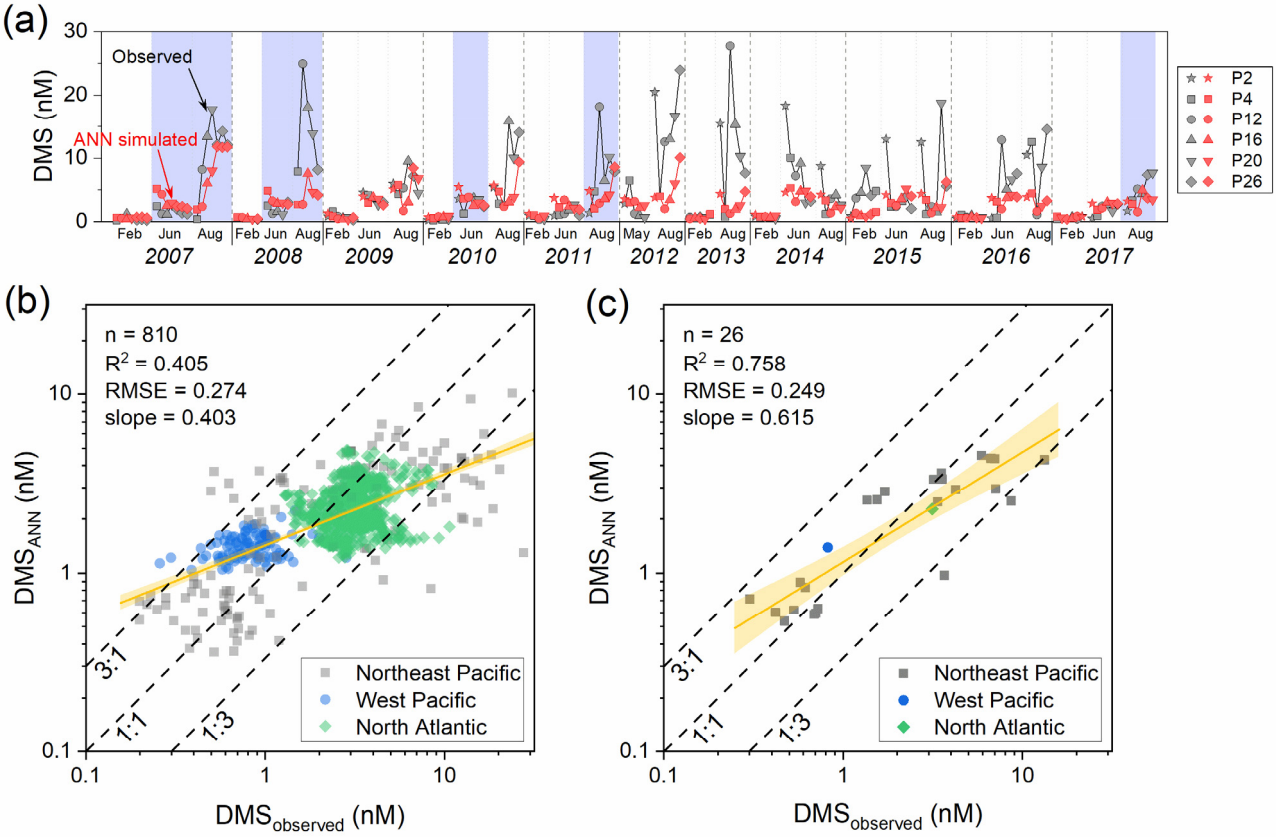

Figure 5. Comparisons between the ANN predictions and observations for fully independent campaigns. (a) Time series of simulation results and DMS observational data obtained from *Line P Program*. The different markers represent different stations of *Line P*. The blue shades cover the data obtained 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 values of R^2, RMSE, and slope 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. 6. Overall, the DMS concentrations in mid- and high-latitude regions exhibit a significant seasonal cycle, peaking in summer and reaching their lowest in winter. This pattern aligns with the results of many prior observational studies. In the northern hemisphere, elevated DMS concentrations (> 2.5 nM) during summer mainly occur in two regions. One is the North Pacific (40°–60° N) where the concentration generally peaks in August, surpassing 10 nM (Fig. 6). The other is the subarctic North Atlantic (45°–80° N). A notable increase of DMS concentration starts around 45°–50° N in May and gradually shifts northward beyond 50° N by July (Fig. 6-7). This spatiotemporal evolution pattern corresponds to the evolution of solar radiation intensity and the spring-summer bloom patterns of phytoplankton (Friedland et al., 2018; Yang et al., 2020). The peak concentration date at the same latitude in the North Atlantic generally precedes that in the North Pacific (Fig. 7). In the southern hemisphere, there is a conspicuous DMS-rich zone near 40° S (where the Subtropical Convergence lies) in summer, delineating a ring-shaped high-concentration band nearly parallel to the latitude. The highest seasonal mean concentration (December–February) occurs at 41.5° S, reaching 3.71 nM (Fig. 9). Southward from this zone, a low-DMS area spans 47°–61° S, where the average concentration is below 2.5 nM across all seasons. However, in the coastal waters of Antarctica (south of 60° S), significantly high concentrations also manifest in summer, surpassing 4.0 nM, even higher than those near 40° S (Fig. 6 and 9). In addition to the above regions, several typical upwelling zones also exhibit relatively higher DMS concentrations, such as the eastern Pacific and the Southeast Atlantic. The former, situated at lower latitudes, shows no significant seasonal variation, while the latter exhibits higher concentrations from October to February. The high nutrient concentrations in upwelling areas can bolster primary productivity, intensifying biological activities and augmenting the production of biogenic sulfur.

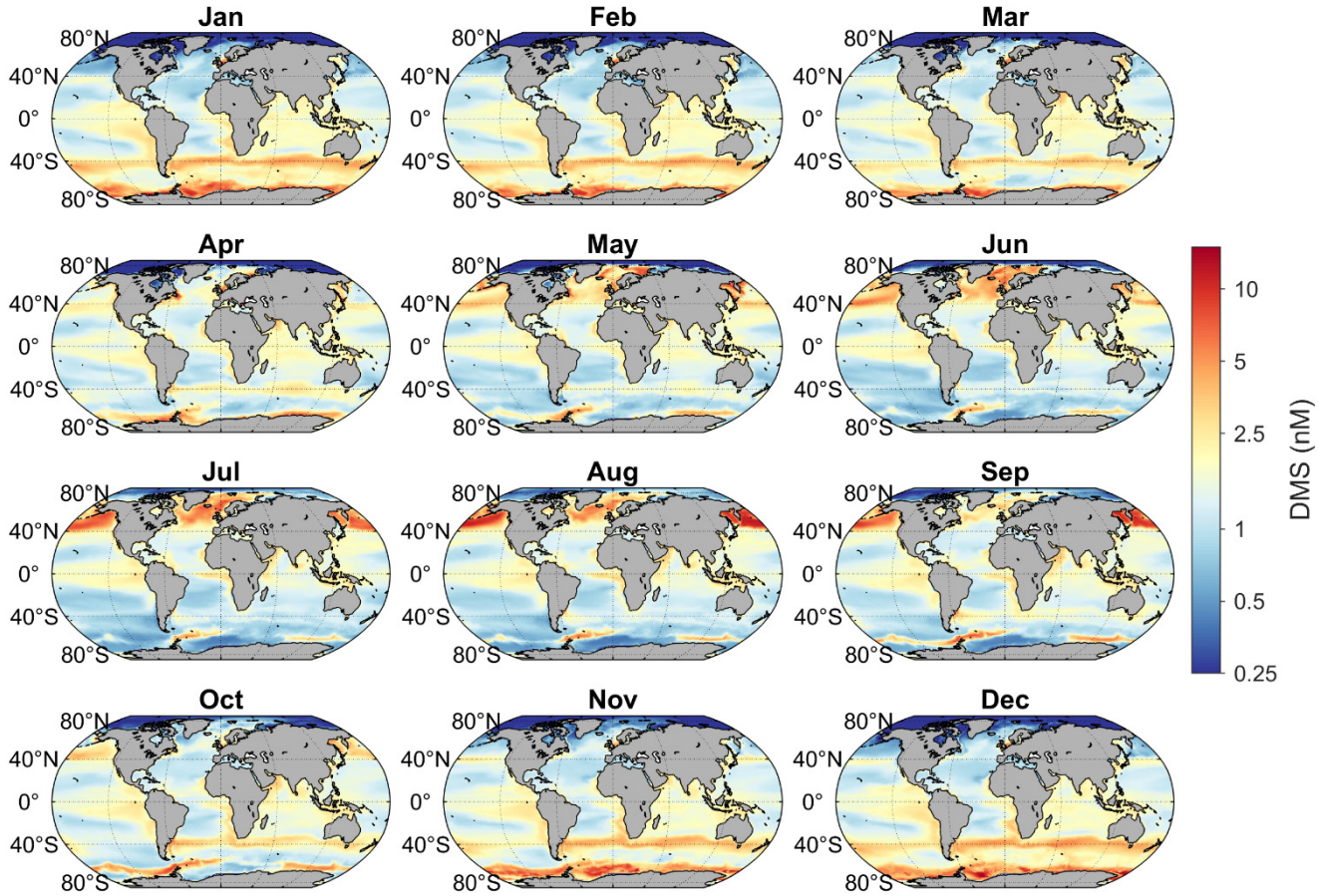

Figure 6. Monthly climatology of global sea surface DMS concentration during 1998 to 2017.

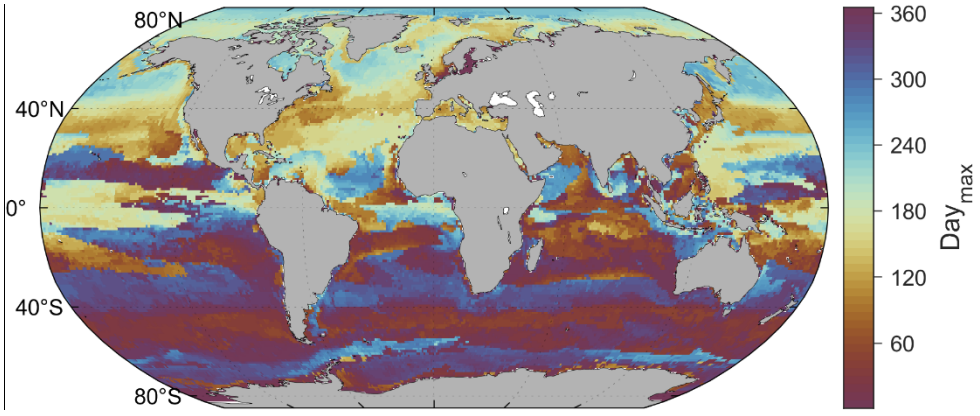

Figure 7. The day of the year with the highest sea surface DMS concentration for each grid point.

The spatiotemporal variation of DMS emission flux is generally consistent with that of concentration. As shown in Fig. 8, DMS fluxes are also significantly higher in summer across most mid- and high-latitude regions, and the high-flux regions generally overlap with the hot spots of DMS concentration. This indicates that the distribution of sea surface DMS concentration is the main factor controlling the monthly variation pattern of DMS emissions at the global scale, and the effect of transfer velocity is secondary. However, certain regions present inconsistencies between DMS flux and concentration dynamics. For instance, in the Arabian Sea and the central Indian Ocean, elevated transfer velocities (Fig. S10) during the June to September, driven by heightened wind speeds, markedly enhance emission fluxes, despite comparatively lower concentrations than other months. In polar regions, especially along the coast of Antarctica, although the DMS concentration is high in summer, sea ice coverage significantly impedes DMS release, thus the emission flux remains at a low level.

As shown in Fig. 9, the higher wind speeds in autumn and winter at mid- and high-latitudes result in higher total transfer velocities, leading to smaller summer-to-winter ratios of DMS emission flux compared to that of DMS concentration. In low latitudes, the existence of the trade wind zones in both hemispheres further leads to two high-flux bands. The emission fluxes in the equatorial region between these two trade zones are significantly lower. Although the latitudinal distributions of mean DMS emission fluxes in the southern and northern hemispheres are almost symmetrical, the huge difference in ocean area between the two hemispheres results in a significantly higher total emission from the southern hemisphere. Since anthropogenic SO_2 emissions are mainly concentrated in the northern hemisphere, oceanic DMS plays a much more important role in the southern hemisphere, especially over the regions south of 40° S where the DMS emission is high and the perturbation of anthropogenic pollution is low.

According to our newly built DMS gridded dataset, the global area-weighted annual mean concentration of DMS at the sea surface from 1998 to 2017 was ~1.71 nM (1.67–1.75 nM), which is within the range among the values (1.6 to 2.4 nM) obtained by various methods in previous studies (Tesdal et al., 2016). The global annual mean DMS emission to the atmosphere was 17.2 TgS yr^{-1} (16.9–17.5 TgS yr^{-1}), with 10.3 TgS yr^{-1} (59.9%) from the southern hemisphere and 6.9 TgS yr^{-1} (40.1%) from the northern hemisphere.

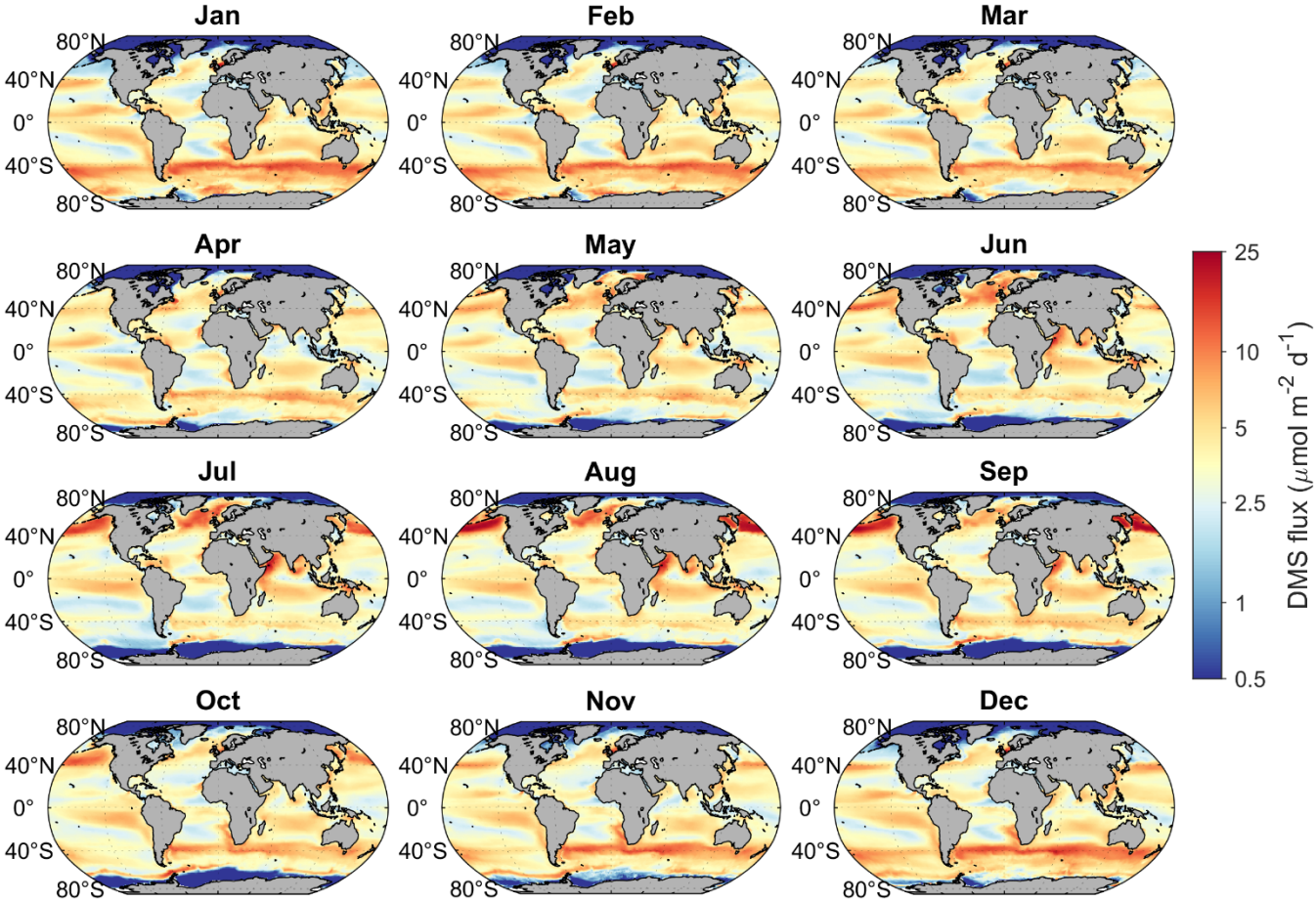

Figure 8. Monthly climatology of global DMS sea-to-air flux from 1998 to 2017.

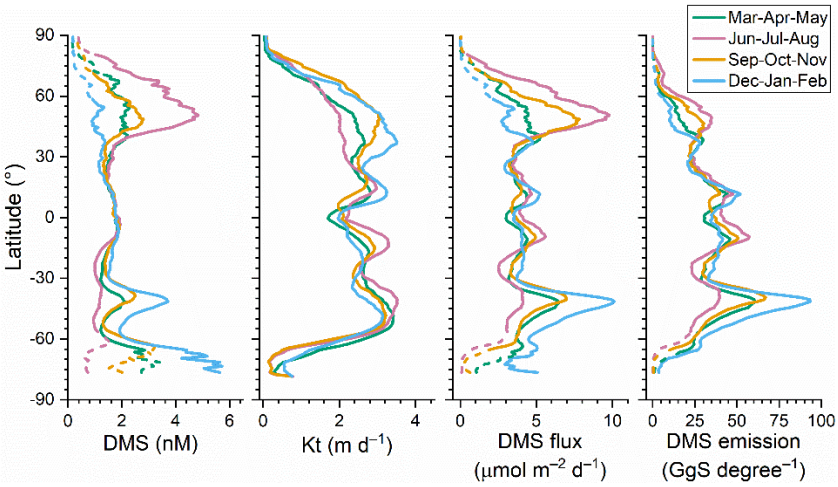

Figure 9. Latitudinal distributions of sea surface DMS concentration, total transfer velocity (Kt), sea-to-air flux, and total emission in different seasons during 1998–2017. The dashed parts of the lines represent the missing ratio of satellite Chl *a* data for DMS simulation is higher than 0.5, thus most Chl *a* data is from CMEMS global biogeochemical multi-year hindcast.

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 (Fig. 10), 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).

Overall, all datasets exhibit the general pattern of high DMS concentration during summer and low concentration during winter, but notable distinctions emerge 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 during summer and in the South Indian Ocean and Southwest Pacific Ocean from December to February (Fig. 10a). Particularly in the southern polar region (Polar_S), latitudinal averages of DMS_{L11} surpass 10 nM during summer, which are 1–3 times higher than DMS_{Z23} (Fig. 10e). However, DMS_{Z23} maintains a similar level around the Antarctic in March compared to summer, and it is significantly higher than DMS_{L11} as well as other three climatologies. DMS_{H22} shows lower disparities with DMS_{Z23} in the Arctic, 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} (Fig. 10b). In contrast, DMS_{H22} in Polar_S from September to November is ~2 nM lower than DMS_{Z23}. The

global area-weighted annual mean DMS concentrations in L11 and H22 are 2.43 nM and 2.26 nM, respectively, which are approximately 42.1% and 32.2% higher than Z23.

G18 exhibits the lowest global annual mean concentration (1.63 nM) among these climatologies, approximately 4.7% lower than Z23. The most notable deviation occurs in the North Pacific during boreal summer and near the Antarctic during austral summer, where DMS_{Z23} is > 3.5 nM (> 100%) higher than DMS_{G18} (Fig. 10c). Conversely, there are high DMS concentrations (> 5 nM) in certain coastal seas (such as the coasts of East and Northeast Asia, the coasts of Patagonia and Peru, the southwestern coast of Africa, and the western coasts of the Sahara Desert and North America) based on the G18 estimate. This characteristic is not fully replicated by other DMS fields, possibly due to the underestimation of DMS by our model and other methods in coastal regions as well as the overestimation of Chl *a* by satellites caused by the interference of colored dissolved organic matters and non-algal detrital particles (Aurin and Dierssen, 2012). W20 exhibits the highest consistency with Z23 in spatiotemporal distribution patterns as well as the lowest difference in global annual mean concentration (1.74 nM, only 1.8% higher than Z23). However, notable discrepancies exist in specific regions. For instance, during summertime, DMS_{Z23} is > 1 nM (> 40%) lower than DMS_{W20} in more than half of the Arctic area, while in North Pacific and Southern Ocean DMS_{Z23} is significantly higher than DMS_{W20} (Fig. 10d). Furthermore, only DMS_{Z23} forms a nearly complete high-concentration annular band at ~ 40° S during austral summer.

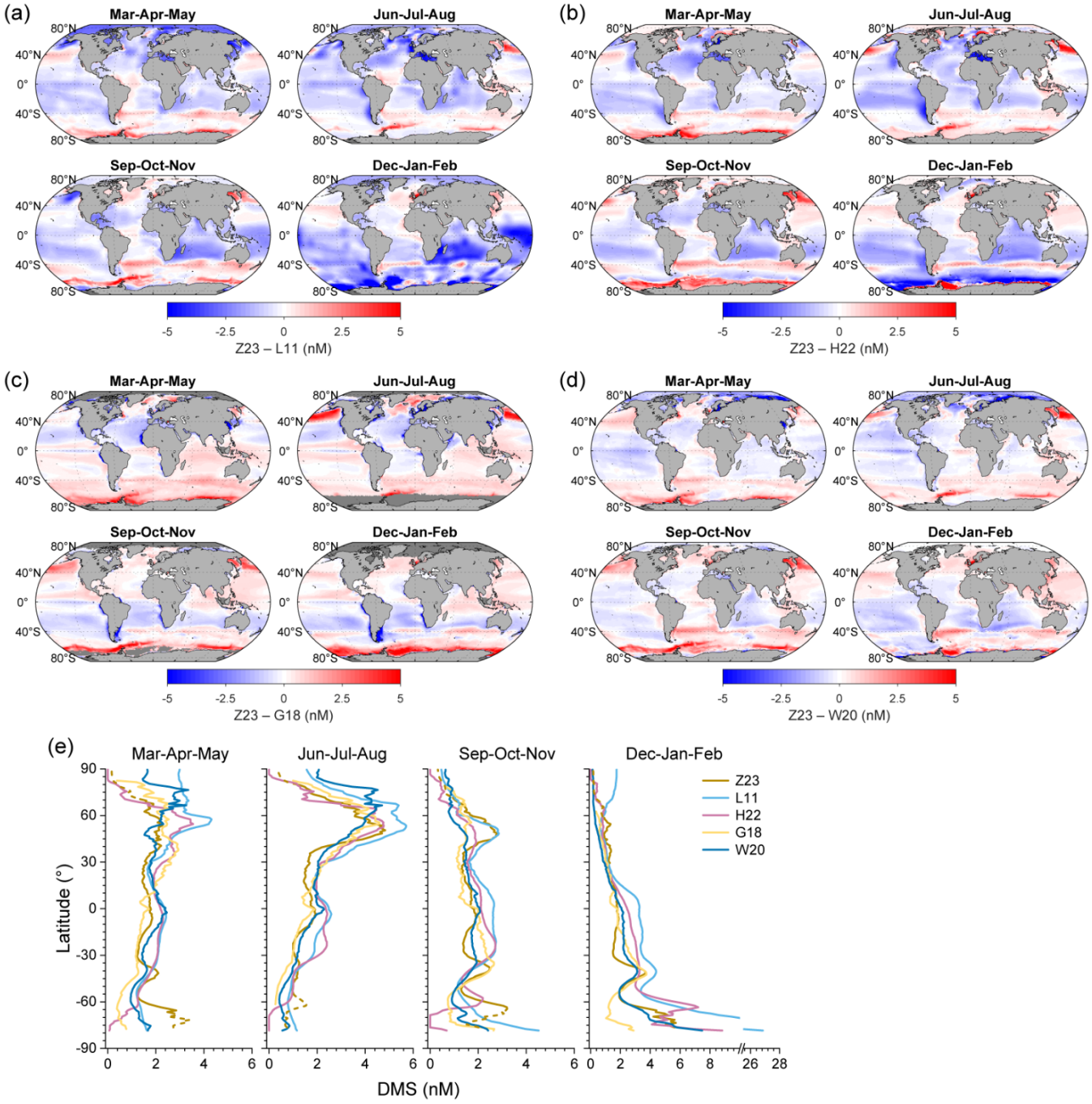

Figure 10. (a–d) The spatial distributions of DMS concentration differences between Z23 and four previously estimated fields across different seasons: (a) L11, (b) H22, (c) G18, and (d) W20. Dark gray regions in the ocean represent data missing in at least one field. (e) Comparisons between the latitudinal distributions of Z23 and four previous DMS fields across different

seasons. The dashed parts of the Z23 lines represent the missing ratio of satellite Chl *a* data for DMS simulation is higher than 0.5, thus most Chl *a* data is from CMEMS global biogeochemical multi-year hindcast.

3.2.3 Decadal changes

One of the advantages of our ANN-derived DMS dataset is its time-resolved nature, which enables us to investigate the interannual variations in sea surface DMS concentration and flux. Here we present the decadal trends of DMS concentration, Kt, and emission flux spanning from 1998 to 2017 at both global and regional scales. Overall, the absolute interannual variability of DMS concentration across most global oceanic regions appears relatively small. 88.4% of the global oceanic area exhibited a range of less than 1 nM between the maximum and minimum annual average concentrations during this 20-year period, particularly evident in tropical and subtropical regions with latitudes between 40° S and 40° N. At latitudes higher than 40° in both hemispheres, notable decadal changes occurred (Fig. 11a). Annual mean DMS concentrations in the Greenland Sea, the North Pacific, and the Southern Ocean exhibited significant decreasing trends with rates exceeding 0.03 nM yr^{-1} ($P < 0.05$). A significant decreasing trend was also noted in the eastern tropical Pacific Ocean, albeit at a much lower absolute rate, primarily below 0.015 nM yr^{-1}. Conversely, there were significant increasing trends in the Labrador Sea, the South Pacific (35° S – 60° S, 150° E – 75° W), and the southeastern Pacific, with the highest rate exceeding 0.02 nM yr^{-1}. The global annual mean concentration exhibited a decreasing trend with a rate of 0.0035 nM yr^{-1} ($P < 0.05$, Fig. 11d). The highest value (1.75 nM) occurred in 1999, and the lowest concentration (1.67 nM) occurred in 2015. Due to the primary influences of increasing WS and secondary impact of rising SST in most mid- and low-latitude regions (Fig. S11), the Kt of DMS also showed an overall increasing trend, especially in the eastern Pacific and Atlantic Ocean (Fig. 11b). The increase in Kt can offset the decrease in DMS concentration to some extent, resulting in no significant trend in global DMS emissions during this 20-year period (Fig. 11d).

In the Arctic region, which stands as one of the most sensitive areas to climate warming (Screen et al., 2012; Serreze and Barry, 2011), the sea ice coverage has undergone significant reduction over the past two decades, particularly noticeable in the Barents Sea and Kara Sea, and further north (> 1% yr^{-1} for annual mean SI, Fig. S11). The retreat of summertime sea ice leads to an expansion of open-sea surface, potentially amplifying DMS emission (Galí et al., 2019). However, despite this trend, there was no significant increase in the annual total emission from the Polar_N region over the same period, primarily due to a decreasing trend in DMS concentration (Fig. 12). On the other hand, the highest emission took place in the last two years (> 0.64 Tg yr^{-1}), attributed to the highest Kt. Thus, it is likely that a rise in DMS emission will appear in future Arctic region with further loss of sea ice coverage (Galí et al., 2019). In contrast to the Arctic, the Southern Ocean has experienced a significant increase in sea ice fraction (Fig. S11), leading to a significant decrease in Kt (Fig. 11b). Coupled with the decreased DMS concentration, it resulted in a substantial decline in the DMS emission flux (Fig. 11c and 12). The highest annual total emission flux in the Polar_S region occurred in 1998 (1.49 TgS), while the lowest occurred in 2013 (1.02 TgS), representing a decrease

of ~32%. Across other oceanic regions, the annual average DMS concentrations in the Westerlies_N_Pacific and Trades_Pacific regions exhibit decreasing trends over the past 20 years, while the concentration in Westerlies_S and Trades_Atlantic has increased ($P < 0.05$, Fig. 12). Regarding DMS flux, the Westerlies_N_Pacific showed a decrease, while the Westerlies_N_Atlantic, Westerlies_S, and Trades_Atlantic showed an increase. There was no significant trend in other low-latitude regions.

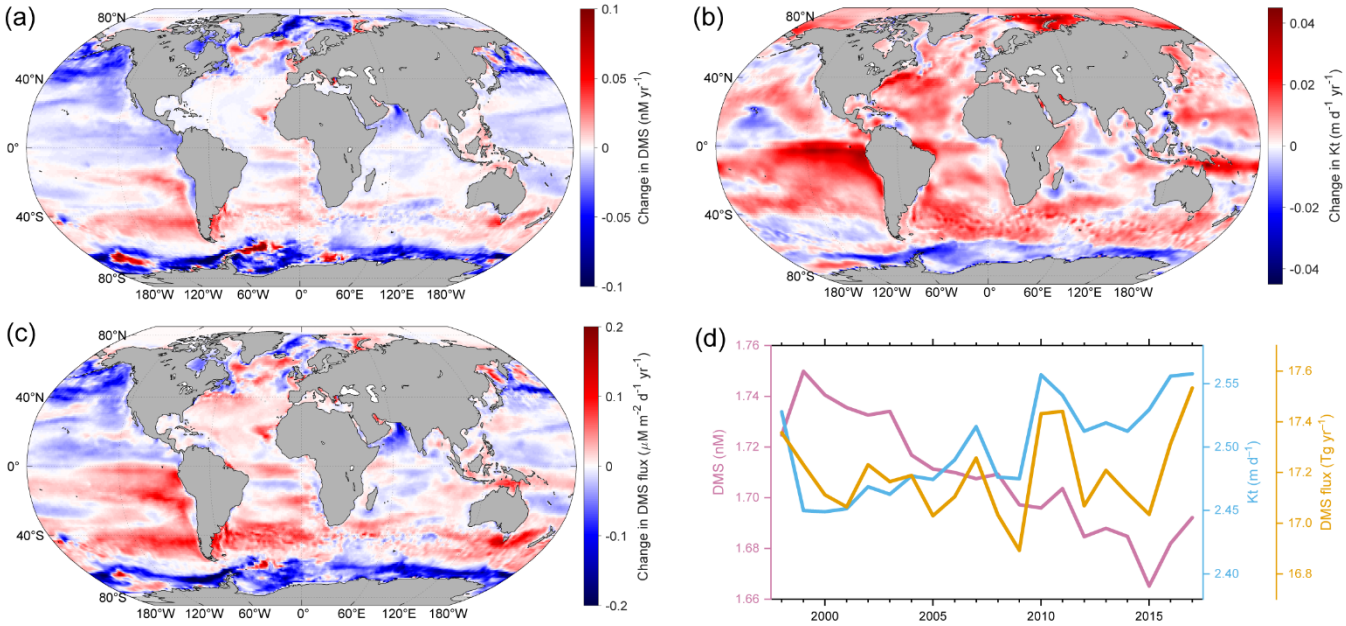

Figure 11. (a–c) The spatial distributions of changes in (a) DMS concentration, (b) Kt, and (c) DMS emission flux from 1998 to 2017. The linear regression slopes for the annual means are taken as the changing rates here. (d) The temporal changes of global annual mean DMS concentration, Kt, and total emission flux from 1998 to 2017.

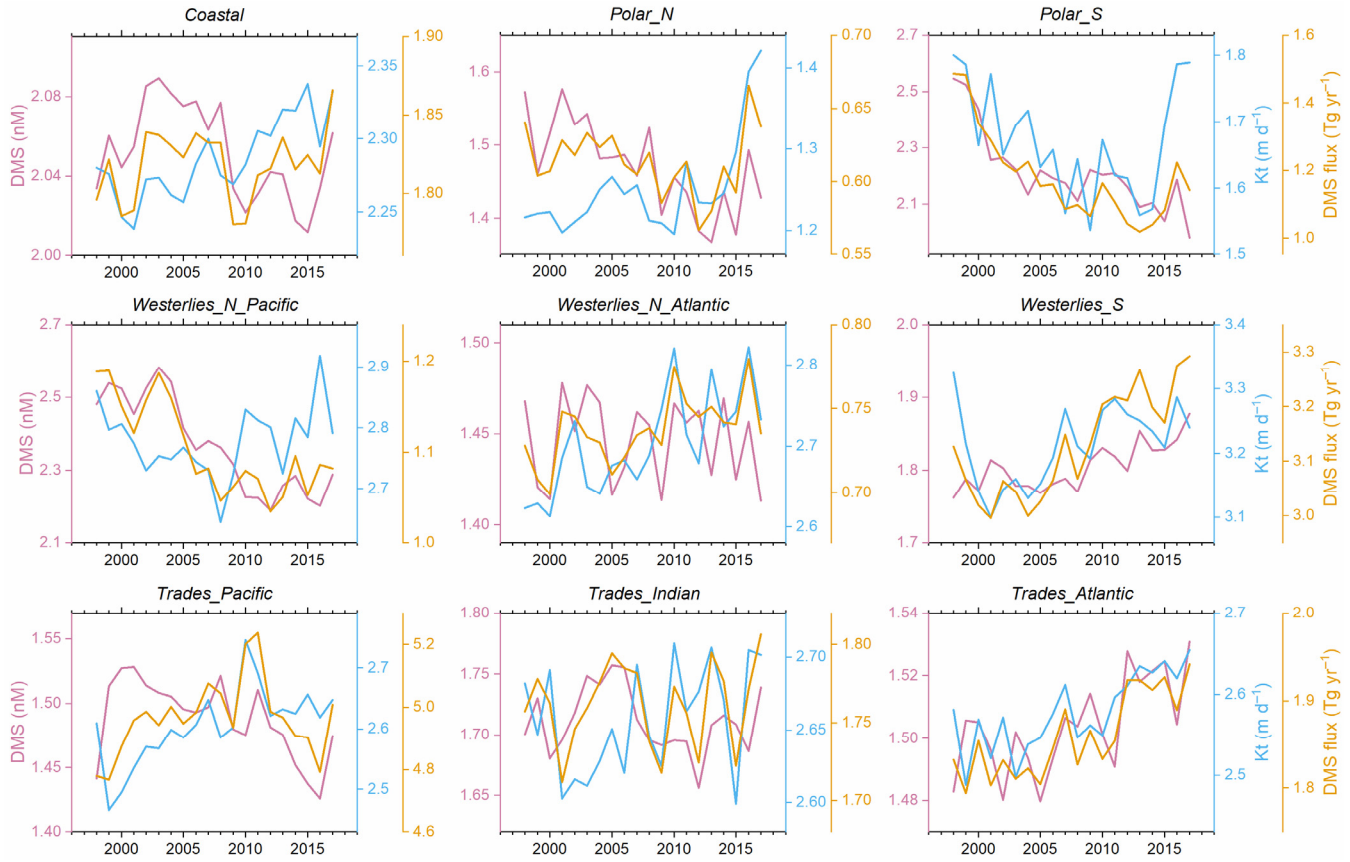

Figure 12. The temporal changes of annual mean DMS concentration, Kt, and total emission flux in different regions from 1998 to 2017.

3.3 Connection with atmospheric biogenic sulfur

One of the primary objectives of developing this daily gridded DMS dataset (Z23) spanning multiple years is to improve the emission inventory of marine biogenic DMS, thereby enhancing the modelling performance for atmospheric sulfur chemistry, especially for simulating sulfur-containing aerosols. To assess whether our newly constructed DMS dataset can reach this objective, we employed a backward trajectory-based method to examine the correlation between sea surface DMS emissions and resulting DMS oxidation products in the atmosphere. The correlation was then compared against those derived from previously reported DMS climatologies (i.e., 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 of the variation of MSA concentration on the DMS emission fluxes. During four transection cruises in the Atlantic conducted by *R/V Polarstern* (20 April – 20 May 2011, 28 October – 1 December 2011, 10 April – 15 May 2012, and 27 October – 27 November 2012), the MSA concentrations in submicron aerosols were measured online using a High-Resolution Time-of-Flight Aerosol Mass Spectrometer. The ship tracks are shown in Fig. S12, and detailed information about the cruises and measurement methodology was provided by Huang et al. (2016). The 72-hour air mass backward trajectories reaching the ship position were calculated every hour by the HYSPLIT model, starting from a height of 100 m (Stein et al., 2015). Subsequently, the air mass exposure to DMS emission (AEDMS), denoting the weighted average of DMS emission flux along the trajectory path, was calculated following the approach of Zhou et al. (2021). We used 5 different DMS gridded datasets, including Z23, L11, H22, G18, and W20. For Z23, the calculated daily DMS fluxes were utilized. For the remaining 4 monthly climatologies, we applied the daily Kt data from Z23 to calculate the DMS fluxes, thus eliminating the potential confounding influences stemming from different Kt parameterizations. In this calculation, the same concentration was assigned to all days within a month without interpolation. Detailed procedures for the calculation of AEDMS are elucidated in Appendix C.

MSA concentrations were significantly higher in late spring than those in autumn for both North and South Atlantic Oceans (Fig. 13a). For example, during the boreal spring cruise in 2011, the average MSA concentration over the North Atlantic (0.068 μg m^{-3}, north of 25° N) was about an order of magnitude higher than the average concentration over the South Atlantic (0.006 μg m^{-3}, south of 5° S). During the boreal autumn cruise in 2011, the average concentration over the South Atlantic (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 to this major seasonal pattern, there was also a minor MSA concentration peak between 5°–15° N in both seasons. The spatial and seasonal variations of AEDMS based on the Z23 dataset (referred to as AEDMS_Z23) largely coincided with these MSA concentration patterns (Fig. 13a). It should be noted that the MSA/AEDMS ratio between 5°–15° N was significantly lower than those in other high-MSA regions, which may result from the DMS simulation biases near the coast of West Africa or the lower DMS-to-MSA conversion yields related with air temperature and oxidant species (Barnes et al., 2006; Bates et al., 1992). There were also several AEDMS peaks in North Atlantic during November 2012, inconsistent with the continuously low MSA concentrations. Given the high precipitation rates along the trajectory (Fig. 13a), a strong wet scavenging process might significantly reduce aerosol concentrations (Wood et al., 2017).

The AEDMS derived from other DMS concentration fields showed similar variations to AEDMS_Z23 (Fig. 13a). It is not surprising since all DMS concentration fields exhibit similar large-scale spatiotemporal patterns, and identical air mass transport path and Kt were applied in different AEDMS calculations. However, due to the lower temporal resolutions and absence of interannual changes in those DMS monthly climatologies, the resulting AEDMS may be less effective in capturing variability at finer scales or across different years. Here we focus on the high-MSA periods to elaborate on this issue, which corresponds to latitudes north of 25° N in boreal spring (S1 and S2 in Fig. 13a), 25° N – 25° S in boreal autumn of 2011 (A1

in Fig. 13a), and south of 5° N in boreal autumn of 2012 (A2 in Fig. 13a). As shown in Fig. 13b, hourly MSA concentrations exhibited significantly stronger correlations with AEDMS_Z23 than with other AEDMS time series in S1 and S2, indicating AEDMS_Z23 can explain more (1.31 – 1.69 times) variance of MSA concentration. During A1 and A2, the correlations between AEDMS and MSA concentration were weaker than those during S1 and S2, possibly due to higher DMS prediction biases in South Atlantic or different influencing factors on atmospheric DMS chemistry across wide spatial ranges. Nonetheless, AEDMS_Z23 still exhibited the highest correlation with MSA (Fig. 13c). This overall stronger connection between Z23 and atmospheric DMS-derived aerosols mainly benefited from the combined effects of higher time resolution and inherent interannual variations. For example, the ratio of average MSA concentration during S1 to that during S2 (S1-to-S2 ratio) was 1.89, and the A2-to-A1 ratio was 1.75. AEDMS_Z23 exhibited a slightly lower but still significant interannual variation degree, where the S1-to-S2 ratio and A2-to-A1 ratio were 1.58 and 1.46, respectively. However, this interannual variation cannot be reproduced by other datasets, where the S1-to-S2 ratio and A2-to-A1 ratio were in the range of 1.08–1.30 and 1.19–1.29, respectively. These results manifest the potential of our newly developed DMS gridded data product to enhance the modeling performance for atmospheric DMS processes compared with previously reported climatologies.

It is worth noting that the satellite-based algorithms of G18 and ANN model of W20 can also be utilized to produce daily multiyear DMS fields as Z23. Future investigations could include comparisons with these fields, facilitating a more comprehensive assessment of the performance of each algorithm/model. Furthermore, the AEDMS method used here is a highly simplified approach without considering the complex DMS chemistry in the atmosphere, and the intercomparisons based on chemical transport models can be used in the future to obtain a more straightforward conclusion.

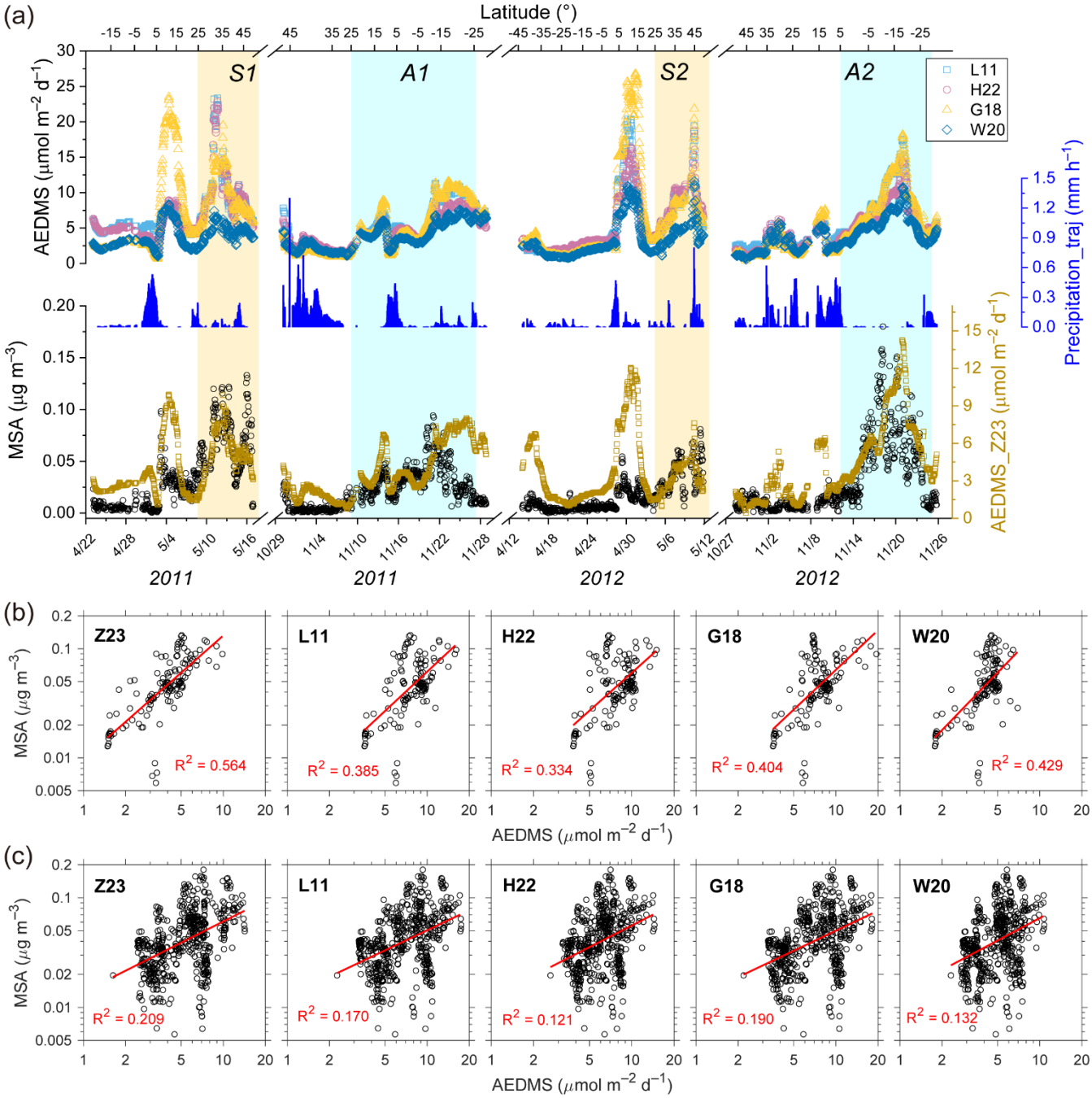

Figure 13. (a) Time series of observed MSA concentration, AEDMS calculated based on different DMS concentration datasets, and average precipitation along the backward trajectory (Precipitation_traj) during four Atlantic cruises in 2011–2012. (b–c) Correlations between hourly MSA concentration and AEDMS based on different DMS concentration datasets (b)

during periods S1 + S2 and (c) during periods A1 + A2. Data points during the periods with air mass time fraction within the boundary layer less than 90% or Precipitation_traj larger than 0.05 mm h^{-1} were removed.

4 Uncertainties and limitations

Although our ANN ensemble model and derived DMS dataset demonstrate certain advantages compared to previous studies, as discussed in Section 3.3, there persist notable uncertainties and limitations, which result in the ~35% uncaptured variance (Fig. 3a) and non-negligible simulation biases, e.g., underestimation of extremely high DMS concentrations and overestimation of low DMS concentrations. Firstly, there is a mismatch in the spatial and temporal scales between the input and target. The target, sea surface DMS concentrations, are obtained from in-situ measurements taken at specific locations and time points. In contrast, the input data are primarily from gridded datasets where each pixel represents an average over a defined spatial and temporal range. This is particularly significant for the ECCO variables, which have the largest spatial grid size of 110 km. Consequently, extreme values at specific locations cannot be accurately captured by the regional averages, resulting in dampened variations among the samples. Secondly, the input data from different sources and the observed sea surface DMS concentrations inherently possess certain uncertainties, which can introduce noises into the ANN learning process. Thirdly, the ANN itself may not be powerful enough to fully capture the complex input-output relationships across different oceanic regions, especially when the samples are scarce under specific environmental conditions. Finally, beyond the 9 variables incorporated in this study, other environmental parameters such as pH (Six et al., 2013; Hopkins et al., 2010) and trace metal elements (Li et al., 2021) can also influence DMS concentration. Not incorporating these factors may introduce additional biases.

The overall bias for $\log_{10}$DMS is at a similar level between high- and low-concentration ends, but the DMS concentration on a linear scale is more underestimated in the high-concentration regime than it is overestimated in the low-concentration regime. As a result, our simulation results may tend to underestimate the annual average DMS concentration and flux. To mitigate this critical bias and reduce model uncertainty, high-quality input datasets with finer spatial resolution are needed in the future. The high-time resolution nature of the resulted daily DMS data product would be more valuable if accompanied by higher spatial resolution. Expanding the data volume is also crucial for improving model performance. Although the current DMS observational data covers all major oceanic basins, certain regions such as the Trades_Pacific remain underrepresented. Advances in online measurement technologies offer promising avenues for acquiring more extensive and convenient observational data (Hulswar et al., 2022). Additionally, incorporating more input features to the model would be beneficial. This necessitates a comprehensive understanding of the spatiotemporal distributions of those input features, and further field measurements are important to this end. Moreover, integrating DMS biogeochemical mechanisms with machine learning technique, i.e., a hybrid model coupling physical processes with data-driven approach, may further improve prediction accuracy, generalization, and interpretability (Reichstein et al., 2019).

When using our newly developed DMS dataset, there are two issues that need to be noted. Firstly, there is a significant portion of missing satellite Chl *a* data during winter in polar regions. In such instances, the modeling data from CMEMS global biogeochemical multi-year hindcast was used, which may introduce higher uncertainty. We have provided the flags indicating the source of Chl *a* data for each grid in the dataset. Nevertheless, given the low phytoplankton biomass and extensive sea ice coverage during winter, DMS emissions are typically at the lowest level of the year, thus the satellite data missing has a relatively small impact on investigating the subsequent effects of DMS emission on atmospheric environment. Secondly, since the ANN ensemble model exhibits limited capacity in accurately reproducing extremely high concentrations of DMS, the DMS concentrations in certain nearshore areas with intensive biological activity may be greatly underestimated.

5 Code and data availability

The generated gridded datasets of DMS concentration, total transfer velocity, and flux have been deposited at https://doi.org/10.5281/zenodo.11879900 (Zhou et al., 2024) and can be downloaded publicly. The ANN model code and the Matlab scripts for data analysis are available from https://doi.org/10.5281/zenodo.12398985 (Zhou, 2024).

6 Conclusions

Based on the global sea surface DMS observations and associated data of 9 relevant environmental variables, an ANN ensemble model was trained. The ANN model effectively captures the variability of DMS concentrations and demonstrates good simulation accuracy. Leveraging this ANN model, a global sea surface DMS gridded dataset with a daily resolution spanning 20 years (1998–2017) was constructed. The global annual average concentration was ~1.71 nM, falling within the range of previous estimates, and the annual total emission was ~17.2 TgS yr^{-1}. High DMS concentrations and fluxes took place during summer in North Pacific (40°–60° N), North Atlantic (50°–80° N), the annular band around 40° S, and the Southern Ocean. With this newly developed dataset, the day-to-day changes and interannual variations can be investigated. The global annual average concentration shows a mild decreasing trend (~0.0035 nM yr^{-1}), while the total emission remains stable. There were more significant decadal changes in certain regions. Specifically, the annual DMS emission in the South Pacific and North Pacific showed opposite trends.

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

Appendix A: Abbreviations

AEDMS	Air mass exposure to DMS emission
ANN	Artificial neural network
BLH	Boundary layer height
CCN	Cloud condensation nuclei
Chl *a*	Chlorophyll *a*
DMS	Dimethyl sulfide
DMSP	Dimethylsulfoniopropionate
DO	Dissolved oxygen
DSWF	Downward short-wave radiation flux
ECCO	Estimating the Circulation and Climate of the Ocean
GSSD database	Global Surface Seawater DMS database
Kt	Total transfer velocity
MLD	Mixed layer depth
MB	Mean bias
MSA	Methanesulfonic acid
MSE	Mean square error
NAAMES	North Atlantic Aerosols and Marine Ecosystems Study
NMB	Normalized mean bias
RMSE	Rooted mean square error
SI	Sea ice fraction
SST	Sea surface temperature
SSS	Sea surface salinity
WS	Wind speed

Appendix B: The weighted resampling strategy

Apart from the data imbalance between coastal and non-coastal regions, there exists an imbalance across different DMS concentration ranges. The majority of DMS concentrations (78.6%) fall within the range of 0.8 to 10 nM ($\log_{10}$(DMS) between -0.1 to 1). Samples with DMS concentrations exceeding 15 nM or falling below 0.3 nM only represent 6.9% of the entire sample set. A weighted resampling strategy was applied to mitigate this imbalance (Fig. S7). We randomly sampled 50,000 samples with replacement from the original sample set. The probability of each sample being selected is proportional to the weighting factor shown as the red dash line in Fig. S7b, which is dependent on its DMS concentration. First, the probability distribution of initial $\log_{10}$(DMS) values was fitted with a gamma distribution, which is given below and displayed as the blue line in Fig. S7b:

$$f(x) = \frac{1}{\Gamma(k)\theta^k}(x+4)^{k-1}e^{-(x+4)/\theta} \quad \text{(B1)}$$

Here k and θ represent the shape parameter and scale parameter, in this case, 100.7 and 0.044, respectively. x is the $\log_{10}$(DMS) value. Since gamma distribution only takes positive values, we added 4 to the original x as the dependent variable for distribution fitting. We then obtained a new gamma distribution function with the same mode but lower shape parameter, in which k = 40 and θ = 0.112. The reciprocal of the new gamma distribution function was taken as the weighting factor. As a result, samples exhibiting high or low DMS concentrations are more likely to be selected, whereas those with intermediate concentrations are less likely to be selected. We also controlled the F_{coastal} value of the resampled data equal to 9.7%. The data distribution of DMS concentrations after the resampling process is shown in Fig. S7c. The fraction of samples with DMS concentrations above 15 nM or below 0.3 nM is elevated to 15.0%. The 50,000 samples were then randomly split to a training set (80%) and a validation set (20%). Since there are duplicate samples in the resampled dataset, the random data split was conducted based on the original sample ID before resampling to ensure that there was no sample overlap between the training and validation sets.

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

Here the AEDMS index followed the similar calculation of the air mass exposure to Chl *a* (AEC) in previous studies (Arnold et al., 2010; Park et al., 2018; Zhou et al., 2021). We adopted the similar approach presented in Zhou et al. (2021) by replacing the Chl *a* concentration with DMS flux, as shown in the following Eq. (C1):

$$AEDMS = \frac{\sum_{i=0}^{72} DMS\ flux_i \cdot e^{-\frac{t_i}{72}} \cdot \frac{600}{BLH}}{\sum_{i=0}^{72} e^{-\frac{t_i}{72}}} \quad \text{(C1)}$$

Here i represents the i-th trajectory point of the 72-hour backward trajectory (0-th for the receptor point). $DMS\ flux_i$ represents the DMS flux of the pixel where the i-th trajectory point locates. $DMS\ flux_i$ is set to zero if the point locates on land or the air mass pressure is below 850 hPa (usually in the free troposphere with little influence of surface emission). t_i 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 closer to the receptor point. To better connect with the atmospheric concentrations in the marine boundary layer, the normalization by boundary layer height (BLH) is added by the $\frac{600}{BLH}$ term. The BLH below 50 m is replaced by 50 m.

Author contributions.

SZ and YC designed the research. SZ, FW, ZX, and KY collected the data and did the data preprocessing. SZ implemented the model development and performed the simulation with assistance from GY, HZ, and YZ. SH, HH, AW, and LP provided the measurement data of atmospheric MSA over the Atlantic Ocean. SZ conducted the data analysis and visualization with advice from YC and XG. SZ and YC wrote the manuscript with inputs from all authors.

Competing interests.

The authors declare that they have no conflict of interest.

Acknowledgements.

We greatly thank National Oceanic and Atmospheric Administration's Pacific Marine Environmental Laboratory for maintaining the Global Surface Seawater DMS Database. We acknowledge Dr. Martin Johnson for sharing the code of DMS transfer velocity calculation. We also thank Dr. Rich Pawlowicz for developing and sharing the M_Map toolbox for Matlab (https://www.eoas.ubc.ca/~rich/map.html), which was used in the mapping of this study. XG was supported by the Research Center for Industries of the Future (RCIF) at Westlake 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.

Aurin, D. A., and Dierssen, H. M.: Advantages and limitations of ocean color remote sensing in CDOM-dominated, mineral-rich coastal and estuarine waters, Remote Sensing of Environment, 125, 181-197, 10.1016/j.rse.2012.07.001, 2012.

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.

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

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

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

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

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

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

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P.: SMOTE: synthetic minority over-sampling technique, Journal of artificial intelligence research, 16, 321-357, 2002.

716 Chen, Q., Sherwen, T., Evans, M., and Alexander, B.: DMS oxidation and sulfur aerosol formation in the marine troposphere:
717 a focus on reactive halogen and multiphase chemistry, Atmos. Chem. Phys., 18, 13617-13637, 10.5194/acp-18-13617-2018,
718 2018.
719 Dubitzky, W., Granzow, M., and Berrar, D. P.: Fundamentals of data mining in genomics and proteomics, Springer Science
720 & Business Media, 2007.
721 Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M., and Wunsch, C.: ECCO version 4: An integrated framework
722 for non-linear inverse modeling and global ocean state estimation, Geosci. Model Dev., 8, 3071-3104, 2015.
723 Friedland, K. D., Mouw, C. B., Asch, R. G., Ferreira, A. S. A., Henson, S., Hyde, K. J. W., Morse, R. E., Thomas, A. C., and
724 Brady, D. C.: Phenology and time series trends of the dominant seasonal phytoplankton bloom across global scales, Global
725 Ecology and Biogeography, 27, 551-569, 10.1111/geb.12717, 2018.
726 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,
727 D. R., Jimenez, J. L., Campuzano-Jost, P., Veres, P. R., Bates, T. S., Shilling, J. E., and Zawadowicz, M.: Exploring dimethyl
728 sulfide (DMS) oxidation and implications for global aerosol radiative forcing, Atmos. Chem. Phys., 22, 1549-1573,
729 10.5194/acp-22-1549-2022, 2022.
730 Galí, M., Devred, E., Levasseur, M., Royer, S.-J., and Babin, M.: A remote sensing algorithm for planktonic
731 dimethylsulfoniopropionate (DMSP) and an analysis of global patterns, Remote Sensing of Environment, 171, 171-184,
732 10.1016/j.rse.2015.10.012, 2015.
733 Galí, M., and Simó, R.: A meta-analysis of oceanic DMS and DMSP cycling processes: Disentangling the summer paradox,
734 Glob. Biogeochem. Cycles, 29, 496-515, 10.1002/2014gb004940, 2015.
735 Galí, M., Levasseur, M., Devred, E., Simó, R., and Babin, M.: Sea-surface dimethylsulfide (DMS) concentration from satellite
736 data at global and regional scales, Biogeosciences, 15, 3497-3519, 10.5194/bg-15-3497-2018, 2018.
737 Galí, M., Devred, E., Babin, M., and Levasseur, M.: Decadal increase in Arctic dimethylsulfide emission, P. Natl. Acad. Sci.
738 USA, 116, 19311-19317, 10.1073/pnas.1904378116, 2019.
739 Garnesson, P., Mangin, A., Fanton d'Andon, O., Demaria, J., and Bretagnon, M.: The CMEMS GlobColour chlorophyll a
740 product based on satellite observation: Multi-sensor merging and flagging strategies, Ocean Science, 15, 819-830, 2019.
741 Haibo, H., Yang, B., Garcia, E. A., and Shutao, L.: ADASYN: Adaptive synthetic sampling approach for imbalanced learning,
742 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008,
743 1322-1328.
744 Hoffmann, E. H., Tilgner, A., Schroedner, R., Bräuer, P., Wolke, R., and Herrmann, H.: An advanced modeling study on the
745 impacts and atmospheric implications of multiphase dimethyl sulfide chemistry, P. Natl. Acad. Sci. USA, 113, 11776-11781,
746 10.1073/pnas.1606320113, 2016.
747 Holder, C., Gnanadesikan, A., and Aude-Pradal, M.: Using neural network ensembles to separate ocean biogeochemical and
748 physical drivers of phytoplankton biogeography in Earth system models, Geosci. Model Dev., 15, 1595-1617, 10.5194/gmd-
749 15-1595-2022, 2022.

Hopkins, F. E., Turner, S. M., Nightingale, P. D., Steinke, M., Bakker, D., and Liss, P. S.: Ocean acidification and marine trace gas emissions, P. Natl. Acad. Sci. USA, 107, 760-765, 10.1073/pnas.0907163107, 2010.

Hopkins, F. E., Archer, S. D., Bell, T. G., Suntharalingam, P., and Todd, J. D.: The biogeochemistry of marine dimethylsulfide, Nature Reviews Earth & Environment, 4, 361-376, 10.1038/s43017-023-00428-7, 2023.

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

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

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

Humphries, G. R. W., Deal, C. J., Elliott, S., and Huettmann, F.: Spatial predictions of sea surface dimethylsulfide concentrations in the high arctic, Biogeochemistry, 110, 287-301, 2012.

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

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

Kettle, A. J., Andreae, M. O., Amouroux, D., Andreae, T. W., Bates, T. S., Berresheim, H., Bingemer, H., Boniforti, R., Curran, M. A. J., DiTullio, G. R., Helas, G., Jones, G. B., Keller, M. D., Kiene, R. P., Leck, C., Levasseur, M., Malin, G., Maspero, M., Matrai, P., McTaggart, A. R., Mihalopoulos, N., Nguyen, B. C., Novo, A., Putaud, J. P., Rapsomanikis, S., Roberts, G., Schebeske, G., Sharma, S., Simo, R., Staubes, R., Turner, S., and Uher, G.: 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, Glob. Biogeochem. Cycles, 13, 399-444, 10.1029/1999gb900004, 1999.

Kloster, S., Feichter, J., Maier-Reimer, E., Six, K. D., Stier, P., and Wetzel, P.: DMS cycle in the marine ocean-atmosphere system–a global model study, Biogeosciences, 3, 29-51, 2006.

Lana, A., Bell, T. G., Simó, R., Vallina, S. M., Ballabrera-Poy, J., Kettle, A. J., Dachs, J., Bopp, L., Saltzman, E. S., Stefels, J., Johnson, J. E., and Liss, P. S.: An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean, Glob. Biogeochem. Cycles, 25, GB1004, 10.1029/2010gb003850, 2011.

Li, H., Zhou, S., Zhu, Y., Zhang, R., Wang, F., Bao, Y., and Chen, Y.: Atmospheric Deposition Promotes Relative Abundances of High-Dimethylsulfoniopropionate Producers in the Western North Pacific, Geophys. Res. Lett., 48, e2020GL092077, 10.1029/2020GL092077, 2021.

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

Lovelock, J. E., Maggs, R. J., and Rasmussen, R. A.: Atmospheric Dimethyl Sulphide and the Natural Sulphur Cycle, Nature, 237, 452-453, 10.1038/237452a0, 1972.

Mansour, K., Decesari, S., Ceburnis, D., Ovadnevaite, J., and Rinaldi, M.: Machine learning for prediction of daily sea surface dimethylsulfide concentration and emission flux over the North Atlantic Ocean (1998-2021), Sci. Total. Environ., 871, 162123, 10.1016/j.scitotenv.2023.162123, 2023.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, R. e.: IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2021.

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 aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo, Science Advances, 1, e1500157, 10.1126/sciadv.1500157, 2015.

McNabb, B. J., and Tortell, P. D.: Improved prediction of dimethyl sulfide (DMS) distributions in the northeast subarctic Pacific using machine-learning algorithms, Biogeosciences, 19, 1705-1721, 10.5194/bg-19-1705-2022, 2022.

McNabb, B. J., and Tortell, P. D.: Oceanographic controls on Southern Ocean dimethyl sulfide distributions revealed by machine learning algorithms, Limnology and Oceanography, 68, 616-630, 10.1002/lno.12298, 2023.

McParland, E. L., and Levine, N. M.: The role of differential DMSP production and community composition in predicting variability of global surface DMSP concentrations, Limnol. Oceanogr., 64, 757-773, 10.1002/lno.11076, 2018.

Moradkhani, H., DeChant, C. M., and Sorooshian, S.: Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, Water Resources Research, 48, 10.1029/2012wr012144, 2012.

Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., Boutin, J., and Upstill-Goddard, R. C.: In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers, Glob. Biogeochem. Cycles, 14, 373-387, 10.1029/1999gb900091, 2000.

Novak, G. A., Fite, C. H., Holmes, C. D., Veres, P. R., Neuman, J. A., Faloona, I., Thornton, J. A., Wolfe, G. M., Vermeuel, M. P., Jernigan, C. M., Peischl, J., Ryerson, T. B., Thompson, C. R., Bourgeois, I., Warneke, C., Gkatzelis, G. I., Coggon, M. 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, E. L., Robinson, C., Thornhill, K. L., Sanchez, K. J., Hall, S. R., Ullmann, K., Dollner, M., Weinzierl, B., Blake, D. R., and Bertram, T. H.: Rapid cloud removal of dimethyl sulfide oxidation products limits SO2 and cloud condensation nuclei production in the marine atmosphere, P. Natl. Acad. Sci. USA, 118, e2110472118, 10.1073/pnas.2110472118, 2021.

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

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

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 Arctic Ocean and Its Relation to Phytoplankton Biomass, Glob. Biogeochem. Cycles, 32, 351-359, 10.1002/2017gb005805, 2018.

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, R., Becagli, S., and Hermansen, O.: Dimethyl Sulfide-Induced Increase in Cloud Condensation Nuclei in the Arctic Atmosphere, Glob. Biogeochem. Cycles, 35, e2021GB006969, 10.1029/2021gb006969, 2021.

Qu, B., Gabric, A. J., Zeng, M., and Lu, Z.: Dimethylsulfide model calibration in the Barents Sea using a genetic algorithm and neural network, Environ. Chem., 13, 413-424, 10.1071/EN14264, 2016.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Wood, R., Stemmler, J. D., Rémillard, J., and Jefferson, A.: Low-CCN concentration air masses over the eastern North Atlantic: Seasonality, meteorology, and drivers, J. Geophys. Res. Atmos., 122, 1203-1223, 10.1002/2016jd025557, 2017.

Woolf, D. K.: Bubbles and their role in gas exchange, in: The Sea Surface and Global Change, edited by: Liss, P. S., and Duce, R. A., Cambridge University Press, Cambridge, 173-206, 1997.

Yang, B., Boss, E. S., Haëntjens, N., Long, M. C., Behrenfeld, M. J., Eveleth, R., and Doney, S. C.: Phytoplankton Phenology in the North Atlantic: Insights From Profiling Float Measurements, Front. Mar. Sci., 7, 10.3389/fmars.2020.00139, 2020.

Yu, L., and Zhou, N.: Survey of imbalanced data methodologies, arXiv preprint arXiv:2104.02240, 2021.

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 degradation to DMS-their roles in the global sulfur cycle, Sci. China Life Sci., 62, 1296-1319, 10.1007/s11427-018-9524-y, 2019.

Zhao, J., Ma, W., Bilsback, K. R., Pierce, J. R., Zhou, S., Chen, Y., Yang, G., and Zhang, Y.: Simulating the radiative forcing of oceanic dimethylsulfide (DMS) in Asia based on machine learning estimates, Atmos. Chem. Phys., 22, 9583-9600, 10.5194/acp-22-9583-2022, 2022.

Zheng, G., Li, X., Zhang, R. H., and Liu, B.: Purely satellite data-driven deep learning forecast of complicated tropical instability waves, Science Advances, 6, eaba1482, 10.1126/sciadv.aba1482, 2020.

Zhou, S.: An artificial neural network ensemble model for sea surface DMS simulation, v3.0, Zenodo [Data set], https://doi.org/10.5281/zenodo.12398985, 2024

Zhou, S., Chen, Y., Huang, S., Gong, X., Yang, G., Zhang, H., Herrmann, H., Wiedensohler, A., Poulain, L., Zhang, Y., Wang, F., Xu, Z., and Yan, K.: A 20-year (1998-2017) global sea surface dimethyl sulfide gridded dataset with daily resolution, v4.0, Zenodo [Data set], https://doi.org/10.5281/zenodo.11879900, 2024.

Zhou, S., Chen, Y., Paytan, A., Li, H., Wang, F., Zhu, Y., Yang, T., Zhang, Y., and Zhang, R.: Non-Marine Sources Contribute to Aerosol Methanesulfonate Over Coastal Seas, J. Geophys. Res.-Atmos., 126, e2021JD034960, 10.1029/2021jd034960, 2021.

Zindler, C., Bracher, A., Marandino, C. A., Taylor, B., Torrecilla, E., Kock, A., and Bange, H. W.: Sulphur compounds, methane, and phytoplankton: interactions along a north–south transit in the western Pacific Ocean, Biogeosciences, 10, 3297-3311, 10.5194/bg-10-3297-2013, 2013.

Zindler, C., Marandino, C. A., Bange, H. W., Schütte, F., and Saltzman, E. S.: Nutrient availability determines dimethyl sulfide and isoprene distribution in the eastern Atlantic Ocean, Geophys. Res. Lett., 41, 3181-3188, 10.1002/2014gl059547, 2014.