



## Diagnosing sea-surface dimethylsulfide (DMS) concentration from satellite data at global and regional scales

Martí Galí<sup>1</sup>, Maurice Levasseur<sup>1</sup>, Emmanuel Devred<sup>2</sup>, Rafel Simó<sup>3</sup>, Marcel Babin<sup>1</sup>

5 <sup>1</sup> Takuvik Joint International Laboratory (Université Laval – CNRS). Biology Department, Université Laval. 1045 Avenue de la Médecine G1V 0A6 Québec (QC) Canada

<sup>2</sup> Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS B2Y 4A2, Canada

<sup>3</sup> Institut de Ciències del Mar (ICM-CSIC). Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain

10 *Correspondence to:* Martí Galí ([marti.gali.tapias@gmail.com](mailto:marti.gali.tapias@gmail.com))

**Abstract.** The marine biogenic gas dimethylsulfide (DMS) can modulate regional and global climate by enhancing aerosol light scattering and seeding cloud formation. However, the lack of time- and space-resolved estimates of DMS concentration and emission hampers the assessment of its climatic effects. Here we present DMS<sub>SAT</sub>, a new remote sensing algorithm that relies on the nonlinear relationship between DMS, its phytoplanktonic precursor dimethylsulfoniopropionate (DMSPt) and plankton light exposure. The DMS<sub>SAT</sub> algorithm is computationally light and can be easily optimized for global and regional scales. At the global scale, it reproduces the main climatological features of DMS seasonality across contrasting biomes with remarkable skill compared to previous algorithms. Shortcomings of the global-scale optimized algorithm are the propagation of regional biases in remotely sensed chlorophyll (causing underestimation of DMS in the Southern Ocean) and the inability to reproduce high DMS/DMSPt ratios in late summer and fall in specific regions (which suggests the need to account for additional DMS drivers). Our work also highlights the shortcomings of interpolated DMS climatologies, caused by sparse and biased in situ sampling. Time series of DMS<sub>SAT</sub> between 2003-2016 in northern subpolar regions show wide interannual variability in the magnitude and timing of the annual DMS peak(s), demonstrating the need to move beyond the climatological view in studies of ocean-atmosphere interactions. By providing time- and space-resolved estimates of DMS emission, DMS<sub>SAT</sub> can leverage atmospheric chemistry and climate models and advance our understanding of plankton-aerosol-cloud interactions in the context of global change.

### 1 Introduction

Ocean-emitted gases and particles control the number, size distribution and composition of aerosols in remote oceanic areas. These aerosols scatter sunlight and can act as cloud condensation nuclei that alter the radiative properties of clouds, both microscopic (cloud droplet number concentration and effective radius) and macroscopic (cloud abundance, albedo and lifetime). Interactions between natural aerosols and clouds are a major source of uncertainty in climate projections, confounding the calculation of natural and anthropogenic radiative forcing and the attribution of anthropogenic climate



change (Carlsaw et al., 2013). Therefore, there is an urgent need to better understand and model the oceanic sources of aerosols, and to better resolve their variations at relevant spatial and temporal scales, from weekly through seasonal and interannual.

5 The gas dimethylsulfide (DMS) is produced by marine microbial food webs in the sunlit layer of the ocean. With its emission currently estimated at 28 Tg S y<sup>-1</sup>, it contributes about 70% of natural sulfur emissions to the global atmosphere and a major portion of the marine emission of organic volatiles (Carpenter et al., 2012; Schlesinger and Bernhardt, 2013; Simó, 2011). The cloud-seeding activity of DMS and its potential role in climate regulation were first postulated three decades ago (Charlson et al., 1987; Shaw, 1983). The so-called CLAW hypothesis (Charlson et al., 1987) proposed that a  
10 negative feedback could operate between marine phytoplankton, DMS emission and cloud albedo, potentially regulating the Earth's climate. Posterior research showed that the mechanisms behind the potential loop are far more complex than initially envisaged. This, and the estimated low sensitivity of each step of the feedback to changes in its forcing factors, led (Quinn and Bates, 2011) to refute the CLAW hypothesis. Nevertheless, recent atmospheric observations –enhanced by new analytical techniques (Kulmala et al., 2014)– and modeling studies have shown instances where marine DMS controls sub-  
15 micrometer aerosol formation in the Arctic (Leaitch et al., 2013), Antarctica (Yu and Luo, 2010) and the tropical South Pacific atmospheres (Modini et al., 2009). Moreover, Quinn et al. (2017) recently reported that non-sea-salt sulfate aerosols, derived from DMS, dominate cloud condensation nuclei populations over most of the global ocean. As a result, the occurrence of a "seasonal CLAW" in remote marine atmospheres is becoming increasingly conceivable (Levasseur, 2013; Vallina and Simó, 2007a).

20

DMS is produced by marine microbial food webs through a complex network of biological interactions and chemical processes (Simó, 2004). Its primary source is the enzymatic and photochemical breakdown of dimethylsulfoniopropionate (DMSP), a multifunctional osmolyte that accumulates at high (mM) intracellular concentrations in some phytoplankton, especially haptophytes, dinoflagellates and some picoeukaryotes (Stefels et al., 2007). DMSP cleavage is catalyzed by a  
25 wide diversity of enzymes, called DMSP lyases, produced by some eukaryotic phytoplankton (Alcolombri et al., 2015) and bacteria (Curson et al., 2011). Breakage of phytoplankton cells through zooplankton grazing, viral attack and autolysis releases DMSP to the algal boundary layer and the dissolved phase and enhances DMS production (Simó, 2004; Stefels et al., 2007). Another process that contributes to DMS production is the diffusive release of DMS from phytoplankton cells, which proceeds almost instantaneously after intracellular DMSP cleavage by DMSP lyases or by photochemically produced  
30 radicals (Lavoie et al., 2015; Spiess et al., 2015). DMS budgets in the upper mixed layer (UML) indicate that, on average, about 90% of dissolved DMS is consumed by bacterial oxidation and UV-driven photolysis, and only 10% is emitted to the atmosphere through turbulent diffusion. The turnover time of DMS is typically faster than 4 days, such that DMS concentration reflects a subtle dynamic equilibrium between production and consumption processes (Galí and Simó, 2015).



Seawater DMS concentration controls the emission flux because the oceanic UML is supersaturated with respect to the atmosphere. This concentration varies over the seasonal cycle in response to the phenology and ecological succession of microbial species and their interplay with physical forcing factors, particularly irradiance and nutrient supply, which are in turn regulated by vertical mixing (Galí and Simó, 2015; Lizotte et al., 2012). Diatom-dominated phytoplankton blooms, typical of nutrient replete conditions at high latitudes, are characterized by low DMSP concentration per unit biomass and low DMS yield. Therefore, extremely high biomass is required for diatom blooms to be significant DMS sources. The opposite is true for microbial communities typical of stratified, nutrient depleted and highly irradiated surface waters, both at low and high latitudes (Galí and Simó, 2010; Lizotte et al., 2012). Under these conditions, two main factors act synergistically to increase DMS concentration (Galí and Simó, 2015; Vallina et al., 2008), namely: the higher contribution of DMSP-rich species to total phytoplankton biomass (Galí et al., 2015; Stefels et al., 2007); and the higher DMSP-to-DMS conversion yield at the microbial community level, possibly caused by the effects of nutrient and irradiance stress on phytoplankton DMS release (Galí et al., 2013; Stefels, 2000; Sunda et al., 2002, 2007; Vallina et al., 2008). The temporal decoupling between phytoplankton biomass, DMSP concentration and DMS concentration, termed the DMS summer paradox (Simó and Pedrós-Alió, 1999), is an essential feature that biogeochemical models strive to reproduce with mixed success (Le Clainche et al., 2010).

With nearly 50,000 DMS measurements taken between 1972 and 2010, the global sea-surface DMS database (<https://saga.pmel.noaa.gov/dms/>) is a valuable resource for model development and validation. Gridded monthly climatologies (Kettle et al., 1999; Lana et al., 2011) calculated from this dataset are the standard DMS product used as input to atmospheric chemistry and climate models, therefore emphasizing the seasonal climatological view (Mahajan et al., 2015; McCoy et al., 2015). At the other end, the climatic role of DMS is often evaluated through climate change projections and extreme sensitivity tests (Grandey and Wang, 2015). In comparison, contemporaneous decadal scale DMS variability has received less attention. This gap can be filled using empirical remote sensing algorithms, a handful of which have been developed since the early 2000s (Tesdal et al., 2015) after some earlier prospective studies (Jodwalis and Benner, 1995; Thompson et al., 1990). Interestingly, data based DMS climatologies and those derived from empirical algorithms or prognostic biogeochemical models exhibit large discrepancies (Tesdal et al., 2015). Although it is tempting to attribute them to the poor predictive skill of the models, discrepancies may also stem from issues in the calculation of the climatology, related to data paucity and the use of some interpolation and extrapolation procedures.

Here we present a new empirical remote sensing algorithm for DMS that proceeds in two steps: (i) estimation of the concentration of the phytoplanktonic DMS precursor, total dimethylsulfoniopropionate (DMSPt), from remotely sensed chlorophyll and light penetration, and from climatological mixed layer depth (MLD); (ii) estimation of DMS concentration from DMSPt and solar irradiance. Since the DMSPt sub-algorithm was described by Galí et al. (2015), here we focus on the relationship between DMS, DMSPt and photosynthetically available radiation (PAR) at the sea surface. We implement our



algorithm to produce a global DMS climatology, which we compare to the current DMS climatology (Lana et al., 2011) and to those derived from other remote sensing algorithms that follow similar rationales (Simó and Dachs, 2002; Vallina and Simó, 2007b). Finally, we implement our algorithm using 14 years of MODIS-Aqua satellite data in the subtropical and the subpolar North Atlantic and in the Northeast Pacific to illustrate and understand interannual DMS variability.

## 5 2 Methods

### 2.1 Datasets used for algorithm development and validation

In situ concentrations of DMS, DMSPt and chlorophyll a (Chl), accompanied by ancillary data (bottom depth, temperature, salinity, wind speed), were downloaded from the global sea-surface DMS database. The latter was complemented with additional datasets recently obtained by the authors' teams. After quality control, the database had 41304, 3700 and 9182 measurements for DMS, DMSPt and Chl, respectively, with 3637 DMS-DMSPt and 8141 DMS-Chl pairs. The in situ database was extended with geophysical and biogeochemical parameters, including satellite matchup data and climatological data following Galí et al. (2015). Detailed information regarding data sources, quality control and processing can be found in the SI and in Tables S1-S3.

We performed satellite matchups using SeaWiFS (1997-2010) and MODIS-Aqua (2003-2012) retrievals of remotely sensed Chl ( $\text{mg m}^{-3}$ , equivalent to  $\mu\text{g L}^{-1}$ ), vertical attenuation coefficient at 490 nm ( $K_{d490}$ ,  $\text{m}^{-1}$ ), particulate inorganic carbon (PIC,  $\text{mol m}^{-3}$ ) and daily photosynthetically available radiation at the sea surface (PAR,  $\text{mol photons m}^{-2} \text{d}^{-1}$ ). To maximize the amount of available matchups, we merged daily and 8-day data from the SeaWiFS (1997-2010) and MODIS-Aqua (2003-2016) following a hierarchical search procedure. These merged satellite variables are hereafter designated with the SAT subscript (e.g.  $\text{Chl}_{\text{SAT}}$ ). Daily and 8-day sea surface temperature ( $\text{SST}_{\text{SAT}}$ ,  $^{\circ}\text{C}$ ) from the AVHRR sensors was also matched to the database.

The database was further extended with monthly climatological data: daily PAR from SeaWiFS (1997-2010 average); mixed layer depth (MLD, m) from the monthly MIMOC climatology (Schmidtke et al., 2013); bottom depth from the General Bathymetric Chart of the Oceans (GEBCO08); and sea-surface nitrate and phosphate concentrations ( $\mu\text{M}$ ) from the World Ocean Atlas 2009 (WOA09). Nutricline depths were calculated from WOA09 vertical profiles as the depth where nitrate and phosphate first exceeded  $1 \mu\text{M}$  and  $0.4 \mu\text{M}$ , respectively. Nutricline depth estimations were robust to changes of  $\pm 50\%$  in these concentration thresholds.

The mean daily PAR in the upper mixed layer ( $\text{PAR}_{\text{MLD}}$ ) was calculated as:

$$\text{PAR}_{\text{MLD}} = [\text{PAR}_{\text{SAT}} / (K_{d490_{\text{SAT}}} \text{MLD})] [1 - \exp(-K_{d490_{\text{SAT}}} \text{MLD})] \quad \text{eq. 1}$$



When satellite matchups were not available (before September 1997), we used climatological PAR from SeaWiFS (1997-2010 average) in order to increase the temporal coverage of the PAR<sub>SAT</sub> and PAR<sub>MLD</sub> variables. Statistical analyses done with climatological or matchup PAR<sub>SAT</sub> gave very similar results. This procedure was not followed with other variables (Chl, PIC, Kd490) that show wider interannual variations.

## 2.2 Statistical analyses and data binning schemes

All statistical analyses were conducted using (i) non-binned data; (ii) data binned by month and 5°x5° latitude-longitude bins (M5x5); and (iii) data binned by month and the 56 Longhurst biogeochemical provinces (MLongh) (Longhurst, 2010). MLongh binned data were further aggregated into six biomes: two Polar biomes (Arctic and Antarctic), two mid-latitude Westerlies biomes (Northern and Southern hemispheres) one Trades biome (tropical latitudes), and one global coastal biome (Fig. 1C). Variables with a right-skewed, approximate lognormal distribution, entered statistical analyses after log<sub>10</sub> transformation: DMS, DMSpt, DMS/DMSpt ratio, Chl, nitrate and phosphate concentrations. We conducted statistical explorations using both bin means and bin medians, generally obtaining similar results.

To develop the DMS algorithm we analyzed the relationship between DMS, the DMS/DMSpt ratio and environmental variables from the extended database. The exploratory analysis included the calculation of pairwise correlation coefficients, followed by stepwise multiple regression. Pearson's linear correlation coefficients calculated on log<sub>10</sub>-transformed variables were higher than those calculated on the same non-transformed variables, and similar in magnitude to Spearman's rank correlations (which are independent of monotonic transformations). Thus, we report only Pearson's linear correlation coefficients (r).

Based on the correlation analysis, we built several regression models where DMS was predicted as a function of in situ DMSpt concentration and one or more additional variables (Table 1). Significant terms were selected using stepwise regression with entrance and removal p-values set at 0.001 and 0.005, respectively. The logic for adding one variable at a time, rather than building a single initial model with all the predictor variables, is that the data matrix is incomplete, such that the size of data subset used for model fitting decreases rapidly when variables with sparse coverage are combined. New predictors were added in order of decreasing data availability, and each set of initial predictors was tested across the three degrees of data binning described above and three degrees of model complexity: linear without interactions, linear with interactions, and quadratic with interactions. This 3x3-nested structure provided a stringent test for the robustness of a given regression model. Improvements in model performance across different levels of data binning and model complexity were assessed based on the increase in adjusted r-square, R<sup>2</sup><sub>adj</sub>, and the decrease in root-mean-square error (RMSE) and the Akaike Information Criterion (AIC).



Predictive models were further optimized for global and regional domains using the bootstrap method followed by nonlinear optimization as described in SI section 4. Selected models were then validated using an independent dataset composed of in situ DMS measurements and satellite matchups (described in section 3.1.3) using a wide array of skill metrics:  $R^2$ , RMSE, the mean absolute percentage error (MAPE), the slope of a major axis (type II) linear regression between observed and predicted fields ( $Slope_{MA}$ ) and the percentage bias. All analyses were carried out using Matlab R2013b.

### 2.3 Algorithm implementation

The newly developed  $DMS_{SAT}$  algorithm (Fig. 2) was implemented to produce (i) a monthly global DMS climatology and (ii) several regional time series with 8-day resolution for the period 2003–2016. Further details and data sources can be found in SI section 5 and Table S2.

Global  $DMS_{SAT}$  fields were computed using ocean color data from SeaWiFS (1997–2010 monthly climatology,  $1/12^\circ$  grid), SST from AVHRR and the MIMOC monthly MLD climatology. We established a reference  $DMS_{SAT}$  run where Chl was computed with a band-ratio algorithm (OC4-OCI standard NASA algorithm) and the euphotic layer depth ( $Ze_{SAT}$ ) was computed as the 1% penetration depth of 490 nm radiation ( $Ze_{SAT} = 4.6/Kd_{490}$ ). The impact of this choice was evaluated with sensitivity tests where  $Chl_{SAT}$  and  $Ze_{SAT}$  were calculated with the semi-analytical algorithms of (Maritorena et al., 2002) and (Lee et al., 2007), respectively, which are more appropriate in optically complex waters. Note also that, since climatological  $Chl_{SAT}$  does not have observation gaps, the equation that estimates  $DMSP_{tSAT}$  from  $PIC_{SAT}$  is not used (Galí et al., 2015). Global monthly  $DMS_{SAT}$  fields were averaged onto  $1^\circ$  and  $5^\circ$  grids for mapping and comparison to other DMS climatologies: the interpolated L11 climatology (Lana et al., 2011), and the climatologies derived with the empirical algorithms of (Simó and Dachs, 2002) (SD02) and (Vallina and Simó, 2007) (VS07). The procedure used to produce the L11 climatology and the bases of the SD02 and VS07 empirical algorithms are briefly described in section 3.2.

Regional  $DMS_{SAT}$  time series between 2003 and 2016 were computed using daily MODIS-Aqua data (4.64 km) combined with the MIMOC MLD climatology. As done for the global implementation, we produced  $DMS_{SAT}$  fields using both band-ratio and semi-analytical Chl products. We also performed a test comparing  $DMSP_{tSAT}$  obtained with the MLD climatology vs. model-derived MLD time series, showing little  $DMSP_{tSAT}$  sensitivity (Fig. S1). Since non-climatological satellite data contain gaps caused by cloudiness, we applied a binning and gap-filling procedure to obtain full coverage, such that the final regional time series had a resolution of 8 days and 27.8 km. We produced  $DMS_{SAT}$  time series for the Bermuda Atlantic Time Series site (BATS;  $31^\circ 40'N$ ,  $64^\circ 10'W$ ) and for the entire northern hemisphere at latitudes  $>45^\circ N$ . The latter dataset was then sampled at selected North Atlantic sites and at the Ocean Station P (OSP) in the NE Pacific ( $50^\circ N$ ,  $145^\circ W$ ). Satellite time series were compared to the L11 climatology and to in situ DMS (and  $DMSP_t$ , if available). These in situ data, kindly



provided by the BATS (Levine et al., 2016) and OSP (<https://www.waterproperties.ca/linep/>) teams, were not used in algorithm development.

### 3 Results

#### 3.1 Development and validation of the DMS sub-algorithm

##### 5 3.1.1 Statistical exploration

We analyzed the correlation between potential predictor variables and  $\log_{10}(\text{DMS})$  or  $\log_{10}(\text{DMS}/\text{DMSpt})$  (Table 1). This analysis systematically showed that (i) DMSpt was the best correlate of DMS ( $r = 0.46$  to  $0.65$ ), and (ii) surface  $\text{PAR}_{\text{SAT}}$  or mean PAR in the upper mixed layer ( $\text{PAR}_{\text{MLD}}$ ) were the best correlates of the DMS/DMSpt ratio ( $r = 0.35$  to  $0.67$ ). These correlation patterns remained across different binning levels, suggesting that DMS can be predicted, to first order, by the concentration of its phytoplanktonic precursor compound and by the PAR-dependent enhancement of DMSpt-to-DMS conversion. It is also noteworthy that the correlation between day length and the DMS/DMSpt ratio was weak or non-significant. This supports the causal relationship between PAR and the DMS/DMSpt ratio and discards other factors that might follow synchronous seasonal cycles.

15 Guided by the correlation patterns, we established a base regression model expressed by the equation:

$$\log_{10}\text{DMS} = \alpha + \beta \log_{10}\text{DMSpt} + \gamma \text{PAR} \quad \text{eq. 2}$$

This model explained between 50% and 57% of  $\log_{10}(\text{DMS})$  variance with an increasing level of data binning, and the corresponding RMSE ranged between 0.35 and 0.21 (Table 2).

We assessed whether the base model could be significantly improved by adding one new variable at a time and/or increasing model complexity. We started by adding a variable X to a linear model without interactions of the form  $\log_{10}\text{DMS} = \alpha + \beta \log_{10}\text{DMSpt} + \gamma \text{PAR} + \delta X$ . X was chosen among the variables showing higher correlations to either DMS or DMS/DMSpt (Table 1): SST, nitrate concentration, nitracline depth, salinity, wind speed and  $\text{PIC}_{\text{SAT}}$ . With non-binned data all these variables entered regression models with significant coefficients, but only salinity, wind speed and  $\text{PIC}_{\text{SAT}}$  produced significant decreases in RMSE and AIC. With MLongh binned data, only SST and  $\text{PIC}_{\text{SAT}}$  entered with significant coefficients. Yet, none of the additional variables improved simultaneously the  $R^2_{\text{adj}}$ , RMSE and AIC skill metrics with respect to the base model. Increasing model complexity through addition of interaction and quadratic terms, or by adding a fourth predictor variable, generally resulted in minor improvements or erratic changes in model performance (results not



shown). Invariably, DMSpt and PAR were the only variables with highly significant coefficients ( $p \ll 10^{-10}$ ) regardless of the binning scheme and the additional predictors included (Table S4).

As a corollary, the use of  $PAR_{ML}$  instead of  $PAR_{SAT}$  slightly degraded the predictive power ( $R^2_{adj} = 0.47$  and  $RMSE = 0.37$  for non-binned data; Table S3). Although  $PAR_{MLD}$  is a priori a more realistic metric of light exposure, it is possible that the use of climatological MLD degraded the  $PAR_{MLD}$  estimations. Another potential explanation is the episodic nature of oceanic vertical mixing, which requires the distinction between the actively mixing layer—defined by turbulence thresholds—and the mixed layer—as detected by regular temperature-salinity profiles (Sutherland et al., 2014). This might imply that, on occasions, mean light exposure at the sea surface is better approximated by surface  $PAR_{SAT}$  than by  $PAR_{MLD}$ . Finally, we cannot discard that the nonlinear relationship between DMS, DMSpt and PAR embodied in eq. 2 (Fig. 2) is implicitly accounting for the effects of vertical mixing and water clarity on plankton light exposure. After these considerations we discarded the use of  $PAR_{MLD}$  as a predictor in our algorithm, and focused on optimizing eq. 2.

### 3.1.2 Implications of the model structure

Here we analyze the physical meaning of eq. 2 coefficients in view of their optimization for predictive purposes (3.1.3). First, it must be noted that the  $\log_{10}DMSpt$  coefficient ( $\beta$ ) is smaller than 1 regardless of the binning applied (Table 2). For a constant PAR, this implies that DMS increases more slowly than DMSpt (Fig. 3A) and that the DMS/DMSpt ratio decreases non-linearly with increasing DMSpt (Fig. 3B). In biogeochemical terms, this implies that the PAR-driven increase of the DMS/DMSpt ratio is stronger in DMSpt-poor conditions. In biogeographic terms (Fig. 1), highest DMS/DMSpt ratios are found in oligotrophic areas of the Trades biome, where low DMSpt concentrations prevail (<20 nM). Low DMSpt concentrations are also found in winter at high latitudes in deeply mixed waters, but the corresponding low irradiance results in  $DMS/DMSpt < 0.05$ . At the high DMSpt concentrations that occur at high latitudes in summer (>100 nM), the DMS/DMSpt ratio is generally <0.1.

Second, we note that the y-intercept ( $\alpha$ ), the  $\log_{10}DMSpt$  coefficient ( $\beta$ ) and the PAR coefficient ( $\gamma$ ) vary in a consistent manner as the binning spatial scale increases (Table 2). To further explore the interrelationship between the model coefficients, we used the bootstrap method to produce  $10^5$  sets of regression coefficients for the MLongh dataset. The scatterplots between  $\alpha$ ,  $\beta$  and  $\gamma$  resulting from the  $10^5$  bootstrapped regressions confirm that covariation between the coefficients is non-random (Fig. S2), and suggest that there are two main avenues for optimizing eq. 2: (i) increasing (decreasing) the PAR coefficient  $\gamma$ , which increases (decreases) the DMS/DMSpt ratio in proportion to PAR regardless of DMSpt concentration; (ii) increasing (decreasing) the DMSpt coefficient, which increases (decreases) DMS more strongly in DMSpt-rich conditions. The intercept ( $\alpha$ ) acts to adjust the magnitude of DMS concentrations by a fixed proportion everywhere. These trade-offs should be kept in mind when optimizing our model for global or regional implementation.



### 3.1.3 Optimization and validation

By definition, least squares regression minimizes the RMSE, but it has been shown that regression models derived in this way do not necessarily have the best predictive skill (Jolliff et al., 2009). Therefore, we devised an alternative nonlinear optimization procedure (SI section 4). To obtain realistic solutions, we constrained the optimized coefficients to the confidence intervals derived from the bootstrapped regressions. The resulting optimal model had higher DMSpt ( $\beta$ ) and PAR ( $\gamma$ ) coefficients and a smaller y-intercept (eq. 2f; Fig. S2), and moved the modeled DMS concentration closer to the 1:1 agreement line without degrading neither RMSE nor  $R^2$  (Table 2).

We validated the different versions of eq. 2 (Table 2) by comparing  $DMS_{SAT}$  against in situ DMS using an independent subset of the database. Since the complete DMS algorithm proceeds in two steps (Fig. 2), its validation must take into account uncertainty in variables used as input to the  $DMSPt_{SAT}$  sub-algorithm. Galí et al. (2015) showed that, apart from the inherent algorithm uncertainty, most uncertainty in  $DMSPt_{SAT}$  ( $RMSE \leq 0.3$  in  $\log_{10}$  space) results from error in  $Chl_{SAT}$ . Thus, the validation subset was defined according to three criteria: (i) satellite match-up data used as input to the algorithm ( $Chl_{SAT}$ ,  $K_{d,490}$ ,  $PAR_{SAT}$  and  $SST_{SAT}$ ) were available (see Fig. 2); (ii) in situ DMSpt was not available —thus excluding the data used for model fitting; (iii) in situ DMS and Chl were available.

We used in situ Chl concentration to constrain the uncertainty in  $Chl_{SAT}$  used as input to the  $DMSPt_{SAT}$  sub-algorithm. Indeed, this procedure progressively reduced the size of the validation subset as the maximum tolerated  $Chl_{SAT}$  error decreased (Fig. 4). Uncertainty arising from  $PAR_{SAT}$  could not be assessed because the current database lacks in situ PAR measurements. Frouin et al. (2003) reported an error of  $\pm 15\%$  ( $< 10\%$  for weekly and monthly periods), with negligible bias for  $PAR_{SAT}$ , suggesting it is a minor source of uncertainty. Fig. 4A summarizes the validation results for the regression model based on MLongh binning (eq. 2e) and the optimized model (eq. 2g). Supporting our assumption, the skill metrics of the  $DMS_{SAT}$  algorithm improved as  $Chl_{SAT}$  RMSE decreased (Fig. 4). Other skill metrics (not shown in Fig. 4) showed comparable trends.

The optimized model coefficients (eq. 2f) reduced the RMSE and increased  $R^2$  with respect to the regression-derived coefficients, achieving a maximal  $R^2$  of 0.52 and minimal RMSE of 0.21 for error-free  $Chl_{SAT}$  (non-binned data; Table S5). The global-scale optimized  $DMS_{SAT}$  had a normalized standard deviation of 1.1 ( $\log_{10}$  space), meaning that the spread of modeled DMS nearly equals that of in situ DMS concentrations. In linear space,  $R^2$  increased from 0.09 to 0.21 and RMSE decreased from 8.5 to 2.8 nM as  $Chl_{SAT}$  error decreased. These linear-space statistics might be interpreted as a sign of poor performance, but it should be noted that they were strongly affected by a small fraction of highly biased estimations. Removing the most biased estimations (the  $< 8\%$  beyond a factor of 3 from real measurements; Fig. 4) increased the linear-space  $R^2$  to 0.42–0.59 and decreased the RMSE to 1.0–2.4 nM across the full range of  $Chl_{SAT}$  error, with MAPE of 30–42%



and relative bias of -3% to 9%. These statistics illustrate the good performance of the algorithm and highlight the better robustness of log-space statistics.

### 3.2 Global climatologies

After verifying the good performance of the algorithm, we implemented it using the global climatology from the SeaWiFS sensor. Maps of the mean seasonal  $DMS_{SAT}$  concentration and the corresponding zonal averages (latitudinal profiles) are displayed in Fig. 5. In each hemisphere, DMS concentrations around ~2.5 nM prevail during the astronomic spring and summer, decreasing to around 1 nM in fall and <1 nM in winter. The seasonal cycle has wider amplitude at high latitudes and is nearly flat in the tropical oceans (Fig. 6 and 7). Regional enhancement of DMS concentrations occurs in some coastal and shelf areas, equatorial and eastern boundary upwellings, close to the subtropical front in austral summer (40°S), and in the subpolar North Atlantic in boreal summer (60°N). The global mean area-weighted  $DMS_{SAT}$  concentration is 1.63 nM. This figure decreases by less than 5% when semi-analytical  $Chl_{SAT}$  and  $Zeu_{SAT}$  products are used instead of our reference products (Table 3).

#### 3.2.1 Comparison to the L11 climatology

The L11 DMS climatology (Lana et al., 2011), as well as prior climatologies (Kettle et al., 1999; Kettle and Andreae, 2000) was calculated using an objective interpolation procedure. An initial template, called first-guess field, was obtained by calculating the monthly mean DMS in each Longhurst province. The gaps were filled through temporal interpolation and, in provinces with too few documented months, the seasonal cycle was extrapolated by scaling that of neighbor provinces. Objective interpolation was then applied by searching measurements within a 555 km radius, weighting them inversely to the distance from a given grid point, and the resulting global fields were repeatedly smoothed.

The global mean area-weighted  $DMS_{L11}$  concentration is 2.44 nM, 1.5-fold higher than  $DMS_{SAT}$  (Table 3). As shown in Fig. 5-7, the disagreement between the  $DMS_{L11}$  and  $DMS_{SAT}$  climatologies varies depending on the regions and the spatial-temporal scales compared. Starting with the coarsest scale, we observe that the seasonal latitudinal profiles (zonal means) of  $DMS_{L11}$  and  $DMS_{SAT}$  have very similar shapes. This can be easily emphasized by multiplying  $DMS_{SAT}$  by 1.5 to "correct" the 50% offset (Fig. 5) (or, what is the same, increasing  $\alpha$  by  $\log_{10}(1.5)$  in eq. 2). The best agreement in the latitudinal profiles is observed in June through August, whereas the strongest disagreement occurs polewards of 50°S during the austral summer, when  $DMS_{SAT}$  is lower than  $DMS_{L11}$  by at least twofold. Comparison of the  $DMS_{SAT}$  and  $DMS_{L11}$  climatologies by means of Hovmöller diagrams (Fig. 7) shows a remarkable qualitative agreement in their month-latitude patterns, except for the polar austral summer. Figs. 6-7 also reveal smaller disagreements in the Arctic Ocean in winter-spring and in the equatorial band during most of the year, with lower concentrations in  $DMS_{SAT}$  in both cases. The most striking disagreement appears when  $DMS_{SAT}$  and  $DMS_{L11}$  are compared by means of seasonal anomaly maps (Fig. 5). The sign of the  $DMS_{SAT}$ -



DMS<sub>L11</sub> anomaly changes from positive to negative in a patchy pattern, often following the boundaries of the Longhurst biogeochemical provinces.

### 3.2.2 Comparison to the SD02 climatology

The SD02 algorithm (Simó and Dachs, 2002) was designed to predict DMS from MLD and Chl<sub>SAT</sub> using two different equations depending on the Chl/MLD ratio:

$$\text{DMS} = -\ln(\text{MLD}) + 5.7 \quad \text{Chl/MLD} < 0.02 \quad (\text{eq. 3a})$$

$$\text{DMS} = 55.8 \text{ Chl/MLD} + 0.6 \quad \text{Chl/MLD} \geq 0.02 \quad (\text{eq. 3b})$$

such that DMS increases linearly with the Chl/MLD ratio in stratified productive conditions (e.g. high latitudes in summer) and inversely with MLD in typical oligotrophic conditions. Validation of SD02 with the same dataset used for DMS<sub>SAT</sub> indicates that it explains less variance ( $\log_{10} R^2$  of 0.22–0.30) but has similar RMSE, MAPE and bias (Table S5). Figs. 6-7 show that the SD02 estimates are in good agreement with the L11 climatology at tropical and temperate latitudes. An exception is found in the Southern Westerlies biome, where prevailing deep vertical mixing and low Chl cause SD02 to underestimate DMS throughout the productive season. A feature of SD02 is the overestimation of DMS in high Northern latitudes through late summer and fall, caused mainly by the shallow MLD due to freshwater-driven stratification. As DMS<sub>SAT</sub>, DMS<sub>SD02</sub> suffers a strong negative bias throughout the Antarctic biome during the productive season (November through February).

### 3.2.3 Comparison to the VS07 climatology

The VS07 algorithm (Vallina and Simó, 2007) relies on the observed linear relationship between DMS concentration and the solar radiation dose (SRD) in the upper mixed layer in the global ocean, according to the equation:

$$\text{DMS} = 0.492 + 0.019 \cdot \text{SRD} \quad (\text{eq. 4})$$

SRD is analogous to PAR<sub>MLD</sub> (eq. 1), but replacing PAR<sub>SAT</sub> by total shortwave irradiance ( $E_{\text{dsw}}$ ;  $\text{W m}^{-2}$ ). Here we implemented VS07 with two variations: (i) we used  $K_{\text{d}490_{\text{SAT}}}$  instead of a fixed  $K_{\text{d}}$  (note that in phytoplankton-rich and continentally-influenced waters,  $K_{\text{d}490_{\text{SAT}}}$  is generally higher than the fixed  $K_{\text{d}} = 0.06 \text{ m}^{-1}$  used by (Vallina and Simó, 2007)); (ii) we estimated  $E_{\text{dsw}}$  from PAR<sub>SAT</sub> by converting the latter to units of  $\text{W m}^{-2}$  (Morel and Smith, 1974) and then applying a constant  $E_{\text{dsw}}/\text{PAR}_{\text{SAT}}$  ratio of 1/0.43 (Kirk, 2011). VS07 shows poorer performance than DMS<sub>SAT</sub> and SD02 when validated with the same dataset (Table S5). Figs. 6-7 show that VS07 produces rather uniform DMS fields compared to the other climatologies. VS07 performs very well in the Westerlies biomes, particularly in the northern hemisphere, but invariably overestimates (underestimates) DMS in the Trades (Polar) biomes.



### 3.3 Regional DMS<sub>SAT</sub> time series

#### 3.3.1 Subpolar Atlantic and Pacific

We used MODIS-Aqua data to produce a 14-year DMS<sub>SAT</sub> time series (and the corresponding climatology) for the northern hemisphere at latitudes  $>45^\circ$  N. In this regional implementation we used a different set of coefficients, obtained from regression of  $M5 \times 5$  binned data restricted to latitudes  $>45^\circ$  N (eq. 2g). In this case, further optimization did not lead to significant improvement. We then sampled the resulting time series in some representative regions: three rectangles with an area of  $\sim 200,000$  km<sup>2</sup> each, located along the  $50^\circ\text{N}$ – $56^\circ\text{N}$  band in the North Atlantic, and the Ocean Station P (OSP,  $50^\circ$  N,  $145^\circ$  W) in the NE Pacific.

We selected the subpolar North Atlantic because it is one of the regions where the algorithm works best (Fig. S3), lending credit to observed variability patterns. Fig. 9 shows DMS seasonal cycles in three selected areas with a relatively high density of in situ DMS measurements: (a) the deep waters of the northwest Atlantic drift, (b) the shelf break west of Ireland, and (c) the shallow Southern North Sea. We observe a good agreement between the 14-year DMS<sub>SAT</sub> climatology and the L11 climatology, except in the Southern North Sea where DMS<sub>SAT</sub> is too high through summer and fall. The most salient result is however the wide interannual variability of the DMS<sub>SAT</sub> seasonal cycles. Mean DMS<sub>SAT</sub> concentration during the productive season can vary by two- to threefold between years (see variability metrics in Fig. 9), and the annual DMS<sub>SAT</sub> peak can occur within a temporal window of 2–3 months. Although years with a major peak in spring-summer are the norm, a second peak in late summer is not unusual. Satellite data also suggest wide geographic variation in the temporal lag between the annual peaks of DMS<sub>SAT</sub>, DMSPT<sub>SAT</sub> and Chl<sub>SAT</sub>, up to four months in the Southern North Sea.

The same MODIS-Aqua dataset was used to analyze the mean seasonal cycle and the interannual variability at Ocean Station P (Figure 10), where DMS has been measured two to three times per year (around February, June and August) since 1996. DMS<sub>SAT</sub> captured well in situ concentrations in February and June but suffered a low bias in August. Examination of August measurements during the 2005–2016 period suggests the existence of two regimes: 8 years have in situ DMS of  $6.6 \pm 1.1$  nM, about twice as high as DMS<sub>SAT</sub>, and 4 years have much higher in situ DMS of  $16.1 \pm 4.8$  nM, about six fold higher than DMS<sub>SAT</sub>. Local tuning of eq. 2 using OSP data cannot increase DMS in August–September without strongly degrading its performance in other months.

Note that these time series were calculated using Chl<sub>SAT</sub> derived from the semi-analytical GSM algorithm. Using the band-ratio OC3 Chl algorithm gave very similar results in the oceanic regions but 70% higher concentrations in the shallow Southern North Sea, possibly due to interference of colored dissolved organic matter and sediments on OC3 Chl retrieval (data not shown).



### 3.3.2 Bermuda Atlantic Time Series

Using the globally tuned coefficients (eq. 2f),  $DMS_{SAT}$  reproduced the shape of the mean seasonal cycle at the oligotrophic BATS station but underestimated DMS by around twofold between June and October (Fig. 10D-E). Fig. 10 F shows that part of this bias can be attributed to the low bias of  $DMS_{PtSAT}$ . However, replacing  $DMS_{PtSAT}$  by in situ  $DMS_{Pt}$  raised  $DMS_{SAT}$  by only 17%, indicating that most of the underestimation is caused by the DMS sub-algorithm. Optimizing the coefficients using local data (eq. 2h; see Appendix A) improved the model-data fit by decreasing the  $DMS_{Pt}$  coefficient (thus weakening the  $DMS_{Pt}$  dependence) and increasing the PAR coefficient. Indeed, different studies have shown that irradiance suffices to explain most of the DMS seasonal cycle at BATS (Galí and Simó, 2015; Toole and Siegel, 2004; Vallina and Simó, 2007). Interestingly, the locally tuned  $DMS_{SAT}$  is in excellent agreement with in situ data throughout 2007, but the underestimation persists in September and October of 2006 and 2008.

## 4 Discussion

Here we explore the strengths and weaknesses of our novel approach focusing on two aspects, from technical to general: the methodological and geo-statistical issues that affect DMS algorithms and climatologies (4.1), and the capacity to account for relevant biogeochemical processes and explain interannual changes using satellite data (4.2).

### 4.1 Geo-statistics, remote sensing algorithms and interpolated climatologies

Global DMS fields estimated by the L11 climatology and by the  $DMS_{SAT}$  algorithm show remarkable geographic differences (Fig. 5). Particularly, changes in the sign of the  $DMS_{SAT}-DMS_{L11}$  anomaly often follow the boundaries of the Longhurst biogeochemical provinces. In our view, the reasons for the disagreement are many fold: (1) the right-skewed distribution of DMS concentrations (Kettle et al., 1999), (2) the small amount of monthly data available in many biogeochemical provinces, (3) the absence of repeat measurements in most oceanic regions, (4) the objective interpolation procedure used to calculate the L11 climatology, (5) the fitting of the  $DMS_{SAT}$  algorithm in log space, and (6) the inherent bias of certain satellite data products in specific regions. In the next paragraphs we show how statistical shortcomings of the in situ DMS database compromise the robustness of interpolated climatologies such as L11, as evidenced by satellite matchup data and satellite-derived DMS gridded fields.

First, the sea-surface DMS database is biased towards productive conditions, which is illustrated by comparing SeaWiFS-retrieved Chl concentration in the global ocean and in the database (Fig. 8A). The SeaWiFS 1997-2010  $Chl_{SAT}$  climatology, which overlaps in time with 55% of the DMS database measurements, has a global median (mean) of 0.17 (0.53)  $\mu g L^{-1}$ . In the subset of the DMS database with available SeaWiFS match-ups,  $Chl_{SAT}$  has a median (mean) of 0.56 (1.10)  $\mu g L^{-1}$ , two or threefold higher than the global statistics. If we assume that satellite matchups represent a random sample of the DMS database, this implies a sampling bias towards high DMS concentrations, given the positive correlation between Chl and



DMS (Fig. 8B). The bias is bigger when the comparison is restricted to the spring-summer semester of each hemisphere, with a median  $\text{Chl}_{\text{SAT}}$  of 0.19 for the SeaWiFS climatology and 0.74 for the SeaWiFS database matchups. This is the period when DMS peaks and has more influence on mean annual DMS concentration.

5 Second, sampling bias is intertwined with the non-normal right-skewed statistical distribution of sea-surface DMS concentrations and the poor spatial resolution of most in situ DMS datasets. Spatial averaging, justified by data scarcity, is appropriate when applied over small or sufficiently homogeneous regions. However, when applied over a large Longhurst province affected by sampling bias, the resulting gridded fields may overestimate DMS over most of the province. This seems particularly problematic in high-latitude biomes and coastal areas with sharp productivity gradients at smaller scales,  
10 where DMS is more tightly correlated to phytoplankton biomass. In low latitude oligotrophic areas where DMS is uncoupled from Chl and has small spatial variability (Royer et al., 2015), the objective interpolation method seems appropriate. As illustrated in Fig. 6, mean DMS concentration in a given month and biome is systematically higher than the corresponding median. In most biomes,  $\text{DMS}_{\text{SAT}}$  tends to follow the monthly medians of in situ data, whereas  $\text{DMS}_{\text{L11}}$  generally follows – by construction – the monthly means. Since  $\text{DMS}_{\text{SAT}}$  has a small positive or negative bias when validated on non-binned data  
15 (Table S5), our analysis suggests that the L11 climatology and its predecessors suffer a global positive bias.

The third major issue is the scarcity of DMS measurements repeated in different years. At the MLongh binning level, 42% of the province-month bins contain measurements from a single year, 21% from two years, and 37% from more than three years. Thus, data from a single or a few years are often assumed representative of the mean ecosystem state in interpolated  
20 climatologies, which is probably not the case in regions with wide interannual variability or long-term trends (Vantrepotte and Mélin, 2011). While this does not necessarily bias global DMS fields, it can produce artificial seasonal cycles. For example, L11 suggests the existence of early spring and fall DMS peaks in the North Atlantic drift area, which result from interpolation from neighbor regions (Fig. 9A). In contrast,  $\text{DMS}_{\text{SAT}}$  suggests these are improbable (spring) or infrequent (fall) features. Another relevant example is found at OSP, where the  $\text{DMS}_{\text{L11}}$  estimates, based on measurements taken before  
25 2003, are in poor agreement with measurements made between 2005-2016. In February and June,  $\text{DMS}_{\text{SAT}}$  is in better accordance with in situ DMS data. Hence, caution has to be taken when comparing DMS measurements, their derived climatological products, and independent model estimates that are not collocated in time. This may partly explain the poor correlation between modeled DMS climatologies, on one hand, and the DMS database and  $\text{DMS}_{\text{L11}}$  climatology, on the other (Tesdal et al., 2015). Note that the latter study compared DMS fields binned into monthly  $5^\circ \times 5^\circ$  boxes (M5x5), such that  
30 82% of the bins contained measurements from a single year.

DMS fields based on satellite-observed properties are in better accordance with natural gradients in plankton abundance (biogeography, phenology) and environmental forcing, as long as the models can account for the driving factors. Indeed, our satellite approach is not devoid of problems. The most obvious of them is the propagation of error in  $\text{Chl}_{\text{SAT}}$  up to  $\text{DMS}_{\text{SAT}}$



estimates. The negative bias in  $\text{Chl}_{\text{SAT}}$  in the Antarctic biome, estimated at more than 50% by Johnson et al. (2013), and 56% in our matchup dataset, causes a large negative bias in  $\text{DMS}_{\text{SAT}}$  (and also  $\text{DMS}_{\text{SD02}}$ ) (Fig. 5-7). This indicates that the  $\text{DMSPt}_{\text{SAT}}$  sub-algorithm, rather than eq. 2 coefficients, would require important tuning for its application in the Antarctic Ocean. Our algorithm also suffers from built-in limitations that are explored below.

#### 5 4.2 How far can we go with empirical remote sensing algorithms?

The  $\text{DMS}_{\text{SAT}}$  algorithm captures in situ variability (Fig. 4) using a small set of predictor variables (Fig. 2). Moreover, it reproduces the mismatch between DMS and Chl such that, at a given  $\text{Chl}_{\text{SAT}}$  concentration, diagnosed DMS can vary by up to 40-fold (Fig. 8D). This mismatch is stronger than that produced by the SD02 or the VS07 algorithms. The correlation between  $\text{DMS}_{\text{SAT}}$  and  $\text{Chl}_{\text{SAT}}$  is 0.34 in the global climatology, similar to that found in the global database ( $r = 0.39$ ), and perhaps more realistic than that between the  $\text{DMS}_{\text{L11}}$  climatology and the SeaWiFS Chl climatology ( $r = 0.15$ ) (Fig. 8). Another positive feature of our algorithm is its capacity to produce a DMS peak in summer across different latitudes, the so-called DMS summer paradox, thanks to the progressive dissociation between  $\text{Chl}_{\text{SAT}}$  and DMS imposed by the two-step structure (Fig. 2) and the nonlinear relationships embodied in eq. 2 (Fig. 3). However, it fails to capture the high DMS/DMSPt ratios that occur in some regions between midsummer and early fall (Figs. 6 and 10), as discussed below.

15

Figs. 6-7 shows that, compared to  $\text{DMS}_{\text{SAT}}$ , the SD02 and VS07 algorithms produce higher DMS (and sometimes too high DMS) well into fall. This suggests that algorithms relying on MLD (SD02) or MLD combined with irradiance and water transparency (VS07) are better able to delay the annual DMS peak with respect to the summer solstice. Examination of the BATS and OSP time series (Fig. 10) gives insights into this issue. At both sites, the summer MLD is stable at about  $\leq 20$  m and deepens slowly in late summer (Levine et al., 2016; Steiner et al., 2012). Together with declining irradiance, this acts to decrease  $\text{PAR}_{\text{MLD}}$ . Thus, using  $\text{PAR}_{\text{MLD}}$  instead of surface PAR would not delay appreciably the decline of modeled DMS/DMSPt ratios through the summer, and other factors need to be invoked.

20

At BATS, some modeling studies proposed nutrient limitation as an important factor behind the seasonal mismatch between DMSPt and DMS, besides irradiance (Vallina et al., 2008). Polimene et al. (2011) proposed a mechanism whereby phosphorus and nitrogen limitation would increase the bacterial DMSP-to-DMS conversion yield in late summer. Belviso et al. (2012) specifically pointed out phosphorus deficiency as the driver of the Sargasso summer paradox through its effects on nanophytoplankton DMSP content and bacterial DMS yield. With these model results in mind, we tried to factor phosphate and nitrate limitation into our regression models using different variables: nutrient concentrations, nutricline depths (Table S4) and nutrient limitation factors estimated according to Michaelis Menten kinetics (not shown). However, none of the tested variables improved the regression models significantly. Moreover, nutrient availability (limitation) terms generally entered regression models with positive (negative) coefficients, even when regressions were restricted to oligotrophic low latitudes. This implies that nutrient limitation of phytoplankton growth globally acts to decrease DMS, offsetting nutrient

30



stress responses that increase DMS. The irregular occurrence of high DMS at BATS in late summer in different years, and the good performance of  $DMS_{SAT}$  in the Southern Westerlies biome (which should also display the effects of nutrient limitation), likely discard nutrient stress as the unique explanation.

5 Analysis of the OSP time series also yields valuable information. While in situ DMS is fairly predictable by  $DMS_{SAT}$  in February and June, the variable DMS peak occurring around August is strongly underestimated. Since OSP is an iron-limited, high-nutrient low-chlorophyll regime, nitrate or phosphate limitation cannot explain the high DMS/DMSPt ratios in late summer. By analyzing pigment data and performing sensitivity tests, Steiner et al. (2012) proposed dinoflagellate abundance, and the related increase in DMSP, as a plausible explanation. Interestingly, the  $DMSP_{tSAT}$  sub-algorithm  
10 produces the annual DMSP peak in August-September at OSP, in phase with the in situ DMS peak, and this despite that  $DMSP_{tSAT}$  does not explicitly resolve phytoplankton taxonomy (Galí et al., 2015).  $DMSP_{tSAT}$  cannot be systematically validated at OSP, but the available DMSPt measurements (Levasseur et al., 2006; Royer et al., 2010) do not suggest strong underestimation. The striking late summer variability at OSP is presently not captured by biogeochemical models (Steiner et al., 2012) or satellite algorithms, and it remains unanswered whether it simply reflects too low sampling frequency, or it is  
15 caused by a mechanism that switches on/off depending on environmental conditions on a given year, or by the variable location of oceanic fronts in response to circulation patterns.

In summary, our analysis indicates that additional factors may be needed to better reproduce DMS/DMSPt ratios in specific regions, but the inclusion of additional terms in eq. 2 lacks strong statistical support when applied globally, at least with the  
20 current dataset. Tuning the eq. 2 coefficients is a workable alternative to better reproduce the mean seasonal cycle in certain regions (Fig. 10), and eq. 2 could perhaps be generalized in a way that allowed its coefficients to vary across different biogeochemical regimes. The addition of new satellite variables as predictors, carrying additional retrieval uncertainty, could increase the  $DMS_{SAT}$  uncertainty. More obviously, climatological variables such as the WOA nutrient concentrations are not appropriate to produce time series, and their use in remote sensing algorithms should be minimized. The only climatological  
25 variable used in our algorithm is MLD, which enters mainly as a categorical variable (Galí et al., 2015), such that  $DMSP_{tSAT}$  is robust to MLD uncertainties (Fig. S1). Biotic interactions like microzooplankton grazing (Steiner et al., 2012) and bacterial metabolism (Levine et al., 2016; Royer et al., 2010) are indeed good candidates to explain strong deviations from the mean relationship between DMS, DMSPt and irradiance. However, they can hardly be included in empirical algorithms.

30 The question of the "optimal model complexity" is a pervasive one in biogeochemistry, and the right answer may depend on the purpose of each study. The algorithms tested here showed improved qualitative and quantitative performance with increasing complexity ( $VS07 < SD02 < DMS_{SAT}$ ).  $VS07$  failed to capture DMS patterns outside the subtropical band, possibly due to its inability to modulate the DMS-irradiance relationship depending on phytoplankton biomass. Inclusion of phytoplankton biomass-dependent predictors ( $SD02$ ), and of implicit taxonomic information through  $DMSP_{tSAT}$  estimation



(DMS<sub>SAT</sub>), improved algorithm skill in productive regions, where DMS shows wider seasonal cycles and sharper spatial gradients. More sophisticated approaches may be needed to achieve significant improvements in model skill, but they also suffer from major uncertainties. For example, neural networks were successfully used to estimate DMS in the Arctic (Humphries et al., 2012), but their predictive power might be compromised by the small training datasets, the use of climatological variables and the lack of a mechanistic basis. Complex biogeochemical models with satellite data assimilation are powerful tools for resolving interannual DMS variations, but reliance on several tens of poorly constrained parameters currently limits their skill (Le Clainche et al., 2010; Galí and Simó, 2015; Tesdal et al., 2015). A pathway of intermediate complexity that deserves further exploration is the remote sensing diagnosis of DMS using a simplified steady-state budget equation, which can account for biotic and abiotic DMS sources and sinks (Galí and Simó, 2015).

## 10 5 Conclusions and outlook

Sensors on polar-orbiting satellites provide synoptic global-scale observations of the ocean surface every few days, and are thus perfectly suited to resolve spatial and temporal variations in DMS concentration as long as they can retrieve the relevant proxy variables. Our algorithm reproduces the main spatial-temporal features of sea-surface DMS(P) concentrations with remarkable skill using only climatological MLD and satellite retrieved Chl, euphotic layer depth and PAR. Yet, it cannot produce high DMS/DMS<sub>Pt</sub> ratios in late summer, which suggests that irradiance cannot fully explain variability in DMS/DMS<sub>Pt</sub> ratios in some regions. In the Antarctic Ocean, bias in satellite retrieved Chl causes a strong negative bias in DMS<sub>SAT</sub>, which should be solved through regional tuning.

When compared against the current L11 DMS climatology, the DMS<sub>SAT</sub> climatology shows similar latitudinal profiles but disagrees in the basin-scale patterns. Examination of spatial DMS statistics highlights possible shortcomings in the L11 climatology caused by the combination of sparse and biased sampling, the right-skewed distribution of DMS, and the interpolation procedures used. High-resolution measurements of DMS(P), if validated against traditional standard techniques (Royer et al., 2014), will help improving interpolated climatologies and models.

The global mean area-weighted DMS<sub>SAT</sub> concentration is 1.63 nM, 33% lower than DMS<sub>L11</sub>. Excluding the Antarctic Ocean, the difference is -22%. Given the linear relationship between global mean DMS concentration and emission (Tesdal et al., 2015), this suggests a global emission of about 16–18 Tg S y<sup>-1</sup>, in the low end of current estimates. Global-scale DMS<sub>SAT</sub> fields are insensitive to the choice of different Chl and euphotic depth satellite products. However, the differences are larger in optically complex waters with continental influence, where semi-analytical products should be used.

30

The main strengths of our approach are its flexibility, allowing for regional tuning, and the minimal computing cost. Unlike climatologies constructed from the database, the satellite-based algorithm allows to explore interannual change. Implementation of DMS<sub>SAT</sub> in the subpolar Atlantic between 2003 and 2016 illustrates the wide interannual variability in the



timing and magnitude of the annual DMS peak(s) over large areas. This opens new avenues for studying the imprint of oceanic aerosol precursors on cloud properties using simultaneous ocean-atmosphere satellite observations (Meskhidze and Nenes, 2006). If coupled to atmospheric measurements and numerical models, DMS<sub>SAT</sub> enables studying the effects of contemporaneous DMS variability on atmospheric chemistry and clouds, which could lead to a better understanding of intricate aerosol-cloud interactions. Further work is warranted to analyze marine DMS emission variability patterns in regions where climate is particularly sensitive to DMS, such as the Southern Ocean and the Arctic.

#### Data availability

The primary database used to develop the DMS<sub>SAT</sub> algorithm is publicly available at <http://saga.pmel.noaa.gov/dms/>. The database extended with satellite matchups and climatological variables can be provided by the authors on request, as well as the global DMS and DMS<sub>Pt</sub> climatologies derived with the DMS<sub>SAT</sub>, SD02 and VS07 algorithms. The L11 DMS climatology and other related documents and datasets can be downloaded from [https://www.bodc.ac.uk/solas\\_integration/implementation\\_products/group1/dms/](https://www.bodc.ac.uk/solas_integration/implementation_products/group1/dms/).

#### Code availability

The code used to perform the data analyses and produce DMS<sub>SAT</sub>, SD02 and VS07 DMS fields can be provided by the authors on request.

#### Supplements

A supplementary information file is available.

#### Author contributions

M.G. designed the study, performed the research and wrote the paper, with input from all coauthors through the different phases. E.D. processed remote sensing reflectance data used as input for the DMS<sub>SAT</sub> northern hemisphere time series.

#### Competing interests

The authors declare that they have no conflict of interest.

#### Acknowledgments

We thank the NASA Ocean Biology Distributed Active Archive Center (OB.DAAC) for access to MODIS and SeaWiFS datasets, T.S Bates (NOAA/ PMEL) for the maintenance of the GSS DMS(P) database and the DMS-GO project for the recent database update (a joint initiative of the SOLAS Integration Project and the EU projects COST Action 735 and EUROCEANS to R.S.); Eric Rehm and Maxime Benoît-Gagné for IT support; Naomi Levine, John Dacey, Nick Bates and Scott



Doney for sharing BATS data; Marie Robert and Michael Arychuk for guidance on access to Ocean Station P public datasets; and Maxime Benoit-Gagné and Eric Rehm for IT support. We acknowledge funding from the Canada Excellence Research Chair in Remote Sensing of Canada's New Arctic Frontier (M.B.), the Canada Research on Ocean Biogeochemistry and Climate (M.L), a NSERC Discovery Grant Program and Northern Research Supplement Program (M.L.), the NETCARE network (funded under the NSERC Climate Change and Atmospheric Research program) and ArcticNet (The Network of Centres of Excellence of Canada). M.G. acknowledges the receipt of a Beatriu de Pinós postdoctoral fellowship funded by AGAUR (Generalitat de Catalunya). This project is a contribution to the research program of Québec-Océan and the Takuvik Joint International Laboratory (CNRS-France & Université Laval-Canada).

## 10 References

- 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(6242), 1466–1469, 2015.
- 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, doi:10.1007/s10533-011-9648-1, 2012.
- Carpenter, L. J., Archer, S. D. and Beale, R.: Ocean-atmosphere trace gas exchange, *Chem. Soc. Rev.*, 41, 6473–6505, doi:10.1039/c2cs35121h, 2012.
- 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(7474), 67–71, doi:10.1038/nature12674, 2013.
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. and Warren, S. G.: Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate, *Nature*, 326, 655–661, 1987.
- Le Clainche, Y., Vézina, A., Levasseur, M., Cropp, R. A., Gunson, J. R., Vallina, S. M., Vogt, M., Lancelot, C., Allen, J. I., Archer, S. D., Bopp, L., Deal, C., Elliott, S., Jin, M., Malin, G., Schoemann, V., Simó, R., Six, K. D. and Stefels, J.: A first appraisal of prognostic ocean DMS models and prospects for their use in climate models, *Global Biogeochem. Cycles*, 24, GB3021, doi:10.1029/2009GB003721, 2010.
- Curson, A. R. J., Todd, J. D., Sullivan, M. J. and Johnston, A. W. B.: Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes, *Nat. Rev. Microbiol.*, 9(12), 849–59, doi:10.1038/nrmicro2653, 2011.
- Frouin, R., Franz, B. and Wang, M.: Algorithm to estimate PAR from SeaWiFS data Version 1.2-Documentation, 2003.
- Galí, M. and Simó, R.: Occurrence and cycling of dimethylated sulfur compounds in the Arctic during summer receding of the ice edge, *Mar. Chem.*, 122, 105–117, doi:10.1016/j.marchem.2010.07.003, 2010.



- Gali, M. and Simó, R.: A meta-analysis of oceanic DMS and DMSP cycling processes: Disentangling the summer paradox, *Global Biogeochem. Cycles*, 29, 496–515, doi:10.1002/2014GB004940, 2015.
- Gali, M., Simó, R., Vila-Costa, M., Ruiz-González, C., Gasol, J. M. and Matrai, P. A.: Diel patterns of oceanic dimethylsulfide (DMS) cycling: Microbial and physical drivers, *Global Biogeochem. Cycles*, 27, 620–636, 2013.
- 5 Gali, M., Devred, E., Levasseur, M., Royer, S.-J. and Babin, M.: A remote sensing algorithm for planktonic dimethylsulfoniopropionate (DMSP) and an analysis of global patterns, *Remote Sens. Environ.*, 171, 171–184, doi:10.1016/j.rse.2015.10.012, 2015.
- Grandey, B. S. and Wang, C.: Enhanced marine sulphur emissions offset global warming and impact rainfall, *Sci. Rep.*, 5, 13055, doi:10.1038/srep13055, 2015.
- 10 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(1–3), 287–301, doi:10.1007/s10533-011-9683-y, 2012.
- Jodwalis, C. M. and Benner, R. L.: On the possibility of remotely sensing global dimethyl sulfide sea-to-air flux, *Polar Rec.*, 31(177), 251–256, 1995.
- Johnson, R., Stratton, P. G., Wright, S. W., McMinn, A. and Meiners, K. M.: Three improved satellite chlorophyll algorithms for the Southern Ocean, *J. Geophys. Res. Ocean.*, 118(7), 3694–3703, doi:10.1002/jgrc.20270, 2013.
- Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A. M., Helber, R. and Arnone, R. A.: Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, *J. Mar. Syst.*, 76(1–2), 64–82, doi:10.1016/j.jmarsys.2008.05.014, 2009.
- Kettle, A. J. and Andreae, M. O.: Flux of dimethylsulfide from the oceans: A comparison of updated data sets and flux models, *J. Geophys. Res.*, 105(D22), 26793, doi:10.1029/2000JD900252, 2000.
- 20 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., Simó, R., Staubes, R., Turner, S. and Uher, G.: A global database of sea surface
- 25 dimethylsulfide (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month, *Global Biogeochem. Cycles*, 13(2), 399–444, doi:10.1029/1999GB900004, 1999.
- Kirk, J. T. O.: *Light and photosynthesis in aquatic ecosystems*, Cambridge University Press, New York, NY, 2011.
- Kulmala, M., Petäjä, T., Ehn, M., Thornton, J., Sipilä, M., Worsnop, D. R. and Kerminen, V.-M.: Chemistry of atmospheric nucleation: on the recent advances on precursor characterization and atmospheric cluster composition in connection with
- 30 atmospheric new particle formation., *Annu. Rev. Phys. Chem.*, 65, 21–37, doi:10.1146/annurev-physchem-040412-110014, 2014.



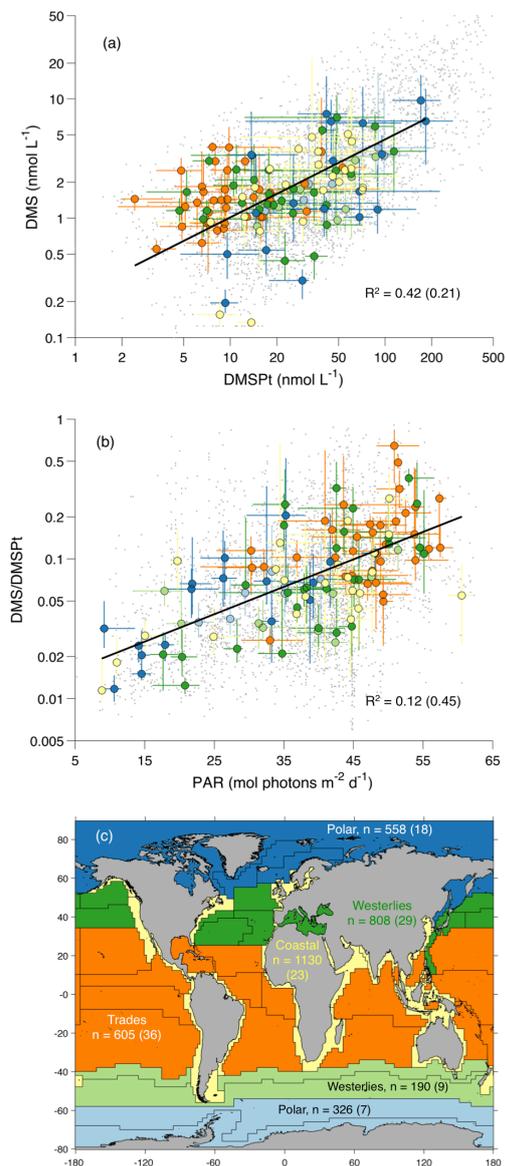
- Lana, A., Bell, T. G., Simó, R., Vallina, S. M., Ballabrera-Poy, J., Kettle, A. J., Dachs, J., Bopp, L., Saltzman, E. S., Stefels, J., Johnson, J. E. and Liss, P. S.: An updated climatology of surface dimethylsulfide concentrations and emission fluxes in the global ocean, *Global Biogeochem. Cycles*, 25, GB1004, doi:10.1029/2010GB003850, 2011.
- Lavoie, M., Levasseur, M. and Sunda, W. G.: A steady-state physiological model for intracellular dimethylsulfoxide in marine phytoplankton, *Environ. Chem.*, doi:10.1071/EN14221\_AC, 2015.
- Leaitch, W. R., Sharma, S., Huang, L., Toom-Sauntry, D., Chivulescu, A., Macdonald, A. M., von Salzen, K., Pierce, J. R., Bertram, A. K., Schroder, J. C., Shantz, N. C., Chang, R. Y. W. and Norman, A.-L.: Dimethyl sulfide control of the clean summertime Arctic aerosol and cloud, *Elem. Sci. Anth.*, 1(1), 17, doi:10.12952/journal.elementa.000017, 2013.
- Lee, Z., Weidemann, A., Kindle, J., Arnone, R., Carder, K. L. and Davis, C.: Euphotic zone depth: Its derivation and implication to ocean-color remote sensing, *J. Geophys. Res.*, 112(C3), C03009, doi:10.1029/2006JC003802, 2007.
- Levasseur, M.: Impact of Arctic meltdown on the microbial cycling of sulphur, *Nat. Geosci.*, 6(9), 691–700, doi:10.1038/ngeo1910, 2013.
- Levasseur, M., Scarratt, M., Michaud, S., Merzouk, A., Wong, C., Arychuk, M., Richardson, W., Rivkin, R., Hale, M. and Wong, S.: DMSP and DMS dynamics during a mesoscale iron fertilization experiment in the Northeast Pacific—Part I: Temporal and vertical distributions, *Deep Sea Res. Part II Top. Stud. Oceanogr.*, 53(20–22), 2353–2369, doi:10.1016/j.dsr2.2006.05.023, 2006.
- Levine, N. M., Toole, D. A., Neeley, A., Bates, N. R., Doney, S. C. and Dacey, J. W. H.: Revising upper-ocean sulfur dynamics near Bermuda: New lessons from 3 years of concentration and rate measurements, *Environ. Chem.*, 13(2), 302–313, doi:10.1071/EN15045, 2016.
- Lizotte, M., Levasseur, M., Michaud, S., Scarratt, M. G., Merzouk, A., Gosselin, M., Pommier, J., Rivkin, R. B. and Kiene, R. P.: Macroscale patterns of the biological cycling of dimethylsulfoniopropionate (DMSP) and dimethylsulfide (DMS) in the Northwest Atlantic, *Biogeochemistry*, 110(1–3), 183–200, doi:10.1007/s10533-011-9698-4, 2012.
- Longhurst, A.: *Ecological Geography of the Sea*, 2nd ed., Academic Press, 2010.
- Mahajan, A. S., Fadnavis, S., Thomas, M. A., Pozzoli, L., Gupta, S., Royer, S., Saiz-lopez, A. and Simó, R.: Quantifying the impacts of an updated global dimethylsulfide climatology on cloud microphysics and aerosol radiative forcing, *J. Geophys. Res. Atmos.*, 120, 2524–2536, doi:10.1002/2014JD022687. Received, 2015.
- Maritorena, S., Siegel, D. A. and Peterson, A. R.: Optimization of a semianalytical ocean color model for global-scale applications, *Appl. Opt.*, 41(15), 2705–2714, doi:10.1364/AO.41.002705, 2002.
- McCoy, D. T., Burrows, S. M., Wood, R., Grosvenor, D. P., Elliott, S. M., Ma, P., Rasch, P. J. and Hartmann, D. L.: Natural aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo, *Sci. Adv.*, 1, e1500157, 2015.
- Meskhidze, N. and Nenes, A.: Phytoplankton and cloudiness in the Southern Ocean, *Science*, 314(5804), 1419–23, doi:10.1126/science.1131779, 2006.



- Modini, R. L., Ristovski, Z. D., Johnson, G. R., He, C., Surawski, N., Morawska, L., Suni, T. and Kulmala, M.: New particle formation and growth at a remote, sub-tropical coastal location, *Atmos. Chem. Phys. Discuss.*, 9(3), 12101–12139, doi:10.5194/acpd-9-12101-2009, 2009.
- Morel, A. and Smith, R. C.: Relation between total quanta and total energy for aquatic photosynthesis, *Limnol. Oceanogr.*, 19, 1974.
- Polimene, L., Archer, S. D., Butenschön, M. and Allen, J. I.: A mechanistic explanation of the Sargasso Sea DMS “summer paradox,” *Biogeochemistry*, 110(1–3), 243–255, doi:10.1007/s10533-011-9674-z, 2011.
- Quinn, P. K. and Bates, T. S.: The case against climate regulation via oceanic phytoplankton sulphur emissions, *Nature*, 480, 51–56, doi:10.1038/nature10580, 2011.
- 10 Royer, S., Mahajan, A. S., Galí, M., Saltzman, E. and Simó, R.: Small-scale variability patterns of DMS and phytoplankton in surface waters of the tropical and subtropical Atlantic, Indian, and Pacific Oceans, *Geophys. Res. Lett.*, 42, 475–483, doi:10.1002/2014GL062543. Received, 2015.
- Royer, S.-J., Levasseur, M., Lizotte, M., Arychuk, M., Scarratt, M. G., Wong, C., Lovejoy, C., Robert, M., Johnson, K., Pena, A., Michaud, S. and Kiene, R. P.: Microbial dimethylsulfoniopropionate (DMSP) dynamics along a natural iron gradient in the northeast subarctic Pacific, *Limnol. Oceanogr.*, 55(4), 1614–1626, doi:10.4319/lo.2010.55.4.1614, 2010.
- 15 Royer, S. J., Galí, M., Saltzman, E. S., McCormick, C. A., Bell, T. G. and Simó, R.: Development and validation of a shipboard system for measuring high-resolution vertical profiles of aqueous dimethylsulfide concentrations using chemical ionisation mass spectrometry, *Environ. Chem.*, 11, 309–317, doi:10.1071/EN13203, 2014.
- Schlesinger, W. H. and Bernhardt, E. S.: The Global Cycles of Sulfur and Mercury, *Biogeochem. An Anal. Glob. Chang.*, 20 469–486, doi:<http://dx.doi.org/10.1016/B978-0-12-385874-0.00012-1>, 2013.
- Schmidtko, S., Johnson, G. C. and Lyman, J. M.: MIMOC: A global monthly isopycnal upper-ocean climatology with mixed layers, *J. Geophys. Res. Ocean.*, 118(4), 1658–1672, doi:10.1002/jgrc.20122, 2013.
- Shaw, G. E.: Bio-controlled thermostasis involving the sulfur cycle, *Clim. Change*, 5(3), 297–303, doi:10.1007/BF02423524, 1983.
- 25 Simó, R.: From cells to globe : approaching the dynamics of DMS(P) in the ocean at multiple scales, *Can. J. Fish. Aquat. Sci.*, 61, 673–684, doi:10.1139/F04-030, 2004.
- Simó, R.: The role of marine microbiota in short-term climate regulation, in *The Role of Marine Biota in the Functioning of the Biosphere*, edited by C. M. Duarte, pp. 107–130, Fundación BBVA, 2011.
- Simó, R. and Dachs, J.: Global ocean emission of dimethylsulfide predicted from biogeophysical data, *Global Biogeochem. Cycles*, 16(4), doi:10.1029/2001GB001829, 2002.
- 30 Simó, R. and Pedrós-Alió, C.: Role of vertical mixing in controlling the oceanic production of dimethyl sulphide, *Nature*, 402, 396–399, 1999.
- Spiese, C. E., Le, T., Zimmer, R. L. and Kieber, D. J.: Dimethylsulfide membrane permeability, cellular concentrations and implications for physiological functions in marine algae, *J. Plankton Res.*, 38(1), 41–54, doi:10.1093/plankt/fbv106, 2015.



- Stefels, J.: Physiological aspects of the production and conversion of DMSP in marine algae and higher plants, *J. Sea Res.*, 43, 183–197, doi:10.1016/S1385-1101(00)00030-7, 2000.
- Stefels, J., Steinke, M., Turner, S. M., 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, 2007.
- 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(1–3), 269–285, doi:10.1007/s10533-011-9669-9, 2012.
- Sunda, W., Kieber, D. J., Kiene, R. P. and Huntsman, S.: An antioxidant function for DMSP and DMS in marine algae, *Nature*, 418(6895), 317–320, doi:10.1038/nature00851, 2002.
- Sunda, W. G., Hardison, R., Kiene, R. P., Bucciarelli, E. and Harada, H.: The effect of nitrogen limitation on cellular DMSP and DMS release in marine phytoplankton: climate feedback implications, *Aquat. Sci.*, 69(3), 341–351, doi:10.1007/s00027-007-0887-0, 2007.
- Sutherland, G., Reverdin, G., Marié, L. and Ward, B.: Mixed and mixing layer depths in the ocean surface boundary, *Geophys. Res. Lett.*, 41, 8469–8476, doi:10.1002/2014GL061939.A, 2014.
- Tesdal, J.-E., Christian, J. R., Monahan, A. H. and von Salzen, K.: Evaluation of diverse approaches for estimating sea-surface DMS concentration and air-sea exchange at global scale, *Environ. Chem.*, 390–412, doi:10.1071/EN14255, 2015.
- Thompson, A. M., Esaias, W. E. and Iverson, R. L.: Two Approaches to Determining the Sea-to-Air Flux of Dimethyl Sulfide: Satellite Ocean Color and a Photochemical Model With Atmospheric Measurements, *J. Geophys. Res.*, D12, 20551–20558, 1990.
- Toole, D. A. and Siegel, D. A.: Light-driven cycling of dimethylsulfide (DMS) in the Sargasso Sea: Closing the loop, *Geophys. Res. Lett.*, 31(9), 5–8, doi:10.1029/2004GL019581, 2004.
- Vallina, S. M. and Simó, R.: Re-visiting the CLAW hypothesis, *Environ. Chem.*, 4, 384–387, doi:10.1071/EN07055, 2007a.
- Vallina, S. M. and Simó, R.: Strong relationship between DMS and the solar radiation dose over the global surface ocean., *Science*, 315(5811), 506–508, doi:10.1126/science.1133680, 2007b.
- Vallina, S. M., Simó, R., Anderson, T. R., Gabric, A., Cropp, R. and Pacheco, J. M.: A dynamic model of oceanic sulfur (DMOS) applied to the Sargasso Sea: Simulating the dimethylsulfide (DMS) summer paradox, *J. Geophys. Res.*, 113, 1–23, doi:10.1029/2007JG000415, 2008.
- Vantrepotte, V. and Mélin, F.: Inter-annual variations in the SeaWiFS global chlorophyll a concentration (1997–2007), *Deep-Sea Res. Part I*, 58(4), 429–441, doi:10.1016/j.dsr.2011.02.003, 2011.
- Yu, F. and Luo, G.: Oceanic dimethyl sulfide emission and new particle formation around the coast of Antarctica: A modeling study of seasonal variations and comparison with measurements, *Atmosphere*, 1(1), 34–50, doi:10.3390/atmos1010034, 2010.

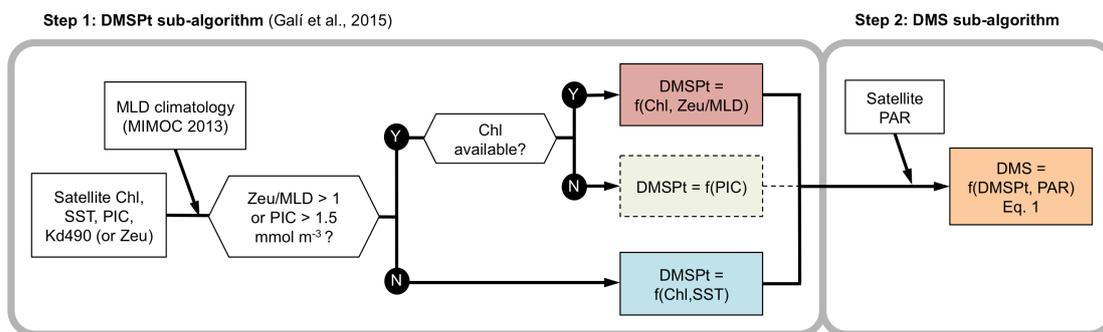


**Figure 1: Relationship between DMS, DMSPt and PAR across oceanic biomes.** (a) DMS vs. DMSPt; (b) DMS/DMSPt vs. mean daily irradiance (PAR) at the sea surface; (c) Longhurst biogeochemical provinces and biomes. In (a) and (b) small

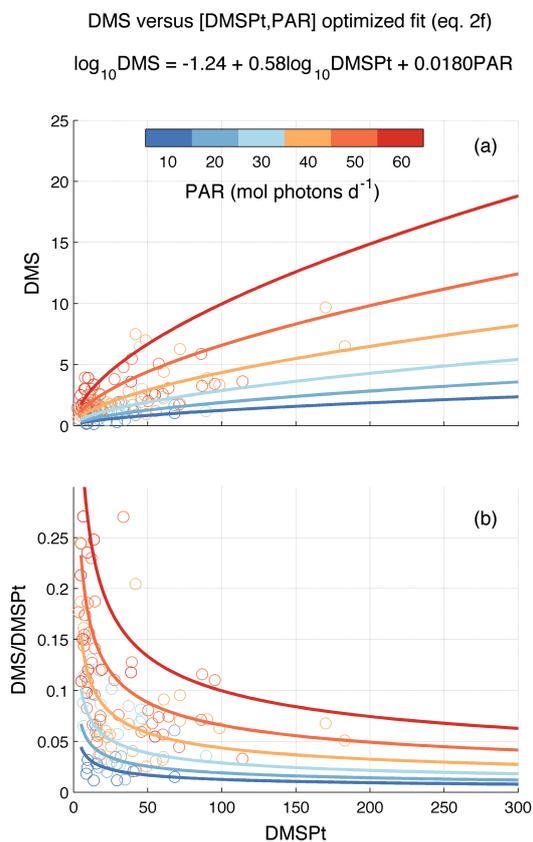


grey dots represent individual data points and large colored dots represent the median in a given Longhurst biogeochemical province and month and the corresponding interquartile ranges. Province-month medians are colored by biome following the map in (c), which also shows the amount of DMS-DMSpt-PAR measurements available in each biome. The  $R^2$  and data counts outside and inside parentheses correspond to non-binned data and province-month binned data, respectively.

5 Regression lines in (a) and (b), calculated with binned data, are only illustrative.

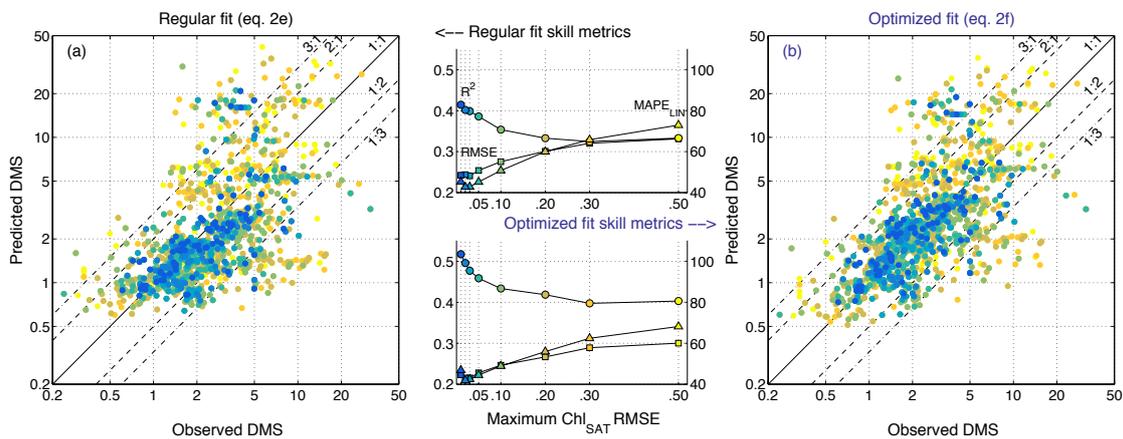


**Figure 2: Scheme of the DMS<sub>SAT</sub> algorithm.** The algorithm proceeds in two steps: the DMSPt sub-algorithm (described by Galí et al., 2015) and the DMS sub-algorithm (described in this study). Dashed lines mark the PIC-based equation of the DMSPt sub-algorithm, which in practice is not used when gap-free satellite Chl fields are used as input.

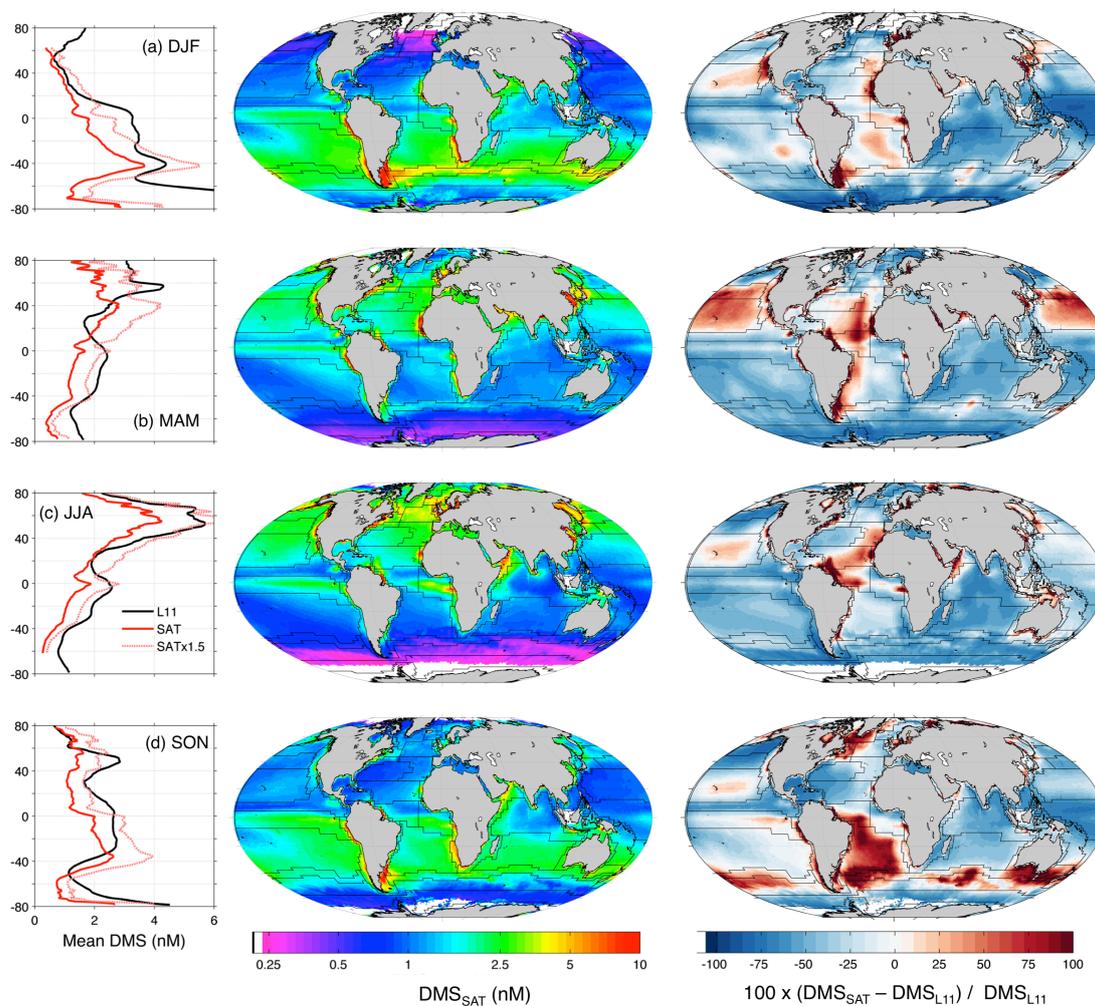


**Figure 3: Relationship between DMS, DMSPt and PAR as represented in the algorithm.** (a) DMS vs. DMSPt as a function of PAR; (b) DMS/DMSPt ratio vs. PAR. Lines correspond to the model predictions for different PAR levels, and colored circles represent the medians of in situ data binned by Longhurst province and month.

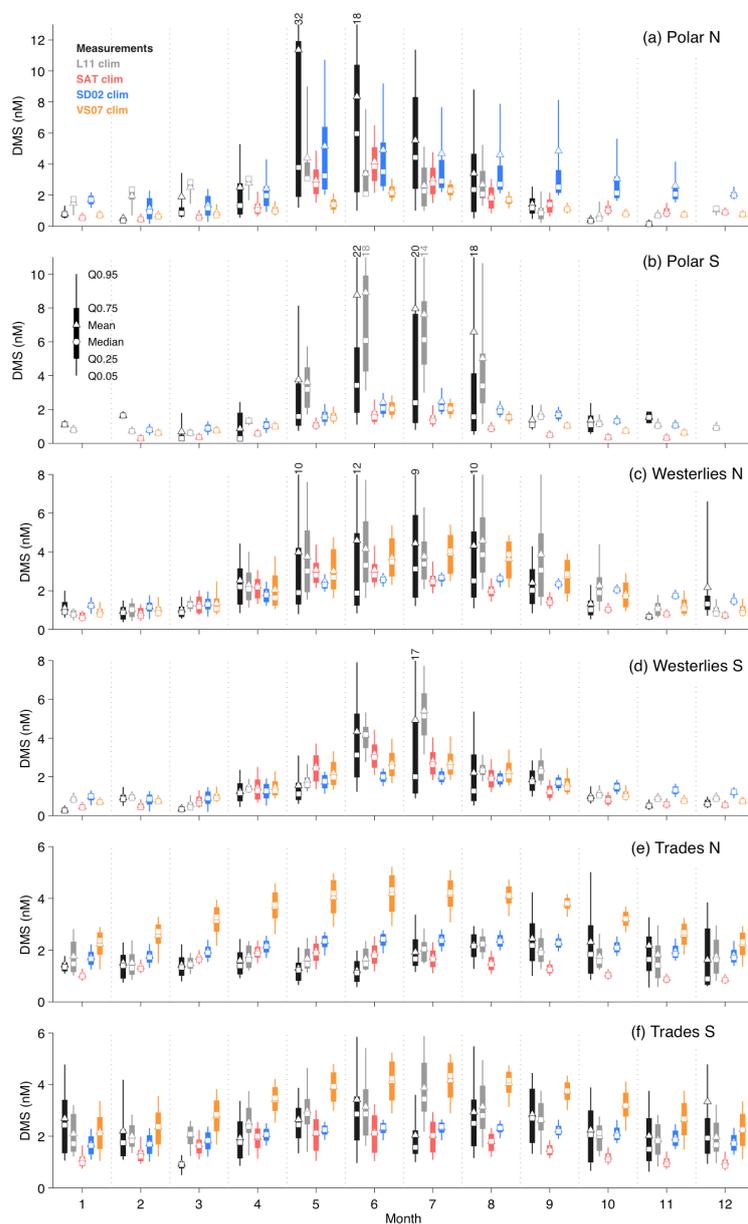
5



**Figure 4: Algorithm validation results constrained by the uncertainty in satellite-retrieved Chl.** (a) eq. 2e, derived from regular multiple regression; (b) eq. 2f, obtained through an optimization procedure. The scatterplots compare non-binned data and model predictions, color-coded depending on the maximum tolerated error in  $\text{Chl}_{\text{SAT}}$  with respect to Chl in situ, as shown in the x-axis of the center plots. The center plots show the performance of the DMS algorithm for increasing error in  $\text{Chl}_{\text{SAT}}$ , evaluated with different skill metrics: the  $\log_{10}$  space  $R^2$  and RMSE (left y-axis) and the linear space MAPE (right y-axis). N increases from 86 to 1293 as the tolerated  $\text{Chl}_{\text{SAT}}$  error increases.



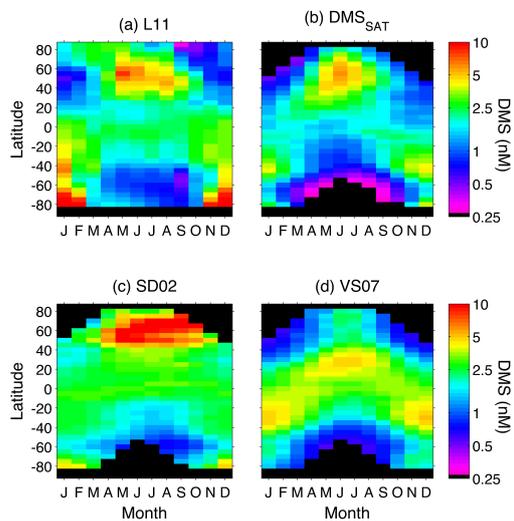
**Figure 5: Global  $DMS_{SAT}$  concentration fields by season.** (a) December-February (DJF); (b) March-May (MAM); (c) June-August (JJA); (d) September-November (SON). Each row contains mean latitudinal profiles for the L11 climatology,  $DMS_{SAT}$ , and  $1.5 \times DMS_{SAT}$  (left);  $DMS_{SAT}$  concentration maps (center); and maps of the % difference between  $DMS_{SAT}$  and the L11 climatology (right).



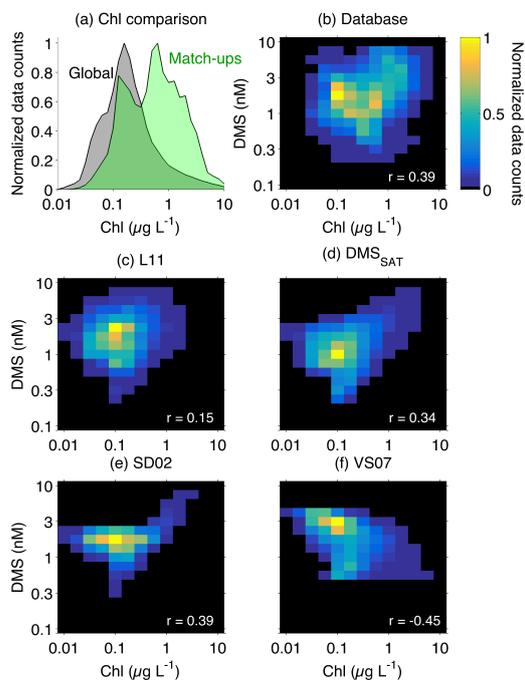
**Figure 6. DMS seasonal cycles by biomes.** The monthly means, medians, interquartile range and 5%-95% percentiles are shown for the in situ database, the L11 climatology, and remote sensing climatologies derived from the  $DMS_{SAT}$ , SD02 and



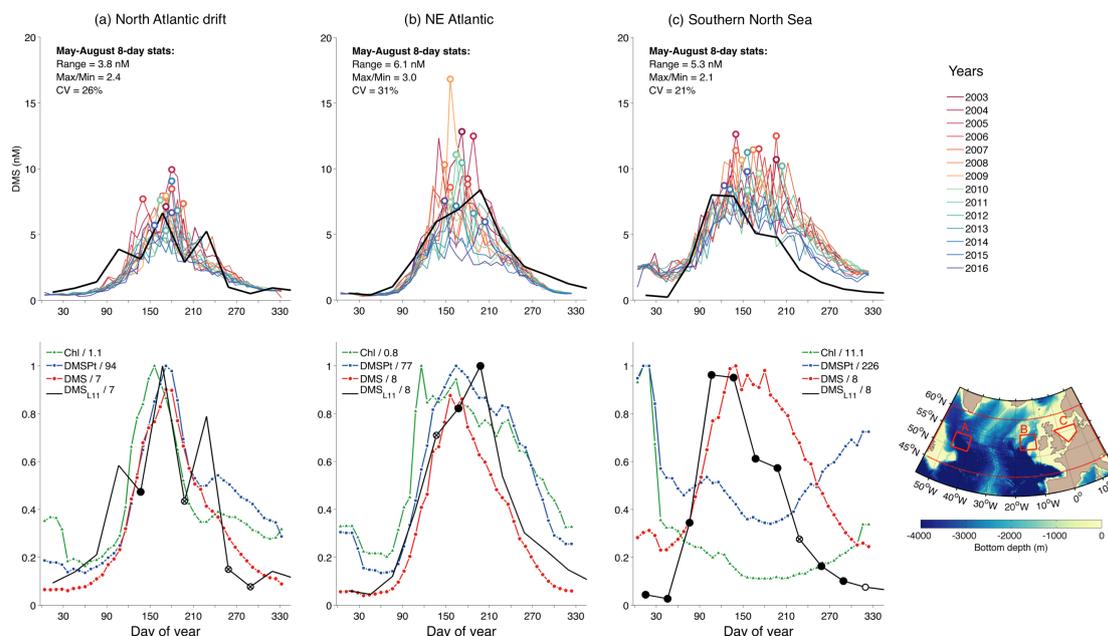
VS07 algorithms. The temporal axis has been shifted by 6 months in the Southern hemisphere, i.e., July is the 1st month and June the 12th.



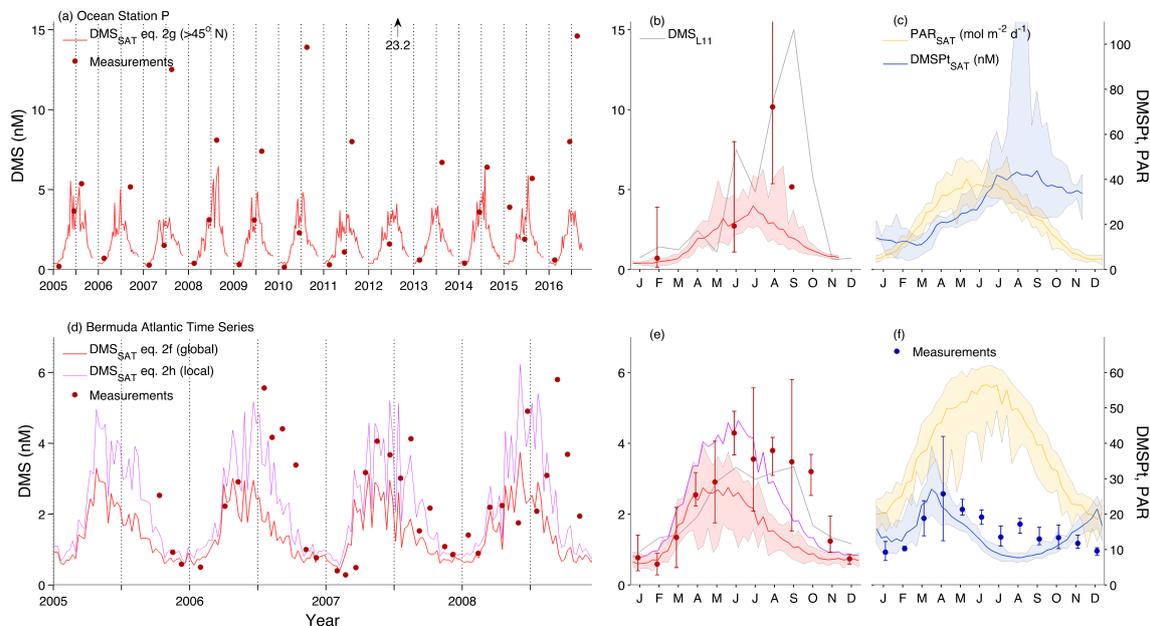
**Figure 7: Hovmöller diagrams comparing climatological DMS fields.** (a) L11 climatology, (a) DMS<sub>SAT</sub> algorithm, (c) SD02 algorithm, and (d) VS07 algorithm.



**Figure 8: Histograms illustrating the relationship between DMS and Chl in the global ocean.** (a)  $\text{Chl}_{\text{SAT}}$  histograms for the global SeaWiFS 1997-2010 climatology and the DMS database SeaWiFS match-ups; (b–f) 2D histograms of the global SeaWiFS 1997-2010 Chl climatology vs. DMS from (b) the in situ DMS database, (c) the L11 DMS climatology, (d)  $\text{DMS}_{\text{SAT}}$ , (e) SD02 and (f) VS07 algorithms.



**Figure 9: Interannual  $DMS_{SAT}$  variability in the subpolar North Atlantic.** (a) northwest Atlantic drift, (b) shelf break west of Ireland, (c) Southern North Sea shelf. The top row shows individual years between 2003 and 2016 diagnosed from 8-day MODIS-Aqua data, marked by colors, and the mean seasonal cycle according to the L11 DMS climatology (black); colored circles mark the peak of each seasonal cycle. The bottom row shows the mean annual cycles of  $Chl_{SAT}$ ,  $DMSPT_{SAT}$ ,  $DMS_{SAT}$  and the L11 DMS climatology; each variable is divided by its maximum, shown by the number in the quotient, and a common scaling factor is used for  $DMS_{SAT}$  and  $DMS_{L11}$ ; markers on the L11 line indicate the amount of in situ data on which the L11 climatology is based in a given month: no data, i.e. month filled through interpolation (no marker); 1–9 measurements on one single year (empty circles);  $\geq 10$  measurements on one single year (crossed circles); and  $\geq 10$  measurements on more than one year (filled black circles). Red polygons on the map show the 3 selected areas and the larger region used for the validation scatterplots (Fig. S3).



**Figure 10: DMS<sub>SAT</sub> vs. in situ data at long-term research stations.** (a–c) Ocean Station P in the northeast Pacific (50°N, 145°W); (d–f) Bermuda Atlantic Time Series station (31°40'N, 64°10'W). (a) and (d) compare DMS<sub>SAT</sub> estimates to in situ measurements; (b) and (e) compare monthly DMS climatologies derived from DMS<sub>SAT</sub> (2003–2016 MODIS-Aqua data), in situ data (available measurements between 2003–2016) and the L11 DMS climatology; (c) and (f) show the corresponding DMSPt<sub>SAT</sub> and PAR<sub>SAT</sub> climatologies (2003–2016) and, at BATS, in situ DMSPt. The periods shown for in situ data are 2005–2016 for OSP (during which 3 measurements per year are generally available) and 2005–2008 for BATS. Shaded envelopes and error bars show the minima and maxima of satellite-derived and in situ data, respectively.

10



**Table 1: Correlation analysis.** Correlation coefficients with p-value < 0.01 are not shown; *italic* marks  $0.0001 < p < 0.01$ ; na: not applicable. "Ratio" refers to  $\log_{10}(\text{DMS}/\text{DMSPt})$ . DMS, DMSPt, Chl, MLD,  $[\text{NO}_3]$  and  $[\text{PO}_4]$  were  $\log_{10}$  transformed. Bottom and nutricline depths have positive sign (deeper is bigger). See the text for other acronyms.

	Non-binned data				Monthly 5°x5° binning (M5x5)				Monthly Longhurst binning (MLongh)			
	DMS	N	Ratio	N	DMS	N	Ratio	N	DMS	N	Ratio	N
<b>In situ data</b>												
DMSPt	0.65	3637	na	1442	0.58	308	na	157	0.46	122	na	87
Chl	0.45	to	-0.33	to	0.37	to	-0.28	to	0.34	to	-0.45	to
SST	-0.02	41304	0.29	3637		1562	0.45	308		322	0.56	119
Salinity	-0.12		0.27								0.32	
Wind speed	-0.12		-0.13		-0.12				-0.20		-0.27	
Bottom depth	-0.19		0.10		-0.12		0.16				0.26	
Day Length	0.42		0.06		0.43				0.49			
<b>Climatological data</b>												
MLD	-0.37	35505	-0.13	3433	-0.32	1474		298	-0.51	312	-0.24	119
$[\text{NO}_3]$	0.06	to	-0.19	to	0.16	to	-0.31	to		to	-0.45	
$[\text{PO}_4]$	0.05	39478	-0.15	3637	0.13	1535	-0.34	308		318	-0.32	
N-cline	-0.14		0.30		-0.22		0.44				0.55	
P-cline	-0.12		0.24		-0.14		0.45				0.37	
<b>Satellite match-up data</b>												
PAR <sub>SAT</sub>	0.32	16411	0.35	1123	0.30	498	0.46	124	0.52	171	0.67	86
PAR <sub>MLD</sub>	0.12	to	0.37	to	0.15	to	0.49	to	0.36	to	0.66	to
Chl <sub>SAT</sub>	0.37	41088	-0.42	3620	0.22	1539	-0.34	307	0.28	321	-0.39	119
PIC <sub>SAT</sub>	0.24		-0.27		0.29		-0.30		0.33			

5



**Table 2: Summary of fitted model coefficients and goodness-of-fit statistics.** Different sets of coefficients were obtained by fitting the model  $\log_{10}DMS = \alpha + \beta \log_{10}DMSpt + \gamma PAR$  to observed DMS, DMSpt and  $PAR_{SAT}$  after applying different binning schemes. Equations 2f and 2h were derived using a different optimization procedure, applied to the global MLongh binned dataset and to the Bermuda Atlantic Time Series local dataset. Shading highlights the models implemented to calculate a global DMS climatology (lighter gray) and regional or local time series (darker gray).

	Binning	Bin metric	Equation	$\alpha$	$\beta$	$\gamma$	$R^2_{adj}$	RMSE	SlopeMA	N
<b>Regression, global scale</b>	Non-binned		2a	-1.213 ± 0.028	0.672 ± 0.012	0.0136 ± 0.0006	0.50	0.35	0.62	3620
	M5x5	Mean	2b	-1.154 ± 0.083	0.669 ± 0.0371	0.0130 ± 0.0015	0.55	0.28	0.67	307
		Median	2c	-1.061 ± 0.084	0.569 ± 0.039	0.0130 ± 0.0015	0.46	0.28	0.58	
	MLongh	Mean	2d	-1.061 ± 0.115	0.583 ± 0.054	0.0155 ± 0.0019	0.57	0.24	0.70	118
		Median	2e	-1.018 ± 0.100	0.452 ± 0.050	0.0163 ± 0.0016	0.57	0.21	0.69	
<b>Optimization, global scale</b>	MLongh	Median	2f	-1.237	0.578	0.0180	0.56	0.22	0.87	118
<b>Regression, regional scale (&gt;45N)</b>	M5x5	Mean	2g	-1.283 ± 0.154	0.670 ± 0.097	0.0186 ± 0.011	0.68	0.28	0.80	87
<b>Optimization, local scale (BATS)</b>	Non-binned		2h	-0.898	0.316	0.0214	0.44	0.26	0.66	35



**Table 3: Global mean area-weighted DMS concentrations calculated with different algorithms.** Different  $DMS_{SAT}$  results were obtained with alternative approaches for retrieving chlorophyll *a* concentration ( $Chl_{SAT}$ ) and the euphotic layer depth ( $Ze_{SAT}$ ) from satellite data. Calculations are based on  $1^\circ \times 1^\circ$  gridded data; na: not applicable.

DMS algorithm or data product	$Chl_{SAT}$ product	$Kd_{SAT}$ or $Ze_{SAT}$ product	Area weighted global DMS mean (nM)
L11 climatology (Lana et al., 2011)	na	na	2.44
SD02 (Simó and Dachs, 2002)	OC4-CI	na	2.12
VS07 (Vallina and Simó, 2007b)	na	$Kd_{490}$	2.71
$DMS_{SAT}$ eq. 2f (this study)	OC4-CI	$Ze_u = 4.6/Kd_{490}$	1.63
		$Ze_{Lee}$	1.58
	GSM	$Ze_u = 4.6/Kd_{490}$	1.58
		$Ze_{Lee}$	1.55

5