Biogeochemistry (2012) 110:287–301
DOI 10.1007/s10533-011-9683-y

Spatial predictions of sea surface dimethylsulfide concentrations in the high arctic

Grant R. W. Humphries · Clara J. Deal · Scott Elliott · Falk Huettmann

Received: 13 February 2011 / Accepted: 17 November 2011 / Published online: 19 January 2012

Abstract The climatically-important compound dimethylsulfide (DMS) has been reported to be abundant in the Arctic, particularly in the marginal sea ice zone. Due to these high concentrations, it may play an important role in climate control. A DMS monthly climatology for July through October was created employing various ocean characteristics and spatial modeling techniques commonly used for describing species distributions in ecology. Comparisons between observed and predicted values of surface seawater DMS concentrations led to r^2 values of 0.61, 0.87, 0.66, and 0.37 for July, August, September, and October, respectively. Measurement data used for model development for July through October were variably distributed spatially. For October only, data were sparse and clustered, resulting in the poor results obtained for this month. Mean sea ice concentration and surface nitrate concentrations were found to be important predictors of surface seawater DMS concentrations. A negative relationship between sea ice concentration and DMS, and a two-phase relationship between nitrate and DMS were found. The two-phase relationship may be indicative of how DMS concentrations are affected when nitrate is the limiting nutrient. From July to September, predicted DMS concentrations were generally lowest under the sea ice. High monthly DMS concentrations (up to 10.7 nM) were predicted in the seasonal ice zone. The highest DMS concentrations in September ($\sim$2.6 nM) were predicted along the ice edge. In order to create more accurate climatologies and to increase our understanding of sulfur cycling in the Arctic, a higher spatial and temporal distribution of DMS measurements is required.

Keywords Dimethylsulfide · Arctic · Sea ice · TreeNet · GIS · Spatial modeling

G. R. W. Humphries (✉)
Center for Sustainability: Agriculture, Food, Energy and Environment, University of Otago, Dunedin, New Zealand
e-mail: humphries.grant@gmail.com

C. J. Deal
International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK, USA

S. Elliott
Division of Computational Sciences, Climate Ocean Sea Ice Modeling, Los Alamos National Laboratory, Los Alamos, NM, USA

F. Huettmann
EWHALE Lab, Institute of Arctic Biology, Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK, USA

Introduction

Dimethylsulfide (DMS) is a marine biogenic compound that is important to climate as a precursor of cloud condensation nuclei (CCN). DMS is formed by the degradation of its precursor, dimethylsulfoniopropionate (DMSP). DMSP is produced by marine phytoplankton, and is hypothesized to act as a stress reduction compound (Sunda et al. 2002), grazing deterrent (Wolfe et al. 1997), foraging cue (Nevitt et al. 1995), cryoprotectant (Karsten et al. 1996), osmolyte (Kirst 1996), and overflow mechanism for reduced sulfur under nutrient limited-unbalanced growth conditions (Stefels et al. 2007). In Arctic waters, *Phaeocystis* spp. and diatoms are significant contributors to the water column DMSP and DMS budgets (Matrai and Vernet 1997). Degradation of DMSP to DMS can occur in healthy cells via lyase activity, or when DMSP is released into the ocean when a cell has been compromised due to natural senescence or grazing (Stefels et al. 2007). DMSP in the dissolved ocean pool can also be converted to DMS by bacterial activity. Part of the DMS is then transferred to the atmosphere where it can form a variety of oxidized sulfur compounds that aid in the growth of aerosols to CCN, thereby influencing cloud formation and albedo (Bates et al. 1987a; Charlson et al. 1987; Andreae and Rosenfeld 2008). The role of DMS in the Arctic is further compounded by the presence of seasonal sea ice, where ice algae are an important source of DMS for the arctic atmosphere (Levasseur et al. 1994). In the marginal sea ice zones biological productivity is enhanced as melting ice releases ice-bottom algae and detritus into the upper water column, leading to an accumulation of dissolved organic matter (Matrai et al. 2007a). Melting sea ice also helps to maintain the pycnocline that regulates surface DMS concentrations through the supply of organic matter from richer water underneath and the balance between competing DMS sinks, such as, DMS photolysis, bacterial consumption, and degassing to the atmosphere (Galí and Simó 2010).

Simulation experiments conducted by Gabric et al. (2005) suggest that the annual DMS flux from the high Arctic may increase to over 80% of present day levels by 2080 under a tripled CO_2 scenario (Gabric et al. 2005; Qu and Albert 2010). The increase in DMS emissions to the atmosphere could impact the overall radiation budget of the Arctic through changes in cloud albedo. The negative feedback of increased solar radiation backscatter caused by increased cloud albedo may mask the decreased albedo from sea ice loss (Gabric et al. 2005).

For the purposes of this study, the Arctic is defined using the astronomical definition of 66.5°N latitude. Seasonal sea ice does not occur south of 66.5°N between July and October, the months that yield the most spatially-distributed data in the high Arctic (with the exception of April).

Until recently, no efforts have been made to create a climatology for Arctic sea surface DMS concentrations as a stand-alone model. However, sulfur cycling in the skeletal sea ice layer has been modeled by the COSIM (Climate Ocean Sea Ice Modeling) group at the Los Alamos National Laboratory. The COSIM model, currently in press, contains a sea ice DMS component. Some climatologies of global DMS concentrations include the Arctic in overall simulations (e.g. Kettle and Andreae 2000; Kettle et al. 1999; Simó and Dachs 2002), but these do not explicitly treat sea ice as a model variable and are meant primarily to study mean global patterns.

The mechanisms of DMS formation and loss are not well understood, making this potentially important compound a prime candidate for spatial modeling techniques. Spatial modeling is commonly used for describing the distribution of various animal and plant species, and combines Geographic Information Systems (GIS) with complex algorithms developed to handle "messy" ecological data (Elith et al. 2006; Craig and Huettmann 2009; Cushman and Huettmann 2010). TreeNet (Salford Systems, 2002) employs an algorithm known as stochastic gradient boosting, which can make predictions on a target variable via associations with up to 50 environmental (predictor) variables. This powerful algorithm allows calculations without a priori assumptions regarding the variables that influence prediction values (Breiman 2001; Friedman 2002). This may be important when examining a system defined by complex relationships (e.g., DMS). Combining TreeNet with GIS and various freely-available environmental data enables efficient, accurate prediction of Arctic sea surface DMS concentrations.

The objective of this study was to create spatial predictions of DMS concentrations at the sea surface during the boreal summer, using associations with ocean characteristics such as nutrients and sea ice. Furthermore, the relationships between these features and DMS were examined via the variable ranking scheme as determined by TreeNet output.

Data and methods

The Pacific Marine Ecological Laboratory (PMEL) database provided 1028 mixed-layer DMS measurements for the defined Arctic region. These data were filtered to the top 5 m of the ocean surface, leaving 854 data points for use in analysis. The measurements spanned the time period from 1985 to 2008 and were taken by 12 contributors (see Table 4 in Appendix). Data were segregated by the months July through October and projected onto the North Pole Stereographic spatial coordinate system. The months July through October were chosen because the extent of the spatial distribution of the data was generally broader during these months than in all other months with the exception of April. April was not chosen for modeling because it does not closely lead or follow the month of September, when sea ice is at its minimum. It is also during July and August when DMS concentrations in the Arctic atmosphere may be highest (Ferek et al. 1995; Lundén et al. 2010).

Spatial distribution of surface seawater DMS measurements was best for July (in relation to all other months), as measured with a nearly-complete cross section of the Arctic Basin. In August, DMS measurements were only collected between the North Pole and Greenland. Most DMS measurements in September were collected in the Beaufort and Chukchi seas, close to land, but a transect of DMS measurements was also available from Svalbard toward the North Pole. The spatial distribution of DMS measurements for October was of the lowest quality (most highly clustered, with few available measurements), with data around Svalbard and in the Chukchi Sea (Fig. 1). In some cases (e.g., July), some DMS measurements are outside the clusters (i.e., one data point in the Chukchi Sea). TreeNet will only take these data into consideration if information useful for reducing model error can be obtained. If the data point does not add any information, then it is treated as an outlier and down-weighted appropriately in the process. We are confident, however, that with the exception of October these data can be used for such a process because TreeNet is able to deal with 'messy' data (Craig and Huettmann 2009), and because similar methods, e.g. the related Random Forests algorithm, have been used successfully with spatially-irregular data (Wei et al. 2010).

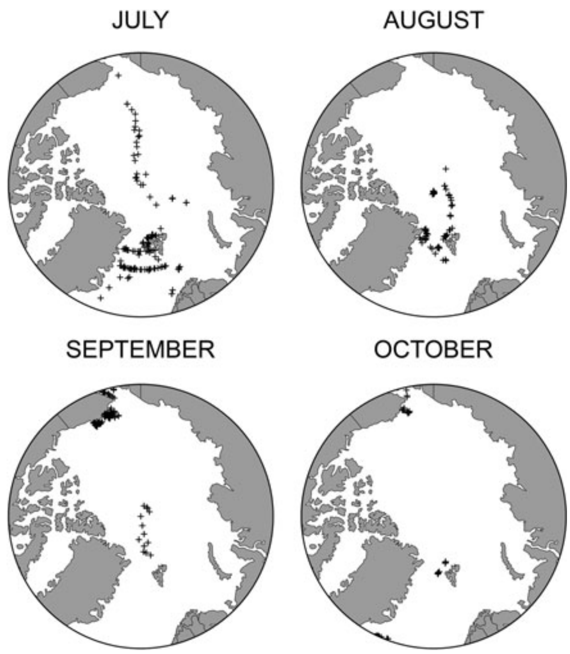


Fig. 1 Spatial distribution of Arctic DMS data from the Kettle database (Kettle and Andreae 2000) for July through October

A list of all environmental data used in this analysis can be found in Table 1. Fourteen environmental variables were used to predict the sea surface DMS concentration. Concentrations of phosphate and nitrate, salinity, and sea temperature at the surface and 10 m depth were downloaded from NOAA's World Ocean Atlas (WOA 2005). For bathymetry, the International Bathymetric Chart for the Arctic Ocean (IBCAO; Holland 2000) was downloaded. Bathymetric slope and distance to shelf (defined as 100 m depth) were calculated in ArcGIS 10.0 from the obtained bathymetry. Mean sea ice concentration was downloaded from the National Snow and Ice Data Center and ice edge, the outer boundary of ice, was located where at least 15% ice concentration occurred. Monthly climatologies of solar radiation dose were calculated according to Vallina and Simó (2007) from mixed-layer depth and irradiance at the top of the atmosphere. Also following Vallina and Simo (2007), mixed-layer depth was defined as the depth where the temperature difference was 0.1°C from the temperature at 5 m. Solar irradiation was estimated using the equations of Brock (1981) with an applied atmospheric loss factor of 50%. Bathymetry, bathymetric slope, and distance to shelf were considered to be invariant over the time period examined.

Table 1 Data sources used for modeling procedure

Dataset	Source	Spatial resolution	Temporal resolution	Description	Units
Dimethylsulfide (DMS)	Pacific marine ecological laboratory (http://www.saga.pmel.noaa.gov)	Points	Points	Point observations of DMS	Nanomoles l^{-1}
Salinity (surface, 10 m)	World Ocean Atlas (www.nodc.noaa.gov)	1°	Monthly	Climatology of salinity	ppm
Nitrate concentration (surface, 10 m)	World Ocean Atlas (www.nodc.noaa.gov)	1°	Monthly	Climatology of nitrate concentrations	Micromoles l^{-1}
Phosphate concentration (surface, 10 m)	World Ocean Atlas (www.nodc.noaa.gov)	1°	Monthly	Climatology of phosphate concentrations	Micromoles l^{-1}
Temperature (surface, 10 m)	World Ocean Atlas (www.nodc.noaa.gov)	1°	Monthly	Climatology of mean temperature	°C
Bathymetry	International Bathymetric Chart of the Arctic Ocean (IBCAO)	2 km	Invariable	Arctic ocean bathymetry	Meters
Solar radiation dose	Calculated as per Vallina and Simo (2007)	1°	Monthly	Calculated from irradiance from the top of the atmosphere and mixed-layer depth	Watts m^{-2}
Monthly sea ice concentration	National Snow and Ice Data Center (www.nsidc.org)	26 km	Monthly	% Sea ice concentration age from SSM/I satellite dsta	%
Distance to ice edge	Calculated in ArcGIS software from mean % sea ice concentration	26 km	Monthly	Distance to sea ice edge (edge defined as 15% sea ice concentration)	Meters
Bathymetric slope	Calculated in ArcGIS software from bathymetry	2 km	Invariable	Slope of ocean floor	Degrees
Distance to shelf	Calculated in ArcGIS software from bathymetry	2 km	Invariable	Distance to shelf (defined as 100 m isodepth)	Meters

Modeling was performed using TreeNet v 2.0, available in the Salford Data Miner suite. This software creates a series of regression trees, using statistical boosting to minimize error. The tree with the lowest error value is used to create predictions of the target variable (Friedman 2002). This differs from classic direct correlation analysis in that there is no emphasis on statistical significance, only on the ability of each variable to lower the amount of error in predictions. Regression tree splits are calculated based on a predictor's ability to lower the variability in observed measurements, thus creating a series of "if", "then" (i.e., conditional) statements to create a rule set for predictions.

To prepare the data for use in TreeNet, the DMS measurements were associated with the environmental data via spatial overlays (i.e., points were associated with pixel values of environmental variables). Spatial overlays were performed with the freely-available Geospatial Modeling Environment (GME; www.spatialecology.com) software.

To insure a good model selection, "battery" tests were performed to choose the settings that provided the best assessment as determined by mean absolute deviation (MAD) provided in TreeNet. Data were randomly split into training and assessment datasets (70 and 30%, respectively) in TreeNet because previous experience has shown that a 70/30 split in data tends to provide robust predictions and assessment (Humphries et al., unpublished). The 70% subset was used to determine the associations between the target and predictors (i.e., "training"). TreeNet then performed predictions on the remaining 30%, and compared the predictions to the observed values to obtain r^2 and MAD values.

A regular grid (i.e. a lattice) of points was created in ArcGIS 10.0 at a resolution of $1 \times 1°$ and was associated with environmental variables via spatial overlays. Using the associations determined in the training phase of TreeNet, DMS concentrations were predicted to the regular grid. The point predictions were interpolated using the inverse distance-weighted function in ArcGIS to generate maps.

TreeNet determines which variables (e.g. nitrate concentration) are the most important target predictors by examining the contribution of each environmental variable to the conditional statements (rules) that are determined. Predictors are scored on a relative index from 0 to 100 in which 100 is the most important predictor. TreeNet also outputs plots of the partial dependence of predictor variables on the target. The partial-dependence plots are visual representations of each variable's contribution to the prediction of DMS concentrations (see Fig. 7 in Appendix for an example from this study). These partial dependence plots are not easy to interpret; therefore, to further examine the relationship between DMS and various important predictors, we plotted the concentrations of measured DMS versus the top two predictors (environmental variables).

Results

Output sea surface DMS maps in a North Pole Stereographic projection are shown in Fig. 2. In July, hot-spots of DMS were found in the Greenland and Chukchi seas. Concentrations in these areas ranged from 4.0 to 7.0 nM and occurred outside the sea ice periphery, while concentrations under the sea ice were low (between 0.4 and 2.0 nM). This trend was also seen in August when higher concentrations of DMS (>5 nM) were predicted near the ice edge, while lower concentrations of DMS were predicted under the sea ice. Overall concentrations of measured DMS in September were low ($\sim$0.9 nM; Figs. 2, 3), leading to lower predicted DMS concentrations than for July and August. Higher DMS concentrations for September were predicted around the sea ice edge, but the DMS concentration in hot-spots was only 2–2.5 nM. There is a clustering of DMS measurements around northern Alaska in September which likely affected the assessment and model output by lowering the variation in the underlying environmental associations. Taking DMS measurements between the North Pole and Greenland may have increased variation and therefore the ability of the model to capture relationships in the data, resulting in better model performance in September than for October. Very low DMS concentrations were predicted for the month of October, corresponding to the low concentrations of measured values in this month ($\sim$0.5 nM). DMS concentrations of $\sim$0.5 nM occur primarily around the periphery of the ice and into the Greenland Sea, with areas of higher concentration under the ice. The best predictor variable for this month was distance to shelf; this explains why the output map resembles the bathymetry. Also, the October data are clustered around a few, very small areas. Due to this spatial

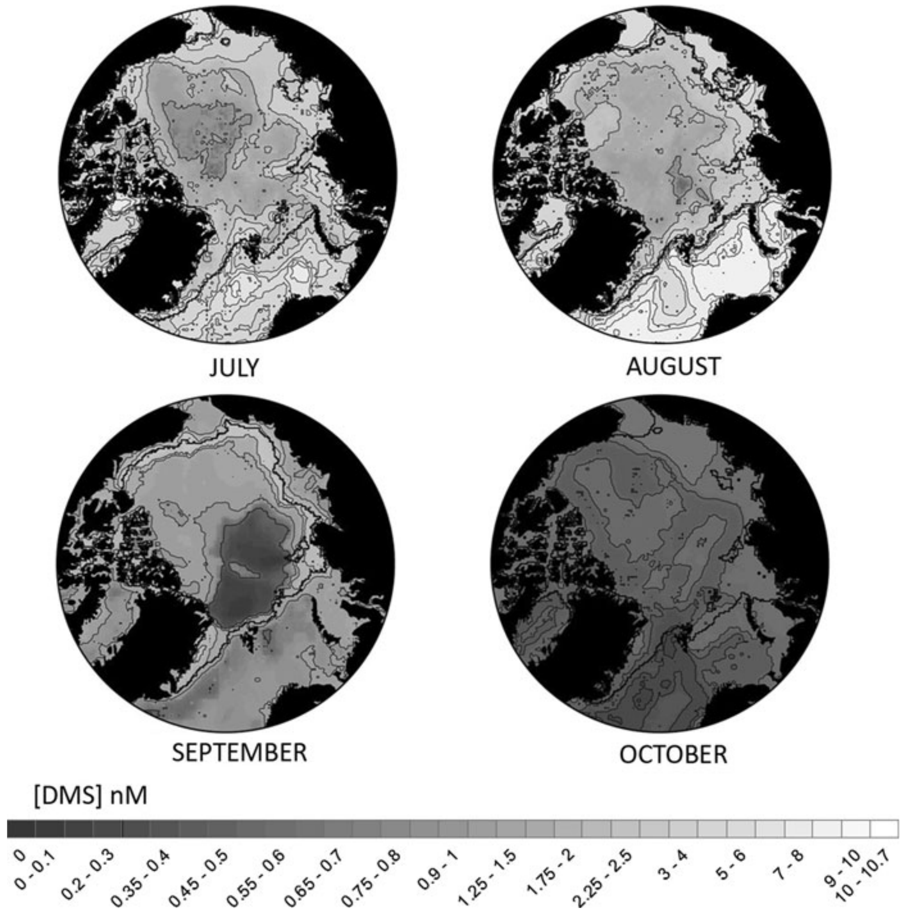


Fig. 2 Monthly mean sea surface DMS concentration (nM) for the months July, August, September, and October, with ice edge (defined as >15% sea ice concentration) overlain

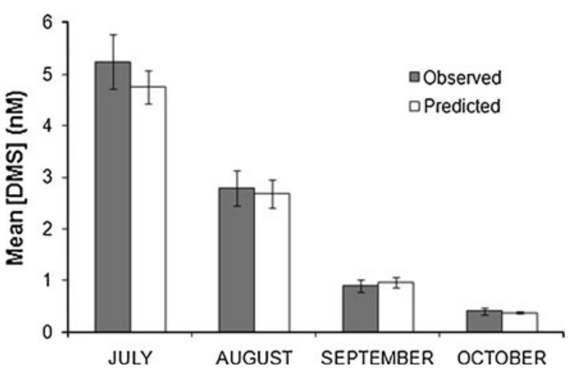


Fig. 3 Mean sea surface DMS concentrations for observed and predicted values with 95% confidence intervals

distortion, October results are probably not accurate (Fig. 1; also note the very low r^2 value shown in Table 2). This illustrates the effects of low-quality spatial data (i.e., very few measurements at spatially-irregular intervals) on data-driven modeling efforts.

Table 2 Assessment values of models from TreeNet as determined by an independent random subset of 30% of observed data

Month	r^2	Mean absolute deviation	# Training points	# Testing points
July	0.61	1.86	209	65
August	0.87	0.70	187	58
September	0.66	0.33	168	54
October	0.37	0.12	90	23

Assessment values for the models with the lowest MAD values are listed in Table 2, as well as the number of training and assessment points used in the analysis. The measured data explained 87% of the variance for August, while 66, 61, and 37% of the variance in the data were explained for September, July, and October, respectively. MAD values were 1.86, 0.70, 0.33, and 0.12 nM for July, August, September, and October, respectively. In order to

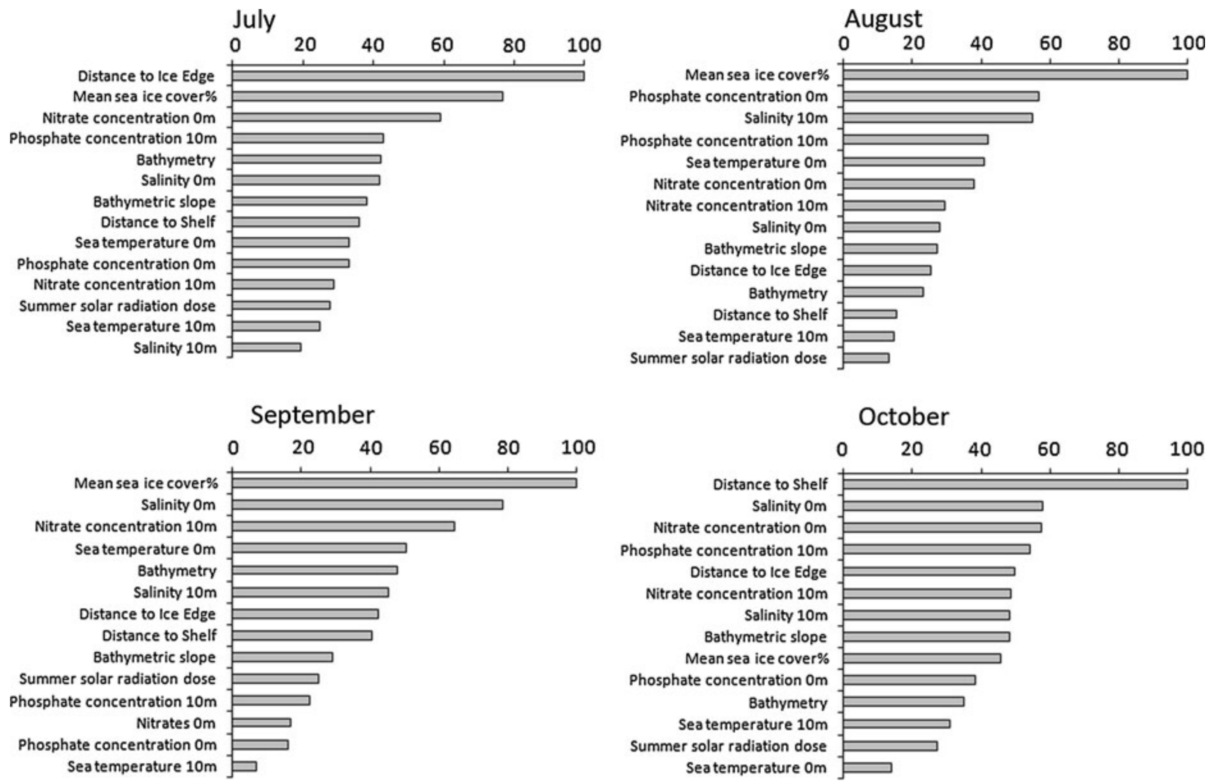


Fig. 4 Relative importance of all variables as determined by TreeNet output

examine whether models were over- or under-predicting values, observed measurements were associated with predicted values via a spatial overlay, and means were compared (Fig. 3). July DMS concentrations were under-predicted; the predicted mean was ~0.49 nM less than the observed mean. The 4.47 standard deviation of observed values in July may explain the slight discrepancy between means. August, September, and October all show high overlap in the confidence intervals between observed and predicted values; predicted means were less than observed by 0.11, 0.06, and 0.03 nM, respectively. September was the only month that slightly over-predicted DMS concentrations.

The relative importance (influence or relevance) of predictors of DMS concentrations for the top models is shown in Fig. 4. The relative importance is a scale from 0 to 100 which represents the ranking of the variables for prediction (e.g., a value of 100 is the most important). For July, August, and September, mean sea ice concentration was among the top two predictor variables. Distance to shelf was the top predictor for October; however, due to the spatial distribution of the data (which degraded the modeling procedure) we cannot be confident of this result. Due to the potential for variable importance to "flip" (e.g., on rare occasions, variables with an initially low importance may become highly important when model settings are changed), which can occur with highly-correlated variables, the robustness of relationships was tested by performing a model run across the entire temporal domain. It was found that mean sea ice concentration was identified as the top predictor, followed by nitrate concentrations at the surface and distance to ice edge (Table 3); similarly, mean sea ice concentration emerges as one of the top three variables in monthly results for July, August, and September, while nitrate concentrations at the surface were among the top three variables for July and October. Galí and Simó (2010) report that solar radiation dose may be an important factor in determining DMS concentrations in the Arctic, however, it was not a highly ranked predictor for any of the model runs. Salinity was one of the top three predictor variables for September and October, but it was not examined further as the relationship was not supported in the model run across the entire

Table 3 Relative importance of variables based on a model run using all data from July, August, September and October

Variable	Relative importance
Mean sea ice concentration	100.00
Nitrate concentration (surface)	95.82
Distance to ice edge	85.03
Sea surface temperature	75.76
Phosphate concentration (surface)	70.17
Phosphate concentration (10 m)	61.27
Surface salinity	59.74
Bathymetry	57.40
Nitrate concentration (10 m)	54.04
Bathymetric slope	52.52
Salinity (10 m)	44.49
Distance to shelf edge	42.46
Temperature (10 m)	40.84
Solar radiation dose	35.02

temporal domain (Table 3). Distance to ice edge was among the top three predictor variables for July only, and it was not examined further, in contrast to surface nitrate concentrations and mean sea ice concentration.

The relationships between measured DMS and mean sea ice concentration as well as between measured DMS and surface nitrate concentrations are shown in Figs. 5a and 6a, respectively. Mean sea ice concentration was binned into groups of 0, 5–20, 30–50, 60–90, and 95% sea ice concentration; these bins were chosen due to the categorical nature of the data. When these bins were plotted against mean observed DMS in each bin, it was found that as sea ice concentration increased, concentrations of DMS decreased. Mean DMS concentration under 95% sea ice concentration for all months was ~1 nM.

DMS and surface nitrate concentrations were plotted and fit using a LOWESS smoother to illustrate the potential relationship between these variables. The LOWESS smoothed line is nonlinear and may be more useful for determining relationships between variables than simple linear smoothers (Cleveland 1979); however, statistics on the fit of the line are generally uninformative in this case and we use the line simply as an illustrative guide. At nitrate concentrations below ~2.0 mmol/m^3, DMS concentrations tend to be higher, showing a slightly positive trend. This relationship does not appear to exist above nitrate concentrations of ~4.0 mmol/m^3. Once this upper

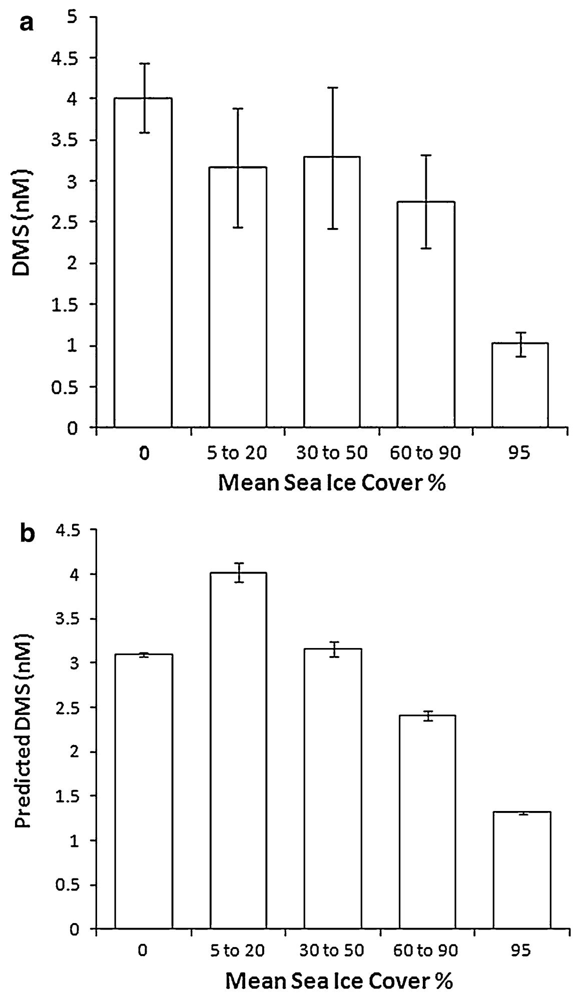
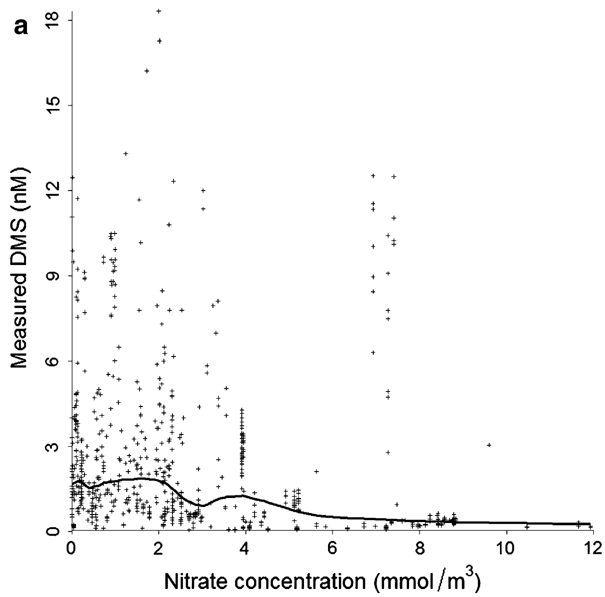


Fig. 5 Mean sea ice concentration (%) plotted against mean measured DMS concentrations for July through October (**a**), and mean sea ice concentration (%) plotted against mean predicted DMS concentrations for July through October (**b**)

threshold is reached, there is a negative relationship between nitrate and DMS concentrations.

The relationships between predicted values of DMS and mean sea ice and surface nitrate concentrations remain robust (in comparison to the relationships between measured DMS, mean sea ice concentration, and surface nitrate concentrations) across the entire spatiotemporal domain (Figs. 5b, 6b). In general, a similar pattern is seen in both Figures compared to Figs. 5a and 6a. The relationship between predicted DMS and sea ice in Fig. 5b differs slightly from Fig. 5a but only because the mean values of DMS differ between Fig. 5a, b. The general negative trend between DMS and sea ice is still apparent, with a



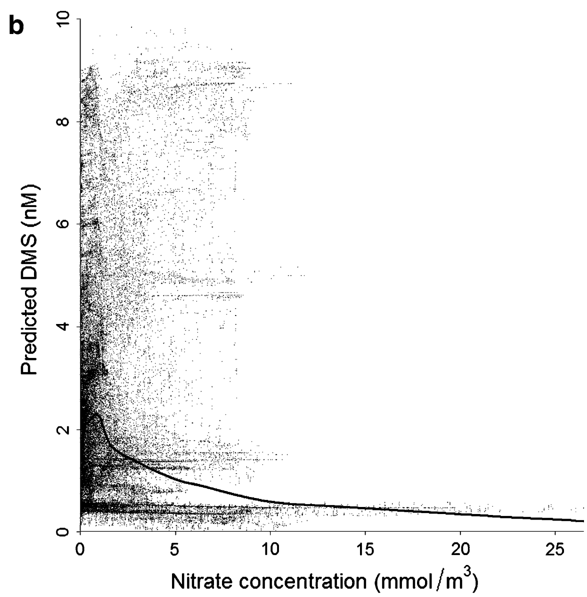


Fig. 6 Surface nitrate concentrations from climatology plotted against measured sea surface DMS concentrations (nM) fitted with a LOWESS smoothed line (**a**), and surface nitrate concentrations from climatology plotted against predicted sea surface DMS concentrations (nM) fitted with a LOWESS smoothed line (**b**)

minimum mean of 1.31 nM at 95% sea ice concentration. Predicted DMS against surface nitrate concentrations in Fig. 6b differs slightly from Fig. 6a in that the threshold of the negative relationship at low concentrations of predicted DMS is approximately 2.0 nM in Fig. 6b. Both surface nitrate concentrations and sea ice exhibited robust relationships with predicted DMS values; this was not surprising, because both surface nitrate concentrations and sea ice were considered top predictors when a model run was performed over the entire temporal domain. Figure 7 in Appendix shows the TreeNet partial dependence plot for nitrate concentrations versus the partial dependence of predictions of DMS. Partial dependence is an indicator of the strength of potential mechanistic relationships with DMS as a function of local values of the predictor variable (e.g., nitrate concentrations). For example, when nitrate concentrations are below 2.0 mmol/m^3, DMS predictions tend to be higher, while above nitrate concentrations of 2.0 mmol/m^3 they tend to be lower.

Discussion

For the 4 months examined, 854 data points were available in our study region. Of those 854 measurements only 113 were available for October, and the October data points were not evenly distributed over the study region. The October associations that are determined from TreeNet are, therefore, based on only a small number of data from very confined regions. The spatial distribution of these data is important because the clustering of data points will affect variation in the associations with environmental layers; because of the spatial resolution of the environmental variables, heavy clustering will lead to many DMS measurements being associated with very few environmental feature values. High clustering and few data lead to lowered variation, which degrades the modeling results. It is, therefore, unlikely that these October DMS data are truly representative of patterns across the entire Arctic region, and it is not surprising that an r^2 value of 0.37 was obtained for October. Based on these findings we are not confident that the model has produced accurate October DMS patterns. To alleviate this problem, a more spatially-uniform sampling of the Arctic in October must be undertaken.

October DMS concentrations are relatively low, with a mean of only 0.9 nM. These low measured concentrations have, therefore, led to low predicted concentrations. It is possible that higher concentrations of DMS would be observed if more measurements, with a more uniform spatial distribution, were made during October. It is also possible that these low concentrations could be confirmed as a general trend. Assessments of July, August, and September showed higher r^2 values. The spatial distributions of data for these months, as well as the number of data available, are much better than for October. This spread of data has enabled us to more accurately capture the relationships between DMS and the underlying predictor variables. It has also led to predicted outputs that seem much more realistic than those for October. Statistical modeling carries a limitation of being driven wholly by the data and, therefore, we are making generalizations about the predictive performance of environmental variables across the entire Arctic based only upon information from those areas that are represented by our data. If measurement data are not representative of the entire area of study, then misrepresentation of the predictions can occur.

For the months of July through September, mean sea ice concentration was determined to be at least the second most important variable. As a way of confirming these findings, the model was run using all data. Mean sea ice concentration emerged as the most important variable when the analysis was performed this way. The relationship between sea ice concentration and measured DMS values was examined further, and it was found that as sea ice concentration increased, mean surface DMS decreased. Prediction maps (except for October) show that the highest concentrations of DMS occur in the seasonal ice zone (in July and August) and close to the ice edge (in September) where elevated concentrations of DMS have been reported (Ferek et al. 1995; Matrai and Vernet 1997). It is thought that the melting of sea ice (in months like July and August) sets the stratification of the vertical column, which drives biological productivity and exposure to solar radiation, and hence, influences DMS concentrations (Galí and Simó 2010). Also, Matrai et al. (2007b) report that there is a high potential benefit to examining sea ice (i.e., ice edge) as a predictor of high DMSP concentration. These findings are corroborated by the results obtained in this modeling study.

Surface nitrate concentrations were also determined to play an important role in predicting DMS concentrations, though nitrate was not as strong a predictor as mean sea ice concentration. Nitrate concentration was among the top three variables for all months except July and October, and was ranked the second most important variable in the run that used all data. Nitrate affects apparent quantum yields (AQYs) and rate constants for DMS photolysis (Bouillon and Miller 2004; Toole et al. 2004). Deal et al. (2005) suggest that although the AQY may increase with increasing nitrate concentrations, dissolved organic matter, rather than nitrate, is likely driving the photolysis of DMS in the Bering Sea. In Antarctic waters, observed DMS photolysis rates increased with added nitrate, although chromophores other than nitrate were primarily responsible for the photolysis of DMS (Toole et al. 2004). The results obtained here suggest a two-phase relationship between surface nitrate concentrations and DMS (both observed and predicted values). This may be due to nitrate acting as a limiting nutrient in the system, perhaps resulting in stronger extracellular release of DMSP (Matrai et al. 2007a) with subsequent conversion to DMS. When nitrate concentrations are higher than ~ 1 mmol/m^3 it may not be limiting, and DMS concentrations do not follow any noticeable pattern with nitrate. A similar two-phase relationship was found between DMSP and nitrate concentrations by Nian-zhi et al. (2003). It could be, therefore, that the relationship determined by Nian-Zhi et al. (2003) is due to the effects of nitrate on extracellular release of DMSP.

The highest predicted DMS concentrations for July, August, and September were found in the Greenland and Barents seas. These areas are generally characterized by lower salinities (due to melting sea ice) and warmer sea surface temperatures than the rest of the Arctic basin, which stratifies the water column in the summer months (Loeng 1991). At low nitrate concentrations (<4 mmol/m^3), measured DMS concentrations can be as high as 18 nM (Fig. 6a). It is possible that nitrate may be a limiting nutrient in these areas (see Lara et al. 1994 for the Greenland Sea), and slight changes in nitrate concentration would place stress on DMS producers, increasing extracellular release of DMSP. Under Arctic sea ice (in the Arctic Basin) DMS concentrations are lower (Leck and Persson 1996), most likely a direct result of the lack of sunlight to promote phytoplankton growth. Although, it is interesting to note that in late winter high concentrations of DMS have been measured in the Barents Sea and suggested to have a heterotrophic source (Matrai et al. 2007b). The Chukchi Sea is characterized by mid-level DMS predictions (1.5–6.0 nM). This sea is an entry point for warmer water into the Arctic Basin, but due to dilution with nutrient-poor water and biological utilization, primary production is lower than in the Greenland or Barents seas (Cooper et al. 1997), possibly factoring into the generally lower DMS concentrations in the Chukchi than in the Greenland or Barents seas. Dominant phytoplankton taxa may be another important factor. In peripheral arctic seas, such as the Greenland, Bering, and Barents seas, colonies of the important DMS/DMPp-producer *Phaeocystis pouchetti* can reach high biomasses, whereas documented phytoplankton blooms in the Arctic Ocean are dominated by diatoms (Tremblay and Gagnon 2009). However, Matrai and Vernet (1997) suggest that the physiological stage of the bloom may be even more important than species dominance in Arctic waters. Ice algae also produce significant amounts of particulate DMSP and dissolved DMSP plus DMS (Levasseur et al. 1994), which may at least in part explain the higher DMS concentrations observed and predicted close to the ice edge and in the seasonal ice zone (perhaps due to seeding of the ice edge phytoplankton bloom).

Due to the influence of DMS on climate (Bates et al. 1987b; Charlson et al. 1987), and the role of the Arctic in global climate systems (McGuire et al. 2006), the ability to accurately predict potential changes in Arctic DMS concentrations could fill a major gap in global climate scenarios. Spatial modeling exercises similar to the technique used in this study are at a disadvantage when asked to predict the future for an area where no assessment data exist (Huettmann and Gottshalk 2010). To perform an accurate future prediction, those associations determined by TreeNet (i.e., the conditional statements) would have to be applied to future scenarios for the predictor variables used. For example, we would require future scenarios of variables such as sea ice concentration and nitrate concentration in order to predict future DMS concentrations. Future scenarios of variables like nitrate concentrations, phosphate concentrations, sea ice concentration and sea surface temperature (for example) may be obtained from climate earth systems models. It may therefore be possible with these scenarios to build the conditional statements from TreeNet directly into earth system simulations to obtain future predictions of DMS concentrations. Also, models created in this study only span the months of July through October, and in order to obtain year-round DMS for future scenarios, model predictions would have to be extended to all months. However, for this to be successful, further DMS sampling in the Arctic must occur at spatially- and temporally-regular intervals.

When using spatial data to create predictions, as in this study, there are large uncertainties, including accuracy of observations and scale of data. Here, it is assumed that all measurements and global positioning system (GPS) coordinates are accurate. In reality, differences may exist in both measurement techniques and in shipboard GPS systems. These differences raise questions about the comparability of data between multiple studies. This could be corrected for in the modeling process by determining the techniques used in each study and then using comparable measurements only. We did not perform such a correction in this study because the number of data available for the Arctic was already so low. Also, machine learning techniques, such as TreeNet, are rather robust when dealing with erroneous data, and are able to determine the prominent signals that occur in data beyond any statistical noise (Craig and Huettmann 2009). Latitude and longitude values associated with shipboard GPS also carry an associated error. This can pose the largest problem when using high-resolution environmental variables because slight geographic errors could place a data point in the wrong pixel. Because the environmental variables used in this study were relatively coarse, it was assumed that GPS error would have little or no effect on the results.

The model predictions for July were generally underestimated in comparison to observations. Model predictions for August through October, however, seemed to be on the same order as observed values (Fig. 2). The higher variability of observed data in July led to model predictions that were, on average, farther away from observed values than in the other months in which observed variability was much lower. This is a known problem with statistical models; greater variability in the original data causes models to make predictions that are further away from observed measurements. It is possible that the variability that has not been captured by these models could be explained by adding more environmental predictors. For the purpose of this study, biological components which may control DMS processes were ignored because these variables are not readily available in a suitable format. The distribution and concentration of phytoplankton as well as species composition may be important in determining DMS concentrations (Levasseur et al. 1994; Gosselin et al. 1997; Vila-Costa et al. 2006; Matrai et al. 2007b). If biological features were included in future runs of these models, more of the variability in the observed data could be accounted for.

Because of the nature of an a priori method of analysis (like the exercise in this study), there is the danger of misinterpreting mechanisms in the system. If predictive ability is high, and there is little information on how a variable would logically affect the target variable, then further research is required. This approach could be described as using inductive thought processes (as in this study) to drive deductive reasoning (experimental research). The interest in inductive modeling is primarily focused on the output and assessment of the model, with some allowance for mechanistic inferences. In the case of this research, we have made inferences about the effect of nitrate concentrations and sea ice on DMS concentrations. Without experimental research, there is no optimal way to determine whether our findings are due to random correlation, or to actual mechanisms. However, we are confident in our interpretation due to the effects of sea ice and nitrate concentrations on DMS that have been previously elucidated by other studies (Matrai et al. 2007a, b; Galí and Simó 2010).

The Arctic ecosystem is changing quickly, and the need for quickly produced, accurate model scenarios is obvious. At the same time, these models must be able to take into account the many environmental conditions, and their interactions, which control the systems in question. Spatial modeling techniques can deal with messy environmental data; these models run quickly, and they can help to elucidate complicated interactions. For this study, the relationships between ocean characteristics and DMS in the high Arctic during summer months were examined using spatial modeling techniques that are commonly implemented in ecological niche modeling. Sea ice concentration and nitrate concentrations at the sea surface seem to be important in accounting for variance in sea surface DMS concentrations. These relationships can be applied to future scenarios to examine how DMS concentrations might change. In order to ensure that these models are accurate, and to increase overall knowledge of systems that control DMS concentrations, further DMS measurements are needed. The Arctic is under-sampled both spatially and temporally, and therefore our ability to predict changes in this climatically-important area of our planet is limited. We recommend that further DMS measurements, with a more uniform spatial and temporal distribution, be made in the Arctic. The application and evaluation of spatially-explicit models like these will help us to understand biogeochemical systems in the Arctic and, potentially, how Arctic climate will change. Such understanding will offer the potential for scientifically-driven decision making to determine the future of this important region.

Acknowledgments We would like to thank all contributors to the DMS database (Please see Table 4 in Appendix for a full list of contributors to Arctic DMS measurements and associated references). Sincere thanks go to Dr. S. Vallina for providing mixed-layer depth and irradiance values at the top of the atmosphere layers which were used to calculate solar radiation dose. We would also like to thank T. Alton and C. O'Connor for thoughtful comments and draft editing, I. Rutzen for original data manipulations, and the EWHALE lab and Salford Systems Ltd. for support with software licensing. Partial funding for this project was from NSF ARC-0652838. This is EWHALE lab publication 110.

Appendix

See Table 4 and Fig. 7.

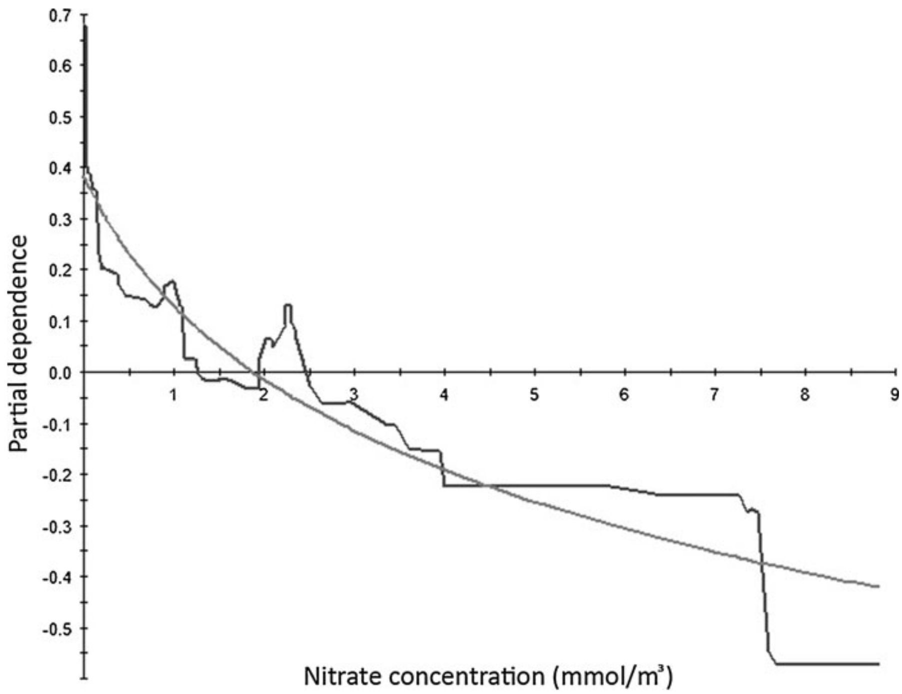
Table 4 Contributors of DMS measurements to the PMEL database initiated by Kettle et al. (2000)

Contributor	Platform	Region	Period	Reference(s)
Bates	Discoverer	Pacific, Arctic	Sep–Oct 1985	Bates et al. (1990, 1987a, b)
Staubes	Polarstern	Greenland Sea	Jul–Aug 1990	Staubes-Diederich (1992, 1993a, b)
Leck	Oden	Arctic	Aug–Oct 1991	Leck and Persson (1996)
Sharma	Polar Sea	Atlantic, Arctic, Pacific	Jul 1994	Sharma et al. (1999)
Leck	Odin	Arctic	Jul–Aug 1996	Unpublished

Table 4 continued

Contributor	Platform	Region	Period	Reference(s)
Belviso	Polarstern	Atlantic	Oct 1996	Belviso et al. (2000)
Lee, DiTullio and Hutchins, Leblanc and Wilhelm	Seward Johnson	North Atlantic Ocean	July 2005	Unpublished
Simo	Hesperides	Arctic	July 2007	Unpublished
Matrai	Oden	Arctic Ocean	July 2001	Matrai et al. (2007a, b)

Fig. 7 Partial dependence plot of surface nitrate concentrations against the function of DMS predictions fitted with a logistic curve to show partial dependence pattern



References

Andreae MO, Rosenfeld D (2008) Aerosol-cloud-precipitation interactions: the nature and sources of cloud active aerosols. Ear Sci Rev 89:13–41

Bates TS, Charlson RJ, Gammon RH (1987a) Evidence for the climatic role of marine biogenic sulphur. Nat 329:319–321

Bates TS, Cline JD, Gammon RH, Kelly-Hansen SR (1987b) Regional and seasonal variations in the flux of oceanic dimethylsulfide to the atmosphere. J Geophys Res 92(6): 2930–2938

Bates TS, Johnson JE, Quinn PK, Goldan PD, Kuster WC, Covert DC et al (1990) The biogeochemical sulfur cycle in the marine boundary layer over the northeast Pacific Ocean. J Atmos Chem 10:59–81

Belviso S, Morrow R, Mihalopoulos N (2000) An Atlantic meridional transect of surface water dimethyl sulfide concentrations with 10-15 km of horizontal resolution and close examination of ocean circulation. J Geophys Res 105(Dll):14423–14431

Bouillon R, Miller W (2004) Determination of apparent quantum yield spectra of DMS photo-degradation in an in situ iron-induced northeast Pacific Ocean bloom. Geophys Res Lett 31:1–4

Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16(3):199–215

Brock TD (1981) Calculating solar radiation for ecological studies. Eco Mod 14(1):1–19

Charlson RJ, Lovelock JE, Andreae MO, Warren SG (1987) Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. Nat 326:655–661

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836

Cooper LW, Whitledge TE, Grebmeier JM, Weingartner T (1997) The nutrient, salinity, and stable oxygen isotope composition of Bering and Chukchi Seas waters in and near the Bering Strait. J Geophys Res 102:12563–12573

Craig E, Huettmann F (2009) Using “Blackbox” Algorithms such as TreeNet and Random Forests for Data-Mining and for finding meaningful patterns, relationships and outliers in complex ecological data: An overview, an example using Golden Eagle satellite data and an outlook for a promi. In: Wang HF (ed) Intelligent data analysis: developing new methodologies through pattern discovery and recovery. Idea Group Inc., Hershey

Cushman SA, Huettmann F (2010) Spatial complexity, informatics, and wildlife conservation. Springer, New York

Deal C, Kieber DJ, Toole DA, Stamnes K, Jiang S, Uzuka N (2005) Dimethylsulfide photolysis rates and apparent quantum yields in Bering Sea seawater. Cont Shelf Res 25:1825–1835

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A et al (2006) Novel methods improve prediction of species distributions from occurrence data. Ecography 2:129–151

Ferek RJ, Hobbs PV, Radke LF, Herring JA, Sturges WT, Cota GF (1995) Dimethyl sulfide in the arctic atmosphere. J Geophys Res 100(D12):26093–26104. doi:10.1029/95JD02374

Friedman J (2002) Stochastic gradient boosting. Comp Stat Data Anal 38(4):367–378. doi:10.1016/S0167-9473(01)00065-2

Gabric AJ, Qu B, Matrai P, Hirst AC (2005) The simulated response of dimethylsulfide production in the Arctic Ocean to global warming. Tellus B 57(5):391–403. doi:10.1111/j.1600-0889.2005.00163.x

Galí M, Simó R (2010) Occurrence and cycling of dimethylated sulfur compounds in the Arctic during summer receding of the ice edge. Mar Chem 122:105–117

Gosselin M, Levasseur M, Wheeler P, Horner R, Booth B (1997) New measurements of phytoplankton and ice algal production in the Arctic Ocean. Deep-Sea Res Pt II 44(8):1623–1644. doi:10.1016/S0967-0645(97)00054-4

Holland D (2000) Merged IBCAO/ETOPO5 topography (AOMIP). Center for Atmosphere–Ocean Studies (CA/OS) of the Courant Institute of Mathematical Sciences Tech. Rep. Available online at http://efdl.cims.nyu.edu/project_aomip/forcing_data/topography/merged/overview.html

Huettmann F, Gottshalk T (2010) Simplicity, model fit, complexity and uncertainty in spatial prediction models applied over time: we are quite sure, aren't we? In: Drew CA, Wiersma Y, Huettmann F (eds) Predictive species and habitat modeling in landscape ecology: concepts and applications. Springer, New York

Karsten U, Kuck K, Vogt C, Kirst GO (1996) Dimethylsulfoniopropionate production in phototrophic organisms and its physiological function as a cryoprotectant. In: Kiene RP, Visscher P, Kirst GO, Keller M (eds) Biological and environmental chemistry of DMSP and related sulfonium compounds. Plenum Press, New York

Kettle AJ, Andreae MO (2000) Flux of dimethylsulfide from the oceans: a comparison of updated data sets and flux models. J Geophys Res 105(D22): 26,793–26,808

Kettle AJ, Andreae MO, Amouroux D, Andreae TW, Bates TS, Berresheim H et al. (1999) A global database of sea surface dimethylsulfide (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month. Global Biogeochemical Cycles 13(2):399–444

Kirst GO (1996) Osmotic adjustment in phytoplankton and macroalgae. In: Kiene RP, Visscher P, Kirst GO, Keller M (eds) Biological and environmental chemistry of DMSP and related sulfonium compounds. Plenum Press, New York

Lara RJ, Kattner G, Tillmann U, Hirche H-J (1994) The north east water polynya (Greenland Sea) II. Mechanisms of nutrient supply and influence on phytoplankton distribution. Polar Biol 14:483–490

Leck C, Persson C (1996) The central Arctic Ocean as a source of dimethyl sulfide seasonal variability in relation to biological activity. Tellus 48:156–177

Levasseur M, Gosselin M, Michaud S (1994) A new source of dimethylsulfide (DMS) for the arctic atmosphere: ice diatoms. Mar Biol 121(2):381–387. doi:10.1007/BF00346748

Loeng H (1991) Features of the physical oceanographic conditions of the Barents Sea. In: Sakshaug E, Hopkins CCE, Oritsland NA (eds) Proceedings of the Pro Mare Symposium on Polar Marine Ecology, Trondheim, 12–16 May 1990, Polar Research IO(1)

Lundén J, Svensson G, Wisthaler A, Tjernström M, Hansel A, Leck C (2010) The vertical distribution of atmospheric DMS in the high Arctic summer. Tellus B 62(3):160–171. doi:10.1111/j.1600-0889.2010.00458.x

Matrai P, Vernet M (1997) Dynamics of the vernal bloom in the marginal ice zone of the Barents Sea: Dimethyl sulfide and dimethylsulfoniopropionate budgets. J Geophys Res 102(C10):22965–22979. doi:10.1029/96JC03870

Matrai P, Tranvik L, Leck C, Knulst JC (2007a) Are high Arctic surface microlayers a potential source of aerosol organic precursors? Mar Chem 108(1–2):109–122. doi:10.1016/j.marchem.2007.11.001

Matrai P, Vernet M, Wassmann P (2007b) Relating temporal and spatial patterns of DMSP in the Barents Sea to phytoplankton biomass and productivity. J Mar Syst 67(1–2):83–101. doi:10.1016/j.jmarsys.2006.10.001

McGuire D, Chapin FS, Walsh JE, Wirth C (2006) Integrated regional changes in arctic climate feedbacks: implications for the global climate system. Ann Rev Environ Resour 31(1):61–91. doi:10.1146/annurev.energy.31.020105.100253

Nevitt GA, Veit RR, Karieva P (1995) Dimethyl sulphide as a foraging cue for Antarctic procellariiform seabirds. Nature 376:680–682

Nian-Zhi J, Cheng-Zheng L, Hua-Sheng H, Harada S, Koshikawa H, Watanabe M (2003) Dynamics of dimethylsulfide and dimethylsulfoniopropionate produced by phytoplankton in the Chinese seas—distribution patterns and affecting factors. Acta Botanica Sin 45(7):774–786

Qu B, Albert GJ (2010) Using genetic algorithms to calibrate a dimethylsulfide production model in the Arctic Ocean. Chin J Oc Lim 28(3):573–582. doi:10.1007/s00343-010-9062-x

Salford Systems (2002) TreeNet version 2.0 data mining software documentation, San Diego. http://www.salford-systems.com/. Accessed 20 Sept 2010

Sharma S, Barrie L, Plummer D, McConnell JC, Brickell PC, Levasseur M et al. (1999) Flux estimation of oceanic dimethyl sulfide around North America. J Geophys Res 104(D17):21327–21342. doi: 10.1029/1999JD900207

Simó R, Dachs J (2002) Global ocean emission of dimethylsulfide predicted from biogeophysical data. Glob Biogeochem Cycles 16(4):1078. doi:10.1029/2001GB001829

Staubes R, Georgii HW (1993a) Biogenic sulfur compounds in seawater and the atmosphere in the Antarctic region. Tellus B 45:127–137

Staubes R, Georgii HW (1993b) Measurements of atmospheric and seawater DMS concentrations in the Atlantic, the Arctic, and the Antarctic region. In: Restelli G, Angeletti G (eds) Dimethylsulfide: oceans, atmosphere, and climate. Germany, Dordrecht, pp 95–102

Staubes-Diederich R (1992) Veneilung von dimethylsulfid, carbonylsulfid und schwefelkohlenstoff in ozean und mariner atmosphere. Dissertation, Johann Wolfgang Goethe Universitat

Stefels J, Steinke M, Turner S, Malin G, Belviso S (2007) Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. Biogeochemistry 83(1–3):245–275. doi:10.1007/s10533-007-9091-5

Sunda WG, Kieber DJ, Kiene RP, Huntsman S (2002) An antioxidant function for DMSP and DMS in marine algae. Nature 418:317–320

Toole DA, Kieber DJ, Kiene RP, White EM, Bisgrove J, del Valle DA et al. (2004) High dimethylsulfide photolysis rates in nitrate-rich Antarctic waters. Geophys Res Lett 31(L11307). doi:10.1029/2004GL019863

Tremblay J-E, Gagnon J (2009) The effects of irradiance and nutrient supply on the productivity of Arctic waters: a perspective on climate change. In: Nihoul JCJ, Kostianoy AG (eds) Influence of climate change on the changing Arctic and Subarctic conditions. Proceedings of the NATO advanced research workshop, Liege, Belgium, 8–10 May 2008. Springer Verlag, Dordrecht

Vallina SM, Simó R (2007) Strong relationship between DMS and the solar radiation dose over the global surface ocean. Science 315(5811):506–508. doi:10.1126/science.1133680

Vila-Costa M, Simó R, Harada H, Gasol JM, Slezak D, Kiene RP (2006) Dimethylsulfoniopropionate uptake by marine phytoplankton. Science 314(5799):652–654. doi:10.1126/science.1131043

Wei C-L, Rowe GT, Escobar-Briones E, Boetius A, Soltwedel T et al (2010) Global patterns and predictions of seafloor biomass using random forests. PLoS ONE 5(12):e15323

WOA (2005) World Ocean Atlas, NODC. http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html. Accessed 30 Oct 2008

Wolfe GV, Steinke M, Kirst GO (1997) Grazing-activated chemical defense in a unicellular marine alga. Nat 387: 894–897