ORIGINAL PAPER

Species distribution modeling of storm-petrels (*Oceanodroma furcata and O. leucorhoa*) in the North Pacific and the role of dimethyl sulfide

G. R. W. Humphries · F. Huettmann ·
G. A. Nevitt · C. Deal · D. Atkinson

Abstract Storm-petrels have been shown to use dimethyl sulfide (DMS) as a foraging cue, suggesting that this compound may be used to predict their distribution. We describe a new distribution model that employs machine learning software and geographic information systems to model storm-petrel distribution. We used environmental predictor variables that included newly available climatologies of sea surface DMS concentrations to construct distribution maps of fork-tailed storm-petrel (*Oceanodroma furcata*) and Leach's storm-petrel (*O. leucorhoa*) in the North Pacific and Bering Sea. Model accuracy was assessed by (1) using the area under the receiver operating characteristic curve (AUC) values and (2) comparing predicted distributions to presence and non-detection data from two opportunistic pelagic surveys performed in summer 2008. Models using all predictor variables gave AUC values of 0.89 and 0.75, sensitivity values of 0.73 and 0.61, and specificity values of 0.83 and 0.73 for fork-tailed and Leach's storm-petrel, respectively. Models using all predictor variables except DMS gave AUC values of 0.87 and 0.74, sensitivity values of 0.81 and 0.60, and specificity values of 0.77 for fork-tailed and Leach's storm-petrel, respectively. The large-scale link between DMS and how storm-petrels use it to locate foraging areas was reinforced by the partial dependence of DMS on the relative index of occurrence (RIO) of storm-petrels, and by a decrease in AUC values when removing DMS as a predictor. This work is a preliminary step toward linking seabird distribution to globally important infochemicals and should be a basis for further study.

Keywords Dimethyl sulfide · Storm-petrels · North Pacific · Bering Sea · GIS · TreeNet

G. R. W. Humphries (✉)
Department of Zoology, University of Otago,
9016 Dunedin, New Zealand
e-mail: humphries.grant@gmail.com

F. Huettmann
EWHALE Lab, Department of Biology and Wildlife,
University of Alaska Fairbanks, Fairbanks, AK 99775, USA

G. A. Nevitt
Section of Neurobiology, Physiology and Behavior,
College of Biological Sciences, University of California,
Davis, CA 94923, USA

C. Deal
International Arctic Research Center, University of Alaska
Fairbanks, Fairbanks, AK 99775, USA

D. Atkinson
Department of Geography, University of Victoria,
Victoria, BC V8P 5C2, Canada

Introduction

Dimethyl sulfide (DMS) is a catabolic breakdown product in marine phytoplankton and is released into the ocean during both senescence and grazing. The air–sea exchange of DMS is primarily controlled by gas concentration differences between air and sea and by turbulence in the lower atmosphere (Charlson et al. 1987). Once particulate DMS is transferred to the atmosphere due to wind action, it begins to form sulfate aerosols via oxidation and becomes climatically active (Charlson et al. 1987). DMS is linked to areas of high productivity where macrozooplankton (e.g., euphausiids) may be located (Andreae and Raemdonck 1983), and it has been shown to function as a foraging cue for various marine organisms (Nevitt 2011).

🍁 Springer

Experimental evidence suggests that storm-petrels are responsive to DMS deployed at sea (Nevitt et al. 1995) and in colonies (Nevitt and Haberman 2002), and that some petrel species can detect DMS at picomolar concentrations (Nevitt and Bonadonna 2005). Other species (e.g., African penguins (*Spheniscus demersus*)) can detect DMS when it is experimentally deployed in breeding colonies (Cunningham et al. 2008). Reef fishes (DeBose et al. 2008) and copepods (Steinke et al. 2006) also use dimethylsulfoniopropionate, the precursor to DMS, as infochemicals in trophic interactions. Fork-tailed storm-petrels (*Oceanodroma furcata*) and Leach's storm-petrels (*O. leucorhoa*) feed on euphausiids (krill) and other planktonic organisms, and it has been suggested that they use DMS to find productive foraging areas (Nevitt et al. 1995, 1999).

Fork-tailed storm-petrels nest sympatrically with Leach's storm-petrels during the breeding season in the North Pacific. Both species leave their burrows at night to forage at sea for several days before returning to their colonies (Wilbur 1969; Boersma et al. 1980). The use of olfactory cues in burrowing seabirds like storm-petrels may be due to evolutionary pressures forcing these birds to capitalize on patchy prey distribution more quickly than other species (Van Buskirk and Nevitt 2007). Like other procellariiformes, these birds have large olfactory bulbs (Bang 1966) and can track food-related odors including DMS (Grubb 1979; Nevitt and Haberman 2002; Nevitt 2008).

The North Pacific Pelagic Seabird Database (NPPSD) contains survey data from the 1970 s to the early 2000 s and shows fork-tailed storm-petrel mostly distributed in the Bering Sea, the Gulf of Alaska, and south along the western North American coast, with some sightings off the coast of Japan (Drew and Piatt 2005). The NPPSD shows Leach's storm-petrel extending south of (or around) the Aleutian islands, through the Gulf of Alaska, down the western coast of North America and off the coast of Japan. Winter distribution of fork-tailed storm-petrel is assumed to be limited to its summer distribution in the North Pacific with recorded sightings on the ice edge during the wintertime in the Bering Sea (Onley and Scofield 2007). Leach's storm-petrels are thought to move further south during the winter months, with the majority of sightings occurring around Hawaii and in the Eastern Pacific (Huntington et al. 1996; Onley and Scofield 2007). Species distribution modeling of storm-petrels in the North Pacific, which could greatly increase our knowledge of what drives the distribution of these procellarids, has not yet been performed.

Pelagic seabird studies frequently examine the relationships between presence/absence data and ocean characteristics such as sea surface temperature, chlorophyll a, and salinity, with less importance being placed on other variables (Tremblay et al. 2009). Many of the ocean parameters used, however, do not directly reflect the sensory mechanisms that seabirds, themselves, use to locate foraging areas. Using DMS as a predictor variable is a likely starting place.

Ecological niche modeling has become an established method for determining species distributions in terrestrial environments. This method is also important for conservation management for examining the relationship between the environment and the distribution of a particular species. The goal of ecological niche modeling is to predict species occurrence based on georeferenced presence and non-detection (absence) points, which, in turn, correlate to some environmental feature. One of the major advantages of this type of modeling is that it enables the efficient creation of large-scale models at both regional and global spatial scales.

Investigating possible variables that may help to understand the foraging ecology in pelagic seabirds is important because of the complexity of ecological systems. Olfactory cues (such as DMS) as represented by climatologies that can be easily associated with presence/absence data may add important information toward the predictability of the pelagic distribution of storm-petrels. The goal of this study was to examine if information could be added to pelagic distributional models of storm-petrels by including DMS as a predictor variable.

Materials and methods

We obtained presence and non-detection data for both species of storm-petrel for May to August for the years 1974 through 2002 in version 1 of the NPPSD (alaska.usgs.gov/science/biology/nppsd/index.php) and then projected these in the World Geodetic System (WGS) 1984 geoid using ArcGIS Desktop 10.0. Only data taken from large ships using strip transect methods binned by 10-minute intervals were used to perform modeling. Each data point reflected a count of each species located within the 10-minute bin. A presence point had a count of Leach's or fork-tailed storm-petrel of one or more. Non-detection points had a zero count for either Leach's or fork-tailed storm-petrel. Non-detection points may not represent true absences as they can occur if a bird is not sighted due to observer fatigue, or because a bird may simply not be in the vicinity at that time. Barbet-Massin et al. (2012) removed absence points that occurred within the known range of a species to create true absences. Because the nature of our goal was to examine whether information is added to our model by using DMS as a predictor variable, and because DMS is patchy within the known range of both species, it was important to keep some absence points in the known storm-petrel domain. We therefore chose to remove absence points from within a 40 km radius of presence points

in order to reduce the number of potential false absences. We chose 40 km as a radius based on tests with various buffer sizes (e.g., 0, 20, 40, 80, and 160 km) to determine which would provide the best models while still allowing models to resolve (i.e., the larger the buffer, the lower the variation in the modeling data, thus causing models to not converge on a solution). Presence data were not excluded 40 km from each other to match the radius for absence points because even though there is pseudo replication within bins, we are able to gain important information from knowing where clusters of presences are occurring. It is also likely that dynamic oceanographic variables (such as DMS), and how storm-petrels are affected by them, operate at very fine spatial scales in the near-shore regions, which cannot be adequately reproduced with the data we have available. To take into account the effect of central place foragers with respect to the scale of the variables available, we buffered all data to 200 km from colonies. This also allowed us to make a more targeted assessment of how DMS affects storm-petrels in pelagic regions. Colony data were obtained from the North Pacific Seabird Colony Database (alaska.fws.gov/mbsp/mbm/northpacificseabirds/colonies/default.htm). The training data were split geographically at 180° longitude, where all data east of 180 were used to train the algorithm and all data west of 180 were used to test the performance of the model. The data were also divided temporally into the months of May through August in order to include month as a predictor variable in the models.

We downloaded a suite of monthly climatologies ("Appendix") for the months of May through August. The years that are represented by these climatologies differ between data sources. Data downloaded from the World Ocean Atlas represent all oceanographic information from all in situ measurements collected by the Global Oceanographic Data Archaeology and Rescue project from 1977 to 2004 (Levitus et al. 2005), the IOC World Ocean Database project, and the IOC Global Temperature Salinity Profile project (IOC 1998). Data from the Oceancolor project represent composite means of data from 2000 to 2009. The climatologies were projected into WGS 1984 then clipped to the study area (36–66° latitude in the North Pacific). All of these climatologies represent a suite of environmental variables, which represent physical oceanic processes. We did not include static layers such as bathymetry, or distance to shoreline or colonies because we removed all data points within a 200 km radius of colonies. Layers like bathymetry or distance to coastline are important for near-shore distributions; however, we were only interested in the information gain of DMS in offshore distributions. It is also often the case that when layers such as bathymetry or distance to coast (or colony) come up as the most important predictors, that unusual patterns are found in the output maps (e.g., regular circular patterns or distributions that follow bathymetry exactly).

The DMS layer was generated using similar spatial modeling techniques used in this study. DMS was computed on a global spatial scale and monthly temporal scale using open access DMS point measurements and a suite of oceanographic variables (Humphries 2008). All of the oceanographic layers used in this modeling procedure were used to derive the DMS layer; however, when examining the correlations of all the oceanographic layers compared to DMS via a correlation matrix, we found that there were no strong linear relationships between DMS and the other variables. We also computed interactions between variables within the models and found no strong interactions between DMS and the other predictor variables, thus justifying the use of DMS in conjunction with all other variables in these models.

Autocorrelation occurs when ecological processes may be expressed as a function of spatial location or time between samples (e.g., how closely samples are correlated to one another) (Cushman 2009). Spatial autocorrelation in a species dataset (e.g., how closely birds flock together in space) can influence apparent relationships between environmental variables (Huettmann and Diamond 2006) and model evaluation (Hijmans 2012). We calculated Moran's I values for our data and determined that the spatial data were highly autocorrelated (Moran's I: 0.99, $p \ll 0.001$). We calculated the autocovariate on spatial autocorrelation in the residuals (RAC) using program R (v2.11; www.rproject.org) as per Crase et al. (2012). The RAC value was integrated into the model as a predictor variable to account for autocorrelation.

Filtered presence and non-detection data were overlaid with all environmental variables using the geospatial modeling environment (GME; version 0.5.2). Presence and non-detection data for each month were associated with the environmental values for their respective months, and ESRI shapefiles from this process were converted to text format for modeling. Presence or non-detection data represented our response variable, while our explanatory variables in the model were represented by the environmental variables ("Appendix"), RAC values, and month. Because of the high number of predictor variables, we chose a machine learning technique for modeling.

Traditionally, generalized linear (GLMs) or additive models (GAMs) have been used to analyze and predict species distributions, but more recently a variety of more sophisticated algorithms have been developed and applied (Elith et al. 2006, 2008; Craig and Huettmann 2009; Nur et al. 2011; Oppel et al. 2011). Many of these algorithms, such as boosted regression trees (brt), "learn" the relationship between a target and the many different predictor variables. GLMs or GAMs require a priori assumptions of

a data model (Breiman 2001; Elith et al. 2006, 2008). When we take a statistical approach with no a priori assumptions regarding what may control the distribution of these species at sea, the modeling process gains a greater degree of flexibility. That is, we can perform complex multi-hypothesis tests on a large variety of predictor variables that define the region of interest and can, from there, make conclusions on the distribution of these species. The brt algorithm uses regression trees to derive the relationships between a series of predictor variables and a target (response) variable. The algorithm is nonparametric and does not require any a priori assumptions about the relationships in the data and therefore allows for great flexibility in model creation (Breiman 2001; Elith et al. 2008). Another advantage of this tool is that overfitting can be avoided by cross-validation of the data, which also boosts prediction power (Friedman 2002; Elith et al. 2006). We used the "dismo" package in R to implement the brt algorithm, "gbm.step" (Elith et al. 2008).

The training data were modeled with all of the variables and then all variables excluding DMS. This design facilitated a targeted assessment of the information added by DMS as a predictor of storm-petrel distribution. This was first done by using the informed default settings (Elith et al. 2008), which is found to be useful in getting fast, accurate results (Craig and Huettmann 2009). We therefore used a tree complexity of 5 to ensure reasonable complexity in our model and a learning rate of 0.01 with all other settings at default. The learning rate was then decreased to 0.001, as a model solution was not reached at the previous learning rate. Model accuracy was assessed by predicting the relative index of occurrence (RIO) as calculated by the gbm.step algorithm to data west of 180° longitude. RIO is, in essence, the likelihood of locating the focal species within a map cell. RIO values were compared to presence and non-detection points within the test data to calculate assessment values. The most accurate models were those with the largest area under the receiver operator curve (AUC) (Fielding and Bell 1997). We also examined the sensitivity and specificity of the model (i.e., the percentage of correctly classified presences and absences, respectively). Sensitivity and specificity are straightforward and easy to interpret metrics, which can help to determine where models may be failing (e.g., poor prediction of presences or absences). These values are calculated from independent data, which is a preferred way to examine the accuracy of models generated by machine learning algorithms (Bradley 1997; Hegel et al. 2009).

In order to determine the temporal scale on which to model (i.e., monthly or seasonal), we examined the relative importance of month briefly. We did this by adding month as a variable in the model and then examining the partial dependence plots and the variable importance values. We found that

month was not important and no pattern existed in the partial dependence plots, we therefore chose to model on a seasonal temporal scale. This was not surprising as there is no evidence to assume that the distribution of either species would change over the breeding season when they are raising chicks. Summer model predictions were represented by visualizing the relative index of occurrence (RIO) as calculated by the gbm algorithm. A regular grid of data points at a resolution of 10 km × 10 km was created in the study area in ArcGIS, and environmental variables were averaged to derive summer (May to August) climatologies. The regular grid was then overlaid with the derived summer climatologies. The trained model was then used to create predictions to each of the points in the regular grid.

Model performance was also assessed via two opportunistic surveys we performed in summer 2008 (July and August). One survey was performed aboard the T/S Oshoro-Maru in the northeastern Bering Sea. The second survey was performed aboard the M/V Tiglax between Homer, Alaska, and Adak Island, Alaska. Many of the survey data aboard the M/V Tiglax were within 200 km of known colonies, which falls within the domain of excluded occurrence data. We opted to keep this data in the evaluation as it helped to make inferences about how well pelagic oceanographic variables could model the distribution in the near-shore region. All data were collected using distance sampling methods (Thomas et al. 2002). Presences of all species within 400 m of the ship were recorded using a hand-held GPS. These data were organized into presence and non-detection of fork-tailed and Leach's storm-petrel. Because transects were <40 km length, we mapped the centroids of all transects and categorized them as presence or non-detection, thus reducing the number of data points to 28 transects. Due to there being only 28 data points, we opted not to calculated AUC values and examined mean RIO of presences and absences.

We also examined the role of DMS toward the predictions of RIO by examining the partial dependence plots generated in the "dismo" package. The partial dependence is not a direct relationship but is a representation of the relative contribution of DMS to the function, which calculates the RIO. Partial dependence is calculated as an approximation of the function to predict RIO while taking into account interactions with all the other predictor variables (Friedman 2001). These plots can be viewed as additive or multiplicative with other variables (i.e., other partial dependences) to create predictions of RIO and are a useful way of interpreting effects of predictors.

Results

Model AUC values were 0.89 and 0.75 for fork-tailed and Leach's storm-petrel, respectively, when using all predictor

Table 1 Assessment scores (AUC) for models using all predictor variables based on an independent test set west of 180° longitude

	FTSP			LESP		
	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
w/DMS	0.73	0.83	0.89	0.61	0.73	0.75
wo/DMS	0.81	0.77	0.87	0.60	0.77	0.74

variables. Sensitivity and specificity values were 0.73 and 0.83, respectively, for fork-tailed storm-petrel, while sensitivity and specificity values were 0.61 and 0.73, respectively, for Leach's storm-petrel when using all predictor variables. We found that when we ran models without DMS as a predictor variable, there was some loss of information in the model. AUC values for both fork-tailed and Leach's storm-petrel models decreased to 0.87 and 0.74, respectively. Sensitivity increased to 0.81 for fork-tailed storm-petrel and decreased to 0.60 for Leach's storm-petrel, while specificity values decreased for fork-tailed storm-petrel to 0.77 and increased for Leach's storm-petrel to 0.77 (Table 1).

We briefly examined the variable importance values determined by the algorithm; however, correlation between variables may lead to more than one predictor explaining the same amount of variation as another, thus depressing the importance of one of those environmental factors causing importance values to be swapped. We found that there were no significant linear relationships between DMS and the other predictor variables via a correlation matrix; however, this does not take into account non-linear relationships. We therefore elected not to consider variable importances in our analysis. We instead focused on the effect of DMS on the predictions of RIO via partial dependence plots.

Partial dependence plots of DMS to the RIO of fork-tailed and Leach's storm-petrels show trends of higher RIO values being found at higher concentrations of DMS. The partial dependence for fork-tailed storm-petrel begins to increase drastically at approximately 1 nM and plateaus around 10 nM. For Leach's storm-petrel, there is a similar pattern where partial dependence values begin to increase at approximately 3 nM and then plateaus at approximately 30 nM (Fig. 1). Information gained from the modeling process was used to create distribution maps for both species of storm-petrel while keeping DMS as a predictor variable.

Distribution maps for fork-tailed and Leach's storm-petrels were produced at a resolution of $1 \times 1°$ to match the same resolution of the predictor variables with an extent of 36–66° latitude and 140 to −122° longitude (Fig. 2). Models were poorly resolved around coastlines due to the coarseness of the predictor variables. The models represent distribution of both species for the breeding season (i.e., the months of May through August) and show relatively ubiquitous distributions within their
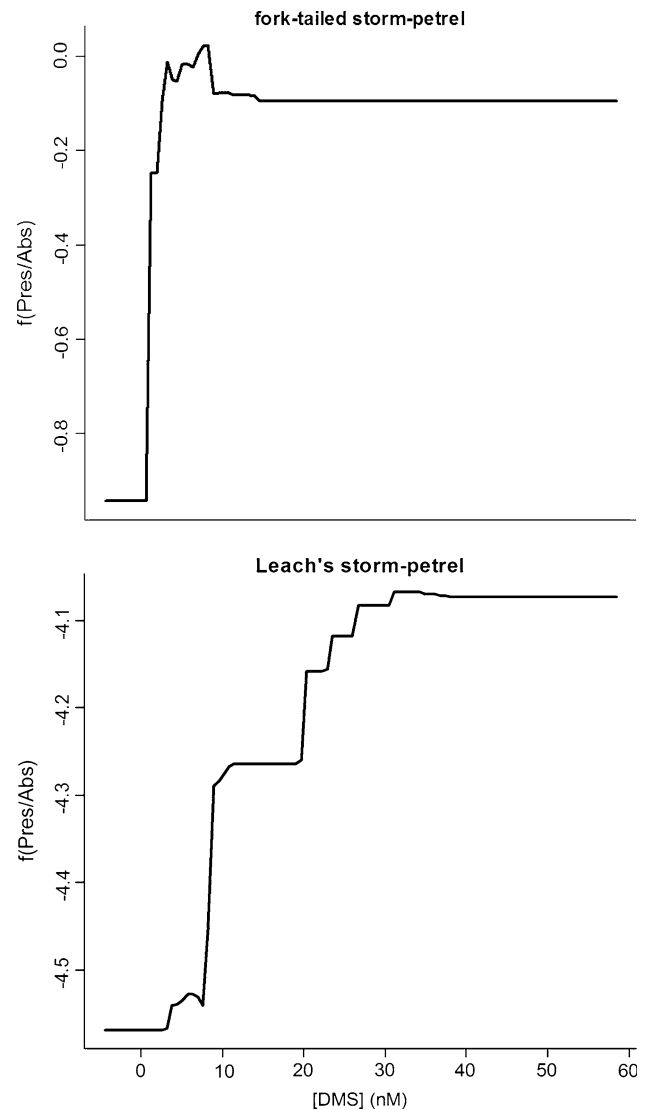
Fig. 1 Partial dependence plots of the approximation of the fitted function of the predictions to concentrations of DMS for fork-tailed and Leach's storm-petrel

individual ranges. Fork-tailed storm-petrels occur much further north, extending into the Aleutian basin and along the Eastern coast of the Kamchatka peninsula. Higher RIO values (>0.90) were found around the edges of the Aleutian basin, around the Kuril Islands and in the Gulf of Alaska. There seems to be a definitive boundary to their southern distribution ($\sim 43.5°$ N), which is corroborated by the lack of presences south of this border (Fig. 3). However, the
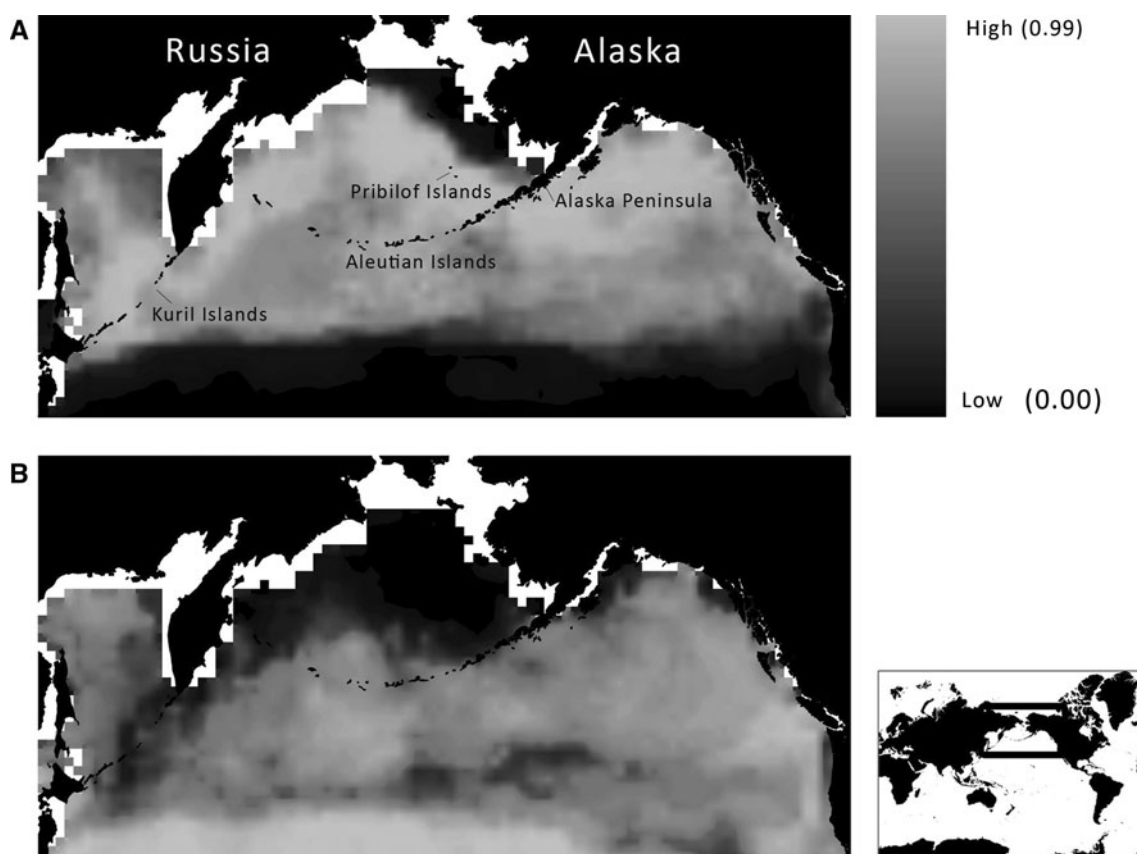
Fig. 2 Predicted distribution of fork-tailed (**a**) and Leach's (**b**) storm-petrel as calculated by the gbm.step algorithm using all predictor variables and presence/non-detection data from the North Pacific Pelagic Seabird Database (Drew and Piatt 2005). Values in the maps represent the relative index of occurrence (RIO) of the focal species

model predicts medium–low RIO values off the coast of Oregon and Washington in the United States. High (>0.70) RIO values for Leach's storm-petrel were not predicted further north than ∼57° N in the Aleutian basin, corroborated by few presences of Leach's storm-petrel recorded in the Aleutian basin (Fig. 4). The range of Leach's storm-petrel extended much further south with the highest RIO values predicted at the southern reaches of the spatial extent of our models. High RIO values were also predicted in the northeastern Pacific off the coast of Vancouver Island, Washington, and Oregon. Both species had high RIO values predicted throughout the Sea of Okhotsk in Russia. Also, high RIO values were predicted around the Aleutian Islands for both species (within 200 km from shore), in areas with known colonies. Despite the fact that we excluded all occurrence records 200 km from known colonies, our pelagic models were able to replicate what is essentially known about storm-petrel distribution in these areas (i.e., high likelihood of finding an individual near the colonies). Some basic ground truthing of the models from shipboard surveys performed in July and August 2008 help

to corroborate how well the pelagic models we created could be extended to near-colony regions.

The centroids of the ship surveys show that in 2008, fork-tailed storm-petrel sightings occurred aboard both vessels on 10 different transects North and South of the Alaska Peninsula in the Aleutian Islands. No fork-tailed storm-petrel was detected North of 56° latitude, coinciding with areas that were predicted to have low RIO with the exception of three transects between 56 and 58° latitude where moderate RIO values were predicted (Fig. 5). Mean RIO for presences of fork-tailed storm-petrel was 0.94, while mean RIO of absences was 0.49 (Fig. 6).

Leach's storm-petrel sightings during the summer of 2008 were limited to only the M/V Tiglax, south of the Alaska Peninsula and occurred in areas where the model predicts moderate RIO for this species. Surveys aboard the T/S Oshoro-maru (where no Leach's storm-petrels were observed) were north of the Aleutians in the Bering Sea (over the Bering shelf) in areas where the model generally predicts low (<0.10) RIO. Transects south of the Pribilof Islands contained no observations of Leach's storm-petrels
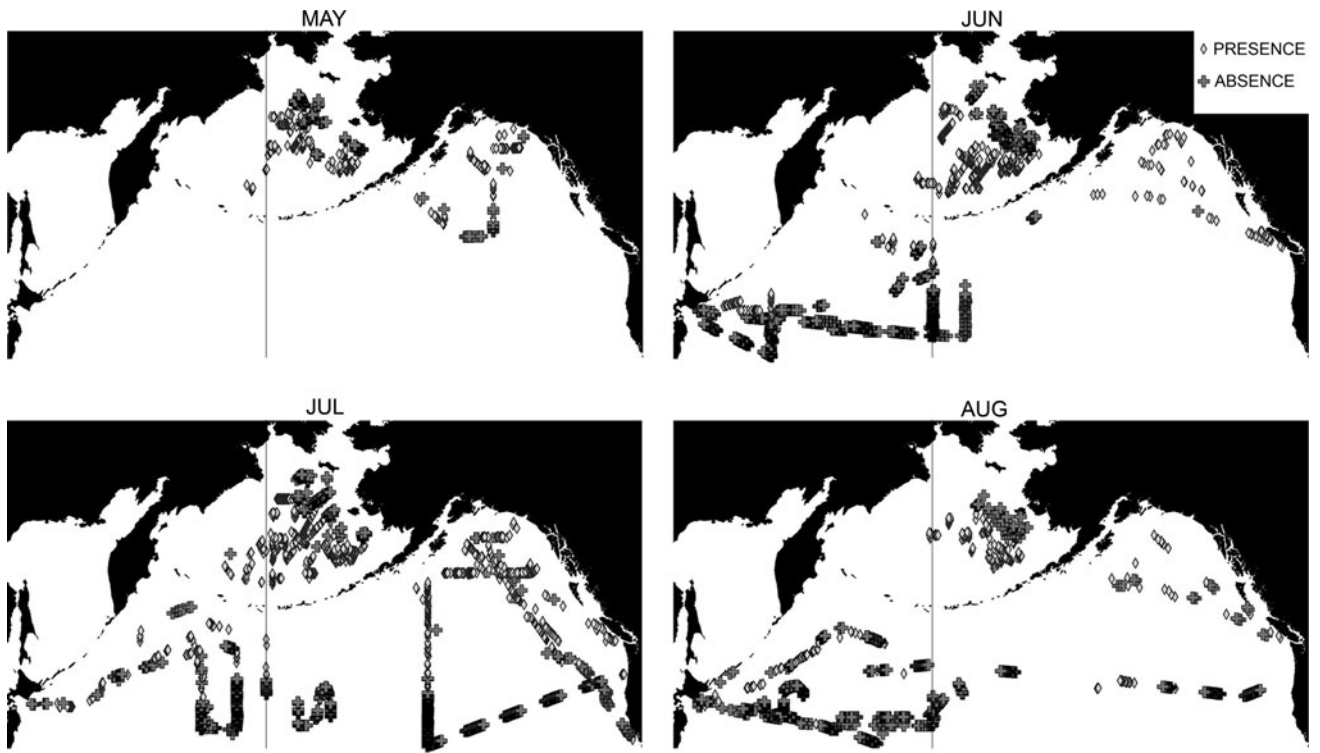
Fig. 3 Distribution of presence and non-detection (absence) points by month of fork-tailed storm-petrel after filtering process from North Pacific Pelagic Seabird Database (Drew and Piatt 2005). The line at 180° longitude represents the split for the training and testing data. Data West (to the *left*) of the line are the testing data, while data East (to the *right*) of the line are the training data
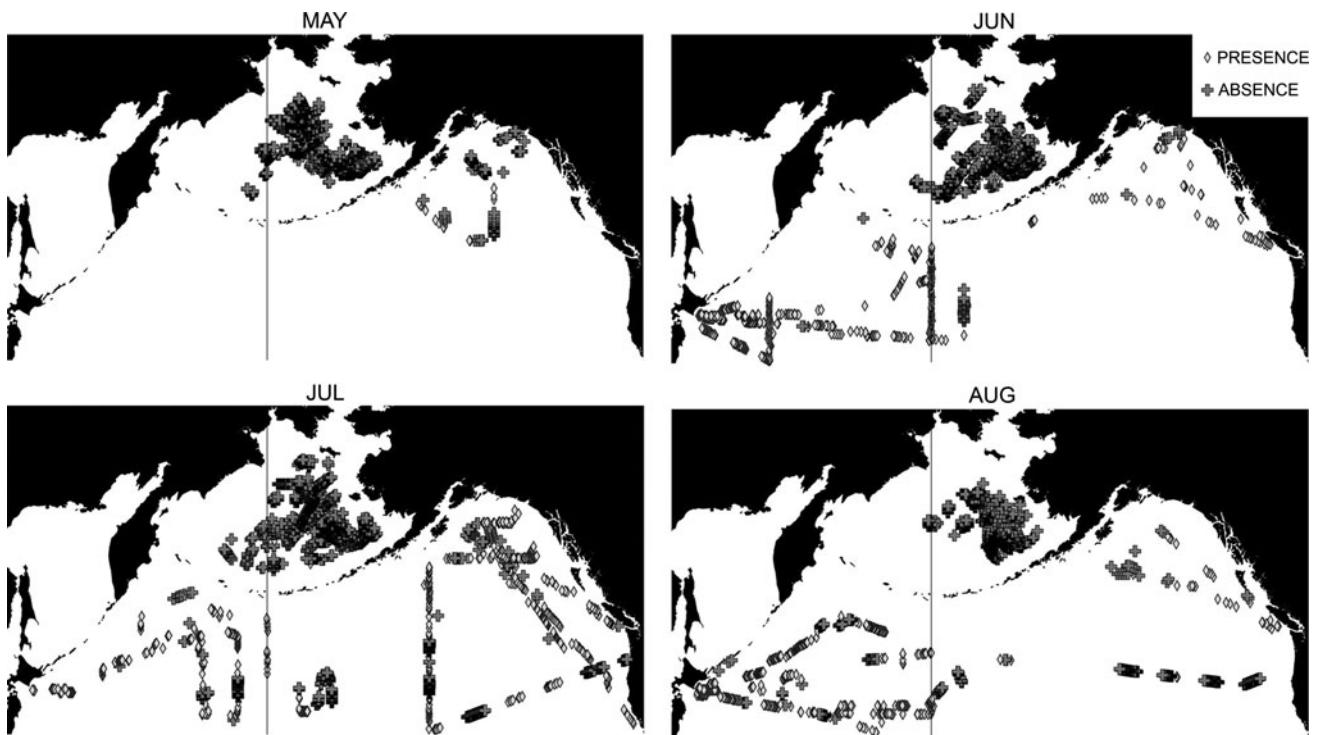

Fig. 4 Distribution of presence and non-detection (absence) points by month of Leach's storm-petrel after filtering process from North Pacific Pelagic Seabird Database (Drew and Piatt 2005). The line at 180° longitude represents the split for the training and testing data. Data West (to the *left*) of the line are the testing data, while data East (to the *right*) of the line are the training data
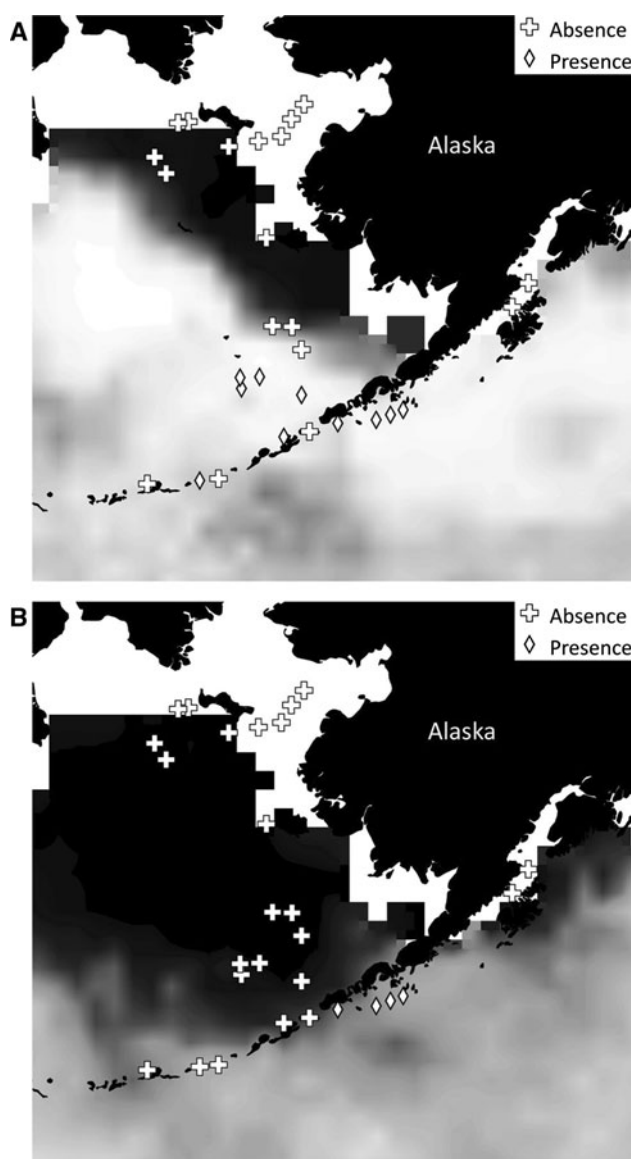
Fig. 5 Confirmed presence and non-detection (absence) points for fork-tailed storm-petrel (**a**) and Leach's storm-petrel (**b**) from at-sea transects performed in summer 2008

Fig. 6 Mean relative index of occurrence values of fork-tailed and Leach's storm-petrel compared to centroids of transects performed in the summer of 2008 as an assessment of model performance

despite higher RIO values predicted for those areas. No Leach's storm-petrel sightings were recorded west of $-164°$ longitude, in disagreement with high RIO values found in the model (Fig. 5). Mean RIO for presences of Leach's storm-petrel was 0.31, while mean RIO for absences was 0.09 (Fig. 6).

Discussion

Our goal was to determine whether DMS could add information to a model of pelagic storm-petrel distribution in the North Pacific. Our results show that the addition of DMS to predictive models of fork-tailed and Leach's
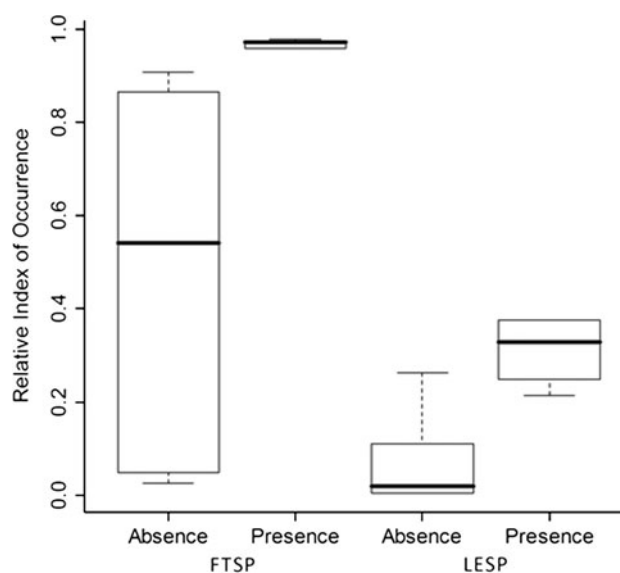
storm-petrel distribution do add information by increasing model accuracy. We were also able to confirm a link between storm-petrels and DMS at biologically appropriate concentrations by examining patterns in the partial dependence plots and the AUC values when removing DMS as a predictor variable. We were also able to create pelagic distribution models for both Leach's and fork-tailed storm-petrels in the North Pacific using all downloaded environmental variables with model accuracies from 0.75 to 0.89.

We ran these models with and without DMS specifically to investigate its potential as a predictor variable. When DMS was not included in the model, performance was lower than when all predictors were used. This indicates that it is likely that DMS is adding a certain amount of information to the model (slight increase in predictive performance). Because a derived variable such as DMS may not act as a powerful predictor on its own, our results are not entirely surprising. This was confirmed by a brief examination of the variable importance plot showing that DMS was not the most important predictor in these models. A correlation matrix of all the predictor variables indicated no significant linear correlations between DMS and other predictor variables. Correlation between predictor variables may depress variable importance as one predictor explains the same variation as another (Grömping 2009). We measured linear and non-linear interactions between DMS and the other variables, but because the DMS model was created using all of the same predictor variables as used in this analysis, there may have been some other effect on the importance of DMS. Because of this, we ignored variable

importances and focused on the partial dependence plots to examine what information DMS was adding to the model.

We examined the relationship between DMS and the predicted RIO of both species of storm-petrel via the partial dependence plots. The partial dependence plots of DMS show that as concentrations of DMS increase, the partial dependence of predicted RIO to DMS increases. This is true for both fork-tailed storm-petrel and Leach's storm-petrel. In both cases, there is an increase in the partial dependence between 1 and 3 nM of DMS. This indicates that there is more likely to be a presence of either species of storm-petrel at higher concentrations of DMS. The concentrations of DMS, which represent higher partial dependence values, are higher than the oceanic means of ~1–3 nM (Stefels et al. 2007), thus corroborating the result that storm-petrels were predicted to occur in areas of high concentrations of DMS. Storm-petrels can track upwind odor sources (Nevitt 2008), and it is likely that, even though turnover time for DMS is quite rapid, birds may track toward broad areas of high productivity (corresponding to higher average DMS concentrations) and then increase track resolution once arriving in these areas. This tracking may explain positive relationship in the partial dependence plots. However, DMS concentrations in the upper mixed layer can change in a matter of hours or days (Yang and Tsunogai 2005; Stefels et al. 2007), and the underlying models of DMS concentrations do not take this into consideration. Also, only presence and non-detection data were used in our model and no behavioral information (i.e., foraging) was incorporated. Because it is theorized that DMS is used to detect foraging areas and presences likely include birds that are in transit to foraging sites, the importance of DMS for predicting storm-petrel distribution could be depressed.

Many species–environment associations can change based on the spatial domain chosen (Schneider and Piatt 1986; Huettmann and Diamond 2006). For this study, we chose the extent to be between 36 and 66° North latitude, comprising the northern halves of the North Pacific Transition Zone Province and the Kuroshio Current Province where 36° N lies between the subtropical and subarctic fronts and 66° N is the Arctic Circle (Longhurst 1998). The presence and non-detection data were further clipped to the extent of the most coarse data layers (i.e., DMS and salinity), which led the final models having an extent stretching only to 63° N. It is possible that biases might exist by excluding presence or non-detection points due to clipping since many of the excluded data were in Prince William Sound, Cook Inlet and in the Shelikof Strait, where storm-petrels have been observed foraging. We have also not included in this analysis any physical predictor variables such as distance to coast, distance to colony, bathymetry, or seamounts. This was done purposefully due to the exclusion of data up to 200 km from

the nearest colony, which was performed to better examine the effects of pelagic oceanographic conditions in our model and to deal with a possible "colony effect" (i.e., many birds being detected simply due to proximity to a colony, not representing the pelagic distribution of the birds). In initial model runs, it was found that these physical variables were controlling the predictions and masking any effects of pelagic oceanographic variables, which were important to the particular question we were addressing. It is unclear as to the advantages or disadvantages to the method we chose; however, distribution models yielded high assessment values (i.e., AUC values of 0.74–0.89) and match what is known about storm-petrels in the North Pacific. Also, despite having excluded occurrence records 200 km from colonies, pelagic models were able to be extended to near-shore regions accurately. The high RIO values predicted around the Aleutian Islands would be expected as the likelihood of finding an individual near a colony is quite high in reality. This could mean that the broad oceanographic conditions near the colonies dictate why storm-petrels occur in these regions (e.g., as opposed to quality of soils or plant structure), or simply that oceanographic conditions are very similar between the pelagic and near-shore zones in the Aleutian Islands. We feel that our models were accurate because they not only yielded high assessment values, but were also substantiated by our summer transects in that presences occurred in areas with higher predicted RIO values than where no birds were observed.

Survey data agree with the summer modeled distributions, indicating that the variables that control storm-petrel distribution have been well captured. It is of importance here to note that being in an area of high RIO (based on the model) does not guarantee a storm-petrel sighting (e.g., lowered detection rate due to weather or observer fatigue). The ship could also be passing through at a time when there are simply no storm-petrels (e.g., perhaps due to time of day). This is one reason why non-detections can potentially cause model disagreement. This is demonstrated in Fig. 6, with a high variability in the mean RIO compared to the absences of fork-tailed storm-petrel. It is clear, however, that the RIO compared to presences is high, while RIO compared to absences is low.

Most of the survey data were taken within the known distribution of fork-tailed storm-petrels (Boersma and Silva 2001; Onley and Scofield 2007). Presences of fork-tailed storm-petrels were only found in areas of predicted high RIO, while many of the non-detections were found in areas of predicted low RIO. Our model shows Leach's storm-petrels to be distributed primarily south of the Aleutians (Fig. 2b), which also agree with other accounts (Huntington et al. 1996; Onley and Scofield 2007). The modeled distribution of Leach's storm-petrels is also generally substantiated by survey data that confirmed presences were

located south of the Aleutians in areas of predicted high RIO. Although no Leach's storm-petrels were noted north of the Aleutians, data from the NPPSD show sightings there in June and July (Fig. 4).

Our study successfully created predictive models of the distribution of Leach's and fork-tailed storm-petrel and suggests that DMS merits further research as a predictor variable. Currently, DMS predictions may be too coarse spatially and temporally to accurately define how storm-petrels are using this compound to locate foraging areas on a small scale. Future work on the role of DMS as a predictor of seabird distribution may help us to create more accurate models for conservation benefits.

Appendix

See Table 2.

Table 2 Sources of predictor variables

Dataset (shorthand)	Source	Resolution
Dimethyl sulfide (dms)	Humphries (2008) Thesis Chapter 1	1°
Salinity (sal)	World Ocean Atlas (www.nodc.noaa.gov)	1°
Dissolved oxygen (do)	World Ocean Atlas (www.nodc.noaa.gov)	1°
Oxygen utilization (o2u)	World Ocean Atlas (www.nodc.noaa.gov)	1°
Nitrates (nit)	World Ocean Atlas (www.nodc.noaa.gov)	1°
Phosphates (pho)	World Ocean Atlas (www.nodc.noaa.gov)	1°
Silicates (sil)	World Ocean Atlas (www.nodc.noaa.gov)	1°
Mean sea surface temperature (sstmean)	NASA—Oceancolor project (www.oceancolor.gsfc.nasa.gov)	1.1 km
Standard deviation of sea surface temperature (sststdev)	NASA—Oceancolor project (www.oceancolor.gsfc.nasa.gov)	1.1 km
Mixed layer depth (mld)	Provided by Dr. Sergio M. Vallina	1°
Mean chlorophyll a (chlmean)	NASA—Oceancolor project (www.oceancolor.gsfc.nasa.gov)	1.1 km
Standard deviation of chlorophyll a (chlstdev)	NASA—Oceancolor project (www.oceancolor.gsfc.nasa.gov)	1.1 km
Chromophoric dissolved organic matter (cdom)	NASA—Oceancolor project (www.oceancolor.gsfc.nasa.gov)	1.1 km
Photosynthetically active radiation (par)	NASA—Oceancolor project (www.oceancolor.gsfc.nasa.gov)	1.1 km

References

Andreae MO, Raemdonck H (1983) Dimethyl sulfide in the surface ocean and the marine atmosphere: a global view. Science 221: 744–747

Bang BG (1966) The olfactory apparatus of tubenosed birds (Procellariiformes). Acta Anat 65:391–415

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Method Ecol Evol 93:679–688

Boersma PD, Silva MC (2001) Fork-tailed storm-petrel (*Oceanodroma furcata*). In: Poole A (ed) The Birds of North America, No. 569. The Academy of Natural Sciences, Philidelphia

Boersma PD, Wheelwright NT, Nerini MK, Wheelwright ES (1980) The breeding biology of the fork-tailed storm-petrel (*Oceanodroma furcata*). Auk 97:268–282

Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 30:1145–1159

Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16:199–231

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

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 promising future. In: Wang HF (ed) Intelligent data analysis: developing new methodologies through pattern discovery and recovery. Idea Group Inc, Hershey

Crase B, Liedloff AC, Wintle BA (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography 35:001–010

Cunningham GB, Strauss V, Ryan PG (2008) African penguins (*Spheniscus demersus*) can detect dimethyl sulphide, a prey-related odour. J Exp Bio 211:3123–3127

Cushman SA (2009) Animal movement data: GPS telemetry, autocorrelation and the need for path-level analysis. In: Cushman SA, Huettmann F (eds) Spatial complexity, informatics, and wildlife conservation. Springer, New York, pp 131–150

DeBose JL, Lema SC, Nevitt GA (2008) Dimethylsulfoniopropionate as a foraging cue for reef fishes. Science 319:1356

Drew GS, Piatt JF (2005) North Pacific Pelagic Seabird Database (NPPSD): compiling datasets and creating an archive, accessible database, and Pelagic Seabird Atlas. US Geological Survey, Anchorage

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Ani Ecol 77:802–813

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absences models. Env Conserv 24:38–49

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

Friedman JH (2002) Stochastic gradient boosting. Comp Stat Dat Anal 38:367–378

Grömping U (2009) Variable importance assessment in regression: linear regression versus random forest. Am Stat 63:308–318

Grubb TC (1979) Olfactory guidance of Leach's storm petrel to the breeding island. Wilson Bull 91:143–145

Hegel TM, Cushman SA, Evans J, Huettmann F (2009) Current state of the art for statistical modelling of species distributions. In: Cushman SA, Huettmann F (eds) Spatial complexity, informatics, and wildlife conservation. Springer, New York, pp 273–309

Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial bias and calibration with a null model. Ecologicals 93:679–688

Huettmann F, Diamond AW (2006) Large-scale effects on the spatial distribution of seabirds in the Northwest Atlantic. Landsc Ecol 21:1089–1108

Humphries GRW (2008) The ecological niche of storm-petrels in the North Pacific and a global model of dimethylsulfide concentration. University of Alaska Fairbanks, Dissertation

Huntington CE, Butler RG, Mauck RA (1996) Leach's storm petrel (*Oceanodroma leucorhoa*). In: Poole A (ed) The Birds of North America, No. 233. The Academy of Natural Sciences, Philadelphia

IOC (1998) Global Temperature-salinity profile programme (GTSPP)—overview and future. IOC technical Series, 49, Intergovernmental Oceanographic Commission, Paris

Levitus S, Sato S, Maillard C, Mikhailov N, Caldwell P, Dooley H (2005) Building Ocean Profile-Plankton databases for climate and ecosystem research. NOAA Technical Report NESDIS 117, US Government Printing Office, Washington

Longhurst AR (1998) Ecological geography of the sea. Academic Press, San Diego

Nevitt G (1999) Olfactory foraging in Antarctic seabirds: a species-specific attraction to krill odors. Mar Ecol Prog Ser 177:235–241

Nevitt G (2008) Sensory ecology on the high seas: the odor world of the procellariiform seabirds. J Exp Bio 211:1706–1713

Nevitt G (2011) The neuroecology of dimethyl sulfide: a global-climate regulator turned marine infochemical. Integr Comp Biol. doi:10.1093/icb/icr093

Nevitt GA, Bonadonna F (2005) Seeing the world through the nose of a bird: new developments in the sensory ecology of procellariiform seabirds. Mar Ecol Prog Ser 287:292–295

Nevitt GA, Haberman K (2002) Behavioral attraction of Leach's storm-petrels (*Oceanodroma leucorhoa*) to dimethyl sulfide. J Exp Bio 206:1497–1501

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

Nur N, Jahncke J, Herzog MP, Howar J, Hyrenbach DK, Zamon JE, Ainley DG, Wiens JA, Morgan K, Balance LT, Stralber D (2011) Where the wild things are: predicting hotspots of seabird aggregations in the California current system. Ecol App 21(6):2241–2257

Onley D, Scofield P (2007) Albatrosses, petrels & shearwaters of the world. Princeton University Press, Princeton

Oppel S, Meirinho A, Ramirez I, Gardner B, O'Connell A, Miller PI, Louzao M (2011) Comparison of five modeling techniques to predict the spatial distribution and abundance of seabirds. Biol Conserv. doi:10.1016/j.biocon.2011.11.013

Schneider DC, Piatt JF (1986) Scale-dependent correlation of seabirds with schooling fish in a coastal ecosystem. Mar Ecol Prog Ser 32:237–246

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 83:245–274

Steinke M, Stefels J, Stamhuis E (2006) Dimethyl sulfide triggers search behavior in copepods. Lim Ocean 51:1925–1930

Thomas L, Buckland ST, Burnham KP, Anderson DR, Laake JL, Borchers DL, Strindberg S (2002) Distance sampling. Encyc Env 1:544–552

Tremblay Y, Bertrand S, Henry RW, Kappes MA, Costa DP, Shaffer SA (2009) Analytical approaches to investigating seabird-environment interactions: a review. Mar Ecol Prog Ser 391:153–163

Van Buskirk RW, Nevitt GA (2007) The influence of developmental environment on the evolution of olfactory foraging behavior in procellariiform seabirds. J Evol Bio 21:67–76

Wilbur HM (1969) The breeding biology of Leach's petrel *Oceanodroma leucorhoa*. Auk 86:433–442

Yang G-P, Tsunogai S (2005) Biogeochemistry of dimethylsulfide (DMS) and dimethylsulfoniopropionate (DMSP) in the surface microlayer of the western North Pacific. Deep-Sea Res I 52:553–567