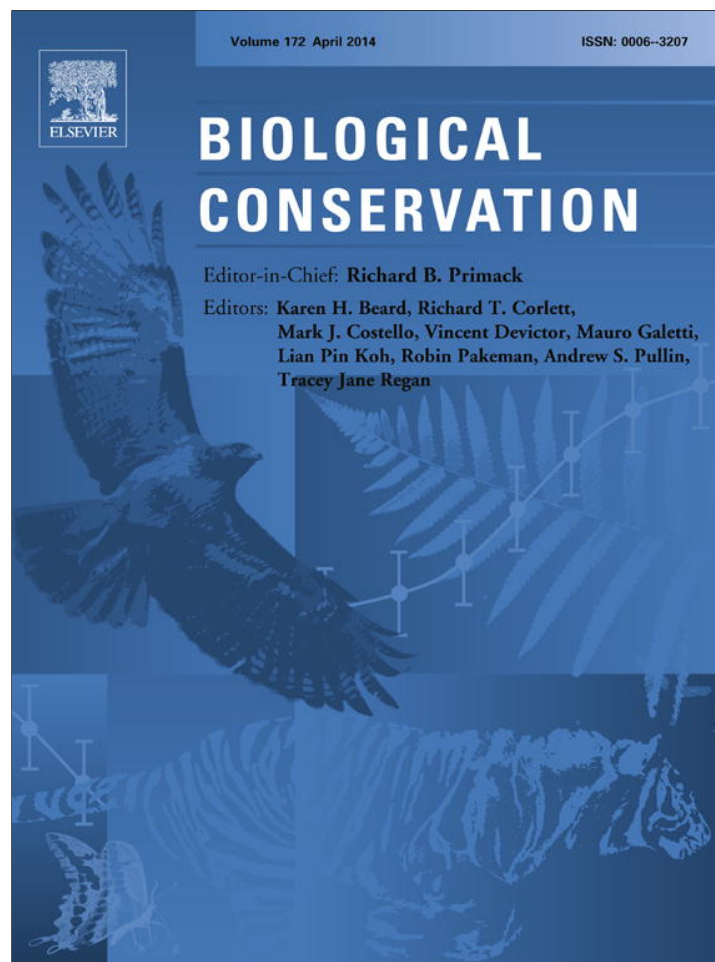


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Maps, models, and marine vulnerability: Assessing the community distribution of seabirds at-sea

David J. Lieske^{a,*}, David A. Fifield^{b,1}, Carina Gjerdrum^c^a Department of Geography and Environment, Mount Allison University, 144 Main Street, Sackville, NB E4L 1A7, Canada^b Environment Canada, Canadian Wildlife Service, 6 Bruce St., St. John's, NL A1N 4T3, Canada^c Environment Canada, Canadian Wildlife Service, 45 Alderney Drive, Dartmouth, NS B2Y 2N6, Canada

ARTICLE INFO

Article history:

Received 12 August 2013

Received in revised form 5 February 2014

Accepted 10 February 2014

Available online 4 March 2014

Keywords:

Species distribution modelling

Multi-species distributions

Ensemble prediction

Seabirds

Negative binomial GLM

Hurdle

Random forest

Maritime Canada

ABSTRACT

Whether considering the cumulative impact of chronic, small-size oil discharges or accidents associated with marine traffic and offshore oil and gas development, seabirds face a variety of threats in the marine environment. Assessing the vulnerabilities of seabirds to maritime hazards requires an understanding of their species distribution, and a means for combining this information across groups. Using at-sea survey data gathered as part of a regional monitoring program, an efficient framework for integrating multi-species data was developed. Survey data was incorporated within a distance sampling framework to generate bias-corrected seabird densities for an area of over 730,000 km² size, which were used to construct multiple species distribution models (SDMs). The structural difficulties associated with sampling sparsely distributed individuals that also occur in large, localized concentrations led to the use of three modelling techniques potentially well suited for this type of data: negative binomial, “hurdle”, and random forest methods. Predicted abundances were combined to produce an ensemble forecast, which met or exceeded the accuracy of predictions from the individual models. Multi-species potential sensitivity maps were developed to identify core areas, confirming the general importance of physiographic features such as the shelf break and bathymetry. Distribution was also seasonally influenced, with spring and winter standing out as periods of peak importance. When combined with an oil pollution layer derived from aerial surveillance, vulnerability was highest in the vicinity of major ports (e.g., Halifax and Sydney, Nova Scotia). However, the vulnerability map also indicated lower but widespread levels of oiling risk throughout the shelf, presumably associated with persistently high levels of shipping traffic and ongoing petroleum extraction and exploration. Outside of this region, migratory connectivity is expected to expose seabirds to a wider network of hazards and further underscores the need for the coordinated and routine collection of marine hazard data alongside at-sea distributional data.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Successful conservation planning hinges on an ability to identify hazards, as well as an awareness of species vulnerability. In marine contexts, threats originate from a number of factors, including: small-scale, chronic oil discharges associated with maritime traffic (Wiese and Ryan, 2003); wide-scale accidental releases of oil (Henkel et al., 2012); fisheries bycatch (Tasker et al., 2000; Hamel et al., 2009); collisions with and loss of habitat associated with offshore wind farms (Exo et al. 2003); and negative interactions with offshore oil drilling platforms (Wiese et al.,

2001). As a result of their low fecundity, seabirds are vulnerable to hazards that impact adult mortality rates (Votier et al., 2005). Furthermore, heavier-bodied, diving species are particularly at risk to oil pollution given the amount of time they spend interacting with the sea-surface–air interface (Camphuysen, 1998). Mitigating anthropogenic risks first requires information about organisms’ usage of the marine environment, in both time and space, so that priority areas can be identified and effective management strategies developed. Unfortunately, the very large area of ocean habitat potentially utilized by seabirds makes it logistically difficult to enumerate these species (Brown, 1980; McKinnon et al., 2009). Pelagic surveys rely heavily on moving platforms (e.g., research/fishery patrol vessels, ferries, sailboats) which may or may not be primarily tasked with gathering seabird information. Large gaps in knowledge inevitably occur, especially during the

* Corresponding author. Tel.: +1 5063642315; fax: +1 5063642625.

E-mail address: dlieske@mta.ca (D.J. Lieske).¹ Current address: Department of Environment and Conservation, Government of Newfoundland and Labrador, P.O. Box 8700, St. John's, NL A1B 4J6, Canada.

non-breeding season when, no longer constrained by central place foraging within range of their colonies, species are free to roam over much larger distances (Huettmann and Diamond, 2001; Oppel et al., 2012).

In terrestrial contexts, there are many successful applications of species distribution modelling (SDM, Guisan and Zimmermann, 2000; Franklin, 2009) to predict habitat usage in unsurveyed areas (Austin, 2002; Shriner et al., 2002). SDMs work by combining empirical patterns of occurrence with Geographic Information System (GIS)-derived information about environmental conditions, and are a valuable tool for conducting conservation planning (Jones-Farrand et al., 2011). Applying this methodology in marine contexts can be quite difficult, however, as marine environments are highly dynamic (Haney, 1989). Processes such as front formation, important for nutrient entrapment and elevated biological productivity (Schneider, 1990; Bakun, 1996), can shift position over a period of hours (Durazo et al., 1998). Nevertheless, upwelling events and nutrient entrapment also associate with persistent physiographic features such as shelf breaks and sea banks (Smith and Petrie, 1982; Hannah et al., 2001), making these useful indicators of marine productivity at broader scales. It can be expected that foraging seabirds will shift their at-sea distributions to benefit from these associations, displaying sparse distributions over wide areas, and larger, more localized concentrations in productive foraging areas (Briggs et al., 1987; Hunt and Schneider, 1987; Fauchald et al., 2002; Clarke et al., 2003). This produces data distributions with both a large number of non-observations (the so-called 'zero-inflation' problem) and occasionally large concentrations of highly variable abundance (the so-called 'overdispersion' problem). Taken together, zero inflation and overdispersion pose serious problems for SDMs based on traditional techniques (Hilbe, 2008).

Over the past 20 years, there has been a notable growth in the number of algorithms used to model animal occurrence and abundance (Hegel et al., 2010). Such methods range from presence-only techniques (e.g., maximum entropy or MAXENT, Phillips et al., 2006; ecological niche factor analysis or ENFA, Hirzel et al., 2002) to presence-absence (logistic generalized linear or logistic generalized additive models), to abundance (Poisson or negative binomial regression). Machine learning methods such as random forest (Breiman, 2001) can be applied in any of these cases. However, the very large number of available methods renders it difficult to choose a technique (Jones-Farrand et al., 2011), so a decision is often based on investigator preference or familiarity (Araújo and New, 2007; Jones-Farrand et al., 2011). Current research, however, points to the advantages of combining the predictions of different modelling methods, a process referred to as ensemble modelling (Araújo and New, 2007; Oppel et al., 2012) or decision fusion (Das et al., 2008). The implicit assumption is that while the true underlying process is unknown, different forecasting models are able to capture different components of the underlying signal (Clemen, 1989). The advantages are expected to be maximized when different, but complementary, modelling algorithms are employed (Clemen, 1989).

Marine SDMs are rare (Robinson et al., 2011), and SDMs based on shipboard surveys are even rarer (Oppel et al., 2012). Due to almost four years of pelagic surveys conducted as part of the Eastern Canada Seabirds at Sea monitoring program, this study was able to extend the approach developed by Oppel et al. (2012) and examine the at-sea distribution of a group of seabirds that occur on or near the Scotian Shelf of Nova Scotia. Survey data was incorporated within a distance sampling framework to generate bias-corrected seabird densities for an area of over 730,000 km² size. Three different modelling approaches were selected for their theoretical suitability to deal with zero-inflated, overdispersed data: negative binomial (NB) generalized linear modelling, Hurdle modelling,

and random forest (RF) machine learning. Results from these methods were combined to produce an ensemble prediction, which was used to produce a composite oil-pollution sensitivity map. The sensitivity layer was then combined with information about chronic, small-scale oil pollution to reveal patterns of oiling vulnerability, illustrating that the results of disparate monitoring programs, with different objectives, can be usefully combined to assist regional conservation planning.

2. Methods

2.1. Study area

The focal area was the Scotian Shelf with a maximum depth of about 200 m, extending 150–250 km offshore of the province of Nova Scotia, Canada. Also included were portions of the Bay of Fundy as well as the Cabot Strait (Fig. 1). At the shelf break sea depth increases rapidly to abyssal levels exceeding 3000 m. The topography of the shelf, combined with tidal forces (e.g., tidal mixing), wind, and fluctuations of offshore currents associated with the Gulf Stream, result in upwelling and partial gyres capable of retaining nutrients and concentrating marine biomass (Smith and Petrie, 1982; Hannah et al., 2001; Nova Scotia Museum of Natural History, 2010). A number of banks occur throughout the shelf, some of which (e.g., Browns Bank, Sable Island Bank, Georges Bank) experience tidally-produced gyres of sufficient strength to also retain nutrients and concentrate prey (Hannah et al., 2001).

2.2. At-sea survey data

Three focal species were the subject of study: (1) Black-legged Kittiwake (*Rissa tridactyla*); (2) Dovekie (*Alle alle*); and (3) Northern Fulmar (*Fulmarus glacialis*). This group is not only a significant presence on the Scotian shelf, but they also spend a significant portion of their time interacting with the sea-surface-air interface, which elevates their risk of oil exposure (Brown, 1980; Camphuysen, 1998). Adopting the terminology of Zacharias and Gregr (2005), these species can be described as *sensitive* to stresses related to sea-surface oil pollution, primarily as a result of their foraging modes: surface feeding in the case of Black-legged Kittiwake and Northern Fulmar, and deep-water pursuit diving in the case of Dovekie (Brown, 1980; Camphuysen, 1998).

Observations from 76 shipboard surveys were gathered from March 1, 2006 to October 31, 2009 as part of the Eastern Canada Seabirds at Sea (ECSAS) program (Fifield et al., 2009). The ECSAS survey protocol (Gjerdrum et al., 2012) incorporates the recommendations of Tasker et al. (1984), as well as modern distance sampling techniques (Buckland et al., 2001). Observations were conducted near the front of the vessel during 5-min periods called watches, but only when vessel speed was between 4 knots (7.4 km/h) and a maximum of 19 knots (35.2 km/h). For each watch, the date, time, start and end positions, course, speed, weather, visibility, and sea and ice conditions were recorded. When visibility was poor (due to rain or fog) the actual width of the visible transect (e.g., 200 m) was recorded. Swimming birds were continuously recorded during 5-min watches, while flying birds were sampled "instantaneously" roughly every 300 m.

Swimming birds were identified on one side of the vessel, and counted and assigned to one of four distance classes perpendicular to the vessel track (0–50 m, 51–100 m, 101–200 m, 201–300 m). A distance gauge was used when necessary. Due to ongoing development of the survey methodology, distances to flying birds were measured using one of three different methods: (1) with no distances measured and an assumed uniform detection function of probability one (9% of surveys by length of track, from October to

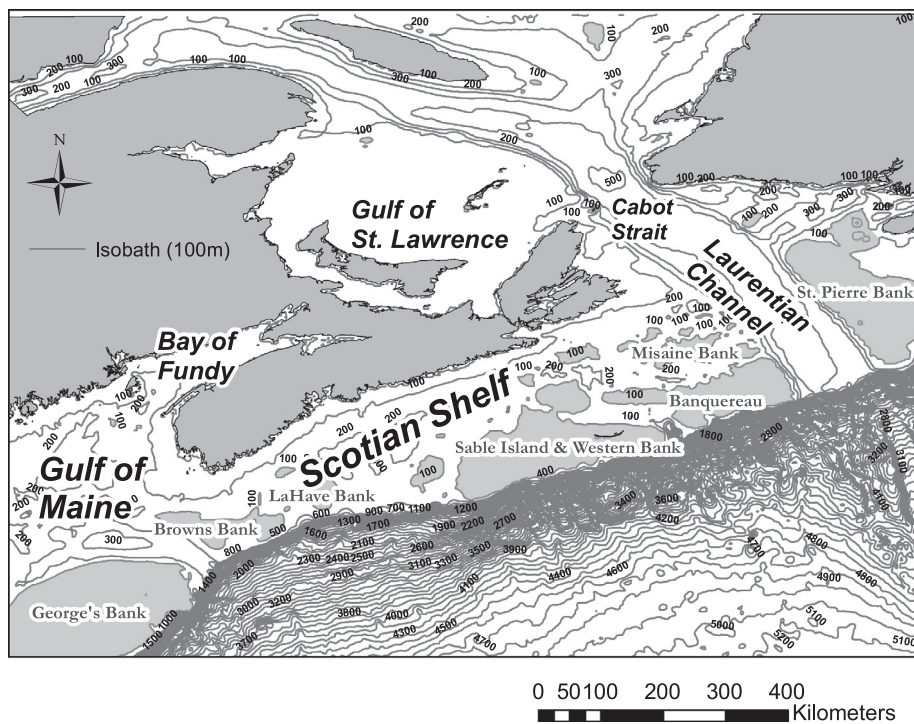


Fig. 1. Coastline (dark gray shading), bathymetry (indicated by 100 m isobath lines), and the location of key seabanks (100 m depth, light gray shading) within the Canadian Maritimes.

December, 2007); (2) perpendicularly to the vessel track, as with swimming birds (56% of track lengths, between January and July, 2008); or (3) radially as point counts (35% of track lengths, after July, 2008). As the detection process differed depending upon method, density of flying and swimming birds were estimated separately and summed to compute total density. As $1^0 \times 1^0$ grid cells were used to summarize densities (see below), many cells combined density estimates based on multiple methods for measuring distance to flying birds. With those blocks, estimates were averaged, weighted by survey effort (using kilometres surveyed).

2.3. Density estimation

Distance 6 Release 2 (Thomas et al., 2010) was used to estimate seabird densities, stratified by season and $1^0 \times 1^0$ grid cell. The choice of $1^0 \times 1^0$ grid was a trade-off between the desire for geographic detail and the need to maintain good estimator precision and uniform coverage of the study area while avoiding the creation of cells with too few observations. Additionally, this also ensured adequate overlap with oil pollution data (Section 2.8). To our knowledge, this is the first attempt to apply a distance sampling framework over such a wide area, and necessitated the computation of 72 separate projects using program Distance 6 (3 species \times 4 seasons \times 3 survey types \times 2 components). We feel that the acquisition of seasonally-specific seabird density estimates, simultaneously controlling for the detection bias induced by differences in sea conditions, justified the effort. Starting with basic key functions (half-normal), optional series expansion terms were chosen from one of three families (cosine, hermite, or polynomial). Visual inspection and the χ^2 goodness-of-fit test were used to assess detection function model fit, and then an attempt was made to improve fit by either choosing a different key function (hazard rate or uniform with optional series expansion terms) or by including explanatory covariates (such as wind speed, sea state, wave height and/or observer) using the multi-covariate distance sampling engine (Marques et al., 2007). Analyses were conducted

separately for each of the sampling regimes (birds on water vs. flying birds under three different distance measurement protocols, see Section 2.2), and yielded separate estimates of detection probability (\hat{p}), mean flock size (\bar{s}), as well as estimated density of flocks ($\hat{D}_f \pm SE$). Using \hat{p} , as well as information about average flock size (\bar{s}), seabird densities were corrected for the proportion of birds present but not observed and densities estimated (by program Distance) as: $\hat{D} = \hat{D}_f \bar{s}$. Grid cells with density estimates based on multiple methods for measuring distance to flying birds were averaged, weighted by survey effort (using kilometres surveyed).

2.4. Environmental data

A wide range of environmental variables have been implicated to explain seabird distribution. Previous studies have either focused on modelling distributional patterns purely as a function of geographic location (e.g., Certain et al., 2007), as a mixture of geographic location and oceanography and physiography (e.g., Spear and Ainley, 2005) or more commonly, purely as a function of oceanography and physiography (e.g., Huettmann and Diamond, 2001; Yen et al., 2004a; Yen et al., 2004b; O'Hara et al., 2006; Oppel et al., 2012; Renner et al., 2013; Chivers et al., 2013). The scale at which models are constructed are also highly variable, with grids ranging from 1 km^2 (e.g., Skov et al. 2008), to 400 km^2 (e.g., Fauchald et al., 2002), to $1^0 \times 1^0$ (Huettmann and Diamond, 2001) and even $2^0 \times 2^0$ (Spear and Ainley, 2005).

The $1^0 \times 1^0$ analysis grid (Section 2.3) combined 3.5 years of survey data within each grid cell. Therefore, we sought environmental variables that matched this scale of spatial and temporal resolution (Table 1). Shelf breaks and associated frontal zones can lead to local concentrations of seabirds (Skov and Durinck, 1998). In the study area, the shelf break is a prominent physical feature with a well documented impact on the distribution of local seabirds (Brown, 1988a, 1988b; Huettmann and Diamond, 2001), acting as a proxy for persistent hydrographic features such as fronts (Nur et al., 2011). This shelf break was identified using the

Table 1
Environmental variables used to construct seabird species distribution models.

Abbreviation	Variable	Unit	Data provider	Website address
DEPTH	Mean seadepth	m	National Geophysical Data Center (NOAA)	http://www.ngdc.noaa.gov/mgg/global/etopo2.html
PBANK	Percentage of area with seabank	%	GIS-derived from 100 m seadepth contours	NA
RUGGED	Melton ruggedness index	Standardized variation	GIS-derived product of DEPTH data	Melton (1965)
SEASON	Categorical variable (fall, spring, summer, winter)	Dummy indicator variable	NA	NA
SHELFDIST	Mean distance to shelfbreak (300 m isobath)	m	GIS-derived product	NA

300 m isobath, and the “Euclidean Distance” tool in ArcGIS 9.3 (ESRI, 2008) used to produce the variable SHELFDIST. As with the shelf break, sea banks are important sources of nutrient upwelling and are well documented as primary spawning areas for many species of commercial fish (Ashmole, 1971; Brander and Hurley, 1992). To model the potential influence of sea banks on seabirds, 100 m isobath polygons were identified within the GIS, corresponding to the banks labeled in Fig. 1. The proportion of each $1^\circ \times 1^\circ$ prediction cell intersecting with sea bank polygons was then calculated using the “Zonal Statistics” tool in ArcGIS 9.3 (ESRI, 2008), resulting in variable PBANK. Bathymetry and seafloor ruggedness are also persistent physical features with meaningful variation at $1^\circ \times 1^\circ$ scales. Data from the ETOPO2v2c data set of 2-min resolution distributed by the National Geophysical Data Centre (NGDC 2009) were integrated within a GIS to produce the variable DEPTH. In keeping with earlier findings of the potential importance of sea floor “roughness” (e.g., Yen et al., 2004b) a ruggedness index (RUGGED) was created using the method of Melton (1965):

$$\text{RUGGED} = \frac{\max(\text{DEPTH}) - \min(\text{DEPTH})}{\sqrt{\text{Area}}} \quad (1)$$

Earlier atlases (Lock et al., 1994) have revealed seasonal patterns in broad-scale distribution. We captured this effect using variable SEASON, which was defined using the following categories: “spring” (March–April), “summer” (May–August), “fall” (September–October) and “winter” (November–February). In the case of negative binomial and hurdle generalized linear models (Section 2.5), fall densities served as the reference category for assessing seasonal effects.

2.5. Model construction

Species distribution models (SDMs) were constructed using the base library of the R Statistical Package (R Development Core Team, 2009) as well as the MASS library of Venables and Ripley (2002). The `glm.nb` function (Venables and Ripley, 2002) was used to construct negative binomial, generalized linear models (NB GLMs). This approach uses a combination of maximum likelihood to estimate the overdispersion parameter (θ) and iteratively-reweighted least squares to estimate the model coefficients. Because of the overdispersion parameter, NB GLMs are more flexible than Poisson GLMs in dealing with extra count variance (Hilbe, 2008).

Hurdle NB models (referred to henceforth as ‘Hurdle’ models) were implemented using the `pscl` library of Zeileis et al. (2008). These two-component models apply a binomial GLM to the ‘zero’ (absence) vs. non-zero ($y > 0$) observations, and a NB GLM to the non-zero ($y > 0$) count data. In effect, an observation must be non-zero in order to pass over the “hurdle” (Zuur et al., 2009) and be modelled as a count-generating process. Predicted seabird density was generated for each location as:

$$y_{\text{exp}} = \mu * \frac{\pi}{p_0} \quad (2)$$

where y_{exp} is the expected number of seabirds, μ is the expected number based on the NB count process, π is the probability of a non-zero count based on the binomial occurrence model, and p_0 is the probability of a non-zero count based on the negative binomial count model (using the notation of Zuur et al., 2009). Final Hurdle models, therefore, involve estimation of two sets of coefficients: one for the binomial (the presence-absence part), and one for the NB GLM (the abundance part).

No variable selection algorithms were used to determine the covariates to retain/discard in the final NB GLM and Hurdle models. Instead, predictive models were constructed using an information-theoretic approach similar to that of Gray et al. (2010) and Lieske et al. (2012). An all-combinations algorithm, implemented in the R Statistical Package by D. Lieske (available upon request), computed predictive models for all possible-combinations and sorted them (lowest to highest) on the basis of Akaike Information Criterion (AIC) values. Only the top 95% of models were retained, and the weighted average of each coefficient was computed using the AIC weight, W_i (Burnham and Anderson, 2002: 152). In the case of the hurdle algorithm, model-averaged coefficients were computed for both the binomial and NB GLM components. The advantage of this approach is that it eliminates the arbitrariness associated with stepwise model selection, and incorporates uncertainty directly into the averaging of coefficients via AIC-based weightings. It also reduces reliance on a single set of estimates from one particular model, which is advantageous when estimates differ markedly across models (Burnham and Anderson, 2002: 150). In the authors’ experience, it is practical to conduct model-averaging with as many as 16 covariates, though the processing time exceeded 11 min on a 64-bit Intel Core i5-3317U CPU with 1.70 GHz clock speed and 6 GB of random access memory. Certain types of models are not amenable to an AIC-based model averaging approach, e.g., non-linear models or pseudo-likelihood models that do not calculate AIC values.

Random forests (RF, Breiman, 2001) is a tree-based method that employs bagging to aggregate the results of multiple, independently generated classification trees (Hastie et al., 2001). An extra level of randomness is introduced by randomly selecting the covariates used to construct each tree (Breiman, 2001). This approach constitutes a machine-learning alternative to NB and Hurdle methods, and generates novel predictions by “dropping” new inputs into the RF classifier to determine the majority class or value across the trees (Breiman, 2001). The *randomForest* library of Liaw and Wiener (2002) was used to construct RF models. Library *randomForest* also provides information on variable importance, which is determined by how much prediction error increases when testing data for that variable is permuted while all others are left unchanged (Liaw and Wiener, 2002). Empirical evidence suggests the random forest is sensitive to the choice of the number of covariates (Oppel and Huettmann 2010), but the number of trees comprising the RF classifier can be influential. For each species we examined variable combinations of the number of covariates (1–4) and number of trees (500, 1000, and 2000) and inspected the predictive accuracy of the results (Section 2.6). On the basis

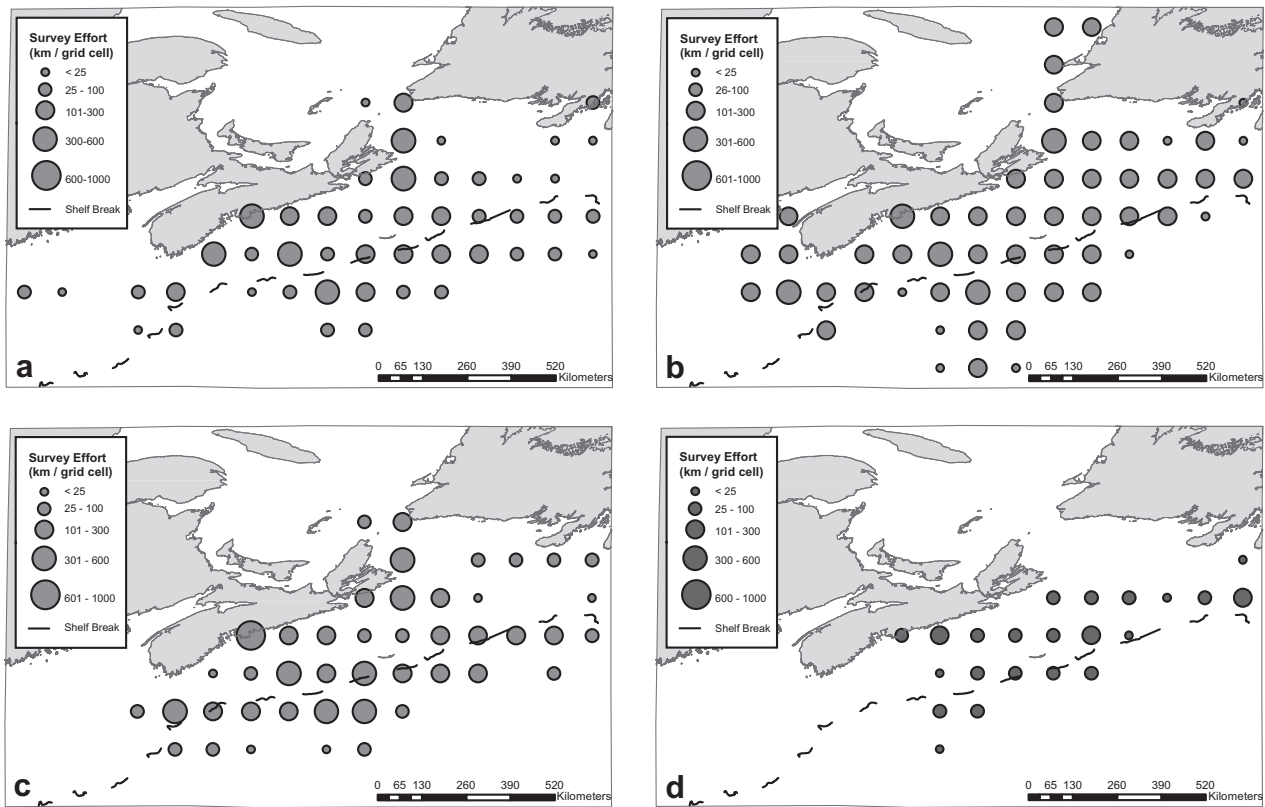


Fig. 2. Seasonally-specific survey effort, indicated by the total kilometers of transects surveyed in each $1^{\circ} \times 1^{\circ}$ grid cell over the period March 1, 2006 to October 31, 2009 (a: March–April “spring”, b: May–August “summer”, c: September–October “fall”, d: November–February “winter”).

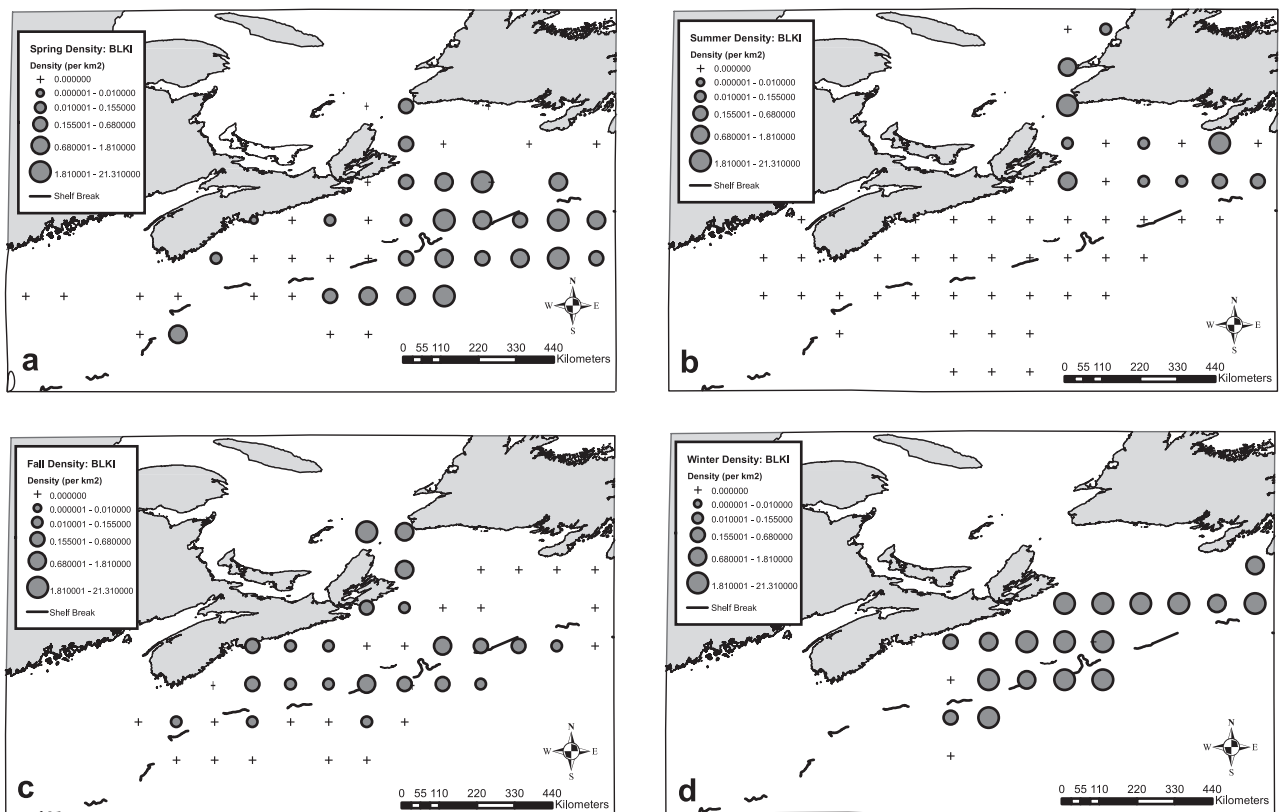


Fig. 3. Density of Black-legged Kittiwake (per km^2) for $1^{\circ} \times 1^{\circ}$ survey grid cells, for spring (a, March–April), summer (b, May–August), fall (c, September–October), and winter (d, November–February).

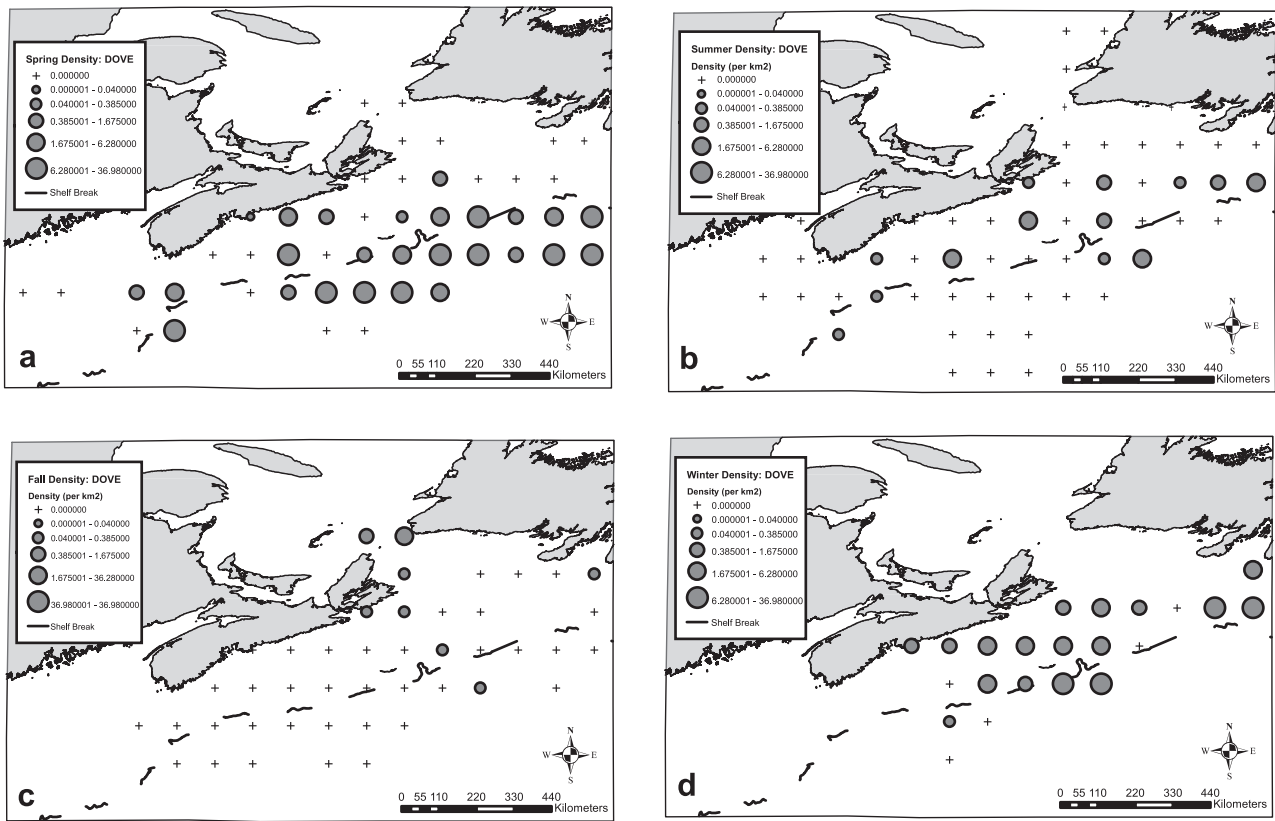


Fig. 4. Density of Dovekie (per km²) for 1° x 1° survey grid cells, for spring (a, Mar–April), summer (b, May–August), fall (c, September–October), and winter (d, November–February).

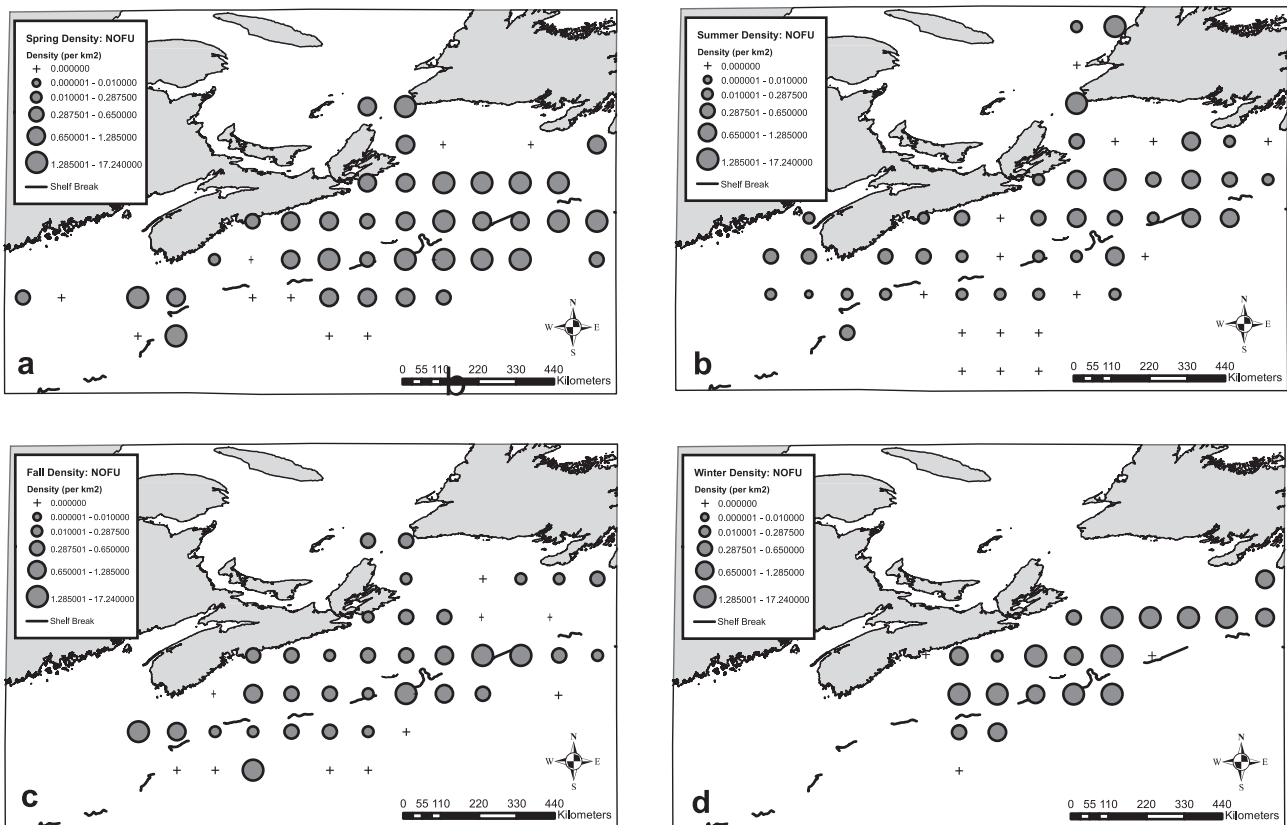


Fig. 5. Density of Northern Fulmar (per km²) for 1° x 1° survey grid cells, for spring (a, March–April), summer (b, May–August), fall (c, September–October), and winter (d, November–February).

of these results, 500 trees were constructed for NOFU, and 1000 trees for both BLKI and DOVE. One covariate was sampled for NOFU, four for BLKI, and three for DOVE.

Ensemble predictions were generated by taking the arithmetic mean of the predictions from the NB GLM, Hurdle GLM, and RF models (the so-called “committee averaging” method, Araújo and New, 2007; Das et al., 2008), weighted by the predictive power of each method (Section 2.6). A recent comparative analysis performed by Marmion et al. (2009) indicated that other measures of calculating consensus (e.g., the use of median values) performed about the same or worse, which is consistent with the general findings reported in the ensemble forecasting literature (Clemen, 1989).

2.6. Model evaluation

Observed (‘optimistic’) predictive power was assessed using the correlation coefficient (r_{obs}) of observed and predicted densities. To correct for model overfitting (Harrell, 2001), we constructed 100 independent, random samples of training and testing data (80% and 20% of original data, respectively), and measured bias by calculating the average difference in correlation coefficient:

$$bias = \frac{1}{n} * \sum_{iter=1}^n r_{obs} - r_{iter} \quad (3)$$

Bias-corrected correlation coefficients (r_{corr}), equivalent to $r_{obs} - r_{adj}$, defined the weights for calculation of the ensemble prediction (Section 2.5).

2.7. Assessment of potential sensitivity

The ensemble predictions from each of the species distribution models were combined to produce seasonally-specific sensitivity maps. Potential sensitivity was quantified for each location in the prediction surface as the sum of the relative importance of that spatial unit (R) for each of the k species:

$$Sensitivity\ index = \sum_{i=1}^k R_i \quad (4)$$

where R was defined as predicted abundance, standardized by scaling from 0 to 1 to allow comparisons between species:

$$R = \frac{lbound}{\max(lbound)}, \text{ where } lbound = y_{pred} - \min(y_{pred}) \quad (5)$$

The scaling of ensemble model predictions (Eq. (5)) was based on all values pooled, not just those from specific seasons. Adopting this approach ensured that seasonal shifts in seabird community usage patterns could be readily visualized and detected. It should be noted that when data is available to justify the use of species-specific sensitivities to particular hazards, Eq. (4) can be modified to incorporate species-specific weightings.

2.8. Assessment of oiling risk

Detection of chronic, small-scale oil pollution is challenging given the size of the marine area and the fact that the probability of detecting oil discharges is low and highly variable. Using aerial and satellite surveillance data gathered from offshore Nova Scotia between June, 2003 and September, 2008, previous work calculated an oil loading index (Lieske et al., 2011):

$$Oil\ loading_i = \text{Number of oiling events}_i / \text{number of surveys}_i \quad (6)$$

for each 50 km × 50 km grid cell, i (738 in total). While seabird (Section 2.2) and oil pollution monitoring was analyzed at slightly different scales ($1^0 \times 1^0$ vs. 50 km × 50 km), both information

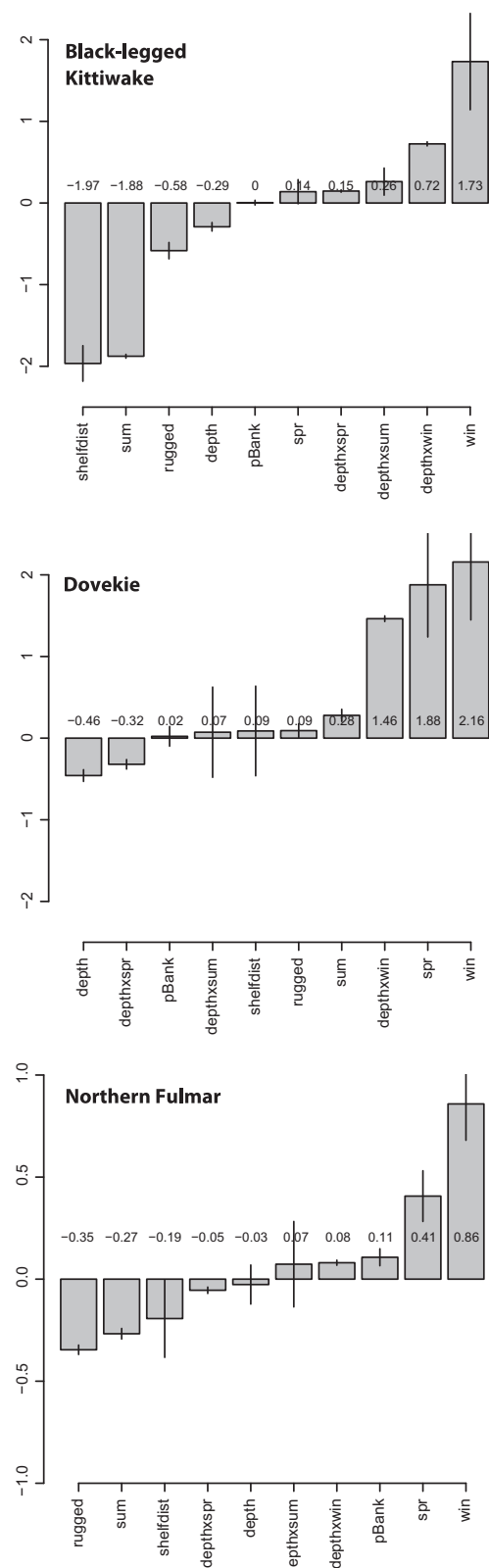


Fig. 6. Variable importance for negative binomial (NB) generalized linear-based SDMs, as indicated by regression coefficients for standardized covariates. Also shown are the estimated confidence intervals (95%) derived from testing and training data.

sources complemented each other in the sense of being spread over the same spatial region as well as overlapping in time. Oil loading information from Lieske et al. (2011) was used to produce the oil risk map.

2.9. Assessment of oil pollution vulnerability

'Vulnerability' is a function of the presence of a sensitive ecological feature as well as the presence of a risk or hazard (Turner et al.,

2003; Zacharias and Gregr, 2005). Within this framework, a species can be sensitive to oil pollution because of its foraging mode, but not vulnerable if it forages in a marine area devoid of oil pollution. For a GIS-based analysis spread over a very large area, indices are a

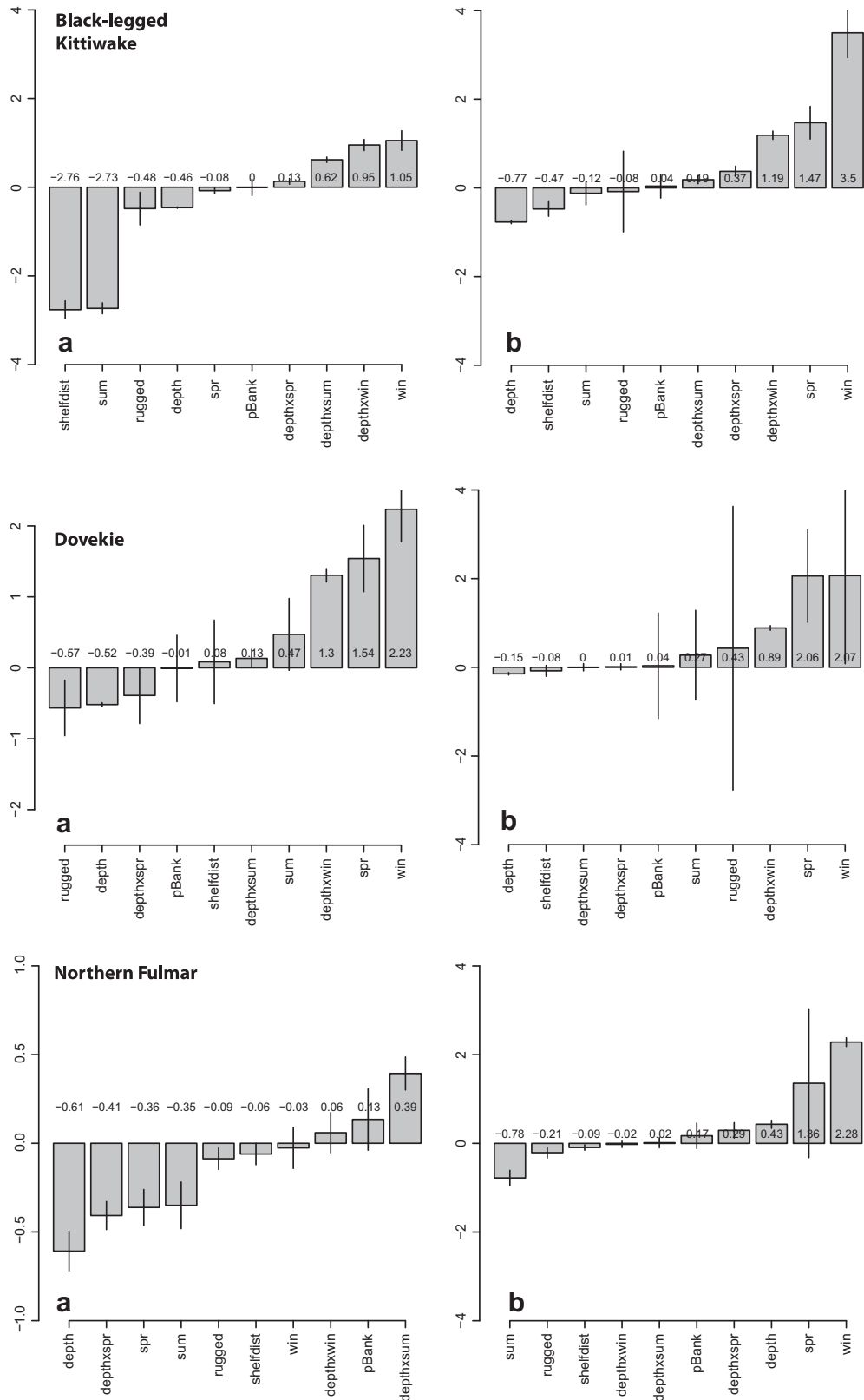


Fig. 7. Variable importance for the hurdle-based SDMs, as indicated by the regression coefficients for standardized covariates. Results are presented for both the presence component (a, binomial distribution) and the count component (b, negative binomial distribution). Also shown are the estimated confidence intervals (95%) derived from testing and training data.

tractable way to approximate oil-pollution vulnerability at particular locations. For instance, Zacharias and Gregr (2005) produced vulnerability maps by multiplicatively overlaying information about sensitivity (regression model predictions of whale presence) and stressor intensity (hypothesized distance-decay relationship characterizing the impact of acoustic pollution). We have applied a similar approach, but instead used relative abundance as a measure of sensitivity (scaled from 0 to 1 to make it possible to compare between species), and an oil-pollution index from Lieske et al. (2011, Section 2.8) to represent risk:

$$\text{Vulnerability index} = \text{Sensitivity index} * \text{Oil Loading index} \quad (7)$$

3. Results

3.1. Species observations

Distance-sampling methods allowed for the estimation of seabird densities, by species group and season, at a $1^{\circ} \times 1^{\circ}$ resolution. Survey effort varied somewhat from season to season (Fig. 2), with sampling most extensive during the summer period and least during the winter period. The bulk of the surveys occurred across the Scotian Shelf, though the Cabot Strait, the offshore of southwest Nova Scotia, and the Laurentian Channel were also surveyed. Inspection of the resulting density distributions suggest significant offshore aggregations of Black-legged Kittiwake in the winter and spring seasons (Fig. 3), and concentrations of Dovekie on the shelf and shelf break during the winter-spring period (Fig. 4). Northern Fulmar occurred throughout the year, but were also most abundant during the winter-spring (Fig. 5).

3.2. Model construction and evaluation

Season was, in general, an important covariate (Figs. 6 and 7), though it played a minor role in RF predictions for the Northern Fulmar (Fig. 8). Relative to fall occurrences, all three species showed strong winter responses under NB models (Fig. 6). The same was true for Hurdle models (Fig. 7), though uncertainty in the coefficients for the count-component tended to be higher. In the case of the Northern Fulmar, positive association with winter manifested in the count rather than the presence portion of the model (Fig. 7). Black-legged Kittiwake were negatively associated with distance to shelf break (SHELFDIST, Figs. 6 and 7), though the covariate played a lesser role in RF models (Fig. 8). DEPTH was universally important regardless of modelling method, while sea floor ruggedness (RUGGED) was important in some combinations (e.g., Northern Fulmar RF model, Fig. 8).

No one algorithm (NB, Hurdle, RF) exhibited consistently superior performance over all species, though the Hurdle models tended to under-perform relative to NB and RF (Table 2, Fig. 9). Ensemble predictions, weighted by bias-corrected r_{corr} values, performed the best for the Black-legged Kittiwake and Dovekie. For the Northern Fulmar, agreement was about the same for NB and ensemble approaches. Overall explanatory power was highest for the Dovekie, which yielded a $r = 0.632$ ($r^2 = 0.40$), followed by the Black-legged Kittiwake ($r^2 = 0.30$) and Northern Fulmar ($r^2 = 0.18$).

3.3. Assessment of species vulnerability

Composite species sensitivity maps revealed some striking seasonal relationships (Figs. 10a–d and 11). Composite abundance peaked in spring and winter, shifting into the pelagic zone and away from the shelf break in spring (Fig. 10a), but concentrating over the shelf during winter (Fig. 10d). The presence of relatively fewer individuals in the summer and fall would serve as a

protective factor in the face of oil pollution risk during those times (Fig. 11), though this group may still be exposed to oiling risks in breeding and migratory waters. The known oiling hazard, based on the combination of aerial and satellite surveillance, revealed relatively heavy pollution in the vicinity of major ports (e.g., Halifax and Sydney, Nova Scotia), and lower but persistent and widespread pollution throughout the shelf (Fig. 10e). The analysis of oiling hazard was based on data that was not tabulated by season, but it still illustrated important patterns of risk. Winter composite vulnerability was approximated by multiplying winter sensitivity (Fig. 10d) with oiling hazard (Fig. 10e), resulting in the composite vulnerability map of Fig. 10f.

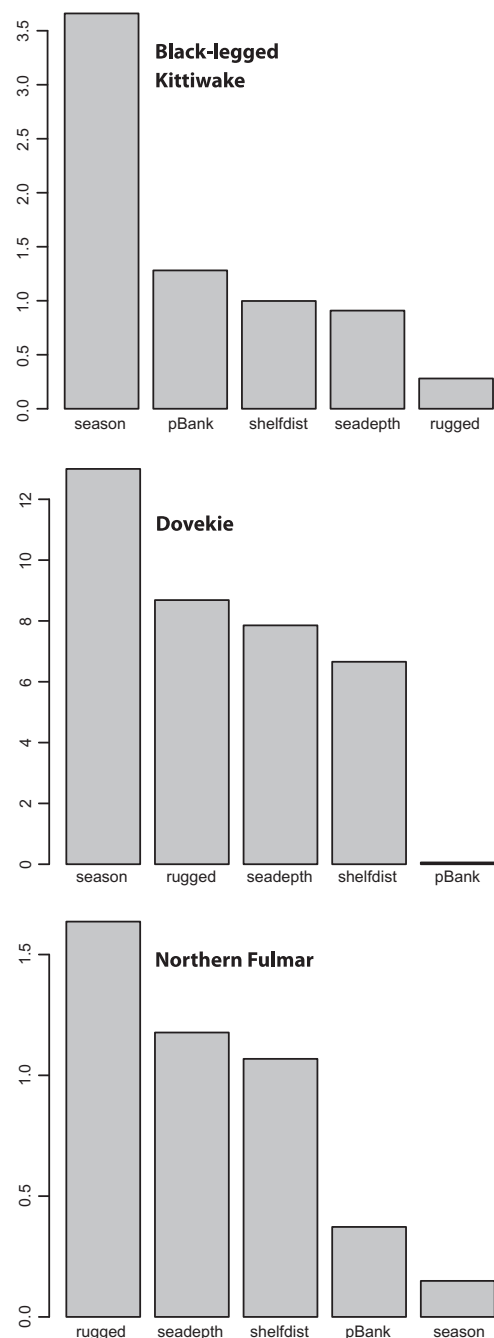


Fig. 8. Variable importance for species distribution models developed using random forest machine learning. "Importance" was assessed as the percentage change in prediction error when each variable is permuted out of the set of covariates used by the algorithm classifier.

4. Discussion

Conservation planning is a spatially-explicit exercise (Magness et al., 2011). The effective management of human activity in marine environments, e.g., through the designation of marine protected areas (MPAs), requires a synthesis of all available information (Araújo and New, 2007; Ronconi et al., 2012), and SDMs offer a powerful way to combine biological surveys with environmental information to better understand habitat usage through space and time. This study demonstrates an efficient framework for integrating multiple species data from two different monitoring programs across several implementation steps: GIS-based data management, state-of-the-art modelling, and production of a key spatial planning product. Each aspect of this framework will be discussed in turn.

Through a regular, grid-based approach (“Eulerian” data, see Tremblay et al., 2009) data gathered through ship-based surveys

were aggregated across a $1^0 \times 1^0$ grid to allow accurate density estimation. Factors which impact seabird detection, e.g., sea condition, or whether a bird was in the air or on the water, were statistically accounted for as covariates using distance sampling (Buckland et al., 2001; Marques et al., 2007). While unable to account for all the variability in distribution, season, sea depth and distance to shelf breaks were important and influential predictors. Proximity to the shelf break, for instance, stood out as an important factor influencing Black-legged Kittiwake numbers, while seafloor ‘ruggedness’ (RUGGED) with associated with Northern Fulmar RF models. Predictive power was highest for two of the three species’ ensemble models, yielding correlation coefficients as high or higher as SDM studies conducted at finer scales (e.g., Yen et al., 2004b; Potts and Elith, 2006).

Despite an extensive history in other disciplines (e.g., macroeconomics and psychology, see Clemen, 1989), ensemble prediction/forecasting has only recently been applied in SDM

Table 2
Agreement between predicted and observed seabird densities, as measured by the correlation coefficient (r). One hundred iterations of training and testing data were used to assess the degree of over- (or under-) optimism in agreement measures (bias), which was used to correct the original correlation measures (r_{corr}). The r_{corr} measures were used to define the weights for calculation of the ensemble model predictions.

Method	Species								
	Black-legged Kittiwake			Dovekie			Northern Fulmar		
	r	bias	r_{corr}	r	bias	r_{corr}	r	bias	r_{corr}
NB GLM	0.544	0.026	0.518	0.604	0.020	0.584	0.431	0.13	0.305
Hurdle	0.533	0.10	0.430	0.576	0.018	0.558	0.371	0.089	0.282
RF	0.508	-0.003	0.512	0.614	-0.003	0.617	0.357	-0.069	0.426
Ensemb.	0.549			0.632			0.430		

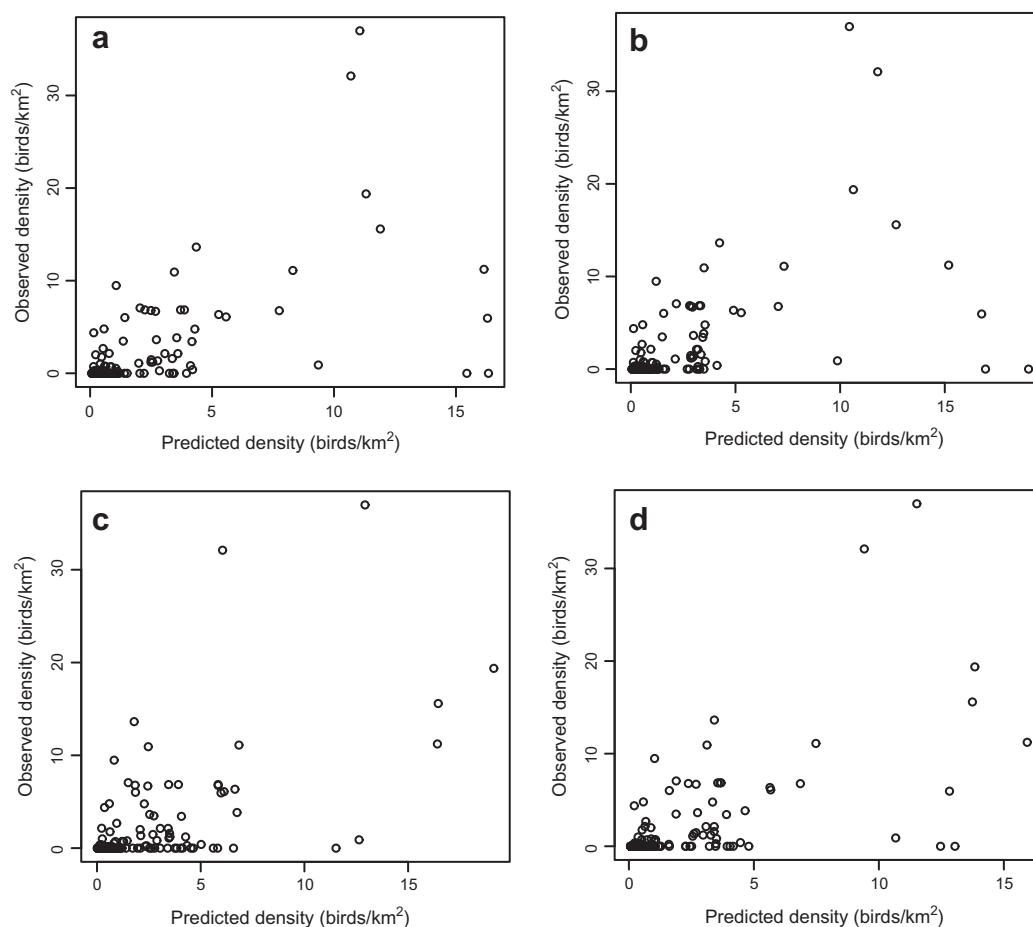


Fig. 9. Scatterplot of observed versus predicted Dovekie densities (birds/km²), based on (a) negative binomial (NB) glm, (b) Hurdle, (c) Random Forest (RF), and (d) ensemble methods.

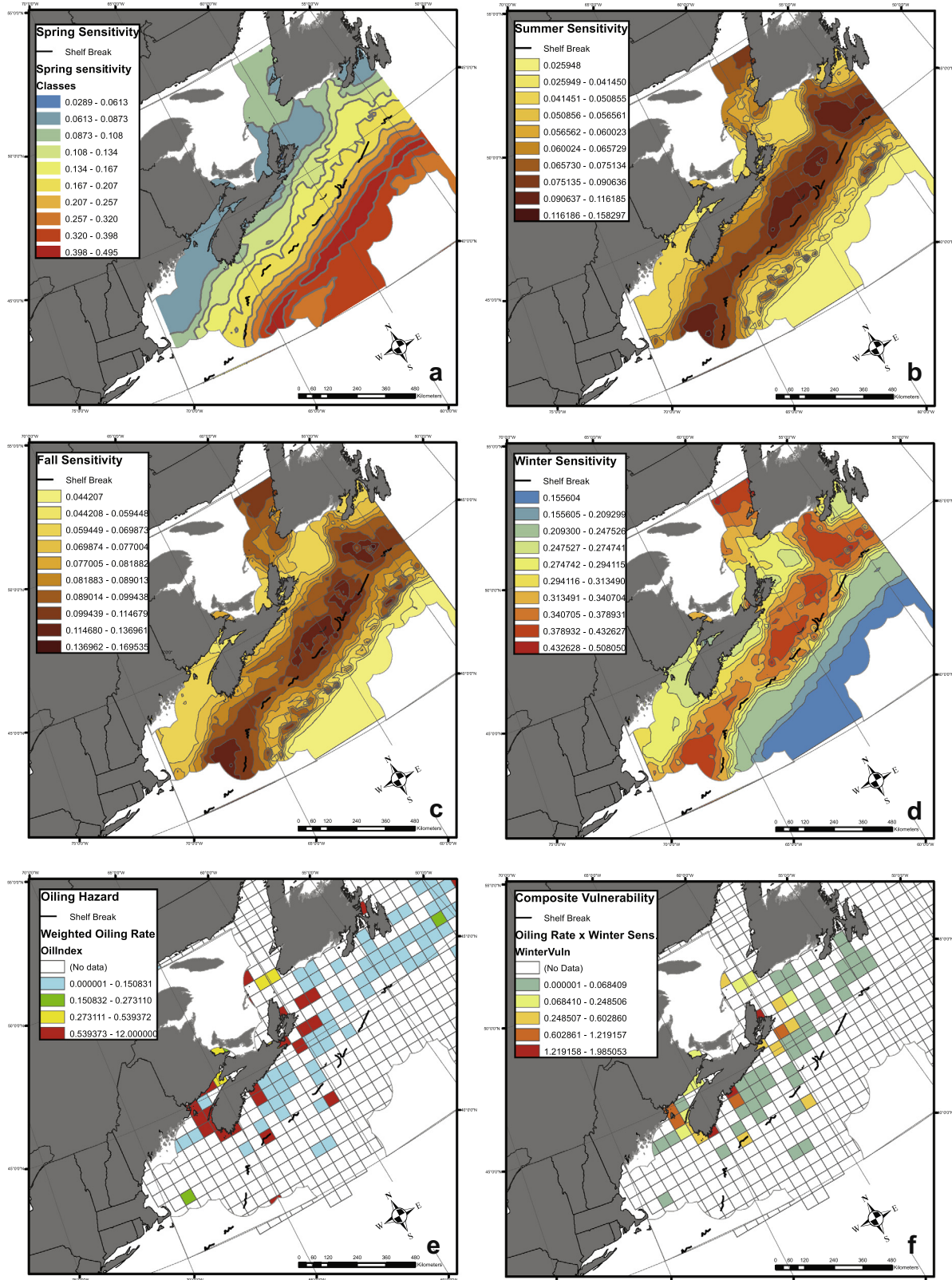


Fig. 10. Seasonally-specific sensitivity maps, based on the sum of the relative densities derived from ensemble model predictions for the three species. Seasons were defined as: (a) spring (March–April), (b) summer (May–August), (c) fall (September–October), and (d) winter (November–February). Also shown is overall oiling hazard (e), reproduced from [Lieske et al. 2011](#)) and the composite vulnerability index for winter (f).

([Araújo and New, 2007](#); [Oppel et al., 2012](#)). The focus of most SDM papers published in the first decade of the 21st century has concentrated on comparing relative performance (e.g., [Elith et al.,](#)

[2006](#)). As a consequence, the SDM literature is replete with a wide array of single-method modelling approaches where no attempt is made to gather the “consensus opinion” presented by a diversity of

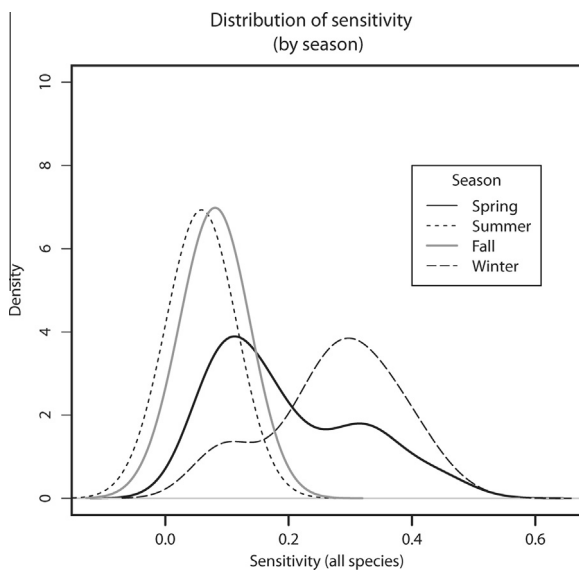


Fig. 11. Cross-seasonal summary of overall species sensitivity, illustrating peak sensitivity in spring and winter periods when species are the most common.

algorithms. While individual methods might be well calibrated and offer high predictive accuracy for the data at hand, predictive power may be substantially poorer in unsampled locations, or under changing environmental conditions (Marmion et al., 2009). Jones-Farrand et al. (2011) argue that basing conservation decisions on the basis of predictions from a single model is “risky”, in that problems are formulated under a specific set of objectives, constraints and assumptions that may not apply more generally. Ensemble approaches have an excellent track record for producing robust predictions, helping to reduce uncertainty and, in the process, increasing confidence in the decisions derived from them. In this study, no one algorithm (NB, Hurdle, RF) exhibited consistently superior performance over all species, though the Hurdle models tended to under-perform relative to NB and RF. Simple averaging of the NB, Hurdle and RF predictions – weighted by bias-corrected predictive accuracy – produced ensemble predictions that were as good or better than the best individual modelling method.

Deriving inferences from model predictions requires useful spatial products. In this study, predicted abundance based on the ensemble model was used to define the relative importance of $n = 1243$ locations in the study area for each species independently. Scaling was applied, based on the entire range of predicted abundances across all seasons, in order to rank the relative usage of each of these locations. Potential sensitivity to chronic oil pollution was then assessed as the sum of the relative usage across species, enabling the simultaneous visualization of the distributional patterns for the group. The composite picture that emerges is one of shifting temporal and spatial patterns of abundance. During the winter and spring, the seabird community is enlarged by the influx of non-breeding individuals from multiple regions: Dovekies from Arctic breeding grounds, particularly Greenland (Gaston and Jones, 1998), Black-legged Kittiwake from the eastern North Atlantic (Bogdanova et al., 2011), and Northern Fulmar from the Canadian High Arctic (Mallory et al., 2008).

When the combined winter distribution of this group was multiplied with oiling hazard, vulnerability was revealed to be highest in the vicinity of major ports (e.g., Halifax and Sydney, Nova Scotia). But the vulnerability map also indicated lower but widespread levels of oiling risk throughout the shelf, presumably associated with persistently high levels of shipping traffic and

ongoing petroleum extraction and exploration. Given the association between sea bird distribution and the shelf break, we recommend that the shelf break region receive more intensive monitoring to characterize offshore oil pollution risk at finer spatial and temporal scales. Such information would help improve understanding of the risk exposure of birds foraging in offshore Nova Scotia. However, it should be noted that Scotian Shelf seabirds are not just vulnerable to oil pollution in Canadian maritime waters; migratory connectivity (Webster et al., 2002) will expose seabirds to a wider network of hazards throughout their range.

The value of at-sea surveys are clear when one considers how difficult it can be to assess the potential vulnerability of seabirds outside the breeding season, when species are widely distributed throughout remote locations (Votier et al., 2005) and others may be present as transient non-breeders. Future work should pursue greater integration of models from different data sets, which would allow broad as well as fine-scale patterns to be explored and important patterns to be detected. For instance, combining telemetry tracking information with broad-scale distributional information would allow for analysis of seabird activity at multiple scales, widening the range of hypotheses that could be tested. The techniques developed in this paper will not only enhance risk assessment at “local” scales, but can be also used to tie in risks that are transmitted over a much larger area through migratory connectivity. But for benefits to be maximized, seabird research programs need to be supported by the routine, coordinated acquisition of marine hazard data (e.g., fishing activity, oil discharges).

Acknowledgements

We acknowledge the financial support of the Environmental Studies Research Funds (ESRF, <http://www.esrfunds.org>), the Environmental Damages Fund of Environment Canada, the Science Horizons Youth Internship Program of Environment Canada, and the Marjorie Young Bell Faculty Fund of Mount Allison University. We thank A. Miller for GIS data management and analysis, as well as the EWHALE lab of the University of Alaska, Fairbanks for advice on the use of random forest machine learning. Access to vessels was kindly provided by the Canadian Coast Guard, Department of Fisheries and Oceans, and the Environmental Studies Research Fund; we wish to thank science staff, ships’ officers, and personnel for their generous support, in particular the Atlantic Zone Monitoring Program participants. Finally, we thank all the observers who participated in the ECSAS program, and braved often challenging conditions to gather the data reported here.

References

- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47.
- Ashmole, N.P., 1971. Sea bird ecology and the marine environment. *Avian Biol.* 1, 223–286.
- Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecol. Model.* 157, 101–118.
- Bakun, A., 1996. Patterns in the Ocean: Ocean Processes and Marine Population Dynamics. La Paz, BCS, Mexico, Centro de Investigaciones Biológicas del Noroeste.
- Bogdanova, M.I., Daunt, F., Newell, M., Phillips, R.A., Harris, M.P., Wanless, S., 2011. Seasonal interactions in the black-legged kittiwake, *Rissa tridactyla*: links between breeding performance and winter distribution. *Proc. Roy. Soc. B* 278, 2412–2418.
- Brander, K., Hurley, P.C.F., 1992. Distribution of early-stage Atlantic cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), and witch flounder (*Glyptocephalus cynoglossus*) eggs on the Scotian Shelf: a reappraisal of evidence on the coupling of cod spawning and plankton production. *Can. J. Fish. Aquat. Sci.* 49, 238–251.
- Breiman, L., 2001. Random forests. *Machine Learn.* 45, 5–32.
- Briggs, K., Tyler, W.M.B., Lewis, D.B., Carlson, D.R., 1987. Bird Communities at Sea off California: 1975 to 1983. *Studies in Avian Biology*, p. 11.
- Brown, R.G.B., 1980. Seabirds as Marine Animals. In: Burger, J., Olla, B.L., Winn, H.E. (Eds.), *Behavior of Marine Animals*, vol. 4. New York, pp. 1–39.

- Brown, R.G.B., 1988a. The influence of oceanographic anomalies on the distributions of Storm-Petrels (Hydrobatidae) in Nova Scotian waters. *Colon. Waterbirds* 11, 1–8.
- Brown, R.G.B., 1988b. Oceanographic factors as determinants of the winter range of the Dovekie (*Alle alle*) off Atlantic Canada. *Colon. Waterbirds* 11, 176–180.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L., 2001. Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Oxford University Press, Oxford, UK.
- Burnham, K., Anderson, D.R., 2002. Model Selection and Multimodal Inference: a Practical Information-theoretic Approach. Springer Verlag, New York.
- Camphuysen, C.J., 1998. Beached bird surveys indicate decline in chronic oil pollution in the North Sea. *Mar. Pollut. Bull.* 36, 519–526.
- Certain, G., Bellier, E., Planque, B., Bretagnolle, V., 2007. Characterising the temporal variability of the spatial distribution of animals: an application to seabirds at sea. *Ecography* 30, 695–708.
- Chivers, L.S., Lundy, M.G., Colhoun, K., Newton, S.F., Houghton, J.D.R., Reid, N., 2013. Identifying optimal feeding habitat and proposed Marine Protected Areas (pMPAs) for the black-legged kittiwake (*Rissa tridactyla*) suggests a need for complementary management approaches. *Biol. Conserv.* 164, 73–81.
- Clarke, E.D., Spear, L.B., McCracken, M.L., Marques, F.F.C., Borchers, D.L., Buckland, S.T., Ainley, A.D., 2003. Validating the use of generalized additive models and at-sea surveys to estimate size and temporal trends of seabird populations. *J. Appl. Ecol.* 40, 278–292.
- Clemen, R.T., 1989. Combining forecasts: a review and annotated bibliography. *Int. J. Forecast.* 5, 559–583.
- Das, S.K., Chen, S., Deasy, J.O., Zhou, S., Fang-Fang, Y., Marks, L.B., Marks, L.B., 2008. Combining multiple models to generate consensus: application to radiation-induced pneumonitis prediction. *Med. Phys.* 35, 5098–5109.
- Durazo, R., Harrison, N.M., Hill, A.E., 1998. Seabird observations at a tidal mixing front in the Irish Sea. *Estuar. Coast. Shelf Sci.* 47, 153–164.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.
- ESRI. 2008. ArcGIS Version 9.3. Environmental Systems Research Institute, Redlands.
- Exo, K.-M., Hüppop, O., Garthe, S., 2003. Birds and offshore wind farms: a hot topic in marine ecology. *Wader Study Group Bull.* 100, 50–53.
- Fauchald, P., Erikstad, K.E., Systad, G.H., 2002. Seabirds and marine oil incidents: is it possible to predict the spatial distribution of pelagic seabirds? *J. Appl. Ecol.* 39, 349–360.
- Fifield, D., Lewis, K.P., Gjerdrum, C., Robertson, G.J., Wells, R., 2009. Offshore Seabird Monitoring Program: Final Report. Canadian Wildlife Service, Atlantic Region, St. John's, Newfoundland.
- Franklin, J., 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, New York.
- Gaston, A.J., Jones, I.L., 1998. The Auks: Alcidae. Oxford University Press, New York.
- Gjerdrum, C., Fifield, D.A., Wilhelm, S.L., 2012. Eastern Canada Seabirds at Sea (ECSAS) Standardized Protocol for Pelagic Seabird Surveys from Moving and Stationary Platforms. Canadian Wildlife Service Technical Report Series No. 515. Atlantic Region. vi + 36 pp.
- Gray, T.N.E., Phan, C., Long, B., 2010. Modelling species distribution at multiple spatial scales: gibbon habitat preference in a fragmented landscape. *Anim. Conserv.* 13, 324–332.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Hamel, N.J., Burger, A.E., Charleton, K., Davidson, P., Lee, S., Bertram, D.F., Parrish, J.K., 2009. Bycatch and beached birds: assessing mortality impacts in coastal net fisheries using marine bird strandings. *Mar. Ornithol.* 37, 41–60.
- Haney, J.C., 1989. Iterative techniques for characterizing marine bird habitats with time-series of satellite images. *Colon. Waterbirds* 12, 78–89.
- Hannah, C.G., Shore, J.A., Loder, J.W., 2001. Seasonal circulation on the western and central Scotian Shelf. *J. Phys. Oceanogr.* 31, 591–615.
- Harrell Jr., F.E., 2001. Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer, New York.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.
- Hegel, T.M., Cushman, S.A., Evans, J., Huettmann, F., 2010. Current state of the art for statistical modelling of species distributions. In: Cushman, S.A., Huettmann, F. (Eds.), Spatial Complexity, Informatics, and Wildlife Conservation. Springer, pp. 273–312.
- Henkel, J.R., Sigel, B.J., Taylor, C.M., 2012. Large-scale impacts of the Deepwater Horizon oil spill: can local disturbance affect distant ecosystems through migratory shorebirds? *Bioscience* 62, 676–685.
- Hilbe, J.M., 2008. Negative Binomial Regression. Cambridge University Press, New York.
- Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83, 2027–2036.
- Huettmann, F., Diamond, A.W., 2001. Seabird colony locations and environmental determination of seabird distribution: a spatially explicit breeding seabird model for the Northwest Atlantic. *Ecol. Model.* 141, 261–298.
- Hunt, G.L., Schneider, D.C., 1987. Scale-dependent processes in the physical and biological environment of marine birds. In: Croxall, J.P. (Ed.), Seabirds: Feeding Ecology and Role in Marine Ecosystems. Cambridge University Press, pp. 7–41.
- Jones-Farrand, D.T., Fearer, T.M., Thogmartin, W.E., Thompson III, F.R., Nelson, M.D., Tirpak, J.M., 2011. Comparison of statistical and theoretical habitat models for conservation planning: the benefit of ensemble prediction. *Ecol. Appl.* 21, 2269–2282.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2/3.
- Lieske, D.J., Mahoney, M., Wilhelm, S.L., Weir, L., O'Hara, P., 2011. Eyes in the sky: a comparative spatial analysis of aerial and satellite surveillance of East Coast Canadian oil pollution. In: Proceedings of the 34th AMOP Technical Seminar on Environmental Contamination and Response, pp. 244–258.
- Lieske, D.J., Pollard, B., Gloutney, M., Milton, R., Connor, R., Dibblee, R., Parsons, G., Howerter, D., 2012. The importance of agricultural landscapes as key nesting habitats for the American Black Duck in Maritime Canada. *Waterbirds* 35, 525–534.
- Lock, A.R., Brown, R.G.B., Gerriets, S.H., 1994. Gazetteer of Marine Birds in Atlantic Canada: An Atlas of Seabird Vulnerability to Oil Pollution. Canadian Wildlife Service.
- Magness, D.R., Morton, J.M., Huettmann, F., 2011. How spatial information contributes to the conservation and management of biodiversity. In: Cushman, S.A., Huettmann, F. (Eds.), Spatial Complexity, Informatics, and Wildlife Conservation. Springer, pp. 429–444.
- Mallory, M., Akearok, J.A., Edward, D.B., O'Donovan, K., Gilbert, C.D., 2008. Autumn migration and wintering of Northern Fulmars (*Fulmarus glacialis*) from the Canadian high Arctic. *Polar Biol.* 31, 745–750.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* 15, 59–69.
- Marques, T.A., Thomas, L., Fancy, S.G., Buckland, S.T., 2007. Improving estimates of bird density using multiple-covariate distance sampling. *Auk* 124, 1229–1243.
- McKinnon, L., Gilchrist, H.G., Fifield, D., 2009. A pelagic seabird survey of arctic and sub-arctic Canadian waters during fall. *Mar. Ornithol.* 37, 77–84.
- Melton, M.A., 1965. The geomorphic and paleoclimatic significance of alluvial deposits in southern Arizona. *J. Geol.* 73, 1–38.
- NGDC, 2009. ETopo2v2 Global Gridded 2-minute Database. <<http://www.ngdc.noaa.gov/mgg/global/etopo2.html>> (accessed March 2010).
- Nova Scotia Museum of Natural History, T6.1. Ocean Currents. <<https://museum.gov.ns.ca/mnh/nature/nhns/t6/t6-1.htm>> (accessed March 2010).
- Nur, N., Jahncke, J., Herzog, M.P., Howar, J., Hyrenbach, K.D., Zamon, J.E., Ainley, D.G., Wiens, J.A., Morgan, K., Ballance, L.T., Stralberg, D., 2011. Where the wild things are: predicting hotspots of seabird aggregations in the California Current System. *Ecol. Appl.* 21, 2241–2257.
- O'Hara, P.D., Morgan, K.H., Sydean, W.J., 2006. Primary producer and seabird associations with AVHRR-derived sea surface temperatures and gradients in the southeastern Gulf of Alaska. *Deep-Sea Res.* 53, 359–369.
- Oppel, S., Huettmann, F., 2010. Using a random forest model and public data to predict the distribution of prey for marine wildlife management. In: Cushman, S.A., Huettmann, F. (Eds.), Spatial Complexity, Informatics, and Wildlife Conservation. Springer, New York, pp. 151–163.
- Oppel, S., Meirinho, A., Ramirez, I., Gardner, B., O'Connell, A.F., Miller, P.I., Louzao, M., 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biol. Conserv.* 156, 91–104.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Potts, J.M., Elith, J., 2006. Comparing species abundance models. *Ecol. Model.* 199, 153–163.
- R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <<http://www.r-project.org>> (accessed September, 2009).
- Renner, M., Parrish, J.K., Piatt, J.F., Kuletz, K.J., Edwards, A.E., Hunt Jr., G.L., 2013. Modeled distribution and abundance of a pelagic seabird reveal trends in relation to fisheries. *Mar. Ecol. Prog. Ser.* 484, 259–277.
- Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P., Richardson, A.J., 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Glob. Ecol. Biogeogr.* 20, 789–802.
- Ronconi, R.A., Lascelles, B.G., Langham, G.M., Reid, J.B., Oro, D., 2012. The role of seabirds in marine protected areas identification, delineation, and monitoring: introduction and synthesis. *Biol. Conserv.* 156, 1–4.
- Schneider, D.C., 1990. Seabirds and fronts: a brief overview. *Polar Res.* 8, 17–21.
- Shriner, S.A., Simons, T.R., Farnsworth, 2002. A GIS-based habitat model for Wood Thrush, *Hylocichla mustelina*, in Great Smoky Mountains National Park. In: Scott, J.M., Heglund, P.J., Morrison, M.L. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 529–535.
- Skov, H., Durinck, J., 1998. Constancy of frontal aggregations of seabirds at the shelf break in the Skagerrak. *J. Sea Res.* 39, 305–311.
- Skov, H., Humphreys, E., Garthe, S., Geitner, K., Grémillet, D., Hamer, K.C., Hennenke, J., Parner, H., Wanless, S., 2008. Application of habitat suitability modelling to tracking data of marine animals as a means of analyzing their feeding habitats. *Ecol. Model.* 212, 504–512.
- Smith, P.C., Petrie, B.D., 1982. Low-frequency circulation at the edge of the Scotian Shelf. *J. Phys. Oceanogr.* 12, 28–46.
- Spear, L.B., Ainley, D.G., 2005. At-sea distributions and abundance of tropicbirds in the eastern Pacific. *Ibis* 147, 353–366.

- Tasker, M.L., Hope-Jones, P., Dixon, T.J., Blake, B.F., 1984. Counting seabirds at sea from ships: a review of methods employed and a suggestion for a standardized approach. *Auk* 101, 567–577.
- Tasker, M.L., Camphuysen, C.J., Cooper, J., Garthe, S., Montevecchi, W.A., Blaber, S.J.M., 2000. The impacts of fishing on marine birds. *ICES J. Mar. Sci.* 57, 531–547.
- Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Bishop, J.R.B., Marquez, T.A., Burnham, K.P., 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. *J. Appl. Ecol.* 47, 5–14.
- Tremblay, Y., Bertrand, S., Henry, R.W., Kappes, M.A., Costa, D.P., Schaffer, S.A., 2009. Analytical approaches to investigating seabird-environmental interactions: a review. *Mar. Ecol. Prog. Ser.* 391, 153–163.
- Turner II, B.L., Kasperson, R.E., Matson, P.S., McCarthy, J.J., Corell, R.W., Christensen, L., Eckley, N., Kasperson, J.X., Luers, A., Martello, M.L., Polsky, C., Pulsipher, A., Schiller, A., 2003. A framework for vulnerability analysis in sustainability science. *Proc. Nat. Acad. Sci.* 100, 8074–8079.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Springer, New York.
- Votier, S.C., Hatchewell, B.J., Beckerman, A., McCleery, R.H., Hunter, F.M., Pellatt, J., Trinder, M., Birkhead, T.R., 2005. Oil pollution and climate have wide-scale impacts on seabird demographics. *Ecol. Lett.* 8, 1157–1164.
- Webster, M.S., Marra, P.P., Haig, S.M., Bensch, S., Holmes, R.T., 2002. Links between worlds: unraveling migratory connectivity. *Trends Ecol. Evol.* 17, 76–83.
- Wiese, F.K., Ryan, P.C., 2003. The extent of chronic marine oil pollution in southeastern Newfoundland waters assessed through beached bird surveys 1984–1999. *Mar. Pollut. Bull.* 46, 1090–1101.
- Wiese, F.K., Montevecchi, W.A., Davoren, G.K., Huettmann, F., Diamond, A.W., Linke, J., 2001. Seabirds at risk around offshore oil pollution platforms in the North-west Atlantic. *Mar. Pollut. Bull.* 42, 1285–1290.
- Yen, P.P.W., Huettmann, F., Cooke, F., 2004a. A large-scale model for the at-sea distribution and abundance of Marbled Murrelets (*Brachyramphus marmoratus*) during the breeding season in coastal British Columbia, Canada. *Ecol. Model.* 171, 395–413.
- Yen, P.P.W., Sydeman, W.J., Hyrenbach, K.D., 2004b. Marine bird and cetacean associations with bathymetric habitats and shallow-water topographies: implications for trophic transfer and conservation. *J. Mar. Syst.* 50, 79–99.
- Zacharias, M.A., Gregr, E.J., 2005. Sensitivity and vulnerability in marine environments: an approach to identifying vulnerable marine areas. *Conserv. Biol.* 19, 86–97.
- Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression models for count data in R. *J. Software* 27 (8).
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York.