

Manuscript Number:

Title: Maps, Models, and Marine Vulnerability: Assessing the Community Distribution of Seabirds at-Sea

Article Type: Full Length Article

Keywords: Species distribution modelling; multi-species distributions; ensemble prediction; seabirds; negative binomial GLM; hurdle, random forest; Maritime Canada

Corresponding Author: Dr. David John Lieske, Ph.D.

Corresponding Author's Institution: Mount Allison University

First Author: David John Lieske, Ph.D.

Order of Authors: David John Lieske, Ph.D.; David A Fifield; Carina Gjerdrum

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. Density estimates, derived from counts of species known to be particularly vulnerable to oil pollution, were used to construct multiple species distribution models (SDMs). The structural difficulties associated with sampling sparsely distributed individuals that can 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, and multi-species potential vulnerability maps developed to identify core areas. The maps confirmed the general importance of the shelf break and sea banks, but distribution was highly seasonally specific. A pattern of shifting presence emerged, with some species disappearing from the region (e.g., storm-petrels and shearwaters in winter) and others increasing at the same time (e.g., Black-legged Kittiwakes, *Rissa tridactyla*). Across all species, spring (March - April) stood out as a period of peak importance, though seasonal variation in the usage of the offshore marine environment may render the seabird community vulnerable to different threats throughout the year. This study illustrates how multi-species distribution modelling can predict seasonal usage patterns, and contribute to an awareness of the collective vulnerability of groups of seabirds to hazards distributed over wide geographic areas.

Dr. David Lieske
Department of Geography and Environment
Mount Allison University
144 Main Street
Sackville, New Brunswick, Canada
E4L 3G1

August 7, 2013

To the Editor(s),

On behalf of my colleagues and myself, please find attached the manuscript entitled "Maps, Models, and Marine Vulnerability: Assessing the Community Distribution of Seabirds at-Sea" for consideration in *Biological Conservation*.

This paper is an original application of ensemble modelling to assess the distribution of a community of seabirds that occur in Maritime Canada. As this is an area subject to significant marine traffic, chronic small-scale oil pollution, and ongoing oil and gas extraction and exploration, this group of seabirds (Black-legged Kittiwakes, Dovekie, murre, Northern Gannet and storm-petrels) constitute a vulnerable community. By combining the predictions from three modelling methods (negative binomial GLM, hurdle GLM, and random forest machine learning) to produce an ensemble prediction, we assessed the predictive accuracy of these techniques and demonstrate how multi-species vulnerability maps can be generated to identify core areas of conservation concern.

This work builds nicely on papers recently published in *Biological Conservation*, e.g., Oppel et al. (2012): Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds, *Biol. Cons.* 156, 91-104, and feel that the results will be of wide interest to its readership.

The manuscript is slightly less than 7500 words in length and has not been submitted for publication elsewhere.

Should you have any questions, please do not hesitate to contact me at the following address. I look forward to hearing from you in due course.

Sincerely,



David Lieske, Ph.D.
Department of Geography and Environment, Mount Allison University,
144 Main Street, Sackville, New Brunswick,
E4L 1A7, Canada: (506) 364-2315; fax: (506) 364-2625
dlieske@mta.ca

1 **Highlights**

2

- 3 ▪ Distribution models were constructed for a vulnerable group of seabirds.
- 4 ▪ Generalized linear, hurdle and random forest modelling methods were used.
- 5 ▪ Predicted seabird abundances were combined to form ensemble estimates.
- 6 ▪ Relative community importance for each regional $1^0 \times 1^0$ unit was calculated.
- 7 ▪ Community vulnerability maps showed strong seasonal and spatial patterns.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Article Type: Research Article

**Maps, Models, and Marine Vulnerability:
Assessing the Community Distribution
of Seabirds at-Sea**

David J. Lieske^{1,2}, David A. Fifield³, Carina Gjerdrum⁴

¹ Department of Geography and Environment, Mount Allison University, 144 Main Street, Sackville, NB, Canada, E4L 1A7.

² Corresponding author: dlikes@mta.ca

³ Department of Environment and Conservation, Government of Newfoundland and Labrador, P.O. Box 8700, St. John's, NL, Canada, A1B 4J6.

⁴ Environment Canada, Canadian Wildlife Service, 45 Alderney Drive, Dartmouth, NS, Canada, B2Y 2N6.

Keywords: Species distribution modelling; multi-species distributions; ensemble prediction; seabirds; negative binomial GLM; hurdle, random forest; Maritime Canada

1
2
3
34
5
35
7
36
9
37
10
38
12
39
14
40
15
41
17
42
19
43
20
44
22
45
24
46
25
47
27
48
29
49
30
50
32
51
34
52
35
53
37
54
39
55
41
56
43
57
44
58
45
59
46
60
48
61
50
62
52
63
54
64
55
65
56
57
66
59
60
61
62
63
64
65

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. Density estimates, derived from counts of species known to be particularly vulnerable to oil pollution, were used to construct multiple species distribution models (SDMs). The structural difficulties associated with sampling sparsely distributed individuals that can 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, and multi-species potential vulnerability maps developed to identify core areas. The maps confirmed the general importance of the shelf break and sea banks, but distribution was highly seasonally specific. A pattern of shifting presence emerged, with some species disappearing from the region (e.g., storm-petrels and shearwaters in winter) and others increasing at the same time (e.g., Black-legged Kittiwakes, *Rissa tridactyla*). Across all species, spring (March - April) stood out as a period of peak importance, though seasonal variation in the usage of the offshore marine environment may render the seabird community vulnerable to different threats throughout the year. This study illustrates how multi-species distribution modelling can predict seasonal usage patterns, and contribute to an awareness of the collective vulnerability of groups of seabirds to hazards distributed over wide geographic areas.

1
2
3
67
5
68
7
69
8
70
10
71
11
72
12
73
13
74
15
75
16
76
17
77
18
78
19
79
20
80
21
81
22
82
23
83
24
84
25
85
26
86
27
87
28
88
29
89
30
90
31
91
32
92
33
93
34
94
35
95
36
96
37
97
38
98
39
99
40
100
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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). Heavier-bodied, diving seabirds are at greater risk of exposure to these hazards by virtue of the amount of time they spend interacting with the sea-surface-air interface (Camphuysen, 1998) but they are also vulnerable as a result of low fecundity (Votier et al., 2005). 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 ('ships of opportunity') which may or may not be primarily tasked with gathering seabird information. Large gaps in knowledge inevitably occur, especially during the 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, 2011). Applying this methodology in marine contexts can be quite difficult, however. Marine environments are highly dynamic (Haney, 1989), and difficult to represent within a standard GIS. For instance, fronts, widely recognized as important areas of nutrient entrapment and increased biological productivity (Bakun, 1996), can shift position over a period of hours (Durazo et al., 1998). Species may respond by rapidly shifting their distributions to benefit from these changing patterns of food abundance. Additionally, seabirds simultaneously display sparse distributions over wide areas, as well as a tendency to form large localized aggregations 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). Take together, zero inflation and overdispersion pose serious problems for SDMs based on traditional techniques (Hilbe, 2008).

1
2
3
10¹⁴ Over the past 20-years, there has been a notable proliferation of algorithms for modelling animal occurrence and
5
10¹⁶ abundance (Hegel et al., 2010). Such methods range from presence-only techniques (e.g., maximum entropy or
7
10¹⁸ MAXENT, Phillips et al., 2006; ecological niche factor analysis or ENFA, Hirzel et al., 2002) to presence-absence
9
10²⁰ (logistic generalized linear or logistic generalized additive models), to abundance (Poisson or negative binomial
11
10²² regression). Machine learning methods, such as random forest (Breiman, 2001), can be applied in any of these cases.
13
10²⁴ The very large number of available methods renders it difficult to choose a technique (Jones-Farrand et al., 2011), so is
15
10²⁶ a decision often based on investigator preference or familiarity (Araújo and New, 2007; Jones-Farrand et al., 2011).
17
10²⁸ Current research, however, points to the advantages of combining the predictions of different modelling methods, a
19
10³⁰ process referred to as ensemble modelling (Araújo and New, 2007; Opperl et al., 2012) or decision fusion (Das et al.,
21
10³² 2008). This approach allows models to reinforce each other when consensus exists, while dampening divergent
23
10³⁴ predictions when they are discrepant. The implicit assumption is while the true underlying process is unknown, different
25
10³⁶ forecasting models are able to capture different components of the underlying signal (Clemen, 1989). The advantages
27
10³⁸ are expected to be maximized when different, but complementary, modelling algorithms are employed.

10⁴⁰ Marine SDMs are rare (Robinson et al., 2011), and SDMs based on shipboard surveys are even rarer (Opperl et al.,
42
10⁴² 2012). Due to four years of pelagic surveys conducted as part of the east coast Eastern Canada Seabirds at Sea
44
10⁴⁴ monitoring program, this study to was able to extend the approach developed by Opperl et al. (2012) and examine the at-
46
10⁴⁶ sea distribution of a range of seabirds that occur on or near the Scotian Shelf of Nova Scotia. Density data were
48
10⁴⁸ corrected for variation in detectability, and three different approaches selected for their theoretical suitability to deal
50
10⁵⁰ with zero-inflated, overdispersed data: negative binomial (NB) generalized linear modelling, Hurdle modelling, and
52
10⁵² random forest (RF) machine learning. Results from each of these methods were combined to produce an ensemble
54
10⁵⁴ prediction, the accuracy of which were carefully assessed and compared. In addition to providing a case-example of
56
10⁵⁶ how multi-method, multi-species models can be informatively combined, seasonally-specific community vulnerability
58
10⁵⁸ maps were generated to assist regional conservation planning.

10⁵⁹ 2. Methods 10⁶⁰

10⁶¹ 2.1. Data collection 10⁶²

10⁶³ The focal area was the Scotian Shelf with a maximum depth of about 200 m, extending 150-250 km offshore of the
10⁶⁴ province of Nova Scotia, Canada. Also included were portions of the Bay of Fundy as well as the Cabot Strait (Figure
10⁶⁵ 1). At the shelf break sea depth increases rapidly to abyssal levels exceeding 3000 m. The topography of the shelf,

1
2
3
135⁴ combined with tidal forces, wind, and fluctuations of offshore currents associated with the Gulf Stream, result in
136⁵ upwelling and partial gyres capable of retaining nutrients and concentrating marine biomass (Smith and Petrie, 1982;
137⁶ Hannah et al., 2001; NSMN, 2010). A number of banks occur throughout the shelf, some of which (e.g., Browns Bank,
138⁷ Sable Island Bank, Georges Bank) experience tidally-produced gyres of sufficient strength to also retain nutrients and
139⁸ concentrate prey (Hannah et al., 2001)

140⁹
141¹⁰
142¹¹
143¹²
144¹³
145¹⁴
146¹⁵ 2.2. *At-sea survey data*

147¹⁶
148¹⁷
149¹⁸
150¹⁹ The focal species of this study were: (1) Black-legged Kittiwake (*Rissa tridactyla*); (2) Dovekie (*Alle alle*); (3) murre
151²⁰ (Thick-billed, *Uria lomvia*; Common Murres, *U. aalge*; unidentified *Uria* sp.); (4) Northern Gannet (*Morus bassanus*);
152²¹ (5) storm-petrels (Wilson’s Storm-Petrel, *Oceanites oceanicus*; Leach’s Storm-Petrel, *Oceanodroma leucorhoa*;
153²² unidentified sp.) and (6) shearwaters (Greater Shearwater, *Puffinus gravis*; Sooty Shearwater, *P. griseus*; Manx
154²³ Shearwater, *P. puffinus*; Cory’s Shearwater *Calonectris diomedea*; Audubon’s Shearwater, *P. lherminieri*; unidentified
155²⁴ sp.), also grouped to enhance sample size and maximize precision of the density estimates (see Section 2.3). As a group,
156²⁵ these species exhibit high vulnerability to sea-surface oil pollution (Brown, 1980; Camphuysen, 1998), are a significant
157²⁶ presence on the Scotian Shelf, and encompass a cross section of ecological niches: e.g., surface feeders (Black-legged
158²⁷ Kittiwake, Leach's Storm-Petrel), deep-water pursuit divers (murre, Dovekie), plunge divers (Northern Gannet), and
159²⁸ surface pursuit divers (shearwaters).

160²⁹
161³⁰
162³¹
163³²
164³³
165³⁴
166³⁵
167³⁶
168³⁷ Observations from 76 surveys were gathered from March 1, 2006 to October 31, 2009 as part of the Eastern Canada
169³⁸ Seabirds at Sea (ECSAS) program (Fifield et al., 2009). The ECSAS survey protocol (Gjerdrum et al., 2012)
170³⁹ incorporates the recommendations of Tasker et al. (1984), as well as modern distance sampling techniques (Buckland et
171⁴⁰ al., 2001), and relies on ships of opportunity to conduct the survey.

172⁴¹
173⁴²
174⁴³
175⁴⁴
176⁴⁵
177⁴⁶
178⁴⁷
179⁴⁸
180⁴⁹
181⁵⁰
182⁵¹
183⁵²
184⁵³
185⁵⁴
186⁵⁵
187⁵⁶
188⁵⁷
189⁵⁸
190⁵⁹
191⁶⁰
192⁶¹
193⁶²
194⁶³
195⁶⁴
196⁶⁵
All birds were counted in a 90° arc to one side of the vessel, out to a maximum distance of 300 m. Birds on the water
were counted and assigned to one of four distance classes (0-50 m, 51-100 m, 101-200 m, 201-300 m) on the basis of
perpendicular distance from the vessel path, confirmed (when necessary) using a distance gauge. Flying birds were
sampled using a “snapshot” methodology such that all flocks present within the 90° arc 300 m to one side and 300 m
ahead of the ship were recorded. The application of the “snapshot” methodology was necessary to offset the tendency to
over estimate the abundance of flying birds, which enter and re-enter the field of view (Tasker et al., 1984). Due to
ongoing development of the survey protocol, there was some variation in the way in which distance was estimated to
flying birds. From October to December, 2007, distances were not measured for flying birds. Between January and July,
2008, flying birds were assigned to distance classes based on perpendicular distance to the vessel path. After July, 2008,

1
2
3
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205

distances to flying birds were measured radially from the vessel. As a result of this variation in sampling, there were really three different protocols applied to flying birds, necessitating three separate density estimates be calculated in program Distance 6 Release 2 (Thomas et al., 2010) and combined at the end of the analysis (Fifield et al., 2009).

Observations were only conducted when the ship was travelling at a minimum speed of 4 knots (7.4 km/h) and a maximum of 19 knots (35.2 km/h). When visibility was poor (due to rain or fog) the actual width of the visible transect (e.g., 200 m) was recorded.

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. We started with basic key functions (half-normal), and optional series expansion terms 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 ($\hat{D}_s \pm$ 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: $\hat{D} = \hat{D}_s * \bar{s}$. Final densities were computed as the weighted average of the three different sampling regimes, based on kilometers surveyed during each survey method.

2.4. Environmental data

A wide range of oceanographic and biological variables have been implicated to explain seabird distribution (Table 1). Broadly speaking, previous studies have either focused on modelling distributional patterns or relating occurrence to oceanographic and bathymetric features or processes. For instance, Certain et al. (2007) used GAM models to produce trend surfaces of seabird occurrence purely as a function of latitude and longitude (although predictions were further refined using residual kriging) but no oceanographic information informed the resulting prediction surface. In contrast, Yen et al. (2004a) integrated factors such as sea-surface temperature (averaged over a 5-month period) and an index of herring spawning to produce predictive maps.

1
2
3
2024
5
2036
7
2048
9
2059
10
2061
12
2073
14
2085
15
2096
17
2108
19
2120
21
2132
22
2144
25
2156
27
2168
29
2180
31
2192
32
2204
35
2216
37
2228
39
2240
41
2252
42
2264
45
2276
47
2288
49
2300
51
2312
52
2324
54
2336
57
2348
59
2360
61
62
63
64
65

Given the prominence of the Scotian Shelf in the study region, as well as previous work documenting the influence of shelf-break and sea bank processes on seabird distribution in this area (e.g., Brown, 1988a; Brown 1988b; Huettmann and Diamond, 2001), regional predictive models were constructed with a focus on bathymetric properties (Table 2). Data from the ETOPO2v2c data set of 2-minute resolution distributed by the National Geophysical Data Centre (NGDC 2009) were integrated within a GIS and used to produce the variable DEPTH. The shelf break was identified using the 300 m isobath, and distances to this linear feature measured using the "Euclidean Distance" tool in ArcGIS 9.3 (ESRI 2008) as variable SHELFDIST. In keeping with earlier findings of the potential importance of sea floor "roughness" (e.g., Yen et al., 2004b) the coefficient of variation of sea depth (DEPTH_CV) was also calculated. 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 Figure 1. The proportion of each 1° x 1° 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. Finally, preliminary analysis, as well as earlier atlas work (Lock et al., 1994), indicated that broad-scale distributional patterns were highly seasonal. To capture seasonal variation in abundances, variable SEASON 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 pscI library of Zeileis et al. (2007). 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. Predictions at any given location were

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

generated as the expected number of seabirds (via the NB GLM) times the fraction of the probability of a non-zero count binomial to a non-zero count negative binomial (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). In the case of the hurdle algorithm, model averaged coefficients were computed for both the binomial and NB GLM components.

Random forests (RF, Breiman, 2001) constitutes a machine-learning alternative to NB and Hurdle methods, and supports its own internal ensemble learning algorithm to construct multiple tree-structured classifiers (Liaw and Wiener, 2002). Using the RF approach, novel predictions are obtained by "dropping" new inputs into the RF classifiers and then combining the "votes" to determine the most popular class or value (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).

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). 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

Ten sets of randomly chosen training and testing data (90% and 10% of original data, respectively) were used to iteratively construct and test the predictions from NB, Hurdle, and RF models. Prediction error was assessed using root mean square error (RMSE):

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{obs} - y_{pred})^2}{n}} \tag{1}$$

2.7. Assessment of potential seabird vulnerability

The predictions from each of the species distribution models were combined into a single, seasonally-specific vulnerability map. Potential vulnerability (V) was quantified for each location in the prediction surface as the sum of the relative importances (R) for each of the k species:

$$V = \sum_{i=1}^k R_i \tag{2}$$

Relative importance was derived from quintile classification of the SDM abundance predictions, but with the break points determined using data pooled across all seasons. Adopting this approach ensured that seasonal shifts in seabird community usage patterns could be readily visualized and detected. It should be noted that while specific hazards were not introduced into equation 2 (e.g., offshore wind farms), they could easily be incorporated by applying species-specific weighting to reflect differing species sensitivity. Furthermore, when regional maps of the distribution and intensity of particular hazards are available, the vulnerability score could be multiplied by the hazard risk to produce a composite risk map.

3. Results

3.1. Species observations

Distance-sampling methods allowed for the estimation of seabird densities, by species group and season, at the level of $1^0 \times 1^0$ survey blocks. Preliminary inspection of the resulting density distributions suggest significant offshore aggregations of Black-legged Kittiwake in the winter and spring seasons (Figure 2a), concentrations of Dovekie on the shelf and shelf break during the winter-spring period (Figure 2b), significant near-shore abundances in murrelets (Figure 2c) and Northern Gannet (Figure 2d) during the spring season, and noteworthy aggregations of storm-petrels (Figure 2e) and shearwaters (Figure 2f) throughout the region during the summer (see also Fifield et al., 2009). Storm-petrels are entirely absent from the study area during winter.

1
2
3
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

3.2. Model construction and evaluation

Species distribution models were constructed, for each species, using NB, Hurdle, and RF algorithms. In the case of NB and Hurdle models, an all-combinations approach was applied and standardized coefficients obtained using AIC-weighted model averaging (see Section 2.5). This process was repeated for 10 sets of training-testing data in order to generate 95% confidence intervals around the model-averaged coefficients obtained using the complete data set.

Physical variables were then sorted from largest amplitude, negative-value to largest amplitude, positive-value coefficients, for both the NB and Hurdle algorithms (Figures 3 and 4, respectively). For RF models, variable importance was assessed using the importance values generated by the *randomForest* software (Figure 5).

Season was a key covariate for all species, regardless of modelling algorithm, though it varied in importance. Relative to fall occurrences, Black-legged Kittiwake, Dovekie, and murre showed strong winter responses under NB models (Figure 3). The same was true for Hurdle models (Figure 4), though uncertainty in the coefficients for the count-component tended to be higher. Storm-petrels and shearwaters showed strong summer responses under NB and Hurdle models, while densities of the Northern Gannet peaked in the fall. Depth (DEPTH) and distance to shelf break (SHELFDIST) were influential for some species, e.g., Black-legged Kittiwake, Dovekie, and Northern Gannet. DEPTH and season tended to be the most influential covariates in RF models (Figure 5), with SHELFDIST most important for Northern Gannet. Variation in seadepth (DEPTH_CV) also played an important role in storm-petrel and shearwater RF models.

No one algorithm (NB, Hurdle, RF) exhibited consistently superior performance over all species and iterations, but occasionally, NB and Hurdle algorithms failed to compute (Figure 6). RF models appeared better able to cope with sparse data and, in some cases, to also generate lower prediction error (e.g., Northern Gannet, Figure 6). More often, however, variation in performance meant that methods alternated in predictive power, working better in some iterations than others. Simple averaging of the NB, Hurdle and RF results produced ensemble predictions that either tracked the consensus or constituted a compromise value.

3.3. Assessment of species vulnerability

Vulnerability maps (Figure 7) represented the influence of season and bathymetric factors on the relative distribution of all species combined. Potential vulnerability for this group of species were highest in the spring, when large numbers of

1
2
3
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

seabirds were widely distributed throughout the Nova Scotian shelf and shelf break. This contrasts with the fall season, when vulnerability was distributed along a narrower front concentrated primarily over the shelf break. There was a near-shore shift in vulnerability in the summer, presumably reflecting the central place foraging limitations imposed on species such as the Northern Gannet. But it is important to note that this was counteracted by the presence of large numbers of non-breeding species such as shearwaters, which would not be similarly range restricted. Potential vulnerability was lowest during the winter season (Figures 7 and 8) when patchier distributions combined with the absence of storm-petrels, for example, lowered the vulnerability scores.

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 (Araujo 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 an ongoing monitoring program 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., 2000) data gathered through opportunistic surveys was aggregated at 1° x 1° survey blocks 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' (DEPTH_CV) was associated with shearwaters and Northern Gannet. Adopting an aggregated, grid-based framework meant that seabird counts had to be pooled over the entire study period, but as fine-scale marine processes were not available contemporaneous for the entire study area, this was a sensible approach that permitted a focus on stable marine features (e.g., bathymetry). This approach did allow for seasonal summaries of abundance, which was important given that season was influential regardless of modelling method.

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 (Opper et al., 2012; Araujo and New, 2007). The focus of most SDM papers published in the first decade of the 21st century focused on comparing relative performance (e.g., Elith et al., 2006). As a consequence, the SDM literature is replete with an enormous array of single-method modelling

1
2
3
3694
5
3706
7
3718
9
3722
10
3731
12
3743
14
3754
15
3766
17
3778
19
3789
20
3791
22
3803
24
3814
25
3826
27
3838
29
3849
30
3851
32
3863
34
3874
35
3886
37
3898
39
3910
40
3911
42
3923
44
3934
45
3946
47
3958
49
3970
50
3971
52
3983
54
3994
55
4006
57
4018
59
4029
60
61
62
63
64
65

approaches where no attempt is made to gather the "consensus opinion" presented by a diversity of algorithms. While individual methods might be well calibrated and offer high predictive accuracy for the data at hand, there is no way to assess how realistic the predictions will be 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 single algorithm (NB, Hurdle, RF) exhibited consistently superior performance over all species and iterations, though occasionally, NB and Hurdle algorithms failed altogether. RF models appeared better able to cope with sparse data and, in some cases, to also generate lower prediction error (e.g., Northern Gannet, Figure 6). In general, variation in performance mean that methods alternated in predictive power, working better in some iterations than others. Simple averaging of the NB, Hurdle and RF results produced ensemble predictions that either tracked the consensus or constituted a compromise value.

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. Quintile classes were assigned, based on the entire range of predicted abundances across all seasons, in order to rank the relative usage of each of these locations. Potential vulnerability to anthropogenic disturbance was then assessed as a function of the sum of the relative usage across species, enabling the visualization of changing patterns of usage across the entire community. The composite picture that emerges is one of shifting presence, with some species disappearing from the region (e.g., petrels and shearwaters in winter) and others increasing at the same time (e.g., Black-legged Kittiwake). Spring (March - April) stood out as a period of peak usage by the entire community, which corresponded with a wide geographic distribution throughout the shelf and shelf break. As this is an area exposed to extensive shipping traffic and ongoing petroleum extraction and exploration, there remains an ongoing potential impact by acute (e.g., accidental) and chronic (e.g., illegal) discharges of oil pollution. A previous study indicated that near-shore areas, particularly in the vicinity of harbours, tend to suffer the highest pollution loading but elevated levels also occurred within and beyond the shelf break (Lieske et al., 2011).

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 and 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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Furthermore, seabird research programs would greatly benefit from the coordinated acquisition of marine hazard data (e.g., fishing activity, oil discharges) at regional scales, which would support geographic analysis of the potential impact of specific risks.

Acknowledgements

We acknowledge the financial support of the Environmental Studies Research Fund (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 while a Science Horizons intern, as well as the EWHALE lab of the University of Alaska, Fairbanks for advice on the use of random forest machine learning. Finally, we thank all the observers who participated in the ECSAS program, and braved often challenging conditions to gather the data reported here.

Figures

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

Figure 2a. Density of Black-legged Kittiwake (per km²) for 1⁰ x 1⁰ survey blocks, for spring (a, Mar - Apr), summer (b, May - Aug), fall (c, Sep - Oct), and winter (d, Nov - Feb).

Figure 2b. Density of Dovekie (per km²) for 1⁰ x 1⁰ survey blocks, for spring (a, Mar - Apr), summer (b, May - Aug), fall (c, Sep - Oct), and winter (d, Nov - Feb).

Figure 2c. Density of murrelets (per km²) for 1⁰ x 1⁰ survey blocks, for spring (a, Mar - Apr), summer (b, May - Aug), fall (c, Sep - Oct), and winter (d, Nov - Feb).

Figure 2d. Density of Northern Gannet (per km²) for 1⁰ x 1⁰ survey blocks, for spring (a, Mar - Apr), summer (b, May - Aug), fall (c, Sep - Oct), and winter (d, Nov - Feb).

Figure 2e. Density of storm-petrels (per km²) for 1⁰ x 1⁰ survey blocks, for spring (a, Mar - Apr), summer (b, May - Aug), fall (c, Sep - Oct), and winter (d, Nov - Feb).

Figure 2f. Density of shearwaters (per km²) for 1⁰ x 1⁰ survey blocks, for spring (a, Mar - Apr), summer (b, May - Aug), fall (c, Sep - Oct), and winter (d, Nov - Feb).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 3. Variable importance for species distribution models developed using the negative binomial (NB) generalized linear model as indicated by regression coefficients for standardized covariates. Also shown are the estimated confidence intervals (95%) derived from testing and training data.

Figure 4. Variable importance for species distribution models developed using the hurdle model, 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.

Figure 5. 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.

Figure 6. Variation in prediction error, as measured by root mean squared error (RMSE), across 10 iterations of randomly selected training and testing data modelled using: negative binomial (NB) glm, Hurdle, Random Forest (RF), and ensemble.

Figure 7. Seasonally-specific potential vulnerability maps, based on the sum of the relative densities derived from ensemble model predictions for each of six species (Black-legged Kittiwake, Dovekie, murre, Northern Gannet, storm-petrels, and shearwaters). Seasons were defined as: spring (Mar - Apr), summer (May - Aug), fall (Sep - Oct), and winter (Nov - Feb).

Figure 8. Density distribution of the cell-specific vulnerability scores derived from the seasonally-specific vulnerability maps (Figure 7).

References

Araújo, M.B., New, M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22, 42-47

Ashmole, N.P. 1971. Sea bird ecology and the marine environment. *Avian Biology* 1, 223-286.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecological Modelling* 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.

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. *Canadian Journal of Fisheries and Aquatic Sciences* 49, 238-251.

Breiman, L. 2001. Random forests. *Machine Learning* 45, 5-32.

1
2
3
483⁴ Briggs, K., Tyler, W.M.B., Lewis, D.B., Carlson, D.R. 1987. Bird communities at sea off California: 1975 to 1983.
484⁵ Studies in Avian Biology 11.
485⁷ Brown, R.G.B. 1980. Seabirds as marine animals, in: Burger J., Olla, B.L., Winn, H.E. (Eds.), Behavior of Marine
486⁹ Animals, vol. 4. New York, pp. 1-39.
487¹⁰ Brown, R.G.B. 1988a. The influence of oceanographic anomalies on the distributions of Storm-Petrels (Hydrobatidae)
488¹² in Nova Scotian waters. Colonial Waterbirds 11, 1-8.
489¹⁴ Brown, R.G.B. 1988b. Oceanographic factors as determinants of the winter range of the Dovekie (*Alle alle*) off Atlantic
490¹⁵ Canada. Colonial Waterbirds 11, 176-180.
491¹⁷ Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., & Thomas, L. 2001. Introduction to
492¹⁹ distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford, UK.
493²⁰ Burnham, K., Anderson, D.R. 2002. Model selection and multimodal inference: a practical information-theoretic
494²² approach. Springer Verlag, New York.
495²⁴ Cairns, D.K., Schneider, D.C. 1990. Hot spots of cold water: feeding habitat selection by Thick-billed Murres. Studies
496²⁵ in Avian Biology 14, 52-60.
497²⁷ Camphuysen, C.J. 1998. Beached bird surveys indicate decline in chronic oil pollution in the North Sea. Marine
498²⁹ Pollution Bulletin 36, 519-526.
499³⁰ Cayula, J.-F., Cornillon, P. 1992. Edge detection algorithm for SST images. Journal of Atmospheric and Oceanic
500³² Technology 9, 67-80.
501³⁴ Certain, G., Bellier, E., Planque, B., Bretagnolle, V. 2007. Characterising the temporal variability of the spatial
502³⁵ distribution of animals: an application to seabirds at sea. Ecography 30, 695-708.
503³⁷ Chivers, L.S., Lundy, M.G., Colhoun, K., Newton, S.F., Houghton, J.D.R., Reid, N. 2013. Identifying optimal feeding
504³⁹ habitat and proposed Marine Protected Areas (pMPAs) for the black-legged kittiwake (*Rissa tridactyla*) suggests
505⁴⁰ a need for complementary management approaches. Biological Conservation 164, 73-81.
506⁴² Clarke, E.D., Spear, L.B., McCracken, M.L., Marques, F.F.C., Borchers, D.L., Buckland, S.T., Ainley, A.D. 2003.
507⁴⁴ Validating the use of generalized additive models and at-sea surveys to estimate size and temporal trends of
508⁴⁵ seabird populations. Journal of Applied Ecology 40, 278-292.
509⁴⁷ Clemen, R.T. 1989. Combining forecasts: a review and annotated bibliography. International Journal of Forecasting 5,
510⁴⁹ 559-583.
511⁵⁰ Das, S.K., Chen, S., Deasy, J.O., Zhou, S., Fang-Fang, Y., Marks, L.B., Marks, L.B. 2008. Combining multiple models
512⁵² to generate consensus: application to radiation-induced pneumonitis prediction. Medical Physics 35, 5098-5109.
513⁵⁴ Durazo, R., Harrison, N.M., Hill, A.E. 1998. Seabird observations at a tidal mixing front in the Irish Sea.
514⁵⁵ Estuarine, Coastal and Shelf Science 47, 153-164.
515⁵⁷ Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick,
516⁵⁹ J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y.,
517⁶⁰

1
2
3
4
517⁴ McC. Overton, J., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J.,
5
518⁵ Williams, S., Wisz, M., and Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions
6
519⁶ from occurrence data. *Ecography* 29, 129-151.
7
8
520⁸ ESRI. 2008. ArcGIS version 9.3. Environmental Systems Research Institute, Redlands.
9
10
521¹⁰ Etnoyer, P., Canny, D., Mate, B., Morgan, L. 2004. Persistent pelagic habitats in the Baja California to Bering Sea
11
522¹¹ (B2B) ecoregion. *Oceanography* 17, 90-101.
12
13
523¹³ Exo, K.-M., Hüppop, O., Garthe, S. 2003. Birds and offshore wind farms: a hot topic in marine ecology. *Wader Study*
14
524¹⁴ Group Bulletin 100, 50–53.
15
16
525¹⁶ Fauchald, P., Erikstad, K.E., Systad, G.H. 2002. Seabirds and marine oil incidents: is it possible to predict the spatial
17
526¹⁷ distribution of pelagic seabirds? *Journal of Applied Ecology* 39, 349-360.
18
19
527¹⁹ Fifield, D., Lewis, K.P., Gjerdrum, C., Robertson, G.J., Wells, R. 2009. Offshore seabird monitoring program: final
20
528²⁰ report. Canadian Wildlife Service, Atlantic Region, St. John's, Newfoundland.
21
22
529²² Franklin, J. 2009. Mapping species distributions: spatial inference and prediction. Cambridge University Press, New
23
530²³ York.
24
25
531²⁵ Gjerdrum, C., Fifield, D.A., Wilhelm, S.I. 2012. Eastern Canada Seabirds at Sea (ECSAS) standardized protocol for
26
532²⁶ pelagic seabird surveys from moving and stationary platforms. Canadian Wildlife Service Technical Report
27
533²⁷ Series No. 515. Atlantic Region. vi + 36 pp.
28
29
534²⁹ Gray, T.N.E., Phan, C., Long, B. 2010. Modelling species distribution at multiple spatial scales: gibbon habitat
30
535³⁰ preference in a fragmented landscape. *Animal Conservation* 13, 324-332.
31
32
536³² Guisan, A., Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147-
33
537³³ 186.
34
35
538³⁵ Hamel, N.J., Burger, A.E., Charleton, K., Davidson, P., Lee, S., Bertram, D.F., Parrish, J.K. 2009. Bycatch and beached
36
539³⁶ birds: assessing mortality impacts in coastal net fisheries using marine bird strandings. *Marine Ornithology* 37,
37
540³⁷ 41-60.
38
39
541³⁹ Haney, J.C. 1989. Iterative techniques for characterizing marine bird habitats with time-series of satellite images.
40
542⁴⁰ *Colonial Waterbirds* 12, 78-89.
41
42
543⁴² Haney, J.C., McGillvray, P.A. 1985. Aggregations of Cory's Shearwaters (*Calonectris diomedea*) at Gulf Stream fronts.
43
544⁴³ *Wilson Bulletin* 97, 191-200.
44
45
545⁴⁵ Hannah, C.G., Shore, J.A., Loder, J.W. 2001. Seasonal circulation on the western and central Scotian Shelf. *Journal of*
46
546⁴⁶ *Physical Oceanography* 31, 591-615.
47
48
547⁴⁸ Hegel, T.M., Cushman, S.A., Evans, J., Huettmann, F. 2010. Current state of the art for statistical modelling of species
49
548⁴⁹ distributions, in: *Spatial Complexity, Informatics, and Wildlife Conservation*, Cushman, S.A., Huettmann, F.
50
549⁵⁰ (Eds.), Springer, pp. 273-312.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
550⁴ Henkel, J.R., Sigel, B.J., Taylor, C.M. 2012. Large-scale impacts of the Deepwater Horizon oil spill: can local
551⁵ disturbance affect distant ecosystems through migratory shorebirds? *Bioscience* 62, 676-685.
552⁷ Hilbe, J.M. 2008. *Negative binomial regression*. Cambridge University Press, New York.
553⁹ Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-
554¹⁰ suitability maps without absence data? *Ecology* 83, 2027-2036.
555¹² Huettmann, F., Diamond, A.W. 2001. Seabird colony locations and environmental determination of seabird distribution:
556¹⁴ a spatially explicit breeding seabird model for the Northwest Atlantic. *Ecological Modelling* 141, 261-298.
557¹⁵ Hunt, G.L., Schneider, D.C. 1987. Scale-dependent processes in the physical and biological environment of marine
558¹⁷ birds, in: Croxall, J.P. (Ed.), *Seabirds: feeding ecology and role in marine ecosystems*. Cambridge University
559¹⁹ Press, pp. 7-41.
560²⁰ Jones-Farrand, D.T., Fearer, T.M., Thogmartin, W.E., Thompson III, F. R., Nelson, M.D., Tirpak, J.M. 2011.
561²² Comparison of statistical and theoretical habitat models for conservation planning: the benefit of ensemble
562²³ prediction. *Ecological Applications* 21, 2269-2282.
563²⁴ Law, A., Wiener, M. 2002. Classification and regression by randomForest. *R News* 2/3.
564²⁵ Lieske, D.J., Mahoney, M., Wilhelm, S.I., Weir, L., O'Hara, P. 2011. Eyes in the sky: a comparative spatial analysis of
565²⁷ aerial and satellite surveillance of East Coast Canadian oil pollution. *Proceedings of the 34th AMOP Technical*
566²⁹ *Seminar on Environmental Contamination and Response*, 244-258.
567³⁰ Lieske, D.J., Pollard, B., Gloutney, M., Milton, R., Connor, R., Dibblee, R., Parsons, G., Howerter, D. 2012. The
568³² importance of agricultural landscapes as key nesting habitats for the American Black Duck in Maritime Canada.
569³⁴ *Waterbirds* 35, 525-534.
570³⁵ Lock, A.R., Brown, R.G.B., Gerriets, S.H. 1994. *Gazetteer of marine birds in Atlantic Canada: an atlas of seabird*
571³⁷ *vulnerability to oil pollution*. Canadian Wildlife Service.
572³⁹ Magness, D.R., Morton, J.M., Huettmann, F. 2011. How spatial information contributes to the conservation and
573⁴⁰ management of biodiversity, in: *Spatial complexity, informatics, and wildlife conservation*, Cushman, S.A.,
574⁴² Huettmann, F. (Eds.), Springer, pp. 429-444.
575⁴⁴ Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W. 2009. Evaluation of consensus methods in
576⁴⁵ predictive species distribution modelling. *Diversity and Distributions* 15, 59-69.
577⁴⁷ Marques, T.A., Thomas, L., Fancy, S.G., Buckland, S.T. 2007. Improving estimates of bird density using multiple-
578⁴⁹ covariate distance sampling. *Auk* 124, 1229-1243.
579⁵⁰ McKinnon, L., Gilchrist, H.G., Fifield, D. 2009. A pelagic seabird survey of arctic and sub-arctic Canadian waters
580⁵² during fall. *Marine Ornithology* 37, 77-84.
581⁵⁴ NGDC, 2009. ETopo2v2 global gridded 2-minute database. <<http://www.ngdc.noaa.gov/mgg/global/etopo2.html>>
582⁵⁵ (accessed March 2010).
583⁵⁷ NSMN. T6.1. Ocean currents. <<https://museum.gov.ns.ca/mnh/nature/nhns/t6/t6-1.htm>> (accessed March 2010).
584⁵⁹
585⁶⁰

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

O'Hara, P.D., Morgan, K.H., Sydeman, W.J. 2006. Primary producer and seabird associations with AVHRR-derived sea surface temperatures and gradients in the southeastern Gulf of Alaska. *Deep-Sea Research II*, 53, 359-369.

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. *Biological Conservation* 156, 91-104.

Phillips, S.J., Anderson, R.P., Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190, 231-259.

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. *Marine Ecology Progress Series* 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. *Global Ecology and Biogeography* 20, 789-802.

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).

Ronconi, R.A., Lascelles, B.G., Langham, G.M., Reid, J.B., and Oro, D. 2012. The role of seabirds in marine protected areas identification, delineation, and monitoring: introduction and synthesis. *Biological Conservation* 156: 1-4.

Schneider, D.C. 1990. Seabirds and fronts: a brief overview. *Polar Research* 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, D.C., pp.529-535.

Skov, H., Durinck, J. 1998. Constancy of frontal aggregations of seabirds at the shelf break in the Skagerrak. *Journal of Sea Research* 39, 305-311.

Skov, H., Humphreys, E., Garthe, S., Geitner, K., Grémillet, D., Hamer, K.C., Hennicke, 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. *Ecological Modelling* 212, 504-512.

Smith P.C., Petrie, B.D. 1982. Low-frequency circulation at the edge of the Scotian Shelf. *Journal of Physical Oceanography* 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., Camphuysen, C.J., Cooper, J., Garthe, S., Montevecchi, W.A., Blaber, S.J.M. 2000. The impacts of fishing on marine birds. *ICES Journal of Marine Science* 57, 531-547.

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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

618⁴ Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Bishop, J.R.B., Marquez, T.A.,
619⁵ Burnham, K.P. 2010. Distance software: design and analysis of distance sampling surveys for estimating
620⁷ population size. *Journal of Applied Ecology* 47, 5-14.

621⁹ Tremblay, Y., Bertrand, S., Henry, R.W., Kappes, M.A., Costa, D.P., Schaffer, S.A. 2009. Analytical approaches to
622¹⁰ investigating seabird-environmental interactions: a review. *Marine Ecology Progress Series* 391, 153-163.

623¹² Venables, W.N., Ripley, B.D. 2002. *Modern Applied Statistics with S*. Springer, New York.

624¹⁴ Votier, S.C., Hatchewell, B.J., Beckerman, A., McCleery, R.H., Hunter, F.M., Pellatt, J., Trinder, M., Birkhead, T.R.
625¹⁵ 2005. Oil pollution and climate have wide-scale impacts on seabird demographics. *Ecology Letters* 8, 1157-1164.

626¹⁷ Wiese, F.K., Montevecchi, W.A., Davoren, G.K., Huettmann, F., Diamond, A.W., Linke, J. 2001. Seabirds at risk
627¹⁸ around offshore oil pollution platforms in the North-west Atlantic. *Marine Pollution Bulletin* 42, 1285-1290.

628¹⁹ Wiese, F.K., Ryan, P.C. 2003. The extent of chronic marine oil pollution in southeastern Newfoundland waters assessed
629²² through beached bird surveys 1984-1999. *Marine Pollution Bulletin* 46, 1090-1101.

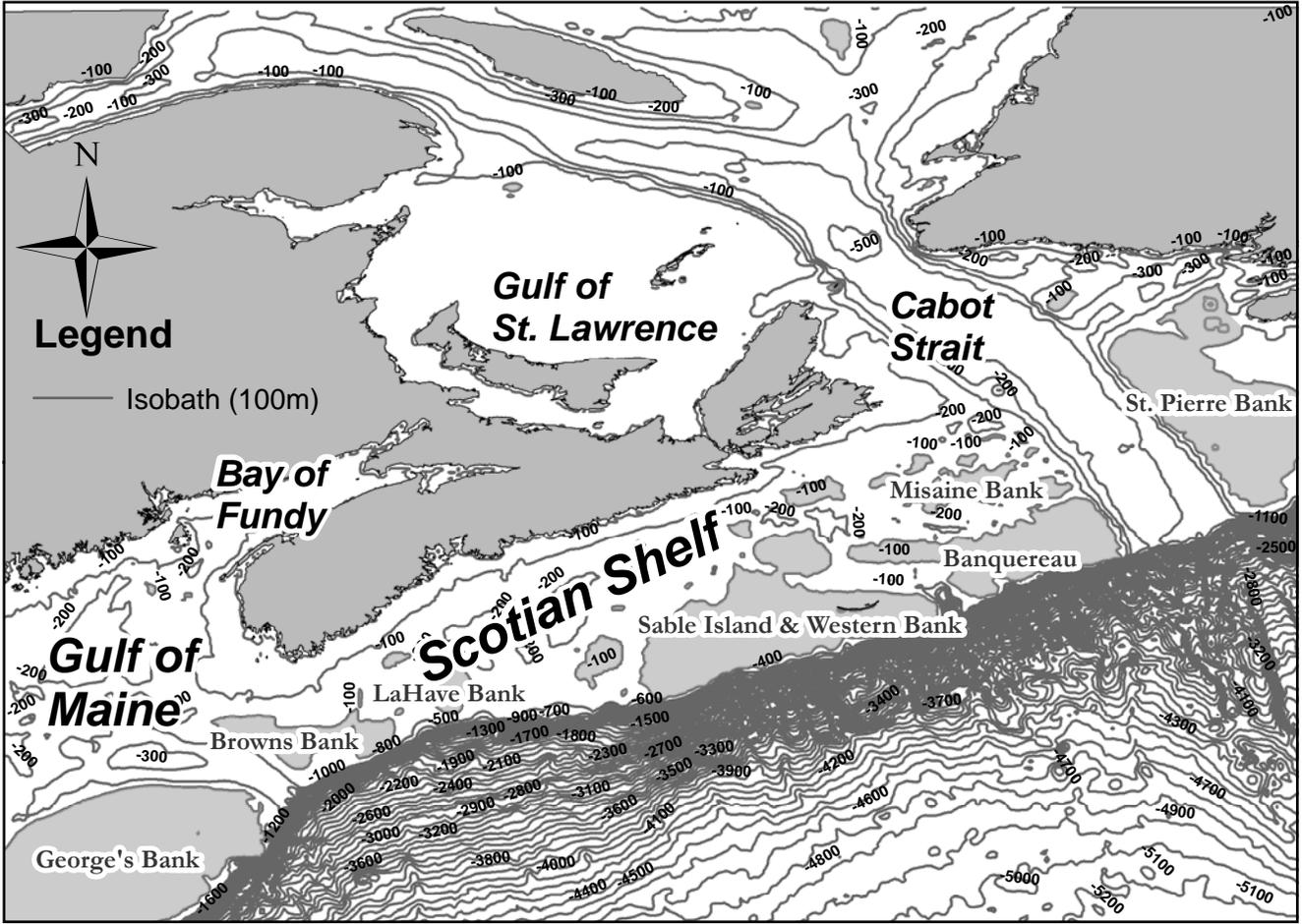
630²⁴ Yen, P.P.W., Huettmann, F., Cooke, F. 2004a. A large-scale model for the at-sea distribution and abundance of Marbled
631²⁵ Murrelets (*Brachyramphus marmoratus*) during the breeding season in coastal British Columbia, Canada.
632²⁷ *Ecological Modelling* 171, 395-413.

633²⁹ Yen, P.P.W., Sydeman, W.J., Hyrenbach, K.D. 2004b. Marine bird and cetacean associations with bathymetric habitats
634³⁰ and shallow-water topographies: implications for trophic transfer and conservation. *Journal of Marine Systems*
635³² 50, 79-99.

636³⁴ Zeileis, A., Kleiber, C., Jackman, S. 2008. Regression models for count data in R. *Journal of Software* 27, Issue 8.

637³⁶ Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M. 2009. *Mixed effects models and extensions in*
638³⁷ *ecology with R*. Springer, New York.

Figure1



0 50 100 200 300 400 Kilometers

Figure2a

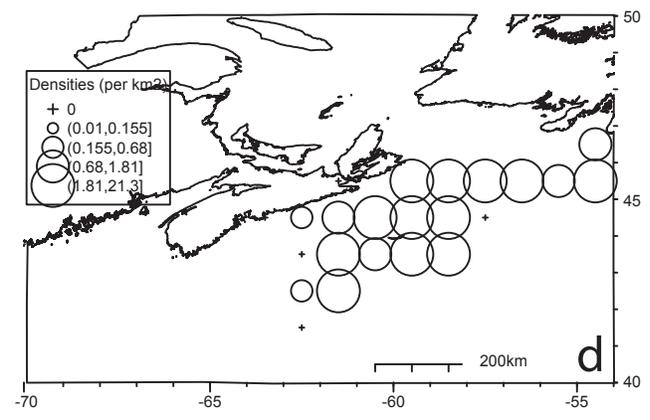
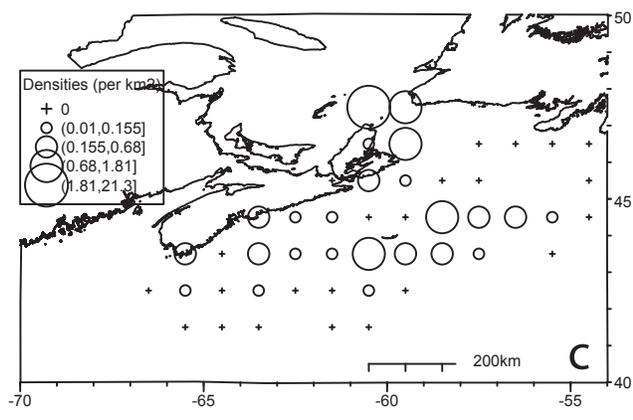
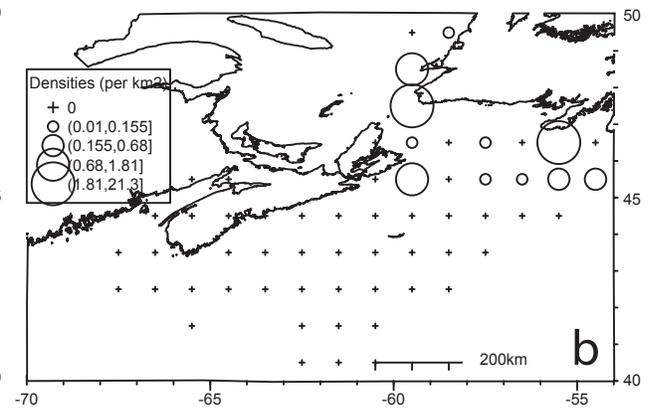
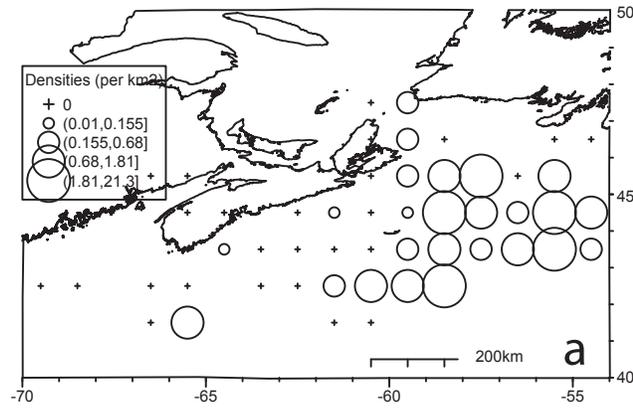


Figure2b

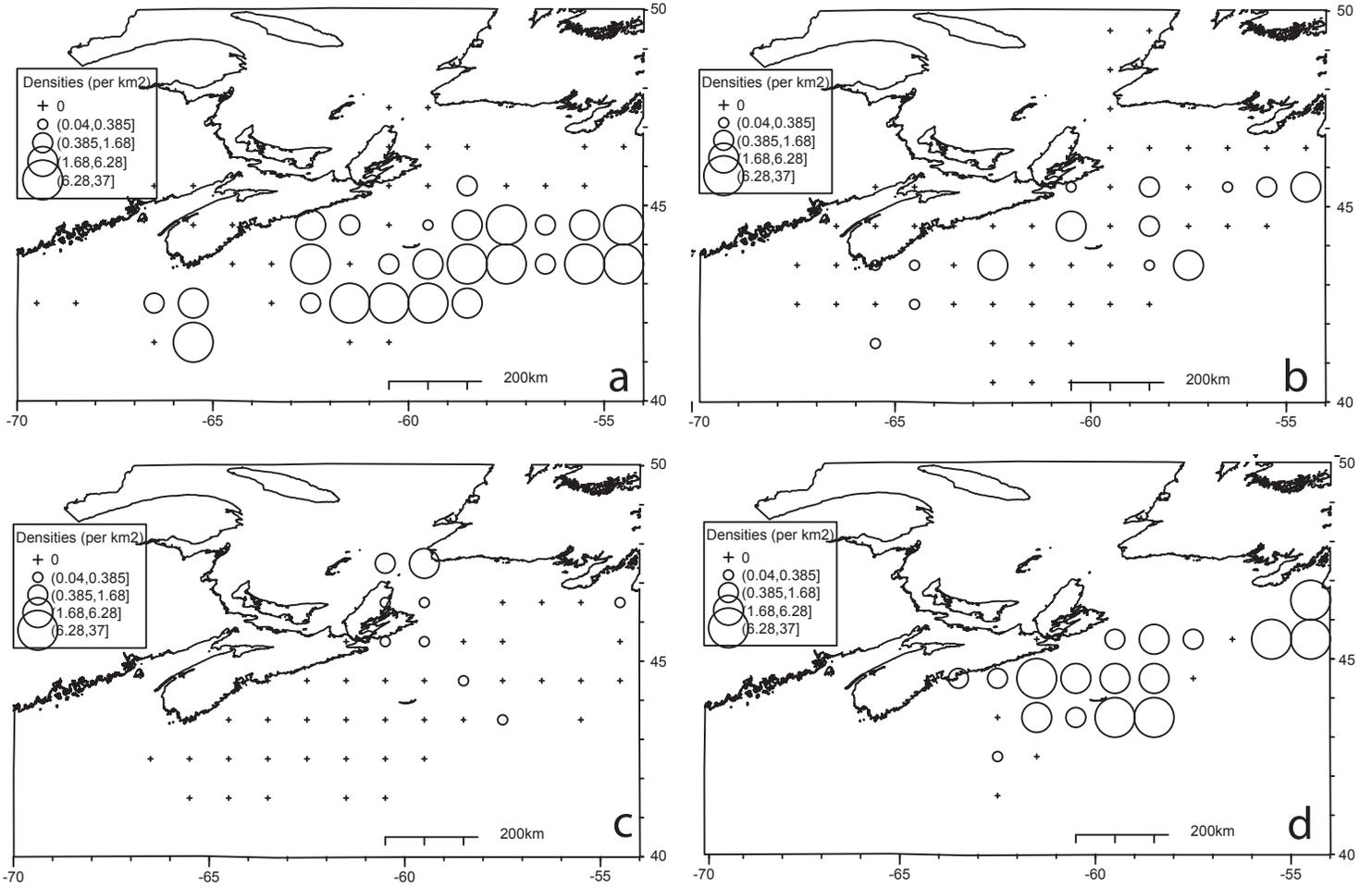


Figure2c

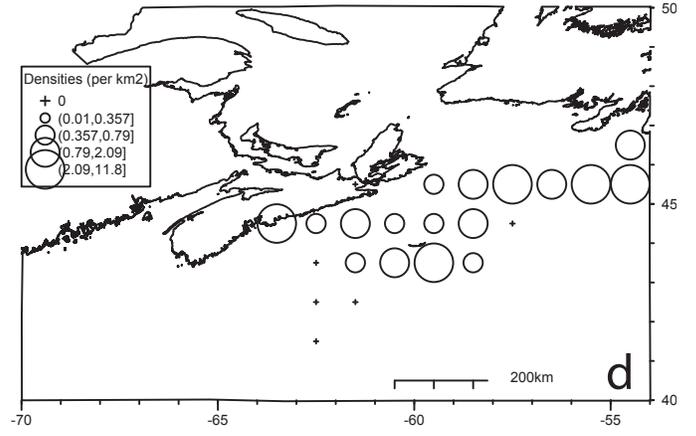
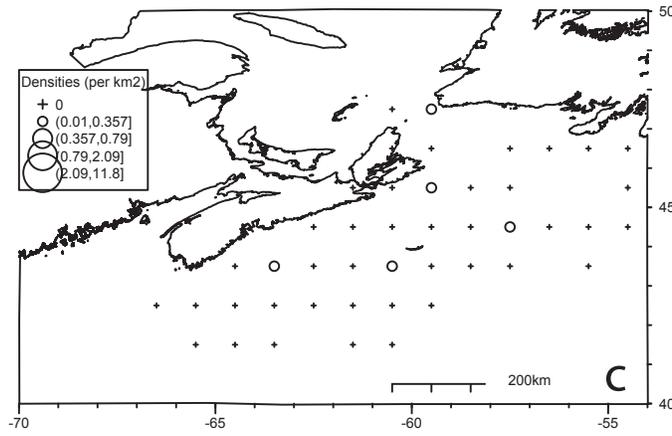
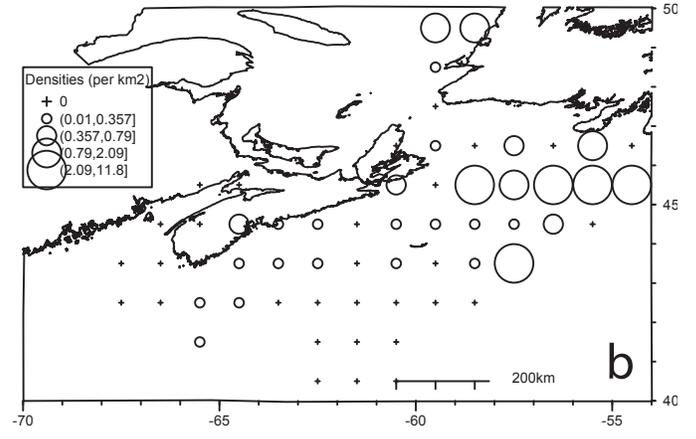
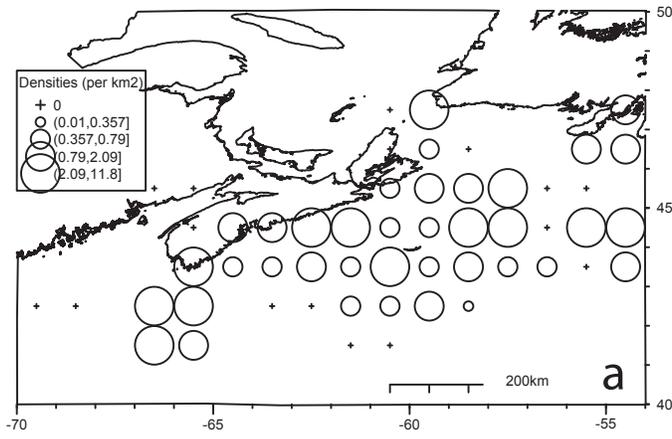


Figure2d

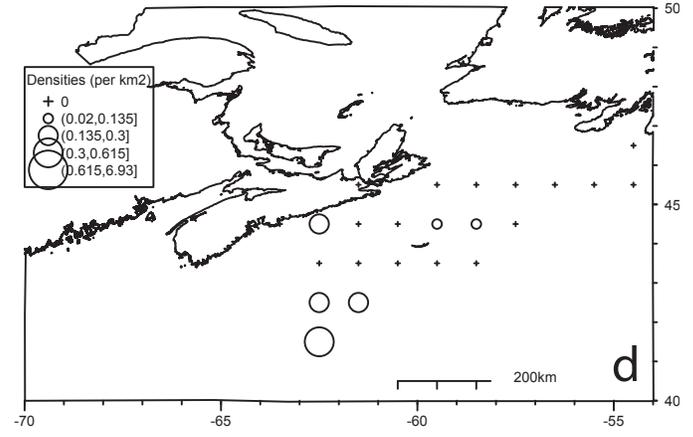
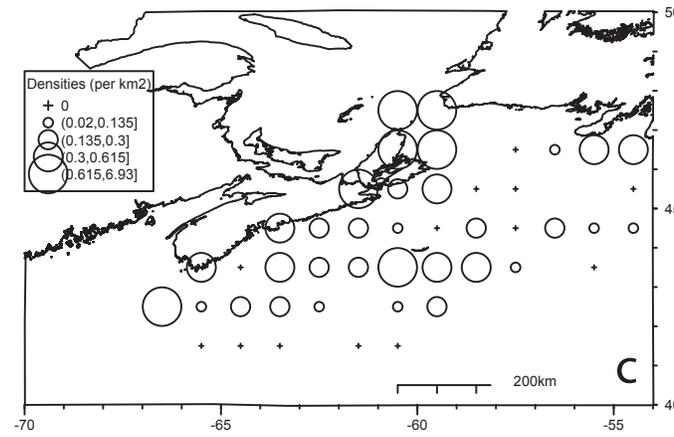
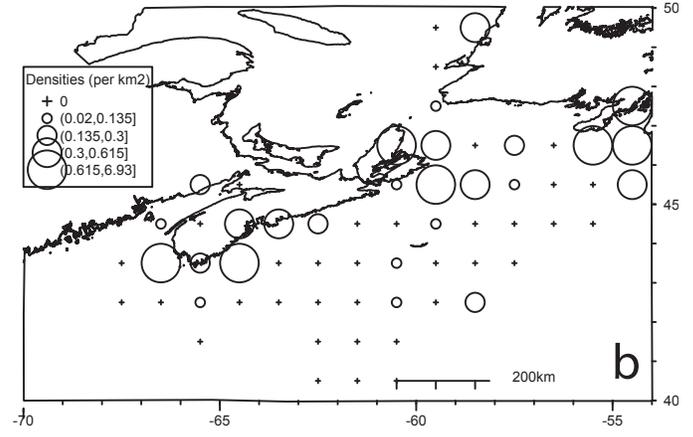
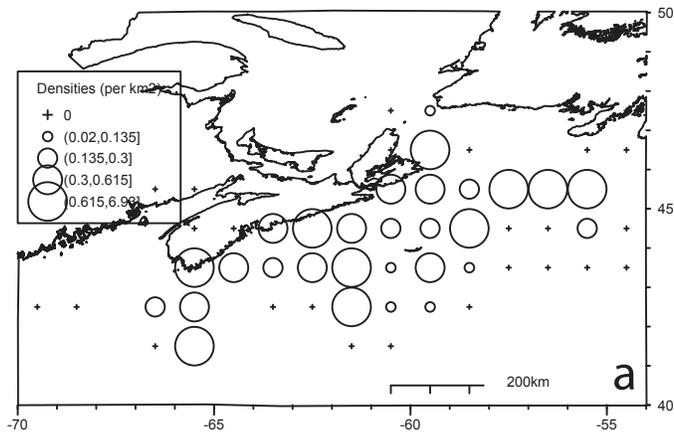


Figure2e

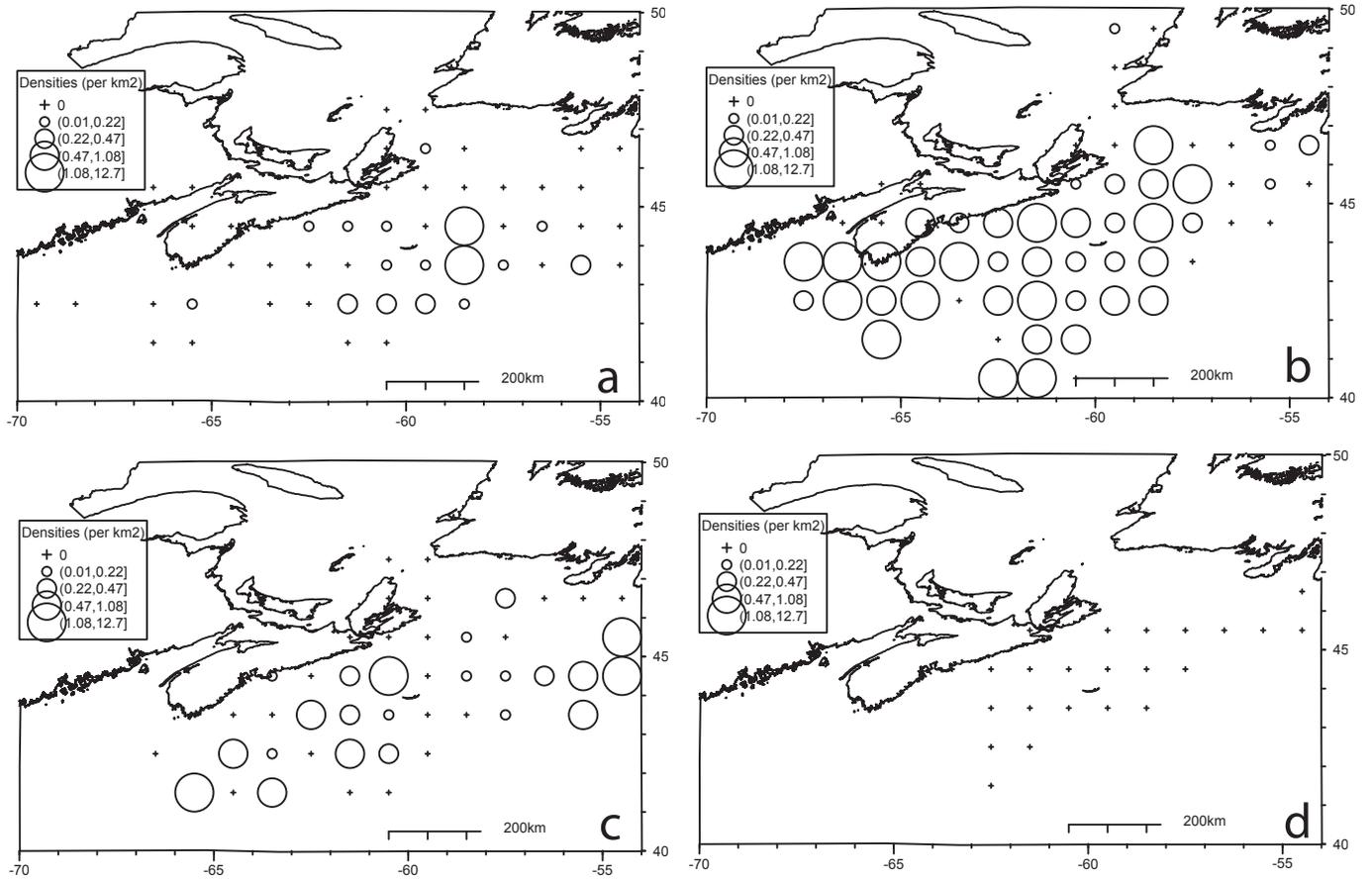


Figure3

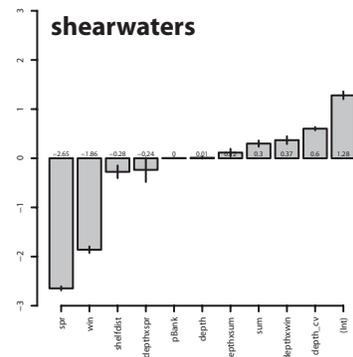
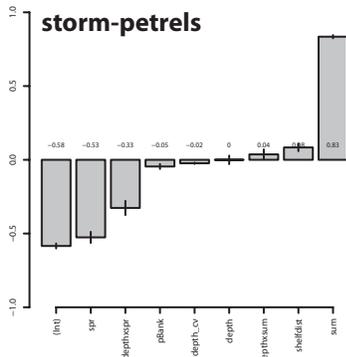
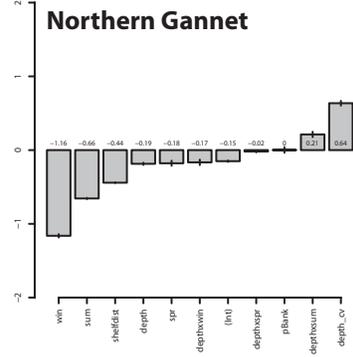
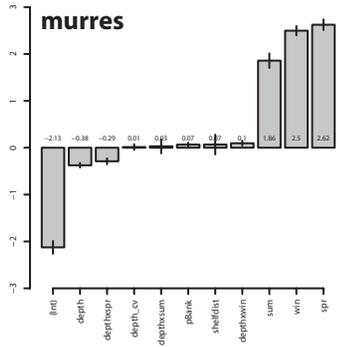
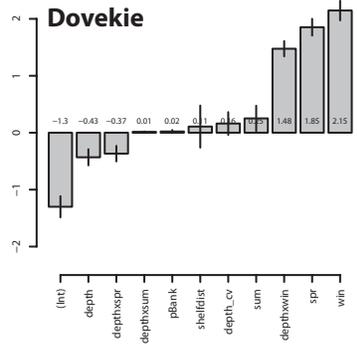
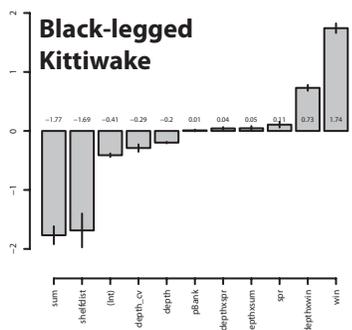


Figure4

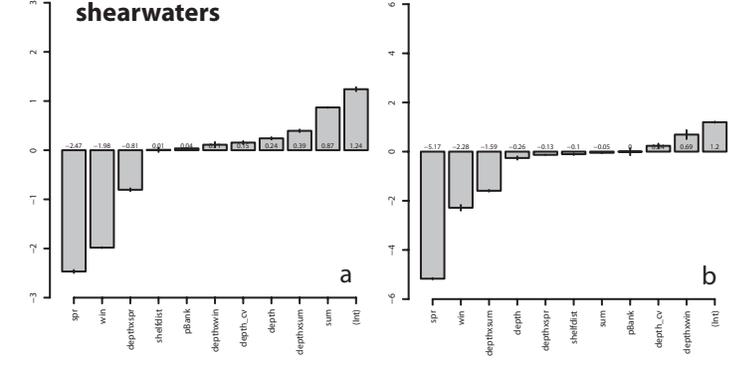
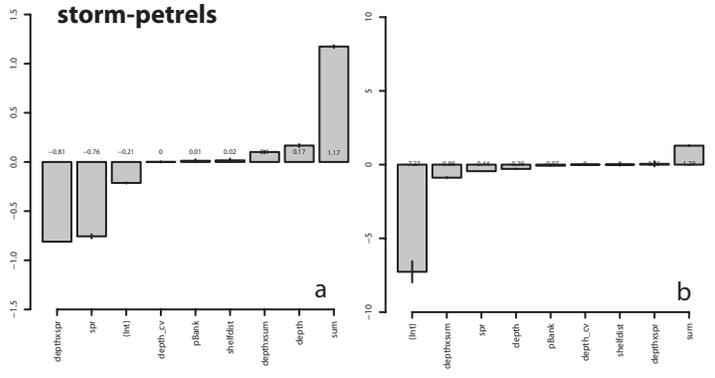
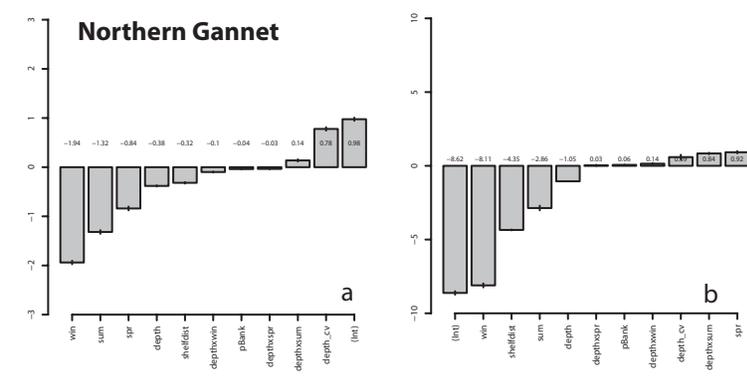
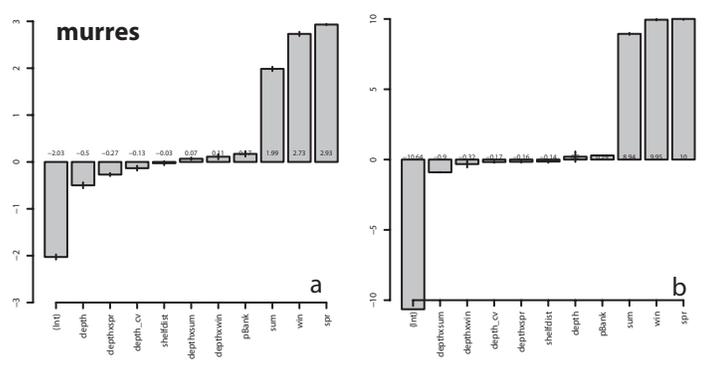
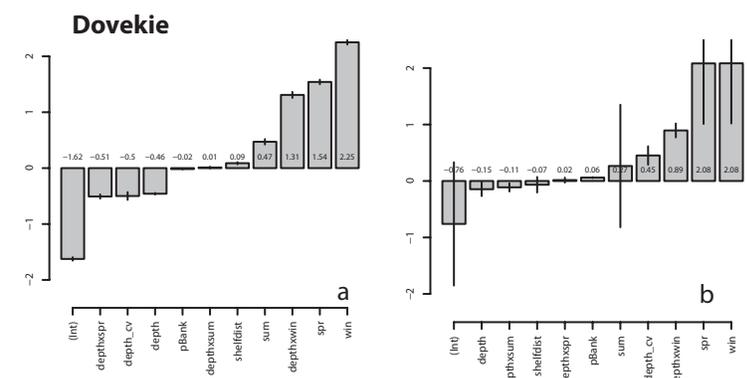
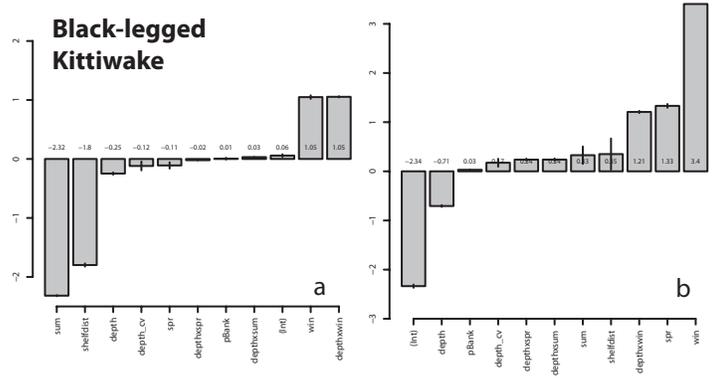


Figure5

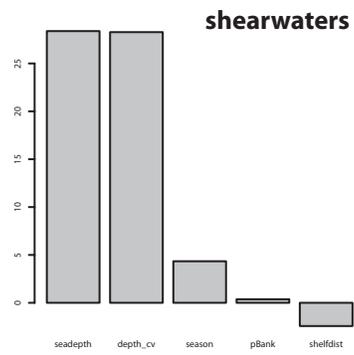
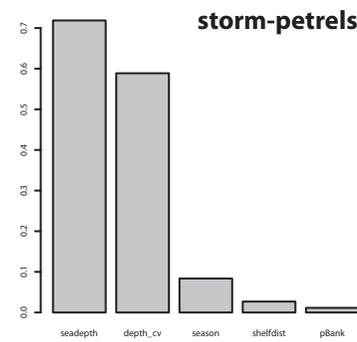
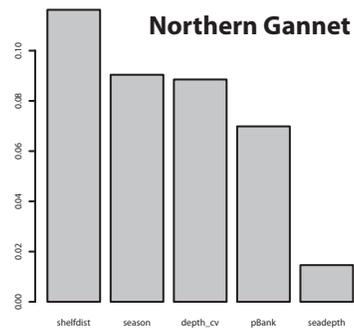
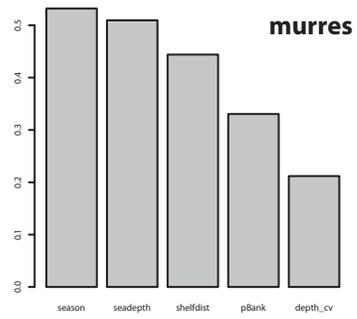
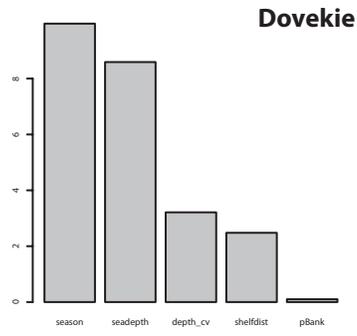
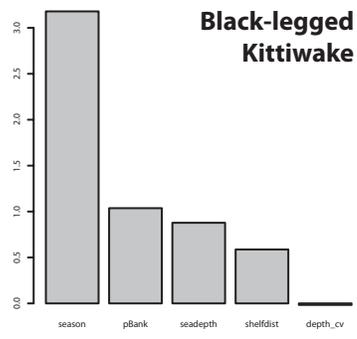
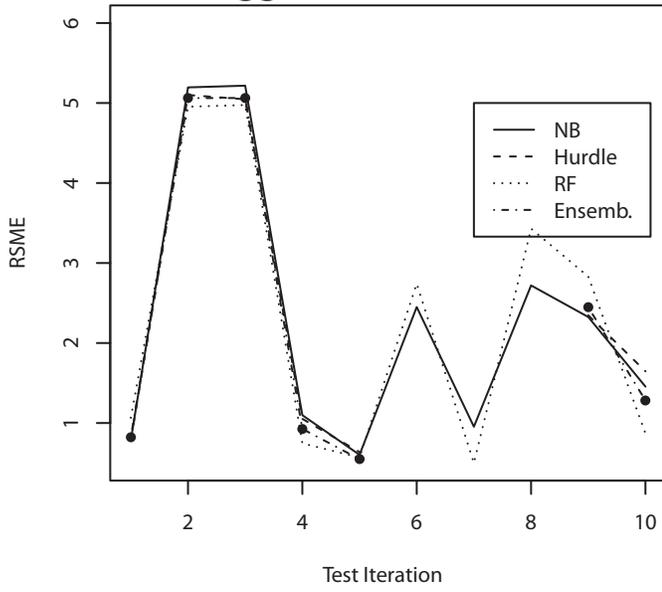
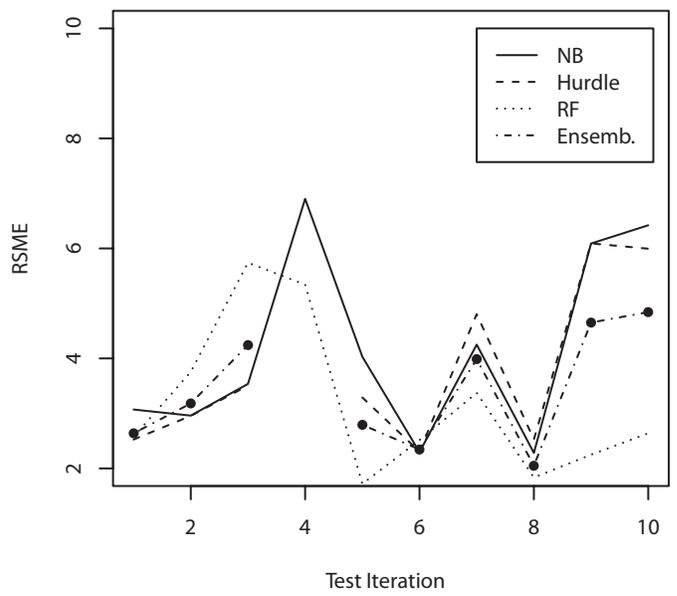


Figure6

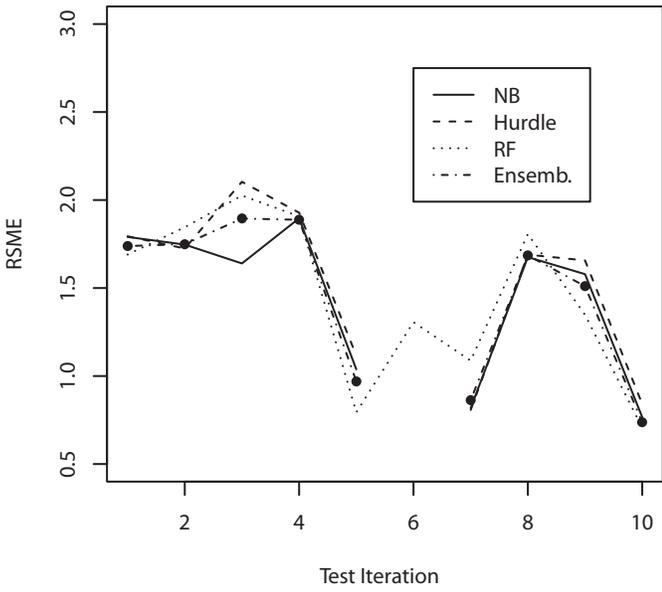
Black-legged Kittiwake



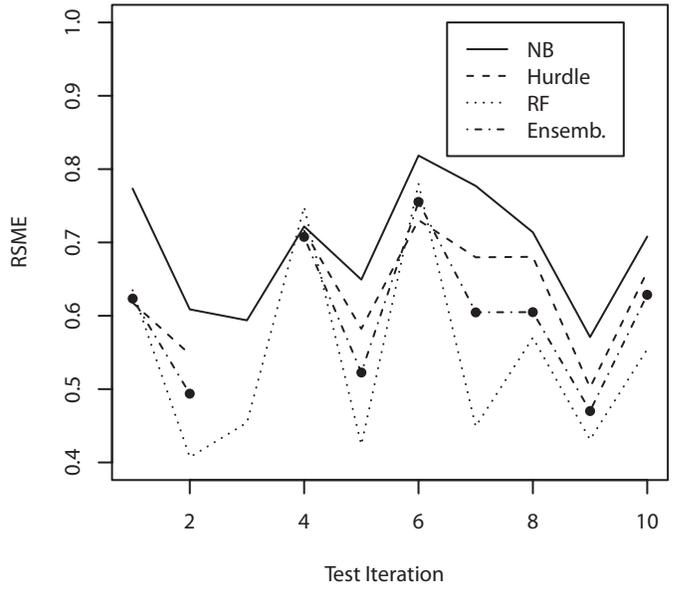
Dovekie



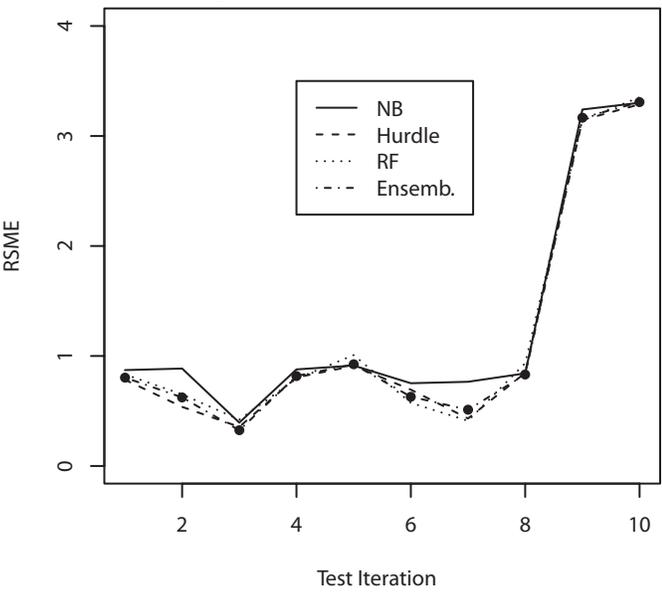
murre



Northern Gannet



storm-petrels



shearwaters

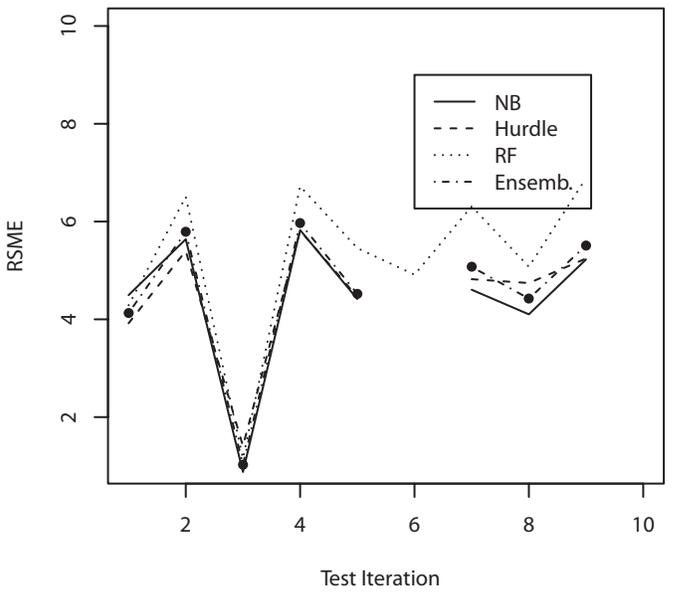
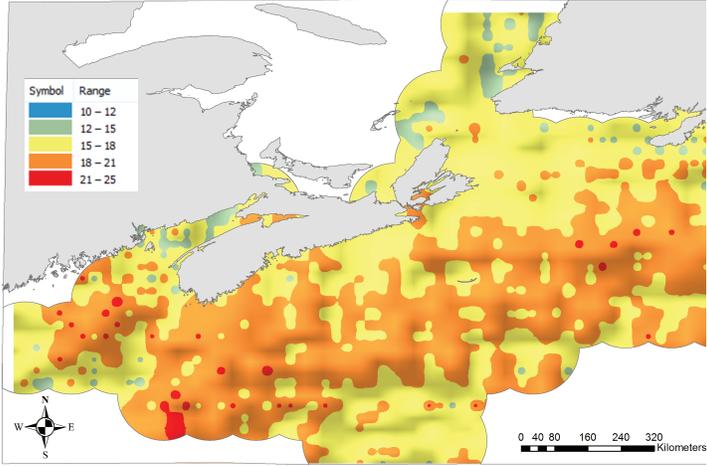
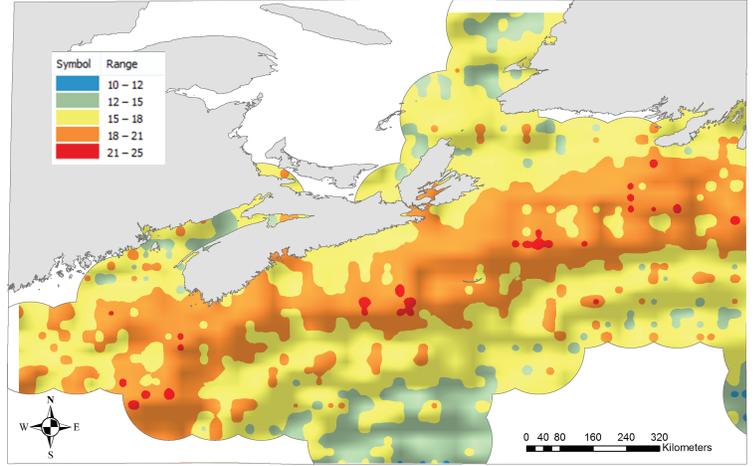


Figure7

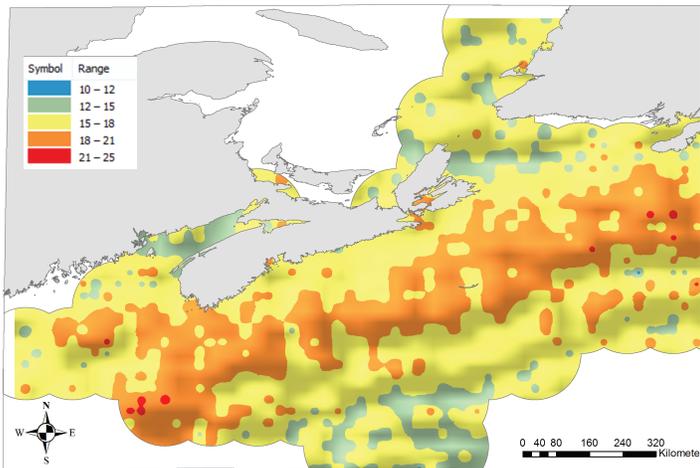
spring



summer



fall



winter

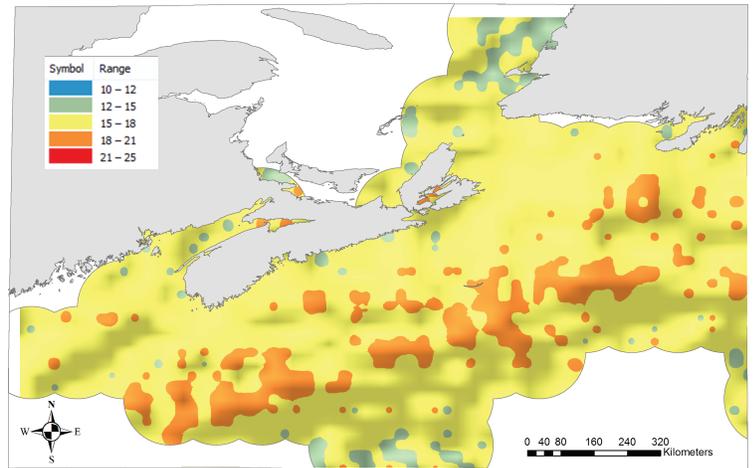


Figure8

Distribution of Spatial Vulnerability

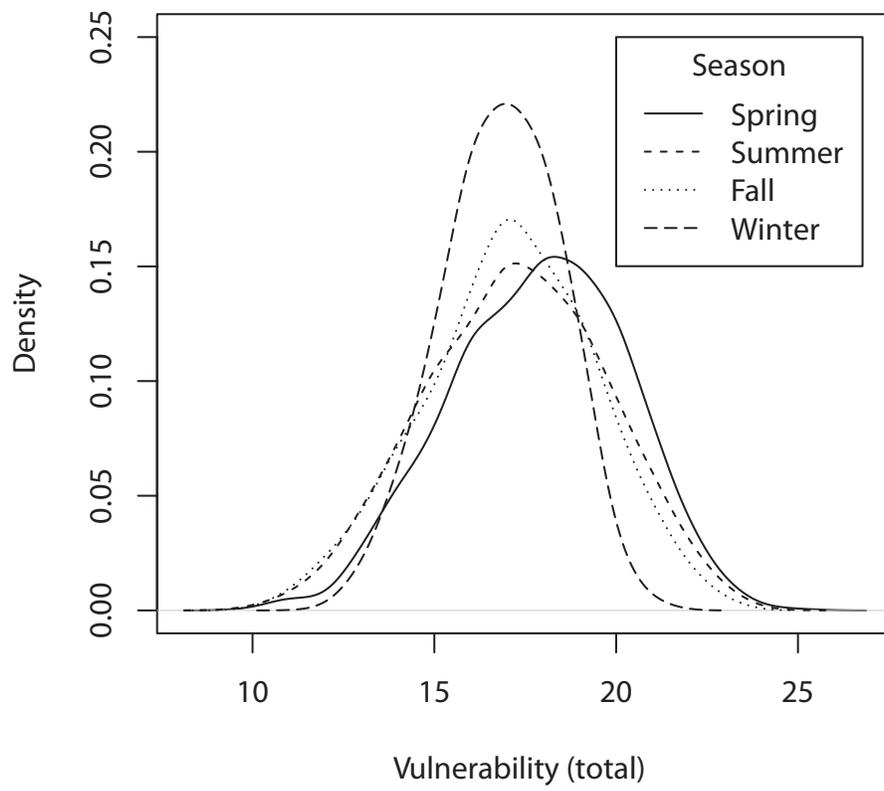


Table 1. Summary of the results of studies with the expressed goal of associating seabird distribution with oceanographic and biological factors.

Species	Scale	Period	Location	Oceanographic & Biological Factors¹	Source	Model?
Storm Petrels (Hydrobatidae)	~ 3.7-km x 0.40-km	1980, 1985, 1986	Scotian shelf, Cabot strait	Shelf break (shelf slope), sea banks (e.g., Georges Bank)	Brown (1988a)	No
Pelagic birds	Mega: >3000 km Macro: 1000-3000 km Meso: 100-1000 km Coarse: 1-100 km Fine: 1-100 m	1987	Global	Mega: major current systems and water masses; margins of oceans Macro: gyres, zonal bands, fronts Meso: rings, eddies, jets, broad upwelling events, freshwater plumes, fronts Coarse: island wakes, ice edge, sea mounts, shelf edges Fine: wind streaks	Hunt & Schneider (1987)	No (review)
Marbled Murrelet (<i>Brachyramphos marmoratus</i>)	10 km shore lengths	Densities: 1972-1993 Counts: 1922- 1989	British Columbia, Canada	Average sea-surface temperature (SST), Apr. to Aug.; herring spawn index; estuary locations; distribution of sand and fine gravel substrate; proximity to glaciers	Yen et al. (2004a)	Yes (Poisson GLM ² , CART ³ , Tree ⁴ , MARS ⁵ ANN ⁶)

Cassin's Auklet (<i>Ptychoramphus aleuticus</i>); Common Murre (<i>Uria aalge</i>); Sooty Shearwater (<i>Puffinus griseus</i>); Phalaropes (<i>Phalaropus</i> spp.)	~ 1 km ²	1996-1997, 2001-2002	California	Median depth; contour index (benthic "roughness"); distance to continental shelf break (200-m isobath); distance to banks, canyons and/or archipelago	Yen et al. (2004b)	Yes (ordinal logistic regression)
Western Gulls (<i>Larus occidentalis</i>); Common Guillemot (<i>Uria aalge</i>); Waves Albatrosses (<i>Phoebastria irrorata</i>)	15-min. survey time period x strip width of 300-600 m	1985-1994, continuously for gulls and guillemots; discontinuously for albatrosses	Farallon Is., California; Galapagos	Latitude; longitude; distance to mainland; distance to breeding colony; depth (model selection results not provided)	Clarke et al. (2003)	Yes (GAM ⁷)
Guillemots (<i>Uria</i> spp.)	20 km x 20 km grid	1986-1994	Barents Sea, Norway	Sea-surface temperature (SST); Sea-surface salinity (SSS); variance in SST; variance in SSS	Fauchald et al. (2002)	Yes (linear regression)
Northern Gannets (<i>Morus bassanus</i>); Large gulls; Auks; Kittiwakes	10 km x 10 km grid	2001-2002	Bay of Biscay, France	Latitude and longitude	Certain et al. (2007)	Yes (Two-stage modelling: GAM ⁷ + kriging of residuals)
Northern Gannets; Black-legged Kittiwake (<i>Rissa tridactyla</i>); Northern Fulmar (<i>Fulmaris</i>)	1 ⁰ x 1 ⁰ grid	1966-1992	Atlantic Canada	Sea surface temperature (SST); sea surface salinity (SSS); water temperature at 30m depth; salinity at 30m	Huettmann and Diamond (2001)	Yes (Two-stage modelling: logistic)

glacialis); Atlantic Puffin (*Fratercula arctica*); Thick-billed Murre (*Uria lomvia*); Common Murre

depth; air pressure at sea level (asl); standard deviation (STD) of atmospheric pressure; windspeed; difference between air and sea temperature; distance to coastline; shelf aspect; shelf edge; sea mount; sea depth; slope; STD of seafloor; aspect of sea floor

GLM + CART³)

Black-footed Albatross (*Phoebastria nigripes*); Northern Fulmar; Sooty Shearwater; Fork-tailed Storm-Petrel (*Oceanodroma furcata*); Leach's Storm Petrel (*O. leucorhoa*); phalaropes; Heermann's Gull (*Larus heermanni*); California Gull (*L. californicus*); Glaucous-Winged Gull; Sabines' Gull (*L. sabini*); Black-legged Kittiwakes; Arctic Tern; Common Murre; Cassin's Auklet (*Ptychoramphus aleuticus*); Rhinoceros Auklet (*Cerorhinca monocerata*); Tufted Puffin (*Fratercula cirrhata*)

4 km x 4 km grid

1996-2004

Southeastern Gulf of Alaska

SST (from AVHRR sensor); SST gradient (using edge detection algorithm of Etnoyer et al. 2004); Chlorophyll a (from SeaWiFS sensor)

O'Hara et al. (2006)

Yes (logistic GLM)

Cory's Shearwater (<i>Calonectris diomedea</i>)	Continuous sampling along transect (sampled in 15 minute increments)	1983	South Atlantic Bight off northeast Florida and southeast Georgia	Distance from Gulf Stream front; change in SST; change in fluorescence	Haney and McGillivray (1985)	Yes (Pearson's <i>r</i>)
Thick-billed Murre	50 m, 100 m, 250 m, 500 m, 1000 m and 2000 m transect segments	1982-1983	Nuvuk Is., Hudson Bay	Depth; depth gradient (maxdepth – mindepth); depth x depth gradient (interaction); nekton density (via acoustic sampling)	Cairns and Schneider (1990)	Yes (Pearson's <i>r</i>)
Dovekie (<i>Alle alle</i>)	Continuous sampling along transect (sampled in 10 minute increments)	1969-1983	Atlantic Canada	Distance to shelf break	Brown (1988b)	Yes (Chi-square)
Red-tailed Tropicbird (<i>Phaethon rubricauda</i>); White-tailed Tropicbird (<i>P. lepturus</i>); Red-billed Tropicbird (<i>P. aethereus</i>)	2 ⁰ x 2 ⁰ grid	1980-1995	Pacific Ocean	Latitude; longitude; shortest distance to mainland; ocean depth; distance to primary breeding colony	Spear and Ainley (2005)	Yes (GAM ⁷)
Manx shearwater (<i>Puffinus puffinus</i>); Guillemot; Razorbill (<i>Alca torda</i>); Kittiwake; Fulmar	~ 2 km resolution (transect sampled in 10-minute intervals)	1990	Irish Sea	SST; change in SST; SSS	Durazo et al. (1998)	Yes (Kruskal-Wallis ANOVA)
Northern Gannet	1 km x 1 km grid	2003	Western North Sea	Distance to colony; coastal water mass; front	Skov et al. (2008)	Yes (ENFA ⁸)

Northern Fulmar, Dovekie	Continuous sampling along 6-km length transects	1986-1990	Skagerrak, North Sea	Shelf break front	Skov and Durinck (1998)	Yes (ANOVA)
Balearic Shearwater	4 km x 4 km	2004-2009	Southwest Iberia	Distance to coast; mean depth; depth gradient; SST; Chl-a; sea surface height (SSH); density of ocean fronts; mean distance to nearest ocean front; 3-month SST, Chl-a, SSH; change in SST; change in Chl-a	Oppel et al. (2012)	Yes (GLM, GAM, BRT ⁹ , Maxent, random forest, ensemble)
Northern Fulmar	3 km x 3 km grid	1975-2009	Bering Sea and Aleutian Islands	Depth; slope; bathymetric features (e.g., plains, ridges); distance to land; colony effect; SST; primary productivity; fish catch	Renner et al. (2013)	Yes (GLMM, GAM, MARS, random forest, ensemble)
Black-legged Kittiwake	1 km x 1 km grid	2009-2010	Lambay and Raithlin Is., Republic of Ireland	Distance to colony; distance to land; depth; slope; Chl-a; SST; July fronts	Chivers et al. (2013)	Yes (Maxent)

¹ Factors which were either implicated by qualitative association, or were selected as significant predictors through formal model selection procedures.

² GLM = Generalized Linear Model.

³ CART = Classification and Regression Tree.

⁴ Tree (SPLUS).

⁵ MARS = Multivariate and Adaptive Regression Splines.

⁶ ANN = Artificial Neural Network.

⁷ GAM = Generalized Additive Model.

⁸ ENFA = Ecological Niche Factor Analysis.

⁹ BRT = Boosted Regression Tree.

Table 2. 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/eto_po2.html
DEPTH_CV	Coefficient of variation in seadepth	Standardized variation	GIS-derived product of DEPTH data	NA
PBANK	Percentage of area with seabank	%	GIS-derived from 100 m seadepth contours	NA
SEASON	Categorical variable (fall, spring, summer, winter)	dummy indicator variable	NA	NA
SHELFDIST	Mean distance to shelfbreak	m	GIS-derived product	NA