
Faculty of Social Sciences

Faculty Publications

Predictions from machine learning ensembles: marine bird distribution and density on Canada's Pacific coast

C. H. Fox, F. H. Huettmann, G. K. A. Harvey, K. H. Morgan, J. Robinson, R. Williams, and P. C. Paquet

2017

© 2017 Fox et al. This is an open access article distributed under the terms of the Creative Commons Attribution License. <http://creativecommons.org/licenses/by/4.0>

This article was originally published at:
<https://doi.org/10.3354/meps12030>

Citation for this paper:

Fox, C.H.; Huettmann, F.H.; Harvey, G.K.A.; Morgan, K.H.; Robinson, J.; Williams, R.; & Paquet, P.C. (2017). Predictions from machine learning ensembles: Marine bird distribution and density on Canada's Pacific coast. *Marine Ecology Progress Series*, 566, 199-216. <https://doi.org/10.3354/meps12030>



Predictions from machine learning ensembles: marine bird distribution and density on Canada's Pacific coast

C. H. Fox^{1,2,6,*}, F. H. Huettmann³, G. K. A. Harvey^{1,2}, K. H. Morgan⁴, J. Robinson^{2,7},
R. Williams⁵, P. C. Paquet^{1,2}

¹Department of Geography, University of Victoria, Victoria, BC V8W 2Y2, Canada

²Raincoast Conservation Foundation, Sidney, BC V8L 3Y3, Canada

³EWHALE lab, Biology and Wildlife Departments, Institute of Arctic Biology, University of Alaska-Fairbanks, Fairbanks, AK 99775, USA

⁴Environment and Climate Change Canada, Sidney, BC V8L 4B2, Canada

⁵Oceans Initiative, Seattle, WA 98102, USA

⁶Present address: Department of Oceanography, Dalhousie University, Halifax, NS B3H 4R2, Canada

⁷Present address: Habitat Acquisition Trust, Victoria, BC V8W 3S2, Canada

ABSTRACT: Increasingly disrupted and altered, the world's oceans are subject to immense and intensifying anthropogenic pressures. Of the biota inhabiting these ecosystems, marine birds are among the most threatened. For conservation efforts targeting marine birds to be effective, quantitative information relating to their at-sea density and distribution is typically a crucial knowledge component. In this study, we generated predictive machine learning ensemble models for 13 marine bird species and 7 groups (representing 24 additional species) in Canada's Pacific coast waters, including several species listed under Canada's Species at Risk Act. Predictive models were based on systematic marine bird line transect survey information collected in spring, summer, and fall on Canada's Pacific coast (2005–2008). Multiple Covariate Distance Sampling (MCDS) was used to estimate marine bird density along transect segments. Spatial and temporal environmental predictors, including remote sensing information, were used in model ensembles, which were constructed using 4 machine learning algorithms in Salford Systems Predictive Modeler v7.0 (SPM7): Random Forests, TreeNet, Multivariate Adaptive Regression Splines, and Classification and Regression Trees. Predictive models were subsequently combined to generate seasonal and overall predictions of areas important to marine birds based on normalized marine bird species or group richness and densities. Our results employ open access data sharing and are intended to better inform marine bird conservation efforts and management planning on Canada's Pacific coast and for broader-scale geographic initiatives across North America and elsewhere.

KEY WORDS: Marine birds · Ensemble models · Density and distribution estimates · Line transect survey · Machine learning · North Pacific Ocean

INTRODUCTION

The world's oceans are increasingly subject to a multitude of anthropogenic pressures, with continental shelf ecosystems ranking among the most heavily affected (Halpern et al. 2008, Humphries & Huett-

mann 2014). In addition to other negative consequences, anthropogenic pressures drive population declines and stress, distribution changes, declines in marine species diversity, and elevate species extinction risks (e.g. Lotze et al. 2006, Butchart et al. 2010). In turn, these biodiversity losses influence ecosystem

*Corresponding author: carolinehfox@gmail.com

© The authors 2017. Open Access under Creative Commons by Attribution Licence. Use, distribution and reproduction are unrestricted. Authors and original publication must be credited.

structure, process, resilience, stability, and trophic cascades (e.g. Hooper et al. 2005, Worm et al. 2006).

Of the world's vertebrate taxa, marine birds—and seabirds in particular—rank among the most threatened (Croxall et al. 2012). Marine birds, defined for purposes here to include seabirds, marine waterfowl, and shorebirds, consist of a highly diverse suite of upper trophic-level consumers in marine and coastal ecosystems. Significant, widespread population declines have been documented in marine birds (e.g. shorebirds, Butchart et al. 2010; seabirds, Żydelis et al. 2013, Paleczny et al. 2015). Further, it has been predicted that by the end of this century, 6 to 14% of all bird species will be extinct and 7 to 25% will be functionally extinct, with marine birds anticipated to experience elevated extinctions (Şekercioğlu et al. 2004).

A variety of conservation and management actions are undertaken to benefit marine birds and their habitats. These include marine protected areas (MPAs) and marine spatial planning (MSP), as well as broader initiatives (e.g. reduction of human consumption and global economic change; see Huettmann 2012 and references therein). However, the identification of areas important to marine birds remains a key topic. To be successful, conservation and management activities also require information regarding the spatiotemporal distribution and abundance of marine birds, in addition to information relating to areas that support high marine bird diversity. On Canada's Pacific coast, such information is particularly relevant because the region is subject to often-intense anthropogenic pressures (e.g. Ban & Alder 2008, Ban et al. 2010, Clarke Murray et al. 2015) and because outstanding knowledge gaps regarding at-sea marine bird distribution and abundance persist (see Yen et al. 2004a, Kenyon et al. 2009).

Predictive modeling represents one component in the combined effort to provide information regarding spatiotemporal distribution and abundance of marine bird species in aid of conservation and management efforts (e.g. Elith et al. 2006, Drew et al. 2011; for seabirds see Huettmann & Diamond 2001, Yen et al. 2004a, Oppel et al. 2012). In addition to information relating to distribution, measures of abundance should also be prioritized, as previous analyses have shown that the use of range maps in spatial planning may misdirect spatial conservation prioritization efforts toward marginal habitats (Williams et al. 2014).

In this study, we generated predictive ensemble models for 13 marine bird species and 7 groups (representing 24 species), including several species listed under Canada's Species at Risk Act (SARA).

Predictive models were based on systematic marine bird line transect survey information gathered in the Queen Charlotte Basin (QCB) region on Canada's Pacific coast (2005–2008) in spring, summer, and fall and were constructed using 4 machine learning algorithms: Random Forests (RF), TreeNet (TN), Multivariate Adaptive Regression Splines (MARS), and Classification and Regression Trees (CART). Here we focus on Cassin's auklet *Ptychoramphus aleuticus*, black-footed albatross *Phoebastria nigripes*, and red-necked phalarope *Phalaropus lobatus*, as these 3 species represent a range of marine bird guilds and model performances. However, all species and group-related results are reported in the extensive online Supplement available at www.int-res.com/articles/suppl/m566p199_supp.pdf. Predictive models were also subsequently combined to generate estimates of seasonal and cumulative or overall areas of importance to marine birds using normalized marine bird species or group richness and density.

The development of our approach builds on, and is informed by, previous marine bird modeling efforts (e.g. Huettmann & Diamond 2001, Yen et al. 2004a, Oppel et al. 2012). Our study presents a quantitative baseline of marine bird information for multiple species and groups in the QCB, and includes the cumulative generation of areas important to marine birds. The synthesis of our results is intended to inform marine bird conservation efforts on Canada's Pacific coast, and for use by broader geographic initiatives that aim to protect marine birds along the Pacific coast of North America and beyond.

MATERIALS AND METHODS

Line transect surveys

The 62 976 km² study area encompasses much of the QCB, which is a coastal, continental shelf region that includes Dixon Entrance (DE), Hecate Strait (HS), Queen Charlotte Sound (QCS), and Queen Charlotte Strait (QCSt) in British Columbia (BC), Canada (Fig. 1). The region hosts millions of breeding seabirds each year, represents a portion of the Pacific Flyway for migratory birds and supports long-distance migrants from across the Pacific Ocean and beyond.

Using Distance software (Thomas et al. 2010), a line transect survey in the QCB was designed and undertaken over 3 seasons (spring, summer, and fall) in 4 years (2005–2008), in conjunction with more extensive marine mammal surveys (see Williams & Thomas 2007). Marine bird surveys took place in

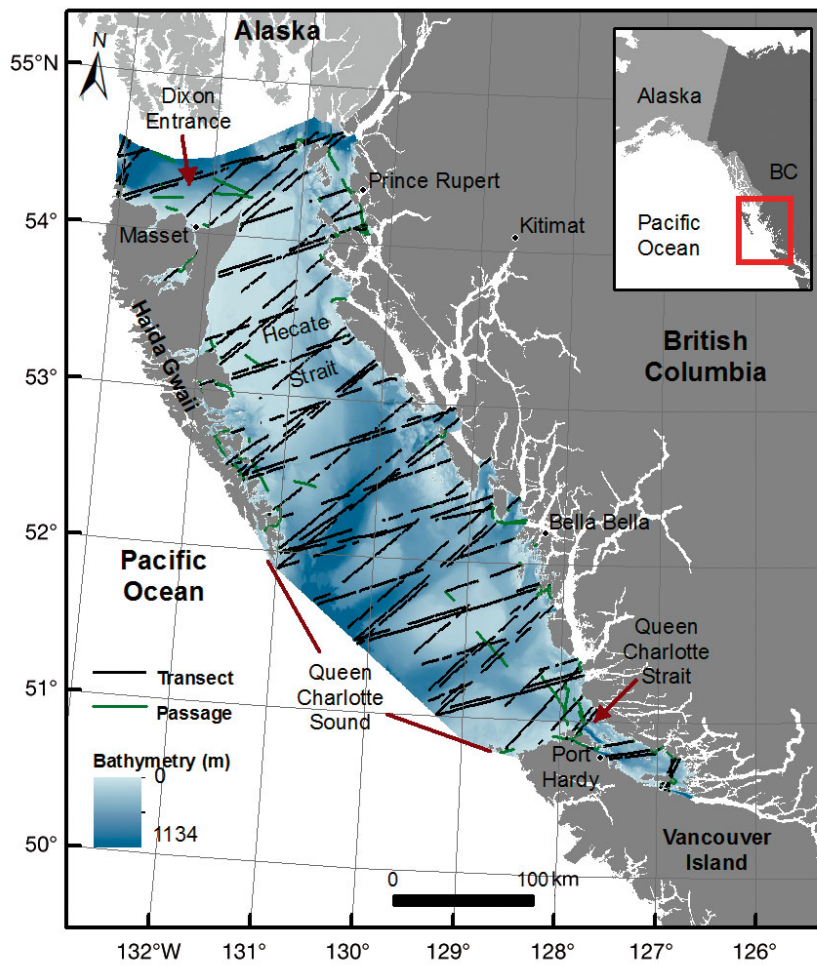


Fig. 1. Queen Charlotte Basin, British Columbia, Canada, study area (2005–2008) with realized marine bird transects (black), on passage transects (green), and shaded bathymetry (m)

spring (April and May 2007; June 2008), summer (August 2005, 2006, and 2008), and fall (October and November 2007). Winter surveys were not undertaken due to the severity of environmental conditions during this season; these conditions also limited the fall 2007 survey. Within the study area, 4729 km of planned transect and 824 km of 'on passage' transect were surveyed for marine birds. We also note that survey coverage for marine birds was not exhaustive, in part due to environmental conditions but also to allow observers rest periods (see Fig. 1 for realized transects).

In brief, line transects were placed against the BC coastline with a random starting point. A zigzag design was adopted, which compensates for excess travel time between transect lines without sacrificing uniform spatial coverage. For survey design details, see Williams & Thomas (2007) and Thomas et al.

(2007). In 2005, the survey was conducted aboard the 'Gwaii Haanas' (20 m powerboat) and from 2006 to 2008, surveys were conducted aboard the 'Achiever' (22 m motorized sailboat). Average running speed during surveys was approximately 15 km h^{-1} . Vessel routes were tracked by a Global Positioning System (GPS), with position and speed recorded automatically every 10 s using the software Logger 2000 (International Fund for Animal Welfare, www.ifaw.org).

Marine birds were surveyed along line transects using Distance Sampling, which can be an effective approach for estimating the density and/or subsequent size of wildlife populations (e.g. Buckland et al. 2001, Buckland 2004). The line transect method, which allows the probability of detection to decrease with distance from the transect line, uses a derived detection function, $g(x)$, to estimate animal densities. Assumptions for line transects following Buckland et al. (2001) and applicable (although often violated) here are (1) transects are randomly or systematically placed throughout the study area independently of the distribution of the survey population; (2) all individuals on the line are detected with certainty ($g(0) = 1$); (3) animal movement is slow compared with observer movement; and (4) distances are measured without error.

Trained marine bird observers conducted surveys at eye heights of 2.0 to 3.5 m above sea level. Observations were conducted by a single observer using the naked eye and binoculars (8×42) to scan 180° ahead of the vessel's course; for $<5\%$ survey effort, a second observer assisted with data entry and identifications. The collected marine bird survey data included detection time, distance between observer and the detected object (individual bird or group of birds), angle of the object from the track line, species, group size, and position (object in flight or on water). Observers estimated distance and angle; to improve accuracy, observers were trained and calibrated on distance and angle estimates before and during survey cruises. Species were identified with estimated $>95\%$ observer confidence; otherwise, higher taxa or group specifications were applied.

Detection function and abundance estimation

The detection function for a given species or group, $g(x)$, was modeled using MCDS in the software program Distance 6.0 release 2 (Thomas et al. 2010), the package MRDS v2.1.4 (Laake et al. 2013), and the software program R v3.0.2 (R Core Team 2013). Covariates included survey season (spring, summer, fall), survey month, log group size (e.g. Marques & Buckland 2004), and sightability (ranked 1–5 using cumulative environmental conditions including Beaufort sea state, glare, and light conditions), which were identified on an *a priori* basis (see Williams & Thomas 2007 for details).

Detection functions for birds on the water and in flight were modeled separately due to differences found during preliminary data inspection and because, with few exceptions, flying birds move faster than the observer, which violates a basic assumption of Distance Sampling (Buckland et al. 2001). Species observed in sufficient numbers for analyses, along with their taxonomic and conservation statuses, are listed in Table 1. Of the species analyzed in this study, 45% of individuals were observed while airborne. We report estimates only for birds on the water, with exceptions for gulls (*Larus* spp. and *Xema* spp.), black-legged kittiwake *Rissa tridactyla*, fork-tailed storm-petrel *Hydrobates furcatus*, Leach's storm-petrel *H. leucorhous*, black-footed albatross, red-necked phalarope, and pink-footed shearwater *Ardenna creatopus*, where information for birds observed both on the water and in flight are provided. Flying northern fulmars *Fulmarus glacialis* and dark shearwaters (*Ardenna* spp.) are also included because both groups are highly abundant and often observed flying. Justification for including information on birds in flight involves ecological and technical rationales. Several bird species or groups routinely forage while flying (e.g. gulls, kittiwakes, and storm-petrels; Jahncke et al. 2008, Nur et al. 2011), others fly at relatively slow flight speeds (e.g. storm-petrels often fly at slower linear speeds than the survey vessel; C. Fox pers. obs.), and certain species were almost exclusively detected in flight (e.g. the majority of pink-footed shearwaters). Note that because certain density estimates and subsequent model predictions involve flying birds, these reported values are undoubtedly inflated and should be interpreted as relative values (see Tasker et al. 1984).

For each species or group, we followed the options available in the Distance software and selected a half normal or hazard rate function. Right truncation of the estimated distance was applied where necessary.

Table 1. Associated taxonomic and conservation status for marine bird species analyzed in this study that relies on line transect marine bird information from British Columbia (BC), Canada (2005–2008). Br: breeding population, Nb: non-breeding population, Mi: migrant (species occurring regularly during migration), IUCN: International Union for the Conservation of Nature, COSEWIC: Committee on the Status of Endangered Wildlife in Canada, NA: not assessed. For the BC provincial conservation status, where a range of conservation ranks was provided for a given species, the lower conservative status is reported

Taxonomic order, family, common name	Scientific name	Global: IUCN	National: COSEWIC	Provincial: BC
Anseriformes, Anatidae				
Scoters ^a				
Black scoter	<i>Melanitta americana</i>	Near Threatened (2013)	NA	Nb: vulnerable (2015)
Surf scoter	<i>Melanitta perspicillata</i>	Least Concern (2012)	NA	Br: vulnerable, Nb: apparently secure (2015)
White-winged scoter	<i>Melanitta deglandi</i>	Least Concern (2013)	NA	Br: apparently secure (2015)
Charadriiformes, Alcidae				
Ancient murrelet	<i>Synthliboramphus antiquus</i>	Least Concern (2012)	Special concern (2014)	Br: imperiled, Nb: apparently secure (2015)
Cassin's auklet	<i>Ptychoramphus aleuticus</i>	Near Threatened (2015)	Special concern (2014)	Br: vulnerable, Nb: apparently secure (2015)
Common murre	<i>Uria aalge</i>	Least Concern (2015)	NA	Br: imperiled, Nb: vulnerable (2015)
Marbled murrelet	<i>Brachyramphus marmoratus</i>	Endangered (2012)	Threatened (2012)	Br: special concern, Nb: vulnerable (2015)
Pigeon guillemot	<i>Cepphus columba</i>	Least Concern (2012)	NA	Br: apparently secure (2015)
Rhinoceros auklet	<i>Cerorhinca monocerata</i>	Least Concern (2012)	NA	Br: apparently secure (2016)
Tufted puffin	<i>Fratercula cirrhata</i>	Least Concern (2012)	NA	Br: imperiled, Nb: apparently secure (2014)

(continued on next page)

Table 1 (continued)

Taxonomic order, family, common name	Scientific name	Global: IUCN	National: COSEWIC	Provincial: BC
Charadriiformes, Laridae				
Large gulls ^a				
California gull	<i>Larus californicus</i>	Least Concern (2012)	NA	Br: imperiled (2015)
Glaucous-winged gull	<i>Larus glaucescens</i>	Least Concern (2015)	NA	Br: apparently secure (2015)
American herring gull	<i>Larus smithsonianus</i>	Least Concern (2014)	NA	Br: secure (2015)
Thayer's gull	<i>Larus thayeri</i>	Least Concern (2015)	NA	M: secure (2015)
Small gulls ^a				
Black-legged kittiwake	<i>Rissa tridactyla</i>	Least Concern (2012)	NA	Nb: critically imperiled (2015)
Bonaparte's gull	<i>Larus philadelphia</i>	Least Concern (2012)	NA	Nb/Br: secure (2015)
Mew gull	<i>Larus canus</i>	Least Concern (2015)	NA	Br: apparently secure (2015)
Sabine's gull	<i>Xema sabini</i>	Least Concern (2015)	NA	M: unranked (2015)
Charadriiformes, Scolopacidae				
Red-necked phalarope	<i>Phalaropus lobatus</i>	Least Concern (2012)	Special concern (2014)	Br: vulnerable (2015)
Gaviiformes, Gaviidae				
Loons ^a				
Common loon	<i>Gavia immer</i>	Least Concern (2012)	Not at risk (1997)	Br: secure (2015)
Pacific loon	<i>Gavia pacifica</i>	Least Concern (2012)	NA	Br: apparently secure, Nb: vulnerable (2015)
Red-throated loon	<i>Gavia stellata</i>	Least Concern (2012)	NA	Br: apparently secure (2015)
Yellow-billed loon	<i>Gavia adamsii</i>	Near Threatened (2015)	Not at risk (1997)	Nb: imperiled (2015)
Pelecaniformes, Phalacrocoracidae				
Cormorants ^a				
Brandt's cormorant	<i>Phalacrocorax penicillatus</i>	Least Concern (2012)	NA	Br: critically imperiled, Nb: apparently secure (2015)
Double-crested cormorant	<i>Phalacrocorax auritus</i>	Least Concern (2012)	Not at risk (1978)	Br: vulnerable (2015)
Pelagic cormorant	<i>Phalacrocorax pelagicus</i>	Least Concern (2012)	NA	Br: apparently secure (2015)
Podicipediformes, Podicipedidae				
Grebes ^a				
Horned grebe	<i>Podiceps auritus</i>	Vulnerable (2015)	Special concern (2009)	Br: apparently secure (2015)
Red-necked grebe	<i>Podiceps grisegena</i>	Least Concern (2015)	Not at risk (1982)	Br: secure (2015)
Western grebe	<i>Aechmophorus occidentalis</i>	Least Concern (2012)	Special concern (2014)	Br: critically imperiled, Nb: imperiled (2015)
Procellariiformes, Diomedidae				
Black-footed albatross	<i>Phoebastria nigripes</i>	Near Threatened (2014)	Special concern (2007)	Nb: vulnerable (2015)
Fork-tailed storm-petrel	<i>Hydrobates furcatus</i>	Least Concern (2012)	NA	Br: apparently secure (2015)
Leach's storm-petrel	<i>Hydrobates leucorhous</i>	Least Concern (2012)	NA	Br: apparently secure (2015)
Procellariiformes, Procellariidae				
Northern fulmar	<i>Fulmarus glacialis</i>	Least Concern (2015)	NA	Br: critically imperiled, Nb: apparently secure (2015)
Pink-footed shearwater	<i>Ardenna creatopus</i>	Vulnerable (2012)	Threatened (2004)	Nb: vulnerable (2015)
Dark shearwaters ^a				
Flesh-footed shearwater	<i>Ardenna carneipes</i>	Least Concern (2012)	NA	Nb: vulnerable (2015)
Short-tailed shearwater	<i>Ardenna tenuirostris</i>	Least Concern (2012)	NA	M: unranked (2015)
Sooty shearwater	<i>Ardenna grisea</i>	Near Threatened (2015)	NA	M: unranked (2015)

^aSpecies that were analyzed as a group

The χ^2 test, Kolmogorov-Smirnov (K-S) test, and the Cramér-von Mises (C-vM) test were used to assess model fit. Akaike's information criterion (AIC) was used for model selection, including inclusion of covariates (Burnham & Anderson 2002). Detection functions, $g(x)$, were generated using pooled observations (gathered on transect and on passage), with abundance estimates generated using transect data only. For species that were low in sample size (<60; Buckland et al. 2001) and/or which had poor model performances, the detection function was generated using a larger taxonomic group (e.g. grebes, cormorants, and loons). Several species were also grouped due to difficulties associated with identification at sea (e.g. large *Larus* spp. gulls and dark *Ardenna* spp. shearwaters).

Density (birds km⁻²) estimates were generated for 13 marine bird species and 7 groups representing an additional 24 species in the QCB at a spatial resolution of 1 km segmented transect lengths (Table 1). The 1 km transect segment length was determined prior to this analysis, in part due to the spatial resolution of GIS and other remote sensing data products that were available at relatively fine spatial scales. Our 1 km segment lengths are the same as other studies (e.g. Buckland et al. 2012, Bradbury et al. 2014) but shorter than others (e.g. 3 km; Yen et al. 2004b, Nur et al. 2011). Previous studies have identified certain segment lengths or 'bins' as an appropriate spatial resolution, due to relatively low spatial autocorrelation at that scale (e.g. 3 km; Yen et al. 2004b). However, for predictive models such as ours, which rely on classification and regression trees instead of linear regression, spatial autocorrelation is not a major concern (see e.g. Betts et al. 2009, Drew et al. 2011, Nur et al. 2011).

Distribution and density modeling

Distribution and density models for 13 bird species and 7 groups were generated using an ensemble model approach that relied upon estimated bird density along transect segments (birds km⁻²) as the response variable, and 27 static, dynamic, and climatological environmental variables (Table 2). Environmental variables were selected based on previous studies (e.g. Yen et al. 2004a, Nur et al. 2011) and the availability of information relevant to our study area. Environmental variables were spatially joined to transect segments, and environmental variable values averaged at transect segment midpoints in ArcGIS 10.1 (ESRI). Marine bird transect segment

densities were similarly represented as a single value at each transect segment midpoint.

Marine bird densities were predicted on a seasonal basis: spring (April, May, and June), summer (August), and fall (October and November). Each survey was completed in approximately 2 mo, which was a primary rationale for combining marine information for 2 mo model projections (spring and fall). Summer surveys could be projected for individual years, but given the number of marine bird species, groups, and uneven effort (e.g. poor survey coverage in 2006), a composite approach for summer was judged the most efficient and further, related well to spring and fall model projections. June, originally categorized as summer (2008 survey), was reclassified as 'spring' due to the breeding chronologies of marine birds in the region.

Overall, we follow the approach of Breiman (2001a,b), which first places emphasis on generalization obtained through robust predictions, and inference second. We elected to use an ensemble model approach, which is based on the concept that many weak learners may be combined to create 1 strong learner (Friedman 1999, 2002, Araújo & New 2007). Although generalized additive models (GAMs) remain a popular choice in studies that use marine megafauna predictive modeling (e.g. Dalla Rosa et al. 2012, Best et al. 2015), machine learning approaches are becoming more common (e.g. Huettmann & Diamond 2001, Yen et al. 2004a, Nur et al. 2011, Oppel et al. 2012) and are increasingly recognized as robust alternatives to traditional statistical approaches (e.g. Elith et al. 2006, Drew et al. 2011). Further, machine learning approaches tend to be less sensitive to the effects of variable collinearity (Dormann et al. 2013). Our reliance on ensemble modeling and the number of species and groups involved also required that efficiency be prioritized alongside the use of robust modeling algorithms. Using a single software program, Salford Systems Predictive Modeler v7.0 (SPM7; see also Drew et al. 2011), which offers sophisticated optimizations and summary graphics not always available in equivalent R software, we relied on 4 machine learning approaches: RF, TN, MARS, and CART.

Of the 4 algorithms, RF and TN can be considered ensemble models in their own rights, containing combined decision trees that use boosting and/or bagging, here meaning the independent construction of many successive trees, and the optimization, as well as random selection of predictors (Breiman 2001a, Friedman 2002). TN, also known as stochastic gradient boosting, is a boosting and bagging algo-

Table 2. Description and information source of static, dynamic, and climatological environmental variables used as predictors for the machine learning model ensemble approach using line transect marine bird information from coastal British Columbia, Canada (2005–2008)

Environmental variable	Type	Unit	Resolution	Source
Static				
Year	Temporal	Year	1 yr	Survey information
Month	Temporal	Month	1 yr	Survey information
Season	Temporal	Season	Spring, Summer, Fall	Survey information
Survey	Temporal	Survey trip	Approx. 2 mo	Survey information
Latitude	Spatial	m	50 m	Survey information
Longitude	Spatial	m	50 m	Survey information
Bathymetry	Spatial	m	100 m	SciTech Environmental Consulting and Living Oceans Society; www.bcmca.ca
Slope	Spatial	°	100 m	Modified from bathymetry
Ruggedness	Spatial	Proportion	100 m	Modified from bathymetry
Distance to coast	Spatial	m	50 m	Modified from DataBC – Freshwater Atlas Coastlines; apps.gov.bc.ca
Distance to current	Spatial	m	50 m	Modified from DataBC – Benthic Marine Ecomunits; apps.gov.bc.ca
Tidal current	Spatial	m s ⁻¹	500 m	Root Mean Square Tidal Current Speed, M. Foreman; www.bcmca.ca
Distance to estuary, marsh and lagoon	Spatial	m	50 m	Modified from DataBC – estuary, marsh, and lagoon polygons, provided by C. Ogborne
Distance to glacier	Spatial	m	50 m	Modified from DataBC – Freshwater Atlas Glaciers; apps.gov.bc.ca
Distance to continental shelf	Spatial	m	50 m	Modified from DataBC – Benthic Marine Ecomunits; apps.gov.bc.ca
Distance to town	Spatial	m	50 m	This paper
Distance to colonies for select species or groups ^a	Spatial	m	200 m	Modified from Canadian Wildlife Service, Washington Department of Fish and Game, and Seabird Information Network
Dynamic				
Chlorophyll <i>a</i>	Spatiotemporal	mg m ⁻³	0.05°, monthly	Modified from AquaMODIS; http://coastwatch.pfeg.noaa.gov
Chlorophyll <i>a</i> gradients	Spatiotemporal	Proportion	0.05°, monthly	Modified from AquaMODIS; http://coastwatch.pfeg.noaa.gov
Sea surface temperature (SST)	Spatiotemporal	°C	0.05°, monthly	Modified from AquaMODIS; http://coastwatch.pfeg.noaa.gov
Distance to fronts	Spatiotemporal	m	0.05°, monthly	Modified from AquaMODIS; http://coastwatch.pfeg.noaa.gov
Front probability index (FPI)	Spatiotemporal	Proportion	0.05°, monthly	Modified from AquaMODIS; http://coastwatch.pfeg.noaa.gov
Sea surface wind speed	Spatiotemporal	m s ⁻¹	0.125°, monthly	Modified from NASA QuikSCAT; http://coastwatch.pfeg.noaa.gov
Sea surface height absolute (SSH)	Spatiotemporal	m	0.25°, monthly	Modified from AVISO program; http://coastwatch.pfeg.noaa.gov
Sea surface height deviation (SSHd) or sea level anomaly	Spatiotemporal	m	0.25°, monthly	Modified from AVISO program; http://coastwatch.pfeg.noaa.gov
Climatological				
Long-term salinity (LT-Sal)	Spatiotemporal	ppm	1.0°, monthly	World Ocean Atlas; www.nodc.noaa.gov
Long-term SST (LT-SST)	Spatiotemporal	°C	1.0°, monthly	World Ocean Atlas; www.nodc.noaa.gov

^aSeabird colony information was available for cormorants (Brandt's, double-crested, and pelagic), large gulls (herring and glaucous-winged), small gulls (Bonaparte's and mew), ancient murrelet, Cassin's auklet, common murre, fork-tailed storm-petrel, Leach's storm-petrel, northern fulmar, pigeon guillemot, rhinoceros auklet, and tufted puffin; species binomials are given in Table 1

rithm that randomly selects subsets of the data to build successive average decision trees (Friedman 1999, 2002; see www.salford-systems.com for additional details). MARS is a nonlinear regression algorithm that relies on regression trees, spline fitting, and linear basis functions (Friedman 1991). Lastly, CART is also a decision tree algorithm that relies on a recursive partitioning procedure (Breiman et al. 1984) and comes from an earlier methodology. With CART, trees are grown to a maximum size, pruned, and each tree is evaluated as a candidate for the optimal tree.

For all models, density of bird species or groups along transect segments was used as the response variable. Model settings were optimized for performance on a species- or group-specific basis (e.g. tree number, learn rate). For most species/groups, least squares was selected as the loss criterion for TN. For several species/groups, issues relating to zero inflation and variable model performances using held-out test data were encountered. Instead of held-out test data, out-of-bag (RF) and cross-validation (TN, MARS, and CART), which are both internal model performance approaches, were used. Individual model performances were evaluated using root-mean-square error (RMSE), mean-square error (MSE), and normalized R^2 . RMSE and MSE are directly related metrics and may be used to evaluate model errors for a given species or group, but should not be used for between species or group comparisons. The relative variable importance rank is also reported.

Although based on the same absolute response variable (bird density along transect segments), predictions derived from different machine learning algorithms generate predictions of relative indices of bird density, not absolute values (e.g. Hardy et al. 2011). As such, prior to integration into an ensemble prediction, adjustment of the predicted output before integration into an ensemble may improve ensemble performance. Calibration can be used to evaluate model performance but also to adjust predictions to reflect absolute predicted values. Linear regressions using paired observed (y -axis) density and predicted (x -axis) density for each bird species/group were generated to achieve that. The regression slope and intercept were used to adjust predicted responses generated by individual models prior to incorporation into an ensemble. Negative values were retained in all models and subsequent model ensembles. For mapping purposes, negative values were displayed in the GIS maps as 0.

Calibrated marine bird species and group model predictions were combined in an ensemble using

equal weighting across seasons. For regression model ensembles, although model weighting approaches have been developed (e.g. Opper et al. 2012), the strengths and weaknesses of these techniques remain relatively unexplored. In an assessment of distribution model ensemble approaches, equal weighting (mean) and weighted average generated the most robust predictions, over single and other ensemble methods (Marmion et al. 2009; see Hardy et al. 2011 and Kandel et al. 2015 for examples); because of these findings, we elected to use equal weighting. To express uncertainty, ensemble variance, meaning the variance between the 4 RF, TN, MARS, and CART model predictions, was also calculated. For each species or group, the predicted ensemble bird density and variance was mapped across seasons using natural breaks in 10 groups (with the exception of density values <50 where 6 groups were used). Transect densities estimated using MCDS were spatially overlaid across the seasonal model predictions to allow for a visual assessment of ensemble model performance. Here, we present and discuss detailed information regarding Cassin's auklet, black-footed albatross, and red-necked phalarope, as these species reflect a range of marine bird guilds and model performances. Information regarding the remaining species and groups are available in the Supplement.

Environmental variables

To capture the complex marine environment, 27 static, dynamic, and climatological (long-term) variables were assembled (e.g. Yen et al. 2004a); 26 were universally available for all bird species and groups (Table 2 and see Figs. S1 & S2 in the Supplement). The 27th environmental variable, viz. distance to nearest breeding colony, was only available for a subset of bird species that breed in and adjacent to the study area. Spatial and temporal variables derived directly from survey trip information included year (2005–2008), month (April, May, June, August, October, and November), season (spring, summer, and fall), survey (~2 mo surveys in 2005, 2006, 2007a, 2007b, and 2008), latitude, and longitude (Table 2). All other variables were obtained from a variety of sources (Table 2). All variables with a spatial component were projected using NAD83 BC Albers, which is the standard for the region. Unless noted, all modification of environmental variables and mapping were completed in ArcGIS 10.1 (ESRI). Lastly, we note that the environmental variables in this model-

ing framework are used for the generation of predictive surfaces; bird locations and densities are associated with a suite of environmental variables that combine to provide a proxy for the ecological niche of a given bird species, which is then used to predict the density of birds across the study area.

Static environmental variables

An existing bathymetry raster was obtained from the BC Marine Conservation Analysis (SciTech Consulting and Living Oceans Society; www.bcmca.ca). Within the study area, 2 small regions in the north-east and northwest were missing data. Values in these areas were populated using nearest-neighbor interpolation. Slope of the ocean floor and benthic terrain ruggedness (hereafter 'ruggedness') were derived from the interpolated bathymetry raster. Ruggedness, which is the variation in the 3-dimensional orientation of grid cells within a specific neighborhood, was generated using the Benthic Terrain Modeler extension (Wright et al. 2012). A moving window of 13 cells was used following visual inspection of various neighbor cell windows.

A number of variables were derived based on the distance to certain environmental features or definitions using the Euclidean distance metric. Distance to coast (distance to closest coastline feature) was calculated for the study area using coastline data from DataBC (Freshwater Atlas Coastlines; apps.gov.bc.ca). Distance to high-current regions (>3 knot current), hereafter 'distance to current' was generated using the Benthic Marine Ecounits—Coastal Resource Information Management System (DataBC; apps.gov.bc.ca). Tidal current speed (root mean square of average tidal speed, m s^{-1}) was generated by a 3D circulation model for the Northeastern Pacific Ocean and provided as a raster (Foreman et al. 2000; www.bcmca.ca). Distance to estuary, marsh, or lagoon, hereafter 'distance to estuary,' was generated by first creating a shapefile from polygon and polyline information regarding estuary, marsh, and lagoon locations (DataBC; apps.gov.bc.ca). Distance to estuary was created from a composite of polygon and line datasets that denoted estuary, marsh, and lagoon locations. Distance to glaciers was created using the provincial glacier shapefile 'Freshwater Atlas Glaciers' (DataBC; apps.gov.bc.ca). Distance to continental shelf was created using the Benthic Marine Ecounits – Coastal Resource Information Management System (DataBC; apps.gov.bc.ca), with the continental shelf break classified as 200–1000 m depth and 5–20° slope. Distance

to town was generated by identifying the locations of human settlements with a population of >1000 near the study area using Google search (www.google.ca). Lastly, distance to colony for individual species or groups was generated based on marine bird breeding colonies for the subset of colonial seabird species that breed within and/or adjacent to the study area (Alaska, British Columbia, and Washington). Colony information (location and species present) from 3 sources (Canadian Wildlife Service, Washington Department of Fish and Game, and the Seabird Information Network; axiom.seabirds.net/portal.php) was combined for the following species or groups: ancient murrelet *Synthliboramphus antiquus*, Cassin's auklet, common murre *Uria aalge*, fork-tailed storm-petrel, Leach's storm-petrel, pigeon guillemot *Cephus columba*, rhinoceros auklet *Cerorhinca monocerata*, tufted puffin *Fratercula cirrhata*, northern fulmar, cormorants (Brandt's *Phalacrocorax penicillatus*, double-crested *P. auritus*, and pelagic *P. pelagicus*), small gulls (black-legged kittiwake, Bonaparte's gull *Larus philadelphia*, and mew gull *L. canus*), and large gulls (American herring *L. smithsonianus* and glaucous-winged *L. glaucescens*).

Dynamic spatiotemporal environmental variables

Due to frequent cloud cover, monthly composites were used for all spatiotemporal variables. When data were missing for small areas of a monthly composite, values were interpolated across the monthly composite using the nearest neighbor values. Unless noted, all remotely sensed environmental variables were obtained from CoastWatch (<http://coastwatch.pfeg.noaa.gov>).

Remotely sensed AquaMODIS chlorophyll *a* (chl *a*) concentrations (mg m^{-3}) were collected by the NASA Earth Observing System program. Chl *a* gradients (proportion), which indicate spatial gradients (rate of change), were derived from the chl *a* raster surface:

$$\text{Chl } a \text{ gradients} = \frac{[(\text{maximum chl } a - \text{minimum chl } a) \times 100]}{\text{maximum chl } a} \quad (1)$$

where maximum chl *a* represents the largest monthly mean chl *a* value within a 3×3 pixel moving window; minimum chl *a* represents the smallest monthly mean chl *a* value within a 3×3 pixel moving window. Remotely sensed daytime sea surface temperature (SST) information was collected by the NASA Earth Observing System program. Distances to persistent SST fronts, hereafter 'distance to fronts' were calcu-

lated from the SST rasters using the MGETs Cayula-Cornillon Single Image Edge Detection algorithm (v0.8a53). A threshold of 0.375°C was applied (e.g. Dalla Rosa et al. 2012). A reduced window size of 16 and a histogram window stride of 4 were used; other settings were default values. The front detection algorithm results in a binary raster; all pixels with a value of 1 were classified as front features and subsequently used to produce the distance to front raster. The front probability index (FPI; Breaker et al. 2005), which identifies the probability of front presence over the month time period, was derived from 'front count' and 'candidate count' outputs generated by the MGETs Cayula-Cornillon Single Image Detection algorithm (v0.8a53). These output values were used to calculate the FPI using the raster calculator:

$$\text{FPI} = \text{front count/candidate count} \quad (2)$$

Monthly average sea surface wind speed (m s^{-1}) was obtained from the SeaWinds sensor on the NASA QuikSCAT platform, which measures directional and speed properties of wind over the surface of the world's oceans. The reference height for wind speeds is 10 m above the ocean surface.

Sea surface height variables are represented by 2 measurements: sea surface height absolute (SSHA) and sea surface height deviation (SSHD). SSHA represents the average monthly sea surface height deviation plus the long-term mean dynamic height. SSHD is the measured monthly average deviation from the mean geoid (at 0 elevation) as measured from 1993 to 1995. Both SSHA and SSHD monthly data were sourced from the AVISO program.

Climatological environmental variables

Two long-term climatological environmental variables were also used. Monthly averaged long-term salinity (LT-Sal) and monthly averaged long-term SST (LT-SST) using all available data (multi-decadal averages, 1955–2006) were obtained from the World Ocean Database (www.nodc.noaa.gov).

For model prediction over the surface of the study area, we relied on 13.86 km² hexagonal grid, in part to ensure future integration into Environment Canada's marine management and conservation activities in Canada's Pacific coast Exclusive Economic Zone (EEZ). All environmental predictors with a spatial component were resampled to 15 × 15 m cell sizes. Within each hexagon, the values of all cells were averaged and provided as a single value per hexa-

gon. This approach was greatly informed by previous efforts, including Huettmann & Diamond (2001), Yen et al. (2004a), Drew et al. (2011), and others.

Areas important to marine birds

In order to identify areas predicted to be important to marine birds (identified here as areas that support an elevated richness and density of marine bird species or groups), seasonal marine bird species/group density values were normalized by dividing the predicted density value within each polygon by the maximum predicted density value for a given bird species/group over the 3 seasons. As such, diversity is defined as marine bird species/group normalized density and richness. Prior to normalization, predicted negative density values were forced to 0. Normalized values were then summed across all species or group on a seasonal and cumulative (overall) basis to illustrate important areas for marine birds.

Information sharing

We encourage the use and further assessment of our data and generated models, as per Zuckerberg et al. (2011 and references therein), and make marine bird information openly available online (OBIS-SEAMAP, <http://seamap.env.duke.edu/dataset/1458>; dSPACE, access granted by the corresponding author upon request). Information sharing allows for an open assessment, transparent and repeatable science, and ensures the availability of valuable information for future assessments. For example, this information is now available for the generation of a relative index of occurrence (RIO) and a relative index of diversity.

RESULTS

Sixty-nine (69) bird species (64 marine and 5 'land' bird species) were observed in the study area (Table S1 in the Supplement). Of these, 13 species and 7 groups (representing 24 species) were seen in sufficient numbers to generate density estimates using MCDS (Table 1). MCDS model information, including detection widths and sample sizes for species or groups are available (Table S2) as are the detection functions and quantile-quantile (q-q) plots (Fig. S3).

Highest densities of Cassin's auklets occurred in the southwestern part of the study area (QCSO) in

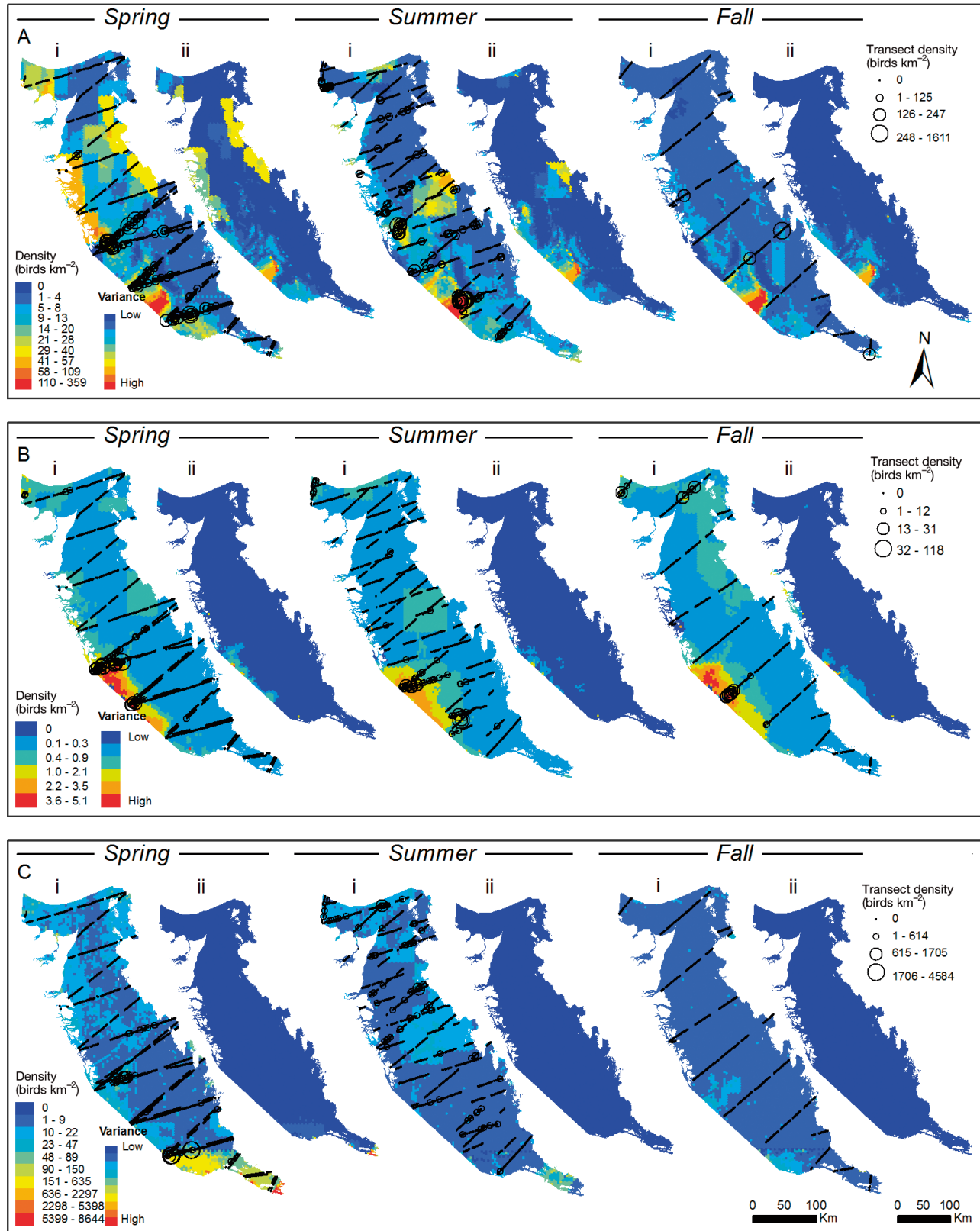


Fig. 2. Seasonal (spring, summer, and fall) predicted density (birds km⁻²) ensemble distribution and density models (i), with associated ensemble model variance (ii) for (A) Cassin's auklet *Ptychoramphus aleuticus* (on water), (B) black-footed albatross *Phoebastria nigripes* (in flight and on water), and (C) red-necked phalarope *Phalaropus lobatus* (in flight and on water). Ensemble models incorporate Random Forests (RF), TreeNet (TN), Multivariate Adaptive Regressive Splines (MARS), and Classification and Regression Trees (CART). Black circles represent bird densities (birds km⁻²) along transect lines, estimated using Distance Sampling. Marine bird information was obtained from line transect surveys in coastal British Columbia, Canada (2005–2008)

Table 3. Distribution and density model performance summary statistics for Cassin's auklet *Ptychoramphus aleuticus* (on water), black-footed albatross *Phoebastria nigripes* (in flight and on water), and red-necked phalarope *Phalaropus lobatus* (in flight and on water), including root-mean-square error (RMSE), mean-square error (MSE), R^2 normalized using Random Forests (RF), TreeNet (TN), Multivariate Adaptive Regressive Splines (MARS), and Classification and Regression Trees (CART). Model performance was evaluated using out-of-bag (RF) and cross-validation (TN, MARS, and CART). Marine bird information was obtained from line transect surveys in coastal British Columbia, Canada (2005–2008)

Species	Model performance summary statistics			
	RF	TN	MARS	CART
Cassin's auklet				
RMSE	23.17	24.14	24.67	26.63
MSE	537.00	582.97	608.58	709.19
R^2 normalized	0.41	0.36	0.34	0.26
Black-footed albatross				
RMSE	1.68	1.41	1.69	1.75
MSE	2.81	1.99	2.87	3.05
R^2 normalized	0.15	0.15	0.14	0.07
Red-necked phalarope				
RMSE	99.70	97.72	104.08	108.22
MSE	9939.63	9530.60	10832.99	11710.54
R^2 normalized	0.22	0.25	0.16	0.08

spring and summer and were largely absent in fall (Fig. 2A). The predictive models for Cassin's auklets were similar and well supported, although RF outperformed the other models (Table 3). The ensemble model predicted highest densities of Cassin's auklets to occur in outer QCSO, near the shelf break, with moderate densities predicted to occur in areas of HS (Fig. 2A). Model variance, here meaning the variance among the 4 models, was highest in the areas predicted to host highest densities, and otherwise was moderate (Fig. 2A). Important variables differed across the 4 models, with static predictors such as distance to estuary and distance to shelf leading in importance (Table 4).

Black-footed albatrosses were present in the study area in all seasons, with highest densities observed in spring and summer (Fig. 2B). Spatially, black-footed albatrosses occurred almost exclusively in outer QCSO, near the shelf break, with additional observations in DE and HS (Fig. 2B). Although the predictive performance of the 4 models was relatively poor, the ensemble model predictions of black-footed albatrosses demonstrated similarity with the line transect distribution and densities, with greatest densities of black-footed albatrosses predicted to occur in outer QCSO (Fig. 2B). Model variance was highest in the

areas predicted to support elevated densities, and otherwise relatively low (Fig. 2A). Important variables differed across the 4 models, with spatiotemporal (i.e. FPI, SSHD) and spatial predictors (i.e. distance to shelf and distance to town) leading in importance (Table 4).

Red-necked phalaropes were found in the study area in spring and summer but were absent in fall (Fig. 2C). In spring, large densities mainly occurred in outer QCSO, but in summer, red-necked phalaropes also occurred throughout DE and HS (Fig. 2C). The predictive performance of the 4 models was moderate (TN and RF) to poor (MARS and CART; Table 3), and the ensemble model predicted elevated densities of phalaropes in QCSO and QCSt in spring, with relatively low densities predicted to occur elsewhere in the study area in spring, summer, and fall (Fig. 2C). Model variance was relatively high overall. Important variables differed across the 4 models, with spatiotemporal (chl *a*, chl *a* gradient), spatial (latitude), and temporal (month) predictors leading in importance (Table 4).

Overall, model performance, as evaluated using RMSE, MSE, and R^2 , was variable across the bird species/groups (Table S3). Among species and groups, zero inflation was commonly encountered. Seasonal ensemble model predictions for the remaining species/groups are available (Figs. S4–S10), as is information regarding environmental variable rank importance (Table S4).

Marine bird ensemble predictions were normalized and subsequently overlaid to allow for an identification of areas important to marine birds, using diversity as a metric (species/group richness and normalized density). Seasonal differences in marine bird diversity were apparent, with high relative importance of DE, HS, outer QCSO, and parts of QCSt in spring, with relatively lower diversity in summer and fall (Fig. 3). Consistently important areas, meaning important areas that persisted throughout all seasons, were found in the outer QCSO, near the shelf break, and smaller areas within QCSt and HS (Fig. 3).

DISCUSSION

In this study, we collected and analyzed line transect survey information regarding 13 marine bird species and 7 marine bird groups representing 24 species in the QCB on Canada's Pacific coast. We followed recommended procedures (e.g. Gottschalk & Huettmann 2011, Zuckerberg et al. 2011) and used

Table 4. Distribution and density model rank variable importance (Var. Rank) and percent importance (% Imp.) for Cassin's auklet *Ptychoramphus aleuticus* (on water), black-footed albatross *Phoebastria nigripes* (in flight and on water), and red-necked phalarope *Phalaropus lobatus* (in flight and on water) using Random Forests (RF), TreeNet (TN), Multivariate Adaptive Regressive Splines (MARS), and Classification and Regression Trees (CART); other abbreviations as in Table 2. Marine bird information was obtained from line transect surveys in coastal British Columbia, Canada (2005–2008)

Rank	RF		TN		MARS		CART	
	Var. Rank	% Imp.	Var. Rank	% Imp.	Var. Rank	% Imp.	Var. Rank	% Imp.
Cassin's auklet								
1	Distance to estuary	100.00	Distance to estuary	100.00	Distance to shelf	100.00	Distance to glacier	100.00
2	Latitude	87.61	Distance to coast	91.87	Survey	91.76	Longitude	98.25
3	Distance to shelf	74.86	Month	75.98	Latitude	55.14	Distance to coast	97.99
4	LT-SST	62.29	Distance to glacier	69.40	Ruggedness	48.01	Distance to colony	89.91
5	Distance to coast	58.03	Latitude	62.13	SSHD	26.92	Distance to estuary	30.21
6	Distance to glacier	57.05	SST	61.29	Chl a gradient	26.36	Distance to current	18.52
7	SSHD	55.66	SSHD	59.87			Distance to front	13.79
8	Longitude	43.79	Longitude	57.20			Chl a	9.41
9	LT-Sal	39.39	Distance to current	56.04			Chl a gradient	7.02
10	SST	39.09	Distance to shelf	54.20			Bathymetry	7.02
Black-footed albatross								
1	FPI	100.00	SSHD	100.00	Distance to shelf	100.00	Distance to shelf	100.00
2			Distance to shelf	94.75	SSHD	89.05	Distance to town	46.59
3			Bathymetry	86.30	Distance to front	51.01	Distance to glacier	33.50
4			Distance to coast	58.60	Bathymetry	48.66	Chl a	32.85
5			Distance to town	57.99	SST	28.39	Latitude	32.02
6			SSHA	55.15	Slope	21.94	Bathymetry	31.67
7			LT-SST	49.66	Distance to coast	20.36	Distance to estuary	25.97
8			Chl a gradient	44.29			Wind	23.09
9			SST	43.90			LT-SST	21.89
10			Month	43.79			Month	21.11
Red-necked phalarope								
1	Chl a gradient	100.00	Chl a	100.00	Latitude	100.00	Latitude	100.00
2	Month	70.26	Latitude	96.06	Month	100.00	Chl a	95.02
3	Chl a	37.23	Month	73.62	Chl a gradient	89.92	Distance to front	40.60
4	SSHD	34.94	Chl a gradient	66.63	Distance to coast	44.01	Distance to estuary	29.93
5	Latitude	32.05	Distance to coast	53.85			Distance to current	29.78
6	Distance to coast	29.86	Distance to glacier	45.49			Bathymetry	26.59
7	Tidal current	9.08	Bathymetry	39.31			Tidal current	26.46
8	SSHA	6.59	Slope	35.55			LT-SST	22.17
9	Longitude	6.55	Distance to front	34.80			Distance to shelf	21.96
10	Distance to shelf	6.38	Distance to shelf	34.73			SSHD	20.30

Distance software to obtain density estimates. Using information on environmental conditions, we then predicted seasonal distributions and densities for numerous marine bird species and groups, with a focus on Cassin's auklet, black-footed albatross, and red-necked phalarope, as these 3 species represent a range of marine bird guilds and model performances. These efforts were achieved using an ensemble of 4 machine learning algorithm outputs (RF, TN, BRT, and CART) embedded within SPM7 software, in conjunction with ArcGIS and Distance software. Marine bird predictions were subsequently combined to identify areas of elevated marine bird importance, based on estimates of diversity (richness and normalized density). Marine bird survey information, in-

cluding predicted line transect densities and ensemble model estimated surfaces, are provided open-access for public and scientific consumption (see 'Results' for details).

Distributions and densities

Similar to our findings, previous surveys aboard ships-of-opportunity in Canada's EEZ documented elevated abundances of Cassin's auklets in spring as opposed to other seasons (Kenyon et al. 2009). In particular, highest densities were documented 50 to 75 km northwest to southwest of Triangle Island (Kenyon et al. 2009), which lies just outside our study

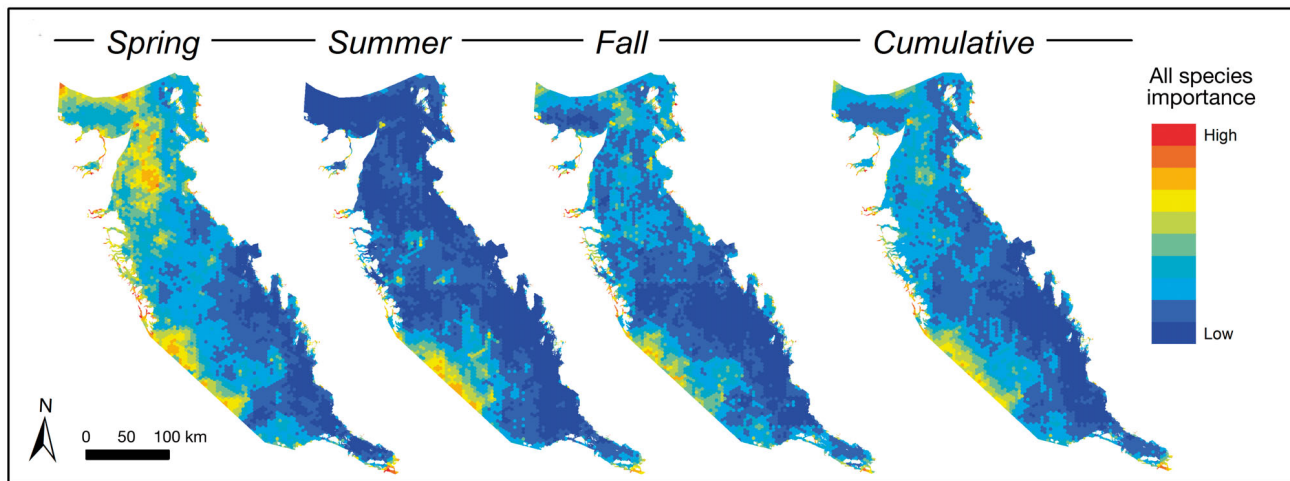


Fig. 3. Normalized seasonal (spring, summer, and fall) and cumulative marine bird species and group importance ranked from low to high (equal weighting). Marine bird information was obtained from line transect surveys in coastal British Columbia, Canada (2005–2008)

area. Our survey offers complementary information, but was also more intensive than those completed by Kenyon et al. (2009) in the outer QCSo region in spring. Our surveys documented elevated densities of Cassin's auklets throughout this area in spring and to a slightly lesser degree, in summer as well. By fall, Cassin's auklets were relatively uncommon, as was similarly reported by Kenyon et al. (2009). However, the predicted density of auklets in fall was relatively high in certain areas, which represents an example of the temporal limitations of even strong performing models. For additional detail on auks and auk sightings, see e.g. McFarlane Tranquilla et al. (2003) and Kenyon et al. (2009).

Although the predictive models generated for black-footed albatrosses performed relatively poorly, the predicted distribution component equated well with our survey distributions and with existing knowledge regarding spatial associations of this species with outer continental shelf and slope regions along the Pacific coast of North America (e.g. Briggs et al. 1987, Day 2006, Kenyon et al. 2009). At least some of the association between black-footed albatrosses and the outer continental shelf and slope region is likely attributable to the distribution of the commercial fishing fleet and albatross vessel-attendance behaviors (see e.g. Wahl & Heinemann 1979, Hyrenbach 2001). Similarly, large aggregations of black-footed albatrosses were commonly associated with actively fishing commercial vessels during our surveys and at times, with our own survey vessel (C. Fox pers. obs.). Black-footed albatrosses were previously reported as the most common longline bycatch

on Canada's Pacific coast (Smith & Morgan 2005); our results, in combination with other at-sea information (i.e. Kenyon et al. 2009), fishing effort, and bycatch estimates, could allow for the identification of areas and time periods where additional mitigation measures could be implemented. Black-footed albatrosses occur year-round in BC's coastal waters (Kenyon et al. 2009) and were similarly documented across all seasons during our surveys. Highest survey and predicted densities occurred in outer QCSo, in particular a smaller area southeast of Cape St. James (southern tip of Haida Gwaii); elevated concentrations of black-footed albatrosses have been previously noted in this area (Kenyon et al. 2009).

Similar to black-footed albatrosses, our surveys revealed highest densities of red-necked phalaropes occurring in outer QCSo waters in spring. Although this result contrasts with Kenyon et al. (2009), who reported highest densities of red-necked and red phalaropes in summer, seasonal designations differed slightly between our study and that of Kenyon et al. (2009), which likely influenced at least some of these differences. Our spring surveys documented large aggregations in 2 key areas of QCSo; east of Cape St. James and north of the Scott Islands, with the models predicting elevated densities in the southern portion of QCSo, and QCSt. Red-necked phalaropes were much more spatially distributed throughout DE, HS, and QCSo in summer. In addition, although the model prediction estimated elevated densities in QCSt, smaller areas of HS and DE were also predicted to host elevated densities of red-necked phalaropes.

Areas important to marine birds

Areas identified in this study as important to marine birds – and potential priority areas for conservation – were located within all 4 major water bodies: DE, HS, QCSt, and QCSo. Some of the marine areas identified in this study as important to marine birds match those previously identified as Ecologically and Biologically Significant Areas (EBSAs) by Clarke & Jamieson (2006) or as specifically being areas important to marine birds by Kenyon et al. (2009). However, it is important to recognize that a comparison between areas important to marine birds identified in this study and by Kenyon et al. (2009) represents a comparison between predictive model surfaces and opportunistic line transect survey information combined without correction for survey effort. On a similar note, the identification of EBSAs was based on the contributions of regional scientific experts as well as physical and oceanographic features (Clarke & Jamieson 2006).

Areas identified as important to marine birds on a cumulative basis in common with Kenyon et al. (2009) include much of the outer QCSo near the shelf break and smaller areas in DE. However, we note that Kenyon et al. (2009) identified the southern portion of outer QCSo near Triangle Island as being of high importance, whereas our study only identified the central and northern outer QCSo regions and the Cape St. James area as being of high importance on a cumulative basis. In terms of the EBSAs identified by Clarke & Jamieson (2006), areas that spatially overlapped with our identified areas important to marine birds included Hecate Strait Front, Dogfish Bank, Cape St. James, and the Shelf Break, which includes the troughs of QCSo, and other areas. The underlying environmental drivers contributing to the persistent importance of the outer QCSo region and the Cape St. James area are likely related to upwelling and elevated productivity associated with the continental shelf break (e.g. Whitney et al. 2005, Foreman et al. 2011), in addition to seabird colony proximity, including Triangle Island and the Kerouard Islands, the latter of which lie off the southern tip of Haida Gwaii. However, additional explanations, including formation of Haida Eddies (e.g. Crawford et al. 2005), are of interest as well.

Our study also reveals dynamic and seasonal shifts in areas important to marine birds. Although certain areas were consistently important to marine birds across all seasons (i.e. outer QCSo, southern QCSt, and margins adjacent to the lands), seasonally important areas were apparent. In particular, much of Dog-

fish Banks in HS and large areas of DE were of importance to marine birds in spring. The importance of Dogfish Banks is of particular interest, as the area is potentially slated to host an offshore wind farm but due to its shallowness, has not been adequately surveyed for marine birds (e.g. Kenyon et al. 2009). The importance of Dogfish Banks in spring is largely attributable to the elevated richness and often extremely high density of marine bird migrants present on the water during surveys in the area (C. Fox pers. obs.). On a seasonal basis, summer (August only) appeared to be of least importance to marine birds, which may be partly attributable to the overall reduction in migrant marine birds present in the region during this period. Fall areas identified as important to marine birds were similar to spring, just less pronounced, with areas of HS and DE considered somewhat seasonally important. However, given the timing of our fall surveys (October and November), we likely failed to capture a portion of the southbound migratory movement that occurs in September (C. Fox pers. obs.).

Currently, the majority of marine areas identified in this study as important to marine birds are granted no additional protection beyond existing legislation for wildlife and their habitats in Canada. Importantly, no marine Critical Habitat has yet been identified for the numerous SARA-listed marine bird species present in the study area. However, a notable exception to additional habitat protections includes the Gwaii Haanas National Marine Conservation Area. Efforts to designate the waters encompassing the Scott Islands as a marine National Wildlife Area are currently underway and further, the current Canadian federal government has committed to meeting the Aichi Biodiversity Targets, one of which specifies that 10% of coastal and marine areas will be protected by 2020 (Convention on Biological Diversity 2013).

Study limitations and future directions

Surveying birds at sea is inherently challenging, and limitations exist with respect to this research design, analysis, and subsequent predictive marine bird information generated. Due to limitations related to the survey and the coarseness of certain environmental information, the spatiotemporal scales applied in this analysis undoubtedly influenced our model-based predictions and model performances. Not all potentially relevant environmental variables were available either, with an absence of environmental information relating to prey (e.g. forage fish and zoo-

plankton), bird behavior, contamination, and water column—as opposed to remotely sensed surface—characteristics (e.g. prey abundance and depth of chlorophyll maximum; Tremblay et al. 2009). Many of the challenges encountered in this study have already been noted by other researchers; for example, data for many species or groups demonstrated zero inflation, as is relatively common for at-sea surveys of marine birds (e.g. Oppel et al. 2012). Further, the collection and subsequent use of flying bird information within a line transect distance sampling analytical framework remains uncommon, despite significant numbers of marine birds being airborne in marine environments. While this issue awaits improvement, the analysis and use of information relating to birds on the water remains a meaningful metric.

Despite the aforementioned limitations and challenges, for many species and groups, this study provides the ‘best available’ information regarding their estimated at-sea distributions and densities in the study area. In conjunction with information on breeding colonies, MPAs, important bird areas, and other areas important for marine birds, this information can be used to inform emergency responses, assessments of risk (e.g. chronic oil pollution; Fox et al. 2016), MPA planning, MSPs, including the Marine Planning Partnership for the North Pacific Coast, and more. Although a first quantitative baseline and useful for management and conservation, this information should nonetheless be considered ‘snapshots’ of marine bird distribution and density. Future marine bird surveys are required and should be used in conjunction with other types of information (e.g. electronic tracking and radar surveys) to iteratively update and improve our knowledge of marine bird distribution, density, and spatiotemporal dynamics in Canada’s Pacific coast waters.

Climate change, loss and degradation of habitat, and other anthropogenic impacts represent significant cumulative and synergistic threats to marine birds that inhabit naturally dynamic marine ecosystems. At-sea surveys of marine birds are often considered prohibitively costly. However, when linked with predictive methods like machine learning, this approach represents a powerful, complementary conservation tool (e.g. Huettmann & Diamond 2001, Yen et al. 2004a, Oppel et al. 2012). Canada’s Pacific coast is subjected to significant anthropogenic pressures (Ban & Alder 2008, Clarke Murray et al. 2015), many of which are projected to increase over time (e.g. climate change, IPCC 2014; shipping traffic, Nuka RPG 2013). As such, improving our understanding of the spatiotemporal distribution and abun-

dance of the marine avifauna, their key habitats, and the ecological processes they rely upon, are essential components of wise stewardship of Canada’s Pacific coast marine ecosystems.

Acknowledgements. Support for marine bird surveys by Raincoast Conservation Foundation (RCF) and subsequent data analysis was provided by the Gordon and Betty Moore Foundation, the Marisla Foundation, the McLean Foundation, the Bullitt Foundation, Mountain Equipment Co-op, Patagonia, the Conservation Alliance, the Vancouver Foundation, the Russell Family Foundation, Environment and Climate Change Canada (ECCC), and RCF donors, volunteers, and others. H. Krajewsky, M. Price, and other marine bird observers and survey members are acknowledged for their contributions. We also thank D. Kawai for his contributions to data preparation and P. O’Hara and N. Serra-Sogas for the study hexagons and advice on environmental variables. Salford Systems Ltd provided SPM7 for this research via the EWHALE lab license to F.H. C.H.F was supported by an NSERC IRDF postdoctoral fellowship, G.K.H. and J.R. by RCF and the ECCC Science Horizons program, P.C.P. by RCF, and K.M. by ECCC. F.H. appreciates the support by UAF for the EWHALE lab, as well as S. Linke, H. Berrios Alvarez, and the project team of co-authors.

LITERATURE CITED

- ✦ Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol* 22:42–47
- ✦ Ban N, Alder J (2008) How wild is the ocean? Assessing the intensity of anthropogenic marine activities in British Columbia, Canada. *Aquat Conserv* 18:55–85
- ✦ Ban NC, Alidina HM, Ardron JA (2010) Cumulative impact mapping: advances, relevance and limitations to marine management and conservation, using Canada’s Pacific waters as a case study. *Mar Policy* 34:876–886
- Best BD, Fox CH, Williams R, Halpin PN, Paquet PC (2015) Updated marine mammal distribution and abundance estimates in British Columbia. *J Cetacean Res Manag* 15: 9–26
- ✦ Betts MG, Ganio LM, Huso MM, Som NA, Huettmann F, Bowman J, Wintle BA (2009) Comment on ‘Methods to account for spatial autocorrelation in the analysis of species distributional data: a review’. *Ecography* 32:374–378
- ✦ Bradbury G, Trinder M, Furness B, Banks AN, Caldow RW, Hume D (2014) Mapping seabird sensitivity to offshore wind farms. *PLOS ONE* 9:e106366
- Breaker LC, Mavor TP, Broenkow WW (2005) Mapping and monitoring large-scale ocean fronts off the California Coast using imagery from the GOES-10 geostationary satellite. California Sea Grant College Program, San Diego, CA
- ✦ Breiman L (2001a) Random forests. *Mach Learn* 45:5–32
- ✦ Breiman L (2001b) Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat Sci* 16:199–231
- Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC Press, Boca Raton, FL
- Briggs KT, Tyler WB, Lewis DB, Carlson DR (1987) Bird communities at sea off California: 1975 to 1983. *Stud Avian Biol* 11:1–74

- Buckland ST (2004) Advanced distance sampling. Oxford University Press, Oxford
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling. Oxford University Press, Oxford
- ✦ Buckland ST, Burt ML, Rexstad EA, Mellor M, Williams AE, Woodward R (2012) Aerial surveys of seabirds: the advent of digital methods. *J Appl Ecol* 49:960–967
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York, NY
- ✦ Butchart SHM, Walpole M, Collen B, van Strien A and others (2010) Global biodiversity: indicators of recent declines. *Science* 328:1164–1168
- Clarke CL, Jamieson GS (2006) Identification of ecologically and biologically significant areas in the Pacific North Coast Integrated Management Area: Phase II – Final Report. *Can Tech Rep Fish Aquat Sci* 2686:1–25
- ✦ Clarke Murray C, Agbayani S, Alidina HM, Ban NC (2015) Advancing marine cumulative effects mapping: an update in Canada's Pacific waters. *Mar Policy* 58:71–77
- Convention on Biological Diversity (2013) Quick guides to the Aichi biodiversity targets, version 2. United Nations. www.cbd.int (accessed on 15 December 2015)
- ✦ Crawford WR, Brickley PJ, Peterson TD, Thomas AC (2005) Impact of Haida eddies on chlorophyll distribution in the eastern Gulf of Alaska. *Deep-Sea Res II* 52:975–989
- ✦ Croxall JP, Butchart SHM, Lascelles B, Stattersfield AJ, Sullivan B, Symes A, Taylor P (2012) Seabird conservation status, threats and priority actions: a global assessment. *Bird Conserv Int* 22:1–34
- ✦ Dalla Rosa L, Ford JK, Trites AW (2012) Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters. *Cont Shelf Res* 36:89–104
- Day RH (2006) Seabirds in the northern Gulf of Alaska and adjacent waters, October to May. *West Birds* 37:190–214
- ✦ Dormann CF, Elith J, Bacher S, Buchmann C and others (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46
- Drew A, Wiersma YF, Huettmann F (2011) Predictive species and habitat modeling in landscape ecology. Springer, New York, NY
- ✦ Elith J, Graham CH, Anderson RP, Dudík M and others (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- ✦ Foreman MGG, Crawford WR, Cherniawsky JY, Henry RF, Tarbotton MR (2000) A high-resolution assimilating tidal model for the northeast Pacific Ocean. *J Geophys Res* 105:28629–28651
- ✦ Foreman MGG, Pal B, Merryfield WJ (2011) Trends in upwelling and downwelling winds along the British Columbia shelf. *J Geophys Res* 116:1–11
- ✦ Fox CH, O'Hara P, Bertazzon S, Morgan K, Underwood FE, Paquet PC (2016) A preliminary spatial assessment of risk: marine birds and chronic oil pollution on Canada's Pacific coast. *Sci Total Environ* 573:799–809
- ✦ Friedman JH (1991) Multivariate adaptive regression splines. *Ann Stat* 19:1–67
- Friedman JH (1999) Stochastic gradient boosting. Tech Rep. Department of Statistics, Stanford University, Stanford, CA
- ✦ Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38:367–378
- ✦ Gottschalk T, Huettmann F (2011) Comparison of distance sampling and territory mapping methods for birds in four different habitats. *J Ornithol* 152:421–429
- ✦ Halpern BS, Walbridge S, Selkoe KA, Kappel CV and others (2008) A global map of human impact on marine ecosystems. *Science* 319:948–952
- ✦ Hardy SM, Lindgren M, Konakanchi H, Huettmann F (2011) Predicting the distribution and ecological niche of unexploited snow crab (*Chionoecetes opilio*) populations in Alaskan waters: a first open-access ensemble model. *Integr Comp Biol* 51:608–622
- ✦ Hooper DU, Chapin FS III, Ewel JJ, Hector A and others (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol Monogr* 75:3–35
- Huettmann F (2012) Protection of the three poles. Springer, Tokyo
- ✦ Huettmann F, Diamond AW (2001) Seabird colony locations and environmental determination of seabird distribution: a spatially explicit breeding seabird model for the north-west Atlantic. *Ecol Model* 141:261–298
- Humphries GRW, Huettmann F (2014) Putting models to a good use: a rapid assessment of Arctic seabird biodiversity indicates potential conflicts with shipping lanes and human activity. *Divers Distrib* 20:478–490
- ✦ Hyrenbach KD (2001) Albatross response to survey vessels: implications for studies of the distribution, abundance, and prey consumption of seabird populations. *Mar Ecol Prog Ser* 212:283–295
- IPCC (Intergovernmental Panel on Climate Change) (2014) Climate Change 2014: Synthesis report. Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva
- ✦ Jahncke J, Vlietstra LS, Decker MB, Hunt GL (2008) Marine bird abundance around the Pribilof Islands: a multi-year comparison. *Deep-Sea Res II* 55:1809–1826
- ✦ Kandel K, Huettmann F, Suwal MK, Regmi GR and others (2015) Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: red panda (*Ailurus fulgens*) in the Hindu-Kush Himalaya region. *Biol Conserv* 181:150–161
- Kenyon JK, Morgan KH, Bentley MD, McFarlane Tranquilla LA, Moore KE (2009) Atlas of pelagic seabirds off the west coast of Canada and adjacent areas. Tech Rep Ser No. 499. Canadian Wildlife Service, Environment Canada, Delta
- Laake J, Borchers D, Thomas L, Miller D, Bishop J (2013) MRDS: Mark-Recapture Distance Sampling. R package version 2.1.4. www.CRAN.R-project.org/package=mrds (accessed on 10 December 2013)
- ✦ Lotze HK, Lenihan HS, Bourque BJ, Bradbury RH and others (2006) Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* 312:1806–1809
- ✦ Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. *Divers Distrib* 15:59–69
- Marques FFC, Buckland ST (2004) Covariate models for the detection function. In: Buckland, ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (eds) Advanced distance sampling. Oxford University Press, Oxford, p 31–47
- McFarlane Tranquilla L, Huettmann F, Loughheed C, Loughheed LW, Kaiser G, Parker N (2003) Sightings of vagrant Pacific alcids in Desolation Sound, British Columbia. *Can Field Nat* 117:53–56

- Nuka RPG (Research and Planning Group) (2013) West coast spill response study, Vol 2. Vessel traffic study. Report to the British Columbia Ministry of Environment. Nuka RPG, Seldovia, AK
- ✦ Nur N, Jahncke J, Herzog MP, Howar J and others (2011) Where the wild things are: predicting hotspots of seabird aggregations in the California Current System. *Ecol Appl* 21:2241–2257
- ✦ Oppel S, Meirinho A, Ramírez I, Gardner B, O'Connell AF, Miller PI, Louzao M (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biol Conserv* 156:94–104
- ✦ Paleczny M, Hammill E, Karpouzi V, Pauly D (2015) Population trend of the world's monitored seabirds, 1950–2010. *PLOS ONE* 10:e0129342
- R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- ✦ Şekercioğlu ÇH, Daily GC, Ehrlich PR (2004) Ecosystem consequences of bird declines. *Proc Natl Acad Sci USA* 101:18042–18047
- Smith J, Morgan KH (2005) An assessment of seabird by-catch in longline and net fisheries in British Columbia. Tech Rep Ser No. 401, Canadian Wildlife Service, Pacific and Yukon Region, Delta
- Tasker ML, Jones PH, Dixon T, Blake BF (1984) Counting seabirds at sea from ships: a review of methods employed and a suggestion for a standardized approach. *Auk* 101:567–577
- Thomas L, Williams R, Sandilands D (2007) Designing line transect surveys for complex survey regions. *J Cetacean Res Manag* 9:1–13
- ✦ Thomas L, Buckland ST, Rexstad EA, Laake JL and others (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 RW, Kappes MA, Costa DP, Shaffer SA (2009) Analytical approaches to investigating seabird–environment interactions: a review. *Mar Ecol Prog Ser* 391:153–163
- ✦ Wahl TR, Heinemann D (1979) Seabirds and fishing vessels: co-occurrence and attraction. *Condor* 81:390–396
- ✦ Whitney FA, Crawford WR, Harrison PJ (2005) Physical processes that enhance nutrient transport and primary productivity in the coastal and open ocean of the subarctic NE Pacific. *Deep-Sea Res II* 52:681–706
- Williams R, Thomas L (2007) Distribution and abundance of marine mammals in the coastal waters of British Columbia, Canada. *J Cetacean Res Manag* 9:15–28
- ✦ Williams R, Grand J, Hooker SK, Buckland ST and others (2014) Prioritizing global marine mammal habitats using density maps in place of range maps. *Ecography* 37: 212–220
- ✦ Worm B, Barbier EB, Beaumont N, Duffy JE and others (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* 314:787–790
- Wright DJ, Pendleton M, Boulware J, Walbridge S and others (2012) ArcGIS Benthic Terrain Modeler (BTM), v. 3.0. Environmental Systems Research Institute, NOAA Coastal Services Center, Massachusetts Office of Coastal Zone Management. <http://esriurl.com/5754> (accessed on 10 November 2013)
- ✦ Yen P, Huettmann F, Cooke F (2004a) Modelling abundance and distribution of marbled murrelets (*Brachyramphus marmoratus*) using GIS, marine data and advanced multivariate statistics. *Ecol Model* 171:395–413
- ✦ Yen PP, Sydeman WJ, Hyrenbach KD (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
- Zuckerberg B, Huettmann F, Friar J (2011) Proper data management as a scientific foundation for reliable species distribution modeling. In: Drew CA, Wiersma Y, Huettmann F (eds) Predictive species and habitat modeling in landscape ecology. Springer, New York, NY, p 45–70
- ✦ Żydelski R, Small C, French G (2013) The incidental catch of seabirds in gillnet fisheries: a global review. *Biol Conserv* 162:76–88

Editorial responsibility: Rory Wilson,
Swansea, UK

Submitted: July 15, 2016; Accepted: December 21, 2016
Proofs received from author(s): February 14, 2017